École Doctorale de Polytechnique

Thèse de doctorat

Discipline : Mathématiques Appliquées

présentée par

Thibaud Mercier

Assimilation de Données et Mesures Primaires REP

---

dirigée par Houssem Haddar

Soutenue le 8 Septembre 2015 devant le jury composé de :

Pr. Olivier Talagrand	CNRS et Ecole Normale Supérieure (LMD)	président
Pr. Rolland Potthast	University of Reading	rapporteur
Pr. Clémentine Prieur	CNRS, Inria et Université Joseph Fourier	rapporteur
Pr. Houssem Haddar	Inria et Ecole Polytechnique (CMAP)	directeur
M. Jean-Philippe Argaud	EDF R&D (SINETICS)	examinateur
Pr. Philippe Courtier	Ministère de l'énergie	examinateur
M. Jean-Mélaine Favennec	EDF R&D (STEP)	examinateur
M. Alexandre Girard	EDF R&D (STEP)	examinateur

STEP
EDF R&D
6 quai Watier
78401 CHATOU CEDEX

CMAP
Ecole Polytechnique
Route de Saclay
91128 PALAISEAU

A mon filleul Basile,
né le même jour que cette thèse...

Remerciements

Je tiens à remercier en premier lieu mon directeur de thèse, le professeur Houssem HADDAR, ainsi que mes tuteurs d'EDF R&D Jean-Mélaine FAVENNEC et Alexandre GIRARD, qui m'ont conseillé, aidé et soutenu tout au long de ces trois années.

Toute ma gratitude également à celles et ceux qui ont bataillé dans les méandres labyrinthiques de la maison-qui-rend-fou administrative pour que cette thèse puisse être formellement contractualisée : Alexandre, Jean-Mélaine, Aurore, Khadra, Maryté, Marta et Stéphanie, côté EDF, et Houssem et Jessica, entre autres, côté Inria.

Il me faut aussi rendre hommage à la dynastie des stagiaires JMF, dont j'ai d'ailleurs fait partie un temps : Guillaume, Thierry, Thibaut, Giorgio, Marie-Jocelyn et Abdu, pour ne citer qu'eux. Leurs rapports de stages et les données qu'ils ont contribué à extraire et à traité m'ont été fort utiles pour mon propre travail de thèse.

Merci à Jennifer, Sylvain, David, Aurélien, Yvan, Julien, Yulu, Grégory, Olivier, Véronique, Daniel, Baligh, Pascal, Amine, Françoise, Bruno, Hoël, Audrey, Annie, Cyrille, Sarah, Michel, Corinne, Thierry, Philippe, et tous mes collègues du département STEP ou d'ailleurs, pour leur accueil chaleureux ; merci également à Gwenaël et tous les thésards du CMAP d'avoir accepté d'héberger une fois par semaine le thésard fantôme que j'étais de leur point de vue.

Un grand merci aux professeurs Vivien MALLET et Marc BOCQUET, pour m'avoir ouvert les portes de leur cours d'Assimilation de Données du Master M2 OACOS et de l'Ensta, à Jussieu ; au professeur Rolland POTTHAST pour sa Summer School sur l'Assimilation de Données à Reading, dont je garde un excellent souvenir ; au professeur Gilbert SAPORTA pour ses conseils avisés et son excellent ouvrage.

Je remercie également le Professeur Roland GLOWINSKI, au Professeur Dmitri KUZMIN et à Luc VAN DEN DURPEL, qui ont été les premiers à me suggérer de me lancer dans l'aventure d'une thèse, et à Jérôme ADNOT, qui m'a mis en contact avec EDF R&D.

Merci enfin à ma famille et à mes chers parents, pour leur tendre soutien. Mes félicitations à mon nouveau camarade Amaury ; mon affection à mon adorable filleul Basile, frère jumeau de cette thèse ; et un remerciement tout particulier à mon oncle Michel BERA, ainsi qu'à Nanabozo les Grandes Oreilles, qui se reconnaîtra.

Résumé

Résumé

Le circuit primaire d'un réacteur à eau pressurisée (REP) est un système thermohydraulique complexe, présentant des champs hétérogènes de puissances, températures et débits dans des conditions extrêmes : l'instrumentation in situ est limitée en nombre, localisation et précision. De ce fait, la connaissance de ces paramètres de fonctionnement est impactée par des incertitudes de représentativité notamment, prises en compte dans la conception et intégrées dans les protections d'exploitation.

Dans ce contexte, EDF R&D cherche à évaluer les apports potentiels de l'Assimilation de Données, un ensemble d'outils mathématiques très utilisé en géosciences qui permettent de corriger des estimations issus de modèles à l'aide de mesures provenant du système réel modélisé, pour une meilleure caractérisation (justesse, incertitudes) des paramètres du point de fonctionnement primaire nominal.

Dans cette thèse, nous définissons un modèle semi-empirique 0D adapté au niveau de description usuellement choisi par les exploitants, et développons une méthodologie (de type Monte-Carlo et inspirée des Méthodes d'Ensemble) pour utiliser ce modèle dans un cadre d'Assimilation de Données. L'application de cette méthodologie à des données simulées permet d'évaluer la réduction des incertitudes pesant sur les paramètres-clés : les résultats sont au-delà des espérances initiales mais nécessitent des hypothèses fortes impliquant un prétraitement soigneux des données d'entrée.

Mots-clefs

Assimilation de Données ; REP ; Circuit Primaire ; BLUE ; Filtre de Kalman ; Méthode Monte-Carlo ; Modélisation 0D/1D ; Expériences Jumelles ; Systèmes de Mesures

Data Assimilation and PWR Primary Measurement

Abstract

A Pressurized Water Reactor (PWR) Reactor Coolant System (RCS) is a highly complex physical process: heterogeneous power, flow and temperature distributions are difficult to be accurately measured, since instrumentations are limited in number, thus leading to the relevant safety and protection margins.

EDF R&D is seeking to assess the potential benefits of applying Data Assimilation to a PWR's RCS (Reactor Coolant System) measurements, in order to improve the estimators for parameters of a reactor's operating setpoint, i.e. improving accuracy and reducing uncertainties and biases of measured RCS parameters.

In this thesis, we define a 0D semi-empirical model for RCS, satisfying the description level usually chosen by plant operators, and construct a Monte-Carlo Method (inspired from Ensemble Methods) in order to use this model with Data Assimilation tools. We apply this method on simulated data in order to asses the reduction of uncertainties on key parameters : results are beyond expectations, however strong hypothesis are required, implying a careful preprocessing of input data.

Keywords

Data Assimilation ; PWR ; Reactor Coolant System ; BLUE ; Kalman Filter ; Monte-Carlo Method ; 0D/1D Model ; Twin Experiments ; Measurement Systems

Table des matières

1 Circuit Primaire d'un REP **15**

1.1 Principes . . . 15
1.1.1 Trois circuits successifs . . . 15
1.1.2 Enceinte de Confinement . . . 18
1.2 Structure . . . 18
1.2.1 Cuve . . . 19
1.2.2 Générateurs de Vapeur . . . 19
1.2.3 Circuits auxiliaires et circuits de sauvegarde . . . 21
1.3 Cœur et Combustible . . . 23
1.4 Cycle . . . 26
1.4.1 Généralités . . . 26
1.4.2 Arrêt du réacteur . . . 27
1.4.3 Vidange et ouverture du circuit primaire . . . 27
1.4.4 Déchargement du cœur . . . 28
1.4.5 Rechargement du cœur . . . 28
1.4.6 Fermeture et remplissage du circuit primaire . . . 29
1.4.7 Chauffage du fluide primaire . . . 29
1.4.8 De l'Arrêt à Chaud au fonctionnement en puissance . . . 30
1.4.9 Fonctionnement en puissance - suivi de charge . . . 30
1.4.10 Fonctionnement en prolongation de cycle . . . 31
1.5 Enjeux . . . 31
1.5.1 Connaissance du Point de Fonctionnement Primaire . . . 31
1.5.2 Complexité de la modélisation physique . . . 32
1.5.3 Limitation de l'instrumentation in situ . . . 33

2 Données d'entrée **35**

2.1 Notations . . . 35
2.2 Mesures . . . 36
2.2.1 Bases de Données . . . 36
2.2.2 Débit Primaire . . . 38
2.2.3 Températures Primaires . . . 43
2.2.4 Puissance de la chaudière . . . 48
2.3 Estimateurs . . . 54
2.3.1 TBC Enthalpique . . . 55
2.3.2 Cartographie des $THAC$ . . . 56
2.3.3 Moyennes par quarts de cœur . . . 59
2.4 Modèles . . . 61
2.4.1 Codes industriels . . . 61
2.4.2 Modèle Semi-Empirique . . . 65

2.5 Bilan . . . 75

3 Assimilation de Données **77**
3.1 Principes généraux . . . 77
3.1.1 Histoire . . . 77
3.1.2 Exemple scalaire . . . 79
3.1.3 Formalisation du problème . . . 80
3.1.4 Analyse de Cressman . . . 82
3.2 Algorithmes . . . 85
3.2.1 Le BLUE . . . 87
3.2.2 Le 3DVAR . . . 89
3.2.3 Le Filtre de Kalman et le 4DVAR . . . 93
3.2.4 Le Filtre de Kalman d'Ensemble . . . 96
3.3 Construction des matrices de covariance . . . 97
3.3.1 Enjeux . . . 97
3.3.2 Méthodes Empiriques . . . 98
3.3.3 Principe de la Méthode de Monte-Carlo . . . 99
3.3.4 Cas Linéaire . . . 100
3.3.5 Cas non linéaire . . . 106
3.3.6 Calibrage du Randomisateur . . . 107

4 Application et Résultats **109**
4.1 Démarche . . . 109
4.1.1 Construction des opérateurs . . . 109
4.1.2 Construction des paramètres du BLUE . . . 113
4.1.3 Expérience Jumelle . . . 116
4.2 Effets de l'Ebauche . . . 122
4.2.1 Calibrage des PIPAN . . . 122
4.2.2 Analyse sur les paramètres . . . 128
4.2.3 Bilan . . . 131
4.3 Effets des Observations . . . 135
4.3.1 Résistance aux biais . . . 135
4.3.2 Perte d'instrumentations . . . 137
4.4 Cadre dynamique . . . 138
4.4.1 Modèle d'évolution . . . 139
4.4.2 Cas normal . . . 144
4.4.3 Cas d'observations biaisées . . . 149
4.4.4 Bilan . . . 152
4.5 Synthèse . . . 153
4.6 L'AD : un outil polyvalent . . . 155
4.7 Cadre semi-empirique pour le Circuit Primaire . . . 156
4.8 Résultats des Expériences Jumelles . . . 156
4.9 Perspectives . . . 158

A Structure des Programmes **160**
A.1 Modules importés . . . 160
A.1.1 NumPy . . . 160
A.1.2 xlrd et xlwt . . . 160
A.1.3 Matplotlib . . . 161
A.1.4 SciPy . . . 161

A.1.5 RANDOM . . . 161
A.2 Classes Principales . . . 161
A.2.1 Modèle . . . 161
A.2.2 Mesure . . . 162
A.2.3 Observations . . . 162
A.2.4 Echantillon . . . 162
A.3 Autres modules . . . 165
A.3.1 auxiliaires . . . 165
A.3.2 importationRES . . . 165
A.3.3 globales . . . 165
A.3.4 thermodynamique . . . 165
A.3.5 BLUE . . . 165

B Publications **166**

Introduction

Une grande majorité des réacteurs en fonctionnement dans le monde est constituée de réacteurs de type REP (Réacteur à Eau Pressurisée). Ce genre de centrale nucléaire n'est structurellement pas très différent des centrales thermiques classiques : sa seule différence fondamentale est que l'énergie thermique provient de la fission nucléaire, et non de la combustion chimique classique. Le reste est semblable à une machine à vapeur traditionnelle : l'énergie thermique est véhiculée par l'eau et la vapeur, puis transformée en énergie mécanique via l'énergie de la vapeur d'eau pressurisée ; cette énergie mécanique est communiquée à une turbine, qui la transforme en énergie électrique grâce à un alternateur.

Dans un REP, pour limiter la diffusion de la radioactivité, un premier circuit d'eau sous pression (circuit dit « primaire ») récupère la chaleur produite par le combustible et la transmet à un deuxième circuit eau-vapeur (dit « secondaire ») via les générateurs de vapeur. C'est ce circuit secondaire qui alimente la turbine. Un troisième circuit, dit de refroidissement, évacue la chaleur résiduelle de ce circuit secondaire vers la source froide (atmosphère via des tours aéroréfrigérantes, ou vaste volume d'eau comme une mer ou un grand fleuve). Dans un réacteur de type REP, l'eau du circuit primaire ne joue pas seulement le rôle de caloporteur : elle sert également de modérateur, capturant et ralentissant les neutrons issus de la fission, ce qui permet d'entretenir le régime critique de la réaction en chaîne.

Les REP constituent la filière de réacteurs nucléaires la plus répandue dans le monde [1] : EDF en exploite 58 en France, et s'intéresse donc tout particulièrement à l'amélioration constante de leur sûreté et de leurs performances. Or les paramètres liés au circuit primaire (températures, puissances et débits au sein du circuit primaire) jouent un rôle primordial sur le fonctionnement normal du réacteur, et doivent respecter de strictes marges de sûreté. L'établissement de ces marges est effectué à partir de calculs conservatifs, qui tiennent compte des incertitudes pesant sur la connaissance de ces paramètres primaires : à sûreté égale, plus l'incertitude est élevée, plus la puissance produite est réduite en conséquence.

Le retour d'expérience des exploitants et des ingénieries du parc EDF en exploitation a montré que le point de fonctionnement du circuit primaire n'était pas connu aussi précisément que souhaité. Ainsi, depuis la mise en place des chargements à faible fluence (vers 2000), on constate une hétérogénéité accrue de température en branche chaude (BC), ce qui affecte la représentativité de la mesure de Température Branche Chaude (T_{BC}). Ceci impacte directement l'incertitude liée au débit-boucle primaire (noté Q), qui n'est par ailleurs pas accessible en continu, mais calculé une fois pour toutes en début de cycle, lors de l'essai périodique (EP) RCP114 par une méthode de bilan enthalpique (qui est le stan-

1. En 2011, environ les deux tiers des réacteurs nucléaires de puissance en fonctionnement dans le monde étaient des REP.

dard international pour les REP). Partant de ces demandes des exploitants, EDF R&D a lancé en 2011 le projet PAPRICA (« ex**P**loiter les **A**méliorations pour la maîtrise du point de fonctionnement **PRI**maire par le **CA**lcul ») [Bri14].

Ce projet s'articule en deux volets. Le premier part du constat qu'on n'envisage pas l'ajout de nouvelles instrumentations sur le circuit primaire : il propose donc des méthodes nouvelles pour mieux exploiter les instrumentations existantes. Le second volet s'intéresse à la caractérisation en termes de faisabilité et de précision de nouvelles instrumentations non intrusives, en vue de les installer occasionnellement pour des expérimentations.

C'est dans ce projet et dans sa continuité que s'inscrit cette thèse. Dans le cadre de PAPRICA, plusieurs nouveaux estimateurs plus précis du point de fonctionnement primaire ont été conçus, exploitant les données de process [2], les données d'essais [3], les résultats de cartes de flux [4], ainsi que d'autres relevés de valeurs sur site, à différents régimes de fonctionnements (à 100% de puissance nominale (PN), aux Arrêts A Chaud (AAC), aux paliers de montée en température ...), et élaborant de nouvelles méthodes et de nouveaux modèles. En vue de développer une synergie entre ces différents nouveaux outils, le Département STEP (Simulation et Traitement de l'Information pour l'Exploitation des systèmes de Production) a recouru dans un premier temps à la Réconciliation de Données pour le circuit secondaire, et à une première application scalaire d'Assimilation de Données (*"Best Estimate"* pour TBC). Par ailleurs, des techniques sophistiquées d'Assimilation de Données ou de statistiques s'y apparentant ont été appliquées avec succès au cours de ces dernières années par différents départements d'EDF R&D, sur des problèmes de neutronique [Pon09] ou de thermohydraulique [Gir12], par exemple. L'Assimilation de Données a donc été envisagée comme une réponse possible aux problématiques visées par PAPRICA, et un prolongement logique des approches effectuées jusqu'alors.

L'Assimilation de Données (AD), telle qu'utilisée en météorologie par exemple, consiste à recaler des modèles sur des observations pour obtenir l'estimation la plus précise possible de l'état du système étudié, étant donné l'ensemble des informations disponibles. Ainsi, au lieu de considérer séparément chacun des paramètres primaires, notre ambition est de considérer le vecteur les réunissant tous, et de procéder à son estimation précise en assimilant simultanément l'ensemble des données relatives au circuit primaire.

L'objectif de la thèse est ainsi de rendre plus performants les nouveaux estimateurs déjà créés dans le cadre de PAPRICA, en réduisant leurs incertitudes par l'Assimilation de Données, et en aboutissant ainsi à une caractérisation du point de fonctionnement primaire qui soit plus complète, plus fréquente et plus précise que celle dont disposait jusqu'alors l'exploitant.

Les apports potentiels à terme pour les exploitants sont (à sûreté égale) :

— une puissance productible optimisée ;

— une capacité accrue de détection d'erreurs ou de dysfonctionnements.

2. Base de Données ORLI
3. Base de Données MOLENE
4. Fichiers RES de la base de données SINFONIE/Musicale

L'exposé est divisé en quatre grandes parties :

— la première partie décrit le Circuit Primaire d'un REP ;

— la deuxième partie détaille les données d'entrée disponibles (mesures, estimateurs et modèles) ;

— la troisième partie porte sur la théorie de l'Assimilation de Données et les considérations guidant le choix des options d'AD au problème posé dans PAPRICA ;

— la quatrième partie réunit les différentes applications mises en œuvre et les résultats correspondants.

Nomenclature

AAC	Arrêt A Chaud
AAD	pompe de démarrage et d'arrêt
AC	Assemblage Combustible
AD	Assimilation de Données
APA	ensemble motopompes alimentaires
ARE	alimentation normale des GV
ASG	circuit d'eau Alimentaire de Secours des Générateurs de vapeur
BC	Branche Chaude
BF	Branche Froide
BLUE	Best Linear Unbiased Estimator
BR	Bâtiment Réacteur
CEF	Contrôle Economique en Fonctionnement (type d'essai périodique)
CFM	Chambres à Fission Mobiles
CFD	Computational Fluid Dynamics
CNPE	Centre National de Production d'Electricité
DA	Data Assimilation
EAS	Enceinte : circuit d'Aspersion Sécurité
EDF R&D	Electricité De France, Recherche et Développement
EP	Essai Périodique
EPR	European Pressurized Reactor / Evolutionary Power Reactor
GMPP	Groupe MotoPompes Primaires
GV	Générateur de Vapeur
IO	Interpolation Optimum
MOX	Mélange d'OXydes (type de combustible utilisé par EDF)
MW	Mégawatt (généralement thermique)
MWe	Mégawatt Electrique
OI	Optimum Interpolation (=IO)
PAPRICA	exPloiter les Améliorations de la connaissance du point de fonctionnement PRImaire par le CAlcul
$PGMPP$	Puissance thermique générée par l'ensemble du Groupe des MotoPompes Primaires
$PIPA$	Puissance Intégrée Par Assemblage (en MW)
$PIPAN$	Puissance Intégrée Par Assemblage Normalisée (notation interne)
PL	Puissance Locale (en MW) (notation interne)
PLN	Puissance Locale Normalisée (notation interne)
PN	Puissance Nominale
$PTHGV$	Puissance Thermique d'un Générateur de Vapeur (notation interne)
$PTHREAC$	Puissance Thermique totale du Réacteur (notation interne)
PWR	Pressurized Water Reactor

Q	Débit-boucle primaire massique (notation interne)
RCP	Réacteur : Circuit Primaire
RCS	Reactor Coolant System
REP	Réacteur à Eau Pressurisée
RES	Document texte contenant les résultats du calcul de carte de flux
RCV	Réacteur : circuit de contrôle Chimique et Volumétrique
RIC	Réacteur : InCore
RIS	Réacteur : circuits d'Injection de Sécurité
RPN	Réacteur : Puissance Nucléaire (chambres de mesure de la puissance neutronique)
RRA	circuit de Refroidissement du Réacteur à l'Arrêt
STEP	Simulation et Traitement de l'Information pour l'Exploitation des systèmes de Production
TBC	Température de Branche Chaude
TBF	Température de Branche Froide
TC	Thermocouple
TEP	Traitement des Effluents Primaires
$THAC$	Température en Haut de l'Assemblage Combustible

Chapitre 1

Circuit Primaire d'un REP

L'objectif de ce chapitre est de proposer une description technique sommaire de notre objet d'étude : l'état thermohydraulique du circuit primaire d'un REP.

1.1 Principes

Comme nous l'avons dit en introduction, le Réacteur à Eau Pressurisée (REP) est le réacteur électrogène le plus fréquent parmi les réacteurs nucléaires : ce réacteur utilise un combustible à oxyde d'uranium légèrement enrichi, ou un combustible mixte composé d'un mélange d'oxyde d'uranium appauvri ou faiblement enrichi et d'oxyde de plutonium (combustible MOX).

La Figure 1.1 décrit schématiquement ce type de réacteur, tandis que la Figure 1.2 donne une idée de l'agencement général des différents composants dans l'espace.

1.1.1 Trois circuits successifs

On se réfère à [NR08].

On remarquera que la centrale nucléaire est conçue autour de trois circuits successifs :

— Le **circuit primaire** est un circuit d'eau restant normalement liquide grâce à une pression de quelques 150 bars assurée par le pressuriseur placé sur l'une des branches de ce circuit. Refoulée depuis les Branches Froides (BF) par les Pompes Primaires (PP), l'eau entre dans le cœur à environ $290°C$ et s'y échauffe d'une quarantaine de degrés en montant dans le cœur le long des rayons des assemblages. A la sortie du cœur l'eau est répartie entre trois et quatre boucles selon qu'il s'agit d'un réacteur de 900 MWe, 1300 MWe ou 1450 MWe[1], placées respectivement à 120 ou 90 degrés, chaque Branche Chaude (BC) allant vers un Générateur de Vapeur (GV) (échangeur thermique).

1. Au cours du travail de thèse, nous avons appliqué notre méthodologie exclusivement à des réacteurs 1300 MWe, pour lesquels le circuit primaire contient quatre boucles. MWe signifie « Mégawatts électriques », par opposition aux Mégawatts thermiques habituellement notés $MWth$ que nous noterons simplement MW.

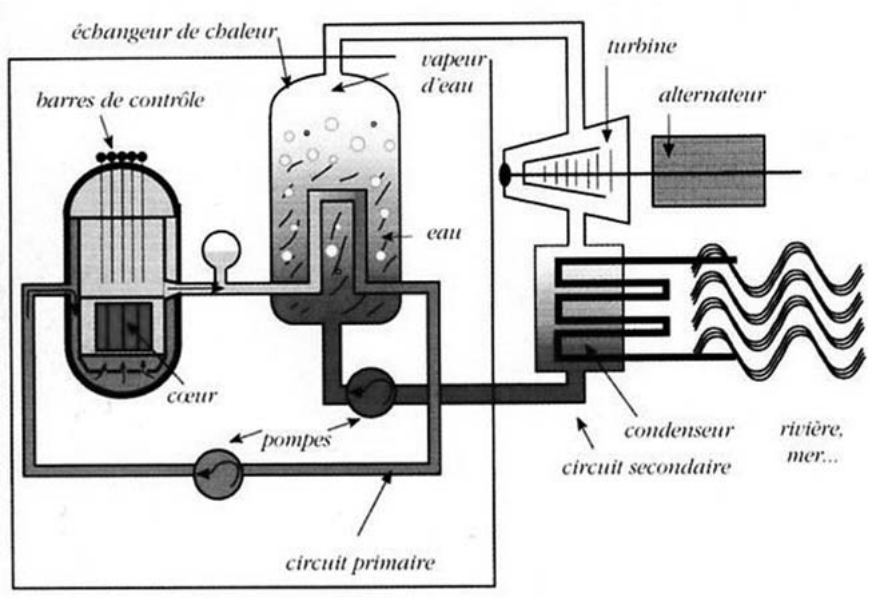


FIGURE 1.1 – Schéma de principe d'un REP (source : Collection Génie Atomique)

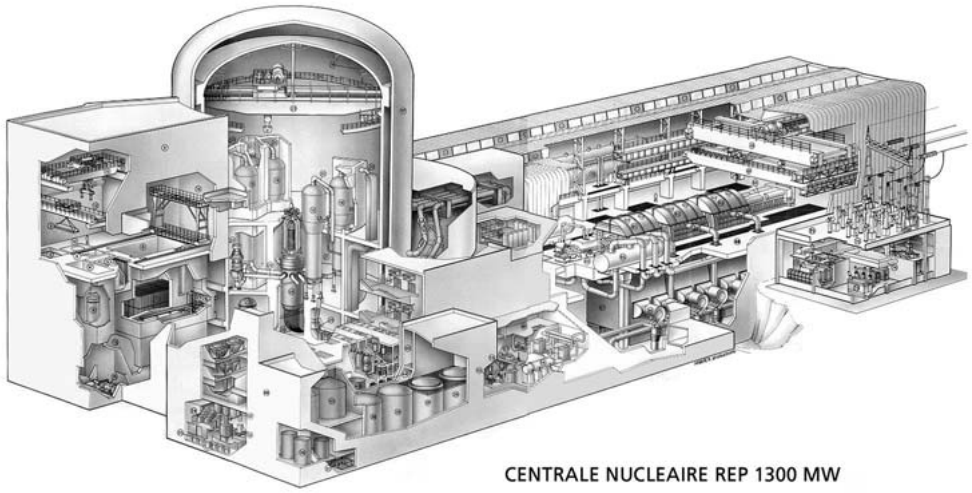


FIGURE 1.2 – Schéma général d'un REP de Framatome (AREVA NP) (source : Collection Génie Atomique)

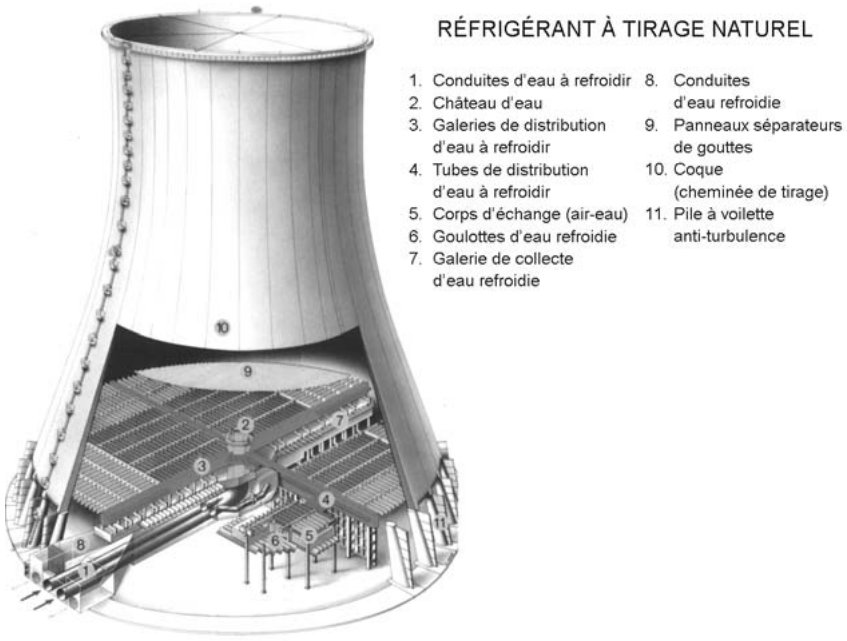


FIGURE 1.3 – Tour aéroréfrigérante d'un REP de Framatome (source : Collection Génie Atomique)

— Le **circuit secondaire** alimente en eau liquide chacun des générateurs de vapeur où la chaleur s'échange entre l'eau du circuit primaire (passant par environ quatre mille tubes en forme de U renversé) et celle du circuit secondaire passant entre ces tubes. Cette dernière se vaporise à la traversée du générateur de vapeur. A la sortie des turbines basse pression, la vapeur est envoyée dans le condenseur où elle repasse en phase liquide avant d'être reprise par des pompes alimentaires et renvoyée dans les générateurs de vapeur après réchauffage par échange avec de la vapeur soutirée des turbines Haute Pression et Basse Pression.

— Contrairement aux deux précédents, le **circuit tertiaire**, qui refroidit le condenseur, est un circuit ouvert sur l'environnement. Suivant les sites, il s'agit de l'eau d'un fleuve, de la mer ou d'un circuit partiellement fermé, lui-même refroidi par des aéroréfrigérants (grandes tours creuses en forme d'hyperboloïdes jouant le rôle de cheminées (voir Figure 1.3).

Environ un tiers de l'énergie produite par les fissions est converti en électricité et les deux tiers sont évacués au niveau de ce troisième circuit. L'échauffement de l'eau du circuit tertiaire n'excède pas quelques dizaines de degrés, et cette chaleur de faible qualité thermique a peu d'usages possibles, hormis des serres de maraîchers ou des piscines d'élevage.

Globalement, le cycle thermodynamique d'un REP est le cycle de Rankine. [2]

2. Les REP construits en Occident sont issus des filières américaines PWR développées par Westinghouse et Combustion Engineering. La filière russe VVER constitue un autre type de REP.

1.1.2 Enceinte de Confinement

La partie classique (turbines, alternateur et condenseur), qui ne présente pas derisque radiologique, est placée dans la salle des machines ; la partie proprement nucléaire (cœur, générateurs de vapeur et organes associés tels le pressuriseur et les pompes primaires) se trouve dans le bâtiment du réacteur (on dit souvent aussi BR pour « Bâtiment Réacteur »). Celui-ci est formé par une enceinte béton doublée d'une enveloppe métallique dite « liner » (ou une double enceinte béton pour les centrales les plus récentes) de forme cylindrique, étanche et capable de résister à la surpression (environ 5 *bar*) qu'occasionnerait la vaporisation de l'eau du circuit primaire suite à un un accident de fusion du cœur.

Ainsi l'accident de Three Mile Island, aux États-Unis, survenu en 1976, au cours duquel pratiquement aucune radioactivité n'a été relâchée dans l'environnement malgré la fusion d'une proportion importante du cœur, a montré l'efficacité de cette troisième et ultime barrière après ruptures de la première barrière représentée par les gaines des crayons de combustible et de la seconde barrière représentée par le circuit primaire. A contrario, l'accident de Tchernobyl a eu des conséquences beaucoup plus graves à cause de l'absence d'une telle enceinte de confinement ; et celui de Fukushima également, du fait du grave endommagement de ladite enceinte.

Accolé au bâtiment du réacteur se trouve un bâtiment nucléaire auxiliaire destiné à l'entreposage des assemblages de combustible neuf (avant chargement en cœur) et irradié (« refroidissement » pendant au moins quelques mois avant envoi au retraitement).

1.2 Structure

Le circuit primaire est composé des éléments suivants [DEL08] :

— la **cuve** qui contient le cœur ;

— les **générateurs de vapeur**, échangeurs de chaleur convertissant l'énergie apportée par le cœur (du circuit primaire) en vapeur disponible pour la turbine (du circuit secondaire) ;

— les **pompes primaires**, qui fournissent le travail nécessaire pour assurer un bon débit de refroidissement du cœur ;

— le **pressuriseur** qui sert au maintien et au contrôle de la pression ;

— les **tuyauteries** de circulation de l'eau (branches chaudes et branches froides) ;

— l'eau confinée dans l'ensemble de ces composants.

Si le nombre d'éléments peut changer d'un type de REP à l'autre[3], au cours de notre travail de thèse, qui s'est focalisé sur des paliers EDF de 1300 MWe, nous ne rencontrerons que des réacteurs à 4 boucles.

1.2.1 Cuve

Les cuves des différents concepts de REP sont sensiblement identiques. Les différents volumes de la cuve sont les suivants (voir Figure 1.4)

— l'*espace annulaire* depuis la bride de la cuve des branches froides, jusqu'au fond de cuve, bordé par les parois internes de la cuve et l'enveloppe du cœur : l'eau y pénètre par les ajutages des branches froides et descend jusqu'au plénum inférieur ;

— le *plénum inférieur* constitué par la partie basse de la cuve entre le fond de cuve hémisphérique et la plaque de supportage du cœur ;

— le *plénum supérieur* constitué par l'espace entre la plaque supérieure du cœur et la bride : les ajutages des branches chaudes débouchent dans ce plénum ;

— le *dôme* est la partie située entre le couvercle de la cuve et la plaque de supportage des tubes guides ;

— les *internes supérieurs* sont constitués essentiellement des guides de grappes dans lesquels coulissent les grappes de commande ;

— les *internes inférieurs* comprennent les éléments de supportage du cœur, et pour les réacteurs actuels, les tubes d'instrumentation qui traversent le fond de cuve[4].

1.2.2 Générateurs de Vapeur

Il y a trois types principaux de générateurs de vapeur sur les réacteurs de vapeur sur les réacteurs à eau sous pression, mais les plus utilisés dans les différentes conceptions de REP sont les *générateurs de vapeur à tube en U* (cf. Figure : 1.5).

Ils fonctionnent avec *recirculation* : l'eau froide du circuit secondaire pénètre dans le générateur de vapeur dans un espace annulaire ; elle entre dans la partie basse de l'échangeur mais elle n'est pas entièrement vaporisée à la sortie du faisceau tubulaire (environ 40% du débit d'eau est vaporisé). Un premier séchage de la vapeur est assuré par des *séparateurs de vapeur* (cyclones), pièces mécaniques fixes, les condensats étant dirigés vers l'espace annulaire où ils se mélangent avec l'eau d'alimentation (recirculation). La vapeur en sortie des séparateurs subit une légère détente dans un *sécheur* qui effectue une dernière séparation

3. La configuration de l'AP-1000 est différente de celle de l'EPR. Doté uniquement de deux boucles, le réacteur AP-1000 a une branche froide dédoublée et quatre pompes : cela permet l'utilisation de pompes de plus petite taille à rotor noyé, la réduction de la taille du circuit primaire, ainsi que la minimisation des conséquences de la perte d'une pompe ou de la rupture d'une branche froide primaire. Pour les réacteurs de type VVER, les principales différences, outre le cœur et le combustible, résident dans l'utilisation de générateurs de vapeurs horizontaux et dans les niveaux différents des branches chaudes et froides

4. cette conception est abandonnée pour les réacteurs de type EPR où l'instrumentation pénètre par le couvercle de la cuve

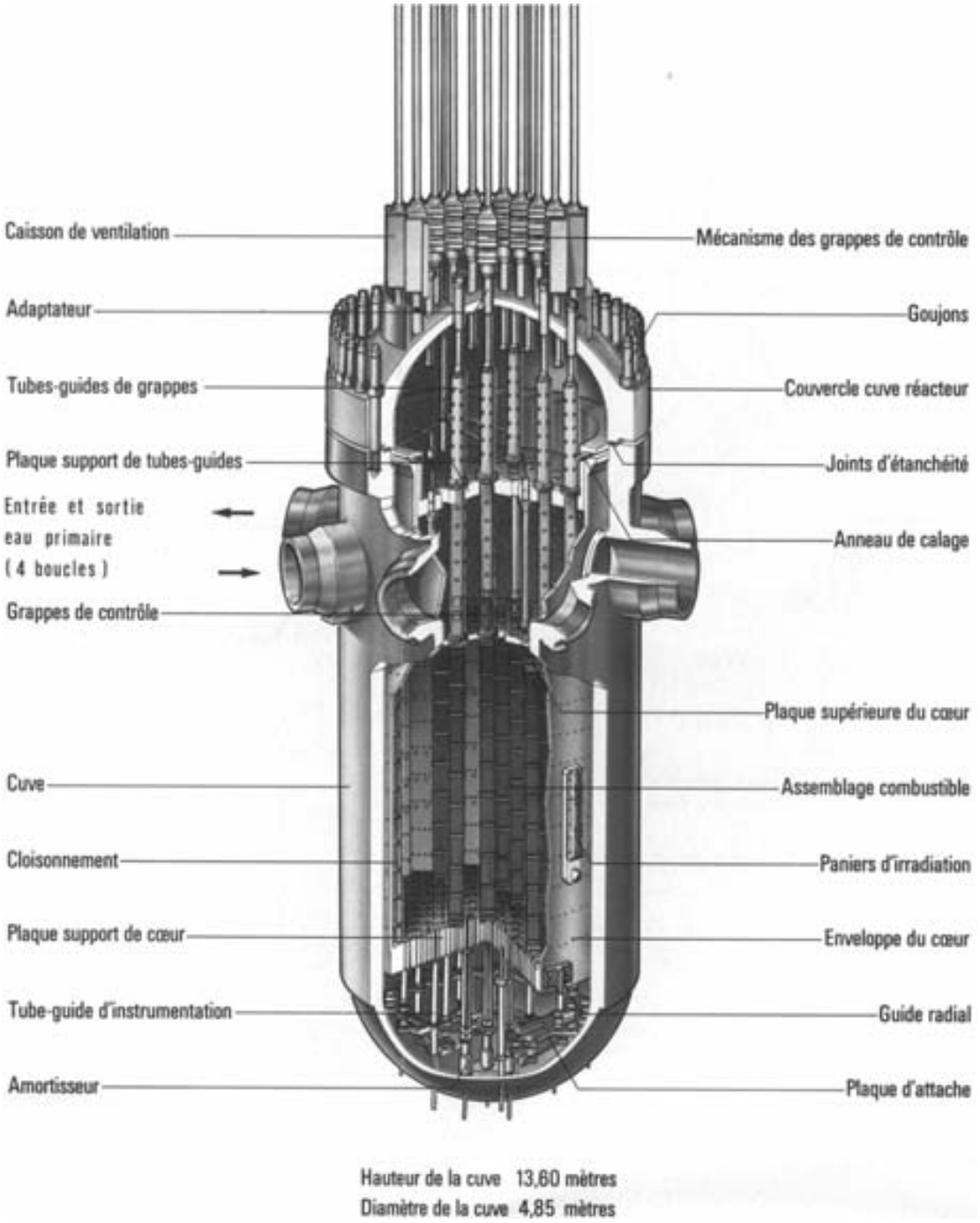


FIGURE 1.4 – Cuve d'un REP de Framatome (source : Collection Génie Atomique)

eau/vapeur avant de sortir en partie haute du générateur de vapeur vers la turbine.

L'avantage de ce principe réside dans le fait que l'échange thermique est réalisé par ébullition autour des tubes. Ce mode d'échange thermique est très efficace et permet donc d'améliorer la compacité de l'échangeur. Certains GV sont dotés d'*économiseurs*, les deux parties de faisceau de tubes sont séparées par une paroi centrale, et des chicanes sont disposées transversalement pour favoriser l'échange sur la partie descendante des tubes contenant le fluide primaire le plus froid. Par leur conception ces générateurs de vapeur ne permettent pas la surchauffe de la vapeur.

1.2.3 Circuits auxiliaires et circuits de sauvegarde

Par souci d'exhaustivité, nous décrivons à présent brièvement les circuits auxiliaires et les circuits de sauvegarde qui sont représentés Figure 1.6.

Circuits auxiliaires

Le circuit de refroidissement du réacteur à l'arrêt (RRA) a pour fonction, lors de la mise à l'arrêt normal du réacteur, d'évacuer la chaleur du circuit primaire et la puissance résiduelle du combustible, puis de maintenir l'eau primaire à basse température pendant toute la durée de l'arrêt. Le circuit RRA sert également à vidanger la piscine du réacteur après rechargement du combustible.

Le circuit de contrôle chimique et volumétrique du réacteur (RCV) permet, pendant le fonctionnement de la chaudière :

— d'ajuster la masse d'eau primaire en fonction des variations de température ;

— de maintenir la qualité de l'eau primaire, en réduisant sa teneur en produits de corrosion et de fission, et en injectant des produits chimiques (inhibiteurs de corrosion, par exemple) ;

— de reprendre et de compenser les fuites normales des joints de pompes primaires ;

— de régler la concentration en acide borique.

Circuits de sauvegarde

Le rôle des circuits de sauvegarde est de maîtriser et de limiter les conséquences des incidents et accidents en assurant les fonctions de sûreté suivantes :

— maintien d'un inventaire en eau suffisant dans la cuve pour refroidir le combustible ;

— évacuation de la puissance résiduelle ;

— maintien de conditions dans l'enceinte assurant l'étanchéité du confinement.

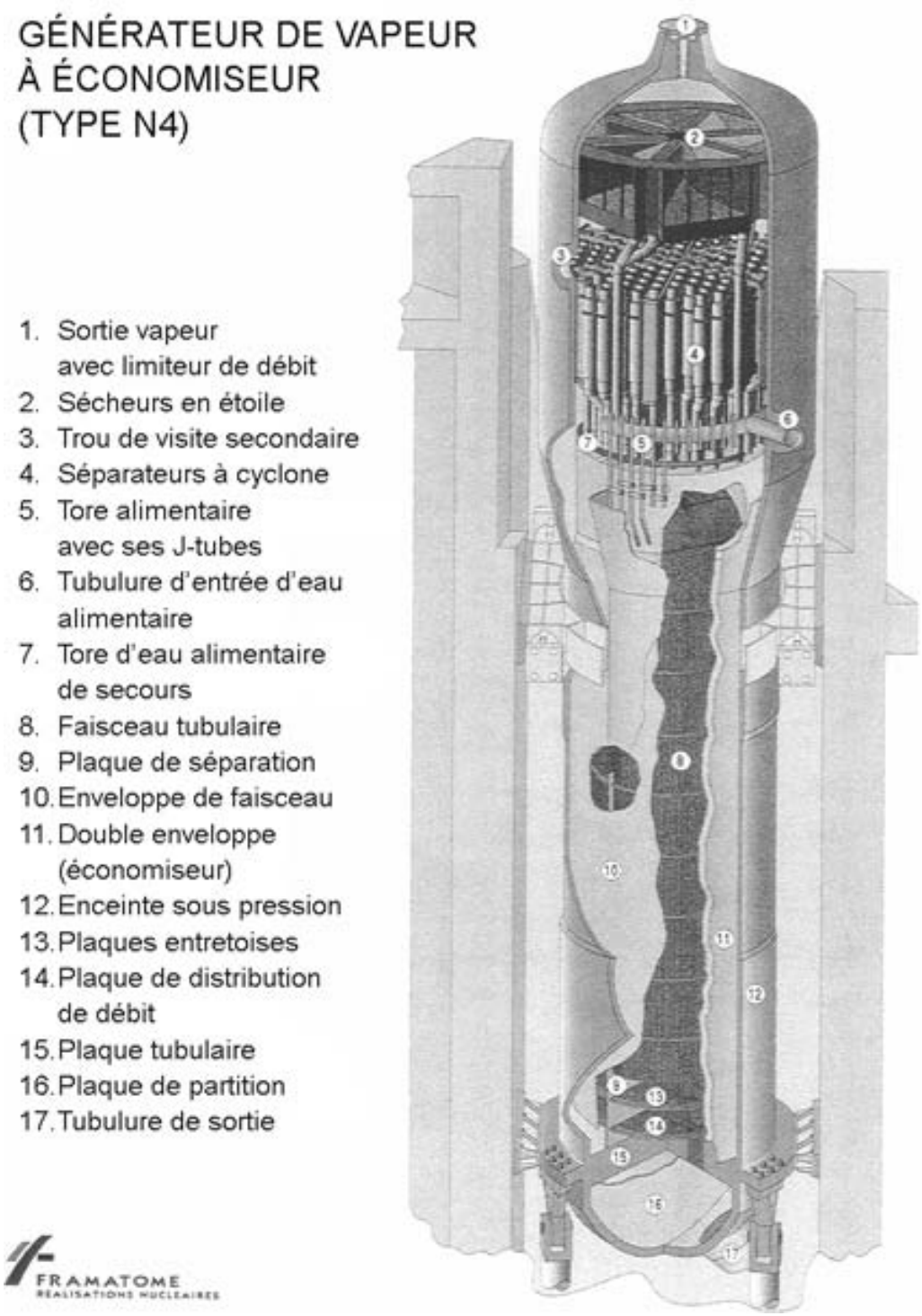


FIGURE 1.5 – Générateur de Vapeur d'un REP de Framatome (source : Collection Génie Atomique)

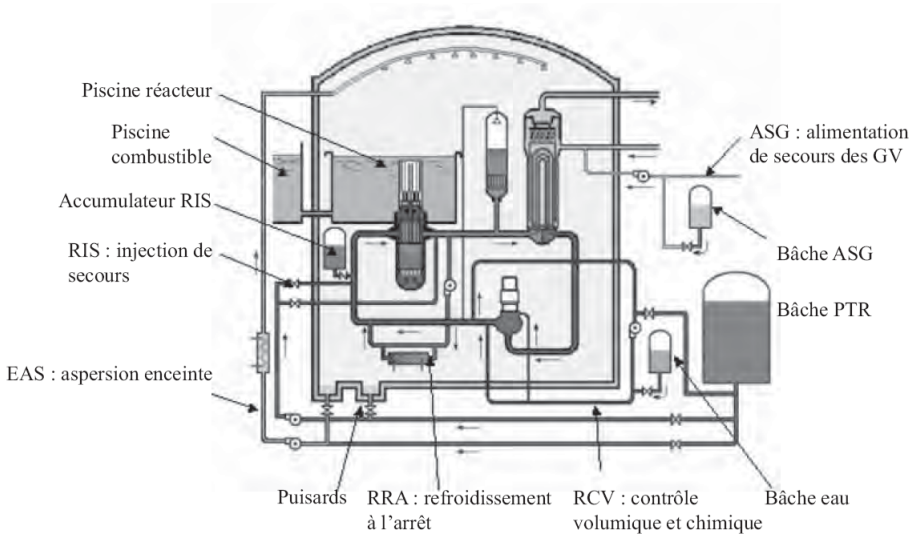


FIGURE 1.6 – Principaux circuits de sauvegarde du primaire d'un REP (source : Collection Génie Atomique)

Les circuits de sauvegarde sont essentiellement les circuits d'injection de sécurité (RIS), le circuit d'aspersion dans l'enceinte (EAS) du bâtiment réacteur, et le circuit d'eau alimentaire de secours des générateurs de vapeur (ASG).

Les circuits RIS injectent de l'eau borée dans le cœur du réacteur en cas d'accident afin d'évacuer la puissance résiduelle et d'empêcher tout retour à la criticité du cœur. Ils sont constitués d'accumulateurs sous pression, qui fonctionnent de manière passive, et de pompes, qui ont des débits et des pressions de refoulement différents pour répondre aux différents types d'accidents. Ces pompes aspirent l'eau d'un réservoir de 2000 m^3 environ. Lorsque ce réservoir est vide, elles sont connectées aux puisards du bâtiment du réacteur, où est recueillie l'eau pulvérisée par le système EAS, ainsi que l'eau qui s'échappe du circuit primaire dans les cas de rupture sur ce circuit.

Le circuit EAS, en cas de rupture du circuit primaire ou d'une tuyauterie de vapeur à l'intérieur de l'enceinte, pulvérise de l'eau additionnée de soude dans l'enceinte, afin de diminuer la pression qui y règne et de rabattre au sol les aérosols radioactifs éventuellement disséminés. Il contient un échangeur de chaleur qui permet l'évacuation de la puissance résiduelle hors de l'enceinte par le circuit de refroidissement intermédiaire (RRI).

Le circuit ASG permet de maintenir le niveau d'eau secondaire dans les GV et donc de refroidir l'eau du circuit primaire en cas de panne de leur circuit d'alimentation normale (ARE) et lors des phases d'arrêt et de démarrage du réacteur.

1.3 Cœur et Combustible

Situé au sein de la cuve, l'élément central du circuit primaire, le **cœur** est composé d'assemblages de barreaux combustibles 1.7.

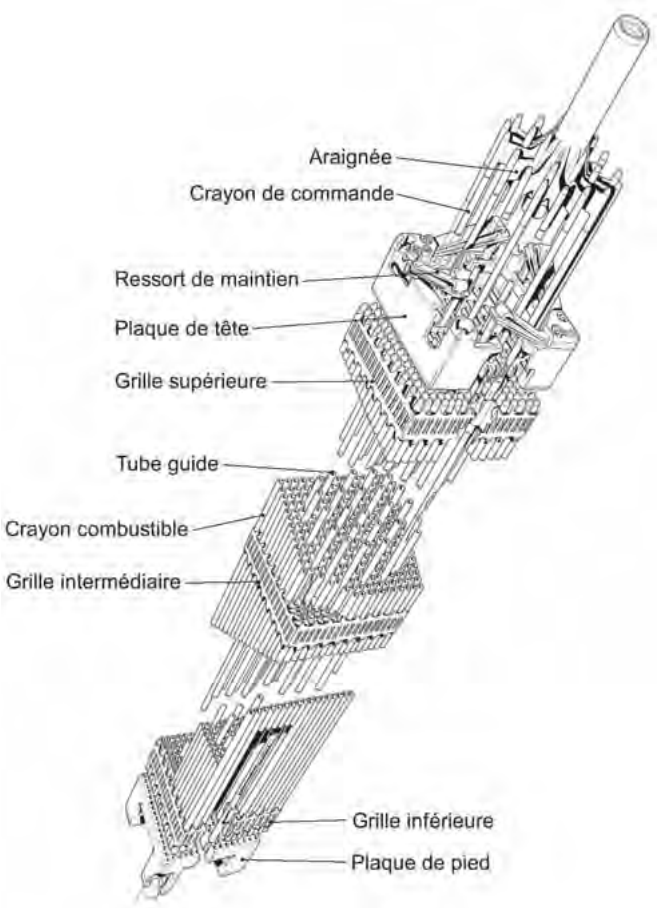


FIGURE 1.7 – Assemblage 17×17 d'un REP (source : Collection Génie Atomique)

Initialement, les **Assemblages de Combustibles (AC)** étaient entourés d'un boîtier et le contrôle de la puissance était assuré par des barres cruciformes s'insérant entre les assemblages. La conception du cœur et des assemblages de REP a évolué dans les années 1960 dans le but de réduire l'absorption de neutrons liés aux boîtiers et les coûts de fabrication, mais surtout pour éliminer les pics de puissance en périphérie des assemblages liés à la réactivité qu'apporte localement le canal d'eau inter-assemblage. Actuellement, les assemblages combustibles des REP sont constitués d'un faisceau à pas carré de crayons combustibles et de tubes guides permettant le passage des barres de commande et de l'instrumentation du cœur.

Le *crayon combustible* est constitué d'un empilement de pastilles d'UO_2 faiblement enrichi (de 3 à 5%) ou d'un mélange de PuO_2 et d'UO_2 (combustible MOX). La teneur initiale en plutonium est de 5.3%, et résulte d'une équivalence énergétique avec du combustible UO_2 initialement enrichi à 3.25% en ^{235}U. Pour certains assemblages qui permettent d'atteindre des taux d'irradiation élevés, afin de compenser l'excès de réactivité en début de vie, on insère dans la matrice de la pastille un absorbant neutronique qui disparaîtra avec l'irradiation. Ce poison consommable est soit de l'oxyde de gadolinium Gd_2O_3, soit de l'erbium Er.

Le *gainage* du combustible est assuré par un alliage de zirconium (Zircaloy) qui, outre de bonnes propriétés thermiques et mécaniques, n'absorbe pas les neutrons. La tenue du gainage est essentielle pour assurer le confinement des produits de fission et des actinides mineurs dans la plupart des situations accidentelles. Un grand nombre des critères de sûreté qui devront être respectés dans les situations incidentelles et accidentelles permettant de dimensionner les systèmes assurant la sûreté du réacteur sont liés à la tenue thermomécanique de ce matériau.

Le maintien et l'espacement des crayons combustibles sont assurés par des *grilles*. Ces grilles sont dotées d'ailettes sur leur partie supérieure. Ces ailettes permettent la déviation des filets fluides le long des crayons et favorisent ainsi le mélange turbulent et l'uniformisation des températures du fluide dans les assemblages [5], l'eau circulant entre les crayons, du bas vers le haut, à environ 4 m/s.

Les mécanismes actionnant les grappes de commande sont placés sur le couvercle de la cuve, et le système d'instrumentation interne du cœur permet de faire pénétrer à travers la partie inférieure de la cuve des « chambres à fission » pouvant mesurer, en marche, le flux neutronique du cœur (voir la sous-section 2.2.4).

La Figure 1.8 montre de façon schématique la disposition des assemblages combustibles dans le cœur. Leur nombre dépend de la puissance du réacteur : il y en a 157 pour un réacteur de 900 MWe, 193 pour un réacteur de 1300 MWe et 205 pour un réacteur de 1450 MWe, ce qui est le cas de la figure 1.8). Globalement, le cœur est à peu près un orthocylindre de 4 m de hauteur, et de 4 m de diamètre environ.

5. La conception des cœurs et assemblages combustibles des réacteurs VVER est différente : si, comme pour les REP occidentaux, le combustible utilisé est constitué par un empilement de pastilles de combustible oxyde dans une gaine de Zircaloy, les crayons sont rangés selon un pas triangulaire dans des assemblages hexagonaux entourés par un boîtier. Le contrôle du cœur est assuré à la fois par des éléments de contrôle hexagonaux s'insérant dans des emplacements du cœur non occupés par des AC et par des barreaux absorbants coulissant dans les tubes guides des assemblages.

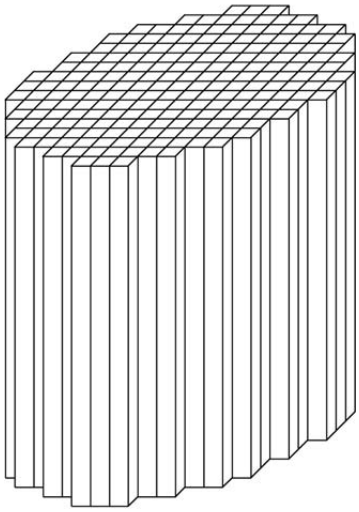


FIGURE 1.8 – Représentation schématique de la disposition des AC dans le cœur (source : Collection Génie Atomique)

On remarquera qu'il n'y a pas de cloisonnement entre les assemblages, mais seulement un faible jeu de l'ordre du millimètre, suffisant pour permettre l'insertion ou l'extraction d'un assemblage.

Les réacteurs industriels de type REP sont gérés par fraction de cœur (tiers ou quart) selon une fréquence approximativement annuelle ou un peu plus longue, par périodes de fonctionnement appelées **cycles**. En effet, une cuve fermée et pressurisée ne permet pas d'envisager un renouvellement du combustible en marche.

Lors de chaque arrêt (quelques semaines, pendant lesquelles on profite de l'occasion pour effectuer également des opérations de maintenance), la cuve est ouverte et le cœur entièrement déchargé. Dans une gestion par tiers, par exemple, un tiers des assemblages – ceux qui ont subi trois cycles et sont les plus irradiés – est déchargé définitivement, et les deux autres tiers sont remis en cœur, mais chaque assemblage dans une position différente de celle qu'il occupait précédemment. Le chargement est complété par un tiers d'assemblages neufs. Le plan de rechargement est soigneusement élaboré pour optimiser la distribution du flux neutronique, donc de la puissance, au cours du cycle à venir.

1.4 Cycle

Dans cette section, nous décrivons schématiquement le déroulement d'un cycle normal d'un réacteur EDF. Le protocole écrit ici est relatif à l'EPR.[Div06]

1.4.1 Généralités

La conduite normale comprend :

— le fonctionnement en puissance et ses transitoires normaux d'exploitation programmés tels que la montée en charge, baisse de charge, suivi de charge, arrêt ou démarrage de la tranche ;

— les fonctionnements particuliers dus à des événements non programmés (exemple : îlotage, perte de sources, etc. ...).

Hormis pour les arrêts pour rechargement, la tranche peut être arrêtée pour des arrêts de plus ou moins courte durée pour des opérations de maintenance ou de réparation, d'économie de combustible ou de gestion du réseau électrique. L'état d'arrêt, Arrêt à Chaud (AAC) ou Arrêt sur RIS-RA, dépendra alors de la nature de l'intervention et de la durée d'arrêt. En cas de maintien en Arrêt A Chaud, la concentration en bore du circuit primaire permettant d'assurer la marge d'antiréactivité est adaptée et dépend de l'épuisement du combustible ainsi que de la durée de cet arrêt.

Le passage en arrêt à froid est réalisé afin d'effectuer le rechargement du combustible ou de permettre des opérations de maintenance ou de réparation nécessitant cet état de tranche.

Les principales opérations de conduite sont décrites dans cette section de manière chronologique, depuis l'arrêt du réacteur en fin de cycle combustible jusqu'au fonctionnement en puissance au début du cycle suivant. Le fonctionnement en prolongation de cycle est également présenté.

1.4.2 Arrêt du réacteur

A la fin d'un cycle combustible, on procède à l'arrêt de tranche : les grappes d'arrêt (différentes des grappes de commandes) sont insérées manuellement, et une séquence automatique assure le refroidissement et la dépressurisation simultanés du circuit primaire ; parallèlement, une borification est effectuée afin de respecter la marge d'antiréactivité requise.

Le nombre de pompes primaires en service est ajusté afin de permettre un refroidissement efficace : la dernière pompe est arrêtée lorsque les critères radio-chimiques sont corrects et que la température du couvercle cuve est inférieure à $70°C$: la température du circuit primaire est alors à $55°C$. Tout au long du refroidissement du primaire, sa contraction est compensée par les pompes des circuits auxiliaires qui aspirent dans les réservoirs de stockage d'acide borique et d'eau déminéralisée : la borification vers la concentration de bore requise en Arrêt à Froid pour rechargement est effectuée en parallèle.

Après l'arrêt du dernier GMPP, la pression primaire est réduite à 5 bars et cette valeur de pression est maintenue par appoint d'azote pendant le refroidissement final. Le pressuriseur est ainsi maintenu en mode diphasique.

1.4.3 Vidange et ouverture du circuit primaire

Le primaire est vidangé jusqu'au niveau 3/4 de boucle par les circuits auxiliaires, et le volume de fluide primaire est transféré vers les réservoirs de stockage du TEP pour

recyclage : une régulation du niveau d'eau à 3/4 des boucles primaires permet de prévenir un découvrement du cœur, et de garantir un fonctionnement sûr des circuits auxiliaires.

Avant rupture de l'intégrité du primaire, un balayage est effectué par injection d'azote via les corps des pompes primaires et l'évent du couvercle de la cuve, l'évacuation étant réalisée par la pompe à vide reliée à l'évent du pressuriseur. Un balayage en air est ensuite effectué. Le niveau d'eau du RCP est maintenu et régulé en automatique au niveau 3/4 de boucle avant l'ouverture de la cuve.

Les connexions électriques des mécanismes de commande de grappes et de l'instrumentation du cœur sont déposées. Les étanchéités mécaniques sont déposées, et l'intégrité du primaire est alors rompue : après la dépose du calorifuge, la machine de serrage et desserrage des goujons est mise en place pour les opérations d'ouverture du couvercle cuve.

1.4.4 Déchargement du cœur

Pendant la dépose du couvercle cuve, les compartiments de la piscine du Bâtiment Réacteur (BR) sont remplis avec de l'eau borée : lorsque la piscine est pleine, l'instrumentation du cœur, dont les systèmes de mesure du flux neutronique, est déposée et les mécanismes de commande des grappes sont déconnectés. Les équipements internes supérieurs sont alors retirés et positionnés dans leur piscine de stockage, et les opérations de déchargement combustible peuvent commencer (environ 70 heures, voire moins, après le découplage). La température du fluide primaire est maintenue inférieure à $50°C$. Le déchargement du combustible dure environ 40 heures.

La puissance résiduelle des éléments combustibles déchargés vient s'ajouter à celle des éléments combustibles déjà stockés dans la piscine de désactivation du bâtiment combustible : il devient ainsi nécessaire de la refroidir pour en maintenir la température inférieure à $50°C$ (du début du déchargement du cœur jusqu'à la fin du rechargement).

Une fois le cœur déchargé, la suite des opérations dépend des travaux programmés. Le RCP est vidangé jusqu'au niveau de la génératrice inférieure par les pompes de purification, et les travaux de maintenance prévus sont réalisés dans cet état de « Réacteur Complètement Déchargé » : contrôle des tubes des GV, permutations des grappes de commande, contrôle des éléments combustibles, etc. ...

1.4.5 Rechargement du cœur

Après fermeture des composants primaires, le compartiment cuve est rempli en eau borée, et le combustible est chargé en cuve : la température du RCP est toujours maintenue inférieure à $50°C$. Les opérations de chargement et de cartographie du cœur durent environ 45 heures.

Une fois le chargement effectué, le tube de transfert est fermé. Les internes supérieurs sont remis en position, les grappes de commande sont reconnectées, et l'instrumentation du cœur réinstallée.

1.4.6 Fermeture et remplissage du circuit primaire

Les compartiments piscines du BR sont vidangés via les pompes de purification, jusqu'au plan de joint cuve.

Le fond de la piscine cuve et le plan de joint cuve sont nettoyés. La cuve du réacteur est fermée, les étanchéités des pénétrations couvercle sont remontées et l'évent du couvercle cuve est fermé. Lorsque le circuit primaire est pressurisable, les circuits auxiliaires sont protégés par les soupapes du pressuriseur. Les connexions électriques des mécanismes de commande des grappes et de l'instrumentation du cœur sont installées.

Au cours de ces opérations, la température du réfrigérant primaire est toujours régulée par les circuits auxiliaires.

La vidange du primaire est ensuite reprise jusqu'au niveau 3/4 de boucle afin de mettre en communication les phases gazeuses de la cuve, du pressuriseur et des tubes GV. Le niveau RCP est réglé automatiquement, toujours afin de prévenir tout découvrement du cœur et garantir un fonctionnement sûr des circuits auxiliaires.

Le circuit primaire est ensuite mis sous vide : la pression obtenue est d'environ 200 $mbar$ (absolus), afin de minimiser la teneur en oxygène du fluide primaire. La température est maintenue inférieure à $40°C$ durant cette phase, afin d'assurer une marge à la saturation suffisante du circuit primaire.

Le RCP est ensuite rempli, et dégazé à grand débit : le remplissage est arrêté lorsque le niveau pressuriseur atteint 7m. Le démarrage de la tranche est réalisé directement en diphasique.

1.4.7 Chauffage du fluide primaire

Après l'arrêt du dispositif de mise sous vide, la pression primaire est augmentée par la mise en marche des chaufferettes du pressuriseur, sans dépasser les 30 $bars$. Il n'y a pas de passage du RCP en monophasique. Les pompes primaires sont démarrées à une pression minimale de 25 $bars$, et la première pompe est démarrée avant l'atteinte des $65°C$ au primaire. Après le démarrage des GMPP, la pression du RCP est placée sous régulation automatique : la puissance fournie par les quatre pompes primaires et la puissance résiduelle du combustible permettent le réchauffage progressif du fluide primaire. Le gradient de chauffe est limité à $50°C/h$ (valeur maximale, si la puissance ainsi fournie le permet).

En parallèle, la partie secondaire est également rendue disponible : en particulier, audessus de $120°C$, les quatre GV sont remplis jusqu'à leur niveau de consigne à charge nulle. Lorsque la température du RCP atteint $120°C$ les circuits auxiliaires encore connectés sont isolés, et le contrôle de la température est alors assuré par les GV.

Durant la chauffe, le volume excédentaire dû à la dilatation du primaire est évacué vers les réservoirs de stockage TEP : parallèlement, la pression est progressivement et automatiquement augmentée jusqu'à l'atteinte progressive des conditions de pression et température souhaitées.

Le RCP est alors en état d'Arrêt à Chaud.

1.4.8 De l'Arrêt à Chaud au fonctionnement en puissance

A ce stade, en préalable à la recherche de criticité, toutes les fonctions de sûreté doivent être disponibles. En Arrêt à Chaud, divers essais sont réalisés, tels que le temps de chute des grappes, par exemple. La température du primaire est réglée automatiquement. Les essais à puissance nulle sont réalisés et le primaire est dilué par injection d'eau déminéralisée. La puissance est ensuite augmentée par la régulation du niveau de flux : les générateurs sont alimentés par la pompe de démarrage et d'arrêt (AAD), puis par les pompes alimentaires (APA) via l'alimentation normale des générateurs de vapeur (ARE).

La turbine est lancée, l'alternateur couplé au réseau principal, et la puissance ensuite progressivement augmentée. Le mode normal de régulation chaudière (priorité à la turbine) est substitué à la régulation du niveau de flux. A ce niveau de puissance, toutes les régulations du RCP sont en automatique, et la puissance est progressivement augmentée jusqu'à 100%.

1.4.9 Fonctionnement en puissance - suivi de charge

En fonctionnement de base, seuls les effets de réactivité à très long terme (épuisement du combustible, accumulation de samarium...) doivent être compensés par une dilution progressive du fluide primaire jusqu'à pratiquement 5 à 10 *ppm* de bore à la fin de cycle combustible.

Si l'équilibre production-consommation du réseau le requiert, la centrale peut être amenée à réduire sa puissance, puis à revenir à pleine puissance quelques heures plus tard. Comme mentionné ci-dessus, toutes les régulations chaudière sont en automatique. Les grappes de commande sont insérées ou extraites par les régulations de température et de distribution de puissance pour compenser les variations rapides de réactivité. Les variations lentes (évolution Xénon) sont compensées par une modification de la concentration en bore ou par un mouvement des grappes de commande.

En plus des besoins en borification et dilution, le fluide primaire est traité afin de respecter les critères chimiques ou d'activité primaire. Les volumes correspondants peuvent être recyclés, notamment les concentrats (acide borique). Une partie des distillats (eau déminéralisée) peut être envoyée vers le traitement des effluents primaires afin d'éviter l'accumulation de tritium dans le circuit primaire.

Dans cette longue période de fonctionnement en puissance (plus d'un an), si la norme reste le régime hyperstable à 100% de puissance nominale, il existe ainsi des fluctuations de température, puissance et de débit, de même que de légers échanges de fluides entre le circuit primaire et ces circuits auxiliaires, pour les régulations radiochimiques ou limiter les effets de contraction/dilution.

1.4.10 Fonctionnement en prolongation de cycle

Avec l'augmentation de l'épuisement combustible, la concentration en bore est continuellement réduite. La fin du cycle est atteinte lorsque la concentration en bore atteint une valeur minimale proche de zéro.

Afin de poursuivre le fonctionnement en puissance au-delà de la fin naturelle du cycle, la baisse de réactivité due à l'usure du combustible est compensée par la réduction de la température primaire.

Avec les grappes de commande presque totalement extraites et les vannes d'admission turbine ouvertes en grand, le niveau de puissance de l'installation est fixé par le bilan de réactivité du cœur et les caractéristiques de la turbine.

Comme il n'y a plus de réserve de réactivité disponible pour assurer une température moyenne primaire constante, la température moyenne primaire et la puissance du réacteur ainsi que la pression vapeur décroissent régulièrement. La masse de fluide primaire est maintenue constante pendant le fonctionnement en puissance. La baisse de la température du primaire entraîne donc un réajustement des principaux paramètres (niveau du pressuriseur, température de référence, signal de protection, etc...).

Le fonctionnement en prolongation de cycle (ou *stretch*), basé sur des ajustements répétés de points de consigne, consomme les réserves de réactivité jusqu'à épuisement.

1.5 Enjeux

1.5.1 Connaissance du Point de Fonctionnement Primaire

Comme nous l'avons expliqué en introduction, le retour d'expérience des exploitants et des ingénieries de centrales montre que le point de fonctionnement du circuit primaire à puissance nominale – ou autour de la puissance nominale – n'est pas connu aussi précisément que souhaité, ce qui est pris en compte pour les marges de sûreté.

En cours de cycle, l'estimation aussi précise du point de fonctionnement primaire s'effectue notamment lors d'Essais Périodiques (EP), au cours desquelles la tranche est placée en régime ultra-stable et un relevé détaillé de mesures est effectué, qui est utilisé pour vérifier le respect des différents critères de sûreté régissant strictement le fonctionnement du réacteur. Pour chaque grandeur, une incertitude se traduit ainsi par une marge importante pour le pilotage de la tranche.

Ainsi, en tout début de cycle, l'Essai Périodique RCP114 détermine détermine, à partir des mesures de température de Branche Chaude et des bilans enthalpiques au secondaire, une estimation aussi précise que possible des débits-boucles et de l'incertitude correspondante, estimation qui est conservée une fois pour toute pour l'ensemble du cycle. Or la représentativité de la mesure de température de Branche Chaude est affectée par une hétérogénéité du champ de température en Branche Chaude, ce qui impacte l'incertitude associée à l'estimation des débits-boucles par RCP114. La mesure de puissance thermique au secondaire est par ailleurs affectée par différents phénomènes (encrassements, colmatages, etc...).

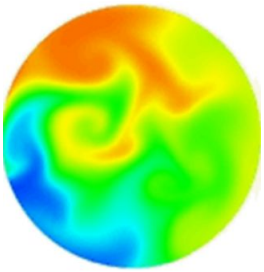


FIGURE 1.9 – Hétérogénéité du champ de Température en Branche Chaude (obtenue par simulation *Code_Saturne*®)

D'une manière générale, ces estimations ne sont disponibles au mieux qu'aux dates des Essais Périodiques, c'est-à-dire une fois par mois environ : entre ces dates, en temps continu, moins de mesures sont disponibles et leur précision est moindre.

Par conséquent, le projet d'EDF R&D PAPRICA et son successeur SOPRANO, dans lesquels s'inscrit cette thèse, ont conçu de nouveaux estimateurs plus précis et plus fiables dans le but d'améliorer cette connaissance du point de fonctionnement primaire, que ce soit en termes de représentativité ou en termes de réductions d'incertitude.

On se limitera ainsi, dans tout le travail de thèse, à l'étude d'un réacteur en **fonctionnement normal et à puissance nominale ou quasi-nominale** : en particulier, on n'évoquera pas ici des situations dégradées autres que des pertes temporaires d'instrumentations de mesure.

1.5.2 Complexité de la modélisation physique

Comme nous l'avons vu, un REP est un process industriel complexe et fortement hétérogène. En particulier, au sein du circuit primaire, les conditions sont extrêmes :

— températures : $290°C$ en entrée et $330°C$ en sortie ;
— pression : 155 bars ;
— vitesse : 1 seconde pour 4 m de hauteur cœur ;

— irradiation forte in-core.

Le nombre de Reynolds de l'écoulement du fluide primaire est approximativement de 10^7, et l'écoulement est donc fortement turbulent : de plus, la structure solide du cœur (agencement des Assemblages Combustibles, grilles, ailettes...) est conçue pour accroître cette turbulence et favoriser ainsi le brassage de la température.

De fait, modéliser la physique fine du circuit primaire n'est pas un problème facile, et de nombreux codes de calcul industriels ont été développés pour reconstruire les champs de débits, températures et flux neutroniques au sein des différentes parties du circuit primaire. Ces codes de calcul, qui mélangent équations théoriques et corrections empiriques, sont le fruit de nombreuses années de recherche et de retours d'expérience, et sont régulièrement validés à partir à partir d'expérimentations sur maquette ou sur site ; néanmoins, outre le fait qu'ils sont parfois protégés par le secret industriel, leurs coûts de calcul sont parfois

rédhibitoires. Si l'on souhaite obtenir *in fine* une description exhaustive 3D réaliste de l'ensemble du circuit primaire, utiliser ces codes semble pourtant indispensable.

Mais doit-on se fixer pour objectif un tel niveau de description pour la conduite d'un réacteur ? Les opérateurs des tranches adoptent quotidiennement une description plus simple du circuit primaire, en utilisant seulement quelques scalaires pour représenter l'état physique de chacun des secteurs, et quelques équations simples pour relier ces différents scalaires entre eux. Le besoin concret des exploitants, en ce qui concerne l'estimation du point de fonctionnement primaire, semble ainsi relever plus de la modélisation 0D que de la modélisation 3D : obtenir des paramètres d'entrée plus précis pour les codes 3D (comme par exemple une cartographie de température 2D à l'interface cœur / plénum supérieur) demeure cependant un objectif également intéressant aux yeux de l'exploitant.

L'un des enjeux de la thèse est donc de définir un niveau de modélisation adapté aux besoins des exploitants, tout en bâtissant une méthodologie qui puisse être par la suite applicable à des codes industriels plus complexes.

1.5.3 Limitation de l'instrumentation in situ

L'un des premiers constats du projet PAPRICA est que l'on n'envisage pas l'ajout de nouvelles instrumentations intrusives sur le circuit primaire, du moins pas sur les réacteurs actuels. Les raisons avancées sont principalement les suivantes :

— tout ajout de matériel à l'intérieur de l'enceinte de confinement nécessite des études de sûreté importantes ;

— un point de mesure étant une faiblesse structurelle supplémentaire potentielle, il est nécessaire limiter le nombre de ces points de mesures sur le circuit primaire et le compromis actuel entre représentativité et « légèreté » du système de mesures convient à l'exploitant.

L'amélioration de la connaissance du point de fonctionnement primaire passe ainsi essentiellement par l'utilisation de modèles et l'exploitation optimale de l'instrumentation déjà présente in situ (et parfois utilisées uniquement pour des situations accidentelles), à savoir :

— les ΔP-Coude permettant de détecter des pertes de GMPP (avec un seuil de -10% du débit) ;

— les sondes de température en Branche Froide ;

— les sondes de température en Branche Chaude, potentiellement affectées par une hétérogénéité de température ;

— les thermocouples RIC, prévus pour les situations accidentelles et dont les valeurs brutes sont réputées peu fiables car soumises à une forte irradiation dans le cœur causant biais et dérives ;

— les CFM (chambres à fission mobiles) utilisées pour réaliser les cartes de flux ;

— les BIL100/BILORLI réalisés à partir de l'instrumentation située sur le circuit secondaire et permettant d'estimer la puissance thermique transmise du primaire au secondaire au niveau des GV ;

— éventuellement les chambres d'ionisation externes (RPN) qui évaluent en temps réel la répartition de l'activité neutronique dans différents grands secteurs du cœur.

Une description plus approfondie de ces différentes mesures est proposée au Chapitre 2.

A cette liste réduite de systèmes de mesure, il convient d'ajouter différents estimateurs – notamment pour la température de branche chaude et le débit-boucle – qui ont été développés ces dernières années par STEP pour pallier les problèmes d'hétérogénéité de Branche Chaude, estimateurs souvent interdépendants et de précisions diverses.

Un enjeu important de la thèse sera ainsi d'utiliser en synergie toutes ces instrumentations et tous ces estimateurs disponibles, afin d'obtenir une estimation plus juste de l'ensemble du circuit primaire avec le niveau de description adapté choisi.

Chapitre 2

Données d'entrée

Dans ce chapitre, nous effectuons l'inventaire des données d'entrée disponibles pour le travail de thèse : les mesures disponibles au niveau du circuit primaire et les chaînes d'acquisition correspondantes ; les différents estimateurs conçus pour notamment la température de branche chaude TBC et le débit-boucle Q ; les différents modèles et codes de calcul que l'on peut utiliser pour calculer le point de fonctionnement primaire.

2.1 Notations

Pour toute la suite, nous adopterons les notations suivantes. Pour la boucle numéro i :

— Q_i le débit-boucle massique (en tonnes par secondes $(t.s^{-1})$) ;

— TBF_i la température en Branche Froide (en $°C$) ;

— TBC_i la température en Branche Chaude (en $°C$) ;

— $PTHGV_i$ la Puissance Thermique transférée au secondaire via le Générateur de Vapeur (en MW).

Pour l'Assemblage Combustible numéro j (en numérotant ligne par ligne, de L1 à E15, cf Figure 2.7) et Figure 2.11) :

— $PIPA_j$ la Puissance Intégrée Par Assemblage (en MW) ;

— $PIPAN_j$ la Puissance Intégrée par Assemblage normalisée (sans dimension) ;

— PL la Puissance Locale au niveau de l'échelon numéro h (en MW) ;

— $PLN_{j,h}$ la Puissance Locale (linéique) Normalisée au niveau de l'échelon numéro h (sans dimension) ;

— $THAC_j$ la Température en Haut de l'Assemblage Combustible (au niveau de l'interface cœur - plénum supérieur) (en $°C$).

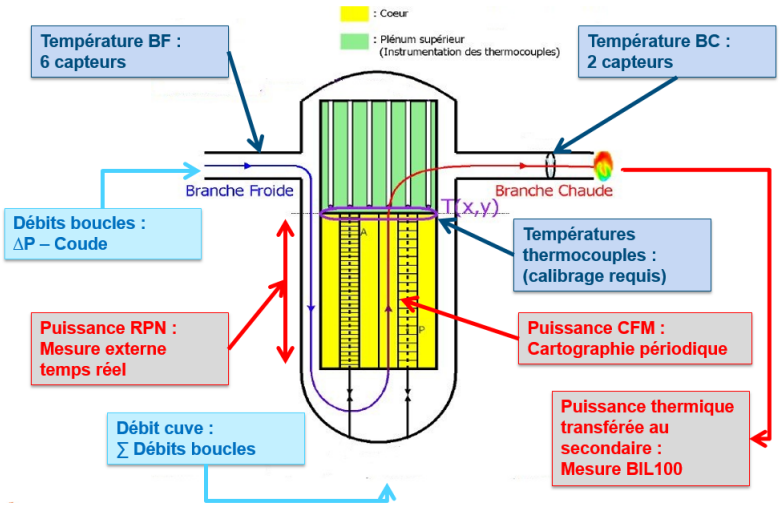


FIGURE 2.1 – Résumé des systèmes de mesure disponibles sur le circuit primaire

Enfin, pour la totalité du circuit primaire :

— $PTHREAC$ la puissance thermique du cœur (en MW) ;

— $PGMPP$ la puissance des pompes primaires (en MW).

2.2 Mesures

Dans cette section, nous détaillons l'ensemble des mesures réalisées autour du circuit primaire sur le REP, et enregistrés dans les différentes bases de données d'EDF. Une bonne synthèse en a été effectuée par [Dum10].

2.2.1 Bases de Données

On distingue différents types de bases de données, chacune ayant une fonction différente.

ORLI

L'application ORLI contient toutes les données brutes provenant des signaux délivrés en temps continu par les différents capteurs installés sur site (données de type M, issues du contrôle-commande).

Chaque instrument de mesure est représenté par une chaîne de caractères constituée :

— de deux lettres majuscules pour indiquer le CNPE correspondant (exemple : NO pour Nogent-sur-Seine, BE pour Belleville, etc...) ;

Trigramme	Sous-Système
AAD	Ensemble Motopompe Alimentaire d'arrêt et de démarrage
APA	Ensemble Motopompe Alimentaire (y compris graissage
ARE	Alimentation Normale des GV
ASG	Alimentation Auxiliaire de secours des GV
RCP	Circuit Primaire
RCV	Contrôle Chimique et Volumétrique
RIC	Instrumentation Interne du Cœur
RIS	Injection de sécurité
RPE	Purges, évents et exhaures nucléaires
RPN	Mesure de la Puissance Nucléaire
TEP	Effluents Liquides Primaires
VVP	Circuit Vapeur Principal, soupapes, évents et vannes GV

TABLE 2.1 – Exemples de trigrammes du système élémentaire

— d'un chiffre pour le numéro de la tranche correspondante ;

— d'un trigramme pour indiquer le sous-système concerné (voir Table 2.1) ;

— de trois chiffres pour indiquer le numéro de l'instrument (lorsque le système est installé sur une boucle en particulier, le premier chiffre est celui de la boucle ;

— d'un suffixe indiquant le type de grandeur mesurée ('MT-' pour une mesure de température, 'MD-' pour une mesure de débit, etc....).

Les données provenant d'ORLI sont à utiliser avec précaution : en effet, à l'état brut, les signaux sont enregistrés en temps quasi-continu (2 secondes pour les puissances, pressions, débits, niveaux et 10 secondes pour les températures) dans la base de données et lorsque l'instrument concerné est indisponible, un signal aberrant (e.g. 549302 pour une mesure de température en $°C$) est enregistré à la date concernée. Il peut également arriver que l'acquisition soit figée pendant un certain temps, la mesure enregistrée chaque seconde étant alors la même au cinquième chiffre significatif près... Il est donc indispensable lors de l'extraction de données de procéder à un prétraitement, d'une part en procédant à un échantillonnage (typiquement par intervalles de vingt minutes), d'autre part en éliminant systématiquement les valeurs aberrantes ou figées.

MOLENE

MOLENE est la base de données dans laquelle sont archivées les données dites de type Y (mesures d'essais), faisant référence aux capteurs d'essais utilisés notamment pendant les EP (BIL100, CEF...) : ces capteurs sont particulièrement précis mais l'accès à leurs mesures en temps réel est actuellement non disponible. Il faut qu'un opérateur lance un EP pour que les valeurs mesurées par ces capteurs soient relevées et consignées dans la base de données après vérification : il n'existe donc a priori aucune valeur aberrante ou

figée enregistrée dans MOLENE.

Les données de type Y enregistrées dans MOLENE sont ainsi plus fiables mais moins nombreuses que les données de type M enregistrées dans ORLI.

MUSICALE

La base de données MUSICALE contient l'ensemble des cartes de flux obtenues à partir des Chambres à Fission Mobile et des codes CABRI et CAMARET. Chaque carte de flux prend la forme d'un document texte (de type RES) contenant une liste d'informations diverses liées à l'état de la tranche et du cœur : la structure du document étant fixe, il est possible de programmer une routine informatique qui permette d'extraire les données numériques souhaitées de chaque document RES.

SYGMA

La base de données SYGMA contient des informations de maintenance en exploitation, dont les données relatives aux calibrages de certaines instrumentations : nous nous intéresserons uniquement aux valeurs dites ΔP_{CAL} qui servent à calibrer la méthode de ΔP-Coude (voir la sous-section 2.2.2) qui y sont consignées.

2.2.2 Débit Primaire

Il existe deux méthodes de mesures de débit primaire : la méthode RCP114, mesure de « débit enthalpique », et la méthode de mesure directe par ΔP-coude.

RCP114

L'Essai Périodique **RCP114** est un essai réalisé une fois par cycle au redémarrage de la tranche, après atteinte des $100\%PN$ et qui vise à calculer les débits boucles et débit cuve par bilan enthalpique afin de vérifier les critères de sûreté et les débits de conception. Lors du premier démarrage d'une tranche, cette mesure de débit est opérationnelle à partir de $30\%PN$ jusqu'à $100\%PN$; pour les cycles suivants, elle est effectuée uniquement après atteinte des $100\%PN$.

Le principe de cette mesure est l'utilisation du bilan enthalpique entre le primaire et le secondaire :

Théorème 1 (Bilan Enthalpique GV côté primaire)**.**
Au niveau du générateur de vapeur de la boucle i, la puissance thermique cédée par le fluide primaire est égale à la puissance thermique reçue par le fluide secondaire, soit :

$$PTHGV_i = Q_i \times (h(TBC_i) - h(TBF_i)) + \frac{PGMPP}{4} \tag{2.1}$$

où :

— *$PTHGV_i$ est la puissance transférée au secondaire à travers le générateur de vapeur ;*

— *Q_i est le débit-boucle massique (en $t.s^{-1}$ (tonnes par seconde)) ;*

— *$h(T)$ est l'enthalpie massique de l'eau à la température T ($°C$) et à la pression $P = 155$ bar (en $kJ.kg^{-1}$) ;*

— *TBC_i est la température de Branche Chaude ;*

— *TBF_i est la température de Branche Froide ;*

— *$PGMPP$ est la puissance de l'ensemble des pompes primaires (approximativement 18 MW).*

Ce théorème peut être considéré comme une définition de TBC.

Lors de l'essai RCP114, $PTHGV$ est évalué pour chaque boucle à partir des mesures effectuées sur le circuit secondaire (mesure dite de BIL100, bilan enthalpique à $100\%PN$: voir plus loin le Théorème 2), et les températures de branche chaude et de branche froide sont mesurées au niveau des bypass directement à partir de la sonde de réserve, ce qui permet de s'affranchir de l'incertitude de la chaîne d'acquisition. En inversant le bilan enthalpique du théorème 1, on peut alors obtenir une évaluation du débit-boucle Q_i par la formule suivante :

$$Q_i = \frac{PTHGV_i - \frac{PGMPP}{4}}{h(TBC_i) - h(TBF_i)} \tag{2.2}$$

Le débit obtenu est le débit massique : on obtient le débit volumique en branche froide en divisant le débit massique par $\rho(TBF_i)$, la masse volumique de l'eau à pression 155 *bar* et à température TBF_i.

Ce calcul s'effectue au moyen des valeurs mesurées moyennées sur 20 minutes, pour chacun des paramètres. La puissance est, au moment de l'essai, stabilisée autour de la puissance nominale pour que l'évaluation du débit-boucle puisse être validée. Cette valeur est ensuite conservée pour tout le reste du cycle.

L'étude des sensibilités de cette formule permet de mettre en évidence que les paramètres les plus influents pour le calcul de débit RCP114 sont la mesure de Température de Branche Chaude (TBC_i) et la mesure de Température de Branche Froide (TBF_i) : cette estimation pour le débit-boucle est donc impactée par l'existence potentielle d'un biais d'hétérogénéité de température en Branche Chaude.

Au final, en appliquant cette méthode à un palier 1300 MW, avec des valeurs de référence pour les différents paramètres, on obtient une valeur de référence pour le débit RCP114 de l'ordre de 5.0 $t.s^{-1}$ en débit massique (tonnes par seconde), soit en débit volumique branche froide un débit de l'ordre de 24000 $m^3.h^{-1}$, et l'incertitude finale attribuée à cette mesure est approximativement de $\pm 4.2\%$.

$$Q_{RCP114} = 5.0 \pm 0.21 \ t.s^{-1} = 24000 \pm 1008 \ m^3.h^{-1} \ \textbf{(soit } \pm 4.2\% QN\textbf{)}$$

Le débit-cuve étant la somme des débits-boucles, et avec une hypothèse d'indépendance des erreurs de mesures de débits RCP114 entre les boucles (hypothèse non absurde puisque chaque boucle possède un système d'acquisition qui lui est propre) l'incertitude sur le débit cuve est alors la somme quadratique des incertitudes des débits-boucles, d'où le résultat :

$$Q_{cuve,RCP114} = 20.0 \pm 0.42\ t.s^{-1} = 96000 \pm 2016\ m^3.h^{-1}\ \textbf{(soit } \pm 2.1\% Q_{cuve} N\textbf{)}$$

ΔP-Coude

On cite notamment [DA11].

La méthode de mesure de débit par ΔP-Coude consiste à mesurer une différence de pression entre l'intrados et l'extrados du coude en sortie de Générateur de Vapeur (voir Figure 2.2). Cette différence de pression résulte des efforts – différents entre l'extrados et l'intrados – exercés par le fluide circulant dans le tuyau sur ses parois, et il a été démontré qu'on peut la relier au débit-boucle via la formule suivante :

$$Q_{vol} = K D^{\frac{3}{2}} R_c^{\frac{1}{2}} \sqrt{\frac{\Delta P}{\rho}} \tag{2.3}$$

où :

— Q_{vol} est le débit-boucle volumique en branche froide exprimé en $m^3.s^{-1}$;

— K est un coefficient de proportionnalité dépendant de l'écoulement, de la géométrie du coude et de la position du capteur sur la paroi du coude ;

— D est le diamètre du cube exprimé en m ;

— R_c est le rayon de courbure du coude exprimé en m ;

— ΔP est la différence de pression mesurée entre intrados et extrados mesurée par le capteur, exprimée en $mbar$;

— ρ est la masse volumique du fluide exprimée en $kg.m^{-3}$ (approximativement 745 $kg.m^{-3}$ en branche froide).

Pour mettre en place cette méthode de mesure, il faut ainsi d'une part disposer d'un capteur qui mesure la valeur absolue ΔP, en $mbar$, et d'autre part connaître les paramètres D, R_c et surtout K.

Les capteurs de pression sont situés dans un plan à 22.5° du coude de 40° en sortie des boîtes à eau des GV. Les paliers 900 MW possèdent un piquage à l'extrados et trois piquages à l'intrados (−15°,0° et 15°) tandis que les paliers 1300 MW et N4 possèdent un piquage à l'extrados et quatre piquages à l'intrados (−22.5°,−7.5°,7.5° et 22.5°) (voir Figure 2.4). Ils ne permettent pas d'obtenir directement une valeur en $mbar$. L'intensité aux bornes des capteurs est mesurée puis convertie en tension à l'aide d'une résistance, et celle-ci est convertie en $\%QN$, une grandeur arbitraire représentant le débit nominal (voir Figure 2.3).

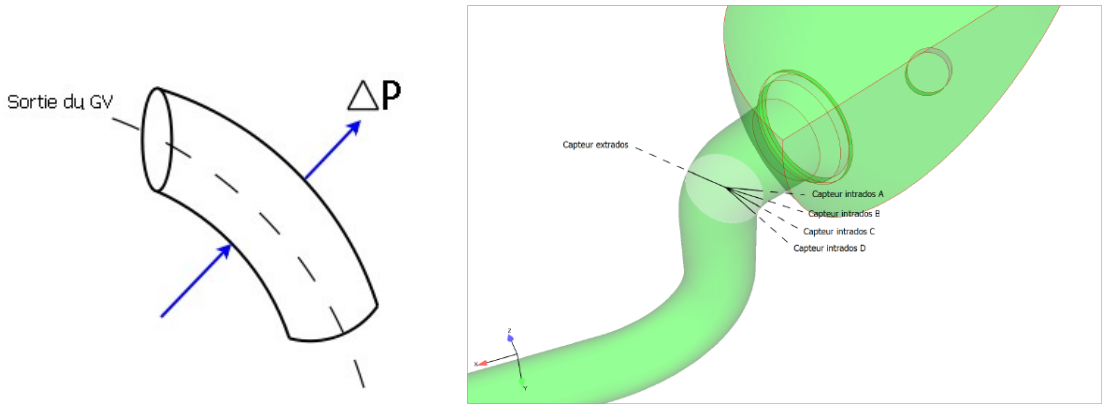


FIGURE 2.2 – Principe de la mesure par ΔP-Coude

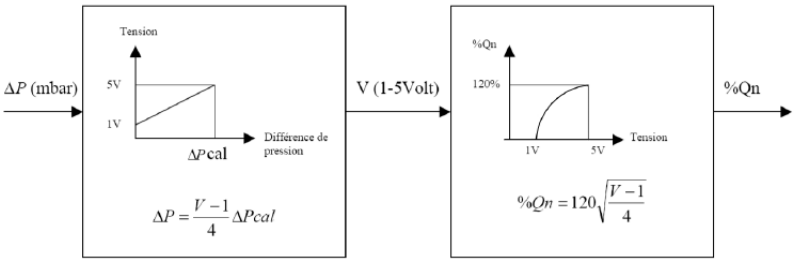


FIGURE 2.3 – Chaîne d'acquisition et de traitement de la mesure de débit primaire par ΔP-Coude

La valeur de différence de pression ΔP absolue en *mbar* se déduit donc du signal en $\%QN$ par la formule suivante :

$$\Delta P = \Delta P_{cal}(\frac{Q_n}{120})^2 \tag{2.4}$$

Les valeurs de recalage ΔP_{cal} de chaque capteur sont évaluées lors d'un étalonnage et enregistrées dans la base de données SYGMA.

Pour les tranches sur lesquelles nous avons pu travailler, peu de valeurs de ré-étalonnage sont enregistrées dans la base de données SYGMA : soit les recalages sont effectivement rares (aux arrêts de tranche décennaux par exemple), soit les valeurs de recalage (les ΔP_{cal}) ne sont pas systématiquement consignés dans la base de données SYGMA. L'incertitude correspondante pour l'évaluation de débit demeurait amplement suffisante pour la fonction originelle de ces capteurs, à savoir une fonction de surveillance pour la détection d'un éventuel sous-débit de -10%. Mais depuis que la méthode ΔP-Coude a été perfectionnée par EDF R&D pour obtenir une mesure absolue, plus fiable et de précision garantie, des directives ont été données aux exploitants pour remédier à ce problème en procédant plus régulièrement à des recalages soigneux et en enregistrant systématiquement les valeurs de recalages correspondantes.

L'enjeu principal de la méthode ΔP-Coude demeure la détermination du coefficient K. Historiquement, une méthode fut mise en place par Westinghouse utilisant la mesure de débit RCP114 du premier cycle de la centrale :

$$K = \frac{Q_{RCP114}}{D^{\frac{3}{2}} R_c^{\frac{1}{2}} \sqrt{\frac{\Delta P}{\rho}}} \tag{2.5}$$

en reprenant les notations précédentes. Cette méthode permet ainsi de déterminer en début de cycle une valeur de référence pour le coefficient K associé. Néanmoins, ce faisant, on lie la mesure ΔP-Coude à la mesure $RCP114$ et on l'affecte potentiellement des mêmes sources d'erreur, notamment l'erreur d'hétérogénéité de température de branche chaude. Cette méthode de Westinghouse suppose de disposer des mesures des premiers cycles, ce qui n'est pas systématiquement le cas.

EDF R&D (STEP avec l'appui de MFEE, développeur du code CFD 3D utilisé) a donc développé une méthode pour déterminer les valeurs de ces différents coefficients K à partir de mesures effectuées sur la boucle de débitmétrie EVEREST [1] de STEP et de résultats de simulations numériques CFD effectuées avec *Code_ Saturne*®. De premières études ont révélé une remarquable convergence de ces valeurs avec celles collectées sur les sites N4, ce qui a conduit à utiliser cette méthode (brevetée) peu onéreuse pour déterminer un coefficient K unique par palier.

Le coefficient K est donc déterminé par simulation numérique CFD avec *Code_ Saturne*® pour chaque palier. La démarche opératoire consiste à créer le maillage de la boîte à eau et de la branche en U (entre le GV et le GMPP et donc juste avant la branche froide) selon les plans de chaque palier. Ensuite, la simulation numérique est lancée en indiquant les

1. maquette à l'échelle 1/4

	K_a	K_b	K_c	K_d	R_c	D
Valeur	8.5934	8.0650	7.9594	8.1765	0.7874 m	1.454 m
Incertitude	2.897%	1.501%	1.948%	0.753%	7 mm	15 mm

TABLE 2.2 – Paramètres pour le ΔP-Coude

propriétés du fluide et le débit en entrée selon les conditions nominales de fonctionnement de la centrale. Les coefficients K (un coefficient pour chaque capteur) sont obtenus en utilisant la formule de l'équation 2.5 et en connaissant le débit. L'incertitude pesant sur ces coefficients est donnée notamment grâce à la comparaison avec les résultats [2] sur la boucle EVEREST.

Les valeurs des coefficients K (et leurs incertitudes) pour le cas d'un palier 1300 MW sont données dans le Tableau 2.2. On y donne également les valeurs des paramètres R_c et D.

Finalement, en utilisant ces capteurs pour obtenir une mesure du débit-boucle, on obtient des incertitudes entre ±3.6% et ±2.2%, ce qui conduit à une estimation du débit plus précise que le débit RCP114, qui plus est en temps réel, et non impactée par le phénomène d'hétérogénéité de température en branche chaude.

Les signaux délivrés par ces capteurs et enregistrés sous ORLI peuvent ainsi être transformés en mesures absolues du débit-boucle. Il est à noter que, d'une boucle à l'autre (et potentiellement d'une tranche à l'autre), la numérotation du capteur dans ORLI (01 ,02, 03, 17) ne correspond pas nécessairement au même emplacement géographique pour le capteur (A, B, C ou D). Le Tableau 2.3 donne un exemple des étiquettes ORLI des 16 capteurs ΔP-Coude et de leurs emplacements respectifs.

2.2.3 Températures Primaires

Températures de Branche Froide et de Branche Chaude

Chaque branche chaude et chaque branche froide est équipée d'un système de mesure de températures par bypass, qui est constitué de 3 écopes au milieu de la branche, disposées avec un angle de 120° entre elles, et chacune percée de 5 orifices orientés vers l'intérieur de la branche, par lesquels sont prélevés en continu des échantillons de fluide. Les distances de ces orifices aux parois internes sont respectivement de 37, 69, 101, 133 et 165 mm (voir Figure 2.5). Les écopes sont reliées dans une tuyauterie dite de bypass, où est réalisée la mesure de température. Le débit total des trois écopes est de 8 $l.s^{-1}$, ce qui représente à peine plus de 0.1% du débit volumique de la branche, et ré-injecté en branche chaude.

La température mesurée par le système d'écopes est donnée par la moyenne ainsi constituée au niveau du bypass.

2. Campagnes 2000,2010

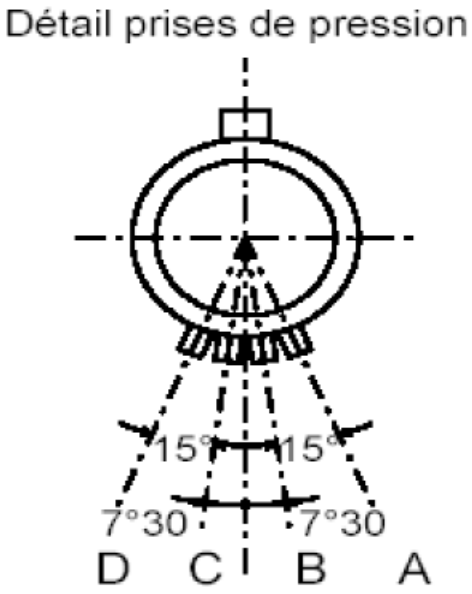


FIGURE 2.4 – Détail de l'emplacement des piquages pour capteurs ΔP-Coude

Etiquette	**Boucle**	**Emplacement**
RCP101MD-	1	C
RCP102MD-	1	B
RCP103MD-	1	D
RCP117MD-	1	A
RCP201MD-	2	C
RCP202MD-	2	B
RCP203MD-	2	D
RCP217MD-	2	A
RCP301MD-	3	C
RCP302MD-	3	B
RCP303MD-	3	A
RCP317MD-	3	D
RCP401MD-	4	C
RCP402MD-	4	B
RCP403MD-	4	D
RCP417MD-	4	A

TABLE 2.3 – Emplacements des différents capteurs ΔP-Coude

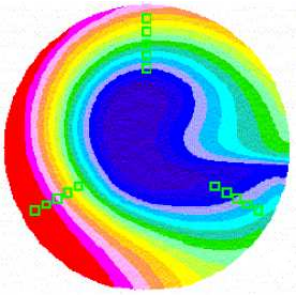


FIGURE 2.5 – Vue du système d'écopes et de l'hétérogénéité du champ de température en Branche Chaude (dans un REP 1300 MW)

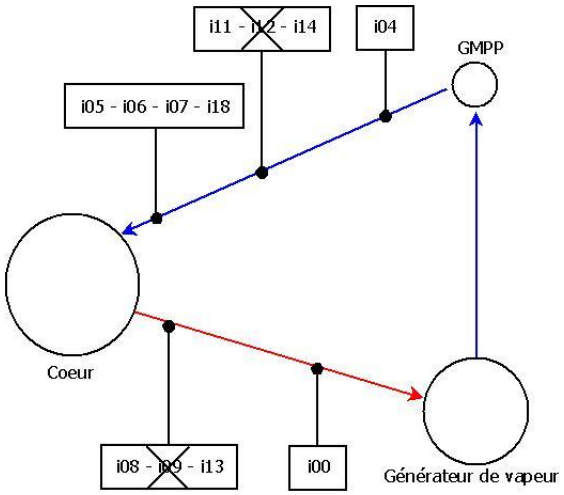


FIGURE 2.6 – Schéma représentant l'emplacement des différentes sondes RCP de mesures de température

Une des mesures de température (en BC et en BF) est faite par une sonde de réserve et n'est pas reliée à la chaîne d'acquisition KIT et à ORLI. Seules deux mesures de températures, placées dans le bypass, sont donc disponibles en temps réel. De plus, on dispose sur chaque boucle en branche froide de quatre mesures de températures gamme semi-rapide placées dans des doigts de gant. Une représentation des différents capteurs est donnée Figure 2.6.

Soit, au total, pour chaque boucle, deux mesures de TBC_i et six mesures de TBF_i, dont les étiquettes et les incertitudes sont données Table 2.4.

Ces sondes de températures peuvent être légèrement recalées si besoin lors de l'Arrêt à Chaud (voir plus loin)

Etiquette	Type	Incertitude
RCP105MT	Gamme semi-rapide BF	$\pm 0.8°C$
RCP106MT	Gamme semi-rapide BF	$\pm 0.8°C$
RCP107MT	Gamme semi-rapide BF	$\pm 0.8°C$
RCP118MT	Gamme semi-rapide BF	$\pm 0.8°C$
RCP111MT	Gamme étroite dans le bypass BF	$\pm 0.6°C$
RCP114MT	Gamme étroite dans le bypass BF	$\pm 0.8°C$
RCP108MT	Gamme étroite dans le bypass BC	$\pm 1.1°C$
RCP113MT	Gamme étroite dans le bypass BC	$\pm 1.1°C$

TABLE 2.4 – Capteurs de températures RCP pour la boucle 1

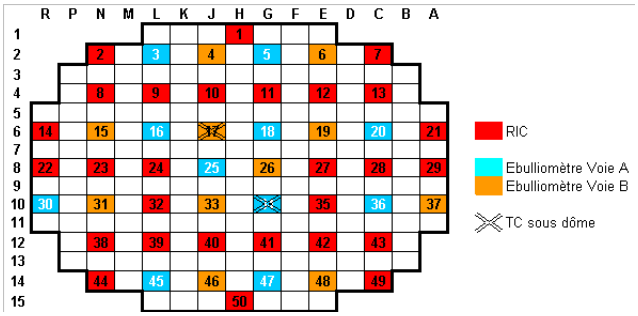


FIGURE 2.7 – Emplacements et fonctions des thermocouples au sein du cœur d'un palier 1300 MW

Températures In-Core

Plusieurs études ont été effectuées à propos de l'utilisation optimisée des thermocouples in-core, comme par exemple [Esc11], [Vai12], ou [Com13].

50 thermocouples sont implantés dans la cuve : 48 sont placés au niveau de l'interface cœur - plénum supérieur, et mesurent donc la température en haut de l'assemblage combustible sur lequel ils sont situés ; les deux derniers sont situés sous le dôme et ne sont donc pas utilisés (voir Figure 2.7).

Les thermocouples étaient peu utilisés en centrale, ne servant que de garde-fou en cas de défaillance des systèmes de mesures de l'activité neutronique RPN (voir plus loin). Pourtant, ils étaient utilisés dans les réacteurs UNGG pour piloter la centrale : cependant, les dérives observées dues à la forte irradiation à laquelle ils sont constamment soumis du fait de leur implantation ont conduit au quasi-abandon de cette instrumentation hors usage de sûreté. Un effort de recherche a donc été nécessaire pour s'assurer que grâce à des techniques innovantes de recalage et d'étalonnage, une utilisation de la mesure fournie par ces thermocouples via ORLI peut être fiable.

La mesure de température de l'eau en sortie de cœur est assurée par des thermocouples de type K. Les thermocouples fournissent une mesure de température grâce à l'effet Seebeck : le flux de chaleur dans le thermocouple crée un mouvement d'électrons et engendre donc une force électromotrice. La relation entre la force électromotrice fem et la température de l'eau mesurée T est assez complexe :

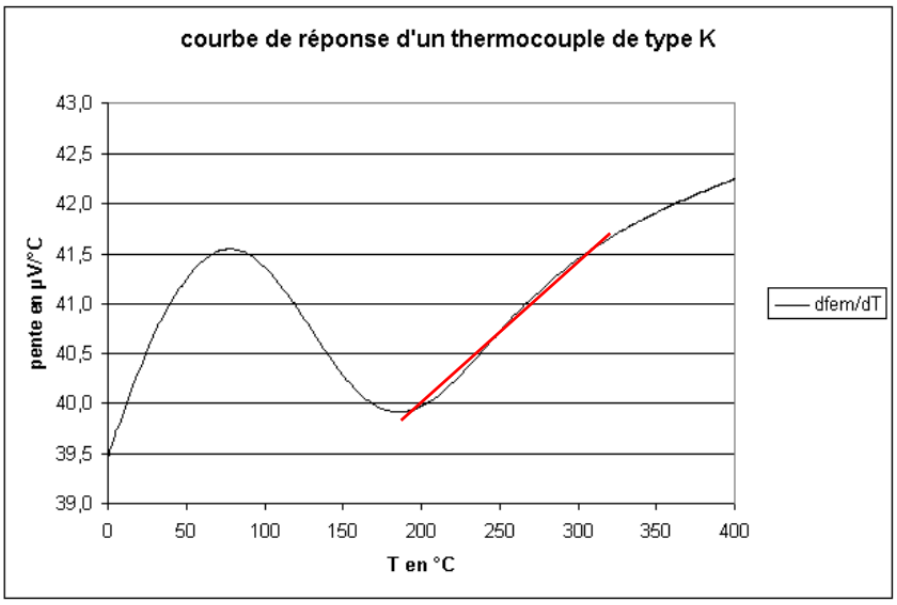


FIGURE 2.8 – Courbe de réponse d'un thermocouple de type K (approximation affine pour la plage de température $200°C - 350°C$

$$fem = \sum_{i=0}^{9} b_i T^i + c_0 e^{c_1(T-T_0)} \tag{2.6}$$

où les coefficients b_i,c_i et T_0 dépendent du type de thermocouple.

Les thermocouples sont disposés de manière à quadriller le cœur. Sur la Figure 2.7 on représente cette disposition, ainsi que leur répartition en trois catégories : voie A ou voie B, RIC ou ébulliomètre. Ces différences relèvent de la chaîne de mesure. Les 28 thermocouples RIC ont le même boîtier de soudure froide ainsi que la même chaîne de mesure. Leur incertitude est estimée à $\pm 2.42°C$. Les 22 thermocouples de type ébulliomètre sont séparés en deux voies A et B qui ont des chaînes de calcul différente, et leur incertitude est estimée à $\pm 2.44°C$. Chaque thermocouple ébulliomètre possède son propre boîtier de soudure froide : ainsi, en règle générale, les thermocouples de type RIC sont susceptibles de fournir des températures légèrement corrélées.

Le rôle des ébulliomètres est de déterminer la marge à la saturation ainsi que le taux de plein de la cuve. A contrario, les thermocouples de type RIC n'ont pas d'utilisation particulière sur le plan de la sûreté. En cas de défaillance d'un ébulliomètre, il peut ainsi arriver qu'on change la chaîne de mesure du thermocouple RIC le plus proche afin de lui faire jouer le rôle d'un ébulliomètre.

L'arrêt à chaud (AAC) est un palier idoine pour mettre en évidence le biais des thermocouples : en effet, durant cette phase du cycle, la puissance thermique est quasi-nulle et la différence entre les températures des branches froides et chaudes est de l'ordre du dixième de degré, de sorte que l'ensemble du fluide primaire peut être considéré comme faisant office de bain d'étalonnage à $297°C$ environ. Les essais réalisés en centrale sur les

thermocouples le sont lors de ce palier. En effet, la moyenne des thermocouples selon leur type est comparée à un seuil : $10.5°C$ pour les thermocouples de type RIC, $5.5°C$ pour les ébulliomètres. La correction n'est faite que s'il y a dépassement du seuil.

Le département STEP a mis au point [Com13] [Esc11] le procédé suivant pour le recalage à l'AAC :

1. Pour chaque thermocouple, on enregistre la mesure du thermocouple lors de l'arrêt à chaud (TC_{AAC}).

2. On relève la température moyenne du fluide lors de cet arrêt à chaud ($T_{moy,AAC}$), obtenue en effectuant la moyenne des mesures (gamme étroite et gamme semi-rapide) de branche chaude et de branche froide, et on compare la mesure du thermocouple pour en déduire le biais à soustraire.

3. En fonctionnement nominal, la mesure du thermocouple étant TC, on supprime le biais qui a été constaté lors de l'AAC ($TC_{AAC} - T_{moy,AAC}$).

4. On effectue ensuite une seconde correction, liée à la pente de la courbe de réponse du thermocouple, en supprimant le biais correspondant au changement de domaine de température ($k \times (TC - TC_{AAC})$).

soit en résumé :

$$TC_{corrige} = TC - (TC_{AAC} - T_{moy,AAC})k \times (TC - TC_{AAC}) \tag{2.7}$$

Le coefficient k de la deuxième correction appliquée peut doit être déterminé en utilisant les premiers paliers de montée en température du cycle, qui précèdent l'Arrêt à Chaud. Dans la Tranche $TR1$ sur lesquels cette méthode a été appliquée, ces paliers se situaient à $176°C$, $202°C$ et $297°C$ (AAC). Or dans la plage de température $[175°C, 330°C]$, la courbe de réponse des thermocouples de type K peut être approchée par une droite, sans dégradation gênante de la précision (voir Figure 2.8). En effectuant ainsi un recalage à chaque palier, et en les gardant en mémoire, il est ainsi possible de déterminer le coefficient de correction de pente k à apliquer (dans notre cas, 0.022).

En résumé, on suppose que le biais de chaque thermocouple dépend linéairement de la température, avec un taux d'accroissement identique pour l'ensemble des thermocouples ; on suppose que le fluide constitue un biais d'étalonnage homogène lors des différents paliers, et utilisant à chaque fois les sondes de température gamme large des différentes boucles pour évaluer la température de ce biais d'étalonnage, on calcule le biais constaté à chaque palier pour chaque thermocouple ; on en déduit la valeur du biais de pente et du biais d'arrêt à chaud pour chaque thermocouple, qu'on peut donc éliminer de la température mesurée en temps réel.

2.2.4 Puissance de la chaudière

Afin de conduire la tranche, il est indispensable de connaître avec exactitude la puissance du cœur. En temps réel, deux estimateurs sont disponibles en temps réel : la puissance

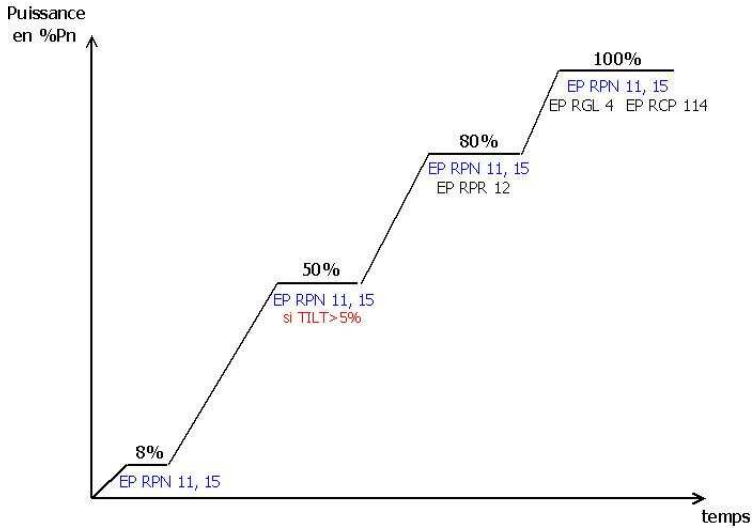


FIGURE 2.9 – Paliers de puissance suivis lors de la montée en charge du réacteur, en début de cycle

neutronique RPN, et le BILKIT.

La puissance neutronique RPN est mesurée en temps réel grâce à des capteurs neutroniques : des Chambres Externes de Puissance. Cette mesure est très sensible aux variations rapides de puissance : elle est donc adaptée à la protection du réacteur, et ce bien qu'elle puisse être très imprécise. Cette imprécision est due à la dérive des capteurs RPN qui sont, de facto, soumis en permanence à une irradiation intense.

Le calcul du BILKIT s'affranchit en partie de cette forte irradiation, car il consiste en le bilan de puissance thermique entre chaque branche chaude et chaque branche froide du circuit primaire : cela revient à effectuer la somme des bilans aux bornes de chaque générateur de vapeur, présenté Théorème 1. La puissance totale transférée via les générateurs de vapeurs est en effet celle de la chaudière, au terme de puissance corrigée des pompes primaires près. L'incertitude sur ce BILKIT est de $\pm 1.1\%$ à 95% de confiance : cette incertitude résulte en partie du phénomène d'hétérogénéité de température en branche chaude causé par les plans de chargements à faible fluence.

BIL100/BILORLI

L'essai périodique BIL100 est réalisé tous les mois afin de mesurer précisément la puissance de la tranche. Il s'agit d'un bilan thermique aux bornes du Générateur de Vapeur effectué côté secondaire (contrairement au BILKIT qui est réalisé côté primaire) lors d'une période stable à 100% de puissance du cœur, avec une exactitude plus importante encore que le BILKIT.

C'est le BIL100, via la mesure de débit RCP114, qui permet de calibrer le BILKIT en début de cycle : le BILKIT requiert en effet de connaître le débit-boucle, et la valeur utilisée dans le calcul du bilan enthalpique est celle du RCP114.

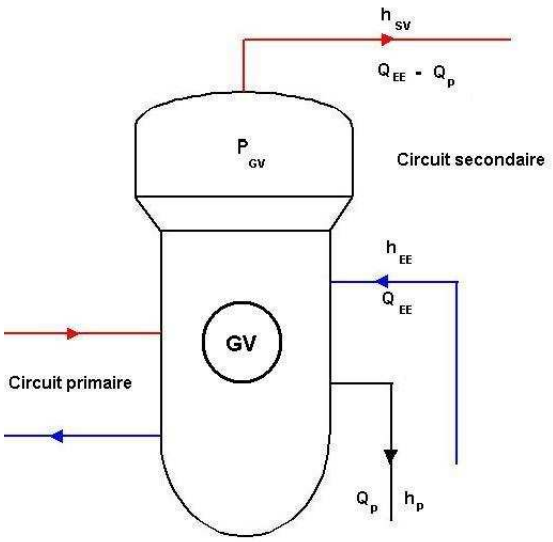


FIGURE 2.10 – Schéma de principe d'un BIL100

Les principales règles pour la mesure et le calcul du BIL100 sont les suivantes :

— être au-dessus de $87\%PN$;

— réaliser une acquisition au moins toutes les 5 secondes pendant au moins 20 minutes ;

— être en régime stable.

On réalise un bilan enthalpique sur chaque GV au secondaire pour obtenir la puissance thermique transféré du primaire au secondaire à travers le GV ($PTHGV$). Pour se ramener à la puissance thermique du cœur, la puissance des pompes ($PGMPP$) ainsi que la puissance des chaufferettes du pressuriseur ($PTHCAL$) sont soustraites à la puissance thermique des GV.

Les détails de la réalisation d'un tel BIL100 sont explicités dans la note [Pie09].

Théorème 2 (Bilan Enthalpique GV côté secondaire)**.** *Au niveau des Générateurs de Vapeur, la puissance thermique cédée par le fluide primaire est égale à la puissance thermique reçue par le fluide secondaire, soit :*

$$PTHGV = Q_{EE} \times (H_{SV} - H_{EE}) - Q_P \times (H_{SV} - H_P) \tag{2.8}$$

où :

— *$PTHGV$ est la puissance transférée depuis le primaire à travers le générateur de vapeur ;*

— *Q_{EE} est le débit massique d'eau alimentaire à l'entrée du GV ;*

— *Q_P est le débit massique des purges GV ;*

— *H_{SV} est l'enthalpie massique du mélange en sortie de dôme GV ;*

— *H_{EE} est l'enthalpie massique de l'eau alimentaire à l'entrée du GV ;*

— *H_P est l'enthalpie massique de l'eau à la purge des GV.*

Ce théorème peut être vu comme une définition de $PTHGV$.

En sortie du GV, on peut prendre en compte le taux de primage pour calculer H_{SV} lorsqu'il est disponible [3] : à défaut, on considère que le mélange eau-vapeur n'est en fait que de la vapeur, ce qui constitue une surestimation de la puissance. H_P est l'enthalpie de l'eau liquide saturée à la pression P_{GV} : le débit de purge Q_P est équilibré entre l'ensemble des GV. Enfin, H_{EE} est l'enthalpie de l'eau à pression P_{GV} et à la température T_{EE}.

La somme des puissances des pompes primaires et de la puissance des chaufferettes du pressuriseur est prise constante, à 18 MW environ pour un palier 1300 MW, et est répartie entre les boucles côté primaire.

Pour le BIL100, on utilise l'ensemble des mesures Y du circuit secondaire pour évaluer chacun des paramètres du bilan enthalpique côté secondaire. Les paramètres à surveiller le plus précisément possible sont le débit d'eau en entrée (débit ARE noté Q_{EE}), la pression sortie vapeur (pression VVP, notée P_{GV}, utilisée pour le calcul des enthalpies), et le débit purge.

Le débit d'eau alimentaire Q_{EE} est mesuré au moyen d'un organe déprimogène : un diaphragme à prises de pression différentielle. Cette mesure peut être détériorée en cas d'encrassement du diaphragme (des dispositions de surveillance et de mesure du biais du à cet encrassement réduisent ce risque).

Pour les purges des générateurs de vapeur, le débit total est déterminé par une mesure d'exploitation est en faisant l'hypothèse que le débit total est équiréparti entre GV, ce qui a pu être vérifié en effectuant une mesure locale du débit purge pour chaque file.

Pour la mesure de pression au niveau du dôme GV, on dispose de capteurs de pression en aval du dôme. Il est alors nécessaire de prendre en compte une perte de charge afin d'obtenir la pression en amont du dôme, où est réalisé le bilan de puissance :

$$P_{GV} = P_{GV,aval} + \Delta P_0 (\frac{Q_v}{Q_{v0}})^2 \tag{2.9}$$

où :

— $P_{GV,aval}$ est la pression mesurée par les capteurs (4 capteurs par boucle) ;

— Q_v est le débit vapeur, égal au débit ARE moins le débit purge ($Q_v = Q_{EE} - Q_P$) ;

3. Les valeurs « nominales » sont de $0,1\%$, ce qui conduit à un biais de $+0,08\%$ sur la puissance finale

— ΔP_0 et Q_{v0} sont des paramètres de calibrage (égaux respectivement à 0.9 bar et 531 $kg.s^{-1}$ dans le cas du palier P'4.

Les incertitudes à 95% de confiance sont connues pour chaque capteur et afin d'obtenir l'incertitude résultante sur le BIL100 de l'ensemble du réacteur, on applique classiquement la formule de propagation des incertitudes par cumul quadratique. On obtient finalement, pour le palier P'4, une incertitude de ± 16 $MWth$, soit $\pm 0.42\%$. L'incertitude sur la puissance d'une boucle est estimée à environ $\pm 0.8\%$.

Hors date d'Essai Périodique, il est possible d'utiliser les données M du système d'information ORLI pour acquérir en temps réel les paramètres de la formule du BIL100, obtenant ainsi un estimateur en temps réel de la puissance de la chaudière, complémentaire du BILKIT et équivalent du BIL100 en quasi-continu : le BILORLI. Les données M étant moins fiables que les données Y, il est nécessaire de les utiliser avec un esprit critique, et de les traiter au préalable avant de les exploiter.

Pour pouvoir réaliser le calcul du BILORLI, on doit se situer dans des conditions de stabilité pour la tranche proches de celles des essais BIL100, utiliser les capteurs disponibles pour les différents paramètres et effectuer quelques hypothèses simplificatrices par exemple sur les différents titres vapeur et sur l'emplacement des capteurs M pour le calcul de la perte de charge. En recalant les différents capteurs utilisés sur les valeurs du BIL100 en début de cycle, on peut obtenir un BILORLI d'une incertitude estimée à ± 34 $MWth$, soit $\pm 0.85\%$ (deux fois moins précis que le BIL100). L'incertitude sur la puissance d'une boucle est alors estimée à environ $\pm 1.5\%$.

Chambres à Fission Mobile

Les chambres à fission mobiles (CFM) sont des chambres d'ionisation, à dépôt d'uranium, qui sont ponctuellement introduites dans le cœur pour fournir une mesure du flux neutronique, selon une périodicité d'une carte de flux tous les 30 JEPP [4], soit à peu près au même rythme que les EP BIL100. Six Chambres à Fission Mobiles sont introduites simultanément dans le cœur au moment de la mesure, par le fond de la cuve. Les détecteurs sont fixés au bout d'une tige souple (téléflex) qui est introduite dans le doigt de gant par le dessous du cœur. Ils sont poussés jusqu'en haut du cœur puis redescendus par pas de 8 mm environ, fournissant une série de mesures du flux neutroniques (616 mesures par assemblage). Au total, 58 assemblages sont ainsi parcourus par les 6 détecteurs, avec un temps d'exploration de l'ordre de 10 minutes (voir Figure 2.11).

Les incertitudes de mesures de puissance sont évaluées à $\pm 1.9\%$ sur les assemblages instrumentés. Une interpolation via le code CABRI-CAMARET fournit des mesures de puissance pour les assemblages non instrumentés avec une incertitude évaluée à $\pm 3.2\%$. Le code CABRI-CAMARET est complexe et utilise des paramètres divers comme les déséquilibres de température constitués via les thermocouples RIC, la position des grappes de commande, les informations relatives au plan de chargement, etc. , ainsi que divers raisonnements semi-empiriques, comme l'utilisation de symétries ou de bilans de puissances.

Les cartes de flux reconstruites par le code CABRI-CAMARET à partir des mesures

4. Jours Equivalents Pleine Puissance

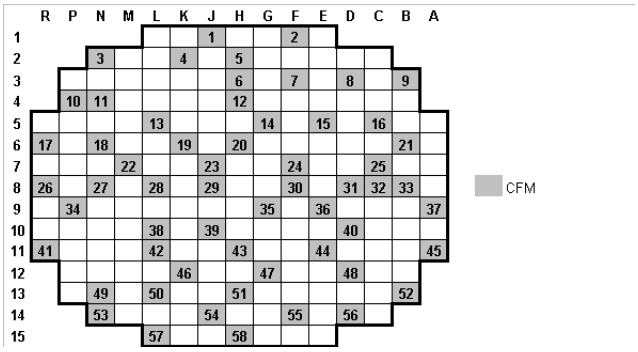


FIGURE 2.11 – Emplacements des Assemblages Combustibles parcourus par les CFM

CFM et RIC sont enregistrées dans des fichiers texte de type RES, lesquels sont stockés dans la base de données MUSICALE. On y trouve ainsi :

— une évaluation de la puissance thermique du réacteur $PTHREAC$ via un BIL KIT, avec les mesures correspondantes (débit RCP114, températures RCP) ;

— les mesures des thermocouples RIC disponibles (recalées) ;

— les valeurs des Puissances Locales Normalisées PLN, 248 valeurs pour chacun des 193 AC[5], telles que $PLN_{j,h} \times \frac{PTHREAC}{193 \times 248}$ soit la puissance thermique locale absolue en MW de la portion numéro h de l'AC numéro j ($PL_{j,h}$) ;

— les valeurs des Puissances Intégrées Par Assemblage Normalisées $PIPAN$, 1 valeur pour chacun des 193 AC, telle que $PIPAN_j \times \frac{PTHREAC}{193}$ soit la Puissance Intégrée Par Assemblage pour l'assemblage numéro j ($PIPA_j$) en MW.

Remarque 1.

On notera que :

— *la somme des PL le long de chaque AC vaut la PIPA dudit AC, et la somme des PIPA sur l'ensemble du cœur vaut exactement PTHREAC ;*

— *les moyennes des PLN et des PIPAN respectivement, sur l'ensemble du cœur, valentt exactement* 1 *;*

— *la moyenne des PLN le long de chaque AC vaut la PIPAN dudit AC.*

Chambres RPN

Le système élémentaire RPN (Réacteur Puissance Neutronique) assure la surveillance permanente de la puissance neutronique du réacteur, de son évolution et de sa distribution, à partir de mesures de flux neutronique effectuées par l'intermédiaire d'une série de détecteurs disposés à l'extérieur de la cuve. Le système RPN fournit au SPIN (Système de Protection Intégrée Numérique) les signaux qui serviront à élaborer les arrêt d'urgence du

5. Pour un palier 1300 MW

réacteur par flux nucléaire élevé et par variation rapide de flux.

Il existe trois types de chaînes d'instrumentation nucléaire fournissant trois niveaux distincts de protection :

— La gamme de niveau source (CNS) comporte quatre chaînes identiques et indépendantes, utilisables à l'arrêt et au démarrage, pour une gamme de puissance de $10^{-9}\%PN$ à $10^{-3}\%PN$ environ. Le capteur est un compteur proportionnel à dépôt de bore.

— La gamme de niveau intermédiaire (CNI) comporte 4 chaînes identiques et indépendantes pour des puissances entre $10^{-6}\%PN$ et $100\%PN$. Le capteur est une chambre d'ionisation à dépôt de bore, compensée aux rayons gamma.

— La gamme de niveau de puissance (CNP) comporte 4 chaînes identiques et indépendantes, pour des puissances entre $10^{-1}\%PN$ et $120\%PN$. Le capteur est une chambre d'ionisation à dépôt de bore non compensée aux rayons gamma. Ce capteur est composé de 6 sections sensibles (3 sections hautes et 3 sections basses).

Les quatre chambres CNP sont disposées autour du cœur (voir Figure 2.11) et mesurent la puissance des assemblages périphériques. On définit ainsi une puissance par quart de cœur $PRPN_i$ correspondant à chaque chambre ainsi qu'une puissance moyenne du cœur égale à leur moyenne.

Les chambres RPN sont des outils indispensables pour la sûreté de la tranche, car ils permettent de suivre en temps réel et de façon directe l'activité neutronique. Néanmoins, dans le cadre de notre étude, ils sont difficilement exploitables. Non seulement leur incertitude est grande au regard de la précision que nous souhaitons atteindre pour l'estimation du point de fonctionnement primaire, mais il resterait de plus à estimer précisément le lien entre l'estimation de la puissance mesurée par chaque carte de cœur et la carte de flux reconstituée par le code CABRI-CAMARET via les mesures par CFM. Par conséquent, pour le cadre de cette étude, nous n'avons pas jugé nécessaire d'inclure les mesures RPN dans notre système d'observation : ceci pourrait faire l'objet d'une étude ultérieure.

2.3 Estimateurs

Le Théorème 1 permet de lier fortement les trois paramètres clés que sont la puissance de la chaudière (via $PTHGV$), le débit primaire (via les débits-boucles Q) et la température de branche chaude TBC : en en mesurant deux, on en déduit le troisième, et c'est ainsi qu'est conçu l'estimateur RCP114 pour le débit. Un calcul dit « de débit enthalpique » reproduisant en tous points le principe de l'estimateur RCP114 en utilisant non plus des BIL100 mais des BILORLI, et les sondes RCP108MT-, RCP113MT- pour les températures de branche chaude, permet d'obtenir une estimation du débit-boucle en temps réel : mais, comme nous l'avons dit, cet estimateur est potentiellement impacté par le phénomène d'hétérogénéité de température en Branche Chaude. Un enjeu important pour EDF R&D STEP a donc été, dans le cadre du projet PAPRICA, de concevoir des estimateurs de TBC qui ne soient pas impactés par ce phénomène. Les stages [Com13],[Esc11], ou [Vai12] ont ainsi consisté en la conception de nouveaux estimateurs pour TBC, afin de disposer

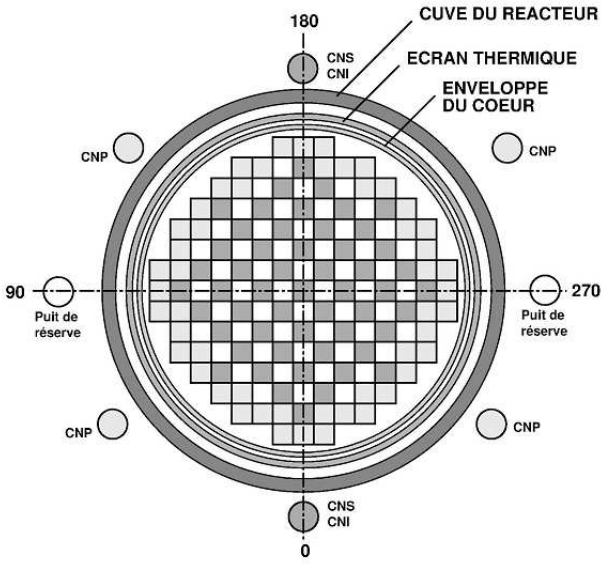


FIGURE 2.12 – Position des chambres RPN autour du cœur

d'une estimation plus précise à la fois du paramètre TBC en lui-même, mais également du phénomène d'hétérogénéité qui biaise sa mesure directe.

2.3.1 TBC Enthalpique

La mise en œuvre de la mesure de débit par ΔP-Coude permet de transformer le bilan enthalpique aux bornes des GV côté primaire (eq. 2.1) en un estimateur pour la température de branche chaude :

$$TBC_i = h^{-1}(h(TBF_i) + \frac{PTHGV_i - \frac{PGMPP}{4}}{Q_i}) \tag{2.10}$$

où h^{-1} est la bijection réciproque de h, c'est-à-dire la fonction telle que $h^{-1}(H)$ est la température de l'eau qui possède une enthalpie massique H à la pression 155 bar.

Ce nouvel estimateur pour TBC possède une incertitude du même ordre de grandeur que celle des mesures directes par les sondes RCP ($\pm 1.1°C$), et en est de plus indépendante, et non impactée par le phénomène d'hétérogénéité en Branche Chaude.

En fait, l'équation 2.10 pourrait même être considérée comme une définition exacte pratique de TBC, plus qu'une simple mesure indirecte : en effet, du fait de l'hétérogénéité du champ de température de Branche Chaude, définir formellement ce que doit être le « vrai » paramètre TBC est délicat :

$$TBC = \frac{1}{V} \int \int \int_V T dV \tag{2.11}$$

où V est l'ensemble du volume de la Branche Chaude,

$$TBC = \frac{1}{S} \int \int_S T d\sigma \tag{2.12}$$

ou encore

$$TBC = \frac{1}{S} \int \int_S \frac{q}{Q} T d\sigma \tag{2.13}$$

où q est la valeur absolue du débit local, sont autant de possibilités pour définir théoriquement TBC à l'aide de formules intégrales, mais la formule de l'équation 2.10 est nettement plus immédiate.

2.3.2 Cartographie des $THAC$

Les cartographies de températures en Branche Chaude sont réalisées via un calcul CFD 3D dans le plénum supérieur, l'un des principaux paramètres d'entrée du calcul étant ainsi la cartographie (2D) de température à l'interface cœur / plénum supérieur, que l'on peut choisir de modéliser par la connaissance de la température en haut de chacun des Assemblages Combustibles, soit, pour le cas d'un palier 1300 MW, la connaissance des 193 $THAC$.

Or seulement 48 mesures de différents $THAC_j$ sont disponibles via les Thermocouples RIC et les ébulliomètres, et extrapoler ces mesures pour reconstruire, par krigeage par exemple, le reste de la carte de température semble délicat.

Une alternative a été d'utiliser les cartes de flux réalisées par le code CABRI-CAMARET à partir des mesures CFM pour réaliser une série de bilans enthalpiques au sein du cœur, et déduire ainsi, à partir des puissances locales thermiques (PL) l'évolution de l'enthalpie massique et donc de la température le long de chacun des assemblages combustibles.

Les hypothèses à la base de ce raisonnement sont les suivantes :

— régime stationnaire (débits, températures et puissances constantes dans le temps) ;

— débits transverses négligés (chaque AC étant donc assimilable à un canal chaud) ;
— équirépartition du débit entre les différents AC ;

— température à la base des AC égale à la moyenne des températures de branche froide (TBF_{MOY}).

Si l'on effectue de plus l'hypothèse que la fonction h peut-être linéarisée :

$$h(T_1) - h(T_2) \approx C_p(T_{MOY})(T_1 - T_2) \tag{2.14}$$

alors on obtient aisément une relation de proportionnalité formelle entre le gradient de température haut-bas AC (soit $THAC_j - TBF_{MOY}$) et la puissance intégrée le long de l'assemblage (soit $PIPA_j$) :

$$THAC_j - TBF_{MOY} = \frac{1}{q_m \times C_p(T_{MOY})} PIPA_j \tag{2.15}$$

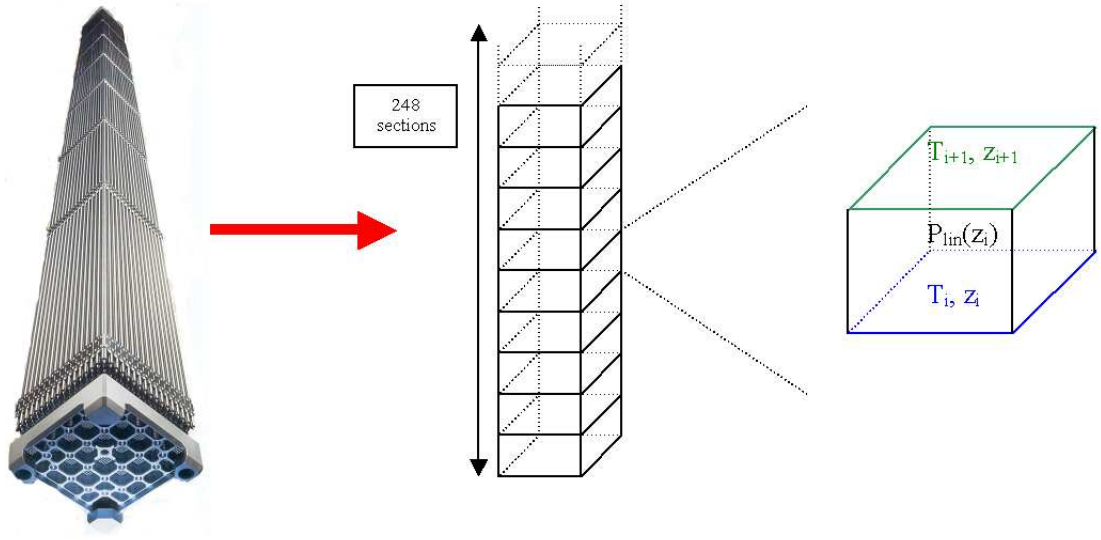


FIGURE 2.13 – Méthode dite des « bilans enthalpiques »

une première application numérique dans le cas d'un 1300 MW (en supposant que $q_m = \frac{\Sigma Q_i}{193}$, et que $T_{MOY} = 305.0\ °C$), donne pour coefficient de proportionnalité théorique $1.724\ °C.MW^{-1}$.

Néanmoins, la plupart des hypothèses nécessaires pour obtenir sont résultats ont peu de chances d'être vérifiées en pratique :

— sur la plage de températures concernée, h n'est pas linéaire ;

— les débits ne sont pas équirépartis entre les différents AC ;

— la totalité du débit-cuve ne passe pas par les AC, car il existe un bypass au sein du cœur qui est utilisé notamment pour refroidir le dôme ;

— il existe nécessairement des débits transverses au sein du cœur : ils sont même favorisés par la présence d'ailettes, en vue de mieux brasser le fluide réfrigérant.

La répartition qualitative des débits dans le cœur est donnée par la Figure 2.14, obtenue via un couplage THYC/COCCINELLE. On observe d'après les résultats de cette simulation qu'en bas des assemblages combustibles, le débit massique est plus important au centre qu'aux extrémités, puis qu'il s'harmonise peu à peu pour devenir presque homogène à mi-hauteur et qu'enfin il devient plus important aux bords lorsqu'on s'approche du haut des assemblages (le fluide est « poussé » par le centre en bas, et aspiré par les boucles en haut).

Ces tendances sont confirmées par les résultats de tests effectuée sur maquette.

Des variantes ont été testées pour améliorer la méthode des bilans enthalpiques en prenant en compte au moins dans une certaines mesures les débits transverses, mais ceux-ci sont trop mal connus pour pouvoir obtenir un résultat théorique fiable simplement : la complexité du problème est telle qu'on serait rapidement ramené à devoir reconstruire

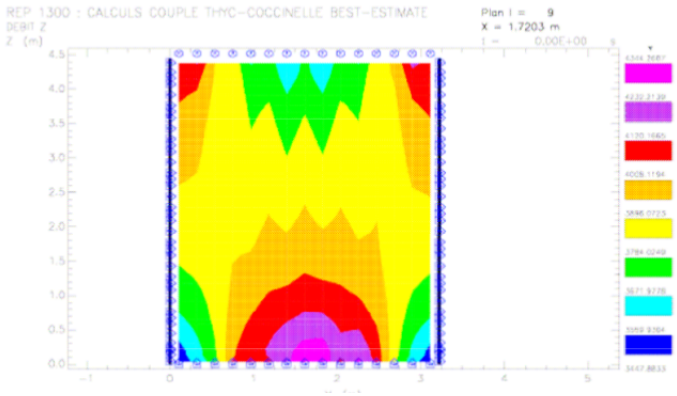


FIGURE 2.14 – Coupe des flux de masse Q_z à travers le cœur obtenue via THYX/COCCINELLE

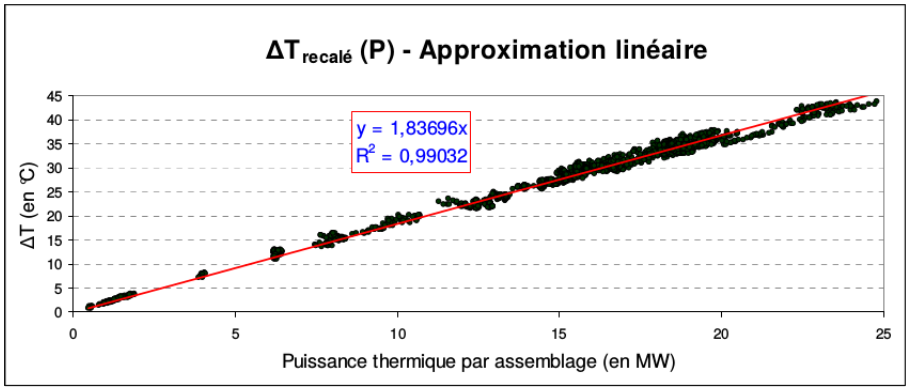


FIGURE 2.15 – Relation de proportionnalité constatée entre les gradients de température haut-bas AC et les $PIPA$s, pour les AC équipés de thermocouples, pour le cycle 16 de la tranche AA

un code de calcul semblable à THYC. Par conséquent, au lieu de chercher à obtenir une formule théorique reliant le gradient de température le long des AC à la puissance linéique des AC, on a chercher à établir une formule semblable à la loi de proportionnalité de l'équation 2.15 :

$$THAC_j = TBF_{MOY} + a \times PIPA_j \tag{2.16}$$

avec a un coefficient de proportionnalité à déterminer de manière empirique.

Pour déterminer le coefficient a, on utilise les thermocouples disponibles (soigneusement recalés) en confrontant, le long des assemblages combustibles équipés et aux dates des différentes cartes de flux, les valeurs des gradients de températures effectivement mesurés aux $PIPA$s correspondantes (voir Figure 2.15).

On constate effectivement une loi de proportionnalité respectée très fidèlement, avec un coefficient de proportionnalité a du même ordre de grandeur que le coefficient $\frac{1}{qm \times Cp}$ que nous avions estimé précédemment. La valeur de ce coefficient de proportionnalité a varie peu selon le cycle considéré, ou même selon le palier 1300 MW étudié.

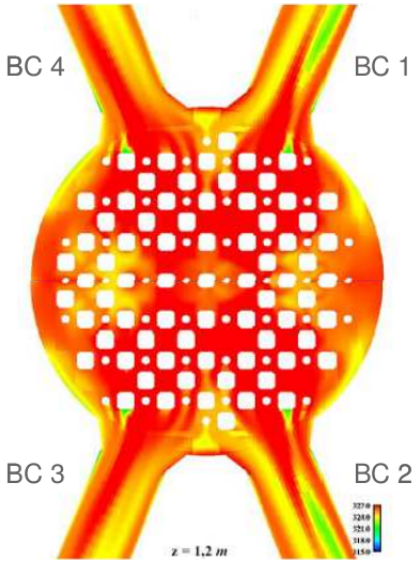


FIGURE 2.16 – Emplacements des BC (cas d'un 1300 MW) : vue du dessus du plénum supérieur, tel que simulé par CODE SATURNE

Ainsi, à partir d'une carte de flux et plus particulièrement de l'ensemble des Puissances Intégrées par Assemblages, la loi de proportionnalité de l'équation 2.16 nous permet d'obtenir une estimation de la température en haut de l'assemblage considéré. L'incertitude associée à cette estimation a été évaluée à $\pm 1.9°C$.

2.3.3 Moyennes par quarts de cœur

Transporter la cartographie 2D des $THAC$ à travers le Plénum Supérieur pour obtenir une évaluation des TBC est un problème sophistiqué de thermohydraulique : le logiciel de CFD 3D *Code_Saturne*® effectue ce calcul, mais le coût est lourd et des paramètres supplémentaires (comme par exemple la cartographie des débits en haut du cœur) sont nécessaires.

La position des branches chaudes autour du haut du cœur est représentée Figure 2.16.

En pratique, on effectue souvent une moyenne par quarts de cœur. Le principe est d'attribuer à chaque boucle un cadran du cœur (celui qui lui est géographiquement le plus proche) comme sur la Figure 2.17, et d'évaluer la température moyenne de ce quart de cœur pour en déduire une estimation de TBC. Dans le choix de segmentation que nous avons adopté, chaque quart de cœur comprend 56 AC (les quarts se recoupent donc légèrement deux à deux).

On note THC_i la température moyenne du quart de cœur (en haut des assemblages combustibles) associé à la boucle numéro i. On a donc :

$$THC_i = \frac{1}{56} \sum_{j \in C_i} THAC_j \tag{2.17}$$

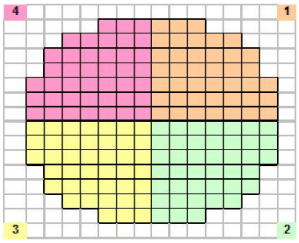


FIGURE 2.17 – Segmentation du cœur en quatre quarts

où C_i est le quart de cœur associé à la boucle i.

On constate en pratique que la température THC_i est systématiquement plus élevée que la température TBC_i telle qu'estimée par les mesures RCP ou par la méthode de température enthalpique. L'explication est double.

— D'une part, la totalité du fluide ne sert pas à évacuer la chaleur dégagée par les AC : une part du débit ne traverse pas le cœur, et est donc plus froide que le fluide en sortie de cœur, ce qui contribue à refroidir légèrement la branche chaude.

— D'autre part, le débit est plus important au bord du cœur, où les AC sont généralement de puissance moindres et les $THAC$ par conséquent plus faibles. L'équation 2.17, qui ne prend pas en compte cet effet, induit donc une surévaluation de la température du quart de cœur.

Par conséquent, il est nécessaire d'introduire un coefficient ϵ qui effectue une pondération entre la température calculée par moyenne simple de quarts de cœur via l'équation 2.17 et la température moyenne de branche froide :

$$TBC_i = \epsilon TBF_{MOY} + (1-\epsilon)THC_i \tag{2.18}$$

soit finalement :

$$TBC_i = \epsilon TBF_{MOY} + \frac{1-\epsilon}{56}\sum_{j\in C_i} THAC_j \tag{2.19}$$

Le coefficient ϵ a été initialement pris égal à la valeur de conception du débit de bypass au sein du cœur, qui est évalué à 7%, et a par conséquent été appelé dans plusieurs travaux « coefficient de bypass ». Mais nous remettons en cause dans cette étude cette appellation et cette valeur par défaut, comme nous l'expliquerons plus loin (voir la sous-section 2.4.2.

Remarque 2.
Hors dates des cartes de flux, la température moyenne du quart de cœur associée à la boucle i (le terme THC_i) ne peut pas être estimé à l'aide de la loi PIPA. On peut, à défaut, effectuer une moyenne de la quinzaine de mesures de températures via thermocouples disponibles dans le quart de cœur. Cependant, du fait de la forte hétérogénéité du champ des $THAC$, il n'est pas certain que la bonne répartition des thermocouples en question suffise à garantir que la moyenne de cette quinzaine de mesures soit effectivement proche de la moyenne de la totalité des 56 $THAC$ du quart de cœur considéré.

Cette méthode aboutit néanmoins à un autre estimateur pour la température de branche chaude, qu'on peut qualifier d'estimateur RIC, dont l'incertitude est du même ordre que les autres estimateurs de températures de branche chaude (entre $\pm 1.0°C$ et $\pm 1.5°C$).

2.4 Modèles

Nous avons jusqu'à présent décrit les mesures présentes in-situ, et les différents estimateurs utilisés couramment par les exploitants ou développés notamment par le département STEP pour évaluer certains paramètres primaires à partir de certaines mesures et de lois semi-empiriques. A chaque fois, l'objectif est de se concentrer sur l'estimation d'un paramètre primaire en particulier (Q dans le cas du RCP114 ou du ΔP-Coude, $PTHGV$ dans le cas du BILKIT ou du BIL100/BILORLI, $THAC$ dans le cas de la loi PIPA, TBC dans le cas de la température enthalpique ou de la moyenne par quarts de cœur (que ce soit en utilisant les $THAC$ obtenus par loi PIPA, ou les quelques $THAC$ mesurés par les thermocouples).

L'objectif de cette section est de faire un bilan des modélisations existantes pour le circuit primaire, et de construire un modèle physique simplifié qui fasse une synthèse des estimateurs développés dans le cadre de PAPRICA notamment, qui puisse être exploité dans un cadre d'Assimilation de Données.

2.4.1 Codes industriels

Nous avons évoqué l'existence de plusieurs codes de calculs utilisés dans différentes simulations pour le calcul précis des champs 3D du point de fonctionnement primaire : leur agencement est résumé Figure 2.18.

Nous en proposons une présentation succincte.

COCCINELLE

COCCINELLE (acronyme entomologiste signifiant Code de Cinétique Neutronique à Eau Légère) constitue le code de cœur 3D de référence d'EDF. Il est écrit en Fortran et sa première version date du milieu des années 80. EDF développe actuellement un nouveau code destiné à lui succéder.

L'objet de COCCINELLE est le calcul de la distribution volumique de puissance dans un cœur 2D ou 3D en statique ou en cinétique. Pour ce faire, COCCINELLE intègre :

— un module de neutronique résolvant en statique ou cinétique les équations de la diffusion neutronique à 2 groupes d'énergie ;

— un module de thermohydraulique (modèle 1D axial) et de thermique du combustible ;

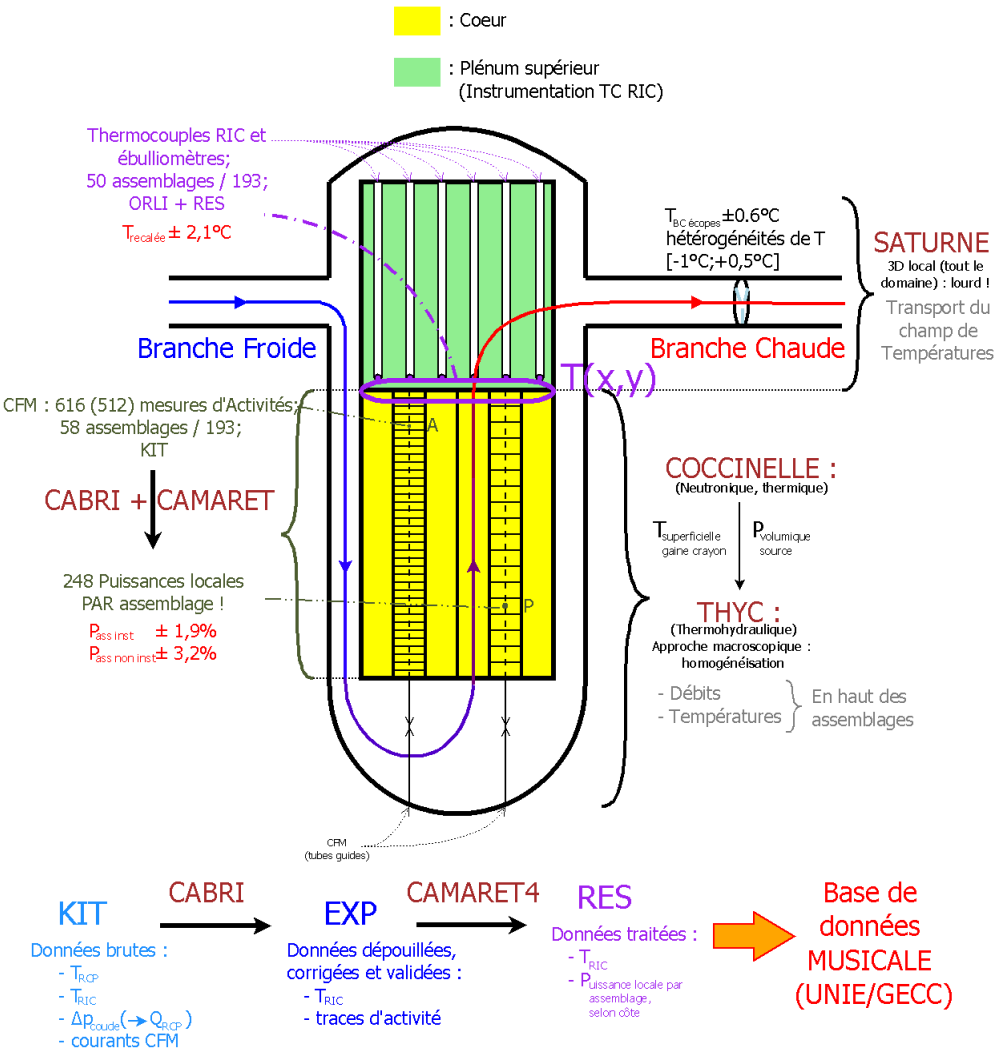


FIGURE 2.18 – Agencement des codes de calcul au sein du cœur

— un module de contre-réactions basé sur une analyse physique du couplage de la neutronique à la thermique du combustible et la thermohydraulique du modérateur ;

— un module d'évolution isotopique du combustible pour rendre compte de l'épuisement du combustible en matières fissiles.

Il existe, dans COCCINELLE, deux niveaux de description de la géométrie et appelés « modes » : le mode crayon par crayon consiste à individualiser le traitement de chaque crayon composant les assemblages du cœur par opposition au mode assemblage où assemblage est vu comme une pâte homogène. Le mode assemblage nécessite un post-traitement sur la puissance volumique calculée afin de rendre compte des hétérogénéités au sein de l'assemblage.

COCCINELLE est au centre de nombreuses applications, comme par exemple la recherche des plans de chargement (c'est-à-dire la recherche d'un positionnement efficace des assemblages dans le cœur en fonction de leur degré d'irradiation) ou les études d'accident.

THYC

THYC [TDR94][ACNN89] (autre acronyme entomologiste signifiant *Therma Hydraulic Code*) est un code généraliste qui a été développé par EDF en 1986 pour étudier les écoulements monophasiques et diphasiques tridimensionnels à l'intérieur des faisceaux de tubes (cœurs de réacteurs, générateurs de vapeur, condenseurs, échangeurs de chaleur...).

Deux types de constituants sont pris en compte dans le domaine 3D : le fluide et les solides (approche du milieux poreux). Le modèle THYC est obtenu en moyennant en temps et en espace les équations instantanées de masse, de quantité de mouvement et d'énergie de chaque phase sur un volume de contrôle incluant du fluide et des solides. Il existe plusieurs variantes de THYC : la version THYC-ECHANGEUR, par exemple, résout de trois à cinq équations du fluide extérieur aux tubes, plus l'équation d'énergie du fluide interne aux tubes. Le code est ainsi capable de modéliser tous les types d'échangeurs de chaleur, de l'échangeur monophasique aux composants impliquant l'ébullition ou la condensation.

De ce fait, THYC est extrêmement utilisé pour les études relatives au circuit secondaire : mais, en prenant les puissances thermiques comme paramètre d'entrée, il peut également être utilisé pour évaluer l'écoulement au sein du cœur.

CODE SATURNE

Code_Saturne® est un logiciel libre de simulation informatique en mécanique des fluides et thermohydraulique. Développé depuis 1997 par EDF R&D, il a été placé sous la GNU GPL en mars 2007.

Basé sur une approche volumes finis co-localisés qui accepte des maillages de tout type (structuré, structuré par bloc, non structuré, hybride, conforme, non conforme...) et contenant tout type d'élément (tétraèdre, hexaèdre, pyramide, polyèdre quelconque...), *Code_Saturne*® permet de modéliser les écoulements incompressibles ou dilatables, avec ou sans turbulence ou transfert de chaleur. Des modules dédiés sont disponibles pour des

physiques particulières comme le transfert de chaleur par rayonnement, la combustion (gaz, charbon pulvérisé, fioul lourd), la magnéto hydrodynamique, les écoulements compressibles, les écoulements polyphasiques (approche Euler/Lagrange avec couplage inverse), ou bien des extensions à des applications spécifiques (par exemple *Mercure_ Saturne* pour les écoulements atmosphériques.

Code_ Saturne® peut de plus être couplé à d'autres codes, de thermique solide ou de mécanique des structures par exemples.

CABRI et CAMARET

Les codes CABRI et CAMARET [Bod04] permettent d'exploiter les mesures in-core réalisées lors des essais périodiques grâce aux CFM et aux thermocouples RIC. Le code CAMARET est destiné au dépouillement des mesures expérimentales issues de l'instrumentation interne des cœurs après un traitement effectué sur ces mesures par le code CABRI.

Les valeurs expérimentales obtenues lors des essais périodiques, notamment celles provenant de l'instrumentation interne RIC, sont transcrites sur un support magnétique et peuvent être traitées localement sur station de travail, ou être envoyées directement sur les calculateurs centraux du centre de Pacy sur Eure. Un premier traitement est effectué par CABRI, puis elles sont utilisées par le code CAMARET.

Le fonctionnement du cœur CAMARET peut se résumer ainsi :

1. prétraitements (recalage, calcul de bruit de fond, calcul de dérive de puissance, troncature de haut de cœur, adaptation des assemblages MOX, suppression des points bas du cœur, calcul des coefficents de calibration) ;
2. affectation des formes d'activités ;
3. lissage ;
4. calcul des puissances thermiques ;
5. calcul des critères de conception et de sûreté ;
6. exportation des résultats.

L'algorithme suivi par CAMARET utilise plusieurs raisonnements de symétries et de proximités ainsi que des paramètres expérimentaux comme la position des grappes de commandes, les coefficients de calibration de chaque détecteur lors des passes de calibration, etc.

Modelica

Le langage de modélisation orienté objet Modelica est destiné à permettre une modélisation pratique de systèmes complexes, par exemple des systèmes comportant des com-

posantes mécaniques, électriques, hydraulique, thermique...Il décrit un système sous la forme d'un ensemble d'équations, et associe un simulateur qui a pour tâche de résoudre ce système d'équations à chaque pas temporel.

Le langage, les bibliothèques et certains outils de simulation de Modelica sont libres, prêts à l'emploi et issus d'une demande industrielle. Le développement et la promotion de Modelica est organisée par l'association à but non lucratif *Modelica Association*.

EDF R&D utilise Modelica notamment via le logiciel Dymola, logiciel de modélisation et de systèmes complexes développé par la société Dassault Systèmes AB [6] : basé sur Modelica, ce logiciel offre aux utilisateurs le choix entre créer leurs propres bibliothèques de composants, ou adapter des bibliothèques prêtes à l'emploi pour répondre à leurs besoins. Une fois un modèle créé par assemblage de composants, Dymola permet ensuite de compiler le code, de simuler le modèle et de visualiser ou exporter les résultats.

EDF R&D a également développé la bibliothèque ThermoSysPro (sous l'environnement Dymola), qui vise à modéliser les systèmes de production toutes énergies confondues (nucléaire, thermique à flamme, cycles combinés, etc. ...) et contient aussi bien les blocs fonctionnels de contrôle-commande que les différents constituants du process physique.

Ces outils sont notamment utilisés par et pour le projet MODRIO (Model Driven Physical Systems Operation), qui cherche à répondre à plusieurs verrous technologiques de la modélisation 0D/1D : modéliser les propriétés d'un système, recaler les modèles sur les états réels des tranches, simuler les systèmes multi-modes, etc. Le projet MODRIO rassemble un grand nombre de partenaires européens et est dirigé par Daniel BOUSKELA, d'EDF R&D, département STEP.

2.4.2 Modèle Semi-Empirique

Structure

Les différents codes de calculs industriels que nous venons de décrire permettent tous de modéliser le comportement de telle ou telle partie du circuit primaire, voire de sa totalité. Néanmoins, notre objectif pour la présente étude est de disposer d'un modèle qui puisse être utilisé dans un cadre d'Assimilation de Données, pour une première évaluation du potentiel de l'Assimilation de Données pour l'amélioration de la connaissance du point de fonctionnement primaire. Il faut donc, au moins dans un premier temps, disposer d'un modèle au coût de computation faible et qui puisse être facilement utilisable en interaction avec les programmes d'assimilation de données, tout en demeurant suffisamment représentatif de la réalité de la physique du circuit primaire.

Idéalement, on souhaiterait pouvoir obtenir un modèle pour l'ensemble de la région 3D du circuit primaire en couplant les différents codes de calcul de la façon suivante :

1. Calcul le la carte de puissance neutronique / thermique au sein du cœur, via les codes CABRI+CAMARET ou COCCINELLE ;

6. filiale Dassault Systèmes depuis 2006 à Lund en Suède

2. Calcul de la thermohydraulique au sein du cœur, puis du Plénum Supérieur, via les codes THYC ou Code_Saturne ;

3. Calcul des échanges de chaleur au niveau des GV, via THYC.

Outre les lourds coûts de calcul que cela implique, il demeure également l'obstacle purement informatique de réussir à faire interagir les différents codes entre eux. Enfin, certains de ces codes sont protégés par un secret industriel et par conséquent difficiles d'accès.

Par conséquent, le niveau de modélisation qui nous a semblé le plus adapté pour ce travail de thèse est celui de la modélisation 0D/1D. Le langage Modelica et la bibliothèque ThermoSysPro de Dymola aurait ainsi pu être utilisé : néanmoins, l'utilisation de Modelica / Dymola dans un cadre d'Assimilation de Données n'est pas immédiate (des études sont actuellement en cours pour rendre cela possible).

Pour toutes ces raisons, il a été décidé, dans un premier temps du moins, d'utiliser un modèle 0D/1D simplifié, construit à partir des différents estimateurs conçus et éprouvés dans le cadre du projet PAPRICA et décrits plus hauts.

Ce modèle est conçu en trois grandes étapes, suivies d'une quatrième étape qui consiste en l'identification du paramètre ϵ de la formule 2.19. Il a été l'un des objets des publications [MFG14], pour les grandes lignes du modèles, et [MFG15], pour l'équation de bouclage et l'identification par recherche dichotomique du coefficient dit « de bypass » intervenant dans la moyenne par quarts de cœur.

Génération de chaleur

Principe

La première étape du modèle vise à représenter l'élévation de température au sein du cœur, en utilisant la « loi PIPA » résumée par l'équation 2.16 : en prenant pour paramètre d'entrée la cartographie des Puissances Intégrées par Assemblages ($\{PIPA_j\}$) et la température moyenne de Branche Froide TBF_{MOY}, on peut reconstituer la carte 2D des températures en haut de l'assemblage combustible ($\{THAC_j\}$). Soit, pour mémoire :

$$\forall j THAC_j = TBF_{MOY} + a \times PIPA_j \tag{2.20}$$

Forme Vectorielle

Cette formule peut s'écrire sous forme vectorielle. Si l'on nomme :

— $THAC$ le vecteur colonne 193×1 des températures en haut des AC ;

— TBF le vecteur colonne des 4 températures de branche froide ;

— PL le vecteur colonne 47864×1 des puissances locales : $47864 = 193 \times 248$ est le nombre total de tronçons des AC : on range ces puissances locales dans l'ordre $PL[248 \times j + h] = PL_{j,h}$ la puissance du tronçon numéro h (en partant du bas) de

l'AC j (toujours en classant les AC dans l'ordre standard $L1 \to E15$) ;

alors, en posant $MOY_{193,4}$ la matrice de taille 193×4 dont chaque coefficient est égal à $1/4$, et F la matrice de taille 193×47864 – aisément constructible par blocs – qui effectue la sommation, le long de chaque AC, des PL pour en déduire la $PIPA$, alors :

$$THAC = MOY_{193,4} \times TBF + a.F \times PL \tag{2.21}$$

Lien avec un schéma d'advection-diffusion

La méthode de « Bilans Enthalpiques » qui nous a menés à cette loi de proportionnalité nous conduit à voir en le coefficient a un terme d'**advection**, représentant le transport de la chaleur au sein du fluide via son mouvement purement vertical, le long de chaque AC. Or nous savons qu'il existe des débits transverses entre assemblages : le brassage du fluide à l'intérieur du cœur est même favorisé par la présence d'un réseau d'ailettes au sein de la structure métallique du cœur.

Par conséquent, nous avons cherché à introduire un terme de **diffusion** ϵ dans cette formule, ce qui induirait une « interaction » entre assemblages : $THAC_j$ ne dépendrait ainsi plus seulement de la puissance le long de l'AC numéro j, mais également des puissances le long de ses voisins.

Concrètement, en nommant $T(j,h)$ la température au niveau la partie inférieure du segment h de l'assemblage j, et en nommant V_j l'ensemble des indices des N_j voisins de cet assemblage, on définit le schéma suivant :

$$T(j,h+1) = (1-\beta)T(j,h) + \frac{\beta}{N_j}\left[\sum_{j' \in V_i} T(j',h)\right] + \alpha(1-\beta)PL_{j,h} + \frac{\alpha\beta}{N_j}\left[\sum_{j' \in V_i} PL_{j',h}\right] \tag{2.22}$$

où α est le paramètre d'advection et β le paramètre de diffusion.

En posant M_β la matrice de taille 193×193 définie par :

$$M_{\beta,i,j} = \begin{cases} 1-\beta & si \quad i = j \\ \frac{\beta}{N_i} & si \quad j \in V_i \\ 0 & sinon \end{cases}$$

alors le schéma peut s'écrire matriciellement échelon par échelon. En posant $T(h)$ le vecteur colonne de la carte de température à l'échelon h et $PL(h)$ le vecteur des puissances à ce même échelon, on montre par récurrence, pour $h > 0$:

$$T(h) = MOY_{193,4} \times TBF + \alpha \sum_{h'=0}^{h-1} M_\beta^{h-h'} PL(h') \tag{2.23}$$

On pose alors la matrice F_β :

$$F_\beta = \begin{pmatrix} M_\beta^{248} & 0 & \dots & 0 \\ 0 & M_\beta^{247} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & M_\beta \end{pmatrix}$$

On obtient alors, avec cette notation :

$$THAC = MOY_{193,4} \times TBF + \alpha.F_{beta} \times PL \tag{2.24}$$

On retrouve une expression vectorielle affine du modèle, de la même forme que l'équation 2.21 : la correspondance est exacte lorsque $\alpha = a$ et $\beta = 0$. Il est apparu, en cherchant à identifier les meilleurs paramètres (α, β) de notre modèle, à l'aide de cartes de flux disponibles, que le caractère diffusif du modèle est extrêmement négligeable devant l'advection, et que le paramètre β peut être presque systématiquement négligé. On est donc ramené in fine à la loi PIPA telle que décrite par l'équation 2.20.

Les puissances locales, de même que les puissances intégrées par assemblages, sont souvent renormalisées à l'aide de la puissance thermique du réacteur, de sorte qu'on peut exprimer les expressions de la loi PIPA sous la forme scalaire normalisée :

$$\forall j, THAC_j = TBF_{MOY} + a\frac{PTHREAC}{193}.PIPAN_j \tag{2.25}$$

ou sous la forme vectorielle normalisée :

$$THAC = MOY \times TBF + \alpha\frac{PTHREAC}{193 \times 248}F_\beta \times PLN \tag{2.26}$$

Résumé

La génération de chaleur au sein du cœur est représentée par un modèle direct simplifié :

$$THAC = f_{gen,a}(TBF, PIPAN, PTHREAC) \tag{2.27}$$

où :

— $THAC$ est le vecteur des températures en haut des AC ;

— TBF est le vecteur des températures de branche froides ;

— $PIPAN$ est le vecteur des Puissances Intégrées par Assemblage Normalisées ;

— $PTHREAC$ est la puissance thermique du réacteur, en MW ;

— $f_{gen,a}$ est la « Loi Pipa » décrite plus haut ;

— a est un paramètre scalaire, dont la valeur varie peu selon les cycles ou les tranches 1300 MW étudiées.

Cette expression peut être généralisée en utilisant les PLN au lieu des $PIPAN$:

$$THAC = f_{gen}(TBF, PLN, PTHREAC) \tag{2.28}$$

la fonction f_{gen} pouvant être alors :
— le modèle PIPA sous sa forme vectorielle ;
— le schéma d'équation d'advection diffusion ;
— un code de calcul industriel, comme THYC ou Code_Saturne, indifféremment.

Brassage dans le plénum supérieur

Principe

La seconde étape du modèle vise à représenter le mélange du fluide dans le plénum supérieur et sa répartition entre les quatre branches chaudes, en utilisant la moyenne par quarts de cœurs avec pondération dite de « bypass » résumée par l'équation 2.19 : en prenant pour paramètre d'entrée la cartographie 2D des températures en haut des AC, ($\{THAC_j\}$) et la température moyenne de Branche Froide TBF_{MOY}, on peut obtenir une évaluation des quatre températures de branche chaude ($\{TBC_i\}$). Soit :

$$\forall i, TBC_i = \epsilon TBF_{MOY} + \frac{1-\epsilon}{N_i} \sum_{j \in C_i} THAC_j \tag{2.29}$$

où C_i est l'ensemble des indices j (au nombre de N_i) tels que l'assemblage combustible numéro j est situé dans le quart de cœur géographique attribué à i : ces ensembles sont à construire en fonction de la forme des quarts de cœurs choisis. Le choix du paramètre ϵ est plus complexe : plusieurs études antérieures ([Vai12], [Esc11], [MFG14])) utilisaient la valeur de conception du débit de bypass, c'est-à-dire la fraction de fluide primaire qui ne passe pas à travers le cœur, mais ce choix est à la fois arbitraire et approximatif, et le paramètre ϵ doit être ajusté, comme nous l'expliquons plus loin.

Forme vectorielle

La formule 2.29 se met facilement sous la forme vectorielle suivante :

$$TBC = \epsilon MOY_{4,4} \times TBF + (1-\epsilon) L \times THAC \tag{2.30}$$

où $MOY_{4,4}$ est la matrice de taille 4×4 dont tous les coefficients sont $1/4$, et la matrice L ne dépend que du choix de la forme des quarts de cœur. Nous avons opté pour des quarts de cœur se recoupant au niveau de la ligne et de la colonne centrales, soit des quadrants de taille 56, comme celle des études EDF R&D ayant précédemment utilisé cette formules.

Résumé

Le mélange du fluide primaire dans le plénum supérieur et sa répartition entre les différentes boucles est représentée par un modèle direct simplifié :

$$TBC = f_{mix,L,\epsilon}(TBF, THAC) \tag{2.31}$$

où :

— TBC est le vecteur des températures de branches chaudes ;

— TBF est le vecteur des températures de branche froides ;

— $THAC$ est le vecteur des températures en haut des AC ;

— ϵ est un paramètre scalaire à identifier, pouvant être initialisé par une valeur par défaut comme la valeur de conception du débit de bypass ;

— L est une matrice de moyenne par quarts de cœur simple, ne dépendant que de la forme des quarts de cœur adoptés ;

— $f_{mix,L,\epsilon}$ est la fonction de moyenne par quart de cœur pondérée décrite plus haut.

Ce modèle peut être remplacé par un code de calcul industriel comme Code_Saturne, par exemple.

Transfert de chaleur au secondaire

Principe

La troisième étape du modèle concerne l'échange de chaleur entre le fluide primaire et le fluide secondaire au niveau des GV, en utilisant simplement le bilan enthalpique du théorème 1 : en prenant pour paramètre d'entrée les quatre températures de branche chaude TBC_i, les quatre températures de branche froide TBF_i, et les quatre débits-boucles Q_i, on en déduit la puissance transférée au secondaire au niveau de chaque GV, $PTHGV_i$. Nous rappelons ce bilan enthalpique :

$$PTHGV_i = Q_i \times (h(TBC_i) - h(TBF_i)) + \frac{PGMPP}{4} \tag{2.32}$$

où $h(T)$ est l'enthalpie massique de l'eau à la température T et à la pression $P = 155\ bar$ (en $kJ.kg^{-1}$) et $PGMPP$ est la puissance de l'ensemble des pompes primaires (approximativement 18 MW).

Forme vectorielle

Cette étape-ci du modèle n'a pas de forme vectorielle évidente, contrairement aux autres étapes du modèles : même en linéarisant la fonction h, le terme Q_i reste en facteur et il

subsiste donc un produit croisé qui rend cette formule non linéaire. On peut chercher à linéariser cette formule, mais l'approximation induite est trop grande pour que le modèle reste fidèle.

Résumé

Le mélange du fluide primaire dans le plénum supérieur et sa répartition entre les différentes boucles est représentée par un modèle direct simplifié :

$$PTHGV = f_{GV,PGMPP}(Q, TBF, TBC) \tag{2.33}$$

où :

— Q est le vecteur des débits-boucles massiques ;

— TBC est le vecteur des températures de branches chaudes ;

— TBF est le vecteur des températures de branche froides ;
— $PGMPP$ est la puissance cumulée des pompes primaires et des chaufferettes au sein du circuit primaire ;

— $f_{GV,PGMPP}$ est la fonction correspondant au bilan enthalpique aux bornes des GV rappelé ci-dessus.

Ce modèle peut éventuellement être remplacé, si on souhaite raffiner la connaissance du champ 3D au sein du GV, par le code de calcul THYC-GV, mais il faut dans ce cas se pencher plus attentivement sur le circuit secondaire, et l'intérêt d'une telle complexification ne semble pas majeur.

Bouclage du modèle

Principe

A l'aide des trois étapes décrites ci-dessus, on peut reconstituer l'état de l'ensemble du circuit primaire à partir des paramètres d'entrées $Q, TBF, PIPAN$ (ou PLN), $PTHREAC$ et ϵ le coefficient de pondération de la moyenne par quarts de cœur. En particulier, on peut en déduire les puissances thermiques des GV.

Nous l'avons dit, parmi ces paramètres d'entrée, le seul que nous ne connaissons pratiquement pas est le coefficient ϵ. Cela confère à notre modèle simple un défaut important : il peut exister un déséquilibre entre la puissance produite au primaire et celle transférée au secondaire. Concrètement, en effectuant un bilan de puissance sur le circuit primaire, on devrait avoir l'équation suivante :

$$PTHREAC + PGMPP = \sum_i PTHGV_i \tag{2.34}$$

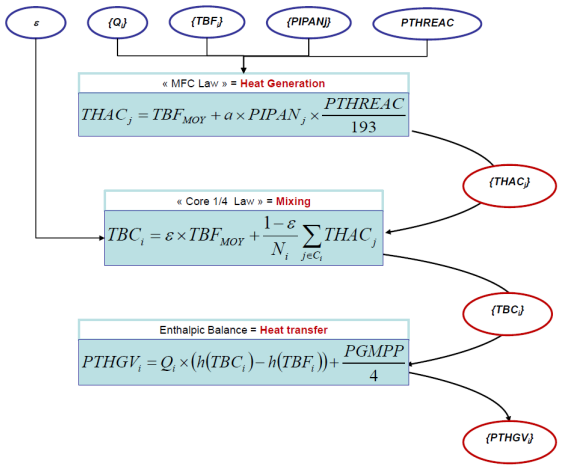


FIGURE 2.19 – Modèle Semi-Empirique première version

Dans notre modèle simplifié, ce bilan énergétique devient une équation de bouclage qui détermine le paramètre ϵ.

Développement

En effet, en combinant la loi PIPA et la moyenne par quarts de cœur, on obtient :

$$TBC_i = (1-\epsilon)\frac{1}{N_j}\sum_{j \in C_i}\left(TBF_{MOY} + aPIPAN_j\frac{PTHREAC}{193}\right) + \epsilon TBF_{MOY} \tag{2.35}$$

On sort de la moyenne tous les termes qui ne dépendent pas de l'indice j :

$$TBC_i = TBF_{MOY} + (1-\epsilon)a\frac{PTHREAC}{193}\frac{1}{N_j}\sum_{j \in C_i} PIPAN_j \tag{2.36}$$

On pose $k_i = \frac{1}{N_j}\sum_{j \in C_i} PIPAN_j$, terme proche de 1 (puisque la moyenne des $PIPAN$ sur l'ensemble du cœur vaut par définition 1 qui ne dépend que de la forme des quarts de cœur et de la forme de la carte de flux neutronique (plus précisément du déséquilibre entre quadrants du flux neutronique). L'expression des TBC_i se réduit alors simplement à :

$$TBC_i = TBF_{MOY} + (1-\epsilon)a\frac{PTHREAC}{193}k_i \tag{2.37}$$

On injecte alors cette expression dans les bilans enthalpiques GV :

$$PTHGV_i = Q_i \times (h(TBF_{MOY} + (1-\epsilon)a\frac{PTHREAC}{193}k_i) - h(TBF_i)) + \frac{PGMPP}{4} \tag{2.38}$$

Et il ne reste plus qu'à sommer cette équation pour obtenir l'Equation de Bouclage Développée du modèle simple :

Théorème 3 (Equation de Bouclage Développée)**.** *Les paramètres Q, TBF, $PIPAN$, $PTHREAC$, a et ϵ du modèle simplifié doivent vérifier l'équation suivante :*

$$PTHREAC = \sum_i Q_i \left(h(TBF_{MOY} + (1-\epsilon)a\frac{PTHREAC}{193}k_i) - h(TBF_i) \right) \tag{2.39}$$

où TBF_{MOY} est la moyenne des températures de branche froide et k_i la moyenne des $PIPAN$ sur le quart de cœur i.

Valeur de « bypass-fit » Dans l'équation 2.39, le seul terme pour lequel on ne dispose d'aucune mesure – et qui, en fait, ne correspond pas à un paramètre physique clairement défini – est le coefficient de la moyenne par quarts de cœur, ϵ. Au vu de la monotonie de la fonction h, il existe une seule et unique valeur de ce coefficient pour laquelle l'équation de bouclage est respectée. Nommons ϵ_{fit} cette valeur : a priori, il dépend de tous les autres paramètres d'entrée du modèle ($PTHREAC$, Q, TBF, $PIPAN$ et a).

On peut effectuer une première application numérique en effectuant plusieurs grosses hypothèses simplificatrices :

— débits-boucles et températures de branches froides uniformes et prenant leurs valeurs de conception (5.0 t/s et 290°C respectivement) ;

— carte de flux uniforme ($k_i = 1$) et puissance nominale ($PTHREAC = 3800\ MW$) ;

— linéarisation de h : $\Delta h \approx C_p \Delta T$ avec $C_p \approx 5.6\ kJ/K/kg$.

L'équation de bouclage s'inverse alors facilement, et on obtient :

$$\epsilon_{fit} \approx 1 - \frac{1}{\frac{4}{193}Q.C_p.a} \approx 6.14\% \tag{2.40}$$

Sous ces hypothèses simplificatrices, donc, la valeur du coefficient ϵ qui permet de « fitter » la puissance thermique générée au primaire et celle transférée au secondaire, selon notre modèle simplifié, est donc proche – par coïncidence – de la fraction de débit de bypass au sein du cœur.

Recherche dichotomique

On ne peut pas prendre une valeur constante pour ϵ_{fit} : en effet, on constate dans les faits que cela conduirait, notamment lors de variations de puissance de tranche, à amener un déséquilibre pouvant dépasser la centaine de MW dans certains cas. On définit alors une méthode pour évaluer ϵ_{fit} à partir d'un jeu de paramètres d'entrées du modèle.

Pour ce faire, on procède par recherche dichotomique : on définit les bornes de l'intervalle de recherche, ϵ_{min} et ϵ_{max} (par exemple respectivement 0.0% et 10.O%) et la précision souhaitée pour le résultat final, $\delta\epsilon$ (par exemple 0.0001%). Ensuite, tant que $\epsilon_{max} - \epsilon_{min} < \delta\epsilon$, on prend $\epsilon = \frac{\epsilon_{max} - \epsilon_{min}}{2}$ et on évalue le terme de droite de l'équation de bouclage :

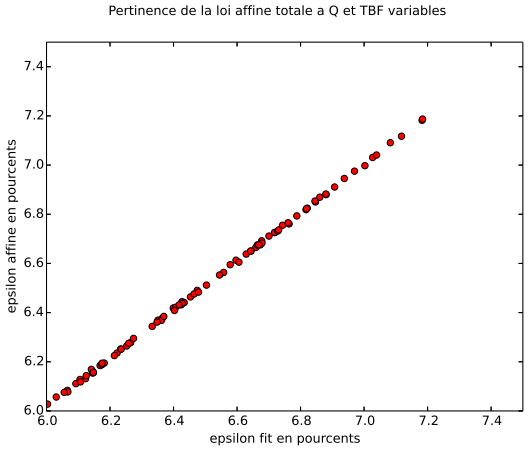


FIGURE 2.20 – Confrontation du résultat de l'évaluation du coefficient ϵ par recherche dichotomique et par loi affine

— s'il est supérieur au terme de gauche, c'est-à-dire si la puissance totale des GV est trop élevée, alors cela signifie qu'ϵ est trop faible, et on pose $\epsilon_{min} = \epsilon$;

— s'il est inférieur au terme de gauche, c'est-à-dire si la puissance totale des GV est trop basse, alors cela signifie qu'ϵ est trop fort, et on pose $\epsilon_{max} = \epsilon$.

Au bout d'un nombre fixe d'itérations (17, avec nos critères de recherche), l'algorithme s'arrête, et on dispose d'une évaluation correcte pour notre coefficient dit de bypass :

$$\epsilon_{fit} = \frac{\epsilon_{max,final} - \epsilon_{min,final}}{2} \pm \frac{\delta\epsilon}{2} \tag{2.41}$$

Approximation Affine

Une étude de sensibilité du calcul de ϵ_{fit} aux paramètres $PTHREAC$, k_i, Q et TBF permet de montrer que, pour des petites variations autour d'un état de référence, ϵ_{fit} peut être évalué sous la forme affine suivante :

$$\epsilon_{fit} \approx \frac{1}{100}\left[7.09 + 92\overline{k_i - 1} + 81\overline{\frac{Q_i}{Q_0} - 1} + 190\overline{\frac{TBF_i}{TBF_0} - 1} + 11\left(\frac{PTHREAC}{PTHREAC_0} - 1\right)\right] \tag{2.42}$$

où Q_0, TBF_0 et $PTHREAC_0$ sont respectivement les valeurs moyennes des débits-boucles, des températures de branche froide et de la puissance thermique du réacteur à l'état de référence, et où l'opérateur $\bar{\cdot}$ est la moyenne sur l'ensemble des boucles.

De sorte qu'une diminution de 1% de la puissance thermique du réacteur provoquera une diminution de 0.11% (e.g. de 7.10% à 6.99%) du coefficient ϵ_{fit}, etc.

En pratique, nous utiliserons la recherche dichotomique pour calculer systématiquement le coefficient ϵ_{fit} et permettre ainsi à l'état calculé par notre modèle semi-empirique de vérifier l'équation de bouclage : ce coefficient dit « de bypass », qui intervient dans la

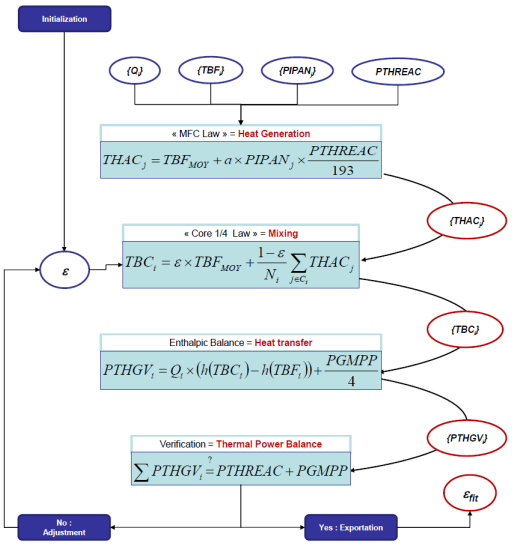


FIGURE 2.21 – Modèle Semi-Empirique avec bouclage par fit de coefficient de bypass

moyenne par quarts de cœur, varie donc de façon importante selon les valeurs des autres paramètres primaires (débits-boucles, températures de branche froide, forme de la carte de flux et puissance thermique totale du réacteur). S'il ne correspond pas à un paramètre physique clairement identifié (en particulier, il n'existe pas d'argument valable permettant d'affirmer qu'il s'agit bien de la fraction du débit de bypass au sein du cœur), le coefficient ϵ_{fit} joue ainsi le rôle d'une variable d'ajustement importante, et peut donc être considéré comme un paramètre primaire à part entière, pour lequel il n'existe aucune mesure directe.

2.5 Bilan

De nombreux outils sont disponibles autour du circuit primaire pour procéder à l'estimation, directe ou indirecte, des principaux paramètres primaires. Notre objectif est d'utiliser au mieux l'ensemble de ces données pour les assimiler simultanément afin de produire une estimation globalement plus fiable de l'ensemble des paramètres primaires.

Pour ce faire, nous effectuons le distinguo entre les mesures directes disponibles, et les estimateurs qui utilisent des liens physiques (théoriques ou empiriques) entre les différentes variables qui décrivent l'état du circuit primaire :

— les mesures directes définissent ainsi un système d'observations, qui constitue notre « porte d'entrée » au sein du circuit primaire réel ;

— les principes d'estimations sont quant à eux regroupés dans un modèle physique de type 0D/1D (que l'on peut résumer par la Figure 2.21) qui rassemble donc toute l'information que nous possédons *a priori* sur le fonctionnement du circuit primaire, sans avoir besoin d'observations pour nous le confirmer.

Comme nous l'avons pu voir tout au long de ce chapitre, les différents capteurs utilisés doivent souvent être soigneusement recalés, notamment en début de cycle lors de l'Arrêt

A Chaud, et certains estimateurs ont besoin d'être paramétrés :

- mesures de températures RCP : recalage à l'Arrêt à Chaud ;
- mesures de débit par ΔP-Coude : calibrage à l'aide des valeurs de ΔP_{cal} enregistrées sous SYGMA, identification de la position des capteurs dans le tube, récupération des coefficients K obtenus par simulation CFD (ou, à défaut, utilisation de la méthode Westinghouse pour obtenir une valeur de départ) ;
- mesures de température in-core : recalage des thermocouples ;
- mesure BIL100/BILORLI : recalages des capteurs utilisés du circuit secondaire et calcul ;
- estimateur « Loi Pipa » : identification du paramètre a (ou, dans le cas du modèle type advection-diffusion des paramètres α, β.

Ces prétraitements sont un préliminaire indispensable à l'utilisation de tout algorithme d'assimilation de données : en assimilant des données de mauvaise qualité, on obtient généralement une analyse de qualité au mieux médiocre.

Enfin, concernant le modèle que nous avons construit, on remarquera que sa conception modulaire – par étapes successives – a été choisie dans la perspective de laisser ouverte la possibilité d'utiliser à la place des codes industriels plus complexes, sans que cela change le raisonnement général et la méthodologie d'utilisation.

Chapitre 3

Assimilation de Données

L'Assimilation de Données est définie comme l'ensemble des techniques statistiques qui permettent d'améliorer la connaissance (passée, présente ou future) de l'état d'un système physique, à partir de la donnée concomitante d'observations expérimentales sur le système et de sa connaissance théorique ou a priori (i.e. un modèle physique) [Boc12].

3.1 Principes généraux

3.1.1 Histoire

On se réfère notamment à [Tal97].

Le terme d'« Assimilation de Données » est assez générique, et désigne un type de Problème Inverse rencontré dans divers domaines scientifiques et technologiques. Ainsi, les problèmes de navigation (aérienne, navale ou spatiale), dans lesquels on cherche à déterminer à tout instant, aussi précisément que possible, la position et la vitesse d'un véhicule « observé » à travers divers instruments, relèvent de l'Assimilation de Données. Toute forme de traitement du signal est aussi un problème mathématique de nature semblable, et on peut noter que le Filtre de Kalman – l'un des outils de base de l'Assimilation de Données en météorologie ou en océanographie – trouve son origine dans le domaine de l'électrotechnique, et a été développé pour résoudre des problèmes d'estimations de trajectoire pour le programme spatial APOLLO.

En fait, nombre de Problèmes Inverses présentent des similarités, dans leur formulation et dans leurs outils mathématiques, avec l'Assimilation de Données. En physique des plasmas, on cherche souvent à connaître l'état interne d'un système physique à partir d'observations pratiquées à sa surface (et, de plus, à contrôler le système en agissant sur sa surface). La Géophysique est un autre exemple : presque tout ce que l'on connaît de la structure interne de la Terre vient de l'inversion de signaux (sismiques, essentiellement) observés à sa surface.

Tous ces exemples consistent finalement en des problèmes d'estimation, dans lesquels on cherche à déduire l'état interne d'un système physique à partir d'informations qui peuvent être extrêmement diverses en termes d'origine, de nature et de précision, et qui peuvent très bien n'être que très indirectement reliées aux quantités à estimer.

En dépit du caractère très généraliste de cette formulation, et de la très grande diversité

des sciences qu'elle concerne, les méthodes utilisées pour résoudre ces problèmes sont curieusement très proches dans leurs principes, bien qu'elles aient toutes été développées de façon indépendante. Elles s'expriment toutes (ou, du moins, peuvent toutes être exprimées) en des termes probabilistes, relevant ainsi de l'Estimation Bayésienne, et visent, dans les faits, à déterminer une approximation convenable de la distribution de probabilité conditionnelle de l'état du système, étant donné l'ensemble des informations disponibles. Les algorithmes numériques qui opèrent leurs résolutions sont très proches les uns des autres, et fondamentalement indépendants des équations qui gouvernent la physique du système considéré.

Ainsi, on peut déceler les prémices de l'Assimilation de Données dans les premiers articles sur la théorie de l'Estimation, à travers la Méthode des Moindres Carrés, développée par Legendre en 1805 [Leg05] et Gauss en 1809 [Gau09], pour des calculs de mécanique céleste. L'optimalité de cette méthode fut démontrée par Gauss en 1826 [Gau26], résultat qui constitue l'un des théorèmes de base de l'Assimilation de Données, le BLUE (*Best Linear Unbiased Estimator*), connu également sous l'appellation « Théorème de Gauss-Markov ». L'apposition du nom de Markov [Mar99] à ce théorème vient du fait qu'il a retrouvé les mêmes résultats que Gauss, un siècle plus tard, avec des résultats relevant plutôt de l'Estimation Bayésienne. Cette dernière avait été développée à partir des résultats du statisticien anglais Thomas Bayes (1702-1761) [Bay63] notamment par Laplace [Lap74], sensiblement à la même époque que Legendre et Gauss (première moitié du XIX^{e} siècle), pour des applications aussi diverses que la mécanique céleste, la médecine, la fiabilité des machines, voire la jurisprudence. L'estimation au sens des moindres carrés et l'estimation au sens bayésien se sont ainsi rejointes dans le théorème du BLUE, et cette équivalence – sous certaines hypothèses – des deux philosophies est un aspect important de toute l'Assimilation de Données par la suite.

A partir des années 1950, plusieurs algorithmes voient le jour : l'Analyse de Cressman (1959) [Cre59], l'Interpolation Optimum (OI) (1963) [Gan63], et le Filtre de Kalman (1960) [Kal60]. L'application de ces trois algorithmes à la météorologie marquera un tournant. Les prédictions effectuées peuvent désormais être globalement corrigées grâce à l'ensemble des observations disponibles, fussent-elles dispersées dans le temps ou l'espace. Parallèlement, des approches variationnelles équivalentes ou presque (toujours sous certaines hypothèses) au BLUE, le 3DVAR [Lor81], et au Filtre de Kalman, le 4DVAR, sont développées pour des applications associées à la météorologie comme la prévision des trajectoires des ouragans : mais il faudra attendre les années 1990 pour que ces schémas variationnels soient effectivement opérationnels sur les grands modèles météorologiques utilisés par exemple par Météo-France, le NMC (*National Meteorological Center*, Etats-Unis), ou l'ECMWF (*European Center for Medium-Range Weather Forecasts*). L'utilisation de ces techniques d'Assimilation de Données dans les chaînes de calcul, associée à l'augmentation croissante du nombre d'observations prises en compte et à l'accroissement constant de la puissance de calcul des superordinateurs, est à l'origine de la majeure partie des gains obtenus sur la qualité des prévision météorologiques.

Les Géosciences ont constitué au cours des dernières décennies le principal moteur et le principal objet de recherche de l'Assimilation de Données : néanmoins, ses racines profondes et ses principes fondamentaux demeurent très généralistes. Récemment, plusieurs disciplines ont redécouvert l'Assimilation de Données, et constaté sa parenté étroite avec d'autres types d'algorithmes de Problèmes inverses [Tar87] qu'elles utilisaient jusqu'alors,

découvrant ainsi, tels M. Jourdain [1], qu'ils avaient fait de l'Assimilation de Données sans le savoir.

3.1.2 Exemple scalaire

L'un des principes essentiels de l'Assimilation de Données est de combiner un maximum d'informations disponibles, en accordant un poids plus grand aux informations les plus fiables.

Prenons un exemple scalaire très simple : on cherche à estimer une grandeur scalaire inconnue, par exemple une température T pour laquelle on dispose de deux estimations T_1 et T_2 d'incertitudes respectives σ_1 et σ_2, provenant par exemple respectivement d'un modèle et d'un thermomètre réel. Formellement, on écrit ainsi :

$$T_1 = T + \epsilon_1$$
$$T_2 = T + \epsilon_2$$

où ϵ_1 et ϵ_2 sont deux bruits blancs, c'est-à-dire deux variables aléatoires d'espérances nulles et d'écarts-types respectifs σ_1 et σ_2. En termes d'invervalles de confiance à 95%, si les distributions sont gaussiennes, on notera ainsi $T_1 = T \pm 2\sigma_1$ et $T_2 = T \pm 2\sigma_2$.

On cherche à combiner ces deux données, en accordant un poids plus élevé aux informations les plus fiables. On cherche ainsi à bâtir un estimateur x_a (a pour « analyse ») défini par :

$$T_a = kT_1 + (1-k)\,T_2 \tag{3.1}$$

où k est un coefficient qui représente la précision relative des deux estimateurs :

- si T_1 est infiniment précis comparé à T_2 ($\sigma_1 \ll \sigma_2$), alors on aura $k = 1$ et donc $T_a = T_1$;
- inversement, si T_2 est infiniment précis comparé à T_1 ($\sigma_2 \ll \sigma_1$), alors on aura $k = 0$ et donc $T_a = T_2$;
- si T_1 et T_2 sont de précisions identiques ($\sigma_1 = \sigma_2$), alors on aura $k = \frac{1}{2}$ et $T_a = \frac{T_1+T_2}{2}$.

Si l'on nomme ϵ_a la variable aléatoire égale à l'erreur commise par l'estimateur T_a, on a :

$$\epsilon_a = T_a - T = kT_1 + (1-k)\,T_2 - T = k\epsilon_1 + (1-k)\,\epsilon_2 \tag{3.2}$$

On a ainsi :

$$E\left(\epsilon_a\right) = 0 \tag{3.3}$$

et, sous réserve que les deux erreurs commises par les deux estimateurs, ϵ_1 et ϵ_2, soient indépendantes :

$$\sigma_a^2 = E\left(\epsilon_a^2\right) = k^2\sigma_1^2 + (1-k)^2\,\sigma_2^2 \tag{3.4}$$

1. « Par ma foi! Il y a plus de quarante ans que je dis de la prose sans que je n'en susse rien, et je vous suis le plus obligé du monde de m'avoir appris cela. » Molière, Le Bourgeois Gentilhomme

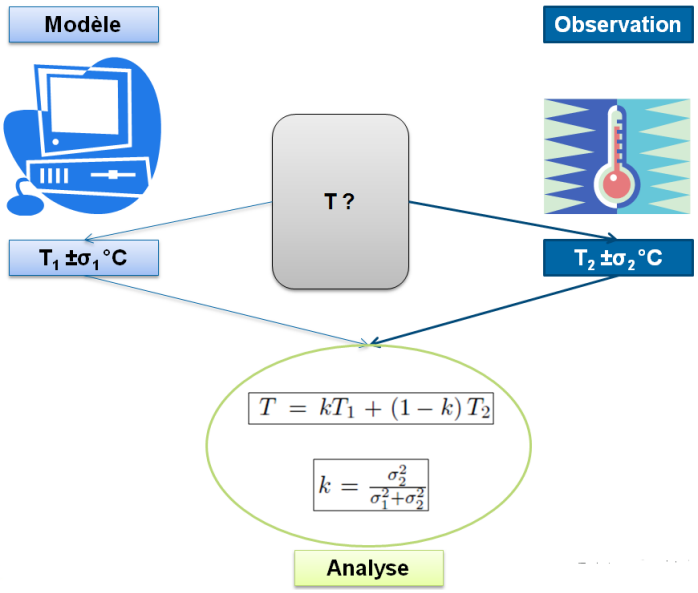


FIGURE 3.1 – Exemple scalaire simple de problème d'assimilation de données

Le coefficient k qui minimise l'incertitude associée à l'estimateur T_a peut donc se calculer avec une simple condition du premier ordre, $\frac{\partial(\sigma_a^2)}{\partial k} = 0$, qui donne pour résultat :

$$k = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \tag{3.5}$$

Et, dans ce cas :

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \tag{3.6}$$

Ainsi, la variance de l'erreur commise par notre analyse est le double de la moyenne harmonique des variances des erreurs initiales (et leur est strictement inférieure). Une autre façon de l'exprimer est, en nommant « précision » d'un estimateur l'inverse de la variance de l'erreur qu'il commet, qu'en assimilant nos deux estimateurs initiaux T_1 et T_2, on a additionné leurs précisions (ou moyenné harmoniquement leurs « incertitudes » .

Ce résultat n'est autre que le BLUE dans son expression la plus simple. Il est aisément généralisable, et l'un de ses corollaires est un résultat bien connu des statisticiens : si l'on dispose de N estimateurs identiques et indépendants pour une même grandeur, on obtient, en effectuant leur moyenne, un estimateur – qu'on peut qualifier de « Best Estimate » – qui « divise l'incertitude » par $\sqrt{N}$ (hypothèses de quasi-ergodicité).

3.1.3 Formalisation du problème

On suppose que le système physique peut être représenté par les composantes d'un vecteur (colonne) x de $\mathbb{R}^n$. Choisir ce vecteur constitue en soi une première approximation : une description exhaustive du système suppose généralement une infinité de données (fonctions continues dans l'espace et le temps, distributions de probabilité, prise en compte de la totalité des phénomènes physiques entrant en jeu, etc.). Il faudrait donc, concrètement, opter pour un champ vectoriel spatio-temporel avec autant de composantes qu'il y a

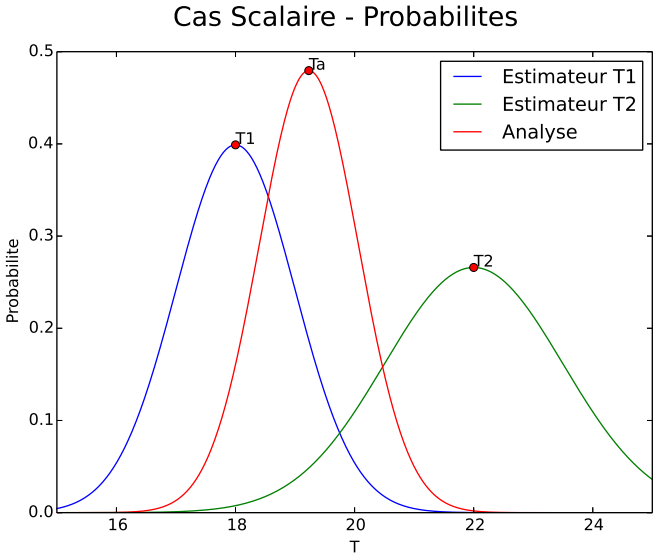


FIGURE 3.2 – Cas scalaire : distributions de probabilités des estimateurs initiaux et de l'analyse

d'observables potentielles en chaque point. Cependant, dans le cadre d'une simulation numérique, une discrétisation est indispensable : il faut ainsi mailler l'espace, échantillonner dans le temps, approximer les distributions statistiques, se désintéresser de certains aspects de la physique du système, jusqu'à trouver le meilleur compromis entre représentativité et complexité. L'espace vectoriel correspondant à ce choix sera nommé **Espace d'Etat** et on notera n sa dimension.

x peut ainsi être, en météorologie, un vecteur constitué des valeurs des températures moyennes en les points d'un maillage donné ; en balistique, les coordonnées cinématiques (position, vitesse, angles, vitesses de rotations, etc.) du mobile étudié ; en chimie atmosphérique, les concentrations des différents polluants en les points d'un maillage, etc. ...

Le problème d'estimation consiste donc à identifier le vecteur dit **réel** x_t (t pour « *True* ») qui représente le mieux la réalité de l'objet physique étudié. Pour ce faire, on dispose de deux types d'informations :

— des estimations théoriques (ou a priori) provenant de modèles numériques ;
— des observations provenant directement de l'objet physique réel.

On nommera ainsi x_b (b pour « *background* », ce qui signifie « **ébauche** ») le vecteur d'état qui constitue une connaissance a priori (sans avoir pris en compte les observations) du système, à partir d'un modèle régissant sa structure ou son évolution, ou à partir d'une estimation effectuée par le passé et qu'on cherche à mettre à jour. Les observations, quant à elles, sont réunies dans un vecteur dit **vecteur d'observations** y, lequel évolue dans un espace différent de l'Espace d'Etat, qu'on nommera **Espace des Observations** et dont on notera p la dimension. Notre objectif est de combiner ces deux informations disponibles, selon leur fiabilité, pour produire une **analyse** x_a, qui « s'approche le mieux » du vecteur d'état réel x_t .

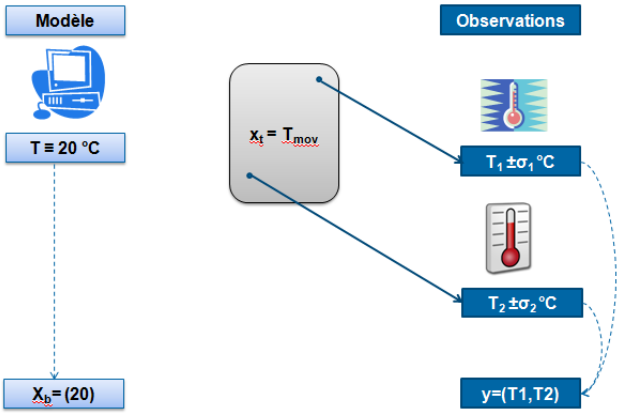


FIGURE 3.3 – Exemple de problème simple d'Assimilation de Données : estimation de la température moyenne d'une pièce

Cette formulation est volontairement ambiguë et soulève d'emblée deux questions :

— comment comparer et combiner x_b et y, qui n'évoluent pas dans le même espace ?
— comment quantifier la précision d'une estimation ?

Répondre à ces deux questions est tout l'objet de l'Assimilation de Données.

Exemple 1. *On étudie la température d'une pièce dans laquelle se trouvent deux thermomètres.*

Le choix le plus simple possible est de représenter la température de la pièce par un seul scalaire, la température moyenne en degrés Celsius :

$$x_t = (T_{moy}) = \left(\frac{1}{V} \iiint T(M) dM \right)$$

On suppose que la température « normale » d'une pièce est de $20°C$ *: l'ébauche sera donc* $x_b = (20)$*. Le vecteur d'observations, quant à lui, sera* $y \in \mathbb{R}^2$ *composé des valeurs, en* $°C$*, indiquées par les deux thermomètres.*

3.1.4 Analyse de Cressman

Une première réponse est apportée par l'utilisation de critères objectifs (c'est-à-dire sans utiliser de modèle physique proprement dit) pour produire l'analyse : l'Analyse de Cressman en est un exemple [BC02]. Dans le cas où le vecteur d'état est égal aux valeurs d'un champ en les points d'un maillage de l'espace, et que les observations correspondent aux mesures de ce champ en certains points de l'espace, alors une analyse simple est obtenue en corrigeant l'ébauche via l'équation de mise à jour suivante :

$$x_a(i) = x_b(i) + \frac{1}{\sum_{j=1}^{p} \omega_{i,j}} \sum_{j=1}^{p} \omega_{i,j} (y(j) - H_j[x_b]) \tag{3.7}$$

où H_j est un opérateur d'interpolation permettant d'évaluer la valeur du champ x_b au point spatial de l'observation $y(j)$; et $\omega_{i,j}$ est un poids positif, fonction décroissante de la distance $d_{i,j}$ entre le point de l'espace correspondant à $x(i)$ et celui correspondant à $y(j)$. Cette fonction peut prendre plusieurs formes, par exemple :

$$\omega_{i,j} = max\left(0, \frac{R^2 - d_{i,j}^2}{R^2 + d_{i,j}^2}\right) \tag{3.8}$$

ou encore :

$$\omega_{i,j} = exp\left(\frac{-d_{i,j}^2}{2R^2}\right) \tag{3.9}$$

où R est un paramètre correspondant à un « rayon d'influence » des observations, au-delà duquel une observation n'est plus valide. Ici, la précision d'une information est donc directement associée à sa proximité spatiale avec le point du maillage considéré. On remarque que l'algorithme suppose notamment que $\forall i, \sum_{j=1}^{p} \omega_{i,j} \neq 0$, c'est-à-dire que tout point du maillage peut être « relié » à une observation.

Concrètement, le raisonnement de l'Analyse de Cressman est le suivant :

1. pour une observation y_j, on constate localement un décalage entre le champ prévu par le modèle et l'observation réelle, cette différence prenant la forme d'une information nouvelle (« innovation ») égale à $c_j = y_j - H_j[x_b]$;
2. cette correction se répercute à l'ensemble du maillage, mais avec d'autant moins d'importance qu'on est éloigné du point de mesure (et qu'elle est donc moins pertinente) : la correction c_j sera ainsi appliquée au point $n°i$ du maillage, mais affectée du poids variable $\omega_{i,j}$;
3. une fois toutes les corrections calculées, on effectue en chaque point du maillage la moyenne des corrections à faire, pondérées par leurs poids : en termes de barycentre, la correction appliquée au point $n°i$ sera $Bar\{c_j, \omega_{i,j}\}$.

Exemple 2. *On reprend l'exemple de la température dans une pièce : au lieu de prendre pour vecteur d'état un simple scalaire correspondant à la température de la pièce, on effectue un maillage de la pièce et on prend pour vecteur d'état la température moyenne au voisinage de chacun de ces points. On dispose de deux thermomètres, situés chacun en un point différent du maillage.*

Prenons un cas 1D : une pièce linéaire (un couloir) de $10m$ *de long.*

- ***Espace d'état :*** *on discrétise le segment en prenant un maillage de* 1000 *points. Notre espace d'état est donc ici* $\mathbb{R}^{1000}$.
- ***Ebauche :*** *on suppose a priori que la température est homogène dans le couloir :* $\forall i, x_b = 20°C$
- ***Observations :*** *les points d'observation sont situés aux points d'abscisses* $2m$ *et* $8m$*, et on imagine qu'ils mesurent respectivement* $y_1 = 23°C$ *et* $y_2 = 18°C$.
- ***Pondération :*** *on utilise la fonction de poids de type « max » définie par* (3.8) *ou la fonction de poids de type « exp » définie par* (3.9)*, en prenant pour paramètre* $R = 2.5m$.

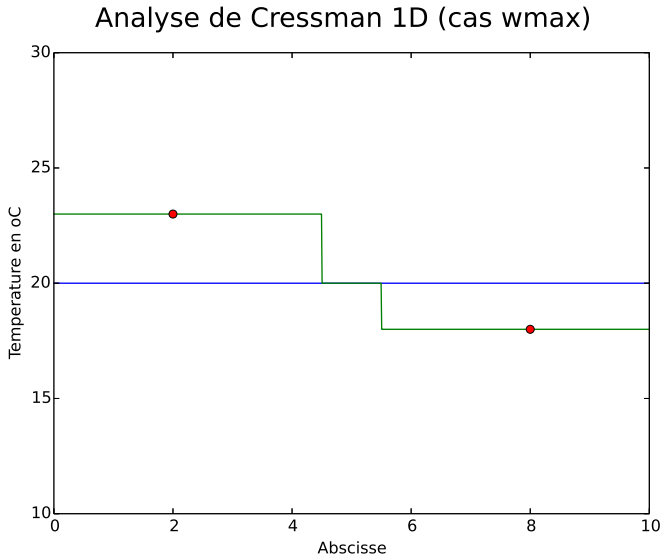


FIGURE 3.4 – Exemple d'Analyse de Cressman dans un cas 1D avec une fonction poids de type max

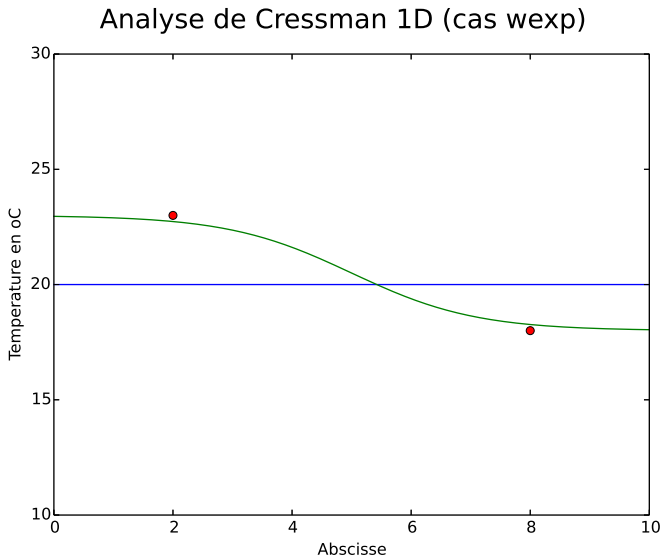


FIGURE 3.5 – Exemple d'Analyse de Cressman dans un cas 1D avec une fonction poids de type exp

Pour le cas max, on voit Fig. 3.4 que chaque observation n'induit une correction que sur son voisinage direct. En revanche, pour le cas exp, chaque observation induit une correction sur l'ensemble du vecteur : on voit ainsi Fig. 3.4 qu'à l'emplacement de l'observation y_1, la valeur du champ analysé n'est pas exactement égale à celle mesurée en ce point.

L'Analyse de Cressman est un algorithme particulièrement simple, mais qui présente plusieurs défauts :

— tout le raisonnement se fonde sur la spatialisation : la précision d'une observation est liée à sa proximité. Si le vecteur d'état est composé non pas de valeurs au point d'un maillage dans l'espace, mais de valeurs hétérogènes, l'algorithme est plus difficilement applicable et bien moins pertinent ;

— l'algorithme ne permet pas de trier rigoureusement les informations en fonction de leur précision : notamment, on ne considère pas l'ébauche et les observations selon leurs précisions relatives. Si on dispose d'une ébauche précise, on n'a pas intérêt à la détériorer en la corrigeant fortement par des observations imprécises, fussent-elles directes et proches ;

— enfin, le vecteur analysé doit satisfaire à plusieurs propriétés physiques connues (équations de conservation, de continuité, etc.), ce que l'Analyse de Cressman ne garantit pas du tout.

Toutefois, l'algorithme introduit deux notions qui sont à la base de l'Assimilation de Données :

— un **Opérateur d'Observation** H est indispensable pour relier l'espace d'état et l'espace d'observations : concrètement, si le vecteur d'état est x, on est censé observer les observations $H[x]$ (si l'opérateur est linéaire, H est tout simplement une matrice de taille $p \times n$) ;

— inversement, le vecteur d'innovation $y - H[x_b]$ (qui se situe dans l'espace d'observations) doit être utilisé et transformé pour induire une correction sur l'ensemble du vecteur d'état : dans le cas de l'analyse de Cressman, cela s'opère linéairement, via une matrice dite de gain, notée K, de taille $n \times p$:

$$K_{i,j} = \frac{\omega_{i,j}}{\sum_{l=1}^{p} \omega_{i,l}} \tag{3.10}$$

Ainsi, l'Analyse de Cressman est un bon point de départ pour se familiariser avec le vocabulaire d'Assimilation de Données, et, dans certains cas spécifiques, permet de produire rapidement des résultats exploitables. Néanmoins, pour réellement utiliser un modèle physique dans un cadre d'Assimilation de Données, il est nécessaire d'introduire plus de rigueur mathématique et de considérations physiques dans le raisonnement.

3.2 Algorithmes

Pour appliquer les principes de l'Assimilation de Données, il convient de définir rigoureusement les incertitudes associées aux différentes données. Ainsi, on définit l'erreur

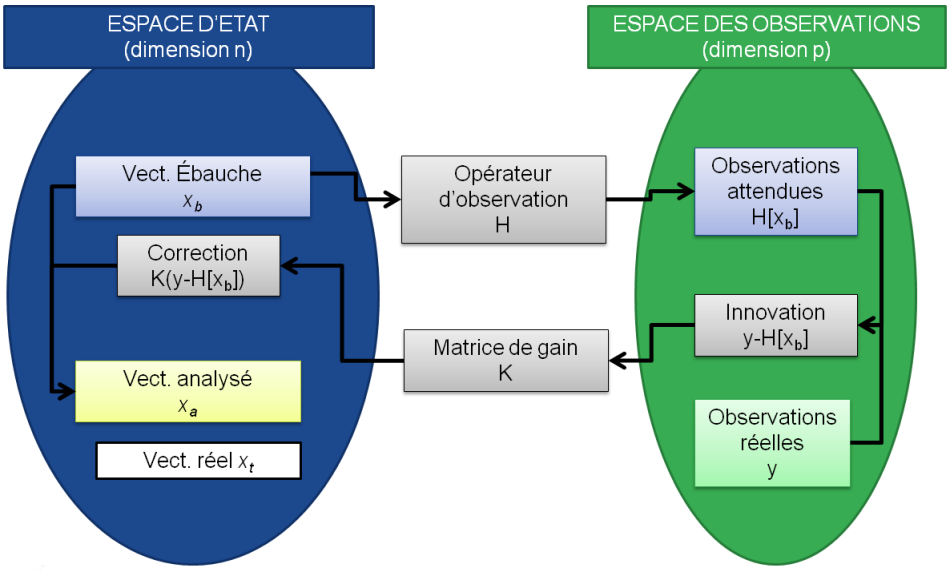


FIGURE 3.6 – Schéma de principe d'une Analyse de Cressman

d'ébauche par :

$$\epsilon_b = x_b - x_t \tag{3.11}$$

et l'erreur d'observation par :

$$\epsilon_{obs} = y - H[x_t] \tag{3.12}$$

où x_b est le vecteur d'ébauche, y le vecteur des observations, et H l'opérateur d'observations.

Le vecteur réel x_t est, par définition, inconnu, et les erreurs d'ébauche ϵ_b et d'observation ϵ_{obs} le sont par conséquent aussi. Pour quantifier les incertitudes associées, il faut ainsi évaluer leurs distributions statistiques. Cette question est épineuse, et constitue en soi un champ de recherche pour lequel la littérature est abondante.

Néanmoins, une approximation courante est de supposer – au moins dans un premier temps – que les erreurs sont sans biais :

$$E[\epsilon_b] = 0$$
$$E[\epsilon_{obs}] = 0$$

On associe par ailleurs à chaque erreur sa matrice de covariance :

$$\begin{aligned} B &= E[(\epsilon_b - E[\epsilon_b]) \cdot (\epsilon_b - E[\epsilon_b])^T] = E[\epsilon_b \cdot \epsilon_b^T] \\ R &= E[(\epsilon_{obs} - E[\epsilon_{obs}]) \cdot (\epsilon_{obs} - E[\epsilon_{obs}])^T] = E[\epsilon_{obs} \cdot \epsilon_{obs}^T] \end{aligned}$$

Il est également fréquent de supposer que les erreurs sont de distribution gaussienne. On notera ainsi :

$$\begin{aligned} \epsilon_b &= NG(0, B) \\ \epsilon_{obs} &= NG(0, R) \end{aligned}$$

et, dans ce cas, la distribution de chaque composante des vecteurs d'erreur sera gaussienne, et l'intervalle de confiance associé sera $[-2\sigma, 2\sigma]$, où σ^2 sera sa variance (le terme diagonal correspondant de la matrice de covariance associée au vecteur d'erreur).

Dans toute cette section, nous introduisons quelques algorithmes simples et très utilisés d'Assimilation de Données.

3.2.1 Le BLUE

Le **BLUE** (« *Best Linear Unbiased Estimator* ») est un résultat fondamental de l'Assimilation de Données, et vient assez naturellement.

On a supposé que les erreurs d'ébauche et d'observation étaient non biaisées : il est normal de chercher une analyse qui ne le soit pas non plus (critère *Unbiased*). Soit en notant ϵ_a l'erreur d'analyse :

$$E[\epsilon_a] = E[x_a - x_t] = 0 \tag{3.13}$$

On suppose à présent que l'opérateur d'observation est linéaire, et, de même que pour le cas scalaire présenté au début de ce chapitre, on cherche l'analyse comme une combinaison linéaire de l'ébauche et de l'observation (critère *Linear*) :

$$x_a = L \cdot x_b + K \cdot y \tag{3.14}$$

Dans ce cas, la condition de non biais devient :

$$\begin{aligned} 0 &= L \cdot E[x_b - x_t + x_t] + K \cdot E[y - H \cdot x_t + H \cdot x_t] - E[x_t] \\ 0 &= L \cdot E[\epsilon_b] + K \cdot E[\epsilon_{obs}] + (L + KH - I)E[x_t] \end{aligned}$$

où I est la matrice identité (de taille $n \times n$). Comme on a supposé que les erreurs étaient sans biais, une condition suffisante est donc $L + KH - I = 0$, soit $L = I - KH$. La nouvelle expression pour le vecteur analysé est ainsi :

$$x_a = x_b + K \cdot (y - H \cdot x_b) \tag{3.15}$$

On retrouve la même forme d'expression que pour l'Analyse de Cressman à la section précédente. Il s'agit à présent d'identifier la matrice K pour laquelle l'analyse est la meilleure possible (critère *Best*). Pour ce faire, évaluons l'erreur d'analyse :

$$\begin{aligned} \epsilon_a &= x_a - x_t \\ &= x_b - x_t + K \cdot (y - H \cdot x_t + H \cdot x_t - H \cdot x_b) \\ &= \epsilon_b + K \cdot (\epsilon_{obs} - H \cdot \epsilon_b) \end{aligned}$$

Par construction, elle est d'espérance nulle. Evaluons à présent la matrice de covariance associée :

$$\begin{aligned} A &= E[\epsilon_a \cdot \epsilon_a^T] \\ &= E[(\epsilon_b + K \cdot (\epsilon_{obs} - H \cdot \epsilon_b))(\epsilon_b + K \cdot (\epsilon_{obs} - H \cdot \epsilon_b))^T] \end{aligned}$$

Lors du développement, on voit apparaître un terme croisé $E[\epsilon_b \epsilon_{obs}^T]$: on fait l'hypothèse supplémentaire que les deux erreurs d'ébauche et d'observation sont décorrélées entre elles, c'est-à-dire :

$$\forall i, \forall j, E[\epsilon_{b,i} \epsilon_{obs,j}] = 0 \tag{3.16}$$

Il vient alors :

$$\begin{aligned} A &= (I - KH)E[\epsilon_b \epsilon_b^T](I - KH)^T + KE[\epsilon_{obs} \epsilon_{obs}^T]K^T \\ &= (I - KH)B(I - KH)^T + KRK^T \end{aligned}$$

Le critère « *Best* » est atteint lorsque l'erreur scalaire commise dans l'analyse, c'est-à-dire $Tr(A)$, la somme des variances des erreurs composantes, est minimale. On étudie ainsi une variation de $Tr(A)$ sous une variation infinitésimale δK de K :

$$\delta(Tr[A]) = Tr[-\delta K H B (I - KH)^T + (I - KH)B(-\delta K H)^T + \delta K R K^T + K R \delta K]$$

où on a utilisé la linéarité (et continuité) de la Trace et la dérivation du produit, et négligé les termes de second ordre ;

$$\delta(Tr[A]) = Tr[(-(I - KH)B^T H^T - (I - KH)BH^T + KR^T + KR)\delta K^T]$$

où l'on a utilisé le fait que la trace d'une matrice est égale à celle de sa transposée ;

$$\delta(Tr[A]) = 2Tr[(-(I - KH)BH^T + KR)\delta K^T]$$

où l'on a utilisé la symétrie de B et de R, et car A est symétrique positive. Ainsi, la condition de premier ordre devient :

$$\begin{aligned} -(I - K^\star H)BH^T + K^\star R &= 0 \\ K^\star(R + HBH^T) &= BH^T \end{aligned}$$

La matrice $R + HBH^T$ est la somme de deux matrices symétriques positives, dont l'une des deux (R) est systématiquement prise comme définie positive ; elle est donc également inversible, et l'on peut multiplier à droite par son inverse, ce qui donne :

$$K^\star = BH^T(R + HBH^T)^{-1}$$

La matrice A s'écrit alors :

$$\begin{aligned} A &= (I - K^\star H)B(I - K^\star H)^T + K^\star R(K^\star)^T \\ &= (I - K^\star H)B + [K^\star R - (I - K^\star H)BH^T](K^\star)^T \\ &= (I - K^\star H)B \end{aligned}$$

car l'expression entre crochets est nulle d'après la condition du premier ordre.

En résumé :

Théorème 4 (BLUE). *Si les erreurs d'ébauche et d'observation sont indépendantes et sans biais, et que l'opérateur d'observation est linéaire, alors la meilleure estimation s'écrit sous la forme :*

$$x_a = x_b + K^\star \cdot (y - H \cdot x_b) \tag{3.17}$$

où x_b est le vecteur d'ébauche, y le vecteur d'observations, H l'opérateur d'observation, et où $K^\star$ est une matrice nommée ***Gain de Kalman****, et qui a pour expression :*

$$K^\star = BH^T(R + HBH^T)^{-1} \tag{3.18}$$

où B et R sont les matrices de covariance associées respectivement à l'erreur d'ébauche et à l'erreur d'observations. Dans ce cas, la matrice de covariance associée à l'erreur d'analyse a pour valeur :

$$A = (I - K^\star H)B \tag{3.19}$$

La matrice de Gain de Kalman $K^\star$ et la matrice de covariance associée à l'erreur d'analyse A ne dépendent ni de x_b, ni de y. C'est un avantage indéniable : on peut ainsi calculer ces matrices avant de les appliquer à un jeu de données, ce qui permet de maîtriser le temps et le coût de calcul.

3.2.2 Le 3DVAR

L'Estimateur des Moindres Carrés Ordinaire, lequel est à l'origine de l'Assimilation de Données, suit une démarche différente du BLUE. Supposons qu'on ne dispose pas d'ébauche, et qu'on cherche à assimiler directement les observations seules : un moyen intuitif est de chercher à minimiser la fonctionnelle suivante :

$$J(x) = \frac{1}{2}(y - H[x])^T R^{-1}(y - H[x]) \tag{3.20}$$

Concrètement, on a défini une norme sur l'espace des observations :

$$\forall a \in \mathbb{R}^p, \|a\|_R^2 = a^T R^{-1} a \tag{3.21}$$

qui permet de trier les observations selon leurs précisions, et on cherche à minimiser le résidu $y - H[x]$ pour cette norme-ci.

La convergence numérique de cette minimisation dépend de la forme de l'opérateur d'observation H. Dans le cas particulièrement simple où chaque observation est une mesure

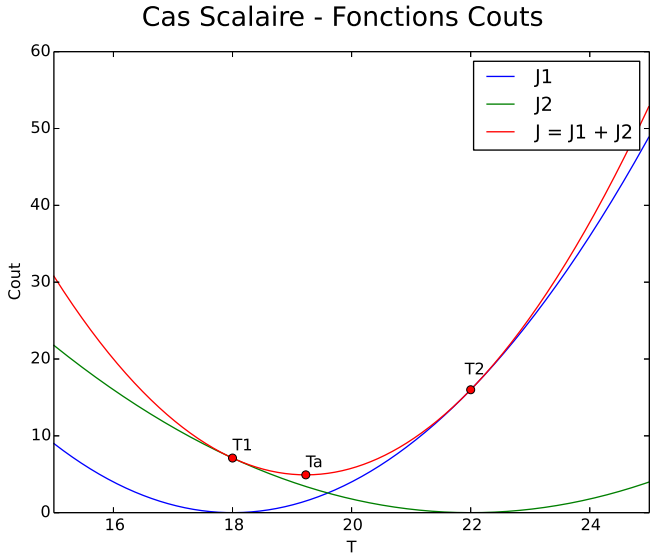


FIGURE 3.7 – Exemple de fonction coûts de type 3DVAR pour un cas scalaire

directe d'une des composantes du vecteur d'état, et où chaque composante du vecteur d'état est mesurée au moins une fois, alors la fonction J définie par (3.20) est strictement convexe et il existe bien une unique solution au problème de minimisation. Si R est diagonale (ce qui est souvent, si ce n'est systématiquement, le cas) on peut l'effectuer composante par composante :

$$x_{a,i} = argmin_{x_i} \sum_{j \in \mathcal{H}_i} \frac{(y_j - x_i)^2}{\sigma_j^2} \tag{3.22}$$

Naturellement, les composantes étant estimées séparément, rien ne garantit (comme pour l'Analyse de Cressman) que le vecteur estimé satisfasse tel ou tel critère physique de continuité, conservation, etc. ...On peut alors choisir de procéder à une minimisation sous contrainte :

$$x_a = argmin_x \|y - H[x]\|_R^2, F(x) = 0 \tag{3.23}$$

Ce type de démarche s'apparentant plus à ce qu'on appelle « Réconciliation de Données » [2].

Lorsque l'opérateur d'observation est de forme plus quelconque, ou lorsque toutes les composantes du vecteur d'état ne sont pas accessibles par la mesure, le problème de minimisation de la fonctionnelle définie à (3.20) est mal posé : il peut exister une infinité de vecteurs x qui satisferont la minimisation.

Le principe de la régularisation Tykhonoff consiste à introduire un terme supplémentaire strictement convexe dans la minimisation :

$$J(x) = \frac{1}{2}\|y - H[x]\|_R^2 + \frac{1}{2}\|x - x_b\|_B^2 \tag{3.24}$$

2. EDF R&D utilise la Réconciliation de Données dans le cadre de modèles 0D/1D dans différents contextes, par exemple pour la détection de mesures biaisées ou de capteurs défaillants (dérive physique)

où la norme $\|\cdot\|_B$ est définie sur l'Espace d'Etat de la même manière que la norme $\|\cdot\|_R$ l'a été sur l'Espace des Observations :

$$\forall x \in \mathbb{R}^n, \|x\|_B^2 = x^T B^{-1} x \tag{3.25}$$

On remarque qu'on suppose ici explicitement que la matrice B soit définie positive. L'ajout du terme d'ébauche $\|x - x_b\|_B^2$ dans la fonctionnelle de coût améliore le conditionnement du problème, permettant de trouver une solution numérique.

Minimiser la fonction $J(x)$ définie en (3.24) consiste à identifier le vecteur x qui constitue le meilleur compromis entre l'ébauche x_b et les observations y. Si on suppose à présent que l'opérateur H est linéaire, le gradient de la fonction coût s'écrit alors :

$$\nabla J(x) = B^{-1}(x - x_b) - H^T R^{-1}(y - Hx) \tag{3.26}$$

La condition d'extremum du premier ordre $\nabla J(x) = 0$ devient donc :

$$B^{-1}(x^\star - x_b) - H^T R^{-1}(y - Hx^\star) = 0 \tag{3.27}$$

qui se réécrit sous la forme :

$$x^\star = x_b + K^\circ(y - Hx_b) \tag{3.28}$$

avec :

$$K^\circ = (B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} \tag{3.29}$$

Or :

$$\begin{aligned}
K^\circ &= (B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} \\
&= (B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1}(R + HBH^T)(R + HBH^T)^{-1} \\
&= (B^{-1} + H^T R^{-1} H)^{-1}(H^T + H^T R^{-1} HBH^T)(R + HBH^T)^{-1} \\
&= (B^{-1} + H^T R^{-1} H)^{-1}(B^{-1} + H^T R^{-1} H)BH^T(R + HBH^T)^{-1} \\
&= I \cdot BH^T(R + HBH^T)^{-1} \\
&= K^\star
\end{aligned}$$

où $K^\star$ est le Gain de Kalman tel qu'obtenu lors du BLUE. Ce résultat constitue le Théorème de Sherman-Morrison-Woodbury, et nous permet d'obtenir l'équivalence entre l'approche variationnelle définie ci-dessus et le BLUE.

Sous cette formulation variationnelle, cette estimation est appelée 3DVAR.

Théorème 5 (3DVAR)**.** *Soit la fonctionnelle de coût :*

$$J(x) = \frac{1}{2}(y - H[x])^T R^{-1}(y - H[x]) + \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) \tag{3.30}$$

où x_b est le vecteur d'ébauche, y le vecteur d'observations, H l'opérateur d'observation, et B et R sont les matrices de covariance associées respectivement à l'erreur d'ébauche et à l'erreur d'observations.

Alors le 3DVAR est le vecteur x_a qui est le résultat du problème de minimisation de la fonction coût :

$$x_a = argmin_x J(x)$$

Si H est linéaire, x_a est le vecteur obtenu par la formule du BLUE, et sous les hypothèses du BLUE, constitue donc l'estimation optimale.

Remarque 3. *Dans ce cas linéaire, à partir de l'expression du gradient de J*

$$\nabla J(x) = B^{-1}(x - x_b) + H^T[R^{-1}(y - H[x])] \tag{3.31}$$

on obtient facilement l'expression de la Hessienne $\mathbb{H}_x$:

$$\mathbb{H}_x = B^{-1} + H^T R^{-1} H \tag{3.32}$$

Or, si l'on est dans le cadre des hypothèses du BLUE, la matrice de covariance d'erreur d'analyse est donnée par (3.19)*, que l'on peut transformer, d'après la formule de Sherman-Morrison-Woodbury, en :*

$$\begin{aligned} A &= (I - (B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} H) B \\ &= (B^{-1} + H^T R^{-1} H)^{-1}((B^{-1} + H^T R^{-1} H) - H^T R^{-1} H) B \\ &= (B^{-1} + H^T R^{-1} H)^{-1} \end{aligned}$$

Ce qui donne finalement :

$$A^{-1} = B^{-1} + H^T R^{-1} H \tag{3.33}$$

ce qui n'est autre que l'expression de la hessienne obtenue en (3.32)*.*

Ce résultat se comprend très simplement : la qualité de l'analyse est proportionnelle à la convexité de la fonction de coût J. Plus le minimum est piqué, meilleure est l'analyse.

Le résultat (3.33) *est l'équivalent du résultat obtenu en* (3.6) *obtenu pour le cas scalaire : si l'on nomme* ***précision*** *l'inverse des matrices de covariances d'erreurs, alors les précisions de l'ébauche et de l'observation se sont additionnées pour donner une meilleure précision qui est celle de l'analyse.*

Contrairement au BLUE, le 3DVAR peut être obtenu sans utiliser de formule matricielle et donc même si l'opérateur d'observation H n'est pas linéaire. Naturellement, si l'opérateur d'observation est particulièrement complexe, la fonctionnelle de coût J peut présenter des minima locaux. On introduit alors l'opérateur linéaire tangent ∇H_x de l'opérateur H au point x, défini par le développement limité :

$$H[x + \delta x] = H[x] + \nabla H_x \delta x \tag{3.34}$$

∇H_x est ainsi une application linéaire de l'espace d'état $\mathbb{R}^n$ dans l'espace d'observation $\mathbb{R}^p$. L'expression du gradient de J s'écrit alors :

$$\nabla J(x) = B^{-1}(x - x_b) + (\nabla H_x)^T[R^{-1}(y - H[x])] \tag{3.35}$$

Ainsi, la condition du premier ordre donne un résultat identique au cas linéaire en remplaçant H par $\nabla H_{x^\star}$:

$$x^\star = x_b + (B^{-1} + \nabla H_{x^\star}^T R^{-1} \nabla H_{x^\star})^{-1} \nabla H_{x^\star}^T R^{-1} (y - H[x_b]) \tag{3.36}$$

On notera que cette équation ne donne pas explicitement $x^\star$, car l'opérateur linéaire tangent dépend a priori de lui.

3.2.3 Le Filtre de Kalman et le 4DVAR

L'Analyse de Cressman, le BLUE et le 3DVAR sont tous des algorithmes statiques : à un instant donné, on corrige notre ébauche avec les observations disponibles. Et jusqu'à présent nous n'avons pas introduit de dimension temporelle à notre problème d'estimation. Néanmoins, la grande généralité des problèmes typiquement traités par l'Assimilation de Données (météorologie, géophysique, balistique, etc.) comportent une composante dynamique : on cherche non seulement à réaliser une analyse en temps réel, mais aussi à effectuer une succession d'analyses, au fur et à mesure de l'évolution de la situation. Les données disponibles se répartissent en effet dans le temps, et l'état correspondant à ces données évolue.

On échantillonne ainsi le problème d'estimation dans le temps, en définissant une succession de dates $t_0, t_1, \ldots, t_k, \ldots, t_n$ séparant les différentes dates auxquelles l'état doit être à chaque fois estimé : et on suppose l'existence d'un modèle régissant l'évolution temporelle de l'état :

$$x_{t,k+1} = M_{k+1}[x_{t,k}] + \epsilon_{m,k} \tag{3.37}$$

où, $\forall k$:

— $x_{t,k}$ est le vecteur d'état réel à l'instant t_k ;

— M_{k+1} est le modèle d'évolution qui régit l'évolution du système entre les instants t_k et t_{k+1} ;

— $\epsilon_{m,k}$ est l'erreur associée à ce modèle.

On suppose systématiquement que ces erreurs modèles sont non-biaisées, décorrélées dans le temps, et de matrices de covariances notées Q_k, de sorte que :

$$\forall k, E[w_k] = 0 \tag{3.38}$$

et

$$\forall k, l, E[w_k w_l^T] = Q_k \delta_{k,l} \tag{3.39}$$

$\delta_{k,l}$ étant le symbole de Kronecker.

L'opérateur d'observation peut lui aussi varier dans le temps (par exemple, parce que toutes les informations ne sont pas toujours disponibles au même moment) :

$$y_k = H_k[x_{t,k}] + \epsilon_{obs,k} \tag{3.40}$$

où, $\forall k$:

— y_k sont les observations correspondant à l'instant t_k ;

— H_k est l'opérateur d'observation correspondant à l'instant t_k ;

— $\epsilon_{obs,k}$ est l'erreur d'observation.

On supposera de même que pour les erreurs modèles, que les erreurs d'observation sont décorrélées dans le temps, et de matrice de covariance notées R_k.

En se replaçant dans les hypothèses du BLUE (opérateurs (d'observation H et d'évolution M) linéaires, erreurs non biaisées et décorrélées entre elles, analyse recherchée comme combinaison linéaire des observations et de l'ébauche initiale, trace minimale de la covariance d'analyse), on obtient alors son équivalent dynamique, le Filtre de Kalman Linéaire, qui est défini séquentiellement de la façon suivante :

Théorème 6 (Filtre de Kalman Linéaire)**.**

1. ***Initialisation*** *: La prédiction initiale est $x_{f,k}$ (l'indice « f »est pour **forecast**, prévision) et la matrice de covariance d'erreur qui lui est associée est $P_{f,0} = B$.*
2. *Si pour t_k, on dispose d'une prédiction $x_{f,k}$ et de sa matrice de covariance d'erreur associée $P_{f,k}$, et d'un vecteur d'observations y_k, alors :*
   — ***Analyse**, calcul du Gain de Kalman :*

$$K_k^\star = P_{f,k} H_k^T (H_k P_{f,k} H_k^T + R_k)^{-1} \tag{3.41}$$

   — *calcul de l'estimé de l'analyse :*

$$x_{a,k} = x_{f,k} + K_k^\star (y_k - H_k x_{f,k}) \tag{3.42}$$

   — *calcul de la matrice de covariance d'erreur d'analyse :*

$$P_{a,k} = (I - K_k^\star H_k) P_{f,k} \tag{3.43}$$

   — ***Prévision**, calcul de l'estimé de prévision :*

$$x_{f,k+1} = M_{k+1} \cdot x_{a,k} \tag{3.44}$$

$$P_{f,k+1} = M_{k+1} P_{a,k} M_{k+1}^T + Q_k \tag{3.45}$$

Dans le cas où toutes les erreurs sont des bruits blancs, les vecteurs $x_{a,k}$ sont les meilleures estimations possibles étant donné x_b et toutes les observations y_l, pour $l \leq k$.

De même que pour le BLUE, il existe un pendant variationnel pour le Filtre de Kalman (ou plutôt pour une de ses variantes qui est le Lisseur de Kalman) : le 4DVAR. On ne cherche plus à recaler l'état à chaque instant, mais à lisser une trajectoire en tenant compte des observations disponibles dans la fenêtre d'assimilation. Chaque état étant recalé en fonction des observations présentes, passées et futures, le 4DVAR n'est pas formellement équivalent au BLUE. On suppose cette fois-ci que les modèles M_k sont parfaits et on note $M_{0\to k} = M_k \circ M_{k-1} \cdots \circ M_0$ le modèle permettant de déduire l'état à l'instant t_k de l'état à l'instant t_0.

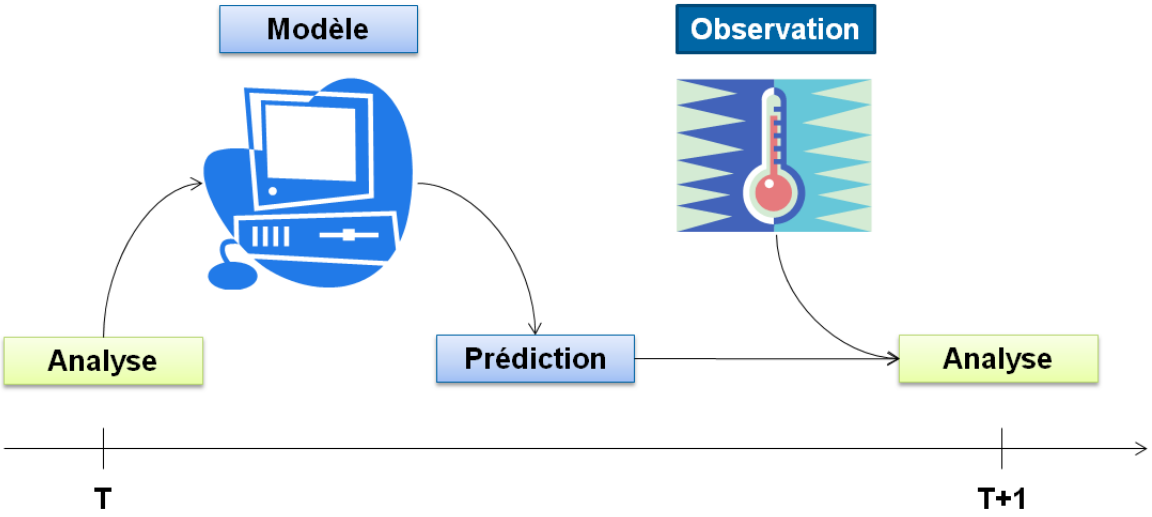


FIGURE 3.8 – Schéma de principe du Filtre de Kalman

Théorème 7 (4DVAR).
On définit la fonctionnelle de coût suivante :

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}\sum_{k=0}^{K}(y_k - H_k[M_{0\to k}])^T R_k^{-1}(y_k - H_k[M_{0\to k}]) \quad (3.46)$$

Alors l'analyse 4DVAR utilisant les observations $y_0, \ldots, y_k$ est définie par :

$$\forall k, x_{a,k} = M_{0\to k}[ArgminJ(x)] \quad (3.47)$$

De même que pour la méthode 3DVAR, la minimisation de la fonction coût du 4DVAR nécessite le calcul de son gradient, ce qui requiert à la fois le calcul des adjoints des opérateurs d'observations H_k et ceux des opérateurs d'évolutions $M_{0\to k}$, mais ne requiert pas obligatoirement leur linéarité, contrairement au Filtre de Kalman Linéaire. D'autre part, $x_{a,k}$ sera le vecteur le plus probable étant donné l'ébauche initiale x_b et toutes les informations disponibles y_l : présentes, passées, mais aussi futures.

L'un des avantages du Filtre de Kalman Linéaire réside en sa facilité d'implémentation et son faible coût de calcul. Les matrices de Gain de Kalman $K_k^\star$ et les matrices de covariances d'erreurs $P_{a,k}$ et $P_{f,k}$ ne dépendent pas de la valeur des vecteurs d'état ou d'observations, uniquement des matrices R_k, H_k, M_k et B, qui sont des paramètres d'entrée du modèle. Il est donc possible d'effectuer séparément tous les calculs matriciels – qui sont les plus coûteux en termes de stockage et de temps de calcul – avant d'appliquer le filtre à une suite de données réelles. Dans un cadre de prévision en temps réel, c'est un avantage précieux.

Il existe un très grand nombre de variantes des algorithmes que nous venons de décrire, particulièrement pour le Filtre de Kalman. La plupart du temps, ces variantes visent à passer outre les hypothèses les plus contraignantes comme celle de linéarité des opérateurs.

3.2.4 Le Filtre de Kalman d'Ensemble

Le Filtre de Kalman d'Ensemble (EKF) a été proposé par Geir Evensen en 1994 et corrigé en 1998 [Eve09]. Il s'inspire du principe des filtres particulaires et l'adapte à l'approche séquentielle du Filtre de Kalman : plutôt que de propager l'estimation et l'incertitude correspondante sous la forme d'un vecteur et d'une matrice de covariance d'erreur associée, on utilise ici un nuage de points qui sont propagés grâce au modèle, sans linéarisation.

La justification semi-empirique du Filtre de Kalman d'Ensemble comporte des points de détails un peu déroutants, mais la séquence de l'algorithme ressemble beaucoup à celui du Filtre de Kalman.

Théorème 8 (Filtre de Kalman d'Ensemble)**.**

1. ***Initialisation*** *: La prédiction initiale est $x_{f,k}$ (l'indice « f »est pour* ***forecast****, prévision) et la matrice de covariance d'erreur qui lui est associée est $P_{f,0} = B$.*
2. *Si pour t_k, on dispose d'une prédiction $x_{f,k}$ et de sa matrice de covariance d'erreur associée $P_{f,k}$, et d'un vecteur d'observations y_k, alors :*
   - ***Observation****, création d'un jeu d'observations perturbées autour de l'observation y_k :*
   $$\forall j = 1 \dots N, z_j = y_k + u_j \tag{3.48}$$
   où les perturbations u_j correspondent à des erreurs d'observations simulées, sans biais et de matrice de variance-covariance R_k :
   $$\begin{aligned} \sum_{j=1}^{N} u_j &= 0 \\ \frac{1}{N-1} \sum_{j=1}^{n} u_j u_j^T &= R_k \end{aligned}$$
   - ***Analyse****, calcul du Gain de Kalman :*
   $$K_k^\star = P_{f,k} H_k^T (H_k P_{f,k} H_k^T + R_k)^{-1} \tag{3.49}$$
   - *calcul des estimés de l'analyse pour chacune des observations perturbées :*
   $$x_a^j = x_{f,k} + K_k^\star (z_j - H_k x_{f,k}) \tag{3.50}$$
   - *calcul de l'analyse, comme moyenne des estimés d'analyse :*
   $$x_{a,k} = \frac{1}{N} \sum_{j=1}^{N} x_a^j \tag{3.51}$$
   - *calcul de la matrice de covariance d'erreur d'analyse, comme moment statistique d'ordre 2 des estimés d'analyse :*
   $$P_{a,k} = \frac{1}{N-1} \sum_{j=1}^{N} (x_a^j - x_{a,k})(x_a^j - x_{a,k})^T \tag{3.52}$$

— ***Prévision***, *calcul des estimés de prévision :*

$$x_f^j = M_{k+1}[x_a^j] \tag{3.53}$$

— *calcul de la prévision, comme moyenne des estimés de prévision :*

$$x_{f,k} = \frac{1}{N}\sum_{j=1}^{N} x_f^j \tag{3.54}$$

— *calcul de la matrice de covariance d'erreur de prévision, comme moment statistique d'ordre 2 des estimés de prévision :*

$$P_{f,k} = \frac{1}{N-1}\sum_{j=1}^{N}(x_f^j - x_{f,k})(x_f^j - x_{f,k})^T \tag{3.55}$$

Le Filtre de Kalman d'Ensemble utilise des observations perturbées pour représenter l'incertitude comme un nuage de points, ce qui s'apparente ainsi à une méthode de Monte-Carlo. L'un des avantages principaux est qu'il n'est pas besoin de supposer une linéarité des opérateurs d'évolution M_k, ni de calculer leurs adjoints, ce qui était nécessaire pour propager dans le temps les matrices de covariance.

3.3 Construction des matrices de covariance

3.3.1 Enjeux

Au centre des techniques d'Assimilation de Données se trouve la notion de quantification des incertitudes, qui s'incarne dans les matrices de covariances (B, R et A pour le BLUE et le 3DVAR ; $P_{f,k}$, R_k, Q_k et $P_{a,k}$ pour le Filtre de Kalman Linéaire...). La plupart sont des paramètres d'entrée des algorithmes d'Assimilation de Données décrits à la section précédente, et il est donc nécessaire de les construire.

Le rôle des matrices de covariance est double :

— les termes diagonaux contiennent les variances des erreurs composantes, et sont donc une indication de l'incertitude de chaque paramètre : elles permettent donc de comparer entre elles des estimations, et déterminent ainsi la pondération à effectuer pour obtenir l'analyse ;

— les termes non diagonaux apportent des informations sur les corrélations entre erreurs composantes, et permettent aux informations de « diffuser » une correction à effectuer sur un ensemble de composantes du vecteur d'état (via la matrice de Gain de Kalman ou via une fonction de coût).

Le choix de R et B, dans le cas du BLUE ou du 3DVAR, est donc tout sauf anodin. Il influe directement sur le résultat qui sera produit par l'analyse, et il est donc obligatoire de bien visualiser et contrôler les hypothèses qui leur sont sous-jacentes.

Pour commencer, il convient de rappeler que les termes diagonaux sont égaux au variances composantes :

$$R_{i,i} = \sigma^2_{obs}(i) = E[(y(i) - H_i[x_t])^2] \tag{3.56}$$

$$B_{i,i} = \sigma^2_b(i) = E[(x_b(i) - x_t(i))^2] \tag{3.57}$$

Reste à interpréter et construire la forme de ces matrices, c'est-à-dire les matrices de corrélation correspondantes.

3.3.2 Méthodes Empiriques

L'une des difficultés vient du fait qu'on n'a jamais accès directement au vecteur d'état réel x_t. Au mieux, un jeu de données réelles peut nous donner accès à un échantillon de vecteurs d'innovations $y - H[x_b]$, dont on pourra étudier la statistique :

$$\begin{aligned} \epsilon &= E[y - H[x_b]] \\ C &= E[(y - H[x_b])(y - H[x_b])^T] \end{aligned}$$

En théorie, si H est linéaire :

$$\begin{aligned} \epsilon &= E[(y - Hx_t - H(x_b - x_t)) \\ &= E[\epsilon_{obs}] - HE[\epsilon_b] \\ &= 0 \end{aligned}$$

Si notre vecteur d'ébauche a été correctement construit et que l'erreur d'ébauche est non biaisée, et que les observations ont été correctement recalées de sorte à être également non biaisées, alors on devrait on devrait observer un moment d'ordre 1 ϵ nul pour le vecteur d'innovation.

$$\begin{aligned} C &= E[(y - Hx_t - H(x_b - x_t))(y - Hx_t - H(x_b - x_t))^T] \\ &= E[(\epsilon_{obs} - H\epsilon_b)(\epsilon_{obs} - H\epsilon_b)^T] \\ &= E[\epsilon_{obs}\epsilon^T_{obs}] + HE[\epsilon_b\epsilon^T_b] - HE[\epsilon_b\epsilon^T_{obs}] + E[\epsilon_{obs}\epsilon^T_b]H^T \\ &= R + HBH^T \end{aligned}$$

car les termes croisés sont nuls par décorrélation des erreurs ϵ_b et ϵ_{obs}. A supposer – comme on le fait classiquement – que la matrice R est diagonale, les termes non diagonaux observés sur le moment d'ordre 2 du vecteur d'innovation (la matrice C) devraient donc être attribuables à la matrice HBH^T. En revanche, chaque terme diagonal étant la somme de celui correspondant de R et celui correspondant de HBH^T, il resterait à séparer le terme attribuable à HBH^T et celui attribuable à R.

Ceci effectué, il resterait également à reconstruire une matrice B à partir de HBH^T.

Il existe des méthodes qui permettent de reconstruire empiriquement B et R à partir de données réelles. En météorologie, on utilise ainsi une méthode observationnelle (dite de Hollingworth-Lonnberg), qui repose sur un raisonnement spatial semblable à celui employé pour l'Analyse de Cressman, partant du principe que deux points proches dans l'espace sont fortement corrélés pour retrouver la part du terme diagonal attribuable à l'ébauche.

Ensuite, la matrice B est reconstruite suivant des raisonnements complexes, subtils et sophistiqués, basés sur la physique du système et visant à respecter des contraintes mathématiques et computationnelles strictes. On peut citer également des méthodes dites ALS (Autocovariance Least-Squares) qui utilisent des données opérationnelles pour proposer des matrices de covariances de bruit déduites de modèles de contrôle prédictifs ; ou encore des méthodes utilisant des algorithmes de krigeage issus de la géostatistique.

Ces méthodes sophistiquées ont été généralement conçues pour des problèmes bien spécifiques, souvent associés aux géosciences, et utilisant la plupart du temps des raisonnement de spatialité qui excluent de fait nombre de problèmes d'estimation plus généraux.

3.3.3 Principe de la Méthode de Monte-Carlo

Il est classique de prendre R diagonale : construire R revient donc à déterminer la variance associée à chaque instrument de mesure utilisé. Nous nous intéresserons par la suite essentiellement à la construction de la matrice de covariance d'erreur d'ébauche B.

Comme vu dans l'équation (3.13), x_b et B sont des termes statistiques qui nous servent à quantifier l'incertitude associée à l'ébauche, c'est-à-dire la distribution de probabilité associée à l'erreur commise. Concrètement, en supposant que cette distribution de probabilité soit gaussienne, cela revient à considérer x_t comme un vecteur aléatoire dont nous approximons la distribution de probabilité par une loi gaussienne d'espérance x_b et de covariance B :

$$x_t = NG(x_b, B) \tag{3.58}$$

Construire les termes d'ébauche revient ainsi à utiliser notre connaissance *a priori* de la physique du système pour donner une estimation de la distribution de probabilité de x_t. On suppose disposer d'un modèle de type boîte noire qui nous permette de reconstituer le vecteur x_t à partir d'un certain nombre de paramètres réunis dans un vecteur u_t :

$$x_t = \Phi[u_t] \tag{3.59}$$

On peut ainsi mettre en place une méthode de Monte-Carlo, qui s'inspire des Méthodes d'Ensemble décrites à la fin de la section précédente, en randomisant les paramètres :

1. on simule un échantillon de vecteurs paramètres $\{u_t\}$ (ce qu'on suppose naturellement plus facile que de simuler directement des vecteurs d'état) ;
2. on utilise notre modèle pour calculer un échantillon correspondant de vecteurs d'état $\{x_t\} = \{\Phi[u_t]\}$;
3. on calcule les moments d'ordre 1 et 2 associés à cet échantillon :

$$\begin{aligned} x_b &= \frac{1}{N}\sum_{t=1}^{N} x_t \\ B &= \frac{1}{N-1}\sum_{t=1}^{N}(x_t - x_b)\cdot(x_t - x_b)^T \end{aligned}$$

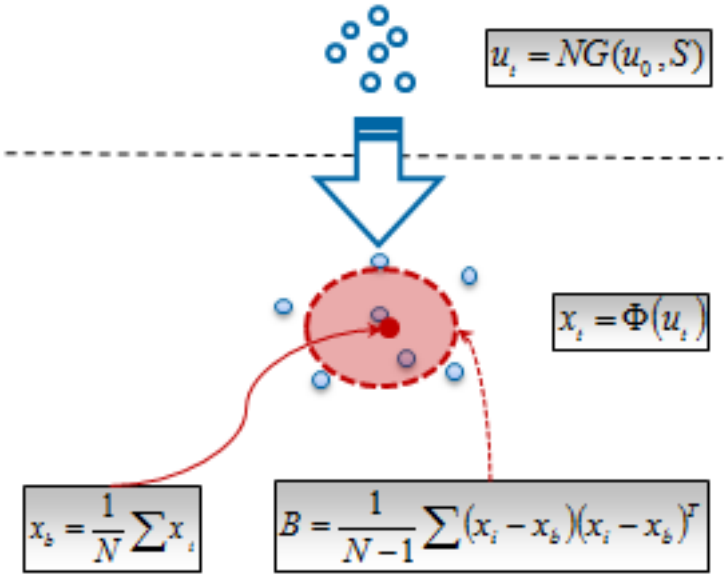


FIGURE 3.9 – Schéma de principe de la méthode de Monte-Carlo pour le calcul des termes d'ébauche

Remarque 4.
La taille N de l'échantillon nécessaire peut être grande, et stocker la totalité de l'échantillon peut s'avérer laborieux. Un moyen de l'éviter est de procéder à une moyenne par étapes :

$$\begin{aligned} m_0 &= 0 \\ m_{t+1} &= \frac{t.m_t + x_t}{t+1} \end{aligned}$$

m_t est ainsi la moyenne des t premiers vecteurs de l'échantillon, et $x_b = m_N$;

$$\begin{aligned} M_0 &= 0 \\ M_{t+1} &= \frac{1}{t+1}[t.M_t + (x_t - x_b)(x_t - x_b)^T] \end{aligned}$$

$\frac{t}{t-1}M_t$ est ainsi, pour $t > 1$, l'estimation sans biais de la variance associée aux t premiers vecteurs de l'échantillon, et $B = \frac{N}{N-1}M_N$.

3.3.4 Cas Linéaire

Expression littérale
Supposons que notre modèle boîte-noire Φ soit linéaire. Dans ce cas, en supposant que le vecteur de paramètres suit la loi $NG(u_0, S)$, on a exactement :

$$\begin{aligned} x_b &= \Phi u_0 \\ B &= \Phi S \Phi^T \end{aligned}$$

et la distribution de x_t (et de ϵ_b, par conséquent) est également gaussienne.

Forme Spectrale de B
Suivant les dimensions relatives de l'Espace d'Etat (de dimension n) où évolue x, et de l'Espace des Paramètres (dont on notera m la dimension) où évolue u, et selon la forme spectrale de Φ, B peut très bien être singulière, même si S est inversible.

S est symétrique positive : il existe donc une matrice L telle que $S = LL^T$. Donc, $B = \Phi LL^T\Phi^T$. On utilise alors la décomposition en valeurs singulières (SVD) de la matrice ΦL : il existe deux matrices unitaires U de taille $n \times n$ et V de taille $m \times m$ telle que :

$$\Phi L = U\Sigma V^T \tag{3.60}$$

où la matrice Σ est de taille $n \times m$, avec des coefficients diagonaux positifs ou nuls et des coefficients non diagonaux nuls, et de rang égal à celui de ΦL. Alors :

$$\begin{aligned} B &= U\Sigma V^T(U\Sigma V^T)^T \\ &= U\Sigma\Sigma^T U^T \end{aligned}$$

On peut facilement vérifier que $\Sigma\Sigma^T$ est de taille $n \times n$, diagonale et de rang égal à celui de Σ, c'est-à-dire à celui de ΦL. En particulier, on a le résultat suivant :

Théorème 9.
Si la matrice S est inversible (ce qui est le cas si tous les paramètres ont été randomisés de façon indépendante), alors le rang de B est exactement celui de Φ.

B est donc nécessairement non inversible si la dimension de l'Espace des Paramètres est inférieure à celle de l'Espace d'Etat. Ceci s'explique par le fait que le vecteur $x_t - x_b$ évolue dans $Im(\Phi)$, qui est un sous-espace strict de l'Espace d'Etat.

Exemple 3.
Prenons pour cas d'application un bilan enthalpique linéarisé, celui aux bornes du générateur de vapeur, par exemple :

$$PTHGV = Q.C_p.(TBC - TBF) \tag{3.61}$$

En prenant pour paramètre :

$$u = \begin{pmatrix} TBF \\ TBC \end{pmatrix}$$

et pour vecteur d'état :

$$x = \begin{pmatrix} TBF \\ TBC \\ PTHGV \end{pmatrix}$$

on a alors transformé notre équation en un modèle direct Φ, avec :

$$\Phi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -Q.C_p & Q.C_p \end{pmatrix}$$

On peut alors randomiser le modèle en posant $TBF = TBF_0 + \sigma_F \epsilon_F$ *et* $TBC = TBC_0 + \sigma_C \epsilon_c$*, avec* ϵ_F *et* ϵ_C *deux variables aléatoires indépendantes de lois centrées réduites. Autrement dit :*

$$u_0 = \begin{pmatrix} TBF_0 \\ TBC_0 \end{pmatrix} \quad , \quad S = \begin{pmatrix} \sigma_F^2 & 0 \\ 0 & \sigma_C^2 \end{pmatrix}$$

On obtient alors, avec notre méthode, pour termes d'ébauche :

$$x_b = \begin{pmatrix} TBF_0 \\ TBC_0 \\ Q.C_p(TBC_0 - TBF_0) \end{pmatrix} \quad , \quad B = \begin{pmatrix} \sigma_F^2 & 0 & -Q.C_p\sigma_F^2 \\ 0 & \sigma_C^2 & Q.C_p\sigma_C^2 \\ -Q.C_p\sigma_F^2 & -Q.C_p\sigma_C^2 & Q^2.C_p^2(\sigma_F^2 + \sigma_c^2) \end{pmatrix}$$

En simplifiant et adimensionnant le problème ($\sigma_F = \sigma_C = 1$*,* $Q.C_p = 1$*), le modèle se résume à :*

$$x = \begin{pmatrix} TBF \\ TBC \\ TBC - TBF \end{pmatrix}$$

soit

$$\Phi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}$$

Et la matrice B *s'écrit alors :*

$$B = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & 2 \end{pmatrix}$$

B *est bien symétrique, non inversible et de rang 2 (valeurs propres* 0, 1 *et* 3*) :* $Im(\Phi)$ *est tout simplement le plan de* $\mathbb{R}^3$ *d'équation* $x_3 = x_2 - x_1$*, qui est bien de dimension 2.*

Conséquence pour le BLUE La formule du BLUE ne requiert pas l'inversion de B, seulement celle de $R + HBH^T$, qui est toujours inversible dès le moment où R au moins est prise inversible.

On remarque que dans ce cas, la formule du BLUE donne :

$$\begin{aligned} x_a &= x_b + BH^T(R + HBH^T)^{-1} \\ &= \Phi(u_0 + S\Phi H^T(R + HBH^T)^{-1}) \end{aligned}$$

donc $x_a \in Im(\Phi)$. Autrement dit :

Théorème 10. *Dans le cas où le modèle physique utilisé est linéaire, l'information contenue dans la matrice de covariance d'ébauche obtenue par la méthode Monte-Carlo est suffisante pour que l'analyse BLUE produise systématiquement une estimation qui soit dans l'image du modèle, c'est-à-dire qui vérifie ses équations, et soit donc physiquement cohérent.*

Exemple 4. *On reprend l'exemple du bilan enthalpique adimensionné. On suppose que TBF, TBC et $PTHGV$ sont tous mesurés directement, de sorte que $H = I_3$, et on suppose pour simplifier que $R = I_3$. Dans ce cas, la formule de Kalman donne :*

$$K^\star = B(I_3 + B)^{-1} = \frac{1}{8}\begin{pmatrix} 3 & 1 & -2 \\ 1 & 3 & 2 \\ -2 & 2 & 4 \end{pmatrix}$$

L'Analyse BLUE donne ainsi pour estimation :

$$TBF_a = TBF_0 + \frac{3}{8}(TBF_{mes} - TBF_0) + \frac{1}{8}(TBC_{mes} - TBC_0) - \frac{2}{8}(PTHGV_{mes} - (TBC_0 - TBF_0))$$

$$TBC_a = TBC_0 + \frac{1}{8}(TBF_{mes} - TBF_0) + \frac{3}{8}(TBC_{mes} - TBC_0) + \frac{2}{8}(PTHGV_{mes} - (TBC_0 - TBF_0))$$

$$PTHGV_a = PTHGV_0 + \frac{-2}{8}(TBF_{mes} - TBF_0) + \frac{2}{8}(TBC_{mes} - TBC_0) + \frac{4}{8}(PTHGV_{mes} - (TBC_0 - TBF_0))$$

On remarque :
- *qu'effectivement le vecteur analysé vérifie bien le modèle ($PTHGV_a = TBC_a - TBF_a$) ;*
- *qu'ici pour chaque composante, la correction la plus importante est induite par la mesure directe (toutes les mesures étant de précisions égales).*

Conséquence pour le 3DVAR En ce qui concerne le 3DVAR en revanche, on ne peut reprendre telle quelle la fonction de coût définie par (3.30), car elle nécessite explicitement l'inversibilité de B. Une première solution consiste à procéder à une perturbation du modèle.

Notre modèle, qui reconstitue notre vecteur d'état à partir d'un certain nombre de paramètres de départ et d'équations, est, comme tout modèle, imparfait. Certains paramètres que nous avions considérés comme constants dans notre méthode de Monte-Carlo (par exemple, dans notre cas d'application, QC_p) sont en fait légèrement variables autour de leurs moyennes ; certaines équations peuvent également n'être qu'approximativement vérifiées (peut-être y-a-t-il des fuites d'énergie qui font que $P = QC_p\Delta T$ n'est pas tout à fait exact : ou on peut se rappeler qu'écrire $\Delta H = C_p\Delta T$ constitue également, en soi, une approximation). On peut ainsi choisir de pénaliser notre modèle en augmentant la dimension de l'Espace des Paramètres.

Concrètement, on peut rajouter une petite perturbation de type « bruit blanc » au vecteur calculé par le modèle :

$$x = \Phi(u) + v \tag{3.62}$$

où v est un vecteur aléatoire gaussien de taille n, d'espérance nulle et de matrice de covariance définie positive S_v : on suppose $\|S_v\| \ll \|B\|$. De cette manière, en posant $w = (u, v)$, on définit un nouveau modèle numérique par :

$$x = \Psi(w) = \Phi(u) + v \tag{3.63}$$

L'opérateur Ψ défini de cette manière est linéaire si Φ l'est aussi. Dans ce cas, on a naturellement toujours pour nouveau vecteur d'ébauche $x'_b = \Phi(u_0) = x_b$, et par ailleurs, on montre facilement :

$$B' = \Phi S \Phi^T + S_v = B + S_v \tag{3.64}$$

S_v étant définie positive, et B étant positive, la matrice B' est donc définie positive, et le 3DVAR est ainsi applicable. Si l'on est dans le cadre des hypothèses du BLUE (opérateur H linéaire, etc.), la formule de Sherman-Morrison-Woodbury nous permet de nous ramener au cas du BLUE appliqué avec les termes d'ébauche x'_b et B'. Dans les deux cas, l'analyse s'écrit donc :

$$\begin{array}{rcl} x'_a & = & x'_b + (B+S_v)H^T + (R + HBH^T + HS_vH^T)^{-1} \\ & = & x_b + (B+S_v)H^T(R+HBH^T)^{-1}(I + (R+HBH^T)^{-1}HS_vH^T)^{-1} \\ & \underset{S_v \to 0}{\longrightarrow} & x_b + BH^T(R+HBH^T)^{-1} = x_a \end{array}$$

Ainsi, dans ce cas linéaire, en perturbant très légèrement le modèle Φ, afin d'obtenir une matrice de covariance d'erreur d'ébauche B' inversible, on peut appliquer un 3DVAR qui donnera un résultat quasi-identique au BLUE pratiqué avec la matrice B non-inversible.

Changement de Variable Une autre solution consiste à changer d'espace d'état, et d'assimiler directement sur l'espace des paramètres, en absorbant le modèle physique dans l'opérateur d'opération. Concrètement, on prendra ainsi :

— u_0 comme vecteur d'ébauche ;

— S comme matrice de covariance d'erreur d'ébauche ;

— y comme observations ;

— $H \circ \Phi$ comme opérateur d'obervations ;

— R comme matrice de covariance d'erreur d'observations.

La fonctionnelle coût du 3DVAR s'écrivant ainsi :

$$J'(u) = \frac{1}{2}(u-u_0)S^{-1}(u-u_0)^T + \frac{1}{2}(y - H[\Phi(u)])R^{-1}(y - H[\Phi(u)])^T \tag{3.65}$$

en supposant que la matrice S soit inversible.

La formule de Kalman nous donne, en supposant Φ et H linéaire :

$$\begin{array}{rcl} u_a & = & S(H\Phi)^T(R + (H\Phi)S(H\Phi)^T)^{-1}(y - (H\Phi)u_0) \\ & = & S\Phi^T H^T(R + H(\Phi S \Phi^T)H^T)^{-1}(y - H(\Phi u_0) \\ & = & S\Phi^T H^T(R + HBH^T)^{-1}(y - Hx_b) \end{array}$$

et, par conséquent,

$$\begin{aligned}\Phi u_a &= \Phi S \Phi^T H^T (R + HBH^T)^{-1}(y - Hx_b) \\ &= BH^T(R + HBH^T)^{-1}(y - Hx_b) \\ &= x_a\end{aligned}$$

Le changement de variable n'a donc pas modifié le résultat de l'analyse.

Théorème 11. *Dans le cas où le modèle physique et l'opérateur d'observation sont linéaires, assimiler directement sur les paramètres est équivalent à assimiler dans l'espace d'état avec un terme d'ébauche obtenu par la méthode de Monte-Carlo.*

Exemple 5. *Toujours dans l'exemple du bilan enthalpique adimensionné, comme $H = I_3$, $H \circ \Phi$ est simplement égal à Φ. Dans ce cas, la fonction coût $J(u)$ s'écrit :*

$$\begin{aligned}J(TBF, TBC) = \frac{1}{2} \quad [\quad & (TBF - TBF_0)^2 + (TBC - TBC_0)^2 + (TBF - TBF_{mes})^2 \\ & + (TBC - TBC_{mes})^2 + ((TBC - TBF) - PTHGV_{mes})^2]\end{aligned}$$

La nouvelle matrice de Kalman est de format 2×3 et s'écrit :

$$K^\star = I_2 \Phi^T (I_3 + \Phi\Phi^T)^{-1} = \frac{1}{8}\begin{pmatrix} 3 & 1 & -2 \\ 1 & 3 & 2 \end{pmatrix}$$

L'analyse BLUE (ou le 3DVAR) sur l'espace des paramètres donne donc pour résultat :

$$TBF_a = TBF_0 + \frac{3}{8}(TBF_{mes} - TBF_0) + \frac{1}{8}(TBC_{mes} - TBC_0) - \frac{2}{8}(PTHGV_{mes} - (TBC_0 - TBF_0)$$

$$TBC_a = TBC_0 + \frac{1}{8}(TBF_{mes} - TBF_0) + \frac{3}{8}(TBC_{mes} - TBC_0) + \frac{2}{8}(PTHGV_{mes} - (TBC_0 - TBF_0)$$

et on retrouve donc bien le même résultat qu'avec l'analyse BLUE sur l'espace d'état effectuée précédemment.

L'avantage d'assimiler directement sur les paramètres est qu'on travaille généralement dans un espace de dimension beaucoup plus petit, ce qui allège notamment les coûts de stockage des différentes matrices intervenant. D'autre part, la formulation variationnelle donnée par (3.65) montre explicitement les rôles respectifs de l'opérateur d'observation H et du modèle physique Φ, sans les « dématérialiser » à travers les matrices B ou K : on voit donc directement nos critères de recherche.

Néanmoins, la minimisation de la fonctionnelle coût (3.65) nécessitera le calcul de l'adjoint de $H \circ \Phi$, ce qui, dans le cas de modèles compliqués, peut s'avérer coûteux et difficile. En calculant préalablement la matrice B, on peut éviter d'avoir à appeler le modèle en cours d'assimilation : les informations contenues dans le modèle sont résumées dans la forme spectrale de B, et, en un sens, l'étape du calcul de l'adjoint de Φ correspondra donc au calcul de B (ce qu'on peut visualiser, dans le cas linéaire, par la présence de Φ^T dans $B = \Phi S \Phi^T$).

3.3.5 Cas non linéaire

Lorsque le modèle physique Φ utilisé dans la méthode de Monte-Carlo n'est pas linéaire, on peut, dans le cas de très petites variations par rapport à l'ébauche, effectuer un développement limité à l'ordre 1 autour de u_0 :

$$\Phi(u + \delta u) = \Phi(u_0) + \nabla\Phi_{u_0} \cdot \delta u + o(\|\delta u\|) \tag{3.66}$$

De sorte que :

$$x_b \approx \Phi(u_0) \tag{3.67}$$

$$B \approx \nabla\Phi_{u_0} S \nabla\Phi_{u_0}^T \tag{3.68}$$

De sorte que, en supposant H toujours linéaire, la formule de Kalman donne :

$$x_a = \Phi(u_0) + \nabla\Phi_{u_0} \cdot (S\nabla\Phi_{u_0}^T H^T (R + H\nabla\Phi_{u_0} S \nabla\Phi_{u_0}^T H^T)^{-1}) \tag{3.69}$$

Autrement dit, x_a se trouvera nécessairement dans l'espace linéaire tangent.

D'autre part, on peut appliquer directement un 3DVAR dans l'espace des paramètres, en prenant pour opérateur d'observation (non linéaire) l'opérateur $\mathcal{H} = H \circ \Phi$. L'opérateur linéaire tangent correspondant s'écrivant :

$$\nabla\mathcal{H}_u = H \circ \nabla\Phi_u \tag{3.70}$$

la condition du premier ordre (3.36) donne donc, en supposant que la fonction $u \to \nabla\Phi_u$ varie peu au voisinage de u_0,

$$u^\star - u_0 = (S^{-1} + \nabla\Phi_{u_0}^T H^T R^{-1} H \nabla\Phi_{u_0})^{-1} \nabla\Phi_{u_0}^T H^T R^{-1} (y - H[x_b]) \tag{3.71}$$

de sorte que la formule de Sherman-Morrisson-Woodbury donne :

$$u^\star - u_0 = S\nabla\Phi_{u_0}^T H^T (R + H\nabla\Phi_{u_0} S \nabla\Phi_{u_0}^T H^T)^{-1} (y - H[x_b]) \tag{3.72}$$

En multipliant par $\nabla\Phi_{u_0}$, en ajoutant $\Phi(u_0)$ et en reprenant le développement limité (3.66) en u_0, on obtient finalement :

$$\Phi(u^\star) \approx \Phi(u_0) + \nabla\Phi_{u_0} S (\nabla\Phi_{u_0}^T H^T (R + H\nabla\Phi_{u_0} S \nabla\Phi_{u_0}^T H^T)^{-1} (y - H[x_b]) \tag{3.73}$$

on retrouve donc bien le résultat de l'analyse BLUE effectuée avec les termes d'ébauche obtenus par la méthode de Monte-Carlo.

Théorème 12. *Lorsque le modèle physique Φ n'est pas linéaire, mais qu'on suppose de petites variations des paramètres autour de leur valeur de référence u_0, suivant une loi statistique de matrice de covariance S, alors on peut se ramener au cas linéaire en utilisant pour termes d'ébauche :*

$$x_b = \Phi(u_0) \tag{3.74}$$

$$B = \nabla\Phi_{u_0} S \nabla\Phi_{u_0}^T \tag{3.75}$$

et ces termes peuvent être obtenus par méthode de Monte-Carlo.

3.3.6 Calibrage du Randomisateur

En principe, simuler l'échantillon des paramètres d'entrée du modèle $\{u_t\}$ s'avère plus facile que de simuler directement un échantillon de vecteurs d'états $\{x_t\}$: en effet, la plupart de temps les paramètres peuvent être considérés, faute d'information plus précise, comme indépendants deux à deux (S est prise diagonale), et sont donc simulés de manière indépendante. Cependant, il peut s'avérer que certains paramètres soient ostensiblement corrélés, car eux-mêmes sont les résultats d'un sous-modèle « caché » qui, comme nous l'avons expliqué précédent, les fait évoluer dans un sous-espace plus petit égal à l'image de ce sous-modèle. Nous ne pouvons pas accéder à ce sous-modèle, seulement à un échantillon de ses sorties, et nous ne pouvons pas l'utiliser en soi. On cherchera donc à concevoir un méta-modèle utilisable, qui nous permettra de générer à loisir un échantillon de paramètres présentant une distribution statistique semblable.

Avec une autre formulation, le problème est ainsi le suivant : étant donné un vecteur u_0 de taille m et une matrice S, symétrique positive, mais non nécessairement définie positive, de taille $m \times m$, comment simuler une variable aléatoire de distribution $NG(u_0, S)$?

Ce problème est standard, et la solution que nous décrivons ici provient de l'Analyse en Composante Principales (ACP) [Sap11].

S étant symétrique positive, elle est diagonalisable en base orthonormale, c'est-à-dire qu'il existe une matrice orthogonale P (i.e. telle que $P^T = P^{-1}$) telle que :

$$S = PDP^T \tag{3.76}$$

avec :

$$D = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \lambda_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} = \begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix}$$

où $\lambda_1 \geq \cdots \geq \lambda_r$ sont les valeurs propres strictement positives de S. On introduit :

$$\begin{aligned} F &= \begin{pmatrix} I_r \\ 0 \end{pmatrix} \\ \Lambda &= F^T P^T \end{aligned} \tag{3.77}$$

alors on a :

$$\begin{aligned} S &= \Lambda^T D_r \Lambda \\ D_r &= \Lambda S \Lambda^T \end{aligned}$$

On introduit un vecteur α de dimension r, défini par :

$$\alpha = \Lambda(u - u_0) \tag{3.78}$$

Alors, si u suit la loi $NG(u_0, S)$, α suit la loi $NG(0, D_r)$. C'est naturellement la réciproque qui nous intéresse : D_r étant diagonale, chacune des composantes du vecteur α peut ainsi être simulée séparément. On peut ainsi simuler un échantillon $\{\alpha_t\}$, et ensuite poser :

$$u_t = u_0 + \Lambda^T \alpha_t \tag{3.79}$$

pour obtenir un échantillon de vecteurs de paramètres $\{u_t\}$ qui ont la distribution statistique voulue $NG(u_0, S)$.

Avec ce point de vue, Λ^T est un méta-modèle linéaire qui émule le sous-modèle « caché », qui avait été utilisé pour obtenir un premier échantillon de vecteurs de paramètres u et d'en déduire une estimation de u_0, S. En supposant que ce sous-modèle caché soit lui aussi linéaire, et qu'on se situe dans les hypothèses du BLUE, alors on peut facilement montrer, avec des calculs quasiment identiques à ceux effectués dans le paragraphe « Changement de Variable » de la sous-section précédente, que les matrices obtenus par la méthode de Monte-Carlo appliquée avec le sous-modèle initial et son émulateur seront identiques, et que le résultat du BLUE s'en trouvera donc inchangé.

Chapitre 4

Application et Résultats

Dans ce chapitre, nous mettons en application la méthodologie développée au chapitre précédent et la testons sur des données réelles et sur des données simulées. L'agencement général des programmes est décrit en Annexe A.

4.1 Démarche

La première étape consiste d'abord en formuler notre problème d'estimation des Paramètres Primaires REP en termes d'Assimilation de Données, pour pouvoir appliquer les méthodes et algorithmes décrits au chapitre précédent.

4.1.1 Construction des opérateurs

Deux opérateurs doivent être construits : le modèle physique Φ, qui va être utilisé pour construire la matrice de covariance d'ébauche B par méthode de Monte-Carlo, et l'opérateur d'opération H qui permet de relier le vecteur d'état aux observations. A ces opérateurs correspond un choix pour les différents espaces vectoriels :

— $\mathbf{U} \equiv \mathbf{R}^m$, l'espace des paramètres ;

— $\mathbf{X} \equiv \mathbf{R}^n$, l'espace d'état ;

— $\mathbf{Y} \equiv \mathbf{R}^p$, l'espace des observations.

Premier choix pour l'Espace d'Etat

Le choix pour l'espace d'état est le plus déterminant, car il a des conséquences directes à la fois sur Φ et sur H. Il s'agit, nous l'avons dit, de définir un bon compromis entre représentativité et complexité, et d'adopter la description du circuit primaire qui nous paraît le mieux répondre à nos besoins. Plusieurs choix sont possibles, mais, dans l'optique d'un suivi en ligne d'une tranche nucléaire, une première option correspondant le mieux à la vision qu'ont les opérateurs de la tranche semble être la suivante :

— Débits-boucles Q_i (en t/s), pour $i = 1 \dots 4$;

— Températures de Branche Froide TBF_i (en $°C$) pour $i = 1 \dots 4$;

— Températures de Branche Chaude TBC_i (en $°C$) pour $i = 1 \dots 4$;

— Températures en haut des AC $THAC_j$ (en $°C$) pour $j = 1 \dots 193$;

— Puissances thermique des GV $PTHGV_i$ (en MW) pour $i = 1 \dots 4$;

— Coefficient de pondération bypass ϵ_{fit}.

En prenant à chaque fois les valeurs moyennées (typiquement avec un échantillonnage de vingt minutes) autour d'une date donnée. Avec ce choix, la dimension de l'Espace d'Etat est donc $n = 210$.

Construction de Φ

On reprend alors le cadre semi-empirique défini au chapitre 2, en définissant le vecteur de paramètres suivants :

— Débits-boucles Q_i (en t/s), pour $i = 1 \dots 4$;

— Températures de Branche Froide TBF_i (en $°C$) pour $i = 1 \dots 4$;

— Puissances Intégrées par Assemblage, normalisées $PIPAN_j$ (en $°C$) pour $j = 1 \dots 193$;

— Puissance thermique du réacteur $PTHREAC$ (en MW).

Avec ce choix, la dimension de l'Espace des Paramètres est donc $m = 202$.

Le modèle Φ est ensuite la succession de la loi PIPA et de la moyenne par quarts de cœur avec « fit de bypass ». A partir du vecteur des paramètres u, on construit ainsi le vecteur d'état x qui vérifie les propriétés physiques du circuit primaire :

1. On utilise les TBF_i, les $PIPAN_j$ et $PTHREAC$ pour appliquer la « loi PIPA » et obtenir les 193 $THAC_j$.

2. On initialise une valeur pour le coefficient de pondération bypass ϵ.
3. Des TBF_i, $THAC_j$ et du coefficient ϵ, on calcule les TBC_i par moyenne par quart de cœur avec pondération bypass.

4. Des Q_i, TBC_i, TBF_i on évalue les $PTHGV_i$ grâce à un bilan enthalpique aux bornes de chaque GV.

5. On évalue la différence entre la puissance transférée au secondaire et celle générée au primaire ($PTHREAC + PGMPP$, avec $PGMPP = 18\ MW$) : s'il existe une différence, on ajuste ϵ en fonction et on reprend à partir de l'étape 3. Si la différence est inférieure à $0.01MW$, alors on accepte $\epsilon_{fit} = \epsilon$ et l'algorithme s'arrête.

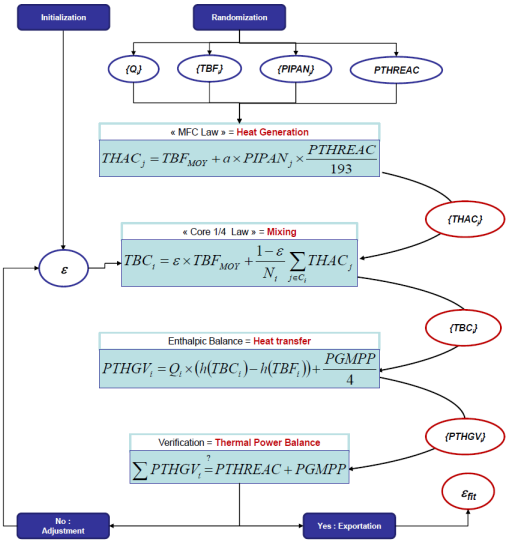


FIGURE 4.1 – Schéma de principe de l'utilisation du modèle-fit en mode direct avec la méthode Monte-Carlo

Construction de H

La construction de l'opérateur d'observation et de l'espace d'observations est à la fois plus simple et plus compliquée.

Plus simple, parce que, étant donné notre choix pour l'espace d'état, la plupart des (si ce n'est toutes les) observations que nous souhaitons assimiler sont chacune mesure directe d'une composante du vecteur d'état :

— une mesure ΔP-Coude, initialement en pourcentage de débit nominal (%QN) peut être transformée en mesure du débit-boucle absolu Q_i (t/s) ;

— une sonde de température RCP, ou un thermocouple RIC est une mesure directe, selon le cas, de TBC_i, TBF_i ou de $THAC_j$;

— une carte de flux extraite d'un fichier RES peut être transformé en une mesure directe de la carte des 193 $THAC_j$, en utilisant la loi PIPA ;

— les puissances GV enregistrées dans MOLENE constituent des mesures directes des $PTHGV_i$.

H est donc une matrice qui est aisément constructible : le coefficient H_{ij} vaut 1 si y_i est une mesure directe de la composante x_j, et 0 sinon.

Ce qui est plus compliqué, en revanche, c'est que toutes ces données ne sont pas disponibles (ou ne sont volontairement pas prises en compte) au même moment. Les données ORLI sont théoriquement disponibles en continu, mais il arrive fréquemment que certains capteurs soient momentanément indisponibles (ce qui est enregistré par une valeur totalement aberrante qui doit être supprimée), ou qu'il doive être écarté parce qu'il est

manifestement défaillant. Les données RES et MOLENE, quant à elles, ne sont disponibles qu'aux dates des EP, soit à peu près une fois par mois. De sorte que l'opérateur d'observation H et l'espace d'observations lui-même sont différents à chaque date.

On procède ainsi de la façon suivante :

1. On définit un cadre d'observations, c'est-à-dire l'ensemble des mesures assimilables.

2. Pour chaque mesure y_i, on construit l'opérateur partiel d'observation H_i, qui est un vecteur ligne tel que $y_i = H_i \cdot x$.

3. A chaque date, on recense les observations disponibles, on sélectionne celles qu'on souhaite utiliser, et on construit le vecteur d'observation y et, ligne par ligne, la matrice d'observation correspondante.

Dans la plupart des cas d'application ou de simulation, on se place aux dates des essais périodiques : on utilise ainsi toutes les mesures disponibles :

— 4 mesures pour chaque Q_i (ORLI) ;

— 6 mesures pour chaque TBF_i (ORLI) ;

— 2 mesures pour chaque TBC_i (ORLI) ;

— 1 mesure pour 48 $THAC_j$ pour les AC équipés de TC bien localisés (ORLI) ;

— 1 mesure pour chaque $THAC_j$ (RES via PIPA) ;

— 1 mesure pour chaque $PTHGV_i$ (MOLENE).

Dans ce cas, la dimension de l'espace d'observations est ainsi $p = 293$.

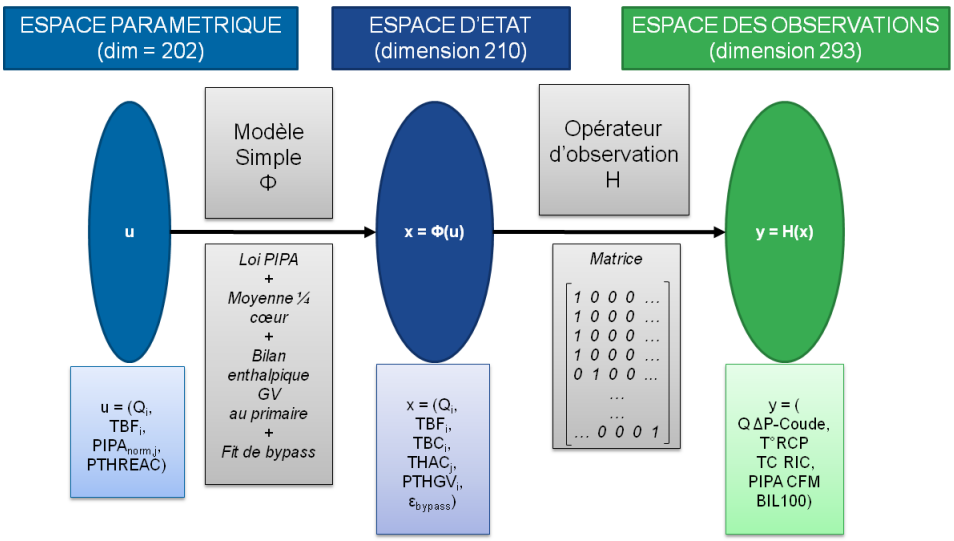


FIGURE 4.2 – Schéma des différents espaces vectoriels et des deux opérateurs principaux

Grandeur	Valeur	Incertitude ($\pm 2\sigma$)
Q	5.0 t/s	± 0.1 t/s
TBF	290.0 $°C$	± 0.4 $°C$
$PIPAN$	1.0	$\pm 10\%$
$PTHREAC$	3800.0 MW	$\pm 2\%$

TABLE 4.1 – Configuration par défaut de la randomisation des paramètres

4.1.2 Construction des paramètres du BLUE

Pour construire les termes d'ébauche, on applique la méthodologie de type Monte-Carlo développée au chapitre précédent. Pour ce faire, on doit randomiser les paramètres d'entrée du modèle.

La première méthode de randomisation qui a été utilisée est la plus intuitive : on simule toutes les composantes du vecteur de paramètres indépendamment les uns des autres. Les valeurs moyennes sont prises homogènes pour les quatre boucles ou les 193 assemblages combustibles. Les valeurs configurées sont présentées Table 4.1.

Les résultats de la randomisation (avec un échantillon de taille 1000) sont résumés Table 4.2 : pour chaque type de grandeur, on donne la valeur de la composante du vecteur d'ébauche, du terme diagonal de la matrice B, et de l'incertitude 2σ qui lui correspondent. Les résultats sont, comme attendu, symétriques pour les quatre boucles et les 193 AC.

La troisième colonne de la Table 4.2 indique l'incertitude 2σ déduite du terme diagonal de B : on constate effectivement que cette valeur correspond bien à la dispersion statistique

Grandeur	$xb[i]$	**Incertitude** $(\pm 2\sigma)$ **d'après** B
Q	5.0 t/s	± 0.10 t/s
TBF	290.0 $°C$	± 0.4 $°C$
TBC	323.7 $°C$	± 0.8 $°C$
$THAC$	326.1 $°C$	± 3.7 $°C$
$PTHGV$	954.5 MW	± 29 MW
ϵ_{fit}	0.067	± 0.11

TABLE 4.2 – Résultats de la méthode Monte-Carlo avec la randomisation par défaut

réelle de la composante considérée du vecteur d'état.

La matrice B obtenue est inversible, quoique mal conditionnée. La Figure 4.3 montre le spectre de la matrice de corrélation qui lui est directement associée (on trace $log(\frac{\lambda}{\lambda_{max}})$) :

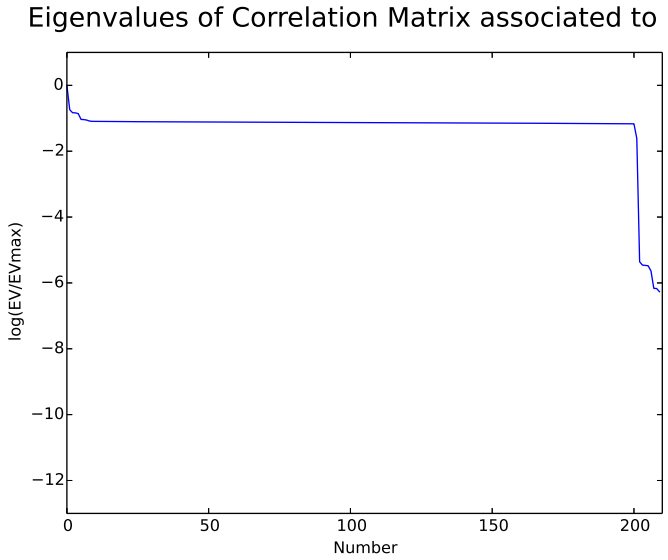


FIGURE 4.3 – Forme spectrale de la matrice de corrélation associée à la matrice de covariance d'erreur obtenue avec la configuration par défaut pour la randomisation des paramètres

— 202 valeurs propres sont supérieures à 10^{-2} de λ_{max} ;

— 8 valeurs propres sont inférieures à 10^{-5} de λ_{max}.

La coupure est nette et semble correspondre assez bien à la dimension de l'espace des paramètres ($202 = 1 + 193 + 4 + 4$), c'est-à-dire au nombre de paramètres qui ont été simulés indépendamment les uns des autres.

La construction de R dépend du choix de l'espace d'observations. Pour chaque mesure du système d'observations, on enregistre l'incertitude (2σ) attribuée à cette mesure ; ensuite,à chaque date, en fonction des mesures disponibles, on construit R diagonalement, mesure par mesure, en utilisant les variances.

Instrument	**Biais**	**Incertitude** ($\pm 2\sigma$)	$R[i,i]$ **correspondant**
Q par ΔP-Coude	0.0 t/s	± 0.15 t/s	5.6×10^{-3}
TBF par sondes RCP	0.0 $°C$	± 1.0 $°C$	2.5×10^{-1}
TBC par sondes RCP	0.0 $°C$	± 1.5 $°C$	5.6×10^{-1}
$THAC$ par sondes RIC	0.0 $°C$	± 2.0 $°C$	1.0×10^{0}
$THAC$ par estimation PIPA	0.0 $°C$	± 3.0 $°C$	2.3×10^{0}
$PTHGV$ par BIL100 ou BILORLI	0.0 $°C$	± 10.0 MW	2.5×10^{1}

TABLE 4.3 – Configuration des incertitudes de mesure

Remarque 5.
En principe, chaque instrument de mesure a une incertitude qui lui est spécifique, et ces incertitudes ne sont pas nécessairement centrées (par exemple dans le cas des sondes RCP de TBC). Néanmoins les ordres de grandeurs restent proches. On décide donc de grouper les instruments par catégories principales et d'accorder une incertitude type pour chaque catégorie, comme décrit dans la Table 4.3 : les incertitudes choisies sont une majoration raisonnable des incertitudes réelles.

On rappelle le côté arbitraire de la construction des matrices de covariance : elles ne font que définir les pondérations qui vont être effectuées entre les différents estimateurs. De plus, supposer une distribution gaussienne pour les incertitudes, l'indépendance des erreurs, et leur absence totale de biais, etc. sont autant d'approximations qui ont déjà été effectuées, et on prend en compte ces approximations dans ce choix majoré pour les incertitudes de mesures.

4.1.3 Expérience Jumelle

Nécessité

Nous avons construit tous les paramètres nécessaires pour appliquer un BLUE : reste à choisir des données à analyser.

Nous disposons de plusieurs échantillons de données réelles, extraites et prétraitées comme décrit dans le Chapitre 2 : notamment celles, pour lesquelles plusieurs études ont déjà été réalisées, des deux tranches d'une même centrale que nous noterons AA (tranches notées AA1 et AA2), pour les cycles 16, 17 et 18, aux dates des cartes de flux. Ce sont à ces données, et plus spécifiquement à celles de AA116 (cycle 16 de la tranche 1 de la centrale AA), que nous appliquerons nos techniques d'Assimilation de Données pour en tester l'efficacité.

Un problème essentiel se pose toutefois rapidement lorsqu'on travaille avec des données réelles : ne connaissant pas le vecteur d'état réel, qui est la cible que nous tentons d'estimer, il est impossible de vérifier l'efficacité de notre analyse. La théorie nous donne explicitement la matrice de covariance d'erreur d'analyse A, et donc l'incertitude 2σ associée à l'analyse de chaque composante, mais la pratique se conforme-t-elle à la théorie ? Tout au plus peut-on vérifier le recoupement des barres d'incertitudes des différents estimateurs, composante par composante, ce qui permet de dégager des tendances et de conjecturer par exemple la défaillance manifeste de telle ou telle sonde. (voir Fig. 4.4)

Pour cette raison, il est standard, pour des problèmes d'estimation, de commencer par valider les performances d'un algorithme en travaillant sur des données pseudo-réelles, c'est-à-dire des données simulées : on simule un échantillon de vecteur d'états réels, qui constituent la « cible » et les observations correspondantes ; on procède ensuite à l'analyse des observations et on examine directement si les analyses sont, dans les faits, plus proches de la cible que les estimations antérieures.

On parle alors d'**Expérience Jumelle**, car on applique la méthode sur un échantillon considéré comme jumeau de l'échantillon réel intangible : on fait l'hypothèse, fondamentale, que cette gémellité est suffisamment effective pour que les performances de notre méthode soient les mêmes sur l'échantillon jumeau que sur l'échantillon réel.

Dans tout le cas statique, la gémellité qu'on recherche porte sur le comportement statistique, et l'échantillon de données simulées reproduit ainsi le comportement statistique que nous supposons être également celui des données réelles. On génère ainsi un nuage de vecteurs d'état pseudo-réels – et de vecteurs d'observations correspondants – et nous faisons l'hypothèse que la forme de ce nuage (visible) est suffisamment proche de celle du nuage (invisible) de vecteurs réels à analyser. Notamment, on remarque ici qu'il ne s'agit pas de simuler de trajectoire vectorielle, c'est-à-dire de simuler des vecteurs agencés les uns avec les autres dans le temps, comme ce sera le cas plus loin dans le cadre dynamique.

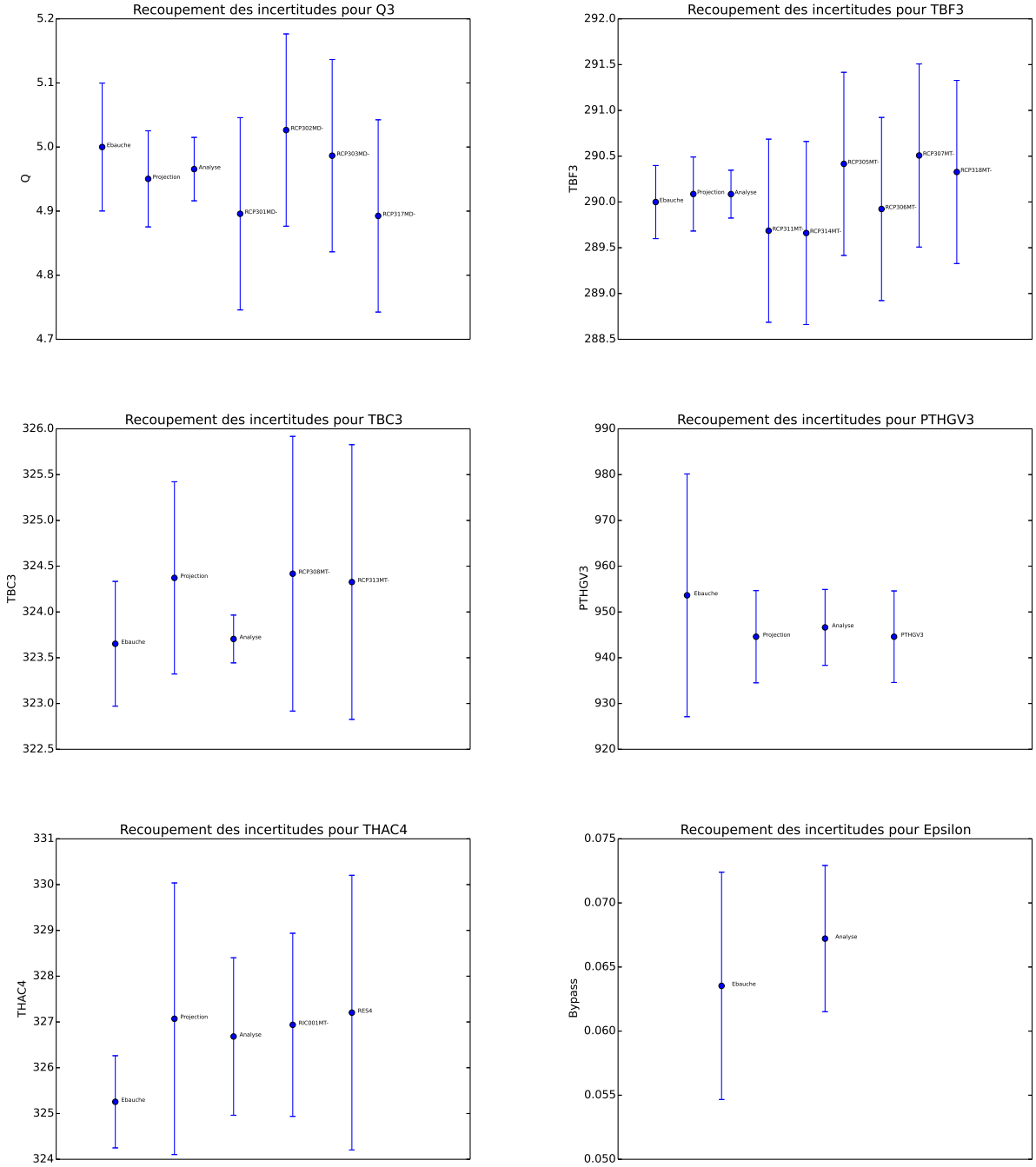


FIGURE 4.4 – Exemple de recoupement des barres d'incertitudes (Tranche NO1, cycle 16, le 9 Octobre 2008)

Processus

Pour effectuer les expériences jumelles décrites dans ce chapitre, nous avons procédé de la façon suivante :

1. On choisit la taille N de l'échantillon à simuler.

2. On crée une liste de « dates » de taille N (bien que l'ordre des vecteurs n'ait aucune importance) : à chaque date va être associée un vecteur d'état réel et ses différentes estimations.

3. On simule les vecteurs réels (en fait pseudo-réels) x_t, en utilisant le modèle physique et en randomisant ses paramètres, de même que lors de la Méthode de Monte-Carlo pour le calcul de l'ébauche.

4. Pour chaque date, on paramètre le cadre d'observations : si on le souhaite, on peut ainsi sélectionner des mesures indisponibles, biaiser volontairement certaines sondes, etc....pour étudier l'effet. Ceci effectué, on simule une observation bruitée du vecteur d'état réel enregistré pour cette date.

5. Pour chaque date, on enregistre trois estimations pour l'état : l'**ébauche**, obtenue à partir du seul modèle physique ; la **projection**, obtenue à l'aide des seules observations ; l'**analyse**, obtenue par assimilation de données.

6. On procède finalement à un « débriefing » global : sur l'ensemble des dates, on calcule les erreurs effectivement commises par les différents estimateurs ; on évalue leur dispersion statistique, et donc l'incertitude qui leur est associée.

Les échantillons ainsi simulés et testés sont enregistrés dans des feuilles de calcul EXCEL. On peut ainsi vérifier à la main que les incertitudes évaluées automatiquement par nos codes de calculs sont correctes, et contrôler ainsi la validité des performances annoncées pour l'approche d'Assimilation de Données.

Les résultats des « **débriefings** » sont résumés dans une feuille de calcul séparée : on évalue les erreurs et les incertitudes commises pour certaines composantes du vecteur d'état, choisies par l'utilisateur : on choisit une boucle i et un Assemblage Combustible j, et les composantes étudiées seront alors Q_i, TBF_i, TBC_i, $PTHGV_i$, $THAC_j$ et le coefficient de pondération bypass ϵ.

Remarque 6.
Si aucun numéro de boucle ou d'AC n'est requis par l'utilisateur, le programme procède un peu différemment : à chaque date et pour chaque estimateur, on effectue la moyenne des erreurs commises sur chacune des quatre boucles (ou, dans le cas des $THAC$, des 193 assemblages combustibles), et c'est de ces erreurs moyennées qu'on évalue le comportement statistique.

On calcule enfin, pour chaque estimateur, un **score absolu** en calculant la trace de la matrice de covariance d'erreur associée. Ensuite, en divisant chaque score absolu par le

Grandeur	Erreur d'Ebauche	Erreur de Projection	Erreur d'Analyse
Q	$0.00 \pm 0.10\ t/s$	$0.00 \pm 0.07\ t/s$	$0.00 \pm 0.05\ t/s$
TBF	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.3\ °C$
TBC	$0.0 \pm 0.8\ °C$	$0.0 \pm 1.1\ °C$	$0.0 \pm 0.3\ °C$
$THAC$	$0.0 \pm 3.7\ °C$	$0.0 \pm 3.0\ °C$	$0.0 \pm 2.3\ °C$
$PTHGV$	$0 \pm 29\ MW$	$0 \pm 10\ MW$	$0 \pm 8\ MW$
ϵ_{fit}	$0.0 \pm 1.1\ \%$	N/C	$0.0 \pm 0.7\ \%$
Score Absolu	1501	466	290
Score Relatif	100%	31%	19%

TABLE 4.4 – Résultats pour une Expérience Jumelle standard avec des paramètres homogènes

score absolu maximal, on évalue son **score relatif** (l'estimateur le moins précis se voit ainsi attribuer *de facto* un score relatif de 100%).

Résultats dans le cas standard

Les principaux résultats de l'Expérience Jumelle réalisés avec les paramètres décrits en Table 4.1 et en Table 4.1 (c'est-à-dire des incertitudes par défaut et des grandeurs homogènes sur les quatre boucles et les 193 Assemblages Combustibles) sont présentés Table 4.4.

Le premier enseignement est que, dans tous les cas, l'incertitude d'analyse est fortement réduite comparée à celles de l'ébauche et de la projection. En terme de score (c'est-à-dire d'erreur quadratique globale), l'analyse est naturellement l'estimateur le plus précis, et d'assez loin.

L'incertitude pesant sur le débit primaire est notablement réduite, et passe à 0.00 ± 0.05 t/s, soit $\pm 1\% QN$, après analyse (voir Fig. 4.5). Comparé aux ± 0.15 t/s, soit $\pm 3\% QN$, que nous avions attribués à chaque mesure ΔP-Coude, l'incertitude a ainsi été divisée par 3.

En ce qui concerne l'autre paramètre-clé qu'est la Température de Branche Chaude, l'incertitude post-analyse descend à $\pm 0.3\ °C$ (voir Fig. 4.6). L'incertitude sur les sondes RCP de branche chaude ayant été prise égale à $\pm 1.5\ °C$, l'incertitude sur TBC a donc été divisée par 5.

Enfin, l'incertitude pesant sur les $PTHGV$ est passée de $\pm 10\ MW$ (mesure BIL100/BILORLI) à $\pm 8\ MW$, ce qui est un gain limité, mais dû au fait que le BIL100 est déjà très précis au départ.

Ces résultats sont ainsi vivement encourageants quant au potentiel de l'Assimilation de Données pour la réduction des incertitudes pesant sur nos paramètres primaires. Naturellement, ils sont conditionnés à des hypothèses sous-jacentes fortes. Essentiellement :

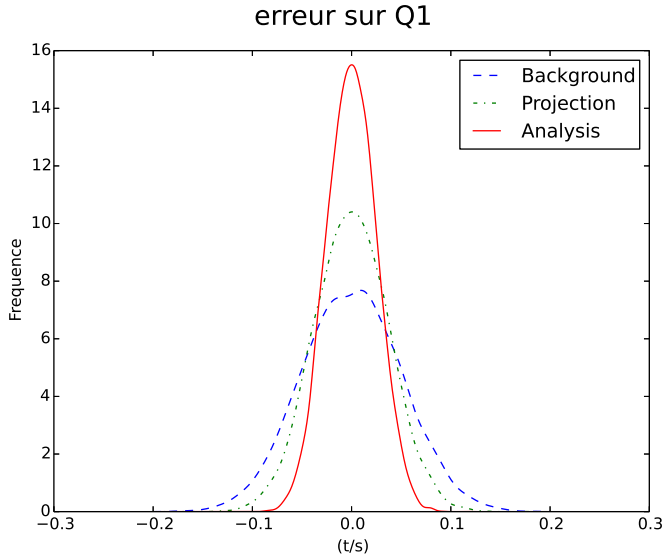


FIGURE 4.5 – Expérience Jumelle Standard avec paramètres homogènes : distribution statistique des erreurs commises pour $Q1$ (débit-boucle primaire boucle 1)

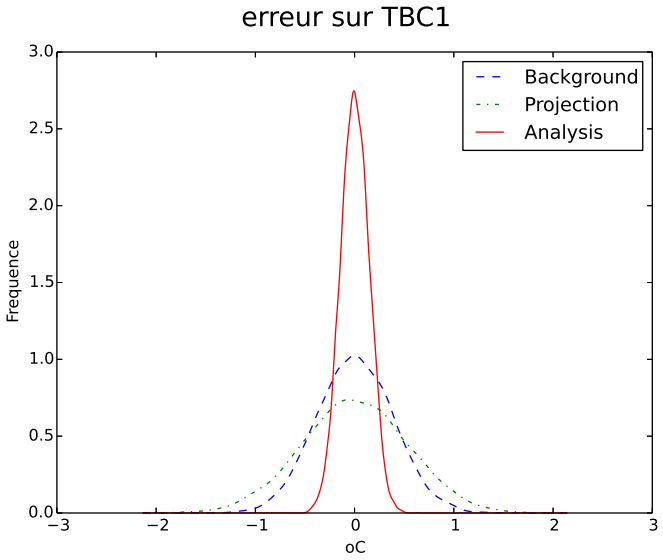


FIGURE 4.6 – Expérience Jumelle Standard avec paramètres homogènes : distribution statistique des erreurs commises pour $TBC1$

— l'hypothèse que le modèle Φ reflète suffisamment la réalité de la physique du circuit primaire ;

— l'hypothèse que le comportement statistique réel des paramètres d'entrée du modèle peut effectivement être modélisé comme on l'a fait lors de la randomisation (petites variations centrées autour d'une valeur nominale) ;

— l'hypothèse que toutes les erreurs d'observations soient décorrélées et sans biais.

La première hypothèse est fondamentale : une expérience jumelle n'apporte aucune information effective sur la performance de notre analyse sur des données réelles, si on ne considère pas avoir restitué le comportement de la réalité avec une fidélité suffisante dans notre modèle et dans le choix des vecteurs décrivant l'état du système.

En ce qui concerne la pertinence de nos choix relatifs à la construction des paramètres d'ébauche et d'observations, nous les discutons dans les sections suivantes.

4.2 Effets de l'Ebauche

Dans cette section, nous examinons les choix effectués pour la construction de la matrice de covariance d'erreur d'ébauche B, les remettons en question, et évaluons leur influence sur le résultat de l'analyse.

4.2.1 Calibrage des PIPAN

Nécessité

Lors de la construction de B effectuée en sous-section 4.1.2, nous avons notamment choisi de simuler nos paramètres de façon indépendantes, c'est-à-dire, en reprenant les notations du chapitre précédent, de définir le vecteur de paramètres u comme une variable aléatoire gaussienne de matrice de covariance S prise diagonale. Examinons la pertinence de cette hypothèse.

Les débits-boucles sont imposés par les Pompes Primaires, qui fonctionnent automatiquement et indépendamment les unes des autres ; les Températures de Branches Froides sont maintenues quasi-constantes (en fonctionnement nominal, et donc après la divergence lors de la montée en température) par régulation thermique : si elles sont corrélées entre elles, la corrélation n'est pas évidente et de toutes façons mal connue. La Puissance thermique du réacteur, quant à elle, est un facteur extérieur, lié au pilotage de la tranche par l'opérateur.

Les Puissances Intégrées par Assemblages Normalisées (les $PIPAN$), en revanche, sont nécessairement corrélés entre eux, car liés notamment par les équations de la neutronique. D'un point de vue pratique, nos données relatives aux $PIPAN$ sont obtenues par le processus suivant : une soixantaine d'Assemblages Combustibles sont équipées de Chambres à Fission Mobiles (CFM) qui mesurent leurs activités neutroniques. Le code CAMARET combine ces mesures avec un certain nombre d'autres paramètres dont certains mal connus, pour produire *in fine* une cartographie 3D des puissances neutroniques sur la totalité du cœur. De fait, nos $PIPAN$ sont ainsi les outputs d'un sous-modèle physique « caché », dont CAMARET effectue la résolution numérique. Les simuler de façon indépendante est ainsi forcément une approximation qui pourrait s'avérer excessive.

Nous appliquons ainsi aux $PIPAN$ la méthode développée au chapitre précédent pour le calibrage de la randomisation des paramètres.

Choix de calibrage

On dispose des fichiers RES calculés pour neuf tranches différentes 1300 MW, sur plusieurs années, ce qui constitue plus d'un millier de cartes de flux différentes. Ces données nous permettent donc d'évaluer le comportement statistique d'un échantillon d'outputs du code CAMARET. Reste à choisir cet échantillon, et plusieurs choix sont possibles, par ordre décroissant de taille :

— le choix « Total » qui consiste à prendre pour échantillon l'intégralité des cartes disponibles ;

Grandeur	TOT	AA1	AA116	DEF
PIPAN minimum	0.37	0.38	0.41	1.0
PIPAN maximum	1.23	1.23	1.22	1.0
Variation minimum	±0.05	±0.04	±0.01	±0.10
Variation maximum	±0.21	±0.28	±0.16	±0.10
Variation moyenne	±0.12	±0.13	±0.07	±0.10
Rang de S	193	149	95	193
Seuil 2	24	10	6	-
Seuil 4	69	31	15	-
Seuil 6	166	76	15	-
Seuil 8	193	92	15	-

TABLE 4.5 – Statistiques des différents échantillons de PIPAN

— le choix « Tranche » qui consiste à prendre pour échantillon les cartes disponibles pour la tranche étudiée ;

— le choix « Cycle » qui consiste à prendre pour échantillon les cartes disponibles pour le cycle actuel de la tranche étudiée.

Reprenons pour exemple la tranche 1 de la centrale AA, pour le cycle numéro 16 : on peut choisir de simuler les $PIPAN$ suivant la statistique constatée sur l'ensemble des cartes de flux disponibles pour les 1300MW, ou suivant celle constatée sur l'ensemble des cartes disponibles pour la tranche AA1, ou sur l'ensemble des cartes du cycle AA116. Il reste également le choix par défaut de les simuler de manière indépendante. Nous notons ces quatre choix respectivement TOT, AA1, AA116 et DEF.

Pour chacun de ces choix, on étudie la statistique du vecteur des $PIPAN$: le moment d'ordre 1 u_0, c'est-à-dire la carte de flux moyenne de l'échantillon, et le moment d'ordre 2 S, c'est-à-dire leur matrice de covariance. Le tableau 4.5 présente les résultats de cette étude :

— la valeur minimale atteinte et la valeur maximale atteinte de la carte de flux moyenne ;

— les variations minimum, maximum et moyenne des $PIPAN$ (on calcule, pour chaque AC, la variation 2σ constatée des $PIPAN$ (σ étant le terme diagonal de S correspondant) pour l'AC en question ; on prend ensuite respectivement le minimum, le maximum et la moyenne de l'ensemble de ces variations) ;

— le rang de la matrice S calculée ;

— les différents « seuils » de la matrice de corrélation : le seuil k est l'indice de la première valeur propre inférieure à $10^{-k}\lambda_{max}$.

La première constatation qui s'impose est que l'hypothèse des $PIPAN$ homogènes est naturellement erronée : d'une manière générale, les cartes de flux ont des amplitudes as-

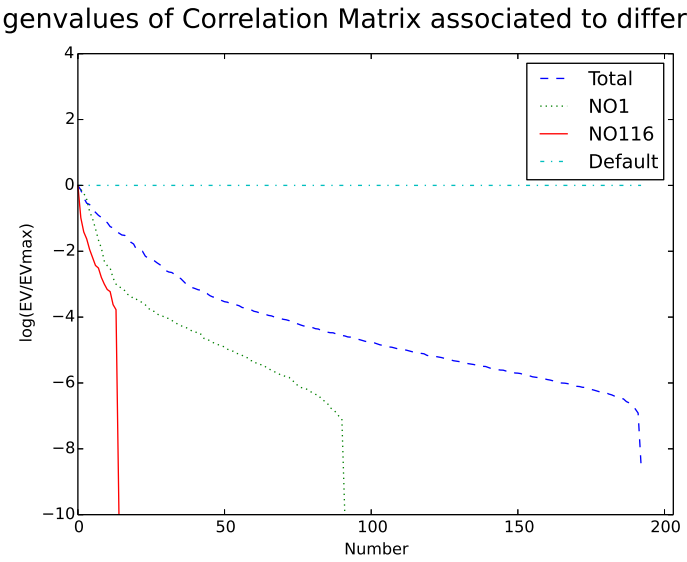


FIGURE 4.7 – Formes spectrales des matrices de corrélation associées aux matrices S obtenues pour les différents choix d'échantillons

sez proches (0.4 pour les AC les plus froids, souvent aux bords 1.2 pour les AC les plus chauds, plutôt vers le centre) ; de plus, les variations d'une carte de flux à l'autre sont plus ou moins importantes selon l'AC considéré. On constate toutefois que ces variations sont en moyenne proches des $\pm 10\%$ qui avaient été choisies pour la randomisation par défaut.

Les spectres obtenus à partir des données réelles, présentés Figure 4.7 sont bien moins plats que le spectre que nous avions supposé au départ : dans le cas TOT, où nous avons utilisé la totalité des cartes de flux à disposition – donc, a priori, le plus de données indépendantes entre elles – la matrice S est certes inversible, mais le spectre demeure néanmoins assez piqué, et seules 69 valeurs propres sont supérieures à 10^{-4} de la valeur propre maximale, ce qui signifie qu'on pourrait simuler un nuage de vecteurs $PIPAN$ imitant assez bien la statistique constatée sur des données réelles avec seulement 69 paramètres indépendants.

Comme attendu, on obtient une corrélation plus forte en restreignant les cartes de flux à celles d'une même tranche, et plus forte encore en restreignant les cartes de flux à celles d'un même cycle : en effet, les plans de chargements varient alors moins d'une carte à l'autre : les seules variations sont causées par l'épuisement du combustible et les légères fluctuations dues au pilotage de la tranche par l'opérateur. Néanmoins, en choisissant de construire notre ébauche (i.e. notre vision *a priori* de l'état) à partir de données réelles directement liées à la tranche, et donc issues de mesures (et donc *a posteriori*), on court le risque de fausser l'hypothèse d'indépendance des données d'ébauche et de données de mesure. Le cas TOT est obtenu sur un vaste ensemble de situations différentes, et peut donc être appliqué en toute généralité sans faire d'hypothèse forte sur l'état actuel du système à analyser.

A chaque couple $(PIPAN, S)$ correspond une loi de distribution de probabilité pour le paramètre aléatoire $PIPAN$ en entrée du modèle physique, que ce soit pour créer un échantillon de données pseudo-réelles, ou pour appliquer la méthode de Monte-Carlo et

DEF, $x_{b,DEF}, B_{DEF}$	**Erreur d'Ebauche**	**Erreur de Projection**	**Erreur d'Analyse**
Q	$0.00 \pm 0.10\ t/s$	$0.00 \pm 0.07\ t/s$	$0.00 \pm 0.05\ t/s$
TBF	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.3\ °C$
TBC	$0.0 \pm 0.8\ °C$	$0.0 \pm 1.1\ °C$	$0.0 \pm 0.3\ °C$
$THAC$	$0.0 \pm 4.3\ °C$	$0.0 \pm 3.1\ °C$	$0.0 \pm 2.6\ °C$
$PTHGV$	$0 \pm 30\ MW$	$0 \pm 10\ MW$	$0 \pm 9\ MW$
ϵ_{fit}	$0.0 \pm 1.2\ \%$	N/C	$0.0 \pm 0.7\ \%$
Score Absolu	1723	465	311
Score Relatif	100%	27%	18%
TOT, $x_{b,TOT}, B_{TOT}$	**Erreur d'Ebauche**	**Erreur de Projection**	**Erreur d'Analyse**
Q	$0.00 \pm 0.10\ t/s$	$0.00 \pm 0.08\ t/s$	$0.00 \pm 0.05\ t/s$
TBF	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.3\ °C$
TBC	$0.0 \pm 0.7\ °C$	$0.0 \pm 1.0\ °C$	$0.0 \pm 0.2\ °C$
$THAC$	$0.0 \pm 6.7\ °C$	$0.0 \pm 2.9\ °C$	$0.0 \pm 0.8\ °C$
$PTHGV$	$0 \pm 27\ MW$	$0 \pm 10\ MW$	$0 \pm 8\ MW$
ϵ_{fit}	$0.0 \pm 1.1\ \%$	N/C	$0.0 \pm 0.5\ \%$
Score Absolu	1619	464	103
Score Relatif	100%	29%	7%

TABLE 4.6 – Résultats pour les analyses droites

obtenir ainsi des termes d'ébauche (x_b, B). On peut ainsi varier à loisir le mode de randomisation des $PIPANS$.

Effet d'une interversion des ébauches

On cherche à présent à étudier l'effet d'une erreur dans le choix de l'ébauche sur l'analyse. On simule ainsi deux échantillons différents (de taille 1000 chacun), l'un imitant la statistique constatée avec l'échantillon total (cas TOT), et un autre correspondant à la calibration précédente par défaut (cas DEF) : on simule également les observations correspondantes, avec les mêmes incertitudes d'observations pour les deux échantillons (totalité des mesures disponibles, incertitudes d'observations comme décrites Table 4.3. On peut maintenant, dans chaque cas, choisir d'utiliser pour l'analyse les termes d'ébauche obtenus avec le mode de randomisation TOT ou avec le mode de randomisation DEF. On compare ainsi six analyses différentes :

- les analyses « droites », où les échantillons sont analysés avec les bons termes d'ébauche : échantillon TOT analysé avec $x_{b,TOT}$ et B_{TOT}, d'une part, et échantillon DEF analysé avec $x_{b,DEF}$ et B_{DEF}, d'autre part ;
- les analyses « croisées », où les échantillons sont analysés avec les mauvais termes d'ébauche : échantillon TOT analysé avec $x_{b,DEF}$ et B_{DEF}, d'une part, et échantillon DEF analysé avec $x_{b,TOT}$ et B_{TOT}, d'autre part ;
- les analyses « semi-croisées », où les échantillons sont analysés avec les bons vecteurs d'ébauche mais les mauvaises matrices de covariances d'erreurs d'ébauche : échantillon TOT analysé avec $x_{b,TOT}$ et B_{DEF}, d'une part, et échantillon DEF analysé avec $x_{b,DEF}$ et B_{TOT}, d'autre part.

DEF, $x_{b,TOT}, B_{TOT}$	**Erreur d'Ebauche**	**Erreur de Projection**	**Erreur d'Analyse**
Q	$0.00 \pm 0.10\ t/s$	$0.00 \pm 0.07\ t/s$	$-0.02 \pm 0.05\ t/s$
TBF	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.3\ °C$
TBC	$0.1 \pm 0.8\ °C$	$0.0 \pm 1.1\ °C$	$0.2 \pm 0.4\ °C$
$THAC$	$2.7 \pm 4.3\ °C$	$0.0 \pm 3.0\ °C$	$0.9 \pm 4.1\ °C$
$PTHGV$	$4 \pm 30\ MW$	$0 \pm 10\ MW$	$5 \pm 9\ MW$
ϵ_{fit}	$-0.2 \pm 1.2\ \%$	N/C	$0.1 \pm 0.9\ \%$
Score Absolu	1723	465	813
Score Relatif	100%	27%	48%
TOT, $x_{b,DEF}, B_{DEF}$	**Erreur d'Ebauche**	**Erreur de Projection**	**Erreur d'Analyse**
Q	$0.00 \pm 0.10\ t/s$	$0.00 \pm 0.08\ t/s$	$0.00 \pm 0.05\ t/s$
TBF	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.3\ °C$
TBC	$-0.1 \pm 0.7\ °C$	$0.0 \pm 1.1\ °C$	$0.0 \pm 0.3\ °C$
$THAC$	$-2.8 \pm 6.7\ °C$	$0.0 \pm 2.9\ °C$	$-1.0 \pm 3.1\ °C$
$PTHGV$	$-4 \pm 27\ MW$	$0 \pm 10\ MW$	$0 \pm 8\ MW$
ϵ_{fit}	$0.2 \pm 1.2\ \%$	N/C	$0.3 \pm 0.6\ \%$
Score Absolu	1619	464	313
Score Relatif	100%	29%	19%

TABLE 4.7 – Résultats pour les analyses croisées

Les résultats de ces analyses droites, croisées et semi-croisées sont présentés en Tables 4.6, 4.7 et 4.8. On rappelle qu'il s'agit des statistiques constatées des différents estimateurs, pour la boucle 2 et l'assemblage combustible 4 (pour lequel on dispose d'une mesure RIC).

On constate tout d'abord que dans le cas des analyses croisées, l'interversion des vecteurs d'ébauche induit des biais : on ne se trouve ainsi plus sous le cadre du BLUE. Ces biais d'ébauche ne disparaissent pas nécessairement à l'analyse, notamment pour l'échantillon DEF. L'échantillon TOT, lui, résiste un peu mieux à ces biais.

Le second constat est que les dégradations (biais et variances) semblent porter essentiellement sur les $THAC$: les paramètres primaires principaux sont finalement assez peu détériorés lors des analyses croisées ou semi-croisées : TBC et $PTHGV$ sont effectivement moins précis après analyse (et, dans le cas des analyses croisées, potentiellement biaisées) mais l'analyse demeure au moins aussi précise que la projection (observations seules). Pour les THAC, en revanche, l'erreur commise à l'ébauche n'est pas rattrapée à l'analyse, et peut même être amplifiée. Cependant, la principale source d'erreur semble être liée au mauvais vecteur d'ébauche x_b : les analyses semi-croisées, où le vecteur d'ébauche x_b a été correctement pris, mais les matrices de covariance B échangées, demeurent dans un cas comme dans l'autre de bonne qualité, surtout pour les termes outcore.

Enfin, on remarque que pour chaque type d'analyse, l'échantillon TOT – qui correspond plus à la réalité physique des réacteurs réels – est mieux analysé que l'échantillon DEF : et que lorsque les termes d'ébauche ont été correctement choisis, on atteint une précision impressionnante sur l'ensemble des composantes du vecteur d'état (score absolu d'analyse minimal), et notamment, ce qui était prévisible, en ce qui concerne la cartographie des $THAC$ (incertitude d'analyse de $\pm 0.8\ °C$ pour $THAC_4$ dans ce cas-ci, par exemple).

DEF, $x_{b,DEF}, B_{TOT}$	**Erreur d'Ebauche**	**Erreur de Projection**	**Erreur d'Analyse**
Q	$0.00 \pm 0.10\ t/s$	$0.00 \pm 0.07\ t/s$	$0.00 \pm 0.05\ t/s$
TBF	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.3\ °C$
TBC	$0.0 \pm 0.8\ °C$	$0.0 \pm 1.1\ °C$	$0.0 \pm 0.4\ °C$
$THAC$	$0.0 \pm 4.3\ °C$	$0.0 \pm 3.0\ °C$	$0.0 \pm 4.1\ °C$
$PTHGV$	$1 \pm 30\ MW$	$0 \pm 10\ MW$	$0 \pm 9\ MW$
ϵ_{fit}	$0.0 \pm 1.2\ \%$	N/C	$0.0 \pm 0.8\ \%$
Score Absolu	1723	465	813
Score Relatif	100%	27%	47%
TOT $x_{b,TOT}, B_{DEF}$	**Erreur d'Ebauche**	**Erreur de Projection**	**Erreur d'Analyse**
Q	$0.00 \pm 0.10\ t/s$	$0.00 \pm 0.08\ t/s$	$0.00 \pm 0.05\ t/s$
TBF	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.3\ °C$
TBC	$0.0 \pm 0.7\ °C$	$0.0 \pm 1.1\ °C$	$0.0 \pm 0.3\ °C$
$THAC$	$0.0 \pm 6.7\ °C$	$0.0 \pm 2.9\ °C$	$0.0 \pm 3.1\ °C$
$PTHGV$	$0 \pm 27\ MW$	$0 \pm 10\ MW$	$0 \pm 8\ MW$
ϵ_{fit}	$0.0 \pm 1.1\ \%$	N/C	$0.0 \pm 0.6\ \%$
Score Absolu	1619	463	313
Score Relatif	100%	29%	19%

TABLE 4.8 – Résultats pour les analyses semi-croisées

En résumé : avoir configuré correctement le simulateur de $PIPAN$ pour obtenir des termes d'ébauches plus conformes à la réalité de l'échantillon permet d'améliorer l'analyse par rapport à l'utilisation de l'ébauche calculée par défaut ; prendre une mauvaise ébauche détériore le résultat, mais cette détérioration porte surtout sur les termes in-core et demeure limitée, particulièrement si le vecteur d'ébauche au moins a été « correctement » choisi.

4.2.2 Analyse sur les paramètres

Comme décrit au chapitre précédent, on peut choisir de se passer de la méthode de Monte-Carlo pour calculer l'ébauche en assimilant directement sur l'espace des paramètres. L'opérateur d'observation H « absorbe » ainsi le modèle Φ, et on fait ainsi porter l'essentiel de l'information sur la physique du système physique par l'opérateur H et non plus par la matrice B.

Si construire la matrice B devient alors beaucoup plus simple – on postule l'indépendance deux à deux des paramètres, sauf pour les $PIPAN$ pour lesquels on récupère directement la matrice S choisie lors du calibrage, et la matrice B est ainsi diagonale par blocs – l'opérateur H devient en revanche bien plus compliqué et en particulier non linéaire. Or la linéarité de l'opérateur d'observations est une hypothèse fondamentale du BLUE : on doit donc soit trouver un moyen de linéariser H, soit préférer une approche variationnelle. Nous privilégions cette option.

Pour la formulation variationnelle, on a besoin de l'inversibilité de B : or à présent, B est diagonale par blocs et est donc inversible si et seulement si chaque bloc l'est. Or l'un des blocs est la matrice S : et, comme vu précédemment, celle-ci est, sinon singulière, du moins mal conditionnée, ce qui rend la minimisation de la fonction coût J délicate. D'autre part, même dans le cas DEF où S est inversible, minimiser une fonction de coût de 202 paramètres présente un coût de calcul trop élevé par rapport à nos moyens immédiatement disponibles (algorithmes intégrés Python : scipy.minimize).

On procède donc à une réduction du nombre de paramètres, en effectuant une Analyse en Composantes Principales de S. En reprenant les notations du chapitre précédent :

1. on part d'une carte $PIPAN_0$ et d'une matrice S donnée, correspondant à une statistique de $PIPAN$;

2. on diagonalise S en base orthonormée, et on obtient ainsi une matrice orthonormale P et une matrice diagonale D, telle que $S = PDP^T$, et telle que la diagonale de D contient les valeurs propres de S, triées par ordre décroissant ;

3. on choisit un seuil η, et on compte le nombre r de valeurs propres λ de S (et de D) telles que $\frac{\lambda}{\lambda_{max}} \geq \eta$, où λ_{max} est la plus grande valeur propre ;
4. on construit alors D_r la matrice diagonale contenant les r plus grandes valeurs propres de S et $F = \binom{I_r}{0}$ matrice de taille $193 \times r$; et on pose $\Lambda = F^T P^T$, matrice de taille $r \times 193$;
5. on a ainsi l'approximation $S \approx \Lambda^T D_r \Lambda$, et on définit un nouvel espace contenant des paramètres α, de taille r, qui sont reliés aux paramètres $PIPAN$ par la relation $\alpha = \Lambda \cdot (PIPAN - PIPAN_0)$ et $PIPAN = PIPAN_0 + \Lambda^T \cdot \alpha$.

Ainsi, avec un seuil $\eta = 10^{-2}$ et en partant de la matrice S obtenue pour le cas TOT, on ne conserverait que 24 valeurs propres de S : l'espace des paramètres α serait donc de dimension 24, et simuler ce petit vecteur de paramètres de manières comme un vecteur gaussien de loi $NG(0, D_r)$ permet de reproduire la statistique globale des $PIPAN$ d'une façon qu'on peut estimer suffisante.

Grandeur	Erreur d'Ebauche	Erreur de Projection	Erreur d'Analyse
Q	$0.00 \pm 0.10\ t/s$	$0.00 \pm 0.07\ t/s$	$0.00 \pm 0.05\ t/s$
TBF	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.4\ °C$	$0.0 \pm 0.3\ °C$
TBC	$0.0 \pm 0.7\ °C$	$0.0 \pm 1.1\ °C$	$0.0 \pm 0.3\ °C$
$THAC$	$0.0 \pm 6.6\ °C$	$0.0 \pm 3.0\ °C$	$0.0 \pm 0.9\ °C$
$PTHGV$	$0 \pm 27\ MW$	$0 \pm 10\ MW$	$0 \pm 8\ MW$
ϵ_{fit}	$0.0 \pm 1.1\ \%$	N/C	$0.0 \pm 0.5\ \%$
Score Absolu	1637	467	114
Score Relatif	100%	28%	7%

TABLE 4.9 – Résultats pour un 3DVAR pratiqué sur l'espace des paramètres obtenu après une Analyse en Composantes Principales de S pour le cas TOT

On définit alors le problème d'assimilation de données suivant :

— espace d'état constitué des vecteurs formés des Q_i, TBF_i, α_j et $PTHREAC$ (ce qui donne, dans le cas TOT avec un seuil $\eta = 10^{-2}$, un espace de dimension 33) ;

— une matrice de covariance d'erreur d'ébauche diagonale, constituée des variances des paramètres (D_r pour la partie correspondant aux α_j ;

— y comme observations ;

— un opérateur d'observations constitué de la juxtaposition de Λ^T, Φ et H (à partir des α on retrouve les $PIPAN$ via Λ^T ; avec les $PIPAN$ et les autres paramètres et via le modèle Phi on calcule le vecteur d'état du circuit primaire complet x correspondant, et via H on en déduit les observations attendues) ;

— R comme matrice de covariance d'erreur d'observations.

Ce problème peut se résoudre par un 3DVAR.

Nous avons testé cette démarche en utilisant le mode de calibrage TOT pour les $PIPAN$ et en partant du $(PIPAN_0, S)$ correspondant, ce qui, comme mentionné plus haut, correspond à un espace de paramètres de dimension 33 (vecteur α de taille 24). Le temps de calcul demeure plus long que pour une analyse BLUE standard. Les résultats sont présentés Table 4.9.

Ces résultats sont à rapprocher de celui obtenu précédemment et présenté Table 4.6, celui de l'analyse BLUE « droite » TOT, c'est-à-dire l'analyse du même échantillon (simulé avec le calibrage TOT pour les $PIPAN$) auquel nous avions appliqué un BLUE avec les termes d'ébauches obtenus par méthode de Monte-Carlo avec le même calibrage TOT : on retrouve les mêmes performances spectaculaires, particulièrement pour les $THAC$. Le « Score Global » (écarts quadratiques cumulés constaté par rapport au vecteur réel) pour l'analyse BLUE était de 115, celui pour l'analyse 3DVAR sur les paramètres après ACP est de 126 : ainsi, la qualité de l'estimation après ACP – en ne conservant que la partie supérieure du spectre de la matrice S de calibration des $PIPAN$ – n'a quasiment pas été

dégradée par cette approximation.

Cette proximité des résultats tend également à valider la méthode de Monte-Carlo qui a été utilisée pour obtenir (x_b, B) dans le BLUE, et que nous n'avons pas utilisée pour le 3DVAR.

4.2.3 Bilan

Choisir une ébauche adaptée au système d'étude est délicat, car il faut à la fois la construire du mieux possible – en particulier, avec le moins de biais possible, ce qui nécessite de disposer d'une première estimation de $PIPAN$ et de $PTHREAC$ – tout en conservant l'indépendance entre l'ébauche et l'observation.

La construction de la matrice de covariance d'ébauche via la méthode de Monte-Carlo nécessite de définir une statistique de simulation pour les paramètres du modèle, en particulier pour les $PIPAN$. On a ainsi étudié différents échantillons de cartes de flux, pour lesquels on obtient autant de formes moyennes de cartes de flux différentes, et autant de matrices de covariance S, associées aux $PIPAN$, différentes. On peut les utiliser tant pour simuler des échantillons de données pseudo-réelles plus conformes à la réalité que pour construire des termes d'ébauche (x_b, B) correspondants.

Si se tromper grossièrement de vecteur d'ébauche x_b, c'est-à-dire introduire un biais d'ébauche, détériore la qualité de l'analyse en introduisant notamment des biais non rattrapés, se tromper de matrice de covariance d'ébauche B prête moins à conséquence, en particulier si les données analysées ont été simulées de manière à reproduire la statistique de cartes de flux réelles (et donc d'émuler les lois neutroniques sous-jacentes). On présente Table 4.10 les performances des analyses appliquées à un même échantillon de type TOT, utilisant le même vecteur d'ébauche $x_{b,TOT}$ mais différentes matrices de covariance d'ébauche B :

— B_{TOT}, obtenue par méthode de Monte-Carlo avec la bonne statistique pour les $PIPAN$;

— B_{DEF}, obtenue par méthode de Monte-Carlo avec la randomisation par défaut ($PIPAN$ indépendants) ;

— B_{ACP}, obtenue par Analyse en Composantes principales de B_{TOT} (on ne conserve que les 25 premières composantes principales de B, c'est-à-dire celles pour lesquelles la valeur propre correspondante est supérieure à 10^{-4} de la valeur propre maximale) ;

— B_{DIAG}, obtenue en ne conservant que les termes diagonaux de B_{TOT} (on élimine donc tous les termes non diagonaux, soit les corrélations entre erreurs composantes, liées au modèle physique).

— B_{INF}, égale à 1000 fois la matrice B_{DIAG} : on pénalise infiniment l'ébauche pour lui accorder un poids minimal, on s'attend donc à ce que l'analyse soit quasiment égale à la projection.

Par construction, dans tous les cas, l'ébauche est la même et la projection est la même ; et toutes les erreurs sont non biaisées. On ne présente donc dans ce tableau que les incertitudes 2σ constatées. L'échantillon TOT est cette fois-ci de taille 10000.

L'analyse avec la matrice B_{DIAG} est instructive : cela correspond en fait, d'une certaine manière, à une Analyse de Cressman : pour chaque composante du vecteur d'état, on

	Q	TBF	TBC	$THAC$	$PTHGV$	ϵ_{fit}	Sc. Abs.	Sc. Rel
Ebauche	0.10 t/s	0.4 $°C$	0.7 $°C$	6.6 $°C$	27 MW	1.1 %	1637	100%
Projection	0.07 t/s	0.4 $°C$	1.1 $°C$	3.0 $°C$	10 MW	N/C	467	28%
Analyse (B_{TOT})	0.05 t/s	0.3 $°C$	0.2 $°C$	0.8 $°C$	8 MW	0.5 %	106	6%
Analyse (B_{DEF})	0.05 t/s	0.3 $°C$	0.3 $°C$	3.0 $°C$	9 MW	0.6 %	318	19%
Analyse (B_{ACP})	0.06 t/s	0.4 $°C$	0.2 $°C$	0.8 $°C$	9 MW	0.5 %	107	7%
Analyse (B_{DIAG})	0.06 t/s	0.3 $°C$	0.6 $°C$	2.8 $°C$	10 MW	1.1 %	321	20%
Analyse (B_{INF})	0.07 t/s	0.4 $°C$	1.1 $°C$	3.0 $°C$	10 MW	1.1 %	460	28%

TABLE 4.10 – Résultats des différentes analyses pratiquées sur l'échantillon TOT avec des matrices de covariance d'erreur d'ébauche B différentes

corrige l'ébauche à l'aide des seules mesures directes de la composante en question. Même dans ce cas très simple, l'analyse est de plutôt bonne qualité : l'incertitude d'analyse sur les Q_i et sur les $PTHGV_i$ n'est presque pas dégradée, et celle sur les TBC_i demeure correcte. L'estimation du coefficient ϵ_{fit} n'est pas corrigée, car il n'en existe pas de mesure directe, et la valeur d'analyse reste égale à celle de l'ébauche (et l'incertitude associée également). Les performances sont proches de celles avec la matrice B_{DEF}, même légèrement meilleure en ce qui concerne les $THAC$.

L'analyse avec la matrice B_{ACP} présente quasiment les mêmes excellentes performances qu'avec l'ébauche optimale B_{T0T} : ce résultat est à rapprocher des performances du 3DVAR sur les paramètres après ACP (voir Table 4.9). En l'occurrence, on ne gagne pas ici grand-chose en termes de coût de calcul à remplacer B_{TOT} par B_{ACP}. Mais cette dernière matrice étant de rang 25, on peut envisager d'appliquer la méthode décrite en 3.3.6 et de construire un méta-modèle de notre modèle physique Φ, afin de travailler dans un espace d'état de dimension réduite (25 ici, au lieu de la dimension 210 actuelle pour l'espace d'état). Le résultat de l'analyse sur l'espace réduit sera par construction strictement identique à celle pratiquée sur l'espace actuel avec B_{ACP}, c'est-à-dire toujours très performant. Si les gains de ce changement d'espace en coût de calcul et de stockage est en l'occurrence minime, ce raisonnement peut resservir dans d'autres applications où l'espace d'état sera de dimension bien plus grande (et la matrice B par conséquent plus lourde).

L'analyse avec la matrice B_{INF}, c'est-à-dire une erreur d'ébauche largement surévaluée, donne le résultat attendu : l'analyse est effectivement quasiment égale à la projection, excepté le terme ϵ_{fit}, non évalué par la projection, et qui ne change pas entre l'analyse et l'ébauche.

L'enseignement principal est finalement le suivant : disposer d'un vecteur d'ébauche x_b de bonne qualité, bien représentatif de la valeur moyenne de l'état réel, est le plus important. On peut améliorer l'analyse en cherchant à obtenir une matrice B la plus représentative de la dispersion statistique de l'état réel, et des corrélations induites par le modèle, mais même l'analyse utilisant la matrice diagonale B_{DIAG}, construite seulement à partir des variances estimées pour chaque erreur composante, nous a permis de réduire l'incertitude de façon intéressante sur des paramètres-clés.

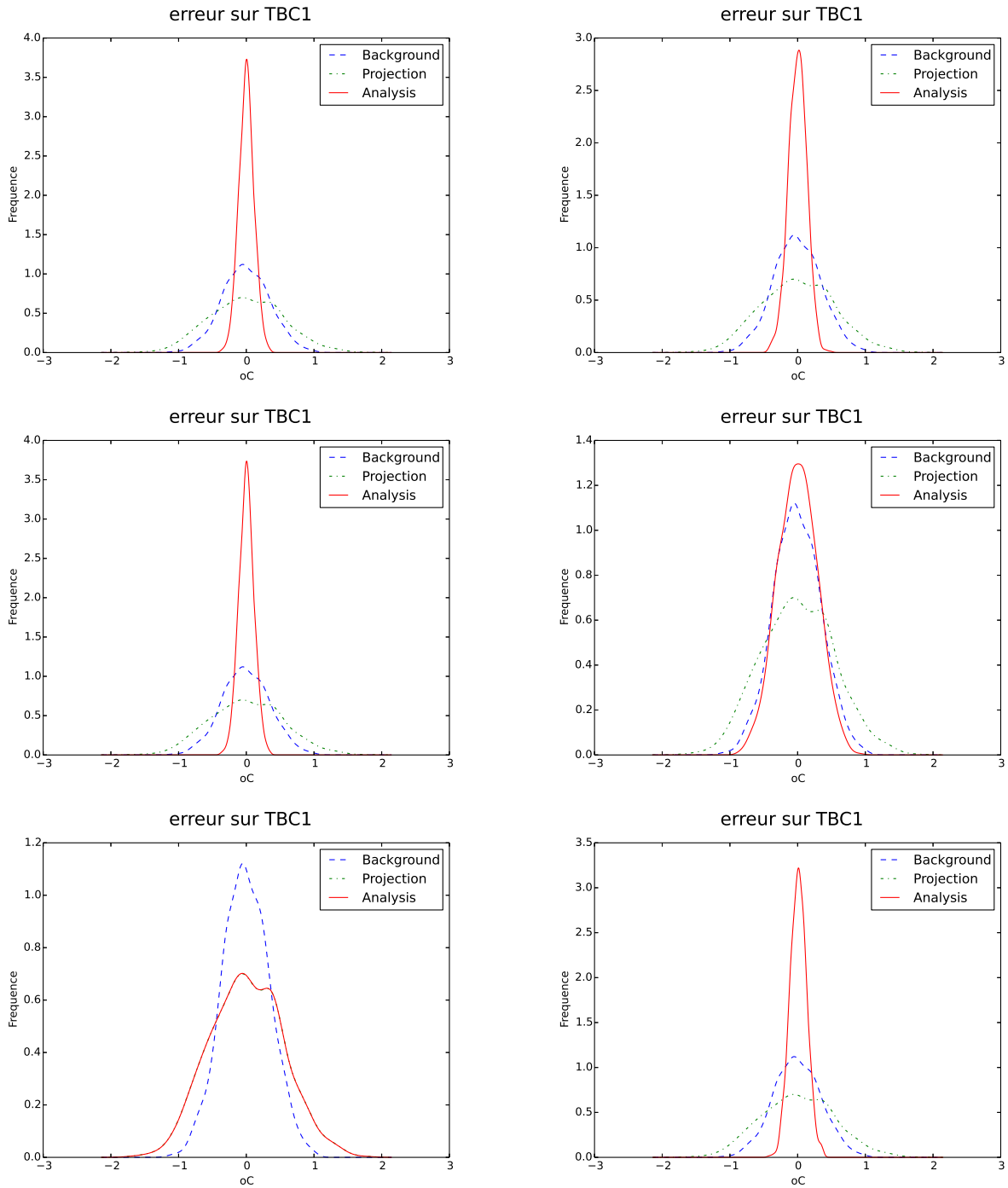


FIGURE 4.8 – Dispersion Statistique des Erreurs d'ébauche, de projection et d'analyse BLUE pour différents choix de B (TOT, DEF, ACP, DIAG et INF) ou d'analyse 3DVAR sur les paramètres après ACP

Cependant, le problème essentiel est que nous devons avoir confiance en notre ébauche, et être convaincus que le comportement statistique du vecteur d'état réel peut bien être modélisé par la loi normale gaussienne d'espérance x_b et de covariance B. Grâce à la méthode Monte-Carlo, cela revient à avoir confiance en la pertinence de notre mode de randomisation des paramètres d'entrée de notre modèle physique, et en notre modèle lui-même. De sorte que le principe fondamental des expériences jumelles, qui est de supposer que les données simulées sont suffisamment jumelles des données réelles pour que les performances de notre algorithme sur les données réelles soient les mêmes que sur les données simulées, est également le principe qui doit nous guider dans la construction de l'ébauche.

Finalement, paramétrer notre randomisateur de manière à le rendre plus proche de la réalité du système, et choisir ainsi les meilleurs termes d'ébauche possibles, permet donc non seulement d'améliorer les performances de l'analyse sur des données simulées, mais surtout de s'assurer que ces performances seront identiques sur des données réelles.

4.3 Effets des Observations

Nous avons étudié l'influence des hypothèses sur l'ébauche : nous allons à présent considérer l'influence des hypothèses sur les observations.

Nous prenons pour cas d'étude un échantillon de type TOT (c'est-à-dire un échantillon simulé avec des $PIPAN$ reproduisant la corrélation constatée sur la totalité des cartes de flux disponibles). Le vecteur d'ébauche x_b utilisé sera dans tous les cas le « bon », c'est-à-dire celui obtenu par méthode de Monte-Carlo utilisant le même randomisateur TOT. La matrice de covariance d'erreur d'ébauche B utilisée sera soit la matrice B_{TOT}, obtenue par cette même méthode, soit la matrice B_{DIAG}, formée de ses termes diagonaux.

4.3.1 Résistance aux biais

Parmi les hypothèses du BLUE, l'hypothèse de non-biais des erreurs (« Unbiased ») semble être la plus forte, en particulier pour les observations. Même en procédant aux soigneux recalages décrits en 2.5, on ne peut être assuré d'avoir éliminé tous les biais, notamment lorsqu'il s'agit de biais de représentativité causés un effet de l'état du système qui n'est pas restitué par la description que nous avons adoptée (effets liés au champs 3D de température de branche chaude, par exemple) et dont la variabilité temporelle est lente par rapport à la fréquence des observations. Quels seront les effets de ces biais éventuels sur la qualité de l'analyse ?

On simule à partir des vecteurs réels de l'échantillon TOT simulé des observations correspondantes comprenant des biais :

— mesures ΔP-Coude biaisées (de quelques $\pm\%QN$) ;

— mesures TBC RCP baisées (de $\pm 0.5 - 1°C$) ;

— mesures $THAC$ par TC RIC biaisées (de quelques dixièmes de degrés $°C$ (données provenant de biais réels estimés après divergence, après recalage AAC) ;

— mesures $PTHGV$ par BIL100 ou BILORLI biaisées (de quelques MW).

On analyse ensuite respectivement ces données biaisées, en utilisant pour termes d'ébauche $(x_{b,TOT}, B_{TOT})$ ou $(x_{b,TOT}, B_{DIAG})$ respectivement : l'erreur liée à l'ébauche est donc non biaisée. On étudie ensuite la statistique de l'erreur d'analyse, et notamment les biais qui subsistent.

On constate que si les biais sont presque effacés en utilisant B_{DIAG} comme matrice de covariance d'ébauche, ce n'est en revanche pas le cas en utilisant B_{TOT}. Les biais liés aux TBC sont certes réduits, mais ceux liés aux débits-boucles Q subsistent, et ceux liés aux $PTHGV$ peuvent même être amplifiés (c'est le cas pour $PTHGV_1$ par exemple, dans notre cas d'application).

Grandeur	Q	TBF	TBC	$THAC$	$PTHGV$	ϵ_{fit}
Ebauche	0.00 t/s	0.0°C	0.0°C	0.0°C	0 MW	0.0 %
Projection	0.05 t/s	0.0°C	1.0°C	-0.6°C	1 MW	N/C
Analyse TOT	0.00 t/s	0.1°C	0.3°C	-0.3°C	6 MW	-0.9 %
Analyse DIAG	0.03 t/s	0.0°C	0.3°C	-0.7°C	0 MW	0.0 %

TABLE 4.11 – Biais après analyse d'un échantillon de type TOT avec des observations biaisées, en utilisant respectivement les matrices B_{TOT} et B_{DIAG}, pour la boucle 1 et l'AC 4

L'explication est simple. Nos observations biaisées vérifient $y' = b+y$, où y est l'observation non biaisée : par conséquent, l'analyse réalisée avec les observations biaisées vérifiera :

$$\begin{aligned} x'_a &= x_b + K \cdot (y' - H[x_b]) \\ &= x_b + K \cdot (y - H[x_b]) + K \cdot b \\ &= x_a + K \cdot b \end{aligned}$$

Soit finalement :

$$x'_a - x_a = K \cdot b \tag{4.1}$$

où x_a est l'analyse effectuée avec des données non biaisées. On peut ainsi calculer explicitement le biais post-analyse, simplement en utilisant la matrice K.

Or l'action de la matrice K dépend directement des termes non-diagonaux de la matrice de covariance d'erreur d'ébauche B utilisée : la Figure 4.9 la contribution de chaque composante du vecteur d'innovation $y - H[x_b]$ pour la correction appliquée au vecteur x_b lors de l'analyse BLUE. Avec B_{DIAG}, comme il n'existe aucune colinéarité entre les différentes composantes de l'erreur d'ébauche, chaque composante est corrigée en utilisant exclusivement les observations qui la mesurent directement : Q est corrigé seulement avec les ΔP-Coude, $PTHGV$ seulement avec les mesures BIL100/BILORLI, TBC seulement avec les mesures RCP de branche chaude, etc. . .et le coefficient de bypass n'est pas corrigé du tout. En revanche, avec B_{TOT}, presque toutes les mesures disponibles sont combinées pour corriger chaque composantes.

Ainsi, Q et $PTHGV$ d'une même boucle étant étant étroitement liés par l'équation contenue dans le modèle $PTHGV = Q\Delta H$, et ce lien étant contenu dans la matrice B_{TOT}, l'analyse BLUE utilise de façon importante les mesures de ΔP-Coude dans la correction de l'estimation des $PTHGV$, et inversement les valeurs du BIL100/BILORLI sont utilisées dans la correction de Q_i.

L'analyse BLUE effectuée avec B_{DIAG} effectue donc, schématiquement, une moyenne composante par composante entre le vecteur d'ébauche et le vecteur de projection [1] : or

1. c'est formellement vrai pour les composantes où toutes les observations directes ont la même incertitude associée (dans la matrice R) ; ce n'est pas tout à fait exact sinon (par exemple pour les $THAC$ qui sont mesurés à la fois par un TC RIC et par une évaluation RES, avec des incertitudes différentes).

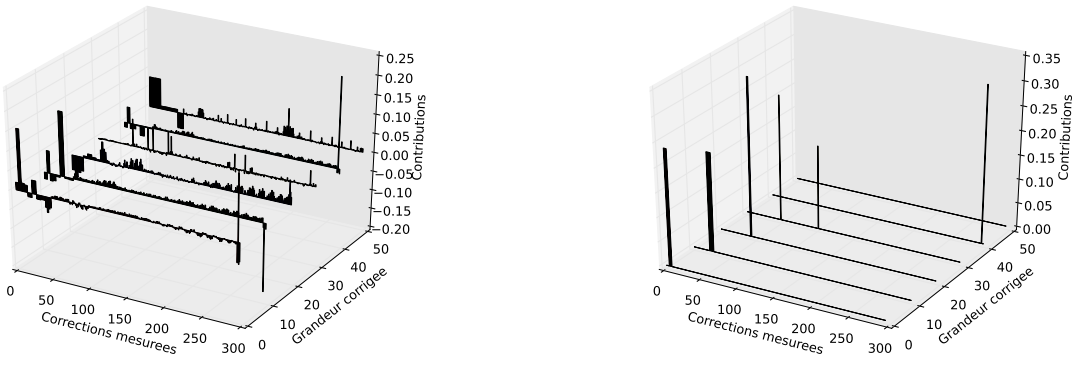


FIGURE 4.9 – Illustration des contributions de chaque composante du vecteur d'innovation pour la correction de l'ébauche effectuée lors de l'analyse BLUE (via les lignes de K) pour le cas B_{TOT} et le cas B_{DIAG}, pour les composantes (de l'avant vers l'arrière) Q_2,TBF_2, TBC_2, $THAC_4$, $PTHGV_2$ et ϵ_{fit}.

l'ébauche étant parfaitement centrée dans notre cas, l'analyse BLUE avec B_{DIAG} diminue automatiquement le biais.

L'analyse BLUE effectuée avec B_{TOT} peut en revanche amplifier des biais : dans le cas de $PTHGV_1$, par exemple, le biais de mesure BIL100/BILORLI était modéré, mais le biais des ΔP-Coude sur la boucle 1 étant élevés, le biais post-analyse sera amplifié. L'effet est d'autant plus important que les erreurs vont dans le même sens : l'effet d'un débit-boucle surestimé par ΔP-Coude et l'effet d'un PTHGV surestimé par BIL100/BILORLI s'additionneront pour amplifier le biais post-analyse. C'est ce qui se produit pour la boucle 1.

Si la matrice B_{DIAG} produit ainsi une analyse mécaniquement plus résistante au biais, il faut rappeler que l'analyse est également totalement dépourvue de cohérence physique : on n'a ainsi aucune garantie que les équations basiques du modèle physique soient vérifiées après analyse, ce qui est en revanche assuré (au moins dans une certaine mesure) par B_{TOT}.

L'enseignement principal est que les biais peuvent avoir un effet important sur l'analyse, notamment pour les ΔP-Coude, et qu'il faut donc être particulièrement prudent dans leur utilisation. Si des biais sont fortement suspectés (et surtout si ces biais sont dans le même sens pour une même boucle), l'analyse peut être fortement détériorée. Dans ce cas, on peut supprimer cet effet pervers en augmentant nettement l'incertitude associée aux mesures qu'on suspecte d'être biaisées, ou tout simplement les retirer du système d'observations.

4.3.2 Perte d'instrumentations

Les expériences jumelles que nous avons effectuées jusqu'ici ont été effectuées en utilisant un espace d'observations de dimension 293, c'est-à-dire avec toutes les observations disponibles. Or la disponibilité des instruments de mesures peut être variable, pour plusieurs raisons : d'une part, comme nous l'avons rappelé au début de ce chapitre, tous les

Analyse	Q	TBF	TBC	$THAC$	$PTHGV$	ϵ_{fit}
Normal	$\pm 0.05\ t/s$	$\pm 0.3°C$	$\pm 0.2°C$	$\pm 0.8°C$	$\pm 8\ MW$	$\pm 0.5\ \%$
Sans ΔP-Coude	$\pm 0.07\ t/s$	$\pm 0.3°C$	$\pm 0.3°C$	$\pm 0.8°C$	$\pm 9\ MW$	$\pm 0.7\ \%$
Sans fichier RES	$\pm 0.05\ t/s$	$\pm 0.3°C$	$\pm 0.3°C$	$\pm 1.0°C$	$\pm 8\ MW$	$\pm 0.6\ \%$
Sans TC RIC ni fichier RES	$\pm 0.05\ t/s$	$\pm 0.3°C$	$\pm 0.3°C$	$\pm 4.7°C$	$\pm 9\ MW$	$\pm 0.6\ \%$
Sans BIL100/BILORLI	$\pm 0.06\ t/s$	$\pm 0.3°C$	$\pm 0.3°C$	$\pm 0.8°C$	$\pm 13\ MW$	$\pm 0.5\ \%$

TABLE 4.12 – Incertitudes (2σ) pour l'analyse TOT d'un échantillon de type TOT pour différentes pertes d'instrumentation.

instruments de mesures ne sont pas disponibles aux mêmes dates (soit structurellement, comme pour les mesures provenant d'essais périodiques, soit par défaillance ponctuelle) ; d'autre part, on peut être amenés à retirer du système d'observations des mesures qui nous semblent ostensiblement biaisées (en constatant un non recoupement des barres d'incertitudes, par exemple).

Naturellement, moins on utilise de mesures, moins l'analyse est performante. On étudie ici la sensibilité de la performance de l'analyse (utilisant les termes d'ébauche ($x_{b,TOT}, B_{TOT}$)) en supprimant différents groupes d'instruments de mesures. On ne fait pas ici d'expérience jumelle, on étudie juste les termes diagonaux de la matrice de covariance d'erreur d'analyse A donnée par la théorie ($A = (Id - KH)B$).

Globalement, les performances de l'analyse restent tout à fait correctes, au regard des performances des deux seuls autres estimateurs disponibles (ébauche et projection). Il est particulièrement remarquable de constater que l'estimation de TBC résiste presque parfaitement à ces différentes pertes d'instrumentation : les informations liées au modèle physique lient si bien TBC à l'ensemble des mesures utilisées que la perte d'un groupe d'instrumentation ne dégrade pas trop la qualité de son analyse. Le fait que l'ébauche soit sans biais joue également dans la performance de l'analyse : simplement en donnant pour information la valeur autour de laquelle TBC est centrée, on modère les variations constatées autour de cette valeur et l'erreur s'en trouve automatiquement réduite. A titre de comparaison, on rappelle que dans le cas d'une analyse DIAG – sans modèle physique donc – (voir Table 4.10) l'incertitude 2σ post-analyse sur TBC était de $\pm 0.6°C$, qui est quasiment l'incertitude liée à l'ébauche ($\pm 0.7°C$, la variabilité telle que nous l'avons paramétrée) : la prise en compte du modèle physique permet ainsi de diviser l'incertitude par trois, pour TBC du moins.

La performance brute de l'analyse résiste donc assez bien à la perte d'instrumentation, ce qui encourage à ne pas hésiter à supprimer une sonde suspecte pour éviter le plus possible d'introduire des biais dans les observations.

4.4 Cadre dynamique

Nous avons jusqu'à présent travaillé dans un cadre statique, les « dates » des échantillons créés lors des expériences jumelles n'étant que des indices sans importance, et chaque

état pseudo-réel simulé à une date étant tiré indépendamment des précédents. Or les données réelles que nous avons à disposition sont elles réparties dans le temps, et l'état que nous cherchons à estimer varie de façon dynamique. Selon l'échantillonnage temporel, on peut supposer une certaine continuité dans cet évolution, et on peut chercher à exploiter cette information en utilisant des algorithmes d'Assimilation de Données dynamiques, comme le Filtre de Kalman ou l'Interpolation Optimum. Notre motivation est de mettre en pratique ces algorithmes simples et extrêmement utilisés notamment en météorologie.

Avant toute chose, il convient d'insister à nouveau sur deux différences majeures entre l'estimation des paramètres primaires d'un REP, et les géosciences :

1. On ne dispose pas de modèle d'évolution sophistiqué qui régisse la dynamique du système, car celle-ci est en majeure partie pilotée par l'opérateur.

2. On ne cherche généralement pas à prédire un état futur, mais plutôt à mieux connaître l'état présent ou passé du réacteur.

Or ces deux points sont précisément ceux qui sont la motivation et la clé du succès des algorithmes d'Assimilation de Données comme le Filtre de Kalman : voilà pourquoi la majeure partie du travail de thèse s'est focalisé sur l'application d'algorithmes statiques comme le BLUE ou le 3DVAR.

Néanmoins, devant l'omniprésence du Filtre de Kalman ou de ses variantes dans la littérature et de sa prédominance dans l'histoire de l'Assimilation de Données, nous avons cherché un moyen de l'appliquer à notre cas d'étude et d'évaluer son potentiel.

4.4.1 Modèle d'évolution

Matrice d'évolution

La première barrière à laquelle on se heurte est ainsi l'absence de modèle d'évolution précis régissant la dynamique du système. Nous l'avons dit, cette dynamique est essentiellement liée au pilotage du réacteur par l'opérateur et aux variabilités structurelles des process de régulations de température et de débit : en fonctionnement normal et à pleine puissance, l'état du réacteur est stable, et oscille de façon à la fois arbitraire et imprévisible autour d'une position d'équilibre.

Faute d'information, donc, il semble donc normal de prendre pour modèle d'évolution la matrice d'identité, soit, en reprenant les notations du Chapitre 3 :

$$\forall k, M_k = I_n \tag{4.2}$$

C'est à dire :

$$\forall k, x_{f,k+1} = x_{a,k} \tag{4.3}$$

La prévision pour l'instant $k+1$ est ainsi égale à l'analyse à l'instant précédent. Cela revient à supposer que les variations de l'état physique du système sont lentes devant le pas de temps choisi.

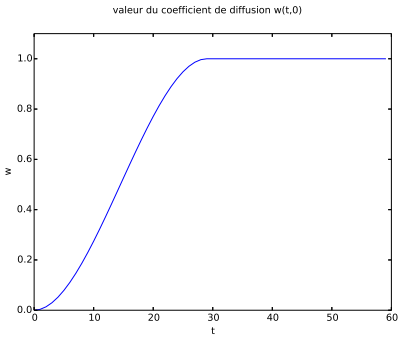


FIGURE 4.10 – allure du graphe de la fonction ω correspondant au coefficient « de diffusion Kalman »

Simulateur d'évolution

Pour pouvoir effectuer des expériences jumelles – toujours dans le but d'évaluer les performances effectives de nos algorithmes d'assimilation de données – il nous faut simuler des échantillons de données dynamiques. On ne peut évidemment pas supposer que notre « modèle d'évolution » est parfait, car cela reviendrait à considérer que l'état du réacteur est parfaitement constant.

Première idée : Marche Aléatoire Une première idée qui vient alors est d'ajouter à chaque pas de temps une petite erreur ϵ_k aléatoire, d'espérance nulle et de matrice de covariance proportionnelle à la matrice de covariance d'erreur d'ébauche B, d'autant plus petite que l'intervalle de temps entre t_k et t_{k+1} est petit, soit :

$$x_{t,k+1} = x_{t,k} + NG(0, \omega(t_{k+1}, t_k)B) \tag{4.4}$$

avec un coefficient « de diffusion » $\omega(t_{k+1}, t_k)$ par exemple égal à :

$$\omega(t_{k+1}, t_k) = \begin{cases} (\frac{t_{k+1}-t_k}{\Delta t})^2(3 - 2(\frac{t_{k+1}-t_k}{\Delta t})) \; si \; (t_{k+1} - t_k) < \Delta t \\ 1 \; sinon. \end{cases}$$

où Δt est le temps caractéristique de l'évolution de l'état.

Le problème est qu'on définit alors un processus de Wiener (mouvement brownien) par définition particulièrement instable, et que les échantillons ainsi simulés sont alors particulièrement instables.

Deuxième idée : Transitions aléatoires successives Une meilleure option semble donc être le processus de simulation suivant, constitué de transitions successives :

1. **Amorçage** On simule un état initial $x_{initial}$ et un état de destination x_{final}, en utilisant le randomisateur du cas statique ($x = \Phi(NG(u_0, S))$) et on pose $x_{t,0} = x_{initial}$. On enregistre la date de départ : $t_{initial} = t_0$.

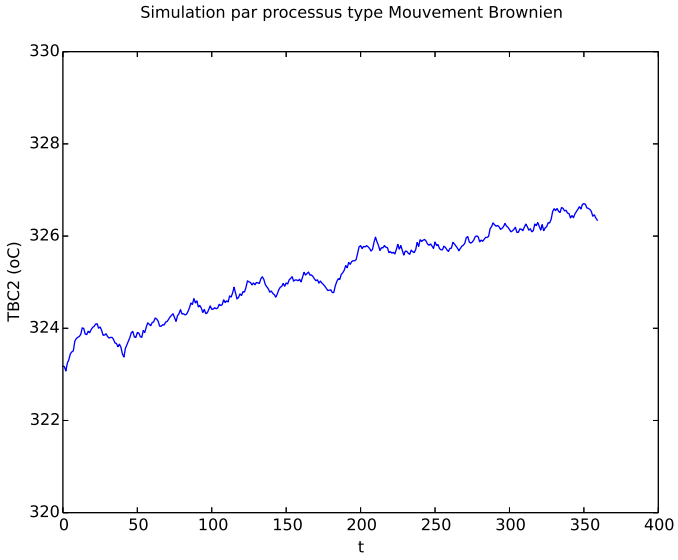


FIGURE 4.11 – Exemple de simulateur d'évolution par marche aléatoire

2. **Transition** Pour l'état numéro k, on effectue une moyenne convexe entre l'état initial et l'état final actuellement enregistrés :

$$x_k = (1 - \omega(t_k, t_{initial}))x_{initial} + \omega(t_k, t_{initial})x_{final} \tag{4.5}$$

3. **Adaptation** On transforme x_k en un vecteur d'état physiquement cohérent : on en extrait les Q_i et les TBF_i, d'une part, et on retrouve $PTHREAC$ et les $PIPAN$ d'autre part : on obtient ainsi un vecteur de paramètres qu'on injecte dans le modèle physique pour reconstruire un vecteur d'état physiquement cohérent $x_{t,k}$.
4. **Réamorçage** Si la transition est terminée (c'est-à-dire si $\omega(t_k, t_{initial})$ est suffisamment proche de 1, selon un critère à définir), alors on effectue un réamorçage : on pose $x_{initial} = x_{final}$, et $t_{initial} = t_k$ puis on simule un nouvel état de destination x_{final} en utilisant le randomisateur du cas statique.

De cette manière, la dynamique ainsi simulée relève bien d'une oscillation imprévisible autour d'un état de référence qui n'est autre que l'ébauche x_b du cas statique.

Nous avons appliqué cette méthode avec le randomisateur statique en mode TOT, un temps caractéristique d'évolution Δt de 29 jours, avec le coefficient de diffusion défini ci-dessus, et en réamorçant dès que $\omega(t_k, t_{initial}) >= 1.0$, soit tous les 30 jours exactement.

De cette manière, on obtient une trajectoire qui ressemble à des « dents de scie » : en effet, l'ordre des transitions est aléatoire et ne suit pas de logique temporelle. De plus, l'amplitude des transitions est importante, notamment en ce qui concerne les $THAC$, ce qui en fait une trajectoire peu réaliste.

Notre simulateur de vecteurs d'état, que nous avons utilisé à la fois pour la méthode de Monte-Carlo pour obtenir les termes d'ébauche et pour les expériences jumelles du cas statiques, nous sert en effet à représenter concrètement des incertitudes en décrivant un espace de vecteurs d'état possible : or lorsqu'on cherche à suivre l'évolution d'un état dans

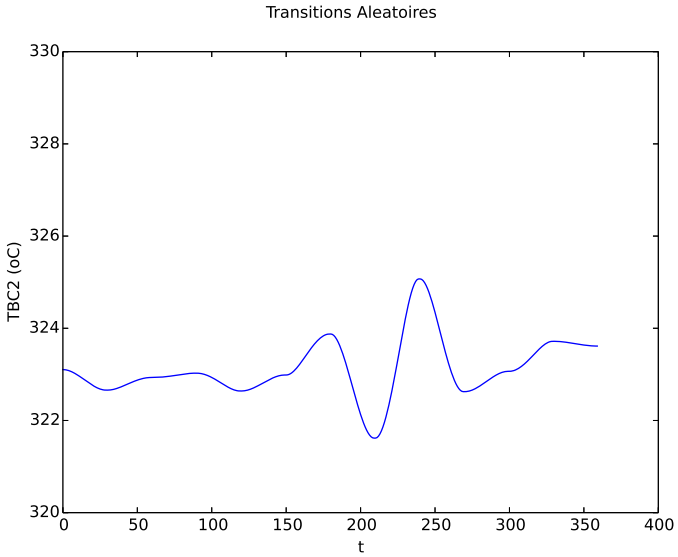


FIGURE 4.12 – Exemple de simulateur d'évolution par transition aléatoire

le temps, on décrit une trajectoire continue incluse dans ce nuage : dans notre cas, où l'on oscille légèrement autour d'un point de fonctionnement nominale, cette trajectoire est donc dans les faits contenue dans un secteur largement réduit de notre « nuage d'incertitude ».

Troisième idée : Interpolation de données réelles On pourrait ainsi recalibrer notre simulateur pour simuler une trajectoire moins ample, autour d'un point de fonctionnement donné : mais – le simulateur étant lui-même calibré à partir d'un jeu de données réelles – il est plus direct de réutiliser des données réelles pour définir les transitions.

On procède donc de la manière suivante :

1. On choisit une tranche et un cycle données, et on procède à un relevé des données disponibles aux dates des cartes de flux.

2. A partir des observations disponibles, on calcule pour chaque date de carte de flux le vecteur projeté x_{proj} (toujours en effectuant, composante par composante, la moyenne arithmétique des mesures directes disponibles).

3. De x_{proj} on extrait les Q_i et les TBF_i, d'une part, et on retrouve $PTHREAC$ et les $PIPAN$ d'autre part : on obtient ainsi un vecteur de paramètres.

4. On utilise le modèle physique pour reconstruire un vecteur d'état complet – physiquement cohérent, et donc potentiellement différent de x_{proj} – à partir dudit vecteur de paramètres.

On obtient ainsi une séquence de vecteurs d'états x physiquement cohérents, reconstruits à partir de données réelles et formant donc une trajectoire plausible d'un réacteur

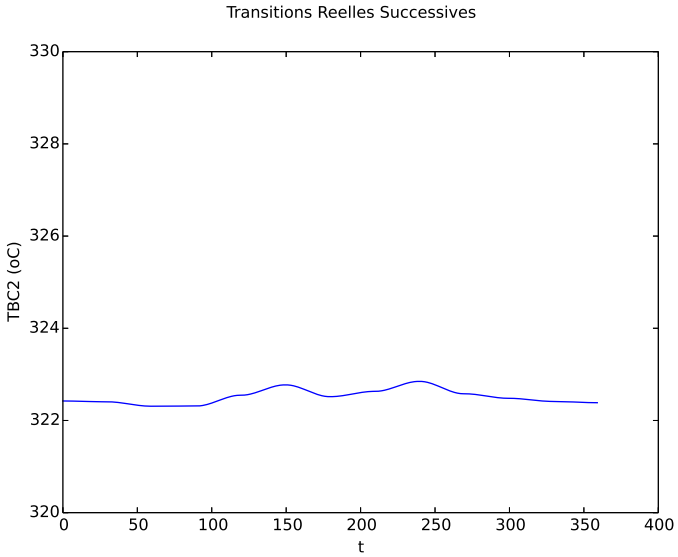


FIGURE 4.13 – Exemple de simulateur d'évolution par transitions réelles successives

au cours d'un cycle. A partir de là, on reprend notre idée précédente d'interpoler entre ces différents états, définissant plusieurs transitions successives entre les étapes :

$$x_{t,k} = (1 - \omega(t_k, t_{initial}))x_{initial} + \omega(t_k, t_{initial})x_{final} \tag{4.6}$$

de même que précédemment, on définit ainsi des transitions continues de 30 jours entre les différents états de la séquence de vecteurs d'états choisie. La trajectoire ainsi définie correspond plus aux critères de réalisme des opérateurs : Q_i et TBF_i sont par exemple très stables dans le temps, et la séquence des $PIPAN$ constatés aux cartes de flux correspond de facto à une séquence réelle constatée au cours d'un cycle.

Modélisation des erreurs d'évolution

Dans le cadre d'un Filtre Kalman, une information nous manque : une estimation de l'erreur modèle Q, qui vient s'additionner à la covariance d'erreur d'analyse à t_{k-1} pour donner la covariance de l'erreur associée à la prévision pour t_k :

$$P_{f,k+1} = M_{k+1}P_{a,k}M_{k+1}^T + Q_k = P_{a,k} + Q_k \tag{4.7}$$

Intuitivement, on devine que cette erreur modèle tend vers zéro si le pas de temps est infiniment petit ; a contrario, si le pas de temps est très grand, on est ramenés au cas statique où l'on tire un vecteur au hasard, ce qui amène à prendre dans ce cas pour Q la valeur de la matrice B de covariance d'erreur d'ébauche du cas statique. Nous avons donc choisi de réutiliser la fonction de « diffusion » ω pour fournir une valeur pour l'erreur modèle entre t_k et t_{k+1} :

$$Q_k = min(\omega(t_k, t_{k+1}), 1)B_{TOT} \tag{4.8}$$

Si les temps (t_k, t_{k+1}) sont très proches (par rapport à l'intervalle de temps caractéristique d'évolution Δt, alors l'incertitude pesant sur la prévision sera ainsi très faible,

quasiment prise égale à l'incertitude de l'analyse précédente. Si au contraire ils sont très espacés, alors la prévision se verra attribuée une analyse élevée, plus élevée encore que l'incertitude d'ébauche du cas statique.

Cette démarche nous permet ainsi d'obtenir les termes nécessaires pour un Filtre de Kalman, dont nous pouvons évaluer l'efficacité par expérience jumelle. On remarquera toutefois que les données simulées par le simulateur construit plus haut ne vérifie pas les hypothèses du filtre de Kalman (les erreurs de prévision sont en effet très probablement fortement corrélées deux à deux) et qu'il n'y a aucune garantie sur le plan théorique que la matrice Q_k que nous avons construite soit une bonne estimation quantitative de l'erreur effective commise lors de la prédiction. Un simulateur qui vérifierait totalement les hypothèses le Filtre de Kalman ainsi implémentait ne serait autre que le processus de Wiener que nous avons précisément choisi de ne pas utiliser pour générer nos données pseudo-réelles.

En fin de compte, en paramétrant de la sorte notre Filtre de Kalman, nous avons choisi d'attribuer à notre connaissance de la dynamique de notre objet physique la même incertitude que pour un mouvement brownien. Ce choix interviendra directement dans la pondération entre les observations et les prédictions, et, de même que pour le choix de la matrice de covariance B pour le BLUE, présente un côté relativement arbitraire.

4.4.2 Cas normal

Algorithmes choisis

Nous cherchons à présent à tester l'efficacité de l'Interpolation Optimum (OI) et du Filtre Kalman, relativement au BLUE et à la projection. Pour ce faire, nous avons utilisé les vecteurs d'état simulés en 4.4.1 (en utilisant l'interpolation entre données réelles) et l'opérateur d'observations non biaisé utilisé dans le cas statique pour obtenir un jeu de données dynamiques.

On applique à ces données quatre techniques d'estimations :

— **Projection** : en effectuant composante par composante la moyenne des observations directement disponibles ;

— **BLUE** : en effectuant à chaque date une analyse BLUE avec des termes d'ébauche constants $(x_{b,TOT}, B_{TOT})$;

— **Interpolation Optimum (OI)** : en utilisant à chaque date une analyse BLUE avec une matrice de covariance d'erreur d'ébauche constante B_{TOT} et pour vecteur d'ébauche le résultat de l'analyse à la date précédente ;

— **Filtre de Kalman** : en utilisant à chaque date une analyse BLUE avec pour matrice de covariance d'erreur de prédiction la matrice de covariance d'erreur d'analyse précédente augmentée de l'erreur d'évolution (voir plus haut) et pour vecteur de prévision le résultat de l'analyse précédente.

On aborde l'analyse de ces données comme s'il s'agissait de données réelles : faute d'information préalable, il convient ainsi de choisir les termes d'ébauche les plus généraux possibles (pour le BLUE ou pour amorcer l'OI et le Kalman) : soit les termes $(x_{b,DEF}, B_{DEF})$, obtenus avec des $PIPAN$ simulés indépendamment les uns des autres, soit les termes $(x_{b,TOT}, B_{TOT})$, obtenus en simulant des $PIPAN$ qui reproduisent la statistique constatée sur l'ensemble des 1300 MW. Nous l'avons vu dans le cas statique, cette dernière option nous paraît un choix plus réaliste et plus performant.

Un bref coup d'œil nous permet cependant de constater immédiatement que l'erreur d'ébauche sera biaisée : la moyenne des vecteurs d'état est en effet différente de $x_{b,TOT}$. De plus, sur la totalité du cycle, la matrice de covariance d'erreur d'ébauche effective sera également différente de B_{TOT}. Pour le BLUE, on se retrouve ainsi dans le cas d'une « analyse croisée » de la section 4.2.1, et on s'attend donc à ce que les biais persistent après analyse et que les résultats s'en trouvent par conséquent dégradée. L'effet sur les algorithmes dynamiques (Kalman et OI) pourrait en revanche être moindre, dans la mesure où les observations sont ici non biaisées, et qu'au fur et à mesure des analyses la prédiction devrait donc être de moins en moins biaisée.

Accessoirement, l'optimalité du Filtre de Kalman requiert de plus l'hypothèse d'indépendance des erreurs de prédiction d'un pas de temps à l'autre : or cette hypothèse est vraisemblablement généralement fausse dans notre cas, du fait de la continuité des variations (en phase ascendante de puissance, par exemple, la prédiction pour les TBC, $THAC$ et $PTHGV$ sera systématiquement une sous-estimation de l'état réel).

D'autre part, contrairement au cas statique, où seul le bypass était totalement indisponible par mesure directe, ici les $THAC$ ne sont dans leur grande majorité mesurés qu'une fois tous les trente jours, aux dates des cartes de flux ; le reste du temps, seules les mesures $TCRIC$ sont disponibles pour la cinquantaine d'AC qui en sont équipés. L'estimateur « Projection » comporte ainsi bien plus de lacunes que les trois analyses.

Evaluation des performances

Enfin, si l'on peut toujours procéder à une étude de la dispersion statistique de l'erreur comme pour le cas statique, c'est un petit peu moins pertinent à présent car les performances des estimateurs varient dans le temps. Pour la Projection et le BLUE, la cause est essentiellement liée au nombres d'observations disponibles, mais pour le Kalman et l'OI, les termes d'ébauche varient également dans le temps, et les performances de l'analyse d'autant plus. On rappelle que, en théorie, les matrices de covariance d'erreur d'analyse sont calculables à l'avance, séquentiellement, dès lors que l'on dispose des termes d'ébauche initiaux, et de la suite des termes liés à l'observation H_k, R_k.

Nous estimons ainsi dans cette section les performances des analyses de façon un peu différente que précédemment : nous n'évaluons pas la dispersion statistique d'une composante donnée d'une vecteur d'état, mais suivons dans le temps l'évolution de l'erreur quadratique, par groupe de composantes. On évaluera ainsi, pour un estimateur x :

— L'erreur quadratique totale : $\Sigma_{i=1}^{n}(x_t[i] - x[i])^2$;

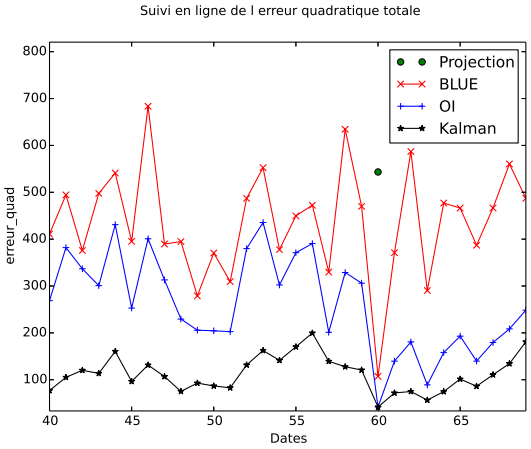


FIGURE 4.14 – Suivi de l'erreur quadratique totale au cours d'un mois de 30 jours (cas sans biais)

— L'erreur quadratique des Q : $\sqrt{\frac{1}{4}\Sigma_{i=1}^{4}(Q_t[i] - Q[i])^2}$;

— L'erreur quadratique des TBF : $\sqrt{\frac{1}{4}\Sigma_{i=1}^{4}(TBF_t[i] - TBF[i])^2}$;

— L'erreur quadratique des TBC : $\sqrt{\frac{1}{4}\Sigma_{i=1}^{4}(TBC_t[i] - TBC[i])^2}$;

— L'erreur quadratique des $THAC$: $\sqrt{\frac{1}{193}\Sigma_{i=1}^{193}(THAC_t[i] - THAC[i])^2}$;

— L'erreur quadratique des $PTHGV$: $\sqrt{\frac{1}{4}\Sigma_{i=1}^{4}(PTHGV_t[i] - PTHGV[i])^2}$;

— L'erreur absolue ϵ_{fit} : $\|\epsilon_{fit,t} - \epsilon_{fit}\|$.

L'erreur quadratique totale est la somme du carré des erreurs composantes ainsi définies.

L'erreur quadratique totale commise par la projection n'est a priori pas calculable (à cause d'ϵ_{fit}), mais aux dates des cartes de flux, les $THAC$ étant tous disponibles par mesure directe, on peut obtenir une erreur quadratique presque totale qui peut servir de bon indicateur. On rappelle que l'objectif essentiel de l'erreur quadratique totale est de nous permettre de vérifier si, dans leur globalité, les analyses sont plus performantes que les observations seules.

On peut également suivre composante par composante chaque estimateur et le comparer à l'évolution de l'état réel ; ou tracer le graphe de l'erreur commise au cours du temps.

Résultats

Ce sont les $THAC$ – par leur nombre et leur incertitude élevée – qui contribuent le plus à l'erreur quadratique. Une erreur quadratique faible montre ainsi généralement une carte des $THAC$ bien évaluée. Sur le graphique 4.14 on voit ainsi à la date de l'EP l'effet de la prise en compte des mesures RES sur les performances du BLUE, qui est moins performant

Grandeur	Projection	BLUE	OI	Kalman
Q	0.03 t/s	0.02 t/s	0.02 t/s	0.01 t/s
TBF	0.2 $°C$	0.3 $°C$	0.1 $°C$	0.1 $°C$
TBC	0.4 $°C$	0.2 $°C$	0.1 $°C$	0.1 $°C$
$THAC$	1.4 $°C$	1.2 $°C$	0.9 $°C$	0.7 $°C$
$PTHGV$	4.8 MW	5.2 MW	3.6 MW	2.5 MW
ϵ_{fit}	N.C.	0.3 %	0.2 %	0.2 %
$Total$	481	412	220	122

TABLE 4.13 – Erreurs quadratiques moyennées sur toute la durée d'un cycle (cas sans biais)

Grandeur	**BLUE**	**OI**	**Kalman**
Q	0.02 ± 0.04 t/s	0.01 ± 0.04 t/s	0.00 ± 0.03 t/s
TBF	-0.3 ± 0.2 $°C$	0.0 ± 0.2 $°C$	0.0 ± 0.2 $°C$
TBC	-0.1 ± 0.2 $°C$	0.0 ± 0.2 $°C$	0.0 ± 0.2 $°C$
$THAC$	-2.1 ± 1.9 $°C$	-0.8 ± 2.1 $°C$	-0.6 ± 1.5 $°C$
$PTHGV$	2 ± 8 MW	0 ± 8 MW	0 ± 7 MW
ϵ_{fit}	0.3 ± 0.5 %	-0.1 ± 0.5 %	-0.1 ± 0.4 %

TABLE 4.14 – Résultats des différentes analyses appliquées aux données dynamiques non biaisées (pour la boucle 1 et l'AC 3)

aux dates normales. En revanche, les performances des algorithmes dynamiques sont plus stables dans le temps et meilleures que celles du BLUE. Il semble que le Kalman soit plus performant que l'Interpolation Optimum.

Comme attendu, les algorithmes dynamiques (OI et Kalman) suivent mieux l'évolution de l'état réel que le BLUE, particulièrement dans le cas du $THAC_3$ qui n'est pas mesuré par un TC RIC. On distingue également clairement un biais pour le BLUE sur TBC_2, conséquence du biais d'ébauche. Les courbes des analyses sont nettement moins lisses que la courbe de l'état réel, conséquence du bruit blanc des observations qui sont prise en compte à chaque pas de temps.

Sur la totalité du cycle, et en moyenne sur les 4 boucles et les 193 AC, les indicateurs d'erreurs quadratiques montrent clairement la nette supériorité des performances des deux algorithmes dynamiques sur l'algorithme statique, tout particulièrement en ce qui concerne les THAC, ce qui constitue un résultat remarquable : même hors EP, on obtient une précision sur l'ensemble de la cartographie 2D des THAC en utilisant exclusivement les TC RIC et les mesures out-core. Le Tableau 4.13 paraît également confirmer une supériorité des performances du Kalman sur celles de l'Interpolation Optimum.

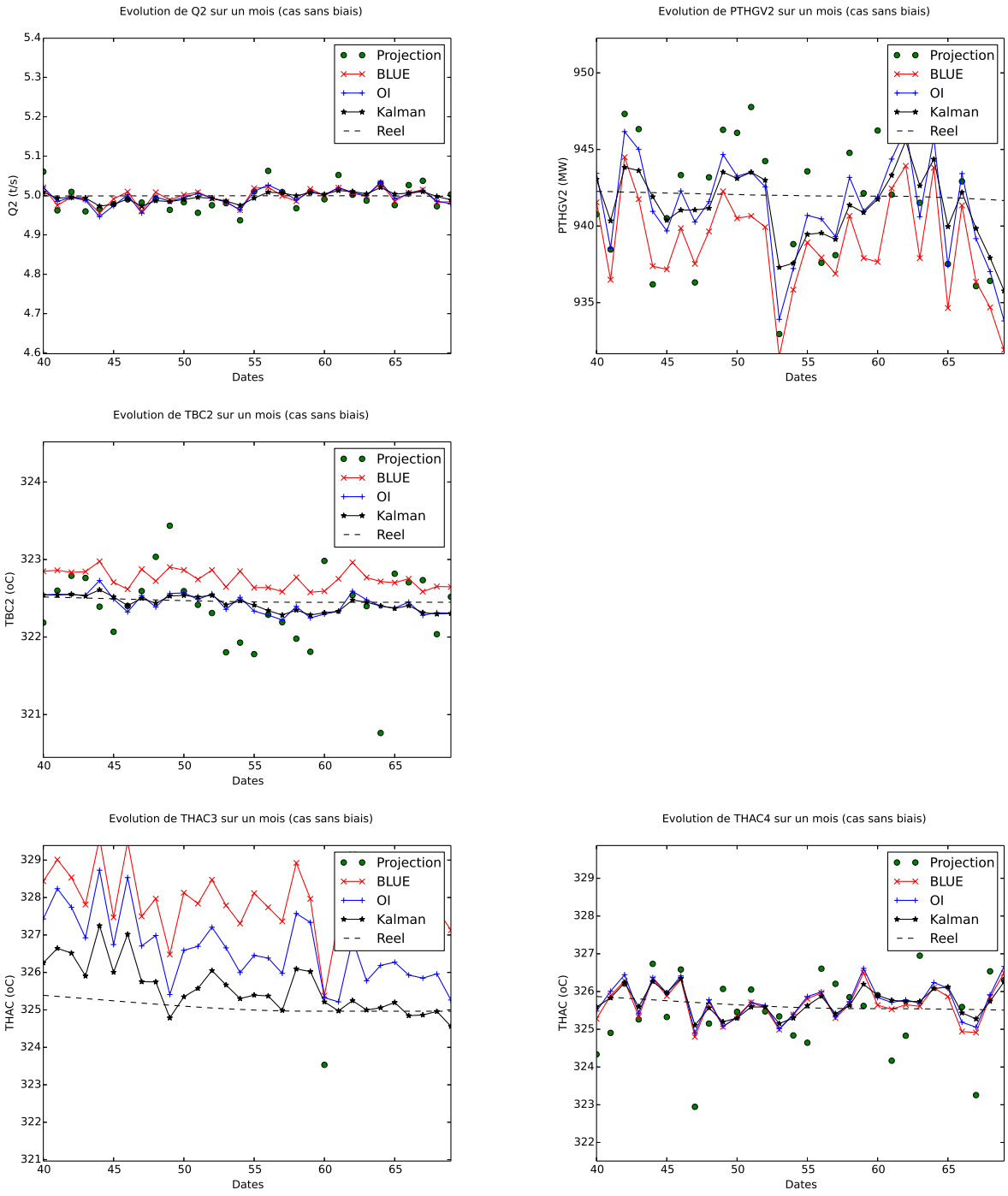


FIGURE 4.15 – Suivi de Q_2,TBC_2, $PTHGV_2$, $THAC_3$ et $THAC_4$ sur un mois de 30 jours par les estimateurs projection, BLUE, OI et Kalman (cas non biaisé)

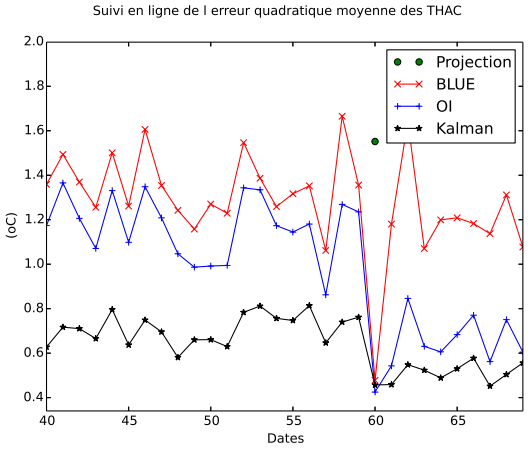


FIGURE 4.16 – Suivi de l'erreur quadratique sur les THAC au cours d'un mois de 30 jours (cas sans biais)

Les résultats de « debriefings » conventionnels confirment à nouveau cette tendance : on constate effectivement que les biais d'ébauche sont presque intégralement effacés par les algorithmes dynamiques, et la réduction d'incertitude finale sur les paramètres-clés (TBC, Q et $PTHGV$) est impressionnante, particulièrement dans le cas du Kalman.

4.4.3 Cas d'observations biaisées

Qu'en est-il à présent si les observations sont biaisées ? On simule un échantillon utilisant la même instance observations biaisées que dans la section 4.3.1. Nous l'avons vu, en ce qui concerne le BLUE, qu'il existe toujours un biais postanalyse (égal à $K\delta y$, où K est la matrice de Gain de Kalman et δy le biais d'observations). Or pour les algorithmes d'Assimilation de Données dynamiques séquentiels, comme l'OI et le Kalman, le résultat de l'analyse devient l'ébauche pour le pas de temps suivant : cette fois-ci, à chaque pas de temps, l'erreur de prédiction incorpore donc au fur et à mesure le biais lié à l'observation. On peut ainsi prédire que l'OI et le Kalman vont moins bien résister que le BLUE à l'insertion de ces biais d'observations.

L'explication est simple : si l'on mesure cent fois une température T avec un thermomètre biaisé de $+1°C$, et qu'on suppose que l'erreur commise est un bruit blanc, alors notre estimation de T va progressivement tendre vers $T+1$, un ensemble de nombreuses observations directes indépendantes étant théoriquement infiniment plus fiable que l'ébauche qu'on avait pu présupposer.

En termes vectoriels, pour l'interpolation optimum par exemple, entre deux EP (avec donc un opérateur d'observation constant H), la matrice de gain de Kalman est constante égale à K, et le biais suivra donc la suite suivante :

$$\delta x_0 = \delta x_b \tag{4.9}$$

$$\delta x_{k+1} = K \cdot \Delta y + (I - KH)\delta y_k \tag{4.10}$$

Grandeur	Projection	BLUE	OI	Kalman
CAS AVEC BIAIS				
Q	0.13 t/s	0.08 t/s	0.10 t/s	0.10 t/s
TBF	0.2 $°C$	0.3 $°C$	0.2 $°C$	0.1 $°C$
TBC	0.8 $°C$	0.5 $°C$	0.5 $°C$	0.5 $°C$
$THAC$	1.4 $°C$	1.2 $°C$	1.0 $°C$	0.8 $°C$
$PTHGV$	5.3 MW	7.1 MW	5.2 MW	4.2 MW
ϵ_{fit}	N.C.	1.2 %	1.3 %	1.2 %
$Total$	494	513	315	199

TABLE 4.15 – Erreurs quadratiques moyennées sur toute la durée d'un cycle (cas avec biais)

où δy est le biais (supposé constant) d'observations.

Remarque 7.

Un biais n'est constant que par rapport à une fenêtre donnée d'étude : on peut notamment remarquer que, étant donné notre mode de simulation de vecteurs d'états réels dynamiques (par transitions successives), au cours d'une période de transition, il existe un biais temporaire lié à la prévision : par exemple, si TBC est en phase ascendante, alors on commet systématiquement une erreur négative en la supposant constante. Sur la totalité du cycle en revanche (les phases ascendantes et descendantes se succédant aléatoirement, pour chaque composante) le biais n'existe pas.

Le comportement de la suite arithmético-géométrique matricielle ainsi définie n'est pas flagrant, d'autant que KH est rarement inversible.

Dans les faits, les algorithmes dynamiques résistent effectivement moins bien à l'introduction d'un biais constant d'observations que le BLUE :

Comme pour le cas statique, le problème principal se situe au niveau des débits-boucles Q et des $PTHGV$: les TBC et les $THAC$ résistent particulièrement bien La dégradation reste cependant minime, et – en moyenne sur les quatre boucles et les 193 AC – les analyses demeurent d'une qualité au moins comparable à la projection [2] et au BLUE. Les résultats sont ainsi assez encourageants.

La présence des biais est clairement visible lorsqu'on suit les trajectoires composante par composante (voir Figure 4.17) mais la précision relative des trois analyses n'est pas

2. On rappelle de plus que, contrairement aux analyses, l'erreur quadratique pour la projection n'est estimée que les jours des EP, une fois tous les trente jours, lors que tous les $THAC$ sont mesurés au moins via les cartes de flux RES. Les erreurs quadratiques pour les analyses sont elles calculées en moyennant sur la totalité des dates.

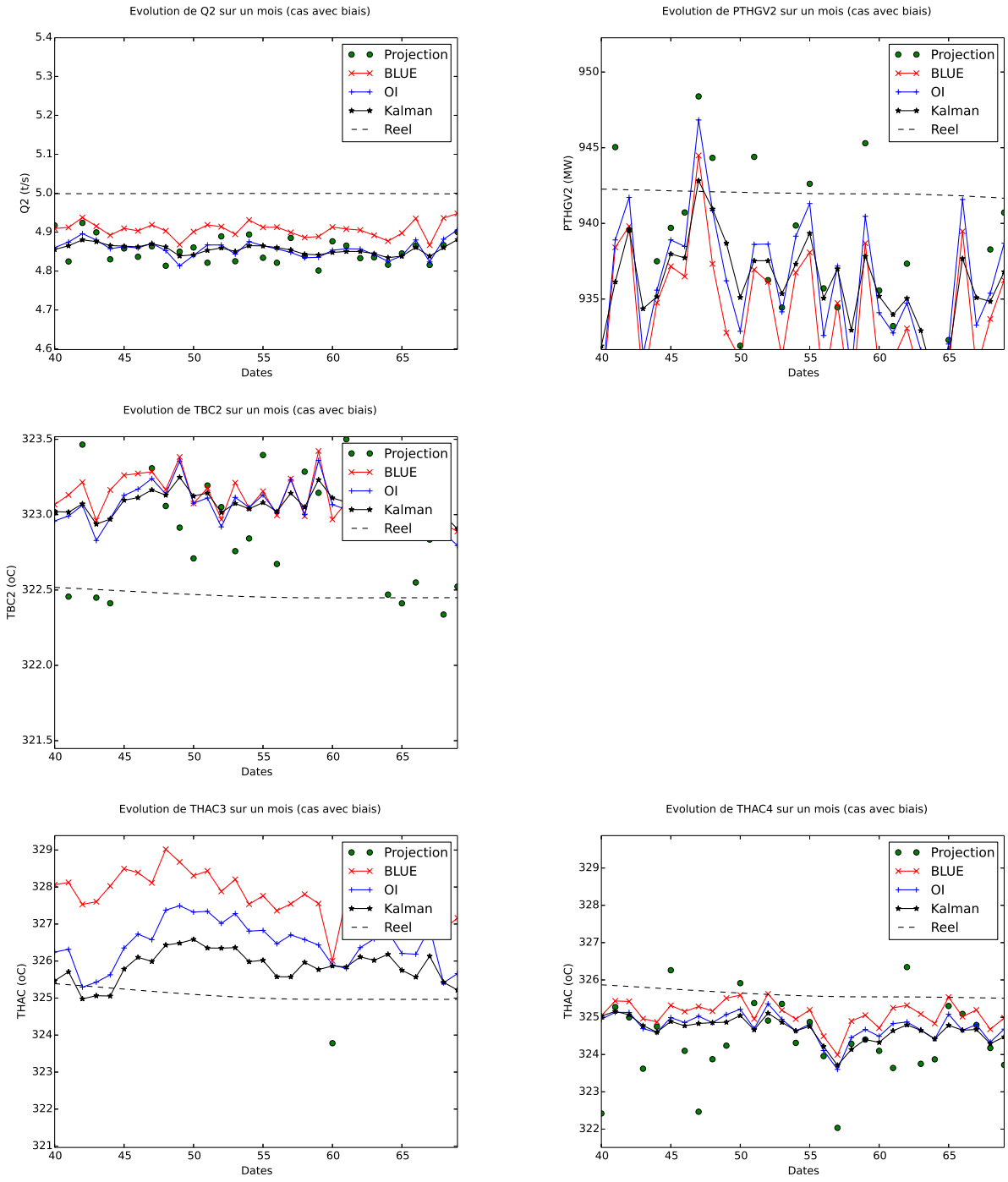


FIGURE 4.17 – Suivi de Q_2,TBC_2, $PTHGV_2$, $THAC_3$ et $THAC_4$ sur un mois de 30 jours par les estimateurs projection, BLUE, OI et Kalman (cas biaisé)

Grandeur	BLUE	OI	Kalman
Q	$0.02 \pm 0.04\ t/s$	$-0.01 \pm 0.04\ t/s$	$-0.01 \pm 0.03\ t/s$
TBF	$-0.2 \pm 0.2\ °C$	$-0.1 \pm 0.2\ °C$	$-0.1 \pm 0.2\ °C$
TBC	$-0.4 \pm 0.2\ °C$	$-0.2 \pm 0.2\ °C$	$-0.2 \pm 0.2\ °C$
$THAC$	$-1.9 \pm 1.8\ °C$	$-0.5 \pm 2.1\ °C$	$-0.4 \pm 1.7\ °C$
$PTHGV$	$-4 \pm 8\ MW$	$-5 \pm 8\ MW$	$-4 \pm 5\ MW$
ϵ_{fit}	$1.2 \pm 0.5\ \%$	$1.3 \pm 0.5\ \%$	$1.2 \pm 0.4\ \%$

TABLE 4.16 – Résultats des différentes analyses appliquées aux données dynamiques biaisées (pour la boucle 1 et l'AC 3)

des plus flagrantes.

Pour finir, on étudie les statistiques des erreurs d'analyses pour les différentes composantes (boucle 1, AC numéro 3).

Les résultats de « débriefing » confirment à la fois la persistance des biais post-analyse et la proximité des performances des trois différentes analyses (pour la boucle et l'AC concerné du moins). Comme pour le cas statique, le principal problème se situe au niveau de l'estimation du débit-boucle Q et de la puissance thermique transférée au secondaire $PTHGV$.

L'effet le plus pénalisant pour le BLUE reste celui du biais d'ébauche : lorsqu'on supprime la majeure partie de ce dernier (en prenant pour paramètres $PTHREAC$ et $PIPAN$ la moyenne de ceux constatés sur l'échantillon de cartes de flux qui nous a servi à simuler les données dynamiques et en les injectant dans le modèle physique avec des données homogènes type pour les débits-boucles Q et les TBF) alors le BLUE devient nettement meilleur que les algorithmes dynamiques. Cependant, dans les faits, on peut avoir des réticences à effectuer un tel calibrage lors de l'étude de données réelles, car cela implique de coupler au moins partiellement les observations et l'ébauche.

Un compromis peut être d'extraire le $PTHREAC$ et le $PIPAN$ de la première carte de flux disponible à pleine puissance, de les utiliser (avec les Q et TBF types homogènes (5.0 t/s et 290.0 ^{o}C pour toutes les boucles) comme paramètres d'entrée pour le modèle phyisuqe 0D, calculant ainsi un vecteur d'état correspondant qui nous servirait de vecteur d'ébauche et qui serait ainsi mieux centré. En retirant les mesures de cette carte de flux de la liste des observations à assimiler, on limiterait dans le même temps la corrélation entre erreur d'ébauche et erreur d'observation.

En suivant cette démarche, le BLUE donne alors dans ce cas des résultats de qualité comparable à ceux de l'OI et du Kalman sur le reste des données dynamiques.

4.4.4 Bilan

Tester l'efficacité des algorithmes d'Assimilation de Données pour la réduction des incertitudes sur les paramètres REP, via des expériences jumelles, est plus délicat que lors du cas statique, car il est nécessaire de simuler des trajectoires-tests de vecteurs d'état, et non

plus de représenter approximativement leurs distributions de probabilités gaussiennes par des échantillons de vecteurs d'état test. Nous avons choisi de construire une trajectoire-test à partir de données réelles, puis de simuler des observations pseudo-réelles correspondantes.

L'avantage principal des algorithmes dynamiques comme le Kalman ou l'Interpolation Optimum par rapport à un algorithme statique comme le BLUE est que le biais éventuel d'ébauche est rapidement effacé au fur et à mesure que la trajectoire est recalée sur des données réelles.

Enfin, si la réduction d'incertitudes permise par les algorithmes dynamiques est particulièrement impressionnante lorsque les observations sont non biaisées, en revanche, l'introduction de biais d'observations les pénalise légèrement par rapport au BLUE, car l'effet des biais s'additionne partiellement à chaque pas de temps : de même que pour le cas statique, les biais sur les débits-boucles et les biais de BIL100 sont ceux qui détériorent le plus les résultats.

L'étude de ce cadre dynamique nous a permis enfin de rappeler que l'Assimilation de Données, en utilisant au mieux les observations disponibles pour corriger une estimation préalable, permet de produire in fine un vecteur d'état physiquement cohérent et globalement plus juste qu'en les utilisant directement et séparément (comme dans le cas de l'estimateur « Projection »), d'autant que les observations ne sont pas toutes disponibles en même temps.

Ainsi, on dispose par exemple grâce à l'Assimilation de Données d'un suivi en ligne précis de la cartographie des $THAC$, même en dehors des cartes de flux où la plupart de ces températures ne sont pourtant pas mesurées.

4.5 Synthèse

Nous avons appliqué avec succès des algorithmes simples d'Assimilation de Données à l'estimation des paramètres primaires d'un REP, en utilisant un modèle physique simple semi-empirique de type 0D et une méthode de Monte-Carlo soigneusement calibrée pour paramétrer les termes d'ébauche correspondant, conformément à la méthodologie développée au Chapitre 3. Nous avons notamment appliqué, dans le cas statique, le BLUE et le 3DVAR, et dans le cas dynamique, le BLUE, l'Interpolation Optimum et le Filtre de Kalman Linéaire.

Bien que la matrice de covariance d'erreur d'analyse puisse être obtenue directement en théorie, il nous est apparu nécessaire d'évaluer la performance de ces analyses via des expériences jumelles, c'est-à-dire en utilisant des données simulées, pour pouvoir confronter directement les résultats des différents estimateurs aux états qu'ils sont censés estimer. Pour simuler ces données pseudo-réelles, il convient d'adopter les mêmes précautions que lors de la méthode de Monte-Carlo qui nous a servi pour calculer les termes d'ébauche, afin de pouvoir considérer que la gémellité entre les données simulées et les données réelles est suffisamment forte pour admettre que les performances des algorithmes sur les unes soient comparables sur les autres.

Dans ce cadre d'expériences jumelles, les résultats de ces premières applications d'As-

similation de Données sont satisfaisantes :

— Globalement, l'erreur d'analyse est effectivement réduite par rapport à l'erreur d'ébauche et l'erreur d'observations, et le vecteur d'analyse vérifie bien les propriétés physiques imposées par le modèle 0D et contenues dans la forme spectrale de la matrice de covariance d'erreur d'ébauche.

— Composante par composante, la réduction de l'incertitude – par rapport à celle de chaque instrument de mesure directe – peut être spectaculaire : c'est notamment le cas pour les paramètres-clés Q (de $\pm 3\% QN$ à $\pm 1\% QN$), TBC (de $\pm 1.5\ ^\circ C$ à $\pm 0.3\ ^\circ C$) et $PTHGV$ (de $\pm 10\ MW$ à $\pm 8\ MW$).

— Les incertitudes sur des composantes non directement mesurées du vecteur d'état sont également – quoique plus modérément – réduites : c'est le cas par exemple du coefficient d'ajustement ϵ_{fit} du modèle 0D, ou des $THAC$ pour les AC non équipés de TC RIC et entre les dates de cartes de flux.

— Dans le cas dynamique, le Filtre de Kalman et l'Interpolation Optimum améliorent encore les performances, notamment en ce qui concerne le suivi en ligne de la cartographie des $THAC$ entre les cartes de flux (erreur quadratique moyenne sur l'ensemble des $THAC$ divisée par deux pour l'analyse par rapport à la projection).

Cependant, la qualité des résultats est nettement dégradée lorsque l'on introduit des biais, que ce soit un biais d'ébauche dans le cas statique, ou un biais d'observations dans les deux cas, statique et dynamique : un biais d'observation sur un capteur $\Delta P - Coude$, de l'ordre de $2\% QN$, détériore ainsi nettement la qualité de l'analyse non seulement pour la composante Q de la boucle concernée, mais aussi pour les composantes $PTHGV$ et TBC de la boucle concernée.

Par conséquent, nous insistons sur la nécessité de prétraiter soigneusement les données d'entrées afin de limiter le plus possible ce type de biais d'observations, quitte à écarter du système d'observations une sonde de mesure qui semble suspecte, d'autant que les performances – même dans le cas sans biais – ne s'en trouvent que modérément dégradées. Concernant le biais d'ébauche, il est possible de le limiter en utilisant une première indication de $PTHREAC$ et du vecteur des $PIPAN$ sur le cycle, par exemple en retirant la première carte de flux à pleine puissance de la liste des données à analyser et en l'utilisant pour calculer ces paramètres, ce qui améliore nettement les performances du BLUE : il est à noter que, dans le cas dynamique, le Kalman et l'Interpolation Optimale sont quant à eux moins sensibles au choix du vecteur d'ébauche initial, qui est corrigé au fur et à mesure par la prise en compte des informations disponibles.

Conclusions

4.6 L'AD : un outil polyvalent

Appliquée le plus souvent à des problèmes de géosciences comme la météorologie ou la chimie atmosphérique, l'Assimilation de Données est un outil polyvalent qui peut être appliqué à d'autres systèmes ou processus, et donc d'autres types de problème d'estimation, comme par exemple l'estimation du point de fonctionnement primaire d'un REP.

L'Assimilation de Données consiste en la correction d'une estimation à l'aide d'observations réelles. Le calcul de la correction est paramétré à partir de la connaissance des statistiques des différentes erreurs, d'une part – ce qui prend généralement la forme des matrices de covariances R et B –, et du lien qui relie l'état du système aux observations attendues – ce qui prend la forme de l'opérateur d'observations H. Evaluer ces paramètres est la clé de la mise en place de l'Assimilation de Données : ce sont eux qui contiennent notamment les informations liées au modèle physique utilisé.

Dans ce travail de thèse, nous avons développé une méthodologie générale pour utiliser un modèle physique direct [3], de type $\Phi : \underline{u} \rightarrow \underline{x}$, dans un cadre d'assimilation de données. Le principe général, lequel s'inspire des méthodes d'ensemble, est d'utiliser une méthode Monte-Carlo pour évaluer la statistique de l'erreur d'ébauche : on simule un jeu de $\{\underline{u}\}$, on en déduit un échantillon de vecteurs d'états $\{\underline{x}\}$, et on approxime la distribution statistique de cet échantillon par une distribution gaussienne d'espérance x_b le moment d'ordre 1 de cet échantillon, et de covariance B le moment d'ordre 2.

Cette méthode utilise notamment le fait que la distribution statistique du paramètre $\underline{u}$ est plus facile à modéliser que celle du vecteur d'état $\underline{x}$, ce qui est le cas lorsqu'on suppose que les différents paramètres scalaires sont indépendants deux à deux. Lorsque ce n'est pas le cas, il est nécessaire de calibrer le randomisateur de paramètres, et nous avons proposé une méthodologie qui le permet, en effectuant en quelque sorte le cheminement inverse de la Méthode Monte-Carlo : à partir d'une matrice de covariance, calculée grâce à un échantillon de paramètres observés dans la réalité, il est possible de concevoir un méta-modèle émulant le modèle physique « caché » dans cette matrice de covariance, et d'utiliser ce méta-modèle pour construire un randomisateur reproduisant la statistique constatée pour ce paramètre.

Nous avons montré l'équivalence de cette méthode Monte-Carlo avec un changement

3. Pour mémoire, ce modèle physique vise à représenter les échanges de chaleur dans un circuit primaire de type REP : production de chaleur, circulation du fluide et échange vers le circuit secondaire (eau/vapeur). Le point de fonctionnement d'un tel circuit est défini par l'ensemble des paramètres mesurables (débits, températures et puissances) plus une variable d'ajustement (le coefficient ϵ dit de bypass).

de variable d'état, dans le cas d'un modèle linéaire ou linéarisable : en effet, dans ce cas, pratiquer l'Assimilation de Données directement sur l'espace des paramètres (en utilisant $H \circ \Phi$ comme opérateur d'observations) conduit formellement au même vecteur d'analyse.

L'avantage de cette méthodologie réside également dans le fait que l'utilisateur visualise mieux les hypothèses relatives aux différentes incertitudes et au modèle physique : en effet, au lieu de « cacher » ces hypothèses dans des matrices de covariances – pour schématiser, les incertitudes dans les termes diagonaux, et le modèle physique dans les termes non diagonaux et donc la forme spectrale –, ces hypothèses sont directement exprimées sous la forme de distributions de probabilités de paramètres scalaires, et un modèle direct.

En résumé, nous avons au cours de ce travail développé une méthodologie facile d'accès, permettant d'effectuer simplement une application des algorithmes d'Assimilation de Données à tout problème d'estimation utilisant un modèle direct.

4.7 Cadre semi-empirique pour le Circuit Primaire

Le modèle direct que nous avons utilisé dans le cadre de cette méthodologie d'Assimilation de Données combine les différents estimateurs semi-empiriques que EDF R&D / Département STEP a développé au cours des années précédentes pour permettre à l'opérateur de réduire l'incertitude pesant sur les paramètres primaires. Chacun de ces estimateurs liaient différents paramètres primaires entre eux, permettant d'estimer l'un à partir de mesures de quelques autres : notre modèle semi-empirique – de type 0D-1D – exploite ces liens pour déduire l'ensemble de l'état du circuit primaire à partir de quelques paramètres d'entrées (débits-boucles, températures de branches froides, carte de flux, puissances).

Nous avons construit ce modèle direct de façon modulaire pour laisser la porte ouverte à l'utilisation ultérieure de codes de calculs industriels (THYC, Code_Saturne...) en remplacement de certains de ces estimateurs.

Au cours du travail de thèse, nous avons notamment mené une réflexion sur la signification et l'estimation du paramètre ϵ de la moyenne par quarts de cœur utilisée pour modéliser le brassage de température au sein du plénum supérieur, et perfectionné notre modèle OD/1D en ajoutant une équation de bouclage déterminant, en fonction des autres paramètres d'entrées, la seule valeur possible pour ce paramètre qui satisfasse le bilan de puissance sur le circuit primaire.

Finalement, ce modèle 0D/1D nous semble un compromis acceptable entre complexité et représentativité de la physique du circuit primaire, et peut être utilisé tel quel par l'exploitant pour étudier, en première approche, l'état de la tranche et détecter, par exemple, telle ou telle dérive de mesure.

4.8 Résultats des Expériences Jumelles

Nous avons testé la méthodologie développée pour appliquer plusieurs algorithmes simples d'Assimilations de Données (BLUE, 3DVAR, Kalman, Interpolation Optimum) au problème d'estimation primaire en utilisant notre modèle semi-empirique. La nécessité

d'évaluer les performances de cette méthodologie via des expériences jumelles est apparue rapidement : bien que l'erreur post-analyse soit calculable théoriquement, il est préférable de confronter directement les résultats de l'analyse aux états qu'ils sont censés estimer, ce qui est impossible en travaillant directement sur des données réelles, mais immédiat en travaillant sur des données simulées.

Les performances calculées lors des expériences jumelles ne peuvent être supposées identiques aux performances sur données réelles que si l'on considère que le degré de gémellité entre données simulées et données réelles est suffisant. Cela suppose notamment que le simulateur utilisé est suffisamment réaliste, tant en termes de vraisemblance physique qu'en terme de dispersion statistique : ces hypothèses étant précisément celles qui sont nécessaires pour valider la méthode Monte-Carlo qui a servi a évaluer les termes d'ébauche.

Dans l'ensemble, les résultats des premières applications d'Assimilation de Données sont plus qu'encourageants : l'erreur d'analyse est globalement réduite par rapport à l'erreur d'ébauche et l'erreur d'observation, et le vecteur d'analyse vérifie bien les propriétés physiques élémentaires qu'on attend de lui (et qui lui sont imposées par le modèle OD via la forme spectrale de la matrice de covariance d'erreur d'ébauche). Les paramètres-clés Q (de $\pm 3\% QN$ à $\pm 1\% QN$), TBC (de $\pm 1.5\ ^\circ C$ à $\pm 0.3\ ^\circ C$) et $PTHGV$ (de $\pm 10\ MW$ à $\pm 8\ MW$). Certaines composantes non directement mesurées peuvent voir également leur estimation améliorée, grâce à la prise en compte de corrections induites par les observations d'autres variables qui leur sont statistiquement corrélées : c'est le cas par exemple des $THAC$ non équipés de TC, entre les cartes de flux.

Nous avons appliqué la méthodologie de calibrage de randomisateur aux $PIPAN$, construisant à partir d'un échantillon de cartes de flux donné un méta-modèle (émulant le code CAMARET dont les cartes de flux sont les outputs) permettant de simuler des puissances neutroniques dont la forme et la statistique correspond mieux à celle constatée sur les tranches réelles, et produire ainsi à la fois des données pseudo-réelles plus proches de la réalité des termes d'ébauches plus pertinents.

En météorologie, l'Assimilation de Données est essentiellement utilisée dans un cadre dynamique : nous avons cherché à nous placer dans un cadre similaire pour pouvoir appliquer notamment le Filtre de Kalman, ou l'Interpolation Optimum, par exemple, à notre problème d'estimation primaire. La difficulté principale à laquelle nous nous sommes heurtés est le manque de connaissance et de prédictibilité de la dynamique du circuit primaire sur les échelles de temps considérées (sur quelques jours ou sur tout un cycle), ce qui rend délicats tant la simulation de données-tests dynamiques que le calibrage des incertitudes paramétrant les différents algorithmes. Les résultats dans ce cadre sont donc essentiellement à interpréter qualitativement : dans ce cadre, les algorithmes dynamiques (Filtre de Kalman et Interpolation Optimum) semblent effectivement plus performants qu'un algorithme statique (comme le BLUE simple), notamment en ce qui concerne le suivi en ligne de la cartographie des $THAC$ entre les cartes de flux.

Dans tous les cas, on constate que la qualité des résultats se trouve nettement dégradée lorsque l'on introduit des biais (d'ébauche ou d'observation) : un biais d'observation sur un capteur de débit ΔP-Coude détériore ainsi nettement la qualité de l'analyse non seulement pour l'estimation du débit-boucle, mais aussi pour les autres composantes de la boucle concernée, comme la température de branche chaude ou la puissance thermique du

GV. C'est l'inconvénient de souder via, le modèle, les composantes du vecteur d'état entre elles : lorsque les mesures sont non biaisées, c'est ce qui contribue à la bonne réduction de l'incertitude, mais en contrepartie il suffit qu'une mesure soit grandement biaisée pour que toute l'estimation s'en trouve faussée.

Nous avons détaillé au Chapitre 2 l'ensemble des prétraitements indispensables pour les données d'entrées (estimateurs et observations) : l'étude de l'influence des biais sur la qualité de notre analyse ne fait que mettre davantage en relief la nécessité de ces prétraitements, afin de limiter le plus possible ce type de biais d'observations. Il est préférable d'écarter du système d'observations un sonde de mesure qui semble suspecte : en effet, même si la sonde n'était pas biaisée, la performance de l'analyse ne s'en trouve généralement que peu diminuée.

Enfin, pour le choix du paramétrage des termes d'ébauche, il faut garder en mémoire l'hypothèse d'indépendance statistique entre erreur d'ébauche et erreur d'observations, qui pousse à utiliser le moins possible les observations présentes de la tranche étudiée pour calibrer la méthode de Monte-Carlo qui permet d'obtenir les termes d'ébauche : s'il est important d'utiliser notre connaissance de l'état de la tranche pour produire les termes d'ébauche les plus plausibles possibles – et limiter notamment les biais d'ébauche – il est tout aussi important de préserver le caractère *a priori* de cette ébauche et son indépendance.

D'une façon plus générale, si l'Assimilation de Données est un outil précieux et peut grandement contribuer à la réduction d'incertitudes, ce n'est pas non plus une solution miracle, et il faut conserver un esprit critique lors de sa mise en application et de l'étude de ses résultats.

4.9 Perspectives

Suite aux avancées effectuées au cours de cette thèse et à la démonstration du potentiel de l'Assimilation de Données pour la réduction des incertitudes, STEP a décidé de poursuivre dans cette voie pour d'autres problèmes d'estimations, notamment ceux relatifs au circuit secondaire. Ainsi, le BIL100 / BILORLI que nous avons considérée, du point de vue de notre circuit primaire, comme une mesure extérieure, correspondrait au résultat d'une analyse par AD de l'ensemble des mesures disponibles côté secondaire, suivant la même méthodologie que celle que nous avons adoptée dans cette étude pour le circuit primaire. On peut également tout à fait envisager d'unifier l'étude du circuit primaire et secondaire, et de procéder simultanément à l'analyse de l'ensemble des données des deux circuits. Naturellement, le modèle physique à utiliser serait d'autant plus complexe.

L'utilisation de codes industriels de calcul pour remplacer ou suppléer le modèle semi-empirique simplifié est également à l'étude. L'un des meilleurs candidats, au moins dans un premier temps, semble être Dymola (sous langage Modélica) : de nombreux modèles 0D/1D utilisés par STEP existent sous Dymola, qui pourraient tout à fait être utilisés dans un cadre d'Assimilation de Données. L'enjeu est de pouvoir réaliser facilement l'interface informatique entre les programmes réalisant l'approche d'Assimilation de Données (par randomisation de paramètres, calcul matriciel, etc) et les codes en langage Modelica qui contiennent le modèle et ses méthodes de résolution. Une fois cette interface effectuée, on

pourra utiliser les modèles Dymola dans un cadre d'assimilation de données, que ce soit en l'utilisant comme modèle direct dans la méthode Monte-Carlo (comme fonction Φ) soit en l'utilisant comme opérateur d'observation non linéaire (comme $H \circ \Phi$) et assimiler sur ses paramètres d'entrées.

Une étude est également en cours visant à mieux évaluer, par Code_Saturne, le transport de la carte de température en haut des Assemblages Combustibles (les $THAC$) en branche chaude, ce qui permettrait soit de remplacer à la méthode de moyenne par quarts de cœur, soit de définir un opérateur d'observations permettant de relier les mesures effectuées au niveau des écopes de la branche chaude (mesures RCP108, RCP113, etc.) au différents $THAC$.

L'estimation détaillée de l'état 3D à l'intérieur du cœur, via l'utilisation soit de THYC, soit d'un modèle type « advection-diffusion » (cf. le modèle α, β brièvement décrit au Chapitre 2) pourrait également faire l'objet d'une approche d'Assimilation de Données. On pourrait ainsi évaluer l'apport de l'assimilation simultanée des mesures RPN, CFM, par TC RIC, etc....en combinaison avec un modèle 3D ou $193 \times 1D$ sur la réduction des incertitudes in-core. On risque toutefois d'être confronté à une explosion de la dimension de l'espace d'état, particulièrement dans le cas de l'utilisation de THYC qui requiert un maillage raffiné, et être donc amené par exemple à des méthodes d'ensemble.

Les hypothèses relatives à la quantification des incertitudes de mesures peuvent être améliorées. Nous avons postulé dans cette étude que les erreurs d'observations scalaires était décorrélées deux à deux, c'est-à-dire que la matrice R pouvait être construite diagonale : cette hypothèse peut vraisemblablement être remise en cause, et mener à la construction de termes non diagonaux reflétant d'éventuelles corrélations connues entre les différentes erreurs de mesures. De plus, le cadre théorique de l'Assimilation de Données et la démonstration de son optimalité reposent sur des hypothèses de gaussiannités des distributions de probabilités des erreurs, et les incertitudes sont directement associées aux moments d'ordre 2 correspondant. L'utilisation d'autres types de distributions de probabilités relève de l'Estimation Bayésienne, et nécessite des connaissances métrologiques approfondies pour l'ensemble des capteurs et de leurs chaînes d'acquisitions respectives.

Enfin, l'Assimilation de Données pourrait être mise à contribution pour une surveillance en continu du cœur à partir des mesures des thermocouples in-core : en effet, ces 48 mesures pourraient être assimilées pour élaborer 193 valeurs sur les assemblages combustible. Cet objectif tirerait profit des techniques mises en œuvre pour améliorer les mesures primaires existantes. Une surveillance nouvelle, proche du temps réel, serait ainsi à expérimenter.

D'une façon générale, notre application de l'Assimilation de Données à un problème industriel comme celui de l'estimation du point de fonctionnement primaire d'un REP constitue un rapprochement entre le monde des Géosciences et celui de l'Ingénierie Industrielle. Trois publications ont été effectuées lors de cette thèse : un poster à l'occasion de la *Summer School* sur l'Assimilation de Données et les Problèmes Inverses à l'Université de Reading (Royaume-Uni) en 2013 [Mer], un article pour la conférence ICONE-22 à Prague (République Tchèque) en 2014 [MFG14], et un article pour la conférence ICONE-23 à Chiba (Japon) en 2015 [MFG15]. A chaque occasion où j'ai pu présenter mes travaux de thèse, j'ai pu mesurer l'intérêt suscité de part et d'autre : chez les spécialistes de l'Assimilation de Données, qui y voient un nouveau panel de cas d'applications, et chez les ingénieurs, qui y découvrent un nouvel outil malléable et susceptible de pouvoir répondre à leurs besoins.

Annexe A

Structure des Programmes

Nous décrivons ici succinctement les programmes PYTHON (version 2.7.6) réalisés pour la thèse. PYTHON est un langage de programmation objet, placé sous une licence libre, et conçu pour optimiser la productivité des programmeurs en offrant des outils de haut niveau et une syntaxe simple à utiliser. PYTHON dispose d'une grande bibliothèque standard, fournissant des outils convenant à de nombreuses tâches diverses, allant de l'algèbre linéaire à l'interface graphique.

A.1 Modules importés

A.1.1 NUMPY

NUMPY est l'une des principales extensions de PYTHON, destinée à manipuler des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux.

Tous les vecteurs colonnes (vecteurs d'état, d'observations, vecteur des PIPA, etc.), ainsi que les opérateurs linéaires d'observation ou les matrices de covariance, sont enregistrés comme des objet NUMPY de type matrices. On utilise d'autre part plusieurs fonctions intégrées NUMPY, comme l'inversion matricielle, le calcul du rang, des fonctions numériques (fonctions $\sqrt{}$, ln, exp, sin, etc.) ou des outils de calculs de moyenne et de covariance.

A.1.2 XLRD et XLWT

Les modules XLRD et XLWT permettent respectivement de lire et d'écrire des feuilles de calcul Excel.

Leur rôle est ainsi primordial : ils permettent de réaliser une interface facile d'utilisation entre les programmes PYTHON et l'utilisateur. Les données extraites des bases de données ORLI ou MOLENE le sont généralement sous forme de tableaux Excel ; de plus, disposer sous forme de feuilles de calcul des différents résultats d'analyse permet de vérifier rapidement la cohérence des estimations.

A.1.3 Matplotlib

Matplotlib est une bibliothèque de modules Python permettant de tracer et de visualiser des données sous formes de graphiques, et de les exporter en de nombreux formats matriciels (PNG, JPEG, etc.) ou vectoriels (PDF, SVG, EPS, etc.).

A.1.4 SciPy

SciPy est un projet visant à unifier et fédérer un ensemble de bibliothèques Python à usage scientifique : SciPy utilise ainsi notamment NumPy et Matplotlib, mais pas seulement.

On utilise essentiellement SciPy pour les modules suivants :

— optimize, pour les algorithmes de minimisation ;

— stats, pour une meilleure estimation des distributions de probabilités d'erreurs ;

— linalg, pour les diagonalisations en base orthonormée.

A.1.5 random

Le module random contient des générateurs de nombres pseudo-aléatoires pour différent type de distributions.

On l'utilisera essentiellement pour obtenir des variables aléatoires scalaires de distributions normales centrées : à partir de cela, on construit « à la main »des simulateurs de variables ou de vecteurs aléatoires gaussiens de distributions souhaitées.

A.2 Classes Principales

Quatre catégories de classes d'objets Python ont été construites : les classes modele, mesure, observations, et echantillon.

A.2.1 Modèle

Les classes modele implémentent les modèles physiques. Schématiquement, leurs méthodes principales s'organisent ainsi :

— modele(...) permet de créer une instance modele à partir de certains paramètres d'entrée ;

— modele.run() permet d'utiliser le modèle physique pour déduire toutes les composantes du modèles à partir des entrées qui y sont enregistrées ;

— modele.random(...) permet de simuler des paramètres d'entrée aléatoires pour le modèle ;

— modele.vecteur_etat() renvoie le vecteur d'état correspondant à l'état actuel du modèle.

A chaque modèle est associé une fonction de type simulation(...), qui permet d'utiliser le modèle pour simuler un vecteur d'état aléatoire, en créant une instance modèle , randomisant ses paramètres, résolvant le modèle, puis extrayant le vecteur d'état correspondant et le renvoyant. Cette fonction simulation(...) est celle utilisée notamment pour le cadre de la méthode Monte-Carlo.

Chaque classe modele correspond à un choix bien défini pour l'espace d'état. Il peut exister plusieurs méthodes de randomisation.

A.2.2 Mesure

La classe mesure permet de restituer le fonctionnement d'un instrument de mesure :
— mesure.etiquette correspond au nom de la mesure (e.g. : RCP114MT-) ;
— mesure.unit est l'unité correspondante (t/s, $°C$,MW,etc.) ;
— mesure.H contient l'opérateur (linéaire) d'observation associé à la mesure et pour un choix donné de vecteur d'état : si le vecteur d'état est x_t, alors la mesure prévue est $mesure.H \cdot x_t$;
— mesure.dispo est un booléen indiquant la disponibilité de la mesure ;
— mesure.valeur contient la valeur de la mesure ;
— mesure.incertitude et mesure.biais contiennent respectivement l'incertitude (2σ) et le biais associés à la mesure ;
— la méthode mesure.simulation(x_t) permet de simuler une mesure aléatoire de l'état donné x_t.

A.2.3 Observations

La classe observations permet de manier un système d'instruments de mesure. Il s'agit simplement d'une liste d'instances mesure, dotée de quelques méthodes particulières. Ainsi, pour une instance observations nommée obs, on a notamment :
— obs.operateur() construit l'opérateur d'observation linéaire correspondant à la totalité des mesures disponibles ;

— obs.projection() construit une estimation de l'état en effectuant, composante par composante, la moyenne des mesures directes ;

— obs.covariance() construit la matrice R de covariance d'erreur d'observations, selon les mesures disponibles et leurs incertitudes ;

— obs.simulation(x_t) simule un vecteur d'observations y correspondant à une observation de l'état x_t avec l'ensemble des mesures disponibles.

A.2.4 Echantillon

les classes de type echantillon sont les plus complexes. Elles collectent l'ensemble des données correspondant à un échantillon de données en cours d'étude et permettent de les

analyser. Une classe de type echantillon se construit à partir d'une classe modele donnée et d'une instance obs de type observations donnée qui sert de référence. Elle correspond donc à un choix déterminé pour l'espace d'état et les espaces d'observations possibles.

Une instance de type echantillon nommée ech contient :

— ech.liste_dates, la liste des dates [1] pour lesquelles il existe des données ;

Quatre dictionnaires ayant pour clés des dates et pour éléments de vecteurs d'états :

— ech.reel, pour les vecteurs d'état réels (x_t) ;

— ech.ebauche, pour les vecteurs d'ébauche (x_b) ;

— ech.projection, pour les vecteurs de projection (x_{proj}) ;

— ech.analyse, pour les vecteurs d'analyse (x_a).

Deux dictionnaires ayant pour clés des dates et pour éléments des listes (de taille fixe, égale au nombre d'instruments de mesures de obs) :

— ech.observations_valeurs, contenant pour chaque date les valeurs des mesures disponibles et 0.0 pour les mesures indisponibles ;

— ech.observations_dispos, contenant pour chaque date les disponibilités des différents instruments de mesures total de obs.

Ce système permet de travailler avec un nombre de mesures variable : à chaque date, on recense les mesures disponibles et leurs valeurs, et on met à jour en fonction l'instance obs pour en déduire le vecteur d'observations y, l'opérateur d'observations H, la matrice de covariance R, etc. . . .correspondant à la date donnée.

Les méthodes des instances de type echantillon sont nombreuses.

Méthodes d'interface

ech.importation(chemin) et ech.exportation(chemin) qui permettent de faire correspondre à l'échantillon ech le classeur Excel enregistré au chemin d'accès chemin. La mise en forme du classeur Excel doit respecter des critères très stricts. Succinctement :

— Sept feuilles de calcul, respectivement pour les dates, le réel, l'ébauche, la projection, l'analyse, les observations type ORLI et les observations type MOLENE ;

— La première ligne de chaque feuille de calcul contient la totalité des « étiquettes »liée soit aux composantes des vecteurs d'états, soit aux instruments de mesure ;

1. en format numérique Excel (nombre de jours depuis le 1er Janvier 1900) le 1er Mai 2012, par exemple, est ainsi le réel 41030.0

— La première colonne de chaque feuille de calcul doit contenir la totalité des dates pour lesquelles il existe des données.

Il suffit de créer un échantillon vide et de l'exporter pour obtenir un classeur Excel qui puisse servir de modèle.

La méthode ech.debriefing(...) est également une méthode d'interface : à partir d'un échantillon complet (dont les dictionnaires reel, ebauche, projection et analyse sont pleins), on produit les graphiques de distribution de probabilité des erreurs commises, ainsi qu'un fichier Excel synoptique résumant les incertitudes et biais constatés pour les différents estimateurs.

Méthodes de simulation

— ech.creation_dates(n) permet de générer une liste de dates de taille n et de l'enregistrer dans ech.liste_dates ;

— ech.simulation_reel(...) permet de simuler un jeu de vecteurs d'état réels pour l'échantillon. Concrètement, on remplit le dictionnaire ech.reel en associant à chaque date de ech.liste_dates un vecteur simulé grâce à la fonction simulation au modele de référence ;

— ech.simulation_reel_kalman(t_0,...) effectue la même opération, mais en simulant une évolution continue de l'état avec des variations de temps caractéristique t_0 ;

— ech.simulation_observations(obs) permet d'utiliser l'instance obs pour simuler un jeu d'observations associés aux vecteurs réels stockés dans l'échantillon : on remplit ainsi ech.observations_valeurs et ech.observations_dispos en fonction.

Ainsi, pour générer un échantillon de données simulées, on créer une instance echantillon, on crée un jeu de dates, on simule un jeu de vecteurs réels, puis on simule un jeu d'observations associées.

Méthodes d'analyse

— Estimateur « ébauche » : ech.calcul_ebauche() calcule le vecteur d'ébauche x_b et la matrice de covariance d'erreur associée B par méthode de Monte-Carlo, en utilisant les vecteurs stockés dans ech.reel, et ech.remplissage_ebauche(x_b) remplit le dictionnaire ech.ebauche avec pour chaque date le vecteur x_b ;

— Estimateur « projection » : ech.remplissage_projection(obs) remplit le dictionnaire ech.projection avec pour chaque date le vecteur x_{proj} calculé grâce aux observations enregistrées ;

— Estimateurs « statiques » : ech.BLUE(B) (resp. ech.analyse_3DVAR(B)) remplit le dictionnaire ech.analyse avec, pour chaque date, le résultat du BLUE (resp. du 3DVAR) pratiqué à partir de l'ébauche et des observations disponibles.

— Estimateurs « dynamiques » : ech.Kalman(...) (resp. ech.OI(...)) remplit le dictionnaire ech.ebauche avec les vecteurs prévisions et le dictionnaire ech.analyse avec les

résultats, pour chaque date, du Filtre de Kalman (resp. de l'Interpolation Optimum).

A.3 Autres modules

A.3.1 auxiliaires

Le module auxiliaires contient, comme son nom l'indique, une multitude de petites fonctions auxiliaires remplissant des rôles aussi divers que des calculs intermédiaires, la production de graphiques, des systèmes de sauvegardes de matrices, etc.

A.3.2 importationRES

Le module importationRES est consacré à l'extraction des données contenues dans les fichiers RES : date, puissance thermique, $PIPA$, $PIPAN$, Puissances Locales Normalisées, signaux des thermocouples, températures de branche chaude et froide, etc.

A.3.3 globales

Le module globales contient les principales variables globales qui sont celles utilisées, par défaut, par les différentes classes principales : il s'agit notamment des valeurs par défaut des principaux paramètres primaires et de leurs incertitudes. Il est ainsi facile de les modifier pour l'ensemble des programmes, simplement en les modifiant dans ce module.

A.3.4 thermodynamique

Le module thermodynamique contient simplement quelques fonctions numériques permettant d'évaluer les valeurs prises par certains coefficients thermodynamiques de l'eau à 155 bar. Ces fonctions polynômiales de degré 2 ou 3 ont été construites par interpolation de valeurs prises dans des tables de référence.

A.3.5 BLUE

Le module BLUE contient les formules matricielles du BLUE : calcul du gain de Kalman, du vecteur d'analyse et de l'erreur associée.

Annexe B

Publications

Nous incluons ici les deux publications qui ont été effectuées dans le cadre de cette thèse.

La première publication [MFG14] a été réalisée pour la conférence ICONE22, qui s'est déroulée à Prague en République Tchèque en Juillet 2014. Elle présente le problème d'estimation des paramètres primaires, le modèle simplifié 0D dans sa première version – sans bouclage du modèle par fit de bypass – et la méthode Monte-Carlo pour l'obtention des termes d'ébauche.

La deuxième publication [MFG15] a été réalisée pour la conférence ICONE23, en Mai 2015 à Chiba au Japon. Elle introduit la méthode de fit de bypass pour améliorer le modèle simplifié 0D, et présente une méthode de calibration du générateur de paramètres aléatoires, ainsi que les résultats de son application sur les Puissances Intégrées Par Assemblages Normalisées.

Il est à noter que le choix des différents paramètres de randomisation et différents points de détails concernant les codes PYTHON ont évolué au fil du travail de thèse, ce qui fait qu'il peut exister de légères différences entre les résultats numériques présentés lors de ces publications et ceux finalement validés dans le cadre du travail de thèse. Ces différences restes minimes et ne remettent pas en cause les conclusions générales des deux articles ni les méthodologies qu'ils présentent.

Proceedings of the 2014 22nd International Conference on Nuclear Engineering
ICONE22
July 7-11, 2014, Prague, Czech Republic

ICONE22-30118

FEASIBILITY OF ENHANCED ESTIMATION OF PWR PRIMARY PARAMETERS: TWIN EXPERIMENTS FOR ASSESSING DATA ASSIMILATION BENEFITS (0D-MODEL WITH A MONTE-CARLO APPROACH ON PSEUDO-REAL DATA)

Thibaud Mercier
PhD Student
EDF R&D (STEP)
6 Quai Watier
78401 CHATOU CEDEX
FRANCE
thibaud.mercier@edf.fr

Jean-Melaine Favennec
Senior Research Advisor
EDF R&D (STEP)
6 Quai Watier
78401 CHATOU CEDEX
FRANCE
jean-melaine.favennec@edf.fr

Alexandre Girard
Research Engineer
EDF R&D (STEP)
6 Quai Watier
78401 CHATOU CEDEX
FRANCE
alexandre-girard@edf.fr

ABSTRACT

A PWR reactor coolant system is a highly complex physical process: heterogeneous power, flow and temperature distributions are difficult to be accurately measured, because instrumentations are limited in number, thus leading to the relevant safety and protection margins. This situation is in many ways similar to climate and weather models: a complex process that is not possible to sample and measure as finely as wanted. Meteorology and climate sciences have adapted and improved the Data Assimilation techniques in order to improve the accuracy of description and prediction in their fields.

EDF R&D is seeking to assess the potential benefits of applying Data Assimilation to a PWR's RCS (Reactor Coolant System) measurements: is it possible to improve the estimates for parameters of a reactor's operating set-point, i.e. improving accuracy and reducing uncertainties of measured RCS parameters?

In this paper we study the feasibility of enhanced estimation of PWR primary parameters, by using twin experiments for assessing Data Assimilation benefits. We simulated test samples with a 0D-Model, and used these samples in a Monte-Carlo approach to get background terms for Data Assimilation.

This successful preliminary study will lead to further assessments with real plant data.

1 INTRODUCTION

Estimating PWR primary system parameters is a non trivial question: performances of operational PWR instrumentation are determined by state of art for models and measurement systems at the plant design phase. Of course, periodic evaluations are performed in order to update and validate those performances and related uncertainties. 3D CFD modeling for instance lead to many improvements. However, a strong structural constraint remains: any instrumentation "point" being a potential weakness, a PWR primary coolant must have a minimal number of measurements locations, and this leads to possible representativeness errors. It is the case especially for highly heterogeneous fields like thermal powers, flows and temperatures (in core or in hot leg (see Figure 1) the limited sampling for those fields is taken into account for bias estimations or field reconstructions, and leads to higher margins for a safe operation of the power plant.

Copyright © 2014 by ASME

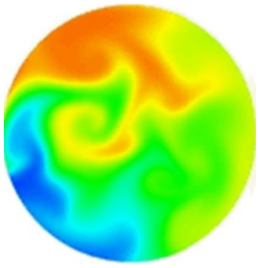


FIGURE 1. Typical hot leg temperature field at nominal power [1]

Last but not least, core loading patterns have evolved during 2000s (LLLPs = Low Leakage Loading Patterns) in order to protect the reactor's vessel lifetime, since the vessel is one of the only components that cannot be replaced: in-core power (and hot leg temperature) heterogeneities have subsequently increased, thus deteriorating uncertainties linked to primary loop-flow (one of the key parameters for safety demonstrations). With Data Assimilation (DA), EDF R&D's main objective is to assess possibilities of improved estimators for these parameters.

In fact, DA answers a very general problem: how to correct a given prediction of a physical state through available observations, giving greater weight to the most reliable data. The more data (models and/or measurements) available for assimilation, the more precise the analysis. Depending on the approach, an assessment of the post-analysis error (and thus of the uncertainty reduction granted by the assimilation) can be computed.

Such tools are obviously valuable to the engineer, and well beyond the scope of geophysics and climate prediction - which has been the most active field for DA developments in the recent decades, which increased computational powers. In this study, we recall the basic principles of Data Assimilation, and the mathematical formulation of BLUE and Kalman Filter [2] [3]; then we use a simple physical model (0D) to represent the primary coolant of a PWR, in order to obtain the necessary parameters for the application of a BLUE-kind approach ; lastly, we evaluate the efficiency of our estimator, by testing it on simulated data.

By using DA on the primary coolant, EDF R&D expects to do more and better with a number-limited (for safety constraints) in-core and out-core instrumentation. Will the combining of measurements and models allow an enhanced estimation of PWR primary parameters? This would be especially convenient in order to diagnose invisible dysfunctions such as instrumentation drifts (i.e. slowly evolving drifts of small amount), representativeness biases... On the pro side, the primary coolant being a "closed" and monophasic system, his thermodynamic equations are constrained: on the con side, it is subject to high heterogeneous fields that disturb measurement processes.

DA will be used on steady operating modes, such as hot shutdowns and 100% power operation, and with averaged (typically on 20 minute stable phases) values, therefore allowing to eliminate statistical uncertainty terms and thermal exchange kinetics.

Otherwise, DA may be seen as extended Data Reconciliation (DR): this method was successfully applied on the secondary side thermal power estimation [4] and allows the operator to follow the drift of feedwater flow rate measurement (and, therefore, of thermal power output measurement) via a model using the many redundant measurements available on the secondary system (water/steam cycle).

In this study, DA's ambition is wider, and may benefit from this "closed" process situation despite of the few available measurement and redundancies.

The outline of the paper is the following. In Section 1, EDF R&D's problematic is explained. In Section 2, Data Assimilation basic principles are reminded, with a focus on BLUE's (or Least Squares Estimator) and Kalman Filter's hypotheses and main results. In Section 3, a simplified model for PWR primary coolant system is built : the model is used with an original Monte-Carlo-like approach in order to get a background vector and the related error covariance matrix, as well as all other parameters required for BLUE implementation. In Section 4, the developed methodology is tested on pseudo-real data (simulated states and observations): it generates a significant reduction of uncertainties, especially for hot leg temperature measurements.

2 DATA ASSIMILATION PRINCIPLES

The material presented in this section is based on [2], [5], and [3].

Copyright © 2014 by ASME

2.1 Mathematical Formulation for the Studied Problem

We have a physical system, supposedly defined by a **true state** vector x_t (t for "true"), and to which we can only access through **observations** collected in a vector y. The **observation operator** H can be defined by:

$$y = H(x_t) + \epsilon_{obs} \tag{1}$$

where ϵ_{obs}, the observational error vector, is supposed to have a null expectation. We call R its covariance matrix.
We also have an *a priori* knowledge, or **background** x_b (b for "background") of the true state vector:

$$x_b = x_t + \epsilon_b \tag{2}$$

where ϵ_b is the background error vector. We call B its covariance matrix.
Our objective is to conciliate those two informations in order to get a new state vector x_a, called **analysis**, that would be the "best" approximation of x_t, which can be interpreted in different ways:

- x_a is the vector minimizing the expectation and the variance of the analysis error vector;
- x_a is the vector with the highest probability knowing y and x_b (Bayesian formulation);
- x_a is the vector minimizing a cost function evaluating the distance to both the observation y and the background x_b (variational formulation).

Under several hypotheses that are those of the BLUE (*Best Linear Unbiased Estimator*), these three formulations are in fact formally equivalent, for they lead to the same formula for x_a, the Kalman Formula [2].

2.2 BLUE (Best Linear Unbiased Estimator)

The BLUE can be obtained by performing the following assumptions:

- **Linearity:** x_a must be a linear combination of the background term x_b and of the observations y, and the observation operator H is supposed to be linear (or linearizable);
- **Unbiased:** error vectors ϵ_b and ϵ_{obs} are supposed to be independent and unbiased, and the related covariance matrices B and R must be definite positive;
- **Best:** $\epsilon_a = x_a - x_t$ must be of null expectation and of minimal variance.

Under those hypotheses, x_a can be written as:

$$x_a = x_b + K^\star \cdot (y_{obs} - H \cdot x_b) \tag{3}$$

$$\text{with } K^\star = BH^T \left(R + HBH^T\right)^{-1} \tag{4}$$

$K^\star$ is called the **Kalman Gain**. The analysis error covariance matrix can be obtained via the following formula:

$$A = (Id - K^\star H)B \tag{5}$$

Here is a simple example: suppose a temperature T is measured with two thermometers, whose measurements T_1 and T_2 are made with error respective standard deviations σ_1 and σ_2. The analysis must be a linear combination of the two measurements: $T_a = kT_1 + (1-k)T_2$, and the Kalman Formula leads to $k = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$.

BLUE and its dynamical version, Kalman Filter (which we do not describe here) are both popular for their easy implementation. However, the hypotheses required in order to prove their optimality are hard to meet, and restrict their scope of application, especially the linearity hypothesis and the knowledge of covariance matrices. For huge systems, the cost of those algorithms can also be very high, and prevents them from being computed in real-time analysis. Many variations of those algorithms have been developed in order to go beyond the scope of those hypotheses or to lower the computation costs, but they are proportionally more sophisticated. [3]

An important further issue is related to the significance of the covariance matrices, and the way to construct them properly. In fact, they can be viewed as partly subjective, as well as the notion of uncertainty associated to a measurement. One can therefore begin with simplified experiments with, for instance, diagonal matrices. But this means throwing away most of the information about the correlation between data, which is most essential and valuable. In this article we give a practical Monte-Carlo-like method that uses a physical model in order to construct both background vector and background error covariance matrix, while quantifying underlying hypotheses.

In order to go beyond the (relatively narrow) scope of application of BLUE, an explicit computation of the gain matrix K can be avoided by choosing a variational approach, called **3DVAR**, for the analysis. The approach

Copyright © 2014 by ASME

mainly consists in minimizing the cost function:

$$J(x) = (x - x_b)^T B^{-1} (x - x_b) + (y - H[x])^T R^{-1} (y - H[x]) \quad (6)$$

Under a few assumptions (linear H, Gaussian uncorrelated errors...) the result is exactly the same as BLUE's one. This variational formulation shows how the BLUE can be related to the Least-Squares Estimator.

There are several implementation versions of these variational approaches, most of them aiming to reduce implementation costs. The main advantage of variational schemes lies in the fact that the linearity hypothesis for observation and evolution operator is no more necessary for solving the problem. Also, the computation and storage of heavy matrices like Kalman Gain can be avoided. However, the cost function gradient calculation requires the adjoint model, which may be complicated to obtain.

Ensemble methods is another class of methods that avoids the linearity hypothesis. Well known by geophysicians, they represent error spreading with random perturbations propagated through times according to observations, instead of mere covariance matrices. The Ensemble Kalman Filter, for instance, has proven to be very efficient on real problems, despite of its simplicity. Data Assimilation is a broad field of Applied Mathematics, and an active research topic: several sophisticated techniques are already available or still being developed, each of them with different features for its own objectives. However, the BLUE, the Kalman Filter and the basic variational schemes are enough in order to quickly get first results on our parameter estimates.

3 SIMPLIFIED MODEL FOR A PWR PRIMARY SYSTEM

If we want to apply these Data Assimilation techniques to our industrial problem - that is, primary coolant operation point estimation -, we first have to clearly define the mathematical objects we are going to work with.

3.1 State vector definition

In theory, the physical state of the primary coolant should be described by the temperatures, flows, pressures and neutron powers fields (3D) value in the whole system at each time t. A state vector x would then collect every valuation (or finite elements valuations) of each of those fields on each point of a given mesh. However, considering

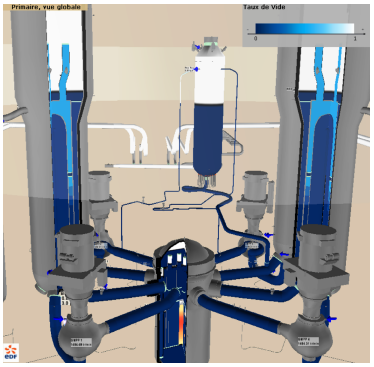


FIGURE 2. 3D view of a PWR primary coolant. The core vessel in the center is about four meter high by four meter wide. Pressurized water circulates at about 17 m/s, at 155 bar pressure and $300°C$ temperature.

how turbulent the flow inside the core and loops is, how few measurements are available, and how uncertain those measurements are, it is impossible to perfectly rebuild those fields except, to some extent, with a complex 3D-CFD model.

For a simplified Model, we can restrict our knowledge of primary coolant functioning point to a few scalars [6] :

- (mass) loop-flows (Q_i), in t/s;
- average cold leg temperatures (TBF_i), in °C;
- average hot leg temperatures (TBC_i), in °C;
- fuel assembly upper temperatures ($THAC_j$), in °C;
- integrated neutron power for each fuel assembly ($PIPA_j$), in MW;
- thermal power transferred to the secondary coolants (through the steam generators)($PTHGV_i$), in MW.

Figure 3 summarizes these choices. The advantage of this description is that it fits with available observations (elbow-taps flowmeters, RCS temperature probes, in-core thermocouples, Mobile Fission Chambers, unit control data...), and therefore with the operator view of the primary coolant.

Here we only study reactors at 100% of their nominal power (typically between 99% and 100% of authorized operation power, with a $\pm 0,4\%$ stability). This nominal power is adjusted as frequently as possible, which means several times every day, and at monthly power stability phases (36 h long) used for evaluating the BIL100 thermal power and the previously mentioned core flux mappings (in-core powers).

Copyright © 2014 by ASME

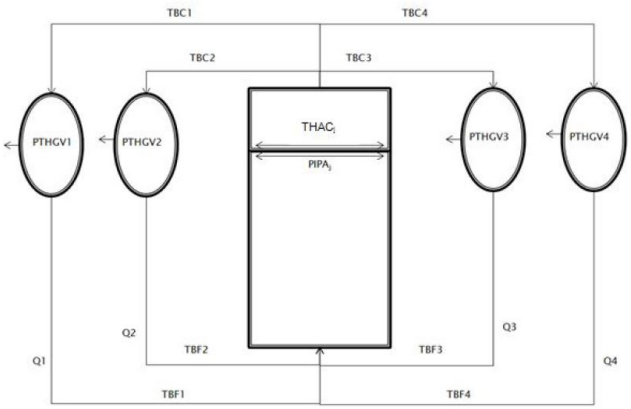


FIGURE 3. Primary coolant model of a PWR unit

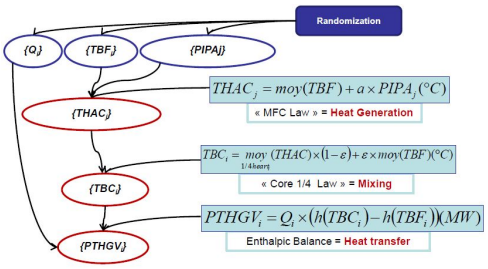


FIGURE 4. Data randomisation process

	Q	TBF	$PTHREAC$	$PIPA_{norm}$
Value	$5.0\ t/s$	$289.5°C$	$3800\ MW$	$1/193$
Variation (2σ)	$\pm 2\%$	$\pm 0.4°C$	$\pm 1\%$	$\pm 10\%$

TABLE 1. Model parameters randomisation

Thus, our DA approach will be only applied, in the end, on real data (averaged and calibrated) from those ultra-stable and repeatable operating modes. We will not study dynamic operating modes: steady-stable periodic phases are sufficient to provide reliable observations for monitoring and diagnosis purposes.

3.2 Sampling simulation

Be it for a BLUE or a Kalman Filter, we need a background vector x_b and the associated error covariance matrix B. They represent more or less the power plant design operation value, and the statistical variation of real values around these design values. Therefore, the idea is to simulate a sample of possible state vectors, then to study their statistical spread. For this purpose:

1. we simulate a set of Q_i and $TBFi$, which stands respectively for primary pumps fluctuation and thermal regulation;
2. we simulate a set of $PIPA_j$, which stands for neutron power fluctuation;
3. we use a simplified enthalpic balance (proportionality between core upper-lower temperature gradient and integrated neutron power, for each fuel rod) to deduce all $THAC_j$;
4. we use a core-quarters averaging in order to deduce all hot leg temperatures TBC_i [7];
5. we use an enthalpic primary/secondary balance around Steam Generators in order to deduce the thermal powers transmitted to secondary water/steam system, $PTHGV_i$.

Thus (See Figure 4 and Table 1) a sample of N simulated vectors x_s can be obtained. We then calculate the background vector and associated error covariance matrix:

$$x_b = \frac{1}{N}\sum_{i=1}^{N} x_{s,i} \tag{7}$$

$$B = \frac{1}{N-1}\sum_{i=1}^{N}(x_{s,i} - x_b)(x_{s,i} - x_b)^T \tag{8}$$

At this point, our physical model is reduced to these two terms, which summarize in a very basic way its statistical behaviour. Their meaning is as follows: if we assume, without any measurement, that the system vector state is x_b, we make a mistake we can suppose to be Gaussian, centered and of variance B. This is of course an approximation, and the true statistical behaviour of our simulated vector is not Gaussian, because some thermodynamical equations are non linear.

With this methodology, it is very easy to quantify and modify all underlying hypotheses (uncertainties, biases, etc.).

3.3 Simulation of observations

In our case, the available observations are the following:

- Elbow-taps flowmeters (4×4 measurements) [8] [9]
- RCS temperature probes (4×6 measurements for cold leg, 4×2 measurements for hot leg)

Copyright © 2014 by ASME

- In-Core Thermocouples (50 measurements)
- Mobile Fission Chambers flux mapping, from which we can extract integrated powers for each fuel rod, which can stand for $THAC$ measurements via our model (193 measurements)
- "BIL100" periodic estimation - or "BILORLI" continuous estimation - for $PTHGV$ (4 measurements)

The observation operator H must allow us to deduce from a state vector x the observation vector y we should obtain: as our space state was designed to be close to available observations, the observation operator H can be easily built: the i-th line contains a 1 at j-th position if the ith observation is a measurement of the j-th physical variable, and zeros elsewhere.

The observation error associated to this operator should be (for BLUE application) close to a centered Gaussian random variable: this is a strong hypothesis, and in order to satisfy it, a careful and systematic calibration of all probes (especially, in our case, elbow-taps, in-core TCs and RCS leg probes) should be performed. For temperature measurement systems, this calibration is done at hot shutdown phases, while thermal power is null and temperature fields are homogeneous. This leads to an elimination of most of the possible biases, which is the most decisive asset for PWR primary real data.

We still have to quantify R, the observation error covariance matrix. It is usually chosen to be diagonal, with diagonal terms equal to estimated measurement error variances. For each observation, if the 95% confidence interval is $\pm\eta$, the chosen variance will be taken equal to $\sigma^2 = \frac{\eta^2}{4}$: the measurement error associated to this observation will then be modelized by a Gaussian white noise of standard deviation σ.

We now have all required terms (x_b, B, H and R) in order to compute the BLUE formula: we now need to apply this formula on y observations. Should we directly use data from real measurements, we would not be able to conclude on our algorithm's efficiency: indeed, with real data, we cannot access to the true state vector x_t and therefore cannot quantify how precise our analysis is. Thus, we need to work with simulated observations.

For a simulated state vector x_s, an associated observation vector y_s can thus be simulated by:

$$y_s = H \cdot x_s + N(0, R) \tag{9}$$

where $N(0, R)$ is the drawing of a random variable following a normal distribution of null expectation and R variance.

4 TWIN EXPERIMENTS (WITH PSEUDO-REAL DATA)

We speak of a **Twin Experiment** when we apply a Data Assimilation technique on simulated data with properties close to real data. With simulated data, analysis error value and statistics can be exactly calculated: therefore quantifying our algorithm performance is possible. The fundamental hypothesis of Twin Experimentation is that our algorithm keeps the same behaviour with real data.

For each simulated vector x_s, we have three estimations:

- **background** x_b, steady and nearly equal to the reactor's conception values;
- **projection** x_{proj}, whose components are the averaging of each available measurement (only possible if every component is measured by at least one sensor);
- **analysis** x_a, result of BLUE application.

Schematically: background is obtained with the model and without measurement; projection with measurements and without model; analysis with both model and measurements. In theory, those different estimations should be compared with their associated error covariance matrix trace: however, from an operator's (such as EDF) point of view, evaluating uncertainty reduction gain component by component seems more direct.

4.1 Studied variants

Several variants of previously described DA framework are possible:

1. **standard :** state vector consists in Q_i, TBF_i, TBC_i, $THAC_j$ and $PTHGV_i$, and all observations are used($PIPA_j$ being transformed into a $THAC_j$ measurement);
2. **without in-core data:** state vector consists in Q_i, TBF_i, TBC_i et $PTHGV_i$, and we exclude PIPA and (in-core thermocouples) of observations;
3. **with bypass :** state vector consists in Q_i, TBF_i, TBC_i, $THAC_j$, $PTHGV_i$, and of ϵ, the bypass coefficient used in the core-quarter averaging, and all observations are used as for standard variant.

To these different variants for state space and observation state correspond different variants for the observation operator and for the model used for the simulation of 1000 vectors. Programmation was done in Python (2.7)

Copyright © 2014 by ASME

4.2 Influence of using in-core parameters : a remarkable improvement

In this experiment, we compare DA efficiency with in-core parameters ($PIPA$, $THAC$ and related measurements) and without. First variant must be more powerful than second variant, as we incorporated two kinds of information included in in-core parameters: first, $THAC_j$ and TBC_i are statistically linked by the physical model, which appears with non-diagonal terms in background error covariance matrix (when the i-th core-quarter temperature is higher, the hot leg temperature TBC_i is also higher); secondly, all in-core measurements (PIPAs and in-core TCs). The question is to quantitatively evaluate the uncertainty reduction brought by these informations.

Thus we simulate a sample of vector states for variant 1, randomizing our physical model as described previously, and generate corresponding observations. Then we reduce those state and observations vectors, in order to exclude in-core terms. This way, we calculate the background terms for each of both variants, and then test the BLUE approach for both formulations.

	Q	TBF	TBC	$THAC$	$PTHGV$
Units	t/s	$°C$	$°C$	$°C$	MW
Typical values	5.0	289.5C	324.1	325.6	981
Errors (2σ) :					
Background	0.10	0.40	0.62	3.63	28.1
Projection	0.06	0.24	0.77	2.06	20.1
Analysis 1	0.05	0.20	0.20	1.64	11.0
Analysis 2	0.05	0.20	0.38	N.C.	13.5

TABLE 2. Errors (95% confidence interval) comparison for in-core parameter influence study. Analysis 1 is performed with in-core terms, analysis 2 without.

These results (Table 2 and Figure 5) have been obtained by averaging errors on the 4 loops or the 193 fuel rods, then by studying their statistics. In this first experiment's framework, all biases are set to be null. The reduction of uncertainty is most spectacular for thermal power transmitted to secondary coolant $PTHGV$ and hot leg temperature TBC :

- the uncertainty on thermal power is nearly divided by two (from ±28.1 MW for the background, and ±20.1 MW for projection, to ±11.0 MW for analysis 1 ("Standard"), and ±13.5 MW for analysis 2 (without in-core terms));
- the uncertainty on hot leg temperature is divided by three for analysis 1, and nearly by two for analysis 2.

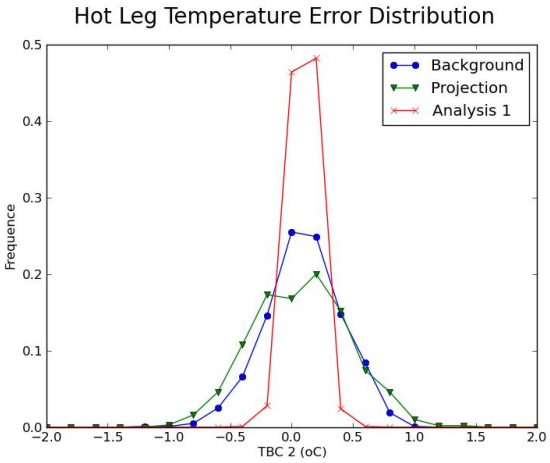


FIGURE 5. Probability distributions for background, projection and analysis 1 errors for TBC_2

The difference between both analysis is clear, as expected, for hot leg temperature : taking into account the in-core parameters divides by two the hot leg temperature uncertainty, even though it was already considerably reduced. Others parameter estimates are not improved between analysis 2 and analysis 1. For cold leg temperature TBF and loop-flow Q, the background error is high compared to observation (projection) error: thus each analysis forsakes the model information and only makes use of measurements. Post-analysis uncertainty is therefore nearly identical to the "projection" (observations only) uncertainty.

In the end, this experiment reveals that the mere combination of available observations with a reliable background leads to a spectacular reduction of uncertainties, especially for some key parameters like thermal powers transmitted to secondary coolant $PTHGV_i$ and hot leg temperatures TBC_i. [10]

In this part of our study, we also made clear that in-core parameters improved the reduction of uncertainties: however, we still need to justify more accurately the diagonal aspect of R, i.e. the fact that all different measurement systems have uncorrelated measurement errors.

4.3 Resistance to biases : promising results

BLUE hypotheses are strong and not likely to be verified for real data, especially the non-bias hypothesis for background and observation errors. Nominal operating

Copyright © 2014 by ASME

mode of a PWR reactor is stable enough to precisely calibrate our physical model, and to assume that background error bias should be low enough to be neglected. However, a lot of measurement systems are still, even after a hot shutdown calibration, subject to nearly steady biases that are hard to quantify. Therefore, we want to test our analysis resistance to known biases.

We keep the same DA parameters as in variant 1 (background terms, observation operator, observation error covariance matrix), and simulate a test sample whose observations are partly biased. We then run our analysis and evaluate both biases and error variances reduction.

	Q	TBF	TBC	$THAC$	$PTHGV$
Units	t/s	$°C$	$°C$	$°C$	MW
Typical values	5.0	290.0	324.6	326.1	985
Errors (Biases) :					
Background	0.00	0.00	0.00	0.0	0.0
Projection	0.02	0.00	0.88	0.15	0.3
Analysis 1	0.01	0.01	0.09	0.14	2.0
Errors (2σ) :					
Background	0.10	0.40	0.62	3.63	28.1
Projection	0.06	0.25	0.77	2.05	20.0
Analysis 1	0.05	0.20	0.21	1.64	11.24

TABLE 3. Errors (biases and 95% confidence intervals) comparison for resistance to biases study.

The biases above (Table 3) are calculated by averaging their absolute values on all 4 loops (or 193 fuel rods for the THAC). In order to simulate observations, we took biases corresponding to real data : some for loop-flows measurements (approximately $\pm 0.05 t/s$, which means $\pm 1\%$ of nominal loop-flow), and more consequent biases for some of the hot leg temperatures probes (approximately $1°C$) as well as for in-core TCs. We can see (Figure 6) that, the background being properly centered, the analysis is nearly unbiased compared to projection, and variances are not very different from the application to unbiased data.

These results are promising in anticipation of an application to real data.

4.4 Incorporation of the bypass parameter identification

Some of the parameters of our physical models are not well-known and can't be directly measured: for instance, the so-called "bypass" coefficient we used for our core-quarter averaging. There is, indeed, a portion of

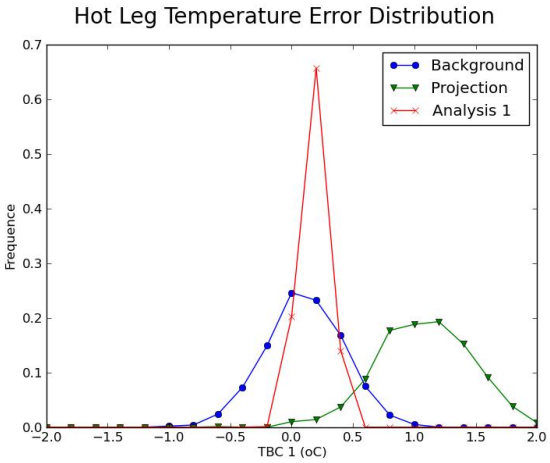


FIGURE 6. Probability distributions for background, projection and analysis 1 errors for TBC_1, with biased observations

the primary coolant that flows around the core, and goes directly from cold leg to hot leg without being heated by fuel rods: because of this bypass, the hot leg temperature is colder than the mere averaging of the core-quarter upper temperatures, thus this correction term in our formula.

Our third DA framework variant incorporates this parameter into our state space for the analysis. We modify our physical model by also randomizing this parameter ($\epsilon = N(0.055, 0.01)$), in order to calculate corresponding new background vector and background error covariance matrix.

We then test our DA approach with these new background terms, but on a test-sample that was simulated with a constant and weaker than expected bypass coefficient (0.050 instead of 0.055). Thus, it is the background error that is biased here. We want to check if available observations lead, via DA, to an accurate correction of bypass parameter estimation.

The background was calculated with a bypass stronger than our test sample: therefore, hot leg temperature (and thermal power transmitted to secondary coolant) estimation was too low.

With this analysis (see Table 4):

- all biases are strongly reduced;
- key parameter TBC variance is only slightly deteriorated in comparison to analysis 1 of first experiment;
- however, uncertainty on bypass parameter increased

Copyright © 2014 by ASME

	Q	TBC	$PTHGV$	ϵ
Units	t/s	$^{\circ}C$	MW	
Conception values	5.0	324.6	985	5.5%
Conception Errors (2σ)	0.10	1.50	59.5	0.5%
Errors (Biases) :				
Background	0.00	-0.19	-5.7	0.51%
Projection	0.00	0.01	0.05	N.C.
Analysis 3	0.00	-0.02	-0.3	0.07%
Errors (2σ):				
Background	0.10	0.62	28.6	0.00%
Projection	0.06	0.79	20.0	N.C.
Analysis 3	0.05	0.29	12.4	0.68%

TABLE 4. Error comparison (biases and 95% confidence intervals) for bypass parameter identification experiment (results for TBF and $THAC$ were omitted for more concision).

after the analysis: in short, our physical model for the background had $\epsilon \in [5.0\%, 6.0\%]$ as 95% confidence interval for this bypass parameter, and our analysis has $\epsilon \in [4.39\%, 5.75\%]$ instead; and we know the real value for this experiment is 5.0%.

Without any direct measurement, using only the systems physical equations, via statistical correlations recorded into background error covariance matrix, we were able to get a corrected estimate for bypass coefficient value.

The lessons from these three "twin experiments" are as follows: merely introducing a background (or *a priori* knowledge) for the primary coolant state, with controlled quantitative hypothesis, leads to a great reduction of measurement uncertainties. If some state variables are well correlated (like for instance TBC_i with $THAC_j$), then uncertainty reduction improves even more.

5 CONCLUSIONS AND PERSPECTIVES

Using a simplified model for a PWR primary coolant system, we were able to formulate our primary system operation point estimation problem in terms of Data Assimilation, and to get, thanks to a simulated sample of possible state vectors for primary coolant, a background vector and an associated error covariance matrix. We tested the methodology defined this way on simulated data, and observed that, despite the relative simplicity of our model and Data Assimilation approach, the analysis reduces greatly the uncertainty associated to some key parameters, like hot leg temperatures or mass loop-flows.

Future works are planned for an application of the same kind of methodology to the in-core part of the primary system, in order to get an improved 2D-mapping of upper fuel rod temperatures. Future work will also include the use real data, and compare this new estimator to already existing ones, and do a cross-validation by verifying their confidence intervals intersection.

These first results on simulated data for PWR primary coolant parameters already give another illustration of the fact that Data Assimilation scope of application is not restricted to geophysical sciences only, and can be of interest for industrial process engineers as well.

NOMENCLATURE

BIL100 Evaluation of Steam Generators Powers using periodic unit control data to compute a thermodynamical balance on the secondary system around Steam Generators

BILORLI Uses the same equations as BIL100 but with data from ORLI (continuous and automatic data recording system) only

BLUE Best Linear Unbiased Estimator

DA Data Assimilation

ϵ Bypass coefficient for core-quarter averaging

$N(x, M)$ drawing of a random variable following a normal distribution of x expectation and M variance.

$PIPA$ Integrated neutron power for a fuel assembly

$PTHGV$ Steam generator thermal power

$PTHREAC$ Reactor full thermal power

PWR Pressurized Water Reactor

Q Mass loop-flows

TBC Hot leg temperature

TBF Cold leg temperature

$THAC$ Upper fuel rod temperature

TC RIC In-core thermocouples

REFERENCES

[1] Bellet, Benhamadouche, and Ferret, 2011. "Secondary flows in PWR hot legs, CFD might bring some light on the flow behavior". *The 14th International Topical Meeting on Nuclear Reactor Thermalhyadraulics NURETH-14, Toronto, Canada.*

[2] Talagrand, 1997. "Data assimilation, an introduction". *Journal of the Meteorological Society of Japan.*

[3] Bocquet, 2012. "Introduction to data assimilation methods and principles in geophysics". *ENSTA and M2 OACOS lecture notes.*

[4] Favennec, Piedfer, and Veau, 2011. "Reactor power measurement: EDF's strategy and ongoing work on measurement issues". *3rd International Workshop on*

Copyright © 2014 by ASME

Metering of Feedwater Flow, Temperature and Thermal Power, Chatou, France.

[5] Bouttier, and Courtier, 2002. "Data assimilation concepts and methods". *Meteorological Training Course Lecture Series ECMWF.*

[6] Duminy, 2010. Innovative estimators for pwr primary system process parameters: integration in a monitoring tool. Tech. rep., EDF R&D.

[7] Vaillant, 2012. Feasibility study of a new hot leg temperature estimator based on in-core temperature and flux sensors in a PWR. Tech. rep., EDF R&D.

[8] Deneux, and Arenas, 2013. "CFD and metrology in flowmetering: RCS flow measurement with elbow taps and its uncertainty".

[9] Deneux, Favennec, Piedfer, Antunes, and Veau, 2013. "Using CFD for metrology prediction in flowmetering : RCS flow measurement in PWRs with elbow taps". *ASME Power Conference, Boston, USA.*

[10] Mercier. Assessment of a data assimilation approach for enhanced estimation of PWR primary parameters. Poster for Summer School on Data Assimilation & Inverse Problems at University of Reading, United Kingdom, July 22-26, 2013.

Copyright © 2014 by ASME

Proceedings of ICONE-23
23th International Conference on Nuclear Engineering
May 17-21, 2015, Chiba, Japan

ICONE23-1442

DATA ASSIMILATION AND PWR PRIMARY PARAMETERS ESTIMATION: ENHANCEMENT OF A 0D MODEL AND CALIBRATION OF RANDOM PARAMETER GENERATOR FOR BACKGROUND TERMS EVALUATION

Thibaud Mercier
PhD Student
EDF R&D (STEP)
6 Quai Watier
78401 CHATOU CEDEX
FRANCE
thibaud.mercier@edf.fr

Jean-Mélaine Favennec
Senior Research Advisor
EDF R&D (STEP)
6 Quai Watier
78401 CHATOU CEDEX
FRANCE
jean-melaine.favennec@edf.fr

Alexandre Girard
Research Engineer
EDF R&D (STEP)
6 Quai Watier
78401 CHATOU CEDEX
FRANCE
alexandre.girard@edf.fr

Keywords: Data Assimilation; PWR; Primary Coolant; Monte-Carlo Method; 0D Model; Measurement Systems

ABSTRACT

EDF R&D is seeking to access the potential benefits of applying Data Assimilation to a PWR's RCS (Reactor Coolant System) measurements, in order to improve the estimators for parameters of a reactor's operating setpoint, i.e. improving accuracy and reducing uncertainties and biases of measured RCS parameters.

In this paper, we use balance (between primary and secondary systems) to improve the simplified semi-empirical 0D Model for RCS, using a "fitting" method for the bypass coefficient related to the part of the flow which is not in contact with the fuel assemblies in each quarter of a four-looped core. Thus, we get a model that can be used to generate state vectors containing most of primary parameters values. Then we describe how to use this model to define a Data Assimilation Approach, by generating random parameters and thus constructing a sample of random state vectors, from which the background vector and related error covariance matrix can be deduced.

Finally, we apply our method with a focus on Normalized Integrated Neutron Powers, using twin experiments to evaluate its performances.

Overall, calibrating the random parameter generator for Neutron Mapping on real data does improve the algorithm performance, though only moderately.

1. INTRODUCTION

Performances of PWR Primary Parameters estimators are determined by the state of art for model and measurement systems at the plant design phase. Periodic evaluations are performed in order to update and validate those performances and related uncertainties, using 3D CFD modelling for instance (Bellet et al, 2011). But strong structural constraints remain: a PWR Reactor Coolant System imposes a limited number of measurement locations, any additional instrumentation "point" being a potential weakness, and thus leading to possible representativeness errors. For highly heterogeneous fields like thermal powers, flows and temperatures, in core or in hot legs, the limited measurement sampling is taken into account for bias estimations or field reconstructions: this leads to higher margins for a safe operation of the power plant (Favennec et al, 2011). Thus, estimating PWR Primary System Parameters with optimal precision is a valuable and non trivial problem.

Moreover, core loading patterns have evolved during 2000s with Low Leakage Loading Patterns, in order to improve the reactor's vessel lifetime (the vessel being regarded as the main primary system component that cannot be replaced). Thus in-core power (and hot leg temperature) heterogeneities have increased, deteriorating uncertainties linked to primary loop flow rates and temperatures (key parameters for safety demonstrations).

EDF R&D seeks to assess possibilities of improved estimators for all primary parameters: Data Assimilation (DA) is one solution currently explored. By combining a physical model (even simplified) with all available observations, a more precise estimation can be computed, furthermore with an assessment of the post-analysis error.

DA techniques have been mostly developed by and for geoscientists: but since their principles are actually quite far reaching, engineers of several different fields have been recently applying them to their own estimation problems, as diverse as neutron mapping, cardiac monitoring, financial markets evolutions, etc.

In a previous work (Mercier et al, 2014), we introduced a Monte-Carlo-like method for an application of a BLUE estimator using a given black-box model, and applied it to a 0D simplified model for RCS, evaluating our algorithm

Copyright © 2014-2015 by JSME

performances with twin experiments. In the present study, we introduce an enhancement for our 0D model; then we recall basics results of DA theory for BLUE estimator, and further explain how to calibrate and use a Random Parameter Generator in order to get all required background terms. Then we apply our calibration method to the randomization of Normalized Integrated Powers for each fuel rod ($PIPA_{norm}$) and evaluate the effect of this calibration on our DA algorithm.

By cautiously calibrating our random parameter generator with real data and/or submodels, we will assess whether our algorithm performance is significantly improved, or not.

2. ENHANCED 0D MODEL

Several models are available for RCS, with different complexity levels. 3D CFD models (such as EDF's THYC or CODE_SATURNE) offer an accurate description of mass flows and temperature fields inside the primary coolant system. However, their computational costs are high, and for our purpose (on-line monitoring of actual power plants, at 100% of their nominal power) a 0D/1D model may prove sufficient.

We therefore restrict our knowledge of primary coolant operation setpoint to a few scalars (as seen in Fig. 1):

- (mass) loop-flows Q_i;
- average cold leg temperatures TBF_i;
- average hot leg temperatures TBC_i;
- fuel assembly upper temperatures $THAC_j$;
- integrated neutron power for each fuel assembly $PIPA_j$;
- thermal power transferred to the secondary coolant system (through the steam generators) $PTHGV_i$.

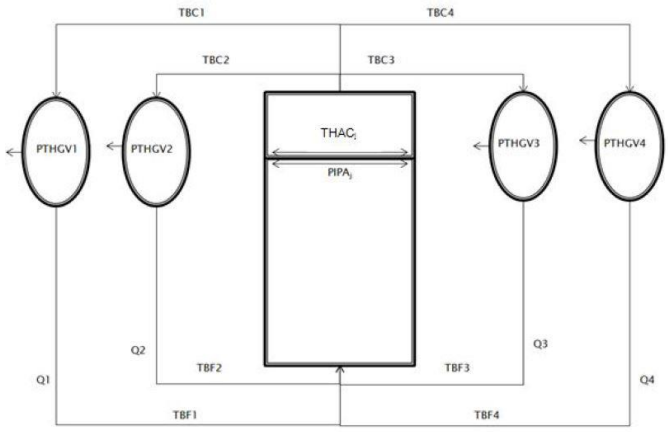


Fig. 1: Primary coolant model of a PWR unit

This description fits with available observations (elbow-tap flowmeters (Deneux et al, 2013), RCS temperature probes, in-core thermocouples, Mobile Fission Chambers, unit control data, etc.), and with actual operator's perspective.

2.1. Semi-Empirical Equations

Our model describes the primary coolant behavior with three semi-empirical equations: one for in-core heat generation, one for top plenum mixing, and one for heat transmission to secondary coolant system.

a) In-core heat generation

The first equation, used to model in-core heat generation, is called "PIPA law" and states a proportionality between core upper-lower temperature gradient and integrated neutron power, for each fuel rod, so that:

$$THAC_j = MOY(TBF) + a \times PIPA_j \qquad \text{(Eq. 1)}$$

Proportionality coefficient a is estimated by confronting neutron powers mapping (obtained with Mobile Fission Chamber measurements and/or interpolation code) at each fuel rod equipped with a thermocouple. This empirical law can be justified when you consider it as an approximated integration of several enthalpic balance equations on the whole fuel rod axis.

b) Top Plenum Mixing

The second equation, called "Core Quarter Averaging", and used to model top plenum mixing, supposes that each hot leg temperature TBC_i can be estimated from the averaging of the 2D temperature field in the related Core Quarter, so that:

$$THC_i = MOY_{j \in C_i}(THAC_j) \qquad \text{(Eq. 2)}$$

However, a fraction of the coolant – called bypass – does not flow through the core but around it, so this approximation is too high and must be corrected:

$$TBC_i = (1-\varepsilon)THC_i + \varepsilon MOY(TBF) \qquad \text{(Eq. 3)}$$

ε is called "bypass coefficient" is supposed to be close to the percentage of mass flow going around the core, which is approximately 5,5%. However, an actual link between those two scalars is not obvious, and the one we use in our formula should be considered as a mere empirical adjustment variable, to be estimated similarly as our 'a' coefficient in "PIPA law".

c) Heat Transmission to Secondary Coolant System

The third equation is an enthalpic balance for the primary coolant around steam generators:

$$PTHGV_i = Q_i.\left(h(TBC_i) - h(TBF_i)\right) \qquad \text{(Eq. 4)}$$

A linear expression for this equation could be obtained as follow:

$$PTHGV_i = Q_i.C_p(T_{MOY}).(TBC_i - TBF_i) \qquad \text{(Eq. 5)}$$

However, this linear equation is not exact, since 'h' mass enthalpy function shape between *TBF* and *TBC* is far from a linear function. We chose to model h as a 3 degree polynomial function (the remaining error between model and actual h value is below 0.1%).

2.2. Energy Balance and Bypass Coefficient Fitting

To complete our model, we write heat flux balance for primary coolant: heat received by primary coolant, via neutron powers and pumps, should be equal to heat transmitted to secondary coolant:

$$\sum_i PTHGV_i = PTHREAC + PGMPP \qquad \text{(Eq. 6)}$$

If we combine Core-Quarter averaging formula with "PIPA law", we get this expression for TBC_i:

$$TBC_i = TBF_i + (1-\varepsilon)a\frac{PTHREAC}{193}k_i \qquad \text{(Eq. 7)}$$

where k_i is the averaging, on each core-quarter, of normalized integrated neutron powers:

$$k_i = MOY_{j \in C_i}(PIPA_{norm,j}) \qquad \text{(Eq. 8)}$$

$PIPA_{norm,j}$ averaging on the whole core is exactly 1, so each k_i should be close to 1. These coefficients depend only from

Copyright © 2014-2015 by JSME

the shape of neutron powers mapping, and from the choice of the core quarter definition. We chose core quarters of 56 fuel rods, a choice that may seem rather arbitrary, but was successfully tested on real data.

Thus, we get the following expression for the power balance equation:

$$PTHREAC + PGMPP = \sum_i Q_i \Delta h_i \quad \text{(Eq. 8)}$$

where

$$\Delta h = h\left(TBF_{MOY} + (1-\varepsilon)a\frac{PTHREAC}{193}\right) - h(TBF_i) \quad \text{(Eq. 9)}$$

In this equation, the only term that cannot be measured or estimated is the bypass coefficient ε. Function h being monotonic, there is only one value for this coefficient that allows this equation to be verified. We note this value ε_{fit}.

With several simplifications (uniform loop-flows, cold leg temperatures, and neutron powers, and linearization for enthalpic difference: $\Delta h_i = Cp\Delta T_i$), we get the following expression for ε_{fit}:

$$\varepsilon_{fit} = 1 - \left(1 + \frac{PGMPP}{PTHREAC}\right) \Big/ \left(\frac{4}{193} Q.C_p.a\right) \quad \text{(Eq. 10)}$$

$$\varepsilon_{fit} = 5.5\% \quad \text{(Eq. 11)}$$

Therefore, with these simplifications, we see that the design value of this bypass coefficient is coincidentally equal to the design value of the bypass flow in the core. However, this is only an approximation. If cold leg temperatures and loop-flows are quasi-steady, however thermal powers and hot leg temperatures fluctuate slightly: thus so does ε_{fit}. Function h being monotonic, we can compute a simple, dichotomic research algorithm that uses (Eq. 9) to calculate a numerical evaluation for ε_{fit}.

If we study our numerical approximation behavior, we can obtain an affine approximation for bypass "fit" coefficient:

$$\varepsilon_{fit} \approx 6.39\% + 0.23\% MOY\left(400(k_i - 1)\right) + 0.11\%\left(100\left(\frac{PTHREAC}{3800} - 1\right)\right) \quad \text{(Eq. 12)}$$

Thus, for a given neutron power mapping (which implies fixed k_i), being 10% below nominal power will cause the bypass "fit" coefficient to decrease by 1.1% (e.g., from 6.4% to 5.3%).

Finally, instead of evaluating a fixed bypass coefficient for our model, we automatically adjust its value in order to satisfy power balance between first and secondary coolant.

2.3. State Vector Construction

We will use our 0D Model as a "Black Box" : $x = \Phi(u)$, where u is an input parameter vector, containing Q_i, TBF_i, $PTHREAC$ and $PIPA_{norm}$, and x is an output state vector, containing Q_i, TBF_i, TBC_i, $THAC_j$, $PTHGV_i$ and bypass coefficient ε.

Function Φ works as follow:

1. we use "PIPA law" to compute all upper fuel rod temperatures $THAC_j$;
2. using a standard value for ε, we use Core Quarter Averaging to obtain TBC_i;
3. we use enthalpy balance around Steam Generators to evaluate thermal powers transmitted to secondary coolant $PTHGV_i$;
4. if power balance equation is sufficiently satisfied (with a precision of ±1 MW for instance) algorithm ends; else we go back to step 2, using a new adjusted value for ε.

With this process, we used all of our semi-empirical equations to construct a "black-box" model that produces physically coherent state vectors for the whole primary coolant system.

3. DATA ASSIMILATION APPROACH

Data Assimilation (DA) is a generic term that refers to several algorithms that might already be known under different names by other fields of mathematics, such as Inverse Problems or Optimization. The main objective is to combine all available information (observations and/or models) of a system considering how accurate it is, in order to get an enhanced estimation of a given physical object within the system. Weather forecasting is the best known application: an evolution model output, obtained with past data, is combined with present measurements, in order to produce an optimized analysis that will be used to predict future weather. However, as said before, DA theorems and statements are actually quite general and can be (and are currently being more and more) used by several kinds of scientists and engineers.

In this section, we present a classic DA tool: the BLUE (Best Linear Unbiased Estimator), and introduce a method for an application to a black-box model, such as our previously described PWR Primary Coolant 0D model.

3.1 BLUE

For general theory on Data Assimilation, see (Bocquet, 2012), (Bouttier and Courtier, 2002) or (Talagrand, 1997)

We can *ex abrupto* summarize the BLUE theorem (sometimes called the Gauss-Markov Theorem) with the following equation:

$$x_a = x_b + K^* \cdot (y_{obs} - H \cdot x_b) \quad \text{with } K^* = BH^T(R + HBH^T)^{-1} \quad \text{(Eq. 13)}$$

where:

- x_b is the background vector (*a priori* estimation of the physical state, without observations);
- y_{obs} is the observation vector (only obtained with measurements, without *a priori*);
- H is the linear observation operator (such as, if x_t is the true state, $y_{obs} \approx H \cdot x_t$);
- B and R are respectively the background and observation error covariance matrices.

BLUE's scope of application is quite narrow: H is supposed to be linear, and x_a is supposed to be a linear combination of x_b and y_{obs} ('L' hypothesis); observation and background errors are all supposed to be both unbiased (so is analysis error) and uncorrelated ('U' hypothesis), and x_a is required to be the best possible estimation ('B' hypothesis), which means minimal variance. However, in most simple cases, these hypotheses can be physically natural and reasonable.

Here is a simple scalar illustration: a temperature T is a priori supposed to be equal to T_b, and we have its measurement equal to T_{obs}, with respective error standard deviations of σ_b and σ_{obs}. Analyzed temperature will be a

Copyright © 2014-2015 by JSME

linear combination of those two data:

$$T_a = kT_b + (1-k)T_{obs}$$
$$k = {\sigma_{obs}}^2/({\sigma_b}^2 + {\sigma_{obs}}^2) \quad \text{(Eq. 14)}$$

Here, all terms of BLUE formula are instantly accessible. However, when our physical state is not a scalar but a vector, potentially large and containing various kinds of physical values, we need much more work. If vector spaces are conveniently built, H, x_b and y_{obs} are most of the time easy to "see" and to build. However, covariance matrices, R and B, are trickier.

R is usually taken as diagonal, since we don't have much information about possible correlation between pure observational errors: diagonal terms are, therefore, variances of scalar error for each measurement taken separately. B, on the opposite, is rarely diagonal. And, looking at BLUE formula, all information about our model behavior must be contained in its shape. Therefore it must be very carefully constructed.

3.2 Background Terms Construction

a) Meaning of Background Terms

Background vector x_b and related covariance matrix B are the two only terms of BLUE formula associated with "model" information: in theory, once we have built them, we will not use our model again. For this reason, we need to better understand what these terms mean, how to obtain them from our model, and how to control underlying hypotheses.

By definition, if x_t is the true vector state:

$$x_b = x_t + \varepsilon_b \quad \text{(Eq. 15)}$$

where $\boldsymbol{\varepsilon}_b$ is the background error vector. Since we only use static DA here (and not dynamic algorithms such as the Kalman Filter), x_b is constant. Basically, it is a mere design value of the physical state, a first guess without any observation. And, by definition, B is the covariance matrix associated with $\boldsymbol{\varepsilon}_b$, in other words, its second order momentum:

$$B = E\left[\left(\varepsilon_b - E(\varepsilon_b)\right)\cdot\left(\varepsilon_b - E(\varepsilon_b)\right)^T\right] \quad \text{(Eq. 16)}$$

B is, therefore, a statistical term, associated to $\boldsymbol{\varepsilon}_b$ probability distribution. Since, for BLUE purpose, $\boldsymbol{\varepsilon}_b$ has to be of null expectancy, and since x_b is constant, we have:

$$x_b = E[x_t]$$
$$B = E[(x_t - x_b)\cdot(x_t - x_b)^T] \quad \text{(Eq. 17)}$$

In other words, with our formulation, true state vector x_t becomes a random variable, and we approximate its distribution by a Normal Gaussian distribution, of x_b expectancy, and B covariance matrix (which we note law $NG(x_b,B)$).

b) Monte-Carlo Method

Therefore, from our point of view, computing these two terms becomes equivalent to evaluating x_t statistical behavior. Thus we set a Monte-Carlo approach and randomize our black-box model using random input parameters:

1. we simulate a sample of random parameters $\{u_t\}$;
2. we use our model to compute an associated sample of random state vectors $\{x_t=\Phi(u_t)\}$;
3. we calculate first and second order momentum associated to this sample:

$$x_b = \frac{1}{N}\sum_{t=1}^{N} x^T$$
$$B = \frac{1}{N-1}\sum_{t=1}^{N}(x_t - x_b)\cdot(x_t - x_b)^T \quad \text{(Eq. 18)}$$

(Saporta, 2011)

N sample size can be great (we used $N=10^6$ in our case), and one might want to avoid storing large vector samples. One way is to use step averaging:

$$m_0 = 0$$
$$m_{t+1} = (t.m_t + x_t)/(t+1) \quad \text{(Eq. 19)}$$

m_t is the averaging of t first simulated vectors of the sample, and $x_b = m_N$;

$$M_0 = 0$$
$$M_{t+1} = \frac{1}{t+1}[t.M_t + (x_t - x_b)\cdot(x_t - x_b)^T] \quad \text{(Eq. 20)}$$

$(t/t\text{-}1)M_t$ is the unbiased estimator of the variance associated to the first simulated vectors of the sample, and $B=(N/N\text{-}1)M_N$.

This method allows us to stock only one simulated vector x_t at all time.

c) Linear case: B spectral analysis

If our black-box model $\boldsymbol{\Phi}$ is linear, then we have exactly, if u_t follows $NG(u_0,S)$:

$$x_b = \Phi u_0$$
$$B = \Phi S \Phi^T \quad \text{(Eq. 21)}$$

and x_t's distribution is also Gaussian. Following our spaces definition, B can be singular: its rank is at best equal to Φ's (if S is invertible). Thus, in our case, since our parameters number is less than our state variables number, B cannot be invertible.

This may be a problem, for instance for variational DA approaches:

$$x_a = \text{Argmin}\,(x - x_b)^T B^{-1}(x - x_b) + (y_{obs} - H\cdot x)^T R^{-1}(y_{obs} - H\cdot x) \quad \text{(Eq. 22)}$$

However, BLUE formula itself does not require B to be invertible: R being invertible is enough. In fact, B singularity coming from $B = \Phi S \Phi^T$ merely implies that x_a will automatically be in $Im(\Phi)$, a subspace strictly included in our State Space. In fact, if we choose to switch variables and practice a Data Assimilation directly on our parameters:

- u_0 as background vector;
- S as background error covariance matrix;
- y_{obs} as observations;
- $H\circ\Phi$ as observation operator;
- R as observation covariance matrix,

BLUE formula will lead to an analyzed parameter state u_a that will exactly satisfy $x_a=\Phi u_a$, where x_a is our "regular" analysis.

To summarize, B eigenvalues and eigenvectors are related to our model image: our state vector sample is entirely included in a subspace $Im(\Phi)$, and BLUE formula automatically use B's spectral shape to compute an evaluation for x_a that meets this implicit model requirement.

Copyright © 2014-2015 by JSME

3.3 Random Parameter Generator Calibration

Simulating our input parameter vector sample $\{u_t\}$ is usually easier than simulating $\{x_t\}$: indeed, most of the time all parameters are independent, and can be simulated separately as uncorrelated scalars, since we have no additional information. However, some parameters may be obviously correlated, because they have been calculated by some "hidden" sub-model that, as said before, causes them to evolve in a subspace equal to this sub-model's image. We cannot get access to this sub-model, only to a sample of its outputs, and we want to compute an algorithm that generates a sample with the same statistical behavior.

With another formulation, the problem is as follow: given a symmetric positive matrix S, how can we generate a Gaussian random vector of law $NG(u_0,S)$? This problem is standard, and the solution described here comes from Principal Component Analysis (PCA) (Saporta, 2011).

S being symmetrical positive, there is an orthogonal matrix P (i.e. with $P^T = P^{-1}$) such that:

$$S = \mathrm{PDP^T}$$

$$\text{with } D = \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{bmatrix} = \begin{bmatrix} \mathrm{D_r} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{(Eq. 23)}$$

where $\lambda_1 \geq \ldots \geq \lambda_r > 0$ are S's positive eigenvalues. If we introduce:

$$F = \begin{bmatrix} I_r \\ 0 \end{bmatrix} \quad \text{(Eq. 23)}$$

$$\Lambda = F^T P^T \quad \text{(Eq. 24)}$$

then we have both:

$$S = \Lambda^T D_r \Lambda \quad \text{(Eq. 25)}$$

$$D_r = \Lambda S \Lambda^T \quad \text{(Eq. 26)}$$

Let us introduce α a vector of r dimension, defined by:

$$\alpha = \Lambda(u - u_0) \quad \text{(Eq. 27)}$$

Then, if u law $NG(u_0,S)$, then α follows $NG(0,D_r)$, and vice versa.

D_r being diagonal, each component of α can be simulated separately. u can then be computed with:

$$u = u_0 + \Lambda^T \alpha \quad \text{(Eq. 28)}$$

From this perspective, Λ^T is a meta-model that emulates the real "hidden" model that was used to generate the sample $\{u_t\}$, from which u_0 and S were calculated (the same way we calculated x_b and B in previous subsection).

Incidentally, if initial model was linear, it can be proven that its emulator has exactly the same image, and a BLUE analysis using the emulator's parameter space (with a variable switch) will lead to exactly the same analysis as the one using the "hidden" model.

To summarize, we used S spectral shape to identify a subspace in which the state vector sample is included, and we built a meta-model that can be used to simulate a random sample with the same statistical distribution.

4. APPLICATION AND RESULTS

4.1 Normalized Integrated Neutron Powers

In our 0D Semi-Empirical Model, some of our input parameters are indeed likely to be correlated: the integrated neutron powers for each fuel rod, the $PIPA_{norm}$, which are physically linked together by neutron power equations.

From a practical point of view, our data for these parameters are obtained with the following process: around 50 fuel rods are actually equipped with Mobile Fission Chambers measuring their neutron activity. CAMARET Code then combines those measurements with several other parameters, to produce a neutron power mapping for the whole core. Therefore, our $PIPA_{norm}$ are indeed the outputs of a "hidden" submodel and are probably correlated.

We estimated this correlation for different samples of neutron power mappings:

- for one reactor cycle (NO116);
- for one reactor available cycles (NO1) ;
- for nine reactor available cycles (Total),

and compared them to our default correlation matrix (identity).

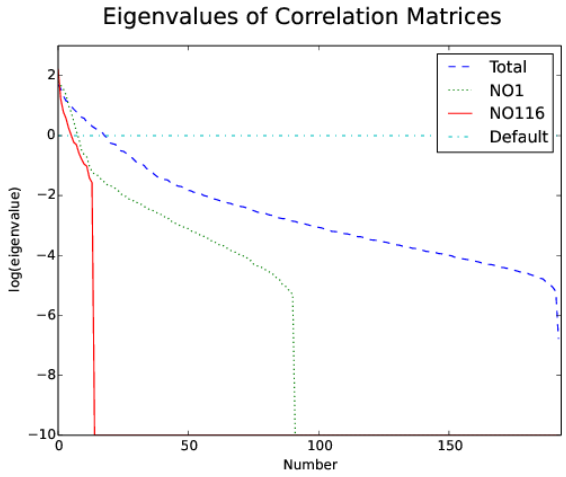


Fig. 2: Eigenvalues of correlation matrices for different sample sizes

As can be seen, for correlation matrices obtained with real data, the largest the sample, the higher the correlation matrix rank. For the sample obtained with all reactors and cycles, the rank is about 50: this means the sample can be generated using merely 50 independent parameters. This is about what we excepted, since we know $PIPA_{norm}$ are obtained by code CAMARET, whose main inputs are measured neutron power activity on the 56 fuel rods equipped with Mobile Fission Chambers. For smallest samples, since they are restricted to one reactor (or one cycle), the fuel loading pattern and thus the neutron mapping is less likely to change from one vector to another. Therefore, a smaller number of parameters is sufficient to compute the sample.

4.2 State Vector Samples comparison

We then calibrate our random parameter generator for $PIPA_{norm}$, as described in section 3.3, and apply our Monte-Carlo approach to compute an evaluation of background terms.

Here we present our results for the "TOT" calibration (for "total", obtained with the whole real data sample for $PIPA_{norm}$) and for the "DEF" calibration (for "default,

Copyright © 2014-2015 by JSME

obtained with independent $PIPA_{norm}$).

To begin with, we want to check if our sampling for our Monte-Carlo method was sufficient. In Fig. 3 and Fig. 4, we represent the computed covariance matrices spectrum for increasing sample size (from 10^1 to 10^6).

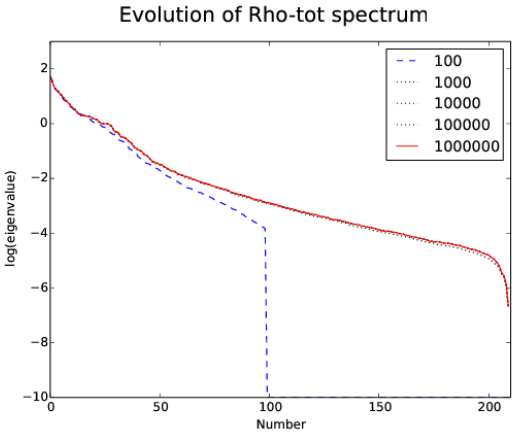


Fig. 3 Evolution of Correlation matrix spectrum with sampling, for TOT calibration

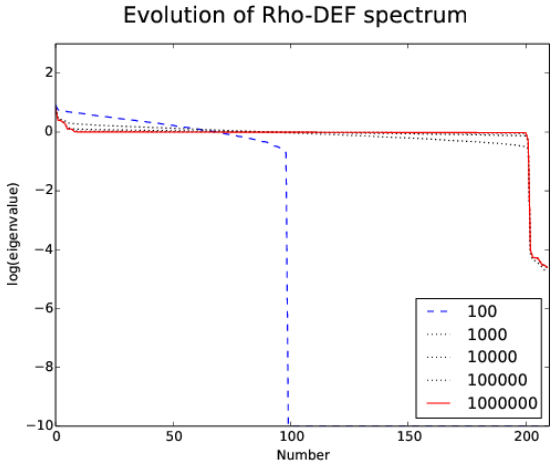


Fig. 4 Evolution of correlation matrix spectrum with sampling, for DEF calibration

Whereas for TOT calibration (correlated $PIPA_{norm}$), the final shape seems to be nearly instantly obtained, for DEF calibration (independent $PIPA_{norm}$), the "convergence" is a bit slower, even though the cut at approximately 200 is obtained already with a sample size of 1000.

If we consider all eigenvalues inferior to 10^{-2} to be null, we read on these figures that our correlation matrices (and thus, B covariance matrices) ranks are respectively around 60 and around 200. This is approximately equal to the number of independent scalar parameters required to compute the random $PIPA_{norm}$, plus the number of other independent input parameters: Q_i, TBF_i and *PTHREAC*.

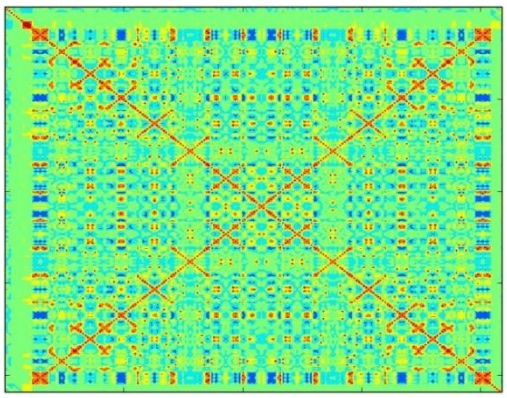


Fig. 5 Background error correlation matrix shape for TOT calibration

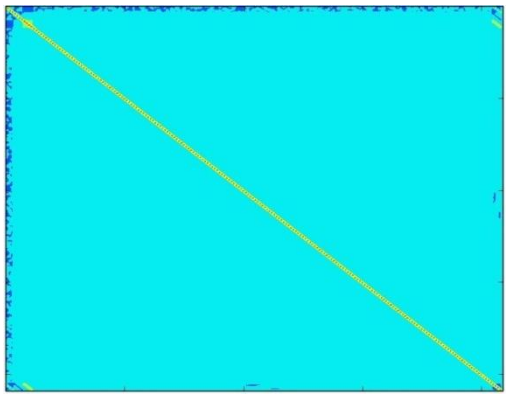


Fig. 6 Background error correlation matrix shape for DEF calibration

In Fig. 5 and Fig. 6 we see the aspects of both correlation matrices. In DEF case (independent input parameters), it is nearly diagonal: we can barely see the correlation between errors for TBC_i (little square around the diagonal, close to the upper left corner) and between errors for TBC_i and $PTHGV_i$, (little stripe in bottom left corner, also in upper right corner, since the matrix is symmetric) that are all modulated by *PTHREAC*. In TOT case (correlated input parameters), those correlations are also present, but the biggest correlation pattern is visible on the central part of the matrix: between errors for *THAC*, who are directly linked to $PIPA_{norm}$ variations around their mean value. There seems to be additional symmetries for this central shape, which is not very surprising, since we know CAMARET code uses geometrical symmetry in its interpolations: thus for instance $PIPA_{193-j}$ should be close to $PIPA_j$, their location in the core being symmetric around the center.

4.3 Influence on BLUE Analysis

Our background error covariance matrices for TOT and for DEF calibrations show clear differences: but is that of consequence for our BLUE analysis?

To study that, we performed "twin experiments": we tested our BLUE algorithm on simulated data, in four different ways:

- data simulated with "S-DEF" calibration, analyzed using "B-DEF" covariance matrix;
- data simulated with "S-TOT" calibration, analyzed using "B-TOT" covariance matrix;

Copyright © 2014-2015 by JSME

- data simulated with "S-DEF" calibration, analyzed using "B-TOT" covariance matrix;
- data simulated with "S-TOT" calibration, analyzed using "B-DEF" covariance matrix.

In all four cases, we used the same background vector. In the two last experiments, we purposely did a mistake by switching the covariance matrix: we thus expect the analysis to worsen, compared to the two first ones.

S-DEF / B-DEF	Q_i *(t/s)*	TBF_i *(°C)*	TBC_i *(°C)*	$THAC_j$ *(°C)*	$PTHGV_i$ *MW*	ε_{fit}
Background	0.10	0.4	1.0	7.3	35	1.7%
Projection	0.07	0.4	1.1	3.0	10	-
Analysis	0.05	0.3	0.3	2.7	9	0.8%

Tab. 1 Errors (2σ) for a BLUE using B-DEF on a sample simulated using S-DEF

S-TOT / B-TOT	Q_i *(t/s)*	TBF_i *(°C)*	TBC_i *(°C)*	$THAC_j$ *(°C)*	$PTHGV_i$ *MW*	ε_{fit}
Background	±0.10	±0.4	±0.7	±6.7	±26	±1.1%
Projection	±0.07	±0.4	±1.1	±3.0	±10	-
Analysis	±0.05	±0.3	±0.2	±0.8	±8	±0.5%

Tab. 2 Errors (2σ) for a BLUE using B-TOT on a sample simulated using S-TOT

S-DEF / B-TOT	Q_i *(t/s)*	TBF_i *(°C)*	TBC_i *(°C)*	$THAC_j$ *(°C)*	$PTHGV_i$ *MW*	ε_{fit}
Background	±0.10	±0.4	±1.0	±7.3	±35	±1.7%
Projection	±0.07	±0.4	±1.1	±3.0	±10	-
Analysis	±0.06	±0.4	±0.6	±6,7	±11	±1.2%

Tab. 3 Errors (2σ) for a BLUE using B-TOT on a sample simulated using S-DEF

S-TOT / B-DEF	Q_i *(t/s)*	TBF_i *(°C)*	TBC_i *(°C)*	$THAC_j$ *(°C)*	$PTHGV_i$ *MW*	ε_{fit}
Background	±0.10	±0.4	±0.7	±6.7	±26	±1.1%
Projection	±0.07	±0.4	±1.1	±3.0	±10	-
Analysis	±0.05	±0.3	±0.3	±2.8	±9	±0.7%

Tab. 4 Errors (2σ) for a BLUE using B-DEF on a sample simulated using S-TOT

For one random loop, and one random fuel rod, we examine the performance of our three estimations:

- the Background vector, which is constant;
- the Projection vector, obtained by averaging, for each component, all available measurements;
- the Analysis vector, which is the output of the BLUE.

We calculate the committed error for each of these estimations, and examine its statistics. Since the biases are all null (or quasi null) we only presented the 95% confidence interval.

A first glance on our results shows one general tendency: the effect on key parameters estimation (Q_i, TBC_i, $PTHGV_i$) is very low. The only components affected by the switching are the $THAC_j$ (since they are directly linked with $PIPA_{norm}$, this was predictable) and also, to some extent, the bypass coefficient ε_{fit}.

In Tab. 2, which means, for the case where we put correlation between $PIPA_{norm}$ and correctly used them with BLUE analysis (using the properly constructed "B-TOT" matrix), the performances are the best, especially for the upper fuel rod temperatures $THAC_j$.

5. CONCLUSIONS

By adjusting the bypass coefficient used in the core-quarter averaging in our model, we obtained a 0D semi-empirical model for the primary coolant.

By calibrating our Random Parameter Generator on real data, we modified our Monte-Carlo method for computing a Background Error Covariance Matrix for BLUE analysis.

Some of our input parameters, the normalized thermal powers for each fuel rod ($PIPA_{norm}$), were indeed outputs of a "hidden" sub-model and thus correlated with each other. Instead of using this "hidden" sub-model, we built an emulator of it that allowed us to build a random vector with similar statistical behavior.

We then computed one covariance matrix for each calibration of our Monte-Carlo Method, and used them for a BLUE analysis on both kinds of simulated data.

Overall, using the right covariance matrix improves greatly the performance for in-core parameter estimation, but using the wrong covariance matrix does not greatly deteriorate the performance for loop component parameter estimation, which are key elements for on-line monitoring.

NOMENCLATURE

BLUE	Best Linear Unbiased Estimator
CAMARET	Semi-Empirical interpolation code for in-core neutron mapping
CFD	Computational Fluid Dynamics
DA	Data Assimilation
DEF	Default configuration for random parameter generator (uncorrelated parameters)
ε_{fit}	Bypass coefficient for core-quarter averaging that satisfy power balance between first and secondary coolant
NG(x,M)	Normal (Gaussian) Distribution of x expectation and M variance)
PIPA	Integrated Neutron Power for a fuel assembly (in MW)
$PIPA_{norm}$	Normalized Integrated Neutron Power for a fuel assembly (of 1.0 mean value)
PGMPP	Primary Pumps Power (in thermal MW)
PTHGV	Steam Generator Thermal Power (in MW)
PTHREAC	Reactor Full Thermal Power (in MW)
PWR	Pressurized Water Reactor
Q	Mass loop-flow (in t/s)
RCS	Reactor Coolant System
TBC	Hot Leg Temperature (in °C)
TBF	Cold Leg Temperature (in °C)
TC RIC	In-Core Thermocouples
THAC	Upper Fuel Rod Temperature (in °C)
TOT	"Total" configuration for random parameter generator (parameters with the same statistical distribution as observed on the whole sample of available real data)

REFERENCES

Bellet, Benhamadouche, and Ferret, 2011. "Secondary Flows in PWR hot legs: CDF might bring some light on the flow behavior". *The 14th International Topical Meeting on*

Copyright © 2014-2015 by JSME

Nuclear Reactor Thermalhydraulics NURETH-14, Toronto, Canada.

Bocquet, 2012. « Introduction to Data Assimilation Methods and Principles in Geophysics ». *ENSTA and M2 OACOS lecture notes.*

Bouttier and Courtier, 2002. "Data Assimilation concepts and methods". *Meteorological Training Course Lecture Series ECMWF.*

Deneux, Favennec, Piedfer, Antunes and Veau, 2013. "Using CFD for metrology prediction in flowmetering: RCS flow measurement in PWRs with elbow taps". *ASME Power Conference, Boston, USA.*

Favennec, Piedfer and Veau, 2011. "Reactor Power Measurement: EDF's strategy and ongoing work on measurement issues". *3rd International Workshop on Metering of Feedwater Flow, Temperature and Thermal Power, Chatou, France.*

Mercier, Favennec and Girard, 2014. "Feasibility of enhanced estimation of PWR Primary Parameters". *ICONE22, n°30118. Prague, Czech Republic.*

Saporta, 2011. "Probability, Data Analysis and Statistics". *Editions TECHNIP.*

Talagrand, 1997. "Data Assimilation, an introduction". *Journal of the Meteorological society of Japan.*

Copyright © 2014-2015 by JSME

Bibliographie

[ACNN89] S. Aubry, J. Cahouet, G. Nicolas et C. Niedergang : A finite volume approach for 3D two phase flows in tuble bundles : the THYC code. *Proceedings of the Fourth International Topical Meeting on Nuclear Reactor Thermal-Hydraulics*, pages 1247–1253, 1989.

[Bay63] T. Bayes : An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, pages 53 :370–418, 1763.

[BC02] F. Bouttier et P. Courtier : Data assimilation concepts and methods. *Meteorological Training Course Lecture Series ECMWF*, 2002.

[Boc12] M. Bocquet : Introduction aux principes et méthodes de l'assimilation de données en géophysique. *Notes du cours de Master M2 OACOS et de l'ENSTA*, 2012.

[Bod04] P. Bodet : CAMARET 4 v.1.5.3. notice utilisateur. Rapport technique, EDF R&D, 2004.

[Bri14] A. Briant : Paprica : Bilan du projet « exploiter les améliorations du point de fonctionnement primaire par le calcul ». Rapport technique, EDF R&D, 2014.

[Com13] M.-J. Comte : Faisabilité de reconstruction de la température en branches chaudes dans un réacteur nucléaire à partir de données réelles. Rapport technique, EDF R&D, 2013.

[Cre59] G.P. Cressman : An operational objective analysis system. *Mon. Wea. Rev.*, pages 87, 367–374, 1959.

[DA11] O. Deneux et M. Arenas : Note de synthèse sur la détermination, par simulation numérique, des paramètres permettant une mesure plus précise du débit primaire à partir du δp-coude existant. Rapport technique, EDF R&D, 2011.

[DEL08] J.-M. DELHAYE : *Thermohydraulique des Réacteurs*. EDP SCIENCES, 2008.

[Div06] Divers : Epr : Rapport préliminaire de sûreté de flamanville 3. Rapport technique, EDF - DPI - CNEN, 2006.

[Dum10] T. Duminy : Circuit primaire d'une centrale nucléaire : conception d'estimateurs innovants des paramètres de fonctionnement pour intégration dans un outil de surveillance. Rapport technique, EDF R&D, 2010.

[Esc11] T. Escabasse : Estimateur de température en branche chaude à partir de l'instrumentation existante. Rapport technique, EDF R&D, 2011.

[Eve09] G. Evensen : *Data assimilation, The Ensemble Kalman Filter, 2nd ed.* Springer, 2009.

[Gan63] L.S. Gandin : Objective analysis of meteorological fields. *Gidrometeor. Izd., Leningrad*, 1963.

[Gau09] C.F. Gauss : *Theoria Motus Corporum Coelestium.* Perthes, Hamburg, 1809.

[Gau26] C.F. Gauss : *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae.* Göttingen, 1826.

[Gir12] S. Girard : *Diagnostic du colmatage des générateurs de vapeur à l'aide de modèles physiques et statistiques.* Thèse de doctorat, EDF R&D, 2012.

[Kal60] R.E. Kalman : A new approach to linear filtering and prediction problems. *J. Basic Eng.*, pages 82D, 35–45, 1960.

[Lap74] P.S. Laplace : Mémoire sur la probabilité des causes par les Évènements. *Savants Étranges*, pages 6 :621–656, 1774.

[Leg05] A.M. Legendre : *Nouvelles Méthodes pour la Détermination des Orbites des Cométes.* Firmin Didot, Paris, 1805.

[Lor81] A. Lorenc : A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, pages 109 :701–721, 1981.

[Mar99] A.A. Markov : *Zakon bol'shik chisel i sposob naimen'shikh kvadratov : Loi des Grands Nombres et Méthode des Moindres Carrés.* Izbrannye trudy, 1899.

[Mer] T. Mercier : Assessment of a data assimilation approach for enhanced estimation of PWR primary parameters. Poster for Summer School on Data Assimilation and Inverse PRoblems at University of Reading, United Kingdom, July 22-26, 2013.

[MFG14] T. Mercier, J.-M. Favenec et A. Girard : Feasibility of enhanced estimation of PWR primary parameters : Twin experiments for assessing data assimilation benefits (0d-model with a monte-carlo approach on pseudo-real data). *Proceedings of the 2014 22nd International Conference on Nuclear Engineering ICONE22, July 7-11 2014, Prague, Czech Republic*, pages ICONE22–30118, 2014.

[MFG15] T. Mercier, J.-M. Favenec et A. Girard : Data assimilation and PWR primary parameters estimation : Enhancement of a 0D model and calibration of random parameter generator for background terms evaluation. *Proceedings of ICONE-23, 23th International Conference on Nuclear Engineering, May 17-21 2015, Chiba, Japan*, pages ICONE23–1442, 2015.

[NR08] G. NAUDET et P. REUSS : *Energie, Electricité et Nucléaire.* EDP SCIENCES, 2008.

[Pie09] O. Piedfer : Doctrine du bilan thermique de référence bil100. Rapport technique, EDF - DPN, 2009.

[Pon09] A. Ponçot : *Assimilation de Données pour la dynamique du xénon dans les cœurs de centrale nucléaire.* Thèse de doctorat, EDF R&D, 2009.

[Sap11] G. Saporta : *Probabilités, analyse de données et statistiques.* Editions TECHNIP, 2011.

[Tal97] O. Talagrand : Data assimilation, an introduction. *Journal of the Meteorological Society of Japan*, 1997.

[Tar87] A. Tarantola : *Inverse Problem Theory.* Elsevier, 1987.

[TDR94] Q. Tinc, F. David et P. Racle : THYC, un code de thermohydraulique 3D pour les générateurs de vapeur, les échangeurs de chaleur et les condenseurs : développements récents et cas de validation. Rapport technique, EDF R&D, 1994.

[Vai12] G. Vaillant : Faisabilité de l'estimation des températures de mélange en branches chaudes des réacteurs à eau pressurisée à partir des capteurs dans le cœur. Rapport technique, EDF R&D, 2012.

Table des figures

1.1 Schéma de principe d'un REP (source : Collection Génie Atomique) 16
1.2 Schéma général d'un REP de Framatome (AREVA NP) (source : Collection Génie Atomique) 16
1.3 Tour aéroréfrigérante d'un REP de Framatome (source : Collection Génie Atomique) 17
1.4 Cuve d'un REP de Framatome (source : Collection Génie Atomique) 20
1.5 Générateur de Vapeur d'un REP de Framatome (source : Collection Génie Atomique) 22
1.6 Principaux circuits de sauvegarde du primaire d'un REP (source : Collection Génie Atomique) 23
1.7 Assemblage 17×17 d'un REP (source : Collection Génie Atomique) 24
1.8 Représentation schématique de la disposition des AC dans le cœur (source : Collection Génie Atomique) 26
1.9 Hétérogénéité du champ de Température en Branche Chaude (obtenue par simulation *Code_Saturne*®) 32

2.1 Résumé des systèmes de mesure disponibles sur le circuit primaire 36
2.2 Principe de la mesure par ΔP-Coude 41
2.3 Chaîne d'acquisition et de traitement de la mesure de débit primaire par ΔP-Coude 41
2.4 Détail de l'emplacement des piquages pour capteurs ΔP-Coude 44
2.5 Vue du système d'écopes et de l'hétérogénéité du champ de température en Branche Chaude (dans un REP 1300 MW) 45
2.6 Schéma représentant l'emplacement des différentes sondes RCP de mesures de température 45
2.7 Emplacements et fonctions des thermocouples au sein du cœur d'un palier 1300 MW 46
2.8 Courbe de réponse d'un thermocouple de type K (approximation affine pour la plage de température $200°C - 350°C$ 47
2.9 Paliers de puissance suivis lors de la montée en charge du réacteur, en début de cycle 49
2.10 Schéma de principe d'un BIL100 50
2.11 Emplacements des Assemblages Combustibles parcourus par les CFM 53
2.12 Position des chambres RPN autour du cœur 55
2.13 Méthode dite des « bilans enthalpiques » 57
2.14 Coupe des flux de masse Q_z à travers le cœur obtenue via THYX/COCCINELLE 58
2.15 Relation de proportionnalité constatée entre les gradients de température haut-bas AC et les $PIPA$s, pour les AC équipés de thermocouples, pour le cycle 16 de la tranche AA 58

2.16 Emplacements des BC (cas d'un 1300 MW) : vue du dessus du plénum supérieur, tel que simulé par CODE SATURNE 59
2.17 Segmentation du cœur en quatre quarts 60
2.18 Agencement des codes de calcul au sein du cœur 62
2.19 Modèle Semi-Empirique première version 72
2.20 Confrontation du résultat de l'évaluation du coefficient ϵ par recherche dichotomique et par loi affine 74
2.21 Modèle Semi-Empirique avec bouclage par fit de coefficient de bypass 75

3.1 Exemple scalaire simple de problème d'assimilation de données 80
3.2 Cas scalaire : distributions de probabilités des estimateurs initiaux et de l'analyse 81
3.3 Exemple de problème simple d'Assimilation de Données : estimation de la température moyenne d'une pièce 82
3.4 Exemple d'Analyse de Cressman dans un cas 1D avec une fonction poids de type max 84
3.5 Exemple d'Analyse de Cressman dans un cas 1D avec une fonction poids de type exp 84
3.6 Schéma de principe d'une Analyse de Cressman 86
3.7 Exemple de fonction coûts de type 3DVAR pour un cas scalaire 90
3.8 Schéma de principe du Filtre de Kalman 95
3.9 Schéma de principe de la méthode de Monte-Carlo pour le calcul des termes d'ébauche 100

4.1 Schéma de principe de l'utilisation du modèle-fit en mode direct avec la méthode Monte-Carlo 111
4.2 Schéma des différents espaces vectoriels et des deux opérateurs principaux . 113
4.3 Forme spectrale de la matrice de corrélation associée à la matrice de covariance d'erreur obtenue avec la configuration par défaut pour la randomisation des paramètres 114
4.4 Exemple de recoupement des barres d'incertitudes (Tranche NO1, cycle 16, le 9 Octobre 2008) 117
4.5 Expérience Jumelle Standard avec paramètres homogènes : distribution statistique des erreurs commises pour $Q1$ (débit-boucle primaire boucle 1) . . . 120
4.6 Expérience Jumelle Standard avec paramètres homogènes : distribution statistique des erreurs commises pour $TBC1$ 120
4.7 Formes spectrales des matrices de corrélation associées aux matrices S obtenues pour les différents choix d'échantillons 124
4.8 Dispersion Statistique des Erreurs d'ébauche, de projection et d'analyse BLUE pour différents choix de B (TOT, DEF, ACP, DIAG et INF) ou d'analyse 3DVAR sur les paramètres après ACP 133
4.9 Illustration des contributions de chaque composante du vecteur d'innovation pour la correction de l'ébauche effectuée lors de l'analyse BLUE (via les lignes de K) pour le cas B_{TOT} et le cas B_{DIAG}, pour les composantes (de l'avant vers l'arrière) Q_2,TBF_2, TBC_2, $THAC_4$, $PTHGV_2$ et ϵ_{fit}. 137
4.10 allure du graphe de la fonction ω correspondant au coefficient « de diffusion Kalman » 140
4.11 Exemple de simulateur d'évolution par marche aléatoire 141
4.12 Exemple de simulateur d'évolution par transition aléatoire 142
4.13 Exemple de simulateur d'évolution par transitions réelles successives 143

4.14 Suivi de l'erreur quadratique totale au cours d'un mois de 30 jours (cas sans biais) 146
4.15 Suivi de Q_2,TBC_2, $PTHGV_2$, $THAC_3$ et $THAC_4$ sur un mois de 30 jours par les estimateurs projection, BLUE, OI et Kalman (cas non biaisé) 148
4.16 Suivi de l'erreur quadratique sur les THAC au cours d'un mois de 30 jours (cas sans biais) 149
4.17 Suivi de Q_2,TBC_2, $PTHGV_2$, $THAC_3$ et $THAC_4$ sur un mois de 30 jours par les estimateurs projection, BLUE, OI et Kalman (cas biaisé) 151

Liste des tableaux

2.1 Exemples de trigrammes du système élémentaire 37
2.2 Paramètres pour le ΔP-Coude 43
2.3 Emplacements des différents capteurs ΔP-Coude 44
2.4 Capteurs de températures RCP pour la boucle 1 46

4.1 Configuration par défaut de la randomisation des paramètres 113
4.2 Résultats de la méthode Monte-Carlo avec la randomisation par défaut . . . 114
4.3 Configuration des incertitudes de mesure 115
4.4 Résultats pour une Expérience Jumelle standard avec des paramètres homogènes 119
4.5 Statistiques des différents échantillons de PIPAN 123
4.6 Résultats pour les analyses droites 125
4.7 Résultats pour les analyses croisées 126
4.8 Résultats pour les analyses semi-croisées 127
4.9 Résultats pour un 3DVAR pratiqué sur l'espace des paramètres obtenu après une Analyse en Composantes Principales de S pour le cas TOT 129
4.10 Résultats des différentes analyses pratiquées sur l'échantillon TOT avec des matrices de covariance d'erreur d'ébauche B différentes 132
4.11 Biais après analyse d'un échantillon de type TOT avec des observations biaisées, en utilisant respectivement les matrices B_{TOT} et B_{DIAG}, pour la boucle 1 et l'AC 4 136
4.12 Incertitudes (2σ) pour l'analyse TOT d'un échantillon de type TOT pour différentes pertes d'instrumentation. 138
4.13 Erreurs quadratiques moyennées sur toute la durée d'un cycle (cas sans biais) 147
4.14 Résultats des différentes analyses appliquées aux données dynamiques non biaisées (pour la boucle 1 et l'AC 3) 147
4.15 Erreurs quadratiques moyennées sur toute la durée d'un cycle (cas avec biais) 150
4.16 Résultats des différentes analyses appliquées aux données dynamiques biaisées (pour la boucle 1 et l'AC 3) 152