HAL
open science

Développement d'une méthodologie pour la garantie de performance énergétique associant la simulation à un protocole de mesure et vérification

Simon Ligier

To cite this version:

Simon Ligier. Développement d'une méthodologie pour la garantie de performance énergétique associant la simulation à un protocole de mesure et vérification. Energie électrique. Université Paris sciences et lettres, 2018. Français. NNT : 2018PSLEM083 . tel-02192468

HAL Id: tel-02192468

https://pastel.hal.science/tel-02192468

Submitted on 23 Jul 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à MINES ParisTech

Développement d'une méthodologie pour la garantie de performance énergétique associant la simulation à un protocole de mesure et vérification

Soutenue par
Simon LIGIER
Le 28 septembre 2018

Ecole doctorale n° 432
Sciences des métiers de l'ingénieur

Spécialité
Energétique et procédés

Composition du jury :

Mme Monika WOLOSZYN, Professeur, Université Savoie Mont-Blanc	*Présidente*
M. David BIGAUD, Professeur, Université d'Angers	*Rapporteur*
M. Stéphane PLOIX, Professeur, INP-Grenoble	*Rapporteur*
M. Laurent MORA, Maitre de conférences, Université Bordeaux	*Examinateur*
M. Christophe GOBIN, Docteur, VINCI	*Examinateur*
M. Bruno PEUPORTIER Directeur de recherche, Mines ParisTech	*Directeur de thèse*
M. Patrick SCHALBART Docteur, Mines ParisTech	*Maître de thèse*

MINES ParisTech | PSL

Remerciements

Je souhaite tout d'abord remercier Bruno Peuportier, mon directeur de thèse, pour la confiance qu'il m'a accordée, d'abord lors de mon arrivée dans l'équipe ETB puis lorsqu'il m'a proposé de me lancer dans cette aventure. Ses conseils et sa vision des différents sujets liés à cette thèse l'ont maintenue sur les bons rails et n'ont rien enlevé à l'indépendance avec laquelle j'ai pu travailler.

Un grand merci à Patrick Schalbart qui m'aura accompagné, conseillé, freiné et parfois relancé durant ces trois ans. Ses remarques et ses analyses, sur le fond comme sur la forme, m'ont permis de construire ce projet et d'en voir le terme. Merci beaucoup pour cet enrichissement professionnel et personnel.

Je remercie sincèrement les membres du jury qui ont accepté d'évaluer ce travail. Merci à David Bigaud et Stéphane Ploix qui ont bien voulu en être les rapporteurs. Merci à Monika Woloszyn, Laurent Mora et Christophe Gobin. Vos retours à tous et les échanges que nous avons pu avoir ont nourri ma réflexion et me permettent aujourd'hui de porter un regard critique sur cette thèse et de mieux l'intégrer à l'écosystème actuel et en développement autour des thématiques étudiées.

De nombreuses personnes ont contribué à cette thèse que ce soit d'un point de vue technique, critique, philosophique, ludique, et sympathique. Je pense d'abord à la Dream Team d'Izuba Energie, à Renaud, Ludo et à Stéphane, pour votre soutien technique et vos conseils. Merci pour tout ce temps que vous m'avez accordé. C'est un plaisir de travailler avec vous. Merci aussi aux cocos de Koclikos, Maxime, Eric, Fabio et Renaud, qui pourraient facilement se retrouver dans trois ou quatre paragraphes de ces remerciements : celui des partenaires (presque) sans faille, des anciens, dont les thèses auront servi de socle et d'inspiration à mes travaux, et des amis bien sûr. Je suis heureux d'avoir assisté aux premiers pas de votre projet dont les avancées techniques se sont faites en parallèle de mes propres travaux. La chaire Eco-conception, au sein de laquelle a été lancée cette thèse, a été un cadre intéressant pour communiquer sous différentes formes sur ce travail et avoir des échanges transdisciplinaires enrichissants. Merci à tous ses membres. Les discussions avec les équipes de VINCI ont aussi enrichi ma vision et ont apporté des retours structurant au début du projet. Merci notamment à Walid Saiem, Christophe Gobin et Nathalie Mehut. Je remercie enfin tous les chercheurs et professionnels avec qui j'ai eu l'occasion d'échanger durant ces trois ans.

Je remercie chaleureusement tous les membres de l'équipe ETB, anciens et nouveaux, qui représenteront sûrement le plus gros contingent de lecteurs de ces remerciements. L'équipe évolue mais il reste cet esprit et cet attachement particuliers, qui viennent aussi d'une volonté de construire sur le travail précédemment réalisé et qui font de ce groupe une équipe à part au sein du centre. Merci donc à Eric, Fabio, Maxime, Charlotte, Guillaume, Thomas, Michaël, Marie-Lise, Marie, Mija, Long, et à tous les stagiaires qui nous ont rejoints et accompagnés. Des amis à garder, mais aussi un réseau qui s'étend à travers la France !

Je remercie aussi les membres du CES que j'ai côtoyés pendant plus de quatre ans. J'ai eu la chance de faire partie d'une belle fournée de doctorants, soudés dans les difficultés mais aussi autour d'un verre. Merci à tous et en particulier à Egoï, Afif, Rami, Iran, Mathildi, Alice ainsi qu'à toute l'équipe administrative du centre. Je garderai en mémoire les séances d'escalade, les soirées jeux et électro, les parties de pétanque et l'inventivité dans les méthodes de choix de restos.

Un grand merci à mes amis qui m'accompagnent, aux colocs de Montrouge et aux copains d'ici et d'ailleurs. Nos relations et toutes nos discussions permettent de tenir dans les moments difficiles et sont source d'inspiration. Merci à ma famille, mes grands-parents, ma sœur et mes parents. Votre confiance et votre soutien m'ont toujours permis de m'exprimer et d'agir librement. Merci enfin à Camille d'être à mes côtés et de m'avoir soutenu tout au long de ce projet.

Table des Matières

Table des Matières **2**

Nomenclature **7**

Chapitre 1. État de l'art – Prévision et mesure des consommations énergétiques des bâtiments, vers la garantie de performance énergétique **11**

1.1. Contexte et enjeux de la prévision des consommations d'énergie 13
1.1.1. La conception éco-énergétique 13
1.1.2. La modélisation d'un système complexe : le bâtiment 14
1.1.3. Les sources d'erreurs et d'incertitudes 17

1.2. Le modèle physico-probabiliste : un changement de paradigme 19
1.2.1. L'exploration de modèles 19
1.2.2. Décrire la variabilité des conditions d'exploitation des bâtiments 25
1.2.3. Abandonner le déterminisme : l'intégration des incertitudes et variabilités 30

1.3. Mesurer pour vérifier la performance 31
1.3.1. Erreurs et incertitudes 32
1.3.2. Protocoles de mesure et vérification 34
1.3.3. État de l'art de la mesure dans le bâtiment 36
1.3.4. Mesurer et inférer l'occupation et les usages 37
1.3.5. Le calibrage de modèle 38
1.3.6. Le constat de l'écart entre prévisions et réalité 40

1.4. La garantie de performance énergétique 42
1.4.1. Enjeux de la garantie de performance 42
1.4.2. Éléments-clés du projet de GPE 45
1.4.3. État de l'art des méthodes 52

1.5. Positionnement de la thèse 55
1.5.1. Objectifs scientifiques 55
1.5.2. Plan du manuscrit 56

1.6. Modèles et cas d'étude 57
1.6.1. Modèles et méthodes utilisés 57
1.6.2. Etude de cas : Rénovation de logements collectifs 60

1.7. Conclusion 64

Chapitre 2. Modèle de génération de données météorologiques pour la prise en compte de la variabilité climatique 67

2.1. Introduction 69

2.2. État de l'art 71

2.2.1. La modélisation des conditions météorologiques 71

2.2.2. L'impact du climat sur les consommations d'énergie des bâtiments 76

2.2.3. Générer la variabilité climatique 79

2.3. Objectifs du modèle développé 83

2.3.1. Propagation des effets de la variabilité climatique dans les modèles SED 83

2.3.2. Apprentissage pour l'entraînement de modèles d'ajustements 84

2.3.3. Limitations et apports scientifiques 85

2.4. Méthodologie pour la génération de données météorologiques 86

2.4.1. Génération de données de température 86

2.4.2. Génération des données de rayonnement global horizontal 91

2.4.3. Génération des données de rayonnement diffus horizontal 95

2.4.4. Variation des parts tendancielles 95

2.5. Évaluation de la fiabilité du modèle 101

2.5.1. Caractéristiques des séries générées 101

2.5.2. Respect de l'intercorrélation entre grandeurs météorologiques 106

2.6. Application à l'étude de l'influence du climat sur les consommations d'énergie d'un bâtiment 107

2.6.1. Propagation de la variabilité 107

2.6.2. Caractérisation de l'impact des conditions météorologiques 109

2.7. Conclusion 113

Chapitre 3. Méthodologie pour la définition d'un engagement de performance énergétique .. 115

3.1. Objectifs de la méthodologie 117

3.2. La simulation du modèle physico-probabiliste 118

3.2.1. Hypothèses et objectifs de la démarche 118

3.2.2. Analyse de sensibilité 118

3.2.3. Propagation des incertitudes et variabilités 119

3.2.4. Développement et calibrage du modèle pre-travaux 120

3.3. Ajustement 123

3.3.1. Formulation et intérêt du modèle 123

3.3.2. Définition et choix des facteurs d'ajustement 123

3.3.3. Méthodes de calcul des modèles d'ajustement 125

3.3.4. GPE absolue : Détermination d'une limite garantie de consommation ajustable 129
3.3.5. GPE relative : détermination du gain et du modèle de référence ajustable 130
3.3.6. Ajustement et vérification de la performance 134

3.4. Intégration de l'influence des incertitudes de mesure 136
3.4.1. Enjeu et approche numérique 136
3.4.2. Estimation des incertitudes 137
3.4.3. Intégration amont au processus 140

3.5. Protocole complet pour la garantie de performance énergétique 142
3.5.1. GPE absolue dans le neuf et la rénovation lourde 142
3.5.2. GPE relative en réhabilitation 143

3.6. Conclusion 147

Chapitre 4. Méthodes d'aide à la mise en œuvre de la méthodologie de GPE 149

4.1. Étude de la robustesse du modèle d'ajustement 151
4.1.1. Variance de l'estimateur de régression quantile 151
4.1.2. Variance statistique du modèle 153
4.1.3. Bootstrap et modification de l'engagement 155
4.1.4. Validation numérique du risque de dépassement 157

4.2. Influence des incertitudes de mesure 158
4.2.1. Caractérisation et intégration 158
4.2.2. Modification des coefficients d'ajustement des modèles 159
4.2.3. Modification de la limite garantie de consommation 161
4.2.4. Application à la comparaison de PMV 162

4.3. Choix et optimisation des paramètres d'ajustement 165
4.3.1. Objectifs 165
4.3.2. Tests de corrélation 166
4.3.3. Qualité du modèle de régression 170
4.3.4. Méthode d'AS par régression 174
4.3.5. Tests statistiques 177
4.3.6. Régression pénalisée 178
4.3.7. Bilan pour l'aide au choix des facteurs d'ajustement 183

4.4. Méthodes alternatives pour les modèles d'ajustement 184
4.4.1. Ajustement de la SED 184
4.4.2. Régression quantile glissante 187
4.4.3. Modèle polynomial 188
4.4.4. Régression multivariée par spline adaptative 191

4.5. Synthèse 195

Chapitre 5. Application de la méthodologie à des projets de construction et de rénovation 197

5.1. GPE en réhabilitation : logements collectifs de FEYZIN 199

5.1.1. Objectifs.....199

5.1.2. Présentation du cas d'étude.....199

5.1.3. Modélisation et simulations.....200

5.1.4. Configuration d'ajustement et plan de mesure 206

5.1.5. Modèles d'ajustement pour la garantie sur l'amélioration de performance 209

5.1.6. Mesures et vérification.....217

5.1.7. Conclusion du cas d'étude de Feyzin.....219

5.2. GPE en construction neuve : bâtiment de bureaux IZUBA 221

5.2.1. Présentation du cas d'étude.....221

5.2.2. Modélisation.....223

5.2.3. Propagation des incertitudes et variabilités.....229

5.2.4. Dispositif de mesure.....230

5.2.5. Engagement de GPE sur les consommations de chauffage et de climatisation ..231

5.2.6. Engagement de GPE sur les consommations électriques totales 238

5.2.7. Vérification des engagements.....240

5.2.8. Bilan du cas d'étude d'Izuba.....246

Conclusion générale 249

Bibliographie 255

Annexes 268

A-1-1 Modèle COMFIE pour la SED 269

A-2-1 Sélection des processus de type ARMA 270

A-2-2 Composantes principales de la décomposition de Fourier de la série tendancielle.....271

A-2-3 Illustration de la méthode de génération des données de RGH 273

A-2-4 Calcul du rayonnement solaire extra-terrestre.....275

A-2-5 Interventions algorithmiques.....277

A-3-1 Incertitude sur les DJU calculés avec la méthode COSTIC.....278

A-5-1 Spécifications nominales des modèles SED de Feyzin avant et après réhabilitation.....279

A-5-2 Comparaison des modèles de $\boldsymbol{LGC5}\%$ pour différents FA et PMV.....282

A-5-3 Schémas des installations techniques du bâtiment Izuba 287

A-5-4 Caractéristiques principales du modèle SED du bâtiment Izuba 288

A-5-5 Dispositif de mesure des performances du bâtiment Izuba 291

A-5-6 Incertitudes de mesure des températures intérieures moyennes 292

A-5-7 Comparaison des modèles de régression quantile 293

A-5-8 Modèles de régression sur l'espérance 294

Nomenclature

NOTATIONS LATINES

C	Amplitude	[-]
$c, cons$	Indice consigne	[-]
cr	Centré réduit	[-]
E	Énergie, consommation d'énergie / Esperance	[kWk]
e, ext	Indice extérieur	[-]
$elec$	Électrique	[-]
$E - T$	Indice de série de rayonnement extra-terrestre	[-]
f	Fréquence	[an^{-1}]
FA	Facteur d'ajustement	[-]
$Fourier$	Indice de série tendancielle (série de Fourier)	[-]
$glob$	Global	[-]
GR	Gain relatif (de consommation post-/pr- réhabilitation)	[%]
h	Horaire	[-]
iid	Indépendant et identiquement distribué	[-]
Im	Partie imaginaire	[-]
int	Intérieur	[-]
IR	Incertitude relative	[-]
j	Journalier	[-]
LGC	Limite garantie de consommation énergétique	[kWh]
m	Moyenne	[-]
mes	Mesuré	[-]
mp	Moyenne pondérée	[-]
n, N	Nombre d'éléments	[-]
N_r	Nombre de répétition (erreurs de mesure virtuelle)	[-]
$norm$	Indice série normalisée	[-]
obs	Observé	[-]
pre	Pre-réhabilitation	[-]
q	Quantile	[-]
r_s	Coefficient de corrélation de Spearman	[-]
RDH	Rayonnement diffus horizontal	[J/cm^2]
Re	Partie réelle	[-]
res	Indice de série résiduelle	[-]

RGH	Rayonnement global horizontal	[J/cm^2]
R^1	Coefficient de régression linéaire quantile	[-]
R^2	Coefficient de régression linéaire sur l'espérance	[-]
s	Coefficients de la série de Fourier discrète	[-]
S	Série de Fourier discrète	[-]
sc	Saison de chauffe	[-]
sf	Saison de refroidissement	[-]
$simu$	Simulation	[-]
sr	Saison de refroidissement	[-]
sys	systématique	[-]
T	Température	[°C/K]
T_c	Température de consigne	[°C/K]
ta	Taux d'accroissement	[-]
ref	Référence	[-]
u	Incertitude	[-]
u_a	Incertitude de type A	[-]
u_b	Incertitude de type B	[-]
X	Variable, variable explicative	[-]
Y	Sortie, variable expliquée	[-]

LETTRE GRECQUES

α	Niveau de risque de défaut de l'engagement de GPE	[-]
β	Coefficient de pente du modèle linéaire	[-]
Δ	Intervalle, différence	[-]
ε	Résidu, série résiduelle	[-]
η	Rendement	[-]
θ	Angle d'incidence, paramètre	[rad/-]
λ	Conductivité thermique	[W/m.K]
μ	Moyenne	[-]
μ^*	Moyenne des valeurs absolues	[-]
ρ	Masse volumique / coefficient du modèle VAR	[kg/m^3]
ρ_p	Coefficient de corrélation de Pearson	[-]
σ	Écart-type	[-]
τ_k	Coefficient de corrélation de Kendall	[-]

ABBREVIATIONS ET ACRONYMES

ACV	Analyse de cycle de vie
AIE	Agence internationale de l'énergie
AMY	*Actual Meteorological Year*
APE	Amélioration de performance énergétique
ARMA	*Auto-regressive and moving average* (modèle autorégressif à moyenne mobile)
AS	Analyse de sensibilité
ASHRAE	*American Society of Heating, Refrigerating and Air-Conditioning Engineers*
BEPOS	Bâtiment à énergie positive
CES	Centre d'efficacité énergétique des systèmes
COP	Coefficient de performance
CPE	Contrat de performance énergétique
DJU	Degrés-jours unifiés
DH	Degrés-heures
ECM	*Energy conservation measure* (cf APE)
ECS	Eau chaude sanitaire
EPBD	*Energy Performance of Buildings Directive*
FA	Facteurs d'ajustement
IPMVP	*International Performance Measurement and Verification Protocol*
GP	*Gaussian process* (processus gaussien)
GPE	Garantie de performance énergétique
GRE	Garantie de résultats énergétiques
LASSO	*Least Absolute Shrinkage and Selection Operator*
LGC	Limite garantie de consommation
LHS	*Latin hypercube sampling*
M&V	Mesure et vérification
MAE	*Mean absolute error*
MARS	*Multivariate adaptive regression splines*
MBE	*Mean biased error*
MC	Monte Carlo / Moindres carrés
MCMB	*Markov Chain Marginal Bootstrap*
MCMC	*Monte Carlo Markov Chain*
OAT	*One-at-a-time*

P+C	Pléiades+COMFIE
PDF	*Probability density function*, fonction de densité de probabilité
PI	Propagation des incertitudes
PIV	Propagation des incertitudes et variabilités
PMV	Plan de mesure et vérification
RDH	Rayonnement diffus horizontal
RGH	Rayonnement global horizontal
RMSE	*Root mean square error*
RQ	Régression quantile
RQL	Régression quantile linéaire
RT	Réglementation thermique
SED	Simulation énergétique dynamique
STD	Simulation thermique dynamique
TMY	*Typical meteorological year*
TRY	*Typical reference year*
VAR	Vecteur autorégressif
VMC	Ventilation mécanique contrôlée
VMI	Ventilation mécanique insufflée

Chapitre 1. État de l'art – Prévision et mesure des consommations énergétiques des bâtiments, vers la garantie de performance énergétique

L'objectif de ce premier chapitre est d'introduire les travaux réalisés dans le cadre de la thèse en présentant à la fois les éléments de contexte au sein desquels ils s'inscrivent et les références bibliographiques sur lesquelles ils s'appuient. Dans un premier temps sont décrits les enjeux associés à la prévision des consommations énergétiques des bâtiments. Les aspects liés à la modélisation de l'ouvrage et de ses usages dans un contexte incertain et imprévisible sont ensuite étudiés. Les méthodes associées à la mesure des performances énergétiques amènent par complémentarité au sujet de la garantie de performance énergétique qui est au cœur de ces travaux. Les méthodologies existantes et leurs limitations sont présentées permettant de définir le positionnement et les objectifs de la thèse. Dans la dernière partie de ce chapitre, les modèles existants et méthodes utilisées sont décrits puis un cas d'étude, servant de fil conducteur lors des différents développements de ces travaux, est présenté.

Sommaire du chapitre 1

Chapitre 1. État de l'art – Prévision et mesure des consommations énergétiques des bâtiments, vers la garantie de performance énergétique 11

Sommaire du chapitre 1 12

1.1. Contexte et enjeux de la prévision des consommations d'énergie 13
1.1.1. La conception éco-énergétique 13
1.1.2. La modélisation d'un système complexe : le bâtiment 14
1.1.3. Les sources d'erreurs et d'incertitudes 17

1.2. Le modèle physico-probabiliste : un changement de paradigme 19
1.2.1. L'exploration de modèles 19
1.2.2. Décrire la variabilité des conditions d'exploitation des bâtiments 25
1.2.3. Abandonner le déterminisme : l'intégration des incertitudes et variabilités 30

1.3. Mesurer pour vérifier la performance 31
1.3.1. Erreurs et incertitudes 32
1.3.2. Protocoles de mesure et vérification 34
1.3.3. État de l'art de la mesure dans le bâtiment 36
1.3.4. Mesurer et inférer l'occupation et les usages 37
1.3.5. Le calibrage de modèle 38
1.3.6. Le constat de l'écart entre prévisions et réalité 40

1.4. La garantie de performance énergétique 42
1.4.1. Enjeux de la garantie de performance 42
1.4.2. Éléments-clés du projet de GPE 45
1.4.3. État de l'art des méthodes 52

1.5. Positionnement de la thèse 55

1.6. Modèles et cas d'étude 57
1.6.1. Modèles et méthodes utilisés 57
1.6.2. Etude de cas : Rénovation de logements collectifs 60

1.7. Conclusion 64

1.1. Contexte et enjeux de la prévision des consommations d'énergie

1.1.1. La conception éco-énergétique

Dans un contexte d'énergie abondante et très bon marché, la forte croissance économique des trente glorieuses entraina une construction massive de logements dont la conception était simplifiée et standardisée, tant en ce qui concerne les procédés constructifs que les matériaux et systèmes énergétiques. L'intégration de la problématique énergétique dans le secteur du bâtiment est consécutive aux chocs pétroliers des années 1970, donnant peu à peu naissance à un nouvel écosystème d'entreprises et de pratiques en constante évolution depuis cette date. Tirée par la question des coûts économiques puis plus récemment du changement climatique, la mutation du secteur du bâtiment s'est concrétisée techniquement sur différents fronts.

Les modes constructifs et matériaux utilisés se sont adaptés aux exigences des réglementations thermiques. L'isolation de l'enveloppe du bâtiment s'est généralisée. Une grande variété de matériaux isolants ont été développés et ce sujet est encore au cœur de différents travaux de recherche (isolants sous vide, aérogel, superisolants…). La maitrise de l'étanchéité de l'enveloppe des bâtiments est aujourd'hui particulièrement travaillée et vérifiée par des tests spécifiques (porte soufflante).

Le développement et l'amélioration de systèmes énergétiques performants ont aussi été portés par cette nécessité d'évolution. Tous les postes de consommation d'énergie sont concernés, qu'il s'agisse de systèmes mécaniques pour la ventilation (hygroréglable, double-flux, VMI), de configurations pour répondre aux besoins de chauffage et de refroidissement (différentes sources énergétiques, vecteurs de chaleur et moyens de distribution) ou de systèmes de production d'eau chaude sanitaire (chaudières gaz, solaire thermique…). La multiplicité des systèmes existants, inventés, et réinventés, comme le montre la cinquantaine de « titres V systèmes » délivrés chaque année (CEREMA, 2016), a pour conséquence un nombre très important de solutions de conception disponibles que ce soit pour les bâtiments neufs ou pour des besoins de rénovation.

On désignera par conception éco-énergétique, l'ensemble de ces choix de conception dans un but d'optimiser le fonctionnement énergétique et de minimiser les consommations. On parle aussi d'écoconception en intégrant la prise en compte des impacts environnementaux liés à l'ouvrage sur son cycle de vie (Peuportier, 2014).

Ces solutions de conception doivent être considérées en fonction des spécificités de chaque projet, des besoins associés, du contexte géographique, climatique et humain. Ce lien entre la conception du bâtiment lui-même, son environnement d'implantation et son usage est au cœur de la

conception éco-énergétique. Afin de résoudre cette problématique aux multiples implications et degrés de liberté, des outils spécifiques basés sur la modélisation des bâtiments ont été développés.

1.1.2. La modélisation d'un système complexe : le bâtiment

1.1.2.1. Pourquoi simuler ?

La modélisation physique et numérique des bâtiments afin d'en simuler le fonctionnement énergétique répond à plusieurs objectifs. Les bâtiments sont des objets complexes dont la « qualité énergétique » est difficile à évaluer au premier abord. Les constantes de temps mises en jeu sont très grandes, qu'elles concernent l'évolution des conditions thermiques au fil des jours ou plus simplement la longue durée de vie des ouvrages. Ce sont par ailleurs des objets qui sont souvent uniques. La notion de prototypage utilisée dans le secteur industriel est ainsi quasiment impossible à mettre en œuvre ; il ne peut y avoir de tests en taille réelle pour chaque projet de bâtiment. En phase de conception, la comparaison de différentes variantes de projets ou de sous-parties de projets doit souvent s'appuyer sur des données provenant d'outils de simulation. La qualité des ouvrages dépend de différents critères et est évaluée de façon complexe. On peut ainsi s'intéresser aux consommations énergétiques futures, aux coûts, au confort thermique, mais aussi à la qualité environnementale de l'ouvrage et aux impacts induits tout au long de sa durée de vie. Enfin les performances dépendent de nombreuses sollicitations dont les effets ne peuvent être facilement appréhendés. Des modèles assez fins sont donc nécessaires pour maitriser le développement des projets de construction et constituent des outils d'aide à la décision en conception.

Les modèles de simulation trouvent aussi une application dans l'exploitation et la gestion en temps réel du fonctionnement énergétique des bâtiments. Ces technologies permettent de commander les systèmes mais aussi de donner des informations aux utilisateurs en fonction de la connaissance apportée par la simulation. Elles participent ainsi à l'optimisation de l'exploitation du bâtiment.

Par ailleurs les performances énergétiques des bâtiments sont réglementées et des outils de modélisation dédiés permettent d'évaluer la qualité des bâtiments par rapport à la réglementation. Différents objectifs normés sont imposés et vérifiés par l'outil de calcul réglementaire. La RT 2012 fixe ainsi un objectif de consommation d'énergie primaire de 50 kWh/m²/an pour les bâtiments neufs. Cette exigence, modulée par différents critères, s'accompagne de limitations sur les besoins d'énergie et sur l'inconfort estival dans les bâtiments non climatisés. Il existe aussi une réglementation thermique pour les bâtiments existants traitant des exigences s'appliquant à la rénovation globale ou par éléments des ouvrages. L'évolution de la réglementation thermique, dont une mise à jour est prévue pour 2020, ne remettra a priori pas en cause cette approche d'évaluation

de la performance énergétique à l'aide d'un modèle de simulation, intégrant même en complément un moteur de calcul des impacts carbone sur le cycle de vie du bâtiment[1].

La connaissance a priori permise par la simulation permet aussi d'évaluer la faisabilité économique d'un projet. Par exemple, pour des travaux de rénovation, une prévision robuste des consommations permet d'anticiper les gains énergétiques et économiques induits. Pour répondre à cet objectif, il est cependant nécessaire d'avoir des résultats de simulation très précis et les retours d'expérience tendent à montrer que l'écart entre performances réelles et prévisions de simulations est souvent important, ces dernières sous-estimant fréquemment les consommations mesurées dans la réalité. La Figure 1-1 présente ainsi un ensemble de retours d'expérience, pour différents types de bâtiments étudiés dans la littérature (van Dronkelaar et al., 2016).

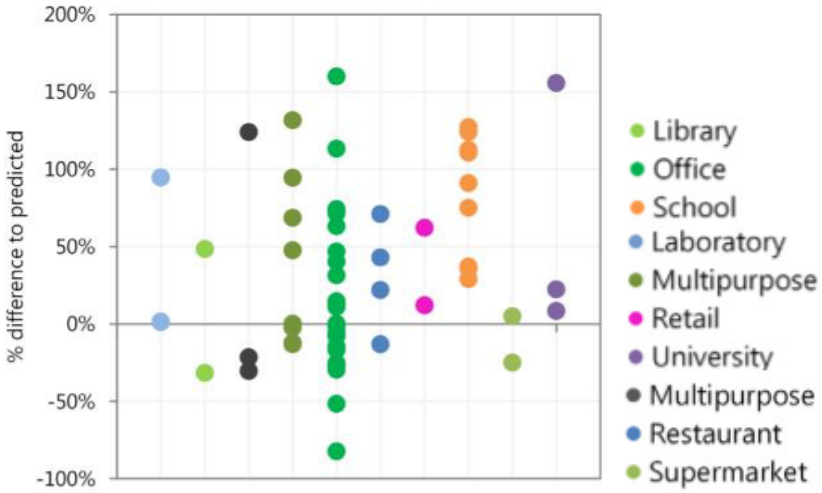

Figure 1-1 : Différences relatives (en pourcentage) entre les consommations énergétiques mesurées et prédites pour différents cas d'étude (van Dronkelaar et al., 2016)

Afin d'encourager les investissements dans la rénovation et la construction vertueuse, les outils doivent donc évoluer pour estimer avec un haut niveau de confiance les consommations futures des bâtiments. Il est aussi primordial d'effectuer des travaux avec le niveau d'ambition adéquat. Afin de ne pas « tuer le gisement » d'économie d'énergie, il est préférable d'investir une fois dans une solution très performante et plus coûteuse que d'envisager une évolution successive de solutions individuellement moins chères et moins performantes (Peuportier, 2014).

L'ensemble de cette connaissance et les applications qui en découlent reposent donc sur la capacité à anticiper par la simulation le fonctionnement des bâtiments. Pour ce faire, différentes méthodes et outils ont été développés.

[1] http://www.batiment-energiecarbone.fr/

1.1.2.2. Les différents types de modèles

Différentes catégories de modèles sont utilisées pour simuler le fonctionnement énergétique des bâtiments et essayer de prévoir leur consommation. On classe ces modèles suivant le type de données nécessaires à leur construction et le niveau de description physique qu'ils cherchent à atteindre. Coakley, et al. (2014a) récapitulent les principaux avantages et inconvénients de ces différentes familles de modèles.

On parle de modèles « boite noire », *Black-box* ou *data-driven*, pour des modèles représentant le fonctionnement d'un système sans décrire les phénomènes physiques de son fonctionnement interne. Ces méthodes reposent sur la prévision de données futures à partir de l'exploitation statistique de données connues. Ce sont des modèles purement mathématiques, rapides à développer et à utiliser. Ils ont en commun de nécessiter très peu de connaissances physiques concrètes sur les bâtiments étudiés mais sont construits à partir de données existantes, déjà mesurées ou provenant d'autres outils de simulation. La qualité de la prédiction effectuée par ce type de modèles dépend donc de la qualité des données utilisées pour les développer et se dégrade rapidement en dehors des plages de variation sur lesquelles ils ont été construits. Suivant les modèles, ces données peuvent provenir de bâtiments similaires à celui étudié ou de mesures réalisées sur des périodes antérieures. Un inconvénient de ce type de méthodes vient du fait que des modifications sur le bâtiment décrit impliquent la nécessité de ré-entrainer le modèle avec de nouvelles données. Ces modèles décrivent en principe les consommations d'énergie des bâtiments à partir d'un nombre limité de variables d'entrée. Li *et al.* (2014) listent le détail des méthodes de ce type couramment utilisées. Des modèles simplifiés sont construits par régression linéaire multiple, mais des formes de métamodèles plus complexes existent tels que les processus gaussiens ou les réseaux de neurones. Une partie de ces modèles, utilisés dans ces travaux, seront décrits plus précisément au paragraphe 1.3.5.

Il existe aussi des modèles hybrides, ou modèles *Grey-Box*, mixant une approche statistique et une approche physique. L'architecture générale est en principe basée sur des lois physiques mais certaines parties du modèle reposent sur l'exploitation mathématique de données connues et constituent des composants assimilables à des *Black-Box*. Ce type de modèles s'appuie donc en partie sur l'accessibilité à des données fiables de construction des modèles statistiques mais aussi sur la connaissance des paramètres d'entrée des modèles physiques.

Les modèles « boite blanche », ou modèles *White-box*, ont l'ambition de décrire le fonctionnement du bâtiment en modélisant directement les lois de la physique qui régissent les phénomènes mis en jeu. Les outils de simulation thermique dynamique (STD), appelés aussi simulation énergétique dynamique (SED), intégrant à l'heure actuelle une prise en compte de phénomènes plus larges que la description des conditions thermiques du bâtiment, font partie de cette catégorie de modèles (Peuportier, 2016).

1.1.2.3. Spécificités de la simulation énergétique dynamique (SED)

Afin de conduire une simulation énergétique dynamique, une description géométrique et physique du bâtiment est d'abord réalisée. L'ensemble des composants de l'enveloppe sont spécifiés ainsi que les différentes sollicitations du bâtiment telles que des données météorologiques, des scénarios de présence, de puissance dissipée, de renouvellement d'air… Lee et al. (2015) listent et détaillent les qualités de différents outils de modélisation, dont une grande partie d'outils de simulation dynamique avancée. Crawley et al. (2008) proposent eux aussi une étude comparative des capacités de prédiction des modèles SED mise à jour récemment (Crawley et Hand, 2017). En France, les principaux logiciels utilisés sont *EnergyPlus*, *TRNSYS* et *Pléiades*+COMFIE (P+C). Les deux premiers modèles sont basés sur des fonctions de transfert. Dans le modèle COMFIE, les échanges de chaleur par conduction, convection et rayonnement sont décrits dans un modèle aux différences finies avec un pas de temps allant de quelques minutes à une heure. Les travaux présentés ici s'appuieront sur ce dernier outil, décrit au paragraphe 1.5.1, en ce qui concerne la simulation énergétique des bâtiments.

Bien que ces modèles revendiquent une description fine des phénomènes physiques, ils nécessitent une grande quantité de données d'entrée concernant le bâtiment en lui-même et la description des sollicitations. Certaines de ces données sont mal connues ou imprévisibles et cette méconnaissance s'ajoute aux propres limites du modèle.

1.1.3. Les sources d'erreurs et d'incertitudes

Il est nécessaire de connaitre au mieux les limitations des outils de modélisation. Ces sources d'incertitudes, amenant à une erreur de prédiction, peuvent être classées en quatre grandes catégories (Fabrizio et Monetti, 2015).

- On peut tout d'abord prendre en compte les **incertitudes de modélisation** survenant lors du développement du modèle à partir des hypothèses, approximations et simplifications considérées (zonage, simplification de la géométrie, phénomènes physiques simplifiés ou négligés, etc.).
- Il existe aussi des **incertitudes de simulation** introduites lors de la discrétisation et de la résolution numérique du modèle.
- On peut aussi considérer les **incertitudes de spécification** (des matériaux, des systèmes, de la géométrie) liées au manque d'informations sur les propriétés exactes des matériaux utilisés, sur la géométrie du bâtiment ou encore sur le fonctionnement des systèmes énergétiques.
- Enfin, les **incertitudes liées aux sollicitations**, aux scénarios d'usage (comportement des occupants) et aux conditions météorologiques ne doivent pas être négligées. La prise en compte de cette source majeure d'écart entre prévisions et mesures fera l'objet d'une étude particulière au sein de ces travaux.

Pour les modèles statistiques construits à partir de données provenant de mesures ou pour l'étude de l'écart entre résultats de simulation et données réelles, il faut aussi intégrer les incertitudes de mesure et d'agrégation des données mesurées utilisées. Ce sujet est abordé au paragraphe 1.3.2.

La gestion des incertitudes de modélisation et de simulation est liée à la question de la validation des modèles. Celle-ci est primordiale dans une démarche scientifique. La modélisation de bâtiments expérimentaux a permis de comparer différents logiciels de SED entre eux et d'opposer leurs résultats aux mesures réalisées. On trouve aussi des cas de validation du modèle COMFIE dans les articles de Recht et al. (2014) et Brun et al. (2009). La Figure 1-2 présente la comparaison de prédictions de puissances et de besoins de chauffage de plusieurs logiciels sur un cas d'étude.

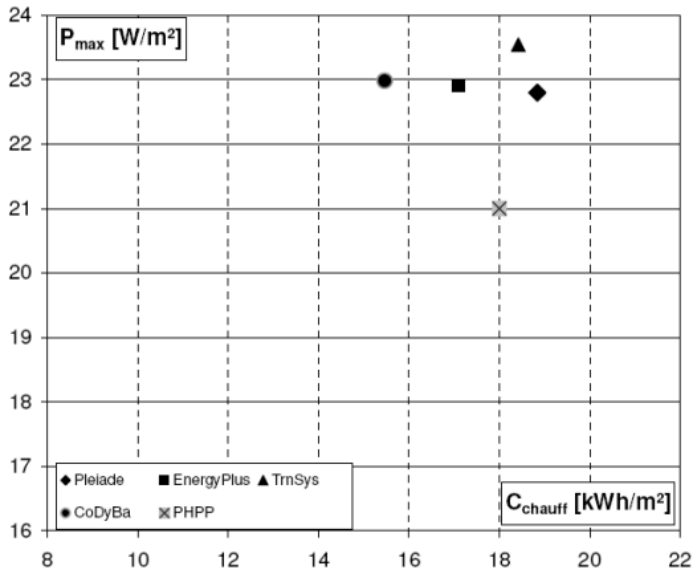

Figure 1-2 : Comparaison des résultats de puissance maximale et besoins de chauffage pour différents codes de calcul de STD (Brun 2009)

Ainsi, pour la suite de ce document, on s'intéressera de manière détaillée aux incertitudes liées aux spécifications techniques des bâtiments et aux entrées dynamiques du modèle liées à l'occupation et à la météorologie. En effet, les résultats de validation des outils SED attestent de leur fiabilité même si de faibles différences persistent entre les résultats des différents modèles. Cette source d'erreur, considérée comme négligeable par rapport aux autres limitations identifiées, ne sera a priori pas prise en compte dans les méthodes développées. Cependant, l'intégration de son influence sera envisagée et discutée dans la suite de ce document. Par ailleurs, on considère qu'il n'y a pas de source d'erreur provenant de la mauvaise utilisation des outils par un opérateur, aspect non négligeable en réalité. Concernant les incertitudes liées aux spécifications et aux sollicitations, cœur du sujet de ces travaux, des revues bibliographiques sont proposées dans la suite de ce chapitre.

1.2. Le modèle physico-probabiliste : un changement de paradigme

1.2.1. L'exploration de modèles

Le constat des différentes sources d'erreurs souligne les limitations des modèles de SED pour un objectif de prédiction des consommations réelles d'énergie. La complexité inhérente au système que l'on souhaite modéliser réclame un changement de vision et la mise en place d'une approche plus globale et intégrée si l'on souhaite progresser vers la compréhension et la prévision des phénomènes énergétiques dans le bâtiment. Pour cela, il est nécessaire d'agrandir le périmètre de l'étude, le système d'intérêt, pour le considérer dans toute sa complexité (Morin, 1990). La nécessité de considérer le bâtiment avec une approche systémique pour répondre aux problématiques énergétiques et environnementales actuelles a aussi été théorisée par Gobin (2014). Dans notre cas, cela se traduit par l'intégration des données d'entrée du modèle thermique (SED) du bâtiment au cœur d'un modèle plus global, où leurs valeurs sont fournies par des modèles statistiques plus ou moins complexes dont les paramètres constituent les nouvelles entrées à spécifier.

Ce changement de paradigme repose donc sur l'intégration du caractère incertain et variable des entrées classiques du modèle énergétique du bâtiment dans un modèle plus large où l'approche statistique est généralisée.

1.2.1.1. Caractériser l'incertitude des paramètres

La caractérisation de l'incertitude associée à la valeur des paramètres d'un modèle physique de bâtiment peut être réalisée avec différents niveaux de détails. Une simple plage de valeurs peut être définie avec des bornes minimales et maximales. Cette description est équivalente à une fonction de densité de probabilité (PDF, *probability density function*) uniforme. Par ailleurs, d'autres PDF peuvent être utilisées pour décrire les valeurs possibles d'un paramètre statique. Celles-ci peuvent être par exemple définies à partir de valeurs mesurées. Les lois les plus couramment utilisées sont les lois de probabilité uniformes, normales et les lois log-normales, tronquées ou non (Figure 1-3).

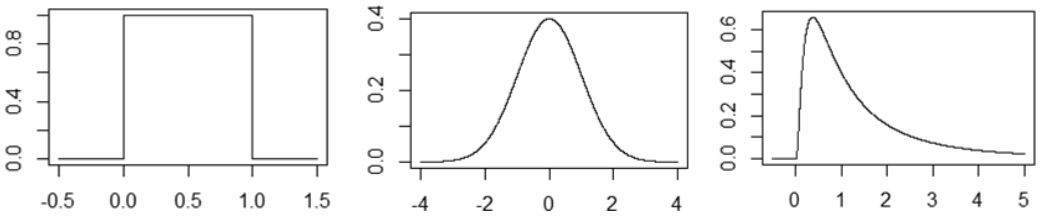

Figure 1-3 : Exemple de lois de distribution, uniforme, normale et log-normale

Ces fonctions sont construites à partir de différentes sources d'information disponibles hiérarchisées selon leurs origines (Silva et Ghisi, 2014). Par exemple, une information obtenue à partir d'une mesure, ou un certificat d'étalonnage, sera jugée plus fiable qu'une information tirée de dires d'experts. Pour les paramètres physiques, des lois normales tronquées ou des lois uniformes sont majoritairement utilisées (Spitz et al. 2012), (Macdonald 2002). Le choix des lois les plus appropriées est un exercice difficile, en partie subjectif (Faivre et al., 2013) et soumis à des limites de validité dont il faut avoir conscience. En cas de difficultés, une approche conservatrice, surestimant l'incertitude, est à préférer.

1.2.1.2. Analyse de sensibilité

L'incertitude sur un paramètre d'entrée du modèle peut avoir un impact plus ou moins important sur les résultats de sortie de simulation énergétique dynamique (SED). Ce lien entre la variabilité induite sur la sortie et l'incertitude sur le paramètre d'entrée est appelé sensibilité. La sensibilité du modèle par rapport à un paramètre donné dépend à la fois de l'incertitude sur ce paramètre mais aussi de l'influence qu'il a sur la sortie d'intérêt. Ainsi, peu importe l'influence d'une entrée si sa valeur est connue de manière précise. Inversement si l'influence d'une entrée est très faible, il n'est pas problématique d'en avoir une connaissance imprécise.

L'analyse de sensibilité (AS) consiste à quantifier l'influence de paramètres inconnus à partir d'une estimation de l'incertitude sur leurs valeurs. Cette méthode permet de mieux connaitre le fonctionnement du modèle et l'importance des liens entre ses entrées et sorties. Les applications concrètes peuvent par exemple concerner la spécification de bonnes pratiques de construction via l'identification de réalisations techniques-clés dont la qualité impacte fortement les performances énergétiques du bâtiment.

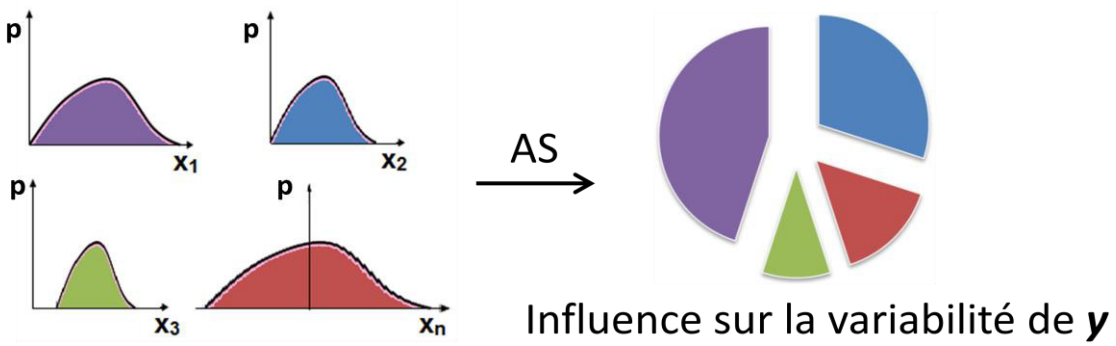

Figure 1-4 : Principe de l'analyse de sensibilité

Différentes méthodes d'analyse de sensibilité ont été développées. Tian (2013) et Iooss (2011a) proposent des revues des méthodes d'analyse de sensibilité, ainsi que des classifications. Le Tableau 1-2 présente les principales méthodes d'analyse de sensibilité et les études associées dans le domaine de la modélisation des bâtiments (Pannier, 2017).

Les méthodes d'analyse de sensibilité (AS), sont directement liées à la problématique de la robustesse des modèles et des résultats de simulation, abordée dans ces travaux. Cependant, la description exhaustive de ces méthodes ne sera pas réalisée ici car les AS sont utilisées de manière limitée et/ou facultative dans les méthodologies qui seront étudiées.

Tableau 1-1 : Classification des méthodes d'analyse de sensibilité (Pannier, 2017)

Type d'AS	Approche	Méthode
AS qualitative	Graphique	Nuage de points (*scatterplot*), *Cobweb plot* (Kurowicka et Cooke, 2006)
AS locale	Locale	Dérivées partielles
Criblage	Criblage à très grande dimension	Plans supersaturés (Satterthwaite, 1959 ; Lin, 1993)
		Criblage par groupe (Dean et Lewis, 2006), bifurcations séquentielles (Bettonvil et Kleijnen, 1997)
	Plans d'expérience	Plan factoriel complet
		Plan factoriel fractionnaire : OAT, plans simplexes, Plackett et Burman, plans composites centrés, Box et Behnken, plans hybrides…
	Effets élémentaires	Méthode de Morris (Morris, 1991)
AS globale	Régression	Régression linéaire : Pearson, SRC, PCC
		Régression basée sur les rangs : Spearman, SRRC, PRCC
	Tests statistiques	Sur les moyennes, médianes, variances des échantillons, AS régionalisée
	Basées sur la décomposition de la variance	Calcul d'indices de Sobol (I. Sobol, 1993)
		Basée sur analyse fréquentielle : FAST (Cukier et al., 1973), E-FAST (Saltelli et al., 1999), RBD-FAST (Tarantola et al., 2006 ; Mara, 2009), EASI (Plischke, 2010), COSI (Plischke, 2012)
Exploration du modèle	Méthodes de lissage	Splines de lissage, polynômes locaux (Veiga et al., 2008), …
	Métamodèles	Polynômes de chaos, krigeage…
Autre	Fiabiliste	FORM et SORM

Pour plus de détails, le lecteur se rapportera aux revues précédemment citées (Iooss, 2011a ; Tian, 2013) ou aux travaux de Pannier (2017). Les méthodes d'AS spécifiquement utilisées dans cette thèse seront décrites au paragraphe 1.5.

1.2.1.3. Analyse d'incertitude

Les méthodes de propagation d'incertitude, ou d'analyse d'incertitude, ont pour objectif d'estimer l'incertitude sur les sorties des modèles, connaissant celles sur les entrées (Figure 1-5). Il existe deux grands types de méthodes : les méthodes ensemblistes, adaptées à des modèles simples et possédant peu de paramètres d'entrée, et les méthodes probabilistes dont le domaine

d'applicabilité est beaucoup plus large. On parlera plus généralement d'approches internes et externes. Le Tableau 1-2 détaille ainsi les principales méthodes rencontrées dans la littérature.

Tableau 1-2 : Classification des méthodes d'analyse d'incertitude proposée par (Pannier 2017)

Approche	Méthode
Interne	Propagation d'incertitude analytique – méthode du cumul quadratique
	Méthode ensembliste : propagation d'intervalles (Moore, 1979 ; Neumeier, 1990)
	Approche possibiliste : ensembles flous (Dubois et Prade, 1988)
Externe	Méthodes basées sur un échantillonnage : aléatoire (Monte Carlo), stratifié (échantillonnage par hypercube latin : LHS), échantillonnage quasi-aléatoire (suites déterministes)
	Probabilité de dépassement d'un seuil : FORM et SORM

Nous nous concentrerons ici sur les méthodes externes utilisées dans le cadre de modèles de SED, caractérisés par un nombre important d'entrées incertaines.

Les méthodes probabilistes s'appuient sur l'échantillonnage de valeurs des entrées dans leur domaine d'incertitude, éventuellement en fonction d'une loi de probabilité. Pour chaque échantillon de valeurs des entrées, une simulation est réalisée. La dispersion des résultats de simulation caractérise directement l'incertitude les concernant.

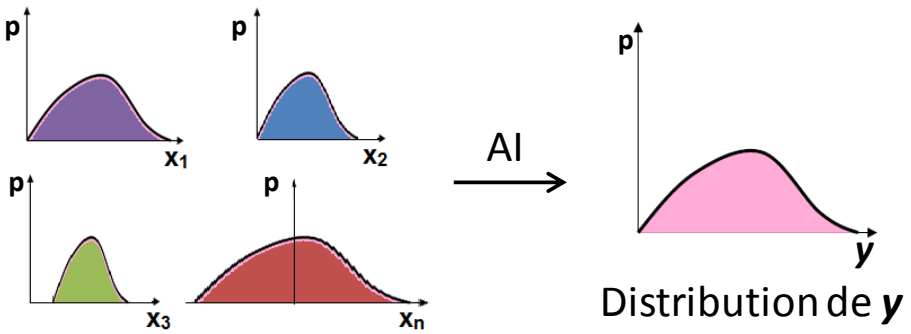

Figure 1-5 : Principe de l'analyse d'incertitude

Différents types de tirages sont mentionnés dans la littérature, l'objectif étant d'atteindre le plus rapidement et précisément possible la convergence des propriétés de la distribution des sorties des modèles.

- *Échantillonnage aléatoire*

La méthode la plus simple et la plus utilisée est l'échantillonnage aléatoire aussi appelée méthode de Monte-Carlo. Elle repose sur la génération aléatoire de vecteurs des paramètres en tenant compte de leur fonction de densité de probabilité. Les performances de cette méthode ne dépendent pas du nombre de variables d'entrée et la convergence de la fonction de densité de probabilité de la sortie s'effectue avec une vitesse de $N^{-1/2}$ où N est le nombre d'échantillons utilisés.

- *L'échantillonnage stratifié*

Pour améliorer la rapidité de la convergence, une meilleure répartition des tirages des valeurs des entrées dans leur domaine de variation peut être étudiée. L'échantillonnage stratifié repose sur le découpage de l'espace des entrées en sous-espaces et sur la réalisation d'un échantillonnage aléatoire dans chacun de ces sous-espaces. Il existe différents moyens de diviser l'espace des entrées et plus de détails sont disponibles dans (Owen, 2003). L'échantillonnage par Hypercube Latin (*Latin Hypercube Sampling* ou *LHS*) est une des méthodes d'échantillonnage stratifié fréquemment utilisées (Spitz, 2012 ; Sun et al., 2016). Elle rajoute une hypothèse d'équiprobabilité des différents sous-espaces générés.

- *L'échantillonnage quasi aléatoire*

Dans les méthodes d'échantillonnage quasi-aléatoire (Quasi Monte-Carlo), une logique déterministe est ajoutée aux suites de tirages réalisés. Une séquence est définie et utilisée pour choisir les variations des valeurs de paramètres d'entrée. Plusieurs méthodes reposant sur ce principe ont été développées. Certaines intègrent une génération aléatoire des séquences déterministes. On parle ainsi de méthodes de Quasi Monte-Carlo Randomisées. La construction des séquences se base sur la minimisation d'un critère de discrépance. Ce type de méthode permet la convergence la plus rapide, de l'ordre de $\frac{\log(\mathrm{N})^{\mathrm{K}}}{\mathrm{N}}$. Cependant, la construction de suites entraine l'apparition de motifs.

La Figure 1-6 présente les résultats de tirages de points dans un espace borné à deux dimensions avec différentes méthodes d'échantillonnage. D'autres types de tirages, reposant sur des hybridations des mécanismes présentés, peuvent être développés permettant de générer des plans d'expérience optimisés (Faivre et al., 2013).

Dans le domaine de la simulation des performances énergétiques des bâtiments, plusieurs études s'appuient sur ces différentes méthodes pour investiguer l'impact des incertitudes sur les spécifications techniques et les autres entrées statiques des modèles de bâtiments. Sun et al. (2016) effectuent une analyse d'incertitude suite à un tirage LHS sur 17 paramètres d'entrée d'un modèle de bâtiment dont les intervalles d'incertitude sont fixés après enquêtes auprès de professionnels du secteur. Suite à l'indentification des paramètres influents, Silva (2014) réalise une propagation d'incertitude par tirage quasi Monte-Carlo (*LHS*) et déduit la fonction de densité de probabilité sur les sorties du modèle. Titikpina et al. (2015a) comparent différentes méthodes d'évaluation de l'incertitude sur les résultats d'un modèle simplifié construit à partir de mesures. La méthode par tirage de Monte-Carlo est comparée à une approche bayésienne et aux techniques d'estimation d'incertitude provenant du *Guide to the expression of uncertainty in measurements* (JCGM, 1995). Les trois approches donnent des résultats très similaires. Sur une maison individuelle modélisée avec EnergyPlus, Silva et Ghisi (2014) mènent une démarche couplant propagation d'incertitude et analyse de sensibilité. Après avoir défini des fonctions de densité de probabilité des paramètres

incertains du modèle, un tirage LHS est réalisé. Des simulations permettent de déterminer les distributions des besoins de chauffage et de refroidissement.

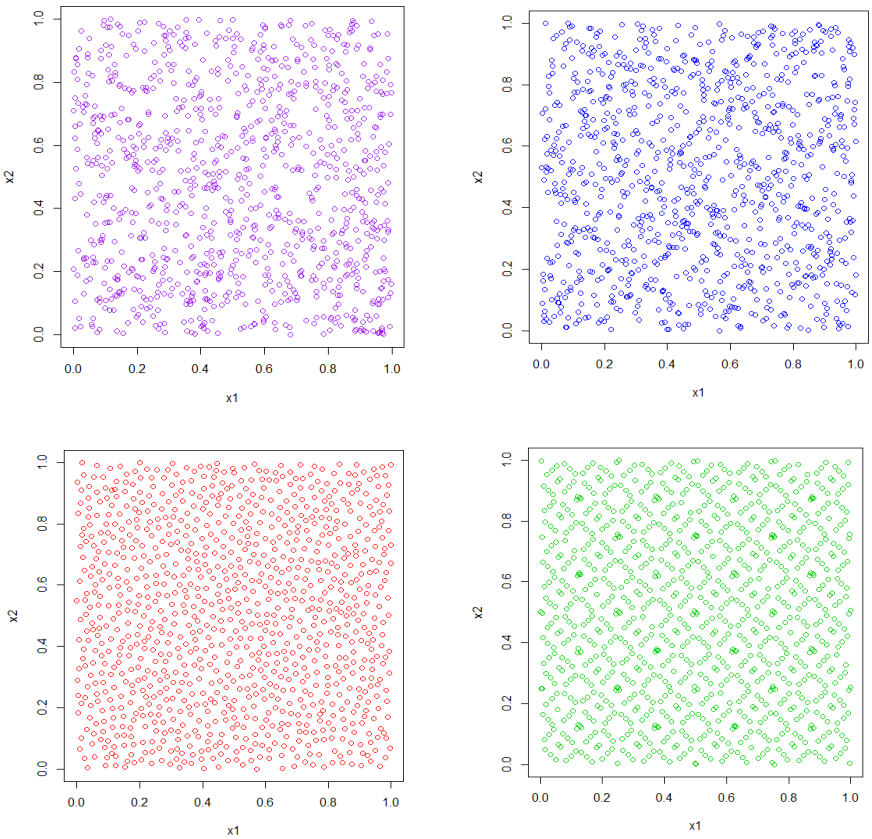

Figure 1-6 : Échantillonnage (1000 tirages) de deux variables aléatoires x_1 et x_2 de loi normale, de type aléatoire (violet), stratifié aléatoire (bleu), stratifié (rouge), quasi-aléatoire par séquences de Sobol (vert) - (Munaretto, 2014)

Ainsi, à partir de l'état de l'art et des différents projets réalisés, on peut proposer un ordre logique d'utilisation commune des méthodes d'analyse de sensibilité et d'incertitude présentées ci-dessus :

- Recenser les paramètres incertains de la simulation.
- Évaluer un domaine d'incertitude pour ces entrées.
- Réaliser une analyse de sensibilité permettant de classer les entrées selon leur influence.
- Effectuer un choix des entrées influentes sur lesquelles poursuivre l'étude avancée.
- Définir précisément des fonctions de densité de probabilité pour ces entrées.
- Réaliser la propagation d'incertitude et les analyses statistiques sur les sorties d'intérêt.

Des variantes de cette méthodologie sont utilisées notamment par Silva et Ghisi (2014) et Lee et al. (2013), dans un objectif d'analyse d'incertitude. Elles est aussi mise en œuvre sur plusieurs cas tests dans le projet (FBE, 2014).

1.2.2. Décrire la variabilité des conditions d'exploitation des bâtiments

Les paramètres considérés au paragraphe précédent sont dits « statiques », ce sont par exemple les grandeurs physiques liées aux matériaux, des valeurs géométriques ou des spécifications techniques des systèmes énergétiques. Ces variables possèdent une valeur unique utilisée durant la SED. Cependant, les simulations nécessitent aussi différents scénarios et sollicitations dynamiques. Ces séries temporelles décrivent notamment la présence des occupants dans le bâtiment, la température de consigne, les puissances dissipées mais aussi les sollicitations météorologiques : la température extérieure et le rayonnement. L'incertitude sur ces données peut être prise en compte grâce à un paramètre unique, un taux de variation du scénario original par exemple, ou un écart absolu constant par rapport à celui-ci (Lee et al., 2013 ; Munaretto, 2014). De cette manière, l'incertitude de ces grandeurs peut être intégrée dans les méthodes de propagation d'incertitude précédemment décrites.

Cette variation relative unique a du sens pour décrire l'incertitude liée à des conditions météorologiques mesurées dont les valeurs sont utilisées a posteriori pour comparer une situation réelle passée à la simulation. Cependant, dans un objectif de prévision a priori, cette approche représente mal la variabilité, naturelle et imprévisible, de ces grandeurs et donc l'impact possible de cette variabilité sur les consommations énergétiques réelles.

Dans la suite de ce document, une différenciation entre les termes d'incertitude et de variabilité sera introduite pour clarifier ce positionnement. Se plaçant dans un contexte de prévision a priori, on parlera d'incertitude pour les paramètres statiques dont la valeur vraie, bien que spécifiée dans le projet de construction ou de rénovation du bâtiment, est incertaine. On considérera une variabilité sur les grandeurs dynamiques dont la valeur vraie future, bien que pouvant être encadrée, est imprévisible. Cette variabilité ne peut être réduite par une meilleure connaissance du système, contrairement à l'incertitude sur les paramètres physiques et techniques du modèle de bâtiment. Par ailleurs, le terme décrit mieux le fait que les conditions ne seront pas identiques d'une année à l'autre durant l'exploitation de l'ouvrage.

1.2.2.1. Les sollicitations météorologiques

Les données météorologiques utilisées dans les simulations sont associées à une position géographique. Des données au pas de temps horaire sur un an sont spécifiées, décrivant différentes grandeurs climatiques. La température extérieure et les composantes du rayonnement solaire sont en particulier nécessaires. Des stations météorologiques couvrent le territoire national et international et leurs relevés permettent de construire des fichiers de données représentant les variations météorologiques sur une année réelle, utilisables pour vérifier les performances d'un

bâtiment sur cette période. Des jeux de données types sont aussi créés à partir de l'étude de données réelles.

L'intégration de la variabilité des conditions météorologiques au modèle physico-probabiliste nécessite une description fidèle de cette variabilité. Cette thématique et les problématiques qui y sont associées sont abordées au chapitre 2, où est présenté un état de l'art spécifique au sujet.

1.2.2.2. Usages et occupation

L'autre source primordiale d'incertitude sur les prévisions de consommations énergétiques provient du comportement des occupants dont l'impact est lié à différents facteurs. On peut ainsi citer les températures de consigne de chauffage et de refroidissement, souvent sous la responsabilité des usagers, les consommations d'électricité spécifiques, fournissant une dissipation d'énergie thermique, les apports de chaleur métabolique, mais aussi la gestion des occultations et des ouvertures de fenêtres, voire éventuellement la régulation de la ventilation. Tous ces éléments sont décrits avec des scénarios horaires ou infra-horaires dans les logiciels de SED. Ils sont eux aussi de nature variable et imprévisible.

La Figure 1-7 illustre l'influence des usages dans les bâtiments. Elle présente la répartition des consommations de chauffage de 290 logements identiques sur plusieurs années de mesure (Andersen, 2012). Les bâtiments étudiés et les conditions météorologiques étant les mêmes pour une année donnée, la très large distribution des consommations est uniquement induite par la différence de comportement des personnes habitant ces logements.

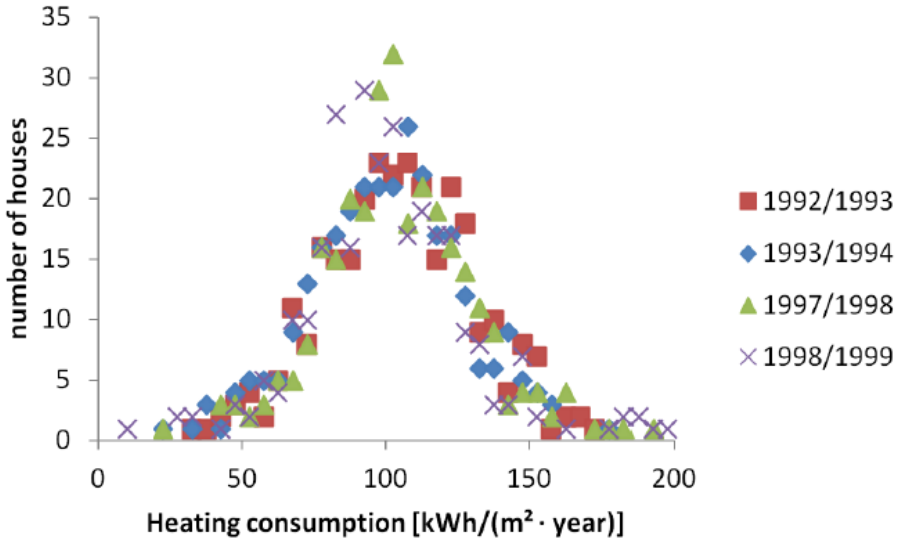

Figure 1-7 : Répartition des consommations énergétiques d'un parc de bâtiment (Andersen, 2012)

Yan et al. (2015) proposent un état de l'art des modèles d'occupation pour les outils de simulation énergétique du bâtiment. Ceux-ci peuvent être séparés en différentes catégories.

- *Modèles de confort avec approche orientée agent*

Les modèles de confort reposent sur l'idée que les actions des occupants ont des causes variées et visent à rétablir une situation de confort. La caractérisation et la modélisation de cette notion de confort sont donc nécessaires. Les sollicitations environnementales constituent les entrées de ces modèles qui proposent en sortie différentes actions afin de rétablir les conditions acceptables de confort. Pour construire ce type de modèles, il faut théoriquement intégrer tous les aspects participant au confort : le confort visuel, acoustique, thermique, la qualité de l'air… Vorger (2014) propose un état de l'art sur ces aspects et leur prise en compte dans les modèles de simulation. Les études concernant le confort thermique sont nombreuses et font apparaître différentes notions-clés. Le bien être thermique ne dépend pas seulement de la température du milieu où l'on se trouve et un bilan énergétique complet est nécessaire. Les mouvements d'air provoquent des refroidissements qui peuvent être locaux, et des effets de perte de chaleur par rayonnement peuvent aussi survenir notamment dus au phénomène de paroi froide. Le métabolisme humain impact aussi le confort ; le contexte météorologique et les variations temporelles des conditions créent des effets adaptatifs. Enfin, les conditions de confort sont variables d'une personne à une autre. Les modèles de confort développés, qu'ils se réclament de l'approche statique ou adaptative, sont donc soumis à des limites de validité importantes sur la description du niveau de confort. Les modèles orientés agent reposent sur la description d'une population occupante et de règles d'interaction entre ces agents et leur environnement et entre ces occupants eux-mêmes. La possibilité de ces interactions va dans un cadre donné permettre à chacun de ces agents d'assurer sa situation de confort suivant des fonctions objectif. Ces modèles, dont des formes actuelles incluent des algorithmes d'auto-apprentissage, nécessitent des connaissances socio-psychologiques poussées. On pourra ainsi citer les travaux de Kashif et al. (2013) sur le développement de modèles causaux sur le comportement des occupants dans des bâtiments résidentiels.

- *Approche déterministe*

Afin de prendre en compte les modes d'occupation, l'approche déterministe consiste à utiliser différents scénarios-types basés sur des études statistiques ou sur des connaissances expertes spécifiques aux bâtiments étudiés. Les logiciels de simulation proposent ainsi de spécifier les scénarios qui seront utilisés dans la simulation. La réglementation impose aussi des scénarios simplifiés sous forme de fonctions temporelles constantes par morceaux pour la réalisation des calculs réglementaires. Des exemples de ces scénarios sont présentés à la Figure 1-8.

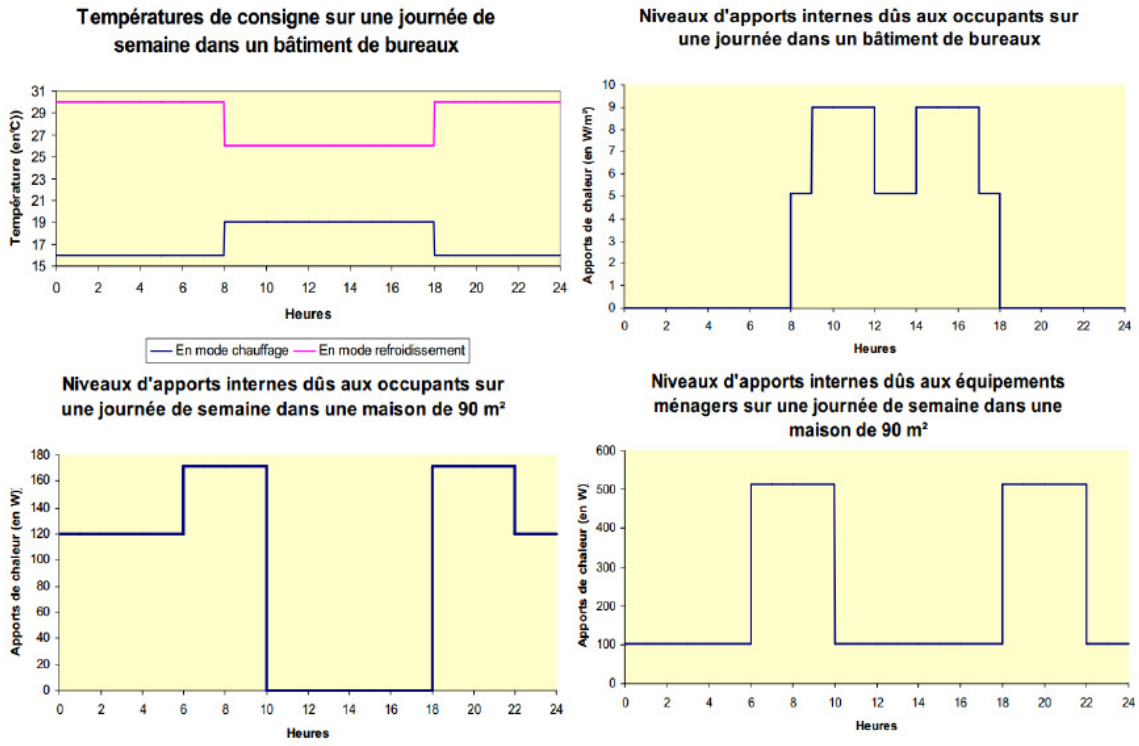

Figure 1-8 : Répartition Exemples de scénarios réglementaires liés à l'occupation d'après les méthodes de Calcul Th-BCE 2012

Les scénarios uniques et simplifiés ne décrivent pas la réalité physique des phénomènes qu'ils souhaitent représenter et encore moins la diversité possible des conditions. L'utilisation de ce type de données d'entrée est donc possible pour certaines applications de la simulation énergétique du bâtiment mais ne peut être intégrée à une prévision fiable des consommations réelles des bâtiments.

- *Modèles stochastiques*

En s'appuyant sur des données statistiques compilant de grandes quantités de données comportementales, on peut construire des modèles reposant sur la notion de probabilité conditionnelle de modes d'usage des bâtiments. Il est par exemple possible d'estimer la probabilité de la taille des ménages suivant la taille de l'appartement considéré, ou l'utilisation des appareil électroménagers suivant l'heure et le profil de présence.

Ce type de modèle s'appuie sur des tirages aléatoires connaissant certaines caractéristiques permettant de se placer dans le cadre d'une étude statistique donnée. Un grand nombre de simulations est réalisé avec des scénarios différents fournis par le modèle stochastique. Ces données proviennent de campagnes de recensement des activités, d'études sociologiques, de mesures effectuées sur un nombre important de bâtiment. Haldi et Robinson (2011b) proposent ainsi un modèle stochastique d'ouverture et fermeture de fenêtres et de gestion des protections solaires permettant d'étudier l'influence sur les consommations. Construit à partir de données provenant d'enquêtes de terrain réalisées sur 8 années, leur modèle impacte les besoins de leur bâtiment test

et les fait varier de plus ou moins 50 %. Afin d'étudier la relation entre les ouvertures de fenêtres et différents paramètres, Herkel et al. (2008) réalisent des mesures sur un bâtiment de bureaux à Fribourg (Allemagne). Un modèle stochastique liant le statut d'ouverture des fenêtres au moment de l'année, à la température extérieure et au statut d'occupation est proposé. Il constate un fort lien entre l'arrivée des personnes dans les pièces et l'ouverture des fenêtres.

Dans ses travaux de thèse, Vorger (2014) développe des modèles stochastiques de comportement des occupants pour des logements et des bureaux. Pour définir les probabilités conditionnelles liées à ses modèles, différentes compilations de données sont utilisées. Elles proviennent de l'enquête emploi du temps de l'INSEE, d'études sur le taux d'équipement provenant de cet organisme de statistiques mais aussi de campagnes de mesures réalisées par le bureau d'études Enertech et de résultats provenant de travaux académiques et industriels (EDF, ADEME…). Ce modèle est plus longuement décrit au § 1.6.1.3.

L'évaluation de l'impact des modes d'occupation sur les consommations énergétiques, qu'elle soit réalisée grâce aux modèles complexes présentés ci-dessus ou par la simple variation de paramètres statiques, a été étudié dans la littérature. Wang et al. (2012) s'intéresse à l'impact de variations des scénarios d'éclairage, de température de consigne, de fonctionnement des systèmes de ventilation, de consommation des équipements sur les consommations d'énergie du bâtiment. En prenant en compte un nombre fini de scénarios correspondant à plusieurs jeux de données d'entrée, il étudie l'influence relative de ces paramètres et identifie en particulier l'importance de l'éclairage, des consommations des équipements et de la température de consigne sur le niveau de consommations. En cumulant les variations possibles, des écarts de -30 % à +80 % sont observés par rapport au scénario de référence. Les travaux de Hoyt et al. (2015) sur l'influence du niveau de température de consigne sur les consommations des appareils de chauffage et ceux de Sun et al. (2014) sur les taux de présence peuvent aussi être cités Ces études s'appuient également sur les résultats de simulation de logiciels STD. Wang et Augenbroe (2015) proposent de modéliser le taux d'occupation et le scénario d'éclairage en générant des séries temporelles à partir de l'identification des caractéristiques du scénario de base. Une décomposition en série de Fourier est réalisée et les résidus sont considérés comme des processus autorégressifs et à moyenne mobile (ARMA). En utilisant les modèles de génération stochastique d'occupants, Vorger et al. (2014b) étudie la répartition probable des besoins de chauffage d'un bâtiment pour deux températures de consigne dont la valeur est fixée. On constate que même avec une température de consigne connue, la variabilité du comportement crée une dispersion des besoins. Par ailleurs, on constate l'impact de la température de consigne sur les besoins énergétiques (Figure 1-9).

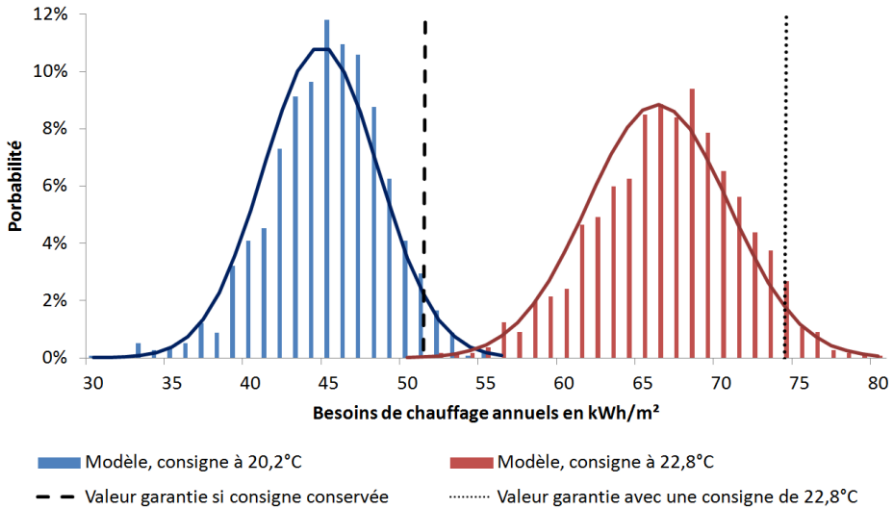

Figure 1-9 : Distributions des besoins de chauffage obtenues avec les modèles d'occupation aléatoires pour deux températures de consigne fixées (Vorger et al., 2014a)

1.2.3. Abandonner le déterminisme : l'intégration des incertitudes et variabilités

L'intégration de la modélisation du caractère incertain et aléatoire des nombreuses entrées des modèles de SED dans un modèle plus global a vocation à mieux décrire et anticiper le fonctionnement énergétique des bâtiments. Le système physico-probabiliste considéré résulte donc de l'agrandissement des frontières du système étudié. Cette nouvelle approche repose ainsi à la fois sur les lois physiques associées à la thermique et à l'énergétique du bâtiment mais aussi sur des méthodes statistiques et probabilistes. Les sorties du modèle physico-probabiliste, besoins ou consommations énergétiques, sont elles aussi exprimées en terme probabiliste, ce qui modifie la façon d'utiliser ces outils mais rend possibles de nouvelles applications (analyse de sensibilité globale, analyse d'incertitude, calibrage bayésien du modèle de bâtiment…). On demande ainsi à l'utilisateur de l'outil de simulation de ne plus s'attendre à obtenir un résultat unique et répétable. Ce nouveau paradigme n'est pas adapté à la résolution de toutes les problématiques et il est nécessaire d'avoir conscience de ses limites de validité. Celles-ci sont notamment liées à la pertinence et à la validité des modèles probabilistes en amont de la SED. Ce sujet est plus largement discuté dans la thèse de Rastogi (2016a). La Figure 1-10 compare schématiquement l'approche classique et l'approche physico-probabiliste.

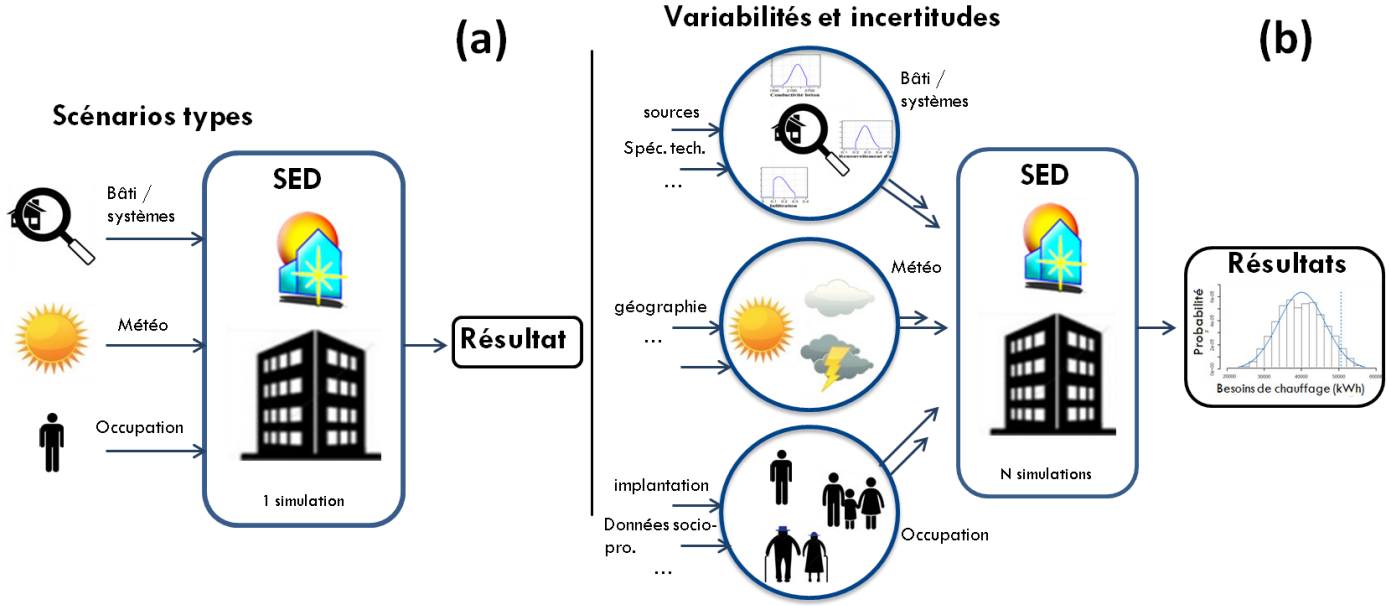

Figure 1-10 : Approches classique (a) et physico-probabiliste (b) pour la prévision par simulation des consommations d'énergie des bâtiments

Ce changement dans les approches de simulation est en partie porté par l'apparition d'une exigence de résultats, allant au-delà de la simple exigence de moyens, sur la performance énergétique des bâtiments. Le développement de méthodologies et de bonnes pratiques liées à la mesure accompagnent elles aussi logiquement cette évolution.

1.3. Mesurer pour vérifier la performance

Au-delà de l'objectif de vérification, de la validation de performances prédites, mesurer les conditions d'usage d'un bâtiment et ses consommations énergétiques permet de répondre à des objectifs variés. Par exemple, le relevé des consommations d'électricité ou de gaz permet simplement de facturer leur utilisation aux consommateurs. D'autres applications nécessitent aujourd'hui de connaitre l'état énergétique des bâtiments, pour le pilotage intelligent des systèmes énergétiques par exemple. Par ailleurs, de nombreuses applications de communication des conditions et des consommations vers les occupants sont aussi en développement autour de la GTB (gestion technique du bâtiment) qui était jusqu'à maintenant principalement utilisée sur les sites tertiaires ou industriels. La combinaison de plusieurs objectifs permet d'exploiter au mieux et de rentabiliser les dispositifs de mesures.

Dans un contexte de faible coût de l'énergie, les consommations énergétiques du bâtiment ont longtemps été une préoccupation mineure, puis les économies d'énergie sont devenues un domaine d'intérêt et la conception des bâtiments a peu à peu intégré la nécessité de construire des bâtiments

limitant les besoins énergétiques. Depuis une dizaine d'années, la nécessité de vérifier les performances attendues s'impose peu à peu. Jens Laustsen, membre de l'agence internationale de l'énergie (AIE) s'étonne du « manque d'intérêt pour les consommations réelles des bâtiments » (Laustsen, 2016) des acteurs du secteur. L'annexe 71 de l'AIE intitulée « *Building Energy Performance Assessment Based on In-situ Measurements* », poursuivant les développements réalisés au sein de l'annexe 58, a pour mission de travailler aux méthodes reposant sur la mesure en exploitation pour la prédiction des performances énergétiques des bâtiments.

La tendance observée ces dernières années consiste en un basculement d'exigences de moyens (sur l'efficacité énergétique des systèmes ou de l'enveloppe) vers des exigences de résultats. Ce changement de pratique est entériné dans la directive européenne EPBD 2002/91/EC, complétée et mise à jour en 2010 et 2013, et dont la vérification des performances attendues est l'objectif central.

1.3.1. Erreurs et incertitudes

1.3.1.1. Caractérisation des mesures

Avant de traiter du sujet de la mesure dans les bâtiments, il est utile de revenir sur un certain nombre de termes et notions associées. Le terme d'**erreur** est défini comme l'écart à la valeur vraie d'une grandeur dont une mesure est réalisée. Il suppose donc la connaissance de cette valeur. Pour la grande majorité des mesures, la valeur vraie est inconnue et il n'y a donc pas de valeur de référence. On parlera alors d'incertitude. Comme le rappelle Goffart (2013a), dans le cas d'une mesure, on peut décomposer l'incertitude totale, en une **incertitude systématique**, aussi appelée **biais**, et une **incertitude aléatoire**. La valeur du biais est liée à la méconnaissance et aux limites de la modélisation du système. Pour une mesure, un biais peut être introduit par un défaut du capteur ou un mauvais calibrage de celui-ci. Par définition, un résultat exact ne présente pas de biais.

Plus l'incertitude systématique est faible plus la mesure est dite juste. La **justesse** caractérise ainsi l'erreur moyenne entre la mesure et la valeur vraie. La **fidélité** de la mesure est liée à la faible dispersion de mesures répétées de la même grandeur. Plus la précision d'un capteur est grande, plus il est fidèle et donc plus la dispersion des mesures est réduite. On cherche donc à obtenir des résultats à la fois exacts (justes) et avec une précision connue et la dispersion la plus faible possible (fidèle). La Figure 1-11 récapitule ces notions.

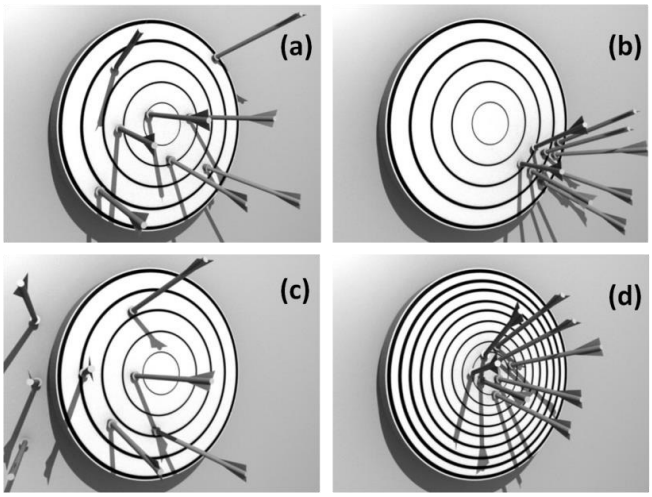

Figure 1-11 : Illustration des notions d'exactitude et de précision d'après (Rouaud, 2012). Mesure juste mais peu fidèle (a), mesure fidèle mais biaisée(b), mesure biaisée et peu fidèle(c), mesure juste et fidèle (d)

1.3.1.2. Méthode d'estimation des incertitudes

L'incertitude associée à une mesure dépend de différents facteurs. Les instruments de mesure utilisés et leur fiabilité ont un impact important mais celle-ci dépend aussi de l'éventuelle répétition des mesures effectuées. La méthode GUM (*Guide to the expression of Uncertainty in Measurement*) proposée par le JCGM (*Joint Committee for guides in Metrology*) (JCGM 1995) est la méthode de référence pour le calcul des incertitudes en métrologie. Deux cas de figure y sont considérés.

Incertitude de type A

Cette méthode est utilisée lorsqu'un échantillon de mesures répétées dans des conditions identiques de la même grandeur est disponible. On s'intéresse à la grandeur X dont n mesures $\{x_1, x_2, \ldots, x_n\}$ sont réalisées. Une approche statistique est considérée pour caractériser le résultat de ces mesures. La moyenne des mesures est considérée comme estimateur de la valeur vraie de la grandeur.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \textbf{(1.1)}$$

L'incertitude associée u_a est définie comme l'écart type expérimental $\sigma_{\bar{x}}$ de la moyenne :

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{\mathrm{N}(\mathrm{N}-1)}\sum_{\mathrm{i}=1}^{\mathrm{n}}(x_i - \bar{x})^2} \qquad \textbf{(1.2)}$$

On parle aussi d'incertitude de répétabilité.

Incertitude de type B

Cette méthode est utilisée lorsque l'approche statistique n'est pas permise du fait de l'absence de mesures répétées. L'évaluation de l'incertitude repose donc sur différentes sources :

- la connaissance générale ou empirique du comportement des instruments utilisés ;
- les spécifications du fabricant ;
- les certificats d'étalonnage ;
- l'incertitude attribuée à des valeurs de référence provenant d'ouvrages, manuels et autres normes.

L'incertitude u_b est construite comme une incertitude composée des différentes composantes $\mathrm{u_j}$ d'incertitudes disponibles :

$$u_b = \sqrt{\sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{u_j}^2} \qquad \textbf{(1.3)}$$

Ainsi pour l'estimation x d'une grandeur X, l'incertitude-type $u(x)$ est évaluée à partir de ces différentes sources. D'une manière générale, si le constructeur fournit l'incertitude-type, on l'utilise directement. Si l'on a très peu d'informations sur une grandeur d'entrée et que l'intervalle de variation supposé de celle-ci est de la forme :

- $\Delta x = \pm a$, alors $u_b(x) = \frac{a}{\sqrt{3}}$;
- $\Delta x = q$, alors $u_b(x) = \frac{q}{2\sqrt{3}}$.

La méthode GUM repose sur l'étude préalable et la caractérisation des causes d'incertitude. L'approche statistique est à préférer lorsqu'un échantillon de mesures est disponible. Les composantes de type A et B peuvent être combinées lorsque les deux types d'informations sont accessibles. On parle d'incertitude composée.

$$u = \sqrt{{u_a}^2 + {u_b}^2} \qquad \textbf{(1.4)}$$

1.3.2. Protocoles de mesure et vérification

On constate aux niveaux national et international une normalisation des méthodologies de mesure et vérification de la performance énergétique des bâtiments. Celles-ci ont pour objectif d'encadrer les pratiques et de mettre en place des cadres communs afin d'assurer des bases d'estimations fiables des performances énergétiques des bâtiments et des économies d'énergie éventuellement réalisées suite à des travaux d'amélioration de la performance énergétique (APE).

La *guideline* 14 publiée par l'ASHRAE (*American Society of Heating, Refrigerating and Air Conditioning Engineers*), a pour but de fournir des standards de procédures de calcul des économies d'énergie. Culp et al. (2005) ont analysé ce document, intitulé *Measurement of energy and demand savings*, datant de 2002, et rédigé par un comité d'experts venant de l'industrie et de la recherche. Y sont notamment spécifiées les approches possibles parmi lesquelles trois principales sont décrites :

- la détermination des performances d'un bâtiment entier après rénovation lorsque des mesures avant et après rénovation sont disponibles ;
- l'estimation des performances d'un équipement ou d'un sous-système dont le fonctionnement peut être mesuré indépendamment du reste du bâtiment ;
- l'étude des performances d'un bâtiment dont la situation de référence est fournie par un modèle de simulation calibré.

Des éléments techniques sur les appareils de mesure, leur calibrage, leur incertitude, leur maintenance sont présentés ainsi que les bases théoriques des méthodes de mesure. Des conseils sur le processus de vérification de la performance sont donnés et une méthodologie est présentée, intégrant la projection d'une situation initiale dans la situation mesurée après travaux.

Développé au niveau international par l'*Efficiency Valuation Organization* (EVO) et reconnu en France par l'ADEME, le protocole IPMVP (*International performance measurement and verification protocol*) est la référence actuelle sur les bonnes pratiques à mettre en œuvre pour évaluer et vérifier les performances énergétiques des bâtiments. Il est en partie inspiré du document de l'ASHRAE mais propose cependant une approche plus concrète et appliquée et est accompagné de différentes annexes techniques donnant des exemples de projets, les spécificités légales des différents pays, les techniques de calcul, etc. Les plans de mesure et vérification (M&V) doivent être précis, exhaustifs, conservateurs (ne pas surestimer les performances), cohérents, pertinents et transparents (EVO, 2012). L'ensemble du respect de ces bases « morales » est au cœur des méthodes de l'IPMVP et explique en partie sa reconnaissance internationale.

Le protocole IPMVP repose sur plusieurs étapes :

- la description précise des actions d'amélioration de la performance énergétique (APE) et le détail du résultat attendu ;
- l'identification et la justification du choix d'une option méthodologique du protocole ;
- la description de la situation de référence et la collecte de données sur celle-ci ;
- l'identification de la période de suivi et la définition des conditions d'ajustement ;
- la spécification de la procédure d'analyse et de la valorisation financière des économies ;
- la spécification du plan de mesures, des responsables du suivi de celles-ci, et de la précision attendue ;
- la définition du budget, des ressources et des rapports attendus.

Nous reviendrons par la suite sur les termes de situation de référence et d'ajustement (paragraphe 1.4.2). Ces bonnes pratiques constituent un cadre méthodologique. Le contenu de chaque étape n'est pas imposé mais les utilisateurs se doivent de respecter ces différentes phases. Pour chacune des étapes du protocole, des documents de référence et des conseils méthodologiques

sont présentés. Il pourra être utile de se rapporter à (EVO, 2012) pour plus de détails. On peut notamment citer une des recommandations du document conseillant d'effectuer des rénovations si elles entrainent des gains supérieurs à 10 % du fait des incertitudes sur les modèles de prévision des consommations post-rénovation.

Le protocole IPMVP propose quatre options méthodologiques. Les options A et B doivent être mises en œuvre lorsque les gains énergétiques spécifiquement dus à l'APE peuvent être mesurés et isolés de la consommation de l'ensemble du bâtiment. L'option A impose la mesure des paramètres principaux et autorise l'estimation d'une partie des paramètres, notamment ceux concernant l'usage. L'option B impose que tous les paramètres soient mesurés et exclut toute estimation.

L'option C s'intéresse aux performances globales du bâtiment avec une rénovation importante. Elle peut s'appuyer sur les éléments des factures des fournisseurs d'énergie. Cette option est la plus couramment utilisée.

L'option D s'appuie sur l'utilisation de logiciels de simulation. Elle est utilisée lorsqu'aucune donnée n'est disponible sur la situation de référence (bâtiments neufs, rénovations lourdes…). Elle impose le calibrage du modèle par des données mesurées.

Dans ces travaux de thèse, liés aux options C et D, on se référera régulièrement à ce protocole afin de se positionner relativement à cette classification des méthodes et des pratiques existantes. Un des mécanismes centraux associés aux protocoles de vérification va maintenant être présenté. Il s'appuie sur l'idée que la performance énergétique est associée à un contexte d'utilisation du bâtiment, imposant ainsi l'obligation de définir ce contexte et éventuellement de le mesurer.

1.3.3. État de l'art de la mesure dans le bâtiment

Différents travaux de recherche traitent de la mesure dans le secteur du bâtiment. Dans une revue sur le sujet Ahmad et al. (2016) s'intéressent à la fois aux aspects sociaux, économiques, législatifs et techniques. Une classification des capteurs est présentée ainsi que les différents protocoles d'acquisition de données. Les principales problématiques liées à la mesure sont abordées : la résolution spatiale et temporelle, la précision, l'exploitation des données, les coûts, etc.

La question du pré-traitement des données et en particulier de la gestion des données manquantes est étudiée par Titikpina et al. (2015b). Les données de température de quatre bureaux sont en partie reconstruites de manière à respecter l'autocorrélation temporelle et l'intercorrélation observées entre les différentes séries.

L'homogénéité des températures au sein d'une même pièce est étudiée dans différents travaux, que ce soit avec une approche numérique (Ham et Golparvar-Fard, 2013) ou expérimentale (Riederer et al., 2001). Des recommandations peuvent être tirées en ce qui concerne la mesure d'une température de pièce. Afin d'être représentative, celle-ci doit être réalisée à une hauteur d'un mètre

cinquante environ, au centre de la pièce ou au moins à distance des appareils de distribution de chauffage ou de ventilation, le capteur ne devant pas être soumis au rayonnement solaire. Cette problématique rejoint celle du maillage spatial des mesures de température à l'échelle d'un bâtiment en lien avec le zonage des modèles en simulation énergétique dynamique. Rivalin et al. (2014) étudient l'influence de ce découpage en zones sur un bâtiment de logements, observant de faibles différences sur les besoins énergétiques dans les différentes configurations étudiées. Cependant, la diversité possible des usages amène à considérer cette question de manière spécifique pour chaque projet étudié. Stefanoiu et al. (2015) caractérisent la corrélation entre les températures de différentes zones d'un bâtiment dans l'objectif de proposer une répartition optimisée des capteurs tout en limitant leur nombre. Dans un rapport du projet OMEGA, Caucheteux (2015) vise un but similaire et détaille les modes de calcul de l'incertitude sur la température ambiante moyenne en fonction des capteurs installés et de leur position. Il souligne l'intérêt d'avoir une répartition des capteurs dans un bâtiment, mais aussi au sein d'une même pièce, pour minimiser l'incertitude.

Jin et al. (2016) s'intéressent aux mesures d'énergie des différents postes de consommations dans les bâtiments. Les bonnes pratiques et le matériel à utiliser sont décrits. L'exploitation de ces données permet notamment de caractériser les usages. L'ensemble du plan de mesure mis en œuvre pour calibrer un modèle d'un bâtiment historique est détaillé dans les travaux de Cornaro et al. (2016). La description de ce projet met en avant les problématiques rencontrées liées à la mesure (températures, énergies…) et les solutions pour y faire face.

1.3.4. Mesurer et inférer l'occupation et les usages

La mesure de l'occupation et des usages dans les bâtiments est un domaine très dynamique à l'heure actuelle qui donne lieu à de nombreux travaux de recherche universitaire et industrielle.

La détermination de la présence des usagers est un des sujets clés donnant lieu à de nombreuses contributions. Ceci s'explique par le grand nombre d'applications découlant de cette connaissance. On peut bien sur citer la gestion et le pilotage énergétique du bâtiment mais aussi l'apport de services variés. Grâce au développement de l'*IoT (Internet of Things*, ou objets connectés) des données provenant de différentes sources peuvent être exploitées et croisées, pour déduire la présence et le nombre de personnes présentes. Des informations peuvent provenir des consommations électriques (Wang et al., 2017), de capteurs de $C0_2$, de température ou d'humidité, de détecteurs de mouvements, des contacteurs d'ouvertures de portes ou de fenêtres, de détections de signaux électromagnétiques, des microphones… Labeodan et al. (2015) proposent notamment une revue des pratiques de mesure de l'occupation dans les bâtiment de bureau. Amayri et al. (2016) étudient l'estimation de l'occupation à partir d'un ensemble de capteurs variés grâce à différents algorithmes de *machine learning* (algorithme d'apprentissage). L'objectif est d'identifié les capteurs

les plus utiles au comptage des occupants. Les détecteurs de mouvement (infrarouge), les capteurs de C02 et les microphones sont identifiées comme les plus intéressants.

La déduction de la présence et des usages par exploitation de données mesurées est rendue possible par l'utilisation d'algorithme d'apprentissage automatique. Ils ont vocation à extraire de l'information à partir d'observations brutes et se divisent en deux catégories : l'apprentissage supervisé et l'apprentissage non-supervisé. Dans le cas supervisé, des données d'entrée et de sortie sont fournies au modèle et l'objectif est que l'algorithme calcule des sorties associées à de nouvelles entrées. Lorsque les sorties sont discrètes, on parle de méthodes classification, et de méthodes régression lorsqu'elles sont réelles. La supervisation suppose d'avoir des données labellisées à fournir à l'algorithme. En non-supervisé, des jeux de données sont fournies aux algorithmes sans spécifier les entrées ou sorties et celui-ci s'applique alors à trouver des logiques en leur sein.

Parmi les travaux liés à la mesure dans les logements, Kazmi et Ploix (2018) cherchent à dé terminer l'occupation et les activités à partir de données sur les consommations d'électricité, d'eau chaude et de taux de $C0_2$ grâce à un réseau bayesien (méthode non supervisée). Un autre modèle non supervisé (Markov caché) est exploité par Candanedo et al. (2017) dans le but de détecter la présence dans un bâtiment résidentiel à partir de mesures de températures, humidité $C0_2$ et luminosité. Ces deux types de modèles sont aussi étudiés dans (Amayri et al., 2017).

Concernant les usages tertiaires, on peut citer les travaux de Candanedo et Feldheim (2016) où à partir du même type de mesures, différentes méthodes de classification sont comparées et leur capacité de prédiction de présence sont comparées. D'Oca et Hong (2015) développent une méthode d'apprentissage pour identifier des scénarios d'occupation et d'usage d'un bâtiment comprenant 16 bureaux. Gilani et O'Brien (2017) proposent une revue des méthodes de mesure liées à l'occupation dans les bâtiments de bureaux.

Ainsi, de nombreuses études traitent de ce sujet, encouragées par l'évolution du matériel disponible, la baisse de leur prix et donc la possibilité d'accès à des données de plus en plus nombreuses et riches. Dans la suite de ces travaux, la problématique de la mesure des usages est associée aux besoins de caractériser l'influence des occupants sur les consommations énergétiques, et d'isoler cette influence.

1.3.5. Le calibrage de modèle

Le calibrage de modèle a pour objectif d'affiner la modélisation d'un ouvrage à partir de l'exploitation de données mesurées caractérisant son fonctionnement énergétique. Ce processus permet de caler les valeurs des paramètres physiques et techniques du modèle de bâtiment de manière à ce que les résultats de simulation soient le plus proche possible du comportement observé. Plus les données mesurées sont riches (nombre de grandeurs mesurées, variations

dynamiques…) plus il est possible de calibrer un modèle avec précision et plus le nombre de paramètres pouvant être recalés est élevé (Robillart, 2015).

1.3.5.1. Indicateurs de proximité

La quantification de la proximité entre des données simulées et des données mesurées passe par le calcul de différents indicateurs, dits de *Goodness-of-fit* (GOF). L'erreur moyenne (MBE –*Mean Biased Error*) est définie comme la moyenne des écarts sur la période considérée. On s'intéresse à n points de données. $\widehat{y_t}$ désigne la valeur calculée par le modèle et y_t la valeur mesurée correspondante :

$$\mathbf{MBE} = \frac{\sum_{t=1}^{n}(\widehat{y_t} - y_t)}{n} \tag{1.5}$$

L'erreur moyenne absolue (MAE –*Mean Absolute Error*) peut aussi être calculée :

$$\mathbf{MAE} = \frac{\sum_{t=1}^{n}|\widehat{y_t} - y_t|}{n} \tag{1.6}$$

Un autre indicateur couramment utilisé est l'erreur quadratique moyenne (RMSE – *Root Mean Square Error*) qui peut aussi être rapportée à la valeur moyenne des données réelles :

$$\mathbf{RMSE} = \sqrt{\mathbf{MSE}} = \sqrt{\frac{\sum_{t=1}^{n}(\widehat{y_t} - y_t)^2}{n}} \tag{1.7}$$

$$\mathbf{cv(RMSE)} = \frac{\sqrt{\frac{\sum_{t=1}^{n}(\widehat{y_t} - y_t)^2}{n}}}{\frac{\sum_{t=1}^{n} y_t}{n}} \tag{1.8}$$

Ces deux derniers indicateurs pénalisent plus lourdement les écarts importants. Les éléments présentés ci-dessus sont ceux étant le plus couramment utilisés mais il existe d'autres indicateurs caractérisant la précision d'un modèle.

1.3.5.2. Méthodes de calibrage

Coakley et al. (2014a) d'une part, et Fabrizio et Monetti (2015) d'autre part, proposent deux revues des méthodes de calibrage de modèles énergétiques de bâtiments. On peut notamment distinguer les méthodes graphiques, l'approche bayésienne, l'utilisation de métamodèles, l'étude de la signature ou l'approche par optimisation. Différents travaux en lien avec le calibrage peuvent être mis en avant (Coakley et al., 2012 ; Harmer et Henze, 2015 ; Nolan et al., 2014 ; Park et al., 2013), appliquant ces différentes approches à une variété de types de bâtiments mais aussi de données et

d'horizons temporels de calibrage. Le calibrage bayésien s'appuie sur la description initiale des paramètres à caler comme des lois de probabilité *a priori*. Des tirages au sort sont effectués dans ces lois, des simulations sont réalisées et leurs résultats sont comparés aux données mesurées. La sélection de valeurs de paramètres satisfaisant à un critère de proximité permet d'affiner, par une mise à jour itérative, les lois de distributions initiales. Cette approche est utilisée dans les travaux de (Heo, Choudhary, et Augenbroe (2012) ou de Robillart (2015). Différents algorithmes et critères de sélection peuvent être utilisés. Chong (2017) propose une comparaison d'algorithmes MCMC (*Monte-Carlo Markov Chain*). Plus de détails sur le calibrage peuvent être trouvés dans les travaux de thèse de Robillart (2015).

Dans une approche de garantie de performance, le calibrage permet de caractériser le plus précisément possible l'état d'un bâtiment, avant une rénovation par exemple. Les prévisions réalisées grâce au modèle après travaux s'appuient sur ce modèle enrichi de la connaissance apportée par des mesures. Dans certains cas, l'instrumentation installée pour caractériser le bâtiment existant est aussi destinée à effectuer des relevés une fois celui-ci réhabilité dans l'objectif de recalibrer des modèles de manière dynamique (Raillon and Ghiaus, 2017) ou de vérifier la performance attendue.

1.3.6. Le constat de l'écart entre prévisions et réalité

L'intérêt croissant pour la concrétisation effective des objectifs de performance et donc pour la vérification des consommations réelles des bâtiments a donné lieu à un nombre important de projets tests, liés à des travaux de recherche mais aussi en vue d'industrialiser ces pratiques. Cette thématique traitant du « *performance gap* », c'est-à-dire de l'écart constaté entre les consommations énergétiques réelles et les estimations a priori s'intéresse à la caractérisation de cet écart, à la recherche des causes et au développement de solutions pour sa réduction, qu'elles soient méthodologiques ou opérationnelles.

Demanuele et al. (2010) étudient l'écart entre prévisions et mesures des consommations énergétiques d'écoles au Royaume-Uni, adoptant une approche probabiliste de modélisation. Ils mettent en avant la responsabilité du comportement des occupants et du pilotage des systèmes énergétiques dans les écarts observés. Burman et al. (2014) proposent une méthode de mesure et vérification pour caractériser les performances du bâtiment. Ils décomposent le « *performance gap* » en deux parties distinctes, elles mêmes liées à différentes causes. Ils distinguent le *procurement gap*, lié à des problèmes de mise en œuvre et l'*operational gap* associé à la conduite des systèmes énergétiques, à l'usage et aux sollicitations. Un plan de commissionnement et de mesure est prescrit suite à ce constat. Kampelis et al. (2017) s'intéressent spécifiquement aux bâtiments à très haute performance énergétique, résidentiels et tertiaires. Une comparaison est réalisée entre les résultats de mesure et de simulation. Un suivi et un accompagnement très performants sur les aspects

énergétiques dans les bâtiments étudiés permet d'obtenir des écarts assez faibles. Une revue des retours d'expérience dans la littérature sur les bâtiments non résidentiels (Royaume-Uni) est proposée par van Dronkelaar et al. (2016). Sur 62 bâtiments considérés, un écart moyen de 34 % est constaté, avec un écart-type de 55 %. L'influence de l'usage est de nouveau mise en avant. (N. Li et al. 2015) posent aussi la question des limites de la simulation pour la prévision des consommations énergétiques futures. Ils soulignent à la fois la responsabilité de la mauvaise modélisation des bâtiments et de leurs usages.

Différents projets tentent de référencer des retours d'expérience. Mise en place par le RIBA (*Royal Institute of British Architects*) et le CIBSE (*Chartered Institution of Building Services Engineers*), la plateforme CarbonBuzz[2] a pour objectif de recenser les retours d'expérience sur les consommations réelles des bâtiments. Ils concluent à des surconsommations de l'ordre de 50 % à 150 %. Les études du *Post Occupancy Review of Building Engineering* (PROBE)[3] et de CBx ont un objectif similaire. En France, les travaux effectués par Enertech[4] depuis plusieurs dizaines d'années ont aussi pour but d'étudier les écarts constatés entre les performances énergétiques attendues et effectives. Le programme PREBAT mis en place par l'ADEME et le CSTB vise l'acquisition d'expérience sur le suivi des consommations énergétiques dans les bâtiments à haute performance dans le but de faire émerger des pratiques de référence.

La caractérisation des écarts de performance observés et l'étude des causes associées ont pour objectif le développement de méthodes pour limiter ces différences. Ceci passe par des simulations plus réalistes et représentatives des conditions opérationnelles futures, par un suivi et un accompagnement de la réalisation et par une maîtrise accrue du fonctionnement du bâtiment en phase de conduite. L'atteinte de ces objectifs peut notamment permettre aux acteurs du secteur du bâtiment une meilleure maîtrise des risques et la mise en place de contrats de performance énergétique ; on parle de garantie de performance énergétique (GPE). Les mécanismes et objectifs liés à la GPE vont maintenant être présentés.

[2] http://www.carbonbuzz.org/

[3] http://www.cibse.org/knowledge/building-services-case-studies/probe-post-occupancy-studies

[4] http://www.enertech.fr/

1.4. La garantie de performance énergétique

1.4.1. Enjeux de la garantie de performance

1.4.1.1. Différentes approches complémentaires

La garantie de performance énergétique (GPE) est une démarche reposant sur l'engagement contractuel d'une structure réalisant des travaux de construction ou de rénovation liés à la sobriété énergétique d'un bâtiment. Une valeur limite de consommation énergétique réelle, ou de diminution de la consommation, est ainsi garantie, au-delà de laquelle un mécanisme de pénalités s'appliquant à la société portant l'engagement est mis en place. La GPE permet ainsi de transférer une partie du risque, associé à d'éventuelles surconsommations, du client contractant la GPE vers la structure proposant l'engagement. Ainsi, ce processus permet d'anticiper les gains énergétiques et donc financiers de travaux d'amélioration de la performance (APE) ce qui est un atout financier majeur. La méthodologie constitue donc un levier de financement de travaux prépondérant, assurant la sécurité et la maitrise du retour sur investissement. La GPE a donc vocation à multiplier les projets de réhabilitation ou de construction à haute performance énergétique. Il existe différentes formes de GPE, terme global, dont les distinctions sont présentées dans (ADEME et FBE 2016).

La garantie de performance énergétique intrinsèque (GPEI) se caractérise par l'évaluation des performances énergétiques d'un bâtiment immédiatement après réalisation des travaux et avant réception selon un protocole court et normalisé. L'absence d'usagers et la maitrise des sollicitations du bâti permettent de calculer, par exploitation des mesures réalisées, les caractéristiques thermiques principales du bâtiment (p. ex. coefficient de déperdition thermique, constantes de temps principales) et de les comparer aux valeurs attendues et garanties. La littérature scientifique et industrielle propose de nombreux travaux en lien avec la GPEI. Ceux-ci reposent en partie sur l'identification de modèles simplifiés de bâtiments de type RC (résistance-capacité) à partir de mesures réalisées *in situ*. La méthode *co-heating* (ou *co-heating test*) prévoit la mise en place d'un protocole expérimental plus long et étoffé, basé sur la sollicitation du bâtiment avec des sources de chaleur et l'acquisition des températures intérieures et extérieures (Sonderegger et Modera, 1979). L'analyse de ces données permet de caractériser la performance intrinsèque du bâtiment. Bauwens et Roels (2014) proposent une revue des travaux basés sur cette approche. Thebault (2017) développe ainsi une méthode d'identification du coefficient de transmission thermique. Les méthodes ISABELE (CSTB) ou QUB (Saint Gobin) ont elles aussi le même objectif.

La garantie de résultats énergétiques (GRE) porte sur les consommations énergétiques réelles du bâtiment en exploitation. Les usages et sollicitations réels impactent donc ces consommations tout comme la conduite effective des équipements et systèmes énergétiques et éventuellement le vieillissement des infrastructures. Différents types de GRE peuvent être distingués, notamment en

fonction des périodes sur lesquelles porte l'engagement et sur les phases du cycle de vie considérées (ADEME et FBE, 2016).

La GRE initiale (GREi) propose un engagement sur une courte durée (de l'ordre de 2 à 3 ans). Elle vise à vérifier la performance effective suite aux travaux et la stabilisation rapide de la qualité de la maintenance et de la conduite du bâtiment.

La GRE exploitation (GREe) porte sur une durée plus longue et vise le maintien ou l'amélioration de la performance stabilisée initialement. Elle s'inscrit dans la continuité de la GREi.

La GRE optimisation (GREo) se caractérise par un réengagement dans un objectif d'amélioration, de remise à niveau des équipements et des systèmes de régulation. Elle peut ainsi être utilisée quand les performances attendues ne sont pas, ou plus, à la hauteur.

Ces différents types de contrats s'enchainent logiquement sur les phases du cycle de vie du bâtiment et peuvent être combinés ou fusionnés. Les travaux présentés dans ce document portent sur l'approche de garantie de résultats énergétiques (GRE), sur les phases initiale et d'exploitation, et c'est dans le cadre de cette approche que doit être envisagée la suite des éléments présentés dans ce manuscrit.

1.4.1.2. Acteurs, projet et organisation

Les contrats de GPE sont proposés à différents types d'acteurs, qui en sont donc les clients. Ces acteurs, que nous appellerons donneurs d'ordre, peuvent être des maitres d'ouvrage (MOA), des propriétaires ou des gestionnaires de bâtiments. La structure contractante peut être un groupement d'entreprises, une société de service énergétique (SSE) unique, ou une maitrise d'œuvre. Différents schémas de contractualisation de GPE sont envisageables. Ils se différencient par la répartition des différentes phases du processus amenant à l'engagement entre le maître d'ouvrage et le porteur de l'engagement mais aussi par la structure proposant la garantie et par les moyens mis en œuvre pour la formaliser. Avant de détailler ces éléments, il est utile de revenir sur les étapes constituant un projet de construction ou de rénovation.

- *Faisabilité du projet et programmation*

Au départ d'un projet de construction, la faisabilité générale de celui-ci est évaluée (Etudes de faisabilité) par le maitre d'ouvrage en collaboration avec un architecte. Une étude d'esquisse (ESQ), ou étude de diagnostic (DIA) en cas de travaux sur un bâtiment existant, est réalisée permettant une première visualisation du projet et une estimation des coûts. Différentes options peuvent éventuellement être proposées à ce stade.

- *Etudes d'avant projet*

Afin de déposer une demande de permis de construire (PC), une description et un chiffrage plus précis des travaux doivent être fournis par les phases d'avant-projet sommaire (APS) puis détaillé (APD).

- *Conception détaillée du bâtiment*

Lors des études de projet (PRO), l'architecte prépare les plans détaillés du bâtiment, des bureaux d'études techniques sont sélectionnés et leur plan d'action est déterminé.

- *Planification de la construction*

Suite aux choix de conception définitifs, l'architecte constitue le dossier de consultation des entreprises (DCE) qui spécifie le détail des caractéristiques techniques et administratives et le budget de chaque lot du projet afin que des entreprises proposent des devis adaptés. Dans le prolongement du DCE, les études d'exécution (EXE) traitent de la réalisation technique du projet fournissant les plans d'exécution et les spécifications nécessaires au chantier.

- *Sélection des entreprises et réalisation du chantier*

L'architecte assiste le MOA pour la sélection des entreprises prestataires allant intervenir sur le chantier. Chaque lot réalise les travaux planifiés dans le programme en respectant les spécifications définies en amont. En complément de la maitrise d'œuvre, des tâches de coordination et de pilotage sont menées durant la mise en œuvre. Le bâtiment est finalement livré après un contrôle des travaux effectués par l'architecte.

- *Exploitation*

La conduite, le pilotage des systèmes énergétiques et la maintenance du bâtiment pendant son exploitation peuvent être sous-traités à un exploitant via un contrat spécifique.

Les contrats de GRE s'inscrivent dans ce cadre organisationnel. Quatre schémas de contractualisation sont proposés par l'ADEME et la fondation Bâtiment-Energie (FBE) dans un ouvrage de référence sur la garantie de résultats énergétiques (ADEME et FBE, 2016), tiré notamment des travaux d'un projet de recherche (FBE, 2014). Dans une première structure d'organisation, c'est la SSE qui s'engage en tout début de projet à la conclusion des études d'opportunité (faisabilité), supportant par la suite la conduite de l'ensemble du projet. Ce schéma a l'inconvénient d'imposer un engagement en amont du projet où très peu d'éléments concrets sont déjà arrêtés sur les travaux qui seront réalisés. Un autre type d'organisation consiste à partager les tâches associées aux spécificités de la GRE entre la MOA et la SSE, qui n'intervient que plus tard dans le projet et en plusieurs phases, proposant un engagement contractuel après avoir réalisé des études techniques plus approfondies. Dans une autre approche, la consultation d'entreprises pour une GRE peut même être lancée après réalisation d'études poussées par la maitrise d'œuvre. Enfin, la GRE peut être portée entièrement par la maitrise d'œuvre qui pilote l'ensemble du projet jusqu'à la

phase d'exécution. La différence principale entre ces approches réside dans l'avancement du projet lors de la contractualisation de l'engagement et donc dans la prise de risque et l'ambition de la GRE. Afin de minimiser les risques et de garantir au plus près de la performance attendue, il est intéressant de concrétiser l'engagement lorsque les plans et spécifications techniques finaux sont connus.

1.4.2. Éléments-clés du projet de GPE

1.4.2.1. Définition des objectifs

Différents objectifs peuvent être associés à la mise en place d'une GRE. Les cas de travaux de construction d'un ouvrage neuf et de réhabilitation d'un bâtiment existant doivent être différenciés. En construction neuve, l'engagement porte nécessairement sur une consommation énergétique maximale exprimée de manière absolue et déterminée à partir de simulations du bâtiment réalisées lors des études techniques. Dans le cas d'une réhabilitation, bien qu'il soit aussi possible de contractualiser un engagement sur les consommations absolues, l'objectif des travaux peut être associé à une diminution des consommations relativement à la situation avant travaux. Il est ainsi exprimé comme un pourcentage de réduction à atteindre.

Cette approche, communément répandue à l'heure actuelle (ADEME et FBE, 2016), impose de prévoir et de mesurer une diminution de consommations énergétiques, c'est-à-dire une non-consommation. Elle repose ainsi sur la caractérisation des performances initiales du bâtiment avant travaux, la situation de référence.

1.4.2.2. Étude énergétique

- *La situation de référence avant travaux*

Une des phases prépondérantes dans un projet de GRE en rénovation est la définition de la des consommations énergétiques de référence avant l'intervention. Celle-ci passe par une phase d'analyse du bâtiment existant, de ses factures énergétiques. Un audit poussé peut être conduit afin de déterminer les caractéristiques de l'enveloppe et des systèmes énergétiques, afin de réaliser un modèle de simulation. Lors de certains projets, une instrumentation est mise en place avant travaux dans l'objectif de caractériser la dynamique des consommations énergétiques et des conditions d'exploitation qui y sont associées.

- *Le calibrage du modèle de référence*

Ces données, qu'elles soient détaillées ou qu'elles proviennent uniquement de factures permettent de calibrer (ou caler) le modèle de référence. Une revue bibliographique des méthodes

de calibrage a été présentée au paragraphe 1.3.5. La détermination de la situation de référence est primordiale pour la normalisation d'un calcul de diminution de consommations d'énergie. Le modèle du bâtiment avant travaux correctement modélisé sert aussi de base à la modélisation de l'ouvrage après intervention. La pertinence du niveau d'engagement dépend donc fortement de la précision et du soin apportés à cette première étape.

- *Le bâtiment suite aux travaux*

Le choix d'une limite de consommation énergétique garantie actée dans le contrat est permis par l'usage des outils de simulation. Le bâtiment construit ou rénové est modélisé à l'aide d'un logiciel de SED. Dans le cas d'un bâtiment neuf, les données techniques de conception, les spécifications des matériaux et des systèmes choisis sont utilisées pour la modélisation de l'ouvrage. La majorité de ces données est accessible, le projet étant en cours de réalisation à un stade plus ou moins avancé. Ceci n'étant pas le cas pour une rénovation, les caractéristiques du bâtiment sont déterminées à partir des documents disponibles, de l'audit réalisé et du modèle calibré. Les actions d'amélioration de performance énergétique (APE) engagées dans la réhabilitation sont intégrées au modèle du bâtiment pour décrire l'ouvrage après travaux.

Dans l'éventualité où la garantie de performance porte sur des consommations absolues, le modèle du bâtiment après travaux permet d'évaluer les performances attendues du bâtiment pour un ensemble de sollicitations donné. Dans le cas où l'engagement porte sur une amélioration de la performance, les modèles avant et après travaux permettent de calculer les gains réalisés pour des sollicitations semblables. Une synthèse des méthodologies accompagnées de schémas illustrant notamment ce processus est proposée au paragraphe 1.4.2.6.

1.4.2.3. Analyse et intégration des risques

Un des objectifs principaux liés aux méthodes pour la GPE est d'être en mesure de quantifier le risque de défaut associé à une valeur garantie de consommations. Celui-ci est directement lié à la connaissance des risques rattachés aux projets de construction ou de rénovation mais aussi à l'exploitation. Ces derniers sont liés à de nombreuses problématiques : un diagnostic imparfait du bâti existant, une mise en œuvre non conforme aux plans de conception, une mauvaise exploitation et maintenance des systèmes énergétiques, des problèmes de pilotage et de régulation, etc.

Le commissionnement (ou *commissionning*) est un processus d'assurance qualité qui a pour but de vérifier que le bâtiment et ses systèmes énergétiques sont conçus, installés et testés pour répondre aux performances et à la qualité exigées. Ce processus permet d'acquérir une connaissance plus précise des caractéristiques de l'ouvrage, des systèmes et des modes d'exploitation du bâtiment. Cette tâche doit être confiée à une instance extérieure et indépendante qui n'est pas partie prenante de la conception et de l'exécution. Les méthodes liées au commissionnement s'intègrent tout au long du projet, dès la programmation. Le rétro-commissionnement s'intéresse à la

qualité des bâtiments existants et vise l'optimisation de leur fonctionnement. Thomare et al. (2017) proposent une méthodologie pour le commissionnement de systèmes CVC basée sur des mesures et des simulations. Une revue des méthodes pour le suivi des performances des bâtiments et systèmes est proposée par Li et al. (2014). L'approche probabiliste en simulation énergétique du bâtiment est une des solutions mises en avant pour intégrer ces risques (§ 1.2.3). L'approche stochastique est largement utilisée (Lee et al., 2013 ; FBE, 2014 ; Rivalin, 2016 ; Titikpina, 2016) dans l'objectif d'associer à des fonctions de densité de probabilité sur les consommations d'énergie, obtenues par simulation intensive, des probabilités d'occurrence et des risques de dépassement. Le quantile à *n* % d'une distribution de probabilité peut être associé à une valeur limite dont la probabilité de dépassement est *1*-n (Figure 1-12). Ce type d'étude permet d'intégrer une quantification du risque dans l'expression d'une performance garantie qu'elle soit absolue ou relative.

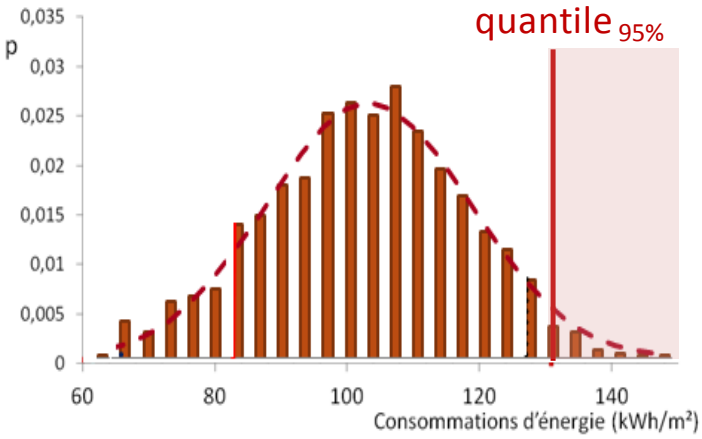

Figure 1-12 : Distribution de probabilité de consommations énergétiques et calcul du quantile à 95 % associé à une limite garantie à 5 % de risque de défaut

La proposition d'une offre de GPE ambitieuse mais réaliste dans un environnement concurrentiel doit s'appuyer sur la maitrise du risque de défaut de l'engagement. Les SSE, ou groupements d'entreprises, portant la GPE doivent être en mesure de maitriser les conséquences financières. Les pénalités en cas de dépassement et les rétributions liées aux sous-consommations, peuvent être des paramètres de modèles financiers et assurantiels permettant le choix de couples optimaux de valeur garantie et de risque.

1.4.2.4. Les mécanismes d'ajustement

- *Un fort impact d'éléments hors responsabilité*

Le lien de dépendance entre les consommations d'énergie et les sollicitations du bâtiment, liées aux usages ou aux conditions météorologiques notamment, rend difficile l'évaluation absolue de la performance énergétique. Par définition, la performance énergétique intègre cette problématique du contexte et est décrite dans l'EPBD 2002/91/CE comme « la quantité d'énergie consommée pour

répondre aux différents besoins liés à une utilisation standardisée du bâtiment ». La notion de conditions standardisées impose la mise en place d'une normalisation vers des conditions communes. Ceci se traduit par la spécification d'un mécanisme d'ajustement dans les contrats de GPE sur les résultats en exploitation. La variabilité des sollicitations météorologiques et des conditions d'usage induit une grande variabilité des consommations énergétiques. Celle-ci impacte la valeur des consommations énergétiques sur laquelle un engagement contractuel est établi et le risque d'excéder cette valeur. En effet, une limite garantie de consommations déterminée pour un risque raisonnablement faible (p.ex. 5 % de risque de dépassement) est élevée et donc peu intéressante à garantir. À l'inverse, une consommation garantie ambitieuse s'accompagne d'une prise de risque importante et est donc peu acceptable d'un point de vue assurantiel.

Ces éléments externes, climat et usages, ne sont aucunement liés à la qualité de la construction, de la rénovation ou de l'exploitation/maintenance du bâtiment sur lequel doit porter la garantie. Il est donc logique que leur impact soit hors de la responsabilité de la structure portant le contrat de garantie. La contextualisation des performances, et donc de l'engagement, s'avère donc nécessaire. Elle peut être prise en compte de manière différente si la performance attendue est absolue ou exprimée comme un pourcentage de réduction des consommations.

Dans cette dernière configuration, l'ajustement se fait par une comparaison des performances réelles après travaux, dans les conditions réelles mesurées, avec la consommation calculée du bâtiment avant travaux en considérant les conditions réelles après travaux. C'est donc la situation de référence qui est ajustée. Soit cet ajustement est réalisé par une simulation du modèle SED du bâtiment avec le détail des sollicitations relevées, soit ce modèle a permis en amont des travaux le calcul d'un modèle d'ajustement simplifié, un polynôme par exemple, associant une valeur des consommations énergétiques à la valeur de différents facteurs d'ajustement. La Figure 1-13 illustre le processus d'ajustement dans le cas d'une rénovation.

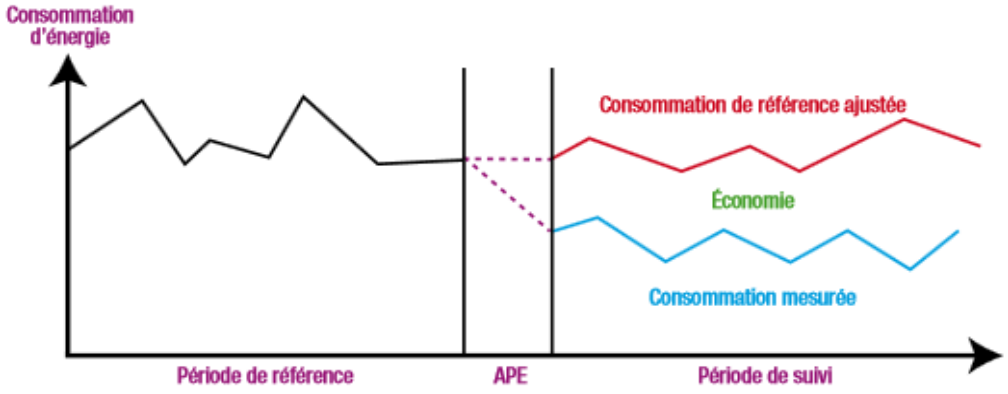

Figure 1-13 : Principe du calcul des économies d'énergie par comparaison à une base de référence suite à une action de performance énergétique (SYPIM 2012)

Dans le cas où l'on s'intéresse à la consommation énergétique absolue du bâtiment, l'ajustement se fait à l'aide du modèle post-travaux. Là aussi, on compare la consommation réelle

mesurée en exploitation associée aux sollicitations réelles, elles aussi mesurées, à une valeur théorique ajustée. De la même manière que précédemment, cette valeur théorique est soit obtenue par la simulation énergétique du modèle détaillé soit par un modèle d'ajustement simplifié spécifié dans le contrat.

La mise à jour du modèle SED avec les sollicitations mesurées et la simulation de ce modèle impose une intervention assez lourde et peu transparente de la SSE à chaque phase de vérification de l'engagement, à la fin de chaque période (souvent annuelle) de référence. L'utilisation d'un modèle d'ajustement simplifié permet de s'affranchir de ces études de post-traitement et garantit une plus grande transparence de la vérification du contrat. La performance théorique, ou la valeur garantie de consommation d'énergie, est calculée grâce à ce modèle.

- *Les facteurs d'ajustement*

Les données temporelles de sollicitations collectées en exploitation, caractéristiques des conditions réelles d'exploitation, permettent le calcul de facteurs d'ajustement (FA). Ceux-ci agglomèrent une information variable dans le temps en une valeur scalaire unique pour une période donnée. Par exemple, des conditions de température extérieure annuelle très différentes peuvent permettre le calcul d'une même valeur moyenne annuelle de la température. Il y a donc une nécessaire perte d'information dans cette description de grandeurs dynamiques, qu'elles soient liées aux conditions météorologiques ou d'usage, par des facteurs d'ajustement. Plusieurs FA différents, décrivant la même grandeur dynamique peuvent être calculés. Par exemple, la température extérieure moyenne ou les degré-jours unifiés (DJU) sont tous deux caractéristiques des températures extérieures. La période temporelle considérée peut elle aussi varier. Les FA les plus pertinents sont ceux expliquant le mieux les variations de la grandeur d'intérêt, c'est-à-dire minimisant la variabilité conditionnelle de la distribution de cette grandeur d'intérêt connaissant la valeur du FA.

- *Les modèles d'ajustement simplifiés*

Les modèles simplifiés lient consommations énergétiques d'intérêt et facteurs d'ajustement. Des modèles linéaires sont majoritairement utilisés du fait de leur simplicité permettant une compréhension par tous les acteurs des projets. Dans la littérature, des modèles *Black-box* (paragraphe 1.1.2.2) plus complexes sont développés pour caractériser les performances des bâtiments en fonction d'un nombre restreint de paramètres. Ils peuvent être non-linéaires et éventuellement non paramétriques. Trois type de modèles sont étudiés par Zhang et al. (2015) : des fonctions linéaires par morceaux, des processus gaussiens et un modèle de mélange gaussien. Des métamodèles de type processus gaussiens sont aussi utilisés par Heo (Burkhart et al., 2014 ; Heo et Zavala, 2012) dans un objectif de mesure et vérification de la performance. Ces formes plus élaborées de modèles visent à décrire avec plus de précision les variations dans les performances énergétiques que ce soit dans un objectif d'ajustement ou de prévision des consommations.

1.4.2.5. Le plan de mesure

La résolution du contrat nécessite la vérification des performances énergétiques réelles du bâtiment en exploitation. Celle-ci est réalisée à la fin de périodes prédéterminées, généralement sur des cycles annuels. Cette vérification repose sur la mesure des consommations énergétiques d'intérêt et sur la mesure des conditions d'exploitation pour lesquelles un processus d'ajustement a été planifié. Les moyens de mesure mis en œuvre pour acquérir ces données doivent donc être spécifiés en amont des projets. Le choix du plan de mesure est interdépendant avec le choix de la performance garantie et des modalités d'ajustement. Ces éléments sont liés à la fois par des facteurs économiques mais aussi par la détermination du risque associé à l'engagement du fait des incertitudes portant sur les mesures. Le protocole IPMVP impose ainsi un calcul précis de ces incertitudes afin qu'elles soient prises en compte dans le processus de vérification. La clarté du plan de mesure, des modalités d'acquisition et de traitement de données et des éventuels post-traitements est nécessaire afin que les parties puissent vérifier l'engagement en toute transparence.

1.4.2.6. Synthèse de la méthodologie

Cette partie résume les différents éléments présentés précédemment. La Figure 1-14 présente le phasage des étapes principales d'un projet de GPE au sein du processus global lié aux travaux et à l'exploitation d'un bâtiment dont on veut garantir la performance.

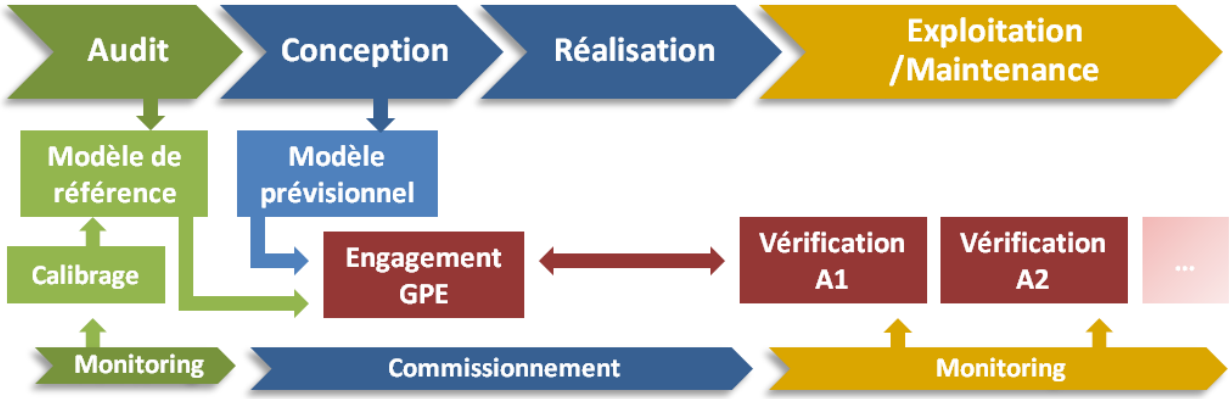

Figure 1-14 : Intégration des différentes phases de la GPE au sein du projet de construction ou réhabilitation.

On peut résumer le déroulement d'un projet de GPE de la manière suivante. Dans le cas d'une réhabilitation, l'engagement porte sur la performance du bâtiment après travaux exprimé comme un pourcentage de réduction par rapport à la situation initiale. Un modèle d'ajustement décrivant cette situation de référence est calculé, et éventuellement calibré. Une étude fournit des informations sur le gain de performance atteignable, permettant de définir le niveau d'engagement. À la fin de chaque année d'exploitation, les consommations réelles sont comparées aux consommations

calculées avec le modèle de référence ajusté dans les conditions réelles, permettant de calculer le gain réel de performance et de vérifier l'atteinte des objectifs. La Figure 1-15 récapitule la logique du processus d'engagement et de vérification et l'exploitation des données disponibles dans le cas d'un projet de rénovation.

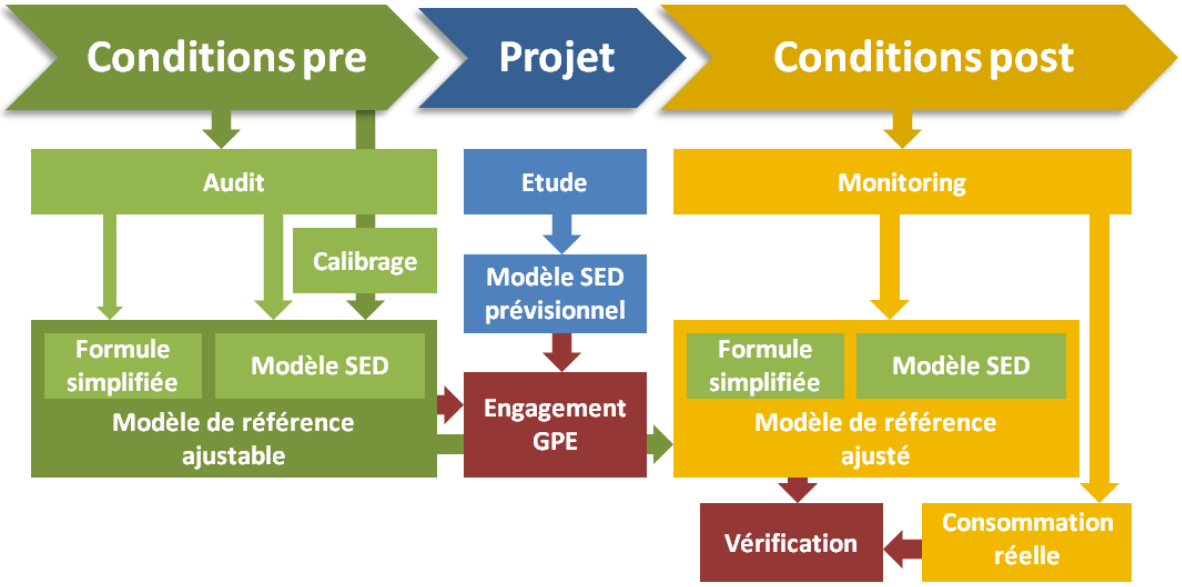

Figure 1-15 : Exploitation et utilisation des données lors des différentes phases d'un projet de GPE pour la réhabilitation

Dans le cas d'une construction neuve, ou d'une réhabilitation très lourde, l'engagement porte sur une consommation énergétique absolue maximale. L'étude technique durant la conception permet au porteur de la GPE, grâce aux outils de simulation, de définir cette valeur limite et un modèle d'ajustement. Les travaux réalisés, à la fin de chaque année d'exploitation, les consommations réelles sont comparées à la valeur limite garantie ajustée en fonction des conditions réelles mesurées (Figure 1-16).

S'intégrant au sein de ces schémas généraux, la littérature scientifique propose différentes méthodes qui ont été développées et mises en œuvre avec une approche globale ou s'intéressant à une étape du processus en particulier.

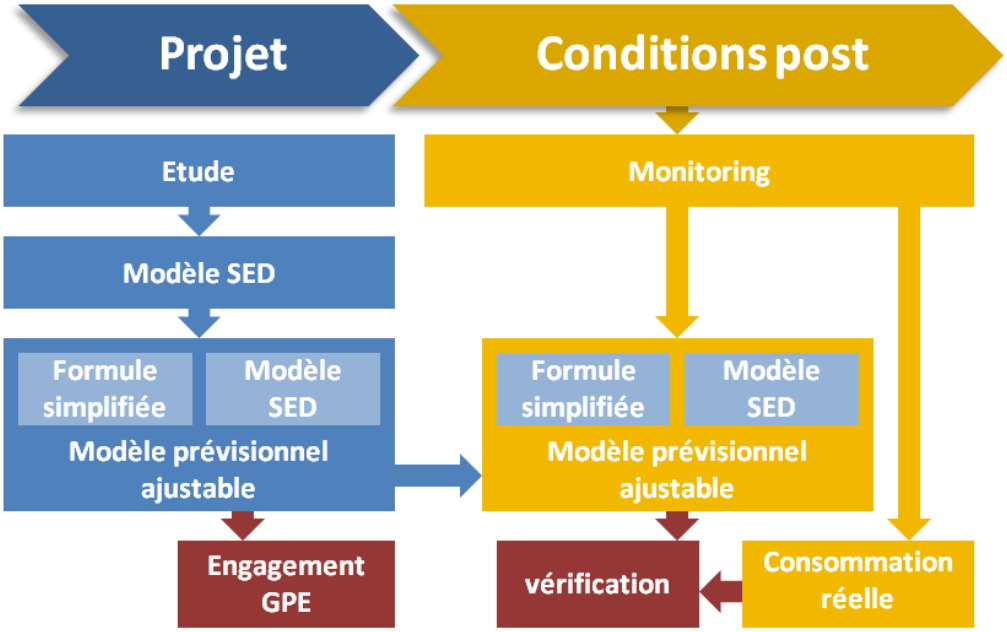

Figure 1-16 : Exploitation et utilisation des données lors des différentes phases d'un projet de GPE pour la construction neuve

1.4.3. État de l'art des méthodes

1.4.3.1. Modélisations et prévisions

Srivastav et al. (2013) se concentrent sur la caractérisation du modèle de bâtiment de référence par le développement d'un métamodèle de type *Gaussian mixture regression* (GMR). C'est un modèle paramétrique considérant une incertitude variable. La comparaison de la précision de ce type de modèle avec d'autres modèles de référence plus classiques est effectuée dans un contexte de GPE sur deux cas d'étude.

Tian et al. (2015) étudient une méthodologie de GPE appliquée à un parc de bâtiments. Des métamodèles d'ajustement prévisionnel des consommations sont développés (régression linéaire et modèles non-paramétriques) et leurs performances sont comparées. Un traitement spécifique de la problématique des paramètres d'ajustement corrélés est réalisé. Ces travaux concluent à la bonne capacité de prédiction des modèles simples et en particulier basés sur la régression linéaire.

Struck et Hensen (2007) décrivent l'évolution de l'incertitude associée à un projet au cours de ses différentes phases. Des analyses de sensibilité sont réalisées mettant l'accent sur les paramètres de conception sur lesquels doit se concentrer la démarche de commissionnement.

P. Lee et al. (2013) développe une méthodologie pour l'évaluation du risque associé aux contrats de performance énergétique (CPE) après rénovation incluant la prise en compte des incertitudes

dans les modèles de simulation. Des modèles du bâtiment pre-travaux et post-travaux sont réalisés et calibrés à partir des données de consommations et de mesures effectuées avant rénovation. Une analyse de sensibilité est par la suite menée permettant de définir des lois de probabilité sur les paramètres statiques réellement influents. Grâce à des tirages de paramètres par la méthode LHS sur les deux modèles, et une prise en compte simplifiée de la variabilité météorologique, la répartition probable des valeurs de gains énergétiques est réalisée. Un risque peut être finalement associé à cette loi de probabilité et à une valeur seuil.

Granderson (Granderson et al., 2015 ; Granderson et Price, 2014) utilise des métamodèles pour décrire les performances pré- et post-rénovation. Cinq modèles non physiques sont entrainés sur des périodes variables et avec des indicateurs de performance différents. Plusieurs périodes d'agrégation des consommations prédites sont aussi considérées. Les résultats obtenus sont comparés. Plus la quantité de données est importante et plus la modélisation est poussée, meilleurs sont les résultats. Les modèles prenant en compte l'évolution temporelle parmi leurs entrées produisent de meilleurs résultats. L'IPMVP recommande que des rénovations soient effectuées si elles entrainent des gains supérieurs à 10 % afin que l'intérêt soit significatif. Du fait du gain en précision permis par les modèles, l'auteur remet en cause l'utilité de cette règle.

Dans des travaux s'appuyant sur l'utilisation du logiciel de simulation *eQUEST* pour décrire un bâtiment tertiaire, Ke et al. (2013) suit la méthodologie de l'Option D de l'IPMVP en étudiant le calibrage du modèle de référence sur différents paramètres. L'éclairage ressort comme l'un des paramètres prépondérant tout comme l'efficacité des équipements et les caractéristiques des vitrages. L'étude s'intéresse à l'impact de la précision du calibrage sur les résultats de la démarche complète de GPE.

Manfren et al. (2013) étudient le processus de calibrage de différents métamodèles décrivant des bureaux avant et après rénovation. Des processus gaussiens et des modèles de régression linéaire par morceaux (*piecewise regression*) sont notamment utilisés. Leurs performances sont comparées aux résultats obtenus par la SED et fournissent des résultats satisfaisants.

Dans ses travaux sur la garantie de performance énergétique, Rivalin (2016) s'intéresse aux méthodes liées à la description et à l'exploitation du modèle physico-probabiliste dans l'objectif de prévoir les consommations énergétiques d'un bâtiment tertiaire. Des métamodèles sont utilisés afin de réduire les temps de calcul des analyses d'incertitudes réalisées.

1.4.3.2. Mesure et vérification de la performance

Brohus et al. (2009) mènent une démarche de caractérisation de différents bâtiments en intégrant des incertitudes aux paramètres des modèles SED réalisés. Des mesures des consommations et des usages sont effectuées et comparées aux hypothèses de simulation faisant

état d'écarts importants dans l'estimation des charges internes liées aux consommations électriques et à l'occupation.

Différentes équipes de recherche ont travaillé de manière intensive autour du développement de méthodes pour la prévision robuste des performances énergétiques et leur M&V. Dans une série d'articles traitant du développement de métamodèles associés à une méthode de M&V, Heo s'intéresse à l'utilisation d'un processus gaussien (GP) pour déterminer les gains énergétiques d'un système modifié ou rénové (Heo et Zavala, 2012 ; Burkhart et al., 2014). Contrairement aux régressions linéaires multiples, les processus gaussiens peuvent traiter les dépendances non-linéaires et mieux estimer les incertitudes associées à la prédiction. Le GP est entrainé sur une période-test où sont mesurées les entrées (variables explicatives) et les sorties du modèle (consommations). L'utilisation de cette technique sur un cas d'étude montre la bonne capacité de prédiction du modèle avec un nombre restreint de données d'entrainement.

Certains travaux académiques s'inscrivent au sein des annexes de l'agence internationale de l'énergie (AIE), et plus particulièrement dans l'annexe 52 intitulée *Towards Net Zero Energy Solar Building*. Un rapport de cette commission sur la M&V des performances des bâtiments à énergie positive propose un état de l'art des publications dans le domaine ainsi qu'un ensemble de méthodes et de bonnes pratiques associées à la mesure.

L'annexe 58*, Reliable Building Energy Performance Characterisation Based on Full Scale Dynamic Measurements*, concerne encore plus directement les travaux présentés dans ces travaux de thèse puisqu'elle intègre la notion de vérification de la qualité des bâtiments et de leurs performances via une approche de mesure globale. Roels et al. (2015) présentent des exemples de résultats des travaux réalisés dans le cadre de cette collaboration. L'annexe 71, *Building Energy Performance Assessment Based on In-situ Measurements*, se déroulant sur la période 2016-2021, s'inscrit dans la suite de ces travaux et a pour objectif de clarifier et d'unifier les méthodes puis de les tester sur des cas d'étude.

Titikpina (2016) développe une méthode pour la sélection d'un plan de mesure optimal. Deux niveaux d'incertitude différents sont considérés sur les paramètres correspondant à une connaissance de conception et à une connaissance de mesure. Les modes de calcul de ces différentes incertitudes sont présentés dans ses travaux. La méthodologie finale repose sur l'identification des paramètres à mesurer prioritairement de manière à minimiser l'incertitude sur la grandeur d'intérêt.

Fennel et al. (2017) étudient l'impact des frontières du système considérées, dépendant de l'option de l'IPMVP choisie, sur les conclusions du processus de mesure et vérification dans une démarche de garantie de performance. Elle souligne le risque de déplacement des consommations et donc de l'importance d'une approche globale à la fois dans l'engagement et dans le processus de M&V.

1.4.3.3. Aide à la décision et maitrise du risque

Dans différentes études, Deng (2014 ; 2015a ; 2015b) s'intéresse aux stratégies de contractualisation des engagements sur la garantie de performance énergétique, travaillant autour des mécanismes de pénalisation et rétribution en fonction des niveaux de performance en s'appuyant sur la simulation. Cette approche financière du problème de la GPE nécessite la caractérisation des risques portant sur la performance énergétique des bâtiments. Des phénomènes comme la détérioration des ouvrages et l'évolution du prix de l'énergie sont pris en compte et modélisés.

Qin et al. (2017) étudient l'impact des différents modèles de CPE dans le contexte chinois et proposent une méthode d'aide à la décision pour choisir un modèle et ses paramètres. Celle-ci quantifie aussi les gains financiers associés au déploiement du contrat.

Ma et al. (2012) proposent un état de l'art des techniques d'aide à la décision sur la réalisation d'actions de rénovation. Des liens sont réalisés entre les aspects liés aux risques, les aspects financiers et les questions de mesure et vérification de la performance réelle. L'approche globale choisie dans cette revue permet d'embrasser l'ensemble des problématiques rencontrées sur ce type de projet.

1.5. Positionnement de la thèse

1.5.1. Objectifs scientifiques

En s'appuyant à la fois sur la littérature scientifique et le contexte industriel, nous pouvons définir les objectifs de ces travaux et leur positionnement par rapport à l'état de l'art. Cette thèse doit proposer une méthodologie globale et innovante permettant de mettre en œuvre des contrats de garantie de performance énergétique des bâtiments. Celle-ci doit unifier différents modèles, techniques et méthodes scientifiques et statistiques solides et validés. Elle repose sur l'exploitation de données provenant d'outils de simulation énergétique dynamique des bâtiments et liées au déploiement de moyens de mesure.

Elle doit intégrer une quantification du risque associé à l'engagement et proposer des modalités d'ajustement permettant de s'affranchir de l'effet des sollicitations externes non liées à la performance du bâtiment : les conditions météorologiques et le comportement des usagers.

La méthodologie doit spécifier les appareils et les modalités de mesure mis en œuvre pour la vérification des performances et leur contextualisation, et prendre en compte l'impact des incertitudes associées sur le niveau d'engagement.

Le cadre méthodologique a vocation à être généralisable à tout type de bâtiment, aux projets de construction neuve et de réhabilitation. La possibilité d'un déploiement industriel rapide constitue aussi un des objectifs de ces travaux qui différencieront ainsi les méthodes aisément compréhensibles par les acteurs et faciles à mettre en œuvre de celles dont le transfert entre le domaine de la recherche et le monde industriel est voué à prendre plus de temps.

1.5.2. Plan du manuscrit

Afin de répondre à ces objectifs, la progression suivante est proposée. Le chapitre 2 est consacré au développement et à l'étude d'un modèle de génération de données météorologiques stochastiques pour la SED. Ce modèle constitue une brique nécessaire à l'approche de simulation physico-probabiliste présentée. Ces développements s'expliquent par l'identification de manques parmi les outils disponibles et par l'opportunité d'améliorer les modèles existants. Un état de l'art spécifique au sujet est présenté puis le modèle proposé est décrit. Une étude est menée afin de valider ces développements. Le modèle est finalement testé sur un cas d'étude afin d'illustrer son intérêt.

Le chapitre 3 présente le cœur de la méthodologie pour la garantie de performance énergétique. Les différentes étapes et leur articulation logique sont décrites. Les cas de l'engagement sur des valeurs de performance énergétique absolues ou relatives sont différenciés. La méthode est ensuite testée sur un cas d'étude simple.

La mise en œuvre de cette méthodologie fait face à plusieurs problématiques et celles-ci sont étudiées au chapitre 4. La stabilité statistique des modèles de régression identifiés est d'abord investiguée. On s'intéresse ensuite à l'influence des incertitudes de mesure lors de la vérification sur le niveau d'engagement. La question du choix des facteurs d'ajustement est ensuite posée. Enfin différentes formes de modèles d'ajustement de la limite garantie de consommations d'énergie sont comparés.

Le chapitre 5 présente deux cas d'étude ou la méthodologie est mise en œuvre. La rénovation d'un bâtiment de logements, décrit dans le paragraphe suivant, est d'abord étudiée. La contractualisation d'un engagement sur la diminution des consommations de chauffage est envisagée. Ensuite, on s'intéresse à la construction d'un bâtiment de bureaux à haute performance énergétique. Sur ce second cas d'étude, on envisage une garantie sur les consommations énergétiques de chauffage et de climatisation fournis par une pompe à chaleur d'une part, puis sur les consommations totales.

1.6. Modèles et cas d'étude

1.6.1. Modèles et méthodes utilisés

Dans cette partie, les modèles et méthodes déjà existants qui seront exploités dans ces travaux de thèse sont présentés et l'intérêt de leur utilisation est justifié.

1.6.1.1. COMFIE : modèle de simulation énergétique dynamique

Le moteur de calcul thermique dynamique COMFIE, issu des travaux de Peuportier et Blanc-Sommereux (1990) est une des briques de base de la méthodologie développée dans cette thèse. Ce modèle repose sur le concept de zone thermique, sous-ensemble du bâtiment à température homogène. Un maillage est réalisé entre les différentes zones (méthode des volumes finis) suivi d'un bilan thermique en considérant une température et éventuellement une capacité associées à chaque maille. L'ensemble des équations peut finalement être représenté sous la forme d'un système différentiel linéaire reliant les sorties, l'état du système et les sollicitations (annexe A-1-1). Une réduction modale permet de limiter les temps de calcul en conservant une bonne précision sur la résolution. Une description plus détaillée de ce modèle énergétique est disponible dans l'annexe A-1-1. Le modèle COMFIE est actuellement intégré au logiciel Pléiades, développé par Izuba Energies.

Plusieurs raisons justifient le choix d'utiliser COMFIE :

- Il a passé avec succès différentes études de validation (Brun et al., 2009).
- L'approche de réduction par analyse modale permet une diminution des temps de calcul pour une bonne précision des résultats. Dans une optique de simulations intensives (méthodes d'analyse de sensibilité, de propagations d'incertitudes ou de calibrage), cette caractéristique permet de s'affranchir de la nécessité de développer des métamodèles, souvent associés à des incertitudes supplémentaires et des hypothèses restrictives.
- Le partenariat historique avec le développeur permet des modifications du modèle énergétique.

1.6.1.2. Méthodes d'analyses de sensibilité et d'incertitude

La principale méthode d'analyse de sensibilité utilisée est la méthode de Morris (1991), brièvement présentée au § 1.2.1. Celle-ci repose sur la répétition de plans d'expérience, dits *one-at-a-time* (OAT), dans l'espace des entrées incertaines. La valeur d'un seul facteur d'entrée est donc modifiée à chaque simulation du plan OAT. La méthode de Morris repose sur l'étude de la dispersion (moyenne et écart-type) des effets élémentaires observés sur la sortie lors des répétitions des variations d'un facteur d'entrée. Cette méthode, peu coûteuse en temps de calcul et ne nécessitant

pas d'hypothèses sur la régularité du modèle, peut prendre en compte des interactions et non-linéarités. Elle permet de classer les facteurs d'entrée en trois catégories, représentées à la Figure 1-17 :

1. les facteurs incertains ayant des effets négligeables ;
2. les facteurs ayant des effets majoritairement linéaires ;
3. les facteurs ayant des effets non linéaires et/ou avec interactions.

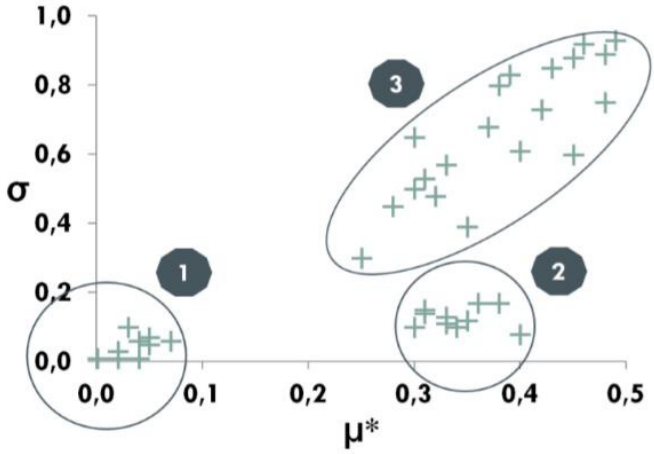

Figure 1-17 : Identification des groupes principaux de facteurs incertains par la méthode d'analyse de sensibilité de Morris (Recht, 2016)

Concernant les propagations d'incertitude, la méthode d'échantillonnage utilisée dans les différents résultats est une méthode Monte Carlo aléatoire si rien n'est précisé, ou un tirage quasi-Monte Carlo *LHS*, pour les paramètres statiques du modèle de bâtiment.

1.6.1.3. Modèle de génération stochastique de comportement des occupants

Dans ses travaux de thèse, Vorger (2014; 2014b), développe un modèle de génération de données liées au comportement des usagers des bâtiments de logements et de bureaux. Ce modèle s'appuie sur l'exploitation d'un grand nombre de données issues de campagnes de mesure, d'enquêtes sociologiques et de la littérature scientifique (Haldi et Robinson 2011a, 2012). Les calculs sont majoritairement réalisés en prétraitement de la SED, excepté pour les ouvertures de fenêtres et occultations. L'outil repose sur une cascade de tirages aléatoires et de liens de cause-conséquence basés sur les différentes sources d'informations.

- *Logements*

Concernant les usages dans les logements, le modèle associe d'abord un ménage à chaque appartement. Le nombre de personnes et leurs caractéristiques personnelles dépendent notamment de la taille du logement. Ces données peuvent être forcées si elles sont connues. Des scénarios

d'activité sont générés ainsi que les équipements du logement. Des scénarios d'occupation des lieux et des usages sont ensuite déduits. En parallèle les scénarios de température de consigne sont construits en fonction des pièces du logement et de ces habitants. Ces tirages permettent de fournir à la SED les données d'entrée nécessaires à la simulation pour chaque logement : des scénarios de température de consigne, d'occupation, de puissances dissipées, de puisages d'ECS. Des modèles adaptatifs pour l'ouverture et l'occultation des fenêtres sont aussi disponibles. Ils interagissent avec la simulation, étant liées aux conditions météorologiques. Le modèle de Vorger impose qu'une zone thermique (maille spatiale de la SED) au moins décrive chaque logement.

- *Bureaux*

Les modèles concernant les usages tertiaires traitent principalement le cas des espaces de bureaux. La aussi des scénarios de présence des usagers sont générés ainsi que la composition du matériel informatique installé. Les actions liées à l'éclairage et à la bureautique sont déterminés. Pour les bureaux, les sorties du modèle (et donc les entrés de la SED) sont les scénarios horaires d'occupation, de puissance dissipée et de puisages d'ECS.

Différents éléments permettent de justifier l'utilisation de ce modèle dans le cadre de ces travaux.

- Le modèle repose sur l'approche stochastique, utilisée dans cette thèse, visant à modéliser la variabilité des sollicitations du bâtiment en générant un échantillon statistiquement représentatif de jeux de données décrivant chacun une situation particulière unique.
- Ce modèle, développé dans l'équipe Éco-conception et Thermique des Bâtiments du CES de l'École des Mines de Paris est couplé au logiciel *Pléiades*+COMFIE et au logiciel d'orchestration de simulations intensives *Amapola*. Il est donc lié aux autres outils et modèles utilisés.

1.6.1.4. Orchestration de la simulation intensive

L'approche stochastique utilisée dans ces travaux, sur laquelle repose la simulation du modèle physico-probabiliste de bâtiments, nécessite de réaliser un grand nombre de simulations (de l'ordre du millier). Une surcouche logicielle permet d'orchestrer ces simulations et les variations paramétriques qui les accompagnent, de piloter de manière automatique le lancement des simulations énergétiques individuelles et de traiter leurs sorties. Cet orchestrateur peut aussi fournir une interface graphique permettant la mise en œuvre des différentes méthodes s'appuyant sur la simulation intensive.

Le logiciel *Amapola*[5] est couplé à la suite logicielle *Pléiades*+COMFIE et a pour objectif de réaliser ces différentes tâches. Il repose sur une architecture codée en langage Python. Son utilisation se

[5] www.kocliko.co

justifie par ce couplage avec le logiciel de SED utilisé et la qualité du travail de maintenance et de développement associé.

1.6.2. Etude de cas : Rénovation de logements collectifs

Cette partie décrit un projet de rénovation d'un bâtiment de logements collectifs servant de cas d'étude pour illustrer les travaux de thèse présentés dans la suite de ce manuscrit. Des résultats associés aux différentes étapes de la méthodologie développée seront présentés sur ce cas d'étude « fil rouge ».

1.6.2.1. Description du bâtiment

Le bâtiment étudié est un immeuble R+3 d'une surface habitable de 1048 m² situé à Feyzin, en banlieue sud de Lyon, comprenant 16 logements (T1, T2, T3 et T4). Le Tableau 1-3 présente les caractéristiques des logements, notamment le nombre d'occupants.

Tableau 1-3 : Caractéristiques des logements

Logement	Type	Étage	Orientation	SHAB (m²)	Nb. occupants
L1	T3	RDC	NE	66	4
L2	T3	RDC	SO	66	6
L3	T4	RDC	NO	86	6
L4	T1	RDC	SE	25	1
L5	T3	R+1	NE	66	1
L6	T2	R+1	SE	57	2
L7	T3	R+1	SO	66	2
L8	T4	R+1	NO	86	3
L9	T3	R+2	SO	66	2
L10	T2	R+2	SE	57	1
L11	T3	R+2	NE	66	3
L12	T4	R+2	NO	86	2
L13	T3	R+3	NE	66	3
L14	T2	R+3	SE	57	2
L15	T3	R+3	SO	66	2
L16	T3	R+3	NO	66	3

http://www.izuba.fr/logiciel/amapola

- *Détails de la rénovation*

La rénovation du bâtiment a concerné en premier lieu une sur-isolation de l'enveloppe. Les systèmes de production et de distribution de chauffage et de préparation d'eau chaude sanitaire (ECS) ont aussi été remplacés, ainsi que les bouches de ventilation. Le Tableau 1-4 récapitule les caractéristiques de l'ouvrage avant et après rénovation.

Tableau 1-4 : caractéristiques énergétiques du bâtiment avant et après rénovation

		Avant rénovation	**Après rénovation**
Enveloppe	Mur extérieur	Béton (15 cm) Polystyrène intérieur (6 cm)	Béton (15 cm) Polystyrène intérieur (6 cm) Polystyrène extérieur (10 cm)
	Plancher bas	Béton (16 cm) Polystyrène extrudé (6 cm)	Béton (16 cm) Polystyrène extrudé (8 cm)
	Toiture terrasse	Béton (20 cm) Polyuréthane (5 cm)	Béton (20 cm) Polyuréthane (10 cm)
Systèmes	Chauffage	Planchers chauffants électriques + appoints	Sous-station chaufferie centrale + radiateurs à eau
	ECS	Cumulus électriques	Système solaire + appoint gaz
	Ventilation	Autoréglable	Hygro. B

1.6.2.2. Modélisation

- *Zonage et caractéristiques thermiques*

Le bâtiment a été modélisé, avec le logiciel *Pléiades*+COMFIE, dans les situations avant et après rénovation. Il est décomposé en 17 zones thermiques correspondant à chacun des appartements et aux parties communes et espaces de circulation. Ce découpage fin est nécessaire à l'utilisation des modèles de variabilité stochastique des modes d'occupation. Des scénarios indépendants sont associés à chacun des logements et l'impact de cette variabilité sur les conditions thermiques et les besoins énergétiques doit être décrit pour chaque appartement. La Figure 1-18 présente une vue 3D du bâtiment et des différentes zones thermiques considérées.

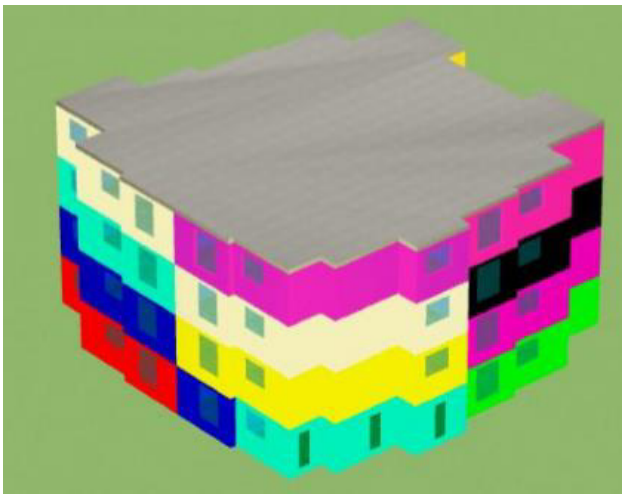

Figure 1-18 : Modèle zoné du bâtiment, logiciel Alcyone

Concernant le renouvellement d'air, un débit de VMC de 0,67 vol/h est considéré ainsi que 0,55 vol/h d'infiltrations dans la situation avant travaux. Après travaux, les spécifications des systèmes sont connues, et un débit de 0,3 vol/h est considéré pour la ventilation. Du fait de la rénovation, les infiltrations ont été réduites et un débit de 0,2 vol/h est pris en compte.

- *Usages*

Dans une démarche de méconnaissance de l'existant et de simulation du modèle physico-probabiliste complet, les scénarios d'usage sont fournis par le modèle de génération stochastique présenté au § 1.6.1.3. Les entrées du modèle stochastique sont constituées de la composition des foyers pour chaque logement et de la spécification de leur statut de locataire en zone urbaine. A partir de ces informations, des scénarios de puissance dissipée, d'occupation, de puisages d'ECS, de température de consigne et d'ouverture/fermeture des fenêtres et occultations sont générés à chaque simulation.

- *Systèmes*

Le chauffage avant travaux est fourni au moyen de planchers chauffants et appoints électriques. Après travaux, une production par chaudière à gaz centralisée permet la fourniture de chauffage à l'aide d'une colonne d'eau irriguant le bâtiment. La Figure 1-19, issue de la documentation technique, décrit ce réseau de distribution (longueurs et diamètres des canalisations) modélisé dans l'outil de simulation.

La production d'ECS, initialement réalisée dans des ballons électriques individuels, est, après rénovation, fournie par un système couplant énergie solaire et gaz. L'architecture de l'installation comprend deux ballons et différentes règles de priorité entre les apports solaires et l'énergie provenant de la boucle d'eau chaude. Étant donné la complexité de cette architecture, des problématiques de mise en œuvre rencontrées lors de la première année d'exploitation (Mora et al., 2016) et des mesures disponibles limitées, ce système n'a pas été modélisé et une prise en compte

simplifiée de ses performances sera considérée. Des apports de chaleur sont en particulier pris en compte par les colonnes d'eau chaude.

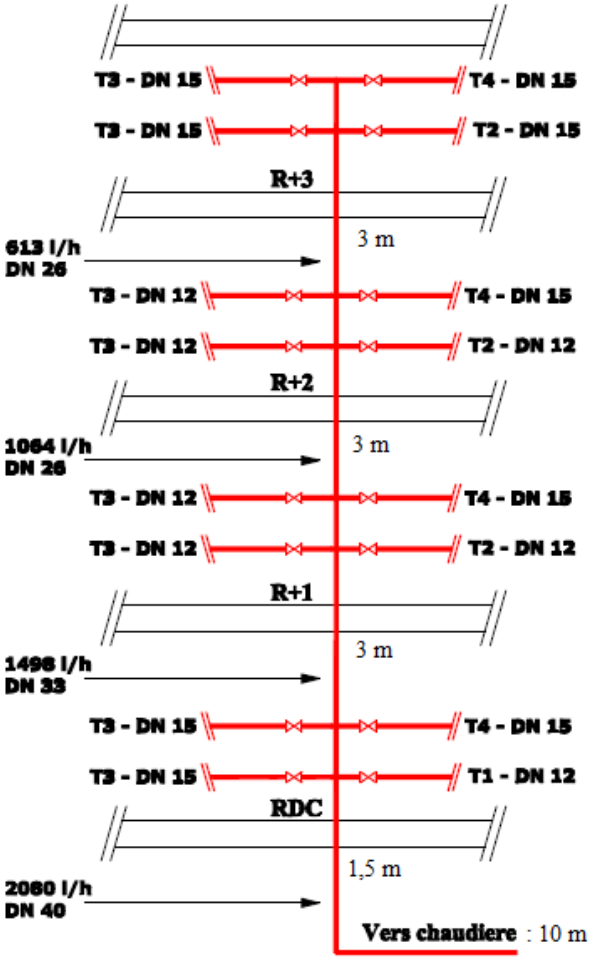

Figure 1-19 : Schéma de l'installation avec diamètres nominaux et débits (issu du CCTP DCE)

1.6.2.3. Mesures

Dans le cadre d'une étude réalisée par (Enertech, 2013), des mesures ont été effectuées comprenant un suivi de différents usages et consommations énergétiques dans les logements. Le fonctionnement détaillé des systèmes de ventilation, d'ECS et de chauffage a été étudié à la fois avant rénovation et durant la première année d'exploitation du bâtiment après la réhabilitation. Ces mesures portent sur les températures intérieures et les compteurs électriques de chaque appartement. Des compteurs d'énergie ont été installés sur l'installation de chauffage après travaux. Le Tableau 1-5 présente les périodes de mesure et de chauffe et les consommations énergétiques de chauffage relevées.

Tableau 1-5 : Consommations énergétiques de chauffage avant et après rénovation

	Avant rénovation	**Après rénovation**
Période mesure	2/11/08 – 2/11/09	13/02/12 – 13/02/13
Saison de chauffe	3/11/08 – 15/05/09 et 15/10/09 – 2/11/09	14/02/12 -15/05/12 et 12/10/12 – 13/02/13
Consommation (kWh)	112030	64700
Consommation (kWh/m²)	106,9	61,7

Des détails supplémentaires concernant le bâtiment et les mesures réalisées peuvent être trouvés dans les rapports d'Enertech (2013) et de Mora et al. (2016), mais aussi dans l'annexe A-5-1. Les données spécifiques nécessaires à l'exploitation du cas d'étude, concernant notamment les conditions d'exploitation, seront présentées au fur et à mesure des développements de cette thèse et accompagneront les résultats auxquels ils se rapportent.

1.7. Conclusion

Ce premier chapitre a permis de poser le cadre de ces travaux de thèse dont l'objectif est le développement d'une méthodologie pour la garantie de performance énergétique (GPE) des bâtiments.

Dans le contexte de la transition énergétique, la GPE est un mécanisme ayant pour but d'encourager et de faciliter le financement de projets de construction ou de rénovation mettant l'accent sur la sobriété énergétique des bâtiments. En effet, l'engagement pris par la structure réalisant les travaux et/ou le mainteneur sur des performances énergétiques réelles permet de mieux anticiper les coûts et les retours sur investissements liés aux travaux.

La méthodologie proposée s'appuie sur la modélisation des ouvrages bâtis et sur les outils de simulation énergétique dynamique (SED). Un état de l'art a été réalisé sur l'exploitation de la simulation pour la prévision des consommations énergétiques. Il décrit les limites des modèles pour représenter la réalité de ces objets complexes et met en avant des marges de progression. Celles-ci reposent notamment sur la modélisation poussée des incertitudes pesant sur les différents paramètres physiques et techniques des bâtiments et des systèmes énergétiques. La variabilité des conditions d'exploitation doit aussi être prise en compte, qu'elle concerne l'usage ou les conditions climatiques. Des modèles spécifiques ont ainsi pour objectif de générer une variabilité réaliste de scenarios. L'articulation entre le modèle énergétique du bâtiment et ces différents sous-modèles constitue le modèle physico-probabiliste global. L'approche stochastique, reposant sur des tirages aléatoires dans les différentes lois et modèles probabilistes, vise à simuler le bâtiment étudié dans un ensemble de configurations statistiquement représentatives.

Les problématiques liées à la mesure des performances énergétiques ont ensuite été abordées. L'exigence de résultats associée à la GPE s'appuie sur la vérification concrète de ces performances. Les mesures des conditions d'exploitation et des consommations énergétiques sont nécessairement entachées d'erreur. Des bonnes pratiques et protocoles encadrent ce processus de mesure et vérification. La connaissance de données d'exploitation peut aussi permettre d'affiner les modèles grâce à des méthodes de calibrage. Un état de l'art clôt cette partie liée aux mesures revenant sur le constat généralisé d'un écart systématique et souvent important entre consommations énergétiques prévues en conception et valeurs effectivement mesurées en exploitation après les travaux.

C'est de ce constat que nait l'intérêt de la garantie de performance énergétique des bâtiments. Les points-clés de la GPE ont ensuite été présentés. Il est dans un premier temps nécessaire de définir le périmètre de l'étude. La méthode d'évaluation permettant le calcul de la performance garantie doit intégrer une quantification des risques associés au projet et in fine, du risque de défaut de l'engagement. Un mécanisme d'ajustement doit aussi être mis en place afin de s'affranchir de l'impact des conditions d'exploitation, climat et usages, qui ne sont pas liés aux performances de l'ouvrage et qui ne doivent pas peser sur la responsabilité de l'entreprise portant le contrat de garantie. Enfin, les moyens de mesure doivent être spécifiés afin d'assurer une vérification transparente pour l'ensemble des parties.

Ces travaux de thèse ont pour objectif de développer un processus fournissant de manière intégrée ces différents éléments et reposant sur la SED et la description du modèle physico-probabiliste. Ils doivent s'appuyer sur des méthodes scientifiques validées pour proposer une méthodologie globale associant une valeur garantie à des modalités d'ajustement et à un risque quantifié de dépassement mais aussi aux moyens de mesure déployés.

Dans la dernière partie du chapitre, plus utilitaire, les modèles et méthodes existants qui seront utilisés dans ces travaux ont été présentés. Un cas d'étude a finalement été décrit ayant vocation à être utilisé tout au long du manuscrit et de ses différents chapitres afin d'illustrer les méthodes développées. Il traite de la rénovation d'un bâtiment de logements.

La prise en compte de l'impact des conditions météorologiques sur le fonctionnement énergétique des bâtiments doit être incluse dans la description du modèle physico-probabiliste et des limitations sont identifiées dans la littérature à ce sujet. Le chapitre suivant aborde cette problématique et a pour objectif de développer un modèle de génération stochastique de données météorologiques.

Chapitre 2. Modèle de génération de données météorologiques pour la prise en compte de la variabilité climatique

Ce chapitre traite de l'influence sur les consommations d'énergie de la variabilité naturelle des conditions météorologiques. La prise en compte de ce facteur dans la simulation des bâtiments passe par une modélisation réaliste de cette variabilité. Un état de l'art est tout d'abord réalisé s'intéressant à la manière dont sont considérées les conditions climatiques en simulation du bâtiment puis aux travaux liés à la quantification de l'influence des variations des sollicitations météorologiques. Les principaux modèles de génération de données météorologiques synthétiques et leurs champs d'application sont aussi détaillés. Ce chapitre présente ensuite un modèle innovant de création de données artificielles de température extérieure et de rayonnement solaire, développé pour différents objectifs : compléter le modèle physico-probabiliste du bâtiment permettant une propagation des incertitudes et variabilités, et étudier le lien entre les conditions météorologiques et les consommations d'énergie dans une perspective de construction de modèles d'ajustement, objet de la suite de ces travaux. Les résultats obtenus sont validés puis le modèle est testé sur un bâtiment où des besoins de chauffage et de refroidissement sont considérés.

Sommaire du chapitre 2

Chapitre 2. Modèle de génération de données météorologiques pour la prise en compte de la variabilité climatique 67

2.1. Introduction 69

2.2. État de l'art 71

2.2.1. La modélisation des conditions météorologiques 71

2.2.2. L'impact du climat sur les consommations d'énergie des bâtiments 76

2.2.3. Générer la variabilité climatique 79

2.3. Objectifs du modèle développé 83

2.3.1. Propagation des effets de la variabilité climatique dans les modèles SED 83

2.3.2. Apprentissage pour l'entraînement de modèles d'ajustements 84

2.3.3. Limitations et apports scientifiques 85

2.4. Méthodologie pour la génération de données météorologiques 86

2.4.1. Génération de données de température 86

2.4.2. Génération des données de rayonnement global horizontal 91

2.4.3. Génération des données de rayonnement diffus horizontal 95

2.4.4. Variation des parts tendancielles 95

2.5. Evaluation de la fiabilité du modèle 101

2.5.1. Caractéristiques des séries générées 101

2.5.2. Respect de l'intercorrélation entre grandeurs météorologiques 106

2.6. Application à l'étude de l'influence du climat sur les consommations d'énergie d'un bâtiment 107

2.6.1. Propagation de la variabilité 107

2.6.2. Caractérisation de l'impact des conditions météorologiques 109

2.7. Conclusion 113

2.1. Introduction

Les conditions météorologiques constituent une part importante des sollicitations thermiques des systèmes « bâtiments ». Les différents phénomènes physiques mis en jeux dans ces systèmes sont quasiment tous reliés à ces sollicitations. Les transferts de chaleur à travers les parois opaques, décomposés en échanges par convection, rayonnement et conduction sont liés à l'écart de température entre l'ambiance intérieure et l'extérieur. L'air extérieur a également un impact via la ventilation et les infiltrations. La maîtrise de ces renouvellements d'air est d'ailleurs un des enjeux principaux pour la performance énergétique des bâtiments. Les conditions de vent ont un impact sur le volume des infiltrations mais aussi sur les échanges convectifs au niveau des parois extérieures. L'énergie solaire apportée à travers les vitrages peut constituer une source de chauffage gratuite non négligeable, en particulier dans les bâtiments actuels ou les pertes thermiques sont réduites. Ces apports sont cependant associés à un risque de surchauffe estivale diminuant le confort des usagers et pouvant engendrer de fortes consommations d'énergie de climatisation. L'impact de l'humidité relative peut aussi être cité du fait des échanges d'énergie liés au changement de phase ou de l'influence sur l'efficacité de certains systèmes énergétiques (p. ex. pompes à chaleur…). La Figure 2-1 schématise les interactions thermiques et énergétiques principales entre le bâtiment et l'environnement extérieur.

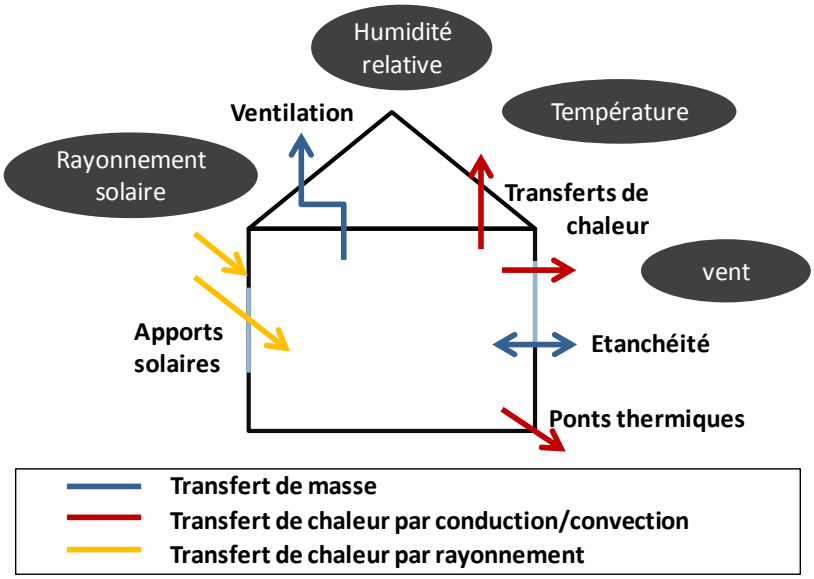

Figure 2-1 : Interactions énergétiques entre les sollicitations météorologiques et le bâtiment

Ces sollicitations ont un tel impact qu'elles orientent aujourd'hui en bonne partie les modes de conception des ouvrages bâtis. La notion d'architecture bioclimatique renvoie ainsi à une volonté d'optimiser les interactions entre les sollicitations climatiques et le système afin de minimiser les besoins énergétiques. Elle a été prolongée ensuite avec les concepts de bâtiments passifs, zéro

énergie et même autonomes. Cette approche repose ainsi sur la maximisation des apports solaires lors des périodes froides, sur la limitation de ces apports en été, sur des solutions liées au refroidissement naturel… Le premier bâtiment autonome en énergie, qui comporte notamment un stockage saisonnier et des murs solaires à isolation transparente, est conçu selon ces principes. La maison Hölken, est située à Fribourg-en-Brisbau (Allemagne) et a été bâtie en 1992 (Figure 2-2). On constate notamment l'importance des surfaces vitrées exposées au sud.

Figure 2-2 : Bâtiment bioclimatique et autonome en énergie : Maison Hölken – Fraunhofer-ISE (1992)

Les grandeurs météorologiques dépendent aussi de la position géographique du système. Le climat tempéré présent sur une partie importante de l'Europe est caractérisé par des taux d'humidité et des vents modérés. Ces grandeurs ne sont donc prises en compte que pour des applications ou des systèmes énergétiques spécifiques et ne constituent habituellement pas une donnée d'entrée des modèles de simulation énergétique dynamique du bâtiment (SED). En se reportant à l'approche de Crawley (1998), seuls la température de l'air et le rayonnement solaire seront modélisés dans nos travaux.

L'ensemble des interactions citées ci-dessus sont dynamiques, à la fois du fait de la variabilité temporelle des sollicitations que des effets capacitifs intrinsèquement associés au système, ce qui rend difficile leur analyse et leur modélisation. L'évolution des conditions météorologiques est un phénomène déterministe mais chaotique et donc très complexe à modéliser, ce qui limite la possibilité de réaliser des prévisions fiables au-delà de quelques jours. Sur des échelles de temps plus longues, de l'ordre de plusieurs années, la modélisation de ces conditions météorologiques doit ainsi être faite selon une approche stochastique. Tous ces facteurs donnent donc une place particulière à l'intégration des grandeurs météorologiques dans les modèles de simulation et l'approche choisie doit être guidée par les objectifs associés aux simulations. Par ailleurs, les méthodes liées à l'analyse de modèles (analyses de sensibilité et d'incertitude, inversion et simplifications de modèles, etc.) doivent être adaptées pour prendre en compte ce type d'entrées dynamiques. Différents travaux se sont intéressés à ces problématiques et une revue de ceux-ci est proposée dans la suite de ce chapitre.

2.2. État de l'art

Afin d'introduire et de contextualiser les travaux présentés dans ce chapitre une revue bibliographique spécifique est tout d'abord présentée. Cet état de l'art aborde les contributions scientifiques ayant trait à la modélisation des conditions météorologiques, à l'étude de l'impact des variations de ces conditions sur les consommations énergétiques des bâtiments, puis s'intéresse aux modèles de génération de données météorologiques synthétiques.

2.2.1. La modélisation des conditions météorologiques

2.2.1.1. Modèles simplifiés

La caractérisation du lien entre conditions météorologiques et besoins énergétiques du bâtiment est rendue difficile par la complexité du système étudié et par la dynamique de son comportement thermique. Cependant, des modèles simplifiés existent et sont soit basés sur des lois heuristiques soit construits à partir de mesures relevées en exploitation.

La définition de variables agglomérées permet de s'affranchir de l'aspect dynamique inhérent aux grandeurs climatiques. Les degrés-jours unifiés (DJU_{Tref}), divisés en degrés-jours de chauffe ou de refroidissement, calculés à partir d'une température de base T_{ref}, font partie des variables communément utilisées. Les DJU de chauffage (2.1) correspondent à une somme des écarts positif entre la température de référence et la température extérieure (moyenne des températures minimale et maximale du jour). Pour les besoins de chauffage, la référence habituelle est de 18°C et la période de sommation des N_j jours considérés est soit l'année entière soit la période de chauffe. En refroidissement, la différence doit être négative et la température de référence est définie en fonction de la consigne de climatisation.

$$DJU_{T_{ref}} = \sum_{\mathrm{j}=1}^{\mathrm{N_j}} \max\left(0\,;\,\left[T_{ref} - \left(\frac{T_{j,ext,max} + T_{j,ext,min}}{2}\right)\right]\right) \quad \textbf{(2.1)}$$

Les degrés-heures DH_{Tref} sont définis de manière semblable, le pas de sommation étant réduit au niveau horaire et la température extérieure considérée étant la moyenne sur chaque heure :

$$DH_{T_{ref}} = \sum_{\mathrm{h}=1}^{\mathrm{N_h}} \max\left(0\,;\,\left[T_{ref} - T_{h,ext,moy}\right]\right) \quad \textbf{(2.2)}$$

Ce type de variable agglomérée permet de construire des modèles simplifiés. L'effet de la température extérieure dans un calcul de besoins énergétiques peut ainsi être pris en compte à partir de la spécification de valeurs de DJU ou de DH en fonction de différentes zones géographiques. Ces valeurs permettent de calculer les déperditions thermiques annuelles par conduction et renouvellement d'air en régime permanent de manière simplifiée. Une prise en compte forfaitaire de la capacité thermique du bâtiment est réalisée à partir d'abaques caractérisant les modes constructifs du bâti. Ce type de méthode permet dans une première approche de faire des estimations grossières des besoins de chauffage ou de refroidissement d'un bâtiment.

Des corrélations sont aussi construites entre les DJU (ou DH) et les consommations d'énergie relevées en exploitation. On identifie par régression un modèle de premier ordre, ou plus rarement de second ordre, parfois appelé signature énergétique (Figure 2-3). Cette caractérisation permet d'observer d'éventuelles dérives de consommation dans le temps en s'affranchissant de la responsabilité des variations de la température extérieure.

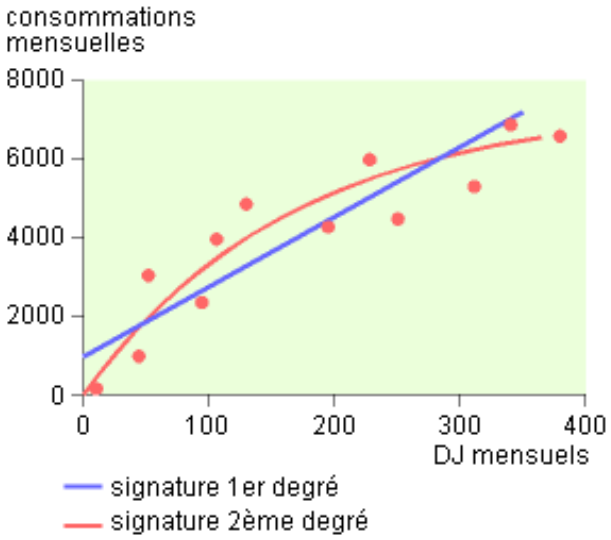

Figure 2-3 : Signature énergétique caractérisant le lien entre consommations d'énergie mensuelles et degré-jours[6]

Cette méthode est assez grossière, le calcul des DJU (ou dans une moindre mesure des DH) entraînant une perte importante d'information sur les variations réelles de la température extérieure et de leurs effets sur le bâtiment. L'impact des apports solaires n'est par ailleurs pas pris en compte ni celui de tous les autres paramètres incertains et variables. Pour aller vers une plus grande précision dans les prévisions et dans l'analyse, il est nécessaire de complexifier les modèles et d'utiliser la simulation énergétique dynamique où les variations temporelles des sollicitations et de l'état du système sont prises en compte.

[6] https://www.energieplus-lesite.be/index.php?id=10016

2.2.1.2. Modélisation des grandeurs météorologiques en simulation énergétique dynamique

Les outils de SED ont été développés dans l'objectif d'étudier le fonctionnement physique du bâtiment en décrivant le plus fidèlement possible ses caractéristiques, son utilisation et son environnement réels. Les données météorologiques utilisées dans les simulations sont donc spécifiques à une position géographique.

Les grandeurs indispensables à l'évaluation du comportement thermique du bâtiment sont la température extérieure et les composantes du rayonnement. D'autres grandeurs, telles la vitesse du vent ou l'humidité relative de l'air, peuvent aussi être utilisées dans les outils de simulation pour plus de précision, dans les calculs aérauliques notamment, ou pour des applications spécifiques.

Ces données météorologiques, décrites sous la forme de séries temporelles, constituent des conditions aux limites au modèle du bâtiment étudié. Une série temporelle (ou série chronologique) à temps discret est une suite finie d'observations de variables aléatoires réelles $(X_t)_{1\leq t\leq n}$, où t représente le temps (en minute, jour, année…). Les outils de simulation nécessitent généralement la spécification de données au pas de temps horaire sur un an de différentes grandeurs climatiques.

Différents formats de fichiers sont utilisés pour ces simulations. Les données les composant proviennent de sources variées et le choix de l'utilisation d'un type de données dépend de l'objectif de l'étude à réaliser. Il sera nécessaire de considérer les données météorologiques réelles, ou AMY (A*ctual Meteorological Year*), dans un objectif d'analyse *a posteriori* du fonctionnement énergétique d'un bâtiment, pour la vérification d'un engagement sur les consommations d'énergie par exemple (OXYGEN). La simulation est ici utilisée afin de représenter le plus finement possible la réalité des sollicitations du bâtiment, monitorées en exploitation, afin d'identifier uniquement les écarts de besoins énergétiques entre résultats réels et simulés, dus aux caractéristiques du bâti. Les données utilisées peuvent provenir de stations météorologiques dédiées présentes directement sur le site ou de stations proches dont les données peuvent être éventuellement corrigées afin de prendre en compte l'éloignement géographique ou la différence d'altitude (Robillart, 2015).

Afin d'étudier la réaction d'un ouvrage ou de systèmes énergétiques à des conditions météorologiques extrêmes, des fichiers météorologiques adaptés doivent être utilisées. Des épisodes climatiques extrêmes passés peuvent servir de référence pour ces études, qu'ils concernent des épisodes hivernaux de froid sévère et prolongé ou des épisodes de canicule. La dynamique des systèmes étudiés est centrale dans ces études et la problématique associée à ces conditions extrêmes est souvent liée à des temps caractéristiques inférieurs à l'année. Les applications principales de ce type d'étude en simulation concernent le dimensionnement de systèmes HVAC ou l'évaluation du confort d'été pendant de fortes chaleurs estivales. Ainsi, dans une étude énergétique menée sur des maisons passives (Peuportier et Thiers, 2009), une période de chaleur extrême correspondant à la canicule de 2003 est simulée afin d'étudier la température intérieure et le

maintien des conditions de confort. Une variante avec l'utilisation d'un puits climatique pré-refroidissant l'air de ventilation est considérée dans ces travaux (Figure 2-4).

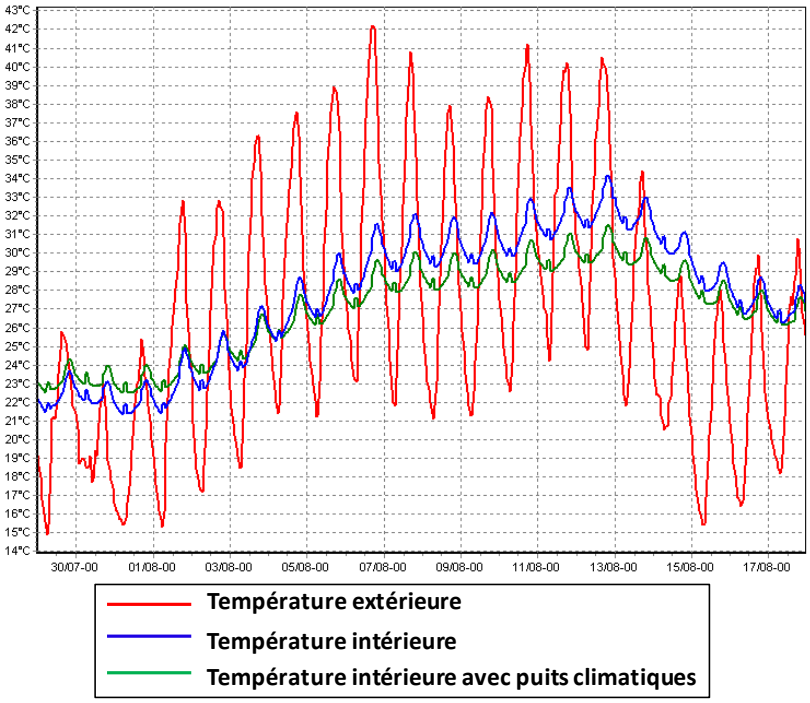

Figure 2-4 : Étude par simulation du confort intérieur d'été avec ou sans puits climatique durant la canicule de 2003 dans les maisons passives de Formerie (Peuportier et Thiers, 2009)

Enfin, les données météorologiques types visent à caractériser le climat typique pour une position géographique donnée. Ce format de données est particulièrement utilisé en phase de conception des ouvrages mais leur utilisation dépasse le cadre de la simulation énergétique des bâtiments (prévision de productions agricoles, de production électrique photovoltaïque, etc.). De nombreux formats de fichiers types ont été développés et la multiplicité des besoins dans différents domaines explique en partie cette diversité. Crawley (1998) et Hong et al. (2013) listent et proposent une étude comparée des formats existants. Ceux-ci incluent les fichiers TRY (T*est Reference Year*), DMY (D*esign Reference Year*), WYEC (W*eather Year for Energy Calculation)*, TMY (*Typical Meteorological Year*). Les données TRY et TMY sont construites à partir de l'étude statistique de plusieurs dizaines d'années réelles. Des mois représentatifs sont sélectionnés parmi les années disponibles et un processus mathématique permet de lier ces mois caractéristiques réels en respectant la logique physique des phénomènes mis en jeu. Différentes méthodes de génération de données types peuvent être trouvées dans la littérature. On pourra citer par exemple la méthode belge développée en 1985 et la méthode Sandia proposée en 1978 dont une comparaison est proposée par Ferrante et al. (2008). La norme européenne EN-ISO 15927-4 décrit aussi une méthode de calcul de ces données types, suivant un principe similaire. Les méthodes de construction de ces fichiers de données ont évolué des années 1970 à nos jours et plusieurs versions de celles-ci existent

pour les différents formats de fichiers (TMY2, TMY3, etc.). Les différences entre ces formats types sont relativement minimes et sont notamment associées au type de données de rayonnement utilisées pour les construire et à des variations dans les processus de sélection (Crawley, 1998). Le logiciel *Pléiades*+COMFIE nécessite par exemple des fichiers sous format TRY pour effectuer les simulations mais d'autres logiciels s'appuient sur leur propre format comme EnergyPlus avec les EPW (*EnergyPlus Weather files*). Les données types reconstruites visent à la fois à respecter une tendance moyenne caractéristique de la position géographique considérée et à représenter les dynamiques d'évolution des grandeurs météorologiques à des échelles de temps courtes. Dans le cadre de la réglementation thermique française de 2012 (RT2012) des fichiers types ont été développés. Ils caractérisent les climats de différentes villes associées à des zones géographiques définies dans cette réglementation (Figure 2-5). C'est sur la base de ces fichiers que sont réalisées les évaluations thermiques du moteur de calcul de la RT2012. Ils ont été construits selon la norme NF-EN-ISO 15927-4. Celle-ci impose un nombre minimal de 10 ans de données à exploiter selon la méthode utilisée pour les fichiers TRY. Ces données météorologiques conventionnelles de référence ont été produites sur la base des observations des années 1994 à 2008.

Figure 2-5. Zones climatiques et villes de référence dans la RT2012[7]

Dans une approche de prévision probabiliste des consommations énergétiques d'un bâtiment, des fichiers artificiels SMY (S*ynthetically modeled Meteorological Year*) peuvent être utilisés.

[7] http://www.rt-batiment.fr/batiments-neufs/reglementation-thermique-2012/donnees-meteorologiques.html

L'objectif n'est pas de créer des données avec une représentativité moyenne mais de générer un ensemble de données représentant chacun un jeu de sollicitations météorologiques possibles et probables. L'avantage de ce type de fichiers réside dans le grand nombre de jeux de données exploitables. En effet, le développement récent des moyens de mesure et de collecte de données ne permet que rarement d'avoir accès à plus d'une quinzaine ou une vingtaine d'années de mesures, ce qui constitue un échantillon statistique limité. Des données artificielles dont la représentativité doit être prouvée permettent de se baser sur un nombre de jeux de données pouvant atteindre plusieurs milliers de configurations pour une même localisation. La pertinence de l'utilisation de ce type de données synthétiques multiples repose sur le réalisme et la représentativité statistique des séries temporelles générées pour la simulation. Une revue bibliographique concernant les méthodes de génération de données synthétiques est présentée au paragraphe 2.2.3 D'autres objectifs et applications rendues possibles par les SMY seront présentés aux paragraphes 2.3 et 2.6.

La modélisation des conditions météorologiques tout comme l'étude de données relevées sur les bâtiments permettent d'étudier et de quantifier l'influence de ces sollicitations sur le fonctionnement énergétique des ouvrages bâtis. Les travaux associés à ces thématiques sont présentés au paragraphe suivant.

2.2.2. L'impact du climat sur les consommations d'énergie des bâtiments

La variabilité au cours des mois des consommations énergétiques liées au chauffage ou au refroidissement est l'indice le plus évident de l'impact des conditions extérieures. Par ailleurs, de nombreuses études s'intéressent aux variations de ces consommations en fonction des positions géographiques. Au-delà de ces considérations évidentes, une variabilité des consommations interannuelle due au climat peut être observée pour un bâtiment donné. Yang et al. (2008) ont comparé ainsi les besoins de chauffage et de refroidissement de bâtiments de bureaux dans cinq villes de Chine à partir de données climatiques relevées sur plusieurs années, observant des variations de 30 à 80 %.

2.2.2.1. Impact de l'évolution des bâtiments

L'impact grandissant des conditions climatiques, et notamment du rayonnement solaire, est lié à une autre évolution du domaine du bâtiment : les maisons passives. Ces bâtiments, très bien isolés et avec une forte inertie, ont des besoins de chauffage réduits voir inexistants. Les apports internes et les apports d'énergie solaire suffisent à compenser les pertes thermiques, mais provoquent en contrepartie des risques de surchauffe estivale. L'influence de l'inertie thermique des bâtiments sur les besoins énergétiques et le confort en lien avec les conditions météorologiques a notamment été

étudié par Munaretto (2014). Concernant les aspects de conception et la géométrie des façades, Rocha et al. (2016) ont étudié la variation de l'ensoleillement sur une surface en fonction des incertitudes sur les paramètres du bâtiment et du dispositif d'ombrage (brise-soleil). Des écarts de l'ordre de 5 % des gains solaires ont été observés pouvant ainsi conduire à des surchauffes mal estimées au sein des bâtiments durant l'été.

2.2.2.2. Prise en compte du changement climatique

L'augmentation moyenne des températures induite par le réchauffement climatique mondial vient elle aussi bouleverser la modélisation du fonctionnement énergétique des bâtiments. Les méthodes basées sur la simulation, liées aux objectifs de prévision future et à long terme des consommations d'énergie, se doivent d'intégrer cette problématique. À partir de différents résultats prospectifs de l'augmentation de température, Chan (2011) réalise une modification de fichiers météorologiques types afin d'y intégrer l'augmentation de température anticipée. Menant une étude basée sur des SED réalisées avec *EnergyPlus*, il constate une augmentation des besoins de climatisation annuels allant jusqu' à 14 % pour les bâtiments de bureaux et 24 % pour les bâtiments résidentiels. Dans une optique d'aide à la conception par analyse de cycle de vie (ACV), Roux (2016) calcule l'influence du réchauffement climatique sur les consommations d'énergie impactant les résultats d'ACV reposant sur un horizon temporel de plusieurs dizaines d'années. Différents scénarios issus des travaux du GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) sont considérés (Roux et al., 2016). Une méthode de morphing permet d'appliquer les projections d'élévation des températures à des données météorologiques types. À l'aide du même modèle, et afin d'étudier la robustesse des solutions de conception identifiés à l'aide d'un algorithme d'optimisation évolutionnaire, Recht (2016) a testé les configurations de conception optimales en prenant en compte cette évolution du climat.

La prise en compte du réchauffement climatique est donc indispensable dans les études s'intéressant à des prévisions à long terme. Ce sujet ne sera cependant pas traité dans ce chapitre bien que le couplage d'un modèle dédié avec le modèle de génération de la variabilité naturelle développé soit possible. En effet, l'échelle temporelle d'un contrat de garantie de résultats énergétiques peut atteindre une dizaine d'années et éventuellement être mis à jour. L'évolution climatique tendancielle est négligeable par rapport à la variabilité naturelle des conditions météorologiques sur une telle période.

2.2.2.3. Îlot de chaleur urbain

Les phénomènes d'îlot de chaleur urbain peuvent aussi avoir un impact sur les écarts entre les consommations réelles et les prévisions. Une augmentation de température localisée dans les zones fortement urbanisées peut ainsi être observée et impacte notamment les besoins de climatisation

dans les bâtiments de bureaux (Chan, 2011 ; Kolokotroni et al., 2012). Des corrections peuvent ainsi être apportées aux données météorologiques utilisées en simulation pour intégrer ce facteur.

La méthode de génération de données synthétiques présentée dans ce chapitre n'inclut pas de modèle spécifique pour traiter de la question des îlots de chaleur urbain. Cependant, ce phénomène pourra être pris en compte soit dans le fichier de données météorologiques types de référence, décrivant précisément la spécificité des conditions météorologiques locales, ou par un post-traitement des données générées pour inclure ces spécificités grâce à une transformation des séries temporelles.

2.2.2.4. Analyse de la variabilité des données météorologiques en SED

Quelle que soit la capacité des données météorologiques types à décrire une évolution annuelle réaliste, leur utilisation constitue l'une des principales limites actuelles des outils de simulation dans un objectif de prévision des consommations énergétiques réelles des bâtiments. Ainsi, Hong et al. (2013) soulignent l'écart entre la moyenne des résultats de consommation d'énergie obtenus par des simulations réalisées avec des données provenant de 30 années réelles et des résultats obtenus par simulation avec des fichiers types au format TMY3. L'étude porte à la fois sur les pics de demande d'électricité dans les bâtiments et sur les besoins énergétiques. Des écarts importants sont observés, de l'ordre de plus ou moins 15 % et pouvant monter jusqu'à 30 % dans quelques configurations, qu'ils expliquent à la fois par la tendance de réchauffement climatique qui n'est pas prise en compte dans le processus de construction de ces données, mais surtout par la variabilité statistique naturelle du climat d'une année sur l'autre. Il constate une tendance à sous-estimer les besoins énergétiques avec l'utilisation de fichiers types. Cette étude menée à large échelle sur des bâtiments de bureaux de performance énergétique variée et sur 17 zones climatiques différentes aux États-Unis, souligne donc l'influence de la variabilité des conditions météorologiques. Par conséquent, elle met en relief les limites de l'utilisation des fichiers météorologiques types en simulation, notamment dans un but de prévision des consommations d'énergie. Wang et al. (2012) ont conduit une étude similaire qui confirme ces conclusions, relevant des fluctuations de l'ordre de -4 à +6 % dans les consommations d'énergie d'un petit bâtiment de bureau obtenues par simulation avec des données météorologiques types et 10 années de données réelles.

2.2.2.5. Estimation de l'influence de la météorologie à partir de données météorologiques synthétiques

À partir de la génération de données météorologiques mensuelles permise par un modèle développé dans le cadre de ses travaux de thèse, Goffart (2013) a réalisé une analyse de sensibilité des besoins de chauffage aux grandeurs climatiques dans une maison individuelle. En étudiant la variance des sorties du modèle en fonction de différents jeux de données météorologiques d'entrée, des indices de Sobol (1993) liés à l'effet de la température extérieure, du rayonnement solaire, de

l'humidité de l'air et du vent, peuvent être calculés à partir d'un plan d'échantillonnage des entrées proposé par Saltelli et al. (Andrea Saltelli et al. 2010). Elle a constaté une influence variable des grandeurs en fonction du bâtiment considéré, du climat étudié et de la grandeur d'intérêt. La variabilité de la température extérieure est l'élément le plus impactant sur les besoins de chauffage, étant responsable de 80 à 90 % de la variance des sorties. Le rayonnement solaire compte pour 10 à 30 % de la variance selon les cas considérés, l'influence augmentant pour des bâtiments bien isolés et à forte inertie. La variabilité observée parmi les grandeurs d'intérêt est supérieure aux résultats de Hong et al. (2013) et Wang et al. (2012), les besoins de chauffage de la maison individuelle étudiée pouvant varier de plus ou moins 40 %.

Cette dernière contribution repose sur l'utilisation de données météorologiques synthétiques multiples pour réaliser une propagation de la variabilité et une analyse de sensibilité et ainsi étudier l'influence du climat sur les consommations d'énergie des bâtiments. Il a été montré que cette approche permettait d'avoir une variabilité statistique des sollicitations qui n'est pas atteignable avec l'utilisation de données réelles provenant de mesures du fait du nombre limité d'années disponibles. Les limites de l'utilisation d'un unique jeu de données typiques, quel que soit le format choisi dans un objectif de prévision des consommations énergétiques, ont clairement été démontrées du fait de l'absence de prise en compte de la variabilité des conditions météorologiques. Les modèles de génération de données artificielles, et leurs domaines de validité et d'application vont maintenant être présentés.

2.2.3. Générer la variabilité climatique

L'utilisation de données météorologiques synthétiques pour la simulation s'est développée parallèlement dans des domaines variés, tel le bâtiment, mais aussi l'agriculture ou la production d'énergie renouvelable. Une approche physico-probabiliste de la prévision des consommations d'énergie des bâtiments constitue une autre utilisation des méthodes de synthèse de données climatiques. L'objectif est de simuler le fonctionnement du bâtiment au sein d'une variété de contextes climatiques et d'en tirer des conclusions statistiques. Différents travaux peuvent être reliés à cette approche méthodologique. La méthode proposée dans la suite de ce chapitre s'inscrit dans cette même démarche. Ainsi, la génération de données synthétiques passe par l'exploitation préalable de données de référence. Cette exploitation consiste en la caractérisation mathématique fine des séries temporelles puis des séries possédant les mêmes caractéristiques mathématiques sont créées. On s'intéresse aux caractéristiques propres des grandeurs étudiées mais aussi à leurs caractéristiques croisées. On parle ainsi d'autocorrélation temporelle et d'intercorrélation entre grandeurs.

Le logiciel *Meteonorm* (Remund et al., 2012a) compile un ensemble de données météorologiques mondiales et propose aussi de générer des données météorologiques associées à une position géographique. Une part importante des modèles utilisés viennent du projet SoDa[8]. La génération de données de rayonnement repose sur les travaux de Aguiar et al. (1988) basé sur l'utilisation de matrices de transition de Markov. Le processus se nourrit aussi de données provenant d'un grand nombre de stations. Une séquence journalière des indices de clarté (2.12) est générée et la variation horaire est calculée à partir de la méthode TAG (*Time dependent, Autoregressive, Gaussian*) proposée par Aguiar et Collares-Pereira (1992). Elle repose sur l'utilisation d'une fonction autoregressive (§ 2.4.1.2, Annexe A-2-1), équation de premier ordre modulant les valeurs journalières moyennes précédemment générées (Remund et al., 2012b). Un modèle autorégressif permet la création de séries de températures moyennes au pas de temps journalier. Les variations horaires sont définies à partir des données de rayonnement solaire selon l'hypothèse que les amplitudes de leurs variations durant la journée sont proportionnelles. Le processus se nourrit aussi de l'exploitation statistique de données de températures observées sur un grand nombre de stations météorologiques (méthodes d'interpolation).

La méthode de génération proposée dans le logiciel *Meteonorm* a été avant tout construite autour de la création de données de rayonnement et c'est pour cette fonctionnalité qu'elle présente les meilleures performances. Le processus de création de données de température est assez simplifié. Remund et al. (2012b) soulignent aussi la limitation du processus dans la représentativité de la variabilité du climat d'une année à l'autre.

Dans ses travaux de thèse, Goffart (2013b) propose un processus de générations de données météorologiques mensuelles à partir de l'exploitation d'un fichier TRY. La variabilité statistique des différents jours d'un même mois est utilisée pour recréer des séries temporelles journalières. Elle fait donc l'hypothèse qu'« un mois est la pseudo-répétition de journées aléatoires par translation », considérant ainsi les jours d'un même mois comme les différentes réalisations statistiques d'une même variable aléatoire : la série horaire des grandeurs météorologiques sur une journée. La méthode proposée prévoit de décomposer ces séries en une part déterministe et une part aléatoire. La part déterministe est définie comme la moyenne horaire sur le mois de la grandeur physique considérée. La part aléatoire est définie par une fonction de covariance (ou autocorrélation) et une fonction de répartition. Ces deux caractéristiques sont extraites à partir de la séquence météorologique mensuelle considérée. La méthode d'Iman et Conover (1982) est utilisée pour re-générer une composante aléatoire à partir des fonctions de covariance et de répartition. Cette composante est ajoutée à la part déterministe identifiée plus tôt. Goffart fait l'hypothèse que seule la part déterministe des grandeurs physiques suffit à décrire l'intercorrélation qui les lie. Les parts aléatoires sont donc créées indépendamment. Dans le cadre de son travail de thèse, l'objectif principal du modèle était de pouvoir réaliser des analyses de sensibilité de résultats de simulation

[8] http://www.soda-pro.com

aux différentes grandeurs météorologiques. Le processus développé étant basé sur l'hypothèse qu'un mois est la pseudo-répétition de journées aléatoires, il n'est valable que sur des mois où les grandeurs météorologiques, notamment la température, sont stables. L'outil ne permet donc pas de conduire des simulations annuelles. L'étude est donc réalisée sur les mois de janvier et de juillet pour les villes de Lyon et Strasbourg. Par ailleurs, l'hypothèse que la part tendancielle suffit à décrire l'intercorrélation ignore la possibilité d'avoir des épisodes météorologiques où les grandeurs dévient conjointement des tendances, ce qui est communément observé dans la réalité. (p ex. une période associant un ensoleillement exceptionnel et des températures au-dessus des normales).

Boland (1995) propose de décomposer une série temporelle de rayonnement solaire existante en une part tendancielle déterministe et une part aléatoire. La série tendancielle est extraite via une transformation de Fourier discrète, équivalent discret d'une décomposition en série de Fourier, équations (2.4) à (2.6). Les fréquences les plus caractéristiques sont celles correspondant à un cycle par an (f_1) et un cycle par jour (f_{365}) décrivant respectivement les variations saisonnières et quotidiennes. La moyenne annuelle (f_0) est elle aussi conservée dans cette série tendancielle. Les signaux de fréquences f_1 et f_0 sont des composantes non négligeables de la décomposition de Fourier et Boland les conserve dans la série identifiée. Après avoir extrait la part tendancielle, la série résiduelle est étudiée et la fonction d'autocovariance est calculée. L'étude du corrélogramme partiel de la série permet d'associer un processus autorégressif (AR) à la série des résidus. Ce type de processus décrit le lien linéaire entre la valeur d'une série à un pas de temps t et sa valeur au pas de temps précédent (il sera plus longuement décrit au paragraphe 2.4.1.2). Boland identifie ce type de processus décrivant la série résiduelle de rayonnement solaire. Des séries aux caractéristiques autorégressives similaires sont générées et ajoutées à la part tendancielle, constituant finalement les séries synthétiques finales de rayonnement.

Boland souligne les limites de la démarche concernant la caractérisation des données de rayonnement solaire. En effet, une normalisation des résidus par les écart-types saisonniers est nécessaire. La stationnarité de la série obtenue, condition indispensable à l'identification d'un processus AR n'est pas assurée. On peut aussi noter que les processus autorégressifs identifiés sont simplifiés. La précision dans la détermination des caractéristiques de ces processus constitue une autre limite de la méthode.

Lee et al. (2012) proposent de modéliser les séries temporelles climatiques à l'aide d'un vecteur autorégressif (VAR), équivalent multidimensionnel du processus AR utilisé par Boland. Cet objet mathématique permet ainsi de prendre en compte l'interdépendance entre les séries. L'exploitation d'une trentaine d'années de données pour chaque phénomène est nécessaire. Ces données sont normalisées via un processus algébrique (transformation de Rosenblatt). Les coefficients du VAR sont identifiés par la méthode des moindres carrés et les caractéristiques (matrices de covariance et moyennes) des résidus du processus sont étudiées. À partir de la génération d'un bruit blanc, l'inversion du processus est réalisée via un traitement de ce bruit blanc puis une transformation

inverse de Rosenblatt afin que les données recréées aient une distribution similaire aux données réelles.

Bien qu'elle permette de générer un grand nombre de phénomènes climatiques de manière simultanée, au pas de temps horaire et avec une bonne représentation de l'intercorrélation, cette méthode repose sur l'exploitation d'un grand nombre de jeux de données en entrée (33 jeux utilisés comprenant la température extérieure, l'humidité, la vitesse du vent, la pression et les composantes du rayonnement). Ces données sont rarement disponibles dans la réalité, pour un usage non expert des outils de SED des bâtiments.

Dans la continuité des travaux précédents, notamment ceux de Boland, Rastogi (2016) a étudié la sensibilité des consommations d'énergie des bâtiments au climat. Ses travaux de thèse présentent un modèle visant à générer des séries temporelles de température, de rayonnement global horizontal et d'humidité relative de l'air. La série des températures extérieures de référence est séparée en une part tendancielle et une part résiduelle via une transformation de Fourier telle que présenté dans les travaux de Boland. Un modèle autorégressif à moyenne mobile avec saisonnalité (SARMA) est identifié à partir de l'étude des résidus. La prise en compte d'une saisonnalité quotidienne (lien entre la température à une heure donnée et la température observée la veille à la même heure) permet d'affiner la caractérisation de l'autocorrélation de la série. Rastogi exploite le bruit issu de l'identification du modèle SARMA, qui est un résidu de second niveau (écart entre les résidus et le modèle). Cette série est séparée en blocs de 3 jours et ré-échantillonnée afin de recréer d'autres séries résiduelles sur lesquelles sont appliquées le modèle SARMA afin de reconstruire des résidus similaires à celui de référence. Cette série résiduelle synthétique est finalement ajoutée à la part tendancielle produisant la série artificielle finale de température. La Figure 2-6, issue de ces travaux récapitule la méthode de génération des données de température extérieures.

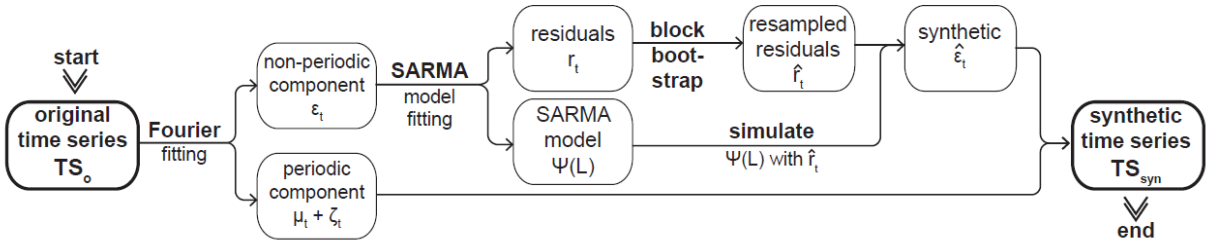

Figure 2-6. Schéma de principe de la méthode de génération proposée par Rastogi (2016)

La création de données de rayonnement obéit à un autre processus. Rastogi étudie la relation dans les données de référence entre la moyenne journalière des températures et la somme quotidienne du rayonnement global. Après avoir généré une série de température, la moyenne journalière des températures est calculée et rapprochée d'une journée réelle du même mois. Les données réelles de rayonnement solaire de cette journée sont finalement associées aux données de

température générée. Cette méthode consiste donc en une redistribution des données réelles de rayonnement au sein de chaque mois pour chaque jeu de données créé.

Les travaux de Rastogi sont originaux sur deux aspects. Le modèle n'a besoin que d'un jeu de données types comme référence et le processus de génération est orchestré par la série de température selon un processus asymétrique. Bien que les séries de température soient générées de manière fine et robuste, on constate que les autres séries, notamment celles de rayonnement, sont créées selon un processus plus grossier, avec une variabilité limitée et moins réaliste. Les niveaux d'intercorrélation entre la température et le rayonnement parmi les données générées sont significativement inférieurs à ceux des données de référence, les coefficients de corrélation de Pearson étant respectivement de 0,43 pour la moyenne des données générées et de 0,52 pour les données originales.

2.3. Objectifs du modèle développé

Le modèle développé doit permettre la génération de séries temporelles horaires annuelles de données météorologiques comprenant la température extérieure, le rayonnement solaire global horizontal et le rayonnement solaire diffus horizontal. Ces grandeurs correspondent aux données météorologiques minimales nécessaires à la réalisation de simulations énergétiques dynamiques.

Les objectifs liés au développement de ce modèle sont présentés par la suite. Ceux-ci sont d'ordre méthodologique et visent aussi à dépasser les limites identifiés dans les modèles de la littérature. Les apports scientifiques associés à ce modèle seront aussi détaillés.

2.3.1. Propagation des effets de la variabilité climatique dans les modèles SED

Le développement d'un modèle de génération de données météorologiques vise tout d'abord à compléter le jeu de méthodes nourrissant l'approche physico-probabiliste des simulations énergétiques du bâtiment. Celles–ci se composent en premier lieu des propagations d'incertitudes sur les paramètres statiques, présentées au chapitre 1 mais aussi de modèles spécifiques décrivant la variabilité du comportement des occupants. Ces modèles stochastiques reposant sur des études sociologiques statistiques seront présentés au chapitre 3. L'approche physico-probabiliste consiste donc à simuler un grand nombre de configurations de bâtiments couplés à une variabilité de sollicitations afin d'étudier la distribution des résultats de consommation d'énergie en sortie du modèle. L'objectif est de tendre vers une exhaustivité dans la prise en compte des configurations possibles. Le modèle développé se propose de fournir la variabilité des conditions climatiques

exploitées dans cette approche. Pour décrire l'utilisation couplée de ces différentes méthodes, on parlera ainsi de propagation des incertitudes et variabilités.

La prise en compte de la variabilité du climat dans les méthodes de propagation d'incertitudes (ou analyse d'incertitudes) est abordée dans plusieurs études grâce à la définition d'une variation constante des grandeurs physiques autour des séries de référence (Munaretto, 2014 ; Recht, 2016 ; Caucheteux et al. 2018) Cette variation peut être absolue ou relative. L'absence de réalité physique associée à cette approche et le choix arbitraire des valeurs retenues soulignent les limites de cette méthode et la pertinence d'utiliser une variabilité réaliste des grandeurs météorologiques.

Les difficultés identifiées pour atteindre ce premier objectif se rapprochent de celles propres aux méthodes de propagation d'incertitudes des paramètres statiques des modèles. La pertinence des résultats obtenus suite à une propagation d'incertitudes, et les caractéristiques de la fonction de densité de probabilité sur la sortie du modèle, dépendent fortement de la spécification des entrées de cette méthode. En effet, la définition des lois de distribution sur les entrées incertaines, que ce soit leurs types et leurs paramètres, est un exercice difficile reposant sur une connaissance et des données souvent inaccessibles. Un parallèle peut être réalisé concernant la propagation de la variabilité climatique en SED. La forme des fonctions de densité de probabilité des paramètres statiques équivaut à la fois à la « dispersion » des séries générées et aux bornes de cette dispersion. Cette délimitation des valeurs des entrées dynamiques doit être réfléchie en termes de valeurs ponctuelles mais aussi de valeurs moyennes et agrégées. Il en va de même pour les distributions probabilistes, la dispersion des grandeurs générées doit être représentative de ces multiples échelles. Le modèle de génération de données météorologiques doit donc respecter une représentativité probabiliste (lois et borne des distributions) la plus réaliste possible à la fois pour des valeurs de pointe, maximales ou minimales, annuelles et mensuelles, que pour des valeurs moyennes. Le parallèle réalisé avec la propagation d'incertitudes fait donc état d'une complexité encore plus grande pour la définition des entrées à propager dans le modèle.

Concernant cette représentativité probabiliste des données générées, la validation du modèle de génération fait face au problème du nombre restreint de données réelles disponibles limitant l'analyse de leur variabilité statistique. Cette limitation doit donc être connue et assumée et amène à interroger l'utilisation de ce type de modèle au regard des objectifs des études réalisées.

2.3.2. Apprentissage pour l'entraînement de modèles d'ajustements

Le développement de cet outil de génération de données climatiques vise aussi un autre objectif s'inscrivant dans le contexte plus global de ces travaux de thèse. On souhaite développer une méthodologie prévoyant un ajustement des prévisions réalisées par simulation en fonction de grandeurs explicatives qui pourront être mesurées pendant l'exploitation du bâtiment. Ce modèle d'ajustement doit décrire le lien entre la grandeur d'intérêt d'une part, telle la consommation

d'énergie annuelle, ou un seuil de consommation maximal garanti pour un risque de dépassement donné, et des variables, dites facteurs d'ajustement, d'autre part. Certaines de ces variables pourront être liées aux grandeurs météorologiques, telle que la température extérieure. Afin de construire ce type de modèle d'ajustement, il est donc nécessaire d'étudier ces liens sur le plus grand nombre de configurations possibles. L'outil de génération de données météorologiques permet donc d'obtenir la variabilité statistique nécessaire pour entraîner les modèles d'ajustement.

Cette approche de construction de modèle boîte noire par exploitation de données obtenues grâce à un modèle détaillé préalablement validé est décrit par Tian et al. (2015) et Coakley et al. (2014b). Elle est en fort développement dans le secteur du *big data*, notamment dans les problématiques liées à l'apprentissage artificiel. De nombreuses familles de modèles, paramétriques et non paramétriques, sont présentées dans ces travaux, tels les modèles de régression, les réseaux de neurones, les vecteurs et processus gaussiens, etc.

Contrairement au premier objectif, on ne s'intéresse donc plus à l'influence absolue des conditions météorologiques sur les consommations mais à la différentielle de cette influence, à la variation de consommations énergétiques associée à une variation des grandeurs climatiques. Ainsi, les questions du réalisme de la répartition probabiliste et du bornage des séries temporelles générées ne sont plus problématiques dans cet objectif, sous l'hypothèse d'une faible non linéarité des effets étudiés.

2.3.3. Limitations et apports scientifiques

Les problématiques identifiées dans l'atteinte des objectifs associés au développement du modèle de génération sont les suivantes :

- réalisme de la répartition probabiliste des données temporelles créées, en termes de forme et de bornage de la répartition, en vue d'un objectif de propagation de la variabilité pour l'estimation des consommations énergétiques des bâtiments ;
- difficulté pour la validation du modèle due au nombre limité de données réelles disponibles et au caractère stochastique des grandeurs décrites.

Étant données les limitations des modèles existants liés aux travaux présentés au paragraphe 2.2.3, les apports scientifiques visés et les intérêts du modèle sont de plusieurs ordres. On souhaite pouvoir générer des séries temporelles horaires de température, de rayonnement global horizontal et de rayonnement diffus horizontal :

- sur une **année entière**, période classique pour la simulation dynamique ;
- à partir d'un **unique fichier type** (format TRY) caractéristique d'une position géographique donnée, et disponible dans les outils de SED ;

- représentant les caractéristiques d'autocorrélation et d'intercorrélation des grandeurs décrites (on considèrera à la fois l'intercorrélation tendancielle moyenne mais aussi **l'intercorrélation résiduelle** caractérisant le lien entre les grandeurs pendant les épisodes météorologiques s'écartant des normales saisonnières) ;
- décrivant de manière réaliste la **variabilité de l'ensemble des grandeurs** physiques grâce à l'analyse préalable d'un échantillon de données d'années passées réelles, permettant de caractériser la variabilité des parts tendancielles « déterministes » des séries.

D'un point de vue technique, le modèle, lié à l'utilisation du logiciel *Pléiades*+COMFIE, vient compléter un ensemble de travaux réalisés dans l'équipe de recherche sur les propagations d'incertitudes et de variabilités (Munaretto, 2010 ; Recht et al., 2014 ; Munaretto et al., 2017) et sa mise en œuvre vient intégrer la plateforme rassemblant les outils associés à ces développements. L'ensemble des entrées statiques et dynamiques, incertaines ou variables, peuvent ainsi être décrites selon une approche physico-probabiliste.

2.4. Méthodologie pour la génération de données météorologiques

2.4.1. Génération de données de température

La température extérieure est la grandeur physique météorologique ayant le plus d'influence sur les besoins énergétiques des bâtiments. En effet, les échanges thermiques par conduction et convection à travers les murs et les surfaces vitrées, mais aussi les pertes par infiltration et renouvellement d'air, constituent les principales sources de déperdition d'énergie calorifique et sont directement reliées à la température de l'air extérieur.

Un des objectifs de l'étude des séries temporelles est de décrire, expliquer et prévoir l'évolution au cours du temps d'un phénomène et c'est au sein de ce triple objectif que se positionne l'outil de génération de données. On va dans un premier temps chercher à générer des séries temporelles de données de température extérieure en s'appuyant sur différentes méthodes reliées à l'étude de traitements de signaux. La série temporelle X_{t} est décrite comme la résultante d'une composante tendancielle déterministe Z_{t} et d'une composante résiduelle ε_t :

$$X_t = Z_t + \varepsilon_t \tag{2.3}$$

En s'appuyant sur les travaux de Boland (1995) on propose de décomposer la série temporelle du fichier TRY en une série tendancielle moyenne, $T_{h.Fourier}$ et une série résiduelle, $T_{h.res}$ représentant la différence entre la série originale et cette série moyenne.

$$T_{h.TRY} = T_{h.Fourier} + T_{h.res} \quad \textbf{(2.4)}$$

Les résultats et figures présentées par la suite illustrent l'utilisation de la méthode à partir du fichier type Mâcon_h1cRT2012.try, composé de données caractéristiques de la ville de Mâcon. Il a été développé pour les besoins de la réglementation thermique RT 2012 pour caractériser la zone géographique H1c et est construit à partir des données météorologiques des années 1994 à 2008.

2.4.1.1. Identification de la série tendancielle

L'identification de la série tendancielle est réalisée par décomposition en série de Fourier de la série initiale. Dans le cas de séries temporelles, on parle de transformée de Fourier discrète. Pour un signal *s* de *N* échantillons, elle est définie comme un ensemble de coefficients $S(k)$ calculés à partir de l'équation suivante :

$$S(k) = \sum_{n=0}^{N-1} s(n).e^{\frac{-2i\pi kn}{N}} \quad \textbf{(2.5)}$$

pour $0 \leq k < N,\ k \in \mathbb{Z}$.

La fonction s peut être approximée à partir des coefficients suivants :

$$s(n) = \frac{1}{N} \sum_{k=0}^{N-1} S(k).e^{\frac{2i\pi nk}{N}} \quad \textbf{(2.6)}$$

pour $0 \leq n < N,\ n \in \mathbb{Z}$.

On définit C_k, l'amplitude de la sinusoïde associée à la fréquence f_k :

$$C_k = \sqrt{\mathrm{Re}\big(S(k)\big)^2 + \mathrm{Im}\big(S(k)\big)^2} \quad \textbf{(2.7)}$$

L'analyse de Fourier fait ressortir des fréquences possédant une amplitude significative, identifiées à partir de la valeur de C_k. La densité spectrale d'énergie du signal décrit les valeurs de cette norme au carré en fonction des fréquences (Figure 2-7).

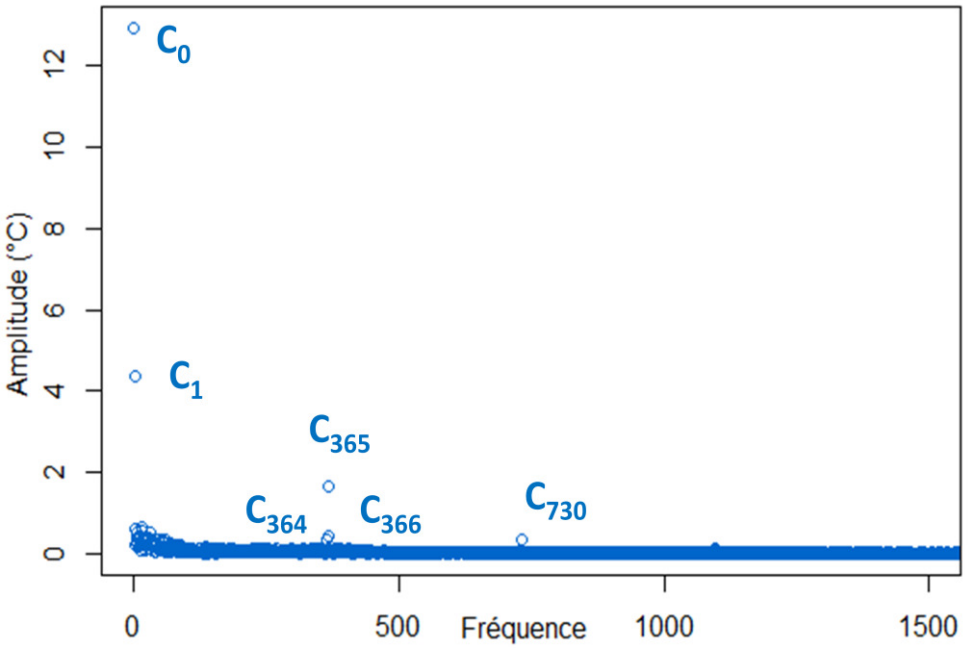

Figure 2-7. Diagramme Amplitude-Fréquence de la décomposition en série de Fourier de la série de température extérieure (Mâcon)

Les composantes caractéristiques de cette série temporelle sont celles de fréquence 0, correspondant à la température moyenne annuelle, de fréquence 1, décrivant les variations saisonnières moyennes et notamment l'écart entre l'hiver et l'été, et enfin le mode 365, caractéristique des variations journalières. En complément, les modes de fréquence 364, 366 et 730 sont retenus du fait de l'amplitude des signaux associés. Les modes de fréquence 364 et 366 permettent de moduler l'amplitude et le déphasage des variations journalières en fonction des saisons. L'amplitude de ces variations est supérieure en été. Le mode de fréquence 730 modifie la forme des pointes et creux de ces variations journalières, décrivant leur évolution moyenne avec plus de réalisme. D'autres modes ont des amplitudes significatives dans le diagramme précédent mais une étude complémentaire, présentée en annexe A-2-2, montre, par comparaison aux résultats de décompositions d'autres séries provenant d'années réelles, que ces modes ne sont pas caractéristiques d'une tendance générale. La série temporelle reconstruite $T_{h.Fourier}$ comme la somme des différents signaux sélectionnés issus de la décomposition de Fourier, décrit un comportement moyenné de la température extérieure, un comportement tendanciel dit de « saison ».

2.4.1.2. Caractérisation de la série résiduelle

La série résiduelle décrit ainsi l'écart de la série initiale à cette température tendancielle. Elle possède une structure d'autocorrélation pouvant être représentée par un modèle ARMA(*p*,*q*), couplage d'un processus autorégressif et d'un processus à moyenne mobile. Ces modèles ne

s'appliquent que pour des séries stationnaires, cette hypothèse étant vérifiée pour la série résiduelle étudiée (annexe A-2-1). La valeur de la série au temps *t*, Z_t, est décrite comme la combinaison linéaire, pondérée par les paramètres φ_i et θ_i , d'un bruit blanc ε_t centré et de variance σ², des valeurs de ce bruit blanc aux *q* pas de temps précédents et des *p* valeurs aux pas de temps précédents des éléments de la série :

$$Z_t - \varphi_1 Z_{t-1} - \cdots - \varphi_p Z_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \quad \textbf{(2.8)}$$

Une corrélation saisonnière est aussi identifiée dans la série résiduelle. Pour traiter cet effet, on peut modéliser le résidu par un processus ARMA qui aura comme unité de temps la période de la saisonnalité (les ordres de ce modèle seront alors notés *P* et *Q*). On notera le processus final : SARMA*(p,q)(P,Q).* La prise en compte de cette saisonnalité permet une caractérisation plus fine des séries étudiées (Rastogi, 2016b) en décrivant toutes les logiques temporelles intrinsèques de la série résiduelle. La série résiduelle privée de ces logiques autocorrélatives doit être identifiable à un bruit blanc.

Les ordres *p* et *q* du modèle SARMA ainsi que les coefficients φ_i et θ_i peuvent être déterminés suivant différents critères et tests statistiques (Robillart, 2015). Ces critères visent à réaliser un compromis entre la fidélité de la description fournie par le modèle et sa complexité. Les critères AIC (*Akaikes's Information Criterion*) et BIC (*Bayesion Information Criterion*) sont définis à partir d'une fonction de vraisemblance et permettent de sélectionner les modèles maximisant la vraisemblance en pénalisant la complexité de ces modèles. Le critère AIC sera privilégié si l'objectif de l'étude est de faire de la prévision et le critère BIC est à préférer si l'objectif est de s'ajuster au mieux à une série observée. Le détail mathématique de ces outils est présenté en Annexe A-2-1. L'estimation correcte du processus d'autocorrélation lié à une série donnée se caractérise par des résidus assimilables à un bruit blanc. Des tests statistiques permettent de vérifier cette condition. Ces tests sont aussi décrits en Annexe A-2-1. L'identification du modèle SARMA est donc rendue possible par l'utilisation des critères de maximum de vraisemblance et des tests de bruits blancs.

Concernant les données étudiées, un modèle ARMA(5,3) permet de décrire avec le plus de pertinence l'autocorrélation horaire de la série résiduelle des températures pour le fichier TRY de Mâcon. Une structure corrélative de période 24 heures est aussi identifiée parmi les résidus traités permettant de définir la valeur *P = 1*. On identifie donc finalement un processus SARMA(5, 3)(1,0).

2.4.1.3. Génération de séries résiduelles

Dans une approche inverse, l'application du modèle SARMA identifié sur différents bruits blancs générés aléatoirement permet de recréer des séries temporelles satisfaisant à ce modèle et possédant les mêmes caractéristiques d'autocorrélation que la série résiduelle initiale $T_{h.res}$. Les séries horaires de température extérieure synthétisées $T^i_{h.gen}$ sont donc finalement constituées de la

somme de la série de Fourier $T_{h.Fourier}$ et de séries résiduelles autocorrélées $T^{i}_{h.res}$ construites avec le modèle SARMA. La Figure 2-8 résume les processus d'analyse de la série de température de référence et de génération de données synthétiques de température extérieure.

Cette première phase du processus de génération est similaire à celle incluse dans les travaux de Boland (1995) et Rastogi (2016b). Nous verrons cependant au paragraphe 2.4.4 qu'une variation des composants de la part tendancielle (composantes de Fourier) sera proposée amenant à une diversité des données plus représentative de la réalité.

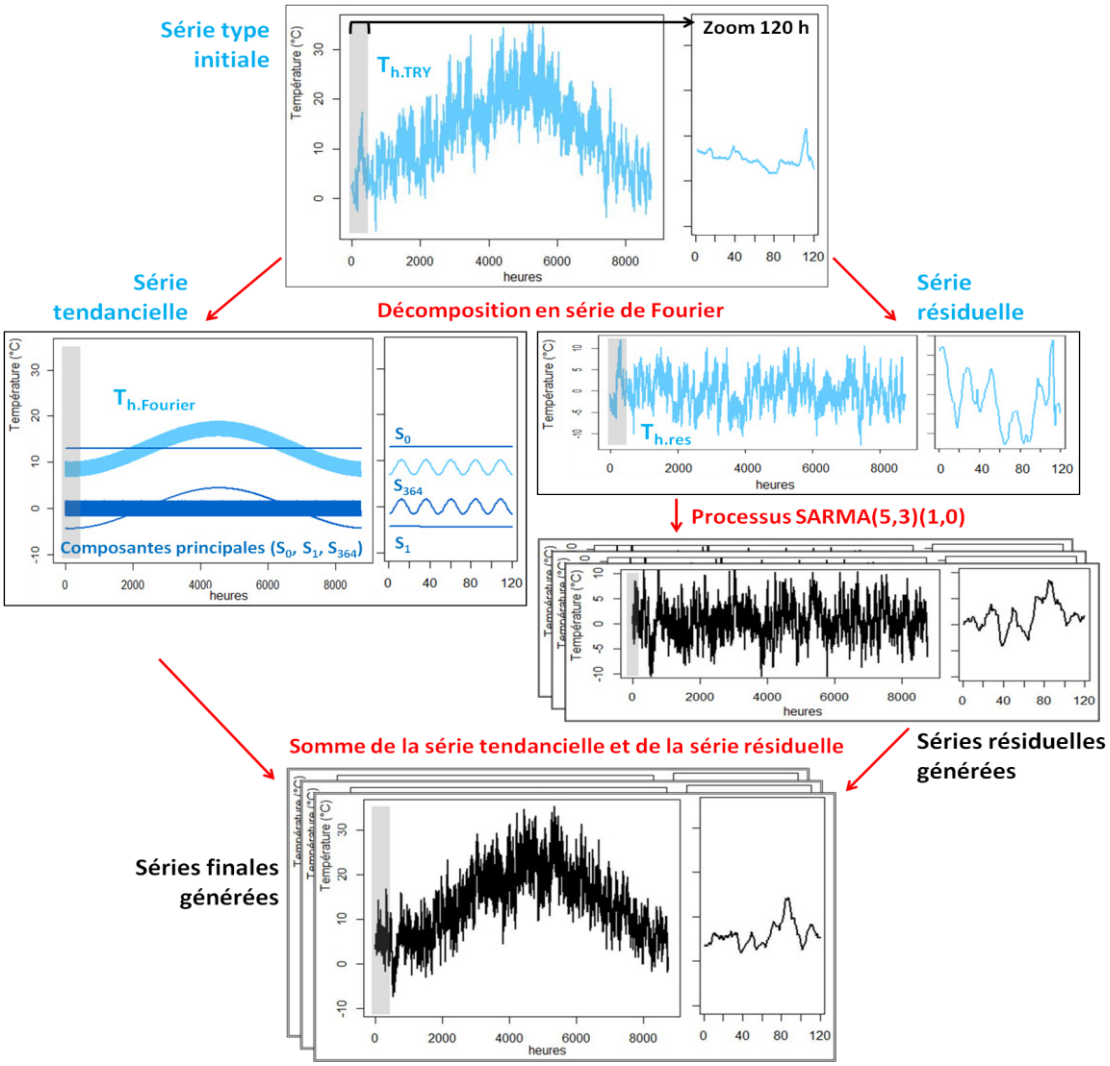

Figure 2-8. Génération des séries de températures extérieures. Séries annuelles et focus sur les cinq premiers jours de l'année.

2.4.2. Génération des données de rayonnement global horizontal

On souhaite générer des données de rayonnement global horizontal (RGH) qui soient réalistes et qui respectent la corrélation naturellement observée avec les données de température. Les séries de rayonnement possèdent quelques particularités : elles sont toujours positives, et nulles durant la nuit.

2.4.2.1. Identification de la série tendancielle

On décompose le signal discret issu du fichier TRY en série de Fourier en gardant là aussi les modes caractéristiques tels que présentés pour les données de température. Les trois modes principaux restent ceux de fréquence f_0, f_1, f_{365} correspondant à 0, 1 et 365 cycles/an. Du fait de la forme plus irrégulière de la série et de la nullité des données la nuit, un plus grand nombre de fréquence voisine et d'harmoniques des modes principaux sont conservées pour caractériser la série tendancielle. En effet, l'amplitude C_k de ces composantes est significative dans la décomposition de Fourier (Figure 2-9). On conserve ainsi les modes de fréquence f_{364}, f_{366}, f_{367}, f_{729}, f_{730}, f_{731}, f_{732}, f_{733}, f_{1092}, f_{1093}, f_{1094}, f_{1095}, f_{1096} et f_{1097}. L'intégration de ces composantes à la série tendancielle est rendue nécessaire par la nécessité d'obtenir une série résiduelle stationnaire pouvant être modélisé par des processus autorégressifs.

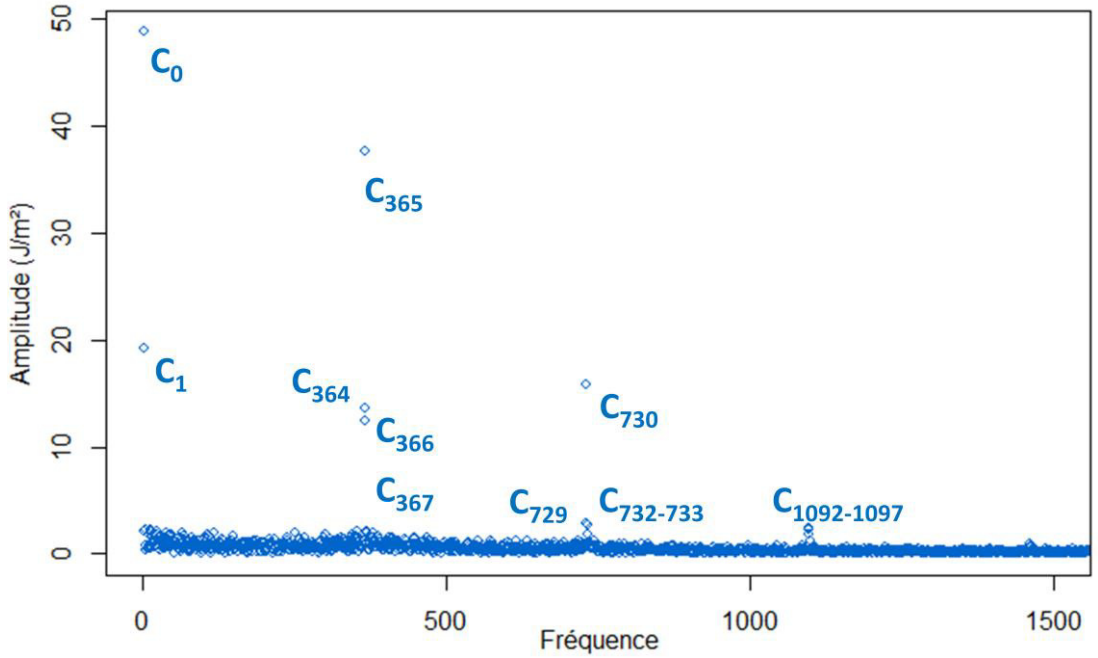

Figure 2-9. Diagramme Amplitude-Fréquence de la décomposition en série de Fourier de la série de rayonnement global horizontal

2.4.2.2. Caractérisation de la série résiduelle

La série résiduelle horaire de rayonnement global horizontal $RGH_{h.res}$ est définie comme la différence entre les données initiales $RGH_{h.TRY}$ et la série tendancielle obtenue par décomposition en série de Fourier $RGH_{h.Fourier}$. Comme précédemment, on cherche à caractériser le résidu issu des données du fichier type en s'intéressant cette fois à ses caractéristiques d'autocorrélation mais aussi à son intercorrélation avec la série résiduelle de température extérieure. Enfin, on souhaite re-générer des résidus de rayonnement solaire global aux caractéristiques similaires, et couplés aux données de température conjointement créées.

La série $RGH_{h.res}$, dont une figure est disponible en Annexe A-2-3, ne peut cependant pas être directement associée à un processus de type ARMA. En effet, plusieurs problématiques sont identifiées :

- Les résidus étant nuls pendant les périodes nocturnes, la série n'est pas stationnaire, empêchant l'identification d'une structure autocorrélative au pas de temps horaire.
- La valeur absolue des résidus en milieu de journée est en moyenne supérieure à celle observée en début et fin de journée.
- L'intercorrélation entre les parts aléatoires (résidus) de température et de rayonnement solaire n'a de sens que durant la journée.

Ces différents facteurs soulignent la difficulté du problème et expliquent les choix simplificateurs réalisés dans les modèles de Boland (Boland 1995) ou Rastogi (Rastogi 2016b). On désire cependant pouvoir générer une grande variabilité de données de rayonnement global horizontal respectant l'autocorrélation temporelle naturelle et l'intercorrélation avec la température extérieure, à la fois tendancielle mais aussi au niveau des résidus.

Afin de s'affranchir des problématiques présentées ci-dessus, on propose une démarche originale consistant à changer d'échelle temporelle pour caractériser l'autocorrélation au pas de temps journalier des résidus de RGH et le lien avec les résidus de température extérieure, puis à re-moduler ce signal suivant une évolution horaire caractéristique.

2.4.2.3. Moyenne journalière et normalisation

La méthode présentée en détail ci-dessous est résumée dans la Figure 2-10 et illustrée à l'annexe A-2-3. À partir de la série résiduelle initiale $RGH_{h.E-T}$, on calcule la série des moyennes journalières des résidus de RGH, nommée $RGH_{j.res}$. Cette série se caractérise par une évolution saisonnière irrégulière, une absence de stationnarité, ses résidus journaliers moyens ayant tendance à être supérieurs durant la période estivale. Afin d'identifier un processus de type ARMA, il est nécessaire de normaliser cette série. La série horaire de rayonnement solaire extraterrestre théorique $RGH_{h.E-T}$ est calculée. Caractéristique d'une position géographique donnée, elle décrit la

quantité de rayonnement théorique atteignant l'atmosphère. Le détail du calcul de cette série est présenté en Annexe C. En soustrayant la série de Fourier $RGH_{h.Fourier}$ à la série du RGH extraterrestre ($RGH_{h.E-T}$), la série des résidus horaires maximaux théoriques, $RGH_{h.res.max}$, et la série des moyennes journalières des résidus horaires maximaux théoriques, $RGH_{j.res.max}$, sont obtenues. Cette dernière permet de normaliser la série des moyennes journalières des résidus, $RGH_{j.res}$, à laquelle on s'intéresse. La série normalisée $RGH_{j.res.norm}$ possède finalement les caractéristiques souhaitées en terme de stationnarité. Un descriptif complet des figures des séries correspondant aux différentes étapes de la méthode est présent en annexe A-2-3.

2.4.2.4. Identification des coefficients d'auto- et d'intercorrélation

Un vecteur autorégressif (VAR) décrit le lien entre deux séries temporelles, caractérisant à la fois l'intercorrélation entre ces séries et leur autocorrélation propre. L'équation (2.9) décrit ainsi un processus VAR(p) associé à un vecteur composé de deux séries temporelles X_1 et X_2 :

$$\begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} = A_0 . \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} + A_1 . \begin{bmatrix} X_1(t-1) \\ X_2(t-2) \end{bmatrix} + A_p . \begin{bmatrix} X_1(t-p) \\ X_2(t-p) \end{bmatrix} + \cdots + \begin{bmatrix} \varepsilon_T(t) \\ \varepsilon_{RGH}(t) \end{bmatrix} \quad \textbf{(2.9)}$$

avec les matrices de coefficient :

$$A_p = \begin{bmatrix} \rho_{p-1} & \rho_{p-12} \\ \rho_{p-21} & \rho_{p-2} \end{bmatrix} \quad \textbf{(2.10)}$$

La méthodologie présentée vise à créer des données de rayonnement associées aux données de température précédemment générées. On s'intéresse donc à l'équation suivante associée à un processus VAR(1) où l'on s'intéresse au rayonnement global horizontal:

$$RGH_{j.res.norm}(t) = \rho_{1-2} . RGH_{j.res.norm}(t-1) + \rho_{0-12} . T_{j.res}(t) + \varepsilon_{RGH}(t) \quad \textbf{(2.11)}$$

L'identification des coefficients ρ_{1-2} et ρ_{0-12} est réalisée par un processus itératif dans le but de respecter l'intercorrélation température / rayonnement des séries résiduelles horaires reconstruites en fin de processus, $T_{h.res}$ et $RGH_{h.res}$. La variable considérée dans la boucle itérative est le coefficient d'intercorrélation entre les deux séries : ρ_{12}. Pour une valeur de ρ_{12} donnée, le coefficient ρ_2 est identifié afin de respecter l'autocorrélation souhaitée en fin de processus. Cette étape permet donc de recréer une série $RGH_{j.res.norm}$ associée à une série précédemment générée de résidus journaliers moyens de température, $T_{j.res}$.

2.4.2.5. Dé-normalisation et modulation horaire

Cette série est ensuite dénormalisée par multiplication terme à terme avec la série utilisée pour la normalisation, $RGH_{j.res.max}$. Cette nouvelle série horaire dénormalisée est ensuite modulée par la

série normalisée des résidus théoriques horaires maximaux obtenus à partir du rayonnement extraterrestre (rapport de $RGH_{h.E-T}$ sur $RGH_{j.E-T}$). On obtient ainsi la série horaire $RGH_{h.res}$ souhaitée. L'intercorrélation entre cette série et celle des résidus de température $T_{h.res}$ est étudiée et comparée à celle observée au sein des données initiales provenant du fichier TRY. La similitude entre ces corrélations est obtenue par le processus itératif optimisant le coefficient du VAR. Le processus complet est décrit à la Figure 2-10 et l'annexe A-2-3 présente des figures des différentes séries intermédiaires obtenues.

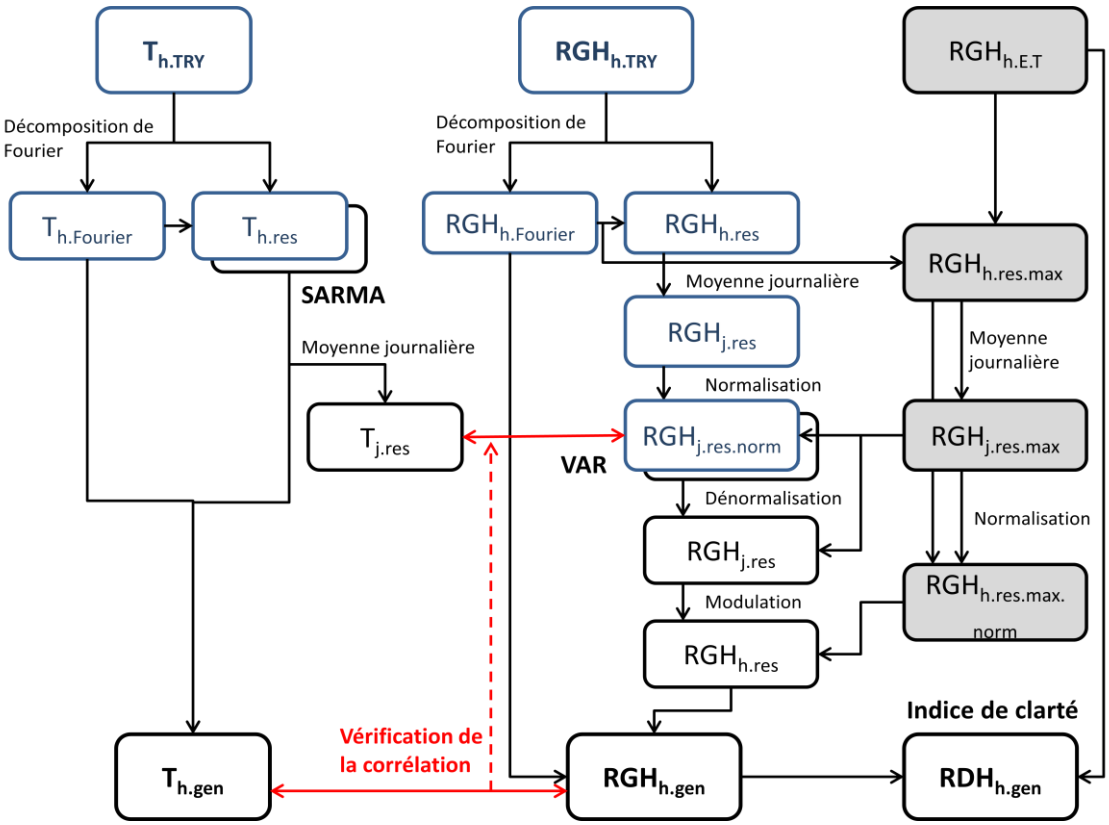

Figure 2-10. Génération des séries de rayonnement solaire

La méthodologie originale présentée ci-dessus permet de créer une variabilité de données de rayonnement solaire, intercorrélées avec une série des températures extérieures précédemment générée. L'intercorrélation est respectée au niveau des parts tendancielles des séries mais aussi dans les parts résiduelles à l'échelle de la journée. Ce respect de l'intercorrélation va de pair avec une dégradation de l'autocorrélation horaire des séries de rayonnement global horizontal. En effet, la part résiduelle n'est correctement corrélée qu'en terme de moyenne journalière. Ce compromis permet cependant une description plus fine des séries temporelles étudiées allant au-delà de la seule considération du lien entre les parts tendancielles.

2.4.3. Génération des données de rayonnement diffus horizontal

Les simulations énergétiques dynamiques nécessitent aussi en entrée la série temporelle horaire du rayonnement diffus horizontal (RDH). Le RDH est physiquement lié au RGH et à l'indice de clarté k_t défini comme le rapport entre le rayonnement global horizontal et le rayonnement global horizontal extraterrestre :

$$k_t = \frac{RGH_h}{RGH_{h.E.T}} \tag{2.12}$$

Cette dernière série décrit la variation temporelle au pas de temps horaire du rayonnement solaire atteignant l'atmosphère pour une position géographique donnée. Le calcul de cette quantité est détaillé dans l'annexe A-2-4. Il est donc possible de recalculer la série de rayonnement diffus horizontal RDH_h à partir de RGH_h d'après la corrélation de Orgill et Hollands (1977) présentée en (2.13). On constate par exemple que les données de référence des fichiers types suivent assez fidèlement cette corrélation.

$$\frac{RDH_h}{RGH_h} = \begin{cases} 1{,}0 - 0{,}249.k_t & si\ k_t < 0{,}53 \\ 1{,}557 - 1{,}84.k_t & si\ 0{,}35 < k_t < 0{,}75 \\ 0{,}177 & si\ k_t > 0{,}75 \end{cases} \tag{2.13}$$

Les limites décrites précédemment dans la création de données de rayonnement global au niveau de la variabilité résiduelle horaire sont aussi vraies pour la série de RDH_h. Cependant, les données sont cohérentes entre elles de par l'utilisation de la corrélation les reliant. L'utilisation de ce modèle déterministe permettant de recréer la série de RDH_h à partir du RGH_h vient donc clore le processus complet de génération de données climatiques.

2.4.4. Variation des parts tendancielles

En étudiant les décompositions en série de Fourier de données de températures provenant d'une dizaine d'années de mesures (2007-2016) réalisées sur le même lieu, on constate une dispersion des différents coefficients de Fourier. On propose donc d'intégrer au processus de génération une variation des coefficients de Fourier de la série tendancielle, grâce à une modification des valeurs obtenues à partir du fichier type, à chaque génération de jeux de données afin de mieux décrire cette possible variabilité de la part tendancielle.

2.4.4.1. Variabilité des coefficients

On s'intéresse uniquement à la variabilité des trois coefficients principaux C_0, C_1 et C_{365} associés à la moyenne annuelle et aux périodes quotidienne et annuelle, du fait de la faible amplitude des autres coefficients retenus suite aux décompositions de Fourier des séries de température et de RGH. L'étude de 10 ans de données annuelles réelles (2007-2016) pour la ville de Mâcon et des séries de Fourier associées permet de caractériser la variabilité de ces coefficients (Tableau 2-1).

Tableau 2-1 : Variabilité des coefficients des transformées de Fourier discrètes identifiées à partir de données réelles de température extérieure à Mâcon (2007-2016)

	Données types	***Données réelles (2007-2016)***			
	H1c_RT2012	***Moyenne***	*Écart-type*	*Min*	*Max*
C_0 (°C)	***12,9***	***12***	*0,67*	*11,1*	*13*
C_1 (°C)	***4,3***	***4,5***	*0,42*	*3,8*	*5*
C_{365} (°C)	***1,6***	***1,7***	*0,11*	*1,5*	*1,8*

On constate ainsi une dispersion de la moyenne annuelle, premier coefficient C_0 de la décomposition de Fourier, entre 11,1 et 13,0 °C avec une moyenne de 12,0 °C et un écart-type de 0,67 °C. Cette valeur moyenne des C_0 est inférieure à la moyenne des températures des données du fichier type (12,9°C) mais supérieure à la valeur moyenne de données types plus anciennes (10,6 °C) reconstruites à partir de mesures réalisées entre 1980 et 1990. En ce qui concerne C_1 et C_{365}, coefficients de Fourier pour les modes de période annuelle et journalière, les valeurs identifiées à partir des années réelles sont proches de celles observées parmi les données types. Des écart-types respectivement de 0,42 °C et de 0,11 °C sont constatés pour C_1 et C_{365}.

Une étude similaire est menée sur les données réelles de rayonnement global horizontal. Les données d'étude disponibles pour la ville de Mâcon vont de 2010 à 2016. Le Tableau 2-2 présente les résultats des décompositions de Fourier réalisées sur les années réelles et compare la dispersion des coefficients aux valeurs identifiées dans les séries de données types. La moyenne annuelle du RGH est plus élevée pour les données réelles que pour celles issues des données types bien que l'écart-type soit de 2,3 J/m². Des écarts sont aussi observés entre les coefficients C_1 et C_{365}.

Tableau 2-2 : Variabilité des coefficients des transformations de Fourier discrètes identifiées à partir de données réelles de rayonnement global à Mâcon (2010-2016)

	Données types	***Données réelles (2010-2016)***			
	H1c_RT2012	***Moyenne***	*Écart-type*	*Min*	*Max*
C_0 (J/cm²)	***48,8***	***51,6***	*2,3*	*48,3*	*55*
C_1 (J/cm²)	***19,2***	***20,4***	*0,8*	*19,5*	*22,1*
C_{365} (J/cm²)	***37,7***	***39,7***	*1,7*	*37,1*	*42,1*

Il semble donc pertinent de considérer une variabilité parmi les caractéristiques des séries tendancielles de température et RGH. Des fonctions de densité de probabilité sont définies à partir des variations observées. On considère ainsi des lois normales de moyenne correspondant à la valeur des coefficients pour le fichier type, d'écart-type égal à celui constaté dans l'analyse et bornée à deux écart-types. Ce choix est justifié par les valeurs maximales et minimales constatées. En complément du processus de génération présenté ci-dessus, un tirage aléatoire est réalisé dans ces lois de probabilité pour chaque série de température générée venant modifier la part tendancielle.

La prise en compte d'une variabilité des caractéristiques de la part tendancielle identifiée et son intégration dans le processus de génération permettent de représenter de manière plus complète la variabilité réelle de la température extérieure et du rayonnement global horizontal.

2.4.4.2. Corrélation entre les coefficients des parts tendancielles

Le Tableau 2-3 présente les coefficients de corrélation entre les coefficients des séries de Fourier identifiées parmi les séries de températures et de rayonnement solaire provenant des données réelles.

Tableau 2-3 : Analyse de la corrélation entre les coefficients de Fourier identifiés parmi 6 ans de données de température et de rayonnement global à Mâcon (2010-2016)

	Coefficient de corrélation $C_i(T_{h.Fourier})/C_i(RGH_{h.Fourier})$
C_0	*0,86*
C_1	*0,18*
C_{365}	*0,93*

On observe une forte corrélation entre les variations des coefficients C_0 et C_{365} parmi les séries tendancielles issues des données de température et de RGH des données réelles de Mâcon. La corrélation est plus faible entre les coefficients C1 des deux séries. Dans la suite, on propose de réaliser des tirages aléatoires corrélés des valeurs de ces coefficients à chaque génération de données synthétiques.

La caractérisation de la dispersion et des corrélations des coefficients principaux des séries tendancielles est étudiée sur un plus large panel de données au prochain paragraphe dans l'objectif de généraliser la méthodologie sur l'ensemble du territoire métropolitain.

2.4.4.3. Généralisation de la variabilité des séries tendancielles

La prise en compte d'une variabilité des coefficients principaux des séries de Fourier est un élément central de la méthodologie présentée et a pour but de mieux décrire la variabilité réelle des grandeurs météorologiques. Dans l'exemple présenté, traitant de la ville de Mâcon, les données provenant de mesures réalisées sur dix années pour les températures et six années pour le rayonnement ont été exploitées afin d'étudier la répartition de ces coefficients des séries de Fourier décrivant la part tendancielle des séries de données (§ 2.4.4). Les résultats de cette étude ont permis de définir des lois de probabilité pour ces coefficients, réparties autour de la valeur identifiée dans les données typiques. Lors de la génération de données, des tirages aléatoires sont réalisés, déterminant la valeur de ces coefficients principaux.

Les lois de probabilité identifiées au paragraphe 2.4.4 reposent sur les données liées à la position géographique de la ville de Mâcon. On propose d'étudier la possibilité d'étendre cette méthode pour n'importe quelle localisation afin de fixer selon des lois préétablies les bornes de variations des coefficients. En effet, l'objectif de la méthode de génération est de ne nécessiter qu'un seul jeu de données de départ, l'obtention et l'exploitation de multiples jeux de données réelles étant un processus long et coûteux. Des données historiques provenant de sept autres villes, propriété de MétéoFrance, ont été obtenues sur la plateforme Publithèque [9]. L'étude réalisée au paragraphe 2.4.4 est reconduite avec ces données. Les villes étudiées sont Agen, Carpentras, La Rochelle, Nancy, Nice, Rennes et Trappes. Ces villes constituent les localisations de référence pour les zones thermiques définies dans la RT 2012 (Figure 2-5).

Les Tableau 2-4 et Tableau 2-5 présentent pour chaque ville la moyenne et l'écart-type des coefficients identifiés sur 10 années. Les données de rayonnement global n'étaient disponibles qu'à partir de l'année 2008, 2009 ou 2010 suivant les villes. En étudiant la variabilité de l'écart-type des coefficients, on constate des variations assez similaires. Ainsi, pour le coefficient C_0 des séries de Fourier issues des températures extérieures, la moyenne des écart-types est de 0,59 °C, légèrement inférieur à l'écart-type de 0,67 °C identifié avec les données de Mâcon. L'écart-type sur ces valeurs est de 0,07 °C montrant ainsi une similarité dans les plages de variation des moyennes annuelles pour toutes les villes. L'écart-type des coefficients C_1 et C_{365} se caractérisent aussi par des variations semblables autour des valeurs de 0,4 °C et 0,1 °C.

[9] https://publitheque.meteo.fr/okapi/accueil/okapiWebPubli/index.jsp

Tableau 2-4 : Variation des coefficients principaux des séries de Fourier des températures des huit villes RT2012 pour les années 2007 à 2016

	C_0 (°C)		C_1 (°C)		C_{365} (°C)	
	Moyenne	*Écart-type*	*Moyenne*	*Écart-type*	*Moyenne*	*Écart-type*
Mâcon	*12,04*	*0,67*	*4,52*	*0,42*	*1,67*	*0,11*
Agen	*13,38*	*0,64*	*4,12*	*0,37*	*2,13*	*0,11*
Carpentras	*14,36*	*0,56*	*4,69*	*0,39*	*2,59*	*0,12*
La Rochelle	*13,18*	*0,62*	*3,41*	*0,41*	*1,27*	*0,05*
Nancy	*10,89*	*0,64*	*4,36*	*0,40*	*1,71*	*0,11*
Nice	*16,32*	*0,43*	*4,01*	*0,29*	*1,22*	*0,11*
Rennes	*11,85*	*0,57*	*3,25*	*0,44*	*1,74*	*0,08*
Trappes	*11,28*	*0,60*	*3,82*	*0,51*	*1,50*	*0,08*
MOYENNE		***0,59***		***0,40***		***0,10***
ÉCART-TYPE		***0,07***		***0,06***		***0,02***

Concernant le rayonnement global, la variabilité des coefficients de Fourier est légèrement plus élevée, même si on retrouve les ordres de grandeur identifiées avec les données de Mâcon. L'écart-type moyen pour les coefficients C_0, C_1 et C_{365} est respectivement de 1,62 J/cm², 0,9 J/cm² et 1,23 J/cm².

Tableau 2-5 : Variation des coefficients principaux des séries de Fourier du rayonnement global horizontal des 8 villes RT2012 pour les années 2007 à 2016

	C_0 (J/cm²)		C_1 (J/cm²)		C_{365} (J/cm²)	
	Moyenne	*Écart-type*	*Moyenne*	*Écart-type*	*Moyenne*	*Écart-type*
Mâcon	*51,64*	*2,26*	*20,36*	*0,82*	*39,68*	*1,71*
Agen	*56,23*	*1,40*	*19,66*	*0,52*	*43,62*	*1,12*
Carpentras	*64,98*	*1,38*	*22,19*	*0,92*	*50,24*	*1,09*
La Rochelle	*56,42*	*2,00*	*20,84*	*1,27*	*43,51*	*1,51*
Nancy	*47,33*	*1,42*	*19,76*	*0,68*	*36,00*	*1,08*
Nice	*64,77*	*1,33*	*22,24*	*0,57*	*50,47*	*1,03*
Rennes	*49,57*	*1,71*	*18,97*	*1,17*	*37,80*	*1,24*
Trappes	*47,89*	*1,46*	*19,28*	*1,22*	*36,34*	*1,04*
MOYENNE (J/cm²))		***1,62***		***0,90***		***1,23***
ÉCART-TYPE (J/cm²)		***0,34***		***0,30***		***0,25***

Suite à cette étude complémentaire, on propose ainsi de fixer les plages de variation des coefficients de Fourier identifiés en fonction des valeurs moyennes obtenues. Des lois normales tronquées sont considérées ; leur moyenne correspond à la valeur du coefficient identifié dans le jeu de données unique de référence, et leur écart-type aux valeurs identifiées ici et récapitulées dans le Tableau 2-6. Les lois sont tronquées à plus ou moins deux fois l'écart-type.

Tableau 2-6 : Moyenne des écart-type des coefficients de Fourier principaux identifiés à partir de l'ensemble des séries de température et de rayonnement global horizontal

	Écart-type		
	C_0	C_1	C_{365}
Température (°C)	*0,59*	*0,40*	*0,10*
RGH (J/cm²)	*1,62*	*0,90*	*1,23*

La corrélation entre les coefficients identifiés parmi les séries de température et de RGH est aussi étudiée dans le but de généraliser la génération conjointe des valeurs de ces coefficients. Le Tableau 2-7 présente les résultats des coefficients de corrélation pour les trois coefficients principaux C_0, C_1 et C_{365} en considérant l'ensemble des données disponibles, toutes années et localisations confondues.

Tableau 2-7 : Coefficients de corrélation entre les coefficients principaux des séries de Fourier identifiés à partir des séries de température et de rayonnement global

	Coefficient de corrélation $C_i(T_{h.Fourier})/C_i(RGH_{h.Fourier})$
C_0	*0,91*
C_1	*0,32*
C_{365}	*0,24*

Ainsi, on constate une forte corrélation des variations des coefficients C_0 des températures et RGH, et une corrélation plus faible pour C_1 et C_{365}. On propose de réaliser des tirages aléatoires (dans des lois normales tronquées dont les écart-types sont présentés dans le Tableau 2-6) venant modifier ces coefficients principaux lors de la génération de données. Un seul tirage commun est réalisé pour les coefficients C_0 (dans une loi centrée réduite) entrainant une corrélation parfaite ($\rho = 1$) et des tirages indépendants sont considérés pour les coefficients C_1 et C_{365} ($\rho = 0$).

Dans cette partie, la méthodologie de génération de données annuelles de température, de RGH et de RDH à partir d'un jeu de données types a été présentée. Elle repose sur l'analyse des séries types initiales et leur décomposition en une part tendancielle moyenne et une part résiduelle. Elle est pilotée en premier lieu par la création de séries de température permettant par la suite la génération de données de rayonnement corrélées. Une variation des parts tendancielles est aussi considérée et a nécessité en amont l'analyse d'une plus grande quantité de données météorologiques caractéristiques du site étudié. L'annexe A-2-5 récapitule l'algorithme de génération et présente les corrections algorithmiques et méthodologiques dues notamment à des problématiques numériques ou liées à l'identification de valeurs extrêmes aberrantes parmi certains processus aléatoires. Le paragraphe suivant a pour objectif l'étude de la qualité des données synthétiques obtenues.

2.5. Évaluation de la fiabilité du modèle

2.5.1. Caractéristiques des séries générées

Afin d'évaluer la qualité du processus de génération, les caractéristiques des données synthétiques sont comparées aux données de référence, données types de la ville de Mâcon, desquelles elles sont issues. Afin d'analyser la dispersion statistique de ces caractéristiques, un échantillon de séries météorologiques provenant de mesures réalisées par Météo-France à Mâcon, de 2007 à 2016 pour la température, et de 2010 à 2016 pour le rayonnement solaire, sont aussi analysées. Il est à noter que la définition des fonctions de densité de probabilité des coefficients des séries de Fourier caractéristiques des parts tendancielles des séries de température et de RGH (paragraphe 2.4.4) provient de l'étude de ces années réelles.

2.5.1.1. Séries de température

Le Tableau 2-8 présente les caractéristiques principales des données de température issues du fichier météorologique type de la ville de Mâcon, utilisé pour la génération de données. La moyenne des caractéristiques des données de dix années réelles d'une part et de 1000 jeux de données générées d'autre part sont comparées. L'écart-type et les extrema de ces caractéristiques sont aussi présentés.

Tableau 2-8 : Comparaison des caractéristiques principales des données horaires de température générées, réelles (2007-2016), et du fichier type de Mâcon

	Données types H1C-RT2012	***Données générées (1000 séries)***				***Données réelles (2007-2016)***			
		Moyenne	*Ecart-type*	*Min*	*Max*	***Moyenne***	*Ecart-type*	*Min*	*Max*
Moyenne annuelle *(°C)*	***12,94***	***12,95***	*1,0*	*10,5*	*15,4*	***12,04***	*0,7*	*11,1*	*13,0*
Écart-type annuel *(°C)*	***7,6***	***7,5***	*0,4*	*6,3*	*8,6*	***7,8***	*0,5*	*6,8*	*8,5*
Minimum *(°C)*	***-6,6***	***-6,6***	*1,9*	*-14,5*	*-1,7*	***-7,7***	*3,2*	*-13,5*	*-4,5*
Maximum *(°C)*	***35,4***	***34,9***	*1,9*	*28,7*	*43,5*	***35,2***	*2,1*	*31,8*	*38,7*
Moyenne hiver *(°C)*	***6,3***	***5,6***	*1,2*	*2,2*	*9,2*	***4,7***	*1,3*	*2,5*	*6,6*
Moyenne printemps *(°C)*	***14,5***	***15,6***	*1,1*	*12,4*	*18,9*	***14,5***	*1,2*	*12,6*	*16,2*
Moyenne été *(°C)*	***21,1***	***20,3***	*1,2*	*16,8*	*23,8*	***20,0***	*1,1*	*18,1*	*21,9*
Moyenne automne *(°C)*	***9,8***	***10,2***	*1,1*	*7,3*	*13,7*	***9,0***	*1,1*	*7,7*	*10,7*
DH_{18} *(°C.h)*	***44349***	***44244***	*8778*	*23051*	*65560*	***52208***	*5833*	*43978*	*60740*

En comparant les attributs des séries de température générées à ceux du fichier TRY ayant servi de référence, on constate que l'écart entre les moyennes annuelles est de 0,01 °C. L'écart-type de la moyenne annuelle est de 1,0 °C ce qui est légèrement supérieur à celui observé pour les 10 années réelles exploitées (0,7 °C). Les écart-types des séries sur l'année sont satisfaisants au regard des caractéristiques de référence, tout comme les extrema. Même si l'étude des données de 10 années réelles permet d'obtenir des ordres de grandeur sur les caractéristiques des séries, un nombre aussi réduit d'échantillon ne permet pas d'estimations statistiquement robustes.

Les moyennes saisonnières de température sont elles aussi étudiées et sont cohérentes. On constate cependant parmi les données générées une moyenne hivernale plus faible (-0,7 °C) que celle du fichier TRY, mais supérieure à la moyenne des températures moyennes hivernales observées sur dix années réelles (+0,9 °C). Cet écart entre les années réelles et l'année type qui considère un hiver très doux, avec une moyenne saisonnière supérieure de 1,6 °C à celles moyennement observées en réalité, pose question sur la représentativité des données types. Cet élément est renforcé par le calcul des DH_{18}. En effet, les DH_{18} des séries générées sont en moyenne proches de ceux des données types qui sont en revanche sous-estimés par rapport aux années réelles considérées.

La Figure 2-11 présente les moyennes horaires de température observées sur chaque mois de l'année pour le fichier type de Mâcon pour 5000 séries générées.

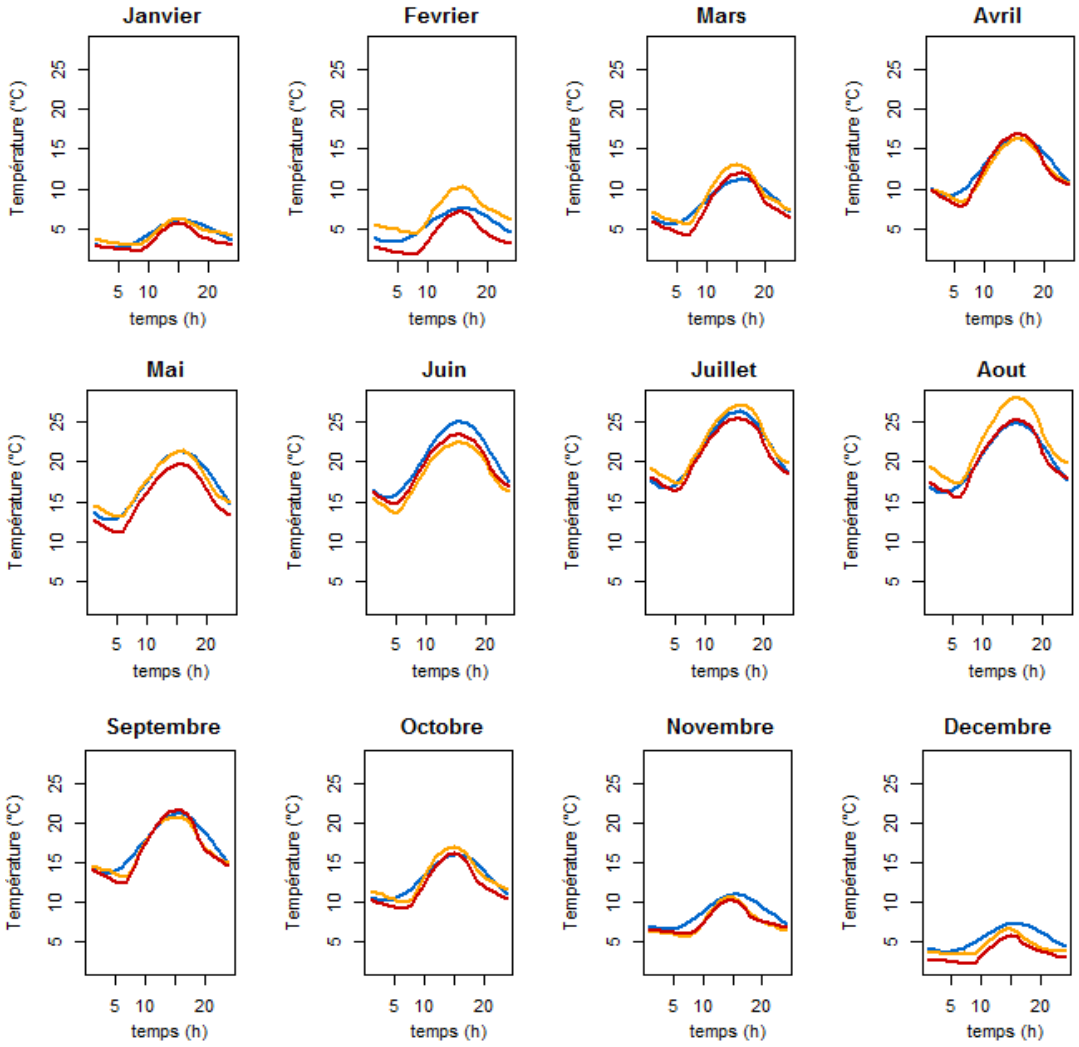

Figure 2-11. Comparaison des moyennes horaires de température par mois pour les données types (orange), réelles (rouge) et la moyenne de 5000 séries générées (bleu)

Il est intéressant de noter que pour les séries générées, cette moyenne horaire réalisée sur 1000 jeux de données correspond à la part tendancielle identifiée. La correspondance est satisfaisante pour la majorité des mois même si des écarts de plusieurs degrés, positifs ou négatifs, sont observés entre données générées et types pour les mois de février, juin et août. Ceci peut en partie s'expliquer par la faible taille de l'échantillon statistique (15 années) utilisée pour construire la série type. En effet, en cas d'épisode météorologique « déviant » des tendances pendant plusieurs jours consécutifs, ce qui peut être observé pour un mois médian dans un si petit échantillon, la moyenne horaire du mois est rapidement affectée. Par ailleurs, on constate un écart important, de l'ordre de 3 °C, entre données réelles et typiques pour le mois de février, les données générées ayant un niveau intermédiaire. Dans une moindre mesure, un décalage similaire est observé au mois d'août. Ces éléments confirment les conclusions effectuées suite à l'étude du Tableau 2-8 des caractéristiques principales des séries.

2.5.1.2. Séries de rayonnement solaire global horizontal

Le Tableau 2-9 présente des informations similaires au Tableau 2-8 pour les données de rayonnement global horizontal typiques, synthétiques et réelles.

Tableau 2-9 : Caractéristiques principales des données de RGH du fichier type de Mâcon, des données générées qui en sont issues, et de données réelles (2010-2016)

	Données types	***Données générées (1000 séries)***				***Données réelles (2007-2016)***			
	H1c_RT2012	***Moyenne***	*Écart -type*	*Min*	*Max*	***Moyenne***	*Écart -type*	*Min*	*Max*
Moyenne annuelle *(J/cm²)*	***48,8***	***49,1***	*2,2*	*42,3*	*55*	***51,6***	*2,3*	*48,3*	*55*
Ecart-type annuel *(J/cm²)*	***78,2***	***76,5***	*2,5*	*68,6*	*84*	***82,8***	*2,2*	*79,6*	*86,1*
Maximum *(J/cm²)*	***335***	***350,9***	*1,6*	*335*	*352*	***350,3***	*9,9*	*339*	*370*
Moyenne hiver *(J/cm²)*	***26***	***25,7***	*2,7*	*17,1*	*34*	***25,8***	*2,3*	*23,2*	*29*
Moyenne printemps *(J/cm²)*	***68,2***	***71,7***	*5*	*56,5*	*91*	***73,9***	*8,7*	*63,4*	*84,9*
Moyenne été *(J/cm²)*	***77,7***	***74,1***	*4,5*	*60,5*	*89*	***81,1***	*4,5*	*74*	*85,1*
Moyenne automne *(J/cm²)*	***23,5***	***25***	*2,6*	*16,6*	*33,5*	***26***	*1,6*	*23,8*	*28,5*

Les moyennes des séries générées sont légèrement supérieures à celle de la série issue du fichier TRY (+0,6 %). L'écart-type des moyennes des séries générées est similaire à celui des années réelles. Les moyennes saisonnières sont elles aussi légèrement décalées, même si l'écart observé est inférieur à 7 %.

La Figure 2-12 présente une comparaison des moyennes horaires par mois pour les données de rayonnement. Les tendances observées dans les séries générées sont conformes à la série type de référence et aux données réelles. Des pointes un peu plus faibles sont tout de même constatées sur certains mois.

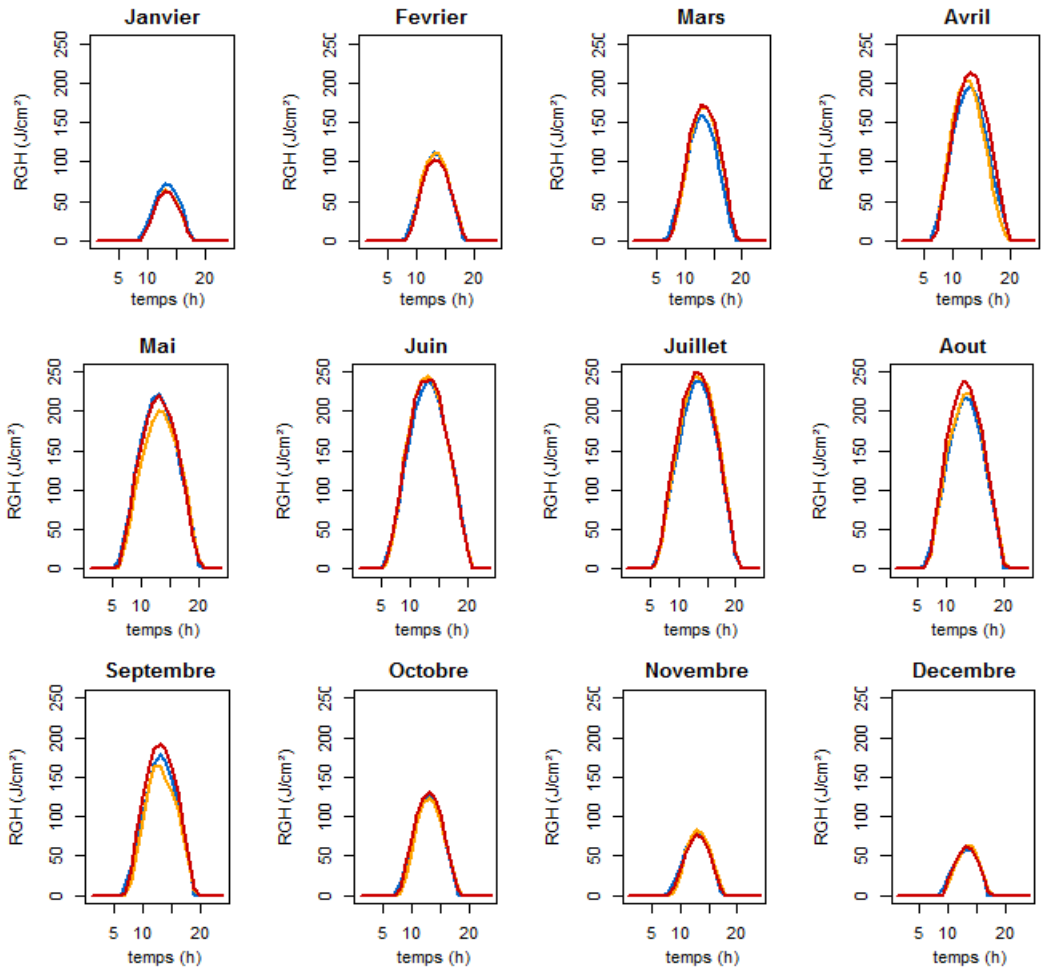

Figure 2-12. Comparaison des moyennes horaires de rayonnement global horizontal par mois pour les données types (orange), réelles (rouge) et la moyenne de 5000 séries générées (bleu)

2.5.1.3. Séries de rayonnement solaire diffus horizontal

Le rayonnement diffus horizontal (RDH) est naturellement très corrélé au RGH (2.13). Les caractéristiques comparées des données types, réelles et générées sont présentées dans le Tableau 2-10. On constate des différences similaires à celles relevées dans le paragraphe précédent. Là-aussi les caractéristiques des séries synthétiques sont souvent légèrement supérieures aux valeurs issues des données réelles.

La comparaison entre un échantillon de données synthétiques et les séries temporelles types à partir desquelles elles sont créées permet de vérifier la cohérence du processus de génération. Les caractéristiques globales et dynamiques sont concordantes et leur dispersion est satisfaisante au regard des années réelles analysées. Ayant étudié indépendamment les séries générées, il est nécessaire de vérifier si l'intercorrélation entre les séries générées est similaire à celle observée dans les séries de références.

Tableau 2-10 : Caractéristiques principales des données de RDH du fichier type de Mâcon, des données générées qui en sont issues, et de données réelles (2010-2016)

	Données types - TRY	***Données générées (1000 séries)***				***Données réelles (2007-2016)***			
	H1c_RT2012	***Moyenne***	*Écart -type*	*Min*	*Max*	***Moyenne***	*Écart -type*	*Min*	*Max*
Moyenne annuelle *(J/cm²)*	***24,8***	***25,7***	*0,4*	*24,7*	*26,8*	***23,3***	*0,3*	*22,8*	*23,5*
Écart-type annuel *(J/cm²)*	***34,6***	***37,8***	*0,4*	*34,6*	*39,1*	***33,2***	*0,4*	*32,7*	*33,7*
Maximum *(J/cm²)*	***147,0***	***144,5***	*0,3*	*142,7*	*147*	***144,5***	*0,2*	*144,3*	*144,8*
Moyenne hiver *(J/cm²)*	***14,4***	***14,8***	*0,9*	*11,6*	*17*	***13,8***	*0,6*	*13,3*	*14,2*
Moyenne printemps *(J/cm²)*	***32,2***	***37,2***	*1*	*33,9*	*40,2*	***32,8***	*0,7*	*31,5*	*33,5*
Moyenne été *(J/cm²)*	***31,3***	***36 ,8***	*1*	*33,3*	*39,9*	***32,6***	*0,7*	*31,5*	*33,6*
Moyenne automne *(J/cm²)*	***13,6***	***14,3***	*0,8*	*11,4*	*16,3*	***13,9***	*0,3*	*13,5*	*14,4*

2.5.2. Respect de l'intercorrélation entre grandeurs météorologiques

On s'intéresse au lien entre les séries générées de température et de RGH, à la fois au sein des parts tendancielles, des parts résiduelles et des séries complètes. Le Tableau 2-11 présente les coefficients de corrélation linéaire calculés pour les données types, synthétiques et réelles.

Tableau 2-11 : Analyse de l'intercorrélation entre les séries de température et de RGH parmi les différents jeux de données

	Données types - TRY	***Données générées (1000 séries)***				***Données réelles (2007-2016)***	
Coefficient de corrélation	***H1c_RT2012***	***Moyenne***	*Écart-type*	*Min*	*Max*	***Moyenne***	*Écart-type*
T_h / RGH_h	***0,54***	***0,52***	*0,02*	*0,47*	*0,56*	***0,54***	*0,02*
$T_{h.Fourier}$ / $RGH_{h.Fourier}$	*0,59*	*0,59*	*0,01*	*0,58*	*0,60*	-	-
$T_{h.res}$ / $RGH_{h.res}$	*0,29*	*0,30*	*0,03*	*0,20*	*0,39*	-	-
RGH_h / RDH_h	***0,81***	***0,83***	*0,01*	*0,79*	*0,87*	***0,78***	*0,01*

En comparant les caractéristiques des données synthétiques à celle des séries typiques, on constate que l'intercorrélation est respectée. On observe que l'intercorrélation finale est due en partie au lien entre les parts tendancielles, qui est fort (ρ = 0,59) et est dégradé par le lien entre les séries résiduelles, naturellement moins corrélées (ρ = 0,3). La corrélation entre les séries complètes est

donc comprise entre ces deux valeurs. Par ailleurs, la corrélation linéaire entre le rayonnement global et diffus est très proche pour les données synthétiques, pour les données types et pour les données réelles.

Les résultats, en terme de respect de l'intercorrélation entre les grandeurs météorologiques, répondent donc aux objectifs fixés, constituant un des apports significatifs du modèle.

2.6. Application à l'étude de l'influence du climat sur les consommations d'énergie d'un bâtiment

2.6.1. Propagation de la variabilité

Un échantillon de 5000 jeux de données météorologiques est généré à partir du fichier type de Mâcon « MaconH1cRT2012.try » disponible dans le logiciel *Pléiades*+COMFIE. Une variabilité des coefficients de la part tendancielle obtenue à partir de l'étude de données météorologiques réelles de la station de Mâcon est aussi considérée. 5000 simulations énergétiques dynamiques du bâtiment étudié sont réalisées possédant chacune en entrée un jeu de données météorologiques différent. Des simulations sont aussi conduites avec le fichier météorologique type, et avec sept ensembles de données réelles (2010-2016). Enfin, une simulation est aussi effectuée avec en entrée les séries de température et de rayonnement correspondant aux parts tendancielles identifiées par décomposition en série de Fourier des séries typiques, sans considérer de parts résiduelles.

Des simulations de l'immeuble de logements collectifs rénové situé à Feyzin (69), et présenté au chapitre 1 sont réalisées. Une température de consigne unique de chauffage de 20 °C est considérée des semaines 1 à 18 et 42 à 52, et une consigne de climatisation de 24 °C, valeur assez faible choisie pour augmenter la sensibilité du bâtiment aux surchauffes estivales, est appliquée des semaines 25 à 32. On s'intéresse ici uniquement à la variabilité météorologique, tous les autres paramètres de la simulation sont constants. L'influence des conditions météorologiques dans un contexte incertain plus global sera étudiée dans les chapitres suivants. On s'intéresse à la variabilité des besoins de chauffage et de refroidissement induite par la variabilité des conditions climatiques : on parle ainsi de propagation de la variabilité. La Figure 2-13 présente les histogrammes de dispersion des besoins de chauffage (a), des besoins de refroidissement (b) et de la somme de ces besoins (c).

Une loi normale équivalente, dont les caractéristiques sont calculées à partir de la moyenne et de l'écart-type de l'échantillon est ajoutée aux figures. Elle permet de définir le quantile à 95 % de la distribution normale. Les besoins énergétiques calculées dans les simulations réalisées avec les données météorologiques des années réelles sont affichés en noir ; la simulation réalisée avec le fichier typique est représentée avec un triangle rouge, et celle réalisée avec des séries de

température et de rayonnement composées uniquement des parts tendancielles identifiées par décomposition en série de Fourier est décrite avec un carré rouge.

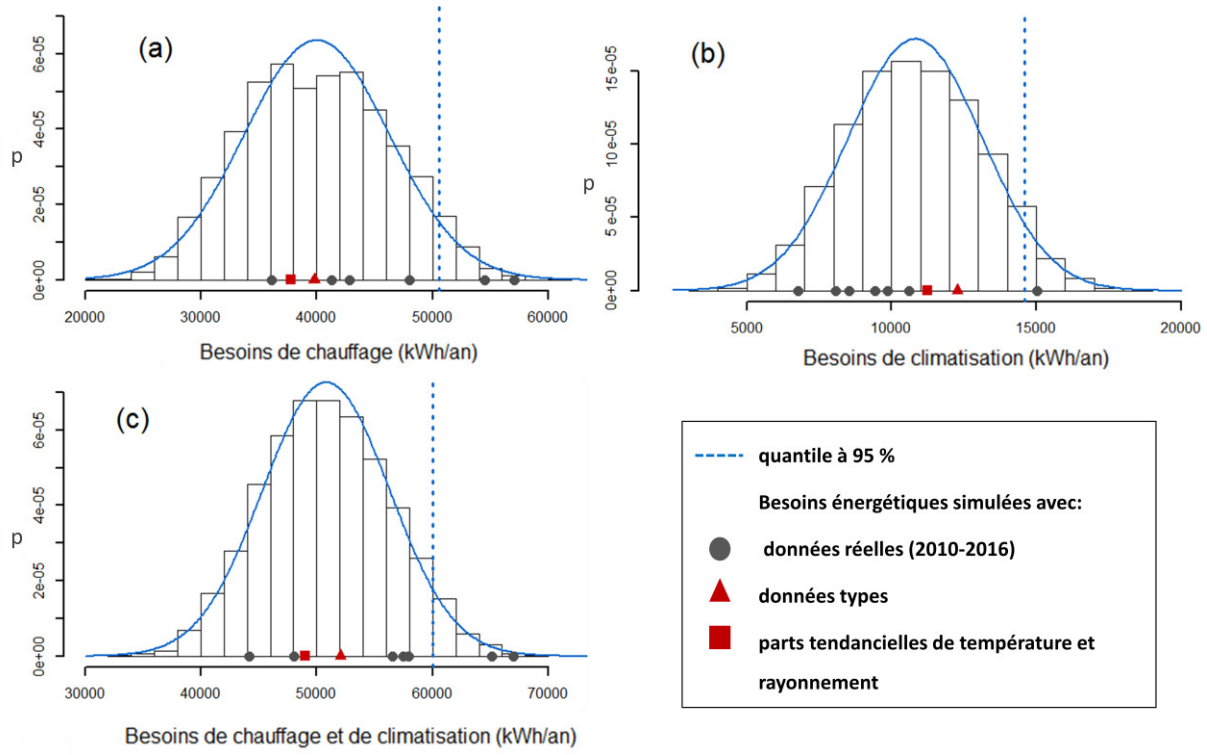

Figure 2-13. Distributions des besoins de chauffage (a), de refroidissement (b) et totaux (c) du bâtiment simulé avec des données météorologiques synthétiques (5000 simulations), types, tendancielles et réelles (2010-2016)

On constate d'abord concrètement la variabilité des besoins induite par les variations climatiques réelles, ce qui justifie l'intérêt porté à la question étudiée dans ce chapitre. Concernant les besoins de chauffage, deux années réelles présentent des besoins très élevés. Ceci s'explique par des hivers 2010 et 2013 particulièrement froids en France et à Mâcon en particulier. De même, l'année 2015 a été marquée par un été en partie caniculaire provoquant des besoins de refroidissement considérés comme extrêmes par le modèle. Les simulations des années 2010 et 2013 montrent des besoins totaux supérieurs au quantile à 95 % de la loi de distribution obtenue.

La variabilité des besoins d'énergie induite par le modèle se caractérise par une dispersion plus large que celles observées dans les études de Hong et al. (2013) et Wang et al. (2012) mais est de l'ordre de celle décrite par Goffart (2013b) où des variations de besoins de chauffage du simple au triple sont constatées.

2.6.2. Caractérisation de l'impact des conditions météorologiques

L'objectif du modèle de création de données météorologiques couplé à la SED est aussi de caractériser le lien entre les conditions climatiques et les besoins énergétiques des bâtiments. La caractérisation de cette dépendance est rendue difficile par l'aspect dynamique des grandeurs étudiées. Les données météorologiques sont modélisées par des séries temporelles mais des variables scalaires peuvent tenter de les caractériser de manière plus ou moins pertinente. De la même manière, les besoins ou consommation de chauffage et de refroidissement peuvent être moyennés ou sommés sur des périodes de référence. On s'intéresse ici au lien entre grandeurs agglomérées caractérisant les consommations énergétiques et les données météorologiques.

2.6.2.1. Variables explicatives des conditions météorologiques

On souhaite étudier la précision avec laquelle des variables scalaires peuvent décrire l'effet des variations météorologiques sur les besoins énergétiques des bâtiments. Définir des variables annuelles à partir de séries temporelles de 8760 points constitue une perte d'information. L'enjeu est de limiter cette perte d'information afin de modéliser plus simplement le lien entre conditions climatiques et besoins énergétiques à partir des résultats de simulation.

Différentes variables peuvent être définies à partir des séries temporelles de température :

- La moyenne annuelle des températures $T_{e.moy}$.
- La moyenne des températures sur les saisons de chauffe et de refroidissement, $T_{e.moy.sc}$ et $T_{e.moy.sr}$. Les saisons de chauffe et de refroidissement sont respectivement définies entre les semaines 42 et 18 et les semaines 25 et 32.
- Les degré-jours unifiés à base de température T_{ref}, $DJU_{T_{ref}}$. La moyenne entre la température minimale et maximale du jour est considérée. Les différences entre la température de référence T_{ref} et cette moyenne sont sommées sur chaque jour de l'année (2.1). Une température de 18 °C est usuellement considérée comme référence pour le chauffage.
- Les degré-heures à base de température T_{ref}, $DH_{T_{ref}}$. La moyenne de la température de l'heure est considérée. Les différences entre la température de référence T_{ref} et cette valeur sont sommées sur chaque heure de l'année (2.2). Une température de 18 °C est usuellement considérée comme référence pour le chauffage. En plus des DH_{18}, on propose de calculer les DH_{T_c} qui ont pour température de référence la température de consigne (éventuellement dynamique) de chauffage dans le bâtiment.

En ce qui concerne l'évaluation de l'impact du refroidissement, on définit de manière similaire les DH_{24} comme étant la somme des écarts négatifs entre la température de référence, 24 °C, et la température extérieure. Il est aussi possible de sommer ces différentes contributions sur une période

réduite de l'année. Les saison de chauffe et de refroidissement étant définies comme ci-dessus, on considère aussi les variables DH_{T_c}, $DH_{T_c,sr}$ et $DH_{24,sr}$.

Des variables concernant les données de rayonnement peuvent aussi être définies :

- la moyenne annuelle du rayonnement global horizontal, RGH_{moy} ;
- les moyennes du rayonnement global horizontal sur les saisons de chauffe et de refroidissement, $RGH_{moy,sc}$ et $RGH_{moy,sr}$.

2.6.2.2. Corrélation linéaire

La Figure 2-14 présente les résultats de besoins de chauffage ((a) à (h)) et de refroidissement ((i) à (l)) de 1000 simulations énergétiques du bâtiment exprimés en fonction de variables météorologiques caractéristiques. Pour chaque simulation, ces variables sont donc calculées à partir des données météorologiques utilisées. Des simulations du bâtiment sont aussi réalisées avec les données météorologiques réelles (rouge) et types (orange).

Ces figures apportent plusieurs informations. Elles permettent tout d'abord de confronter les résultats de simulation obtenus avec les données météorologiques synthétiques et réelles, liant besoins énergétiques et facteurs météorologiques caractéristiques. Enfin, la qualité de la prédiction des différents modèles linéaires est aussi comparée permettant la sélection des variables explicatives les plus pertinentes. Le coefficient de détermination R² décrit la qualité de la prédiction du modèle linéaire identifié par la méthode des moindres carrés. Il est défini comme le rapport de la variance expliquée sur la variance totale :

$$R^2 = \frac{\sum_{i=1}^{n}(\widehat{y_i} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \textbf{(2.14)}$$

En comparant les résultats de simulation réalisées avec les données synthétiques et réelles, on constate que pour la plupart des corrélations étudiées dans la Figure 2-14, les points représentant les simulations avec données météorologiques réelles sont inclus dans le nuage des points liés aux simulations avec données synthétiques (besoins de chauffage en fonction de $T_{e,moy}$, $T_{e,moy,sc}$, DH_{18}, $DH_{18,sc}$, DH_{T_c}, DH_{T_c} et besoins de refroidissement en fonction de $T_{e,moy,sr}$). Pour d'autres variables explicatives, on constate un décalage faible et partiel (DJU_{18}, DH_{24}, et $DH_{24,sr}$, $RGH_{moy,sr}$) ou plus important (RGH_{moy}, $RGH_{moy,sc}$). Ceci signifie que l'effet sur les besoins énergétiques des conditions météorologiques, décrites selon ces variables est différent entre conditions générées et réelles. Cela peut s'expliquer par des variations dynamiques mal représentées entraînant une réponse thermique du bâtiment différente même si les conditions météorologiques moyennes sont similaires.

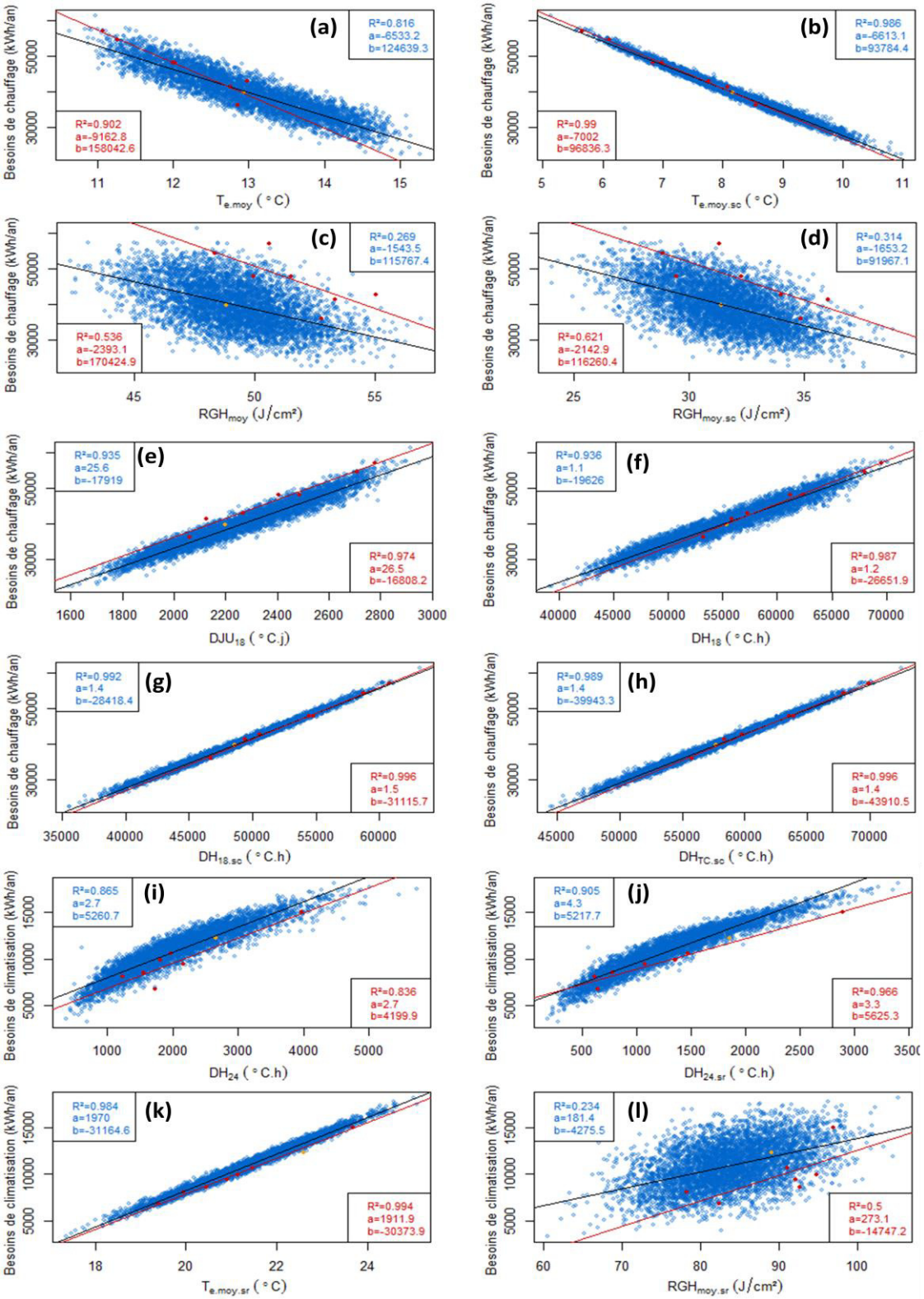

Figure 2-14. Comparaison de modèles linéaires liant différentes variables météorologiques d'ajustement aux besoins de chauffage et de refroidissement; SED avec 1000 jeux de données météorologiques synthétiques (bleu), 7 jeux de données réelles (rouges) et le fichier type (orange)

Ce décalage est souvent plus faible ou inexistant pour le résultat de simulation associé aux données types, à partir desquelles sont recréées les données synthétiques, ce qui questionne encore une fois la représentativité de ces fichiers types. Ceci est particulièrement vrai pour les corrélations impliquant le RGH ou le nuage de points est centré sur résultat de simulation type alors que les simulations avec données réelles sont caractérisées par des valeurs moyennes de rayonnement plus élevées pour un même besoin de chauffage ((c) et (d)).

Par ailleurs, on constate que certaines variables agglomérées ont un pouvoir de prédiction supérieur, elles décrivent mieux les conséquences de la variabilité météorologique sur les besoins de chauffage et de climatisation. Ceci s'observe par des nuages de points plus resserrés et mathématiquement par des valeurs du coefficient de détermination R^2 plus proches de 1. Que ce soit pour les besoins de chauffage ou de climatisation, on constate que des gains en qualité de prédiction sont obtenus en se concentrant sur les saisons de chauffe et de refroidissement, que ce soit pour les $DH_{18,sc}$, $DH_{24,sr}$ ou pour la température extérieure moyenne $T_{e,moy,sc}$ et $T_{e,moy,sr}$. Le choix d'une température de référence correspondant à la température de consigne ($DH_{Tc,sc}$) n'améliore pas significativement la prédiction par rapport au $DH_{18,sc}$. Ceci peut s'expliquer par le fait qu'une consigne constante de chauffage de 20 °C est considérée dans la simulation, limitant l'intérêt de considérer cette température comme référence.

Concernant les besoins de refroidissement, la température extérieure moyenne en période de rafraîchissement, $T_{e,moy,sr}$, est la variable expliquant le mieux les variations observées. Par ailleurs on constate une répartition non linéaire des besoins de refroidissement en fonction des DH_{24} et des $DH_{24,sr}$. Un modèle polynômial d'ordre supérieur, ou une combinaison de segments linéaires, serait plus adapté pour caractériser le lien unissant les besoins de climatisation et ces variables.

Que ce soit pour les besoins de chauffage ou de refroidissement, les variables dérivées du rayonnement solaire ont un mauvais pouvoir de prédiction avec un R^2 faible. En effet, les variations de température on un effet beaucoup plus important que les variations de rayonnement solaire, même si ces deux grandeurs sont corrélées. Cela est propre au cas d'étude utilisé mais correspond à une tendance globalement généralisable (Goffart, 2013b).

Cette étude montre que la connaissance de la valeur de certaines variables scalaires associées aux conditions météorologiques permet une prévision relativement précise des besoins de chauffage et de climatisation. Cette prévision repose sur l'estimation des paramètres d'un modèle simplifié associé à un bâtiment donné, ici une simple loi affine. Cette estimation est rendue possible par la simulation du bâtiment dans un grand nombre de contextes climatiques générés par le modèle développé. La taille des échantillons est garante de la robustesse de la valeur des paramètres du modèle identifié. La création de ce type de métamodèle simplifié et les objectifs qui y sont associés seront plus largement abordés au chapitre 3.

2.7. Conclusion

La modélisation de la variabilité naturelle des conditions météorologiques a pour but d'évaluer son impact sur les variations de consommation d'énergie dans les bâtiments. Après être revenu sur les travaux traitant de la prise en compte du climat dans les méthodes de simulation et sur les études tentant de quantifier l'influence de cette variabilité, ce chapitre présente un modèle de génération stochastique de données de températures pour la SED. Il permet de recréer une grande quantité de séries temporelles au pas de temps horaire de température extérieure, de rayonnement global horizontal (RGH) et de rayonnement diffus horizontal (RDH).

Par rapport aux modèles existants, il a l'avantage de générer des données sur la totalité du cycle annuel et de ne nécessiter en entrée qu'un seul jeu de données de température et de RGH, usuellement disponibles dans les logiciels de SED. Le processus de génération, s'appuyant sur la décomposition des séries en une part tendancielle déterministe et une part résiduelle aléatoire est piloté en premier lieu par la création des données de température. Les séries de RGH sont générées selon un processus stochastique et l'intercorrélation température-rayonnement est modélisée au niveau tendanciel mais aussi au sein de la part résiduelle. Ces deux éléments constituent aussi une originalité du modèle.

Ces développements visent plusieurs objectifs s'inscrivant à différents niveaux au sein de ces travaux de thèse. Le modèle de génération de données météorologiques vient compléter l'ensemble des méthodes développées pour une modélisation physico-probabiliste des bâtiments. Cette approche décrit le bâtiment comme un système incertain avec des sollicitations variables par nature. Elle est basée sur la réalisation d'un grand nombre de simulations dont les résultats constituent un échantillon statistique représentatif de la distribution des réalisations possibles induites par l'ensemble des entrées du système décrites selon des lois et des modèles probabilistes. La généralisation de ce type d'approche constitue l'avenir de la simulation des bâtiments pour la prévision des consommations, dans un contexte ou la puissance de calcul est en augmentation et où l'exploitation de données massives se généralise. Par ailleurs, cet outil a vocation à être utilisé comme un moyen de construire des modèles simplifiés reliant les résultats de SED à des variables scalaires construites à partir des données météorologiques. L'identification de tels modèles de type « boîte noire » nécessite des données d'entraînement obtenues grâce à l'utilisation du modèle de génération de données météorologiques. On s'intéresse ici à l'influence relative des conditions météorologiques et non plus à leur influence absolue.

Un travail de validation du modèle de génération est présenté, soulignant des performances satisfaisantes. Cependant, ces possibilités de validation sont limitées par le nombre de jeux de données réelles disponibles et par le caractère aléatoire des processus décrits. Le contexte de changement climatique pose aussi question sur la modification de la fréquence d'épisodes

climatiques extrêmes et sur des tendances de réchauffement à prendre en compte pour des études intégrant des objectifs de plus long terme, de l'ordre de plusieurs dizaines d'années.

Enfin, une étude reposant sur une propagation de la variabilité des conditions météorologiques présente la manière dont les objectifs cités peuvent être concrétisés dans des études SED. La répartition induite des besoins de chauffage et de climatisation d'un bâtiment de logement se caractérise par une large variabilité, avec des valeurs allant du simple au triple pour le chauffage et d'un rapport de un à dix pour la climatisation. Ce fort impact du climat souligne d'autant plus le fait que les fichiers météorologiques types sont inadaptés dans une approche prédictive et mettent en avant l'intérêt de conditionner les prévisions de consommation d'énergie aux conditions météorologiques. Pour cela, des modèles simplifiés peuvent être développés pour quantifier cette relation de causalité. Les modèles linéaires reliant besoins énergétiques et variables explicatives d'ajustement présentés en fin de chapitre introduisent ainsi la méthodologie d'ajustement présentée au chapitre 3.

Chapitre 3. Méthodologie pour la définition d'un engagement de performance énergétique

Ce chapitre présente une méthodologie permettant de définir un engagement sur la performance énergétique des bâtiments. L'ensemble des éléments associés aux projets de garantie de performance énergétique (GPE) sont interdépendants mais sont souvent traités de façon cloisonnée. La méthodologie développée ici vise à intégrer ces interdépendances. Les différentes étapes du processus et leur articulation logique sont détaillées, allant de la modélisation physico-probabiliste, au traitement statistique des résultats de simulation, et comprenant la prise en compte des problématiques liées à la mesure. Le lien entre processus d'ajustement et caractérisation d'un risque de défaut est un élément central de la problématique. La mise en œuvre d'une méthode intégrée et globale qui soit à la fois robuste et facilement applicable constitue l'objectif de ces travaux. Les cas de GPE absolue et relative, respectivement adaptées aux opérations de construction neuve ou de réhabilitation, sont traités et les méthodes illustrées par un cas test.

Sommaire du chapitre 3

Chapitre 3. Méthodologie pour la définition d'un engagement de performance énergétique.. 115

3.1. Objectifs de la méthodologie 117

3.2. La simulation du modèle physico-probabiliste 118

3.2.1. Hypothèses et objectifs de la démarche 118

3.2.2. Analyse de sensibilité 118

3.2.3. Propagation des incertitudes et variabilités 119

3.2.4. Développement et calibrage du modèle pre-travaux 120

3.3. L'ajustement 123

3.3.1. Formulation et intérêt du modèle 123

3.3.2. Définition et choix des facteurs d'ajustement 123

3.3.3. Méthodes de calcul des modèles d'ajustement 125

3.3.4. GPE absolue : Détermination d'une limite garantie de consommation ajustable129

3.3.5. GPE relative : détermination du gain et du modèle de référence ajustable 130

3.3.6. Ajustement et vérification de la performance 134

3.4. Intégration de l'influence des incertitudes de mesure 136

3.4.1. Enjeu et approche numérique 136

3.4.2. Estimation des incertitudes 137

3.4.3. Intégration amont au processus 140

3.5. Protocole complet pour la garantie de performance énergétique 142

3.5.1. GPE absolue dans le neuf et la rénovation lourde 142

3.5.2. GPE relative en réhabilitation 143

3.6. Conclusion 147

3.1. Objectifs de la méthodologie

L'objectif des travaux présentés dans ce chapitre est de fournir un cadre méthodologique pour pouvoir proposer des contrats de garantie de performance énergétique (GPE). Cette méthodologie doit venir se substituer aux pratiques actuelles, présentées au chapitre 1, et doit dépasser les limitations et faiblesses identifiées en leur sein.

Sa première ambition réside dans la rigueur scientifique des méthodes sur lesquelles elle s'appuie qu'il s'agisse des modèles physiques ou des cadres mathématiques et statistiques utilisés. Ainsi, la quantification et la maitrise du risque associé à l'engagement sur les consommations d'énergie est au cœur de la méthode. En effet, un risque de défaut α de cet engagement doit lui être associé, permettant de réaliser un compromis entre le niveau de risque et la valeur garantie. On souhaite aussi pouvoir s'engager sur des niveaux de consommation énergétique ambitieux afin que le contrat de GPE constitue un levier de financement efficace des travaux de construction ou de rénovation visant la haute performance énergétique des bâtiments.

Comme précisé précédemment, la méthodologie doit conduire à un processus d'ajustement permettant de s'affranchir de l'effet d'éléments impactant fortement les consommations énergétiques sans être directement liés à la performance intrinsèque de l'ouvrage : les sollicitations météorologiques et le comportement des usagers.

La méthode développée doit intégrer l'ensemble des étapes d'un projet de GRE, de l'étude de conception à la vérification de la performance réelle. Cette intégration est rendue indispensable par l'interdépendance entre ces étapes. L'accent sera notamment mis sur l'intégration de l'influence des incertitudes de mesure en amont du processus, celles-ci ayant un impact sur le risque associé à l'engagement.

Une autre des problématiques qu'il est nécessaire de traiter est la transparence et la bonne compréhension des éléments contractuels et des modes de résolution du contrat. La vérification de l'engagement, qui s'appuie sur la mesure et sur la prise en compte d'un ajustement en fonction des conditions réelles d'exploitation, ne doit pas nécessiter de nouvelles interventions lourdes ou de nouvelles simulations mais doit s'effectuer de manière simple et transparente.

Enfin, deux cas de figure seront traités en parallèle dans ce chapitre. On différenciera les méthodologies associées à des cas de GPE absolue ou relative. Dans le premier cas, l'engagement contractuel porte sur une valeur maximale de consommation énergétique exprimée de manière absolue, on parlera de limite garantie de consommations (LGC). Cette configuration est associée à des projets de bâtiment neuf, à des rénovations lourdes de bâtiments existants ou lorsque les performances du bâtiment avant travaux ne sont pas connues. La GPE relative vise à définir une cible de diminution de consommations d'énergie par rapport à une situation précédant les travaux et est donc associée aux projets de réhabilitation. L'engagement est ici lié à la qualité des travaux

d'amélioration de la performance énergétique (APE). La définition de la situation de référence avant travaux est centrale dans le processus de GPE relative.

La méthode décrite dans ce chapitre sera concrètement illustrée sur le projet de réhabilitation d'un bâtiment de logement collectif présenté au chapitre 1. Les cas de GPE absolue et relative seront étudiés sur ce même projet. Cet exemple sera mis en œuvre de manière simplifiée dans un objectif pédagogique et il n'a pas ici vocation à présenter une utilisation réaliste et exhaustive du processus. Des cas d'étude poussés seront détaillés au chapitre 5 de cette thèse dans cet objectif.

3.2. La simulation du modèle physico-probabiliste

3.2.1. Hypothèses et objectifs de la démarche

La modélisation du bâtiment étudié grâce à un outil de simulation énergétique dynamique (SED) constitue la première étape de la démarche. Dans une approche théorique, la valeur de chacun des nombreux paramètres de ce modèle doit être spécifiée sous la forme d'une loi de probabilité, que ce soit sur les paramètres physiques et techniques du bâti ou des équipements énergétiques. En pratique, on verra que seuls les paramètres les plus influents nécessitent d'être décrits de cette manière, les autres pouvant être définis par une valeur nominale.

Cette première phase de la méthodologie a pour but de prendre en compte et de modéliser l'« ensemble des possibles ». En faisant l'hypothèse d'une précision suffisante du modèle physique, les cas simulés, suite à des tirages aléatoires des valeurs des entrées du modèle correspondent à des réalisations concrètes possibles des conditions annuelles d'exploitation d'un bâtiment et c'est de cette manière qu'il faut les considérer dans la suite du déroulement de la méthodologie. Par exemple, toute valeur d'une grandeur d'intérêt issue d'une simulation sera considérée comme une « valeur vraie » mais la connaissance réelle de cette valeur dépendra des moyens de mesure de la grandeur qui sont disponibles. Dans ces « expériences numériques » (Iooss, 2011b), les « conditions d'expérience » sont complètement maitrisées, mais c'est uniquement le modèle qui est caractérisé et non la réalité physique qu'il représente, d'où la nécessaire validation de ces modèles. Cette validité constitue ainsi une des deux hypothèses fortes de la méthodologie avec la juste caractérisation des fonctions de densité de probabilité des différentes entrées.

3.2.2. Analyse de sensibilité

La mise en œuvre concrète de cette modélisation peut être très couteuse en temps du fait des nombreuses informations nécessaires à la caractérisation des incertitudes portant sur les multiples

paramètres associés aux bâtiments. De ce fait, il est pertinent de découper cette phase en deux temps.

Dans un premier temps, des plages d'incertitude grossières (lois uniformes) peuvent être définies pour tous les paramètres avec des valeurs par défaut. Des méthodes d'analyse de sensibilité, et notamment de criblage, permettent d'effectuer un premier tri entre les entrées incertaines en identifiant les variables influentes sur la grandeur d'intérêt de sortie. Les variables dont l'influence est négligeable, selon un seuil et des modalités à définir peuvent être considérées avec une valeur nominale unique. Dans un second temps, un effort plus fourni doit être réalisé pour décrire au mieux les incertitudes portant sur les variables influentes du modèle, en termes à la fois de lois utilisées et de bornes considérées.

Des méthodes d'analyse de sensibilité plus précises mais plus coûteuses en temps de calcul, basées sur l'analyse de la variance peuvent éventuellement être mises en œuvre afin de quantifier avec plus de précision l'influence relative des entrées incertaines. Ceci permet d'obtenir une information plus précise qui peut aussi être utile à d'autres phases du projet, en particulier au cours du commissionnement où le suivi de la qualité de la réalisation peut être orienté vers les éléments critiques.

3.2.3. Propagation des incertitudes et variabilités

L'exploitation du modèle physico-probabiliste passe par la simulation du bâtiment dans un grand nombre de conditions différentes. Pour chaque simulation, des tirages sont réalisés dans les lois de probabilité des paramètres incertains, et des sollicitations d'entrée sont générées. La réalisation de simulations multiples va permettre l'étude statistique des sorties du modèle et des dépendances entre différentes grandeurs d'intérêt. On parle d'analyse d'incertitude ou d'approche Monte-Carlo pour décrire ce processus, mais nous utiliserons plutôt dans ces travaux le terme de propagation des incertitudes et variabilités (PIV).

Le chapitre 1 détaille plus largement ces différents aspects et les raisons de la différenciation entre incertitudes sur les paramètres de conception du bâtiment et variabilité naturelle des sollicitations de l'ouvrage. Cette dernière est considérée grâce à des modèles spécifiques générant des données d'entrée pour le modèle de SED. Ceci concerne les aspects météorologiques sollicitant le modèle via la température extérieure et le rayonnement solaire principalement mais aussi les aspects liés au comportement des occupants. Ceux-ci ont un impact sur l'utilisation d'appareils électriques, sur les puisages d'eau chaude et froide, sur les températures de consigne de chauffage et de refroidissement, sur l'éclairage mais aussi par leurs propres apports de chaleur métabolique. Les modèles doivent pouvoir générer une variabilité de situation représentative des conditions probables qui auront lieu pendant l'exploitation des bâtiments étudiés.

La grandeur d'intérêt de sortie *Y* de la PIV peut être la consommation d'énergie finale totale ou être restreinte à un ou plusieurs postes énergétiques en particulier. On peut aussi imaginer que la garantie porte directement sur des factures énergétiques, *Y* désignant alors le coût. On peut aussi s'intéresser à d'autres grandeurs d'intérêt fournies par les simulations ; on verra que les facteurs d'ajustement sont un exemple de grandeurs d'intérêt issues des PIV. La Figure 3-1 schématise le principe des PIV.

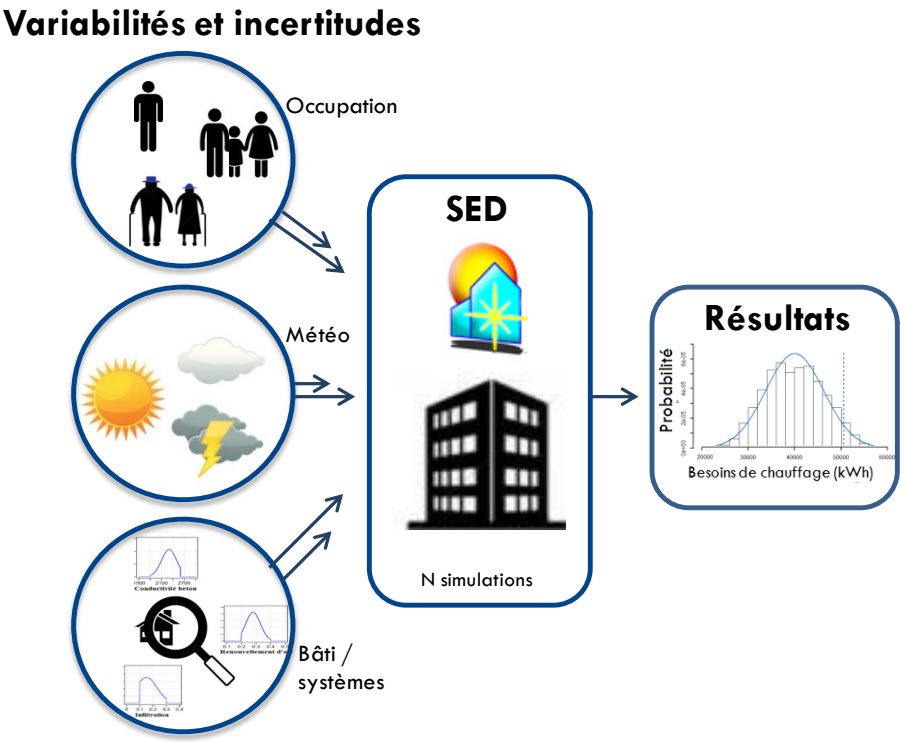

Figure 3-1 : Principe général de la propagation des incertitudes et variabilités (PIV)

3.2.4. Développement et calibrage du modèle pre-travaux

La caractérisation des performances du bâtiment avant une réhabilitation est nécessaire pour un engagement relatif afin notamment de fournir un modèle des consommations énergétiques de référence ajustable. Ce sujet a été abordé au § 1.4.2 du chapitre 1 et sera détaillé au § 3.3 de ce chapitre.

Pour des projets où le bâtiment est assez bien connu, il est recommandé de réaliser une modélisation SED pour étudier au mieux la performance énergétique. Une PIV est réalisée de la même manière que décrite précédemment afin d'obtenir une distribution probabiliste des consommations dans la situation avant rénovation.

Des méthodes de calibrage peuvent aussi être utilisées afin d'affiner le modèle physico-probabiliste d'un bâtiment existant. Ces méthodes sont abordées au chapitre 1, § 1.3.5. Le modèle du bâtiment avant travaux décrit la situation de référence, et si des mesures sont disponibles, il est possible de confronter celles-ci aux résultats du modèle afin d'améliorer la représentativité de ce dernier. En effet, le modèle du bâtiment réhabilité peut être construit à partir de celui avant travaux, une partie du bâti n'étant pas impactée par les actions d'APE. Cette étape de calibrage, si elle est réalisable, permet ainsi d'améliorer les capacités de prédiction du modèle de simulation utilisé pour définir l'engagement de GPE.

Pour illustrer cette première étape de la méthodologie, deux modèles de SED du bâtiment de logements de Feyzin ont été créés correspondant aux états de l'ouvrage avant et après la réhabilitation. La description de ce bâtiment est présentée au § 1.6.2 du chapitre 1. Du fait du peu de données de mesure disponibles avant la réalisation des travaux, le modèle pre-réhabilitation n'a pas été calibré. Des propagations d'incertitude et variabilités ont été effectuées (deux fois 1000 simulations). Le Tableau 3-1 présente les paramètres incertains des modèles SED.

Tableau 3-1 : Bornes des lois de distribution uniformes considérées sur les paramètres incertains des modèles SED avant et après réhabilitation

Paramètres communs	***Borne inf.***	***Borne sup.***
Conductivité matériaux (4-8)	*- 5 %*	*+ 5 %*
Masse volumique matériaux (4-8)	*- 1 %*	*+ 1 %*
Chaleur spécifique matériaux (4-8)	*- 5 %*	*+ 5 %*
Épaisseur matériaux dans parois (10-12)	*- 10 %*	*+ 5 %*
Hauteur parois (4-5)	*- 1 %*	*+ 1 %*
Largeur parois (4-5)	*- 1 %*	*+ 1 %*
Facteur solaire vitrages (3)	*- 5 %*	*+ 5 %*
U vitrages (3)	*- 5 %*	*+ 5 %*
Hauteur vitrages (3)	*- 1 %*	*+ 1 %*
Largeur vitrages (3)	*- 1 %*	*+ 1 %*
Ponts linéiques (6-5)	*- 10 %*	*+ 10 %*
Ponts menuiseries (2)	*- 10 %*	*+ 10 %*
Inflitrations (17)	*- 20 %*	*+ 20 %*
Débits de ventilation (16)	*- 10 %*	*+ 10 %*
Rendement chauffage (post-r)	*0,7*	*0,95*

Des lois uniformes sont considérées et les bornes choisies sont décrites dans le tableau ci-dessus. Le nombre d'éléments concernés est inscrit entre parenthèses pour les modèles pre et post rénovation quand ils diffèrent, donnant une idée de la dimensionnalité de la propagation

d'incertitude. Les valeurs nominales et le détail du modèle de simulation sont présentés en annexe A-5-1. Ce cas d'étude a une vocation illustrative, des lois simples sont considérées et aucune analyse de sensibilité permettant la réduction du problème n'a été menée. Une variabilité météorologique est par ailleurs considérée à partir du fichier météo type de Mâcon. Le modèle de variabilité stochastique des comportements des occupants est aussi utilisé, en renseignant le nombre de personnes par appartement (chapitre 1, § 1.2.2.2).

Dans une perspective de GPE absolue, seuls les résultats des simulations du modèle post-réhabilitation seront exploités. Concernant l'approche en GPE relative, nous nous intéresserons aux deux jeux de résultats. La Figure 3-2 présente les distributions des consommations énergétiques de chauffage avant et après travaux, ainsi que leur moyenne et quantile à 95%.

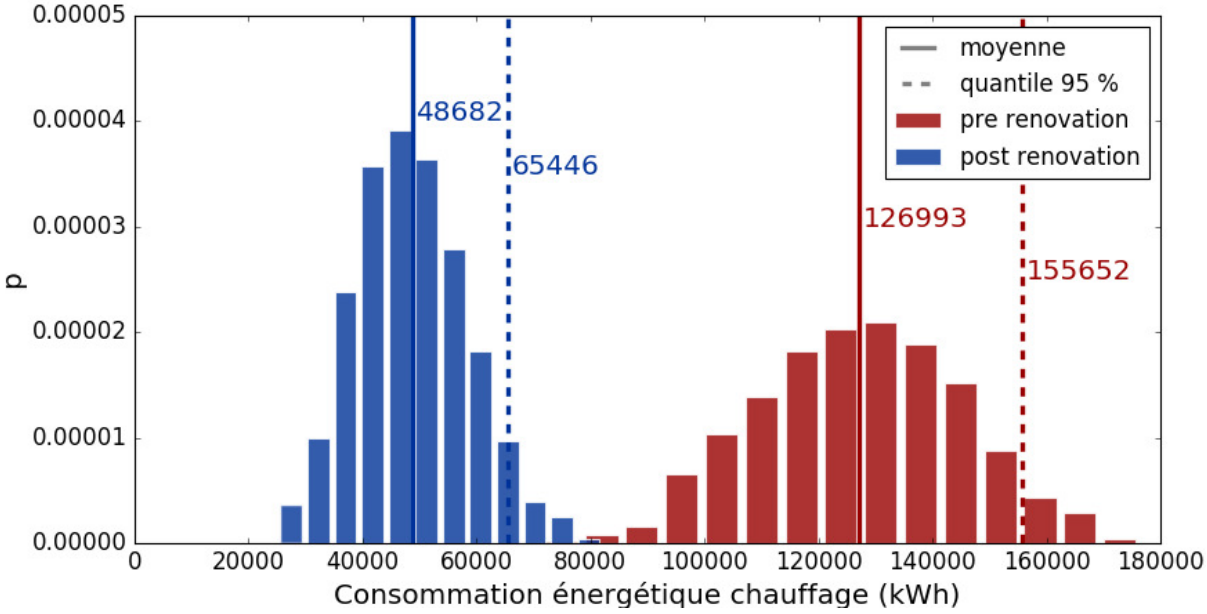

Figure 3-2 : Distribution des consommations énergétiques de chauffage avant et après réhabilitation

La définition et la simulation du modèle physico-probabiliste du bâtiment étudié constitue donc la première phase de la méthodologie. Elle est primordiale car la validité de ces conclusions repose sur la qualité de la description des modèles. Les spécificités différenciant les cas de construction neuve ou de réhabilitation ont été abordées et d'autres divergences seront à noter dans la suite du processus.

3.3. Ajustement

3.3.1. Formulation et intérêt du modèle

La propagation des incertitudes et variabilités au sein du modèle permet de caractériser la réponse probabiliste de celui-ci. Dans une problématique de prévision des consommations d'énergie des bâtiments, la distribution de cette grandeur d'intérêt de sortie se caractérise par une importante dispersion des résultats. Il n'est donc pas pertinent de s'appuyer sur cette distribution statistique brute pour associer une valeur garantie à un risque de défaut car soit cette valeur est trop élevée pour un risque raisonnable, soit le risque est trop grand pour un engagement ambitieux. Il n'est effectivement pas facile de justifier qu'une consommation énergétique garantie soit supérieure de 50 ou 80 % à une valeur nominale fournie habituellement par une SED unique.

La littérature scientifique et de simples études d'analyse de sensibilité montrent qu'une part importante de cette variabilité est due à des facteurs qui ne sont pas liés aux qualités intrinsèques du bâtiment, de ses systèmes, de son exploitation : les conditions météorologiques et d'usage. La GPE devant porter sur la performance énergétique, il est justifié de ne pas intégrer ces éléments dans la prise de risque liée à l'engagement. Ceci passe par la prise en compte d'une contextualisation de l'engagement en fonction de ces facteurs ; on parlera de processus d'ajustement.

Cet ajustement est étudié dans un contexte de GPE absolue ou relative. Dans tous les cas, il passe par la caractérisation du lien entre les consommations d'énergie d'un bâtiment et les sollicitations extérieures dont on veut isoler l'influence. Dans ces travaux, et toujours en s'appuyant sur la notion d'expérience numérique développée précédemment, cette caractérisation est réalisée via l'étude des résultats de simulation et des liens statistiques entre la sortie Y et les grandeurs d'intérêt X liées aux conditions météorologiques et d'usage que nous appellerons facteurs d'ajustement (FA). Ces grandeurs seront rassemblées dans la suite de ce chapitre dans le vecteur X des grandeurs explicatives. On parlera des n-uplet (X, Y) des grandeurs d'intérêt. L'équation 3.1 ci-dessous décrit la logique sur laquelle repose l'ajustement où le fait d'expliquer les variations de Y par les paramètres d'ajustement X permet de s'intéresser aux résidus ε dont la variabilité est plus faible.

$$Y = f(X) + \varepsilon \tag{3.1}$$

3.3.2. Définition et choix des facteurs d'ajustement

Les grandeurs d'intérêt et paramètres d'ajustement considérées sont associés à une période donnée. On s'intéressera majoritairement à des périodes annuelles sur les engagements de GPE dans ces travaux. Les FA sont des nombres scalaires calculés à partir de grandeurs dynamiques, dites

grandeurs d'ajustement, liées au climat et au comportement des usagers. L'information relative à ces grandeurs d'ajustement est ainsi agglomérée en une variable unique, le facteur d'ajustement (FA).

Les facteurs d'ajustement :

- doivent être associés à des grandeurs quantifiables et mesurables (en effet, un plan de mesure est automatiquement associé à un FA, ce qui permettra d'obtenir sa valeur réelle en exploitation) ;
- ne doivent pas être liés à des éléments relevant de la responsabilité de l'entreprise contractante sur la performance énergétique du bâtiment ou des systèmes ;
- doivent être associés à une grandeur ayant un impact important sur la variabilité de la grandeur d'intérêt Y (en effet, plus les grandeurs d'ajustement sont influentes, plus une part importante de la variabilité globale de la grandeur d'intérêt est expliquée par le modèle d'ajustement).

Dans un projet, le choix des facteurs d'ajustement est un exercice difficile, dépendant de différentes contraintes. Certains FA sont plus efficaces pour décrire l'influence d'une grandeur que d'autres, tous n'ont pas les mêmes qualités explicatives de la variabilité sur la sortie d'intérêt. Il sera pertinent de choisir les FA les plus performants de ce point de vue.

Des contraintes sont aussi liées au plan de mesure qui sera mis en œuvre pour vérifier les valeurs réelles des FA, des moyens de collecte, de gestion et de traitement des données. La complexité, le coût, l'acceptabilité par le client peuvent entrer en jeu. Peuvent ainsi se poser des questions liées au respect de la vie privée si des moyens de mesure de la présence ou des activités sont déployés.

La complexité de la procédure d'ajustement, si trop de facteurs sont pris en compte et doivent être mesurés, peut aussi rebuter certains clients. Il sera donc nécessaire de traiter cette question en fonction des spécificités de chaque projet, même si des méthodes d'aide à la décision et d'optimisation peuvent être utilisées (chapitre 4).

Afin de présenter la suite de la méthodologie, deux cas de figure sont considérés. Premièrement (configuration A), un ajustement est envisagé uniquement en fonction des températures extérieures, via les DJU, ce qui correspond à un exemple classique et facile à mettre en place, ne nécessitant éventuellement aucune instrumentation pour la vérification (données Météo-France). Un second cas est étudié (configuration B) avec un ajustement de la garantie en fonction des degrés-heures à base 18 sur la saison de chauffe $DH_{18.sc}$ (théoriquement plus précis que les DJU_{18}) et de la consommation d'énergie électrique moyenne annuelle, impactant les apports de chaleur internes dans les zones chauffées. Cette différenciation va permettre d'illustrer les écarts entre différentes modalités d'ajustement et d'étudier l'influence du nombre de dimensions retenues notamment en ce qui concerne la visualisation des résultats.

3.3.3. Méthodes de calcul des modèles d'ajustement

En se rapportant à l'équation 3.1, le paragraphe précédent traitait des grandeurs explicatives X (facteurs d'ajustement) à prendre en compte dans le processus d'ajustement. La grandeur d'intérêt de sortie Y est liée à la consommation énergétique d'intérêt. Le processus d'ajustement peut être appliqué de manière différente en fonction des configurations retenues pour l'engagement de GPE. Les sorties Y concrètement utilisées pour les cas de GPE absolue et relative, seront décrites au paragraphe suivant. Dans cette partie, on s'intéresse aux modèles f les plus adaptés. Leurs spécificités sont décrites en regard des besoins associés à notre problématique.

3.3.3.1. La régression linéaire - espérance conditionnelle

Les méthodes de régression statistiques ont pour objectif d'analyser les relations entre une variable d'intérêt et différentes variables explicatives. La régression linéaire multiple d'une variable aléatoire Y en fonction du vecteur de variables aléatoires X fournit la moyenne conditionnelle de Y sachant X, décrite selon un modèle linéaire du premier ordre. On peut ainsi décrire l'ensemble des n réalisations de la variable Y comme la somme du résultat du modèle de régression et d'un terme d'erreur.

$$Y = X\beta + \varepsilon \tag{3.2}$$

avec :

- Y vecteur colonne de dimension *n* (nombre de simulations) de la grandeur d'intérêt;
- X matrice des *p* variables explicatives de dimension *n* × *p ;*
- β vecteur colonne de dimension *p* composé des coefficients de régression β_j ;
- ε vecteur colonne des erreurs (résidus), de dimension n.

Pour s'assurer de la pertinence d'un modèle de régression linéaire, il est nécessaire de vérifier la véracité des hypothèses sur les résidus, qui doivent être de moyenne nulle et d'écart-type constant.

La méthode des moindres carrés permet d'estimer la valeur des coefficients du vecteur β. L'estimateur associé $\hat{\beta}$ est défini de la façon suivante :

$$\hat{\beta} = arg \min_{\beta \in \mathbb{R}^p} \|Y - X\beta\| = arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} \beta_j x_{ij})^2 \tag{3.3}$$

Cet estimateur minimise le carré des écarts entre les valeurs des observations de la grandeur à expliquer et l'estimation de ces valeurs par le modèle linéaire. Une solution analytique des valeurs des coefficients peut être calculée et différents algorithmes permettent de résoudre le problème numériquement. Le modèle de régression linéaire multiple fournit un estimateur de la moyenne conditionnelle de la grandeur d'intérêt à partir des observations disponibles. Le coefficient de

détermination R^2 est une mesure de la qualité de la prédiction d'une régression linéaire. Il correspond au rapport entre la variance expliquée par le modèle et la variance totale. Sa valeur est comprise entre 0 et 1, plus il est proche de 1, plus le modèle est bon.

$$R^2 = \frac{\left\|X\hat{\beta} - Y\right\|}{\|Y - \bar{Y}\|} \tag{3.4}$$

3.3.3.2. La régression quantile

Il est cependant possible de ne s'intéresser non pas à la moyenne conditionnelle mais à n'importe quel élément de la distribution conditionnelle de Y sachant X. On souhaite obtenir un quantile de cette distribution conditionnelle afin de caractériser plus finement la dispersion de l'échantillon. (Koenker et Bassett, 1978) proposent une méthode de régression linéaire fournissant les quantiles conditionnels d'une distribution afin d'étudier son hétéroscédasticité par rapport à une variable explicative couplée, c'est-à-dire la dépendance entre la valeur de cette variable explicative et la dispersion conditionnelle de la variables expliquée (constance de l'écart-type). Ces travaux ont donné lieu à de nombreuses applications, notamment en sciences sociales et économiques, et les bases théoriques associées peuvent être trouvées dans l'ouvrage de (Koenker 2005).

On s'intéresse à la fonction de répartition F_Y de la variable Y définie comme :

$$F_Y(y) = P(Y \leq y) \tag{3.5}$$

Le quantile d'ordre τ est alors :

$$q_\tau(Y) = {F_Y}^{-1}(\tau) \tag{3.6}$$

La régression quantile vise à définir un modèle linéaire liant les quantiles de la grandeur d'intérêt connaissant les valeurs des variables explicatives.

$$q_\tau(Y|X) = X\beta_\tau \ + \ \varepsilon \tag{3.7}$$

L'estimateur $\widehat{\beta_\tau}$du vecteur β_τ de la régression quantile est :

$$\widehat{\beta_\tau} = arg \min_{\beta \in \mathbb{R}^p} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i\beta_\tau) \tag{3.8}$$

avec ρ_τ, une fonction test définie par :

$$\rho_\tau(u) = \begin{cases} \tau.u & \{u \geq 0\} \\ (\tau - 1).u & \{u < 0\} \end{cases} \tag{3.9}$$

Par exemple, pour $\tau = 1/2$ correspondant à la médiane, la fonction est équivalente à la valeur absolue. On parle ainsi de l'estimateur des moindres erreurs absolues (*Least Absolute Deviation Estimator*) utilisé comme une alternative aux moindres carrées car moins sensibles aux valeurs extrêmes des observations.

La pertinence des méthodes de régression linéaire est liée à différentes hypothèses. Le non-respect de ces hypothèses peut entrainer l'obtention d'estimateurs biaisés ou non efficaces. La non-colinéarité des variables explicatives et l'indépendance des erreurs résiduelles sont des hypothèses fortes. La moyenne des erreurs doit être nulle. La normalité de leur distribution n'est pas obligatoire mais permet de s'assurer des bonnes propriétés du modèle. L'hypothèse d'homoscédasticité, c'est-à-dire d'indépendance de la variance des résidus par rapport aux variables explicatives est conseillée pour des régressions linéaires. Dans notre cas, on constate la présence de distributions non homoscédastiques et la régression quantile permet justement de capturer ces effets d'hétéroscédasticité. Les modèles de quantile conditionnel étudiés ne sont donc pas nécessairement colinéaires au modèle de régression linéaire sur l'espérance. La Figure 3-3 présente un exemple de régressions linéaires moyenne et quantile ($\tau = 0{,}95$) dans le cas où une seule variable explicative est considérée.

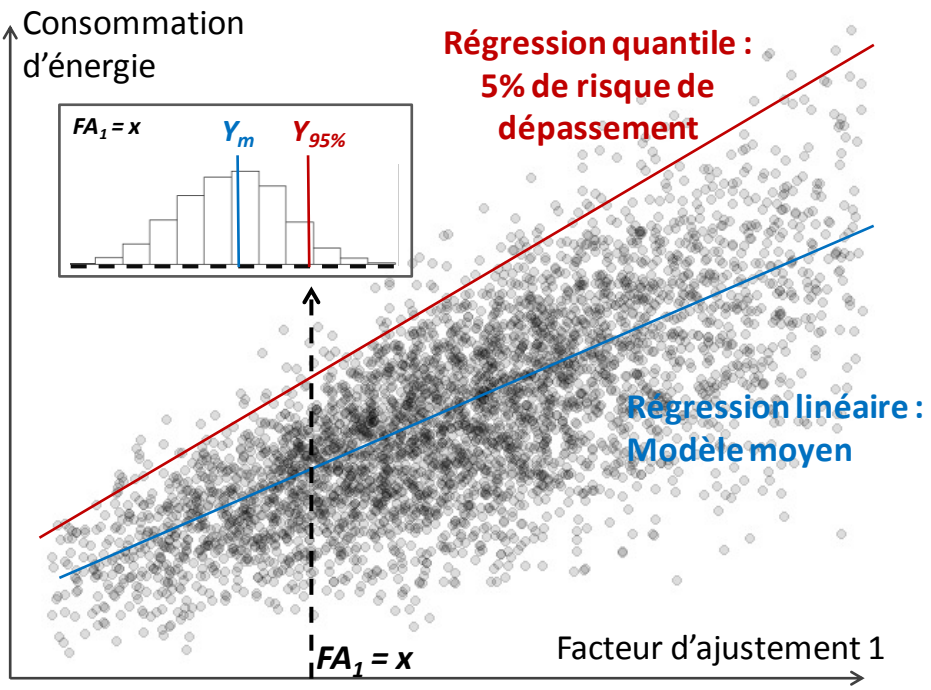

Figure 3-3 : Exemple de modèles linéaires issus de régressions moyenne (bleu) et quantile (rouge)

En se rapportant au contexte de la GPE, on peut écrire le modèle linéaire obtenu par régression linéaire (simple ou quantile).

$$Y = \beta'_0 + \sum_{i=1}^{n} \beta'_i . FA_i \quad \textbf{(3.10)}$$

L'expression d'un modèle d'ajustement sous la forme d'un polynôme du premier ordre doit être justifiée par la vérification de l'hypothèse de linéarité entre la grandeur d'intérêt et les différentes sorties. Dans le cas des consommations énergétiques des bâtiments, on constatera que cette hypothèse est souvent acceptable dans la pratique, du fait des caractéristiques régulières de la majorité des phénomènes physiques mis en jeu. D'un point de vue industriel, elle permet une meilleure compréhension et donc une meilleure acceptabilité du principe de l'ajustement par les cocontractants, donnant une impression de plus grande transparence. On se propose d'exprimer le modèle d'ajustement en fonction à la fois de facteurs d'ajustement mais aussi de leurs valeurs dans une situation de référence :

$$Y = \beta_0 + \sum_{i=0}^{n} \beta_1 . \frac{FA_i - FA_{i-ref}}{FA_{i-ref}} \quad \textbf{(3.11)}$$

Ces valeurs et les conditions associées devront être explicitées dans le contrat. Ceci permet d'exprimer la grandeur d'intérêt Y en fonction du taux d'accroissement des FA par rapport à leur valeur de référence. L'ensemble des coefficients β sont ainsi sans dimension, leur valeur caractérisant la sensibilité de la grandeur d'intérêt aux variations de chaque FA.

La robustesse des modèles de régression identifiés sera discutée au chapitre 4. La détermination d'intervalles de confiance sur les valeurs des coefficients des modèles linéaires sera notamment étudiée.

3.3.3.3. Vers des formes de modèle plus complexes

D'autres types de modèles peuvent être utilisés afin de répondre à la problématique traitée dans ces travaux : la caractérisation d'un lien entre une grandeur expliquée et des facteurs explicatifs et une quantification d'un risque statistique associé.

Les modèles présentés ci-dessus sont paramétriques et linéaires ce qui les rend faciles à comprendre et à utiliser mais des structures de modèles moins courants et applicables industriellement à court terme ont d'autres avantages. Des alternatives à la régression quantile multiple seront présentées au chapitre 4.

3.3.4. GPE absolue : Détermination d'une limite garantie de consommation ajustable

Dans le cas d'une GPE absolue, l'ensemble de l'étude repose sur la modélisation et la simulation énergétique du bâtiment final. L'engagement de la GPE est exprimé comme une valeur limite des consommations d'énergie correspondant à la grandeur d'intérêt Y. On parlera de limite garantie des consommations d'énergie pour un risque de dépassement α de cette limite (LGC_α). L'objectif est ici de déterminer le modèle f_α fournissant la LGC_α en fonction du risque choisi et des FA sélectionnés (équation 3.12).

$$LGC_\alpha = f_\alpha(FA_1, \dots, FA_n) \quad \textbf{(3.12)}$$

En phase de vérification, la consommation réelle mesurée sera comparée à la LGC_α calculée dans les conditions d'ajustement mesurées en exploitation.

Pour le cas d'application en GPE absolue, on souhaite déterminer une limite garantie de consommation pour un risque de 5 %, ajustable en fonction des FA choisis. Les valeurs de référence considérées correspondent à la médiane dans la distribution des résultats de simulation du modèle avant travaux. L'équation ci-dessous présente le modèle ajustable de la $\mathrm{LGC}_{5\%}$ dans la première configuration d'ajustement (cas A).

$$LGC_{5\,\%} = 6{,}012.10^4 + 7{,}29.10^4 \left(\frac{DJU_{18}}{DJU_{18-ref}} - 1\right) \quad \textbf{(3.13)}$$

avec $DJU_{18-ref} = 2292\ °\mathrm{C.h}$, valeur médiane des résultats de la PIV avant travaux

L'équation ci-dessous présente le modèle ajustable de la $\mathrm{LGC}_{5\%}$ dans la seconde configuration d'ajustement (cas B).

$$LGC_{5\%} = 5{,}955.10^4 + 8{,}57.10^4 * \left(\frac{DH_{18.sc}}{DH_{18.sc-ref}} - 1\right) - 1{,}14.10^4 * \left(\frac{E_{elec.sc}}{E_{elec.sc-ref}} - 1\right) \quad \textbf{(3.14)}$$

avec $DH_{18.sc-ref} = 49560\ °C.h$ et $E_{elec.sc-ref} = 25550\ kWh$.

La Figure 3-4 propose une représentation graphique de ces modèles linéaires obtenus par régression quantile, fournissant les quantiles conditionnels des distributions des consommations de chauffage. On constate que les valeurs garanties de référence sont assez proches dans les deux modèles. Ceci s'explique en partie par l'effet limité que semblent avoir les consommations électriques sur les besoins de chauffage. Cela est confirmé par la valeur du coefficient associé à ce FA dans le cas B, dont la valeur absolue est beaucoup plus faible que pour les $DH_{18.sc}$. Visuellement, la pente du modèle linéaire est plus faible dans cette dimension d'ajustement (à droite de la figure).

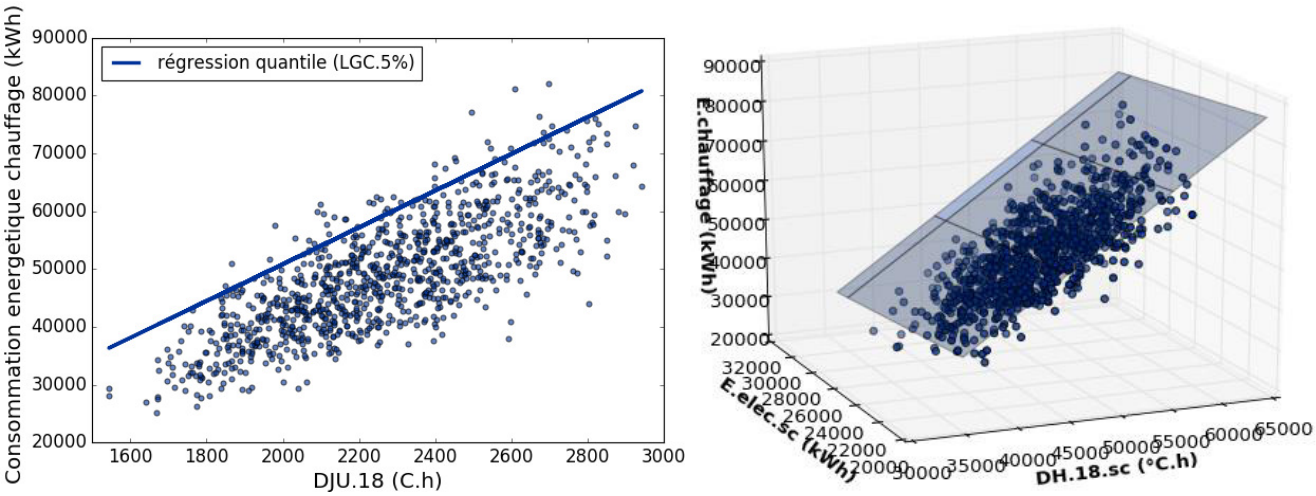

Figure 3-4 : représentation graphique de la $LGC_{5\%}$ sur les consommations énergétiques de chauffage obtenue par régression quantile dans les cas A (gauche) et B (droite)

3.3.5. GPE relative : détermination du gain et du modèle de référence ajustable

La GPE relative se caractérise par un engagement sur une amélioration de la performance énergétique, ou plus concrètement sur une réduction des consommations d'énergie par rapport à la situation antérieure à une action d'amélioration de la performance énergétique (APE). C'est en pratique l'application la plus couramment utilisée de la GPE aujourd'hui (ADEME et FBE, 2016).

3.3.5.1. Le modèle de référence pre travaux ajustable

La problématique de la définition de cette situation antérieure à la réhabilitation, dite « de référence », a été abordée au § 1.4.2 du chapitre 1. La difficulté de cette approche de GPE relative vient de la volonté de comparer une situation post travaux réelle et mesurée à une situation pre travaux théorique dans des conditions identiques.

Pour cela, il est nécessaire de déterminer un modèle fournissant la consommation énergétique attendue (espérée) dans la situation de référence en fonction des éléments dont on veut s'affranchir : les FA. L'utilisation de ce modèle g permettra de calculer la consommation de référence et donc le gain relatif de consommation dans les conditions d'ajustement observées en exploitation.

$$E_{pre} = g(FA_1, \dots, FA_n) \quad \textbf{(3.15)}$$

Ce modèle doit être basé sur les données pre travaux disponibles de consommations d'énergie et des FA que l'on souhaite considérer dans l'étude. Il peut provenir d'un modèle SED

calibré ou être construit directement (modèle statistique, modèle affine ou même linéaire) à partir des données collectées.

Il est important de constater que, même s'il est construit à partir des informations historiques de consommations énergétiques, ce modèle est « arbitraire » et les valeurs de référence calculées après la réhabilitation seront par nature invérifiables. C'est une base de comparaison théorique. Deux niveaux d'engagement sur un gain relatif de performance, provenant de deux offres concurrentes par exemple, ne peuvent être comparés que si les modèles d'ajustement de référence considérés sont équivalents.

En GPE relative, il est nécessaire de fournir le modèle d'ajustement de la situation de référence (espérance conditionnelle de la consommation avant travaux). Les valeurs de référence des facteurs d'ajustement sont ceux définis plus hauts, valeurs médianes des résultats de la PIV post-réhabilitation afin de pouvoir comparer les modèles. L'équation ci-dessous présente le modèle d'ajustement de la consommation de référence dans la première configuration, calculé à partir des résultats de la PIV avant réhabilitation :

$$E_{pre} = 1{,}277.10^5 + 1{,}30.10^5 * \left(\frac{DJU_{18}}{DJU_{18-ref}} - 1\right) \quad \textbf{(3.16)}$$

avec $DJU_{18-ref} = 2292\ \text{°C.h}$, valeur médiane des résultats de la PIV.

Le modèle d'ajustement de la consommation de référence dans la seconde configuration d'ajustement (B) est le suivant :

$$E_{pre} = 1{,}275.10^5 + 1{,}48.10^5 * \left(\frac{DH_{18.sc}}{DH_{18.sc-ref}} - 1\right) - 2{,}78.10^4 * \left(\frac{E_{elec.sc}}{E_{elec.sc-ref}} - 1\right) \quad \textbf{(3.17)}$$

avec $DH_{18.sc-ref} = 49560\ \text{°C.h}$ et $E_{elec.sc-ref} = 25550\ \text{kWh}$, valeurs médianes des résultats de la PIV.

Les coefficients obtenus peuvent être comparés à ceux des équations 3.13 et 3.14 de la $LGC_{5\%}$ après travaux. On constate notamment une plus grande sensibilité aux températures extérieures (DJU et DH) avant travaux, tout comme aux consommations électriques $E_{elec.sc}$. La Figure 3-5 présente les résultats de simulation et les modèles linéaires moyens (E_{pre}) et quantiles ($LGC_{5\,\%}$) associés.

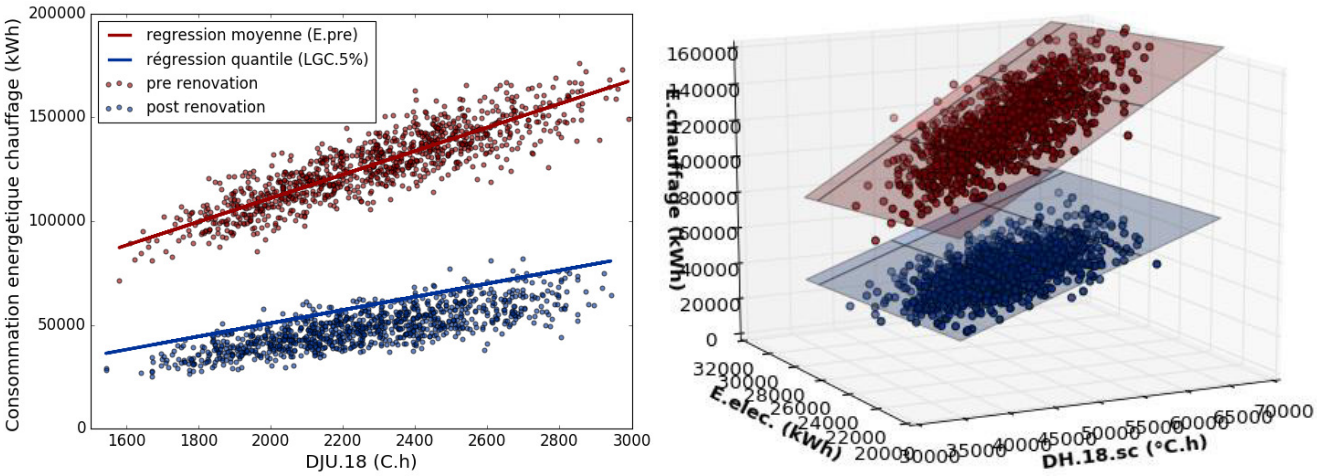

Figure 3-5 : Résultats de simulation et modèles d'ajustement (cas A à gauche, cas B à droite) des consommations de chauffage avant (E_{pre} - rouge) et après ($LGC_{5\%}$ - bleu) réhabilitation

3.3.5.2. Le gain énergétique garanti

Dans la mise en œuvre des contrats de GPE, le gain garanti est usuellement une valeur constante fixée. L'objectif de la méthodologie est d'associer à cette valeur un risque de défaut, qui n'est que rarement calculé et exprimé dans les pratiques actuelles.

Comme nous l'avons décrit précédemment, la grandeur d'intérêt dans cette approche est le rapport de la différence entre la consommation énergétique de référence (E_{pre}) et la consommation réelle (E_{obs}) divisée par la consommation de référence, cette dernière étant calculée grâce aux FA observés (équation 3.18). On nommera cette grandeur d'intérêt le gain relatif (GR).

$$Y = GR = [E_{pre}(FA_{obs}) - E_{obs}] \,/\, E_{pre}(FA_{obs}) \qquad \textbf{(3.18)}$$

On étudie la distribution de cette grandeur d'intérêt par l'exploitation de la PIV du modèle post travaux, le modèle des consommations énergétiques de référence étant obtenu par la PIV du modèle pre travaux.

La validité de l'hypothèse qui suppose que ce gain (associé à un risque donné de défaut) est constant et ne dépend pas des valeurs des FA peut être remise en cause comme le montre la Figure 3-6 introduite par la suite. Le quotient de deux fonctions linéaires n'est pas linéaire et il n'est donc pas pertinent d'utiliser une régression quantile pour étudier la variabilité conditionnelle des gains relatifs en fonction des FA. Cependant, le modèle des consommations d'énergie de référence (E_{pre}) est déterministe ce qui permet donc d'exprimer le quantile conditionnel du GR avec un risque α de dépassement en fonction du quantile conditionnel sur la consommation après travaux (LGC_{α}). Ce seuil garanti de gain relatif pour un risque α (GR_{α}) s'exprime ainsi :

$$GR_{\alpha}(FA) = [\, E_{pre}(FA) - LGC_{\alpha}(FA) \,] \,/\, E_{pre}(FA) \tag{3.19}$$

On pourra donc s'engager, au risque α, sur ce seuil ajustable en fonction des FA ou calculer le quantile global $GR_{\alpha,glob}$ de la distribution des gains relatifs.

$$GR_{\alpha,glob} = q_{\alpha}(GR) \tag{3.20}$$

Dans ce cas, le risque α est un risque global, moyenné, qui ne tient pas compte de la connaissance des valeurs des FA. C'est donc un indicateur dégradé par rapport à GR_{α}.

La Figure 3-6 présente la distribution des gains relatifs (équation 3.18) exprimés en fonction des FA pour les cas A (gauche) et B (droite).

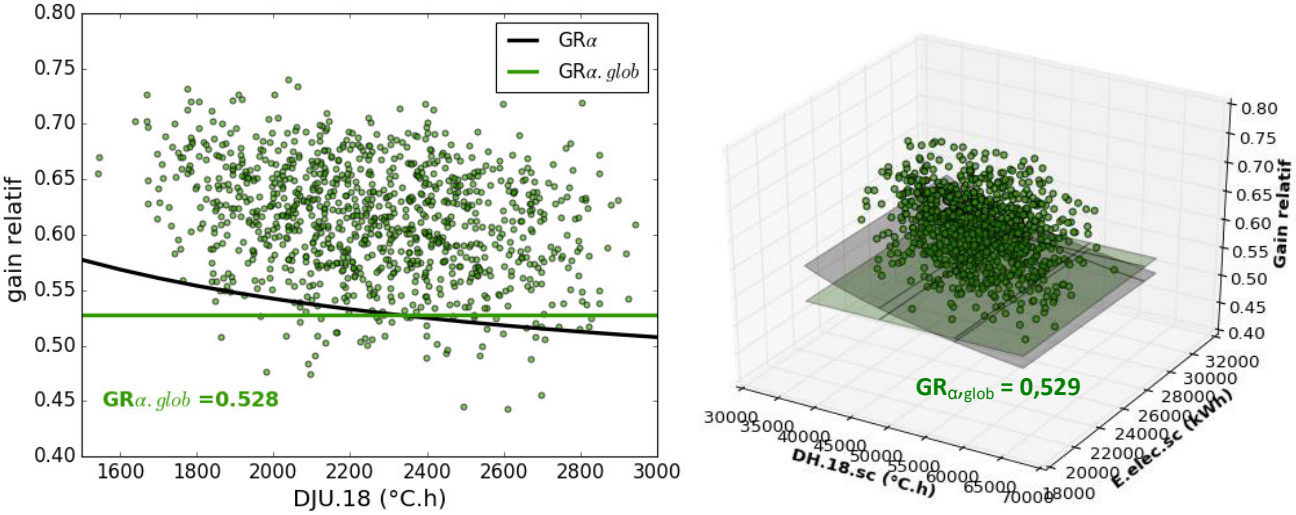

Figure 3-6 : Gains relatifs (vert) et valeurs limites garanties du gain relatif (α = 5 % de risque) global (vert) et ajustable (noir) (cas A à gauche, cas B à droite)

Le seuil garanti de gain relatif ajustable est évalué à partir du modèle linéaire moyen avant travaux (E_{pre}) et du modèle linéaire quantile des consommations après réhabilitation ($LGC_{5\,\%}$). On peut constater sa variabilité en fonction des valeurs des FA. En effet, il varie de 45 à 60 % sur les plages de variation des FA. Pour un risque fixé, la diminution des consommations garantie dépend donc bien des conditions d'exploitation. Ce résultat remet en cause les processus actuels d'engagement dans les CPE en réhabilitation où une valeur unique est considérée.

Le quantile global à 5 % de la distribution des gains relatifs est aussi calculé et fournit une valeur de $GR_{5\,\%,glob}$ moyennée (en vert). Il est possible de ne conserver que cette dernière valeur unique du quantile à 95 % de la distribution de la distribution complète des gains relatifs. Dans ce cas, il faut avoir conscience que le risque associé est un risque global et n'est pas constant sur tout l'espace des facteurs d'ajustement. Ce gain moyen pour un risque de 5 % est de 52,8 % dans le cas A et de 52,9 % dans le cas B.

3.3.6. Ajustement et vérification de la performance

La vérification de la garantie est réalisée après la période d'exploitation correspondant à la période d'engagement. Des compteurs indiquent les consommations d'énergie, et les facteurs d'ajustement peuvent être déterminés par l'exploitation de mesures réalisées en continu sur le bâtiment ou obtenues de manière externe (données météorologiques de Météo-France par exemple).

3.3.6.1. GPE absolue

Dans le cas d'une garantie exprimée en valeur absolue, la consommation énergétique réelle sera comparée à la valeur limite LGC_{α} calculée à partir du modèle et des valeurs des FA relevées. Afin de respecter l'engagement pris, la valeur de consommations réelles doit être inférieure à la LGC_{α} ajustée dans les conditions réelles.

Pour le cas d'étude, les mesures réalisées durant la première année d'exploitation nous renseignent sur les consommations énergétiques liées au chauffage et sur les valeurs réelles des FA (Tableau 3-2).

Tableau 3-2 : Résultats de mesures issues de la première année d'exploitation du bâtiment

Grandeurs d'intérêt	***valeurs***
Consommation énergétique de chauffage (kWh)	*64700*
DJU_{18} *(°C.j)*	*2213*
$DH_{18.sc}$ *(°C.h)*	*50968*
$E_{elec.sc}$ *(kWh)*	*19681*

Dans la première configuration d'ajustement (DJU_{18}), la limite garantie ajustée obtenue avec le modèle de l'équation 3.13 est de 57620 kWh. Dans le cas B (ajustement sur les $DH_{18.sc}$ et $E_{elec.sc}$), la $LGC_{5\%}$ obtenue (équation 3.14) est de 64599 kWh.

La garantie est loin d'être respectée dans la première configuration A. Dans la configuration B, la $LGC_{5\%}$ est légèrement inférieur à la consommation réelle. La consommation de chauffage mesurée correspond en effet aux valeurs extrêmes des résultats de simulation. Cela s'explique en partie par les températures de consigne très élevées lors de cette année d'exploitation (non ajustées ici) et par les faibles apports dus aux consommations électriques, pris en compte dans l'ajustement pour le deuxième cas et permettant ainsi de respecter la $LGC_{5\%}$. On constate donc que l'ajustement se fait à la baisse par rapport aux conditions de référence dans la configuration A, du fait de conditions météorologiques mesurées (DJU_{18}) un peu plus favorables. Pour la configuration B,

l'ajustement fait augmenter la $LGC_{5\%}$ par rapport à sa valeur de référence du fait des faibles consommations électriques mesurées.

3.3.6.2. GPE relative

Si l'engagement porte sur l'amélioration relative de la performance, la consommation énergétique de référence pre travaux est estimée grâce au modèle, à partir des valeurs réelles des FA calculées à partir des données observées sur la période d'exploitation. La consommation énergétique mesurée est comparée à cette situation de référence ajustée ce qui permet de calculer le gain effectif. Celui-ci est finalement comparé au gain garanti. L'ensemble des résultats pour l'engagement de garantie de performance sont présentés sur la Figure 3-7.

En GPE relative, les consommations de référence ajustées sont calculées (équations 3.16 et 3.17). On obtient respectivement des consommations de référence de 123194 kWh et 138131 kWh dans les deux configurations d'ajustement.

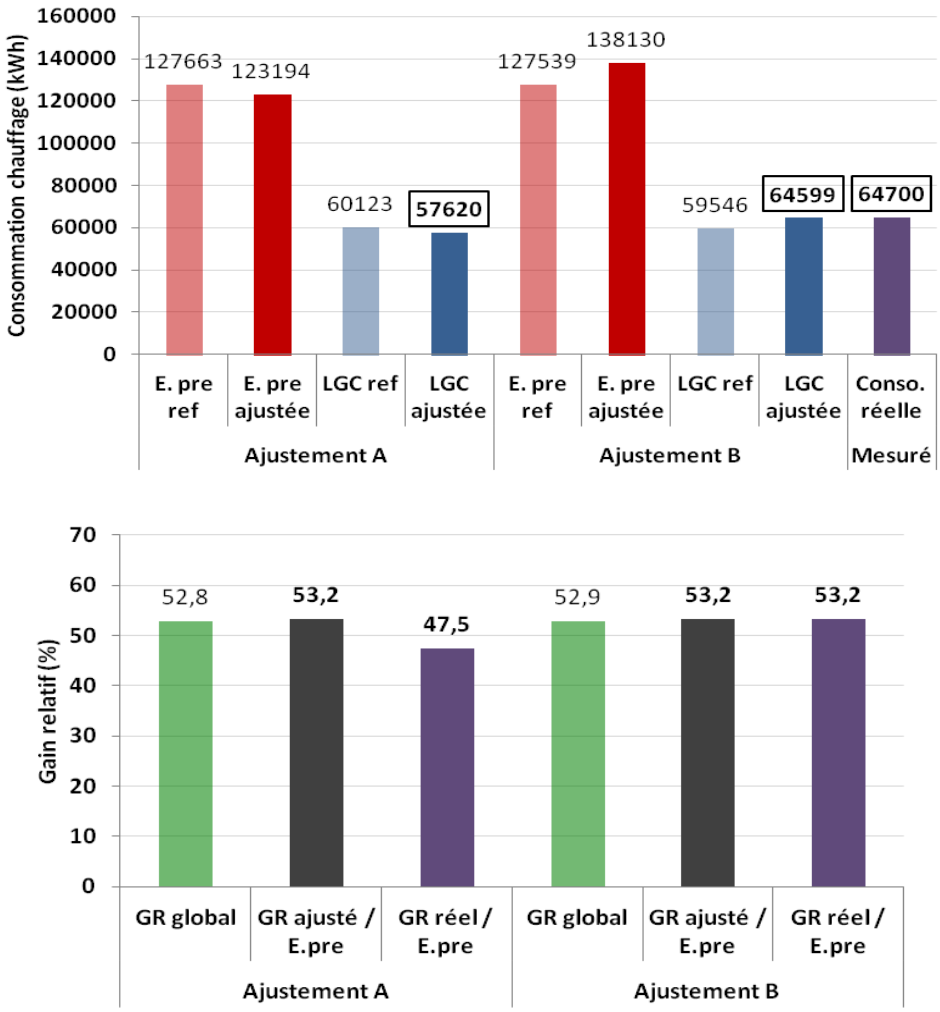

Figure 3-7 : Résultats du processus de garantie pour les deux modèles d'ajustement A et B

Les gains réels obtenus à partir des mesures de consommation énergétiques présentées dans le Tableau 3-2 sont donc pour les configurations A et B de 47,5% et 53,2 %. Ces valeurs peuvent être comparées au $GR_{5\%}$ ajusté en fonction des valeurs de FA. Celui-ci est respectivement de 53,2 % et de 52,3 % pour les deux modèles d'ajustement. On constate qu'ici non plus, la garantie n'est pas respectée dans le cas de l'ajustement aux DJU_{18}. Dans le cas B, le gain garanti est très proche mais légèrement supérieur à la réalité. Le $GR_{5\%,glob}$ est de 52,8 % (cas A) et 52,9 % (cas B) et est donc supérieur au gain réel dans le cas A et inférieur pour la configuration B.

L'impact des incertitudes de mesure lors de la phase de vérification n'est pas pris en compte dans ces résultats. On constate que l'estimation du risque obtenu par l'exploitation des données de simulation dans le cadre de la méthodologie présentée ci-dessus est remise en cause par l'existence d'un risque supplémentaire, lié aux mesures physiques lors de la vérification. En effet, les erreurs sur les mesures des consommations d'énergie et sur les mesures permettant le calcul des FA peuvent être responsables d'un non-respect faussé de la garantie. On se propose d'intégrer l'impact des incertitudes de mesure au sein de la méthodologie.

3.4. Intégration de l'influence des incertitudes de mesure

3.4.1. Enjeu et approche numérique

La caractérisation du risque de défaut repose sur l'hypothèse de la représentativité de la distribution statistique de la grandeur d'intérêt de sortie obtenue suite à la propagation des incertitudes et variabilités. En supposant que cette hypothèse soit respectée, cela permet de considérer les résultats d'une année d'exploitation future du bâtiment comme une réalisation statistique, équivalente à un tirage aléatoire au sein de notre modèle physico-probabiliste. En réalité, les expériences numériques, issues des simulations, et les expériences réelles ne peuvent être directement comparées du fait du filtre supplémentaire que constituent les erreurs de mesure dans le second cas. Ces incertitudes de mesure étant inévitables dans le processus de vérification, on souhaite anticiper leur impact sur le risque lié à la prévision de la performance énergétique. La grandeur d'intérêt dont on souhaite modéliser et étudier la variabilité statistique n'est plus la consommation énergétique effective mais la consommation énergétique mesurée (sous-entendu : connue avec une erreur de mesure).

En partant du principe que les simulations sont des réalisations possibles de la réalité et que les résultats qui en sont issues sont les valeurs vraies des grandeurs d'intérêt et explicatives, on considère que l'obtention de ces valeurs est entachée d'une erreur si elles sont mesurées. On propose d'intégrer à la modélisation la réalisation simulée d'une mesure des grandeurs d'intérêt. Cette mesure simulée, effectuée à chaque simulation énergétique (et éventuellement plusieurs fois

pour chaque simulation individuelle) transforme un ensemble de valeurs vraies en valeurs erronées. Ce bruitage des résultats de simulation est effectué via des tirages aléatoires dans les plages d'incertitude de mesure associées au plan de mesure prescrit, que ce soit pour les consommations énergétiques ou les facteurs d'ajustement.

La méthodologie prévoit donc, via la prescription d'un plan de mesure et d'un cahier des charges d'exploitation des données, la caractérisation des marges d'incertitude associées aux valeurs réelles des FA. On parlera de plan de mesure et vérification (PMV) pour décrire cet ensemble comprenant à la fois l'obtention de mesures brutes mais aussi leur post-traitement en vue d'obtenir les grandeurs finales souhaitées et nécessaires à la vérification de l'engagement.

C'est à partir des résultats de simulation bruités par les erreurs liées aux mesures que sont estimés l'engagement de la GPE et les modalités d'ajustement qui l'accompagnent tel que décrit aux paragraphes précédents. Le protocole complet intégrant cette étape sera résumé au § 4.5. dans les cas de GPE absolue et relative.

3.4.2. Estimation des incertitudes

Le choix des facteurs d'ajustement s'accompagne de la prescription des moyens de mesure qui permettront d'obtenir leurs valeurs réelles pendant l'exploitation du bâtiment. Ce plan de mesure comprend la description du matériel nécessaire, son positionnement dans l'espace, les détails des modes d'acquisition des données (pas de temps, protocole de récupération…) mais aussi le post-traitement des données brutes à partir desquelles sont calculées les valeurs des FA. Il est ainsi possible d'anticiper l'incertitude associée à ces valeurs, conséquence de toute la chaine d'obtention de l'information, des mesures brutes aux calculs.

Les incertitudes sur les grandeurs d'intérêt (consommation énergétique et FA) dépendent du plan de mesure et vérification, et donc à la fois de la qualité des mesures brutes mais aussi des opérations de post-traitement. Ainsi, l'agglomération de données temporelles en une grandeur unique, le FA, passe par différents traitements mathématiques (sommes, moyennes…) qui impactent l'incertitude finale sur cette grandeur d'intérêt. Des règles de calcul des incertitudes existent, elles ont été décrites au paragraphe 1.3 et sont détaillées dans le *GUM* (JCGM, 1995). Le choix d'un facteur d'ajustement, de la complexité des post-traitements des mesures associées, n'est donc pas anodin.

L'estimation des plages d'incertitude s'effectue selon les méthodes de mesures physiques et de calcul des incertitudes combinées. Les travaux de Titikpina et al. (2016; 2015a ; 2015b) et de (Spitz, 2012) traitent aussi de ces question dans le cadre de l'estimation des performances énergétiques des bâtiments. Nous abordons ici les problématiques principales qui peuvent être rencontrées. Les facteurs d'ajustement sont des variables agglomérant des informations dynamiques et sont donc

issues de mesures plus ou moins indirectes. L'équation ci-dessous lie l'incertitude sur une grandeur Y liée par une fonction f à des variables a_i mesurées.

$$u(Y) = \sqrt{\sum_{i=1}^{p} \left[\frac{\partial f}{\partial a_i} . u(a_i)\right]^2} \quad \textbf{(3.21)}$$

L'estimation de l'incertitude sur les grandeurs d'intérêt repose en partie sur cette loi de composition. Quelques cas typiques de grandeurs d'intérêt sont détaillés ci-dessous.

Consommations énergétiques

La mesure des consommations dépend de la source énergétique considérée. Sur des compteurs de gaz et d'électricité, l'incertitude-type est souvent exprimée en pourcentage de la valeur lue. S'il est spécifié qu'il s'agit d'une incertitude-type, un facteur d'élargissement $k = 2$ doit être considéré pour définir l'incertitude correspondant à un intervalle de confiance de 95 %. Les classes de précision des compteurs électriques correspondent à la marge d'erreur (à 95 %) qui leur est associée.

La mesure d'énergie calorifique (boucle d'eau chaude) est dotée d'une incertitude plus difficile à évaluer du fait des différentes mesures nécessaires, des températures et du débit, et de l'intégration de variations temporelles de ces grandeurs. Le calcul des incertitudes composées peut être réalisé mais une classe de précision globale est souvent associée aux systèmes, renseignant sur l'incertitude relative de l'énergie mesurée.

Températures d'air

Les moyennes de températures intérieures ou extérieures sur des périodes d'intérêt peuvent constituer des facteurs d'ajustements. L'incertitude d'une mesure de température ponctuelle dépend du type de capteur employé. Pour les sondes à résistance de platine, l'erreur maximale est exprimée en fonction d'une part fixe et d'une part proportionnelle à la valeur mesurée ; là aussi, des classes normalisées d'appareils existent. Avec un thermocouple, la précision du capteur est un pourcentage de la mesure ou une valeur absolue en Kelvin.

L'évaluation d'une température, si l'on s'intéresse à une température intérieure de zone par exemple, soulève des problématiques de variabilité spatiale de cette grandeur et donc plus concrètement du nombre et du positionnement des capteurs. Cette question a été abordée par (Caucheteux, 2015). L'évaluation de la température moyenne intérieure d'une zone comportant plusieurs pièces est réalisée avec une combinaison de capteurs. Pour minimiser l'incertitude, il conclut qu'il est pertinent de disperser des capteurs dans les différentes pièces, en fonction des ressources disponibles, et de mettre plusieurs thermomètres dans la même pièce.

Pour le calcul d'une moyenne temporelle des températures mesurées, les erreurs sont séparées en parts aléatoire et systématique. À partir de l'équation 3.21, l'influence des erreurs aléatoires indépendantes sur l'incertitude de la moyenne des températures $u_{al}(\overline{T})$ peut s'exprimer selon :

$$u_{al}(\bar{T}) = \sqrt{\sum_{i=1}^{n} \left[\frac{1}{n} u_{al}(T_i)\right]^2} \quad \textbf{(3.22)}$$

L'erreur systématique est constante durant la période de mesure et l'erreur sur la moyenne est donc elle aussi égale à cette erreur systématique

$$u_{sys}(\bar{T}) = u_{sys}(T_i) \quad \textbf{(3.23)}$$

En pratique, on négligera l'effet de l'erreur aléatoire dans le calcul de l'incertitude sur la moyenne. Pour obtenir une température de zone moyennée spatialement et temporellement, il est donc au final nécessaire de composer avec les incertitudes dans ces deux dimensions.

Données météorologiques

Les facteurs d'ajustement relatifs aux conditions météorologiques sont liés à la température extérieure et éventuellement au rayonnement solaire. Ces données peuvent être mesurées *in situ* ou obtenues auprès d'organismes relevant ces données tel Météo-France. Là aussi, l'agrégation temporelle de l'information et les traitements mathématiques (sommes, moyennes) impacte le calcul des incertitudes des grandeurs finales. La question de l'incertitude de la moyenne temporelle d'une température mesurée a été abordée plus haut. Concernant les degrés-heures et degrés jours unifiés, ces facteurs sont calculés avec un effet de seuil ce qui complique l'évaluation de l'incertitude de mesure. En effet, lorsque l'écart entre la température extérieure et la température de référence est inférieur à l'incertitude de la température mesurée, le pas de temps (jour ou heure) pourrait être considéré. L'incertitude sur le facteur final dépend du nombre de pas de temps pris en compte dans la somme.

Pour simplifier le problème, nous ne prenons pas en compte ces cas de recoupement dont l'effet est relativement plus faible que l'incertitude liée à l'erreur systématique de la mesure de température. L'incertitude aléatoire est négligée. Ainsi, l'incertitude des DJU (méthode classique) peut être exprimée par :

$$u_{sys}(DJU) = N_{jour} * u_{sys}(T_{ext}) \quad \textbf{(3.24)}$$

Pour les DJU calculés avec la méthode des professionnels de l'énergie, les résultats sont présentés en annexe A-3-1. De même pour les DH :

$$u_{sys}(DH) = N_{heures} * u_{sys}(T_{ext}) \quad \textbf{(3.25)}$$

Le calcul des incertitudes associées à la mesure et au calcul des différentes grandeurs d'intérêt doit donc respecter les méthodes normalisées. En cas de manque d'information, une approche conservatrice de surestimation de l'incertitude de mesure est à préférer.

3.4.3. Intégration amont au processus

Afin d'intégrer l'effet de ces incertitudes de mesure à la définition du niveau d'engagement et éventuellement du modèle d'ajustement, une étape supplémentaire est ajoutée au sein de la méthodologie précédemment décrite. Après avoir choisi les FA et défini le plan de mesure associé, un post-traitement des résultats de simulation est effectué dans le but de bruiter les valeurs de la grandeur d'intérêt *Y* et des FA.

Chaque composante des n-uplet (X, Y) issus des simulations est modifiée par le tirage aléatoire d'une erreur de mesure qui lui est ajoutée. Ce tirage est réalisé dans la fonction de densité de probabilité associée à l'incertitude de mesure, C'est sur le nouveau jeu de données qu'est ensuite effectuée la détermination du modèle d'ajustement tel que présenté précédemment. L'impact de l'erreur des mesures qui seront réalisées durant la vie de l'ouvrage est donc anticipé et intégré au sein du processus. Ce bruitage des données d'entrainement du modèle d'ajustement de la LGC_α du fait des incertitudes liées aux mesures a pour conséquence une augmentation globale de la valeur limite garantie des consommations pour un même risque α (Figure 3-8 et résultats ci-dessous).

L'évaluation de la plage d'incertitude associée à un FA peut être réalisée en post-traitement pour certaines grandeurs d'intérêt si l'incertitude d'une valeur de FA ne dépend pas de cette valeur (ex : moyenne de températures en négligeant l'erreur aléatoire) ou en dépend de manière linéaire (ex : pourcentage d'erreur pour un compteur d'énergie). Cependant, si l'incertitude varie en fonction des valeurs des FA (ex : degrés-heure), elle doit être estimée à chaque simulation et une fonction de densité de probabilité de l'incertitude liée à la mesure virtuelle est donc associée à chaque simulation. Le schéma de la Figure 3-8 présente ainsi des plages d'incertitude différentes pour chaque point (X, Y).

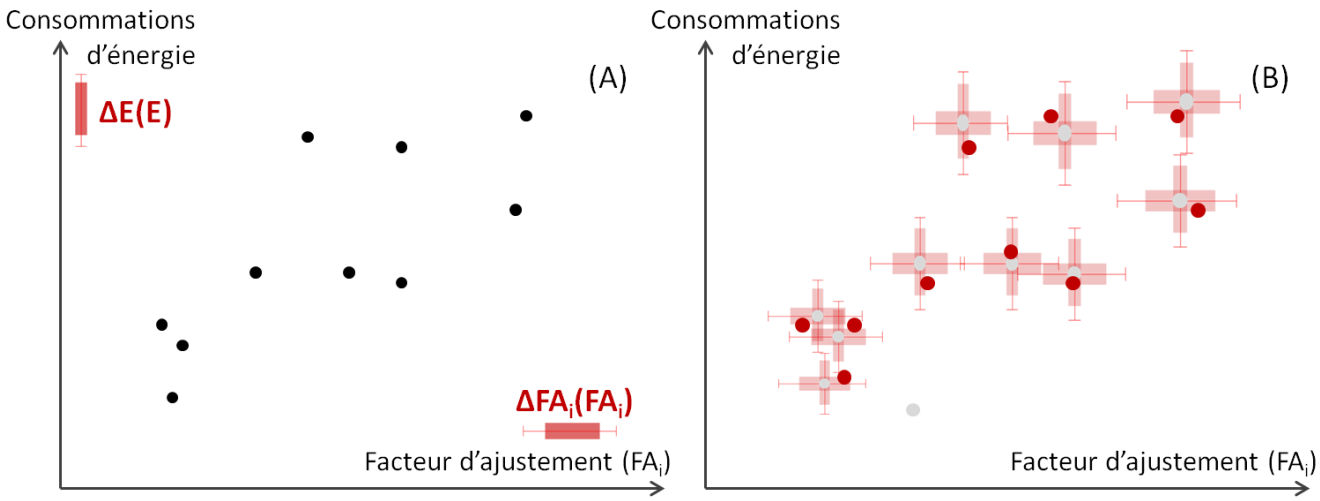

Figure 3-8 : Schéma de l'intégration numérique des incertitudes de mesure. Calcul des incertitudes (A) puis tirage aléatoire des erreurs de mesure modifiant les résultats issus des simulations (B)

Il est aussi possible de multiplier le jeu de données initiales des « valeurs vraies » en réalisant plusieurs tirages d'erreurs de mesure pour chaque jeu (X, Y). Ce processus est en effet peu couteux en temps de calcul et permet d'avoir un plus grand échantillon final et donc une meilleure identification et une meilleure convergence des modèles de régression. Cette question sera traitée au chapitre 4.

Les incertitudes considérées dans l'exemple sont présentées dans le Tableau 3-3. Une incertitude systématique de +/- 0,5 °C est considérée pour la mesure de la température extérieure permettant de calculer les incertitudes portant sur les DJU_{18} et $DH_{18.sc}$.

Tableau 3-3 : Incertitudes de mesure associées aux grandeurs d'intérêt

Grandeurs d'intérêt	***incertitude***
Consommation énergétique de chauffage (kWh)	*+/- 5 %*
T_{ext} *(°C)*	*+/- 0,5*
DJU_{18} *(°C.j)*	*éq. (3.25)*
$DH_{18.sc}$ *(°C.h)*	*éq. (3.26)*
$E_{elec.sc}$ *(kWh)*	*+/- 2 %*

Des tirages aléatoires d'erreur de mesure sont réalisés parmi les lois normales construites à partir de ces incertitudes. Pour chaque jeu de résultat de simulation, cinq tirages correspondant aux mesures virtuelles sont effectuées ce qui multiplie le nombre de données par cinq (5000) pour l'identification de la $LGC_{5\%}$. Les résultats mis à jour et intégrant les incertitudes de mesure sont présentés au paragraphe suivant.

3.5. Protocole complet pour la garantie de performance énergétique

3.5.1. GPE absolue dans le neuf et la rénovation lourde

Une synthèse de la méthodologie présentée dans ce chapitre est schématisée à la Figure 3-9. Elle complète les figures générales 1.11 et 1.12 présentées au chapitre 1 qui décrivent le fonctionnement des engagements de GPE et leur vérification.

En phase d'étude, cette méthodologie repose sur l'exploitation des résultats de simulation issus de la propagation des incertitudes et variabilités (PIV) du modèle physico-probabiliste du bâtiment et sur le choix de facteurs d'ajustement. L'étude statistique des distributions vise à extraire un modèle proposant une limite garantie de consommations énergétiques pour un risque α de dépassement, fonction de facteurs d'ajustement choisis et associés à un plan de mesure. Celui-ci permettra en exploitation de connaitre la valeur effective de ces FA et de vérifier le respect de la LGC_α ajustée.

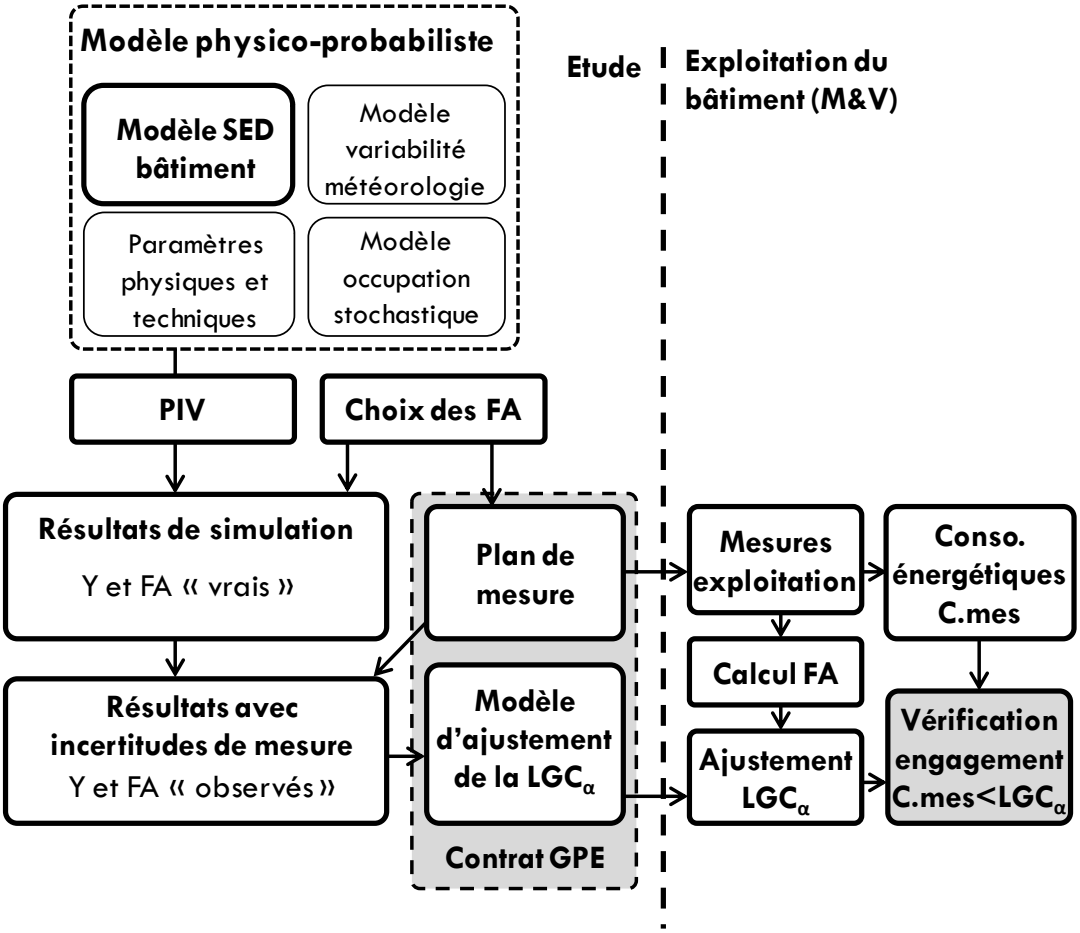

Figure 3-9 : Méthodologie pour un engagement de garantie performance énergétique absolu

En considérant l'ensemble de ces étapes et après l'intégration de l'influence des incertitudes de mesure, les résultats définitifs de l'exemple de GPE absolue sur les besoins de chauffage du bâtiment collectif sont indiqués ci-dessous, et sont rassemblés dans la Figure 3-12. La $LGC_{5\%}$ ajustable pour le cas A est :

$$LGC_{5\%} = 6{,}078.10^4 + 7{,}08.10^4 * \left(\frac{DJU_{18}}{DJU_{18-ref}} - 1\right) \quad \textbf{(3.26)}$$

avec $DJU_{18-ref} = 2292\ \text{°C.h}$, valeur médiane des résultats de la PIV.

Pour le cas B :

$$LGC_{5\%} = 5{,}996.10^4 + 8{,}39.10^4 * \left(\frac{DH_{18.sc}}{DH_{18.sc-ref}} - 1\right) - 1{,}31.10^4 * \left(\frac{E_{elec.sc}}{E_{elec.sc-ref}} - 1\right) \quad \textbf{(3.27)}$$

avec $DH_{18.sc-ref} = 49560\ \text{°C.h}$ et $E_{elec.sc-ref} = 25550\ \text{kWh}$, valeurs médianes des résultats.

Pour une GPE absolue, la vérification réalisée par ajustement de la $\mathrm{LGC}_{5\%}$ fournit des valeurs de 58343 kWh dans la première configuration et de 65356 kWh dans la seconde, soit une valeur supérieure à la consommation mesurée. Même après intégration de l'influence des incertitudes de mesure, l'engagement (à 5 % de risque) n'est toujours pas respecté dans le cas A. En ajustant de manière plus précise et plus complexe, la consommation observée est inférieure à la $LGC_{5\%}$ dans le second cas. On constate que la prise en compte de cette incertitude supplémentaire à fait augmenter les valeurs de référence de la $LGC_{5\%}$ et a légèrement fait varier les valeurs des coefficients d'ajustement. L'influence de ce processus sera plus particulièrement étudiée au chapitre 4.

3.5.2. GPE relative en réhabilitation

Un schéma de synthèse équivalent est présenté à la Figure 3-10 pour le cas de la méthodologie en GPE relative, pour les projets de réhabilitation. Une double PIV est réalisée sur les modèles physico-probabilistes représentant le bâtiment avant et après l'intervention. Suite au choix des FA (contraint par la connaissance disponible avant les travaux), les résultats de simulation pre-réhabilitation fournissent un modèle d'ajustement de la consommation énergétique de référence. Ce modèle est utilisé comme base pour calculer les gains relatifs observés dans les résultats de simulations du modèle post-réhabilitation intégrant les incertitudes de mesure issues des choix sur la prescription du plan de mesure et vérification (PMV). L'étude statistique de cette distribution des écarts relatifs permet de calculer un gain relatif $GR_{\alpha,glob}$ moyenné sur la plage des FA, associé à un risque de dépassement α.

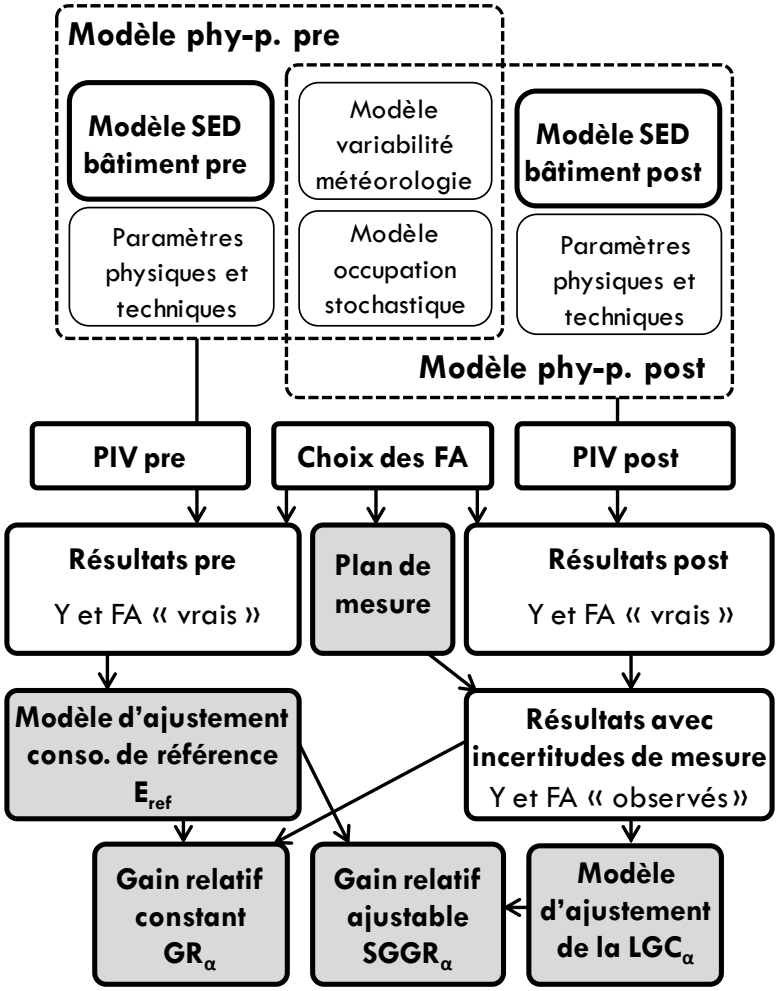

Figure 3-10 : Méthodologie pour un engagement de garantie performance énergétique relatif sur la phase d'étude (suite à la Figure 3-11)

Il est aussi possible de calculer la variabilité de ce gain en fonction des valeurs des FA (GR_α). Cette variabilité, bien que limitée, impacte le niveau de risque en fonction des valeurs effectives des FA. Les éléments grisés sont ceux qui sont directement liés au contrat, le gain relatif contractualisé pouvant être choisi de manière global ou être calculé à partir du modèle de la LGC_α.

La Figure 3-11 illustre le processus de vérification suite à l'établissement des éléments contractualisés issus de la méthodologie. Le gain relatif réel est calculé à partir de la consommation énergétique réelle et de la consommation de référence avant travaux, ajustée en fonction des FA effectifs. Cette valeur est ensuite comparée à la valeur de gain garantie, soit le gain global $GR_{\alpha-glob}$, soit le gain ajusté GR_α (droite).

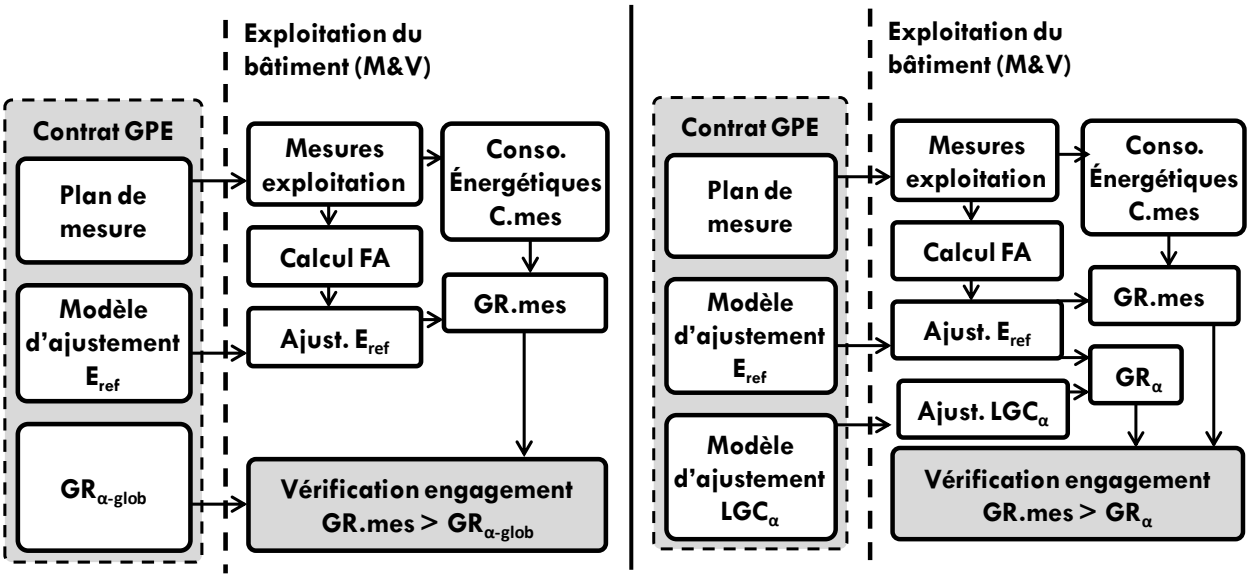

Figure 3-11 : Vérification de l'engagement en GPE relative. Engagement sur un gain relatif global (gauche) ou ajustable (droite)

Après intégration de l'influence des incertitudes de mesure, les résultats définitifs de l'exemple de GPE relative sur les besoins de chauffage du bâtiment collectif sont présentés à la Figure 3-12. Le modèle de la consommation pre réhabilitation ajustable n'est pas modifié, l'équation 3.16 (ou 3.17) est contractuelle et la consommation ajustée en fonction des valeurs de FA mesurées est celle calculée au § 3.3.6.2. L'exploitation des résultats de simulation post travaux intégrant les incertitudes de mesure permet d'obtenir les modèles de $LGC_{5\,\%}$ présentés au chapitre précédent.

Le gain relatif contractuel est ainsi soit exprimé comme le rapport des modèles de $LGC_{\ 5\%}$ et de E_{pre} (GR_{α} ajustable), soit calculé comme le quantile à 5 % de la distribution des écarts entre consommations énergétiques simulées (intégrant l'incertitude de mesure) post-travaux et les valeurs de consommation d'énergie de référence correspondantes, calculées avec les FA intégrant les incertitudes de mesure ($GR_{\alpha,glob}$). Les modèles d'ajustement de la $LGC_{5\%}$ après travaux (équations 3.26 et 3.27) fournissent un seuil de gain relatif garanti ajusté ($GR_{5\%}$) de 52,6 % et de 52,7 % dans les cas A et B. Le gain relatif global garanti $GR_{\alpha,glob}$ est de 52,5 % dans le cas A et de 52,9 % dans le cas B. Ces résultats sont présentés sur la Figure 3-12 qui constitue une mise à jour de la Figure 3-7, où les incertitudes de mesure n'étaient pas prises en compte.

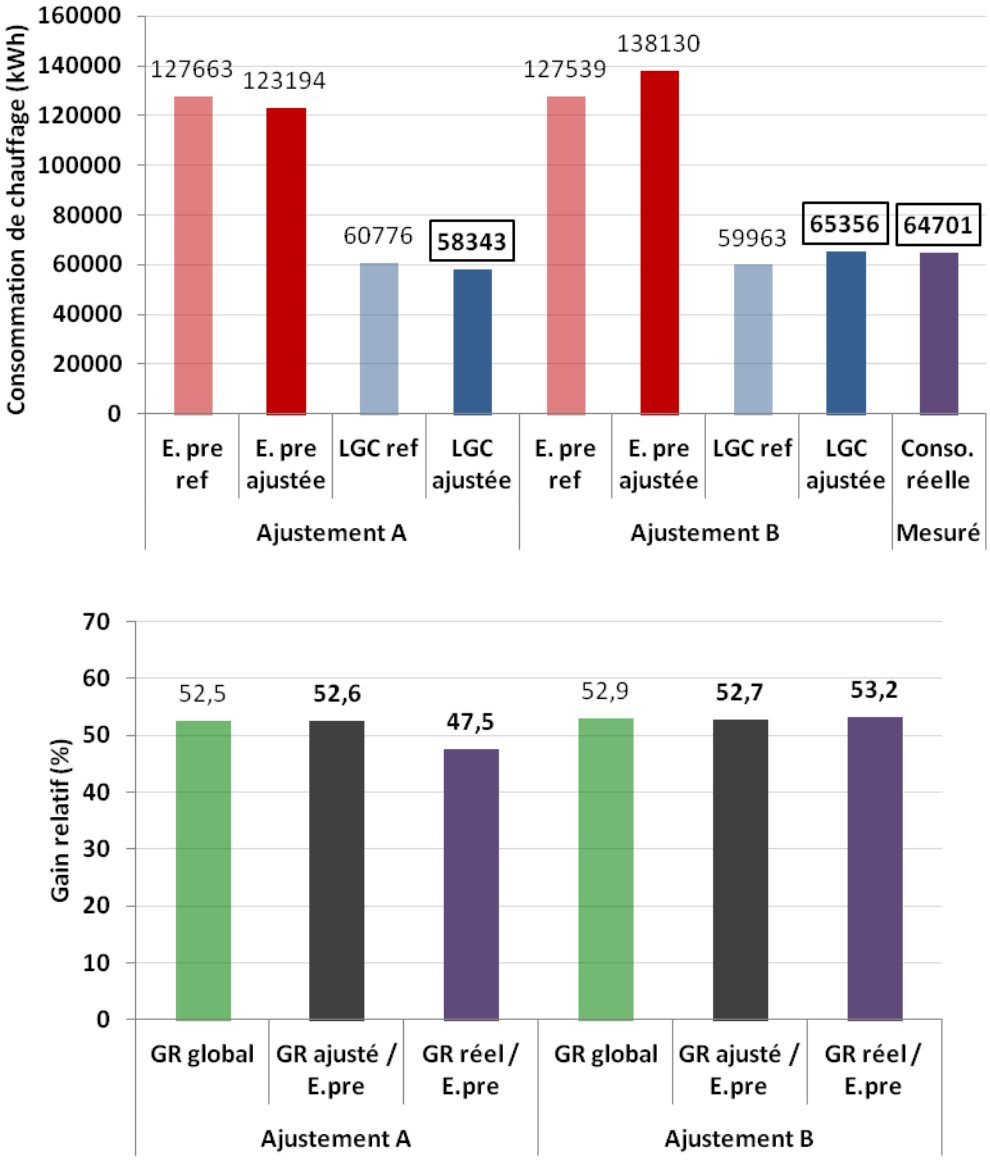

Figure 3-12 : Comparaison entre les performances garanties et mesurées pour les cas d'ajustement A et B pour un engagement absolu (haut) ou relatif (bas)

Le gain observé était de 47,5 % dans la première configuration et de 53,2 % dans la seconde. La garantie n'est donc toujours pas respectée dans le cas A. On constate que l'ajustement en fonction des très faibles consommations électriques fait diminuer le gain relatif garanti dans les cas A et B par rapport au seuil global. Cependant, des facteurs exceptionnellement défavorables (tels que la mesure de températures de consignes particulièrement élevées après réhabilitation) placent les consommations réelles au-delà des engagements.

Même si les modèles d'ajustement identifiés reflètent un même risque de dépassement au regard des résultats de simulation exploités, la prise en compte d'un maximum de FA permet de réduire le risque de conditions d'exploitation extrêmes dont la réalisation statistique dans les

modèles est improbable. Dans cet exemple, la conjonction de consommations électriques très faibles et de température de consignes élevées entraine une consommation énergétique réelle importante dépassant la limite garantie lorsque seul un ajustement en fonction des DJU_{18} est effectué. La comparaison de configurations d'ajustement sera plus systématiquement étudiée au chapitre 4.

3.6. Conclusion

Ce chapitre décrit le cœur de ces travaux de thèse : une méthodologie intégrée pour la mise en place de contrats de GPE. Celle-ci se déploie sur l'ensemble des phases de ce type de projet, de l'étude, en passant par la détermination d'un plan de mesure, à la vérification de la performance, et tient compte des interactions entre ces phases.

Elle repose sur l'utilisation d'outils de modélisation et sur l'exploitation statistique de multiples résultats de simulations intégrant l'incertitude des caractéristiques constructives du bâtiment et des systèmes énergétiques, et décrivant la variabilité possible des sollicitations. La pertinence de la méthodologie repose sur le soin apporté à la description du modèle physico-probabiliste qui regroupe l'ensemble de ces données. Une approche conservatrice, surestimant incertitudes et variabilités, est à préférer si certains éléments sont méconnus. Le raffinement de la description des incertitudes et l'utilisation de modèles de variabilité de sollicitations plus précis et plus spécifiques au cas étudié ont vocation à renforcer la méthodologie présentée en minimisant la prise de risque et en réduisant le niveau de performance garanti.

Un processus d'ajustement permet de s'affranchir d'une partie de l'influence des sollicitations extérieures, non liées à la performance énergétique intrinsèque du bâtiment, en particulier l'effet des conditions météorologiques et des comportements des usagers. Les facteurs d'ajustement choisis seront mesurés en exploitation, permettant de replacer les consommations énergétiques observées dans leur contexte de sollicitations. L'approche de modélisation probabiliste développée dans cette méthodologie a pour objectif de caractériser le lien entre consommations énergétiques d'intérêt et facteurs d'ajustement mais aussi d'étudier la distribution statistique de ces consommations afin de proposer un niveau d'engagement contractuel optimal.

L'incertitude due à la mesure des performances réelles et des facteurs d'ajustement lors de la vérification entraine un décalage entre le comportement énergétique réel, simulé par les modèles, et la connaissance de ce comportement, entachée d'erreurs. L'influence de ces erreurs est prise en compte en amont de la méthodologie via une intégration numérique réalisée par un bruitage des données de simulation représentant leur mesure virtuelle.

L'étude statistique des données obtenues permet de déterminer un niveau d'engagement avec un risque quantifié de défaut. Celui-ci est exprimé comme une limite garantie de consommations absolue ajustable dans le cas d'une GPE pour des bâtiments neufs (ou rénovation lourde), et comme

un gain relatif de performance par rapport à une consommation de référence ajustable pour des projets de GPE relative en réhabilitation.

Les principaux atouts de la méthodologie sont les niveaux ambitieux de consommations d'énergie garantie rendus possible par la précision des méthodes et par la quantification du risque associé, mais aussi la transparence du processus et de la vérification périodique annuelle de la performance garantie pour le client.

Un des partis pris originaux de la méthode proposée réside dans le phasage de l'étude. L'ensemble des simulations et des traitements de données est réalisé avant la période d'exploitation du bâtiment. Ainsi, les éléments fournis par la méthodologie dans le contrat de GPE sont déterminés et actés en amont, avant la livraison du bâtiment. Les mesures réalisées durant l'exploitation de l'ouvrage permettent simplement la mise en œuvre du processus d'ajustement suivant le protocole préétabli. La vérification de l'atteinte des objectifs constituant la résolution du contrat est ainsi simplifiée et ne nécessite pas l'intervention de l'entreprise portant la garantie ni par la réalisation de nouvelles simulations ni par un post-traitement complexe et opaque des données mesurées. Ceci permet donc une transparence totale de la vérification de la performance énergétique.

La prochaine partie présente un ensemble d'études liées à la méthodologie développée dans ce chapitre et s'intéresse notamment à sa robustesse et à la sensibilité des niveaux d'engagement obtenus à différentes critères. L'influence des facteurs d'ajustement choisis et des niveaux d'incertitudes de mesure sont ainsi étudiés tout comme la possibilité d'optimiser le processus.

Chapitre 4. Méthodes d'aide à la mise en œuvre de la méthodologie de GPE

Ce chapitre propose un ensemble d'études liées à la méthodologie présentée précédemment avec pour objectif de répondre à différentes questions pouvant se poser lors de sa mise en œuvre. Ces questions concernent la validité des résultats, le paramétrage des différents éléments de la méthode et les limites de validité de ses hypothèses. La robustesse statistique des résultats fournis par la régression quantile est tout d'abord étudiée et testée afin de valider le niveau de risque associé au modèle. Par la suite, le choix des moyens de mesure mis en œuvre pour la vérification des performances énergétiques et la caractérisation des conditions d'exploitation est investigué dans le but de caractériser l'influence des incertitudes de mesure sur la valeur garantie de la performance. Le choix des facteurs d'ajustement est une des problématiques principales de l'utilisation pratique de la méthodologie et différentes méthodes sont ainsi évaluées pour assister le choix et pour quantifier la qualité du modèle d'ajustement. Enfin, des modèles plus complexes, constituant des alternatives à la régression quantile linéaire, sont étudiés afin d'expliquer au mieux les variations de la grandeur d'intérêt et de son quantile conditionnel en fonction des facteurs d'ajustement.

Sommaire du chapitre 4

Chapitre 4. Méthodes d'aide à la mise en œuvre de la méthodologie de GPE 149

4.1. Étude de la robustesse du modèle d'ajustement 151
4.1.1. Variance de l'estimateur de régression quantile 151
4.1.2. Variance statistique du modèle 153
4.1.3. Bootstrap et modification de l'engagement 155
4.1.4. Validation numérique du risque de dépassement 157

4.2. Influence des incertitudes de mesure 158
4.2.1. Caractérisation et intégration 158
4.2.2. Modification des coefficients d'ajustement des modèles 159
4.2.3. Modification de la limite garantie de consommation 161
4.2.4. Application à la comparaison de PMV 162

4.3. Choix et optimisation des paramètres d'ajustement 165
4.3.1. Objectifs 165
4.3.2. Tests de corrélation 166
4.3.3. Qualité du modèle de régression 170
4.3.4. Méthode d'AS par régression 174
4.3.5. Tests statistiques 177
4.3.6. Régression pénalisée 178
4.3.7. Bilan pour l'aide au choix des facteurs d'ajustement 183

4.4. Méthodes alternatives pour les modèles d'ajustement 184
4.4.1. Ajustement de la SED 184
4.4.2. Régression quantile glissante 187
4.4.3. Modèle polynomial 188
4.4.4. Régression multivariée par spline adaptative 191

4.5. Synthèse 195

4.1. Étude de la robustesse du modèle d'ajustement

Dans cette partie, la stabilité de la méthode de régression quantile qui fournit le modèle contractualisé d'une limite garantie de consommations énergétique ajustable, est étudié. L'objectif est de quantifier l'incertitude sur les paramètres du modèle linéaire, et en particulier sur le premier coefficient, associé à la valeur de LGC_α dans les conditions de référence (équivalent à l'ordonnée à l'origine du modèle linéaire).

4.1.1. Variance de l'estimateur de régression quantile

L'estimateur $\hat{\beta}$ de la solution β de la régression quantile, est évalué à partir d'un échantillon de données. Dans une approche de caractérisation du risque, on s'intéresse à la variance de cet estimateur, dite aussi variance asymptotique. L'objectif est ainsi de calculer des intervalles de confiance autour des résultats de régression. Ceux-ci dépendent de la densité des données autour du quantile d'intérêt. Les propriétés de la variance de l'estimateur sont difficiles à établir car il n'existe pas de forme explicite de $\hat{\beta}$, contrairement à l'estimateur des moindres carrés (Givord et D'haultfoeuille, 2013). Des résultats analytiques et des théorèmes ont été démontrés par Koenker et Basset sur les propriétés de l'estimateur, et l'ensemble de ces développements sont décrits dans l'ouvrage de Koenker (2005). Kocherginsky et al. (2005) proposent une revue des méthodes disponibles pour le calcul d'intervalles de confiance sur l'estimateur de la régression quantile.

4.1.1.1. Méthodes directes

Le calcul de la variance nécessite l'évaluation de la densité des termes d'erreur (écarts au modèle de régression). Sous hypothèse spécifique sur la répartition des erreurs, Koenker (1994) propose des méthodes d'estimation de cette densité en fonction d'un paramètre de lissage (*bandwidth*).

Les principaux travaux sur l'estimation de la valeur de ce paramètre ont été effectués par Bofinger (1975), Hall et Sheather (1988) et Chamberlain (1994). Les modèles qui en découlent sont couramment disponibles dans les modules ou *packages* de régression quantile des outils de statistique Python et R. Les résultats de Hall et Sheather sont admis comme étant les plus pertinents et seront donc utilisés ci-dessous (*HS*).

Trois méthodes d'estimation directes sont identifiées mais il est à noter qu'elles reposent sur un ensemble d'hypothèses dont la validité est rarement vérifiable *a priori* :

- Les erreurs sont considérées indépendantes et identiquement distribuées (méthode *iid*).
- La fonction du τ-quantile conditionnel est supposée linéaire en τ (méthode *nid*).
- Une estimation par noyau de la densité des erreurs est considérée (méthode *ker*).

Les détails de ces éléments sont présentés par Koenker (2005). Ces méthodes ont donc le désavantage de reposer sur des hypothèses non vérifiables et sur le choix d'un paramètre de lissage.

Afin d'éviter une partie de ces limitations, et notamment l'estimation du paramètre de lissage, Koenker (1994) propose une autre approche pour la construction d'intervalles de confiance basée sur les tests de rang (*rank score*). Cette méthode n'est théoriquement valide que dans le cas d'une homoscédasticité des données (variance conditionnelle de Y non dépendante des variables explicatives) et peut s'appuyer sur l'hypothèse d'erreurs identiquement distribuées ou non (*iid/nid*). Elle nécessite des coûts de calcul prohibitifs en grande dimension.

4.1.1.2. Bootstrap

Les techniques de *bootstrap*, popularisées par Efron (1979), reposent sur la répétition des méthodes statistiques (approche Monte-Carlo) avec une variabilité de leurs entrées provenant de ré-échantillonnage au sein des données initiales. Ces échantillons factices sont obtenus par tirages avec remise au sein de l'échantillon disponible. La dispersion des résultats du *bootstrap* permet de caractériser l'incertitude sur les résultats obtenus avec l'échantillon global. Givord et D'haultfoeuille (2013) mettent en avant cette technique pour estimer l'intervalle de confiance sur les coefficients des modèles linéaires obtenus par régression quantile.

Deux approches sont proposées. Premièrement, la variance asymptotique du paramètre d'intérêt parmi les résultats de *bootstrap* est calculée. Les quantiles de la distribution pour la construction de l'intervalle de confiance peuvent être estimés à partir de cette variance en supposant une distribution normale. Il est aussi possible de ne faire aucune hypothèse sur la loi de répartition du paramètre et de s'intéresser directement au quantile de la fonction de répartition empirique.

Kocherginsky et al. (2005) s'intéressent aux méthodes d'échantillonnage les plus efficientes favorisant l'estimation des propriétés de l'estimateur pour un coût de calcul réduit. Les méthodes reposant sur le ré-échantillonnage des erreurs sont distinguées des techniques de ré-échantillonnages des données complètes. Dans le premier cas, les erreurs se doivent d'être indépendantes et identiquement distribuées. La méthode de base du ré-échantillonnage aléatoire des données est dite *bootstrap* par paires (Efron et Tibshirani, 1993) car des paires de variables expliquées et explicatives sont considérées. Des variations de cette approche sont proposées par Parzen et al. (1994) pour la question spécifique de la régression quantile. Kocherginsky et al. (2005) développent eux une technique d'échantillonnage dite MCMB (*Markov Chain Marginal Bootstrap*) permettant de limiter le coût de calcul pour l'estimation des propriétés asymptotiques des grandeurs statistiques d'intérêt.

Suivant les méthodes et les travaux de la littérature considérés, un nombre de ré-échantillonnages très variable est conseillé, allant de quelques centaines à plusieurs milliers. Dans la

plupart des packages statistiques existants proposant des implémentations de ces différentes méthodes, les ré-échantillonnages sont réalisés jusqu'à convergence de la variance de l'estimateur selon un critère prédéfini.

4.1.2. Variance statistique du modèle

Les résultats associés aux différentes méthodes d'estimation de la variance des coefficients du modèle de régression quantile présentés ci-dessus sont comparés sur les deux configurations du cas d'étude. Le Tableau 3-2 présente l'écart-type des coefficients des modèles pour les configurations A et B. Le même jeu de données provenant de 1000 simulations utilisé dans le chapitre 3 est exploité pour calculer ces valeurs. On s'intéresse d'abord au modèle identifié avant intégration des incertitudes de mesure, avec les résultats de simulation initiaux. En effet, même si la régression quantile finale se fait sur un jeu de données intégrant des erreurs de mesure virtuelles de taille plus conséquente, ces données sont dérivées du jeu initial, qui se doit d'être statistiquement assez riche. Le modèle de limite garantie de consommations à 5 % de risque est rappelé ci-dessous. Pour la configuration A :

$$LGC_{5\%} = 6{,}01.10^4 + 7{,}29.10^4 * \left(\frac{DJU_{18}}{DJU_{18-ref}} - 1\right) \quad \textbf{(4.1)}$$

avec $DJU_{ref} = 2292\ \text{°C.j}$, valeur médiane des résultats de la PIV.

Pour le cas B :

$$LGC_{5\%} = 5{,}59.10^4 + 8{,}57.10^4 * \left(\frac{DH_{18.sc}}{DH_{18.sc-ref}} - 1\right) - 1{,}14.10^4 * \left(\frac{E_{elec}}{E_{elec.sc-ref}} - 1\right) \quad \textbf{(4.2)}$$

avec $DH_{ref} = 49560\ \text{°C.h}$ et $E_{elec-ref} = 25550\ \text{kWh}$, valeurs médianes des résultats.

Les valeurs relatives des écart-types sont affichées entre parenthèses à coté des valeurs absolues. On constate une variabilité des résultats obtenus entre les différentes méthodes d'estimation de la variance, les méthodes *iid-HS* et *nid-HS* fournissent des valeurs plus élevées. La variance relative est beaucoup plus grande pour les coefficients de pentes du modèle que pour le coefficient β_0 ($LGC_{5\%-ref}$), lié à l'ordonnée à l'origine, en particulier pour le coefficient associé à la consommation électrique durant la période de chauffe (E_{elec}) dont l'écart-type est de l'ordre de 50 % de la valeur du coefficient. Comme on le verra dans la suite de ce chapitre, la trop faible taille du jeu de données n'est pas la seule explication de cette forte variance statistique. L'influence limitée de ce FA explique aussi cette constatation.

Tableau 4-1 : Écart-types des coefficients du modèle de régression quantile (sans incertitudes de mesure) obtenus avec différentes méthodes de la littérature

Méthodes d'estimation de la variance	***Écart-type***				
	Configuration A		***Configuration B***		
	$\boldsymbol{LGC_{ref}\ (kWh)}$	$\boldsymbol{\beta\ DJU_{18}}$	$\boldsymbol{LGC_{ref}\ (kWh)}$	$\boldsymbol{\beta\ DH_{18.sc}}$	$\boldsymbol{\beta\ E_{elec.sc}}$
COEFFICIENTS	***6,01.10⁴***	***7,29.10⁴***	***5,59.10⁴***	***8,57.10⁴***	***-1,14.10⁴***
Directe (iid-HS)	*341 (0,6 %)*	*2774 (3,8 %)*	*402 (0,7 %)*	*3733 (4,4 %)*	*6662 (58,6 %)*
Directe (nid-HS)	*437 (0,7 %)*	*3334 (4,6 %)*	*438 (0,7 %)*	*3810 (4,4 %)*	*6645 (58,5 %)*
Directe (ker-HS)	*369 (0,6 %)*	*3136 (4,3 %)*	*363 (0,6 %)*	*3226 (3,8 %)*	*6144 (54,1 %)*
Rank score (iid)	*338 (0,6 %)*	*3015 (4,1%)*	*290 (0,5 %)*	*3433 (4,0 %)*	*4411 (38,8 %)*
Rank score (nid)	*340 (0,6 %)*	*3377 (4,6 %)*	*298 (0,5 %)*	*3491 (4,1 %)*	*4418 (38,8 %)*
Bootstrap (xy)	*307 (0,5 %)*	*2437 (3,3 %)*	*361 (0,6 %)*	*2675 (3,1 %)*	*6327 (55,7 %)*
Bootstrap (pwy)	*273 (0,5 %)*	*2662 (3,7 %)*	*336 (0,6 %)*	*2909 (3,4 %)*	*5617 (49,4 %)*
Bootstrap (MCMB)	*351 (0,6 %)*	*2372 (3,3 %)*	*375 (0,6 %)*	*2770 (3,2 %)*	*5917 (52,1 %)*

Concernant le coefficient constant β_0, la Figure 3-6 présente les valeurs des incertitudes sur la $LGC_{5\%-ref}$ pour un intervalle de confiance de 90 % (donc pour un risque de dépassement à droite de 5 %). L'incertitude statistique associée à la $LGC_{5\%-ref}$ du modèle construit avec les données intégrant les erreurs de mesure virtuelles est aussi présentée. Du fait du plus grand nombre de données d'entrainement, celle-ci est inférieure, et de l'ordre de 0,5 % pour toutes les méthodes.

En considérant les travaux de la littérature et les limites de validité des différentes méthodes et de leurs hypothèses, nous retenons les méthodes *bootstrap* et en particulier celle basée sur un tirage MCMB permettant une accélération de la convergence (Kocherginsky et al., 2005). Dans une approche où le risque de surconsommation est considéré, on s'intéresse en particulier à l'incertitude sur le coefficient correspondant à la limite garantie de consommations dans les conditions de référence. Les incertitudes sur les autres coefficients liés aux pentes du polynôme influencent aussi le risque de surconsommation mais de manière centrée. En effet, une sous-estimation de la valeur d'un coefficient positif entrainera le calcul d'une LGC_α ajustée trop faible pour les faibles valeurs du FA associé et trop élevée pour les valeurs élevées, alors qu'une surestimation aura l'effet inverse. Il est donc pertinent de conserver l'estimateur centré dans ce cas.

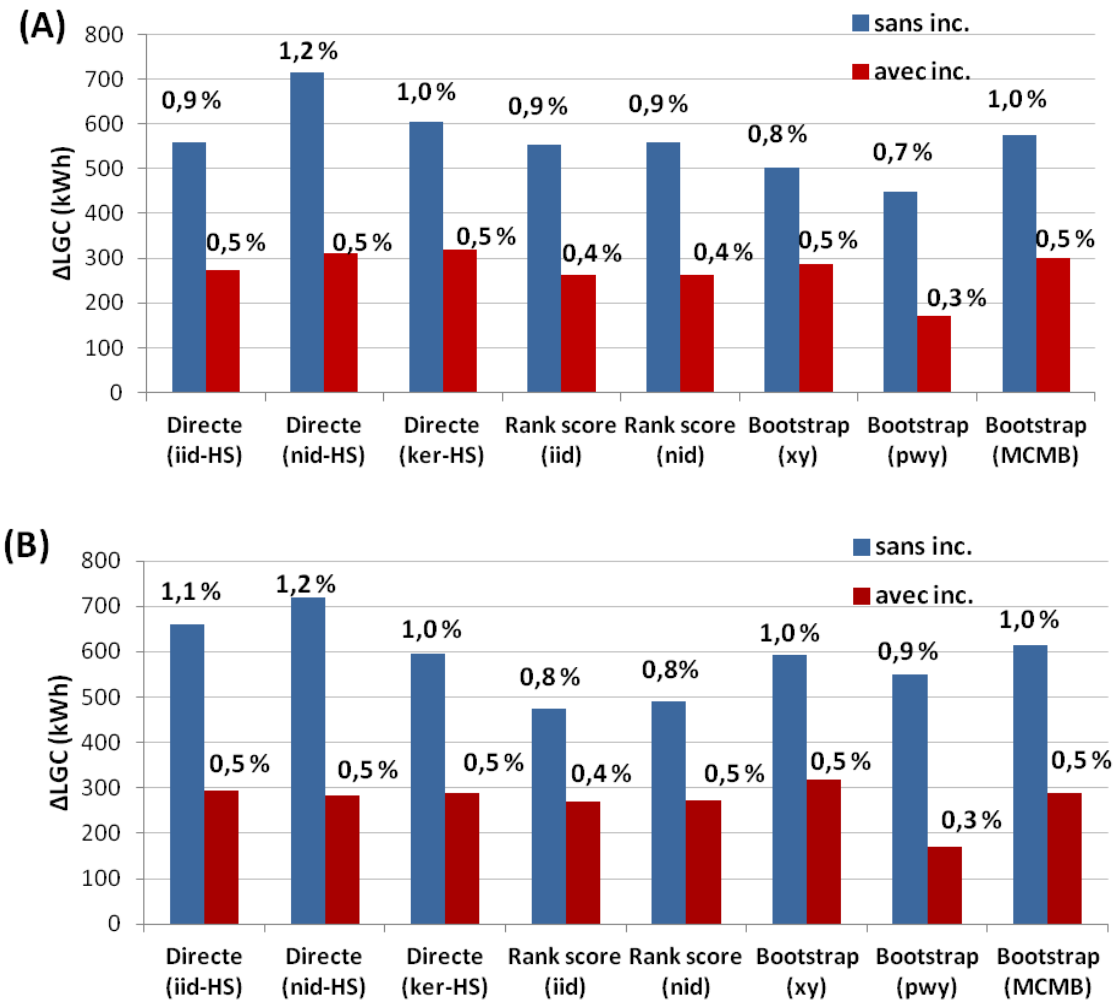

Figure 4-1 : incertitude à 90 % (centrée) sur la $LGC_{5\%.ref}$ avec les différentes méthodes pour les cas A (haut) et B (bas) et pour les régressions avec (5000 s.) et sans (1000 s.) incertitudes

4.1.3. Bootstrap et modification de l'engagement

La Figure 4-2 s'intéresse à l'évolution de la variance estimée de la $LGC_{5\%.ref}$ en fonction du nombre de données utilisées pour construire le modèle. La méthode *Bootstrap*-MCMB est utilisée pour calculer l'écart-type sur l'estimation de cette valeur. Des fractions de 5000 jeux de données sont exploitées. L'influence du nombre Nr de ré-échantillonnages *bootstrap* est aussi étudiée.

On constate une diminution de l'incertitude en fonction du nombre de données de construction du modèle de régression quantile. Le seuil d'incertitude de 1 % est atteint pour un peu moins de 2000 simulations. On constate l'instabilité du calcul de l'incertitude pour des nombres trop faibles de répétition de la méthode *bootstrap*. 2000 à 5000 ré-échantillonnages *bootstrap* semblent appropriés

pour obtenir des résultats stables, même si un nombre supérieur peut être envisagé du fait que cette méthode est peu coûteuse en temps de calcul.

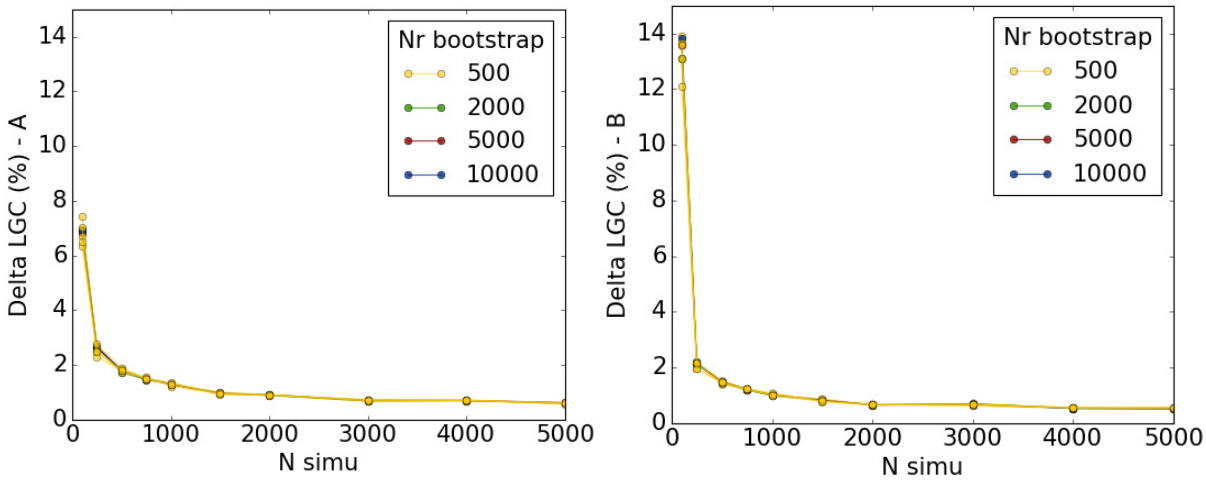

Figure 4-2 : Évolution de l'incertitude sur la LGC calculée par bootstrap-MCMB en fonction du nombre de données utilisées pour la régression quantile et du nombre de répétitions du bootstrap pour les configurations A (gauche) et B (droite)

Afin d'intégrer les apports de ces résultats dans la méthodologie présentée au chapitre 3, on propose de calculer systématiquement la variance des coefficients du modèle et d'ajouter à la valeur de LGC_{α} identifiée, l'incertitude correspondant à un intervalle de confiance de 90 % centré, ce qui correspond à un risque de dépassement à droite de 5 %. Ce choix est légèrement conservateur car le risque associé à une valeur à 95 % de confiance sur un quantile à 95 % est finalement inférieur à 5 %. Par ailleurs, un nombre minimal de 1000 à 2000 jeux de données et donc la réalisation d'autant de simulations initiales est conseillé, même si les résultats associés peuvent être multipliés dans le processus d'intégration des incertitudes de mesure, ou plusieurs mesures virtuelles peuvent être associées à chaque simulation (cf paragraphes 3.4.3 et 4.2). Les modèles (sans incertitudes de mesure, équations 3.13 et 3.14) intégrant l'incertitude statistique sont décrits ci-dessous.

Pour le cas A :

$$LGC_{5\%} = (6{,}012 + 0{,}061).10^4 + 7{,}29.10^4 * \left(\frac{DJU_{18}}{DJU_{18-ref}} - 1\right) \quad \textbf{(4.3)}$$

avec $DJU_{ref} = 2292$ °C.j , valeur médiane des résultats de la PIV.

Pour le cas B :

$$LGC_{5\%} = (5{,}955 + 0{,}060).10^4 + 8{,}57.10^4 * \left(\frac{DH_{18.sc}}{DH_{18.sc-ref}} - 1\right) - 1{,}14.10^4 * \left(\frac{E_{elec}}{E_{elec-ref}} - 1\right) \quad \textbf{(4.4)}$$

avec $DH_{18.sc-ref} = 49560$ °C.h et $E_{elec-ref} = 25550$ kWh, valeurs médianes des résultats.

4.1.4. Validation numérique du risque de dépassement

Afin de valider ces modifications méthodologiques, une expérience numérique est réalisée dans le but d'évaluer le pourcentage de résultats de simulation satisfaisant à l'engagement de GPE calculé avec un jeu de données initiales. Les données de validation sont indépendantes des données de construction du modèle. Les résultats de FA et de consommations de 5000 simulations sont étudiés. Des tirages d'erreurs de mesure virtuelles associés à chaque grandeur d'intérêt sont réalisés. La Figure 4-3 présente la répartition des écarts entre les consommations observées et la $LGC_{5\%}$ ajustée (équations 4.3 et 4.4) avec les valeurs de FA correspondantes. On constate que 4,98 % et 4,66 % des cas correspondent à des défauts de l'engagement pour les cas A et B. Ce taux de défaut aurait été respectivement de 5,90 et de 5,64 sans ajout de l'incertitude statistique sur $LGC_{5\%-ref}$.

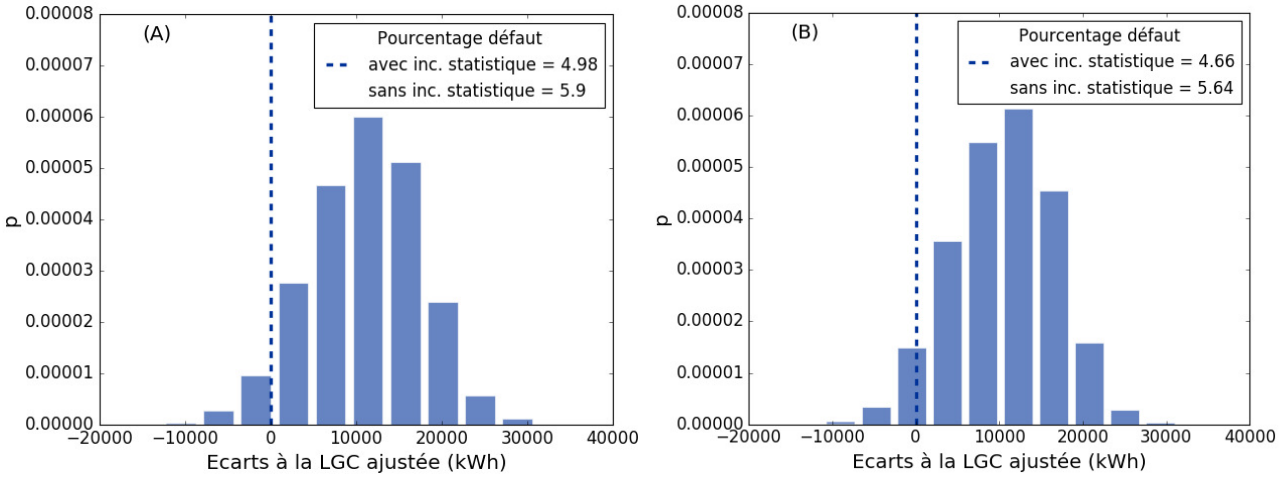

Figure 4-3 : Répartition des écarts entre les valeurs de consommations simulées et les LGC ajustées correspondantes pour les configurations A (gauche) et B (droite)

Une validation expérimentale, bien que souhaitable ne peut être réalisée qu'avec un nombre de données expérimentales statistiquement représentatif, de plusieurs centaines de réalisations. Ceci peut éventuellement être envisageable dans le cas où un grand nombre de bâtiments identiques seraient réalisés, pour des ensembles pavillonnaires par exemple.

Cette première partie a permis de caractériser l'incertitude statistique pesant sur les résultats de la régression quantile. Des adaptations de la méthode sont proposées, un processus *bootstrap* permet de calculer l'incertitude statistique sur la $LGC_{\alpha-ref}$, qui lui est ajoutée pour tenir compte de ce risque supplémentaire. Un nombre minimal de 1000 à 2000 simulations énergétiques est conseillé, et d'au moins 5000 ré-échantillonnages bootstrap pour le calcul d'incertitude statistique. Les incertitudes sur les coefficients d'ajustement peuvent être calculées. Bien qu'étant importantes, elles ne modifient pas les résultats du modèle d'ajustement.

4.2. Influence des incertitudes de mesure

4.2.1. Caractérisation et intégration

La prise en compte de l'impact des incertitudes associées à la mesure et vérification (M&V) des performances réelles en exploitation a été détaillée au paragraphe 3.4 du chapitre 3. Durant la phase d'étude amont, elle passe par différentes étapes :

- la spécification d'un plan de mesure ;
- l'évaluation des incertitudes qui seront associées à la mesure des facteurs d'ajustement choisis et de la grandeur d'intérêt lors de la vérification ;
- l'intégration numérique d'erreurs de mesure virtuelles au sein des résultats de simulation afin d'anticiper l'impact des incertitudes de mesure lors de la vérification.

Comme abordé au paragraphe 3.4, l'évaluation d'une plage d'incertitude associée à la mesure virtuelle des résultats de simulation peut prendre différentes formes en fonction des FA et grandeurs d'intérêt considérées. Ces incertitudes peuvent ou non être indépendantes de la simulation, être exprimées de manière absolue ou relative. Ces éléments dépendent de la grandeur mesurée, du facteur d'ajustement déterminé et des moyens de mesure, et vont impacter de façon différente le bruitage des données et donc les résultats du modèle de régression identifié. Par exemple, la prise en compte d'une incertitude relative sur la grandeur d'intérêt Y peut provoquer une hétéroscédasticité des données. En effet, les erreurs de mesure virtuelles tirées aléatoirement dans cette incertitude relative sont plus petites pour de faibles valeurs de Y que pour des valeurs élevées. La variance conditionnelle de Y en fonction d'un FA explicatif FA_i est donc modifiée, la dispersion de $Y|FA_i$ étant plus importante pour des valeurs de FA_i associées à des Y élevés.

Dans cette partie, on souhaite étudier la sensibilité des modèles pour l'engagement de GPE fournis par la méthodologie au processus d'intégration des incertitudes de mesure. On s'intéresse à la dépendance des coefficients d'ajustement au niveau d'incertitude sur la mesure des facteurs d'ajustement, et à l'impact de ces incertitudes et de celle sur la mesure de la consommation énergétique d'intérêt sur la limite garantie de consommation LGC_α. Cette étude a pour but d'apporter des réponses opérationnelles aux questionnements sur la qualité des plans de mesure à mettre en œuvre dans des processus de GPE.

Par ailleurs, une problématique méthodologique est abordée, liée à la stabilité statistique des résultats suite à cette procédure d'intégration numérique d'erreurs de mesure dans les résultats de simulation. Comme signalé au paragraphe 3.4, il est possible d'associer plusieurs erreurs de mesure au même résultat unitaire de simulation afin d'augmenter la taille de l'échantillon statistique utilisé. Pour chaque jeu (Y, FA_i) issu d'une simulation, un nombre p de jeux d'erreurs $(\Delta Y, \Delta FA_i)$ peuvent être définis (par tirage aléatoire dans les lois de distribution liées aux incertitudes de mesure),

fournissant p jeux $(Y + \Delta Y, FA_i + \Delta FA_i)$. L'échantillon initial de N simulations est remplacé par un échantillon de $N.p$ jeux de données. Cette multiplication des données est envisageable du fait du faible coût de calcul associé aux tirages des erreurs en comparaison du temps de simulation énergétique du bâtiment. La pertinence de cette procédure et son paramétrage seront donc aussi étudiés dans cette partie.

4.2.2. Modification des coefficients d'ajustement des modèles

Les résultats présentés s'appuient sur le cas d'étude utilisé précédemment. Les configurations d'ajustement A (DJU_{18}) et B ($DH_{18.sc}$/ $E_{elec.sc}$) sont étudiées. Pour chaque variable des modèles de régression, des plages d'incertitude de mesure relatives sont considérées selon des lois normales tronquées $N\left(0, X.\frac{IR}{2}, X.(1-IR), X.(1+IR)\right)$, où X est la valeur du facteur d'ajustement ou de la grandeur d'intérêt issue de la simulation et IR l'incertitude relative sur cette valeur. Des modèles de régression quantile sont construits à partir des résultats de simulation intégrant des erreurs de mesure virtuelles tirées aléatoirement dans ces lois de distribution.

La Figure 4-4 présente l'évolution du coefficient d'ajustement β sur les DJU_{18} (configuration A) du modèle de régression quantile pour la détermination de la $LGC_{5\%}$ en fonction de l'incertitude de mesure sur les DJU_{18} et sur la consommation de chauffage. À gauche, seule l'incertitude sur les DJU_{18} est prise en compte ; à droite, l'impact conjoint des incertitudes sur le FA et la grandeur d'intérêt est étudié ($p = 10$). Différentes valeurs p de multiplication du nombre de simulations intégrant les erreurs de mesure sont comparées. Des résultats correspondants aux coefficients des FA de la configuration B sont détaillés à la Figure 4-5.

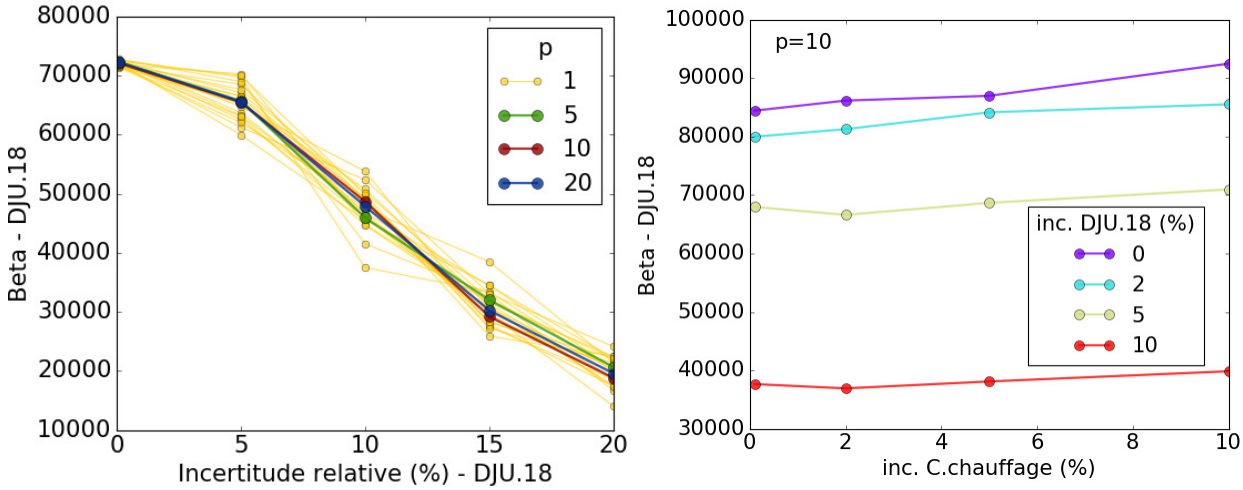

Figure 4-4 : Influence de l'incertitude de mesure des DJU_{18} et des consommations de chauffage sur le coefficient d'ajustement $\beta_{DJU_{18}}$

On peut d'abord noter l'influence importante de l'incertitude de mesure des DJU_{18} sur le coefficient d'ajustement de ce facteur, qui est divisé par 3 pour une incertitude de 20 %. Plus les erreurs de mesure sont importantes, moins marqué est l'ajustement. Ce phénomène s'explique par un étalement plus grand de la distribution des FA du fait des erreurs de mesure qui entrainent aussi une diminution du caractère explicatif de ce FA sur les variations de la grandeur d'intérêt. On constate par contre (à droite) une influence croissante mais assez faible de l'incertitude sur la grandeur d'intérêt sur la valeur du coefficient (phénomène d'hétéroscédasticité décrit précédemment).

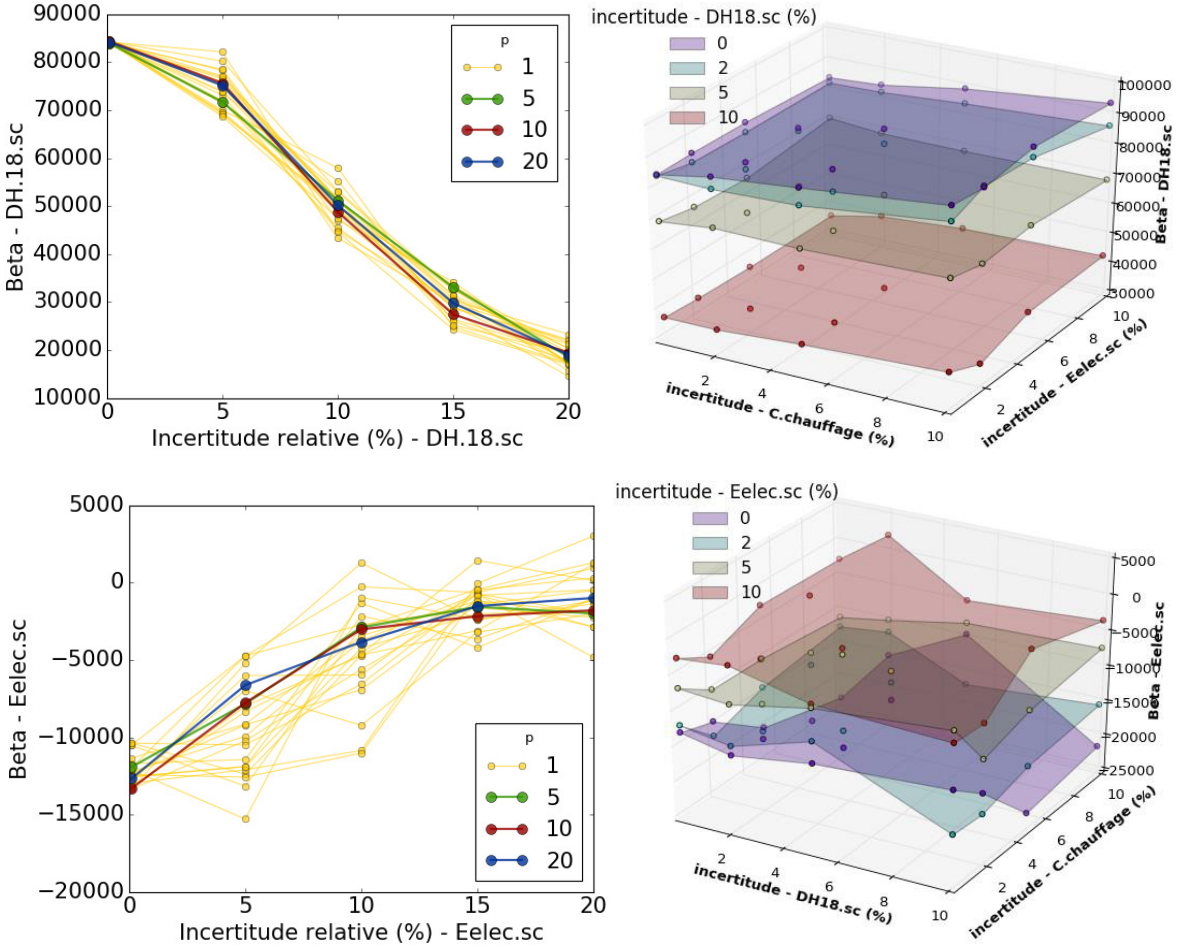

Figure 4-5 : Influence de l'incertitude de mesure des $DHU_{18.sc}$, des $E_{elec.sc}$ et des consommations de chauffage sur le coefficient d'ajustement $\beta_{DH_{18.sc}}$ (haut) et $\beta_{E_{elec.sc}}$ (bas)

Des constations similaires sont faites pour la configuration B. Les valeurs absolues des coefficients du modèle de régression quantile tendent vers 0 quand le bruit autour des valeurs des FA correspondants augmente. Si l'incertitude tend vers l'infini, le FA bruité par les erreurs de mesure n'est plus corrélé à la grandeur d'intérêt Y : le quantile conditionnel de $Y|FA$ est égal au quantile global de Y et le coefficient de pente est nul.

De même que dans la configuration A, l'incertitude de mesure des autres variables du modèle a peu d'influence sur le coefficient de pente d'un FA (figures de gauche).

Les Figure 4-4 et Figure 4-5 permettent de constater l'intérêt de dédoubler le jeu de simulations en associant plusieurs jeux d'erreur de mesure à un même résultat de simulation. On constate en effet l'instabilité du calcul des coefficients lorsque $p = 1$, en particulier pour le facteur $E_{elec.sc}$ dont l'influence est plus faible. Dès $p = 5$, les estimations sont plus précises et les valeurs semblent vraiment se stabiliser pour $p = 10$.

4.2.3. Modification de la limite garantie de consommation

La Figure 4-6 se rapporte à la configuration d'ajustement A et présente l'évolution de la limite garantie de consommation à 5 % de risque, $LGC_{5\%-ref}$, en fonction de l'incertitude de mesure sur le FA et sur la grandeur d'intérêt.

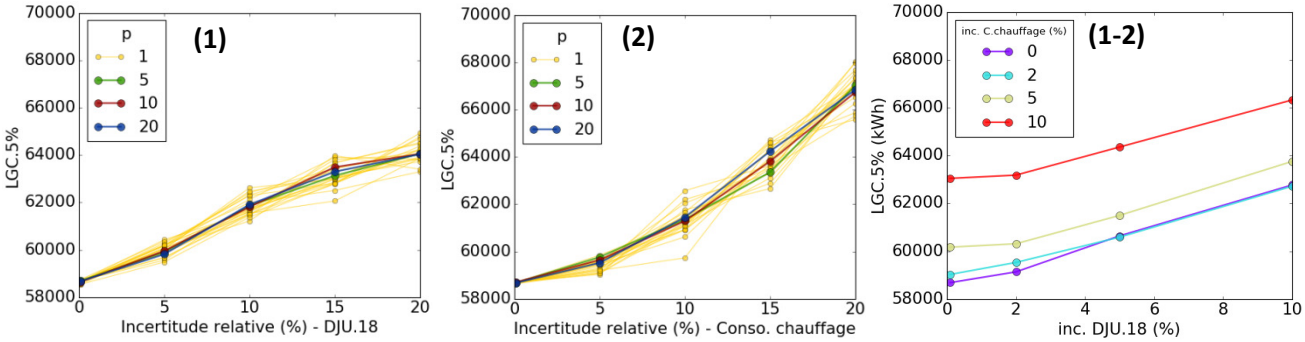

Figure 4-6 : Influence de l'incertitude de mesure des DJU_{18} (1) et des consommations de chauffage (2) sur la valeur de référence de la $LGC_{5\%}$

On constate sur la Figure 4-6-(1) que pour une plage d'incertitude des DJU_{18} de +/- 5 %, l'augmentation de la $LGC_{5\%}$ est de l'ordre de 3 % environ mais atteint plus de 12 % lorsque la plage d'incertitude est de +/- 15 %. On remarque par ailleurs l'influence de l'incertitude de mesure des consommations de chauffage (Figure 4-6-(2)) sur la valeur de l'engagement sur ces consommations, $LGC_{5\%-ref}$. Une incertitude relative de mesure de 10 % entraine une augmentation de la $LGC_{5\%-ref}$ de près de 5%. Les erreurs de mesure provoquent un élargissement de la dispersion des consommations et font donc logiquement augmenter le quantile conditionnel. La Figure 4-6-(1-2) permet de constater l'effet conjoint des incertitudes des différentes variables du modèle sur $LGC_{5\%-ref}$ ($p = 10$).

Les Figure 4-7 et Figure 4-8 sont associées à la configuration B. La consommation électrique étant un FA ayant un caractère explicatif plus faible, le bruitage des valeurs a un effet limité sur la $LGC_{5\%}$ (Figure 4-7-(2)). La Figure 4-8 montre l'effet sur la limite garantie de la variation simultanée de toutes les variables du modèle d'ajustement de la configuration B.

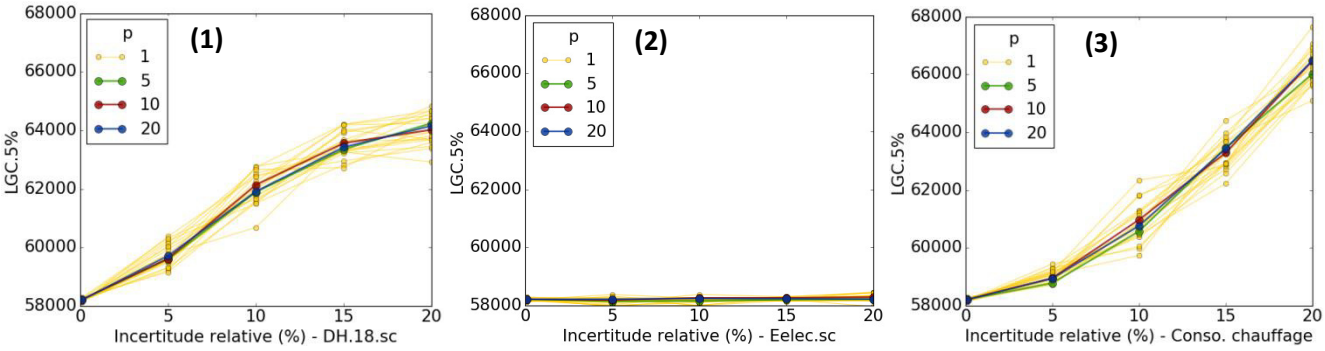

Figure 4-7 : Influence de l'incertitude de mesure des $DH_{18.sc}$ (1), des $E_{elec.sc}$ (2) et des consommations de chauffage (3) sur la valeur de référence de la $LGC_{5\%}$

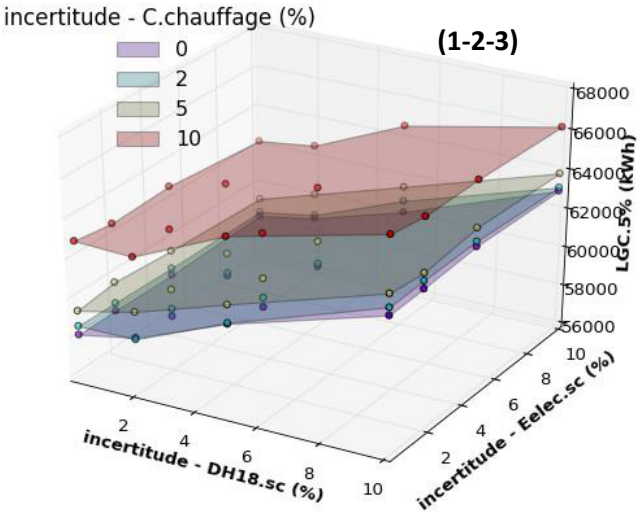

Figure 4-8 : Influence cumulée de l'incertitude de mesure des $DH_{18.sc}$, des $E_{elec.sc}$ et des consommations de chauffage sur la valeur de référence de la $LGC_{5\%}$ (p=10)

4.2.4. Application à la comparaison de PMV

Dans le cadre d'une GPE sur le bâtiment Feyzin, on suppose le choix de l'ensemble d'ajustement suivant pour l'engagement sur une limite garantie de consommation de chauffage suivant : $DH_{18.sc}$, $E_{elec.sc}$, et $T_{mp.cons.sc}$. Ce dernier FA correspond à la température de consigne moyenne sur la saison de chauffe entre les différents appartements pondérée par leur surface. Deux plans de mesure (PMV) sont envisagés. Les grandeurs à estimer sont la température extérieure, la température intérieure, l'énergie électrique et l'énergie de chauffage.

Le premier plan de mesure (PMV1) s'appuie sur l'utilisation d'un capteur de température extérieure dont l'incertitude type (systématique) est de 0,5 °C. L'incertitude-type sur les $DH_{18.sc}$ est calculée en multipliant cette incertitude par le nombre d'heures où la température moyenne horaire est inférieure à la température de référence de 18 °C (N_h). On fait l'hypothèse que les heures

comprises entre 17,5 et 18 °C d'une part et 18 et 18,5 °C d'autre part, qui pourraient être intégrées ou non aux degré-heures suivant l'erreur de mesure, se compensent. Concernant les températures de consigne, on suppose pour ce premier plan de mesure une incertitude type de 0,5 °C, l'erreur systématique étant la même pour tous les appartements (même série d'instruments par exemple). L'énergie électrique est mesurée avec une incertitude relative de 2 % et le compteur d'énergie évaluant la consommation de chauffage possède une incertitude type de 5 %.

Pour le second plan de mesure (PMV2), l'incertitude-type sur la température extérieure est de 0,2 °C. Pour les consignes, les incertitude-types sur chaque température moyenne d'appartement sont de 0,5 °C mais celles-ci sont considérées ici comme indépendantes entre les appartements. L'incertitude sur la moyenne pondérée sur l'ensemble des logements est donc finalement de 0,125 °C. L'énergie électrique est mesurée avec une incertitude relative de 1 % et le compteur d'énergie évaluant la consommation de chauffage possède une incertitude type de 2 %. Ces données sont récapitulées dans le Tableau 4-2.

Tableau 4-2 : Incertitudes de mesure des facteurs d'ajustement et de la consommation de chauffage pour deux plans de mesure différents

Incertitude-type	*PMV1*	*PMV2*
$DH_{18.sc}$	0,5 °C . N_h	0,2 °C . N_h
$T_{mp.cons.sc}$	0,5 °C	0,125 °C
$E_{elec.sc}$	2 %	1 %
$E_{chauffage}$	5 %	2 %

Des erreurs de mesure sont tirées aléatoirement dans des lois normales centrées sur les valeurs vraies, d'écart-type correspondant aux incertitude-types décrites ci-dessus et tronquées à plus ou moins deux fois ces valeurs. Une valeur multiplicative *p* de 10 est utilisée pour les deux configurations d'intégration des incertitudes de mesure. Trois modèles de régression sur l'espérance et de régression quantile à 5 % de risque associés aux différentes configurations de plans de mesure (PMV1,PMV2 et sans incertitudes) sont construits. Leur forme générale est la suivante :

$$LGC_{5\%} = LGC_{ref} + \beta_{DH_{18.sc}}.\left(\frac{DH_{18.sc}}{DH_{18.sc\,ref}} - 1\right) + \beta_{E_{elec.sc}}.\left(\frac{E_{elec.sc}}{E_{elec.sc\,ref}} - 1\right) + \beta_{T_{mp.cons.sc}}.\left(\frac{T_{mp.cons.sc}}{T_{mp.cons.sc\,ref}} - 1\right) \quad \textbf{(4.5)}$$

avec $DH_{18.sc\,ref} = 48766\,°C.h, E_{elec.sc\,ref} = 25455\,kWh\ et\ T_{mp.cons..sc\,ref} = 20{,}64\,°C$ et les coefficients des modèles dans le Tableau 4-3.

Concernant l'utilisation de températures comme FA, il est à noter que la mise en forme (ou changement de variable) consistant à exprimé le FA en taux d'accroissement ($\frac{T-T_{ref}}{T_{ref}}$) par rapport à

une valeur de référence pose un problème de signification physique, la température étant une grandeur exprimée de manière relative en degrés celsius. L'augmentation relative d'une température n'a pas vraiment de sens. Il faut donc considérer ce changement de variable comme une manipulation mathématique, permettant d'obtenir un ensemble de coefficients de régression adimensionnalisés.

Tableau 4-3 : Coefficients des modèles de régression quantile à 5 % de risque pour plusieurs configurations de plans de mesure

Coefficients	***PMV1***	***PMV2***	***sans incertitudes***
LGC_{ref}	5,914.10^4	5,695.10^4	5,623.10^4
$\beta_{DH_{18.sc}}$	7,618.10^4	8,411.10^4	8,818.10^4
$\beta_{E_{elec.sc}}$	-1,123.10^4	-1,754.10^4	-1,801.10^4
$\beta_{T_{mp.cons.sc}}$	8,588.10^4	1,614.10^5	1,619.10^5

On constate tout d'abord une augmentation de la LGC_{ref} avec l'imprécision du plan de mesure considéré (+ 1,3 % pour le PMV2 et +5,2 % pour le PMV1). En effet, les erreurs de mesure virtuelles entrainent une variabilité supplémentaire provoquant une plus grande dispersion des consommations énergétiques et des FA. Par ailleurs, ce bruitage provoque aussi une diminution des valeurs absolues des coefficients multiplicateurs des FA, du fait de cette variabilité supplémentaire.

La Figure 4-9 présente la distribution des résidus autour des modèles de régression sur l'espérance (ou de l'espérance globale pour le cas sans ajustement) pour les deux configurations de plan de mesure et pour les données de simulation brutes, sans erreur de mesure virtuelle.

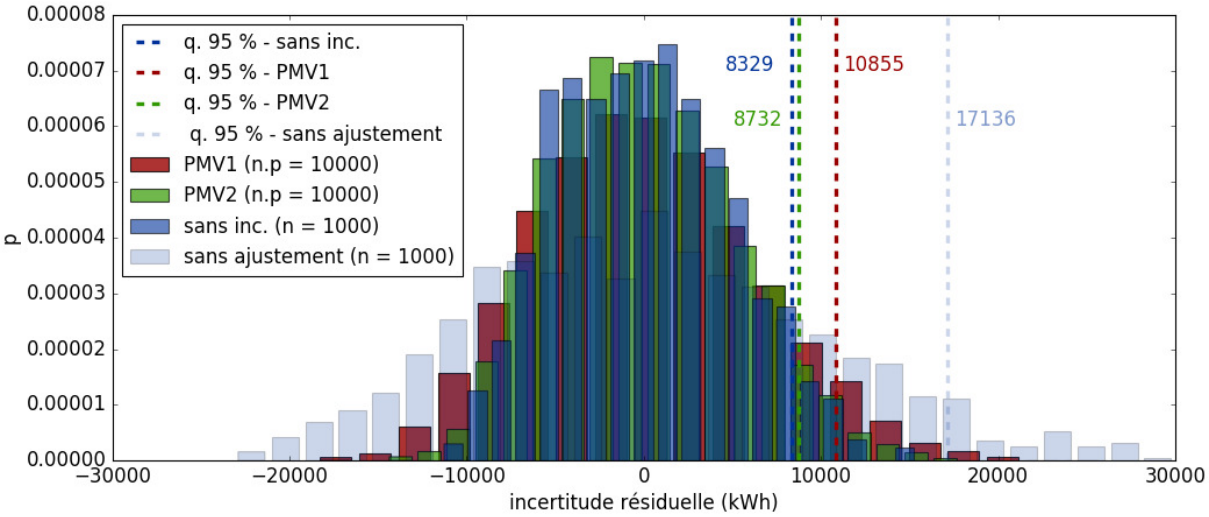

Figure 4-9 : Distribution et quantile à 95 % des résidus des modèles de régression linéaire sur l'espérance

On constate l'influence des erreurs de mesure virtuelles sur la dispersion des distributions résiduelles. Pour chaque configuration, on peut vérifier que la somme du quantile des résidus (en pointillé) et de l'espérance des consommations ajustées dans les conditions de référence (48310 kWh) se rapproche des $LGC_{5\%\,ref}$ fournies par les modèles de régression quantile (Tableau 4-3).

Cette étude théorique a permis de mettre en avant plusieurs phénomènes. Les coefficients d'ajustement tendent vers 0 lorsque l'incertitude des FA associés, donc le bruit autour des « valeurs vraies » de ces FA, augmente. Les incertitudes de mesure des FA et de la grandeur d'intérêt impactent le niveau d'engagement, la $LGC_{5\%}$, du fait de l'augmentation de la variabilité globale qui est engendrée. Il est donc important de prendre en compte ces incertitudes liées à la vérification effective des performances et des conditions d'exploitation et donc de parfaire la qualité des mesures afin de proposer des niveaux d'engagement optimisés.

D'un point de vue méthodologique, il est pertinent d'étendre la taille du jeu de données pour diminuer la variance des estimateurs de régression. Cependant, il est à noter que l'estimation du modèle de régression peut rencontrer des difficultés pour converger lorsque le nombre de données est trop important. Une valeur multiplicative p de 5 voire 10 est conseillée, entrainant des jeux de données de l'ordre de 5000 à 10000 N-uplet (X,FA).

4.3. Choix et optimisation des paramètres d'ajustement

4.3.1. Objectifs

L'intégration de facteurs liés à l'exploitation d'un bâtiment dans un processus d'ajustement fait face à plusieurs problématiques concrètes. Quelles sont les grandeurs les plus influentes à prendre en compte dans le modèle d'ajustement ? Quels sont les FA les plus pertinents pour décrire ces grandeurs ? Combien de FA faut-il considérer ? Les réponses à ces interrogations dépendent de plusieurs critères liés tout d'abord à la faisabilité et au coût des mesures associées mais aussi aux caractéristiques propres des bâtiments et à leurs façons de réagir aux différentes sollicitations dont on veut isoler l'impact via le processus d'ajustement. Les FA à utiliser en priorité sont ceux expliquant le mieux la variabilité de la grandeur d'intérêt de la GPE, la consommation énergétique.

Cette question ne peut être tranchée à l'aide d'une méthode d'analyse de sensibilité telle la méthode de Morris car les FA ne sont pas des entrées du modèle de simulation, mais plutôt des sorties complémentaires. Ils ne peuvent donc être échantillonnés selon un protocole détaillé. Par ailleurs, des entrées temporelles différentes peuvent être associées à une même valeur de FA mais provoquer des consommations énergétiques différentes. On ne peut donc pas étudiés d'effets élémentaires tels que définis dans la méthode de Morris. Dans cette partie, différentes méthodes statistiques sont étudiées afin d'identifier les meilleurs FA à prendre en compte.

4.3.2. Tests de corrélation

4.3.2.1. Indépendance des facteurs d'ajustement

Les tests de corrélation visent à identifier une dépendance entre deux variables statistiques. Dans l'optique de construire des modèles de régression, il est nécessaire de vérifier que les variables explicatives, les facteurs d'ajustement, sont indépendantes pour que les modèles identifiés soient pertinents.

Des coefficients de corrélation peuvent être calculés par différentes méthodes. Les méthodes de Kendall (τ_k) et de Spearman (r_s) sont toutes deux basées sur l'étude des rangs, elles sont non paramétriques. Les séries étudiées sont classées (rangées) pas ordre croissant puis la présence de paires de valeurs de mêmes rang est étudiée. Ces coefficients sont calculés de la manière suivant :

$$\tau_k = 2.\frac{N_{paires\ concordantes} - N_{paires\ discordantes}}{n.(n-1)} \quad \textbf{(4.6)}$$

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(rg(X_i) - rg(Y_i))^2}{n.(n^2-1)} \quad \textbf{(4.7)}$$

avec X et Y les variables statistiques dont la corrélation est étudiée.

La méthode de corrélation de Pearson (ρ_p) est paramétrique et mesure une corrélation linéaire entre les variables étudiées. Cette méthode est donc limitée aux hypothèses habituelles liées à la construction de modèles linaires. La distribution normale des données est ainsi souhaitée. Dans un cas contraire, les méthodes de Kendall et Spearman doivent être préférées. Dans ces trois méthodes, les coefficients sont compris entre -1 et 1. Plus la valeur absolue est grande plus les données sont corrélées. Les valeurs négatives correspondent à une évolution conjointe de sens opposé.

$$\rho_p = \frac{Cov(X,Y)}{\sigma_x\sigma_y} = \frac{E(XY) - E(X)E(Y)}{\sigma_x\sigma_y} \quad \textbf{(4.8)}$$

L'étude de la p-valeur du coefficient de corrélation permet de s'assurer de la significativité statistique du résultat. La p-valeur décrit la probabilité, sous l'hypothèse nulle selon laquelle il n'y a pas de corrélation significative, d'obtenir la valeur de coefficient de corrélation calculée. C'est-à-dire la probabilité d'identifier une corrélation par hasard. Plus la p-valeur est faible, plus on peut considérer avec confiance la corrélation identifiée.

Ces outils statistiques peuvent permettre dans une étude préalable de quantifier l'indépendance entre les FA que l'on souhaite intégrer au modèle d'ajustement. La prise en compte de FA non colinéaires contribue à la robustesse du modèle de régression identifiée et permet de mieux expliquer les variations de la grandeur d'intérêt. Afin d'illustrer ces éléments, des résultats du cas

d'étude de Feyzin sont présentés. 1000 simulations sont réalisées pour lesquels différents FA sont calculés. Dans les résultats ci-dessous, aucune incertitude supplémentaire liée à la mesure n'est prise en compte. La Figure 4-10 présente l'intercorrélation entre différents facteurs d'ajustement.

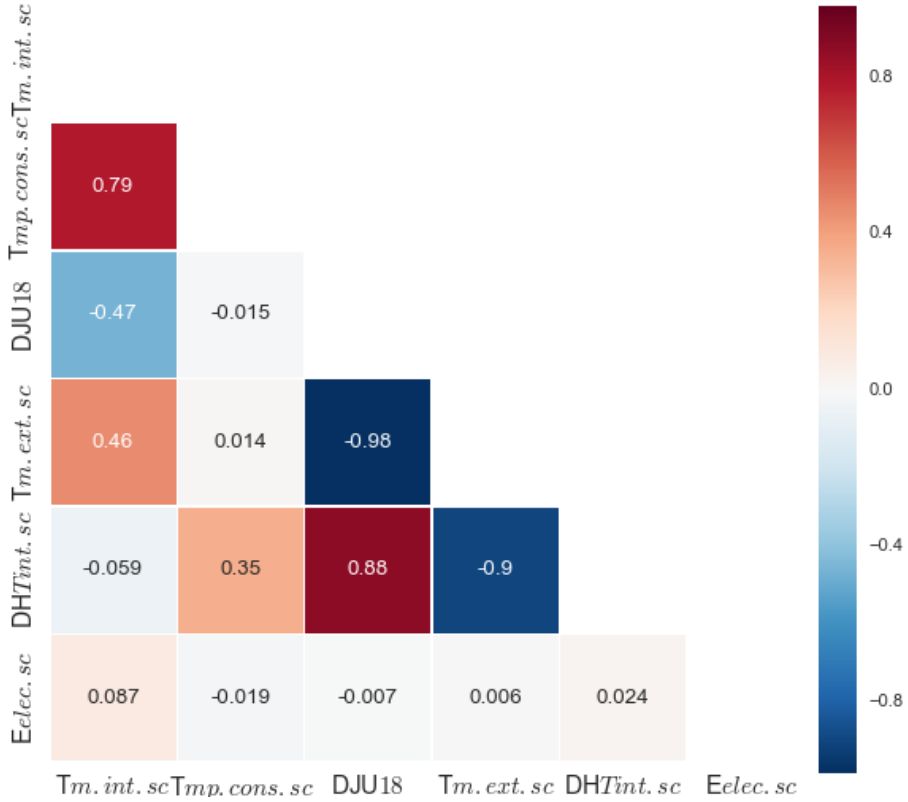

Figure 4-10 : Coefficients de corrélation de Spearman liant différents facteurs d'ajustement

L'étude de cette figure permet de confirmer certains éléments logiques, tels que la corrélation entre la température extérieure moyenne en saison de chauffe $T_{m.ext.sc}$, les DJU_{18} et les degré-heures à base température de consigne $DH_{Tint.sc}$. Ces derniers sont aussi corrélés à la température de consigne moyenne (pondérée) $T_{mp.cons.sc}$ mais assez peu à la température intérieure moyenne $T_{m.int.sc}$ sur la saison de chauffe (*sc*). On constate par ailleurs un fort lien entre la température intérieure moyenne et les FA liés à la température extérieure. Ceci s'explique par la saison de chauffe assez étendue prise en compte, durant laquelle des périodes de forts redoux sont fréquemment observées au sein des données météorologiques générées. Durant ces périodes de redoux, la température intérieure est fortement liée à la température extérieure et non plus à la consigne.

On peut ainsi identifier trois dimensions indépendantes parmi ces FA : une dimension liée à la température extérieure, une autre à la température de consigne et une dernière aux consommations d'électricité spécifique. Des FA comme les $DH_{Tint.sc}$ ou même la température intérieure moyenne, $T_{m.int.sc}$ intègrent à la fois l'impact de la température intérieure et extérieure. L'étude des corrélations entre FA permet donc de choisir un jeu de grandeurs indépendantes pour le processus d'ajustement.

4.3.2.2. Corrélation entre FA et consommations énergétiques

Les tests de corrélation peuvent aussi être utilisés pour caractériser le niveau d'interdépendance entre un facteur d'ajustement et la grandeur d'intérêt, la consommation énergétique de chauffage dans cette étude. Un classement entre différents FA peut ainsi être effectué, nourrissant la décision du choix des FA sélectionnés dans le processus d'ajustement. Dans la Figure 4-11 sont ainsi comparés les coefficients de corrélation de Spearman, de Kendall et de Pearson entre différents facteurs d'ajustement liés à la température extérieure et les consommations de chauffage, triés par ordre croissant (de ρ_p de Pearson).

On constate tout d'abord que les différents coefficients donnent des résultats similaires en ce qui concerne le classement des FA. Deux groupes principaux se dégagent et l'on remarque que la prise en compte des températures uniquement sur la saison de chauffe (sc) permet d'obtenir de meilleures corrélations. Les performances de plusieurs FA calculés sur cette période sont très proches : les degré-heures à base 18, à base 19, la température extérieure moyenne mais aussi les DJU, calculés avec la méthode classique ou la méthode COSTIC (DJU_c), dite aussi « des professionnels de l'énergie ». Celle-ci se différencie de la méthode classique (équation 2.1) par la gestion des journées ou la température de référence est entre les températures minimales et maximales. Le DJU considéré pour ce jour là est alors de :

$$DJU = \left(T_{ref} - T_{ext,min}\right) * \left(0{,}08 + 0{,}42 * \frac{T_{ref} - T_{ext,min}}{T_{ext,max} - T_{ext,min}}\right) \quad \textbf{(4.9)}$$

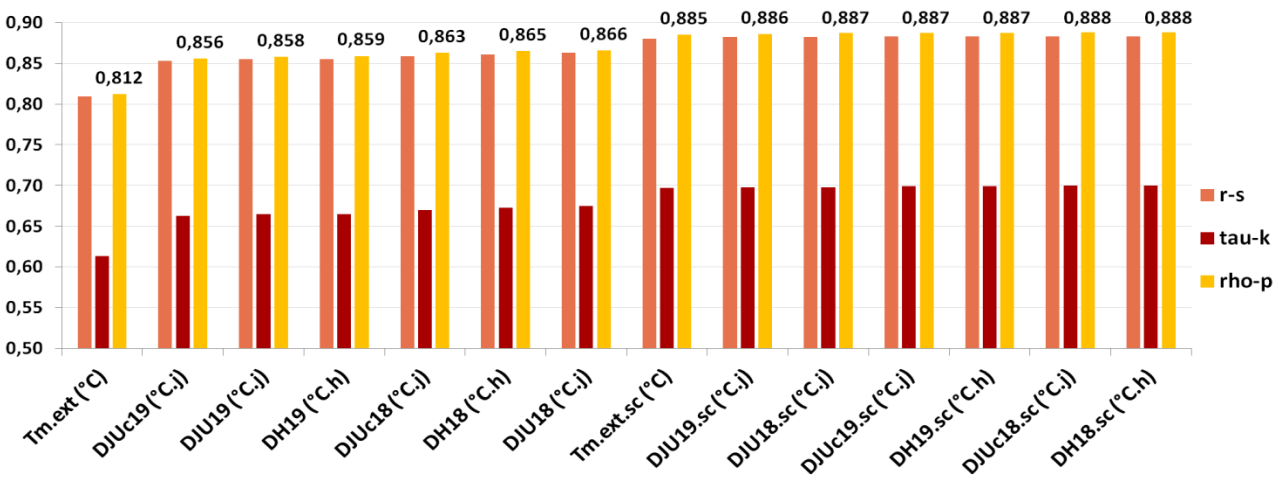

Figure 4-11 : Coefficients de corrélation entre différents facteurs liés à la température extérieure et les consommations de chauffage

Ces résultats semblables signifient que la connaissance de ces différents facteurs est à peu de chose près équivalente pour décrire et expliquer les variations de consommation énergétique de chauffage. Le meilleur FA est les $DH_{18.sc}$ puis les $DJU_{c.sc}$ à base 19 et 18 °C, puis $DH_{19.sc}$. L'indicateur

de degré-heures n'est donc pas significativement plus précis que celui des degré-jours. On peut aussi noter la moins bonne corrélation de la température extérieure moyenne. En effet comme expliqué précédemment, la température extérieure peut excéder la consigne de chauffage, durant les périodes douces en début et fin de saison de chauffe, et faire varier de manière plus ou moins prononcée ce facteur pour des périodes sans besoins de chauffage. Ceci entraine une décorrélation du facteur et de la consommation énergétique.

La Figure 4-12 présente les résultats d'une étude similaire sur des FA liés à la température à l'intérieure des appartements. On s'intéresse aux températures moyennes et aux températures de consigne dans les 16 appartements, pondérées ou non par leur surface, sur la saison de chauffe. Pour comparaison, les valeurs de ces températures sur un échantillon partiel d'appartements (8 logements) est aussi étudié investiguant la possibilité de ne réaliser des mesures que dans une partie des logements.

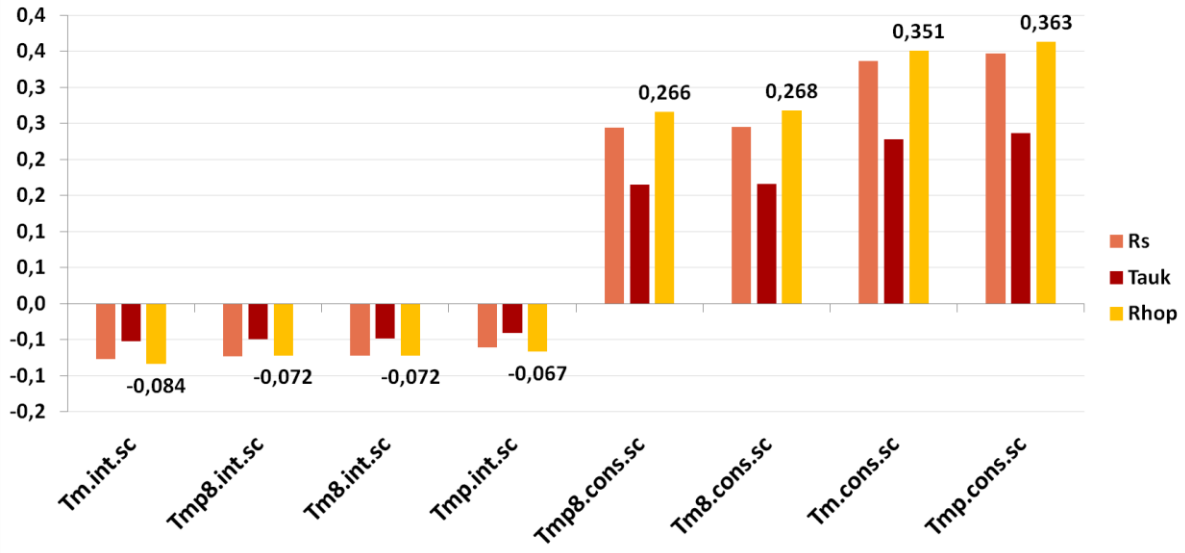

Figure 4-12 : Coefficients de corrélation entre différents facteurs liés à la température intérieure et les consommations de chauffage

Les FA liés à la température intérieure présentent une corrélation négative avec les consommations de chauffage alors que les FA liés à la consigne de chauffage ont une corrélation positive. Ceci s'explique par les périodes de début et de fin de saison de chauffe qui, en fonction les données météorologiques générées et utilisées en simulation, peuvent être plus ou moins chaudes. Ces périodes tirent fortement à la hausse les moyennes de température intérieure. Cette dépendance de la température intérieure moyenne à la température extérieure, s'appliquant simultanément sur tous les logements, est plus forte que la dépendance aux températures de consigne, dont la variabilité est assez faible car déjà foisonnée entre les 16 appartements. La température de consigne moyenne permet de s'affranchir complètement de cet effet lié aux périodes de redoux et ne capture donc que l'action des occupants sur leur consigne de chauffage. La

prise en compte de seulement 8 logements entraine une baisse importante des corrélations observées. Il n'est donc statistiquement pas pertinent ici de prendre un sous-échantillon, le nombre d'appartements étant trop petit. Les températures de consigne moyenne pondérée $T_{mp.cons.sc}$ sont mieux corrélées à la consommation énergétique que la moyenne $T_{m.cons.sc}$.

4.3.3. Qualité du modèle de régression

Le coefficient R^2 est un indicateur usuel de la pertinence d'un modèle de régression linéaire par moindre carrés pour décrire une distribution conditionnelle :

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \quad \textbf{(4.10)}$$

avec Y_i la valeur simulée, $\bar{Y}$ la moyenne de ces valeurs et $\widehat{Y}_i$ la valeur prédite par le modèle de régression. Ce coefficient correspond à la diminution relative de la variabilité autour du modèle linéaire par rapport à la variabilité totale.

Koenker et Machado (1999) proposent un indicateur R^1, adapté à la régression quantile, et qui constitue une mesure locale, dépendante du quantile étudié, de la qualité d'ajustement de la régression quantile linéaire. De manière analogue à R^2, elle est définie à partir du rapport entre les résidus du modèle de régression quantile complet non contraint (avec *p* variables indépendantes) et du modèle contraint (avec seulement une constante décrivant le quantile).

$$R^1(\tau) = 1 - \frac{\min\limits_{b \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i b)}{\min\limits_{b_0 \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(Y_i - b_0)} \quad \textbf{(4.11)}$$

La valeur de R^1, comprise entre 0 et 1 augmente lorsque des variables explicatives sont ajoutées au modèle.

Dans un premier temps on s'intéresse à la problématique du choix du meilleur facteur d'ajustement associé à une grandeur donnée : la température extérieure. Afin d'ajuster la limite garantie des consommations de chauffage en fonction de celle-ci, différents FA sont envisagés et les valeurs de R^2 et R^1 des modèles de régression de l'espérance et du quantile à 95 % sont comparés sur la Figure 4-13.

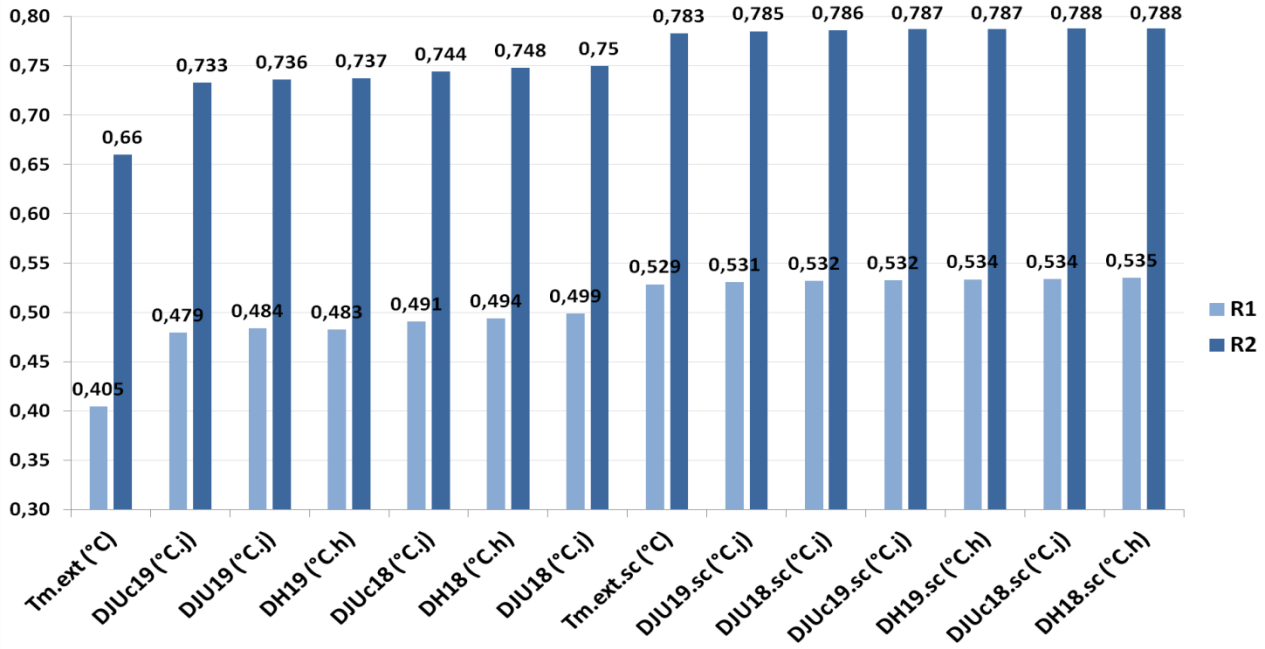

Figure 4-13 : Comparaison des coefficients de régression de différents modèles de régression linéaire et quantile (95 %) en fonction de FA liés à la température extérieure

Les FA sont triés par ordre croissant de R^2. On constate là aussi que les FA calculés sur la période de chauffe ont un meilleur caractère explicatif que les variables annuelles. Les conclusions sont similaires à celles du paragraphe précédent se rapportant à la Figure 4-11. Les meilleurs facteurs sont $DH_{18.sc}$, meilleur sur les deux coefficients de régression et $DJU_{c18.sc}$. La température de référence de 18 ou 19 °C n'a pas grande influence et il est confirmé que les DJU sont un indicateur aussi pertinent que les DH.

La Figure 4-14 présente les résultats pour les indicateurs liés aux températures à l'intérieur des logements.

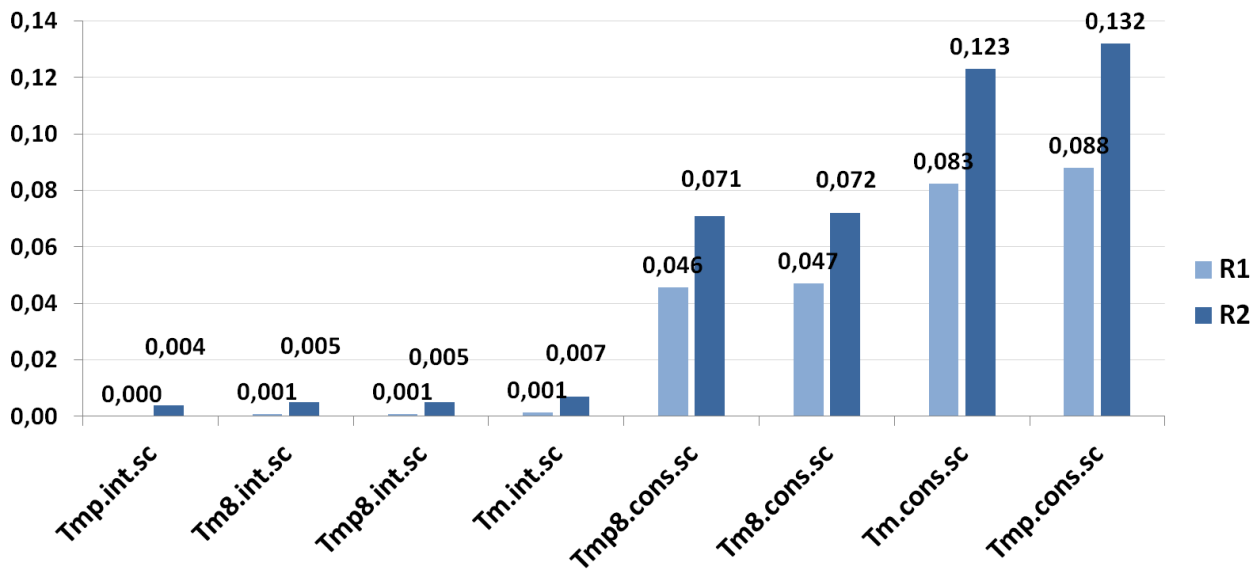

Figure 4-14 : Comparaison des coefficients de régression de différents modèles de régression linéaire et quantile (95 %) en fonction de FA liés aux températures intérieures et de consigne

Les FA sont classés par ordre croissant de R^2. Les résultats sont assez similaires à ceux de la Figure 4-12. Les facteurs liés à la température de consigne produisent de bien meilleurs modèles linéaires que ceux liés à la température intérieure, qui comme nous l'avons dit plus haut, est influencée par la température extérieure en période de redoux. Les valeurs observées des coefficients pour les FA liés aux températures intérieures sont si faibles qu'il n'est même pas pertinent de considérer un lien linéaire entre ceux-ci et les consommations énergétiques de chauffage. Les températures moyennes de consigne pondérées par les surfaces d'appartement produisent de meilleurs résultats que les températures moyennes. La prise en compte de seulement la moitié des appartements réduit fortement la qualité des modèles de régression.

Dans un second temps, différentes configurations sont comparées dans le but d'étudier l'intérêt d'ajouter ou non des facteurs d'ajustement indépendants supplémentaires au modèle. La Figure 4-15 présente les coefficients R^2 et R^1 des modèles de régression de l'espérance et du quantile conditionnels pour plusieurs jeux de FA.

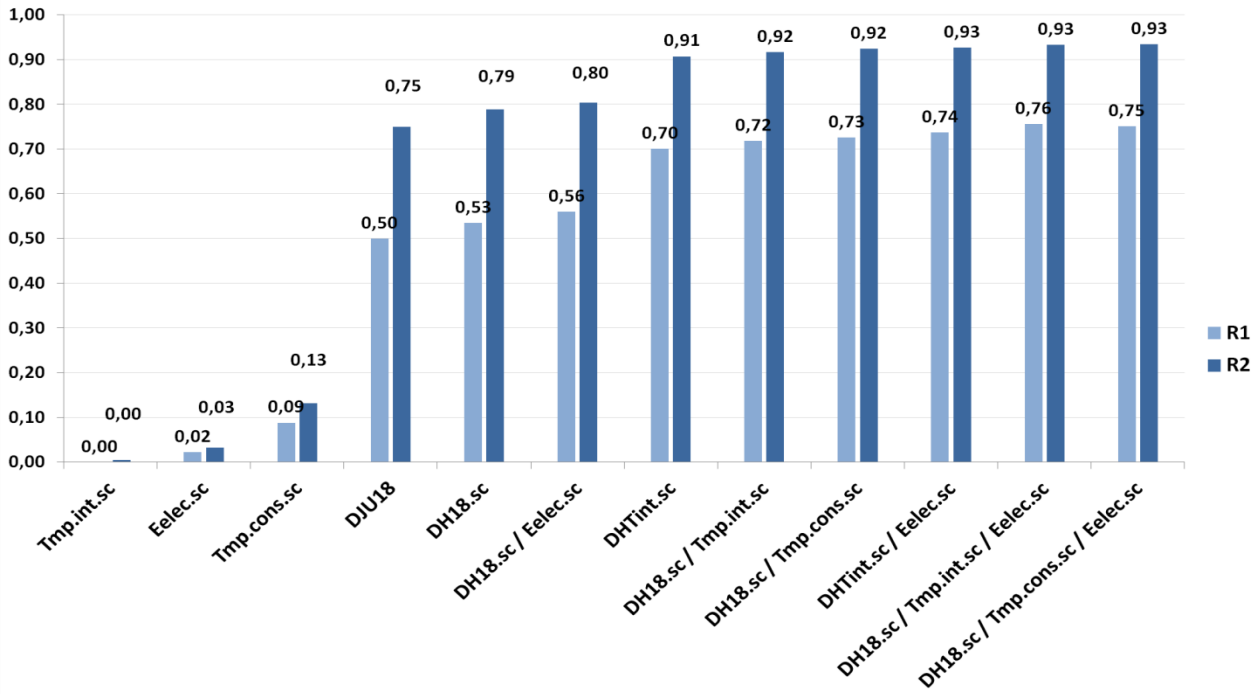

Figure 4-15 : Comparaison des coefficients de régression de différents modèles de régression linéaire et quantile (95 %)

Parmi, les FA uniques, la forte influence de la température extérieure (DJU, DH) est confirmée. L'ajout au $DH_{18.sc}$ de la température de consigne moyenne pondérée $T_{mp.cons.sc}$ constitue un meilleur modèle que l'ajout de la consommation électrique $E_{elec.sc}$. Les degré-heures à base température de consigne sont un bon FA mais reposent sur la connaissance des températures intérieures et extérieures. Le meilleur modèle d'ajustement est ici constitué de $DH_{18.sc}$, $T_{mp.cons.sc}$ et $E_{elec.sc}$. D'après la définition du R^2, l'écart-type résiduel des consommations de chauffage autour du modèle linéaire est réduit de 90 % par rapport à l'écart-type global. Pour le R^1, c'est l' « écart pondéré type » (par la fonction de pondération ρ_τ) autour du quantile conditionnel qui subit une diminution de 69 % par rapport à l'écart pondéré type autour du quantile global de la distribution.

Les distributions des résidus autour des modèles de régression de l'espérance sont présentées à la Figure 4-16 pour trois jeux de FA : les résidus autour de la moyenne globale des consommations (bleu), et les résidus des modèles de régression en fonction des DJU_{18} d'une part (rouge) et de l'ensemble $DH_{18.sc}$ / $T_{mp.cons.sc}$ / $E_{elec.sc}$ d'autre part (vert). Les valeurs des quantiles de ces résidus à 5 % de risque de dépassement sont aussi affichées. Ces différentes distributions caractérisent l'incertitude résiduelle non expliquée par le modèle d'ajustement. Les quantiles à 95 % des résidus sont affichés en pointillés.

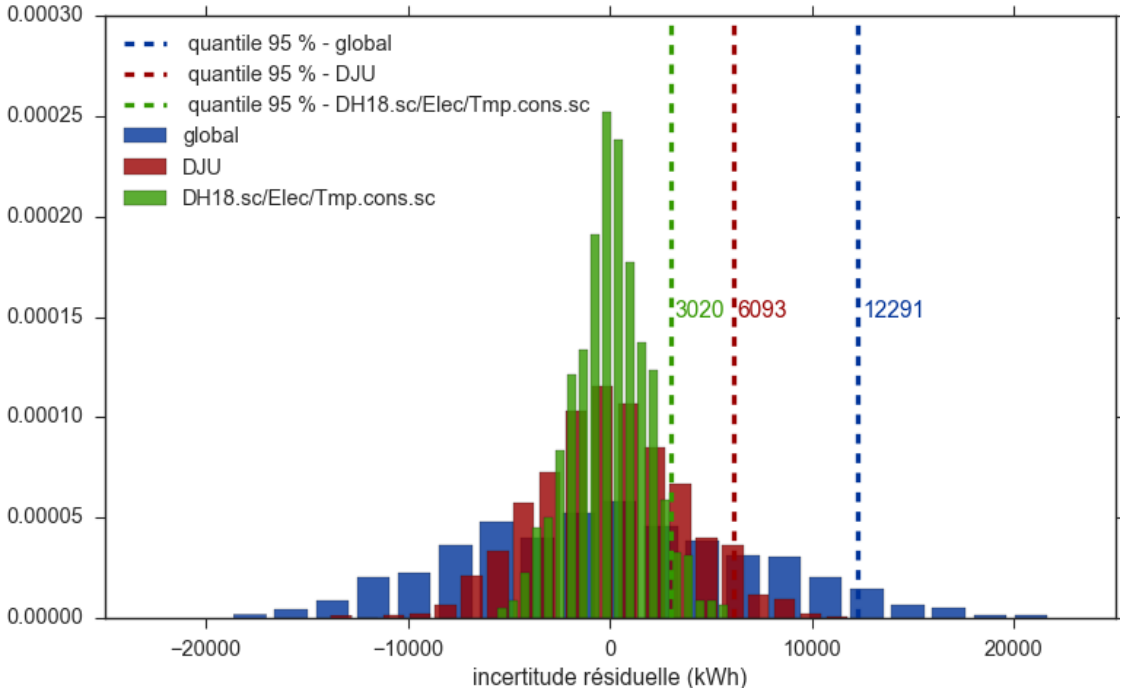

Figure 4-16 : Distribution des résidus des consommations énergétiques autour des modèles de régression sur l'espérance pour différents jeux de facteurs d'ajustement

Les coefficients de régression sont donc des indicateurs pertinents et complémentaires aux coefficients de corrélation pour choisir les FA les plus performants car ils permettent de quantifier la part de la variabilité expliquée par le modèle de régression, et donc de comparer différentes configurations. Le coefficient R^1 répond spécifiquement à cette problématique pour l'estimation de quantiles conditionnels.

4.3.4. Méthode d'AS par régression

Les méthodes d'analyse de sensibilité dont un état de l'art a été présenté au chapitre 1 visent à identifier les paramètres d'entrée d'un modèle dont la variation impacte le plus la variabilité de la sortie.

La régression standardisée est une des méthodes d'analyse de sensibilité communément utilisée en modélisation des bâtiments. Le terme de standardisation est lié à l'utilisation de variables explicatives X et expliquée Y centrées et réduites. Le changement de variable associé est décrit dans l'équation ci-dessous, où Y^{cr} est la variable centrée réduite, $\bar{Y}$ désigne la moyenne de la variable Y et $\sigma(Y)$ son écart-type :

$$Y_i^{cr} = \frac{Y_i - \bar{Y}}{\sigma(Y)} \qquad \textbf{(4.12)}$$

Les coefficients β du modèle sont ainsi adimensionnels. Une augmentation d'un écart type de la variable X_1, $\sigma(X_1)$, entraine une augmentation de $\beta.\sigma(Y)$ de la variable expliquée. L'analyse des coefficients obtenus permet de classer des variables explicatives en fonction de leur influence, les coefficients avec les plus grandes valeurs absolues étant associés aux FA ayant le plus d'impacts sur la sortie.

Le changement de variable associé au modèle de régression proposé (équation 4.13) est différent. Il décrit le taux d'accroissement (Y^{ta}) de la variable Y par rapport à sa moyenne $\bar{Y}$ (ou sa médiane) :

$$Y_i^{ta} = \frac{Y_i - \bar{Y}}{\bar{Y}} \tag{4.13}$$

La variable expliquée (l'espérance ou le quantile de la consommation d'énergie) n'est pas normalisée et est seulement exprimée en fonction d'une valeur de référence (centrée). Dans ce cas, l'augmentation de la variable explicative, le FA, de n % entraine une augmentation de la variable expliquée de $n.\beta$. On rappelle l'aparté fait au paragraphe 4.2.4 concernant le manque de sens physique de ce changement de variable pour les températures, de se rapporter à une température de référence exprimée en degrés celsius et de considérer une augmentation relative de température. Là-aussi, il faut considérer cette manipulation comme une opération mathématique.

Dans le cas centré, les coefficients décrivent donc l'impact d'une augmentation se rapportant à l'écart-type de chaque FA et dans le second à une augmentation relative par rapport à la moyenne de chaque FA. La comparaison des différents coefficients β a donc un sens différent dans ces deux cas et les rapports observés entre ces coefficients ne seront pas identiques.

Le Tableau 4-4 présente les résultats de régression moyenne et quantile avec les deux changements de variable, le taux d'accroissement (Y^{ta}) et la variable centrée réduite (Y^{cr}). Pour le modèle de régression quantile, avec taux d'accroissement, on retrouve les résultats présentés dans le Tableau 4-3.

Tableau 4-4 : Coefficients des modèles de régression de l'espérance et du quantile conditionnel pour différents jeux de facteurs d'ajustement

Coefficients	***Régression moyenne***		***Régression quantile (95 %)***	
	Y^{cr}	Y^{ta}	Y^{cr}	Y^{ta}
β_0	0,00	$4{,}91.10^{4}$	0,84	$5{,}62.10^{4}$
$\beta_{DH_{18.sc}}$	$7{,}98.10^{-1}$	$7{,}57.10^{4}$	$8{,}64.10^{-1}$	$8{,}81.10^{4}$
$\beta_{T_{mp.cons.sc}}$	$3{,}25.10^{-1}$	$1{,}40.10^{5}$	$3{,}87.10^{-1}$	$1{,}66.10^{5}$
$\beta_{E_{elec.sc}}$	$-9{,}31.10^{-2}$	$-1{,}52.10^{4}$	$-1{,}15.10^{-1}$	$-1{,}87.10^{4}$

La Figure 4-17 reprend les valeurs des coefficients ci-dessus mais une normalisation est réalisée, tous les coefficients associés aux $DH_{18.sc}$ étant ramenés à 1. On constate que les deux changements de variables donnent des résultats différents. Pour la variable centrée réduite, le coefficient associé à la température de consigne,$\beta_{T_{mp.cons.sc}}$, est près de deux fois inférieur à celui lié aux degré-heures. Ce rapport s'inverse pour la variable se rapportant au taux d'accroissement. Ceci s'explique par le fait que $T_{mp.cons.sc}$ a une plage de variation relativement réduite dans le jeu de données issu des simulations (du fait du foisonnement de la variabilité entre les 16 appartements notamment) mais qu'une faible variation a tout de même un fort impact sur les consommations. Y^{ta} ne s'intéresse qu'à ce dernier effet alors que Y^{cr} prend aussi en compte la plage de variabilité observée. Ainsi $DH_{18.sc}$ a un effet relatif moindre mais ses variations possibles étant importantes, son influence est mise en avant par le changement de variable centré réduit. Les deux informations sont donc intéressantes et pertinentes.

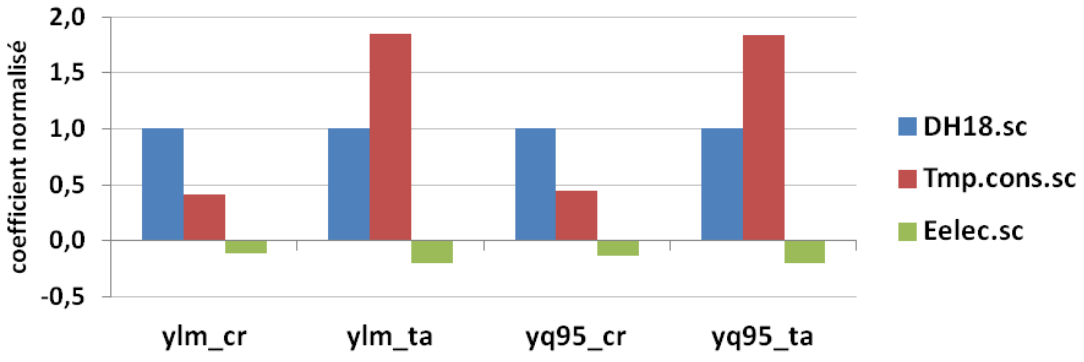

Figure 4-17 : Rapports normalisés entre les coefficients des modèles de régression sur l'espérance (y_{lm}) et sur le quantile (y_{q95}) pour les deux types de normalisation des FA (ta et cr)

Les modèles d'ajustement étant construits sur la base du second changement de variable, on propose d'utiliser directement les coefficients des modèles d'ajustement pour conclure sur l'influence des FA. L'étude des coefficients standardisés peut si besoin compléter l'analyse pour le choix des FA.

4.3.5. Tests statistiques

Dans l'objectif de sélectionner le meilleur jeu de FA pour le processus d'ajustement, des tests statistiques peuvent permettre d'étudier la pertinence d'ajouter de nouvelles variables à un modèle de régression donné. Ce problème statistique consiste à étudier le rejet de l'hypothèse nulle selon laquelle le modèle le plus simple est suffisamment performant par rapport au modèle plus complexe, dans lequel d'autres variables explicatives sont ajoutées. Il a fait l'objet de plusieurs développements.

Koenker et Bassett (1978) proposent une méthode pour tester de nouvelles variables explicatives sur la base du test de Wald consistant à étudier la robustesse des valeurs de coefficients de modèles de régression (méthode *Wald*). D'autres travaux sont effectués par Chen et al. (2008) qui analysent les résidus de la régression quantile (méthode *anowar*). Une troisième forme du test est proposée par Gutenbrunner et al. (1993), basée sur un test de rang (méthode *rank*). Ces différentes formes du test statistique permettent de sélectionner un nombre optimisé de variables. Le Tableau 4-5 présente les valeurs *p* (*p-value*) de ces tests statistiques étudiant l'intérêt d'utiliser un modèle 1 de régression quantile, comprenant un FA supplémentaire par rapport à un modèle 0 initial. Plus la valeur *p* est faible, plus les différences entre les deux modèles sont statistiquement solides et plus l'intérêt de ce FA supplémentaire est assuré. On considère comme trop faiblement significatives des valeurs-p supérieures à 0,05. Sont surlignées en vert, les valeurs inférieures au millième, en orange, inférieures au centième.

Tableau 4-5 : Valeurs *p* associées aux tests statistiques d'analyse de la variance entre deux modèles de régression quantile

Modèle 0	Modèle 1	*p-value*		
		Wald	***anowar***	***rank***
-	$DH_{18.sc}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$
-	$T_{mp.cons.sc}$	$< 10^{-15}$	$< 10^{-15}$	$5{,}2.10^{-5}$
-	$E_{elec.sc}$	$2{,}5\ .10^{-3}$	$2{,}2.10^{-4}$	$1{,}1\ .10^{-3}$
$DH_{18.sc}$	$DH_{18.sc} / T_{mp.cons.sc}$	$< 10^{-15}$	$< 10^{-15}$	$< 10^{-15}$
$DH_{18.sc}$	$DH_{18.sc} / E_{elec.sc}$	$1{,}2.10^{-3}$	$4{,}1.10^{-5}$	$3{,}07.10^{-4}$
$T_{mp.cons.sc}$	$T_{mp.cons.sc} / E_{elec.sc}$	$2{,}5.10^{-3}$	$1{,}7.10^{-3}$	$7{,}5.10^{-3}$
$DH_{18.sc} / T_{mp.cons.sc}$	$DH_{18.sc} / T_{mp.cons.sc} / E_{elec.sc}$	$1{,}1.10^{-4}$	$4{,}0.10^{-4}$	$5{,}1.10^{-5}$

On constate que les trois formes du test statistique donnent des résultats proches. Il y est confirmé que la configuration de FA étudiée est pertinente. On constate cependant que la prise en compte de l'électricité spécifique comme seul FA est considéré comme moins efficace, le quantile

conditionnel obtenu étant moins significativement différent du quantile global. L'usage de ce facteur en complément d'autres FA, $DH_{18.sc}$ et le couple $DH_{18.sc}$ / $T_{mp.cons.sc}$ entraine aussi des p-values plus élevées mais restant significatives.

Ainsi, ces tests permettent de s'assurer de l'intérêt statistique de prendre en compte des FA supplémentaires dans le modèle d'ajustement. Ils permettent donc de créer des modèles efficaces et parcimonieux.

4.3.6. Régression pénalisée

La régression pénalisée repose sur l'ajout d'un terme de pénalisation à la fonction objectif à minimiser. Ce terme pondère les coefficients du modèle linéaire standardisé (variables centrées réduites), venant pénaliser l'utilisation d'un trop grand nombre de variables explicatives. La minimisation permet ainsi d'opérer un tri entre les variables en faisant ressortir celles ayant un fort impact sur les variations de la grandeur d'intérêt. La standardisation des variables explicatives (variable centrée réduite) est nécessaire afin d'adimensionnaliser les coefficients et de pouvoir évaluer de manière pertinente les influences relatives.

Les équations ci-dessous décrivent les modes de calcul des modèles de régression linéaire de l'espérance (4.14) et de la régression quantile (4.15) pénalisées :

$$\beta_{\lambda.lr} = \frac{1}{n}\sum_{i=1}^{n}(y_i - {x_i}^T\beta)^2 + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \quad \textbf{(4.14)}$$

$$\beta_{\lambda.qr} = \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(y_i - {x_i}^T\beta) + \sum_{j=1}^{p} p_\lambda(|\beta_j|) \quad \textbf{(4.15)}$$

avec ρ_τ la fonction poids définie au chapitre 3, $p_\lambda$une fonction de pénalisation dépendant d'un paramètre λ. Ce paramètre correspond au poids de la fonction de pénalisation. L'étude de l'effet de la variation de ce paramètre sur les résultats de régression pénalisée permet d'identifier les variables explicatives les plus rapidement délaissées et donc les moins influentes.

Différentes méthodes de pénalisation ont été développées, se démarquant par la fonction de pénalisation utilisée.

- *LASSO*

La méthode LASSO repose sur une pénalisation exprimée comme la somme des valeurs absolues des coefficients du modèle de régression multipliée par un paramètre de pondération λ.

$$p_y(|\beta_j|) = \lambda|\beta_j| \quad \textbf{(4.16)}$$

- *SCAD*

La méthode *SCAD* a été introduite pour faire face à des problématiques d'élimination de nombreuses variables parasites en grande dimension. Avec la méthode *LASSO*, un coefficient λ élevé est nécessaire pour écarter un grand nombre de variables parasites, biaisant fortement les variables explicatives. La pénalisation *SCAD* vise à pénaliser plus fortement les petits coefficients et à réduire le biais sur les gros coefficients. L'équation ci-dessous décrit la fonction de pondération par morceaux correspondante. Un second paramètre a est introduit dans la méthode *SCAD*. Pour les faibles valeurs de β_j, le poids est linéaire, puis exprimé comme un polynôme du second degré pour les valeurs plus élevées, puis finalement constant.

$$p_\lambda(|\beta_j|) = \lambda|\beta_j| I(0 \leq |\beta_j| < \lambda) + \frac{a\lambda|\beta_j| - ({\beta_j}^2 + \lambda^2)/2}{a-1} I(\lambda \leq |\beta_j| \leq a\lambda) + \frac{(a+1)\lambda^2}{a-1} I(|\beta_j| > a\lambda) \quad \textbf{(4.17)}$$

Avec $a > 2$

Cette fonction pénalisation est tracée sur la Figure 4-18.

- *MCP*

La méthode *MCP* est une autre alternative à la pénalisation *LASSO* et s'appuie sur une autre pondération par morceaux (deux plages) en fonction de la valeur absolue des coefficients du modèle de régression, dans le but de limiter le biais sur les coefficients pertinents. Les fonctions associées sont différentes, de second ordre pour les faibles valeurs de β_j puis constante pour les valeurs plus élevées (équation ci-dessous).

$$p_\lambda(|\beta_j|) = \lambda\left(|\beta_j| - \frac{{\beta_j}^2}{2a\lambda}\right) I(0 \leq |\beta_j| < a\lambda) + \frac{a\lambda^2}{a2} I(|\beta_j| > a\lambda) \quad \textbf{(4.18)}$$

Avec $a > 1$

La Figure 4-18 présente l'évolution de la fonction de pénalisation en fonction des valeurs d'un coefficient du modèle de régression pour $\lambda = 1$ et $a = 1$. Pour ces trois méthodes, l'analyse consiste à étudier la variation des coefficients de régression β_j en fonction de λ, le niveau de pénalisation. Celui-ci augmentant, les variables les moins explicatives, ou apportant une information redondante, sont progressivement éliminées. Le paramètre a des méthodes *SCAD* et *MCP* est par défaut fixé à sa valeur minimale (respectivement 2 et 1) et peut faire l'objet d'une variation paramétrique, en cas de difficulté d'identification des variables par exemple.

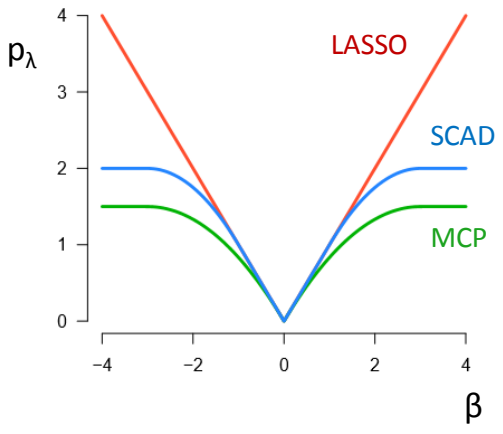

Figure 4-18 : Pénalisation d'un coefficient en fonction de sa valeur pour trois méthodes de régression pénalisée ($\lambda = 1$ et $a = 1$)

La régression pénalisée peut donc être utilisée pour sélectionner les FA les plus influents. Une régression LASSO sur l'espérance (moindres carrés) est effectuée sur un jeu de sept facteurs liés à la température extérieure, aux températures intérieures ou de consigne et aux consommations électriques. Les différents FA ont été standardisés (variable centrée réduite). Les résultats de cette régression LASSO sont présentés à la Figure 4-19. Le paramètre de pénalisation λ est tracé en abscisse en échelle logarithmique. Le modèle est composé de variables explicatives qui ne sont pas indépendantes : trois se rapportent directement à la température extérieure et trois aux températures intérieures. Cette redondance d'information entraine rapidement, pour de faibles valeurs de λ, une élimination de certaines variables ($DH_{Tint.sc}$, DJU_{18} puis $T_{m.ext.sc}$). Pour une plage assez large correspondant à un λ compris entre 10^{-1} et 10^{0}, les FA restants sont, dans l'ordre d'importance, les $DH_{18.sc}$, $T_{mp.cons.sc}$, $T_{m.int.sc}$ *et* $E_{elec.sc}$, et les coefficients correspondants sont assez stables. On constate donc que la connaissance conjointe des moyennes des températures intérieures et des températures de consigne est intéressante et permet de mieux expliquer les besoins de chauffage. Pour une pénalisation encore plus élevée, le facteur $DH_{Tint.sc}$ est de nouveau sélectionné car il agglomère à lui seul de l'information sur les températures intérieures et extérieures, entrainant de ce fait un abandon des variables $T_{m.int.sc}$ puis $DH_{18.sc}$. Cette méthode nous permet de conclure que le meilleur jeu de FA est $DH_{18.sc}$ / $T_{mp.cons.sc}$ / $E_{elec.sc}$. Par ailleurs, $T_{m.int.sc}$, dépendant à la fois des consignes et de la température extérieure dans les périodes de redoux est complémentaire des FA directement associés à ces grandeurs. Si un seul FA devait être choisi, ce serait les degré-heures à base température intérieure moyenne sur la saison de chauffe. Cependant, il est tout de même nécessaire de mesurer l'ensemble des températures pour obtenir ce FA unique et dans ce cas, le couple $DH_{18.sc}$ / $T_{mp.cons.sc}$ est plus performant.

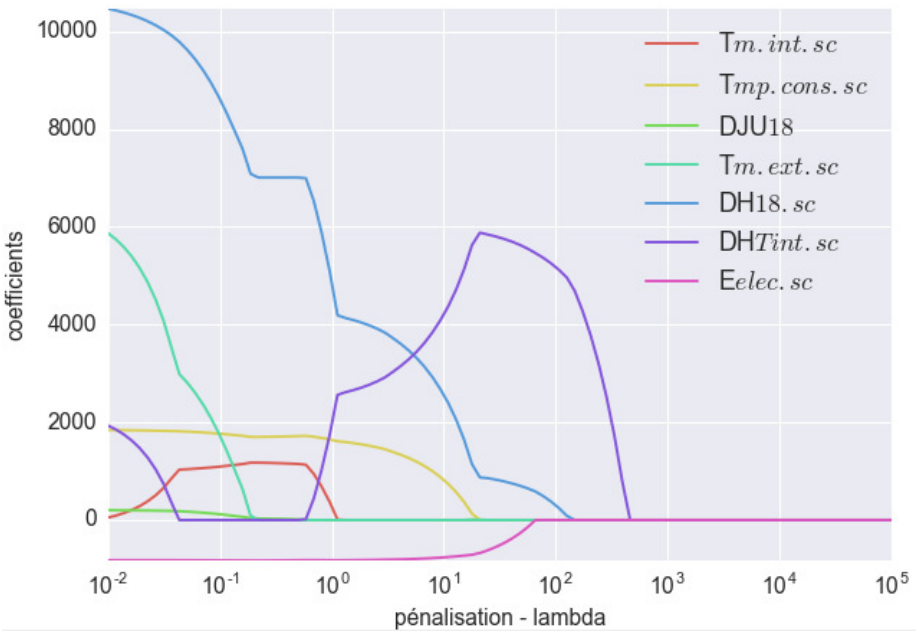

Figure 4-19 : Régression LASSO sur l'espérance pour un ensemble de FA non indépendants

Ces résultats sont comparés avec des méthodes de régression quantile pénalisée décrites précédemment. La Figure 4-20 présente les coefficients de régression quantile LASSO (gauche) et SCAD (droite). Les résultats pour la méthode MCP sont très proches de ceux de la méthode SCAD et ne sont pas affichés ici. Les valeurs des coefficients sont plus instables, la convergence étant plus difficile à atteindre pour l'estimation du quantile que pour celui de l'espérance. L'élimination de variables se fait de manière plus nette pour la méthode SCAD. Les profils observés sont similaires à la régression LASSO sur l'espérance et permet de tirer des conclusions semblables.

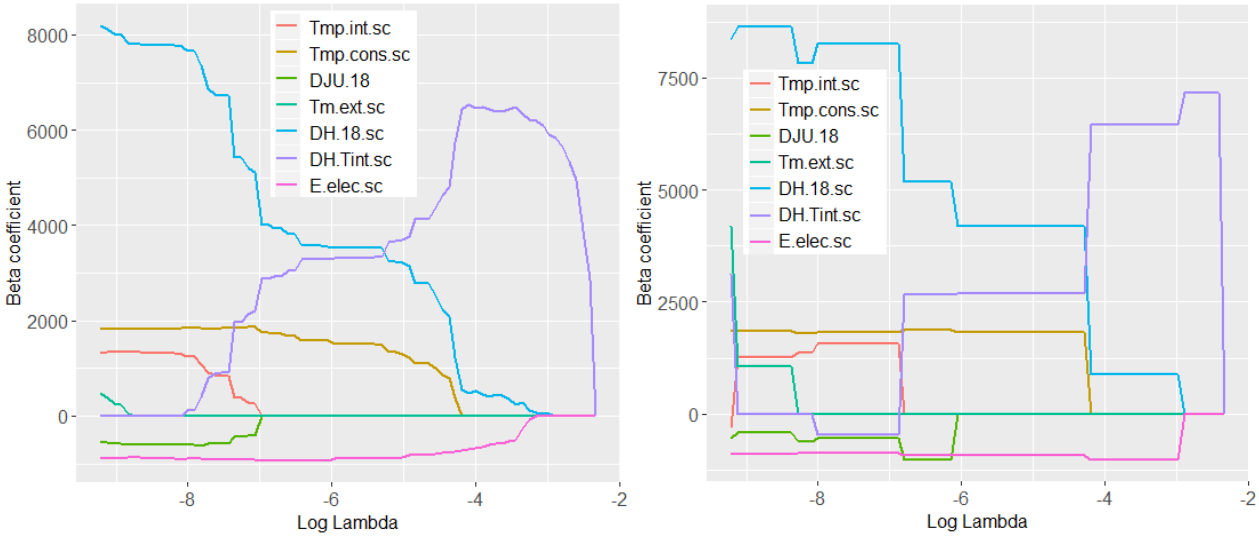

Figure 4-20 : Régression LASSO (gauche) et SCAD (droite) sur le quantile à 95 % pour un ensemble de FA dépendants

Les Figure 4-21 et Figure 4-22 s'intéressent à un ensemble de FA indépendants et à l'impact relatif sur les consommations de chauffage entre ces FA. On constate la prédominance de l'influence des degré-heures devant la température de consigne moyenne pondérée puis des consommations électriques. Ces résultats recoupent ceux de l'analyse de sensibilité par régression pour des variables centrées réduites présentées à la Figure 4-17.

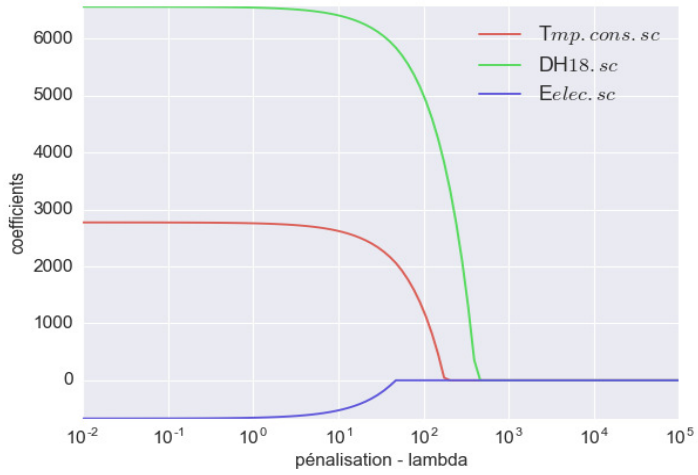

Figure 4-21 : Analyse de sensibilité par régression LASSO sur l'espérance pour un ensemble de FA indépendants

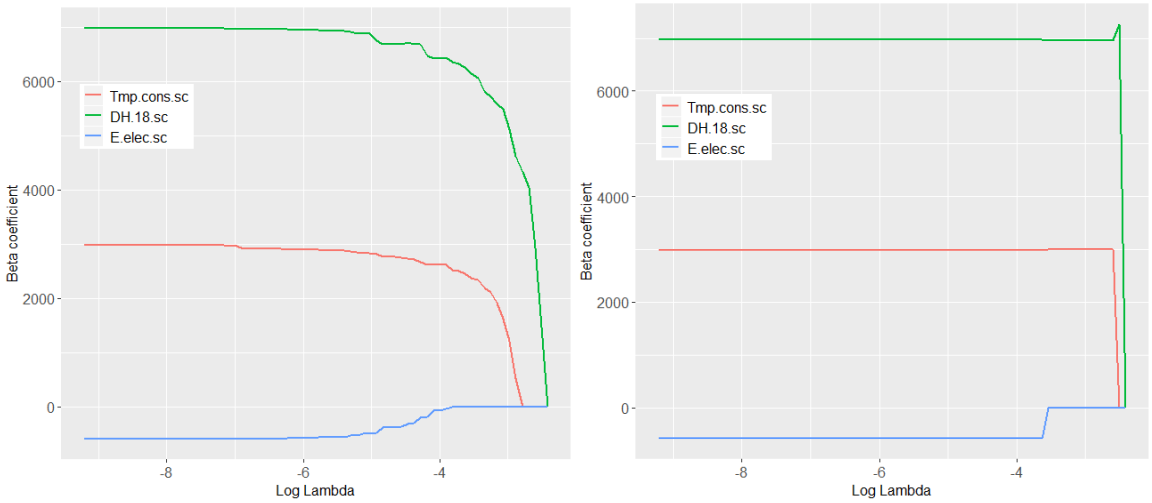

Figure 4-22 : Analyse de sensibilité par régression LASSO (gauche) et SCAD (droite) sur le quantile à 95 % pour un ensemble de FA indépendants

Les méthodes de régression pénalisée permettent à la fois de traiter la question de la sélection de variables dépendantes, par la visualisation des FA conservés lorsque la pénalisation augmente, et la question de l'estimation de l'influence relative des facteurs par le calcul de coefficients standardisés.

4.3.7. Bilan pour l'aide au choix des facteurs d'ajustement

Dans cette partie, différentes méthodes d'analyse des résultats de simulation ont été présentées. L'étude préalable des corrélations entre les facteurs d'ajustement permet de s'assurer de leur indépendance, préalable à la construction de modèles de régression robustes. Le choix des facteurs d'ajustement peut se séparer en deux problématiques : la sélection du FA le plus pertinent associé à une grandeur sur laquelle on souhaite procéder à un ajustement et le choix d'un ensemble optimisé de FA indépendants.

L'étude des corrélations entre les FA et la sortie d'intérêt, et l'étude des coefficients de régression permet de classer les FA selon leur caractère explicatif des variations de la sortie d'intérêt. Ces méthodes permettent donc de répondre aux deux problématiques, via la comparaison de facteurs un à un, ou par l'opposition de différents jeux de FA. Ces méthodes et leurs différentes variantes apportant sensiblement les mêmes informations, on propose d'utiliser prioritairement l'étude des coefficients de régression R^1 du modèle du quantile conditionnel. Elle répond par définition à la question de la quantification de la part de variabilité du quantile expliqué par les facteurs d'ajustement.

Les tests statistiques et les régressions pénalisées fournissent une information sur la constitution d'un jeu de FA complet parcimonieux, qui décrive au mieux les variations de la sortie mais qui soit le plus simple possible. L'analyse des résultats des méthodes de pénalisation, telle que la méthode *SCAD*, semble plus aisée et ces méthodes apportent des connaissances assez larges sur la sensibilité du modèle aux différents FA et sur la priorité à accorder aux différentes variables explicatives.

L'utilisation combinée de ces méthodes peut ainsi permettre la constitution de modèles d'ajustement efficients et d'en fournir une évaluation.

4.4. Méthodes alternatives pour les modèles d'ajustement

Cette partie vise à tester la possibilité d'utiliser des modèles d'ajustement différents de la régression quantile linéaire (RQL) pour caractériser la relation entre le quantile conditionnel de la distribution d'intérêt et les facteurs d'ajustement.

Cette méthode statistique, mise en avant jusqu'ici, possède l'avantage d'être paramétrique et de ne reposer que sur un nombre limité de paramètres et donc d'être facilement compréhensible par tous les acteurs concernés par un projet de garantie de performance. Cependant, d'autres types de modèles peuvent fournir de meilleures performances en termes de caractère explicatif des FA sur la variabilité de la grandeur d'intérêt. Un autre objectif est d'étudier la solidité de l'hypothèse d'une relation linéaire entre le quantile conditionnel de la grandeur d'intérêt et les FA.

En complément des configurations A et B de facteurs d'ajustement utilisées jusqu'ici sur le cas d'étude, une autre application (configuration C) est envisagée : sur le même bâtiment, une consigne de climatisation de 24 °C est considérée entre les semaines 25 et 32 (*sr* : saison de refroidissement). Cette faible consigne est choisie afin d'augmenter la sensibilité du bâtiment aux surchauffes estivales. On s'intéresse aux besoins de climatisation.

4.4.1. Ajustement de la SED

Une des méthodes couramment utilisées dans la littérature (Titikpina, 2016 ; Caucheteux et al., 2018) et dans l'industrie (Méthode OXYGEN) consiste à réaliser un ajustement du fichier de simulation énergétique dynamique après que des grandeurs aient été mesurées en exploitation. Les conditions météorologiques peuvent notamment être mises à jour. Un ajustement théoriquement plus fin et plus précis est permis car des données d'entrée dynamiques peuvent être mesurées puis spécifiées dans le modèle détaillé.

On étudie ce mécanisme en considérant un ajustement des conditions météorologiques sur le fichier de simulation : les séries horaires de température extérieure et de rayonnement. Sur la Figure 4-23 sont comparées les distributions de consommation énergétique de chauffage issues de trois PIV : tout d'abord de la PIV complète avec les incertitudes sur les paramètres statiques et les modèles de variabilité météorologique et d'usage ; puis deux autres distributions issues d'une propagation des incertitudes et variabilités d'usage qui correspondent à des configurations ajustées et n'intégrant donc pas de variabilité des conditions météorologiques. Celles-ci sont décrites par les conditions types (fichier TRY de Mâcon) (en rouge), et par les conditions météorologiques observées lors de la première année (2012-2013) d'exploitation du bâtiment (en vert).

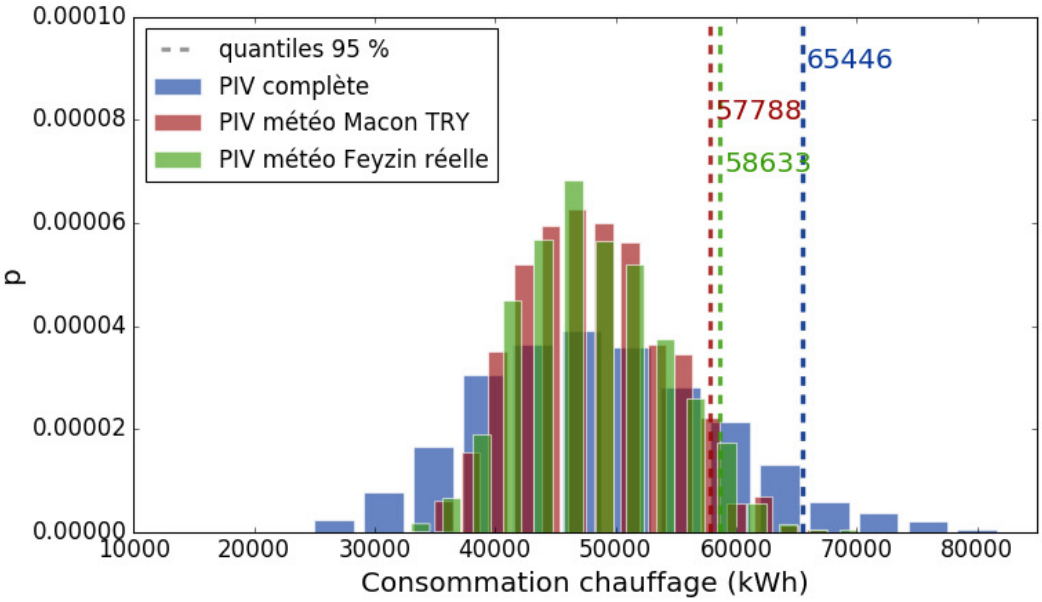

Figure 4-23 : Distributions des consommations de chauffage avec et sans ajustement sur le fichier SED des conditions météorologiques horaires

Le quantile à 95 % pour la météo mesurée est légèrement supérieur à celui pour la propagation avec météo type, passant de 57778 à 58633 kWh. En effet, la météo observée à Feyzin sur l'année d'exploitation, bien que légèrement plus défavorable, est assez conforme aux données types de Mâcon. Le Tableau 4-6 présente une comparaison des résultats d'ajustement sur le modèle SED et avec des modèles linéaires simplifiés issus de régressions quantiles (avec incertitude statistique et sans incertitudes de mesure) construites à partir des résultats de la PIV complète (bleu). Le modèle d'ajustement sur les DJU_{18} (équation 4.3) est utilisé ainsi qu'un autre modèle monodimensionnel, présenté ci-dessous, sur les $DH_{18.sc}$, meilleur FA sur les températures extérieures d'après les conclusions du paragraphe 4.3.

$$LGC_{5\,\%} = (5{,}96 + 0{,}06).10^4 + 8{,}51.10^4 \left(\frac{DH_{18.sc}}{DH_{18.sc-ref}} - 1\right) \qquad \textbf{(4.19)}$$

avec $DH_{18.sc-ref} = 49560$ °C.h

Tableau 4-6 : Comparaison de l'ajustement de la $LGC_{5\,\%}$ sur les conditions météorologiques avec le fichier SED et avec des modèles de régression quantile linéaire (RQL)

	Conditions types Mâcon			**Conditions Feyzin A1**		
Modèles	Fichier SED	RQL DJU_{18}	RQL $DH_{18.sc}$	Fichier SED	RQL DJU_{18}	RQL $DH_{18.sc}$
ajustement	T_{ext} et RGH h.	2212 °C.j	47521 °C.h	T_{ext} et RGH h.	2213 °C.j	50968 °C.h
$LGC_{5\,\%}$ (kWh)	57788	58220	56700	58635	58220	62620

Les DJU_{18} annuels sont très similaires dans les deux fichiers de température extérieure. L'ajustement de la $LGC_{5\,\%}$ sur le modèle linéaire se fait aux alentours de 58200 kWh, légèrement

supérieur au quantile à 95 % obtenu avec les fichiers SED. Les $DH_{18.sc}$sont moins proches et l'ajustement avec les modèles correspondants fournissent des $LGC_{5\,\%}$ de 56700 kWh pour les données types et de 62620 kWh pour les données de Feyzin à comparer respectivement à 57788 et 58635 kWh obtenus par l'ajustement sur le fichier SED. Les niveaux garantis sont donc très proches pour les deux méthodes et même inférieurs dans ce dernier cas pour l'ajustement par régression quantile sur des FA simplifiés.

L'ajustement des variations dynamiques des conditions d'exploitation sur le modèle de SED permet théoriquement d'atteindre un niveau de garantie plus précis, du fait de la description temporelle plus détaillée des conditions d'ajustement et de la plus grande précision du modèle. Cependant, cette méthode n'est pas directement comparable à l'approche décrite dans cette thèse. Tout d'abord, elle suppose de s'engager sur un niveau de performance indirectement défini par un fichier de simulation, dont une partie des entrées devront être mises à jour lors de la phase de vérification. Cette définition indirecte peut poser des problèmes de contractualisation. La réalisation de tests préalables, d'anticipation de l'ajustement, est rendue difficile du fait du grand nombre de degrés de liberté à cause des données temporelles ajustables.

Techniquement, la mise en œuvre de l'ajustement repose sur une intervention experte plus lourde pour mettre à jour les fichiers de simulation. Ceci entraine un manque de transparence dans la vérification du contrat même s'il est possible d'envisager qu'une tierce partie réalise cette étape de manière indépendante, complexifiant cependant le processus.

La connaissance des variations temporelles des grandeurs d'ajustement impose des contraintes spécifiques sur les plans de mesure à prendre en compte. Par exemple l'ajustement en fonction des consommations électriques détaillées n'est pas possible dans ce cas d'étude car nous avons seulement accès aux consommations globales relevées sur compteur à la fin de la période annuelle. Une acquisition dynamique par appartement permettrait d'ajuster le fichier SED.

De plus, ce mécanisme impose une souplesse importante des modèles de variabilité s'il veut tout de même être couplé à l'utilisation d'un modèle physico-probabiliste décrivant incertitudes et variabilité. Par exemple, le modèle de génération stochastique de données liées au comportement des occupants utilisé dans ces travaux résulte d'une chaine causale complexe. La spécification forcée, pour l'ajustement, de données partielles sur l'usage n'est pas compatible avec une description stochastique du reste des données. Par exemple, si le profil de consommation électrique est renseigné (ajusté), on ne peut pas générer toutes les configurations possibles d'occupation potentiellement associées (nombre d'occupants, profil des occupants déterminant la consigne, les puisages d'ECS…).

Cette méthode ne répond donc pas aux objectifs initiaux de détermination d'un modèle d'ajustement estimé en phase amont du projet, contractualisable et transparent. Elle constitue cependant une base de comparaison intéressante car elle fournit une valeur de $LGC_{5\,\%}$ plus précise (et théoriquement la plus faible possible), et permet donc d'évaluer la dégradation liée à l'utilisation

d'un modèle d'ajustement simplifié. Dans les exemples présentés ci-dessus, les $LGC_{5\,\%}$ obtenus avec les deux méthodes sont très proches, celles déterminées par régression quantile étant même inférieures dans une partie des cas. Les pertes d'information due à la prise en compte de FA globaux et de précision du fait de l'utilisation d'un modèle linéaire simplifié sont donc limitées, soulignant la pertinence de l'approche retenue.

Les méthodes alternatives d'ajustement présentées par la suite se replacent dans le contexte d'étude de ces travaux, où une estimation a priori d'un modèle simplifié et opposable doit être déterminé à partir des résultats obtenus par la simulation du modèle physico-probabiliste.

4.4.2. Régression quantile glissante

Dans un premier temps et afin d'étudier la validité de l'hypothèse de linéarité, le quantile conditionnel peut être évalué par une estimation de modèles RQL, entrainés sur une fraction glissante de l'échantillon de données. En N dimensions, l'espace des variables explicatives est divisé en différents sous-espaces sur lesquels sont estimés des modèles de régression différents.

Cette méthode dépend d'un paramètre de lissage qui est le nombre de bandes (b). Celui-ci est exprimé ici en fraction de la plage de variation du FA. L'estimation du quantile conditionnel à une abscisse FA_i est fournie par le modèle de régression quantile construit sur les points dont les valeurs de FA sont situées dans l'intervalle $\left[FA_i - \frac{\Delta FA}{2b}\ ;\ FA_i + \frac{\Delta FA}{2b}\right]$. La valeur du paramètre de lissage impacte l'estimation du quantile glissant comme on le voit à la Figure 4-24 où un ajustement en fonction des DJU_{18} est étudié.

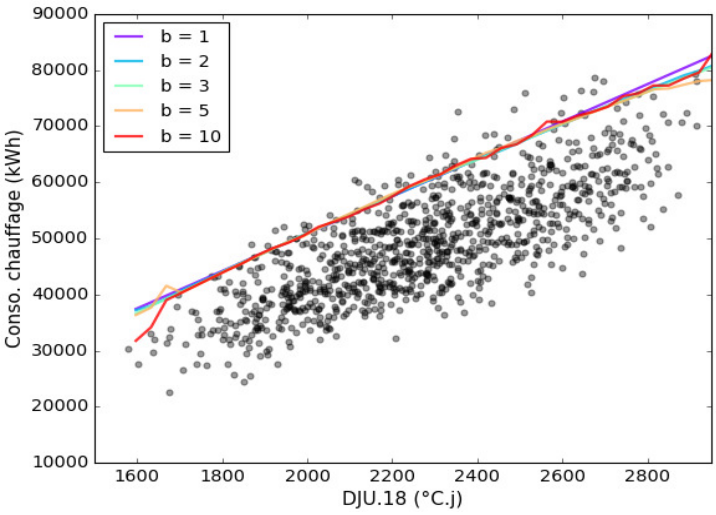

Figure 4-24 : Régression quantile glissante des besoins de chauffage pour la configuration d'ajustement A pour différentes largeurs de bande

Plus b est grand, plus la bande de valeurs glissantes est étroite et rend l'estimation instable et sensible aux irrégularités statistiques. Sur l'exemple ci-dessus, on constate visuellement la pertinence de l'hypothèse de linéarité pour l'ajustement en fonction des DJU_{18}.

La Figure 4-25 présente une estimation glissante du quantile conditionnel des besoins de climatisation en fonction des degré-heures de climatisation à base 24 °C (Configuration C). On y constate que cette dépendance n'est pas linéaire.

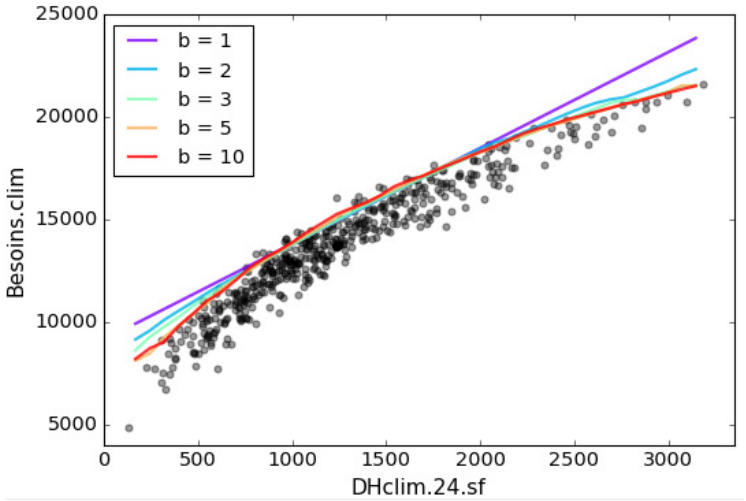

Figure 4-25 : Régression quantile glissante des besoins de climatisation pour la configuration d'ajustement C pour différentes largeurs de bande

La régression quantile glissante ne fournit pas de modèle explicite exploitable mais permet visuellement de constater des non-linéarités du quantile conditionnel en faibles dimensions. Ces non-linéarités peuvent être expliquées par plusieurs raisons :

- des raisons physiques, où des phénomènes non-linéaires ou d'interactions sont impliqués (telles que constatées dans cette étude) ;
- des régulations, venant introduire des changements de comportement ;
- des effets de seuil, liés au dimensionnement des systèmes telles que des puissances limitées…

D'autres méthodes peuvent permettre d'inclure une quantification et une description mathématique des non-linéarités.

4.4.3. Modèle polynomial

Le modèle polynomial de degré $q \geq 2$ est une alternative possible au modèle linéaire pour décrire le quantile conditionnel de la grandeur d'intérêt en fonction des FA. Cette forme, décrite dans l'équation ci-dessous, est paramétrique et permet de capturer les éventuelles dépendances entre FA. Plus l'ordre est élevé, plus le nombre de paramètres à estimer est grand, ce qui peut

entrainer l'identification d'estimateurs peu robustes. L'équation ci-dessous présente la forme générale d'un modèle polynomial multivarié d'ordre 2.

$$\hat{Y} = Y_0 + \sum_i \left(\beta_i x_i + \beta_{ii} {x_i}^2 + \sum_{j>i} \beta_{ij} x_{ij} \right) \quad \textbf{(4.20)}$$

La Figure 4-26 présente les résultats des régressions sur l'espérance et le quantile pour les modèles linéaires et polynomiaux des configurations A et B. Leurs coefficients sont affichés dans le Tableau 4-7.

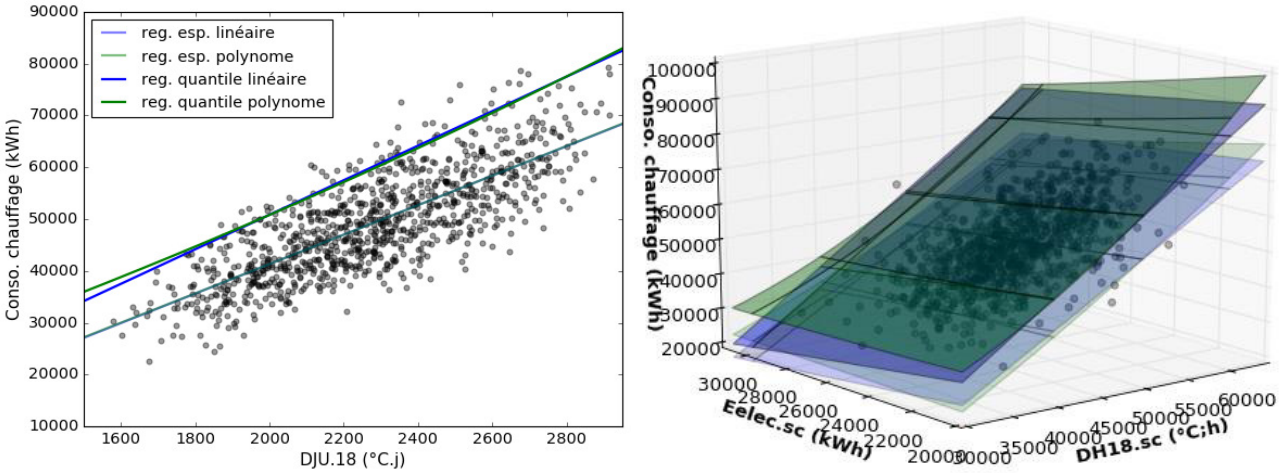

Figure 4-26 : Régressions sur l'espérance et le quantile des consommations de chauffage pour des modèles de degrés 1 (bleu) et 2 (vert) pour les configurations A (gauche) et B (droite)

La linéarité des modèles pour l'espérance et le quantile conditionnels est confirmée, les estimations polynomiales du second degré donnant des résultats très similaires aux modèles linéaires. L'étude des coefficients des modèles exprimés en fonction des taux d'accroissement (équation 4.10) des FA montre en effet des valeurs assez faibles. Par ailleurs la modification de 10 % d'un facteur d'ajustement FA_i par rapport à sa valeur de référence modifie la LGC_α d'un dixième de β_{FA_i} mais seulement d'un centième de $\beta_{(FA_i)^2}$, donnant moins de poids à ces éléments.

Tableau 4-7 : Coefficients des modèles de régression de l'espérance (RE) et de régression quantile (RQ) linéaire (L) ou polynomiale de degré 2 (poly.) pour les configurations A et B

Configuration	*Coefficients*	*REL*	*RE poly.*	*RQL*	*RQ poly.*
A	$\boldsymbol{\beta_{ref}}$	49098	49085	59947	60372
	$\boldsymbol{\beta\ DJU_{18}}$	64649	64642	75715	74125
	$\boldsymbol{\beta\ (DJU_{18})^2}$	-	890	-	-23665
B	$\boldsymbol{\beta_{ref}}$	48940	48724	59521	58751
	$\boldsymbol{\beta\ DH_{18.sc}}$	75243	75307	90041	90652
	$\boldsymbol{\beta\ (DH_{18.sc})^2}$	-	17795	-	2056
	$\boldsymbol{\beta\ E_{elec.sc}}$	-18682	-18954	-22878	-20130
	$\boldsymbol{\beta\ (E_{elec.sc})^2}$	-	-50987	-	-29200
	$\boldsymbol{\beta\ (DH_{18.sc})(E_{elec.sc})}$	-	-888	-	198954

La Figure 4-27 et le Tableau 4-8 présentent les résultats correspondants à la configuration C pour laquelle on constate un lien non-linéaire entre les besoins de climatisation et les $DH_{24.clim.sr}$. Cet effet a déjà été observé et discuté au chapitre 2 (Figure 2-14) où l'on montrait par ailleurs un lien linéaire entre les besoins de climatisation et la température extérieure moyenne sur la saison de refroidissement ($T_{m.ext.sr}$), un indicateur prenant en compte l'ensemble des variations de la température extérieure et non seulement lorsqu'elle dépasse une valeur seuil. En effet, du fait de l'effet inertiel du bâti et des autres apports de chaleur, les besoins de climatisation sont aussi liés à la durée et à l'amplitude des périodes plus fraiches, en particulier nocturnes, participant au refroidissement passif du bâtiment. Plus les $DH_{24.clim.sr}$ sont élevés, plus une augmentation supplémentaire de $DH_{24.clim.sr}$ est associée à une augmentation de $T_{m.ext.sr}$ faible et donc à une augmentation des besoins de climatisation limitée.

On constate sur le croisement des courbes des modèles polynomiaux de l'espérance et du quantile, pour de faibles valeurs du FA notamment, limitant la validité de l'ajustement sur ces extrémités (peu probables) des plages de variation. Il pourrait donc être pertinent d'augmenter l'ordre du modèle polynomial ou de définir un intervalle de validité.

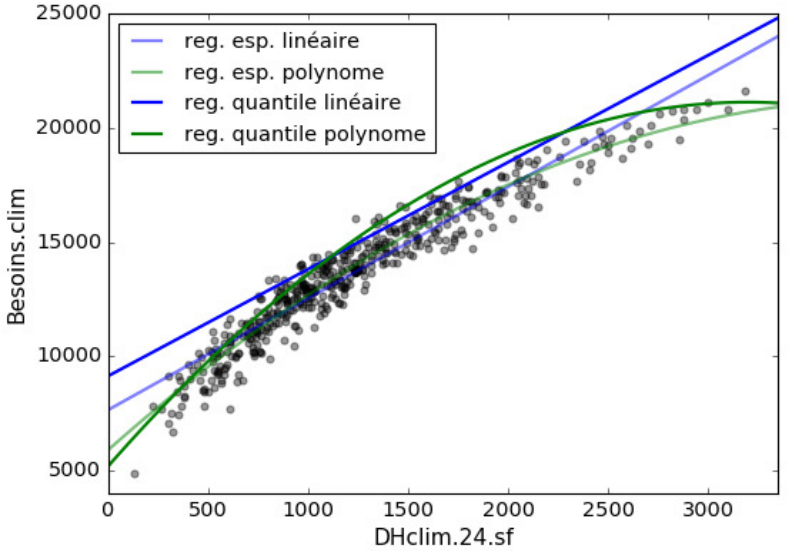

Figure 4-27 : Régressions sur l'espérance et le quantile des besoins de climatisation pour des modèles de degrés 1 (bleu) et 2 (vert) pour la configuration C

Tableau 4-8 : Coefficients des modèles de régression de l'espérance (RE) et de régression quantile (RQ) linéaire (L) ou polynomiale (poly.) pour la configuration C

Configuration	*Coefficients*	*REL*	*RE poly.*	*RQL*	*RQ poly.*
C	$\boldsymbol{\beta_{ref}}$	13178	13433	14431	14661
	$\boldsymbol{\beta\ DH_{24.clim.sr}}$	5522	6275	5293	6054
	$\boldsymbol{\beta\ (DH_{24.clim.sr})^2}$	-	-1256	-	-1309

La mise en œuvre d'une régression quantile polynomiale constitue une proposition originale. La régression polynomiale ($q \geq 2$) est pertinente pour modéliser les non-linéarités et les interdépendances entre FA pour des configurations d'ajustement qui le nécessiteraient. Bien qu'un polynôme soit plus complexe qu'une relation linéaire, cette forme paramétrique demeure transparente et compréhensible. Il est cependant plus difficile d'analyser l'influence des coefficients des termes de corrélation et de second ordre dans une démarche d'analyse de sensibilité.

4.4.4. Régression multivariée par spline adaptative

La régression multivariée par spline adaptative (MARS *: multi-variate adaptative regression splines*) est une méthode de régression non paramétrique développée par Friedman (1991). Elle constitue une extension de la régression linéaire multiple mais modélise aussi les interactions et les non-linéarités. La méthode MARS repose sur l'utilisation de fonctions linéaires dites fonctions charnières (*hinge-function*) qui sont de la forme suivante :

$$C(t) = max(0, t - \theta) \text{ (respectivement } C(t) = max(0, \theta - t)) \quad \textbf{(4.21)}$$

Ce sont donc des fonctions à une variable, nulles sur l'intervalle $]-\infty; \theta]$ (respectivement $[\theta; +\infty[$) et linéaires sur $[\theta; +\infty[$ (respectivement $]-\infty; \theta]$). Le modèle $\hat{Y}$ déterminé par la méthode MARS est de la forme :

$$\hat{Y} = \sum_{j=1}^{k} c_j B_j(X) \quad \textbf{(4.22)}$$

où les c_j sont des coefficients constants pondérant des fonctions B_j, qui sont soit des constantes, soit des fonctions linéaires charnières, soit des produits de fonctions linéaires charnières.

Le nombre de degrés de liberté est donc théoriquement infini. Le modèle est déterminé en deux phases distinctes, comme c'est le cas dans d'autres méthodes statistiques d'apprentissage : une phase « aller » (*forward pass*) et une phase « retour » (*backward pass*). Lors de la phase « aller », des fonctions de base sont ajoutées et testées afin de minimiser l'erreur quadratique du modèle. Des paires de fonctions à droite et à gauche des meilleures valeurs de pivot sont considérées. Lors de la phase « retour », des termes sont supprimés afin de simplifier le modèle et de réaliser un compromis entre complexité et précision. Le critère GCV (*Generalised cross-validation*) est utilisé pour cette opération, proportionnel à la somme quadratique des erreurs mais pénalisé par le nombre de paramètres.

La Figure 4-28 et le Tableau 4-9 présentent les résultats de régression par spline adaptative pour la configuration A fournissant l'espérance conditionnelle de la grandeur d'intérêt.

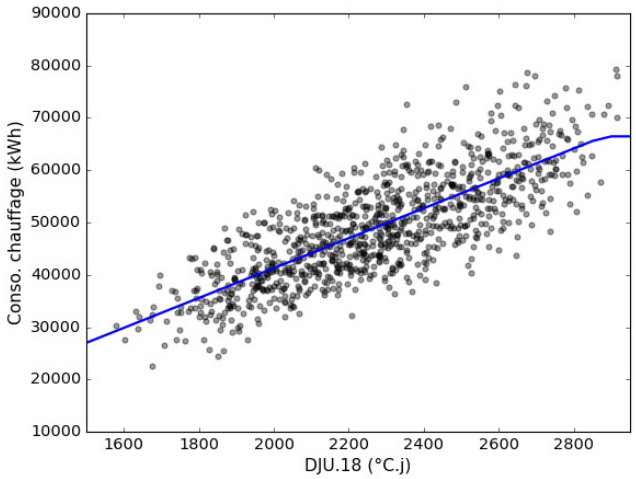

Figure 4-28 : Résultats de la méthode MARS pour l'estimation de l'espérance conditionnelle des besoins de chauffage sur la configuration A

Tableau 4-9 : Coefficients des modèles de régression sur l'espérance par la méthode MARS pour la configuration A

Configuration	*Termes*	*Coefficients MARS*
A	β_0	66456
	$\beta\ (2881 - DJU_{18})$	-28,555

Le modèle MARS identifie une rupture de pente (pente nulle) pour les valeurs élevées de DJU_{18}. Cependant, on constate que cette identification aux limites de la plage de variation repose sur peu de points ce qui limite la validité statistique de ce résultat qui ne repose a priori sur aucune logique physique. Pour la configuration B, la méthode MARS n'identifie aucune non-linéarité et fournit le modèle de régression sur l'espérance du Tableau 4-7. La solidité de l'hypothèse de linéarité entre les FA et les consommations de chauffage est de nouveau confirmée.

Les résultats de la méthode MARS pour la configuration C sont présentés à la Figure 4-29 et dans le Tableau 4-10. Trois branches linéaires sont identifiées dans le modèle décrivant l'espérance des besoins de climatisation en fonction des $DH_{24.sr}$. Les valeurs pivots et coefficients du modèle sont récapitulés dans le Tableau 4-10.

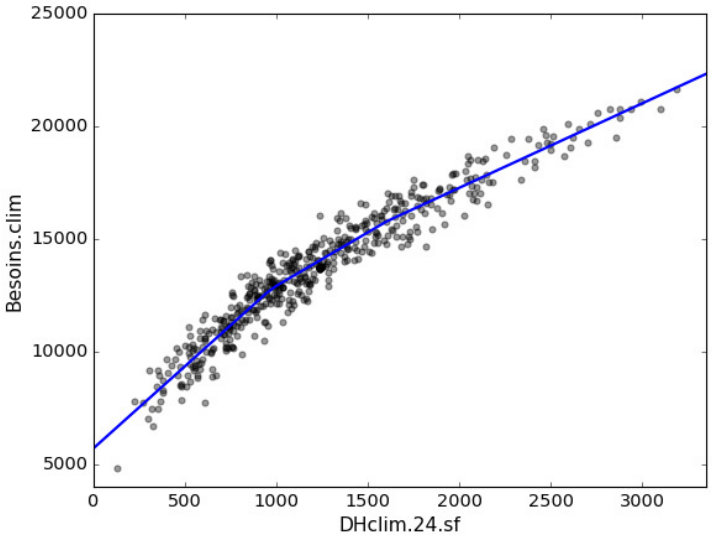

Figure 4-29 : Résultats de la méthode MARS pour l'estimation de l'espérance conditionnelle des besoins de climatisation pour la configuration C

Tableau 4-10 : Coefficients MARS des modèles de régression de l'espérance et du quantile conditionnels des besoins de climatisation pour différents jeux de facteurs d'ajustement

Configuration	***Termes***	***Coefficients MARS***
C	β_0	12790
	$\beta\ (DH_{24.sr} - 972{,}1)$	4,734
	$\beta\ (972{,}1 - DH_{24.sr})$	-7,273
	$\beta\ (DH_{24.sr} - 1605{,}9)$	-0,991

La méthode MARS est une alternative intéressante à la régression linéaire car elle prend en compte les dépendances et non-linéarités mais se compose d'une combinaison de fonctions linéaires simples. Une adaptation de l'algorithme afin de modéliser le quantile conditionnel, nécessaire à la méthodologie globale, pourrait être envisagée. Sinon, les valeurs pivot identifiées pour la régression sur l'espérance définissent des intervalles sur lesquels des modèles de régression quantile linéaire peuvent être calculés.

4.5. Synthèse

Dans ce chapitre, différentes questions ont été abordées de manière indépendante, venant compléter la méthodologie présentée au chapitre 3 en proposant des outils complémentaires afin de mieux maitriser sa mise en œuvre.

Le calcul de variance statistique de l'estimateur de régression correspondant à la valeur de référence de la limite garantie, $LGC_{\alpha-ref}$, à partir d'une méthode *bootstrap* permet l'intégration d'un risque supplémentaire lié à la représentativité statistique des données à partir desquelles est construit le modèle. La prise en compte de cette incertitude supplémentaire sur la $LGC_{\alpha r-ef}$ est nécessaire, celle-ci étant associée à un risque décentré. Les autres coefficients du modèle décrivent la pente d'ajustement autour d'un pivot constitué des valeurs de référence des facteurs d'ajustement. Bien que la variance statistique de ces estimateurs soit une donnée intéressante, l'incertitude est centrée et n'impacte donc pas la valeur des coefficients. Suite à cette étude, un nombre de simulations minimal de l'ordre de 1000 jeux de données est conseillé afin de limiter la variance statistique.

L'intégration d'erreurs de mesure virtuelles aux données de simulations dans le but d'anticiper l'impact des incertitudes de mesure lors de la vérification en exploitation a ensuite été étudiée. Cette variabilité supplémentaire, introduite par le bruitage des données, provoque d'une part l'augmentation de la LGC_{α} et d'autre part la diminution des valeurs absolues des coefficients d'ajustement. Le tirage de 5 à 10 jeux d'erreurs de mesure par jeu de données est conseillé afin de conserver une bonne représentativité statistique avant la détermination du modèle de régression.

Différentes méthodes ont été étudiées afin d'aider au choix du meilleur facteur d'ajustement pour une grandeur donnée et à la sélection du meilleur jeu de FA multidimensionnels. Un indicateur permet de quantifier la part de variabilité du quantile conditionnel expliquée par le modèle de régression quantile. Les coefficients du modèle standardisé ou du modèle avec taux d'accroissement des FA par rapport à leurs valeurs de référence caractérisent l'influence relative des FA. Des tests statistiques peuvent invalider l'intérêt d'ajouter des facteurs supplémentaires au modèle d'ajustement. Enfin des méthodes de régression quantile pénalisée ont été adaptées à l'évaluation du quantile conditionnel et permettent d'identifier les meilleurs FA. Ces méthodes fournissent un ensemble d'outils cohérent et complémentaires pour le choix des FA. Suite à l'étude de ces méthodes et afin de donner des recommandations pratiques, l'utilisation du coefficient de régression R^1 sur le modèle de régression quantile et l'exploitation de la régression pénalisée sont les méthodes à privilégier pour le choix des FA et l'évaluation de la qualité des modèles.

Enfin, l'hypothèse de linéarité peut ne pas être vérifiée suivant les projets étudiés, et la construction de modèle par régression quantile linéaire n'est donc pas toujours pertinente. L'identification de modèles intégrant les non-linéarités et les interactions permet ainsi d'obtenir des modèles au meilleur caractère explicatif. La régression polynomiale et la méthode MARS sont ainsi

des alternatives possibles. La validité de l'hypothèse de linéarité a par ailleurs été vérifiée sur notre cas d'étude traitant des consommations de chauffage.

Les éléments présentés dans ce chapitre s'intègrent donc à l'ensemble méthodologique pour la mise en œuvre d'un processus de GPE. La Figure 4-30 constitue une mise à jour de la Figure 3-9 présentée au chapitre précédent. Elle intègre les développements ci-dessus. Le processus complet va maintenant être testé de manière globale sur différents cas d'étude.

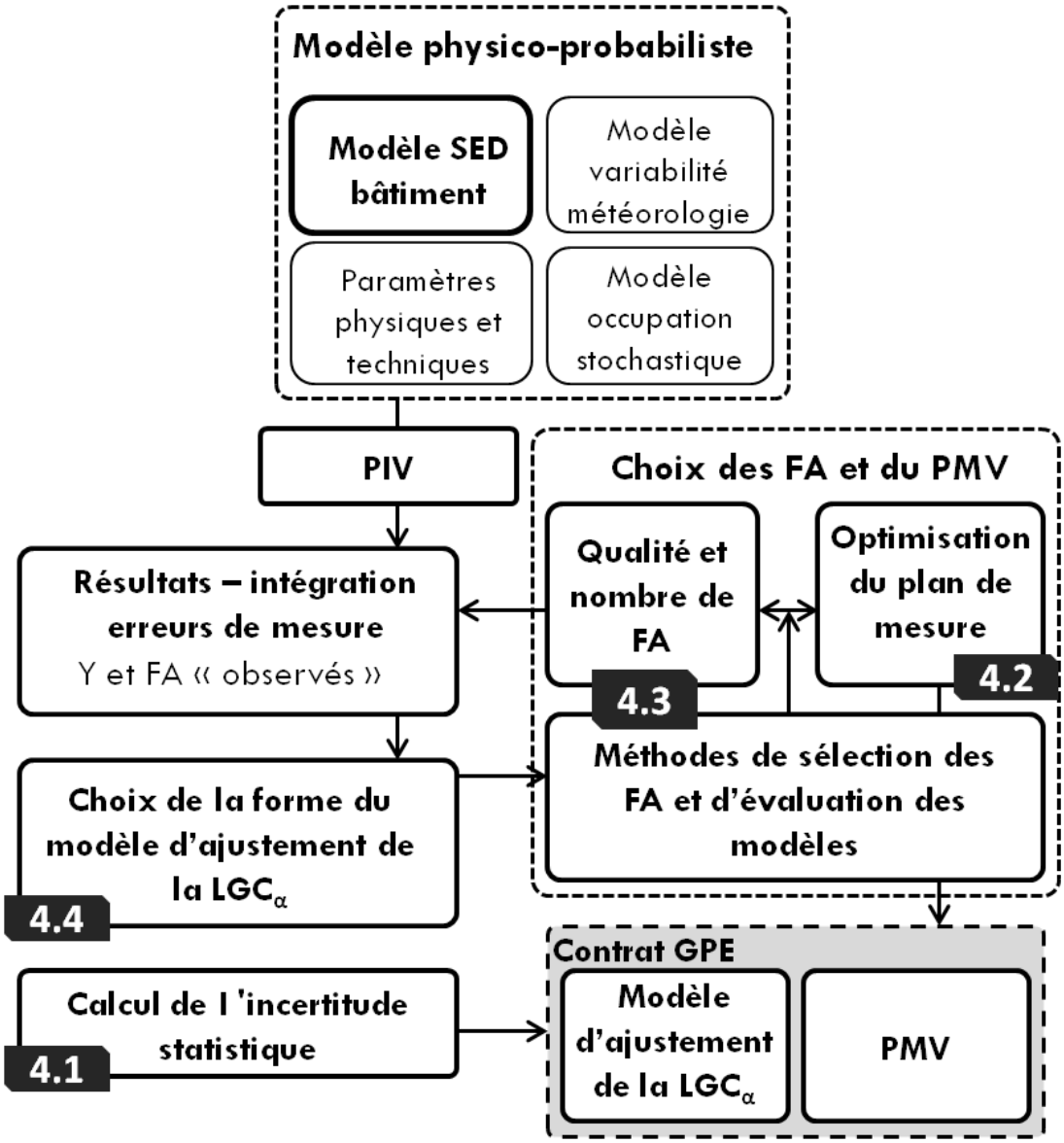

Figure 4-30 : Schéma d la méthodologie pour la GPE et intégration des études du chapitre 4

Chapitre 5. Application de la méthodologie à des projets de construction et de rénovation

Dans ce dernier chapitre, les éléments méthodologiques développés sont appliqués à deux cas d'étude : un bâtiment de logements existant subissant une opération de réhabilitation et un bâtiment de bureaux neuf à haute performance énergétique. Les différentes phases d'une étude sont conduites en prenant en compte les interactions entre ces processus, décrits de manière séquentielle dans les chapitres précédents. Les spécificités des méthodes pour un engagement sur des améliorations de performance énergétique et sur un niveau absolu de consommation sont étudiées. L'objectif est d'illustrer concrètement la mise en œuvre des travaux effectués dans ce manuscrit afin de tester l'applicabilité de la méthodologie.

Sommaire du chapitre 5

Chapitre 5. Application de la méthodologie à des projets de construction et de rénovation ... 197

5.1. GPE en réhabilitation : logements collectifs de FEYZIN ... 199
5.1.1. Objectifs ... 199
5.1.2. Présentation du cas d'étude ... 199
5.1.3. Modélisation et simulations ... 200
5.1.4. Configuration d'ajustement et plan de mesure ... 206
5.1.5. Modèles d'ajustement pour la garantie sur l'amélioration de performance ... 209
5.1.6. Mesures et vérification ... 217
5.1.7. Conclusion du cas d'étude de Feyzin ... 219

5.2. GPE en construction neuve : bâtiment de bureaux IZUBA ... 221
5.2.1. Présentation du cas d'étude ... 221
5.2.2. Modélisation ... 223
5.2.3. Propagation des incertitudes et variabilités ... 229
5.2.4. Dispositif de mesure ... 230
5.2.5. Engagement de GPE sur les consommations de chauffage et de climatisation ... 231
5.2.6. Engagement de GPE sur les consommations électriques totales ... 238
5.2.7. Traitement des données et vérification ... 240
5.2.8. Bilan du cas d'étude d'Izuba ... 246

5.1. GPE en réhabilitation : logements collectifs de FEYZIN

5.1.1. Objectifs

Cette première partie présente l'application des méthodes pour la GPE sur l'opération de réhabilitation du bâtiment de logements de Feyzin, intégrant l'ensemble des phases du projet, avec l'objectif de s'engager sur une diminution des consommations énergétiques de chauffage. Bien qu'envisagée la prise en compte des consommations d'ECS n'a pas été retenue du fait de l'absence de données de consommations mesurées fiables pour effectuer la vérification.

Précédemment dans ce document, ce cas d'étude a permis d'illustrer les différentes étapes du processus et les modèles développés, de manière cloisonnée. L'objectif est ici de mettre en œuvre la méthodologie de manière complète, réaliste et intégrée en illustrant l'interdépendance entre les étapes méthodologiques.

5.1.2. Présentation du cas d'étude

Le projet de rénovation du bâtiment de logements de Feyzin est présenté au chapitre 1 (paragraphe 1.6.2). C'est un bâtiment sur quatre niveaux composé de 16 appartements. La Figure 5-1 présente le plan du rez-de-chaussée et du premier étage et le Tableau 3-2 ses principales caractéristiques thermiques décrites dans le logiciel de SED *Pléiades*. Le détail des interventions de réhabilitation réalisées sont présentées au chapitre 1. Ce bâtiment a fait l'objet d'une étude d'audit puis d'un suivi énergétique sur une année d'exploitation après travaux (Enertech, 2013). D'autres informations sur ce bâtiment et sa modélisation sont disponibles dans le rapport du projet OMEGA (Mora et al., 2016) et dans la thèse de (Vorger, 2014).

Avant travaux, le chauffage est fourni par des planchers chauffants électriques et des postes d'appoints eux aussi fonctionnant à l'électricité. Suite aux rénovations, un réseau d'eau chaude provenant d'une sous-station présente au rez-de-chaussée du bâtiment irrigue les appartements et produit l'ECS et le chauffage. La sous-station est alimentée par une chaudière à gaz commune à plusieurs immeubles. On s'intéresse à l'énergie disponible en sortie de sous-station mesurable par un compteur dédié.

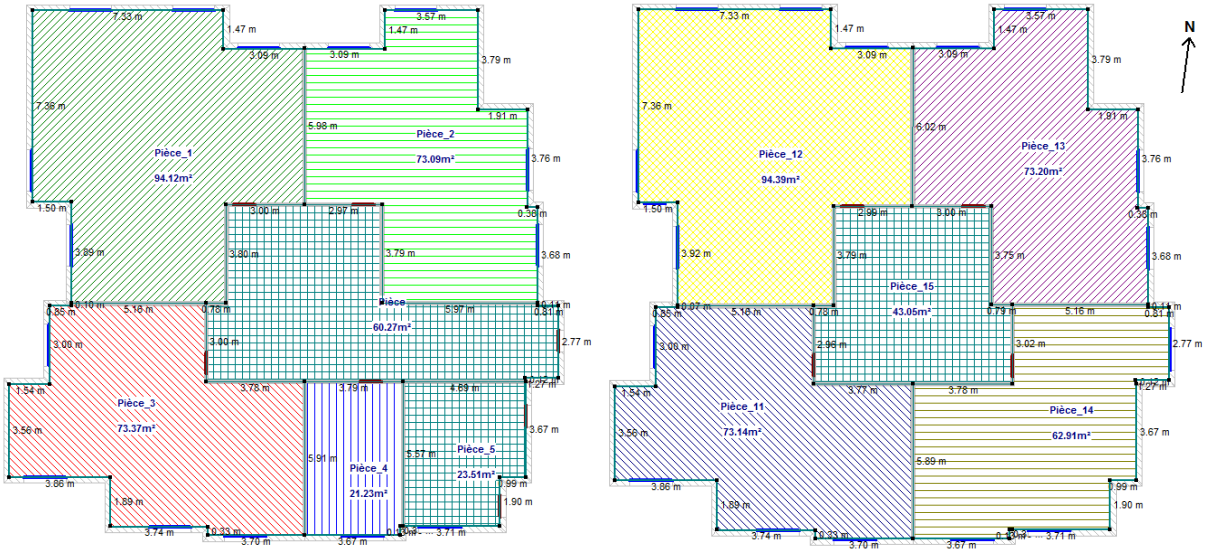

Figure 5-1 : Plans du rez-de-chaussée et d'un étage courant, Modeleur Pléiades

Tableau 5-1 : Principales caractéristiques thermiques du bâtiment rénové

Façades	$\mathbf{U = 0,22\ W/(m^2.K)}$
Plancher bas sur VS	$\mathbf{U = 0,49\ W/(m^2.K)}$
Toiture terrasse	$\mathbf{U = 0,29\ W/(m^2.K)}$
Menuiseries PVC 4/12/4	$\mathbf{U = 1,9\ W/(m^2.K)}$
Total ponts thermiques	$\mathbf{\sum \psi.L = 356\ W/K}$
Débit VMC hygro B	$\mathbf{V_{vmc} = 0,28\ vol/h}$
Débit infiltrations	$\mathbf{V_{inf} = 0,15\ vol/h}$
Taux vitrages (*N,S,E,O*)	$\mathbf{(19, 25, 14, 15)}$ %

5.1.3. Modélisation et simulations

La modélisation du système physico-probabiliste est effectuée en deux étapes successives dont la logique est décrite au chapitre 1. Cette méthode a pour objectif de définir au plus juste les incertitudes sur les paramètres du modèle en optimisant les efforts de modélisation.

Dans un premier temps, des plages d'incertitude correspondant à des lois uniformes sont définies pour chaque paramètre statique incertain dans le but de réaliser une analyse de sensibilité et d'identifier les paramètres les plus influents. Dans cette première étape, aucune variabilité sur les conditions météorologiques ou sur les usages n'est prise en compte, l'influence de ces éléments étant démontrée d'après la littérature comme rappelé au chapitre 1.

Dans un second temps, le modèle physico-probabiliste final est réalisé. Des fonctions de densité de probabilité plus précises sont définies pour les paramètres incertains influents mis en avant par l'analyse de sensibilité. Une recherche spécifique est menée afin de configurer au mieux ces lois de distribution. Les modèles de variabilité du comportement des occupants (Vorger, 2014) et des conditions météorologiques (chapitre 2) fournissent des données d'entrée temporelles stochastiques. Une propagation des incertitudes et variabilité est ensuite réalisée.

5.1.3.1. Analyses de sensibilité et d'incertitudes

Le Tableau 3-1 présente les plages d'incertitude initialement considérées pour les paramètres statiques du modèle du bâtiment (après réhabilitation). Le nombre de paramètres (avant – après réhabilitation) est précisé entre parenthèses. Les spécifications nominales du modèle de bâtiment sont décrites dans l'annexe A-5-1. Les débits d'infiltration et de ventilation sont considérés indépendamment pour chaque appartement. Le rendement de distribution de chauffage n'a pas été inclus dans cette analyse de sensibilité du fait de sa forte influence, qui masque le reste des résultats de l'analyse. Il est cependant considéré dans les paramètres incertains sélectionnés.

Tableau 5-2 : Bornes des lois de distribution uniformes considérées pour les paramètres incertains des modèles SED avant et après réhabilitation

Paramètres communs	***Borne inf.***	***Borne sup.***
Conductivité matériaux (4-8)	*- 5 %*	*+ 5 %*
Masse volumique matériaux (4-8)	*- 2 %*	*+ 2 %*
Chaleur spécifique matériaux (4-8)	*- 5 %*	*+ 5 %*
Épaisseur matériaux dans parois (10-12)	*- 10 %*	*+ 5 %*
Hauteur parois (4-5)	*- 1 %*	*+ 1 %*
Largeur parois (4-5)	*- 1 %*	*+ 1 %*
Facteur solaire vitrages (3)	*- 5 %*	*+ 5 %*
Conductivité vitrages (3)	*- 5 %*	*+ 5 %*
Hauteur vitrages (3)	*- 1 %*	*+ 1 %*
Largeur vitrages (3)	*- 1 %*	*+ 1 %*
Ponts thermiques linéiques (6-5)	*- 10 %*	*+ 10 %*
Ponts thermiques menuiseries (2)	*- 10 %*	*+ 10 %*
Infiltrations (17)	*- 20 %*	*+ 20 %*
Débits de ventilation (16)	*- 10 %*	*+ 10 %*

Les résultats de l'analyse de sensibilité réalisée avec la méthode de Morris sont décrits à la Figure 5-2. Un graphe des écart-types et des moyennes des valeurs absolues des effets élémentaires est présenté dans la partie supérieure, et un classement des distances à l'origine est disponible en

dessous. On trouve parmi les paramètres les plus influents : les ponts thermiques linéiques et des menuiseries, le coefficient U_w des vitrages majoritairement déployés sur le bâtiment, les épaisseurs des matériaux isolants des parois. Les écart-types des effets élémentaires sont assez faibles en comparaison de leur moyenne. Les effets non-linéaires ou d'interaction sont donc très faibles. Pour l'épaisseur de l'isolant (Mat2) du plancher, un écart-type de 55 kWh entre les effets élémentaires de 10 répétitions est cependant identifié. L'influence des échanges d'air est masquée du fait de l'indépendance entre les débits de ventilation et d'infiltration des seize appartements (équipés de bouches hygroréglables d'extraction), entrainant un foisonnement et donc une limitation de la variabilité.

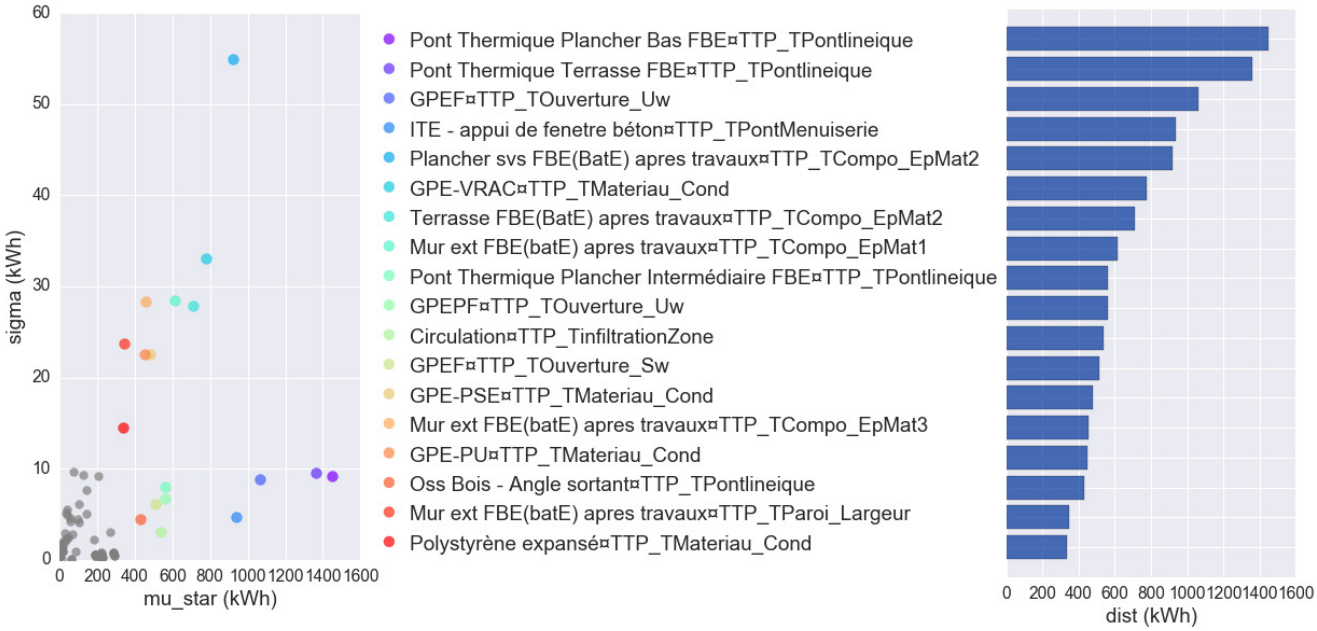

Figure 5-2 : Analyse de sensibilité aux paramètres du modèle SED après réhabilitation (N_r=10)

L'analyse de sensibilité permet de concentrer l'effort de modélisation des PDF sur les paramètres influents. Pour ce cas d'étude, qui date de plusieurs années et dont les détails sont connus à partir de sources déjà établies, il n'est pas possible d'obtenir des informations plus détaillées. Cependant, des lois de probabilité sont redéfinies suite à cette première étape, avec notamment l'utilisation de lois normales tronquées (NT) centrées ou décentrées. Le Tableau 5-3 récapitule les fonctions de densité de probabilité, les moyennes des lois correspondent aux valeurs nominales des paramètres.

Tableau 5-3 : Bornes des lois de distributions normales tronquées considérées sur les paramètres incertains des modèles SED avant et après rénovation

Paramètres communs	*Lois*	*Écart-type*	*Borne inf.*	*Borne sup.*
Conductivité matériaux (4-8)	*NT*	*2 %*	*- 5 %*	*+ 5 %*
Masse volumique matériaux (4-8)	*NT*	*1 %*	*- 2 %*	*+ 2 %*
Chaleur spécifique matériaux (4-8)	*NT*	*2 %*	*- 5 %*	*+ 5 %*
Épaisseur matériaux dans parois (10-12)	*NT*	*5 %*	*- 10 %*	*+ 5 %*
Hauteur parois (4-5)	*NT*	*1 %*	*- 2 %*	*+ 2 %*
Largeur parois (4-5)	*NT*	*1 %*	*- 2 %*	*+ 2 %*
Facteur solaire vitrages (3)	*NT*	*2 %*	*- 5 %*	*+ 5 %*
Conductivité vitrages (3)	*NT*	*2 %*	*- 5 %*	*+ 5 %*
Hauteur vitrages (3)	*NT*	*1 %*	*- 2 %*	*+ 2 %*
Largeur vitrages (3)	*NT*	*1 %*	*- 2 %*	*+ 2 %*
Ponts linéiques (6-5)	*NT*	*5 %*	*- 10 %*	*+ 10 %*
Ponts menuiseries (2)	*NT*	*5 %*	*- 10 %*	*+ 10 %*
Infiltrations (17)	*NT*	*10 %*	*- 20 %*	*+ 20 %*
Débits de ventilation (16)	*NT*	*5 %*	*- 10 %*	*+ 10 %*
Rendement distribution chauffage (1)	*NT*	*10 %*	*-20 %*	*+ 15 %*

5.1.3.2. Propagation des variabilités

Le processus de modélisation repose ensuite sur la paramétrisation et l'utilisation de modèles de génération stochastique de sollicitations dynamiques. Ces modèles fournissent des données météorologiques et d'usages variables et réalistes. Le modèle de génération de données météorologiques présenté au chapitre 2 est utilisé à partir du fichier météo type de Mâcon, référence réglementaire pour la zone ou se trouve la commune de Feyzin (69) sur laquelle est implanté le bâtiment. Une variabilité des indicateurs tendanciels est aussi prise en compte à partir des résultats de l'étude présentée au chapitre 2. Pour les données d'usage, le modèle de génération stochastique de (Vorger, 2014) est utilisé. Le nombre d'occupants de chaque appartement (Tableau 1-3) et leur statut de locataire sont renseignés, ainsi que l'usage des pièces.

La Figure 5-11 présente les distributions des consommations de chauffage associées à quatre propagations d'incertitudes et variabilités (PIV). En gris, seules les incertitudes sur les paramètres statiques, tirées dans les fonctions de densité de probabilité du Tableau 5-3, sont propagées. La variabilité météorologique (rouge) et des scénarios liés aux usages (vert) entrainent un élargissement de la distribution. La variabilité météorologique semble avoir plus d'impact, mais cela s'explique par l'effet de foisonnement de la variabilité des usages sur les seize appartements, qui limite des situations globales extrêmes. En bleu l'ensemble des incertitudes et variabilités est considéré. On

constate que les distributions verte et bleue d'une part et grise et rouge d'autre part sont centrées sur des moyennes différentes. Ceci s'explique par les scénarios d'usage déterministes considérés dans ces deux dernières PIV. Ces scénarios sont basés sur la situation avant travaux et sont différents de l'effet moyen produit par le modèle de variabilité des usages (distributions bleue et verte). La PIV complète de la situation après réhabilitation fournit les résultats exploités pour construire les modèles de consommation énergétique garantie.

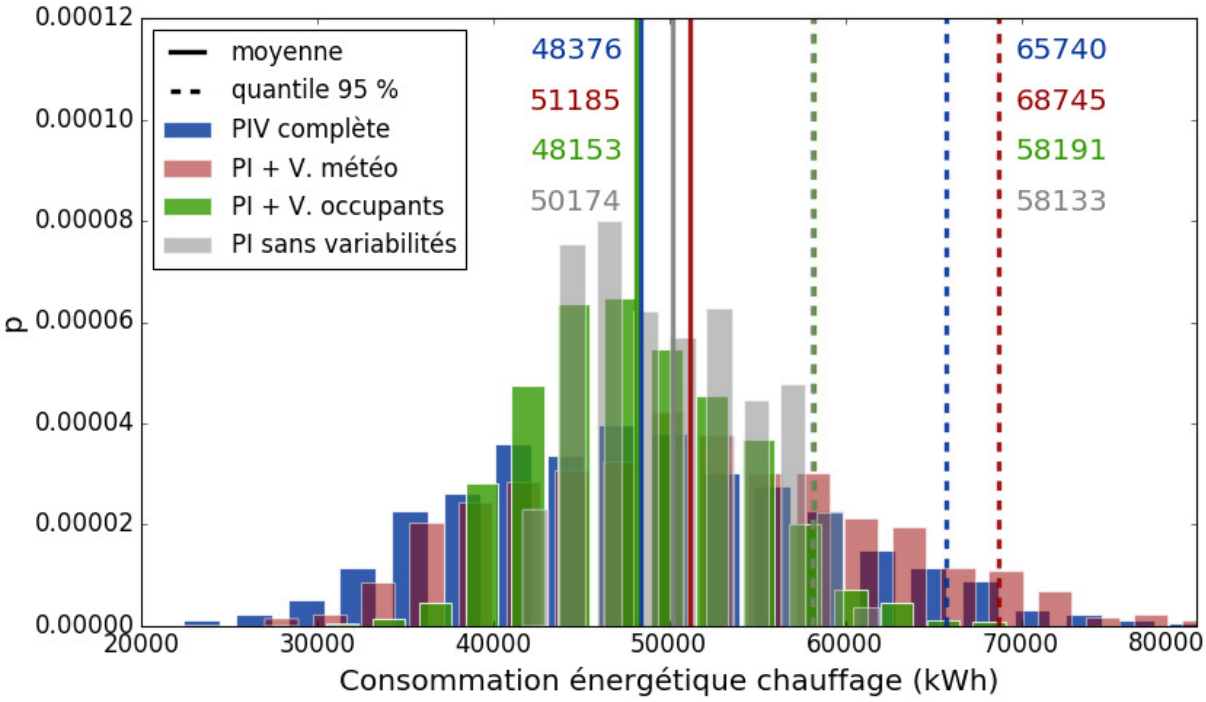

Figure 5-3 : Propagations des incertitudes et variabilités avec les modèles de génération stochastique de sollicitations météorologiques et d'usage

5.1.3.3. Situation de référence avant réhabilitation

Dans ce cas d'étude, il s'agit de s'engager sur la diminution des consommations d'énergie de chauffage. Il est nécessaire de définir un niveau de consommation de référence avant les travaux, éventuellement ajustable en fonction de FA (§ 3.3.5.1). Pour cela, une PIV complète est aussi réalisée sur le modèle de simulation avant travaux. Les détails de cette modélisation sont présentés en annexe A-5-1. La Figure 5-4 présente les distributions des consommations de chauffage avant et après réhabilitation qui seront exploitées dans la suite de ce cas d'étude.

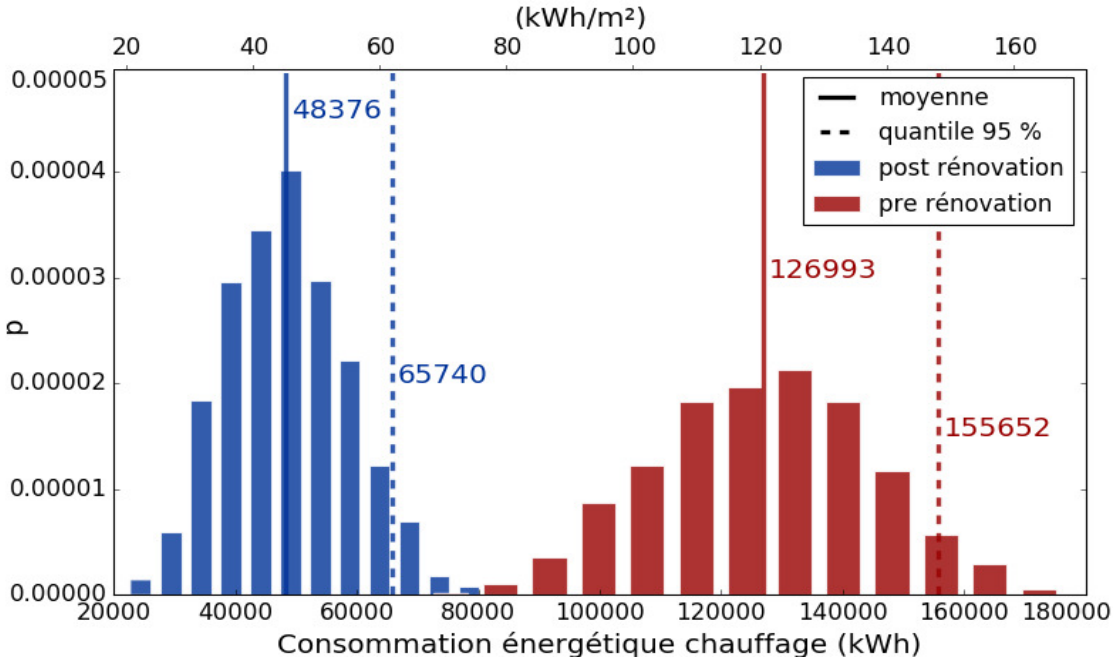

Figure 5-4 : Distributions des consommations de chauffage des propagations des incertitudes et variabilités des modèles SED pre- et post- réhabilitation

5.1.3.4. Synthèse des phases de modélisation et simulation

Cette première partie de l'étude se concentre sur la spécification des entrées des modèles SED, afin de simuler le fonctionnement du bâtiment dans un ensemble de conditions probables statistiquement représentatives. Cette variabilité de conditions concerne l'état du bâtiment lui-même, de son enveloppe et de l'efficacité de ces systèmes, et les sollicitations qu'il subit.

La réelle complexité de cette phase initiale réside dans les recherches nécessaires à la définition poussée des incertitudes des paramètres statiques influents. Une étude de la littérature et des retours d'expérience spécifique à chaque problématique doit être menée. Des recommandations sont fournies par (Mora et al., 2016 ; Titikpina, 2016).

Concernant la simulation, des propagations ont été réalisées sur des ensembles de 1000 simulations énergétiques dynamiques. Une parallélisation (8 *threads*) des calculs de la *BlackBox* (exécutable du modèle SED sans interface) du logiciel SED *Pléiades+COMFIE* est effectuée, orchestrée par des scripts Python du logiciel *Amapola*. Un prétraitement est associé aux modèles de variabilité météorologique et de comportements des occupants. Le Tableau 5-4 récapitule les temps de calcul moyens observés rapportés à une simulation. Le temps unitaire est d'environ 25 secondes, et la réalisation de 1000 simulations prend ainsi environ 7 h.

Les temps de calcul associés à ce type d'étude semblent donc raisonnables. Ils peuvent notamment être effectués de nuit. Par ailleurs, des solutions de parallélisation des simulations sur le

Cloud se développent, permettant de réaliser un nombre quelconque de simulation dans un temps équivalent à une simulation unitaire.

Tableau 5-4 : Temps de calcul rapportés à une simulation pour une PIV complète

Étapes	Temps
Génération des données météorologiques	~ 3,5 s
Génération des données d'occupation	~ 3 s
SED et post-traitements	~ 15 s
Orchestration	~ 4 s

Cette première phase de la méthodologie est coûteuse en temps de modélisation et en temps de calcul. Ce dernier point n'est plus réellement un problème pour la suite, les méthodes mises en œuvre étant moins lourdes mathématiquement.

5.1.4. Configuration d'ajustement et plan de mesure

La détermination d'une diminution garantie des consommations de chauffage, vérifiable en exploitation nécessite la détermination :

- de facteurs d'ajustement et d'un plan de mesure des FA et des consommations énergétiques ;
- d'un modèle de calcul des consommations d'énergie de référence avant réhabilitation, fonction des FA, permettant de calculer un gain relatif ;
- d'un modèle de la limite garantie de consommations après réhabilitation pour un risque de défaut α.

On souhaite déterminer les modalités d'ajustement du niveau d'engagement en fonction des conditions d'exploitation. Cette question, traitée en plusieurs étapes aux chapitres 3 et 4, interconnecte en réalité les problématiques liées au choix des FA, au choix du plan de mesure et vérification (PMV) et au calcul de l'incertitude statistique des modèles identifiés. C'est donc un problème d'optimisation multicritère qui, hors des aspects statistiques, dépend aussi en réalité de considérations techniques et économiques.

On s'intéresse ici au calcul des modèles de $LGC_{5\%}$ après travaux et à la détermination conjointe des FA et des PMV les plus performants. Un ensemble de grandeurs dynamiques mesurables est sélectionné dans le Tableau 4-2 avec pour chacune deux niveaux d'incertitudes de mesure correspondant à la mise en place de moyens d'acquisitions de données plus ou moins performants. Le niveau *PMV-B* correspond à une incertitude de mesure élevée et le niveau *PMV-A* à une plus grande précision. L'objectif de l'étude est d'identifier les meilleurs FA mais aussi les actions

prioritaires à mettre en place, entre l'ajout de FA supplémentaires et l'amélioration du plan de mesure.

Tableau 5-5 : Incertitudes de mesure des grandeurs d'ajustement et de la consommation de chauffage pour deux plans de mesure différents

Incertitude-type	*PMV-B*	*PMV-A*
T_{ext}	0,5 °C	0,2 °C
T_{int}	0,5 °C	0,125 °C[10]
T_{cons}	0,5 °C	0,125 °C
E_{elec}	2 %	1 %
$E_{chauffage}$	5 %	2 %

Les incertitudes de mesure sur les grandeurs de température et de consommations énergétiques du Tableau 4-2 permettent de calculer des plages d'incertitude sur les FA, variables agglomérées calculées à partir des grandeurs dynamiques (Tableau 5-6). Pour les DH_{Tint}, quatre niveaux d'incertitudes sont donc envisageables, ce FA dépendant de la mesure de deux grandeurs. N_j et N_h désignent le nombre de fois ou les compteurs de degré-jours et de degré-heures sont incrémentés.

Tableau 5-6 : Incertitudes de mesure induites sur les facteurs pour deux plans de mesure différents

Incertitude-type (°C)	*PMV-B*	*PMV-A*
DJU_{18}	0,5 N_j	0,2 N_j
DH_{18}	0,5 N_h	0,125 N_h
$T_{mp.int}$	0,5	0,125
$T_{mp.cons}$	0,5	0,125
$T_{m.ext}$	0,5	0,125
DH_{Tint}	0,71 N_h	0,18 N_h
	BA (T_{ext} : B et T_{int}: A) : 0,52 N_h	
	AB (T_{ext} : A et T_{int}: B) : 0,54 N_h	

Des erreurs de mesure virtuelle des FA et consommations de chauffage sont tirées aléatoirement dans des lois normales tronquées $N(0, IT, -2.IT, 2.IT)$ et sont ajoutées aux données issues des simulations après travaux (1000 simulations), où IT désigne l'incertitude-type. Un facteur multiplicatif $p = 10$ (§ 4.2) est pris en compte, signifiant que pour un jeu de données (FA, Y), dix jeux

[10] Les valeurs de température affichées au 1/100 ou 1/1000 °C sont uniquement exploitées de manière calculatoire

d'erreurs de mesure sont générés produisant finalement dix n-uplets (FA', Y') perturbés. L'ensemble statistique contient donc finalement 10 000 jeux de données pour l'identification de modèles.

Des modèles de $LGC_{5\%}$ sont déterminés par régression quantile pour les différentes configurations de FA et de PMV. Les coefficients de régression quantile R^1 sont étudiés afin de comparer les performances des modèles (§ 4.3.3).

L'ensemble des résultats de cette comparaison sont présentés dans l'annexe A-5-2. De manière générale, on constate que le jeu de FA choisi a plus d'importance que les niveaux d'incertitude de mesure pour les meilleurs modèles. On observe ensuite des cas de croisements entre variations des FA et variations des niveaux d'incertitudes de mesure dans le reste du tableau. On constate la prédominance de $T_{m.ext.sc}$ par rapport à $DH_{18.sc}$ et des couples $T_{m.ext.sc}/T_{mp.int.sc}$ et dans une moindre mesure $DH_{18.sc}/T_{mp.int.sc}$ par rapport à DH_{Tint}. Les résultats de R^1 étant très proches, la différence se faisant parfois à 10^{-4} près, des résultats non logiques sont parfois visibles, des modèles équivalents avec des FA plus incertains étant mieux classés que d'autres avec des FA sur lesquels il y a moins d'incertitude. Bien qu'on ne retrouve pas cette configuration pour l'ensemble des jeux de FA, $T_{mp.cons.sc}$, $T_{m.ext.sc}$ et $E_{elec.sc}$ sont les FA pour lequel on observe le plus rapidement un basculement du niveau d'incertitude de *PMV-A* à *PMV-B*, c'est-à-dire que le fait de moins bien les mesurer impacte moins la qualité du modèle identifié. La mesure précise des consommations d'énergie est aussi mise en avant, une dégradation de celle-ci entrainant des valeurs de R^1 significativement plus faibles pour une configuration donnée.

Le Tableau 5-7 regroupe les solutions pour les niveaux les plus performants d'incertitude et permet donc de comparer les configurations de FA. Le meilleur modèle dépend de $T_{m.ext.sc}$, $T_{mp.int.sc}$, $T_{mp.cons.sc}$ et $E_{elec.sc}$. Il a été montré au chapitre 4 que $T_{mp.int.sc}$ est fortement corrélé à $T_{m.ext.sc}$ et $T_{mp.cons.sc}$ mais l'utilisation combinée de ces FA permet de mieux expliquer le quantile conditionnel de la distribution des consommations de chauffage. En particulier, $T_{mp.int.sc}$ qui constitue seul un très mauvais facteur d'ajustement fournit une information équivalente à $T_{mp.cons.sc}$ lorsqu'il est associé à $T_{m.ext.sc}$.

Tableau 5-7 : Classement des modèles de régression quantile en fonction de R^1

DH_{Tint}	$DH_{18.sc}$	$E_{elec.sc}$	$T_{mp.int.sc}$	$T_{mp.cons.sc}$	$T_{m.ext.sc}$	R^1
						0,592
						0,591
						0,579
						0,576
						0,559
						0,552
						0,551
						0,548
						0,545
						0,534
						0,521
						0,509
						0,500
						0,497
						0,472
						0,459
						0,439
						0,437
						0,417
						0,059
						0,024
						0,000

Pour la suite du cas d'étude, le meilleur modèle qui repose sur un ajustement de la $LGC_{5\%}$ en fonction de $T_{m.ext.sc}/T_{mp.int.sc}/T_{mp.cons.sc}/E_{elec.sc}$ sera étudié. On considère les meilleurs niveaux d'incertitude, décrits dans le *PMV-A*.

5.1.5. Modèles d'ajustement pour la garantie sur l'amélioration de performance

5.1.5.1. Modèle de référence avant travaux

- *Régression linéaire multiple sur l'espérance des consommations de chauffage*

Afin de calculer une diminution des consommations d'énergie, il est nécessaire d'établir le modèle de la consommation d'énergie de référence avant réhabilitation en fonction des facteurs d'ajustement (§ 3.3.5.1). On s'intéresse à l'espérance de ces consommations, obtenue par régression linéaire sur l'espérance (moindres carrés). L'équation ci-dessous est obtenue par exploitation des résultats de la propagation des incertitudes et variabilités avant travaux.

$$C_{pre} = 1{,}272.10^5 - 1{,}486.10^4 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right) + 8{,}253.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right) + 6{,}120.10^3 * \left(T_{mp.cons.sc} - T_{mp.cons.sc-ref}\right) - 2{,}005.10^4 * \left(\frac{E_{elec}}{E_{elec.sc-ref}} - 1\right) \quad \textbf{(5.1)}$$

avec
$T_{m.ext.sc-ref} = 8{,}34\ °C$, $T_{mp.int.sc-ref} = 20{,}84\ °C$, $T_{mp.cons.sc-ref} = 20{,}62\ °C\ et\ E_{elec.sc-ref} = 25550\ kWh$.

Les valeurs de référence des FA correspondent aux médianes des distributions issues des simulations et seront conservées pour les modèles après réhabilitation. Pour les raisons de manque de signification physique d'une division par une température, décrites plus en détails précédemment (§4.3.4), les FA qui sont des températures moyennes sont exprimés en variation absolue par rapport à la valeur de référence et les autres FA en variation relative.

- *Calage du modèle*

Afin de déterminer un modèle de référence le plus conforme possible à la réalité avant réhabilitation, un processus de calibrage, ou calage de modèle peut être mis en œuvre si des données de mesure avant travaux sont disponibles.

L'entreprise Enertech a réalisé une campagne de mesures avant la réhabilitation du bâtiment afin de l'auditer. Sur la période allant du 02/11/2009 au 01/11/2010, avant les travaux effectués en 2011, les consommations de chauffage sont connues ainsi que la température intérieure dans une partie des appartements (dix logements sur seize). On fait l'hypothèse que la moyenne des températures relevées sur ces appartements pendant la saison de chauffe est représentative de la moyenne globale des logements. Les consommations d'électricité spécifiques ne sont pas connues. On fait l'hypothèse qu'elles sont identiques à celles mesurées après rénovation. Les données météorologiques *Météo-France* de la station de Feyzin sont exploitées pour calculer la température moyenne sur la saison de chauffe.

Pour pouvoir caler le modèle l'estimation d'une température de consigne moyenne sur la saison de chauffe est nécessaire. Les valeurs de $T_{mp.int.sc}$, $T_{m.ext.sc}$ et $T_{mp.cons.sc}$ sont corrélées et l'on propose d'établir un lien entre ces trois grandeurs en exploitant les résultats de simulation avant rénovation et d'en déduire une valeur probable de $T_{mp.cons.sc}$.

$$T_{mp.cons.sc} = -0{,}5327 + 1{,}0439 * T_{mp.int.sc} - 0{,}0723 * T_{m.ext.sc} \quad \textbf{(5.2)}$$

avec $R^2 = 0{,}98$ et $\sigma(T_{mp.cons.sc}) = 0{,}47$ °C d'où, par définition de R^2, $\sigma_{résidus} = 0{,}07$ °C. L'équation de régression ci-dessus permet d'estimer une température $T_{mp.cons.sc}$ de 20,23 °C pour cette période. Le Tableau 5-8 présente les valeurs des FA représentatives des conditions de la période d'étude avant réhabilitation.

Tableau 5-8 : Mesures avant la réhabilitation

Grandeurs d'intérêt	***valeurs***
Consommation énergétique de chauffage (kWh)	*112 031*
$T_{mp.int.sc}$ *(°C)*	*20,4*
$T_{m.ext.sc}$ *(°C)*	*7,36*
$E_{elec.sc}$ *(kWh)*	*19 681*
$T_{mp.cons.sc}$ *(°C)*	*20,23*

En ajustant le modèle C_{pre} avec les données du tableau ci-dessus, on obtient une consommation de 135 745 kWh soit une surestimation de 23 714 par rapport à la consommation réelle de 112 031 kWh. Cette différence peut notamment s'expliquer par des limitations des hypothèses sur l'estimation des FA ($T_{mp.cons.sc}$ et $E_{elec.sc}$), par de possibles erreurs de mesure et d'estimation sur les consommations, et par une modélisation trop défavorable des caractéristiques du bâtiment (par exemple les infiltrations ajoutées au débit d'air extrait, des ponts thermiques, etc).

On met à jour le modèle de référence en le calant à partir de l'exploitation de ces données. Le nouveau modèle est le suivant :

$$\begin{aligned} C_{pre} = {} & 8{,}850.10^4 - 1{,}486.10^4 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right) \\ & +8{,}253.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right) \\ & +6{,}120.10^3 * \left(T_{mp.cons.sc} - T_{mp.cons.sc-ref}\right) \\ & -2{,}005.10^4 * \left(\frac{E_{elec}}{E_{elec.sc-ref}} - 1\right) \end{aligned} \quad \textbf{(5.3)}$$

Ce calage du coefficient constant (ordonnée à l'origine) du modèle des consommations de référence avant réhabilitation n'avait pas été réalisé au chapitre 3 entrainant le calcul de gains plus élevés. Comme signalé au §3.3.5.1, ce modèle n'impacte pas le risque lié à l'engagement sur la performance énergétique suite à la rénovation, c'est une référence non vérifiable *a posteriori*. Les diminutions de consommation d'énergie et les gains relatifs garantis seront par la suite calculés à partir du modèle de l'équation (5.3).

5.1.5.2. Régression quantile linéaire et polynomiale pour la $LGC_{5\%}$

Afin d'établir un niveau de performance garanti, la limite garantie de consommations à 5 % de risque est calculée par régression quantile sur les résultats de simulation après réhabilitation. L'équation ci-dessous présente le modèle linéaire obtenu.

$$LGC_{5\%} = 5{,}701.10^4 - 9{,}472.10^3 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right) \quad \textbf{(5.4)}$$

$$+6{,}670.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)$$

$$+2{,}354.10^3 * \left(T_{mp.cons.sc} - T_{mp.cons.sc-ref}\right)$$

$$-2{,}967.10^4 * \left(\frac{E_{elec.sc}}{E_{elec.sc-ref}} - 1\right)$$

On constate l'influence conjointe de $T_{mp.cons.sc}$ et de $T_{mp.int.sc}$ même si le coefficient associé à ce dernier FA est près de trois fois plus élevé. Pour une même variation absolue par rapport à la valeur de référence $T_{m.ext.sc}$ a la plus forte influence. L'utilisation de FA fortement corrélés pose la question de la pertinence d'un modèle polynomial d'ordre plus élevé faisant apparaitre des termes corrélatifs couplant plusieurs FA (§4.4.3).

Le Tableau 5-9 présente les résultats de régression quantile pour le modèle linéaire dans la première colonne puis le modèle polynomial d'ordre 2 complet dans la 2ème colonne, et un modèle polynomial réduit aux termes les plus influents dans la 3ème colonne. Les FA principaux sont renommés de manière simplifiée dans les premières lignes du tableau pour plus de clarté. Les coefficients correspondent au changement de variable « relatif », ou « taux d'accroissement », $\frac{FA-FA_{ref}}{FA_{ref}}$ (même pour les températures) afin de pouvoir comparer plus aisément les différents coefficients.

Tableau 5-9 : Coefficients des modèles de régression quantile linéaires (RQL), polynomial et polynomial réduit (RQP)

	RQL	***RQ Poly.***	***RQ Poly. réduit***
R^1	**0,592**	**0,596**	**0,595**
$LGC_{5\%-ref}$	$5{,}701.10^4$	$5{,}65.10^4$	$5{,}655.10^4$
$\beta\ x0_E_{elec.sc}$	$-2{,}967.10^4$	$-2{,}734.10^4$	$-2{,}792.10^4$
$\beta\ x1_T_{mp.int.sc}$	$1{,}390.10^5$	$1{,}294.10^5$	$1{,}136.10^5$
$\beta\ x2_T_{m.ext.sc}$	$-7{,}900.10^4$	$-8{,}089.10^4$	$-7{,}873.10^4$
$\beta\ x3_T_{mp.cons.sc}$	$4{,}853.10^4$	$6{,}189.10^4$	$7{,}373.10^4$

$\boldsymbol{\beta\ x0^2}$	-	-2,635.10^4	-
$\boldsymbol{\beta\ x1^2}$	-	-2,826.10^6	-4,83.10^5
$\boldsymbol{\beta\ x2^2}$	-	-3,795.10^4	-
$\boldsymbol{\beta\ x3^2}$	-	-7,915.10^5	2,8.10^5
$\boldsymbol{\beta\ x0.x1}$	-	5,557.10^4	-
$\boldsymbol{\beta\ x0.x2}$	-	1,054.10^4	-
$\boldsymbol{\beta\ x0.x3}$	-	1,492.10^5	-
$\boldsymbol{\beta\ x1.x2}$	-	9,32.10^5	3,60.10^5
$\boldsymbol{\beta\ x1.x3}$	-	3,546.10^6	3,55.10^5
$\boldsymbol{\beta\ x2.x3}$	-	-6,67.10^5	-2,63.10^5

La comparaison des coefficients entre les termes d'ordre 1 et 2 doit être prudente. En effet, une variation de dix pour cent de chaque FA par rapport à leur valeur de référence entraine une variation d'un dixième de la valeur des coefficients d'ordre 1 mais seulement d'un centième pour les coefficients d'ordre 2. La Figure 5-5 présente les coefficients d'ordre 2 du modèle de régression quantile polynomial. On constate l'importance des termes $x1x3$, $x1^2$ et dans une moindre mesure $x1x2$, $x3^2$ et $x2x3$, où $x1$ désigne le FA $T_{mp.int.sc}$, $x2$ le FA $T_{m.ext.sc}$ et $x3$ $T_{mp.cons.sc}$.

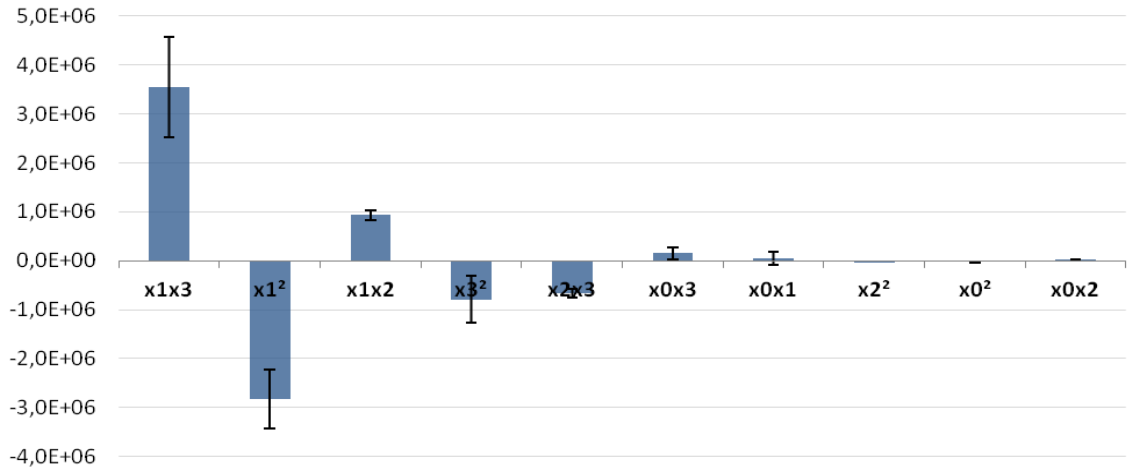

Figure 5-5 : Comparaison des coefficients d'ordre deux du modèle polynomial identifiés par régression quantile linéaire

Ces éléments sont confirmés par l'étude de régression pénalisée (*LASSO*) menée sur le modèle polynomial. La Figure 5-6 décrit l'évolution des coefficients de régression en fonction de la pénalisation λ pour des valeurs standardisées des entrées ($FA_i, FA_i{}^2$ et FA_iFA_j). En étudiant à la fois leurs valeurs initiales (pour $\lambda = 0$) et leur évolution (leur élimination pour des λ croissant), on constate l'importance des coefficients d'ordre 1 par rapport à ceux d'ordre 2 dans le modèle. Les

coefficients des variables $x1x3$ et $x1^2$ présentent des valeurs normalisées non négligeables bien que rapidement éliminées (zoom à droite de la figure).

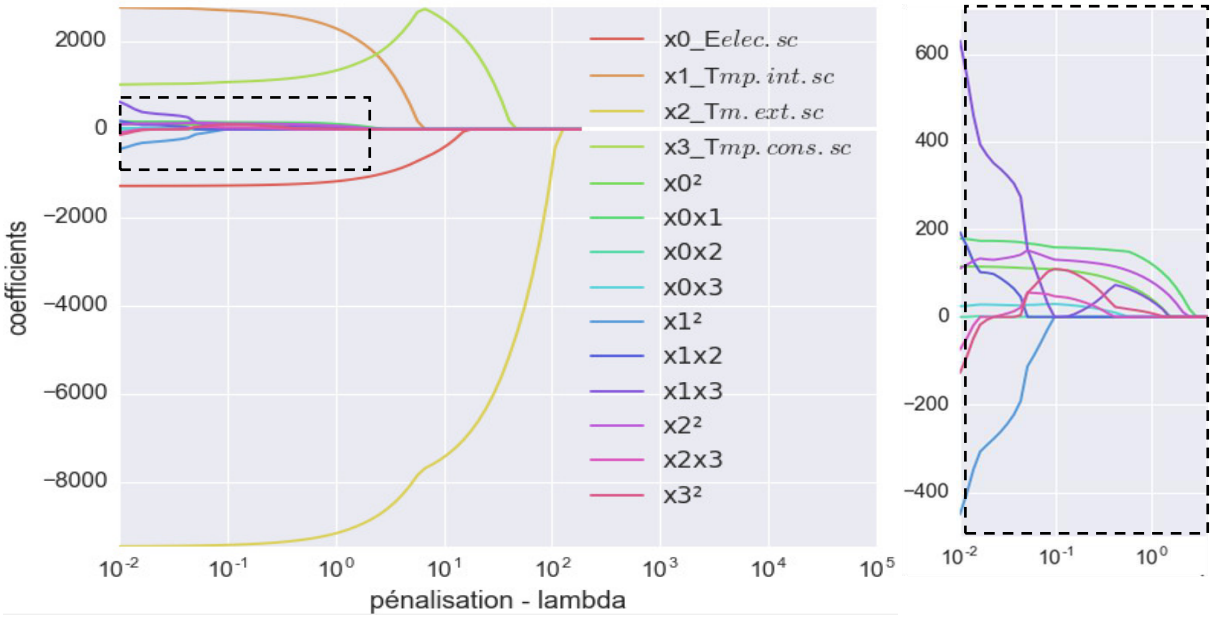

Figure 5-6 : Régression LASSO effectuée sur le modèle polynomial complet. Zoom sur l'évolution des coefficients des termes non linéaires (droite) en fonction de λ

Un modèle polynomial réduit composé des cinq termes influents du second ordre est construit. Le coefficient de régression R^1 associé est très proche de celui du modèle complet. Les coefficients identifiés par régression quantile sont présentés dans la colonne de droite du Tableau 5-9. On constate globalement une diminution de la valeur des coefficients par rapport à ceux du modèle polynomial complet. L'équation ci-dessous reprend le modèle polynomial réduit mis en forme :

$$LGC_{5\%} = 5{,}655.10^4 - 9{,}440.10^3 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right) \quad \textbf{(5.5)}$$

$$+5{,}451.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right) - 1{,}112.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)^2$$

$$+3{,}576.10^3 * \left(T_{mp.cons.sc} - T_{mp.cons.sc-ref}\right) + 6{,}59.10^2 * \left(T_{mp.cons.sc} - T_{mp.cons.sc-ref}\right)^2$$

$$+2{,}071.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)\left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right)$$

$$+8{,}26.10^2 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)\left(T_{mp.cons.sc} - T_{mp.cons.sc-ref}\right)$$

$$-1{,}529.10^3 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right)\left(T_{mp.cons.sc} - T_{mp.cons.sc-ref}\right)$$

$$-2{,}792.10^4 * \left(\frac{E_{elec.sc}}{E_{elec.sc-ref}} - 1\right)$$

Ce modèle constitue un bon compromis entre transparence, simplicité et précision. La Figure 5-7 illustre les variations du modèle de $LGC_{5\%}$ identifié par régression quantile linéaire et du modèle de régression quantile polynomial réduit. Les FA varient ici indépendamment, les autres FA étant considérées à leur valeur de référence.

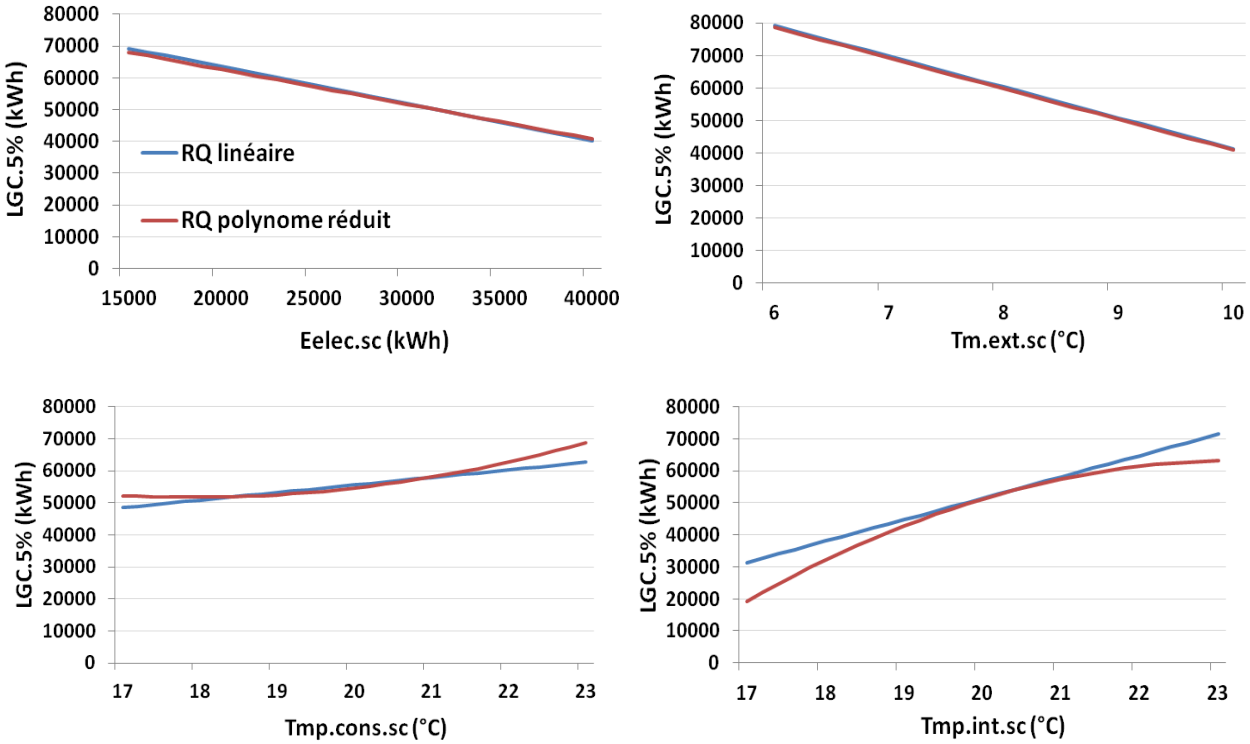

Figure 5-7 : Variations de la $LGC_{5\%}$ pour le modèle linéaire (RQL) en bleu, et polynomial (RQP), en rouge, en fonction des FA, les autres facteurs étant considérés à leur valeur de référence

5.1.5.3. Engagement sur l'amélioration de la performance

Un engagement sur la diminution des consommations d'énergie de chauffage passe par l'expression d'un gain relatif minimal garanti pour un risque de défaut $\alpha = 5\ \%$ défini à partir des modèles décrits précédemment (équation 5.4). La $LGC_{5\%}$ est soit exprimé avec le modèle linéaire (RQL) soit avec le modèle polynomial réduit (RQP).

$$GR_{5\%} = (C_{pre} - LGC_{5\%}) \,/\, E_{pre} \qquad \textbf{(5.6)}$$

Les diminutions de consommations $\mathrm{GR}_{5\%-\mathrm{ref}}$ pour les valeurs de référence des FA sont respectivement de 55,2 % pour le modèle RQL et de 55,5 % pour le modèle RQP. Les modifications du gain relatif garanti pour des variations indépendantes des FA sont présentées à la Figure 5-8. On constate l'effet des termes non-linéaires pour des variations de $T_{mp.int.sc}$ et $T_{mp.cons.sc}$.

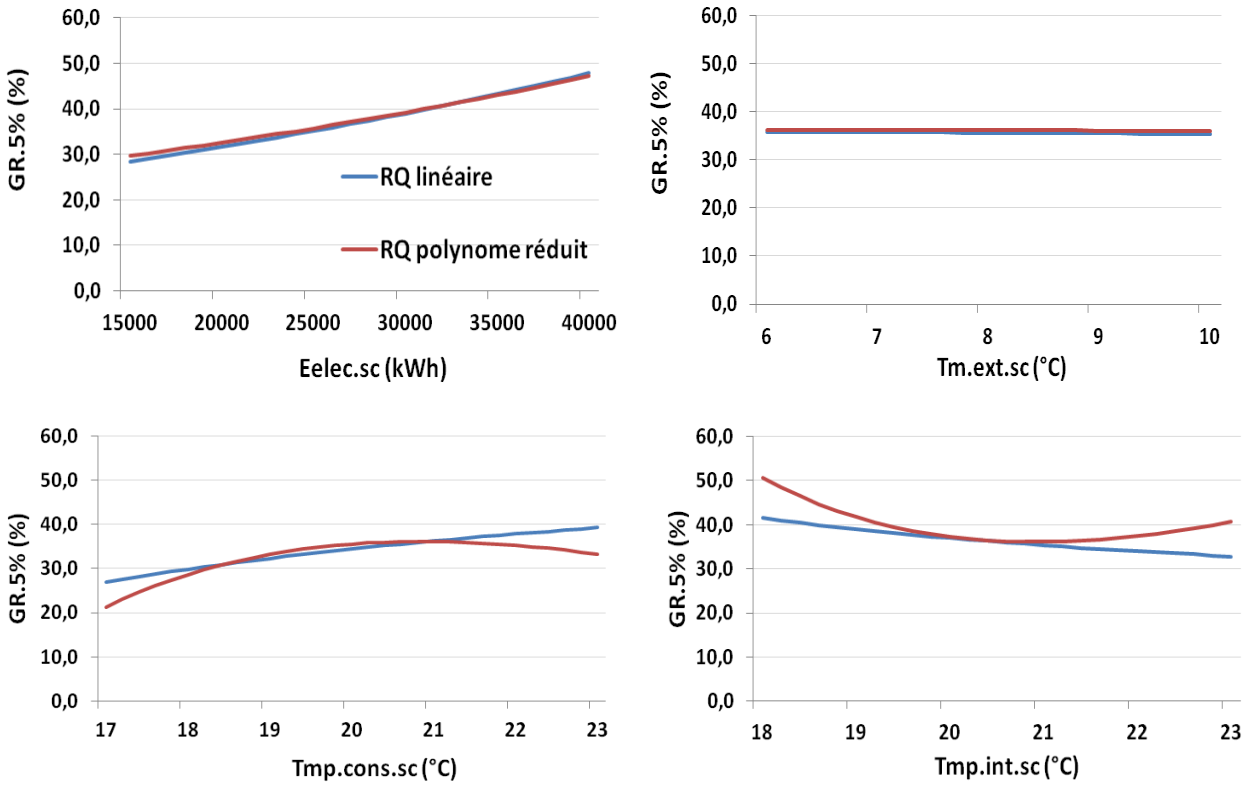

Figure 5-8 : Variations du $GR_{5\%}$ pour le modèle linéaire et polynomial en fonction des FA, les autres facteurs étant considérés à leur valeur de référence

Ces modèles sont entrainés sur un espace de variation de facteurs d'ajustement, délimité par les données de simulation produites. Comme vu précédemment au chapitre 4, leur validité n'est pas assurée au-delà de ces bornes, et en extrapolant, les modèles mathématiques peuvent fournir des résultats non réalistes physiquement. Afin de mieux appréhender les plages de variation possibles de $GR_{5\%}$, la Figure 5-9 présente la distribution des gains relatifs pour les abscisses (jeux de FA) issues des simulations, calculées avec les deux modèles. Ces diminutions garanties s'échelonnent entre 25 et 40 % environ. Si pour chaque jeu de FA on compare le gain relatif effectif (calculé avec la consommation de chauffage obtenue après travaux et le modèle de C_{pre}) et le $GR_{5\%}$ ajusté dans les conditions de FA observées, on peut étudier la distribution des écarts au seuil garanti (Figure 5-9 B).

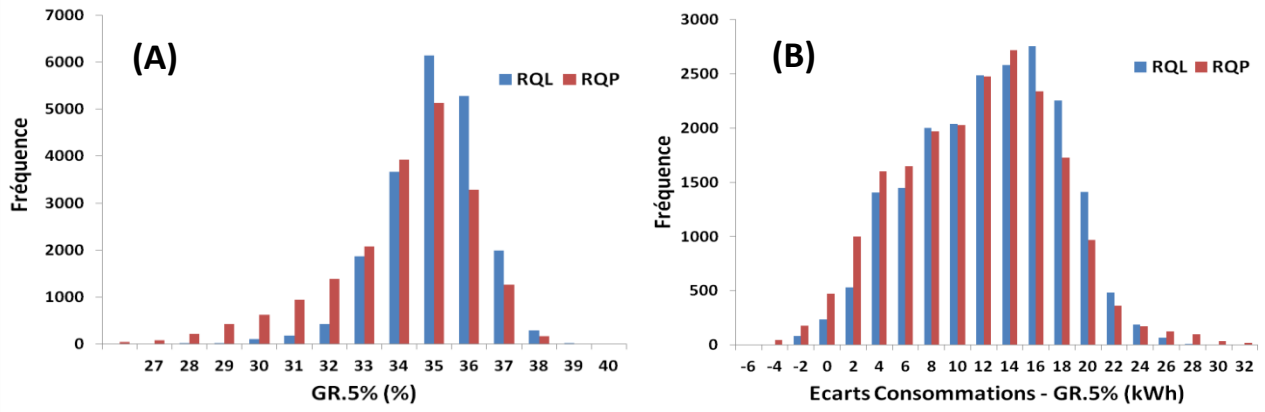

Figure 5-9: Distributions des gains relatifs garantis ajustés dans les conditions observées dans les simulations pour les deux modèles de $LGC_{5\%}$ (A) et de la différence entre les gains issus du modèles d'ajustement et ceux calculés par SED (B)

Le seuil garanti n'est pas atteint (écarts négatif) dans 3,5 % des cas pour le modèle RQL et 3,3 % pour le modèle RQP, ce qui est cohérent avec l'objectif initial de 5 % de risque de défaut, bien qu'un peu faible. Ces valeurs s'expliquent par l'ajout de l'incertitude statistique (évaluée par *bootstrap*) sur le coefficient constant du modèle de régression, qui considère un risque de 5 % sur un quantile à 95 %, limitant finalement le risque statistique global de défaut à une valeur inférieure à 5 % (§ 4.1).

5.1.5.4. Bilan de l'étude amont

L'étude réalisée s'appuyant sur la modélisation et la simulation du bâtiment avant et après réhabilitation permet de définir les éléments nécessaires à la mise en place d'un contrat de garantie de performance énergétique. Ces éléments sont constitués des modèle de consommations de référence avant réhabilitation et de $LGC_{5\%}$ après travaux, ajustable en fonction des FA, de l'équation du $GR_{5\%}$ qui dépend simplement de ces deux derniers modèles, du plan technique de mesure des consommations et des conditions d'exploitation et finalement des modalités de l'exploitation et de la vérification annuelle de l'engagement (mesures et calcul des FA, utilisation des modèles).

5.1.6. Mesures et vérification

Suite à la réhabilitation des logements à Feyzin en 2011, des mesures ont été conduites afin de suivre les performances énergétiques. Les consommations d'énergie de l'ensemble des postes réglementaires et non réglementaires ont donc été suivies, ainsi que d'autres données liées aux conditions d'exploitation. Les températures intérieures ont été mesurées dans les appartements. Les

conditions climatiques sont connues via la récupération des données de la station *Météo-France* de Feyzin.

L'application de la phase d'ajustement effectif et de vérification de l'engagement a été illustrée au § 3.3.6. Les modèles (RQL et RQP) déterminés dans cette partie reposent sur la connaissance des températures intérieures, qui peuvent être obtenues par mesure directe, et des températures de consigne, qui peuvent provenir de l'enregistrement par une GTB/GTC des valeurs fixées sur les thermostats (ce qui n'est pas techniquement possible dans tous les projets). Dans cette étude, les informations des températures de consigne ne sont pas disponibles. Afin d'illustrer la fin de la méthodologie nous allons tenter d'estimer une valeur de température de consigne moyenne sur la saison de chauffe, de manière similaire à ce qui a été fait au § 5.1.5.1 pour le calage du modèle de référence avant réhabilitation. Les valeurs de $T_{mp.int.sc}$, $T_{m.ext.sc}$ et $T_{mp.cons.sc}$ étant fortement corrélées, on caractérise par régression linéaire la dépendance entre ces trois grandeurs en exploitant les résultats de simulation après travaux afin de déduire une valeur probable de $T_{mp.cons.sc}$:

$$T_{mp.cons.sc} = -0{,}4025 + 1{,}0593 * T_{mp.int.sc} - 0{,}2055 * T_{m.ext.sc} \quad \textbf{(5.7)}$$

avec $R^2 = 0{,}874$ et $\sigma(T_{mp.cons.sc}) = 0{,}47\ °C$ d'où, par définition de R^2, $\sigma_{résidus} = 0{,}17\ °C$. On considère une plage probable de $T_{mp.cons.sc}$ comprise entre la valeur du modèle de l'équation 5.5 plus ou moins deux écarts-types des résidus de ce modèle soit $22{,}16$ +/- $0{,}34\ °C$.

Les mesures et traitement de données de la période allant du 13/02/12 au 12/02/13 permettent de déterminer les valeurs des FA correspondant aux conditions d'exploitation de cette première année post-intervention dont les valeurs sont décrites dans le Tableau 5-10.

Tableau 5-10 : Résultats de mesures issues de la première année d'exploitation du bâtiment

Grandeurs d'intérêt	***valeurs***
Consommation énergétique de chauffage (kWh)	*64700*
$T_{mp.int.sc}$ *(°C)*	*22,9*
$T_{m.ext.sc}$ *(°C)*	*8,27*
$E_{elec.sc}$ *(kWh)*	*19681*
$T_{mp.cons.sc}$ *(°C)*	***22,6** - [21,8 ; 22,5]*

Le Tableau 5-11 présente les résultats d'ajustement des différents modèles dans les conditions mesurées. À gauche, les valeurs de référence (centrées sur les conditions médianes observées parmi les simulations) de ces modèles sont rappelées pour constater l'effet induit par le processus. Il est aussi intéressant de comparer les consommations réelles et ajustées aux

caractéristiques de la distribution initiale de la PIV dont la moyenne est de 48376 kWh et le quantile à 95 % de 65740 kWh.

Dans l'optique de vérification de l'engagement, les valeurs ajustées de $LGC_{5\%}$ et de $GR_{5\%}$ sont respectivement comparées aux consommations de chauffage et à la diminution des consommations réelles. On constate que cette consommation est bien inférieure à la $LGC_{5\%}$ et que le gain observé est supérieur à celui garanti, l'engagement est donc respecté, pour les deux modèles possibles de $LGC_{5\%}$. Ceci se vérifie sur la plage probable de $T_{mp.cons.sc}$ identifiée.

Tableau 5-11 : Ajustement et vérification de l'engagement sur la première année d'exploitation

		Référence	***Ajustement***
C_{pre}		*88500*	***120561*** *- [118358 ; 122642]*
$LGC_{5\%}$ (kWh)	RQL	*57100*	***81854*** *- [81006 ; 82654]*
	RQP	*56550*	***79689*** *- [77106 ; 82286]*
$GR_{5\%}$ $(\%)$	RQL	*55,2*	***32,1*** *- [31,6 ; 32,6]*
	RQP	*55,5*	***33,9*** *- [34,9 ; 32,9]*

	Mesuré
$Conso.Ch.$ (kWh)	***64700***
$GR_{réel}$ $(\%)$	***46,3*** *- [45,3 ; 47,2]*

5.1.7. Conclusion du cas d'étude de Feyzin

L'étude complète menée sur ce projet a permis d'illustrer la mise en œuvre de la méthodologie pour la GPE dans le cas d'un engagement relatif sur la diminution des consommations d'énergie suite à une intervention de réhabilitation. La description du modèle physico-probabiliste a été affinée suite à une analyse de sensibilité qui a permis de mettre en évidence les paramètres influents du modèle SED. Une étude exhaustive des modalités d'ajustement a été menée s'intéressant conjointement aux FA à considérer et à la précision du plan de mesure à mettre en œuvre.

Le « meilleur modèle » de limite garantie de consommation d'énergie ($LGC_{5\%}$) a été identifié. Il a été mis en évidence l'intérêt de mesurer à la fois les températures de consigne et les températures intérieures bien que la connaissance seule de l'une de ces deux grandeurs permettent déjà l'obtention d'un modèle de qualité, se prévalant des risques liés à des températures de consigne de chauffage très élevées et très énergivores. Bien que le FA lié à la température de consigne ne soit pas mesurable (et n'ait pas été mesuré) sur ce cas d'étude, il a été conservé dans cet exemple où le modèle le plus performant était recherché.

L'exploitation des résultats de SED du modèle avant réhabilitation a permis la détermination d'un modèle de référence, recalé grâce à des mesures réalisées avant travaux. Pour la limite garantie après travaux, un modèle polynomial a été identifié du fait des interactions entre trois facteurs d'ajustement, les températures moyennes sur la saison de chauffe à l'extérieur, à l'intérieur et des consignes de chauffage.

Le processus de vérification qui est appliqué de manière cyclique durant l'exploitation du bâtiment rénové a été testé avec des données mesurées lors de la première année suite aux travaux. La consommation de chauffage mesurée n'excède pas le niveau d'engagement déterminé en amont et ajusté aux conditions réelles. Le gain relatif minimal garanti est donc respecté.

On constate dans ce cas d'étude que le modèle d'ajustement de la $LGC_{5\%}$ permet en particulier de se prémunir de l'impact des températures de consigne très élevées sur les consommations de chauffage. De telles valeurs extrêmes des moyennes de température, dues ici à un effet rebond suite aux travaux, correspondent à une situation très peu probable du modèle de simulation physico-probabiliste du fait du nombre d'appartements et donc du foisonnement des probabilités. La consommation réelle excède presque le quantile global à 95 % des distributions sans ajustement. Cette étude illustre donc comment le modèle d'ajustement permet de se protéger de ce risque difficile à anticiper.

5.2. GPE en construction neuve : bâtiment de bureaux IZUBA

5.2.1. Présentation du cas d'étude

Dans cette seconde partie, un projet de garantie de performance énergétique est conduit sur un bâtiment de bureaux nouvellement construit, avec pour objectif de s'engager sur plusieurs postes de consommations énergétiques. Dans un premier temps, la grandeur d'intérêt envisagée est la consommation annuelle de la pompe à chaleur qui fournit le chauffage et la climatisation, puis un engagement sur la consommation énergétique totale du bâtiment est envisagé. Ce cas d'étude se caractérise par une modélisation plus exhaustive du fonctionnement énergétique du bâtiment et de ses systèmes.

5.2.1.1. Général

L'entreprise IZUBA Energies, BET et éditeur de logiciels, a fait construire son siège dans la commune de Fabrègues près de Montpellier. Celui-ci est prévu pour pouvoir accueillir une vingtaine de salariés. Les travaux ont eu lieu durant l'année 2014 et la livraison a été effectuée début 2015. Le bâtiment de 396 m² répartis sur deux niveaux a été conçu dans le cadre d'une approche globale visant à quantifier et réduire au maximum les impacts environnementaux sur son cycle de vie. Cela passe notamment par une très haute performance énergétique, la maitrise des consommations, l'utilisation d'énergies renouvelable et par la production locale d'électricité photovoltaïque. Plus de détails sur la réalisation et le fonctionnement énergétique de l'ouvrage sont disponibles sur un site web dédié à ce projet[11].

5.2.1.2. Enveloppe

L'enveloppe se caractérise par des matériaux à faible impact environnemental. Le bâtiment possède une ossature bois et est isolé avec des bottes de paille. Des enduits en terre, et surtout une paroi centrale en terre crue, ont été pensés pour apporter une forte inertie thermique. Concernant les infiltrations, une forte étanchéité est visée. Les prévisions amont prévoyaient des infiltrations de l'ordre de 0,04 vol/h. Des brise-soleil extérieurs permettent de limiter les apports solaires directs en été. La Figure 5-10 montre une vue de la façade sud du bâtiment.

[11] http://batiment.izuba.fr/

Figure 5-10: Vue de la façade sud du bâtiment et des systèmes solaires photovoltaïque et thermique

5.2.1.3. Systèmes énergétiques

- *Production de chaleur et de froid*

Une pompe à chaleur (*Weishaupt* – WWP S 11 IBER) réversible sur sonde géothermique assure la production de chauffage et de climatisation. Le fluide frigorigène utilisé est le R407C. Ses caractéristiques sont présentées dans le Tableau 5-12.

Tableau 5-12 : Caractéristiques de la pompe à chaleur

Mode	***Puissance***	***Performance***	***Conditions***
Chaud	*11,4 kW*	*COP 4,0*	*0 °C amont / 35 °C aval*
Froid	*13,8 KW*	*EER 6,3*	*18 °C amont / 10 °C aval*

L'émission d'énergie se fait par plancher chauffant/rafraichissant. Le régime des températures de départ et retour d'eau est de 30°C/20 °C pour le chauffage, commandé par une loi d'eau et de 18/23 °C pour le rafraichissement. Une régulation locale de l'émission de chauffage et de climatisation est aussi possible, l'échangeur étant partagé en huit branches parallèles (quatre par niveaux). Des ventilo-convecteurs, alimentés sur une branche du circuit de distribution, sont aussi installés dans la salle de réunion du fait de besoins de climatisation ponctuels plus importants. Un schéma de fonctionnement de l'installation technique est disponible à l'annexe A-5-3.

- *Ventilation*

Un système double-flux assure le renouvellement de l'air intérieur. La ventilation des locaux est commandée afin d'en couper le fonctionnement en dehors des plages d'occupation. La centrale est

de marque *SWEGON*, de type *GOLD RX TOP*, le débit maximal est de 1200 m^3/h et le rendement de l'échangeur de chaleur est certifié supérieur à 80 %. Par ailleurs, une ventilation nocturne manuelle est envisagée en période estivale.

- *Eau chaude sanitaire*

L'eau chaude sanitaire est produite dans un ballon de 200 L alimenté en chaleur par un capteur solaire thermique d'une surface d'environ 2 m² et par un système d'appoint électrique.

- *Éclairage*

L'éclairage artificiel a été conçu afin de répondre aux besoins en éclairement spécifiques à chaque pièce et est composé de matériel à faibles consommation d'énergie. Le Tableau 5-13 récapitule les puissances installées.

Tableau 5-13 : Conception de l'éclairage artificiel

Locaux	***Puissance installée (W)***	***Puissance surfacique (W/m²)***
Bureaux	*55*	*5,9*
Postes de travail	*56*	*5,9*
Salle de formation	*420*	*9,3*
Salle de réunion	*196*	*14,6*
Circulations	*270*	*2,8*
Sanitaires	*165*	*7,7*
Pièces de service	*732*	*13,2*

- *PV*

Afin d'équilibrer les bilans énergétiques et environnementaux du bâtiment, une installation photovoltaïque d'une puissance crête de 28,8 kWc est installée en toiture. L'inclinaison est de 10° par rapport au sol et l'orientation est de 10° par rapport au sud, la façade sud étant légèrement tourné vers l'ouest. Le rendement maximal des panneaux *SUNPOWER* installés est de 20,4 %.

5.2.2. Modélisation

Afin de conduire l'étude de GPE, un modèle SED du bâtiment développé en phase de conception avancée a été récupéré. Les caractéristiques de l'enveloppe et des systèmes y sont décrits afin de simuler le fonctionnement énergétique annuel complet de l'ouvrage. Les détails de ce modèle sont disponibles en annexe A-5-4.

5.2.2.1. Modèle SED de base

- *Enveloppe*

L'enveloppe et ses différents constituants sont modélisés à partir des données de conception. Les compositions des parois, menuiseries et ponts thermiques sont décrites. Les 26 pièces du bâtiment sont réparties dans 18 zones thermiques pour la simulation énergétique dynamique. Des ponts thermiques linéiques, intégrés aux éléments de paroi, et de menuiserie sont spécifiés. Une vue 3D du bâtiment sur le logiciel *Pléiades* est présentée à la Figure 5-11. Les débits d'infiltration sont fixés à 0,04 vol/h. Le Tableau 5-14 récapitule quelques caractéristiques globales de l'enveloppe dans le modèle SED. Un scénario d'occultation (intérieure) est appliqué aux fenêtres des semaines 20 à 39 correspondant à une diminution de 50 % de la surface transparente.

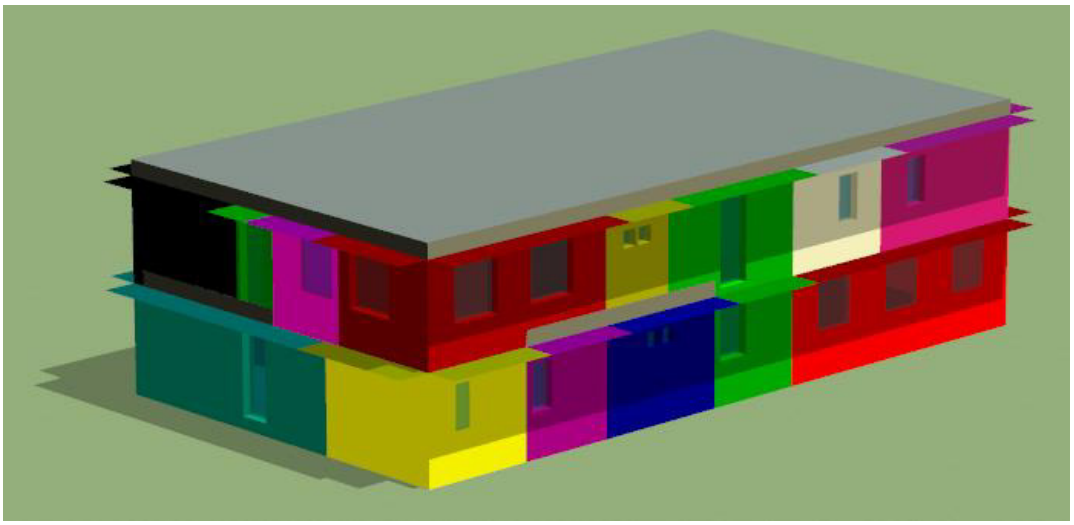

Figure 5-11: Modélisation 3D et zonage du bâtiment (Pléaides)

Tableau 5-14 : Principales caractéristiques thermiques de l'ouvrage

Façades	$\boldsymbol{U = 0,165\ W/(m^2.K)}$
Toiture terrasse	$\boldsymbol{U = 0,155\ W/(m^2.K)}$
Menuiseries PVC 4/12/4	$\boldsymbol{U = 1,85\ W/(m^2.K)}$
Total ponts thermiques	$\boldsymbol{\sum \psi L = 69\ W/K}$
Débit VMC DF	$\boldsymbol{V_{vmc} = 0,9\ vol/h}$
Débit infiltrations	$\boldsymbol{V_{inf} = 0,04\ vol/h}$

- *Systèmes*

Les différents systèmes présentés au paragraphe précédent sont décrits avec le logiciel *Pléiades*. Les modèles de la RT2012 sont utilisés pour décrire le fonctionnement des systèmes de production et de distribution de chauffage et de climatisation, comprenant la description du forage géothermique, de la boucle amont, de la PAC, des boucles de distribution alimentant le plancher chauffant et les ventilo-convecteurs, ainsi que ces émetteurs. Il en est de même pour le montage de production d'ECS avec capteur solaire. Les modèles réglementaires de la méthode ThBCE 2012, qui décrivent le fonctionnement de ces systèmes, sont utilisés en post-traitement de la SED pour calculer les consommations énergétiques associées aux différents postes. Pour la pompe à chaleur, la matrice de performance est calculée à partir d'une « valeur pivot » correspondant aux conditions nominales des sources amont et aval, tel que décrit dans la méthode réglementaire.

- *Sollicitations temporelles*

Le modèle de génération stochastique de données d'occupation dans les bâtiments tertiaires de (Vorger 2014) fournit les données temporelles d'occupation et de puissances dissipées (bureautique, cuisine). L'utilisation de l'éclairage est couplée au modèle d'occupation stochastique et aux scénarios de présence générés.

Des scénarios de température de consigne de chauffage et de climatisation déterministes sont définis indépendamment dans les 8 espaces correspondant aux branches de l'échangeur sous plancher. Les consignes de chauffage nominales sont de 21 °C avec un réduit de 2 °C la nuit (19h-7h) et la climatisation est fixée à 26 °C de 7h à 19h.

Les scénarios de ventilation décrivent un renouvellement d'air en journée compris entre 0,8 et 0,9 vol/h. Des scénarios de surventilation nocturne sont spécifiés (3 vol/h de 19h à 7h) dans toutes les zones en été (semaines 20 à 39).

5.2.2.2. Description physico-probabiliste initiale et analyse de sensibilité

Afin de prendre en compte l'incertitude liée à la réalisation du bâtiment, l'approche probabiliste présentée dans cette thèse est mise en œuvre. Dans un premier temps, des fonctions de densité de probabilité (*PDF*) uniformes sont spécifiées et décrivent les plages de variation possible des paramètres du modèle SED. Les modèles de variabilité des sollicitations météorologiques et d'usage ne sont pas utilisés, du fait de la nature différente de ces paramètres et de leur influence déjà bien documentée.

Pour réaliser l'analyse de sensibilité qui a pour but d'identifier les facteurs influents, les sollicitations (occupation, consommations électriques, éclairage) fournies par le modèle stochastique d'occupation sont fixées à une valeur nominale tout comme les données

météorologiques. Les scénarios de consigne de température étant déterministes, leur variabilité est considérée avec un facteur global (Tableau 5-13) décalant uniformément les températures horaires.

Tableau 5-15 : Bornes des lois uniformes sur les paramètres incertains du modèle SED

Paramètres	***Borne inf.***	***Borne sup.***
MATERIAUX / ÉLEMENTS		
Conductivité matériaux (12)	*- 5 %*	*+ 5 %*
Masse volumique matériaux (12)	*- 2 %*	*+ 2 %*
Chaleur spécifique matériaux (12)	*- 5 %*	*+ 5 %*
PAROIS		
Épaisseur matériaux dans parois (24)	*-1 mm*	*+ 1 mm*
Hauteur parois (8)	*- 1 %*	*+ 1 %*
Largeur parois (8)	*- 1 %*	*+ 1 %*
MENUISERIES		
Facteur solaire vitrages (8)	*- 5 %*	*+ 5 %*
Conductivité vitrages (8)	*- 5 %*	*+ 5 %*
Hauteur vitrages (8)	*- 1 %*	*+ 1 %*
Largeur vitrages (8)	*- 1 %*	*+ 1 %*
PONTS THERMIQUES		
Ponts intégrés (2)	*- 10 %*	*+ 10 %*
Ponts linéiques (5)	*- 10 %*	*+ 10 %*
Ponts menuiseries (3)	*- 10 %*	*+ 10 %*
VENTILATION		
Infiltrations (18)	*- 10 %*	*+ 10 %*
Débits de ventilation mécanique (18)	*- 5 %*	*+ 5 %*
Débit de surventilation (1)	*- 20 %*	*+ 20 %*
SCENARIOS		
Températures (18)	*-2 °C*	*+ 2°C*
Occultation (1)	*- 20 %*	*+ 20 %*
SYSTEMES		
Rendement PAC	*- 5 %*	*+ 5 %*
Températures émetteur	*- 0,5 °C*	*+ 0,5 °C*
Puissance ventilation	*- 5 %*	*+ 5 %*
Rendement DF	*- 5 %*	*+ 5 %*

Deux analyses de sensibilité sont menées avec la méthode de Morris avec comme grandeurs d'intérêt les consommations de la PAC pour le chauffage et pour le refroidissement. Dix répétitions sont effectuées sur la variation des 178 paramètres incertains ce qui nécessite donc la réalisation de 1790 simulations. La Figure 5-12 (consommation de chauffage) et la Figure 5-13 (consommation de climatisation) présentent les résultats de ces analyses de sensibilité où sont mis en avant les paramètres les plus influents sur les grandeurs d'intérêt. L'écart-type σ et la moyenne μ^* (positive) des effets élémentaires sont présentés dans le graphique à gauche. Les distances des points à l'origine du graphe, indicateur courant de la sensibilité avec la méthode de Morris sont présentées dans le diagramme à droite.

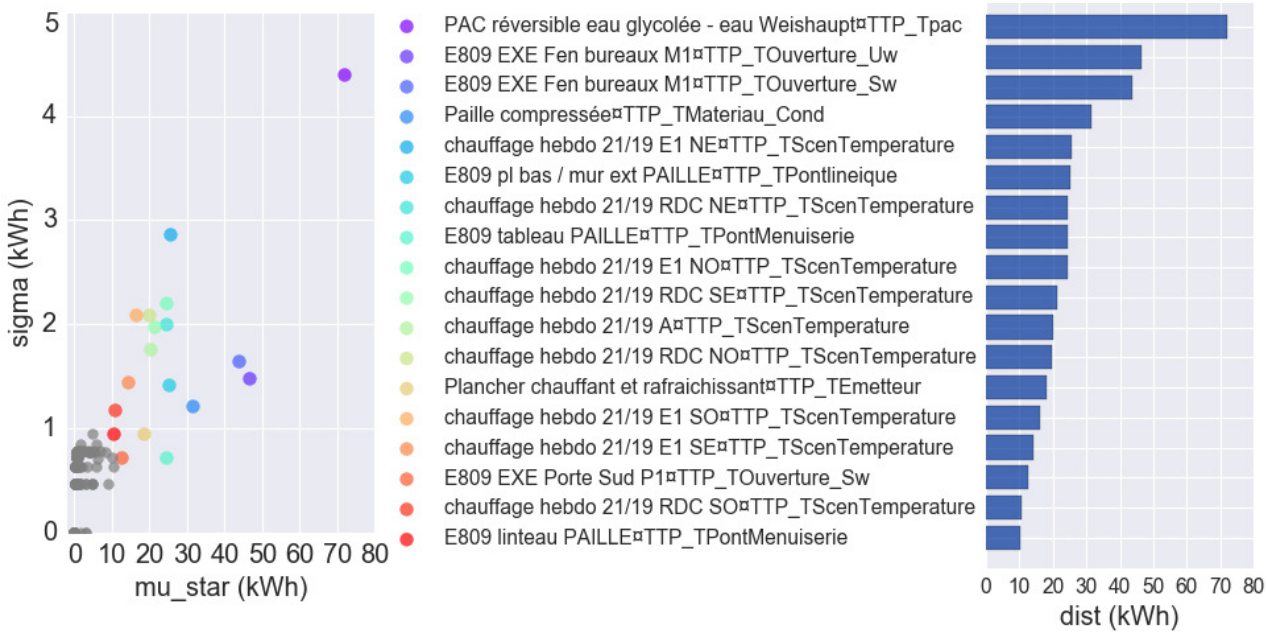

Figure 5-12: Résultats de l'analyse de sensibilité sur les consommations annuelles de chauffage avec la méthode de Morris

Le rendement nominal de la PAC est le facteur dont l'incertitude a le plus d'influence sur les consommations énergétiques de chauffage. On constate parmi les autres paramètres influents sur les consommations de chauffage, les caractéristiques des vitrages, la conductivité de l'isolant en paille, les ponts thermiques, la variation temporelle de température du plancher chauffant et les scénarios de températures de consigne.

Concernant l'impact sur la variabilité des consommations de refroidissement, l'analyse fait ressortir comme principaux paramètres influents le rendement de la PAC et les scénarios de consigne de climatisation, le scénario d'occultation, la variation temporelle de température du plancher chauffant et les caractéristiques de vitrages.

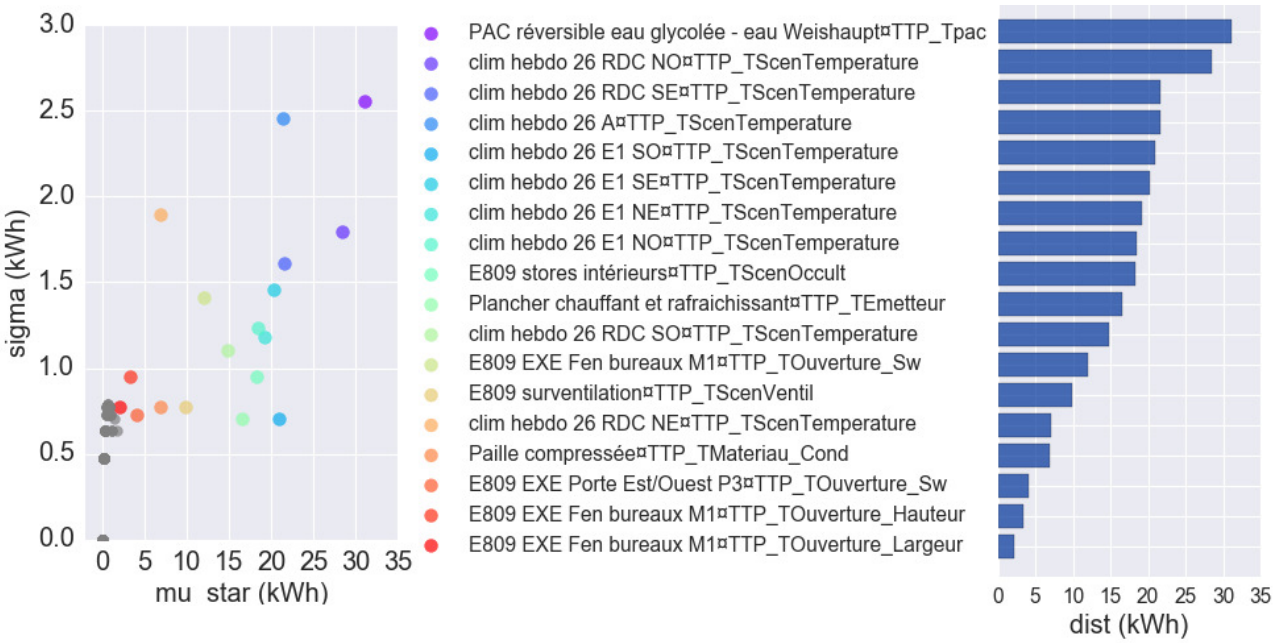

Figure 5-13: Résultats de l'analyse de sensibilité sur les consommations annuelles de climatisation avec la méthode de Morris

L'étude croisée des deux analyses permet d'identifier 32 paramètres dont l'influence (distance à l'origine) est nulle sur les deux grandeurs et 85 dont l'influence est inférieure à 1 kWh/an. Les premiers sont fixés à leur valeur nominale et l'incertitude sur les seconds n'est pas retravaillée et demeure décrite selon les lois uniformes déterminées initialement. Les paramètres restant sont étudiés avec une attention particulière.

5.2.2.3. Modèle physico-probabiliste final

- *Incertitudes sur les paramètres influents*

- Rendement nominal de la PAC

L'incertitude initiale de 5 % sur les rendements de la PAC réversible a une forte influence sur les consommations énergétiques. Les valeurs nominales sont certifiées par le constructeur et il est difficile d'avoir une estimation plus précise de l'incertitude sur ces valeurs. En adoptant une démarche conservatrice on définit une fonction de probabilité décentrée sur l'incertitude relative, une loi normale tronquée $\mathrm{N}(0,\ 5\ \%, -10\ \%, 5\ \%)$ où de fortes dégradations de performance, bien que peu probables peuvent être observées en simulation.

- Scénarios de température

Les consignes de température de climatisation et de chauffage impactent les besoins et sont liées aux réglages de l'installation et aux actions des usagers. Un ajustement en fonction de ces températures, rendu possible par la mesure des températures intérieures ou l'enregistrement des consignes, est à prévoir ce qui limite le risque de défaut d'un engagement de GPE dû à ces éléments.

Leur variabilité étant imprévisible a priori, l'incertitude initiale assez large de plus ou moins 2 °C sur les scénarios nominaux est maintenue mais décrite selon un loi normale tronquée $\mathrm{N}(0, 1, -2, 2)$.

- Surventilation et occultation

Les besoins de climatisation sont dépendants des scénarios de surventilation nocturne et d'occultation estivale des fenêtres, deux paramètres liés aux usagers, dont la commande est prévue pour être manuelle en exploitation, et dont l'intégration dans un modèle d'ajustement est délicate voire impossible. Le risque de surconsommation de climatisation lié à la mauvaise prise en compte de ces éléments est donc fort. L'incertitude sur ces scénarios est élargie, une surventilation nocturne estivale comprise entre 1 et 4 vol/h et des occultations entre 10 et 60 % des surfaces vitrées sont modélisées (lois uniformes).

- Paramètres physiques et techniques

Les bornes d'incertitude (+/- Δ) des autres paramètres influents liés aux caractéristiques des vitrages, des matériaux, aux ponts thermiques, à la distribution et émission de l'énergie sont conservées. Des lois normales tronquées $\mathrm{N}(0, \frac{\Delta}{2}, -\Delta, \Delta)$., plus représentatives de la réalité de l'incertitude sur ces caractéristiques techniques, sont prises en compte.

- *Variabilité des sollicitations*

Pour la réalisation de la propagation d'incertitudes et variabilités complète, les modèles de génération stochastique de sollicitations sont utilisés. Une variabilité de données météorologiques est générée, par le modèle présenté au chapitre 2, à partir du fichier météo type de Nice (zone H3 de la RT2012). Le modèle de génération stochastique de comportement des usagers de bureaux développé par (Vorger, 2014) permet de générer des scénarios de présence, de consommations électriques et de puisages d'ECS dans les zones thermiques du modèle SED.

5.2.3. Propagation des incertitudes et variabilités

Une simulation du modèle physico-probabiliste complet est menée, correspondant à un ensemble de tirages dans les lois et modèles probabilistes pour la réalisation de 2000 simulations SED.

Les distributions des résultats de consommation énergétiques des différents postes sont présentés à la Figure 5-14. Les consommations élevées d'électricité spécifique, de l'ordre de 8500 kWh/an ne sont pas représentées ici. Les consommations de distribution correspondent aux pompes, en particulier celle de la boucle de l'échangeur géothermique en amont de la PAC. Le Tableau 5-16 récapitule les caractéristiques statistiques principales de l'ensemble des distributions.

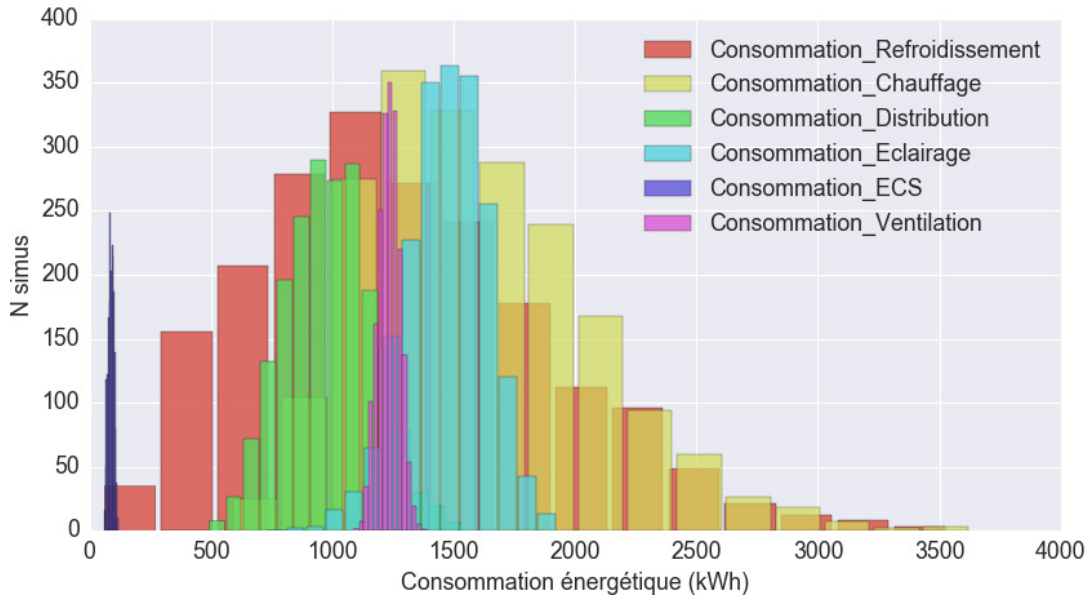

Figure 5-14: Distributions des consommations énergétiques des différents postes (2000 simulations)

Tableau 5-16 : Moyennes, écart-types et quantiles à 95 % des distributions des consommations des différents postes (kWh)

Postes	***Moyenne***	***Écart-type***	***q-95%***
PAC total	2880	649	4081
PAC chauffage	1601	470	2444
PAC refroidissement	1279	600	2360
Distribution	987	180	1284
Éclairage	1471	166	1729
Ventilation	1235	44	1305
ECS	90	12	109
Électricité spécifique	8656	1657	11522
TOTAL	15321	2050	18843

On constate en particulier la grande variabilité des consommations de chauffage, de climatisation et d'électricité spécifique parmi ces résultats.

5.2.4. Dispositif de mesure

La méthodologie prévoit la spécification d'un dispositif de mesure lié aux modalités d'ajustement déterminées pendant l'étude, en phase de conception. Sur ce cas d'étude, comme sur le précédent, le bâtiment est déjà réalisé et livré lors de cette étude, qui se doit tout de même d'être réalisée en « aveugle », sans tenir compte de la connaissance disponible a posteriori. Afin de mener à

son terme la méthodologie et de réaliser le processus de vérification des performances effectives, les facteurs d'ajustement sélectionnés et les capteurs associés sont imposés par les moyens de mesure réellement installés sur l'ouvrage, décrits ci-dessous et présentés à l'annexe A-5-5.

- *Énergie*

Des compteurs différenciés permettent de connaitre les consommations électriques de la PAC, de l'éclairage, de l'ECS (appoint), de la CTA, des appareils de bureautique. Des compteurs d'énergie sont aussi disponibles en amont et en aval de la PAC réversible ce qui permet de différencier le fonctionnement en chauffage et en climatisation. La production photovoltaïque est aussi mesurée.

On considère dans cette étude une incertitude-type IT de mesure de 2 % sur les relevés de compteurs électriques, notamment pour les différents postes qui disposent d'un compteur dédié. Dans le processus d'intégration a priori de l'impact des incertitudes de mesure par bruitage des résultats de simulation (§3.4), des erreurs de mesure virtuelles sont tirées aléatoirement dans des lois $N(0, IT, -2.IT, 2.IT)$.

- *Températures intérieures*

Huit capteurs de température ont été répartis sur les deux niveaux du bâtiment, dans différents bureaux, la salle de formation, la salle de réunion, des espaces associés à toutes les façades du bâtiment. Plus de détails sur ce dispositif sont présentés dans l'annexe A-5-5. Des tirages d'erreurs de mesure virtuelles sur les températures intérieures moyennes sont effectués dans des lois normales $N(0, IT, -2.IT, 2.IT)$ où IT est l'incertitude-type estimée pour chaque simulation en fonction des températures relevées selon les modalités décrites dans l'annexe A-5-6.

- *Données météorologiques*

La température extérieure et le rayonnement solaire global sont mesurés sur le site. On considère pour les calculs une incertitude systématique de 0,125 °C sur la valeur instantanée des températures et donc une incertitude équivalente sur la moyenne des températures sur une période. Le mode de calcul de l'incertitude sur les degré-heures a été présenté aux § 3.4.1 et § 5.1.4.

5.2.5. Engagement de GPE sur les consommations de chauffage et de climatisation

On étudie dans un premier temps la mise en œuvre de la méthodologie de GPE pour un engagement sur les consommations annuelles de la pompe à chaleur fournissant chauffage et climatisation.

5.2.5.1. Évaluation des opportunités et des modalités d'ajustement

La mise en œuvre de la suite de la méthodologie repose sur l'exploitation des résultats de simulation de la propagation des incertitudes et variabilités. La détermination d'un niveau d'engagement maximal sur les consommations de la PAC nécessite tout d'abord le choix de facteurs d'ajustement associés à un plan de mesure.

Etant donné les mesures planifiées sur le bâtiment, il est envisagé d'ajuster la limite garantie des consommations de la PAC en fonction de la température extérieure, des températures intérieures (liées aux consignes de chauffage et de climatisation) et aux consommations d'électricité spécifiques principalement liées à la bureautique qui provoque des apports de chaleur bénéfiques pour le chauffage et pénalisant pour la climatisation.

Il est nécessaire de différencier les FA liés à ces grandeurs impactant chauffage ou climatisation, et de définir de manière pertinente les périodes estivales et hivernales sur lesquels seront agglomérées (mesurées) les informations d'ajustement. Ainsi, la dépendance des consommations d'énergie à la température extérieure moyenne annuelle risque d'être négligeable car de fortes valeurs de $T_{m.ext}$ peuvent signifier des températures estivales élevées entrainant de fortes consommations liées à la climatisation, ou des températures hivernales élevées responsables de faibles consommations sur cette période. Par ailleurs, la période estivale ou hivernale à considérer doit maximiser la dépendance entre le FA et la grandeur d'intérêt. La température extérieure moyenne prise sur une période estivale trop resserrée risque de ne pas intégrer des périodes de forte chaleur la précédant ou la suivant. Sur une période trop large, la température moyenne est moins caractéristique des périodes de forte chaleur et est parasitée par les températures de période de mi-saison ou les besoins de refroidissement sont faibles.

Deux périodes estivales et deux périodes hivernales dont les caractéristiques sont présentées dans le Tableau 5-17 sont envisagées, couvrant des parts élargies ou restreintes des périodes de chauffage et de climatisation.

Tableau 5-17 : Saisons de chauffe et de refroidissement considérées pour l'ajustement

Saison	***Semaines***	***Dates***
sc_42_18	42 à 18	15/10 – 30/04
sc_44_13	44 à 13	1/11 – 31/03
sf_23_35	23_35	1/06 – 31/08
sf_25_32	25_32	15/06 – 15/08

- *Températures intérieures et extérieures*

Dans un premier temps, les FA liés aux températures intérieures et extérieures sont comparés. Les FA liés à la température extérieure qui sont étudiés sont $T_{m.ext.sc}$, $T_{m.ext.sf}$, $DH_{18.sc}$, $DHc_{24.sf}$.

Les températures intérieures moyennes, pondérées ou non, décrivent l'influence des consignes intérieures. L'ensemble de ces FA sont calculés sur les différentes périodes estivales et hivernales.

Les résultats de l'étude systématique des coefficients de régression R^1 des modèles de régression quantile (annexe A-5-7) identifie le meilleur jeu de FA composé de $T_{m.ext.sc_44_13}$, $T_{m.ext.sf_23_35}$, $T_{mp.int.sc_44_13}$ et $T_{mp.int.sf_23_35}$. Le coefficient R^1 pour ce modèle est de 0,385. Ces résultats permettent de constater quelques grandes tendances, la période de chauffe la plus caractéristique est celle comprise entre les semaines 44 et 13. Les limites de la période estivale ont très peu d'influence. Les DHc_{sf} fournissent de meilleurs modèles avec une température de référence de 24 °C plutôt que 26 °C.

- *Consommations d'électricité spécifique*

Dans un second temps on étudie l'intérêt de prendre en compte des FA liés aux consommations électriques dues aux occupants. La Figure 5-15 présente la régression *LASSO* réalisée avec les facteurs d'ajustement supplémentaires. Pour analyser ces figures, on s'intéresse aux valeurs des coefficients normalisés pour $\lambda = 0$, et à l'évolution de ces coefficients pour une pénalisation croissante. On constate ainsi que les valeurs des coefficients normalisés sont faibles pour les deux FA liés aux consommations électriques. Pour une pénalisation lambda croissante, ces deux variables sont rapidement éliminées, en particulier $E_{elec.sf}$.

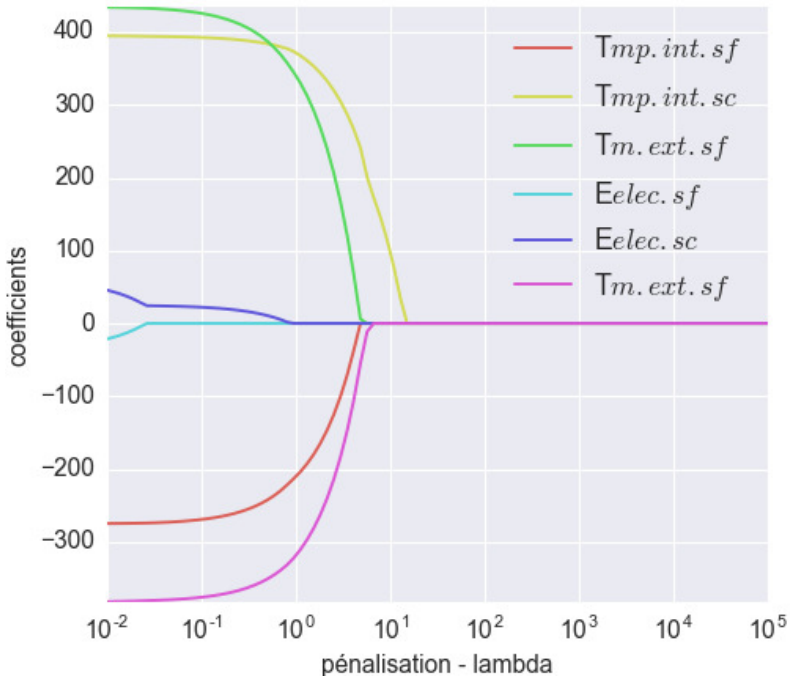

Figure 5-15: Régression LASSO sur les FA normalisés

La prise en compte de ces FA fournit un modèle de régression quantile de coefficient R^1 de 0,391 contre une valeur de 0,385 s'ils ne sont pas intégrés. Cette faible dépendance des consommations annuelles de la PAC aux consommations électriques spécifiques durant les périodes

de chauffage et de climatisation peut s'explique par la forte corrélation entre $E_{elec.sc}$ et $E_{elec.sf}$ au sein des résultats de simulation (Figure 5-16). En effet, la variabilité générée par le modèle occupant fournit des conditions d'usage où la quantité, la qualité du matériel informatique installé, le temps d'utilisation moyen sont déterminés en *pre-process* pour l'année et impactent globalement les consommations sur toute l'année. Des valeurs élevées de $E_{elec.sc}$ provoquent ainsi une baisse des besoins de chauffage mais sont aussi accompagnés de valeurs élevées de $E_{elec.sf}$ qui entrainement elles une hausse des besoins de climatisation. Les effets estivaux et hivernaux de valeurs extrêmes s'annulent. Ces FA ont donc un faible caractère explicatif de la consommation totale de la PAC.

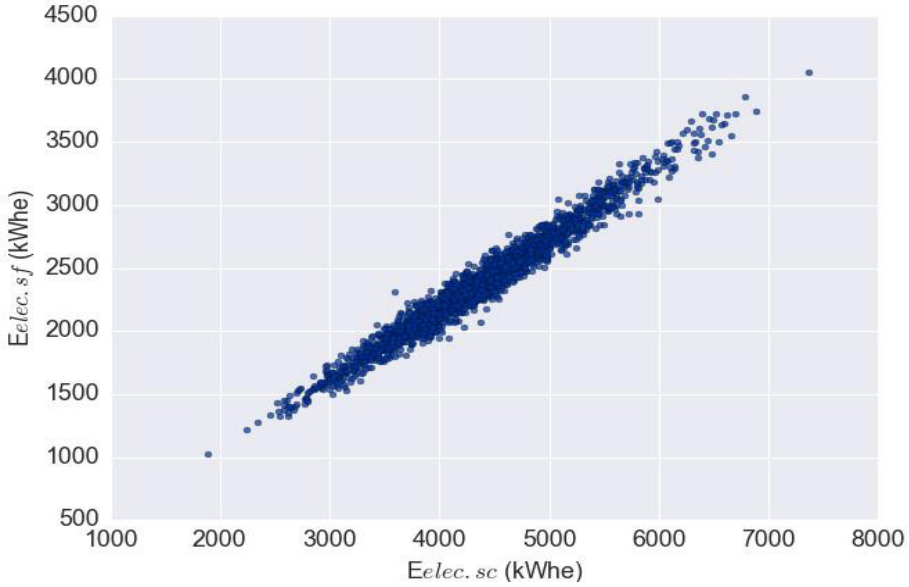

Figure 5-16: Corrélation entre les consommations d'électricité spécifiques en été et en hiver

L'équation ci-dessous est issue de la régression quantile sans FA liés aux consommations d'électricité spécifiques. Les données d'entrainement intègrent les incertitudes de mesure virtuelles ($N_r = 10$). La valeur entre parenthèses correspond à la part liée à l'incertitude statistique sur la $LGC_{5\%-ref}$ estimée par *bootstrap* (cf § 4.1).

$$LGC_{5\%} = 3{,}653(+0{,}012).10^3 - 3{,}34.10^2 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right) \quad \textbf{(5.8)}$$

$$+ 1{,}02.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)$$

$$+ 4{,}08.10^2 * \left(T_{m.ext.sf} - T_{m.ext.sf-ref}\right)$$

$$- 5{,}74.10^2 * \left(T_{mp.int.sf} - T_{mp.int.sf-ref}\right)$$

avec $T_{m.ext.sc-ref} = 10{,}36\ °C$, $T_{m.ext.sf-ref} = 23{,}10\ °C$, $T_{mp.int.sc-ref} = 22{,}18°C$, $T_{mp.int.sf-ref} = 24{,}51\ °C$, valeurs de référence qui correspondent aux médianes des distributions observées au sein des simulations. La forte influence des températures intérieures et extérieures est confirmée, la

variation d'un degré autour des valeurs de référence impactant le $LGC_{5\%}$ de l'ordre de 300 à 1000 kWh. L'écart entre les coefficients des températures moyennes extérieures et intérieures peut s'expliquer par des effets d'interaction de ces dernières. Ce modèle peut être contractualisé et propose une limite garantie des consommations de la PAC avec un risque de défaut de 5 %, ajustable en fonction de facteurs à mesurer sur les périodes décrites précédemment.

La question de la validité de la modélisation de la variabilité des usages de bureautique peut se poser étant donnée ce niveau de corrélation entre $E_{elec.sc}$ et $E_{elec.sf}$. Bien que ces FA dépendent conjointement du matériel informatique installé et du nombre d'usagers, on souhaite se prémunir d'éventuelles surconsommations inhabituelles d'électricité, qui pourraient entrainer une hausse des consommations de climatisation par exemple. Pour cela, on propose de considérer indépendamment les consommations de la PAC liés au chauffage et à la climatisation et de construire deux modèles de régression quantile indépendants, sommés dans l'équation ci-dessous.

$$\begin{aligned} LGC_{5\%} = \; & 2{,}004.10^3(+0{,}016.10^3) - 5{,}02.10^2 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right) \\ & + 7{,}49.10^2 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right) \\ & - 1{,}21.10^3 * \left(\frac{E_{elec.sc}}{E_{elec.sc-ref}} - 1\right) \\ & + 1{,}895.10^3 + \; 5{,}09.10^2 * \left(T_{m.ext.sf} - T_{m.ext.sf-ref}\right) \\ & - 4{,}32.10^2 * \left(T_{mp.int.sf} - T_{mp.int.sf-ref}\right) \\ & + 1{,}65.10^3 * \left(\frac{E_{elec.sf}}{E_{elec.sf-ref}} - 1\right) \end{aligned} \quad \textbf{(5.9)}$$

avec $T_{m.ext.sc-ref} = 10{,}36\,°C$, $T_{m.ext.sf-ref} = 23{,}10\,°C$, $T_{mp.int.sc-ref} = 22{,}18°C$, $T_{mp.int.sf-ref} = 24{,}51\,°C$, $E_{elec.sc-ref} = 4320\,kWhe$ et $E_{elecs.sf-ref} = 2345\,kWhe$. Ceci permet de prendre en compte l'influence spécifique de $E_{elec.sc}$ et de $E_{elec.sf}$ respectivement sur les consommations de chauffage et de climatisation. On remarque que la somme des deux valeurs de $LGC_{5\%}$ dans les conditions de référence ($2004 + 1895 = 3899\,kWh$) est supérieure à la valeur de l'équation précédente (3653 kWh). En effet, la dépendance entre consommations de chauffage et de climatisation, due notamment à celle sur les consommations d'électricité spécifiques, entraine un écart-type de la distribution totale de la PAC inférieur à la somme des écart-types des distributions indépendantes des consommations de chauffage et de climatisation (Tableau 5-16). On peut comparer ces valeurs de référence de la $LGC_{5\%}$ au quantile global de la distribution des consommations de la PAC qui est de 4081 kWh. Par ailleurs, les coefficients des autres FA sont légèrement modifiés par rapport à l'équation (5.8), avec en particulier une hausse de l'influence des températures extérieures. Ceci s'explique de manière similaire aux FA sur les consommations

électriques, par la faible corrélation existant entre $T_{m.ext.sc}$ et $T_{m.ext.sf}$, qui ont un effet inverse sur les consommations de la PAC, et dont l'influence est décorrélée dans l'équation (5.9).

5.2.5.2. Construction de modèles d'ajustement non linéaires

On étudie l'intérêt de prendre en compte un modèle non linéaire et d'ajouter des termes de second ordre pour l'amélioration de la précision du modèle de l'équation (5.8) et la diminution du niveau garanti. Une étude sur la régression *LASSO* (variables standardisées) est effectuée dont les résultats sont présentés sur la Figure 5-17. La partie gauche de la figure présente l'évolution des coefficients pour l'ensemble des variables en fonction de la pénalisation λ et la partie droite présente l'étude avec uniquement les variables les plus influentes. Les facteurs ${T_{mp.int.sc}}^2$, ${T_{mp.int.sf}}^2$, $T_{mp.int.sf} * T_{mp.int.sc}$, $T_{m.ext.sf} * T_{mp.int.sf}$ permettent d'expliquer une part non négligeable des variations de la grandeur d'intérêt.

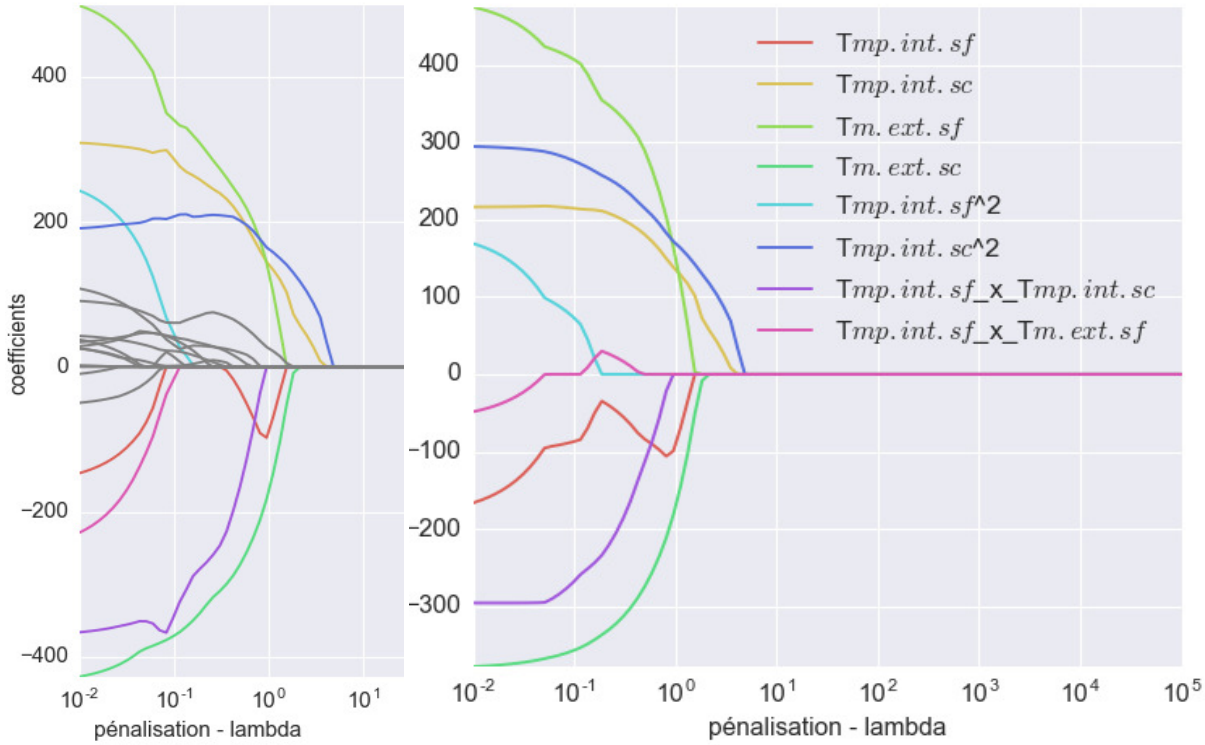

Figure 5-17: Régression LASSO sur les termes du second ordre les FA normalisés, complète (gauche) et uniquement avec les termes les plus influents (droite)

Un modèle de régression quantile est construit avec ces termes additionnels et l'équation obtenue est présentée ci-dessous. On constate que la valeur de $LGC_{5\%}$ dans les conditions de référence est de 3530 kWh pour 3653 kWh dans le modèle linéaire (équation 5.8). Le coefficient R^1 est de 0,419 ce qui constitue une augmentation significative par rapport au modèle linéaire de l'équation (5.8) dont le coefficient était de 0,385.

$$LGC_{5\%} = 3{,}530(+0{,}015).10^3 - 3{,}27.10^2 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right) \quad \textbf{(5.10)}$$

$$+ 1{,}11.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)$$

$$+ 4{,}07.10^2 * \left(T_{m.ext.sf} - T_{m.ext.sf-ref}\right)$$

$$- 5{,}98.10^2 * \left(T_{mp.int.sf} - T_{mp.int.sf-ref}\right)$$

$$+ 2{,}48.10^2 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)^2$$

$$+ 1{,}60.10^2 * \left(T_{mp.int.sf} - T_{mp.int.sf-ref}\right)^2$$

$$- 2{,}81.10^1 * \left(T_{m.ext.sf} - T_{m.ext.sf-ref}\right).\left(T_{mp.int.sf} - T_{mp.int.sf-ref}\right)$$

$$- 4{,}40.10^2 * \left(T_{mp.int.sf} - T_{mp.int.sf-ref}\right).\left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)$$

avec $T_{m.ext.sc-ref} = 10{,}36\ °C$, $T_{m.ext.sf-ref} = 23{,}10\ °C$, $T_{mp.int.sc-ref} = 22{,}18°C$, $T_{mp.int.sf-ref} = 24{,}51\ °C$. Malgré l'amélioration de la qualité statistique du modèle de régression quantile avec les termes polynomiaux, la validité physique de ces effets non linéaires est plus difficile à analyser et à vérifier. Comme discuté précédemment, la question des effets sur les bords des espaces de variation des facteurs d'ajustement se pose là encore, avec des aberrations physiques pour des valeurs très élevées ou très faibles des FA. Il est donc pertinent de borner le domaine de validité des modèles aux valeurs probables des FA, observables dans les résultats de simulation.

5.2.5.3. Bilan de l'engagement sur les consommations de la PAC

Cette étude, où une $\mathrm{LGC}_{5\%}$ des consommations annuelles de chauffage et de climatisation fournis par la pompe à chaleur est déterminée, permet plusieurs constats originaux.

Malgré la fourniture de chauffage et de refroidissement sur une partie importante de l'année, les FA liés aux températures intérieures et extérieures sur la période hivernale expliquent mieux la variabilité des consommations s'ils sont calculés sur une période relativement restreinte. En effet, il est plus intéressant de caractériser ce qui s'est passé sur la période froide la plus intense pour « comprendre » la tendance annuelle sur les consommations de chauffage.

On constate par ailleurs que l'effet des consommations électriques et des apports de chaleur associés est assez limité du fait de la corrélation entre les consommations électriques en période de chauffage et de refroidissement qui ont des effets opposés. Si l'on sépare la grandeur d'intérêt initiale, la consommation totale de la PAC, en consommations associées au chauffage et au refroidissement, on peut capturer l'influence de ces facteurs mais le niveau de consommations garanti est plus élevé.

Le modèle de régression quantile polynomial intégrant les effets non linéaires les plus importants permet d'obtenir un meilleur modèle d'ajustement, mais cette forme de modèle pose des questions de validité sur les bords des domaines de variabilité des FA.

5.2.6. Engagement de GPE sur les consommations électriques totales

Dans une seconde étude, on propose de définir un engagement sur la consommation énergétique totale. Tous les besoins du bâtiment sont fournis par l'énergie électrique. On s'intéresse donc aux consommations de chauffage, de climatisation, de distribution, d'ECS, de ventilation et d'électricité spécifique. L'ensemble de ces consommations est calculé dans les simulations SED réalisées avec le logiciel *Pléiades* selon les modalités décrites au § 5.2.2.1.

5.2.6.1. Ajustement lié aux différents postes de consommation

Une part importante de ces consommations dépend de la qualité de la mise en œuvre et du pilotage des systèmes énergétiques (consommations de ventilation, de distribution), peu impactés par les usagers. Aucun facteur d'ajustement lié à ces postes n'est donc intégrable au modèle. Il est à noter que les consommations de distribution sont corrélées aux consommations de la PAC et que les FA expliquant la variabilité des consommations de la PAC expliquent aussi une part de la variabilité des consommations de distribution.

Concernant les consommations d'énergie de l'appoint pour la production d'ECS, un ajustement sur les volumes d'eau puisés pourrait être envisagé. Cependant, ce poste dépend en partie du taux de couverture des besoins par le système solaire. Par ailleurs les consommations calculées sont très modestes (de l'ordre de 100 kWh/an, cf. Figure 5-14 et Tableau 5-16) et varient très peu. Le niveau des puisages explique donc statistiquement très mal la variation d'énergie totale, fortement impactée par les autres postes, et les incertitudes et variabilités qui influencent ces derniers. Pour ces raisons, on propose de ne pas intégrer de FA liés aux consommations d'ECS.

Les consommations d'éclairage dépendent en partie de l'occupation via la présence dans les pièces et l'utilisation des interrupteurs mais sont impactés par le mode de régulation et les conditions d'éclairement. Il est difficile d'envisager un facteur d'ajustement associé à la responsabilité et à l'impact des usages si sa gestion est automatisée. Dans le cas où les occupants ont la maitrise de celui-ci, on peut considérer qu'ils sont responsables de ce poste de consommations.

Comme montré dans le Tableau 5-16, les consommations d'électricité spécifique sont élevées (8656 kWh/an en moyenne) et largement distribuées (écart-type de 1657 kWh). Celles-ci sont principalement liées aux usages de bureautique et à la cuisine. Cette variation dans les résultats de simulation est issue de la modélisation d'une variabilité stochastique des usages, du temps d'utilisation et du matériel installé. La variabilité des consommations d'énergie totales est fortement

liée à ce poste, comme l'est donc le risque de surconsommation et de non respect d'un engagement sur les consommations globales. La prise en compte d'un ajustement lié à ce poste semble indispensable pour limiter le niveau d'engagement et l'on propose ainsi de considérer la consommation d'électricité spécifique comme un FA.

5.2.6.2. Prise en compte des consommations d'électricité spécifique dans le modèle d'ajustement

Les FA identifiés précédemment, qui expliquent une partie de la variabilité des consommations de la PAC, sont conservés ($T_{m.ext.sc_44_13}$, $T_{m.ext.sf_23_35}$, $T_{mp.int.sc_44_13}$ et $T_{mp.int.sf_23_35}$). On étudie l'intérêt de considérer les consommations électriques annuelles en incluant uniquement l'électricité spécifique (bureautique et cuisine) ou bien la somme de celle-ci et de l'éclairage.

Tableau 5-18 : Coefficients R^1 pour des modèles de régression quantile linéaire sur les consommations totales incluant des FA liés aux consommations électriques des usagers

Configuration		**Modèle RQL**		
	Facteurs d'ajustements supplémentaires	**coefficients**	$\boldsymbol{FA_{ref}}$	$\boldsymbol{R^1}$
a	-	-	-	0,193
b	$E_{elecSpé}$	9549	8507	0,742
c	$E_{elecSpé+Ecl}$	10677	10009	0,745

On confirme d'abord, l'intérêt, et la nécessité, d'ajuster l'engagement sur les consommations totales en fonction des consommations d'électricité spécifique. Le coefficient R^1 passe ainsi de 0,193 à 0,742/0,745. Ce constat est assez logique car ce poste représente plus de la moitié des consommations totales. La prise en compte de l'éclairage dans le FA associé aux consommations électriques des usagers permet d'obtenir un modèle légèrement plus précis.

En supposant que la performance liée à l'éclairage est à la charge de la structure contractante et doit être comprise dans la garantie sans être ajustée, on s'intéresse au modèle (b). Le modèle de $LGC_{5\%}$ associé à cette configuration d'ajustement est décrit dans l'équation ci-dessous. On constate une légère augmentation des coefficients liés aux températures intérieures et extérieurs par rapport aux résultats de l'équation (5.8). Ceci s'explique par le fait que ces FA impactent les consommations de chauffage et de climatisation auxquelles sont corrélées les consommations de distribution. Ces FA expliquent donc aussi une partie des variations de ce poste de consommation.

$$LGC_{5\%} = 16079(+16) - 3{,}23.10^2 * \left(T_{m.ext.sc} - T_{m.ext.sc-ref}\right) \quad \textbf{(5.11)}$$
$$+ 1{,}11.10^3 * \left(T_{mp.int.sc} - T_{mp.int.sc-ref}\right)$$
$$+ 4{,}90.10^2 * \left(T_{m.ext.sf} - T_{m.ext.sf-ref}\right)$$
$$- 3{,}22.10^2 * \left(T_{mp.int.sf} - T_{mp.int.sf-ref}\right)$$
$$+ 9{,}55.10^3 * \left(\frac{E_{elec}}{E_{elec-ref}} - 1\right)$$

avec $T_{m.ext.sc-ref} = 10{,}36\ °C$, $T_{m.ext.sf-ref} = 23{,}10\ °C$, $T_{mp.int.sc-ref} = 22{,}18°C$, $T_{mp.int.sf-ref} = 24{,}51\ °C$ et $E_{elec-ref} = 8507\ kWhe$. Le coefficient associé à la consommation d'électricité spécifique annuelle est supérieur à la valeur de référence de ce FA. Cela signifie qu'une augmentation de 10 kWh d'électricité spécifique entraine une augmentation supérieure de la $\mathrm{LGC}_{5\%}$. Ces consommations impactent en effet les besoins de chauffage et de climatisation et on constate que la surconsommation de climatisation est plus élevée que la sous-consommation liée à la baisse des besoins de chauffage. Consommer plus d'électricité spécifique entraine donc statistiquement une hausse du total des consommations des autres postes.

5.2.7. Vérification des engagements

5.2.7.1. Traitement et analyse des données

Des moyens de mesure et comptages, décrits au § 5.2.4, ont été installés afin de suivre les performances du bâtiment. Des acquisitions de données sont effectuées depuis avril 2015 et des capteurs complémentaires ont été mis en place courant mai 2016. Ainsi, deux années complètes de mesure sont exploitables afin de vérifier le niveau d'engagement de consommation annuelle proposé en phase de conception. La période de vérification est donc établie du 1er mai au 30 avril sur les années 2016-2017 et 2017-2018.

Les données récupérées sont échantillonnées avec un pas de temps de 15 minutes. Des mesures sont ponctuellement manquantes pour des périodes restreintes. Ces plages sont complétées par interpolation, pour des températures ou des données énergétiques cumulées. Il est à noter que ces post-traitements de données peuvent ajouter une part d'incertitude au calcul des consommations ou facteurs d'ajustement effectifs, qui, contrairement aux incertitudes de mesure, n'ont pas été prises en compte en amont du processus de GPE.

Les températures extérieures moyennes sont directement calculées à partir des acquisitions faites sur site. Les températures moyennes pondérées, facteurs d'ajustement choisis dans l'étude

précédente, sont estimés à partir de la moyenne des 8 températures relevées sur le bâtiment, en faisant l'hypothèse qu'une surface d'un huitième (un quart de niveau) du bâtiment est affectée à chaque mesure. Cette hypothèse est acceptable étant donnée la répartition spatiale équilibrée des capteurs. Elle impacte aussi le calcul de l'incertitude de mesure au sein de laquelle sont tirées les erreurs de mesure virtuelle (annexe A-5-6).

Par ailleurs, les informations issues des compteurs électriques différenciés ne permettent pas d'isoler les consommations de distribution, du compteur général. Celles-ci sont donc estimées à partir des informations et hypothèses suivantes :

- la pompe du circuit amont (sonde géothermique) de la PAC constitue l'essentiel de ces consommations.
- La puissance de la pompe est de 130 W.
- Les heures de fonctionnement de la pompe sont obtenues à partir des débits mesurés et sont respectivement de 7312 h et 8219 h sur les deux années de vérification. On constate ainsi que celle-ci fonctionne une très grande partie de l'année, même à faible charge.

Le Tableau 5-19 présente l'ensemble des résultats de mesure provenant du traitement des données relevées sur site. Pour rappel, la saison de calcul pour le chauffage (*sc*) est comprise entre les semaines 44 et 13, et la saison de calcul pour la climatisation (*sf*) entre les semaines 23 et 35.

Tableau 5-19 : Mesures sur les deux années d'exploitation du bâtiment

FA et grandeurs d'intérêt	***2016-2017***	***2017-2018***
$T_{m.ext.sc}$ *(°C)*	*9,50*	*8,40*
$T_{m.ext.sf}$ *(°C)*	*26,07*	*25,53*
$T_{mp.int.sc}$ *(°C)*	*21,92*	*22,01*
$T_{mp.int.sf}$ *(°C)*	*25,18*	*25,55*
$E_{elec.sc}$ *(kWh)*	*3041*	*3085*
$E_{elec.sf}$ *(kWh)*	*1658*	*1517*
Consommation PAC (kWh)	*3409*	*3585*
Consommation Distribution (kWh)	*951*	*1068*
Consommation Éclairage (kWh)	*497*	*521*
Consommation Ventilation (kWh)	*551*	*524*
Consommation ECS (kWh)	*95*	*73*
Consommation Électricité spécifique (kWh)	*6390*	*6397*
Consommation totale (kWh)	*11892*	*12168*

Les mesures des différents postes de consommation peuvent être comparées avec les résultats initiaux de simulation et les distributions obtenues sans ajustement, décrites dans le Tableau 5-16. Les consommations de la PAC sont dans la partie haute de la distribution des résultats de simulation.

Les estimations des consommations de distribution et d'eau chaude sanitaire sont très proches de la réalité. Les consommations d'éclairage et de ventilation étaient largement surévaluées. Concernant la ventilation, cela peut s'expliquer par la mauvaise spécification de la puissance nominale des ventilateurs. Les paramètres des équipements d'éclairage correspondent au matériel installé, mais on peut imaginer que les niveaux d'éclairement à respecter, imposant l'utilisation de l'éclairage en simulation, ont été surévalués, et que l'éclairage a été moins utilisé. Les consommations d'électricité spécifiques, très étalées dans les résultats de simulation sont en réalité assez faibles, de l'ordre de la moyenne moins l'écart-type de la distribution. Les consommations globales sont inférieures à la moyenne des prévisions. La Figure 5-18 présente une comparaison des résultats de simulation sans ajustement et des mesures sur les deux années de vérification.

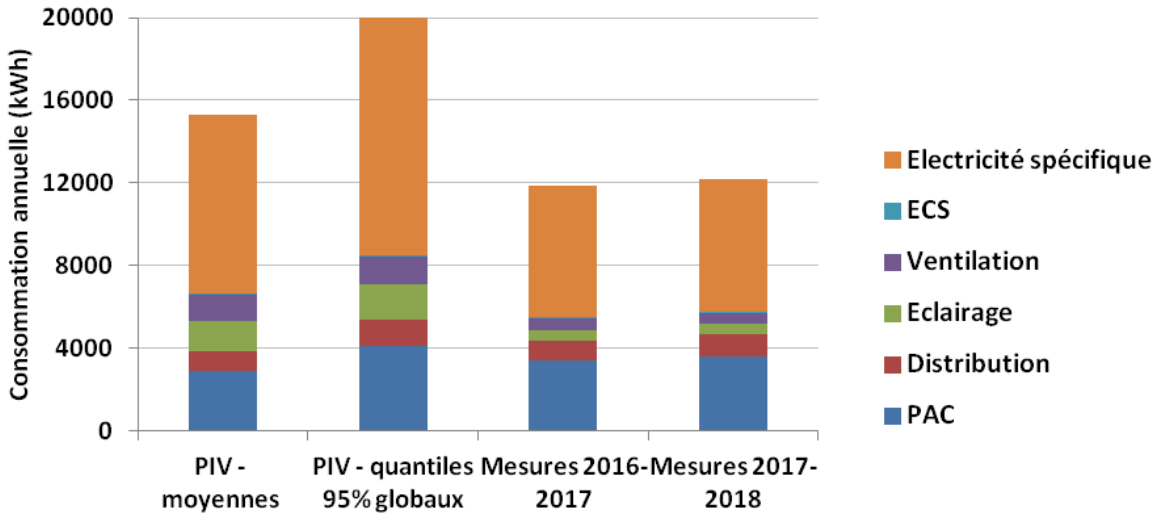

Figure 5-18: Comparaison des moyennes et quantiles globaux des simulations et des mesures pour les différents postes de consommation énergétique

5.2.7.2. Vérification de l'engagement sur les consommations de chauffage et de climatisation

Ces mesures permettent de vérifier l'engagement sur les consommations de la PAC exprimés précédemment selon trois modèles différents de $LGC_{5\%}$: un premier modèle linéaire (équation 5.8) où l'électricité spécifique n'est pas prise en compte, un second modèle linéaire (équation 5.9) construit à partir de deux régressions quantiles distinctes et incluant l'impact estival et hivernal des consommations d'électricité spécifique, et un modèle obtenu par régression quantile polynomiale, sans ajustement en fonction des consommations d'électricité spécifique (équation 5.10). Les valeurs ajustées des limites garanties sont calculées et présentées dans le Tableau 5-20 pour les différents modèles et sur les deux années d'exploitation.

Tableau 5-20 : Ajustement et vérification de l'engagement sur les consommations de la PAC sur les deux années d'exploitation pour les trois modèles proposés

$LGC_{5\%}(kWh)$								
Modèle RQL (5.8)			***Modèle RQL (5.9)***			***Modèle RQP (5.10)***		
Réf.	*Ajustement*		*Réf.*	*Ajustement*		*Réf.*	*Ajustement*	
	2016-2017	*2017-2018*		*2016-2017*	*2016-2017*		*2016-2017*	*2016-2017*
3665	***4520***	***4553***	*3915*	***5246***	***5311***	*3545*	***4451***	***4543***

Mesures (kWh)	
2016-2017	***2017-2018***
3409	***3585***

On constate globalement des ajustements à la hausse par rapport aux conditions de référence. Ceci est dû principalement, à la mesure de $T_{m.ext.sf}$ plus élevée que les valeurs observées en simulation et de $T_{m.ext.sc}$ plus faibles. L'ajustement en fonction des températures intérieures sur les deux saisons a tendance à faire diminuer le niveau garanti.

Les mesures de $E_{elec.sc}$ et $E_{elec.sf}$ sont inférieures aux valeurs médianes issues des simulations (d'un tiers environ). Pour le modèle RQL (5.9), ces mesures provoquent donc respectivement un ajustement à la hausse et à la baisse de la $LGC_{5\%}$ se compensant en partie. En prolongement des interrogations du § 5.2.5.1 sur la modélisation des consommations d'électricité spécifique et de la corrélation entre leur niveau hivernal et estival, la Figure 5-19 présente une comparaison des valeurs de $E_{elec.sc}$ et $E_{elec.sf}$ issues des simulations et mesurées. Comme vu précédemment, les valeurs réelles sont dans la plage basse des résultats issus de la modélisation, mais le rapport entre les valeurs en saisons de chauffe et de climatisation est en adéquation avec les simulations. La prise en comte de ces FA dans le modèle RQL (5.9) n'a pas d'impact important pour ces deux années de mesure.

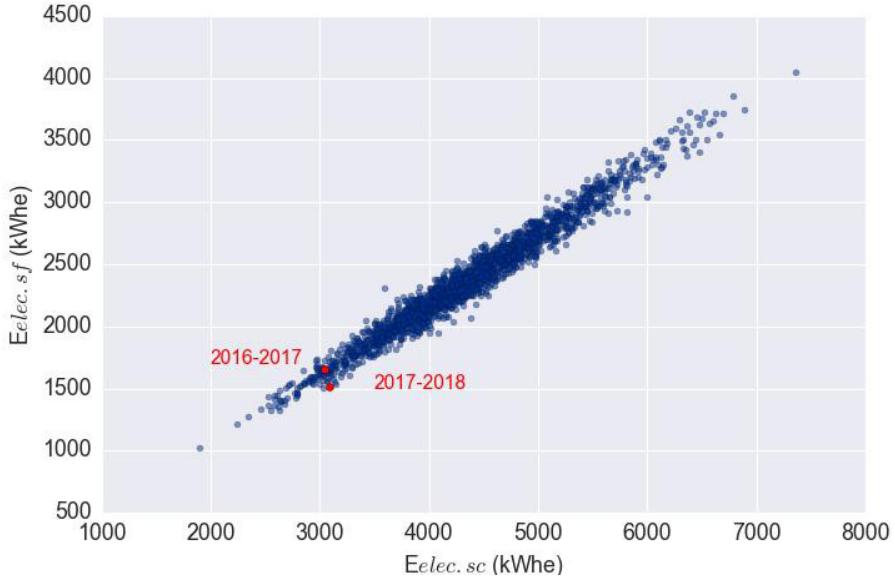

Figure 5-19: Consommations d'électricité spécifique des périodes de chauffe et de climatisation mesurées (rouge) et simulées (bleu)

La comparaison des limites garanties ajustées et des consommations mesurées de la PAC permet de valider l'engagement, quel que soit le modèle contractualisé en amont.

5.2.7.3. Vérification de l'engagement sur les consommations totales

La $LGC_{5\%}$ sur les consommations annuelles globales du bâtiment est ajustée à partir de l'équation (5.11) et des mesures présentées dans le Tableau 5-19. Les valeurs ajustées sont décrites dans le Tableau 5-21. Du fait de la mesure de consommations d'électricité spécifique peu élevées en regard des prévisions, les limites garanties des deux années sont ajustées à la baisse. L'ajustement en fonction des températures extérieures et intérieures moyennes entrainent des effets similaires (et accentués) à ceux décrits précédemment. Les limites garanties ajustées sont supérieures aux consommations totales réelles de 2500 kWh, soit 20 % environ. L'engagement est donc respecté sur les deux périodes.

Tableau 5-21 : Ajustement et vérification de l'engagement (équation 5.11) sur les consommations globales sur les deux années d'exploitation

$LGC_{5\%}(kWh)$			*Mesures (kWh)*	
Référence	*Ajustement*			
	2016-2017	*2017-2018*	*2016-2017*	*2017-2018*
16095	***14670***	***14598***	***11892***	***12168***

5.2.7.4. Retour sur les processus de GPE

Il a été constaté que les différents engagements (à un niveau de risque de 5 %) présentés, sur la consommation de la PAC et sur la consommation globale, sont largement respectés. Ainsi, il est intéressant d'étudier les espérances conditionnelles exprimées en fonction des mêmes jeux de FA. Ces modèles correspondent à une $LGC_{50\%}$, où la valeur obtenue est donc centrée sur l'incertitude résiduelle.

Pour la consommation de la PAC, les modèles d'espérance conditionnelle équivalents aux configurations d'ajustement des modèles RQL (5.8) et (5.9) sont présentés en annexe A-5-8. Ils fournissent des valeurs ajustées, en fonction des conditions observées sur les deux périodes, de 3720 et 3812 kWh pour la configuration (5.8) et de 4175 et 4240 kWh pour le modèle (5.9). Les consommations de la PAC mesurées sont donc inférieures à ces espérances conditionnelles ajustées. Concernant la consommation totale, un modèle de régression sur l'espérance avec la configuration d'ajustement de l'équation (5.11) est décrit dans l'annexe A-5-8. Les espérances conditionnelles ajustées sont de 13797 kWh pour l'année 2016-2017 et 13829 kWh pour 2017-2018. Là-aussi ces valeurs sont supérieures aux mesures (+16 % / + 13 %). Ces résultats sont présentés sur la Figure 5-20.

Le modèle physico-probabiliste est donc globalement décentré, évaluant des consommations trop élevées. Les hypothèses retenues sur les fonctions de densité de probabilité des paramètres, et notamment leur moyenne, sont trop pessimistes. Des lois décentrées par rapport aux valeurs nominales ont en particulier été considérées pour le rendement de la PAC, les débits de surventilation, les scénarios d'occultation, et peuvent en partie expliquer ce constat. L'approche conservatrice est cependant à privilégier dans une démarche de garantie, et permet ici de respecter les engagements déterminés dans l'étude amont.

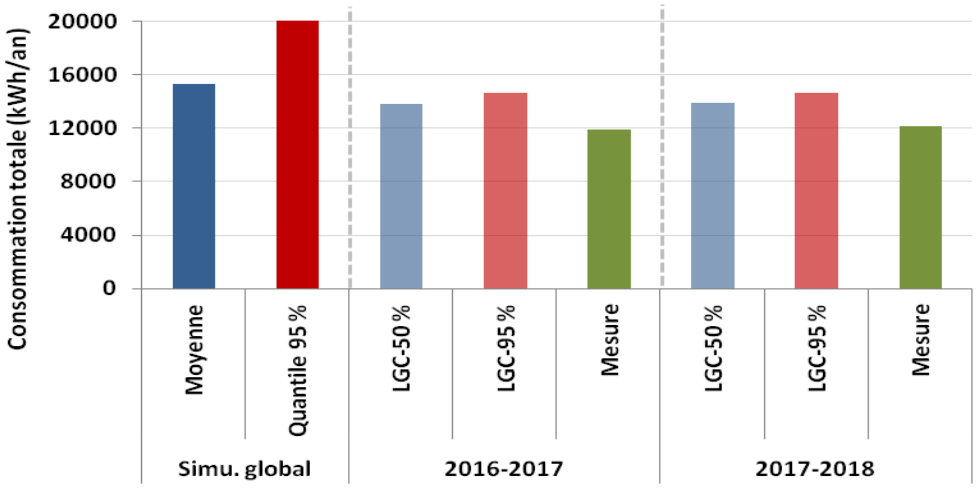

Figure 5-20: Impact du processus d'ajustement (de l'espérance et du quantile à 95 %) sur la prévision des consommations totales et comparaison aux mesures

5.2.8. Bilan du cas d'étude d'Izuba

L'originalité de ce second cas test réside dans l'usage tertiaire du bâtiment étudié, dans sa très bonne performance énergétique et dans la complexité des systèmes HVAC et donc de leur modélisation. Les faibles besoins et consommations liés aux postes conventionnels entrainent ainsi une forte influence des modes d'usage sur les consommations énergétiques. D'autres aspects spécifiques à ce projet sont la prise en compte de multiples postes de consommations dans la garantie et le travail de collecte et d'analyse de données de mesures réalisées sur site.

Un fichier de simulation SED correspondant à la phase de conception du bâtiment a été transmis par l'entreprise Izuba Energie et a permis de constituer un modèle physico-probabiliste complet selon l'approche décrite dans cette thèse. La description simplifiée de plages d'incertitudes sur les paramètres physiques et techniques de l'enveloppe et des systèmes a d'abord permis de conduire une analyse de sensibilité des consommations de chauffage et de climatisation. Les paramètres influents ont ensuite été décrits de manière plus poussés. Le rendement de la PAC, les scénarios de chauffage, de climatisation, d'occultation et de surventilation, mais aussi les caractéristiques des vitrages et de l'isolant en paille, sont les entrées incertaines les plus influentes. Le modèle de génération stochastique de données associées à l'occupation et aux usages tertiaires (apports métaboliques, puisages ECS, consommations spécifiques) a été consolidé et exploité. Les données types RT2012 de Nice ont été le support à la génération de données météorologiques variables.

La propagation des incertitudes et variabilités a fourni un jeu de 2000 simulations exploitées pour proposer des engagements de GPE sur deux périmètres : les consommations annuelles de chauffage et de climatisation associées à la PAC, puis les consommations annuelles globales du bâtiment. Afin d'ajuster la limite garantie des consommations de la PAC, des températures intérieures et extérieures moyennes ont été considérées sur des périodes estivales et hivernales optimisées pour expliquer au mieux la variabilité des consommations et fournir les facteurs d'ajustement les plus performants. Une corrélation négative entre consommations de chauffage et de climatisation a été observée du fait du fort lien entre les consommations spécifiques hivernales et estivales fournies par le modèle de génération stochastique. Des modèles couplés et découplés de $LGC_{5\%}$ ont été envisagés, ainsi qu'un modèle couplé intégrant des termes du second ordre. Le modèle découplé surestime les consommations. Le modèle polynomial explique mieux la variabilité de la grandeur d'intérêt mais pose des questions de validité sur le bord des plages des facteurs d'ajustement. L'engagement sur la consommation totale intègre lui un ajustement en fonction des consommations d'électricité spécifique. L'anticipation de la mesure des FA a permis de déterminer des incertitudes de mesure. Dix tirages d'erreurs de mesure virtuelle ont été associés à chaque réalisation de simulation, fournissant finalement un ensemble statistique de 20 000 points à analyser.

L'analyse des données récoltées sur la GTC du bâtiment depuis deux ans et demi a permis d'estimer les valeurs effectives des FA et des consommations d'intérêt sur deux périodes annuelles sur lesquelles les engagements peuvent être vérifiés. Les limites garanties excèdent significativement les consommations mesurées, pour tous les modèles et périodes. Une analyse complémentaire de l'incertitude résiduelle via l'expression de modèles d'espérance conditionnelle ajustable montre que la modélisation physico-probabiliste initiale était plutôt conservatrice et pessimiste. Par ailleurs les deux années étudiées présentent des caractéristiques assez proches pour tous les facteurs d'ajustement considérés et il serait intéressant de réitérer la vérification sur les années à venir.

En perspective du travail réalisé sur ce cas d'étude, il aurait été intéressant d'envisager un engagement sur le bilan global d'énergie primaire qui inclut la production d'énergie photovoltaïque sur le bâtiment. Un ajustement en fonction des jours d'ensoleillement, ou du rayonnement solaire cumulé, aurait par exemple pu être pris en compte.

Conclusion générale

Bilan du travail effectué

Les enjeux environnementaux actuels poussent à la mise en place de mécanismes d'engagement et de vérification des performances énergétiques effectives des bâtiments. L'ambition supposée de projets de construction et de réhabilitation doit alors être validée en exploitation. La garantie de performance énergétique (GPE) a vocation à répondre à ces enjeux, elle ne doit pas être perçue comme une responsabilité contraignante mais comme un levier qui peut sécuriser les investissements et encourager la multiplication des projets énergétiquement vertueux. Dans ce contexte et conformément aux objectifs présentés, cette thèse s'est intéressée au développement d'une méthodologie pour la mise en œuvre de la garantie de performance énergétique en s'appuyant sur l'exploitation des outils de simulation et sur l'anticipation de mesures lors de l'exploitation du bâtiment. Cette méthodologie doit permettre la détermination d'une limite garantie de consommation énergétique, associée à un risque de défaut quantifié, et doit isoler l'influence de facteurs extérieurs, non liés aux qualités intrinsèques du bâtiment et de ses systèmes énergétiques.

Un état de l'art s'est tout d'abord intéressé aux travaux de la littérature associés aux principales thématiques liées à ces objectifs, avec en premier lieu les méthodes liées aux simulations énergétiques dynamiques. Les analyses de sensibilité et d'incertitude ont pour but la connaissance élargie du bâtiment modélisé. Elles informent respectivement sur les paramètres d'entrée dont l'incertitude peut provoquer une variabilité importante des grandeurs d'intérêt de sortie, et sur la variabilité globale de ces grandeurs d'intérêt induite par le contexte incertain des entrées. Par ailleurs, les conditions aux limites des modèles, liées aux sollicitations dynamiques intérieures et extérieures, impactent fortement les consommations énergétiques simulées. Dans l'objectif de prévision qui est le nôtre, la modélisation et la prise en compte de ces conditions, par nature variables et imprévisibles, est un verrou majeur pour la détermination d'une limite garantie de consommation. Ces sollicitations correspondent aux données météorologiques et aux données d'usage. L'approche stochastique a été identifiée comme pertinente pour intégrer ces phénomènes. Elle vise à créer des données réalistes variables, dont la variabilité est elle-même réaliste. Ces données pseudo-aléatoires sont fournies en entrée des simulations énergétiques et permettent d'inférer la distribution des sorties induite par la variabilité de ces sollicitations. Le terme de propagation de la variabilité a ainsi été introduit. Cette distinction entre incertitude et variabilité résulte de la nature différente des paramètres d'entrée auxquels ces termes se réfèrent. Le premier concerne les paramètres statiques, liés à l'enveloppe et aux systèmes, sur lesquels se pose la question de la qualité dans l'optique d'une GPE. Ces incertitudes sont décrites avec des fonctions de densité de probabilité et peuvent être réduites par une meilleure prévision de la mise en œuvre. La

variabilité se rattache à des paramètres dynamiques qui seront par nature changeants au cours de l'exploitation et qui ne sont pas liés à la qualité du bâtiment et des systèmes. On doit tenter d'isoler leur influence dans la détermination d'un engagement de GPE. La méthodologie repose ainsi sur la propagation des incertitudes et variabilités au sein du modèle physico-probabiliste du bâtiment étudié. Celui-ci désigne l'inclusion, autour du modèle physique de SED, des méthodes fournissant une variabilité stochastique des sollicitations, et des lois de distribution sur les paramètres incertains. Il s'agit ensuite d'étudier les résultats de simulations, non seulement pour caractériser la variabilité absolue de la grandeur d'intérêt mais sa variabilité contextualisée, conjointe à celle d'autres facteurs.

En parallèle à la formalisation de cette approche de simulation, l'étude des retours d'expérience sur la vérification des performances effectives et sur les méthodes pour la GPE a permis d'identifier plusieurs limitations. Une part des travaux existants repose sur la mise à jour d'un fichier SED à partir de données mesurées en exploitation afin de vérifier l'adéquation entre consommations mesurées et résultats de simulation. Cette approche peu transparente nécessite une intervention experte a posteriori et ne permet pas aisément de projeter en amont une valeur limite de consommation. Les problématiques de la prévision d'une valeur garantie, de l'ajustement en fonction de facteurs externes et de la mesure sont par ailleurs considérées de manière dissociée alors qu'il a été montré que ces éléments sont interdépendants. La méthodologie développée s'est donc appliquée à prendre en compter ces interactions et à construire un ensemble cohérent scientifiquement pour la mise en œuvre de la GPE.

Avant de présenter les développements méthodologiques, une étude, présentée au chapitre 2, s'est concentrée sur la réalisation d'un modèle de génération stochastique de données météorologiques pour la SED. Ce modèle est une brique nécessaire au modèle-physico-probabiliste décrit ci-dessus. Ces travaux s'expliquent par l'identification de manques parmi les outils disponibles et par l'opportunité d'améliorer les modèles de ce type déjà existants. Des données horaires de température extérieure, de rayonnement global horizontal et de rayonnement diffus horizontal sont générées, à partir d'un seul jeu de données types, et sur une année entière. Ce processus repose sur l'identification au sein des données d'entrainement de parts tendancielles, déterministes, et de parts résiduelles, pseudo-aléatoires. La génération stochastique d'autres parts résiduelles permet d'introduire la variabilité souhaitée. Les apports du modèle concernent différents points. L'intercorrélation température/rayonnement est considérée non seulement au sein des parts tendancielles des séries mais aussi au sein des parts résiduelles aléatoires. L'étude complémentaire d'un ensemble de données réelles a aussi permis une meilleure description de la variabilité de ces grandeurs. Celle-ci est introduite par la prise en compte d'une variation des caractéristiques des séries tendancielles. Les données météorologiques variables et réalistes produites par ce modèle sont exploitées dans les propagations d'incertitudes et variabilités (PIV) nécessaires à la méthodologie pour la GPE.

Le chapitre 3 décrit la méthodologie développée. Celle-ci repose sur l'exploitation statistique des résultats de simulation issus de la PIV, et sur l'étude des variations conjointes de la consommation énergétique d'intérêt et de variables explicatives, appelées facteurs d'ajustement. Les méthodes de régression quantile permettent de déterminer conjointement un niveau de consommation associé à un risque statistique de dépassement et des modalités d'ajustement de cette valeur en fonction des FA. Les modèles linéaires qui fournissent le quantile conditionnel de la grandeur d'intérêt présentent des pentes différentes de la régression linéaire classique qui calcule l'espérance conditionnelle. Ceci confirme la nécessité de traiter simultanément les problématiques de l'étude de l'influence des incertitudes et de l'ajustement en fonction des variabilités.

L'incertitude liée à la mesure des sollicitations et des consommations énergétiques en exploitation lors de la phase de vérification introduit un risque supplémentaire de défaut de l'engagement, non pris en compte lors de l'étude amont. Le mode d'intégration de cette influence des incertitudes de mesure au sein de la méthodologie constitue un des apports originaux du travail présenté. La simulation physico-probabiliste vise à représenter le comportement énergétique réel d'un bâtiment dans une variabilité statistique probable de conditions d'exploitations. En faisant le constat que la mesure des conditions effectives ne nous donne pas accès à la réalité de ce comportement mais à une estimation composée d'une erreur de mesure, on propose de considérer, en amont lors de notre étude, des erreurs de mesure virtuelles venant entacher les résultats de simulation. Ainsi, il y a équivalence entre le modèle probabiliste et la mesure en condition réelle, cette dernière pouvant être considérée comme une réalisation statistique du modèle. L'exploitation statistique s'applique ainsi sur les résultats de simulation bruités par des erreurs de mesure tirées aléatoirement dans des plages d'incertitude associées à la prévision d'un plan de mesure. Les spécificités des projets de construction ou de réhabilitation ont ensuite été décrites et étudiées spécifiquement. Dans le second cas, il est nécessaire de caractériser la situation avant les travaux avec un modèle de référence dépendant du jeu de facteurs d'ajustement considéré.

Différents développements ont été présentés dans le chapitre 4 afin de fournir des clés pour la mise en œuvre concrète de cette méthodologie. La robustesse statistique de la méthode de régression quantile linéaire a été étudiée. Une méthode de *bootstrap* permet d'évaluer la variance statistique des estimateurs du modèle. La prise en compte d'une incertitude statistique sur le coefficient lié à l'ordonnée à l'origine permet de sécuriser l'expression du quantile conditionnel. L'influence des incertitudes de mesure des consommations et des facteurs d'ajustement est ensuite étudiée, permettant la détermination d'un plan de mesure et vérification optimisé. Différentes méthodes statistiques ont été comparées afin d'identifier les meilleurs facteurs d'ajustement, expliquant les variations de la grandeur d'intérêt et minimisant sa variabilité résiduelle. Il s'agit d'expliquer au mieux le contexte d'exploitation pour réduire l'incertitude résiduelle pouvant provoquer des surconsommations d'énergie. Enfin, des formes alternatives de modèles pour l'expression du quantile conditionnel ont été étudiées. Des modèles non-linéaires qui prennent en compte les interactions ont notamment été testés.

Enfin, le chapitre final propose deux applications : un projet de réhabilitation d'un bâtiment de logements et la construction d'un bâtiment de bureaux à très haute performance énergétique. Sur le premier cas d'étude, un engagement sur la diminution des consommations de chauffage est mis en place. Sur le second, une limite garantie des consommations de la PAC (chauffage et climatisation) puis des consommations totales sont déterminées. Dans les deux projets, les modèles contractuels identifiés sont confrontés aux mesures réalisées lors de l'exploitation des bâtiments. La prise en compte des températures intérieures, liées aux consignes, sur les périodes de besoin de chauffage et de climatisation permet de s'affranchir des risques de surconsommation liés à ces facteurs. L'impact de la température extérieure doit aussi être intégré dans le modèle d'ajustement via des facteurs de type degrés-jours-unifiés ou degrés-heures ou via des moyennes de température sur des périodes clés. Enfin, les consommations d'électricité spécifique, fortement dépendantes des usages, peuvent être l'objet d'un ajustement pour isoler leur influence en tant qu'apport de chaleur ou lorsque les consommations totales sont garanties. Dans les deux applications, les valeurs garanties exprimées ont été respectées après le processus de vérification et l'ajustement en fonction des conditions réelles.

Réflexions et perspectives

Un des constats associés à ces travaux est la grande dépendance du niveau d'engagement à la qualité de la modélisation des incertitudes en amont. Les valeurs nominales et les lois qui décrivent l'incertitude sur les entrées sont difficiles à déterminer. Quelques travaux de recherche traitent de cette problématique centrale. Un dialogue avec les différents métiers en lien avec la réalisation des travaux est nécessaire pour définir au mieux ces incertitudes et c'est un point clé d'amélioration. L'utilisation des analyses de sensibilité couplée à une approche conservatrice et pessimiste peut permettre d'éviter les erreurs de modélisation préjudiciables à la définition d'un engagement fiable.

Une autre piste d'amélioration concerne le paramétrage du nombre de simulations nécessaires à la bonne convergence des modèles établis en post-traitements. Un seuil de variance statistique des coefficients du modèle de régression pourrait être établi afin d'assurer une bonne stabilité de celui-ci. Le problème réside dans le caractère itératif de la méthodologie. En fonction de facteurs d'ajustement choisis et de leur caractère explicatif, plus ou moins de simulations sont nécessaires. Des bouclages du processus pourraient être effectués en plusieurs endroits, pour tester d'autres FA, d'autres plages d'incertitudes de mesure virtuelle, d'autres types de modèles ou seuils de risques. Il peut par exemple être intéressant d'avoir la souplesse de compléter la PIV pour s'assurer du niveau de convergence des nouvelles solutions étudiées. Il faut aussi avoir à l'esprit qu'une partie des informations sont calculées pendant les simulations, notamment les FA, et que l'ajout de nouveaux facteurs nécessite l'implémentation de leur calcul au fil de l'eau de simulations énergétiques.

Par ailleurs, la question de la validation de ce type de méthodes probabilistes doit être posée. La progression globale possède une logique et les principales briques méthodologiques, méthodes

statistiques et modèle de simulation énergétique dynamique, sont validées indépendamment. La description probabiliste des entrées statiques et dynamiques est plus questionnable. Seuls de nombreux retours d'expérience peuvent permettre d'avoir du recul. Il serait idéal de tester la méthodologie sur des projets où un grand nombre de bâtiments semblables sont instrumentés et étudiés sur plusieurs années. L'influence des usages, des conditions météorologiques et les problématiques de mise en œuvre et d'exploitation pourraient être mieux différenciées.

La méthodologie développée n'a pas vocation à être une réponse ultime à la problématique traitée. Elle propose un cadre logique et articule différentes briques de modèles et de méthodes afin de définir des modèles d'engagement de GPE les plus fiables possibles, une prise en compte la plus exhaustive possible des incertitudes et une quantification des risques de défaut de la garantie. Chacune des briques utilisées peut et doit être améliorée, de la définition des incertitudes des paramètres statiques à la description de la variabilité aux modèles de simulation des bâtiments et des systèmes. L'utilisation de modèles générant une variabilité d'usages souples et paramétrables, adaptés à chaque projet constitue un des défis pour l'avenir. Une meilleure prise en compte de l'interdépendance des modèles (enveloppe, sollicitations et systèmes) est possible.

La contractualisation d'une GPE doit aussi s'accompagner d'un processus de commissionnement, de surveillance des dérives en exploitation. En cela, la méthodologie est complémentaire d'autres travaux, liés à l'exploitation continue des données de mesure, que ce soit pour la construction ou la mise à jour de modèles pour le pilotage ou l'identification de malfaçons par calibrage. Des interactions sont aussi à prévoir avec les développements traitant de la compréhension, la modélisation et la prévision des usages à partir de données d'exploitation.

La problématique du vieillissement n'a pas été directement abordée dans ces travaux. La détérioration de l'enveloppe et des systèmes ne risque-t-elle pas d'entrainer des surconsommations non prises en compte dans la méthodologie ? Cette question rejoint celle traitée ci-dessus en lien avec le commissionnement. On a supposé que le temps d'application de la garantie était inférieur à celui correspondant à l'apparition de défauts dus au vieillissement. Une mise à jour du contrat peut cependant être envisagée avec une redéfinition du modèle physico-probabiliste. La prise en compte de ces aspects nécessiterait à la fois l'utilisation de modèles peu fiables et la considération d'une temporalité, d'une évolution de l'engagement au fil du temps, complexifiant significativement le processus.

Il pourrait enfin être intéressant d'envisager l'intégration de la méthodologie dans des travaux plus globaux. On s'est limité ici à une contractualisation concernant la consommation énergétique qui permet d'anticiper les consommations d'énergie, ou la diminution de ces consommations. Grâce à cette anticipation, il est possible d'établir des temps de retour sur investissement L'intégration temporelle de ces méthodes peut aussi permettre un foisonnement des effets des conditions météorologiques. La méthode pourrait être associée à des modèles financiers, d'évolution des prix de l'énergie par exemple. Des évolutions des services en énergétique pour les bâtiments sont en

réflexion, avec notamment l'idée de proposer un service de prise en charge des consommations énergétiques de manière forfaitaire par un prestataire. Ce type d'approches probabilistes peut nourrir le développement de ces solutions, par la prévision des consommations et donc des prix du service, mais aussi par l'évaluation de l'intérêt énergétique et financier de planifier des actions d'amélioration de la performance énergétique.

Bibliographie

ADEME et FBE, 2016. *Méthodes et outils de la garantie de résultats énergétiques*. Le Moniteur. S.l. : s.n.

Aguiar, R. et M. Collares-Pereira, 1992. TAG: a time-dependent, autoregressive, Gaussian model for generating synthetic hourly radiation. In : *Solar energy*. 1992. Vol. 49, n° 3, p. 167–174.

Aguiar, R. J., M. Collares-Pereira et J. P. Conde, 1988. Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices. In : *Solar Energy*. 1988. Vol. 40, n° 3, p. 269–279.

Ahmad, M. W., M. Mourshed, D. Mundow, M. Sisinni et Y. Rezgui, 2016. Building energy metering and environmental monitoring – A state-of-the-art review and directions for future research. In : *Energy and Buildings*. mai 2016. Vol. 120, p. 85‑102. DOI 10.1016/j.enbuild.2016.03.059.

Amayri, M., A. Arora, S. Ploix, S. Bandhyopadyay, Q.-D. Ngo et V. R. Badarla, 2016. Estimating occupancy in heterogeneous sensor environment. In : *Energy and Buildings*. octobre 2016. Vol. 129, p. 46‑58. DOI 10.1016/j.enbuild.2016.07.026.

Amayri, M., Q.-D. Ngo, E. A. E. Safadi et S. Ploix, 2017. Bayesian network and Hidden Markov Model for estimating occupancy from measurements and knowledge. In : [en ligne]. S.l. : IEEE. septembre 2017. p. 690‑695. [Consulté le 29 juin 2018]. Disponible à l'adresse : http://ieeexplore.ieee.org/document/8095179/.

Andersen, R. K., 2012. The influence of occupants' behaviour on energy consumption investigated in 290 identical dwellings and in 35 apartments. In : *10th International conference on healthy buildings* [en ligne]. S.l. : s.n. 2012. [Consulté le 4 mai 2016]. Disponible à l'adresse : http://orbit.dtu.dk/fedora/objects/orbit:113412/datastreams/file_9903893/content.

Bauwens, G. et S. Roels, 2014. Co-heating test: A state-of-the-art. In : *Energy and Buildings*. octobre 2014. Vol. 82, p. 163‑172. DOI 10.1016/j.enbuild.2014.04.039.

Bettonvil, B. et J. P. C. Kleijnen, 1997. Searching for important factors in simulation models with many factors: Sequential bifurcation. In : *European Journal of Operational Research*. 10 janvier 1997. Vol. 96, n° 1, p. 180‑194. DOI 10.1016/S0377-2217(96)00156-7.

Bofinger, E., 1975. Estimation of a density function using order statistics. In : *Australian Journal of Statistics*. 1975. p. 1‑7.

Boland, J., 1995. Time-series analysis of climatic variables. In : *Solar Energy*. 1995. Vol. 55, n° 5, p. 377‑388.

Bourdin, V., 2014. *Calculs astronomiques simplifiés*. 2014. S.l. : Université Paris-Sud Orsay – Master PAM.

Brohus, H., P. Heiselberg, A. Simonsen et K. C. Sorensen, 2009. Uncertainty of energy consumption assessment of domestic buildings. In : *Proceedings of BS2009*. Glasgow Scotalnd : s.n. 2009.

Brun, A., C. Spitz, E. Wurtz et L. Mora, 2009. Behavioural comparison of some predictive tools used in a low-energy building. In : *Eleventh International IBPSA Conference*. Glasgow, Scotland, 27-30 July 2009 : s.n. 27 juillet 2009. p. 27–30.

Burkhart, M. C., Y. Heo et V. M. Zavala, 2014. Measurement and verification of building systems under uncertain data: A Gaussian process modeling approach. In : *Energy and Buildings*. juin 2014. Vol. 75, p. 189‑198. DOI 10.1016/j.enbuild.2014.01.048.

Burman, E., D. Mumovic et J. Kimpian, 2014. Towards measurement and verification of energy performance under the framework of the European directive for energy performance of buildings. In : *Energy*. décembre 2014. Vol. 77, p. 153‑163. DOI 10.1016/j.energy.2014.05.102.

Candanedo, L. M. et V. Feldheim, 2016. Accurate occupancy detection of an office room from light, temperature, humidity and CO 2 measurements using statistical learning models. In : *Energy and Buildings*. janvier 2016. Vol. 112, p. 28‑39. DOI 10.1016/j.enbuild.2015.11.071.

Candanedo, L. M., V. Feldheim et D. Deramaix, 2017. A methodology based on Hidden Markov Models for occupancy detection and a case study in a low energy residential building. In : *Energy and Buildings*. août 2017. Vol. 148, p. 327‑341. DOI 10.1016/j.enbuild.2017.05.031.

Caucheteux, A., 2015. *Incertitudes de mesure de la température ambiante d'un bâtiment pour l'évaluation de sa performance énergétique*. 2015. S.l. : Projet OMEGA.

Caucheteux, A., F. Titikpina, A. Charki et D. Bigaud, 2018. Optimisation des plans de mesure pour la garantie de performance énergétique. In : . 2018. p. 8.

CEREMA, 2016. *Bilan d'activité 2015 des commissions Titre V « Opération » et « Système »*. S.l. CEREMA.

Chamberlain, G., 1994. Quantile Regression, Censoring and the Structure of Wages. In : *Advances in Econometrics*. 1994.

Chan, A. L. S., 2011. Developing future hourly weather files for studying the impact of climate change on building energy performance in Hong Kong. In : *Energy and Buildings*. octobre 2011. Vol. 43, n° 10, p. 2860‑2868. DOI 10.1016/j.enbuild.2011.07.003.

Chen, K., Z. Ying, H. Zhang et L. Zhao, 2008. Analysis of Least Absolute Deviation. In : *Biometrika*. 2008. Vol. 95, n° 1, p. 107‑122.

Coakley, D., P. Raftery et M. Keane, 2014a. A review of methods to match building energy simulation models to measured data. In : *Renewable and Sustainable Energy Reviews*. septembre 2014. Vol. 37, p. 123‑141. DOI 10.1016/j.rser.2014.05.007.

Coakley, D., P. Raftery et M. Keane, 2014b. A review of methods to match building energy simulation models to measured data. In : *Renewable and Sustainable Energy Reviews*. septembre 2014. Vol. 37, p. 123‑141. DOI 10.1016/j.rser.2014.05.007.

Coakley, D., P. Raftery et P. Molloy, 2012. Calibration of whole building energy simulation models: Detailed case study of a naturally ventilated building using hourly measured data. In : *First Building Simulation and Optimization Conference, Loughborough, UK* [en ligne]. S.l. : s.n.

2012. p. 10–11. [Consulté le 20 avril 2016]. Disponible à l'adresse : http://www.ibpsa-england.org/resources/files/bso-2012/1C2.pdf.

Cornaro, C., V. A. Puggioni et R. M. Strollo, 2016. Dynamic simulation and on-site measurements for energy retrofit of complex historic buildings: Villa Mondragone case study. In : *Journal of Building Engineering*. juin 2016. Vol. 6, p. 17‑28. DOI 10.1016/j.jobe.2016.02.001.

Crawley, D. B., 1998. Which weather data should you use for energy simulations of commercial buildings? In : *Ashrae Transactions*. 1998. Vol. 104, p. 498.

Crawley, D. B., J. W. Hand, M. Kummert et B. T. Griffith, 2008. Contrasting the capabilities of building energy performance simulation programs. In : *Building and Environment*. 2008. Vol. 43, n° 4, p. 661‑673. DOI 10.1016/j.buildenv.2006.10.027.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petschek et J. H. Schaibly, 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients : I Theory. In : *Journal of Chemical Physics, Vol. 59*. 1973. p. 3873‑3878. DOI 10.1063/1.1680571.

Culp, C., D. E. Claridge et J. S. Haberl, 2005. ASHRAE's Guideline 14-2002 for Measurement of Energy and Demand Savings: How to Determine what was really saved by the retrofit. In : [en ligne]. 2005. [Consulté le 27 avril 2016]. Disponible à l'adresse : http://oaktrust.library.tamu.edu/handle/1969.1/5147.

Dean, A. et S. Lewis, 2006. *Screening - Methods for Experimentation in Industry, Drug Discovery, and Genetics* [en ligne]. Springer-Verlag. New York : Springer-Verlag. [Consulté le 9 mai 2015]. ISBN 978-0-387-28013-4. Disponible à l'adresse : http://link.springer.com/10.1007/0-387-28014-6.

Demanuele, C., T. Twedell et M. Davies, 2010. Bridging the gap between predicted and actual energy performance in schools. In : *Proceedigs WREC2010*. Abu Dhabi : s.n. 2010.

Deng, Q., X. Jiang, Q. Cui et L. Zhang, 2015a. Strategic design of cost savings guarantee in energy performance contracting under uncertainty. In : *Applied Energy*. février 2015. Vol. 139, p. 68‑80. DOI 10.1016/j.apenergy.2014.11.027.

Deng, Q., X. Jiang, L. Zhang et Q. Cui, 2015b. Making optimal investment decisions for energy service companies under uncertainty: A case study. In : *Energy*. août 2015. Vol. 88, p. 234‑243. DOI 10.1016/j.energy.2015.05.004.

Deng, Q., L. Zhang, Q. Cui et X. Jiang, 2014. A simulation-based decision model for designing contract period in building energy performance contracting. In : *Building and Environment*. janvier 2014. Vol. 71, p. 71‑80. DOI 10.1016/j.buildenv.2013.09.010.

Dickey, D. A. et W. A. Fuller, 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. In : *Econometrica*. 1981. n° 49, p. 1057. DOI 10.2307/1912517.

D'Oca, S. et T. Hong, 2015. Occupancy schedules learning process through a data mining framework. In : *Energy and Buildings*. février 2015. Vol. 88, p. 395‑408. DOI 10.1016/j.enbuild.2014.11.065.

Dubois, D. et H. Prade, 1988. *Possibility Theory - An approach to computerized processing of uncertainty*. Springer US-Plenum Press, New York. S.l. : Springer US - Plenum Press, New York. ISBN 978-1-4684-5289-1.

Efron, B., 1979. Bootstrap Methods: Another Look at the Jackknife. In : *Annals of statistics*. 1979. p. 1‑26.

Efron, B. et R. J. Tibshirani, 1993. *An introduction to the bootstrap*. New York : CRC Press. S.l. : s.n.

Enertech, 2013. *Immeuble d'habitation, rue Vignettes, FEYZIN (69) Campagne de mesures des performances énergétiques avant/après rénovation*. 2013. S.l. : s.n.

EVO, 2012. *International Performance Measurement and Verification Protocol. Vol. 1*. 2012. S.l. : s.n.

Fabrizio, E. et V. Monetti, 2015. Methodologies and Advancements in the Calibration of Building Energy Models. In : *Energies*. 31 mars 2015. Vol. 8, n° 4, p. 2548‑2574. DOI 10.3390/en8042548.

Faivre, R., B. Iooss, S. Mahevas et H. Monod, 2013. *Analyse de sensibilité et exploration de modèles: Application aux sciences de la nature et de l'environnement*. Quae. S.l. : s.n.

FBE, 2014. *Garantie de performance énergétique - Livrable 2.2*. S.l.

Fennel, P. J., P. A. Ruyssevelt et A. Z. Smith, 2017. Exploring the commercial implications of measurement and verification choices in energy performance contracting using stochastic building simulation. In : *Proceedings of BS2017*. San Francisco : s.n. 2017.

Ferrante, P., M. La Genusa et S. Nicolosi, 2008. Comparison of methodologies for test reference year (TRY) generation for mediterranean sites. In : *25th Conference on Passive and Low Energy Architecture*. PLEA, Dublin : s.n. 2008.

Friedman, J. H., 1991. Multivariate Adaptive Regression Splines. In : *The Annals of Statistics*. 1991. Vol. 19, n° 1, p. 1‑67.

Gilani, S. et W. O'Brien, 2017. Review of current methods, opportunities, and challenges for in-situ monitoring to support occupant modelling in office spaces. In : *Journal of Building Performance Simulation*. 2 novembre 2017. Vol. 10, n° 5‑6, p. 444‑470. DOI 10.1080/19401493.2016.1255258.

Givord, P. et W. D'haultfoeuille, 2013. *La régression quantile en pratique*. S.l. INSEE. Méthodologie Statistique.

Gobin, C., 2014. *Le bâtiment comme système*. 2014. S.l. : Techniques de l'ingénieur.

Goffart, J., 2013a. *Impact de la variabilité des données météorologiques sur une maison basse consommation. Application des analyses de sensibilité pour les entrées temporelles.* [en ligne]. S.l. : Université de Grenoble. [Consulté le 4 juin 2014]. Disponible à l'adresse : http://tel.archives-ouvertes.fr/tel-00982150.

Goffart, J., 2013b. *Impact de la variabilité des données météorologiques sur une maison basse consommation. Application des analyses de sensibilité pour les entrées temporelles.* [en ligne]. Thèse de doctorat. S.l. : Université de Grenoble. [Consulté le 4 juin 2014]. Disponible à l'adresse : http://tel.archives-ouvertes.fr/tel-00982150.

Granderson, J. et P. N. Price, 2014. Development and application of a statistical methodology to evaluate the predictive accuracy of building energy baseline models. In : *Energy*. mars 2014. Vol. 66, p. 981‑990. DOI 10.1016/j.energy.2014.01.074.

Granderson, J., P. N. Price, D. Jump, N. Addy et M. D. Sohn, 2015. Automated measurement and verification: Performance of public domain whole-building electric baseline models. In : *Applied Energy*. avril 2015. Vol. 144, p. 106‑113. DOI 10.1016/j.apenergy.2015.01.026.

Gutenbrunner, C., J. Jurečková, R. Koenker et S. Portnoy, 1993. Tests of linear hypotheses based on regression rank scores. In : *Journal of Nonparametric Statistics*. 1 janvier 1993. Vol. 2, n° 4, p. 307‑331. DOI 10.1080/10485259308832561.

Haldi, F. et D. Robinson, 2011a. Modelling Occupants' Presence and Behaviour – Part I. In : *Journal of Building Performance Simulation*. 2011. Vol. 4, n° 4, p. 301‑302.

Haldi, F. et D. Robinson, 2011b. The impact of occupants' behaviour on building energy demand. In : *Journal of Building Performance Simulation*. 2011. Vol. 4, n° 4, p. 323‑338.

Haldi, F. et D. Robinson, 2012. Modelling Occupants' Presence and Behaviour – Part II. In : *Journal of Building Performance Simulation*. 2012. Vol. 5, n° 1, p. 1‑3.

Hall, P. et S. Sheather, 1988. On the distribution of a studentized quantile. In : *Journal of the Royal Statistical Society: Series B*. 1988. p. 381‑391.

Ham, Y. et M. Golparvar-Fard, 2013. EPAR: Energy Performance Augmented Reality models for identification of building energy performance deviations between actual measurements and simulation results. In : *Energy and Buildings*. août 2013. Vol. 63, p. 15‑28. DOI 10.1016/j.enbuild.2013.02.054.

Harmer, L. C. et G. P. Henze, 2015. Using calibrated energy models for building commissioning and load prediction. In : *Energy and Buildings*. avril 2015. Vol. 92, p. 204‑215. DOI 10.1016/j.enbuild.2014.10.078.

Heo, Y., R. Choudhary et G. A. Augenbroe, 2012. Calibration of building energy models for retrofit analysis under uncertainty. In : *Energy and Buildings*. avril 2012. Vol. 47, p. 550‑560. DOI 10.1016/j.enbuild.2011.12.029.

Heo, Y. et V. M. Zavala, 2012. Gaussian process modeling for measurement and verification of building energy savings. In : *Energy and Buildings*. octobre 2012. Vol. 53, p. 7‑18. DOI 10.1016/j.enbuild.2012.06.024.

Herkel, S., U. Knapp et J. Pfafferott, 2008. Towards a model of user behaviour regarding the manual control of windows in office buildings. In : *Building and Environment*. avril 2008. Vol. 43, n° 4, p. 588‑600. DOI 10.1016/j.buildenv.2006.06.031.

Hong, T., W.-K. Chang et H.-W. Lin, 2013. A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data. In : *Applied Energy*. novembre 2013. Vol. 111, p. 333‑ 350. DOI 10.1016/j.apenergy.2013.05.019.

Hoyt, T., E. Arens et H. Zhang, 2015. Extending air temperature setpoints: Simulated energy savings and design considerations for new and retrofit buildings. In : *Building and Environment*. juin 2015. Vol. 88, p. 89‑ 96. DOI 10.1016/j.buildenv.2014.09.010.

Iman, R. L. et W. J. Conover, 1982. A distribution-free approach to inducing rank correlation among input variables. In : *Communications in statistics - Simulation and computation*. 1982.

Iooss, B., 2011a. Revue sur l'analyse de sensibilité globale de modèles numériques. In : *Journal de la Société Française de Statistique*. 2011. Vol. 152, n° 1, p. 3–25.

Iooss, B., 2011b. Revue sur l'analyse de sensibilité globale de modèles numériques. In : *Journal de la Société Française de Statistique*. 2011. Vol. 152, n° 1, p. 3–25.

JCGM, 1995. *GUM, Guide to the expression of uncertainty in Measurement*. S.l. : s.n.

Jin, H.-S., B.-H. Choi, J.-K. Kang, S.-I. Kim, J.-H. Lim et S.-Y. Song, 2016. Measurement and Normalization Methods to Provide Detailed Information on Energy Consumption by Usage in Apartment Buildings. In : *Energy Procedia*. septembre 2016. Vol. 96, p. 881‑ 894. DOI 10.1016/j.egypro.2016.09.161.

Kampelis, N., K. Gobakis, V. Vagias, D. Kolokotsa, L. Standardi, D. Isidori, C. Cristalli, F. M. Montagnino, F. Paredes, P. Muratore, L. Venezia, M. K. Dracou, A. Montenon, A. Pyrgou, T. Karlessi et M. Santamouris, 2017. Evaluation of the performance gap in industrial, residential & tertiary near-Zero energy buildings. In : *Energy and Buildings*. août 2017. Vol. 148, p. 58‑ 73. DOI 10.1016/j.enbuild.2017.03.057.

Kashif, A., S. Ploix, J. Dugdale et X. H. B. Le, 2013. Simulating the dynamics of occupant behaviour for power management in residential buildings. In : *Energy and Buildings*. janvier 2013. Vol. 56, p. 85‑ 93. DOI 10.1016/j.enbuild.2012.09.042.

Kazmi, H. et S. Ploix, 2018. Estimating Occupancy in residential context using Bayesian Networks for Energy management. In : *20th International Conference on Machine Learning for Prediction and control*. London : s.n. 2018.

Ke, M.-T., C.-H. Yeh et J.-T. Jian, 2013. Analysis of building energy consumption parameters and energy savings measurement and verification by applying eQUEST software. In : *Energy and Buildings*. juin 2013. Vol. 61, p. 100‑ 107. DOI 10.1016/j.enbuild.2013.02.012.

Kocherginsky, M., X. He et Y. Mu, 2005. Practical Confidence Intervals for Regression Quantiles. In : *Journal of Computational and Graphical Statistics*. mars 2005. Vol. 14, n° 1, p. 41‑ 55. DOI 10.1198/106186005X27563.

Koenker, 2005. Quantile Regression. In : *Economic Society Monograph Series*. 2005.

Koenker et Machado, 1999. Goodness of fit and related inference processes for quantile regression. In : *Journal of the American Statistical Association*. 1999. Vol. 94, n° 448, p. 1296‑ 1310.

Koenker, R., 1994. Confidence intervals for regression quantiles. In : *Asymptotic statistics*. S.l. : Springer. p. 349–359.

Koenker, R. et G. Bassett, 1978. Regression Quantiles. In : *Econometrica*. 1978. p. 33‑ 50.

Kolokotroni, M., X. Ren, M. Davies et A. Mavrogianni, 2012. London's urban heat island: Impact on current and future energy consumption in office buildings. In : *Energy and Buildings*. avril 2012. Vol. 47, p. 302‑ 311. DOI 10.1016/j.enbuild.2011.12.019.

Kurowicka, D. et R. Cooke, 2006. *Uncertainty Analysis with High Dimensional Dependence Modelling*. Wiley. S.l. : s.n. ISBN 978-0-470-86306-0.

Kwiatkowski, D., P. C. B. Philips, P. Schmidt et Y. Shin, 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? In : *Journal of Econometrics*. 1992. n° 54. DOI 10.1016/0304-4076(92)90104-Y.

Labeodan, T., W. Zeiler, G. Boxem et Y. Zhao, 2015. Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation. In : *Energy and Buildings*. avril 2015. Vol. 93, p. 303‑ 314. DOI 10.1016/j.enbuild.2015.02.028.

Laustsen, J., 2016. *New IEA EBC Annex-proposal : Building energy performance assessment based on optimized in-situ measurements*. 2016. S.l. : s.n.

Lee, B. D., Y. Sun, H. Hu, G. Augenbroe et C. J. Paredis, 2012. A framework for generating stochastic meteorological years for risk-conscious design of buildings. In : *Proceedings of the SimBuild 2012 Conference, Madison, Wisconsin, USA*. S.l. : s.n. 2012. p. 345–52.

Lee, P., P. T. I. Lam, F. W. H. Yik et E. H. W. Chan, 2013. Probabilistic risk assessment of the energy saving shortfall in energy performance contracting projects–A case study. In : *Energy and Buildings*. novembre 2013. Vol. 66, p. 353‑ 363. DOI 10.1016/j.enbuild.2013.07.018.

Lee, S. H., T. Hong, M. A. Piette et S. C. Taylor-Lange, 2015. Energy retrofit analysis toolkits for commercial buildings: A review. In : *Energy*. septembre 2015. Vol. 89, p. 1087‑ 1100. DOI 10.1016/j.energy.2015.06.112.

Li, N., Z. Yang, B. Becerik-Gerber, C. Tang et N. Chen, 2015. Why is the reliability of building simulation limited as a tool for evaluating energy conservation measures? In : *Applied Energy*. décembre 2015. Vol. 159, p. 196‑ 205. DOI 10.1016/j.apenergy.2015.09.001.

Li, Z., Y. Han et P. Xu, 2014. Methods for benchmarking building energy consumption against its past or intended performance: An overview. In : *Applied Energy*. juillet 2014. Vol. 124, p. 325‑ 334. DOI 10.1016/j.apenergy.2014.03.020.

Lin, D. K. J., 1993. A New Class of Supersaturated Designs. In : *Technometrics*. 1 février 1993. Vol. 35, n° 1, p. 28‑ 31. DOI 10.2307/1269286.

Ma, Z., P. Cooper, D. Daly et L. Ledo, 2012. Existing building retrofits: Methodology and state-of-the-art. In : *Energy and Buildings*. décembre 2012. Vol. 55, p. 889‑ 902. DOI 10.1016/j.enbuild.2012.08.018.

Macdonald, I. A., 2002. *Quantifying the effects of uncertainty in building simulation* [en ligne]. S.l. : University of Strathclyde. [Consulté le 26 septembre 2013]. Disponible à l'adresse :

https://www.strath.ac.uk/media/departments/mechanicalengineering/esru/research/phdmphilprojects/macdonald_thesis.pdf.

Manfren, M., N. Aste et R. Moshksar, 2013. Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation. In : *Applied Energy*. mars 2013. Vol. 103, p. 627‑641. DOI 10.1016/j.apenergy.2012.10.031.

Mara, T. A., 2009. Extension of the RBD-FAST method to the computation of global sensitivity indices. In : *Reliability Engineering & System Safety*. août 2009. Vol. 94, n° 8, p. 1274‑1281. DOI 10.1016/j.ress.2009.01.012.

Moore, R., 1979. *Methods and Applications of Interval Analysis*. S.l. : Society for Industrial and Applied Mathematics. Studies in Applied and Numerical Mathematics. ISBN 978-0-89871-161-5.

Mora, L., A. Faggianelli, S. Bontemps, M. Robillart, C. Garnier, B. Peuportier, L. Rivalin, D. Da Silva et A. Caucheteux, 2016. Livrable L2.1-L2.2. : *Projet OMEGA*. S.l.

Morin, E., 1990. *Introduction à la pensée complexe*. S.l. : ESF Editeur. ISBN 978-2-02-066837-8.

Morris, M. D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. In : *Technometrics*. 1 mai 1991. Vol. 33, n° 2, p. 161‑174. DOI 10.1080/00401706.1991.10484804.

Munaretto, F., 2010. *Etude de l'influence de l'inertie thermique sur les performances énergétiques et environnementales des bâtiments*. Paris.

Munaretto, F., 2014. *Étude de l'influence de l'inertie thermique sur les performances énergétiques des bâtiments* [en ligne]. Thèse de doctorat. S.l. : École nationale supérieure des mines de Paris. [Consulté le 9 octobre 2014]. Disponible à l'adresse : http://pastel.archives-ouvertes.fr/pastel-01068784.

Munaretto, F., T. Recht, P. Schalbart et B. Peuportier, 2017. Empirical validation of different internal superficial heat transfer models on a full-scale passive house. In : *Journal of Building Performance Simulation*. 6 juin 2017. p. 1‑22. DOI 10.1080/19401493.2017.1331376.

Neumeier, A., 1990. *Interval Methods for Systems of Equations* [en ligne]. Cambridge University press. S.l. : s.n. [Consulté le 21 décembre 2016]. ISBN 0-521-33196-X. Disponible à l'adresse : http://www.cambridge.org/fr/academic/subjects/mathematics/numerical-analysis/interval-methods-systems-equations?format=PB&isbn=9780521102148.

Nolan, E., J. Allsopp, A. Galata, G. Pedone et Z. Zivkovic, 2014. Calibrating whole building energy model: a case study using BEMS data. In : [en ligne]. 2014. [Consulté le 20 avril 2016]. Disponible à l'adresse : http://eprints.sztaki.hu/8150/1/Nolan_487_2797156_ny.pdf.

Orgill, J. F. et G. T. Hollands, 1977. Correlation Equation for Hourly Diffuse Radiation on a Horizontal Surface. In : *Solar Energy*. 1977. Vol. 19, p. 357‑359.

Owen, A. B., 2003. Quasi-monte carlo sampling. In : *Monte Carlo Ray Tracing: Siggraph*. 2003. Vol. 1, p. 69–88.

Pannier, M.-L., 2017. *Étude de la quantification des incertitudes dans l'analyse de cycle de vie des bâtiments et des quartiers*. S.l. : PSL - Mines ParisTech.

Park, S., V. Norrefeldt, S. Stratbuecker, G. Grün et Y.-S. Jang, 2013. Methodological approach for calibration of building energy performance simulation models applied to a common "measurement and verification" process. In : *Bauphysik*. août 2013. Vol. 35, n° 4, p. 235‑ 241. DOI 10.1002/bapi.201310070.

Parzen, M. I., L. J. Wei et Z. Ying, 1994. A resampling method based on pivotal estimating functions. In : *Biometrika*. 1994. Vol. 81, n° 2, p. 341–350.

Peuportier, B., 2014. *Ecoconception des bâtiments et des quartiers*. 2014. S.l. : Techniques de l'ingénieur.

Peuportier, B., 2016. *Energétique des bâtiments et simulation thermique*. Eyrolles. S.l. : s.n. Blanche BTP.

Peuportier, B. et I. Blanc-Sommereux, 1990. Simulation tool with its expert interface for the thermal design of multizone buildings. In : *International Journal of Solar Energy*. 1990. Vol. 8, n° 2, p. 109‑ 120. DOI 10.1080/01425919008909714.

Peuportier, B. et S. Thiers, 2009. Les maisons passives : sont-elles confortables ? écologiques ? In : *Revue Chauffage, Ventilation et Climatisation*. février 2009. n° 857, p. 22‑ 25.

Plischke, E., 2010. An effective algorithm for computing global sensitivity indices (EASI). In : *Reliability Engineering & System Safety*. avril 2010. Vol. 95, n° 4, p. 354‑ 360. DOI 10.1016/j.ress.2009.11.005.

Plischke, E., 2012. How to compute variance-based sensitivity indicators with your spreadsheet software. In : *Environmental Modelling & Software*. 2012. Vol. 35, p. 188‑ 191.

Qin, Q., F. Liang, L. Li et Y.-M. Wei, 2017. Selection of energy performance contracting business models: A behavioral decision-making approach. In : *Renewable and Sustainable Energy Reviews*. mai 2017. Vol. 72, p. 422‑ 433. DOI 10.1016/j.rser.2017.01.058.

Rastogi, P., 2016a. *On the sensitivity of buildings to climate: the interaction of weather and building envelopes in determining future building energy consumption* [en ligne]. Thèse de doctorat. S.l. : EPFL. [Consulté le 18 novembre 2016]. Disponible à l'adresse : https://pdfs.semanticscholar.org/50af/f6e04c32eb4acb83c1ebe3e155ef8885cfee.pdf.

Rastogi, P., 2016b. *On the sensitivity of buildings to climate: the interaction of weather and building envelopes in determining future building energy consumption* [en ligne]. Thèse de doctorat. S.l. : EPFL. [Consulté le 16 février 2017]. Disponible à l'adresse : https://pdfs.semanticscholar.org/50af/f6e04c32eb4acb83c1ebe3e155ef8885cfee.pdf.

Recht, T., 2016. *Étude de l'écoconception de maisons à énergie positive* [en ligne]. Thèse de doctorat. Paris : École nationale supérieure des mines de Paris. Disponible à l'adresse : https://www.theses.fr/2016PSLEM024.

Recht, T., F. Munaretto, P. Schalbart et B. Peuportier, 2014. Analyse de la fiabilité de COMFIE par comparaison à des mesures. Application à un bâtiment passif. In : *Proceedings IBPSA France*. Arras : s.n. 20 mai 2014. p. 8.

Remund, J., S. Mueller, S. Kunz et C. Schilter, 2012a. *METEONORM Handbook part I : Software technical report*. S.l.

Remund, J., S. Mueller, S. Kunz et C. Schilter, 2012b. *METEONORM Handbook part II : Theory*. S.l.

Riederer, P., D. Marchio, J. C. Visier, A. Husaunndee et R. Lahrech, 2001. Influence of sensor position in building thermal control : development and validation of an adapted zone model. In : *Building Simulation*. Rio de Janeiro : s.n. 2001.

Rivalin, L., 2016. *Vers une démarche de garantie des consommations énergétiques dans les bâtiments neufs : Méthodes d'évaluation des incertitudes associées à la simulation thermique dynamique dans le processus de conception et de réalisation.* Mines ParisTech : s.n.

Rivalin, L., D. Marchio, P. Stabat, M. Caciolo et B. Cogné, 2014. Influence du découpage en zones sur les besoins énergétiques annuels. In : *IBPSA'International Building Performance Simulation Association*" [en ligne]. S.l. : s.n. 2014. [Consulté le 21 septembre 2017]. Disponible à l'adresse : https://hal-ensmp.archives-ouvertes.fr/hal-01299906/document.

Robillart, M., 2015. *Étude de stratégies de gestion en temps réel pour des bâtiments énergétiquement performants*. Thèse de doctorat. S.l. : École nationale supérieure des mines de Paris.

Rocha, A. P. de A., J. Goffart, L. Houben et N. Mendes, 2016. On the uncertainty assessment of incident direct solar radiation on building facades due to shading devices. In : *Energy and Buildings*. décembre 2016. Vol. 133, p. 295‑304. DOI 10.1016/j.enbuild.2016.09.058.

Roels, S., P. Bacher, G. Bauwens et H. Madsen, 2015. Characterizing the actual thermal performance of buildings : current results of common exercises performed in the framework of the IEA EBC Annex 58-project. In : *IBPC 2015*. IBPC : s.n. 2015.

Rouaud, M., 2012. *Probabilités, statistiques et analyses multicritères*. Books on Demand Editions. S.l. : s.n. ISBN 978-2-8106-2461-4.

Roux, C., 2016. *Analyse de cycle de vie conséquentielle appliquée aux ensembles bâtis*. Thèse de doctorat. S.l. : MINES ParisTech.

Roux, C., P. Schalbart, E. Assoumou et B. Peuportier, 2016. Integrating climate change and energy mix scenarios in LCA of buildings and districts. In : *Applied Energy*. 15 décembre 2016. Vol. 184, p. 619‑629. DOI 10.1016/j.apenergy.2016.10.043.

Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto et S. Tarantola, 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. In : *Computer Physics Communications*. février 2010. Vol. 181, n° 2, p. 259‑270. DOI 10.1016/j.cpc.2009.09.018.

Saltelli, A., S. Tarantola et K. P.-S. Chan, 1999. A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. In : *Technometrics*. 1999. Vol. 41, n° 1, p. 39‑56. DOI 10.2307/1270993.

Satterthwaite, F. E., 1959. Random Balance Experimentation. In : *Technometrics*. 1 mai 1959. Vol. 1, n° 2, p. 111‑ 137. DOI 10.2307/1266466.

Silva, A. S. et E. Ghisi, 2014. Uncertainty analysis of user behaviour and physical parameters in residential building performance simulation. In : *Energy and Buildings*. juin 2014. Vol. 76, p. 381‑ 391. DOI 10.1016/j.enbuild.2014.03.001.

Sobol, I., 1993. Sensitivity estimates for nonlinear mathematical models. In : *Mathematical Modelling and Computational Experiment*. 1993. Vol. 1, n° 4, p. 407‑ 414. DOI 1061-7590/93/04407-008$9.00.

Sobol, Ilya Meïérovitch, 1993. Sensitivity estimates for nonlinear mathematical models. In : *Math. Mod. and Comput. Exp.* 1993. DOI 1 :407–414.

Sonderegger, R. et M. Modera, 1979. Electric co-heating : a method for evaluating seasonal heating efficiencies and heat loss rates in dwellings. In : *Second CIB Symposium on Energy Conservation in the Built Environment*. Lawrence Berkeley National Laboratory : s.n. 1979.

Spitz, C., 2012. *Analyse de la fiabilité des outils de simulation et des incertitudes de métrologie appliquée à l'efficacité énergétique des bâtiments* [en ligne]. S.l. : Université de Grenoble. [Consulté le 19 avril 2016]. Disponible à l'adresse : https://tel.archives-ouvertes.fr/tel-00768506/.

Spitz, C., L. Mora, E. Wurtz et A. Jay, 2012. Practical application of uncertainty analysis and sensitivity analysis on an experimental house. In : *Energy and Buildings*. décembre 2012. Vol. 55, p. 459‑ 470. DOI 10.1016/j.enbuild.2012.08.013.

Srivastav, A., A. Tewari et B. Dong, 2013. Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models. In : *Energy and Buildings*. octobre 2013. Vol. 65, p. 438‑ 447. DOI 10.1016/j.enbuild.2013.05.037.

Stefanoiu, A.-M., M. Woloszyn, A. Jay, E. Wurtz et C. Buhé, 2015. A Methodology to Assess the Ambient Temperature of a Building using a Limited Number of Sensors. In : *Energy Procedia*. novembre 2015. Vol. 78, p. 1944‑ 1949. DOI 10.1016/j.egypro.2015.11.377.

Struck, C. et J. Hensen, 2007. On supporting design decisions in conceptual design addressing specification uncertainties using performance simulation. In : *Proceedings of the 10th IBPSA Building Simulation Conference*. S.l. : s.n. 2007. p. 3–5.

Sun, K., D. Yan, T. Hong et S. Guo, 2014. Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration. In : *Building and Environment*. septembre 2014. Vol. 79, p. 1‑ 12. DOI 10.1016/j.buildenv.2014.04.030.

Sun, S., K. Kensek, D. Noble et M. Schiler, 2016. A method of probabilistic risk assessment for energy performance and cost using building energy simulation. In : *Energy and Buildings*. janvier 2016. Vol. 110, p. 1‑ 12. DOI 10.1016/j.enbuild.2015.09.070.

Tarantola, S., D. Gatelli et T. A. Mara, 2006. Random balance designs for the estimation of first order global sensitivity indices. In : *Reliability Engineering & System Safety*. juin 2006. Vol. 91, n° 6, p. 717‑ 727. DOI 10.1016/j.ress.2005.06.003.

Thebault, S., 2017. *Contribution à l'évaluation in situ des performances d'isolation thermique de l'enveloppe des bâtiments*. S.l. : s.n.

Tian, W., 2013. A review of sensitivity analysis methods in building energy analysis. In : *Renewable and Sustainable Energy Reviews*. avril 2013. Vol. 20, p. 411‑419. DOI 10.1016/j.rser.2012.12.014.

Tian, W., R. Choudhary, G. Augenbroe et S. H. Lee, 2015. Importance analysis and meta-model construction with correlated variables in evaluation of thermal performance of campus buildings. In : *Building and Environment*. octobre 2015. Vol. 92, p. 61‑74. DOI 10.1016/j.buildenv.2015.04.021.

Titikpina, F., 2016. *Optimisation des plans de mesure dans un contexte de garantie de performance énergétique des bâtiments: Caractérisation des incertitudes de mesure et propagation au sein de modèles dynamiques*. S.l. : Université Angers.

van Dronkelaar, C., M. Dowson, C. Spataru et D. Mumovic, 2016. A Review of the Regulatory Energy Performance Gap and Its Underlying Causes in Non-domestic Buildings. In : *Frontiers in Mechanical Engineering* [en ligne]. 13 janvier 2016. Vol. 1. [Consulté le 26 juillet 2017]. DOI 10.3389/fmech.2015.00017. Disponible à l'adresse : http://journal.frontiersin.org/Article/10.3389/fmech.2015.00017/abstract.

Veiga, S. D., F. Wahl et F. Gamboa, 2008. Local polynomial estimation for sensitivity analysis for models with correlated inputs. In : *Technometrics*. 2008. Vol. 51, n° 4, p. 452‑463. DOI 10.1198.

Vorger, E., 2014. *Etude de l'influence du comportement des occupants sur la performance énergétique des bâtiments*. Thèse de doctorat. S.l. : Ecole Nationale Supérieure des Mines de Paris.

Vorger, E., P. Schalbart et B. Peuportier, 2014a. Etude de l'influence des occupants sur la performance énergétique des logements par le biais d'une modélisation stochastique globale. In : *IBPSA France*. Arras : s.n. 20 mai 2014.

Vorger, E., P. Schalbart et B. Peuportier, 2014b. Integration of a comprehensive stochastic model of occupancy in building simulation to study how inhabitants influence energy performance. In : *Proceedings PLEA 2014*. Ahmedabad (India) : s.n. décembre 2014. p. 8.

Wang, L., P. Mathew et X. Pang, 2012. Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building. In : *Energy and Buildings*. octobre 2012. Vol. 53, p. 152‑158. DOI 10.1016/j.enbuild.2012.06.017.

Wang, Q., G. Augenbroe et Kim, 2015. A framework for meta-analysis of the tole of occupancy variables in the energy use of commercial buildings. In : *IBPSA*. Hyderabad, India : s.n. 2015.

Wang, S., A. A. Kim et E. M. Johnson, 2017. Understanding the deterministic and probabilistic business cases for occupant based plug load management strategies in commercial office buildings. In : *Applied Energy*. avril 2017. Vol. 191, p. 398‑413. DOI 10.1016/j.apenergy.2017.01.082.

Yan, D., W. O'Brien, T. Hong, X. Feng, H. Burak Gunay, F. Tahmasebi et A. Mahdavi, 2015. Occupant behavior modeling for building performance simulation: Current state and future challenges. In : *Energy and Buildings*. novembre 2015. Vol. 107, p. 264‑ 278. DOI 10.1016/j.enbuild.2015.08.032.

Yang, L., J. C. Lam, J. Liu et C. L. Tsang, 2008. Building energy simulation using multi-years and typical meteorological years in different climates. In : *Energy Conversion and Management*. janvier 2008. Vol. 49, n° 1, p. 113‑ 124. DOI 10.1016/j.enconman.2007.05.004.

Zhang, Y., Z. O'Neill, B. Dong et G. Augenbroe, 2015. Comparisons of inverse modeling approaches for predicting building energy performance. In : *Building and Environment*. avril 2015. Vol. 86, p. 177‑ 190. DOI 10.1016/j.buildenv.2014.12.023.

Annexes

A-1-1 Modèle COMFIE pour la SED 269

A-2-1 Sélection des processus de type ARMA..... 270

A-2-2 Composantes principales de la décomposition de Fourier de la série tendancielle 271

A-2-3 Illustration de la méthode de génération des données de RGH..... 273

A-2-4 Calcul du rayonnement solaire extra-terrestre..... 275

A-2-5 Interventions algorithmiques 277

A-3-1 Incertitude sur les DJU calculés avec la méthode COSTIC..... 278

A-5-1 Spécifications nominales des modèles SED de Feyzin avant et après réhabilitation 279

A-5-2 Comparaison des modèles de $LGC5\%$ pour différents FA et PMV 282

A-5-3 Schémas des installations techniques du bâtiment Izuba 287

A-5-4 Caractéristiques principales du modèle SED du bâtiment Izuba 288

A-5-5 Dispositif de mesure des performances du bâtiment Izuba 291

A-5-6 Incertitudes de mesure des températures intérieures moyennes 292

A-5-7 Comparaison des modèles de régression quantile..... 293

A-5-8 Modèles de régression sur l'espérance 294

A-1-1 Modèle COMFIE pour la SED

COMFIE est le moteur de calcul de simulation énergétique dynamique utilisé dans ces travaux. Il repose sur le concept de zone thermique, partie du bâtiment considérée à température homogène. Selon le principe de la méthode des volumes finis, les murs sont décomposés en mailles, à température uniforme. L'air, le mobilier et les cloisons légères sont regroupés au sein d'une même maille. Un bilan thermique est appliqué sur l'ensemble des mailles, et prend la forme suivante :

$$C_m \frac{dT_m}{dt} = \dot{Q}_G - \dot{Q}_P$$

avec :

- C_m la capacité thermique de la maille,
- T_m la température de la maille,
- $\dot{Q}_G$, les apports de chaleur (transferts thermiques apports solaires et internes, puissances de chauffage et de climatisation)
- $\dot{Q}_P$ les pertes de chaleur (transferts thermiques ,ventilation)

Les phénomènes non stationnaires sont pris en compte dans un vecteur de sollicitations U. Le problème peut ainsi s'écrire sous la forme d'un système d'équations différentielles linéaires :

$$\begin{cases} C\dot{T}(t) = AT(t) + EU(t) \\ Y(t) = JT(t) + GU(t) \end{cases}$$

avec :

- T, le champ discrétisé des températures,
- C, la matrice diagonale des capacités thermiques des mailles,
- U, le vecteur des sollicitations,
- A, la matrice des termes d'échanges entre mailles (matrice d'état),
- E, la matrice des termes d'échanges entre mailles et sollicitations (matrice de commande),
- J, la matrice liant sorties et températures de mailles (matrice d'observation),
- G, matrice liant sorties et sollicitations (matrice d'action directe),
- Y, vecteur des sorties (température de zone).

Le système est linéaire et invariant, car les transferts radiatifs sont linéarisés. Il peut être décomposé en un terme dynamique et un terme de régime permanent :

$$T(t) = T_d(t) + T_p(t)$$

où $T_p(t)$ représente le comportement du système en régime permanent. On a alors :

$$T_p(t) = -A^{-1}EU(t)$$

En intégrant cette décomposition dans le système d'équation, on obtient :

$$\begin{cases} T_d(t) = C^{-1}AT_d(t) + A^{-1}E\dot{U}(t) \\ Y(t) = JT_d(t) + (G - JA^{-1}E)U(t) \end{cases}$$

La matrice $C^{-1}A$ est diagonalisée pour résoudre le système. Un changement de base est effectué en utilisant la matrice de passage P formée des vecteurs propres de $C^{-1}A$. $T_d(t)$ est transformé en un vecteur d'état $X(t)$: $T_d(t) = PX(t)$. Le système s'écrit alors :

$$\begin{cases} \dot{X}(t) = FX(t) + B\dot{U}(t) \\ Y(t) = HX(t) + SU(t) \end{cases}$$

avec $F = P^{-1}C^{-1}AP$, $B = P^{-1}A^{-1}E$, $H = JP$ et $S = G - JA^{-1}E$.

F est une matrice diagonale dont le $i_{ème}$ terme diagonal est $-1/\tau_i$. Ces constantes de temps τ_i sont ordonnées dans la matrice par ordre décroissant. Ce changement de base permet de réduire l'ordre du système, ce qui accélère les calculs. En effet, les faibles constantes de temps correspondent à des termes $X(t)$ atteignant très rapidement leur régime permanent. Ces termes peuvent être négligés dans le calcul de la partie dynamique. On peut ainsi obtenir un modèle réduit fiable en ne conservant que les N plus grandes constantes de temps. La valeur par défaut de N est 3, mais peut être modifiée automatiquement en fonction des systèmes (et des constantes de temps) pour chaque maille. Pour plus de détails sur la résolution du système réduit, il est possible de se référer à Robillart (2015).

A-2-1 Sélection des processus de type ARMA

Stationnarité

La stationnarité d'une série temporelle, c'est-à-dire l'invariance de ses caractéristiques par translation temporelle, est vérifiée à partir de tests. Deux types de tests peuvent être cités, ceux pour lesquels l'hypothèse nulle est la stationnarité de la série, par exemple *Kwiatkowski-Phillips-Schmidt-Shin* (*KPSS*) (Kwiatkowski et al. 1992), et ceux pour lesquels cette hypothèse correspond à l'existence d'une racine unitaire dans la série et donc à une non-stationnarité tel que *Augmented Dickey-Fuller (ADF)* (Dickey et Fuller 1981).

Critères de sélection de modèle ARMA

Soit x_1,…, une série d'observations indépendantes d'un processus aléatoire de loi inconnue et soit (x_1,…,$|\theta$) la densité de probabilité spécifiée par le paramètre θ. On appelle L la vraisemblance du paramètre θ :

$$L(\theta) = \prod_{j=1}^{n} f(x_j|\theta)$$

Afin d'évaluer la correspondance d'une série temporelle à un processus de type ARMA, trois indicateurs principaux peuvent être calculés (Robillart 2015) :

- Le critère AIC (*Akaike's Information Criterion*) :

$$AIC = -2logL(\theta) + 2\omega$$

où θ désigne les paramètres du modèle et ω le nombre de ces paramètres. Ce critère sera utilisé si l'objectif de l'étude est de réaliser des prévisions.

- Le critère AIC corrigé :

$$AICc = AIC + \frac{2\omega(\omega + 1)}{n - \omega - 1}$$

avec *n* le nombre d'observations.

- Le critère BIC (*Bayesian Information Criterion)*

$$BIC = -2logL(\theta) + n\omega$$

Ce critère sera préféré si l'on souhaite ajuster un modèle à une série observée.

Le modèle sélectionné est celui qui minimise le ou les critères étudiés. Tout en pénalisant une sur-complexité du modèle, ces critères permettent la sélection de modèles ayant la vraisemblance maximale. Ils sont calculés automatiquement grâce à des fonctions disponibles dans des bibliothèques *R* ou Python.

A-2-2 Composantes principales de la décomposition de Fourier de la série tendancielle

La Figure A-1 et la Figure A-2 présentent les résultats partiels de la décomposition de Fourier des données de température du fichier-type de Mâcon et des années réelles de 2010 à 2016. On s'intéresse à la représentativité des différents modes comme composantes d'une série tendancielle de température extérieure. Les coefficients $S(k)$ (équation 2.5) peuvent s'écrire sous forme de nombres exponentiels complexes définis par un module (amplitude) et un argument.

Au-delà des modes principaux f_0, f_1 (hors échelle) et f_{365}, on constate que certains modes ont des amplitudes similaires, de l'ordre de 0,2 à 0,6 °C. En étudiant les déphasages associés à ces fonctions, on constate que pour certains modes, ils sont très variables suivant les séries sur lesquelles la décomposition est effectuée. Ces modes sont donc importants dans la description des séries, l'amplitude du signal étant non négligeable, mais ils ne correspondent pas à une part tendancielle. Pour les modes f_{364} et f_{366}, tout comme pour les modes principaux, les déphasages sont similaires,

quelles que soient les années étudiées. La prise en compte de ces signaux est donc importante dans la série tendancielle annuelle de l'évolution des températures extérieures.

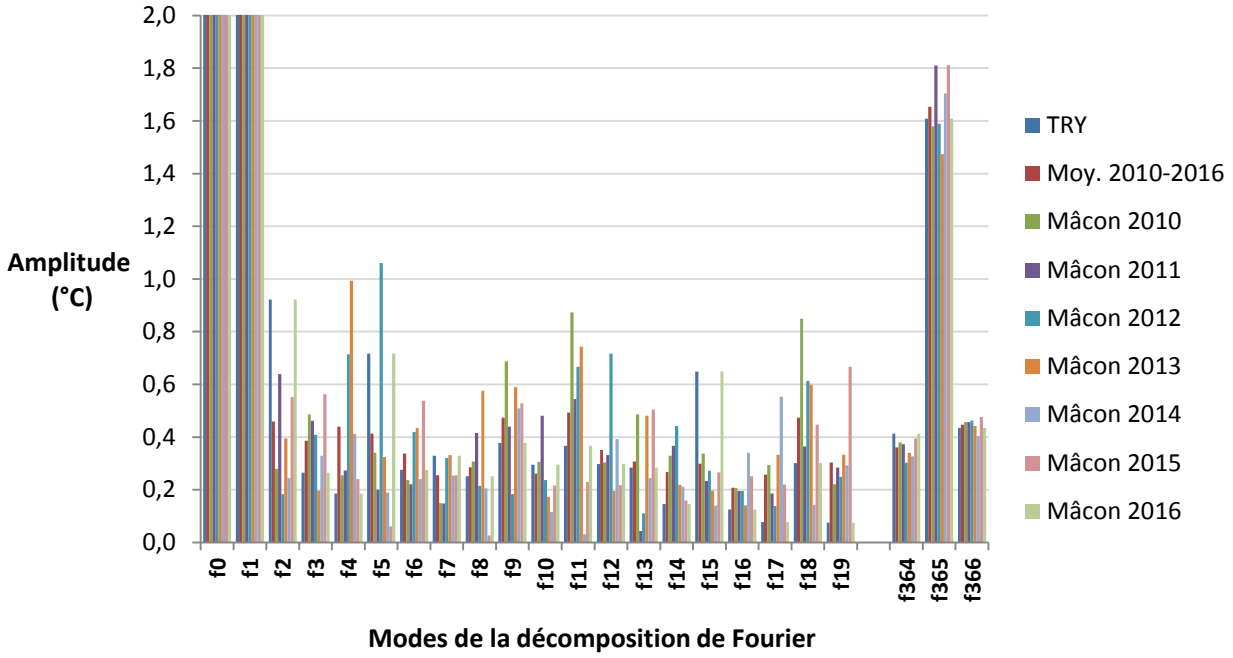

Figure A-1 : Amplitudes des fonctions issues de la décomposition de série de Fourier pour les modes 0 à 19 et 364 à 366

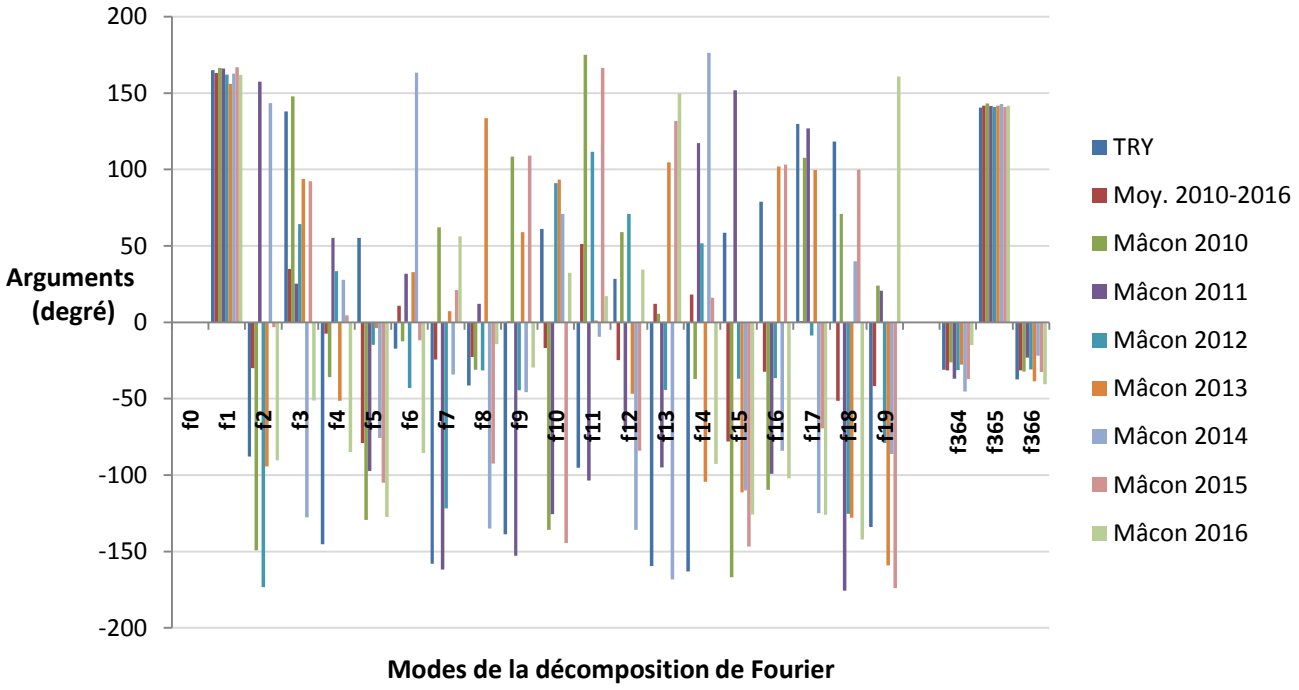

Figure A-2 : Arguments des fonctions issues de la décomposition de série de Fourier pour les modes de 0 à 19 et de 364 à 366

A-2-3 Illustration de la méthode de génération des données de RGH

La Figure A-3, la Figure A-4 et la Figure A-5 illustrent la méthode de génération des données de rayonnement présentée au paragraphe 2.4.2. Suite à la transformation de Fourier discrète effectuée sur la série de RGH horaire, la part résiduelle subit un traitement afin d'obtenir une série stationnaire, modélisable selon un processus autorégressif (Figure A-3). Ce traitement repose en partie sur l'exploitation de la série de RGH extraterrestre de laquelle est déduite la série des résidus maximaux théoriques par soustraction de la série tendancielle identifiée, à chaque heure. En moyennant $RGH_{h.res.max}$ sur chaque jour, on obtient $RGH_{j.res.max}$. En normalisant $RGH_{h.res.max}$ par $RGH_{j.res.max}$, on obtient aussi la série $RGH_{h.res.max.norm}$ qui décrit la forme typique des résidus au pas de temps horaire, leur moyenne journalière (ou intégrale) valant 1. On constate ainsi que les maximums journaliers de cette série sont inférieurs en été. En effet, les temps d'éclairement étant plus importants, les paraboles journalières sont plus larges et donc moins hautes, leur moyenne étant fixée à 1.

Le traitement du résidu passe par le changement de pas de temps et le calcul de la moyenne journalière de ces résidus (série $RGHj_{.res}$) pour s'affranchir des problématiques dues à la nuit (Figure A-4). Du fait de la forme de cette série dont les valeurs estivales sont plus extrêmes, nous la normalisons par la série $RGH_{j.res.max}$ obtenant finalement une série stationnaire exploitable $RGH_{j.res.norm}$.

Ce sont les caractéristiques de cette série qui sont utilisées pour re-générer des séries synthétiques similaires (Figure A-5). Ces séries brutes subissent un traitement inverse, elles sont dénormalisées par une multiplication terme à terme avec la série des $RGH_{j.res.max.}$ pour repasser du pas de temps journalier au pas de temps horaire. Les séries des résidus journaliers moyens modulent la série du résidu maximal horaire normalisé. La forme parabolique typique est ainsi modulée par la moyenne journalière générée pour chaque jour de l'année et d'une manière différente pour chaque génération.

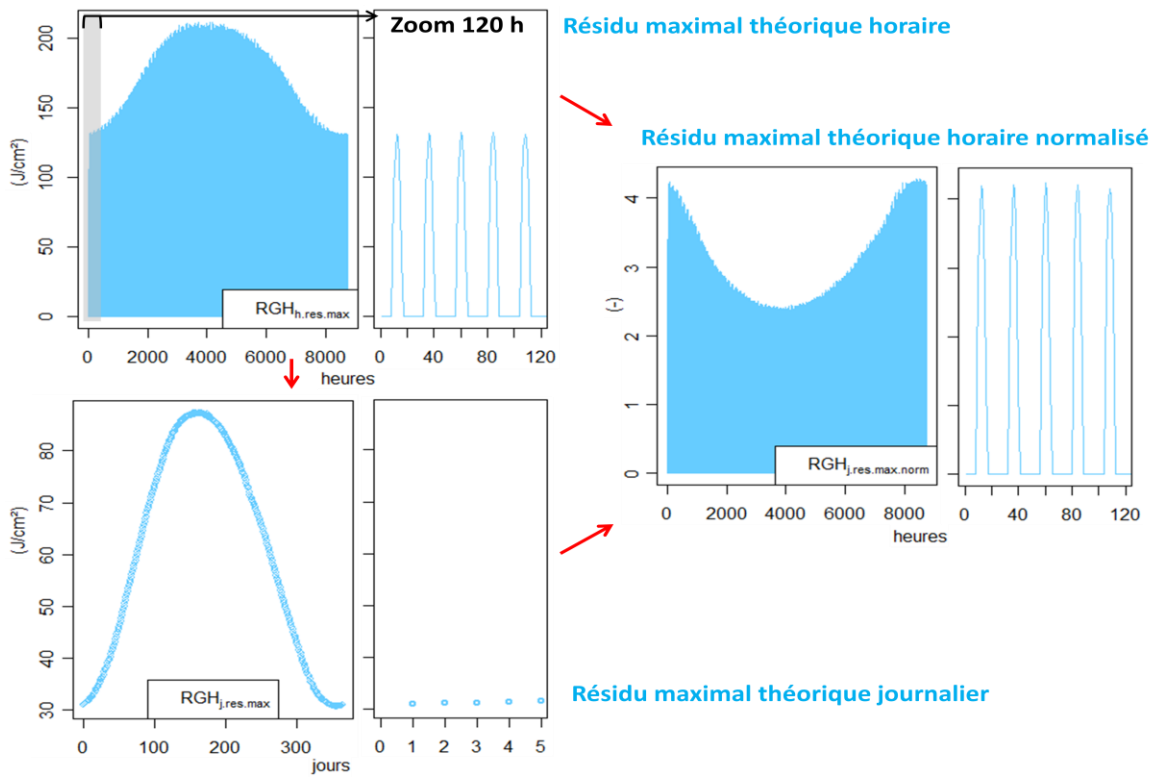

Figure A-3 : Exploitation du rayonnement global extraterrestre afin de calculer la série du résidu maximal théorique horaire, moyennée sur chaque jour, et horaire normalisé

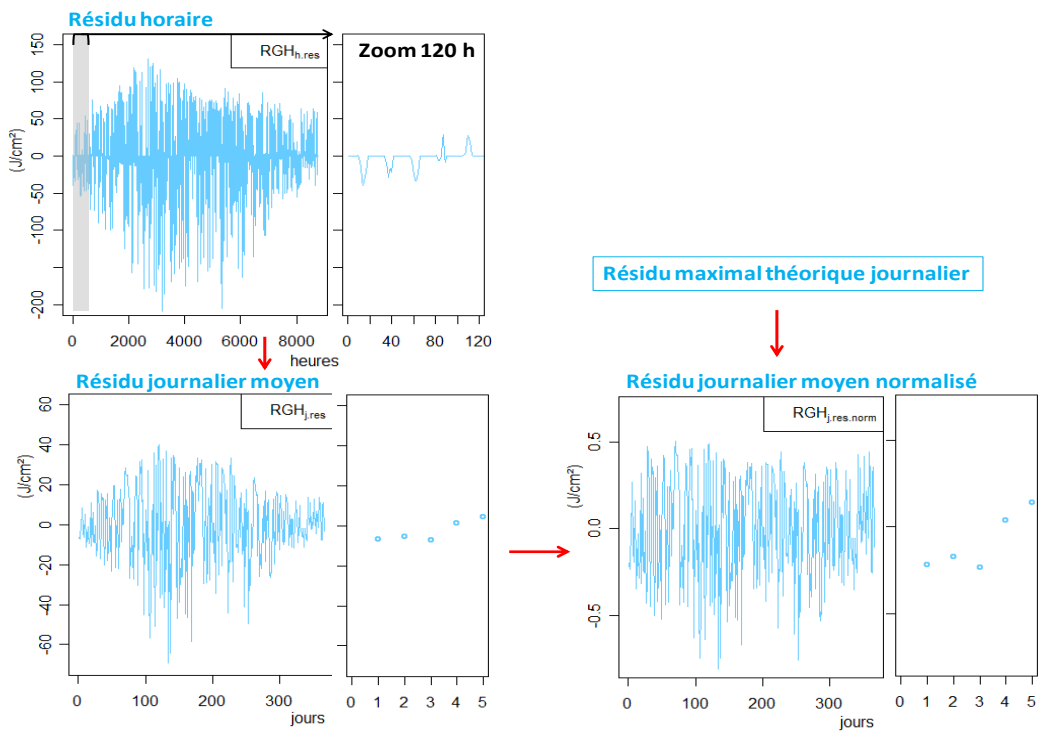

Figure A-4 : Obtention de la série stationnaire des résidus journaliers normalisés identifiable selon un modèle ARMA

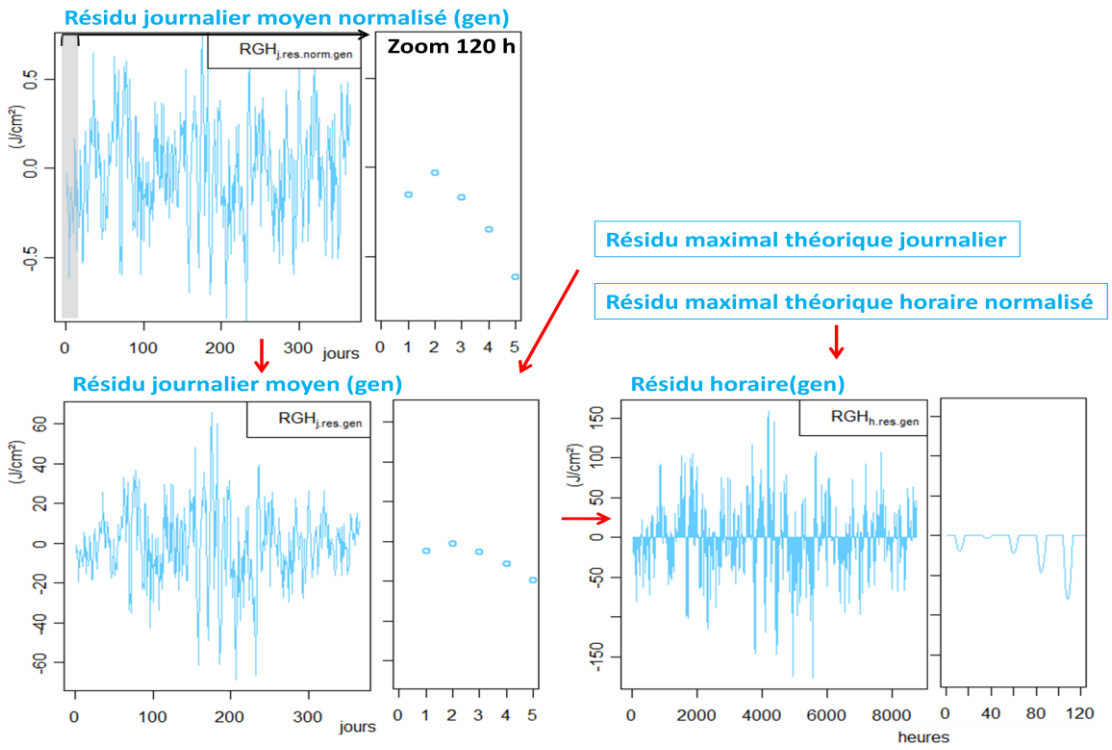

Figure A-5 : Dénormalisation de la série stationnaire générée et calcul de la série résiduelle finale

A-2-4 Calcul du rayonnement solaire extra-terrestre

L'ensemble des formules permettant le calcul de la série de rayonnement global extra-terrestre sont extraites du cours "Calculs astronomiques simplifiés" de l'université Paris-Sud Orsay - Master PAM - Energies renouvelables - Energie solaire thermique de Vincent Bourdin (2014).

Le temps solaire vrai, H, est tout d'abord calculé

$$H = H_l - \Delta H_l + \Delta H_g - E$$

avec H_l l'heure locale administrative, ΔH_l l'avance de l'heure locale sur l'heure normale du fuseau horaire, ΔH_g la correction due à la longitude considérée et E l'équation du temps. L'équation du temps donne, en minutes, l'avance du temps moyen sur le temps solaire, pour une Terre non perturbée :

$$E = 7{,}53\, cos(B) + 1{,}5\, sin(B) - 9{,}87 sin(2B)$$

avec $B = \frac{2\pi(J-81)}{365}$

La déclinaison, δ, est la position angulaire du soleil à midi solaire (i.e. quand le soleil est sur le méridien local) par rapport au plan de l'équateur.

$$\delta = 23{,}45.\,sin(\frac{2\pi(J+284)}{365})$$

L'angle horaire, ω, correspond à la mesure du déplacement angulaire du soleil à l'est ou à l'ouest du méridien local dû à la rotation de la Terre sur son axe de 15° par heure.

$$\omega = 15.\,(12 - H)$$

θ est l'angle entre la direction du rayonnement solaire et la direction d'intérêt :

$$cos(\theta) = cos(\varphi)\,cos(\omega) + sin(\delta)\,sin(\varphi)$$

avec φ la latitude.

$RGH_{E.T.perp}$ est le rayonnement fourni par le soleil sur une surface perpendiculaire hors atmosphère :

$$RGH_{E.T.perp} = G(1 + 0{,}034\,cos\left(\frac{2\pi J}{365}\right))$$

avec G = 1367 W/m², la constante solaire.

En tenant compte de l'angle d'incidence par rapport à la normale au sol, on obtient finalement le rayonnement global horizontal extra-terrestre :

$$RGH_{E.T} = RGH_{E.T.perp}.\,cos(\theta)$$

A-2-5 Interventions algorithmiques

Le processus de génération des données horaires de température, de RGH et de RDH est décrit au paragraphe 2.4. La Figure A-6 récapitule les différentes étapes de l'algorithme.

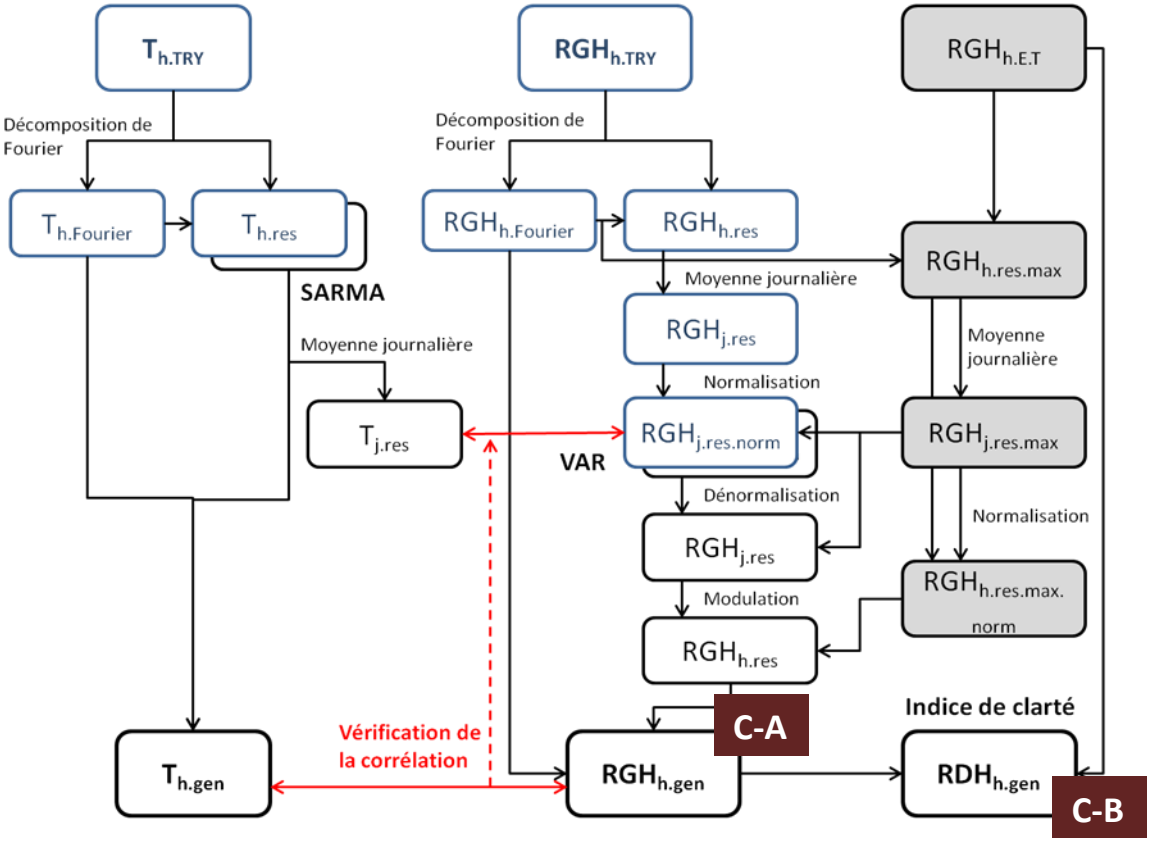

Figure A-6 : Génération des séries de température et de rayonnement solaire

Plusieurs corrections et étapes supplémentaires s'intègrent au processus et doivent être signalées. Elles se situent dans la boucle de génération de données de RGH corrélées aux données de températures précédemment générées (C-A sur la Figure A-6) Ces interventions sont liées à la manipulation de valeurs aléatoires provoquant ponctuellement des données non-physiques. Ainsi, les modifications suivantes sont apportées :

- Si le RGH horaire (tendance + résidus) est négatif, les valeurs horaires correspondantes sont fixées à 0.
- Si le RGH horaire est supérieur à 0,8 fois le FGH extra-terrestre, il est fixé à cette dernière valeur. Ce coefficient correspond à l'indice de clarté maximal, à la part de rayonnement atteignant effectivement la surface terrestre.

Dans le processus de génération de séries de RDH, des corrections sont faites sur les indices de clarté, calculés comme le rapport RGH/RGH_{ext} (C-B sur la figure précédente). De légers décalages temporels peuvent entrainer des divergences de ce rapport aux premières ou dernières heures du jour, lorsque le lever et le coucher du soleil a lieu proche d'un début ou fin de pas de temps. Ainsi, l'algorithme est corrigé de la façon suivante, pour les valeurs divergentes :

- Si au pas de temps précédent, il faisait jour ($k_t > 0$), cette valeur est conservée.
- Si au pas de temps précédent il faisait nuit, la valeur de k_t de l'heure suivante est considérée.

Ces corrections ont un impact négligeable car les quelques pas de temps concernés correspondent à des heures très peu ensoleillées (lever ou coucher du soleil).

A-3-1 Incertitude sur les DJU calculés avec la méthode COSTIC

L'incertitude sur les degrés-jours unifiés calculés avec la méthode COSTIC (§ 4.3.2.2) peut s'exprimer de la façon suivante :

$$u(DJU) = \sum_{i=1}^{N_{jours}} \sqrt{\left[\frac{\partial DJU}{\partial T_{max,i}} . u(T_{max,i})\right]^2 + \left[\frac{\partial DJU}{\partial T_{min,i}} . u(T_{min,i})\right]^2}$$

avec

$$\left[\frac{\partial DJU}{\partial T_{max,i}} . u(T_{max,i})\right]^2 = \left[0{,}08 + \frac{0{,}42}{(T_{max} - T_{min})^2} * (-2T_{ref}T_{min,i} - {T_{ref}}^2 + 2T_{min,i}T_{max,i} - {T_{max,i}}^2)\right]^2$$

$$\left[\frac{\partial DJU}{\partial T_{min,i}} . u(T_{min,i})\right]^2 = \left[0{,}08 + \frac{0{,}42}{(T_{max} - T_{min})^2} * (2T_{ref}T_{max,i} - {T_{ref}}^2 - {T_{max,i}}^2 T_{ref})\right]^2$$

A-5-1 Spécifications nominales des modèles SED de Feyzin avant et après réhabilitation

- *Modélisation du bâtiment avant réhabilitation*

Les tableaux Tableau A-1, Tableau A-2, Tableau A-3, Tableau A-4 et Tableau A-5 décrivent les caractéristiques principales du bâtiment modélisé dans la configuration avant réhabilitation.

Tableau A-1 : Caractéristiques des matériaux utilisés avant réhabilitation

Matériaux	Conductivité (W/m².K)	Masse volumique (kg/m3)	Chaleur spécifique (Wh/kg.K)
Polyuréthane	0,03	35	0,233
Béton lourd	1,75	2300	0,256
Polystyrène expansé	0,039	25	0,383
Platre courant	0,35	1000	0,222

Tableau A-2 : Épaisseur et composition des parois avant réhabilitation

Parois	*Matériau*	*épaisseur (cm)*
Terrasse	Béton lourd	20
	Polyuréthane	5
Plancher svs apres travaux	Beton lourd	16
	Polystyrène expansé	3
Mur ext	Béton lourd	15
	Polystyrène expansé	7
	Plâtre courant	1
Mur int	Béton lourd	15
	Plâtre courant	1

Tableau A-3 : Caractéristiques des menuiseries avant réhabilitation

Menuiseries	*Conductivité (W/m².K)*	*Largeur (m)*	*Hauteur (m)*
Porte	3,5	0,83	2,04
GPEF	1,9	1,6	1,3
GPEDF	1,9	1,6	2

Tableau A-4 : Ponts thermiques avant réhabilitation

Ponts thermiques	*linéiques (W/(ml/K)*	*menuiseries (W/K)*
Oss Bois - Pl haut/Mur	1	
Oss Bois - Angle rentrant	0,06	
Oss Bois - Angle sortant	0,11	
Plancher intermédiaire	1	
Refends en T	1	
Plancher bas	1	
ITE-appui fenêtre béton		0,4
ITE tableau / linteau béton		0,04

Tableau A-5 : Débits d'échanges d'air avant réhabilitation

Echanges d'air (vol/h)	
Infiltrations	0,55
Scénario ventilation	0,67

- *Modélisation du bâtiment après réhabilitation*

Les tableaux Tableau 5-17, Tableau A-7, Tableau A-8, Tableau A-9 et Tableau A-10 décrivent les caractéristiques principales du bâtiment modélisé dans la configuration après réhabilitation. Les lignes grisées correspondent aux modifications induites par les travaux.

Tableau A-6 : Caractéristiques des matériaux utilisés après réhabilitation

Matériaux	*Conductivité (W/m².K)*	*Masse volumique (kg/m3)*	*Chaleur spécifique (Wh/kg.K)*
Polyuréthane	0,03	35	0,233
Béton lourd	1,75	2300	0,256
Polystyrène expansé	0,039	25	0,383
Platre courant	0,35	1000	0,222
GPE-PU	0,03	35	0,233
GPE-VRAC	0,035	35	0,286
GPE-PSE	0,038	50	0,403
GPE-BL	1,75	2300	0,256

Tableau A-7 : Épaisseur et composition des parois après réhabilitation

Parois	*Matériau*	*épaisseur (cm)*
Terrasse apres travaux	GPE-BL	20
	GPE-PU	10
Plancher svs apres travaux	GPE-BL	16
	GPE-VRAC	8
Mur ext apres travaux	GPE-PSE	10
	GPE-BL	15
	Polystyrène expansé	7
	Plâtre courant	1
Mur int	Béton lourd	15
	Plâtre courant	1

Tableau A-8 : Caractéristiques des menuiseries après réhabilitation

Menuiseries	**Conductivité (W/m².K)**	**Largeur (m)**	**Hauteur (m)**
Porte	3,5	0,83	2,04
GPEF	1,9	1,6	1,3
GPEDF	1,9	1,6	2

Tableau A-9 : Ponts thermiques après réhabilitation

Ponts thermiques	*linéiques (W/(ml/K)*	*menuiseries (W/K)*
Oss Bois - Angle rentrant	0,06	
Oss Bois - Angle sortant	0,11	
Plancher intermédiaire	0,1	
Terrasse	0,74	
Plancher bas	0,79	
ITE-appui fenêtre béton		0,4
ITE tableau / linteau béton		0,04

Tableau A-10 : Débits d'échanges d'air après réhabilitation

Echanges d'air (vol/h)	
Infiltrations	0,15
Scénario ventilation	0,28

A-5-2 Comparaison des modèles de $LGC_{5\%}$ pour différents FA et PMV

Le Tableau A-11 ci-dessous présente les coefficients R^1 associés à différents modèles de régression quantile linéaire où deux niveaux d'incertitude de mesure (*PMV-A,* en vert ou *PMV-B* en rouge) ont été pris en compte pour les FA sélectionnés. Les configurations *AB* ou *BA* concernent les DH_{Tint} nécessitant la mesure des températures intérieures et extérieures. Comme signalé au § 5.1.4, des inversions illogiques (où par exemple, pour une même configuration de FA, un PMV-B va être préféré à un PMV-A) peuvent être ponctuellement observées du fait de l'incertitude statistique des coefficients et de faibles différences entre certains modèles.

Tableau A-11 : Comparaison des coefficients R^1 des modèles de régression quantile pour différents jeux de FA et différents niveaux d'incertitudes de mesure virtuelle

Conso. Ch.	DHTint	DH18.sc	Eelec.sc	Tmp.int.sc	Tmp.cons.sc	Tm.ext.sc	R1
A			A	A	A	A	0,592
A			A	A	B	A	0,592
A			B	A	B	B	0,592
A			A	A	A	B	0,592
A			A	A	B	B	0,591
A			B	A	B	A	0,591
A			B	B	B	A	0,591
A			A	A		A	0,591
A			A	A		B	0,591
A			B	A	A	B	0,591
A			A	B	B	A	0,591
A			A	B	A	A	0,590
A			A	B	B	B	0,590
A			B	A	A	A	0,590
A			B	A		A	0,590
A			A	B	A	B	0,590
A			B	B	A	A	0,590
A			B	A		B	0,589
A			B	B	A	B	0,589
A			B	B	B	B	0,589
A			A	B		A	0,588
A			B	B		A	0,588
A			A	B		B	0,587
A			B	B		B	0,587
B			B	B	B	A	0,580
B			B	A	B	B	0,580
B			A	A	B	A	0,580
B			A	A	A	B	0,579
A	A		A		A	A	0,579
A	A		B		A	A	0,579
B			A	A	A	A	0,578
B			B	A	B	A	0,578
B			B	A	A	B	0,578
B			A	A	B	B	0,578
B			A	A		B	0,578
A	AB		A		A	B	0,578

Conso. Ch.	DHTint	DH18.sc	Eelec.sc	Tmp.int.sc	Tmp.cons.sc	Tm.ext.sc	R1
B			A	B	B	A	0,577
B			A	A		A	0,577
B			A	B	B	B	0,577
A	A		A		B	A	0,577
A	AB		B		A	B	0,577
A	BA		A		A	A	0,577
A	A		B		B	A	0,577
B			A	B	A	A	0,577
A	B		A		A	B	0,577
A	BA		B		A	A	0,576
A	BA		B		B	A	0,576
A	B		B		A	B	0,576
B			B	B	A	B	0,576
B			B	A	A	A	0,576
B			B	A		A	0,576
A	B		B		B	B	0,576
B			B	B	B	B	0,576
B			B	A		B	0,576
A			A		A	A	0,576
B			B	B	A	A	0,576
A	AB		B		B	B	0,575
B			A	B	A	B	0,575
A	BA		A		B	A	0,575
A			A		A	B	0,575
A			B		A	A	0,575
A	AB		A		B	B	0,575
B			B	B		A	0,575
A			A		B	B	0,575
B			A	B		B	0,575
A			B		B	A	0,574
B			A	B		A	0,574
B			B	B		B	0,574
A			A		B	A	0,574
A			B		A	B	0,573
A			B		B	B	0,572
B	A		B		A	A	0,569
B	A		A		A	A	0,568
B	AB		A		A	B	0,567
B	BA		A		A	A	0,566
B	B		A		A	B	0,566
B	A		A		B	A	0,566
B	A		B		B	A	0,566
B	BA		B		B	A	0,565
B	B		B		A	B	0,565
B	AB		B		A	B	0,565
B			A		A	A	0,565
B	B		B		B	B	0,565
B			B		A	A	0,564
B	BA		B		A	A	0,564
B			A		A	B	0,564
B	BA		A		B	A	0,564
B			A		B	B	0,564
B	AB		A		B	B	0,564
B	AB		B		B	B	0,563
B			A		B	A	0,563
B			B		B	A	0,562
B			B		A	B	0,562
B			B		B	B	0,561
A	A	A	B		B		0,561
A	A	A	B		A		0,560
A	A	A	A		A		0,558
A	A	A	A		B		0,558

Conso. Ch.	DHTint	DH18.sc	Eelec.sc	Tmp.int.sc	Tmp.cons.sc	Tm.ext.sc	R1
A	BA	A	A		A		0,554
A				A		A	0,553
A	BA	A	A		B		0,552
A				B		A	0,552
A				A		B	0,552
A					A	B	0,552
A					A	A	0,551
A					B	A	0,551
A	BA	A	B		B		0,551
A				B		B	0,551
A	BA	A	B		A		0,550
B	A	A	B		B		0,549
A					B	B	0,549
B	A	A	B		A		0,548
A		A	A		A		0,548
B	A	A	A		A		0,548
B	A	A	A		B		0,547
A		A	A	A	A		0,547
A		A	A	A	B		0,546
B	BA	A	A		A		0,544
A		A	A		B		0,544
B				B		A	0,544
A		A	B		A		0,544
A		A	B	A	A		0,543
B				A		A	0,543
B	BA	A	A		B		0,543
B				A		B	0,543
B					B	A	0,543
A		A	A	B	A		0,542
A		A	B	B	B		0,542
B					A	B	0,542
A		A	B	B	A		0,542
B	BA	A	B		B		0,541
B				B		B	0,541
A		A	B		B		0,541
A		A	A	B	B		0,541
A		A	B	A	B		0,541
B					A	A	0,540
B		A	A		A		0,540
B					B	B	0,539
B	BA	A	B		A		0,538
B		A	A	A	B		0,538
B		A	A	A	A		0,536
B		A	A		B		0,534
B		A	B	A	A		0,534
B		A	B		A		0,534
B		A	A	B	A		0,533
A		A	A	A			0,533
B		A	B	B	A		0,532
B		A	B		B		0,532
B		A	B	B	B		0,532
B		A	A	B	B		0,532
B		A	B	A	B		0,530
A		A	A	B			0,530
A		A	B	A			0,529
A		A	B	B			0,527
A		A			A		0,524
B		A	A	B			0,523
B		A	A	A			0,522
A		A			B		0,521
B		A	B	A			0,520
B		A	B	B			0,518

Conso. Ch.	DHTint	DH18.sc	Eelec.sc	Tmp.int.sc	Tmp.cons.sc	Tm.ext.sc	R1
B		A			B		0,514
B		A			A		0,514
A		A		A			0,509
A	AB	B	B		A		0,507
A	AB	B	B		B		0,506
A	AB	B	A		B		0,506
A		A		B			0,506
A	AB	B	A		A		0,502
B		A		A			0,501
A	A		A		A		0,501
A	A		A				0,500
A	A		B		A		0,499
B	AB	B	B		B		0,499
B	AB	B	B		A		0,498
B		A		B			0,498
A	A		B		B		0,498
B	AB	B	A		B		0,497
A	A		A		B		0,496
A	A		B				0,495
A	B	B	A		A		0,494
B	A		B		A		0,492
B	A		A				0,492
A	B	B	B		A		0,492
B	AB	B	A		A		0,492
A	B	B	B		B		0,492
B	A		A		A		0,491
B	A		B		B		0,490
B	A		A		B		0,487
B	A		B				0,487
B	B	B	A		A		0,486
B	B	B	B		A		0,486
B	B	B	B		B		0,485
A		B	A	A	B		0,475
A	A						0,474
B	A						0,470
B		B	A	A	B		0,470
A		B	A	B	B		0,469
A		B	A	B	A		0,469
A		B	A	A	A		0,469
A		B	B	A	A		0,469
A		B	B	A	B		0,467
A		B	B	B	B		0,467
A		B	B	B	A		0,466
A		B	B		B		0,466
B		B	B	A	A		0,465
B		B	A	B	B		0,464
B		B	A	A	A		0,462
A		B	A		B		0,462
B		B	A	B	A		0,462
B		B	B	A	B		0,461
A		B	A		A		0,460
B		B	B	B	B		0,460
A			A			A	0,459
A			B			B	0,459
B		B	B		B		0,459
A		B	B		A		0,458
A			B			A	0,458
A			A			B	0,458
B		B	B	B	A		0,458
B		B	A		B		0,455
B			A			A	0,454
B		B	A		A		0,454

Conso. Ch.	DHTint	DH18.sc	Eelec.sc	Tmp.int.sc	Tmp.cons.sc	Tm.ext.sc	R1
A			A			A	0,453
A		A	A		B		0,453
A			B			A	0,450
A			A			B	0,450
B		A			B		0,449
B		A			A		0,447
A		A			B		0,443
A		A			A		0,442
B		B	A				0,442
B						B	0,439
B						A	0,438
A		B	A				0,436
B		B	B				0,436
A						B	0,433
A						A	0,432
A		B	B				0,431
B		A	B	A			0,430
B		A	A	A			0,428
B		A	A	B			0,427
A		A	B	A			0,425
A		A	A	A			0,423
B		A	B	B			0,422
A		A	A	B			0,421
B		B					0,417
A		A	B	B			0,414
A		B					0,412
B		A		B			0,404
B		A		A			0,401
A		A		B			0,398
A		A		A			0,395
B		A	A				0,378
A		A	A				0,373
B		A	B				0,373
B	A		A		B		0,369
B	A		B		A		0,369
A		A	B				0,368
B	A		A				0,367
B	A		A		A		0,366
A	A		B		A		0,365
B	A		B				0,365
A	A		A		B		0,364
A	A		A				0,363
A	A		A		A		0,362
B	A		B		B		0,362
A	A		B				0,360
A	A		B		B		0,357
B		A					0,354
A		A					0,349
B	A						0,346
A	A						0,342
B					A		0,060
B					B		0,059
A					A		0,059
A					B		0,058
A			B				0,026
B			B				0,024
B			A				0,023
A			A				0,023
A				A			0,000
B				A			0,000
B				B			0,000
A				B			0,000

A-5-3 Schémas des installations techniques du bâtiment Izuba

La Figure A-7 ci-dessous décrit de manière schématique le système de chauffage et de climatisation du bâtiment Izuba. Le circuit en amont de la pompe à chaleur comporte un échangeur géothermique profond. La PAC réversible alimente un circuit de distribution constitué d'un ballon tampon et de planchers chauffants et rafraichissants. Des ventilo-convecteurs permettent un appoint pour la climatisation de la salle de réunion. La Figure A-8 décrit le système de production d'ECS solaire avec appoint électrique.

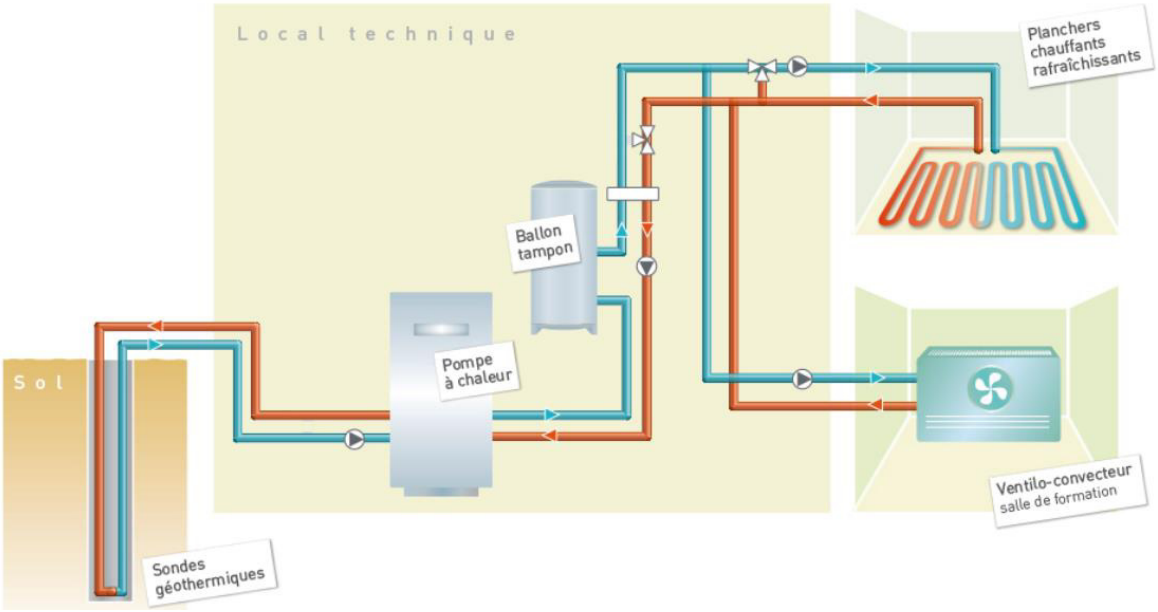

Figure A-7 : Production de chaleur et de froid par PAC sur sonde géothermique et émission par plancher chauffant (chauffage/climatisation) et ventilo-convecteurs (climatisation)

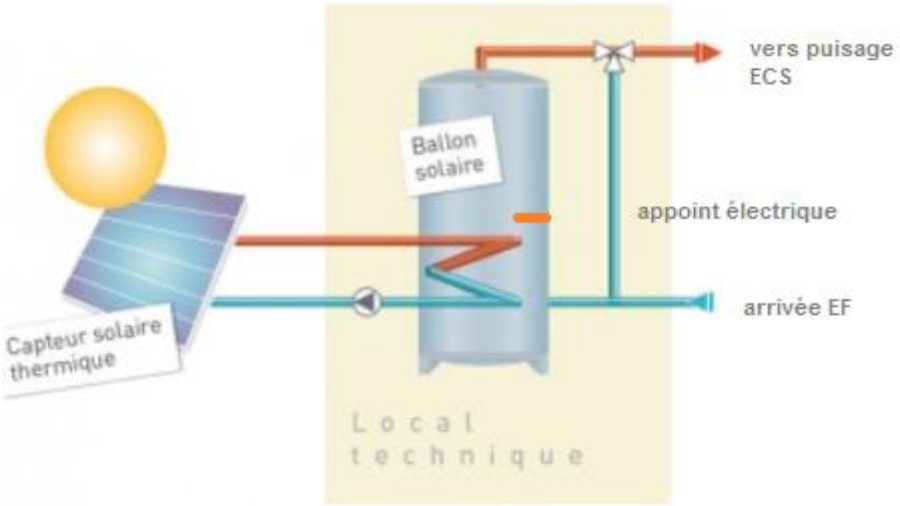

Figure A-8 : Installation de production d'ECS

A-5-4 Caractéristiques principales du modèle SED du bâtiment Izuba

La Figure A-9 présente les plans et le zonage des pièces du bâtiment de bureaux d'Izuba. 18 zones thermiques sont considérées, découpées en fonction des usages et occupations.

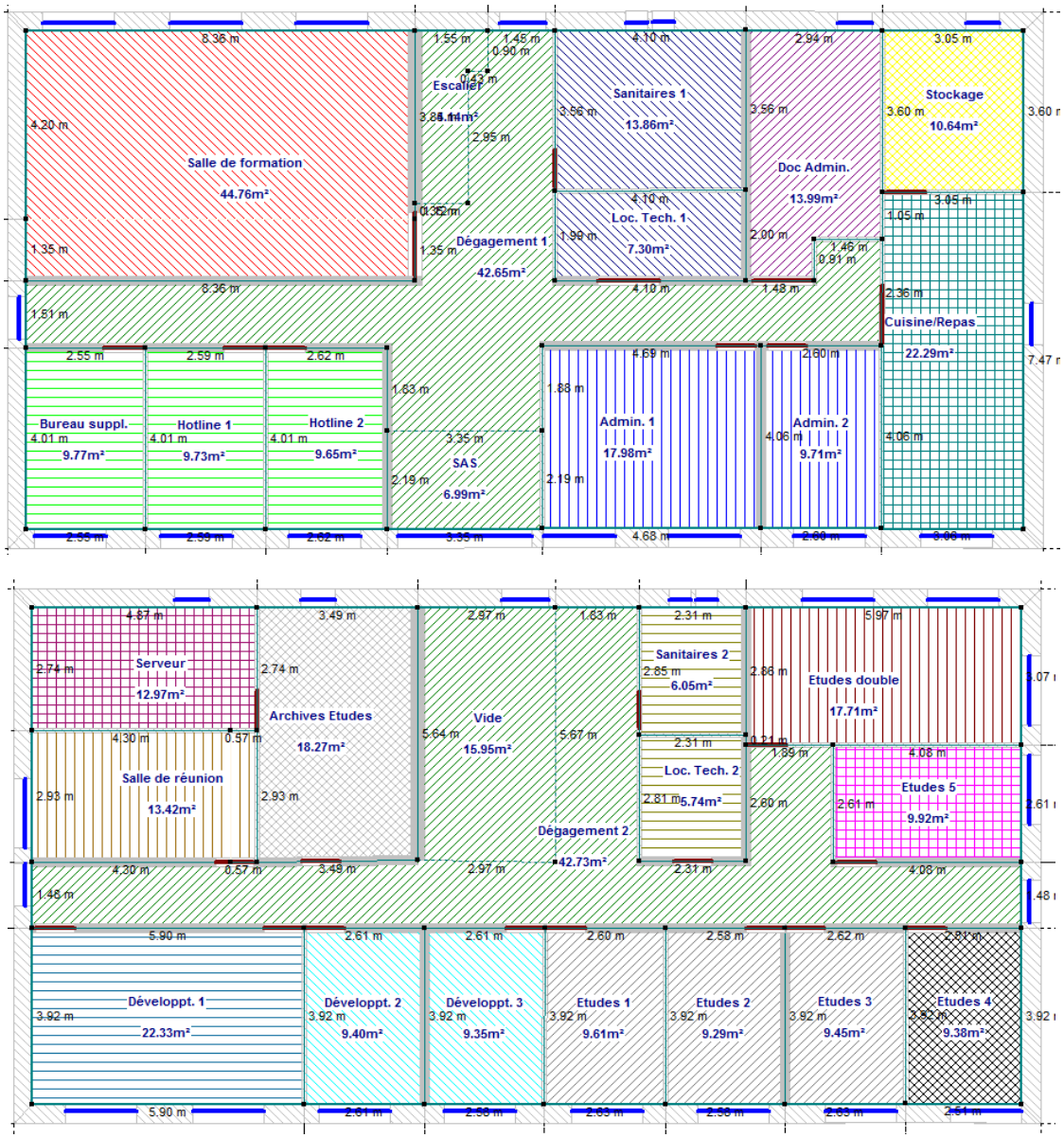

Figure A-9 : Plan et zonage du RDC (haut) et du premier étage (bas) du bâtiment

Les tableaux Tableau A-17, Tableau A-18 et Tableau A-19 décrivent les caractéristiques nominales du modèle de simulation énergétique dynamique.

Tableau A-12 : Caractéristiques des matériaux du bâtiment Izuba

Matériaux	Conductivité (W/m².K)	Masse volumique (kg/m³)	Chaleur spécifique (J/kg.K)
polystyrène expansé	0,029	35	1188
Fibre de bois	0,15	800	2093
Mortier	1,15	2000	840
Béton lourd	1,75	2300	920
Terre crue	0,85	2000	1008
Laine de bois	0,038	40	2100
Enduit chaux	0,7	140	1000
Fermacell	0,36	1150	1623
Panneau OSB	0,13	650	1700
Paille compressée	0,065	120	1400
Bois léger	0,15	500	1200
FibraFutura	0,08	480	1800

Tableau A-13 : Caractéristiques des menuiseries et ponts thermiques du bâtiment Izuba

Menuiseries	Uw horizontal (W/m².K)	Largeur (m)	Hauteur (m)
Porte intérieure	5	0,83	2,04
Porte Sud	1,86	3,15	2,15
Fen. escalier haut	1,862	1	2
Fen. escalier bas	1,85	1	1,96
Porte fen. est/ouest	1,776	1,12	2,49
Fen Sanitaires	1746	0,47	0,65
Fen bureaux	1,845	1,54	1,4
Fen 1 battant	1,785	0,74	1,4

Ponts thermiques (W/m.K)	
angle sortant	0,04
meur refend	0,05
pl bas / mur ext	0,332
pl haut / mur ext	0,065
pl int / mur ext	0,068
appui fenêtre	0,172
linteau	0,167
tableau	0,158
mur ext	0,005
toiture	0,015

Tableau A-14 : Compositions et épaisseurs des parois du bâtiment Izuba

Parois	Matériaux	Epaisseur (cm)
Mur ossature bois (refend)	Fermacell	1
	Laine de bois	12
	Lame d'air	5
	Panneau OSB	1
	Fermacell	1
Mur enduit terre (cloison)	Pans de bois torchis	10
Cloison platre (cloison)	Fermacell	1
	Laine de bois	8
	Fermacell	1
Plancher intermédiaire	Mortier	7
	Fibre de bois	2
	Panneau OSB	1,8
	Lame d'air	6,7
	FibraFutura	2
Plancher bas	Polystyrène expansé	11
	Béton lourd	20
	Mortier	6
Mur terre crue (refend)	Terre crue	29,5
mur isolé (façade)	Enduit chaux	3
	Panneau OSB	2
	Paille compressée	37
	Panneau OSB	2
	Fermacell	1
toiture	Panneau OSB	1,8
	Paille compressée	37
	Bois léger	3
	Lame d'air	1,5
	Fibrafutura	2

A-5-5 Dispositif de mesure des performances du bâtiment Izuba

Les mesures sont disponibles à un pas de temps de 15 min. Une partie des données sont mesurées à partir de mai 2015 mais la plupart sont collectées depuis avril 2016. La majeure partie des acquisitions ont été stoppées involontairement durant quelques semaines de l'été 2016. Le Tableau A-15 détaille les grandeurs mesurées sur le bâtiment.

Tableau A-15 : Dispositif de mesure installé sur le bâtiment

Compteurs électriques	Capteurs PAC	Autres capteurs
Compteur général	Energie chaud sonde	Température intérieure (8)
Compteur général étage	Energie froid sonde	Température extérieure
Compteur général RDC	Energie chaud prod	Rayonnement direct
Energie PAC	Energie froid prod	Temp. Depart plancher
Energie ventilo-convecteur	Puissance sonde	Temp. retour plancher
Energie ECS	Puissance prod	Température sol
Energie éclairage étage	Temp. Entrée PAC	pompe chauffage M13
Energie Eéclairage RDC	Temp. Sortie PAC	Temp.entrée ECS
Energie cuisine	débit sond	Temps. Sortie ECS
energie ventilateur	débit prod	Température extraction DF
energie CTA	Puissance PAC	Température souflage DF

Les températures intérieures et extérieures mesurées, exploitées pour calculer les facteurs d'ajustement, sont présentées à la Figure A-10.

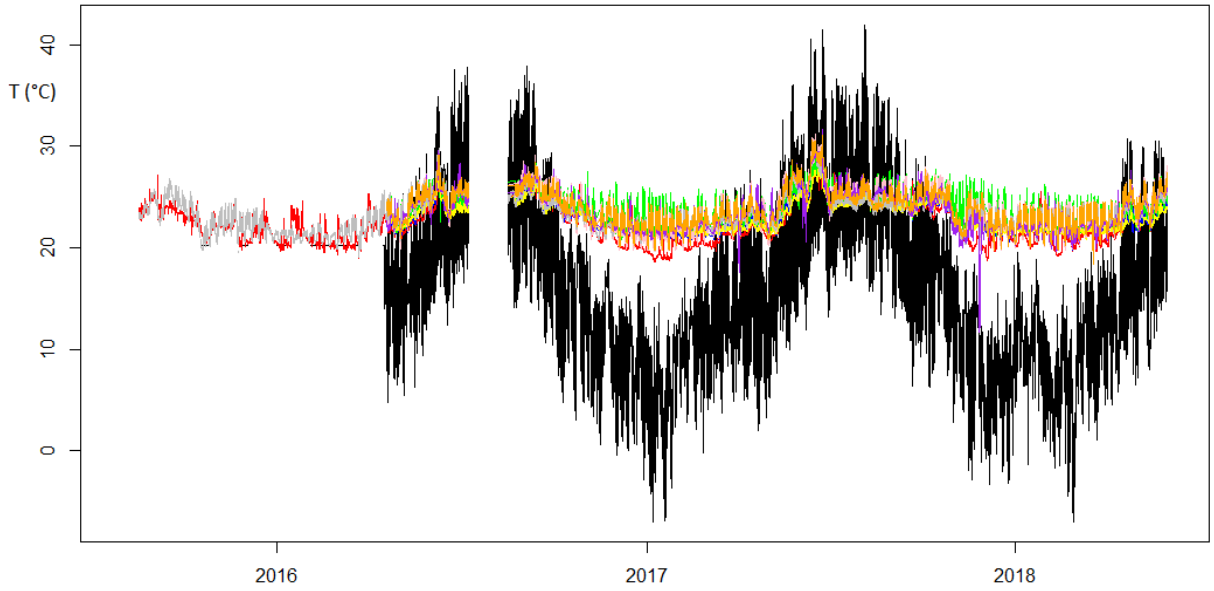

Figure A-10 : Mesures brutes des températures extérieure (noire) et intérieures (couleurs)

A-5-6 Incertitudes de mesure des températures intérieures moyennes

Les calculs présentés ici visent à estimer une incertitude sur la température intérieure moyenne estimée à partir des mesures réalisées sur le bâtiment d'Izuba. Huit relevés de température au pas de temps de 15 minutes sont réalisés. Quatre capteurs sont placés à chaque étage dans des pièces réparties aux quatre coins du bâtiment.

Ces éléments s'appuient sur les recommandations du GUM (JCGM, 1995) et sur les travaux de Caucheteux (2015) au sein du projet Omega. Plusieurs hypothèses peuvent être faites pour simplifier le problème. On suppose en particulier l'absence de stratification et le bon étalonnage des capteurs. On démontre, lorsque le nombre de données n est inférieur à 30 que la variable :

$$T = \frac{T_{bat-ech} - T_{bat}}{\sigma_{ech}/\sqrt{n-1}}$$

suit une loi de Student à $(n-1)$ degrés de libertés. T_{bat} désigne la température moyenne « vraie » du bâtiment. $T_{bat-ech}$ est l'estimation issue de l'échantillonnage calculée comme la moyenne des échantillons. L'incertitude dépend donc du nombre de capteurs et de la dispersion des mesures. Grâce aux tables de la loi de Student, cette formule fournit une estimation de l'incertitude de type A ($u_{a.i}$) de la température moyenne instantanée.

En supposant une pondération surfacique équilibrée entre les huit capteurs, hypothèse acceptable étant donnée leur répartition, on calcule l'incertitude de type A sur la moyenne temporelle comme (§ 3.4.2) :

$$u_a(\overline{T}) = \sqrt{\sum_{i=1}^{n} \left[\frac{1}{n} u_{a.i}(T_i)\right]^2}$$

L'incertitude sur la moyenne temporelle et spatiale de la température intérieure peut ainsi être estimée à chaque simulation.

A-5-7 Comparaison des modèles de régression quantile

Dans le Tableau A-16, une comparaison de différents modèles de régression quantile à 95 % est effectuée. Les coefficients R^1 des modèles caractérisent la qualité de l'ajustement. Cette étude permet donc de déterminer les meilleurs jeux de FA.

Tableau A-16 : Comparaison des coefficients R^1 de différents modèles de régression quantile pour une variété de jeux de facteurs d'ajustement

DH.Tint.sc		DH.Tint.sf		Tm.ext.sc		Tm.ext.sf		Tm.int.sc		Tmp.int.sc		Tm.int.sf		Tmp.int.sf		
42_18	44_13	25_32	23_35	42_18	44_13	25_32	23_35	42_18	44_13	42_18	44_13	25_32	23_35	25_32	23_35	R^1
					■		■				■				■	0,385
					■	■					■			■		0,384
					■		■		■				■			0,383
					■	■			■			■				0,382
					■	■					■			■		0,374
				■		■				■				■		0,372
					■	■			■			■				0,372
				■		■		■				■				0,371
				■			■				■				■	0,371
				■			■				■		■			0,369
					■		■		■						■	0,368
					■		■		■				■			0,368
					■	■					■			■		0,363
					■	■			■			■				0,363
				■		■				■				■		0,361
							■				■				■	0,359
				■		■		■				■				0,359
							■		■				■			0,357
				■			■			■					■	0,355
				■			■	■					■			0,352
				■		■				■				■		0,349
				■		■		■				■				0,348
				■			■			■					■	0,342
				■			■	■					■			0,341
	■	■										■				0,229
	■		■											■		0,228
	■		■										■			0,227
	■	■													■	0,226
■		■										■				0,226
■		■												■		0,225
■			■										■			0,216
■			■												■	0,212

A-5-8 Modèles de régression sur l'espérance

Les tableaux Tableau A-17, Tableau A-18 et Tableau A-19 présentent les résultats de régression sur les configurations d'ajustement présentées au § 5.2. Ces modèles permettent de calculer l'espérance conditionnelle ajustable des consommations d'énergie et peuvent être comparées aux modèles de quantile conditionnel ($LGC_{5\%}$) et aux consommations réelles.

Tableau A-17 : Modèle de régression sur l'espérance correspondant à la configuration (5.8)

	Coefficients	***Références***
$LGC_{50\%-ref}$	2870 (+5)	
$T_{m.ext.sc}$	-344	10,36
$T_{m.ext.sf}$	360	23,10
$T_{mp.int.sc}$	860	22,18
$T_{mp.int.sf}$	-443	24,51

Tableau A-18 : Modèle de régression sur l'espérance correspondant à la configuration (5.9)

	Coefficients	***Références***
$LGC_{50\%-ref}$	2836 (+8)	
$T_{m.ext.sc}$	-502	10,36
$T_{m.ext.sf}$	510	23,10
$T_{mp.int.sc}$	749	22,18
$T_{mp.int.sf}$	-432	24,51
$E_{elec.sc}$	-1207	4320
$E_{elec.sf}$	1652	2345

Tableau A-19 : Modèle de régression sur l'espérance correspondant à la configuration (5.11)

	Coefficients	***Références***
$LGC_{50\%-ref}$	15160	
$T_{m.ext.sc}$	-364	10,36
$T_{m.ext.sf}$	446	23,10
$T_{mp.int.sc}$	1085	22,18
$T_{mp.int.sf}$	-619	24,51
$E_{elecSpé}$	9269	8507

RÉSUMÉ

Les écarts communément observés entre les prévisions de consommations énergétiques et les performances réelles des bâtiments limitent le développement des projets de construction et de réhabilitation. La garantie de performance énergétique (GPE) a pour vocation d'assurer des niveaux de consommations maximaux et donc de sécuriser les investissements. Sa mise en place fait cependant face à plusieurs problématiques, notamment techniques et méthodologiques.
Ces travaux de thèse se sont intéressés au développement d'une méthodologie pour la GPE associant les outils de simulation énergétique dynamique (SED) à un protocole de mesure et vérification. Elle repose d'abord sur la modélisation physico-probabiliste du bâtiment. Les incertitudes sur les paramètres physiques et techniques, et les variabilités des sollicitations dynamiques sont modélisées et propagées dans la SED. Un modèle de génération de données météorologiques variables a été développé. L'étude statistique des résultats de simulation permet d'identifier des modèles liant les consommations d'intérêt à des facteurs d'ajustement, caractéristiques des conditions d'exploitation. Les méthodes de régression quantile permettent de déterminer le quantile conditionnel des distributions et caractérisent donc conjointement la dépendance aux facteurs d'ajustement et le niveau de risque de l'engagement. La robustesse statistique de ces méthodes et le choix des meilleurs facteurs d'ajustement ont été étudiés, tout comme l'influence des incertitudes sur la mesure des grandeurs d'ajustement en exploitation. Leur impact est intégré numériquement en amont de la méthodologie. Cette dernière est finalement mise en œuvre sur deux cas d'étude : la rénovation de logements, et la construction de bureaux.

MOTS CLÉS

garantie de performance énergétique, simulation énergétique dynamique, mesure et vérification, ajustement, régression quantile

ABSTRACT

Discrepancies between ex-ante energy performance assessment and actual consumption of buildings hinder the development of construction and renovation projects. Energy performance contracting (EPC) ensures a maximal level of energy consumption and secures investment. Implementation of EPC is limited by technical and methodological problems.
This thesis focused on the development of an EPC methodology that allies building energy simulation (BES) and anticipation of the measurement and verification (M&V) process. The building parameters' uncertainties and dynamic loads variability are considered using a Monte-Carlo analysis. A model generating synthetic weather data was developed. Statistical studies of simulation results allow a guaranteed consumption limit to be evaluated according to a given risk. Quantile regression methods jointly capture the risk level and the relationship between the guaranteed energy consumption and external adjustment factors. The statistical robustness of these methods was studied as well as the choice of the best adjustment factors to consider. The latter will be measured during building operation. The impact of measurement uncertainties was statistically integrated in the methodology. The influence of M&V process accuracy was also examined.
The complete EPC methodology was finally applied on two different projects: the refurbishment of a residential building and the construction of a high energy performance office building.

KEYWORDS

energy performance contracting, building energy simulation, measurement and verification, parameter adjustment, quantile regression