

Low-flow simulation and forecasting on French river basins : A hydrological modelling approach

Raji Pushpalatha

▶ **To cite this version:**

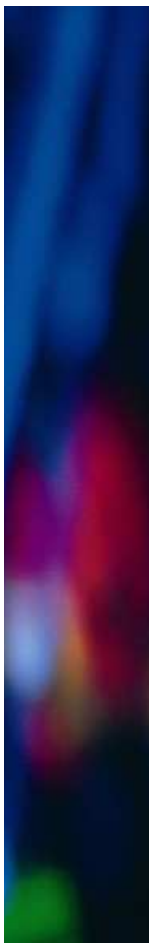
Raji Pushpalatha. Low-flow simulation and forecasting on French river basins : A hydrological modelling approach. Earth Sciences. AgroParisTech, 2013. English. NNT : 2013AGPT0012 . pastel-00912565

HAL Id: pastel-00912565
https://pastel.hal.science/pastel-00912565

Submitted on 2 Dec 2013

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

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



Doctorat ParisTech

École doctorale Géosciences et Ressources naturelles

THÈSE

pour obtenir le grade de docteur délivré par

L'Institut des Sciences et Industries du Vivant et de l'Environnement

(AgroParisTech)

Spécialité: Hydrologie

Raji PUSHPALATHA

18 janvier 2013

Low-flow simulation and forecasting on French river basins: a hydrological modelling approach

Simulation et prévision des étiages sur des bassins versants français : approche fondée sur la modélisation hydrologique

Directeur de thèse : **Vazken ANDRÉASSIAN**

Co-encadrement de la thèse : **Charles PERRIN**



Jury

Mme Florentina MOATAR	Université François Rabelais, Tours	Rapporteur
M. Christophe BOUVIER	UMR Hydrosciences, Montpellier	Rapporteur
M. Laurent PFISTER	Institut Gabriel Lippmann, Belvaux	Examinateur
M. Thibault MATHEVET	EDF-DTG, Grenoble	Examinateur
M. Pascal MAUGIS	ONEMA, Vincennes	Invité

AgroParisTech

Irstea, Unité de recherche Hydrosystèmes et Bioprocédés

1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony cedex

Acknowledgements

I would like to thank several people and organizations for their invaluable helps to fulfil this work. First of all I would like to express my heartfelt thanks to my supervisor Dr. Charles Perrin, Irstea, Antony, France, for his valuable guidance, critical suggestions and immense help throughout my PhD. I have immense pleasure to express my sincere gratitude to Dr. Vazken Andréassian, head of the hydrology group, Irstea, for his support, suggestions and constant encouragements during the course of the project. I am also indebted to Dr. Maria-Helena Ramos, Irstea, for her valid advices at different phases of my work. I am expressing my sincere gratitude to Dr. Cécile Loumagne and Didier Pont, respectively former and current head of the HBAN unit, Irstea, for their support. I wish also to thank Pr. Florentina Moatar, Dr. Christophe Bouvier, Dr. Laurent Pfister, Dr. Thibault Mathevet and Dr. Pascal Maugis, for accepting to be part of the jury and to evaluate this work.

I am thankful to Peschard Julien who was always ready to make series of maps throughout my work. I acknowledge Dr. Nicolas Le Moine and Dr. Thibault Mathevet for their valuable suggestions at different phases of my work.

I am very grateful to *Météo-France* for providing meteorological data and *SCHAPI* for providing flow data necessary to carry out my tests. The last chapter of this thesis could not have been carried out without the help of Thibault Mathevet (EDF-DTG) and Claudine Jost (Seine Grands Lacs) who helped building the database and who are thanked for that. I also wish to thank *DIM ASTREA* of *Région Ile-de-France*, as well as *ONEMA*, for their financial support to conduct this study.

I express my sincere gratitude to Dr. Elisabeth Maltese for the interesting French classes and her guidance.

I express my gratitude to Dr. Lionel Berthet, Dr. Audrey Valery, Marine Riffard and Dr. Mamoutou Tangara for their help and support throughout my work.

I acknowledge my gratitude to my colleagues: Ioanna, Annie, François, Florent, Laurent, Gianluca, Pierre, Pierre-Yves and Damien for their help and support throughout my work period at Irstea.

I am also extremely thankful to my friends: Dr. Agnel Praveen Joseph, Dr. Sunil Kumar, Rubie M Sam, Dr. Sonam Popli, Dr. Debjani Bagchi, Dr. Prasanth Karikkethu and Ananya at Cité Universitaire for their support throughout my stay in Paris.

I am most grateful to my parents and my sister for the love and affection they showered on me and the support they have shown during this long yet rewarding process of finishing my degree.

The words fail me to express my appreciation to one, whose love and persistent confidence in me, has taken the load off my shoulder. I owe Dr. Govindan Kutty, my husband, for being unselfishly let his intelligence, passions, and ambitions collide with mine.

Abstract

Long-term stream low-flow forecasting is one of the emerging issues in hydrology due to the escalating demand of water in dry periods. Reliable long-lead (a few weeks to months in advance) streamflow forecasts can improve the management of water resources and thereby the economy of the society and the conditions for aquatic life. The limited studies on low flows in the literature guided us to address some of the existing issues in low-flow hydrology, mainly on low-flow simulation and forecasting. Our ultimate aim to develop an ensemble approach for long-term low-flow forecasting includes several prior steps such as characterisation of low flows, evaluation of some of the existing model's simulation efficiency measures, development of a better model version for low-flow simulation, and finally the integration of an ensemble forecasting approach.

A set of catchments distributed over France with various hydrometeorological conditions are used for model evaluation. This data set was first analysed and low flows were characterized using various indices. Our objective to better evaluate the models' low-flow simulation models resulted in the proposition of a criterion based on the Nash-Sutcliffe criterion, but calculated on inverse flows to put more weight on the errors on extreme low flows. The results show that this criterion is better suited to evaluate low-flow simulations than other commonly used criteria.

Then a structural sensitivity analysis was carried out to develop an improved model structure to simulate stream low flows. Some widely used models were selected here as base models to initiate the sensitivity analysis. The developed model, GR6J, reaches better performance in both low- as well as high-flow conditions compared to the other tested existing models.

Due to the complexity of rainfall-runoff processes and the uncertainty linked to future meteorological conditions, we developed an ensemble modelling approach to issue forecasts and quantify their associated uncertainty. Thus the ensemble approach provides a range of future flow values over the forecasting window. Here observed (climatological) rainfall and temperature were used as meteorological scenarios fed the model to issue the forecasts. To reduce the level of uncertainty linked to the hydrological model, various combinations of simple updating procedures and output corrections were tested. A straightforward approach, similar to what can be done for flood forecasting, was selected as it proved the most efficient.

Last, attempts were made to improve the forecast quality on catchments influenced by dams, by accounting for the storage variations in upstream dams. Tested on the Seine and Loire basins, the approach showed mixed results, indicating the need for further investigations.

Résumé

La prévision d'étiage à long terme est l'une des questions émergentes en hydrologie en raison de la demande croissante en eau en période sèche. Des prévisions fiables de débits à longue échéance (quelques semaines à quelques mois à l'avance) peuvent améliorer la gestion des ressources en eau et de ce fait l'économie de la société et les conditions de vie aquatique. Les études limitées sur les étiages dans la littérature nous a conduit à traiter certaines des questions existantes sur l'hydrologie des étiages, principalement sur la simulation et la prévision des étiages. Notre objectif final de développer une approche d'ensemble pour la prévision à long terme des étiages se décline en plusieurs étapes préalables, telles que la caractérisation des étiages, l'évaluation de mesures existantes d'efficacité des simulations des modèles, le développement d'une version améliorée d'un modèle de simulation des étiages, et enfin l'intégration d'une approche de prévision d'ensemble.

Un ensemble de bassins distribués partout en France avec une variété de conditions hydro-météorologiques a été utilisé pour l'évaluation des modèles. Cet échantillon de données a d'abord été analysé et les étiages ont été caractérisés en utilisant divers indices. Notre objectif de mieux évaluer les simulations des étiages par les modèles a conduit à proposer un critère basé sur le critère de Nash-Sutcliffe, calculé sur l'inverse des débits pour mettre davantage de poids sur les erreurs sur les très faibles débits. Les résultats montrent que ce critère est mieux adapté à l'évaluation des simulations des étiages que d'autres critères couramment utilisés..

Une analyse de sensibilité structurelle a ensuite été menée pour développer une structure de modèle améliorée pour simuler les étiages. Des modèles couramment utilisés ont été choisis ici comme modèles de base pour commencer l'analyse de sensibilité. Le modèle développé, GR6J, atteint de meilleures performances à la fois sur les faibles et les hauts débits par rapport aux autres modèles existants testés.

En raison de la complexité du processus pluie-débit et de l'incertitude liée aux conditions météorologiques futures, nous avons développé une approche d'ensemble pour émettre des prévisions et quantifier les incertitudes associées. Ainsi l'approche d'ensemble fournit une gamme de valeurs futures de débits sur la plage de prévision. Ici, la climatologie a été utilisée pour fournir les scénarios météorologiques en entrée du modèle pour réaliser les prévisions. Pour réduire le niveau d'incertitude lié au modèle hydrologique, des combinaisons variées de procédures de mise à jour et de corrections de sortie ont été testées. Une approche directe,

similaire à ce qui peut être fait pour la prévision des crues, a été sélectionnée comme la plus efficace.

Enfin, des essais ont été réalisés pour améliorer la qualité des prévisions sur les bassins influencés par les barrages, en tenant compte des variations de stockage dans les barrages amont. Testée sur les bassins de la Seine et de la Loire, l'approche a donné des résultats mitigés, indiquant le besoin d'analyses complémentaires.

Table of contents

Acknowledgements 1

Abstract 3

Résumé 5

Table of contents 7

List of Tables 11

List of Figures 12

General introduction 1

Introduction 3
Evaluation of model's low-flow simulation efficiency 4
Why a general model structure for the French river basins? 4
Long-term forecasting of low flows and possible improvements 5
Exploratory tests to make low-flow forecasts on catchments influenced by dams 5
Organisation of thesis 6

Part I: Stream low flows and their characterization 7

Chapter 1. Low flows and low-flow generating factors in the French context 11

1.1. Introduction 13
1.2. Natural influences on low flows: catchment characteristics 15
1.2.1. Catchment geology and river low flows 16
1.3. Anthropogenic impacts on catchment characteristics and low flows 18
1.3.1. Direct influences 18
1.3.2. Indirect influences 19
1.4. Potential impacts of climate change on future river flows 20
1.5. Impact of low flows on society 21
1.6. Conclusions 21

Chapter 2. Data set and low-flow characterisation 23

2.1. Introduction 25
2.2. Description of the catchment set 25
2.2.1. Catchment selection 25
2.2.1. Data sets 27
2.2.2. Sample catchments 28
2.3. Low-flow measures 30
2.3.1. Low-flow period (low-flow spells) 31
2.3.2. Flow duration curve (FDC) 31
2.3.3. Mean annual minimum flow for different durations 31

2.3.4. Base-flow index (BFI) 31
2.4. Computations of low-flow indices on the study catchment set 32
2.5. Link between low-flow indices 36
2.6. Low-flow indices and catchment descriptors 38
2.6.1. Aridity index (AI) 38
2.6.2. Catchment geology 39
2.7. Conclusions 40

Part II: Stream low-flow simulation 43

Chapter 3. Analysis of efficiency criteria suitable for low-flow simulation 47
3.1. Introduction 49
3.2. Criteria used for the evaluation of low–flow simulation 50
3.2.1. Are existing criteria appropriate to evaluate low-flow simulations? 53
3.3. Scope of the study 54
3.4. Models and evaluation methodology 55
3.4.1. Hydrological models 55
3.4.2. Testing scheme 58
3.4.3. Criteria analysed 58
3.4.4. Approach for criterion analysis 61
3.5. Results and discussion 64
3.5.1. Overall efficiency distribution 64
3.5.2. Criteria interdependency 65
3.5.3. Analysis of model error terms 68
3.5.4. Which criterion should be used for low–flow evaluation? 69
3.5.5. What should be done with zero flows when calculating NSE^*_{iQ} ? 70
3.5.6. Power transformation and criteria 71
3.6. Conclusions 73

Chapter 4. Development of an improved lumped model for low-flow simulation 75
4.1. Introduction 77
4.2. Selection of basic model structures 77
4.3. Model testing 78
4.4. Evaluation of the selected model structures 79
4.5. Development of an improved model structure from the base models 80
4.5.1. Integration of groundwater exchange function (F) 81
4.5.2. Addition of new stores 84
4.6. Results and discussion 86
4.6.1. Can the existing groundwater exchange term in GR5J be improved? 86
4.6.2. Should the volumetric splitting between flow components be adapted to each catchment? 87
4.6.3. Should a new store be added in series or in parallel? 88
4.6.4. Does the formulation of routing stores matter? 89
4.6.5. Comparing the results of GR5J and GR6J 90
4.6.6. Illustration of model's results 91
4.6.7. Parameter stability and identifiability 92

4.1. Conclusions 95

Part III: Stream low-flow forecasting 97

Chapter 5. Low-flow forecasting: implementation and diagnosis of a long-term ensemble forecasting approach 101

5.1. Introduction 103
5.2. Literature on long-term forecasting in the French River basins 104
5.3. Scope of the study 105
5.4. Methodology 105
5.4.1. An ensemble approach towards low-flow forecasting 106
5.4.2. Description of the forecasting approach 107
5.5. How to improve the quality of a model forecast? 108
5.5.1. Output bias corrections 110
5.5.2. Model updating to improve the forecast quality 111
5.6. Assessment of the forecasting system 113
5.6.1. Measure based on error in magnitude 113
5.6.2. Measures based on contingency table 114
5.6.3. Measure based on probability 115
5.7. Results and discussion 117
5.7.1. How long are model forecasts useful (without any corrections)? 118
5.7.2. Influence of simulation error on model forecasts 119
5.7.3. How to improve model's forecast quality? 120
5.7.1. Maximum possible lead-time (MPLT) 128
5.8. Conclusions 135

Chapter 6. Accounting for the influence of artificial reservoirs in low-flow forecasting 137

6.1. Introduction 139
6.2. Scope of the study 140
6.3. Test catchments 140
6.1. Methodology 142
6.1.1. To account for the reservoirs into an R-R model 142
6.1.2. Model testing and evaluation 145
6.2. Results and discussion 145
6.2.1. Can we really improve the simulation performance of models by accounting the reservoirs? 145
6.2.2. Can we improve the forecast quality? 148
6.3. Conclusion 150

General conclusion 151

Synthesis 153
Perspectives 154

References 157

Appendices 173

A Selection of objective function for calibration 175
A.1 Test of a set of criteria 177

B A review of efficiency criteria suitable for evaluating low-flow simulations ... 179

C Description of the hydrological models tested in this study 193
C.1 Model Gardénia (GARD) 197
C.2 GR-series (GR4J & GR5J) 198
C.3 HBV0 200
C.4 IHAC 201
C.5 MOHY 202
C.6 MORD 203
C.7 TOPM 204

D The developed model versions and their performances 205
D.1 Versions of model GR4J 207
D.2 Test results of versions of GR4J 208
D.3 Performance of versions of MORD & TOPM 208

E A downward structural sensitivity analysis of Hydrological models to improve low-flow simulation 211

F Model's sensitivity to temporal variations of potential evapotranspiration 225
F.1 Introduction 227
F.2 Tested models 228
F.3 Results and discussion 229
F.4 Conclusions 230

List of Tables

Table 2.1 Variability in size and the main hydroclimatic conditions over the catchment set 27

Table 2.2: Characteristics of sample catchments..... 29

Table 3.1: Studies that used criteria to evaluate low-flow simulation quality..... 51

Table 3.2: Selection of evaluation criteria and their corresponding formulation and specific values (Q_i and $\hat{Q}_i$ are the observed and simulated flows, respectively, n the total number of time steps, T a low-flow threshold (here $T = Q_{90}$), $\overline{Q}$ the mean of Q, and ε a small constant) ... 60

Table 3.3: Scatter plots of pairs of criteria on the 940–catchment set for the GR4J in validation..... 66

Table 3.4: The Spearman correlation coefficient for the criteria corresponding to the plots in Table 3.3 66

Table 3.5: Scatter plots of pairs of criteria on the 940–catchment set for the MORD in validation 67

Table 3.6: The Spearman correlation coefficient for the criteria corresponding to the plots in Table 3.5 67

Table 4.1: Selected model structures and their number of parameters..... 78

Table 4.2: Average efficiency values in validation for the lumped models for various criteria (criteria on Q and iQ put more emphasis on high flows and low flows, respectively, and criterion on lnQ is intermediate). 79

Table 4.3: Modified versions of the GR5J model in terms of groundwater exchange function..... 83

Table 4.4: Modified versions of the GR5J model and their main characteristics..... 85

Table 4.5: Mean model performance for versions M4 and M5..... 87

Table 4.6: Mean efficiency values for versions M5 and M6..... 88

Table 4.7: Mean model performance for versions M8–M11 (multiple routing stores) and mean relative performance RE* with reference to M8 over the catchment set..... 88

Table 4.8: Mean efficiency values of M8 vs M12..... 90

Table 4.9: Mean performance of GR5J and GR6J and relative performance of GR6J with reference to GR5J over the catchment set for various criteria (criteria on Q and iQ put more emphasis on floods and low flows, respectively). The results were obtained in validation..... 91

Table 5.1: Notations and their definitions (Q - observed streamflow; $\hat{Q}$ - simulated flow; $\hat{Q}f$ - forecast value; $\hat{\hat{Q}}f$ - corrected forecast value) 109

Table 5.2: Formulations of bias corrections and model updating techniques used in this study..... 110

Table 5.3: Formulations of model output updating techniques..... 112

Table 5.4: Contingency table..... 114

Table 5.5: Scores based on the contingency table..... 114

Table 5.6 Median values of the BSS for the different correction procedures 129

Table 6.1: Summary of data set in the Seine basin..... 143

Table 6.2: Summary of data set in the Loire basin..... 143

List of Figures

Figure 1.1: Natural hydrological processes and catchment storages (WMO, 2008) 13

Figure 1.2: Illustration of stream low flows of one of the study catchments (Catchment L'Aber Wrac'h at Le Drennec) for a period of three years (2003-2005) 14

Figure 1.3: Left: Summer low flows of the Garonne River downstream Toulouse (Source: A. Dutartre, Irstea) and right: winter low flows of the Columbia River, Canada (Source: WMO, 2008) 15

Figure 1.4 Geological map of France (Source: Geological Survey of France) 17

Figure 1.5 Total water availability and its utility by different sectors in France in 2009 (Source: Commissariat général au développement durable, 2012) 19

Figure 1.6: Changes in the annual discharges (using SIM model) between the 2046-2065 and 1971-2000 periods averaged over France using the inputs from 14 climate models (Source: Boé et al., 2009) 20

Figure 2.1: Location of 1000 catchments of the data set (the highlighted background indicates the presence of mountains) 26

Figure 2.2: Mean Q/P vs P/PE values for the catchments in the data set 28

Figure 2.3: Illustration of location of sample catchments and their daily flow hydrographs from 2003 to 2005 30

Figure 2.4: Illustration of the period of occurrence of low-flows in the test data set 33

Figure 2.5: Flow duration curve of the sample catchments for a period of 2003 to 2005 (the values below 0.1 mm/d are not shown) 34

Figure 2.6: Illustration of spatial variability of the low-flow indices: (a) - MAM7; (b) - QMNA5; (c) - Q90/Q50; (d) - BFI 36

Figure 2.7: Relation between Q_{90} and MAM_7 37

Figure 2.8: Relation between $QMNA_5$ and MAM_7 37

Figure 2.9: Aridity index vs. low-flow indices 38

Figure 2.10: Boxplots of distribution of Q_{90} values of catchments based on their geology (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles) 39

Figure 2.11: Distribution of BFI values based on geology (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles) 40

Figure 3.1: Illustration of changes in the series of flows, model errors and cumulated squared errors using root square and logarithm transformations for the Moselle River at Custines (Hydro code: A7010610; catchment area: 6830 km²; period: 01/01/1980–31/12/1981). 54

Figure 3.2 Schematic diagrams of the structures of (a) the GR4J model and (b) the MORD model (PE: potential evapotranspiration; P: rainfall; Q: streamflow; X parameters; other letters are internal model variables) 56

Figure 3.3: Box plots of performance (NSE^*_Q in validation) of the GR4J model on the 940 catchments as a function of catchment size (boxplots show the 0.05, 0.25, 0.5, 0.75 and 0.95 percentiles from bottom to top; each box plot contains one tenth of catchments) 57

Figure 3.4: Example of derivation of box plots from the distribution of weighted flows (here weights are the squared errors) to the flow duration curve (distribution of non weighted flows). The same example as in Figure 3.1 is used here (Period: 01/01/1980 – 3112/1981) 62

Figure 3.5: Illustration of flow values that contribute the most of the total model error. The error considered here is the square error calculated on flows. The darker the dots, the larger the contribution to the total error. Horizontal dashed lines correspond to the percentiles of the box plot shown in Figure 3.4. 63

Figure 3.6: Schematic representation of the derivation of master box plot which summarizes the behaviour of a given criterion over the entire catchment set (M stands for the total number of catchments) 63

Figure 3.7: Box plot of distribution of efficiency criteria obtained by the GR4J and MORD models over the entire catchment set in validation (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.05 and 0.95 percentiles) 64

Figure 3.8: Master box plot obtained over the entire catchment set for the GR4J and MORD models (each flow percentile of exceedance corresponds to the median of this percentile over all the box plots obtained on each catchment of the entire set) 68

Figure 3.9: Change in the mean value of NSE^*_{iQ} and $NSE^*_{\ln Q}$ obtained by the GR4J model in validation over the catchment set with different values of ε (fractions of Qm from Qm/10 to Qm/100) 70

Figure 3.10: Comparison of λ values obtained by GR4J (left) and MORD (right) models for the two test periods (P1 and P2) 72

Figure 3.11: Comparison of λ values obtained by the GR4J and MORD models 72

Figure 3.12: Scatter plot of Q_{90}/Q_{50} vs. lambda values (λ) for the two test periods for the GR4J model 73

Figure 4.1: Structural layout of the base models 80

Figure 4.2: Layout of two versions of the GR5J model with different additional stores 85

Figure 4.3: Box plots of NSE*iQ values obtained in validation over the catchment set by GR5J and three model versions with modified groundwater exchange functions (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles respectively 86

Figure 4.4: Comparison of the splitting coefficient (SC) values obtained on the two calibration periods (P1 and P2) in the M4 model version 87

Figure 4.5: Box plots of NSE*iQ values obtained in validation by model versions having different formulations of the additional routing store (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 90

Figure 4.6: Illustration of the hydrographs simulated by the GR5J and GR6J models, with corresponding NSE*iQ efficiency values 92

Figure 4.7: Comparison of the parameter values obtained on the two calibration periods P1 and P2 for the GR5J and GR6J models (1:1 line on each graph) 94

Figure 5.1: An ensemble flood forecast with different levels of warnings (Source: Cloke and Pappenberger, 2009) 106

Figure 5.2: Layout of the probabilistic approach of low-flow forecasting (observed and simulated flow hydrographs of catchment J for the year 2003 is illustrated here) 107

Figure 5.3: Illustration of the identification of maximum possible forecast lead-time with respect to the BSS 117

Figure 5.4: Distribution of Brier skill scores (at Q_{75} and Q_{90}) over the full catchment set at different lead times (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles) (in simulation, i.e. without correction) 118

Figure 5.5: Model performance in simulation (i.e. without correction) mode at Q_{75} and Q_{90} 119

Figure 5.6: The influence of error in simulation mode (a) on model's forecasts (b and c). Boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles 120

Figure 5.7: Comparison of model forecasts with and without bias corrections, BC1 & BC2 (the first column represents the results at threshold Q_{75} and the second column represents the results at Q_{90}) 122

Figure 5.8: Forecast skill score with model output error corrections: EC1 and EC2 123

Figure 5.9: The performance of model forecasts with different beta values and different lead times 124

Figure 5.10: Dependency of beta value with forecast lead-time 125

Figure 5.11: Forecast skill score with lead-time using model updating, EC1' 126

Figure 5.12: Illustration of the model's forecast skill score with the state updating ((a) & (b): updating of routing stores; (c) & (d): updating of the production store) 128

Figure 5.13: POD of forecasts with bias correction (EC1') and POD of forecasts without correction 130

Figure 5.14: NSE^*_{iQ} of forecasts with bias corrected (EC1') and NSE^*_{iQ} of forecasts without correction 131

Figure 5.15: Spatial distribution of the maximum possible lead-time based on BSS_{75} for the data set 132

Figure 5.16: Spatial distribution of the maximum possible lead-time (in days) based on BSS_{90} for the data set 133

Figure 5.17: Distribution of the maximum possible lead time for the catchments on the data set 133

Figure 5.18: Possible link of lead-time with catchment geology 134

Figure 6.1: Location of test catchments and dams in the Seine river basin 141

Figure 6.2: Location of catchments and dams in the Loire River basin 141

Figure 6.3: Illustration of how to account for the artificial reservoir volume variations (ΔV) in the GR6J model 144

Figure 6.4: Illustration of model's efficiency values with (R) and without reservoirs 146

Figure 6.5: Illustration of model's efficiency values with and without reservoirs (R) 146

Figure 6.6: Comparison of model performances with and without the influence of reservoirs using the NSE^*_{LnQ} and NSE^*_{iQ} criterion 147

Figure 6.7: Illustration of the observed and simulated hydrographs (with and without R) of one of the test catchments for a period of four years (2000-2003) and its efficiency values 148

Figure 6.8: Illustration of forecast skill score of the model GR6J in the Seine basin 149

Figure 6.9: Illustration of forecast skill score of the model GR6J in the Loire basin 149

General introduction

Introduction

River low-flow is a seasonal phenomenon and an integral component of the flow regime of any river (Smakhtin, 2001). As a primary water supply system, river flows in most countries need more attention, especially during low-flow periods. Low-flow hydrology is a wide area raising many issues regarding spatial and temporal components of the system. Stream low-flows are influenced by several factors such as the distribution and infiltration characteristics of soils, the hydraulic characteristics and the extent of aquifers, the rate, frequency and amount of recharge, the evapotranspiration rates from the basin, the type of vegetation (land cover), topography, climate conditions and anthropogenic activities such as river abstraction and changes in land use patterns. The occurrence of low flows influences agricultural activities, power generation, navigation and other domestic purposes (see Gazelle, 1979; Mignot and Lefèvre, 1996). Low flows may also result in increased sedimentation that changes the morphology of the stream channel. Another impact may be on the chemical or thermal quality of stream water (e.g. Moatar et al., 2009). This might influence the ecosystem species distribution, their abundance, and more generally the ecological status of the river.

The advanced prediction of stream low flows can result in the development of better water management programs and it can mitigate some of the problems which we stated here. But the existing literature on low-flow hydrology highlights the necessity to carry out studies to develop appropriate methodologies for low-flow simulation as well as forecasting. Hopefully these methodologies should be applicable to a diversity of catchments and conditions.

Low flows are different from drought events. Droughts include low-flow periods, but a continuous seasonal low-flow event does not necessarily constitute a drought event (Smakhtin, 2001). Drought is beyond the scope of this study and for more on this issue, please refer to Duband et al. (2004), Wipfler et al. (2009), Duband (2010), Garnier (2010) or Mishra and Singh (2010). For more on drought and drought modelling, especially in France (where we conducted our study), please refer to Prudhomme and Sauquet (2006), Jacob-Rousseau (2010). However, water scarcity will probably become a growing concern in the next years due to the impacts of climate change. This impact will be severely affecting the river basins in France (Boé et al., 2009). This situation also demands the hydrologists to concentrate more on low-flow hydrology.

Therefore, in our study, we will focus on low-flow simulation as well as forecast. Our research focused on a number of issues that deserve more attention, as described below.

Evaluation of model's low-flow simulation efficiency

Rainfall-runoff (R-R) models are standard tools used today for the investigations in quantitative hydrology. To date, a vast number of R-R model structures (a combination of linear and non-linear functions) has been developed in order to mimic catchment hydrologic behaviour. But these models differ in their performances based on the selected criteria and the catchment characteristics. One idea pushed forward in this study is to try to identify a better performing "one-size-fits-all" model that could be a good compromise between accuracy, robustness and complexity. This question will be addressed in the next section.

A corollary issue to this search for an improved model is to find out appropriate ways to evaluate the selected model structure. The literature shows there are various efficiency criteria available to evaluate model structures, especially in peak-flow conditions. But criteria suitable to evaluate model's performance in peak-flow condition may not be suitable to evaluate the same model in low-flow conditions. The availability of several efficiency criteria in the literature can lead the end-users to draw wrong interpretations if they are not aware of the properties of these criteria. In this context, the present study analyses some of the existing criteria and their inter-dependencies to identify a criterion better suitable for the evaluation of low-flow simulated by hydrological models. Thus this study can give better guidance to the modeller who is interested in the evaluation of model's low-flow simulation efficiency.

Why a general model structure for the French river basins?

The emerging demands of long-term plans for the proper management of water resources and the possible impacts of climate change on river flows incite hydrologists to discuss more on the issues of low-flow simulation and advanced predictions. There are various hydrometeorological processes occurring in a catchment, from the formation of rainfall to the streamflow that finally leaves the catchment through a river, and their modelling is quite complex. This complexity changes from one catchment to another depending on the catchment and climatic conditions. Therefore, a model structure appropriate for a given catchment characteristics and data set may not be the best for another catchment.

However, using a specific model structure for each study catchment raises practical issues (operational use on many catchments, regionalization, etc.). Besides, it remains difficult to identify dominant processes on a catchment, and underground processes that play a key role in low-flow generation are still difficult to observe. This highlights the need to develop more generally applicable model structures that could adapt (through calibration) to various types of catchments and climatic conditions. By keeping this in mind, in this study, we tried to propose a general model structure for low-flow simulation and prediction, by implementing extensive tests of various structures on a large set of catchments distributed over France. At the same time, we kept in mind that the proposed continuous model should remain coherent over the whole hydrological cycle, i.e. not losing its efficiency in high-flow simulation. The main principle behind this study was a structural sensitivity analysis of some widely used models.

Long-term forecasting of low flows and possible improvements

As in simulation, a general approach will be convenient to issue forecast over a large set of catchments. The long-term forecasting of low flows can assist water managers to make appropriate decisions in advance. The main uncertainty associated with long-term forecasting comes from the characteristics of future meteorological conditions. Instead of the deterministic approach, an ensemble approach can provide probabilistic information on the likelihood of future low-flow events. In our study, we will address the suitability of an ensemble approach in low-flow forecasting and evaluate its efficiency with increasing lead-time. We will also focus on some of the possible options to improve the forecast quality by using observed flow information.

Exploratory tests to make low-flow forecasts on catchments influenced by dams

Man-made reservoirs have a major role in changing the streamflow hydrology in the downstream reaches. The integration of information on reservoirs into the R-R model is relevant for proper management of downstream water resources, especially in cases of low-flow augmentation. This indicates the necessity for accounting for reservoirs in hydrological models while modelling streamflow in catchments influenced by dams. Based on the existing

literature, here we will try to improve the model forecasts with the integration of storage information on influenced catchments.

Organisation of thesis

The thesis is organised into three main parts and a general conclusion.

Part I presents the general background and framework of this work. It is subdivided into two chapters. Chapter 1 is an introductory chapter to stream low flows. The main factors influencing the occurrence of stream low flows are detailed with reference to the available literature. Chapter 2 briefly describes the data set used for the entire analysis and the classical low-flow indices used to characterize the flow regime of the data set. Here we present the spatial variation of low flows in the study data set. This chapter also discusses the links between catchment descriptors such as aridity index and catchment geology with low-flow indices, which may be useful to characterize flows in ungauged catchments.

Part II focuses on low flow simulation and includes chapter 3 and chapter 4. Chapter 3 questions the existing criteria used to evaluate low-flow simulation efficiency and proposes another criterion better adapted to low flows. Chapter 4 presents the development of a hydrological model version showing improved low-flow simulation capacity. This is based on an extensive test of alternative model structures starting from existing models.

Part III analyzes the performance of models for forecasting purposes. Chapter 5 investigates how a model structure developed for simulation can be adapted for low-flow forecasting purposes, within an ensemble framework, and links to works on flood forecasting are analyzed. Chapter 6 presents the exploratory tests of the forecasting methodology on catchments influenced by artificial reservoirs

The conclusion chapter summarizes the major outcomes of the thesis and the perspectives for future research.

Part I: Stream low flows and their characterization

This part is divided into two main chapters. Chapter 1 gives the basics of stream low flows in the French context. The second chapter gives more on the characteristics of stream low flows with respect to the study catchments distributed over France.

Chapter 1. Low flows and low-flow generating factors in the French context

1.1.Introduction

River flow derives from three main components: overland flow, throughflow from soils and groundwater discharge (see **Figure 1.1**). The overland flow and through flow respond quickly to rainfall or melting snow, whereas groundwater discharge responds slowly with a time lag of several days, months or years. Catchments dominated by overland flow and through flow are called flashy catchments, and catchments fed primarily by groundwater discharge are slowly responding catchments with a high base-flow. **Figure 1.1** illustrates the hydrological processes and catchment storages which maintain flows in streams throughout the year. Here the river catchment is considered as a series of interconnected reservoirs with three main components: recharge, storage and discharge. Recharge to the whole system is largely dependent on precipitation, whereas storage and discharge are complex functions of catchment physical characteristics. Therefore, these two factors highly influence the flow conditions of a stream.

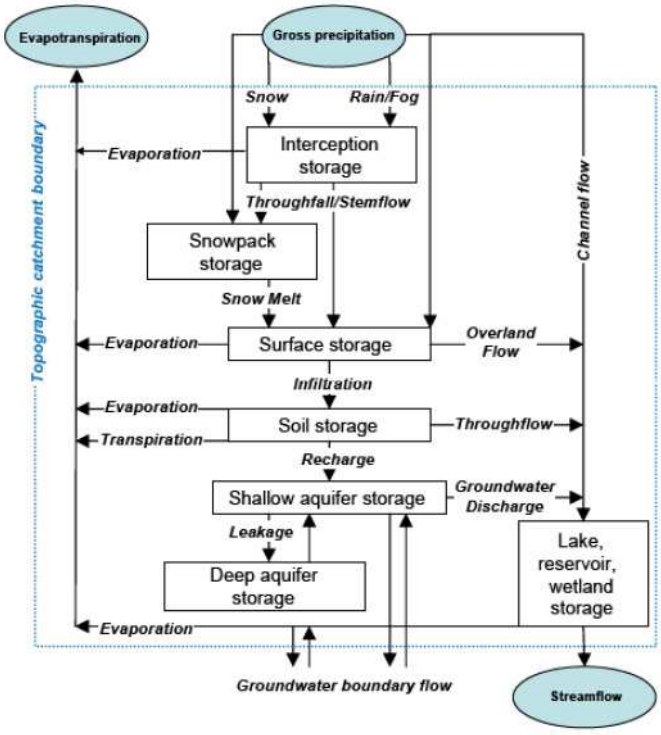


Figure 1.1: Natural hydrological processes and catchment storages (WMO, 2008)

Based on the catchment storage and discharge properties, three flow conditions can be schematically distinguished in streams: low, medium and high-flow conditions. Here our interest is on stream low flows. There are several definitions of low flows available in the literature. In general, stream low-flows indicate the flow of water in a stream during prolonged dry weather and it is a seasonal phenomenon (Smakhtin, 2001).

Figure 1.2 shows an example of the occurrence of low flows in a French catchment in Brittany. The three-year streamflow hydrograph shows that the low-flow occurs in the same season in each year. This information can assist the water managers to develop appropriate plans during this period in advance. In France, many water uses (e.g. agriculture, industry, navigation) partly or significantly rely on rivers, with consequences on their ecological state. Hence the low-flow information is very much needed in the French river basins.

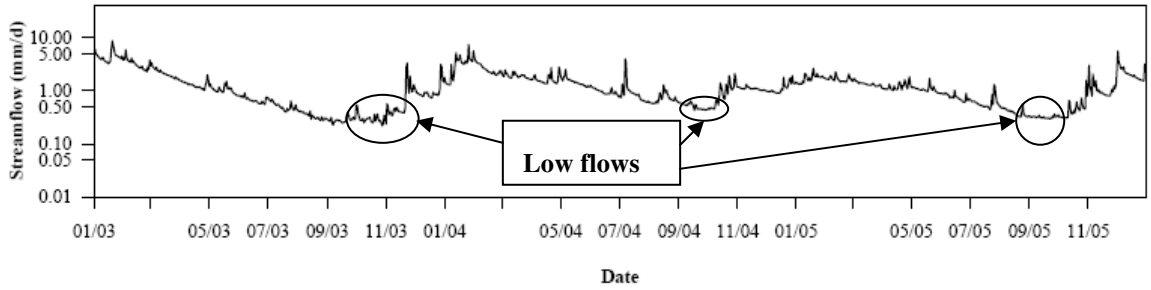


Figure 1.2: Illustration of stream low flows of one of the study catchments (Catchment L'Aber Wrac'h at Le Drennec) for a period of three years (2003-2005)

Several factors (such as land cover or climate) can influence the occurrence of stream low flows. Two main situations generate low-flow events:

- an extended dry period leading to a climatic water deficit when potential evaporation exceeds precipitation (known as summer low flows);
- an extended period of low temperature during which precipitation is stored as snow (known as winter low flows).

Figure 1.3 illustrates summer and winter stream low flows in the Garonne River downstream Toulouse (August 1998) and Columbia River, Golden, Canada (February 2007), respectively.



Figure 1.3: Left: Summer low flows of the Garonne River downstream Toulouse (Source: A. Dutartre, Irstea) and right: winter low flows of the Columbia River, Canada (Source: WMO, 2008)

Low flows are generally fed by groundwater discharge or surface discharge or melting glaciers. In this work, we will concentrate on summer low-flows on which are the main pressure for various uses in France. For a sustainable low-flow, the aquifers must recharge seasonally with adequate amount of moisture and the water table should be shallow enough to be intersected by the stream to sustain the streamflow throughout the year, a slowly responding catchment with a high contribution of base-flow sustains streamflow throughout the year. The magnitude and variability of low flows depends on several natural factors as well as anthropogenic influences (construction of artificial reservoirs, river abstractions for agricultural and industrial operations, aquifer abstractions that consequently modify the surface water – groundwater relationship, water transfers between catchments, urbanization). The natural factors that can have an influence on low flows are the catchment characteristics (infiltration characteristics of soil, aquifer characteristics, topography of the catchment, land use, evapotranspiration) and climate conditions. This chapter summarizes the influence of some of the major factors on the spatial variability of low flows between catchments based on the French context.

1.2.Natural influences on low flows: catchment characteristics

The river basin or catchment is a geographical integrative unit of the hydrological cycle (see **Figure 1.1**). Catchment descriptors that can influence river low flows include catchment area, slope, percentage of lakes, land cover, soil and geologic characteristics, and mean catchment

elevation. The following section gives a brief description of catchment geology, one of the catchment descriptors which has a major role on stream low flows.

1.2.1. Catchment geology and river low flows

Catchment geology is one of the dominant factors controlling the flow regime of a river. Here geology is understood as the first geological layer(s) having a direct impact on surface water. It influences the storage and discharge properties of a catchment. For example, precipitation on an impervious basin of barren rock results in sudden runoff and frequent flood events. But on a basin under similar climatic conditions but underlain by thick permeable material, the infiltration contributes to the groundwater reservoir. Literature analysis shows a direct relationship between catchment geology and discharge rate during low-flow periods, i.e., the low-flow statistics are highly dependent on hydrogeology (Gustard et al., 1992). Ackroyd et al. (1967) clearly explained the influence of catchment characteristics on the groundwater contribution to streamflows. Streams with geology of gravel deposits permit a sustainable flow during dry periods, but streams with glacial deposits show significant flow variations (Schneifer, 1957). Streams with geology of different types of unconsolidated rocks have low yields during low-flow period and streams with geology of metamorphic sedimentary rocks and igneous rocks show high-flow values relative to their catchment size (Smakhtin, 2001).

Figure 1.4 shows the geology of France. The geological information on the data set which we used throughout the thesis was provided by BRGM, the Geological Survey of France. The analysis of individual catchment geology is complex since there are numerous geological compositions within each catchment's geology. An overview of the geological map shows that the major geology of French basins includes crystalline igneous rocks, marls, massive limestone, detrital crystalline rocks, metamorphic rocks, chalk, quaternary volcanic rocks, non-carbonates and flysch sediments. Some of the major geology (e.g. detrital crystalline, marnes, massive limestone, clays and sands) can favour low-flow conditions in the corresponding catchments. Thus the information on geology is highly useful to the flow characterisation in ungauged catchments (Clausen and Rasmussen, 1993; Tague and Grant, 2004) (see next chapter).

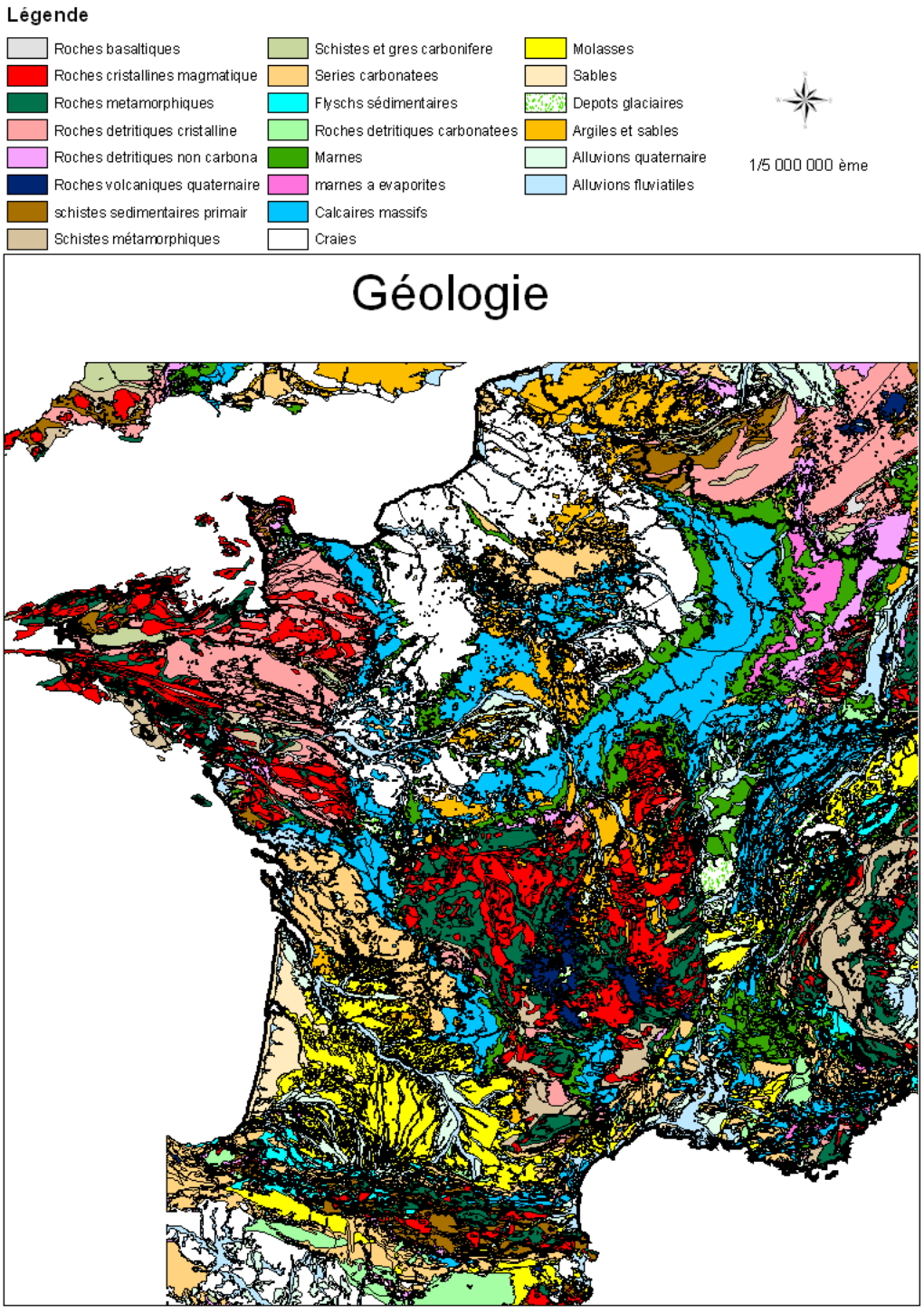


Figure 1.4 Geological map of France (Source: Geological Survey of France)

1.3. Anthropogenic impacts on catchment characteristics and low flows

Anthropogenic activities can cause severe changes in the river hydrology and morphology. There are direct as well as indirect influences of human activities on catchment characteristics and hence on the river low flows.

1.3.1. Direct influences

The direct impacts of human activities on low flows include river water abstraction for industrial as well as agricultural purposes, waste load allocation into rivers, and sand mining operations, construction of reservoirs at the upstream of the river etc. Among these influences, the construction of reservoirs bears a major role on stream low flows which we further discuss in chapter 6. Water abstraction for agricultural purposes (mainly for irrigation) decreases streamflow discharge, especially during dry periods (Eheart and Tornil, 1999; Ngigi et al., 2008). Waste load allocation leads to the pollution of river ecosystem and also increases the river deposits (Despriée et al., 2011; Lemarchand et al., 2011).

Figure 1.5 gives statistics on water use by different sectors in France. Most abstractions originate from rivers. According to the Commissariat général au développement durable (2012), 33.4 billion m^3 of water were withdrawn in France to meet the needs of drinking water, industry, irrigation and electricity production in 2009 (note that the actual water consumption may differ significantly from water withdrawal). These abstractions are likely to increase in the future due to the population growth, climate change and other human activities. Hence to sustain water sources for these mentioned uses especially during dry periods, we need the information of stream low flows in advance. This is also one of the arguments supporting the present study.

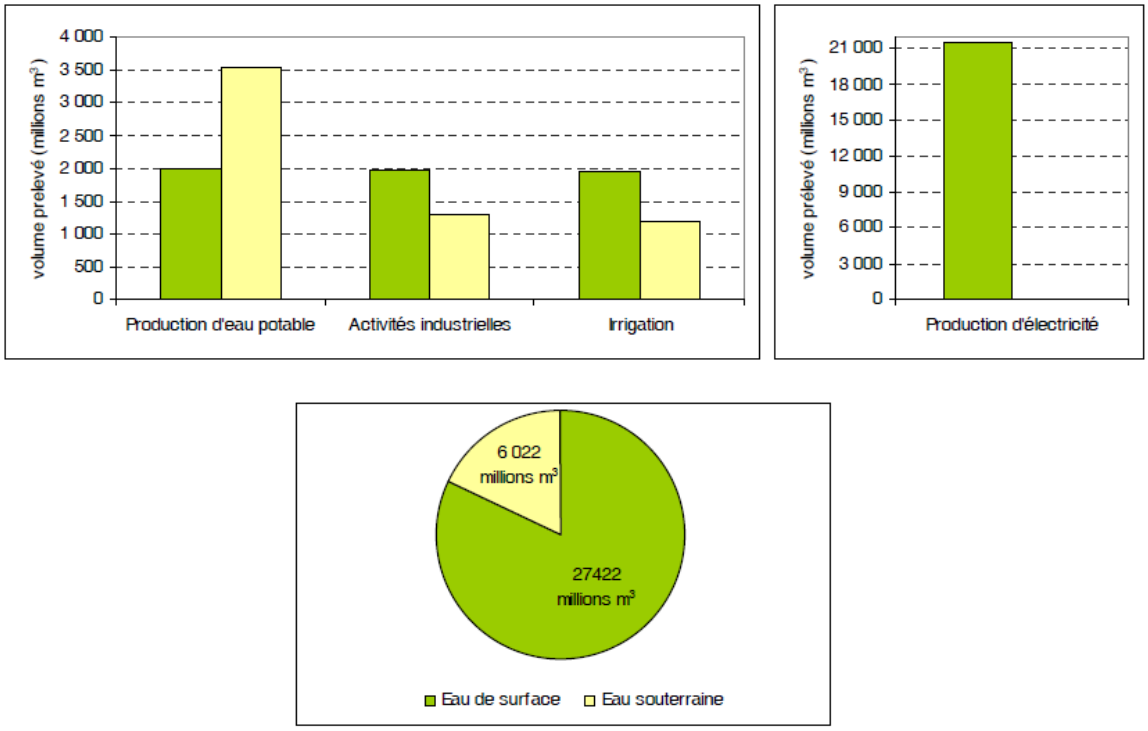


Figure 1.5 Total water availability and its utility by different sectors in France in 2009 (Source: Commissariat général au développement durable, 2012)

1.3.2. Indirect influences

The indirect influences include groundwater abstraction (tube wells), changes in land-use pattern (cropping pattern), afforestation and deforestation of catchments, etc. **Figure 1.5** shows the total groundwater abstractions in France in 2009. These abstractions contribute to the aquifer depletion during the dry periods, which consequently may limit the amount of water available to feed stream low flows.

Afforestation can have a major role on streamflow variability. The forest area in France is 16.1 million ha (Institut National de l'information Géographique et Forestière, IGN), with an increasing trend over the past decades. The forest cover may result in an increase in evapotranspiration, a decrease in groundwater recharge due to the water uptake by the trees during the dry periods. Hence the forestry may reduce the river flows and causes more frequent occurrence of low flows (see Johnson, 1998; Robinson and Cosandey, 2002; Hundecha and Bardossy, 2004; Lane et al., 2005).

However, it is difficult to identify clear links between land cover and flow characteristics (Andréassian et al., 2003). One reason is that the impact of forest cover differs between catchments with specific climatic and pedological contexts (Andréassian, 2004). Similar to afforestation, agricultural land-use changes or urbanization can also have a significant role on stream low flows (see Sharda et al., 1998; Sikka et al., 2003; Oudin et al., 2008; Wang et al., 2008).

1.4.Potential impacts of climate change on future river flows

Climate change may have severe impact on frequency, magnitude, location and duration of hydrological extremes (Hoog, 1995; Cunderlik and Simonovic, 2005; Diaz-Nieto and Wilby, 2005). The decrease in precipitation, increase in temperature and evapotranspiration may cause a decrease in the average discharge during low-flow periods (de Wit et al., 2007; Johnson et at., 2009; Woo et al., 2009). This impact could be severe in Europe (Hisdal et al., 2001; Drogue et al., 2004; Blenkinsop and Fowler, 2007; Calanca, 2007; Feyen and Dankers, 2009), e.g. with more frequent occurrence of low-flow events in the French River basins (Boé et al., 2009) in this century. See Renard et al. (2006), Moatar et al. (2010) for more on the impact of climate change on hydrological extremes in France. In a study Bourgin (2009) also assessed the impact of climate change on water sources in the Loire basin, France.

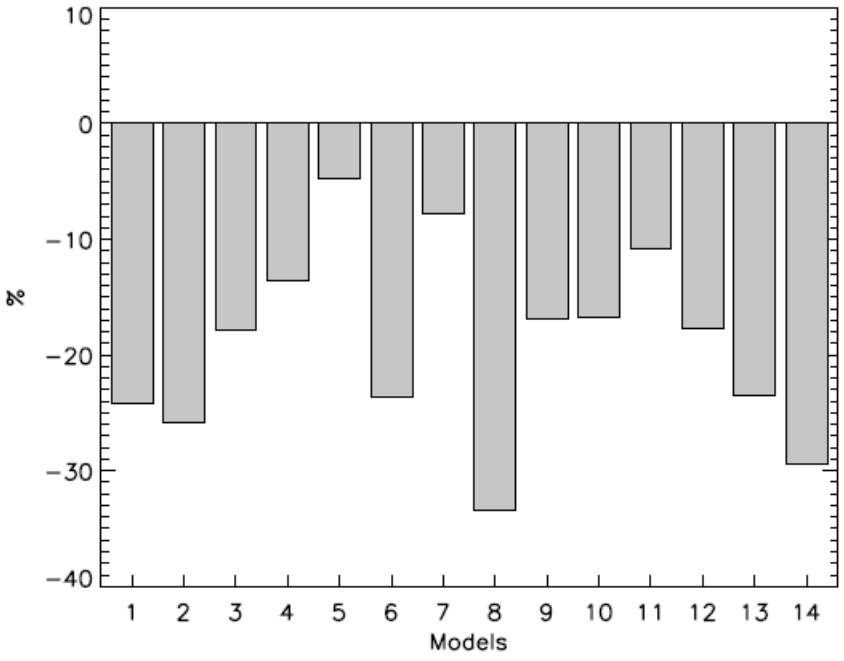


Figure 1.6: Changes in the annual discharges (using SIM model) between the 2046-2065 and 1971-2000 periods averaged over France using the inputs from 14 climate models (Source: Boé et al., 2009)

Figure 1.6 shows the relative annual changes in river discharges averaged over France for different climate models between the periods 2046-2065 and 1971-2000. The changes vary from -5% to -33% between models. This result shows that water scarcity will probably become a growing concern in the next years and it strengthens the necessity to develop a suitable methodology for low-flow prediction to better anticipate these periods.

1.5. Impact of low flows on society

Stream low flows have considerable impact on society due to their significant role in various water management operations (Stromberg et al., 2007). Frequent occurrences of low flows mainly influence domestic purposes such as drinking, agriculture, power generation (see Manoha et al., 2008) and navigation. Stream low flows can also exacerbate the effects of water pollution. Long-term river low flows are synonymous of drought event and they may have more severe consequences than flood events. For example, the cost of damages in the years 1988-1989 in the US due to drought event was approximately US$40 billion and US$18-20 billion during 1993 due to flood event (Demuth, 2005). The extended low flows may also cause changes in the stream ecosystem which influence the abundance and distribution of fish, algae etc., with possible socio-economic impacts (Hamlet et al., 2002).

1.6. Conclusions

In France, the domestic, industrial and agricultural activities are mainly fulfilled using river waters and hence a proper management of these water sources is very important, especially during low-flow periods. This chapter highlighted some of the major factors that influence the occurrence of stream low flows. The different water uses such as domestic water supply, agricultural and industrial uses etc., and the potential decrease in the annual water availability in France due to the negative impact of future climate change on river water sources in the near future are urging us to provide more information in advance on the occurrence and other flow characteristics of rivers during low-flow periods in the French river basins.

The next chapter presents low-flow characteristics in more details and discusses their significance in water management.

Chapter 2. Data set and low-flow characterisation

2.1. Introduction

Smakhtin (2001) indicates that various flow statistics are available to characterise low-flow conditions depending on the purpose and data availability. Low flows can be characterised by their magnitude, frequency, duration, timing and geographic extent. These derived characteristics can be called low-flow indices. A low-flow index refers to the specific values derived from an analysis of low flows WMO (2008). Estimating these indices of river low flows is necessary for the management of surface waters (Dakova et al., 2000; Mijuskovic-Svetinovic and Maricic, 2008). For example, Q_{95} (the flow exceeded 95% of the time) is used in the UK for licensing of surface water extraction (Higgs and Petts, 1988).

There are several studies conducted on characterising stream low flows (for more details please refer to Smakhtin and Toulouse, 1998; Rifai et al., 2000; Reilly and Kroll, 2003; Laaha and Blöschl, 2007; Chopart and Sauquet, 2008; Whitfield, 2008). Each low-flow statistic has its own significance in low-flow hydrology and it depends on the purpose. Because of the diverse applications of these flow statistics, the present study mainly focuses on the spatial variability of low flows in the data set of French catchments we will use. Some of the low-flow indices we discuss in the coming sections are used as threshold to define low flows in our study catchments and to better analyse results.

The following section presents the data set and the different flow statistics chosen to characterise the stream low flows.

2.2. Description of the catchment set

2.2.1. Catchment selection

It is important here to remember the words of Linsley (1982): "Because almost any model with sufficient free parameters can yield good results when applied to a short sample from a single catchment, effective testing requires that models be tried on many catchments of widely differing characteristics, and that each trial cover a period of many years". This suggests that the test of a model on a large set of catchments is a necessary condition to derive appropriate conclusions (see Andréassian et al., 2006a). By keeping these points in mind, a set of 1000 catchments spread over France (**Figure 2.1**) was used for the present study. This

database was built by Le Moine (2008). These catchments represent a wide range of hydro-meteorological conditions, including:

- temperate (characterised by moderate weather events),
- oceanic humid (characterised by plentiful precipitation year around),
- Mediterranean (characterised by a succession of drought and high intensity rainfall periods and by the high spatio-temporal variability of precipitation within and between years)
- and continental (characterised by annual variations in temperature due to the lack of precipitation or water bodies) climate conditions.

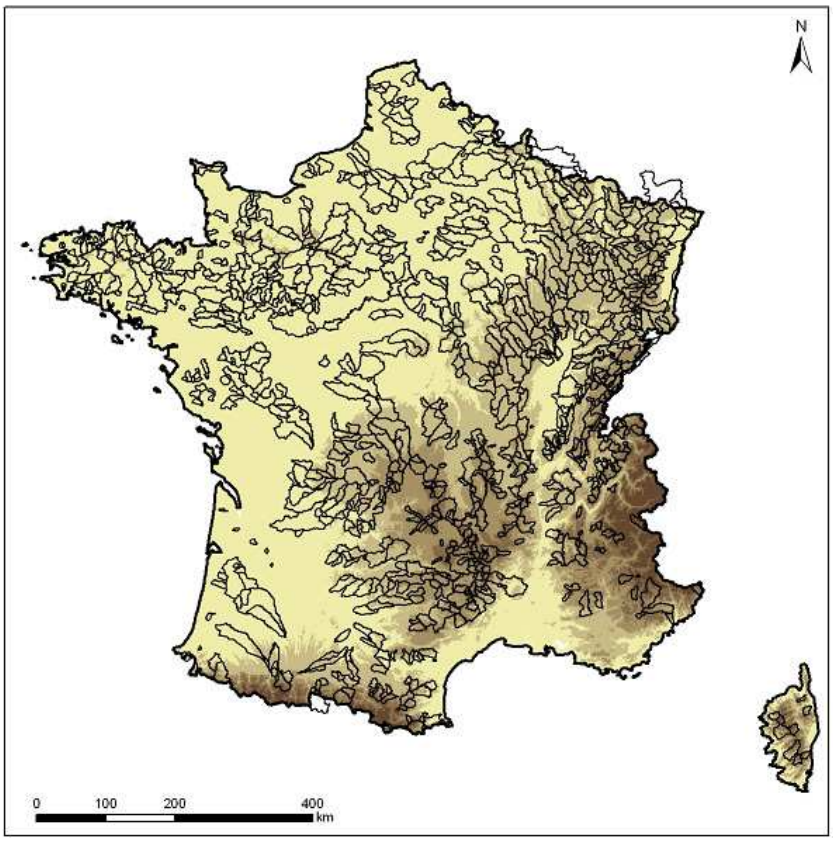


Figure 2.1: Location of 1000 catchments of the data set (the highlighted background indicates the presence of mountains)

Catchments are mostly small to medium-size catchments (median size of 162 km², and 5th and 95th percentiles equal to 28 and 2077 km² respectively). The lack of larger catchments mainly comes from the fact that the catchments selection had been limited to the [10; 10000] km²

range and to catchments without strong artificial influences (here it means without major dams, the level of abstractions could not be checked given the available data).

2.2.1. Data sets

Due to the hydro-meteorological variability and other sources of variability that occur over short time scales, low-flow characteristics estimated from a few years of streamflow data deviate from the long-term average. Because of this, it is usually recommended to use streamflow records of 20 years or more for low-flow estimation (Tallaksen and van Lanen, 2003). Hence, our test catchments are selected in a way that there is at least 20 years of daily data available for the analysis.

Continuous series of precipitation (P), potential evapotranspiration (PE) and streamflow (Q) were available for the 1970–2006 time period, providing good variability of meteorological conditions, with quite severe drought periods (e.g. the years 1976, 1989–1991, 2003 and 2005). Meteorological data come from the SAFRAN reanalysis of Météo-France (Quintana-Segui et al., 2008; Vidal et al., 2010). Daily potential evapotranspiration was estimated using the formulation proposed by Oudin et al. (2005a) based on temperature and extra-terrestrial radiation. Streamflow (Q) data were extracted from the national HYDRO database. Given the size of our data set, we did not perform visual inspection of the quality of flow data retrieved from the HYDRO database. We acknowledge that the quality of flow data for some of these basins may be questioned, but given the size of our catchment set and past experience at Irstea, this is likely to have a limited impact on our overall results.

Table 2.1 summarises the variability of rainfall, streamflow, potential evapotranspiration, and catchment size in the data set.

Table 2.1 Variability in size and the main hydroclimatic conditions over the catchment set

Catchment characteristics	**5th quantile**	**25th quantile**	**50th quantile**	**75th quantile**	**95th quantile**
Rainfall (mm/year)	714	833	967	1155	1636
Streamflow (mm/year)	141	246	366	564	1193
Potential evapotranspiration (mm/year)	532	615	646	688	755
Area (Km^2)	28	77	162	391	2077

The variability of mean streamflow values can be expressed as a function of mean P and mean PE. **Figure 2.2** plots the runoff coefficient (Q/P) as a function of the aridity index (P/PE) (see Le Moine et al., 2007). It illustrates the variability of hydro-climatic conditions in the test catchments. As explained in details by Le Moine et al. (2007), there are many catchments in this data set for which water losses are greater than PE (points lying below the line y=1−1/x), which may be an indication of leaky catchments. There are also a few catchments for which mean flow is greater than mean rainfall (points above the line y=1), which mainly correspond to catchments with karstic influences, i.e. those fed by inter-catchment groundwater flows from surrounding areas. Even though these catchments may prove more difficult to model, they were not discarded from the data set, as advocated by Andréassian et al. (2010).

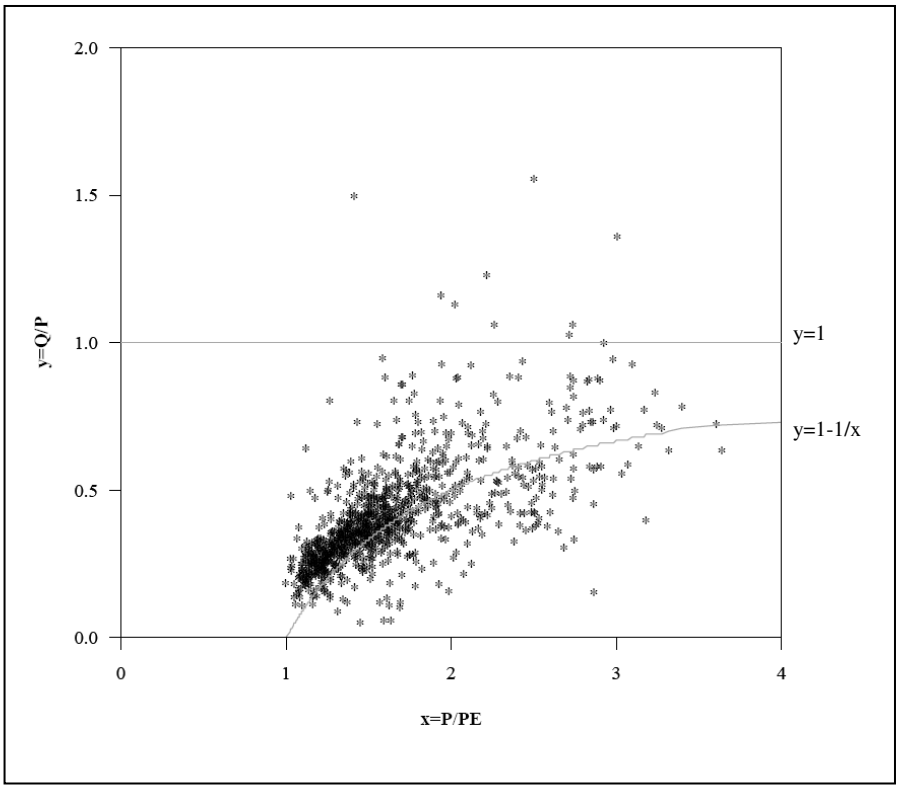


Figure 2.2: Mean Q/P vs P/PE values for the catchments in the data set

2.2.2. Sample catchments

It will not be possible to display the detailed results for all catchments. Hence we selected four sample catchments (listed in **Table 2.2**) representing various conditions. The selection is based on the analysis of some of its flow characteristics such as base-flow index, Q_{90} (flow equalled or exceeded 90% of the time) and climatic characteristics such as the aridity index.

Table 2.2: Characteristics of sample catchments

Code in the study	**A**	**B**	**J**	**K**
HYDRO Code	A7010610	B2220010	J3205710	K6402510
River	Moselle River	Meuse River	Aber Wrac'h River	La Sauldre
Gauging station	Custines	Saint Michel	Drennec	Salbris
Area (Km^2)	6830	2540	24	1200
Mean flow (mm/y)	530	378	593	232
Mean rainfall (mm/y)	1109	954	1087	825
Mean potential evapotranspiration (mm/y)	614	619	643	684
Major geology	Marnes	Massive limestone	Crystalline igneous rocks	Chalks

Table 2.2 gives characteristics of the sample catchments. The selected catchments differ in hydro-climatic as well as geologic conditions. **Figure 2.3** shows the location and the flow hydrographs of these catchments for a period of three years (2003-2005), in which 2003 and 2005 correspond to two strong drought events in France. To improve the readability of low flows, discharge is plotted on a logarithmic scale.

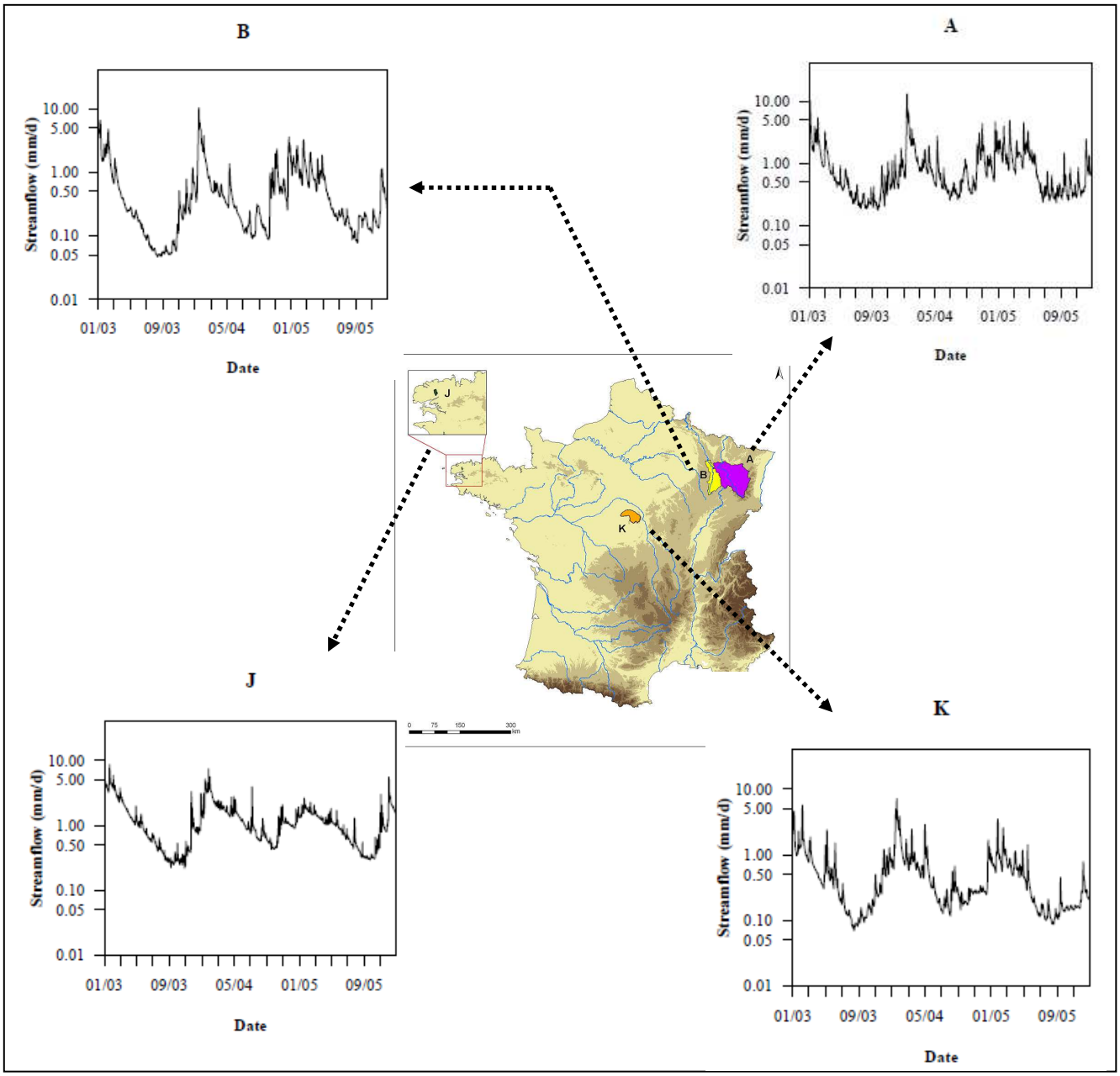


Figure 2.3: Illustration of location of sample catchments and their daily flow hydrographs from 2003 to 2005

2.3.Low-flow measures

The next sections briefly describe some of the low-flow measures used in this study to characterise low flows in the data set. The values taken by these indices on our catchment set will be discussed in section 2.4.

2.3.1. Low-flow period (low-flow spells)

Low-flow period indicates the period for which stream flows are below a certain threshold. The threshold value depends on the purpose. In France, periods of low flows are critical for managing water resources as the society depends on the availability of water through regulated and unregulated river systems. In regulated river systems, reservoirs make a balance between high and low flows. Hence the knowledge of low-flow period is required by engineers and water managers for their design and also for allocations.

2.3.2. Flow duration curve (FDC)

Flow duration curve is an informative way to display the complete range of discharge, from low flows to flood events. It shows the percentage of time a given flow value is exceeded (Iacobellis, 2008). For more information about FDC, please refer to Vogel and Fennessey (1994). The FDC of sample catchments will be discussed in the later section.

2.3.3. Mean annual minimum flow for different durations

The mean annual minima of flow values can be derived from a daily flow series by selecting the lowest flow every year and calculating the mean. Minima of different durations can be determined, for flows commonly averaged over 1, 3, 7, 10, 30 and 90 days. These low-flow indicators are quite widely used in the water resources evaluation, as discussed in section 2.3.1. Among these, mean annual minimum for a consecutive seven-day period (MAM_7) is one of the most commonly used low-flow indices in many countries. It can eliminate the day-to-day variations in the river flow. Smakhtin and Toulouse (1998) also supported the applicability of MAM_7 in the UK as abstraction licensing index. $QMNA_5$ is another commonly used low-flow index in France as a policy threshold. It is the minimum annual monthly flow with a return period of five years. In the present study, we consider MAM_7 and $QMNA_5$ to characterise low flows on the study catchments. These minimum values are used to regulate reservoir operations, irrigation or industrial abstractions, and waste load allocation into streams.

2.3.4. Base-flow index (BFI)

Streamflow hydrograph represents the discharge of the stream as a function of time and it is the combination of different water sources which contributes to the streamflow. The main components which contribute to streamflow are quick flow and base flow. Quick flow is the

direct response to a rainfall event including overland flow (runoff), lateral movement in the soil profile (interflow) and direct rainfall on the stream surface (direct precipitation). Base flow is the delayed discharge from natural storages (groundwater) and it depends on the aquifer characteristics. The groundwater contribution to streams is a critical issue while considering different water management operations. One widely used index to specify the contribution of groundwater to streams is the base flow index (BFI). It is the proportion of base flow in the total runoff. BFI can be used as a descriptor for hydrological modelling, as a tool for selecting analogue catchments, and for estimating annual and long-term groundwater recharge. The literature analysis shows that different methodologies are available to derive the BFI from the streamflow hydrograph (e.g. Nathan and McMahon, 1990; Szilagyi and Parlange, 1998; Furey and Gupta, 2000; Schwartz, 2007; Longobardi and Villani, 2008; Santhi et al., 2008). The present study used the smoothed minima technique (Nathan and McMahon, 1990) to derive the BFI values, because it is easy to implement.

2.4. Computations of low-flow indices on the study catchment set

This section presents a brief description of the computation of the selected low-flow characteristics for the data set.

We used the mean monthly minimum streamflow values to characterise the period of occurrence of low flows (**Figure 2.4**). In most of the catchments, low flows occur during August and September. For a few catchments, low flows occur during November, indicating late flow recessions. In some catchments, low flows occur in February which is typically due to the snow accumulation during winter season.

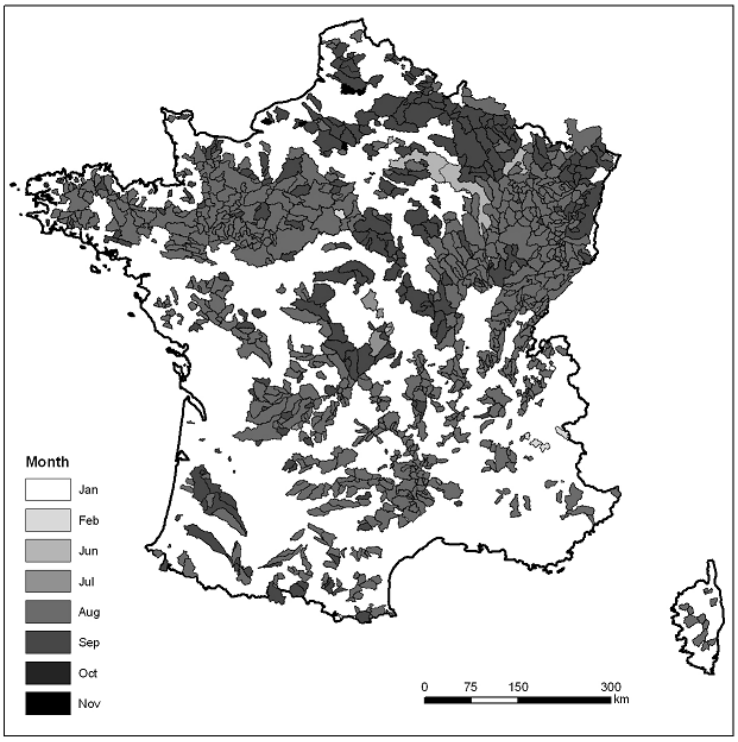


Figure 2.4: Illustration of the period of occurrence of low-flows in the test data set

The flow duration curve of the sample catchments are presented in **Figure 2.5**. The flow values (Q) are plotted in a logarithmic scale. The shape of the FDC is quite different between the sample catchments.

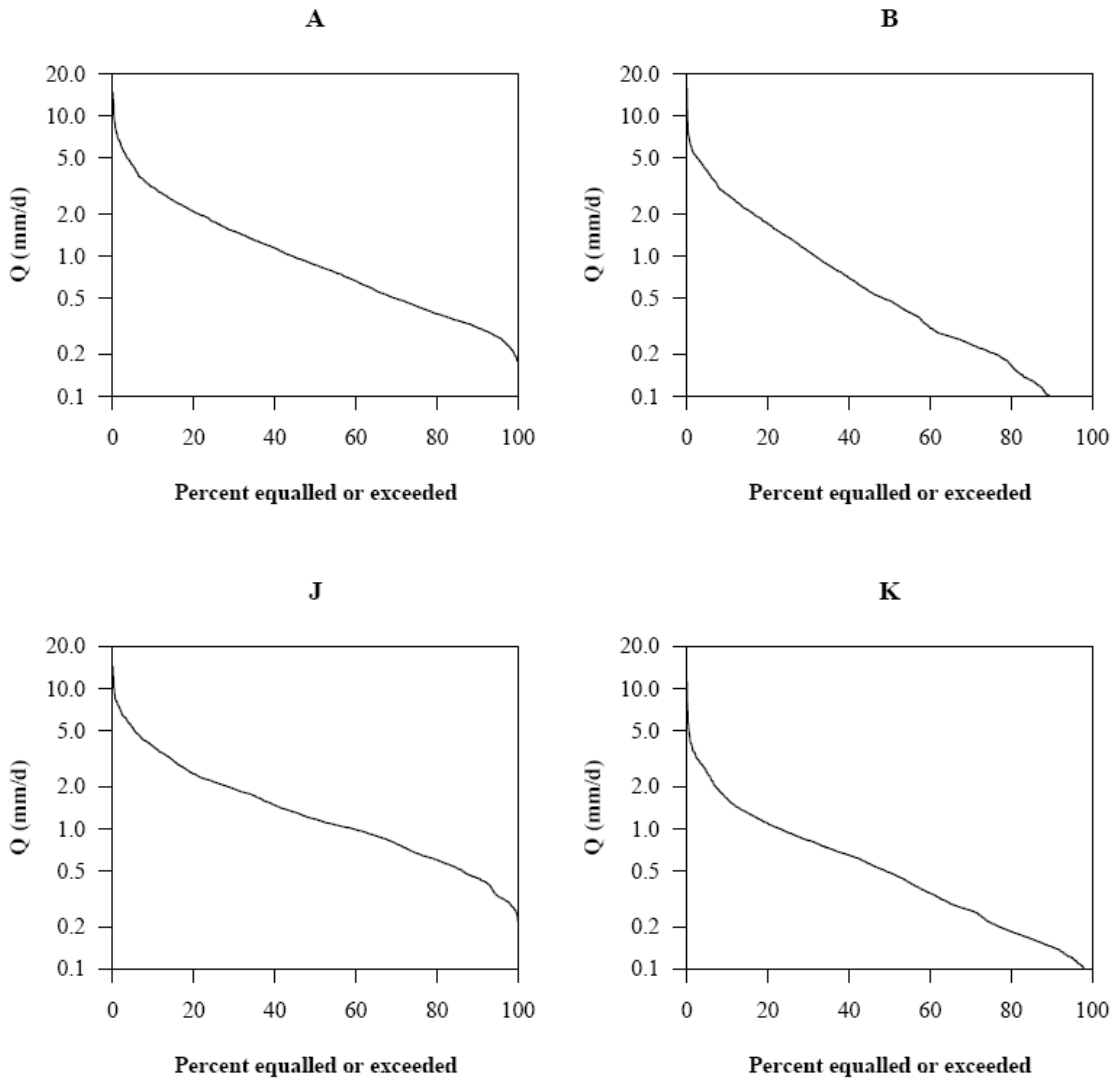


Figure 2.5: Flow duration curve of the sample catchments for a period of 2003 to 2005 (the values below 0.1 mm/d are not shown)

The slope of the FDC gives information about the base-flow contribution: curves with flat slope represent more or less constant contribution of base-flow and curves with steep slope represent a little contribution of groundwater into the stream and hence the low flows are highly variable. In **Figure 2.5**, the low slope of the low-flow part of the FDC shows more or less constant low streamflow in catchments A and J. But in catchment B and K, the FDC shows periods of very low flows, as the river flow is lower than 0.1 mm/day for 15% and 5 % of days respectively. This indicates the occurrence of severe low flows in catchments B and K during the dry periods. Hence FDC is useful to identify the flow conditions of a catchment.

Other significant information can be extracted from the FDC using the Q_{90} and Q_{50} quantiles (Smakhtin, 2001), which are two important flow measures. Q_{50} represents the median flow, which is the flow equalled or exceeded 50% of the time. River flows which are below Q_{50}

represent low-flow conditions and flows which are above Q_{50} represent high-flow conditions. Q_{90} is an indicator of severe low-flow conditions (Smakhtin, 2001). The dynamics of groundwater contribution to streamflow can be characterized by the index, Q_{90}/Q_{50} and is one of the generally used indices for low-flow characterisation.

Figure 2.6 indicates the spatial distribution of MAM_7, $QMNA_5$, Q_{90}/Q_{50} and BFI values. A colour change from white to black represents a gradual increment in the values of corresponding variable in the map. As discussed in section 2.3.3, $QMNA_5$ and MAM_7 show the severity of low-flow conditions of a catchment. A white colour in the maps of $QMNA_5$ and MAM_7 shows severe low-flow condition for that catchment. Compared to these indices, BFI and Q_{90}/Q_{50} show the strength of base flow contribution during the river low-flow period. The colour change from light to dark colour in the BFI map indicates an increase in the base-flow contribution in the North eastern part of France compared to the other regions. This variation may be linked to the hydro-climatic characteristics and hydrogeology. A similar behaviour is visible in the map of Q_{90}/Q_{50}. Similar to the behaviour of these two indices, the spatial distribution (the colour in the map) is similar for both $QMNA_5$ and MAM_7. The analyses of the spatial distribution of these variables show that the MAM_7 and $QMNA_5$ on the one hand and BFI and Q_{90}/Q_{50} on the other hand can be grouped together due to their similar behaviour. This will be discussed in the next section.

The variation of flow conditions from one catchment to another is clearly visible in **Figure 2.6**. Catchment geology influences the spatial variation in stream low flows. The BFI map can be related to catchment characteristics, particularly geology (Gustard et al., 1992). The details on this issue are presented in the later section.

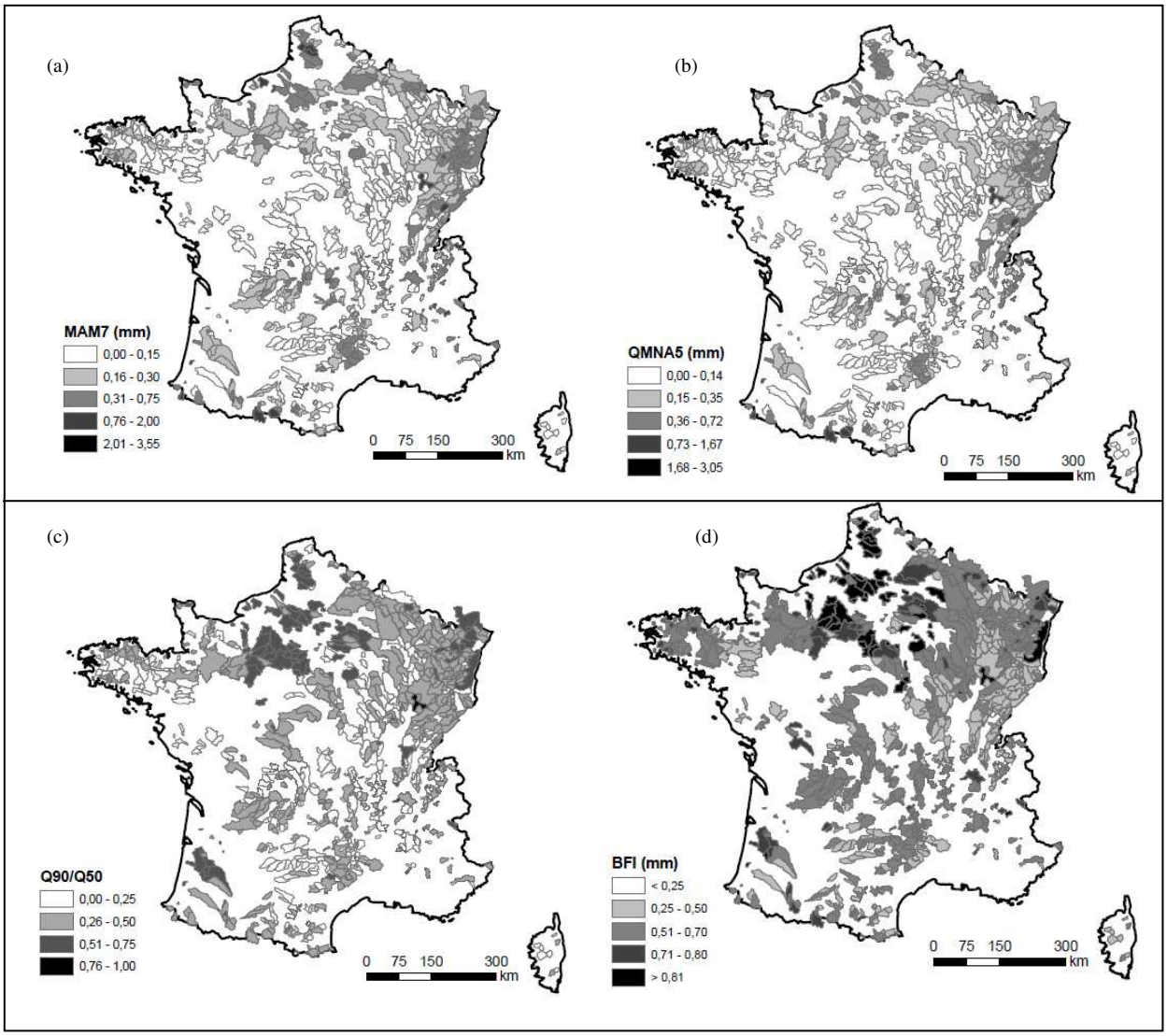


Figure 2.6: Illustration of spatial variability of the low-flow indices: (a) - MAM7; (b) - QMNA5; (c) - Q90/Q50; (d) - BFI

2.5.Link between low-flow indices

This section further investigates the possible links between these indices which can avoid the duplication in the use of indices that bring similar information (for a more complete discussion see, Giuntoli and Renard, 2009). In **Figure 2.7**, MAM_7 and Q_{90} are interrelated with an R^2 value of 0.97 which shows that these two indices bring similar information about low-flow conditions.

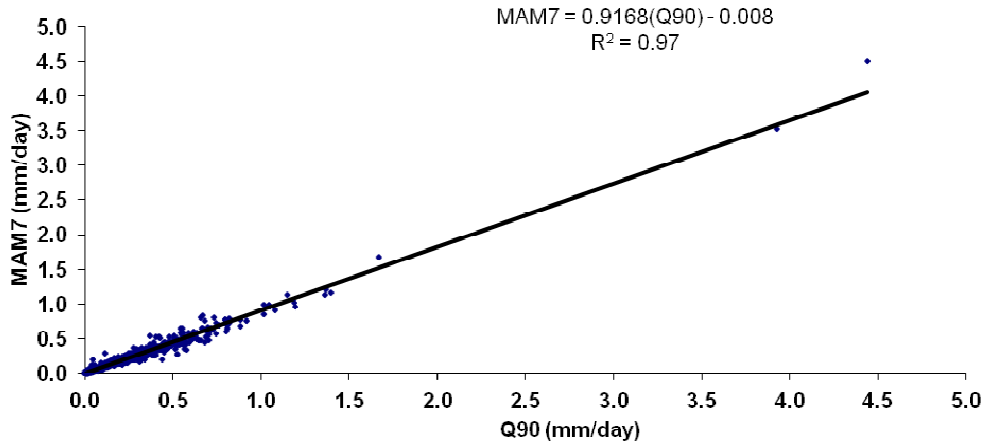


Figure 2.7: Relation between Q_{90} and MAM_7

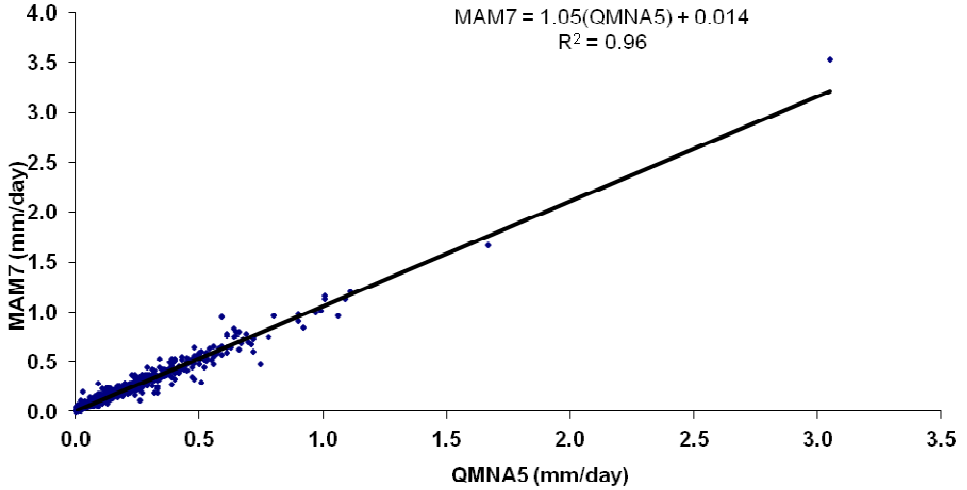


Figure 2.8: Relation between $QMNA_5$ and MAM_7

Similar to **Figure 2.7**, **Figure 2.8** also shows that MAM_7 and $QMNA_5$ are highly correlated with an R^2 value of 0.96. These results suggest that any of the Q_{90}, MAM_7 and $QMNA_5$ descriptors can be used to similarly characterize low-flow severity. In our entire analysis, we consider the Q_{90} as the threshold to define the low-flow severity in the data set, since this statistics is quite widely used in the literature.

The relation between Q_{90}/Q_{50} and BFI is not very strong and hence it is not presented here. Also the high scattering of the plot made between MAM_7 and BFI (not presented here) shows that these two indices represent different characteristics of the low-flow conditions.

2.6.Low-flow indices and catchment descriptors

This section investigates the link between climate and catchment descriptors (such as aridity index and catchment geology) with low-flow indices.

2.6.1. Aridity index (AI)

Aridity index is the degree of dryness of a climate at a given location and is given as:

$$AI = \frac{PE_m}{P_m} \qquad \textbf{Eq. 2.1}$$

where PE_m and P_m are the mean annual potential evapotranspiration (mm) and mean annual precipitation (mm), respectively.

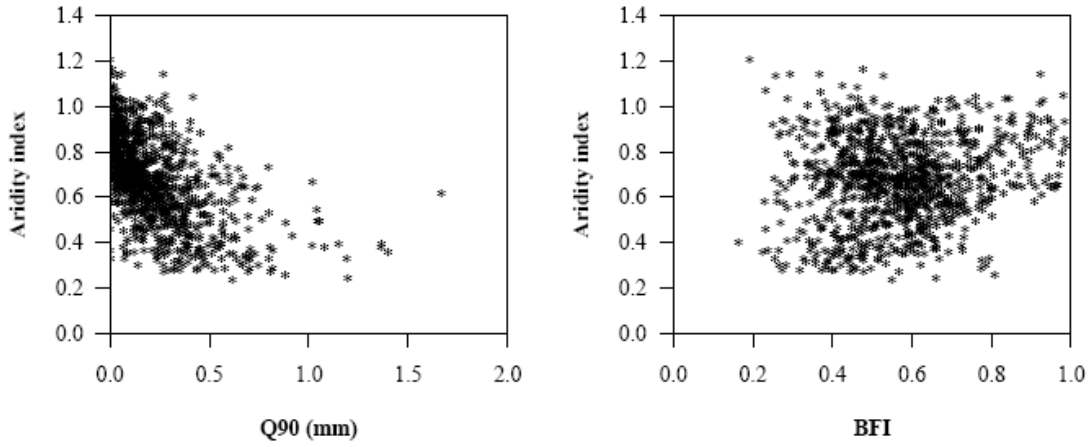


Figure 2.9: Aridity index vs. low-flow indices

We analysed the correlation between aridity index and low-flow indices such as Q_{90} and BFI. In **Figure 2.9**, Q_{90} shows some links with the climatic index, AI. A lower value of Q_{90} tends to correspond to a higher value of AI, i.e. the aridity index of a region is inversely related to this low-flow index with an R^2 value of 0.15. Therefore, low AI values mean more sustainable flow condition. But the link between BFI and the aridity index is much weaker.

2.6.2. Catchment geology

As discussed in the previous chapter, catchment geology can have a major role on the flow regime of a river, especially for a sustained flow condition. We used the Q_{90} and BFI values to make a link with catchment geology. The geological composition of catchments in the French river basins is really complicated and hence we considered only the majority of geology for individual catchments. The fluctuation in the range of values of box plots of the quaternary volcanic in **Figure 2.10** shows that they have highly varying flow conditions.

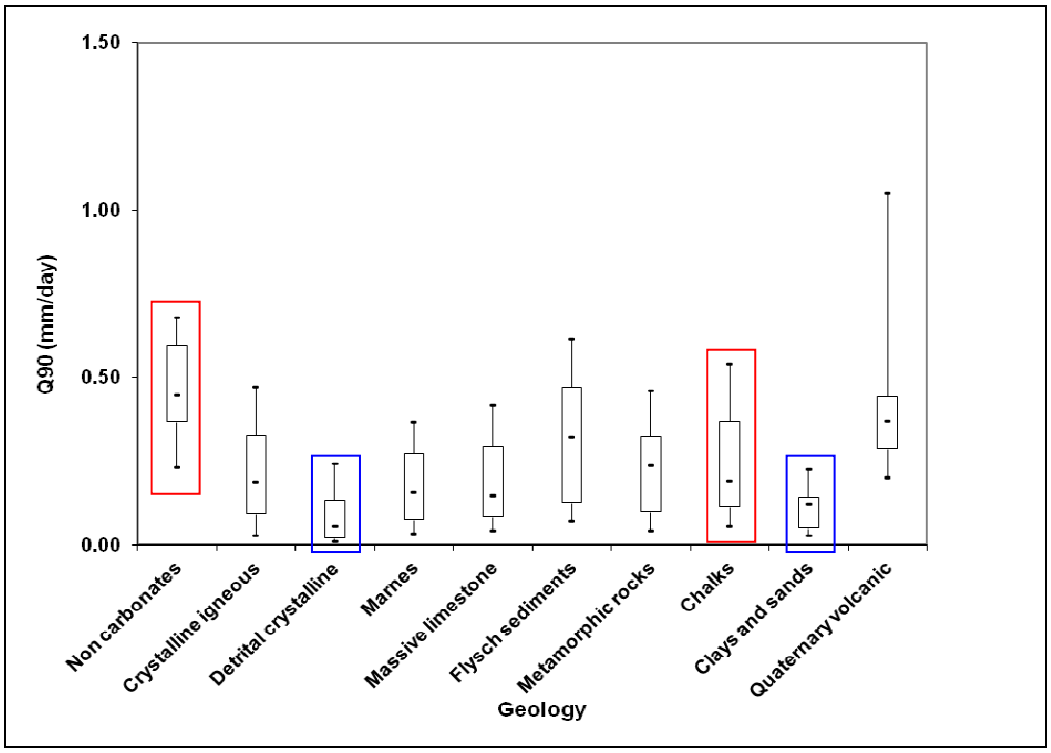


Figure 2.10: Boxplots of distribution of Q_{90} values of catchments based on their geology (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles)

In our data set, catchments with geology of detrital crystalline and clays & sand have comparatively low Q_{90} values which indicate more extreme low flows in those geological conditions.

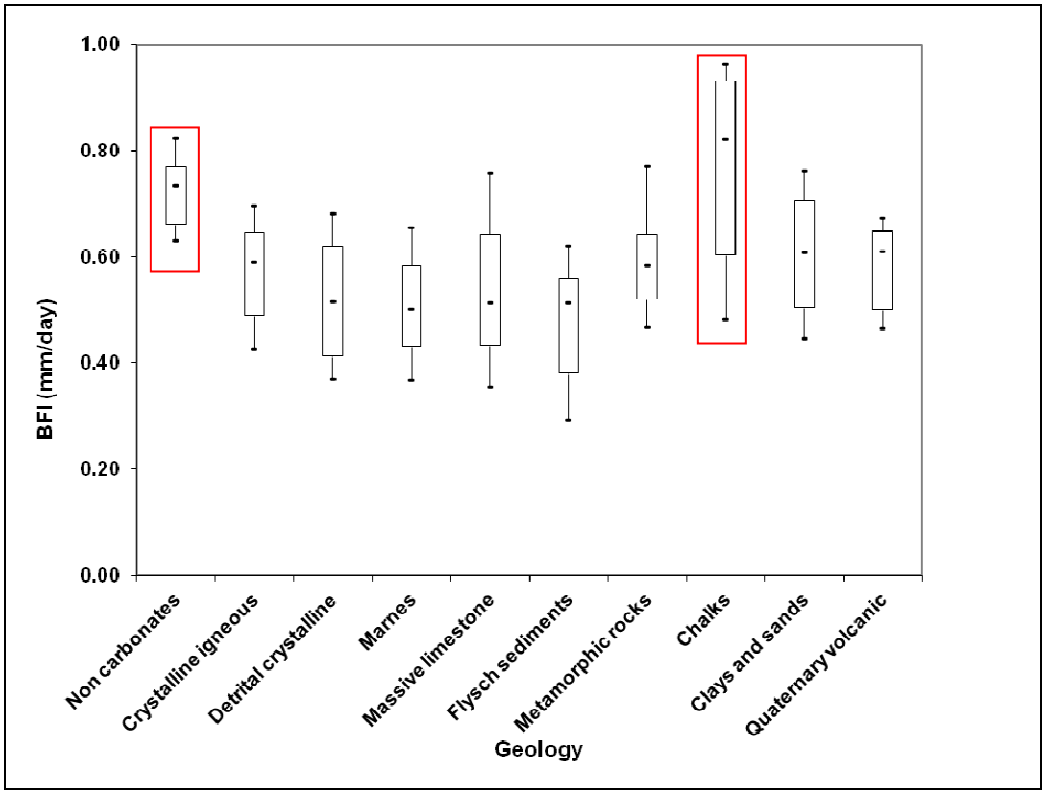


Figure 2.11: Distribution of BFI values based on geology (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles)

Figure 2.11 also indicates that catchments with geology of non-carbonates and chalks (see the box plots highlighted with red colour) have comparatively high base-flow contribution which enables a sustainable flow during the low-flow periods (see the box plots highlighted in red colour in **Figure 2.10**).

2.7. Conclusions

Low-flow indices can provide valuable information on the flow conditions experienced during the dry season. The analysed low-flow measures are FDC (Q_{90}, Q_{50}), MAM_7, $QMNA_5$ and BFI. The detailed analysis of the selected variables shows that Q_{90}, MAM_7 and $QMNA_5$ represent severe low-flow conditions of a catchment. In the following chapters, we use Q_{90} as the low-flow index to characterise low-flow conditions. Q_{50} is the median flow and the analysis of Q_{90} with respect to Q_{50} gives information about the base-flow contribution to total streamflow. BFI gives an idea about the base-flow response in a catchment similar to Q_{90}/Q_{50} and it can also yield the influence from catchment geology on river low flows.

The analysis of aridity index shows that regions with a high value of AI represent extreme low-flow conditions. Low-flow indices are inversely related to the aridity index of a region. The analyses of catchment geology with low-flow indices (Q_{90} and BFI) showed that catchments with geology of chalks and non-carbonates maintained their flows during the low-flow conditions. Due to the subsurface drainage property of massive limestone, they have high fluctuations in the flow conditions throughout the year.

Part II: Stream low-flow simulation

In the previous part we discussed the characteristics of stream low flows based on the analysis of a set of catchments distributed over France. These low-flow characteristics are used in the coming chapters to differentiate the flow regimes (here our interest is on low flows).

In the present part, we focus on stream low-flow simulation. Part II is divided into two chapters:

- Chapter 3 discusses the selection of criteria to evaluate the low-flow simulation models. Here we proposed a criterion which can better evaluate streamflow simulation in low-flow conditions.
- Chapter 4 presents the stepwise development procedures of a better version of low-flow simulation model by the structural sensitivity analysis.

Chapter 3. Analysis of efficiency criteria suitable for low-flow simulation

The content of this chapter corresponds mainly to the following article:

Pushpalatha, R., Perrin, C., Le Moine, N. and Andréassian, C., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. Journal of Hydrology 420-421:171-182.

An edited version of the article is given in Appendix B.

3.1. Introduction

Hydrological modelling aims at understanding and interpreting catchment hydrological behaviour. It is also used to address a number of practical issues, ranging from flood estimation to water resources management and low–flow forecasting. Whatever model is applied, the model user needs appropriate and meaningful indicators informing on the actual capacity of the model.

However, the evaluation of goodness–of–fit is not as straightforward as it may seem at first glance. Of course, model performance can first be evaluated by the visual comparison of the observed and simulated flow hydrographs, but this remains extremely dependent on the evaluator's experience (Chiew and McMahon, 1993; Houghton-carr, 1999). A more objective way to evaluate model performance is to use numerical criteria, but then the user may get lost in a jungle of potential criteria. Why is the choice so difficult? Several reasons can be put forward:

- Flows vary by several orders of magnitude that may not be equally useful for the modeller;
- Hydrological models often produce heteroscedastic errors, i.e. their variance is not independent of the flow value;
- The range of target flows may vary significantly between evaluation periods;
- The model may be used for different applications, which may require specific criteria.

For these reasons, a large variety of criteria have been proposed and used over the years in hydrological modelling, as shown for example by the lists of criteria given by the ASCE (1993), Dawson et al. (2007), Moriasi et al. (2007) and Reusser et al. (2009). Among these criteria, some are *absolute* criteria such as the widely used root mean square error, while others are *relative* criteria (i.e. normalized) such as the Nash and Sutcliffe (1970) efficiency criterion (*NSE*). In the latter, model errors are compared to the errors of a reference or benchmark model (Seibert, 2001; Perrin et al., 2006). This provides a useful quantification of

model performance in that it indicates to which extent the model is better (or worse) than the benchmark. It also facilitates the comparison of performance between catchments. For more on the significance of ranges of values of the *NSE*, please refer to Hall (2001).

However, the choice of a benchmark is difficult: different benchmarks are more or less demanding (and thus the comparison more or less informative) depending on the type of hydrological regime or the type of model application. This sometimes makes it difficult to interpret the relative criteria and a bit puzzling for an inexperienced end–user, who may simply want to know whether the model can be considered as "good", "acceptable" or "bad". Actually, there is no single criterion that can evaluate model performance in all cases (Jain and Sudheer, 2008). Many authors use several criteria simultaneously (see e.g. Efstratiadis and Koutsoyiannis (2010) for a review in the context of multi–objective calibration), but a discussion of these approaches is not within the scope of this study.

The merits and drawbacks of several efficiency criteria have already been discussed and debated in the literature, as well as the links between them. The Nash–Sutcliffe efficiency criterion has probably received the most attention (Garrick et al., 1978; Houghton-carr, 1999; McCuen et al., 2006; Schaefli and Gupta, 2007; Clarke, 2008; Gupta et al., 2009; Moussa, 2010; Gupta and Kling, 2011). Like many other criteria based on the mean model square error, this criterion is known to put greater emphasis on high flows when calculated on a continuous simulation. Although it was shown to have several limitations (such as its sensitivity to the hydrological regime, sample size or outliers), it remains a valuable and popular means to evaluate models for high–flow simulation.

Comparatively little work has been carried out on the meaning and interpretation of the criteria used to evaluate models in low–flow conditions. The following section summarizes the existing studies.

3.2. Criteria used for the evaluation of low–flow simulation

Table 3.1 lists some of the studies discussing performance criteria able to judge low–flow simulations. Note that the various criteria formulations listed here depend on at least three factors:

Table 3.1: Studies that used criteria to evaluate low-flow simulation quality

Reference	Low flow statistics	Residual errors	Standard deviation	Bias	Coeff. of determination	Nash efficiency (all transforms)	Coeff. of variation	Log MSE	Relative efficiency	Index of agreement
Chiew et al. (1993)						x				
Ye et al. (1998)	x	x				x				
Houghton–Carr (1999)	x					x				
Krause et al. (2005)					x	x			x	x
Oudin et al. (2006)		x				x				
Jain and Sudheer (2008)	x		x	x		x	x			
de Vos et al. (2010)								x		

- Calculation period:

Instead of calculating criteria only over the low–flow periods (which generally requires the subjective choice of a low–flow threshold), most of the existing criteria calculate model errors over the entire test period. Thus, they give some weight to the errors in low–flow as well as to the errors in high–flow conditions.

- Target variable:

The second major aspect in calculating criteria is the choice of a target variable. Some authors (e.g. Houghton-Carr, 1999) appeal to statistical measures classically used to characterize low flows, such as the base–flow index, a percentile of the flow duration curve (*FDC*) or some minimum accumulated flows over a continuous period (e.g. 7 days). By calculating the ratio between the simulated and observed values, a relative efficiency criterion is obtained. These criteria are very useful when studies focus on specific aspects of low flows. However, one may also continue calculating the sum of errors over the entire test period, provided that the

appropriate transformation on flows is used (Box and Cox, 1964; Chiew et al., 1993). This transformation helps put more weight on low flows. The root square or the logarithms are among the most widely used transformations on low–flow values. For example, Houghton–Carr (1999), Oudin et al. (2006), Jain and Sudheer (2008) and de Vos et al. (2010) used the sum of squared residuals calculated on the logarithms of flow values in order to reduce the biasing towards peak flows. Chiew et al. (1993) used the root squared transform to evaluate the model's performance in low–flow conditions. Krause et al. (2005) proposed using a relative variable as the ratio between simulation and observation, and calculated the distance of this variable from 1. Le Moine (2008) discussed a generalization of these transformations as a power law transformation with positive or negative exponents and proposed a family of squared criteria based on the power transformation of flows, defined by:

$$RMSE(\lambda) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_i^{\lambda} - \hat{Q}_i^{\lambda})^2} \qquad \textbf{Eq. 3.1}$$

where λ is the power of the flow transformation, n is the number of time steps and Q_i and $\hat{Q}_i$ are the observed and simulated flows, respectively, at time step i. λ is not necessarily an integer and can take positive and negative values. When λ tends towards zero, the transformation tends towards the logarithm transformation. As noted by Le Moine (2008), low and high values of λ will tend to emphasize the model errors on the minimum and maximum flow values, respectively. For example, Chiew et al. (1993) had used a value of λ equal to 0.2 to give more emphasis on low flows.

- Error normalization:

Another aspect that differs between criteria is the type of error used and the way model error is normalized. Most of the criteria are based on the squared residuals, but absolute errors can also be considered. A power of these absolute errors may also be used (see e.g. Krause et al., 2005). In terms of model error normalization, most of the efficiency indexes use the form of the *NSE*. Willmot (1984) proposed the index of agreement as another way to normalize model square error, by dividing it by the potential error.

3.2.1. Are existing criteria appropriate to evaluate low-flow simulations?

Only a few authors have discussed the suitability of the variety of existing criteria to evaluate low-flow simulations. Oudin et al. (2006) compared several objective functions and concluded that the square root transformation provides an all–purpose efficiency measure, not specifically focusing on low flows. Analysing several efficiency indices, Krause et al. (2005) showed that some criteria are closely related while others show very different patterns. They advised using the relative efficiency index for the evaluation of low–flow simulations, noting that the logarithm transformation on flows provides a higher sensitivity to low flows, although they indicate that this criterion remains sensitive to high flows.

Actually, the impact of flow transformations significantly changes the way the hydrograph and model errors are considered in criteria, as shown by Le Moine (2008). This is illustrated in **Figure 3.1** in the case of natural, root square and logarithmic transformed flows.

Figure 3.1 shows the series of flow values and model errors as well as the cumulated quadratic error (to facilitate the comparison between graphs, all series were normalized by the maximum value in the series, thus providing values ranging between 0 and 1). **Figure 3.1** illustrates the significant differences between transformations and the resulting error values: indeed, the part of the hydrograph that produces most of the errors strongly depends on the transformation chosen. Hence, a detailed analysis of power law transformations is needed and will be discussed in section 3.5.6.

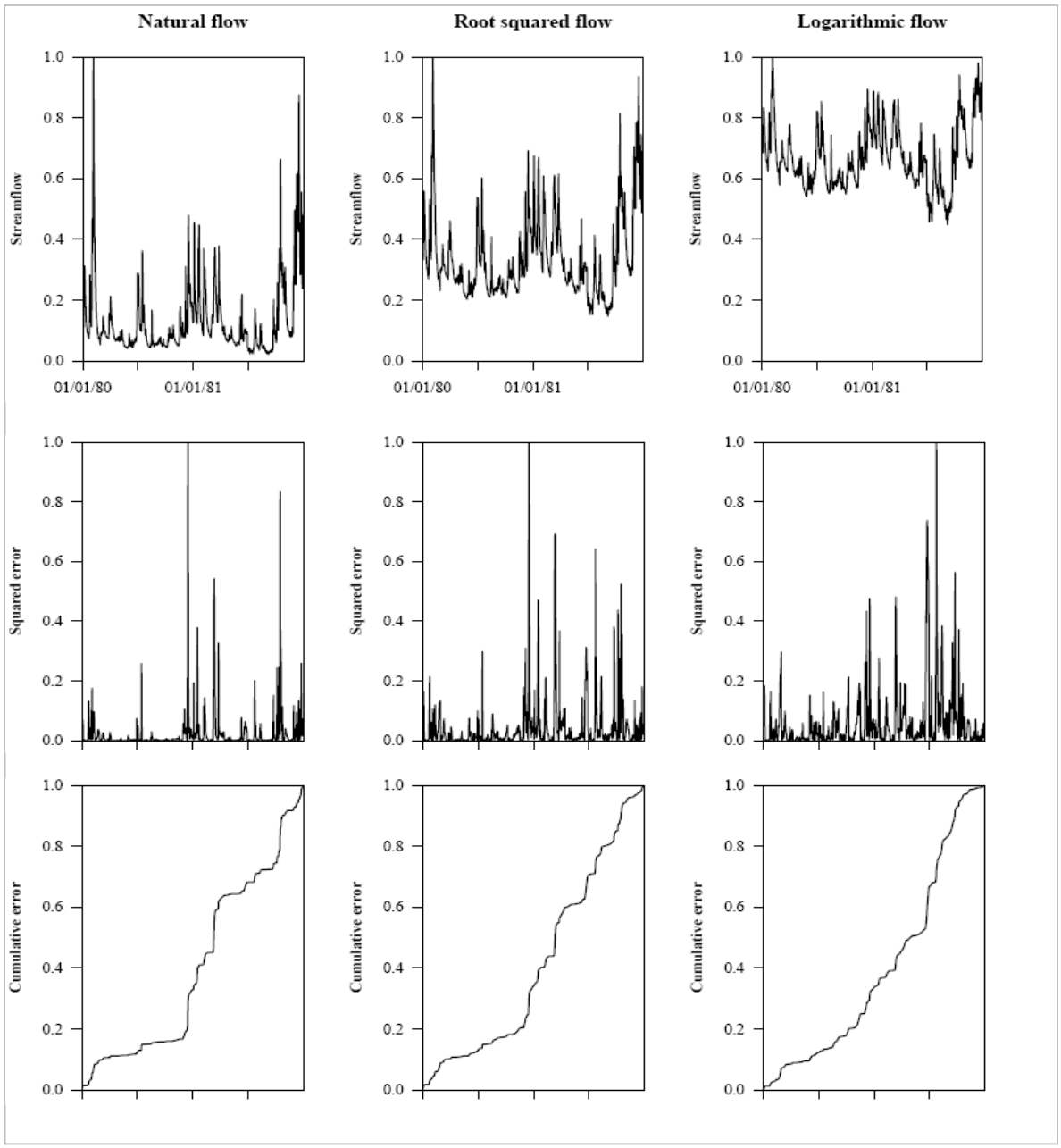


Figure 3.1: Illustration of changes in the series of flows, model errors and cumulated squared errors using root square and logarithm transformations for the Moselle River at Custines (Hydro code: A7010610; catchment area: 6830 km²; period: 01/01/1980–31/12/1981).

3.3.Scope of the study

The present work intends to complement previous studies by objectively identifying which criteria are best suited for the evaluation of low–flow simulations using actually observed data. The two main objectives are listed below:

- First, we wish to discuss the links between the criteria used or proposed to evaluate low–flow simulations.
- Second, we wish to better understand which part of the hydrograph most influences the various criteria used for the evaluation of low–flow simulations.

To meet these objectives, we will evaluate various criteria, mainly pertaining to the family of criteria proposed by Le Moine (2008) in the form of the Nash–Sutcliffe efficiency index. They will be compared to other criteria proposed in the literature for low–flow evaluation on a large set of catchments showing various low–flow characteristics. We will restrict this analysis to (i) the case of non-dimensional criteria (i.e. relative criteria in which the model error is normalized), and (ii) deterministic low–flow simulations. Although several authors recently questioned the relevance of deterministic predictions in an information theory perspective (Weijs et al., 2010), the criteria for the evaluation of probabilistic low–flow predictions are not within the scope of this study (see e.g. Laio and Tamea (2007), Bartholmes et al. (2009), Boucher et al. (2009), Ramos et al. (2010) or Randrianasolo et al. (2010) for discussions on this issue). Last, we will not discuss the issue of model calibration (i.e. the choice of objective function), but concentrate on the evaluation of simulations only.

The next section presents the methodology used in this study. Then the results are detailed and discussed before the concluding remarks.

3.4.Models and evaluation methodology

3.4.1. Hydrological models

The tests were carried out using two continuous lumped rainfall–runoff models: the four–parameter GR4J model (Perrin et al., 2003) and a six–parameter version of the MORDOR model (Garçon, 1999; Mathevet, 2005) called MORD hereafter to avoid confusion with the original model. These models are quite widely used in France for operational purposes. Schematic diagrams of model structures are shown in **Figure 3.2**. For detailed descriptions of the models, please refer to Andréassian et al. (2006b). Here, models were used in the same conditions, i.e. they were fed exactly with the same rainfall and potential evapotranspiration inputs.

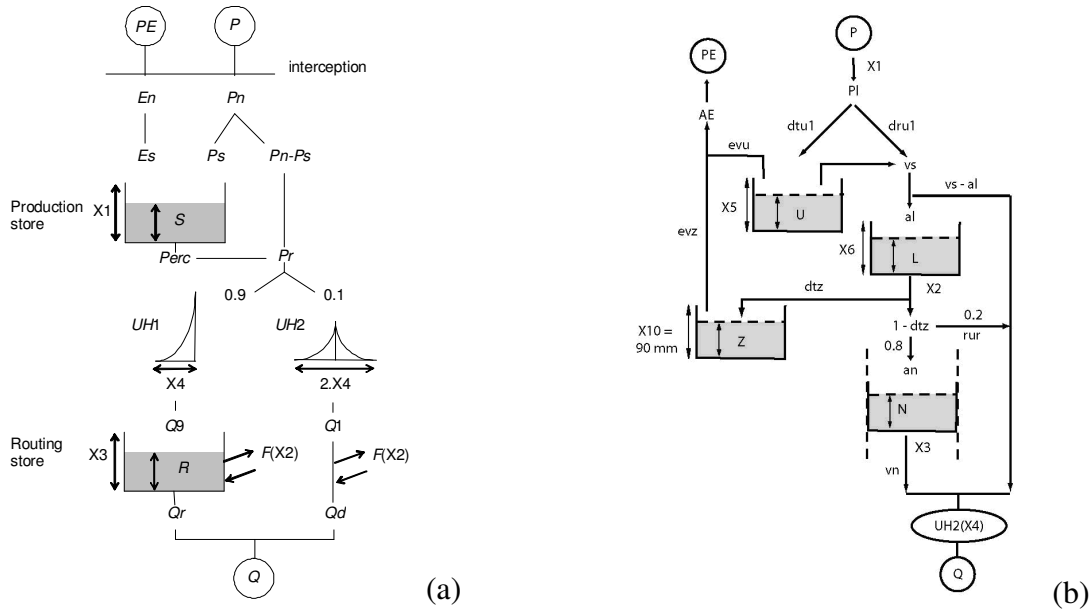


Figure 3.2 Schematic diagrams of the structures of (a) the GR4J model and (b) the MORD model (PE: potential evapotranspiration; P: rainfall; Q: streamflow; X parameters; other letters are internal model variables)

Catchment size is indeed quite variable on our catchment set (5.3 to 9423 km²). The two models used in this study are lumped. Our experience with lumped models is that catchment size is not a strong limitation to apply this type of models. Actually, we noticed on large data sets that the probability of model failure on large catchments is much lower than on small catchments. This pattern is illustrated in **Figure 3.3** for the GR4J model tested on the catchment set used in our study. We drew box plots of efficiencies for the $NSE*_Q$ criterion as a function of catchment size. Each of the ten box plots is built on one tenth of the catchments. We can see that there is an increasing trend in the median values as catchment size increases. We can also see that there is also an increasing trend in the lower percentiles, indicating a lower probability of the model to get bad efficiencies; a similar behaviour is observed with the MORD model. This can be related to the results presented by Merz et al. (2009) who applied a hydrological model (with semi–distributed inputs but lumped parameters) to 269 Austrian catchments ranging in size from 10 to 130,000 km². They concluded that model efficiency tends to increase over the scale range of 10 to 10,000 km².

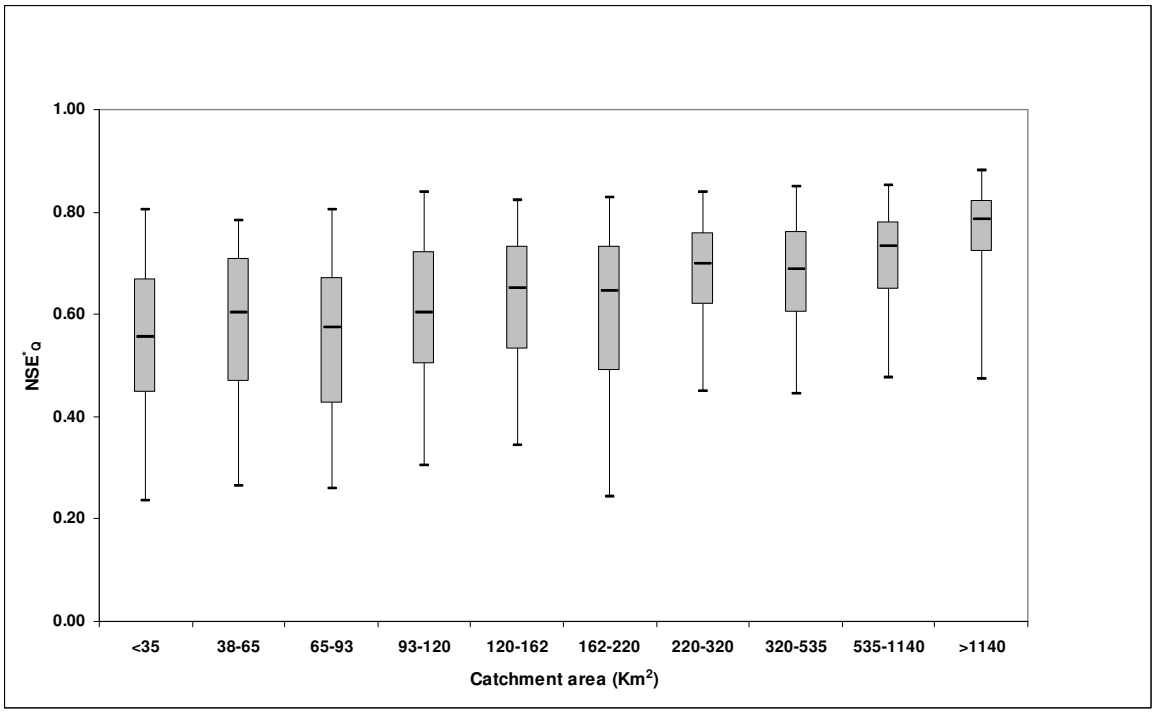


Figure 3.3: Box plots of performance (NSE^*_Q in validation) of the GR4J model on the 940 catchments as a function of catchment size (boxplots show the 0.05, 0.25, 0.5, 0.75 and 0.95 percentiles from bottom to top; each box plot contains one tenth of catchments)

Note that the two models simulate the low flows in a quite different manner: in the GR4J model, low flows are generated by the percolation from the soil moisture store (S), outflow from the routing store (R) and the groundwater exchange function (F) (which simulates possible interactions with deep aquifers or surrounding catchments). However, in MORD, low flows are produced by two routing stores (L and N) (**Figure 3.2**).

The models' parameters were calibrated using a two–step search procedure (a prior gross inspection of the parameter space to identify the most likely zone of convergence, followed by a local search algorithm). This approach proved efficient for these parsimonious models compared to more complex search algorithms (Edijatno et al., 1999; Mathevet, 2005). As an objective function for calibration, we used the root mean squared error calculated on root square flows, which was found by Oudin et al. (2006) to be a good compromise for an all–purpose model (not giving too much emphasis to low or high flows). The performances of the GR4J model for different objective functions are presented in Appendix A and are in agreement with the previous studies.

3.4.2. Testing scheme

This section describes the method used to evaluate criteria. The split sample testing scheme proposed by Klemes (1986) was used to evaluate model performance. For each catchment, the period where rainfall, PE and flow data were available (at most 1970–2006) was split into two halves (P1 and P2) of similar length, alternatively used for model calibration and validation. It means that for each catchment and each tested model, two calibrations and two validations were systematically performed. The first year of each test period was used for model warm–up. To avoid initialization problems for catchments with long–term memories, five years of warm–up were considered in addition to the 1–year warm–up period: they were either the five years of observed data preceding the test period when available, or a mean year repeated five times otherwise. In this study, only performance in validation was considered to evaluate models.

3.4.3. Criteria analysed

The analysis was carried out on the criteria detailed in **Table 3.2** :

- The first five criteria represent different Nash–Sutcliffe efficiencies calculated on flows (NSE_Q), root squared flows (NSE_{sqrtQ}), logarithmic transformed flows (NSE_{lnQ}), inverse flows (NSE_{iQ}) and relative flows (NSE_{rQ}). NSE_{iQ} is not commonly used in the literature and was suggested by Le Moine (2008). The comparison of these criteria will provide information on the effect of prior flow transformation.
- NSE_{uT} is the *NSE* calculated under a threshold taken equal to the 0.9 percentile of the observed flow duration curve of the test period (other thresholds could obviously be considered). Compared to the previous criteria, NSE_{uT} is calculated only on one–tenth of the time steps corresponding to the lowest observed flows and will provide information on the use of a threshold.
- IA_{rel} is the relative index of agreement proposed by Krause et al. (2005). Here, the normalization of the error is different from the one made in the *NSE*.
- B_{uT} is the bias under the same threshold as for NSE_{uT} (0.9 flow percentile). Contrary to the previous criteria that are all based on the squared error, B_{uT} is based on the cumulative error and will indicate the error of water balance in low–flow conditions.

- Last, $RLFD_{uT}$ is the flow deficit under the low–flow threshold (0.9 flow percentile). The formulation of $RLFD_{uT}$ is similar to that of B_{uT} and indicates the error of the water deficit in low–flow conditions based on the threshold.

As these criteria have no lower bounds, they may take strongly negative values in case of model failure. These individual values can severely impact the calculation of statistical moments of the set of criteria (e.g. the mean value). To avoid this bias, we used the mathematical transformation proposed by Mathevet et al. (2006), in which the range of variation of the transformed criterion is]-1;1] instead of]-∞;1]. The transformation is given by:

$$C^* = \frac{C}{2 - C} \qquad \textbf{Eq. 3.2}$$

where C^* is the transformed criterion and C is the initial one. In addition to solving the problem of extreme negative values, the 0 and 1 values remain the same: $C^* = 0$ for $C = 0$ and $C^* = 1$ for $C = 1$, hence keeping the same interpretation relative to the benchmark. Note that for $C > 0$, C^* will be lower than C and the reverse for $C < 0$.

In the next sections, we will use the notations NSE^*_Q, NSE^*_{sqrtQ}, $NSE^*_{\ln Q}$,…, $RLFD^*_{uT}$ for the bounded versions of the criteria listed in **Table 3.2** .

Note that for some criteria (e.g. NSE^*_{lnQ} and NSE^*_{iQ}), it is not mathematically possible to compute the transformed flow values when simulated or observed flows equal zero. However, these time steps should not be ignored during model evaluation, or else it could result in an overly optimistic assessment of model performance. To avoid this, one can add a small quantity to the flow value before computing its transformation. It can be, for example, a small fraction of the mean observed flow over the test period. In the following, ε will be set at one hundredth of the mean flow, but the choice of this value will be further discussed in section 3.5.5. Note that the same comment applies to all criteria derived from Eq. 3.1 with negative λ values.

Table 3.2: Selection of evaluation criteria and their corresponding formulation and specific values (Q_i and $\hat{Q}_i$ are the observed and simulated flows, respectively, n the total number of time steps, T a low-flow threshold (here $T = Q_{90}$), $\overline{Q}$ the mean of Q, and ε a small constant)

Criterion	Usual name	Reference	Mathematical formulation	Range of values	Value of perfect agreement	Target variable	Benchmark model
NSE_Q	NSE calculated on flows	Nash and Sutcliffe (1970)	$1-\frac{\sum_{i=1}^{n}(Q_i-\hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i-\overline{Q_i})^2}$	]-∞,1]	1	Q_i	$\overline{Q_i}$
NSE_{sqrtQ}	NSE calculated on root squared transformed flows	Oudin et al. (2006)	$1-\frac{\sum_{i=1}^{n}(\sqrt{Q_i}-\sqrt{\hat{Q}_i})^2}{\sum_{i=1}^{n}(\sqrt{Q_i}-\overline{\sqrt{Q_i}})^2}$	]-∞,1]	1	$\sqrt{Q_i}$	$\overline{\sqrt{Q_i}}$
$NSE_{\ln Q}$	NSE calculated on log transformed flows	Oudin et al. (2006)	$1-\frac{\sum_{i=1}^{n}(\ln(Q_i+\varepsilon)-\ln(\hat{Q}_i+\varepsilon))^2}{\sum_{i=1}^{n}(\ln(Q_i+\varepsilon)-\overline{\ln(Q_i+\varepsilon)})^2}$	]-∞,1]	1	$\ln(Q_i+\varepsilon)$	$\overline{\ln(Q_i+\varepsilon)}$
NSE_{iQ}	NSE calculated on inverse transformed flows	Le Moine (2008)	$1-\frac{\sum_{i=1}^{n}(\frac{1}{Q_i+\varepsilon}-\frac{1}{\hat{Q}_i+\varepsilon})^2}{\sum_{i=1}^{n}(\frac{1}{Q_i+\varepsilon}-\overline{\frac{1}{Q_i+\varepsilon}})^2}$	]-∞,1]	1	$\frac{1}{Q_i+\varepsilon}$	$\overline{\frac{1}{(Q_i+\varepsilon)}}$
NSE_{rQ}	NSE calculated on relative flows	Krause et al. (2005)	$1-\frac{\sum_{i=1}^{n}((Q_i-\hat{Q}_i)/(Q_i+\varepsilon))^2}{\sum_{i=1}^{n}((Q_i-\overline{Q_i})/(\overline{Q_i+\varepsilon}))^2}$	]-∞,1]	1	$\frac{Q_i}{\overline{Q_i}}$	1
NSE_{uT}	NSE calculated under low-flow threshold		$1-\frac{\sum_{i\in[1,n]\mid Q_{oi}<T}(Q_i-\hat{Q}_i)^2}{\sum_{i\in[1,n]\mid Q_{oi}<T}(Q_i-\overline{Q_i})^2}$	]-∞,1]	1	$Q_i < Q_{90}$	$\overline{Q_i}\ _{(Q_i<Q_{90})}$
IA_{rel}	Relative index of agreement	Krause et al. (2005)	$1-\frac{\sum_{i=1}^{n}((Q_i-\hat{Q}_i)/Q_i)^2}{\sum_{i=1}^{n}((\lvert Q_i-\overline{Q_i}\rvert+\lvert\hat{Q}_i-\overline{Q_i}\rvert)/\overline{Q_i})^2}$	]-∞,1]	1	$\frac{Q_i}{\overline{Q_i}}$	$\overline{Q_i}$
B_{uT}	Bias under low-flow threshold		$1-\frac{\sum_{i\varepsilon[1,n]\mid Q_i<T}\hat{Q}_i}{\sum_{i\in[1,n]\mid Q_i<T}Q_i}$	]-∞,1]	0	$Q_i < Q_{90}$	---
$RLFD_{uT}$	Ratio of Low-flow deficit		$1-\frac{\sum_{i\in[1,n]\mid Q_i<T}Max(0,T-\hat{Q}_i)}{\sum_{i\in[1,n]\mid Q_i<T}(T-Q_i)}$	]-∞,1]	0	$T-Q_i$	---

3.4.4. Approach for criterion analysis

The main objective of this study was to determine which criteria were actually the best adapted to the evaluation of low flows, by an objective analysis on a large set of catchments presented in Chapter 2. To this aim, we identified the part of the hydrograph bearing most of the total model error. The procedure adopted is as follows:

Rank daily flow values over the study period by decreasing order. For each of them, the corresponding model squared error term *SE* can be computed.

Compute the empirical cumulative frequency, k/N, with k the rank of the flow value and N is the total number of time steps.

Compute the weighted empirical cumulative frequency by using the values of errors *SE* as weights w. The weighted cumulative frequency for the k^{th} flow is therefore given by the ratio between the accumulated error for the k smallest flows and the total error (*SSE*) over the N time steps:

$$F_k = \frac{\sum_{i=1}^{k} w_i}{\sum_{i=1}^{N} w_i} = \frac{\sum_{i=1}^{k} SE_i}{\sum_{i=1}^{N} SE_i} = \frac{\sum_{i=1}^{k} SE_i}{SSE} \qquad \textbf{Eq. 3.3}$$

Summarize the weighted cumulative distribution as a box plot showing the flow percentiles of exceedance: 0.05, 0.25, 0.50, 0.75 and 0.95.

Transform back these flow percentiles into percentiles of the flow duration curve. The range between the 0.05 and 0.95 percentiles gives a 90% confidence interval of the range of observed flows contributing the most to the total model error.

An example of weighted and non-weighted distributions and the results of steps 4 and 5 above are illustrated in **Figure 3.4**. **Figure 3.4** clearly explains the derivation of box plot from the distribution of weighted flows for squared errors to the flow duration curve (non-weighted flows) for the Moselle River at Custines for a period of 1980 to 1981. For the same example, **Figure 3.5** illustrates flow values in the hydrograph that contributes most of the model error. In **Figure 3.5**, the darker the observation, the larger the contribution of this time steps to the total model error.

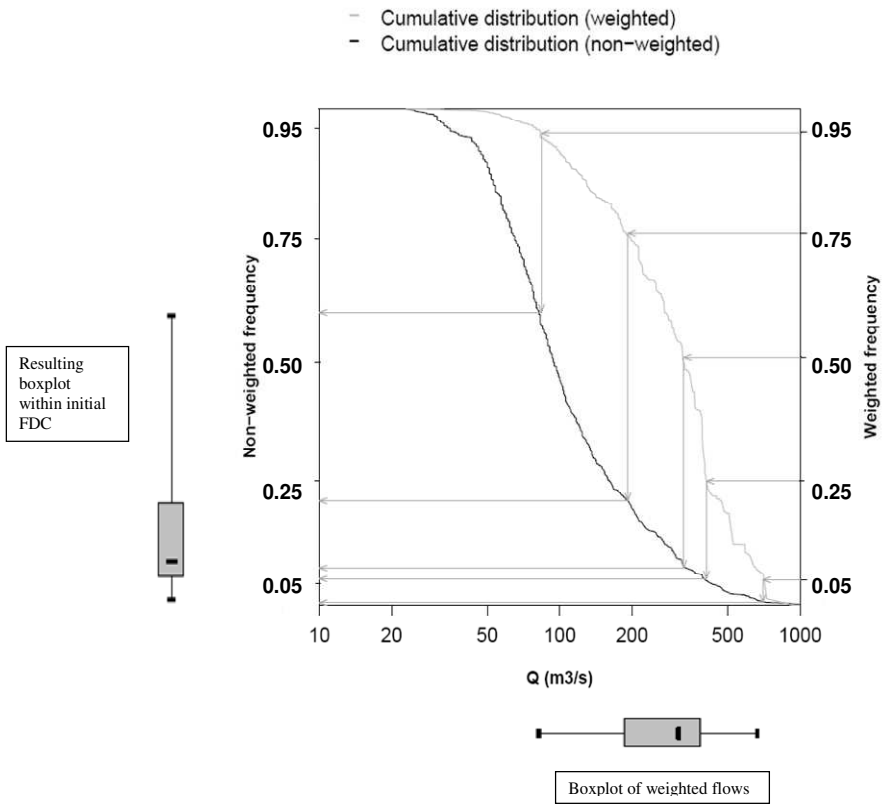


Figure 3.4: Example of derivation of box plots from the distribution of weighted flows (here weights are the squared errors) to the flow duration curve (distribution of non weighted flows). The same example as in Figure 3.1 is used here (Period: 01/01/1980 – 3112/1981)

This was done for each criterion and for each catchment. The resulting box plots can be compared between catchments. We did not analyse the results for each catchment individually, but for the whole catchment set to obtain more general conclusions. To this aim, we will draw a "master" box plot, in which each percentile shown will correspond to the median value of the same percentiles of all the box plots computed over the catchment set. An illustration of the derivation of the master box plot is presented in **Figure 3.6**.

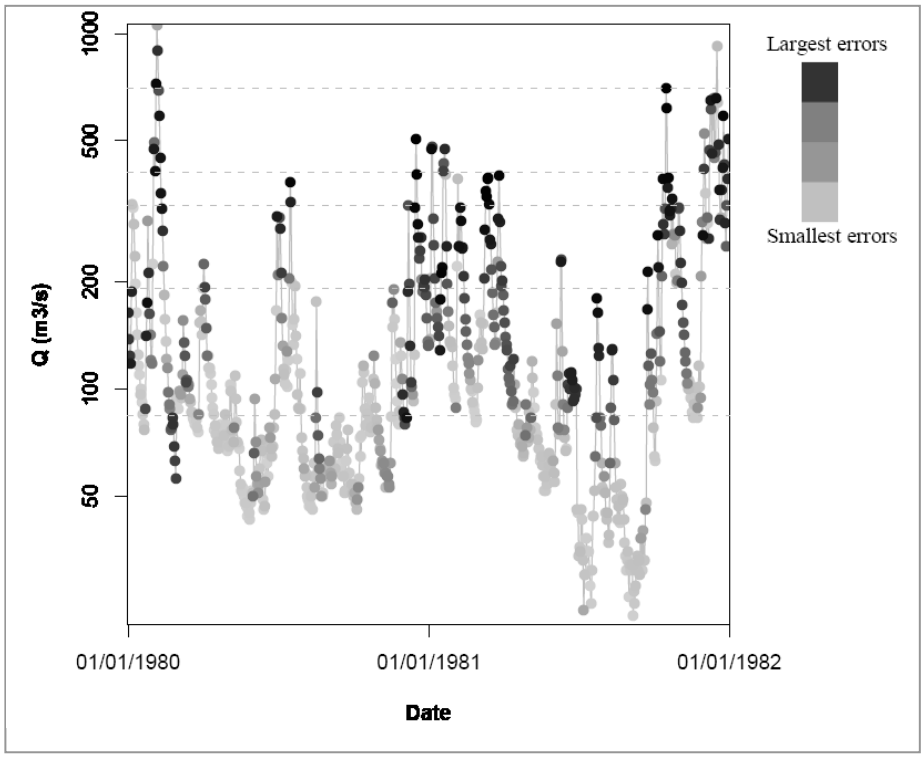


Figure 3.5: Illustration of flow values that contribute the most of the total model error. The error considered here is the square error calculated on flows. The darker the dots, the larger the contribution to the total error. Horizontal dashed lines correspond to the percentiles of the box plot shown in Figure 3.4.

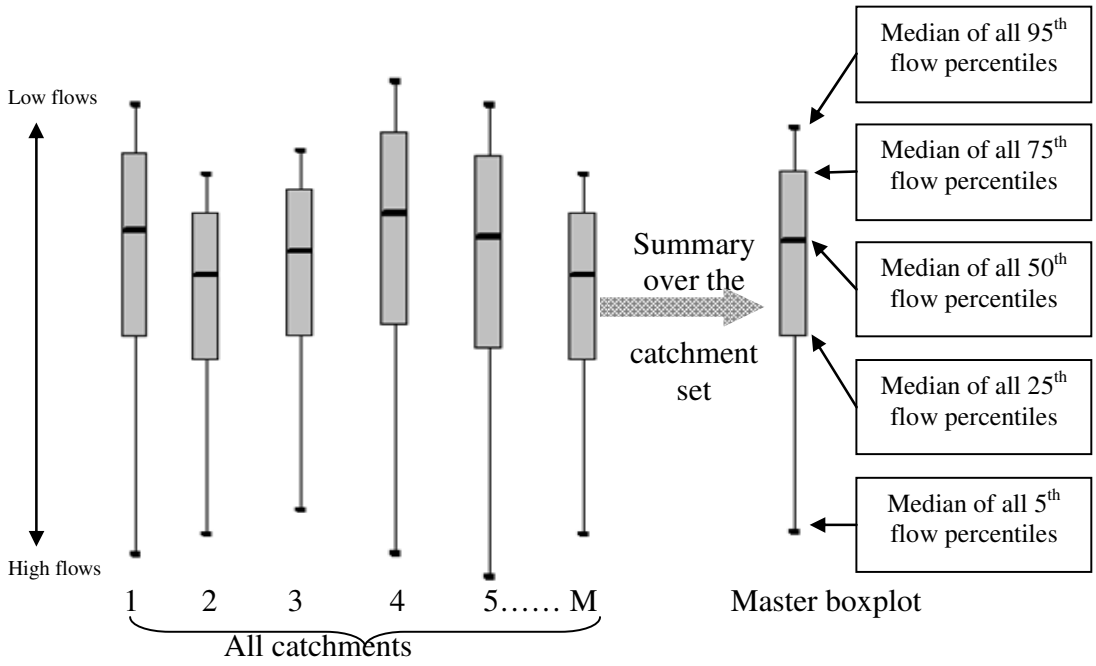


Figure 3.6: Schematic representation of the derivation of master box plot which summarizes the behaviour of a given criterion over the entire catchment set (M stands for the total number of catchments)

3.5. Results and discussion

This section first presents the distributions of the efficiency criteria over the data set and then analyses their interdependency and their behaviour in terms of dependence to low–flow values. This leads to a selection of criteria for the evaluation of low–flow simulation.

3.5.1. Overall efficiency distribution

Figure 3.7 shows the box plots of the distributions of efficiency values for the entire catchment set and for the nine criteria. The two models show quite similar behaviours.

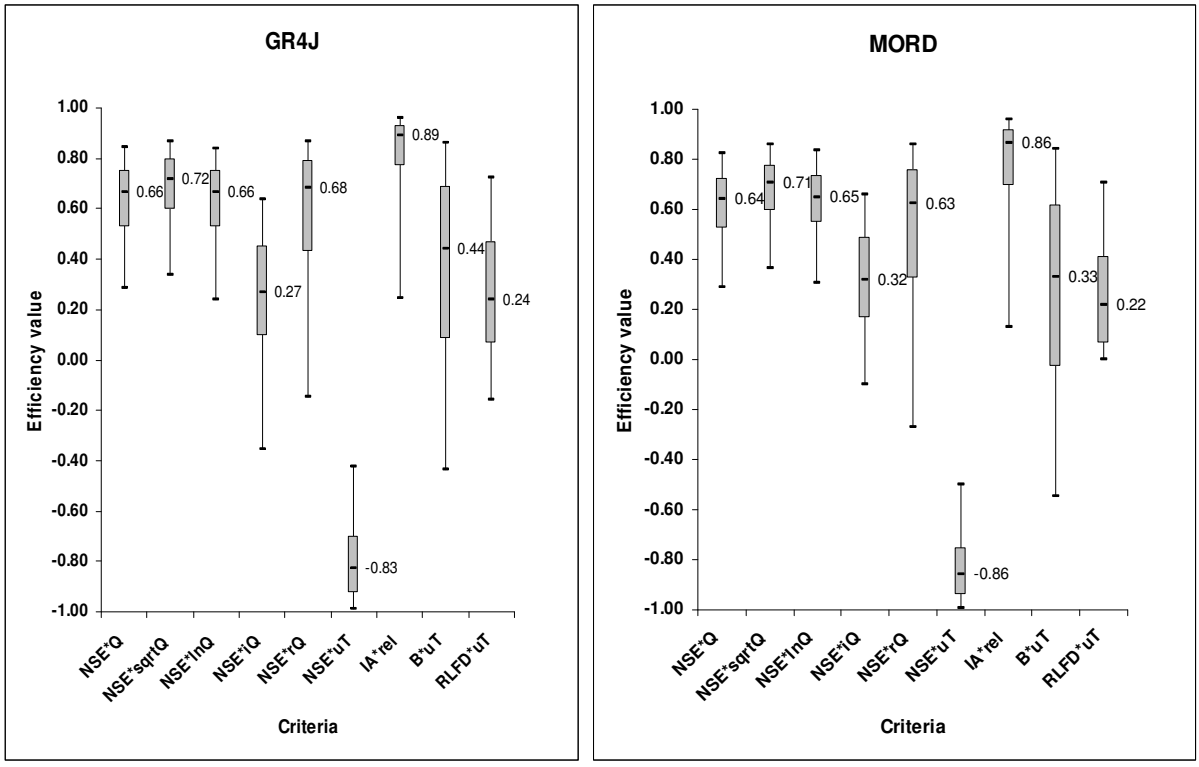


Figure 3.7: Box plot of distribution of efficiency criteria obtained by the GR4J and MORD models over the entire catchment set in validation (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.05 and 0.95 percentiles)

The NSE^{*}_{uT} (*NSE* calculated under 0.9 exceedance flow percentile) exhibits very different behaviour from the other criteria. Its values mostly lie between −1 and 0. This indicates that on most catchments, the two models used are not better than the benchmark that provides the mean observed value under the threshold as the simulation. Although this result is a bit disappointing, it should be remembered that the variability of flow values in these conditions

is limited, so that the benchmark is already a model that is difficult to equal (even if a constant flow is not perfect in terms of dynamics, it is already quite good). Besides, the benchmark is by construction perfect in terms of water balance, which means that no model can be better (see e.g. Perrin et al., 2006). Therefore, although they are clear in terms of the level of performance provided by the models, these mostly negative values make the criteria difficult to interpret: the benchmark is probably too demanding in this case. The distributions of the other efficiency criteria show a much larger share of positive values (**Figure 3.7**).

This variety of efficiency ratings obtained with the different criteria comes clearly from the differences in criteria formulations that put more or less emphasis on different parts of the hydrograph. This indicates that the tendency of modellers to declare that a model is "bad" or "good" based on the interpretation of relative efficiency criteria is often dangerous: it would be more rigorous to always refer to the benchmark used to say to what extent the model is "worse" or "better" than this benchmark.

3.5.2. Criteria interdependency

To analyse the possible interdependency between criteria, we drew scatter plots of pairs of criteria over the catchment set. The scatter diagrams for the possible combinations of criteria are presented in **Table 3.3** and **Table 3.5** for the GR4J and MORD models, respectively (note that patterns in scatter diagrams were similar for the two models tested, which indicates that they are not model–specific). In order to clarify the scatter plots, the corresponding Spearman correlation coefficients of the plots for both the models are presented in **Table 3.4** and **Table 3.6**. The scatter plots of NSE_Q^* vs NSE_{sqrtQ}^*, NSE_{sqrtQ}^* vs $NSE_{\ln Q}^*$, and NSE_{rQ}^* vs IA_{rel}^* illustrate that there is a significant relation between these pairs of criteria. It means that they provide similar information on model efficiency. The relation between NSE_{rQ}^* and IA_{rel}^* could be expected since these criteria are based on the same model error and differ only in their normalization. The larger scatter between $NSE_{\ln Q}^*$ and NSE_Q^* shows that the log transformation substantially changes the information contained in the NSE_Q criterion. This is even more obvious when looking at the absence of relationships between NSE_Q^*, and NSE_{iQ}^* and NSE_{rQ}^* on the peak values. For the remaining pairs of criteria, the relationships are generally weak, indicating that they probably provide complementary information.

Table 3.3: Scatter plots of pairs of criteria on the 940–catchment set for the GR4J in validation

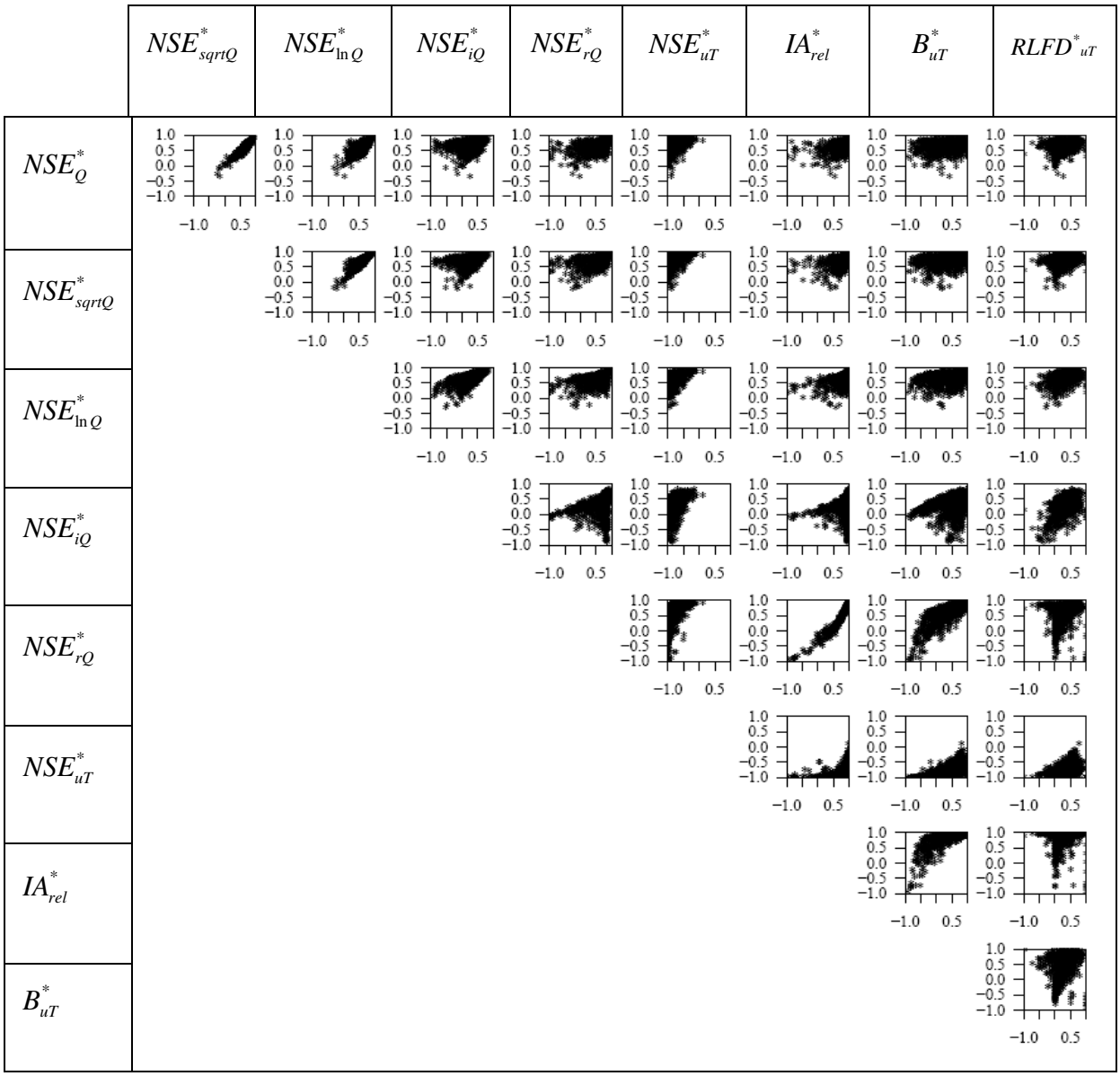


Table 3.4: The Spearman correlation coefficient for the criteria corresponding to the plots in Table 3.3

	NSE^*_{sqrtQ}	NSE^*_{lnQ}	NSE^*_{iQ}	NSE^*_{rQ}	NSE^*_{uT}	IA_{*rel}	$RLFD^*_{uT}$	B^*_{uT}
NSE^*_Q	0.95	0.79	0.26	0.48	0.43	0.44	0.29	0.25
NSE^*_{sqrtQ}		0.90	0.33	0.51	0.40	0.46	0.23	0.28
NSE^*_{lnQ}			0.65	0.50	0.44	0.48	0.24	0.37
NSE^*_{iQ}				0.19	0.42	0.21	0.10	0.48
NSE^*_{rQ}					0.51	0.95	0.80	0.17
NSE^*_{uT}						0.43	0.60	0.47
IA_{*rel}							0.73	0.11
$RLFD^*_{uT}$								0.29

Table 3.5: Scatter plots of pairs of criteria on the 940–catchment set for the MORD in validation

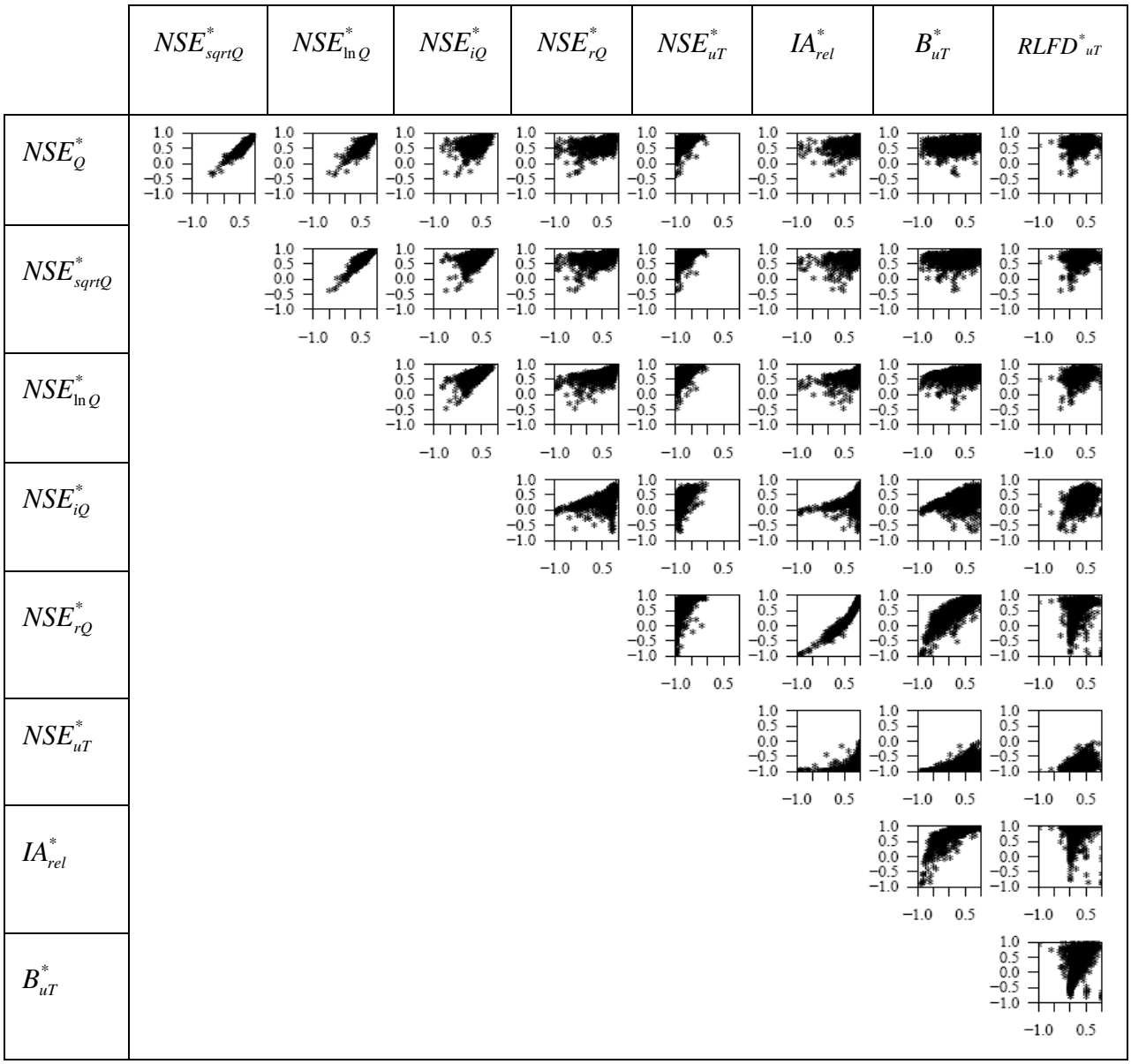


Table 3.6: The Spearman correlation coefficient for the criteria corresponding to the plots in Table 3.5

	NSE^*_{sqrtQ}	NSE^*_{lnQ}	NSE^*_{iQ}	NSE^*_{rQ}	NSE^*_{uT}	IA^*_{rel}	$RLFD^*_{uT}$	B^*_{uT}
NSE^*_Q	0.95	0.83	0.41	0.46	0.38	0.41	0.28	0.26
NSE^*_{sqrtQ}		0.93	0.45	0.47	0.33	0.42	0.21	0.26
NSE^*_{lnQ}			0.69	0.57	0.40	0.52	0.30	0.31
NSE^*_{iQ}				0.48	0.50	0.45	0.38	0.39
NSE^*_{rQ}					0.53	0.96	0.83	0.27
NSE^*_{uT}						0.45	0.63	0.43
IA^*_{rel}							0.76	0.17
$RLFD^*_{uT}$								0.42

3.5.3. Analysis of model error terms

As explained in section 3.4.4, we built master box plots over the entire set of catchments for the nine efficiency criteria to try to identify from which part of the flow duration curve most of the model error comes on average. Here again, the results show similar patterns for the two simulation models, indicating that they are not highly model–dependent. The results are shown in **Figure 3.8**. A brief content in **Figure 3.8** is discussed below.

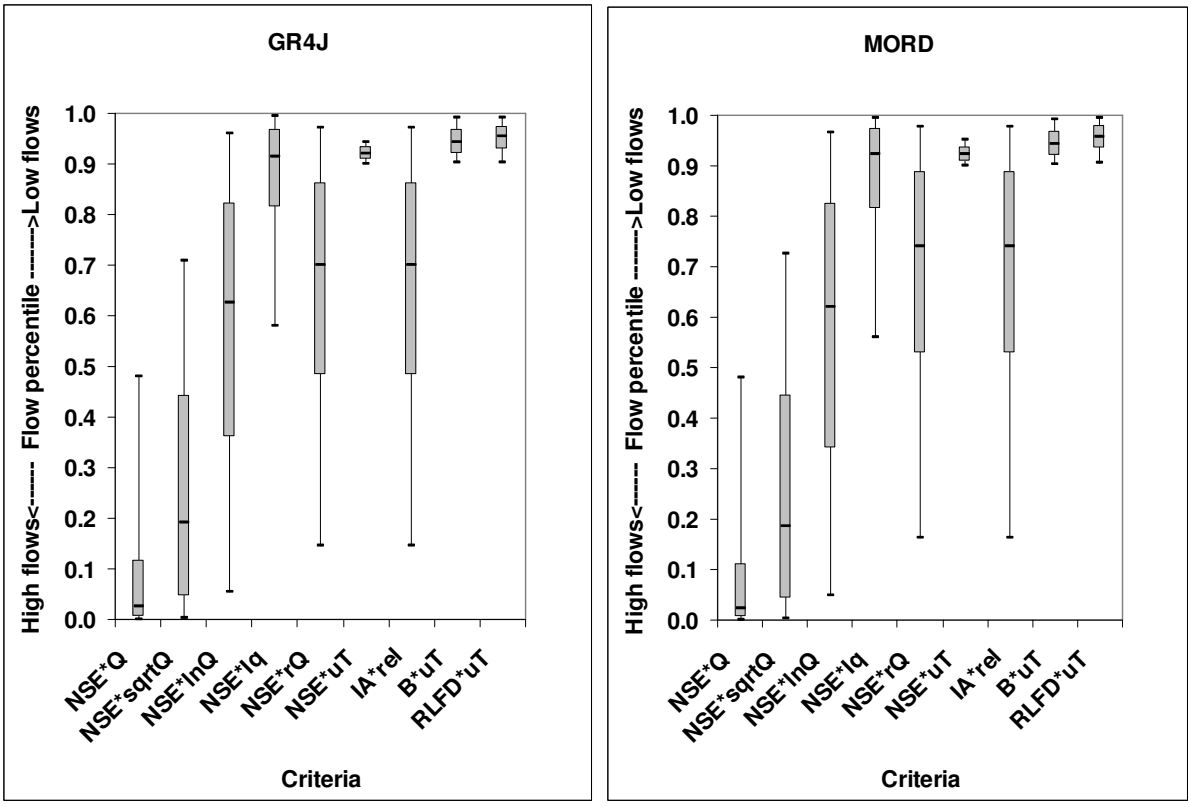


Figure 3.8: Master box plot obtained over the entire catchment set for the GR4J and MORD models (each flow percentile of exceedance corresponds to the median of this percentile over all the box plots obtained on each catchment of the entire set)

As could be expected for NSE_Q^* given the past analyses on this criterion, 90% of the model error accumulates below the 20th percentile of the *FDC*. This confirms the bias of NSE_Q^* towards high flows and strongly supports the suitability of NSE_Q^* for the evaluation of model performance on flood events.

The box plots for the NSE_{sqrtQ}^*, $NSE_{\ln Q}^*$, and NSE_{iQ}^* criteria show that the transformations with decreasing power (root square, logarithm, inverse) progressively put more emphasis on

low flows. The inverse transformation seems the most appropriate if one wishes to concentrate on the very low flows, i.e. roughly the 20% lowest flows in the hydrograph. The box plots for NSE^*_{rQ} and IA^*_{rel} are similar. Ironically, $NSE^*_{\ln Q}$ did not show clear links with NSE^*_{rQ} and IA^*_{rel}, but these criteria seem to concentrate on similar parts of the hydrograph, with a stronger emphasis on low flows. NSE^*_{uT}, B^*_{uT} and $RLFD^*_{uT}$ stress the 10% lowest flows by construction.

3.5.4. Which criterion should be used for low–flow evaluation?

Our results tend to confirm previous comments made in the literature:

- NSE^*_Q should only be regarded as a criterion to evaluate the high–flow simulation efficiency and is of no use for the evaluation of low–flow simulations.
- NSE^*_{sqrtQ} provides more balanced information as the errors are more equally distributed on high– and low–flow parts, in agreement with the study by Oudin et al. (2006).
- $NSE^*_{\ln Q}$ is also intermediate like NSE^*_{rQ} and IA^*_{rel} but with greater emphasis on low flows. However, the sensitivity of $NSE^*_{\ln Q}$ to high flows noted by Krause et al. (2005) is confirmed here.

The *NSE* criterion on inverse flows (NSE^*_{iQ}) appears to be the most useful criterion to evaluate the very low flows among the criteria calculated over the total period, because it shows no sensitivity to high flows, contrary to $NSE^*_{\ln Q}$. This transformation is more relevant than the logarithmic transformation if one wishes to evaluate models in severe low–flow conditions. In comparison with NSE^*_{uT}, B^*_{uT} and $RLFD^*_{uT}$, it avoids the often arbitrary definition of a low–flow threshold: the most widely used low flows range from Q_{70} to Q_{99} (Smakhtin, 2001), but may depend on local conditions and regime type. In addition, it avoids obtaining the often negative values found with NSE^*_{uT}, which are due to an overly demanding benchmark model and are difficult to interpret.

Therefore, we would suggest that NSE^*_{iQ} is a good candidate (the most adequate of the criteria analysed here) to evaluate the low–flow simulation efficiency of a hydrological model. We

wish to emphasize that even if at present $NSE^*_{\ln Q}$ is the most commonly preferred criterion to evaluate low flows, its value is influenced by a wider range of flows.

3.5.5. What should be done with zero flows when calculating NSE^*_{iQ} ?

As mentioned in section 3.4.3, the calculation of the NSE on inverse flows raises a problem when the catchment is non–perennial or ephemeral, i.e. with some days with zero flows. To deal with zero–flow values, it was necessary to add a small constant, ε, to the flow values before calculating their transforms. However, the choice of ε may have an impact on the relative level of performance attributed to the model.

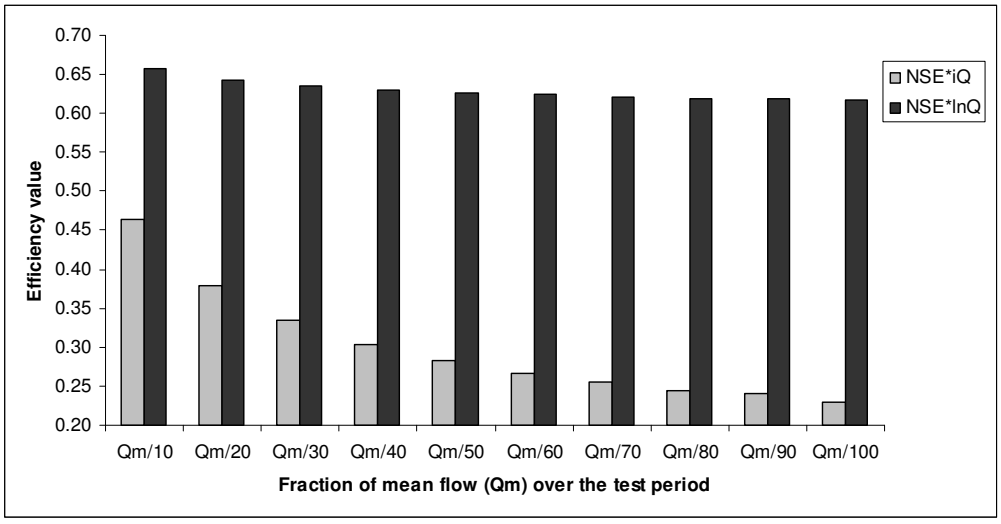


Figure 3.9: Change in the mean value of NSE^*_{iQ} and $NSE^*_{\ln Q}$ obtained by the GR4J model in validation over the catchment set with different values of ε (fractions of Qm from Qm/10 to Qm/100)

As shown in **Figure 3.9** where values of ε ranging from one–tenth to one–hundredth of the mean flow were considered, the mean model efficiency on the catchment set varies significantly. ε should not be too large, or else it would tend to artificially enhance the relative level of model efficiency. Therefore the value of one–hundredth of the mean flow could be advised for ε, as it seems that this value corresponds to a plateau in the efficiency values, i.e. with low sensitivity of model efficiencies.

3.5.6. Power transformation and criteria

The results in section 3.5.3 clearly indicate that transformations on flows with power (λ) values lower than 0 give greater emphasis to low flows. A value of λ equal to -1 (i.e. the inverse transformation which we ultimately recommend) was shown to stress the approximately 20% of lowest flows on average on the catchment set, as shown in **Figure 3.8** . However, the box plots in this figure indicate that this λ value does not emphasize the quality of model fit on exactly the same parts of the hydrographs for all catchments. Indeed, it may depend on regime characteristics, flow variability or model performance on low flows. We tried to quantify the potential variability of λ values from catchment to catchment because we wished to emphasize model errors on the same parts of the hydrograph. For each catchment, we selected the λ value for which 80% of the largest errors were concentrated on the 20% lowest flows at most. **Figure 3.10** compares the λ values obtained on the two test periods for the two models. The range of λ values obtained for the two models is similar but also quite wide, which confirms that a single λ value will not emphasize the same flow range on different catchments. The scatter on the graphs also shows a quite strong dependency of λ values on the test period. A detailed analysis of the catchments where the λ value is very unstable shows that the criterion is very sensitive to a few time steps among the lowest flows in one period where the model makes large errors. **Figure 3.11** compares the λ values obtained by the two models. Interestingly, there is quite good agreement between them, though the scatter is quite wide for decreasing λ values. This is probably the indication that models behave similarly in low–flow conditions, but that their behaviour on extreme low flows is different. These results indicate a dependency on the characteristics of the test period and the model used.

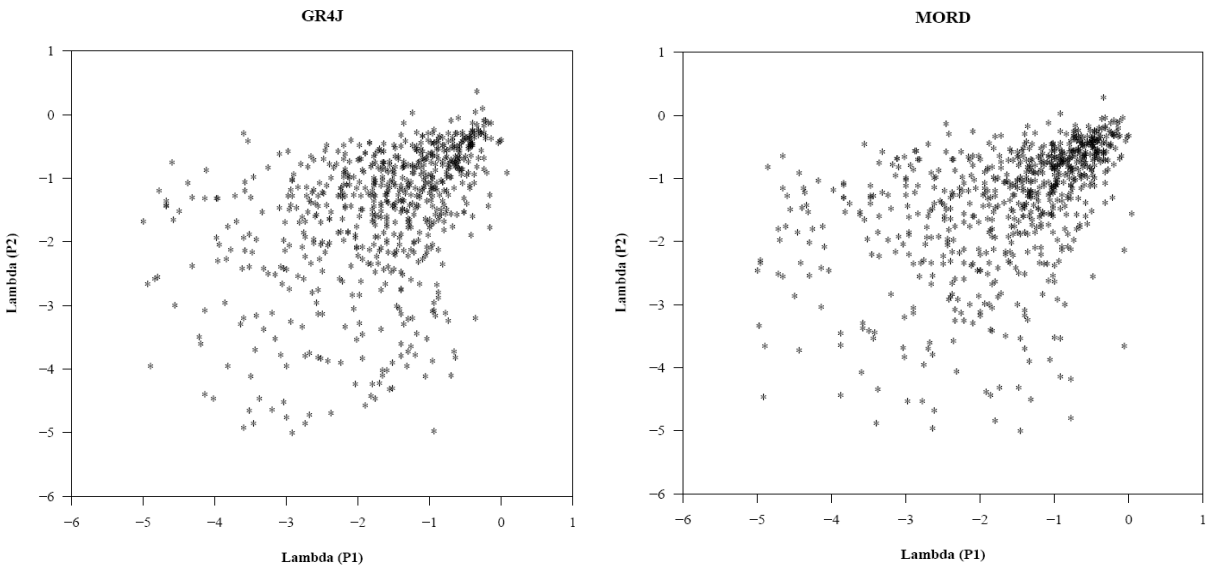


Figure 3.10: Comparison of λ values obtained by GR4J (left) and MORD (right) models for the two test periods (P1 and P2)

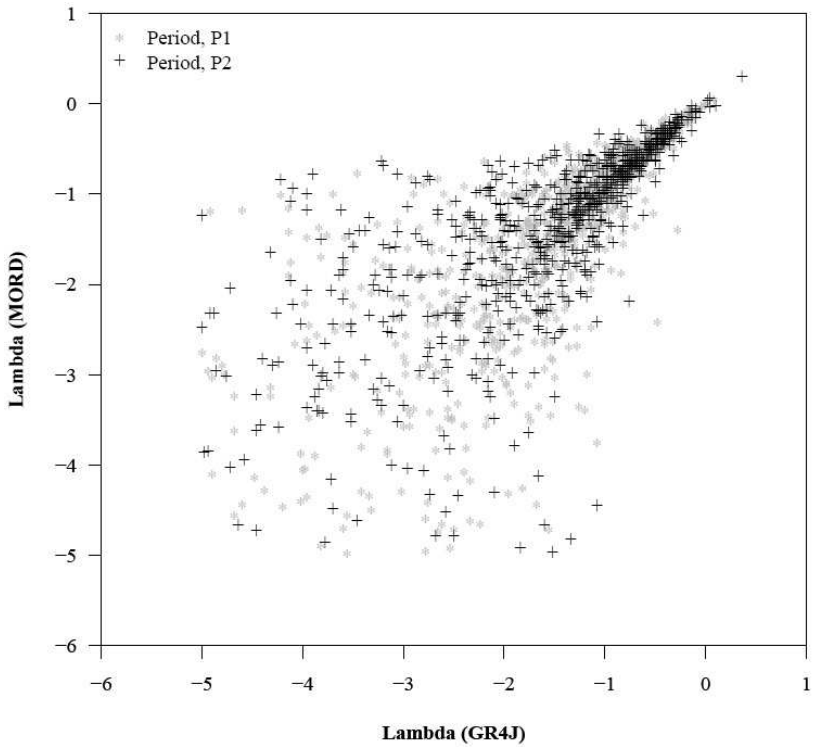


Figure 3.11: Comparison of λ values obtained by the GR4J and MORD models

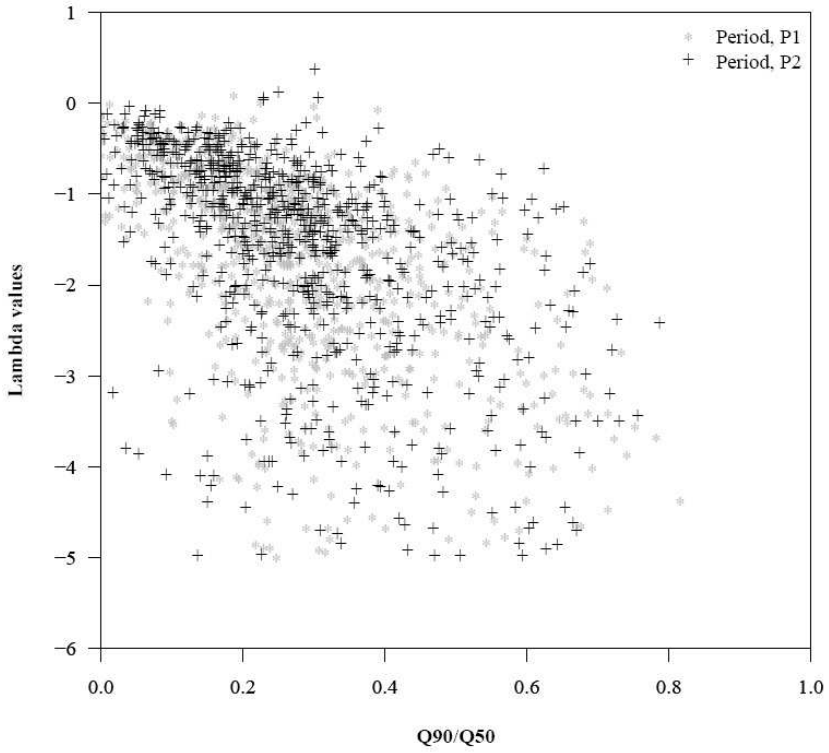


Figure 3.12: Scatter plot of Q_{90}/Q_{50} vs. lambda values (λ) for the two test periods for the GR4J model

The dependency of λ on flow characteristics is illustrated for the GR4J model in **Figure 3.12**. It shows the relationship between λ and the Q_{90}/Q_{50} descriptors that quantify the variability of flows. Although the scatter is large, the value of λ seems to depend on this descriptor: the smoother the catchment response (i.e. the larger Q_{90}/Q_{50}), the lower the λ value.

All these results indicate that the interpretation of the Nash–Sutcliffe criterion based on inverse transformation does require experience and that one must be cautious when comparing values obtained on different periods and/or catchments; however, this was already the case with the original Nash–Sutcliffe criterion (see the comments by Perrin et al. (2006)).

3.6. Conclusions

This study aimed at identifying adequate criteria for evaluating the simulation of low flows using hydrological models. The behaviour of nine criteria was analysed on a set of 940 catchments. We used a master box plot representation to locate which part of the hydrograph contributes most of the error in these criteria, on average on the total catchment set. The

detailed analysis of error distributions shows that the *NSE* calculated on inverse flows is better suited for the evaluation in very low–flow conditions than the classically used *NSE* on logarithm flows, as it does not show sensitivity to high–flow values. It tends to focus on the 20% lowest flows on average, like the *NSE* on natural flows tends to focus on the 20% largest flows. Therefore, we recommend using this criterion calculated on inverse flow values for low–flow studies. However, it should be noted that the part of the flow duration curve that will cause most of the errors when using this transformation may differ from period to period and catchment to catchment, depending on model suitability and flow regime characteristics. This encourages to be cautious on the interpretation of this criterion when changing test conditions.

Chapter 4. Development of an improved lumped model for low-flow simulation

This chapter contributes to the following article:

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, C., 2011. A downward structural sensitivity analysis of hydrological models to improve low-flow simulation. Journal of Hydrology 411: 66-76.

The edited version of this article is given in Appendix E. This chapter presents a brief description of the overall step-by-step analysis conducted to develop the improved model version. Compared to the article, the results of some complementary tests are shown here.

4.1.Introduction

There are various rainfall-runoff (R-R) models available to simulate streamflows, but limited studies are carried out in stream low flows as compared to the peak-flow events. Therefore, the selection of models will probably become a difficult task when end-users need to make simulations over the whole flow range. When working on many catchments at the same time, as it is often the case in operational conditions, using individual models in each catchment is not a practical solution. For example, in our study, we used a set of 1000 catchments and it is not practically possible to use 1000 models based on catchment characteristics and flow conditions. Therefore, the present study tried to develop a more general model structure for low-flow simulation which we can also be used to simulate other flow regimes with better efficiencies on a large set of data irrespective of the catchment and climatic conditions. We started from existing models and tested modified version to try to improve their efficiency. Here our test catchments are gauged stations and for low-flow modelling in ungauged basins, please refer to the recent study by Lang et al. (2010).

The following section briefly explains the selection of base models in order to start the structural modifications.

4.2.Selection of basic model structures

Lumped models are generally simple, have limited input requirements and acceptable performance compared to more complex models, which constitute good enough reasons to use them (Chiew, 2010). They represent the catchment as a single unit and hence their parameters and variables represent average effective values over the entire catchment. In our study, a set of nine lumped hydrological models (see **Table 4.1**) were used to simulate river low flows. Some previous studies showed that these selected models are quite reliable to simulate the

rainfall-runoff transformation. As an example, Mathevet et al. (2004) used some of the listed models to develop a modified hourly version of the GR4J rainfall-runoff model structure.

Some of the listed models in **Table 4.1** are not originally lumped, but we used here lumped versions. Therefore all tested models use the same inputs. The GR4J and MORD models were already presented in Chapter 3. The details on models GR5J and TOPM are available in the coming sections of this chapter. The structural layout and the mathematical formulations of the models used in this thesis are presented in Appendix C.

Table 4.1: Selected model structures and their number of parameters

Model acronym	Reference describing the original version	Number of free parameters
HBV0	(Bergström and Forsman, 1973)	9
IHAC	(Jakeman et al., 1990)	6
MOHY	(Fortin and Turcotte, 2007)	7
MORD	(Garçon, 1999)	6
TOPM	(Beven and Kirby, 1979)	8
GR4J	(Perrin, 2000)	4
GR5J	(Le Moine, 2008)	5
GAR1	(Thiéry, 2009)	6
GAR2	(Thiéry, 2009)	8

As a first step to develop an improved version, we analysed the performance level of these nine models on our catchment set distributed over France (see chapter 2). The objective was to evaluate their strengths and weaknesses and the differences between models. Best performing model structures in these tests could serve as a basis for further model improvements.

4.3.Model testing

The model testing scheme we adopted consists in a split-sample test approach, as already described in Chapter 3. The parameters were calibrated using the NSE^*_{sqrtQ} criterion presented in Chapter 3 which is a good compromise between high and low flows for model calibration. Here, as we focus on low flows, we could have chosen an objective function putting more

weight on low flows. However, it would have been to the detriment of the simulation of high flows. So we preferred to keep this objective function to obtain a general model. This did not prevent us from assessing the model (in validation) over a wider range of criteria.

As discussed in the previous chapter, here we used the NSE^*_Q, NSE^*_{lnQ} and NSE_{iQ} to evaluate model performances in high and low-flow conditions. Performances were evaluated in validation.

4.4.Evaluation of the selected model structures

Results of model testing are presented in **Table 4.2**.

Table 4.2: Average efficiency values in validation for the lumped models for various criteria (criteria on Q and iQ put more emphasis on high flows and low flows, respectively, and criterion on lnQ is intermediate).

Model acronym	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
HBV0	0.546	0.559	0.156
IHAC	0.528	0.556	0.196
MOHY	0.493	0.554	0.229
MORD	0.603	0.616	0.302
TOPM	0.574	0.584	0.216
GR4J	0.621	0.617	0.230
GR5J	**0.629**	**0.648**	**0.346**
GAR1	0.378	0.376	0.164
GAR2	0.538	0.489	0.194

Out of the nine models, GR5J, GR4J, TOPM and MORD models show better efficiency values irrespective of their structural complexities and number of parameters. Some interesting features of these models can be noticed:

- GR4J and GR5J include a rough representation of groundwater exchanges.
- TOPM has an exponential store which can better represent the base flows.

- in MORD model, the low flows are mostly produced by a single routing store since the intermediate routing store is often empty.

The type of routing stores and the way they are organized as well as the presence of a non-conservative groundwater exchange function seem to be key aspects of the low flow simulation.

In the next step, we selected these four models as the base model structures to start the structural modifications.

4.5.Development of an improved model structure from the base models

A number of modifications were made on the base structures to try to improve their efficiency. The structural modifications include integration of additional routing stores, integration or modification in groundwater exchange functions, among others. The details of model development stages are presented in the following sections. Here we presented the structures of the remaining base models (GR5J and TOPM) in **Figure 4.1** . We are not going into the details of the model structures and their parameters, and for more on these model structures, please refer to the corresponding references provided in **Table 4.1**.

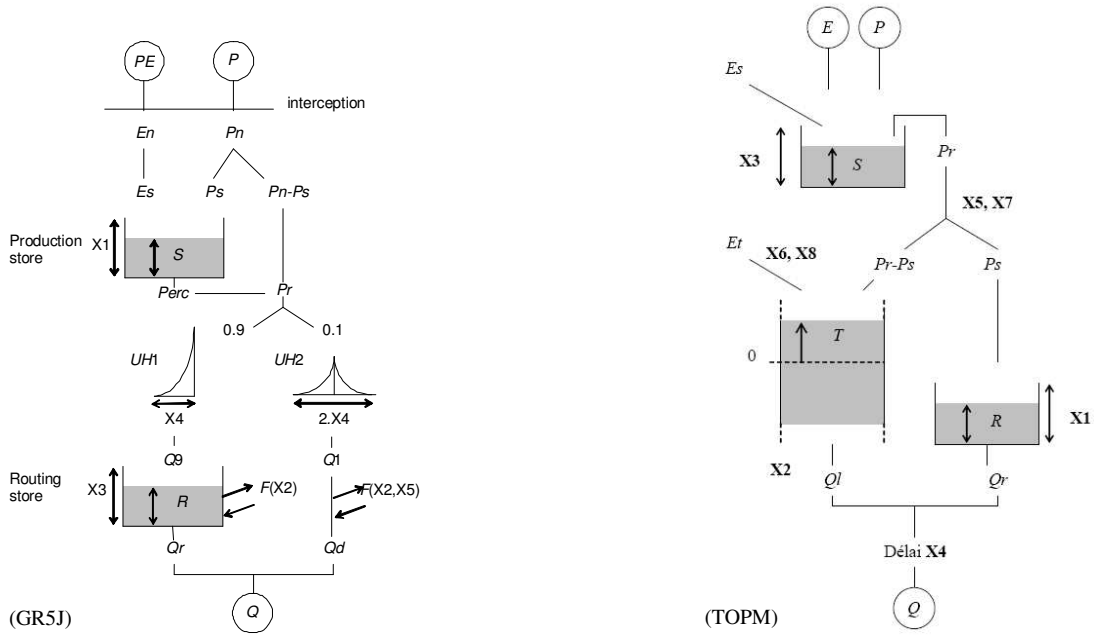


Figure 4.1: Structural layout of the base models

The analysis of the model structures shows that TOPM and MORD (MORD has a rainfall adjustment factor which can be considered as volume adjustment factor) do not have groundwater exchange function in the production or routing module. As groundwater is the major source contributing to river flows during the dry period, the first set of modification corresponds the addition of water exchange function in the production or routing module of TOPM and MORD models. The GR series has the simplest structure compared to other models and hence more trials are possible with their present versions. Along with the groundwater exchange functions, additional stores in the routing module can also enhance model performance, especially in the case of delayed flows (Wagener et al., 2004; Mathevet, 2005). If the exchanges are limited, the model may perfectly simulate base-flow without exchange functions by considering the routing module. Hence, the modifications in GR4J and GR5J include the addition of new routing stores along with the modifications in their existing groundwater exchange functions.

4.5.1. Integration of groundwater exchange function (*F*)

Groundwater (GW) is the main source for river flows during prolonged dry periods. Hence the recharge and release of groundwater is one of the important processes to consider for simulating low flows. Literature shows that there are studies conducted on how to account the groundwater exchange into an R-R model (e.g. Fenicia et al., 2006). Some authors considered only the flow towards the stream using a specific groundwater reservoir (Pointet, 2004; Davison and van der Kamp, 2008), but in the present analysis, we considered an exchange function that can account for both recharge and discharge from the groundwater reservoir, as proposed by Le Moine (2008). During the course of this research, we evaluated the sensitivity of low–flow simulation to various formulations of the existing GW exchange functions.

4.5.1.1. Modification of groundwater exchange function (*F*) in the GR4J and GR5J models

The production function that controls water balance in the GR4J model structure consists of a soil moisture accounting reservoir and a conceptual water exchange function (*F*), expressed as:

$$F = X2.\left(\frac{R}{X3}\right)^{3.5} \qquad \textbf{Eq. 4.1}$$

in which $X2$ (mm) is the "groundwater" exchange coefficient and R and $X3$ (mm) are the water level and the capacity of the routing store, respectively. $X2$ can be positive or negative, meaning that the water exchange function can simulate imports or exports of water with the underground (i.e. connections with deep aquifers or surrounding catchments). Note that $X3$ is also used to parameterize the outflow from the routing store, which limits the interactions that would unavoidably exist between $X2$ and $X3$ if Eq. 4.1 was used alone. The routing part of the structure consists in two flow components routed by two unit hydrographs and a non–linear store. The latter is mainly responsible for low–flow simulations, along with leakage (percolation) from the SMA store. The groundwater exchange term F is added to the two flow components of the routing module.

Mathevet (2005) tested several modified versions of this model, especially by increasing the complexity of the routing part of the model and adding stores in parallel to the existing one. His tests, made at the hourly time step, showed limited sensitivity of model results, but the criteria he used focused more on high flows.

Following this work, Le Moine (2008) investigated the interactions between surface and groundwater and evaluated several modifications of the GR4J model to better account for these exchanges. These included different water exchange functions and the addition of a new store representing long–term memory. He proposed a five-parameter version of the model (GR5J) in which the groundwater exchange function has been modified to:

$$F = X2.\left(\frac{R}{X3} - X5\right) \qquad \textbf{Eq. 4.2}$$

where $X5$ is a dimensionless threshold parameter. It allows a change in the direction of the groundwater exchange within the year depending on the water level R in the routing store compared to this threshold. This model has shown significant performance improvement over the GR4J model, especially in low–flow conditions. It can be noted that the time–varying term F is only a very crude way to simulate groundwater–surface water connections. $X5$ can be seen as the external, quasi–stationary potential of the groundwater system and F is a "restoring flux" acting like a spring device with constant $X2$. Usually, $X2$ is negative: the more $R/X3$ departs from $X5$, the more intense the flux is, which tends to restore its value to $X5$.

During the course of our research, we evaluated the sensitivity of low-flow simulation to various formulations of the existing F in the GR4J and GR5J models (Eq. 4.1 and Eq. 4.2).

We developed several model versions that differ only by their groundwater exchange formulation were tested. Due to the higher performance of GR5J, this chapter mainly discusses the modified versions of the GR5J model (the best among the seven models) and the other model versions are discussed in the Appendix D. **Table 4.3** lists the versions of the GR5J model with respect to the modifications in the groundwater exchange functions.

Table 4.3: Modified versions of the GR5J model in terms of groundwater exchange function

Model version	Characteristics of the groundwater exchange function		Number of routing stores	Number of free parameters
	Eq. 4. 2	Others		
M1	✓	Exchange dependent on SMA store	1	5
M2	✓	Splitting coefficient applied to F	1	5
M3		Formulation of Nascimento (1995)	1	4

4.5.1.2. Integration of F into the TOPM

The structure of TOPM shows that it has no function representing groundwater exchange. Hence, we added a new parameter into the TOPM that represents groundwater exchange function. In this version, the introduced groundwater exchange function is the same way as in Eq. 4.1. The new version has the same structure as TOPM, the only change is the incorporation of a groundwater exchange parameter, and therefore the new version has 9 parameters.

4.5.1.3. Modification in the MORD model

Mathevet (2005) derived some versions of the original MORDOR model. Out of the several versions, MORD (with six parameters) shows better simulation efficiency. Our first evaluation results also suggest the applicability of this model in simulating stream low flows. The structural analysis reveals that one simple option to improve the model performance is to

consider the groundwater exchange function. Therefore the introduction of *F* into MORD transformed the original structure to a new one with seven parameters.

We did not integrate several modifications in TOPM and MORD in terms of the routing module as they already have complex model structures with several parameters. The larger number of parameters increases the model complexity (Perrin et al., 2003), hence we tried to keep simple model structures with avoid overparameterization.

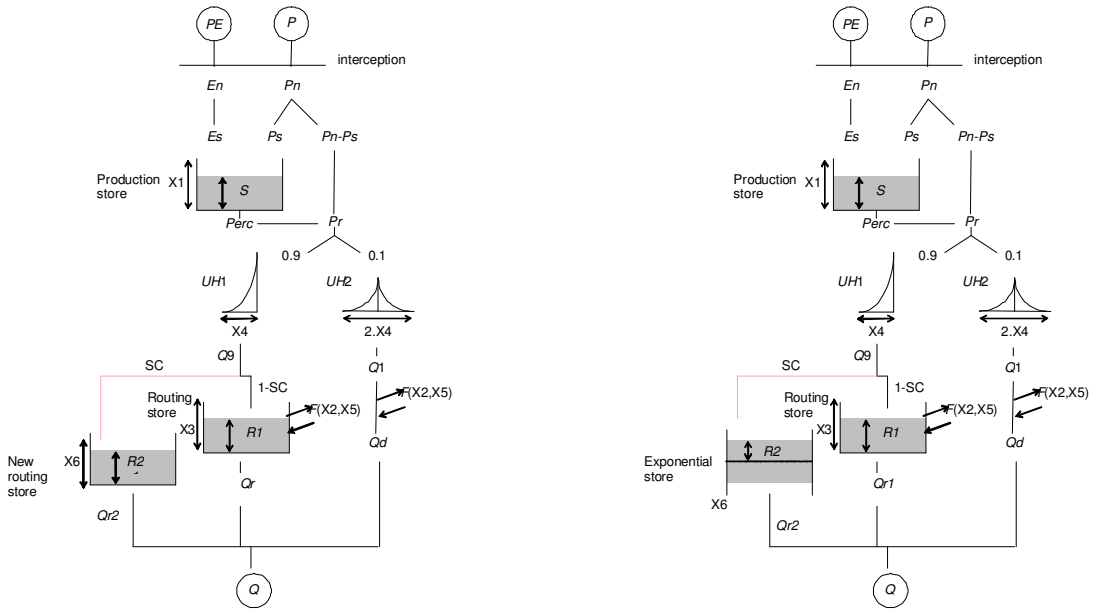
4.5.2. Addition of new stores

Because of complex processes during the transformation of rainfall into runoff at the outlet of a catchment, additional stores may improve model performance (Wagener et al., 2004; Mathevet, 2005). The additional stores can improve the simulation efficiency especially during low flows, as they may help simulating delayed runoff. In the present analyses, additional stores were introduced to the GR4J and GR5J models, the most simple and best performing models among the selected structures.

New versions of the GR4J and GR5J models were derived by the addition of non-linear routing stores (quadratic as well as exponential stores). In both models, 90% (**Figure 4.1**) of the total effective rainfall is diverting to the routing store. Hence, one of the factors to be considered during the addition of new stores in parallel to the initial store is the splitting coefficient of effective rainfall (*SC*) into the existing store. Mathevet (2005) conducted some trials to divide the effective rainfall among the existing and new stores of the GR4J model. In the present study, the *SC* value was selected after successive trials and the same value was used in all catchments. But the suitability of *SC* changes from one catchment to another and hence the single *SC* value may not be suitable for all the catchments. Therefore, the *SC* was also optimized in one of the versions of GR5J model for individual catchments. Some of the versions of the GR5J model are listed in **Table 4.4**. GR4J versions are listed in Appendix D. **Figure 4.2** shows the layout of the structures of two versions of the GR5J model with the addition of power-5 routing store (version M4) and exponential store (M8) respectively.

Table 4.4: Modified versions of the GR5J model and their main characteristics

Model version	Characteristics of the groundwater exchange function		Characteristics of the additional routing stores					Number of routing stores	Number of free parameters
	Eq. 4.2	Others	Power-2 store	Power-5 store	Exponential store	Added in parallel	Added in series		
M4	✓			✓		✓		2	7
M5	✓			✓		✓		2	6
M6	✓			✓			✓	2	6
M7	✓		✓			✓		2	6
M8	✓				✓	✓		2	6
M9	✓			✓	✓	✓		3	7
M10	✓				✓ ✓	✓		3	7
M11	✓		✓		✓	✓		3	7
M12	✓				✓	✓		2	6



M4 M8

Figure 4.2: Layout of two versions of the GR5J model with different additional stores

4.6. Results and discussion

Several model versions were discussed in the previous section. An overview on the performance of these models shows that the GR5J model versions are better than the others. Hence, here we present the results of the GR5J model versions because of their better performance. Then the best version was selected. The other model versions derived from other model structures (GR4J, TOPM and MORD versions) are presented in Appendix D. As the number of modifications is almost infinite, here we chose to present only a few of them to answer a number of simple questions that may arise when discussing the model's structure. Although these questions are sometimes interrelated, they are presented in sequence for the sake of clarity. The following section is adopted from the published article which is presented in Appendix E).

4.6.1. Can the existing groundwater exchange term in GR5J be improved?

Figure 4.3 shows the distribution of performances of the selected versions. It indicates significant sensitivity of model results to this function, which corroborates the findings of Le Moine et al. (2007). The existing exchange function in the base model appears to be the best performing one. This confirms the robustness of the solution proposed by Le Moine (2008). In the following versions, we will stick to this formulation.

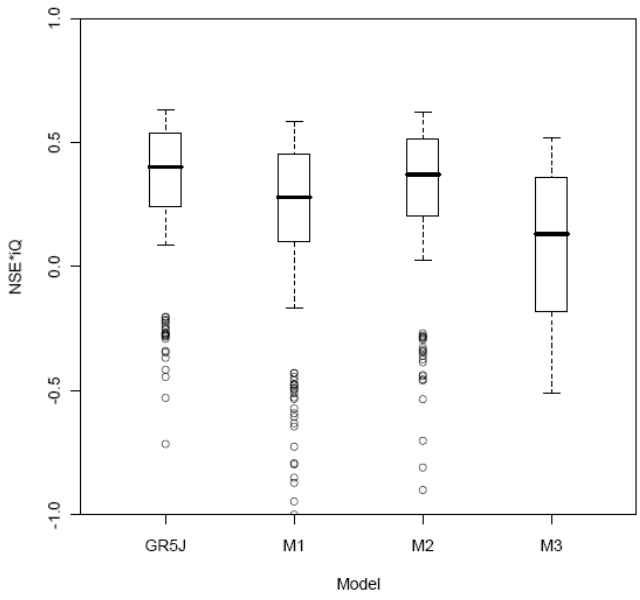


Figure 4.3: Box plots of NSE*iQ values obtained in validation over the catchment set by GR5J and three model versions with modified groundwater exchange functions (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles respectively

4.6.2. Should the volumetric splitting between flow components be adapted to each catchment?

Here we tested two model versions with two stores in parallel, one in which we first optimised *SC* (version M4 in **Table 4.5**) on each catchment and the other in which we fixed *SC* to 0.4 (version M5).

Table 4.5: Mean model performance for versions M4 and M5

	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}	Number of free parameters
M4	0.637	0.661	0.369	7
M5	0.634	0.659	0.365	6

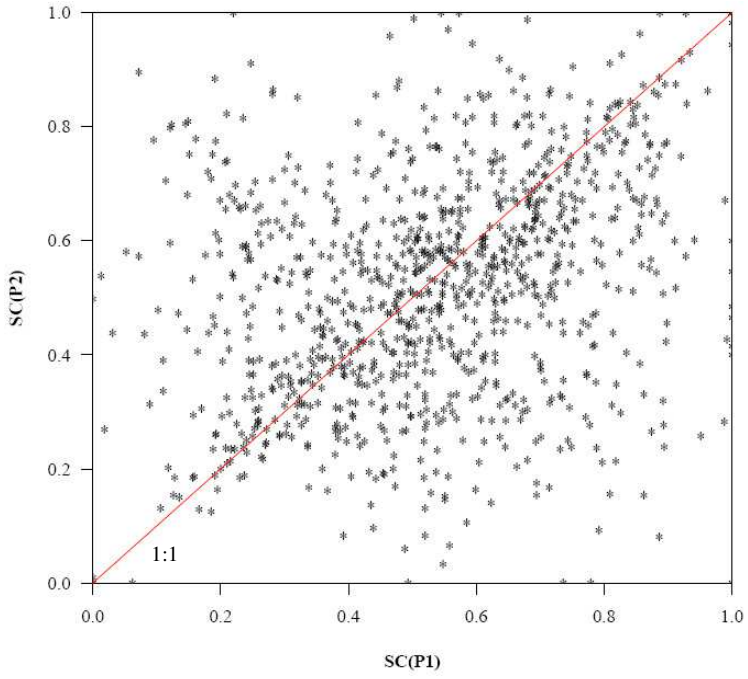


Figure 4.4: Comparison of the splitting coefficient (SC) values obtained on the two calibration periods (P1 and P2) in the M4 model version

The parameter analysis in M4 shows that *SC* values are very sensitive to calibration conditions. This is illustrated in **Figure 4.4** that clearly shows the spread of *SC* values obtained on the two test periods (P1 and P2). This shows that this parameter is poorly defined and it will be difficult to relate it to catchment characteristics. The limited difference in model efficiency between M4 and M5 versions (see **Table 4.5**) shows that *SC* can be fixed without

significant efficiency loss. In the coming sections, we will test versions considering only fixed splitting coefficients.

4.6.3. Should a new store be added in series or in parallel?

Existing models propose a variety of conceptualizations for flow routing, using stores either in series and/or in parallel. Jakeman et al. (1990) discussed this issue in the case of the IHACRES model, in which the routing module is made of linear stores. This model structure can be adapted to have several stores in series or in parallel. Despite this flexibility, the authors indicate that in most cases, having two stores in parallel is the most efficient configuration.

Here we analysed the sensitivity of model performance to the arrangement of routing stores, be they added in parallel or in series to the existing store. Two versions were tested, in which a new store similar to the existing one was added in parallel (version M5) or in series (version M6). **Table 4.6** shows the mean performance of these two versions. The M5 version gets higher efficiency values than M6. We also tried to add one more routing store in parallel, to have a third routed flow component (versions M9 to M11 in **Table 4.4**). Results in **Table 4.7** indicate that improvements for low-flow simulation are not significant, which means that this extra complexity is not warranted by data.

Table 4.6: Mean efficiency values for versions M5 and M6

	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
M5	0.634	0.659	0.365
M6	0.625	0.641	0.31

Table 4.7: Mean model performance for versions M8–M11 (multiple routing stores) and mean relative performance RE* with reference to M8 over the catchment set

	M8	**M9**	**M10**	**M11**
NSE^*_{iQ}	0.384	0.384	0.385	0.384
$RE^*(\%)$	-	0.0	0.1	0.0

This confirms that the best compromise on average is to have two stores in parallel. The series arrangement did not prove to be an efficient option. Therefore, following Jakeman et al. (1990), we suggest that an increased complexity of the routing part of the model should be made by considering two independent flow components. This is a solution to have more varied flow dynamics.

4.6.4. Does the formulation of routing stores matter?

Here we tried to identify the best formulation of routing stores for which the model shows higher efficiency values. There is a variety of possible formulations of routing stores, ranging from linear to non-linear stores, e.g. power law or exponential stores (see Michel et al. (2003) for a correct formulation of this store). In previous work (Edijatno and Michel, 1989; Edijatno et al., 1999), a power-5 non-linear routing store was found as the most efficient. When adding a new store in parallel, another formulation may be interesting to introduce a variety of behaviours in the flow components. Various formulations were tested, among which we give the examples of versions M5, M7 and M8 in **Table 4.4**. **Figure 4.5** shows the corresponding box plots of efficiency values over the catchment set. The box plot of the model version M8 (with an additional exponential store) indicates a higher efficiency. The exponential store is known to be an efficient tool to simulate long recession spells (see Michel et al., (2003)).

As suggested by Le Moine (2008), we also analysed the performance of M8 by removing the direct flow component (version M12). Indeed, the introduction of a new flow component may make this direct flow component not necessary. But results are slightly lower than M8 (**Table 4.8**), so we chose to keep this direct flow in the model. Note that this direct flow does not require free parameters.

Other versions were tested and several gave similar although slightly lower results. Thus, in all our tests, the M8 version revealed to be the most satisfactory one and we chose to select it as a good candidate for providing improved low-flow simulation. We will call it GR6J hereafter (version of GR5J with six parameters).

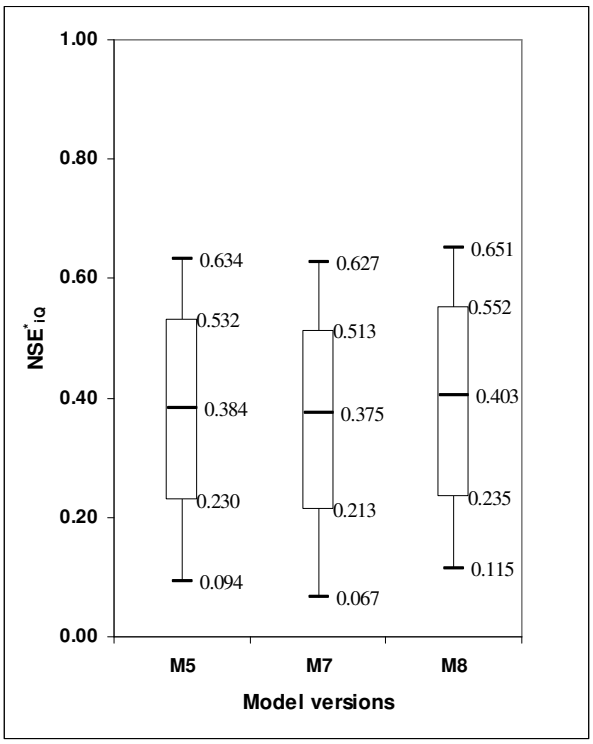


Figure 4.5: Box plots of NSE*iQ values obtained in validation by model versions having different formulations of the additional routing store (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90

Table 4.8: Mean efficiency values of M8 vs M12

	*NSE*Q*	*NSE*lnQ*	*NSE*iQ*
M8	0.634	0.662	0.383
M12	0.631	0.657	0.378

4.6.5. Comparing the results of GR5J and GR6J

This section quantifies the differences in the model's behaviour and performance between the GR5J and GR6J versions in greater detail.

The percentage improvement in the NSE^*_{iQ} values of GR6J are calculated in terms of relative efficiency values (RE^*). The RE^* values of the GR6J model are calculated with reference to the GR5J model. **Table 4.9** shows the average relative performance for the three NSE^* criteria. It indicates a significant improvement in the simulation of low flows without losing

efficiency on high flows. When looking at the criterion on inverse flows, RE^* is positive on a majority of catchments, which means that the additional store improves this set of catchments.

Table 4.9: Mean performance of GR5J and GR6J and relative performance of GR6J with reference to GR5J over the catchment set for various criteria (criteria on Q and iQ put more emphasis on floods and low flows, respectively). The results were obtained in validation

Model	GR5J	GR6J	RE^*(%)
NSE^*_Q	0.629	0.634	0.83
NSE^*_{lnQ}	0.648	0.662	2.45
NSE^*_{iQ}	0.346	0.383	4.26

4.6.6. Illustration of model's results

It is always difficult to select representative examples when working on a large catchment set. However, we wished to illustrate the model's results on a few case studies, by providing simulated hydrographs. We considered streamflow values for a period of 1 year for the four sample catchments. We chose the year 2003, which was one of the driest years over the past decade in France. **Figure 4.6** shows the observed flow series and the flow series simulated by the two models, GR5J and GR6J, for the three catchments. Note that the graphs use logarithmic scales to emphasize differences in low flows.

In catchment A, the performance of the GR6J model is significantly better than GR5J's performance on the very low-flow. We can also see a better performance of the GR6J model in catchments B and K. The GR5J model tends to underestimate these flows, especially in the case of catchment B. In catchment J, the two models give similar results and also similar dynamics in low-flow conditions, which indicates that the introduction of the new store is neutral on this catchment.

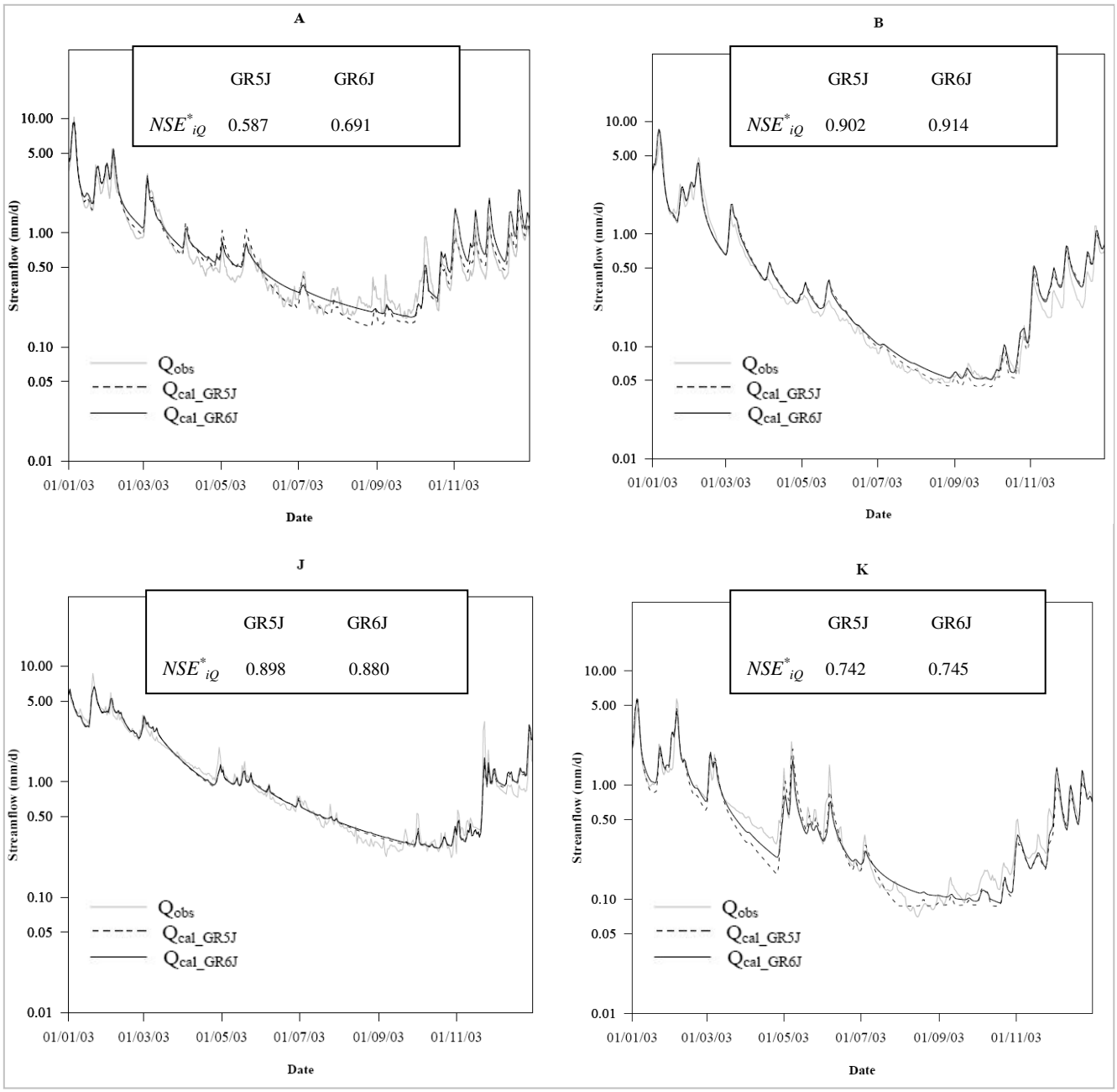


Figure 4.6: Illustration of the hydrographs simulated by the GR5J and GR6J models, with corresponding NSE*iQ efficiency values

4.6.7. Parameter stability and identifiability

Figure 4.7 compares the stability of parameters of the GR5J and GR6J models obtained on the two calibration periods (P1 and P2). In general, there is a quite good agreement between periods, with the parameters showing good stability, which is a desirable property. However, the threshold values for groundwater exchange ($X5$) change significantly for a number of

catchments in the two models. The scatter seems a bit lower with the GR6J model. Interestingly, the reverse is observed for the capacity of the SMA store ($X1$), for which the spread seems greater in the case of GR6J. This means that the introduction of the new routing store impacted the rest of the model's structure to some extent. For the sixth parameter in GR6J which corresponds to the exponential store, there is quite good agreement between periods, showing that this parameter is clearly identifiable.

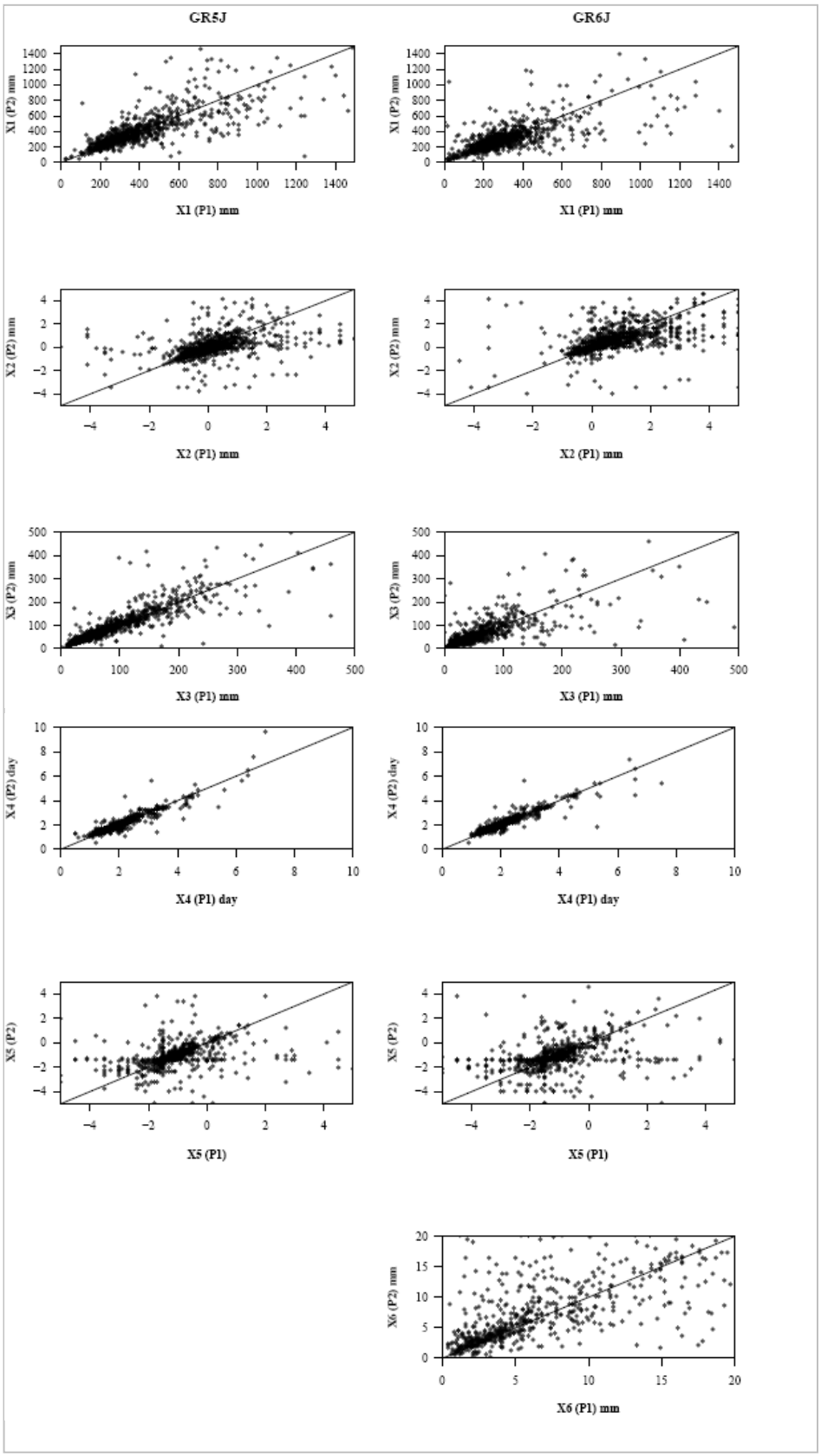


Figure 4.7: Comparison of the parameter values obtained on the two calibration periods P1 and P2 for the GR5J and GR6J models (1:1 line on each graph)

4.1. Conclusions

Improving the low-flow simulation ability without impacting the high-flow simulation ability was one of our objectives in this study. We chose to proceed by trial and error, as recommended by Nash and Sutcliffe (1970) and Michel et al. (2006). Working on a large set of catchments proved to be a good way to prevent undue complexity in the proposed modified versions of the model. Here we started from the simple GR4J model and some of the most widely used models, and tested a number of modified versions, some having higher performance values compared to the initial model structures. The model's performance was not equally sensitive to all the tested modifications. Among the modifications that proved the most robust, the addition of an exponential routing store, in parallel to the existing routing store in the GR5J model, showed improvement in low flows on average, still remaining efficient in high-flow conditions. The complexity added by this modification (an additional free parameter) seems to be warranted by the model's results, as well as by the comparison to other existing models. In spite of this improvement, it is not possible to say that we improved the physical realism of this model, since the initial intention was not to explicitly represent the physical mechanisms. Instead this study focused more on identifying the main features of the rainfall-runoff transformation at the catchment scale and improving the model's predictive power. However, we are convinced that the significant improvement gained on our large data set could not occur by chance, and that using this improved model version provides a better representation of the catchment's hydrological behaviour.

Last, let us note that the level of performance in low-flow conditions seems to remain quite low. This may be for several reasons, including structural model errors, data quality or artificial influences. Nonetheless, it shows that specific research should be continued to improve the efficiency of hydrological models for low-flow simulation, a domain that was probably overlooked in the past.

Part III: Stream low-flow forecasting

In part II, we proposed an improved model version to simulate stream low flows. We can expect that this improvement gained in simulation mode can also bring improvement in forecasting mode. Hence as a continuation, this part focuses on issuing long-term low-flow forecasts. Part III is divided into two chapters:

- Chapter 5 discusses the tested forecasting methodology, the evaluation of the methodology in terms of different criteria (deterministic as well as probabilistic) and the options to improve the forecast quality in terms of bias corrections and updating techniques. It also provides an assessment of the maximum possible lead-time for the tested catchments based on the usefulness of the forecasts.
- Chapter 6 discusses the possible improvements in the forecast quality with the integration of the influence of artificial reservoirs into the R-R model. Here we tested the methodology proposed by Payan et al. (2008) to account for the influence of reservoirs.

Chapter 5. Low-flow forecasting: implementation and diagnosis of a long-term ensemble forecasting approach

5.1. Introduction

Low-flow forecasting is one of the emerging issues in hydrology due to the escalating demand of water in dry periods. The absolute flow variability in a river during dry period is much smaller than during flood events, due to lower water fluxes and slower dynamics of processes involved. Therefore, the lead-time for performing low-flow forecasting can be generally much longer than for flood forecasting, varying from a few days or weeks to a few months. Reliable long-lead (up to a few months in advance) flow forecasts can improve the management of water resources and thereby the economy of the society (Hamlet et al., 2002; Chiew et al., 2003; Letcher et al., 2004; Karamouz and Araghinejad, 2008) along with the protection of natural ecosystems.

There is limited literature available in low-flow forecasting compared to high-flow forecasting. Most of the existing methodologies are for short lead-times and based on recession analysis (e.g. Larras, 1972; Campolo et al., 1999; Garçon et al., 1999; Stravs and Brilly, 2007; Lang et al., 2008). At the same time, Adler and Ungureanu (2006) used multi-model technique to issue mean monthly forecasts for the Oltet Basin, Romania.

Streamflow forecasts issued by relying on teleconnections may provide forecasts with longer lead-times. Teleconnections are highly relevant for streamflow forecasting in countries influenced by El Nino and Southern oscillation and is highly associated with extreme streamflow events such as flood and drought. The literature shows that several studies were already conducted on teleconnections and streamflow forecasting. But there is limited teleconnection signal in Europe (Chiew and McMahon, 2002). Also teleconnections are more interesting for seasonal forecasts (e.g. Rutten et al., 2008), which is not the focus of this study.

The limited literature on low-flow forecasting shows that additional efforts should be made on long-term low-flow forecasts since the consequences will be severe in the near future with the frequent occurrence of low flows. As discussed in the previous chapter, a general methodology should be sought to issue forecasts for various types of catchments. By keeping these points in mind, this chapter mainly discusses how the efficiency of a low-flow forecasting approach can be improved for lead times ranging from a few days to a few weeks. We tested a number of simple approaches in which we tried to incorporate information of model results at the day of issuing the forecast. The forecasting efficiency is evaluated on our data set in the French basins. As a first step, we analysed the feasibility of long-term forecasts

in the study catchments with respect to the available literature. The following section briefly describes some of the related studies in the French basins.

5.2.Literature on long-term forecasting in the French River basins

Literature shows that there are several studies conducted in France to characterize and manage the river low flows (see Moulin and Thépot, 1999; Cavitte and Moor, 2004; Coeur, 2004; Gaume, 2004; Moreau, 2004; Larue and Giret, 2006; Villocel et al., 2010) due to the demand of water sources in dry periods and due to the impact of climate variability. Therefore, low-flow forecasting is not a new issue since the first attempts to produce low-flow forecasts dates back more than 30 years ago (see e.g. Avalos Lingan, 1976; Guilbot et al., 1976; Girard, 1977, Thiéry, 1978). But there are limited studies conducted to issue long-term low-flow forecasts compared to the short-term and seasonal forecasts. Here we analyse some of the recent advances in forecasting over France.

Fifteen years ago, EDF (Garçon, 1996; Garçon et al., 1999) started to produce low-flow forecasts routinely on the basins using rainfall-runoff models. Then Perrin et al. (2001) proposed a probabilistic approach to issue long-term low-flow forecasts in two catchments in France. The approach was based on the joint use of a stochastic rainfall model and a continuous rainfall-runoff model, GR4J. They concluded that it is necessary to test the developed approach on more catchments to find its applicability. Staub (2008) presented a probabilistic approach to analyse the possibility to issue long-term forecasts (up to several weeks in advance) in France using the GR4J model with the incorporation of past rainfall scenarios. The results indicate the possibility to issue long-term forecasts in the studied catchments. Sauquet et al. (2008) also conducted a study to analyse the possibility of seasonal forecasting in the French basins using the GR4J model with the incorporation of the HBV snow module. They included the uncertainty of future rainfall scenarios with the integration of a stochastic rainfall generator. The results indicated that mid-term (up to 30 days) forecasting is possible in the studied catchments. Lang (2007) also proposed a low-flow modelling approach that was included in the PRESAGES platform aiming at forecasting flow recessions in the French part of the Rhin-Meuse basins (Lang et al., 2006).

Mathevet et al. (2007) and Mathevet et al. (2010) discuss the possibility of long-lead forecasting of streamflow especially in the Loire basin which is affected by several influences such as dams. Céron et al. (2010) recently analysed the feasibility of seasonal forecasts using

a hydro-meteorological model, SIM (SAFRAN-ISBA-MODCOU) over the French river basins. They concluded that there is a high possibility for the predictability of hydrological system in France because of the slow evolution of both soil moisture and snowpack. Soubeyroux et al. (2010) developed an approach by coupling the SIM with the forecasting seasonal system- 'Arpege Climat' of Météo-France to issue seasonal forecasts of low-flow event and their approach was solely depends on the climate information. Using the same model SIM, Singla et al. (2011) also studied the seasonal spring predictability of the hydrological system over France over the 1960-2005 period. Their results show that the predictability of hydrological variables in plains primarily depends on temperature and total precipitation, but in mountains, it mainly depends on snow cover.

5.3.Scope of the study

One of the major constraints in issuing long-term forecast is the proper integration of future changes in meteorology, especially in terms of rainfall (and to a lesser extent temperature and evapotranspiration), since future streamflows are mainly dependent on future rainfall conditions (apart from initial catchment states for past conditions). In this case, an ensemble-based approach is appropriate to issue long-term forecasts: it provides an estimate of the uncertainty associated with the future rainfall events. The present study will focus on an ensemble based approach to issue long-term low-flow forecasts over a large set of catchments using a more general model structure (GR6J). The evaluation of forecast quality will be discussed in the later section followed by the discussion of some of the simple methodologies that may improve the forecast quality. At the end of this chapter, we present the possible link between forecast lead-time and the catchment geology. The following section discusses the methodology used in this study.

5.4.Methodology

The data set, model testing and evaluation in simulation mode are already discussed in chapter 2 and in chapter 4. Hence here we only discuss the approach adopted to issue stream low-flow forecasts.

5.4.1. An ensemble approach towards low-flow forecasting

Over the last decade, the operational and research streamflow forecasting systems have been moving towards using ensemble prediction system (EPS) rather than single deterministic forecasts. HEPEX (Hydrologic Ensemble Prediction Experiment) is one of the initiatives set up to investigate the hydrologic ensemble forecasts. This was launched to develop and test procedures to produce reliable hydrologic ensemble forecasts, and to demonstrate their utility in decision making related to the water, environmental and emergency management sectors (for more on EPS, refer to Schaake et al., 2006; Seo et al., 2006; Schaake et al., 2007; Thielen et al., 2008; Żelazinski and Mierkiewics, 2009; Pagano et al., 2012).

The ensemble streamflow forecasts are produced using ensemble members of precipitation created from an analysis of historical observations or from ensemble weather forecasts. This ESP is most widely used in flood forecasting systems and an example of output of such an ESP forecast is illustrated in **Figure 5.1** (Cloke and Pappenberger, 2009). It shows an ensemble "spaghetti" hydrograph for a hindcasted flood event. The plot shows the discharge predicted for each ensemble forecast (solid lines), the observed discharge (dashed black line) and four flood discharge warning levels (horizontal dashed lines).

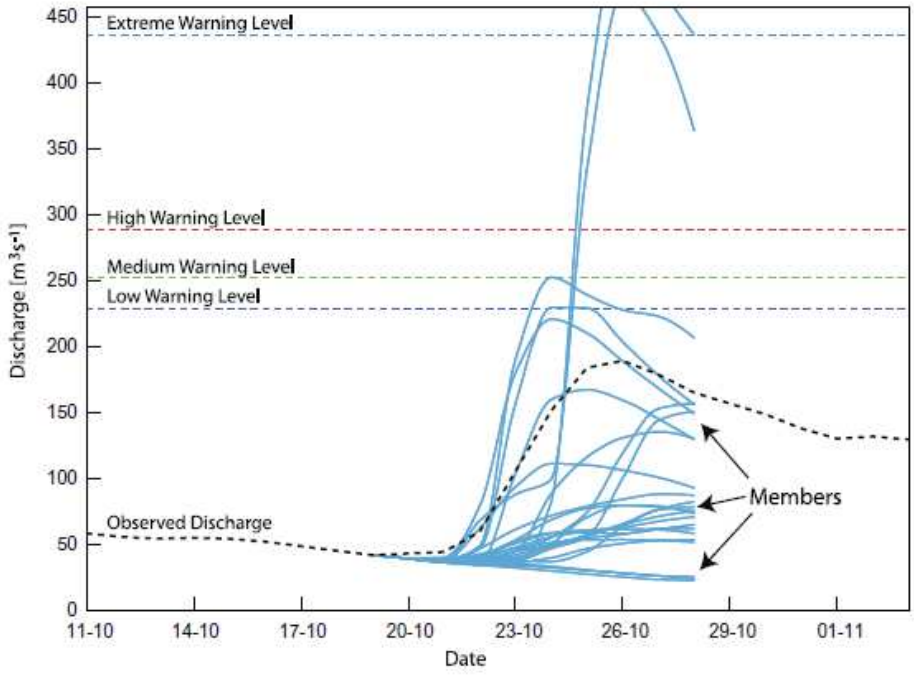


Figure 5.1: An ensemble flood forecast with different levels of warnings (Source: Cloke and Pappenberger, 2009)

The ensemble approach in flood forecasting is also much relevant in low-flow forecasting since it provides a mean to account for the uncertainty in future precipitation, which is crucial for long lead times. The low variability in absolute flow values during low-flow period also helps the ensemble system to issue forecasts with much longer lead compared to the flood events. Hence, based on this concept, we tested our methodology with an ensemble system. As mentioned in the previous literature discussion, the application of ensemble approaches for low-flow forecasting is not new, but the work done on this is much more limited than for short-term flood forecasting. Therefore, it seems necessary to better investigate the specificity of low-flow ensemble forecasting to propose possible improvements.

5.4.2. Description of the forecasting approach

A brief description of the adopted methodology is presented below.

The first step is to produce *n* rainfall scenarios. Rainfall ensembles can be created from archives of past observations, stochastic rainfall generators, and ensemble forecasts issued by meteorological models. If we have ample data at hand, the climatology can act as a good referee to make future decisions. In this study, we have a period of 36 years of meteorological data and we used these data series to create rainfall ensembles. Obviously, other ensemble sources could be adopted, that may provide better results (e.g. outputs of meteorological models) but this was not further investigated here.

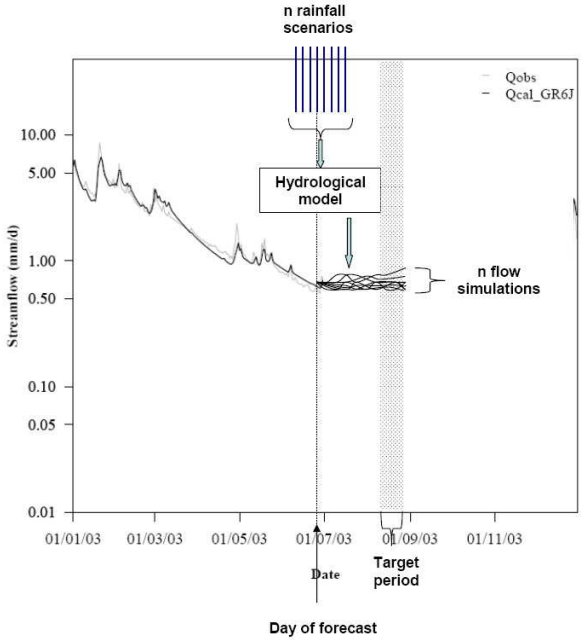


Figure 5.2: Layout of the probabilistic approach of low-flow forecasting (observed and simulated flow hydrographs of catchment J for the year 2003 is illustrated here)

Then these rainfall scenarios are used as inputs to the hydrological model on the day of forecast (see **Figure 5.2**). The hydrological model (here we used the GR6J model) simulates *n* streamflow values corresponding to the *n* rainfall scenarios for the target period (period for which we issue the forecasts).

The accuracy of streamflow predictions derived from combining rainfall predictions with hydrologic models should improve in the future, as there will be more developments in meteorological predictions (see Werner et al., 2005; Thirel et al., 2008; Bravo et al., 2009; Desaint et al., 2009; Cuo et al., 2011).

5.5.How to improve the quality of a model forecast?

The quality of forecasts is influenced by the errors of the hydrological model, which can be seen as error in timing and amplitude (bias). Although it is very important to limit errors both in timing and amplitude in high-flow conditions, the focus is much more on the errors in amplitude (overestimation or underestimation of flow values by the model) in low-flow conditions, due to the much slower dynamics of low-flow periods. Some of the possible reasons for these biases are:

- errors in model input data (typically rainfall, potential evapotranspiration),
- imperfections of the model structure,
- model calibration and identifiability issues,
- errors in the determination of the discharge hydrographs at the gauging station, which may hamper parameter estimation,
- errors associated with future rainfall scenarios in case of forecasting context.

In our case, the objective is to issue long-term forecasts with as good quality as possible by minimising these model errors and biases. The errors due to the uncertainty in future rainfall event increase with forecast lead-time and these errors from the forecast model will propagate to each member in the ensemble, which degrades the quality of resulting forecasts. Therefore these biases need to be minimized to improve forecasts.

Several options are available in the literature to minimise these errors such as bias correction, model error correction, and model updating techniques. These techniques can correct the model inputs, outputs, model states or model parameters and thereby improve the model

forecast quality. The following section briefly explains some of the selected methods to improve the model forecast quality. This work is exploratory in the sense that we tested simple approaches to evaluate to which extent model updating is relevant in the context of low-flow forecasting. Other more advanced techniques (like filters) than those selected here could be used.

Before going into the discussion, **Table 5.1** presents the notations used in the coming sections.

Table 5.1: Notations and their definitions (Q - observed streamflow; $\hat{Q}$ - simulated flow; $\hat{Q}f$ - forecast value; $\hat{\hat{Q}}f$ - corrected forecast value)

Notations	**Description**
j	Julian day (here day of issue of forecasts)
i	Ensemble member used to make the model forecast
L	Forecast lead-time
T	Threshold value (in our case, Q_{75} and Q_{90})
Q_j	Observed streamflow for day *j*
$\hat{Q}_j$	Simulated streamflow for day *j*
$\hat{Q}f^{i}_{j+L\|j}$	Forecast of ensemble member *i* for day *j+L* issued at day *j*
$\hat{\hat{Q}}f^{i}_{j+L\|j}$	Corrected forecast for the member *i* for day *j+L* issued at day *j*
$\overline{Q_j}$	Mean of all observed flows at day *j* on historical records
$\overline{\hat{Q}f_j}$	Mean of forecasts at day *j*
$\overline{Q_{j+L}}$	Mean of all observed flows at day *j+L* on historical records
$\overline{\hat{Q}_{j+L}}$	Mean of all simulated flows at day *j+L* on historical records
$\overline{\hat{Q}f_{j+L}}$	Mean of all forecasted flows (mean of ensemble member values) at day *j+L*

5.5.1. Output bias corrections

Model biases resulting from the input data, the estimated model parameters, or simplifying assumptions used in the model limit the quality of ensemble forecasts. Biases from the forecast model will propagate to each member of the ensemble and this degrades the overall quality of the resulting forecast. One of the options to account for biases is to use a bias correction transformation to adjust all model simulated ensemble members. These corrected members are then used to issue the forecasts. For more on bias corrections, please refer to Smith et al. (1992), Hashino et al. (2007), Brown et al. (2010).

The absolute flow variability in rivers during low-flow period is limited compared to the flood events. Therefore, simple output bias corrections may significantly improve the forecast quality. In our study, we tested two simple bias corrections in order to improve the quality of low-flow forecasts. The mathematical formulations of these corrections are shown in **Table 5.2**. Other more elaborated approaches could obviously be tested (see e.g. the research work by Bourgin at Irstea, ongoing PhD).

Table 5.2: Formulations of bias corrections and model updating techniques used in this study

Bias correction formulation	Corrected forecast	Correction No
$BC1 = \left(\frac{\overline{Q_{j+L}}}{\overline{\hat{Q}_{j+L}}} \right)$	$\hat{\hat{Q}}f^{i}_{j+L\|j} = \hat{Q}f^{i}_{j+L\|j} \times BC1$	**Eq. 5.1**
$BC2 = \left(\frac{\overline{Q_{j+L}}}{\overline{\hat{Q}f_{j+L}}} \right)$	$\hat{\hat{Q}}f^{i}_{j+L\|j} = \hat{Q}f^{i}_{j+L\|j} \times BC2$	**Eq. 5.2**

BC1 and BC2 are the abbreviations used to represent the multiplicative bias corrections. BC1 is derived from the ratio of the mean observed and mean simulated flow of all available flows at day “j+L” on historical record. The formulation of BC2 is similar to BC1, but BC2 uses mean forecasts at day “j+L” as the denominator instead of considering the past simulations. These corrections can give the information on the influence of biases in simulation as well as forecasts.

5.5.2. Model updating to improve the forecast quality

Updating procedures are based on the idea that the forecast is improved when the model is in better agreement with recent observations up to the time of issuing the forecast. Different types of updating procedures are available such as updating of inputs, updating of model state variables, updating of model parameters, and updating of output variables (error prediction). For a brief description of these classifications, refer to Refsgaard (1997), Babovic et al. (2001) or Anctil et al. (2003).

In our study, we selected model output updating with error corrections and model state updating due to their simplicity. Such procedures are also used in the flood forecasting GRP model developed at Irstea (Tangara, 2005; Berthet, 2010).

5.5.2.1. Updating of model output (error correction)

In an R-R model, the output error occurs due the over or under simulation of the streamflow. Therefore, in output error correction, we consider the actual streamflow as the sum of the model output and of an error term. There are simple to sophisticated procedures available in the literature to update the model output especially in flood forecasting systems with shorter lead-times (e.g. Babovic et al., 2001; Xiong and O'connor, 2002; Anctil et al., 2003; Tucci et al., 2003; Valença and Ludermir, 2004; Goswami et al., 2005).

Compared to the peak events, the fluctuations in the flow values will be lower in case of low flows. Hence simple error correction procedures may produce as satisfactory results as the sophisticated methodologies available in the literature. In our study, we present a simple procedure to update the model output to improve the low-flow forecasts. In this procedure, the error term is the error of observed and simulated flow values on the day of issue of forecasts. The formulations are presented in **Table 5.3**.

In Eq. 5.3, we use the last model error to update the model output through a multiplicative correction factor. We update the model output with the model error calculated on the day (j) of issue of forecasts. The β (beta) coefficient in Eq. 5.3 is a parameter which depends on the forecast lead-time and also the catchment characteristics. The value of β varies from 0 (indicating no correction) to 1 or higher values (indicating maximum possible correction).

Table 5.3: Formulations of model output updating techniques

Error correction formulation	Corrected forecast	Correction No
$EC1 = \left(\frac{Q_j}{\hat{Q}_j}\right)^{\beta}$	$\hat{\hat{Q}}f^{i}_{j+L\|j} = \hat{Q}f^{i}_{j+L\|j} \times EC1$	**Eq. 5.3**
$EC2 = \left(Q_j - \hat{Q}_j\right)$	$\hat{\hat{Q}}f^{i}_{j+L\|j} = \hat{Q}f^{i}_{j+L\|j} + EC2$	**Eq. 5.4**

Tangara (2005) used a similar formulation to calculate the past forecast errors to correct the flood forecasts with short lead times. Berthet et al. (2009) and Randrianasolo et al. (2010) also used a similar correction methodology in their studies on flood forecasting. In contrast to these studies, here we used this formulation to update the model output to improve the low-flow forecasts. Instead of selecting a single β value, we tested the forecasts with different β values in order to find out the optimum β for each lead-time by linking it with the forecast lead-time. The concept behind the formulation of Eq. 5.4 is similar to that in Eq. 5.3, but here we use the additive error with a β value of 1 instead of the multiplicative one with optimum β in Eq. 5.3.

5.5.2.2. Updating of model state variable

The term ‘state’ is used to describe a variable of a model which changes between inputs to the model and the model output (Szollosi-Nagy, 1976). Model state updating is quite popular in hydrology (Aubert et al., 2003; Moore et al., 2005; Moore, 2007; Thirel et al., 2010a; Thirel et al., 2010b). The idea behind the state updating is that day-to-day watershed’s state deviate from the average conditions simulated by the model.

In our study, we used a simple updating procedure already tested in flood forecasting (e.g. Berthet et al., 2009; Randrianasolo et al., 2010). In this updating procedure, we use the last observed discharge information to update the state of the model stores. The GR6J model has two routing stores and one production store in its structure (see chapter 4). The routing part of this model consists of two stores and is mainly responsible to generate stream low-flows. Similar to the routing part, the soil moisture store also has a significant role in streamflow in terms of percolation (it contributes to direct runoff). Therefore, the updating of these stores can improve the model's low-flow forecast quality. Hence we tested the model’s forecasts

with the updating of the production store as well as the two stores at the routing part of the model. There is no direct formulation to calculate the store levels (production as well as routing stores) at each time step. But we have a formulation which connects the model output (flow value) and the store levels (see Appendix C). Therefore, with the latest flow value, we calculate back the corresponding store levels. These updated levels are then used to issue the model forecasts.

5.6. Assessment of the forecasting system

Forecast verification is the process of assessing the quality of a forecast. In verification, the forecast is compared with a reference estimate of the true outcome (usually the observation). The verification of forecast with respect to the observed value can help to improve the existing forecasting system. In general, forecast verification includes the calculation of different verification scores over a forecast-observation data set and then the performance of each forecast system is evaluated based on the computed verification scores. There are deterministic as well as probabilistic criteria to evaluate the corresponding forecasts. Deterministic scores are also used to evaluate the ensemble forecasts. Here the ensemble forecasts are converted to deterministic one before the evaluation using the corresponding deterministic criterion. But, finding an efficient way to evaluate hydrological forecasts is a challenging task for hydrologists. In our study, we used some criteria (deterministic as well as probabilistic) to evaluate the stream low-flow forecasts.

5.6.1. Measure based on error in magnitude

Here we propose the Nash-Sutcliffe efficiency calculated on the inverse transforms of flows (NSE_{iQ}) which we already discussed in Chapter 3 to evaluate the low-flow forecasts. Here also the bounded version of the criterion (NSE^{*}_{iQ}) is used. The formulation of this deterministic criterion in the present context is as follows:

$$NSE_{iQ} = 1 - \frac{\sum_{j=1}^{N}\left(\frac{1}{Q_j} - \frac{1}{\hat{Q}f_j}\right)^2}{\sum_{j=1}^{N}\left(\frac{1}{Q_j} - \overline{\frac{1}{Q_j}}\right)^2} \qquad \textbf{Eq. 5.5}$$

$$NSE^{*}_{iQ} = \frac{NSE_{iQ}}{(2 - NSE_{iQ})} \qquad \textbf{Eq. 5.6}$$

The notations used are defined in **Table 5.1**.

5.6.2. Measures based on contingency table

Contingency table indicates, for a given observed or not observed event, the number of times this event was predicted or not (Atger, 2001). In the present analysis, we consider the probability of occurrence of an event at two thresholds (T) such as the observation and forecast at Q_{75} and Q_{90} (the low-flow indices). The structure of a contingency table is shown in **Table 5.4**.

Table 5.4: Contingency table

		Observed (Q)		
		Yes ($Q < T$)	No ($Q > T$)	Total
Forecasted ($\hat{Q}f$)	Yes ($\hat{Q}f < T$)	Hits (a)	False alarms (b)	Forecast yes (a+b)
	No ($\hat{Q}f > T$)	Misses (c)	Correct negatives (d)	Forecast no (c+d)
	Total	Observed yes (a+c)	Observed no (b+d)	

where, Hit (a) - Both forecast and observation occurred

Misses (c) - Only the observation occurred

False alarm (b) - Only the forecast occurred

Correct negative (d) - Both the forecast and observation did not occur

A perfect forecast system would always produce hits and correct negatives. There are a large number of evaluation measures based on the contingency table. **Table 5.5** presents two of the most widely used verification scores based on the contingency table which we used to evaluate the present low-flow forecasting system.

Table 5.5: Scores based on the contingency table

Verification measure	**Equation**	**Range**
Bias (B)	$(a+b)/(a+c)$	[0,∞[, perfect score: 1
Probability of detection (POD)	$a/(a+c)$	[0,1], perfect score: 1

Bias measures the ratio of the frequency of forecast events to the frequency of observed events. It indicates whether the forecast system has a tendency to underforecast (B < 1) or overforecast (B > 1) the events. POD is the probability of detection (hits). The selected two criteria consider hits, misses and false alarms which are more sensitive in a forecasting system than the correct negatives (it is not so dangerous as the observation and forecasts are not occurred).

5.6.3. Measure based on probability

The Brier score (BS) is one of the most widely used criterion to evaluate the accuracy of probabilistic forecasts (Brier, 1950). BS is the mean squared error of the probability forecasts over the verification sample and is expressed as:

$$BS = \frac{1}{N}\sum_{j=1}^{N}(\hat{P}_j - P_j)^2 \qquad \textbf{Eq.5.7}$$

where $\hat{P}_j$ is the probability of the forecast event and P_j is the actual outcome of the event (equal to 1 if the event is observed, equal to 0 if it is not observed). *N* is the number of forecasts (here it corresponds to the number of ensemble members for day j). BS values range between 0 and 1. For a perfect forecast, BS becomes zero and hence the smaller the value of BS, the higher the quality of forecast. The BS is generally used to evaluate the occurrence of an event greater than or equal to a pre-defined threshold, T. In the present analysis, we consider the calculation of BS at two thresholds (BS_{75} and BS_{90}), Q_{75} and Q_{90} which we already used to formulate the contingency table.

Using a relative measure that represents the forecast quality compared to a reference is a better way to assess the forecast quality (Bradley et al., 2004). That is the evaluation of the performance of the model forecast with respect to a reference forecast (here it is the "climatological" forecasts: the streamflow values in the historical data for a particular date at the same thresholds). The formulation of the calculation of daily Brier skill score (BSS) for a single catchment is given by:

$$BSS_{75} = 1 - \frac{BS_{75}}{BS_{c\,75}} \qquad \textbf{Eq. 5.8}$$

where BS_{75} and BS_{c75} are the mean score of the model forecast and the reference forecast respectively. The BS_{c75} is the Brier score of climatological forecast (for each day) which is calculated similar to BS_{75}, but here in climatology, we consider past observations rather than past simulations. From these values we calculated the BSS_{75} for each catchment for each lead-time. In order to make a general conclusion, we illustrate the average BSS_{75} calculated over the entire catchment set for each lead-time. Similarly we calculated the BSS_{90} at Q_{90}.

A score of positive values indicates good performance of the model compared to the reference forecasts and a reverse case with negative values. The BSS calculated at Q_{75} and Q_{90} are represented as BSS_{75} and BSS_{90} respectively. Based on these values (BSS_{75} and BSS_{90}), we will fix the maximum possible lead-time (MPLT) for each catchment in our data set. The MPLT is the forecast lead-time beyond which the forecasts do not bring better information to the end-user than a statistical analysis of historical data. In the present study, we fixed the lower limit of the BSS as 0.2 to fix the MPLT. This says that, for our test catchments, we consider the lead-time for which the BSS falls to 0.2 as the MPLT. This indicates that the lead-time for which the BSS comes below 0.2 is no more useful. The lower limit of BSS to fix the MPLT depends on the purpose of the model output.

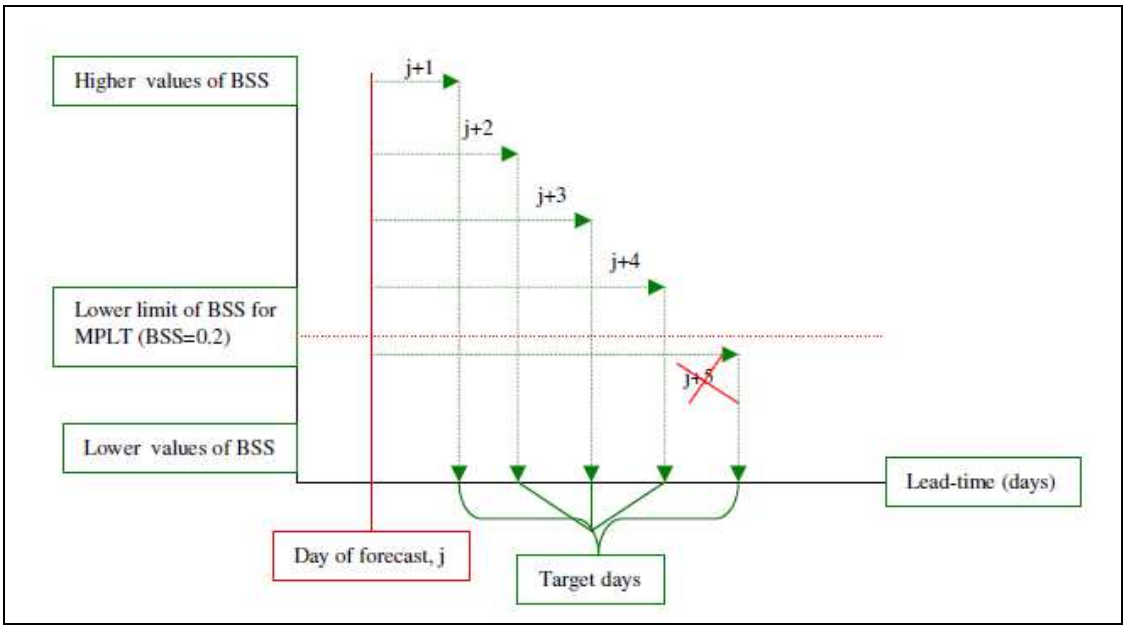


Figure 5.3: Illustration of the identification of maximum possible forecast lead-time with respect to the BSS

Figure 5.3 illustrates how we selected the maximum possible lead-time for each catchment. We consider that forecast corresponding to the lead-time for which the BSS is below 0.2 is no more useful. In this example, the maximum possible lead-time is "4" with a BSS value higher than 0.2. For the fifth day, the BSS reaches a value less than 0.2 and hence we cannot consider this as our MPLT.

Here the lower limit of BSS is kept in such a manner that the values very near to 0 (here less than 0.2) are not so significant. The BSS value greater or equal to 0.2 is considered significant here.

5.7. Results and discussion

This section mainly discusses the results in terms of:

- the skill score of the ensemble forecasts issued by the model GR6J;
- the impact of biases during simulation on model forecasts;
- the improvement of model forecast skill with bias corrections and updating techniques;
- and finally the identification of MPLT for the individual catchments based on the BSS.

5.7.1. How long are model forecasts useful (without any corrections)?

We used the Brier skill score (BSS) to evaluate the quality of GR6J model forecasts. The "climatology" (past flow observations) was used as the reference forecast in the BSS to evaluate the gain or loss in performance of the GR6J model forecasts. The BSS was calculated at two thresholds: Q_{75} and Q_{90}. The BSS at Q_{75} (BSS_{75}) and the BSS at Q_{90} (BSS_{90}) for different lead times are presented in **Figure 5.4**. It shows that the skill score is degrading with the lead-time. The values above zero indicate a better performance of the model forecasts than the reference forecast and a reverse case with negative values. In the very first lead-time itself, for the majority of catchments, the BSS values are below zero in case of BSS_{90} (i.e. the Brier score calculated at a threshold of Q_{90}). This is less pronounced in the case of BSS_{75}. This may be due to the difficulty of the model to simulate the most extreme low-flow conditions. The poor model performance in simulation mode at the extreme low-flow condition (at Q_{90}) compared to the model performance at Q_{75} is illustrated in **Figure 5.5**. The inverse transform is used to evaluate the model performance. This clearly indicates that the BSS or the forecast skill score is somehow related with the simulation efficiency of the model.

The poor BSS values indicate the limited usefulness of the model, which is even truer at the longer lead times. This indicates that the model used in pure simulation mode is not very useful on average for low-flow forecasting. This suggests the necessity to improve the model forecasts and we will discuss this in section 5.6.3.

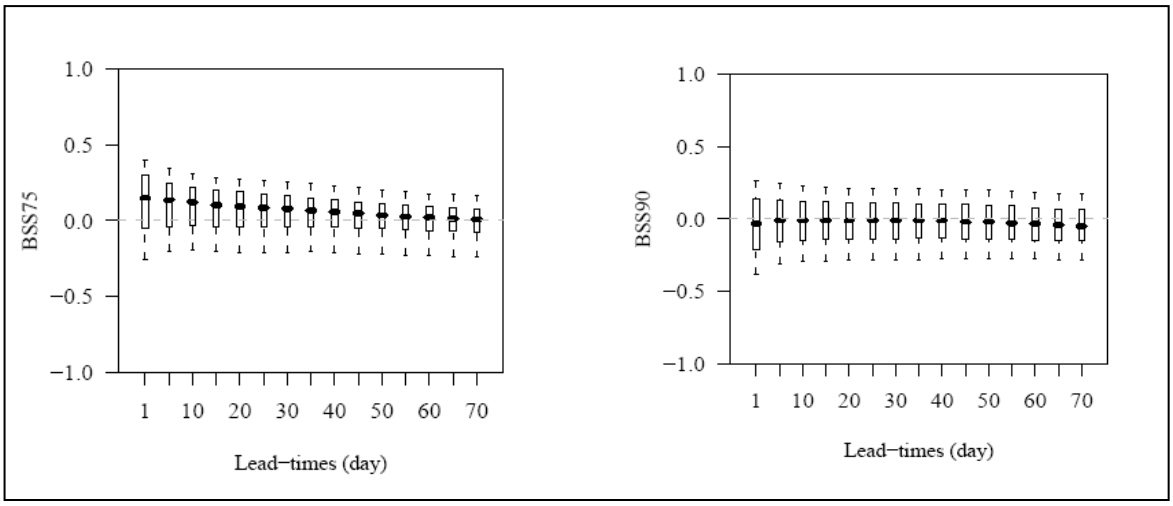


Figure 5.4: Distribution of Brier skill scores (at Q_{75} and Q_{90}) over the full catchment set at different lead times (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles) (in simulation, i.e. without correction)

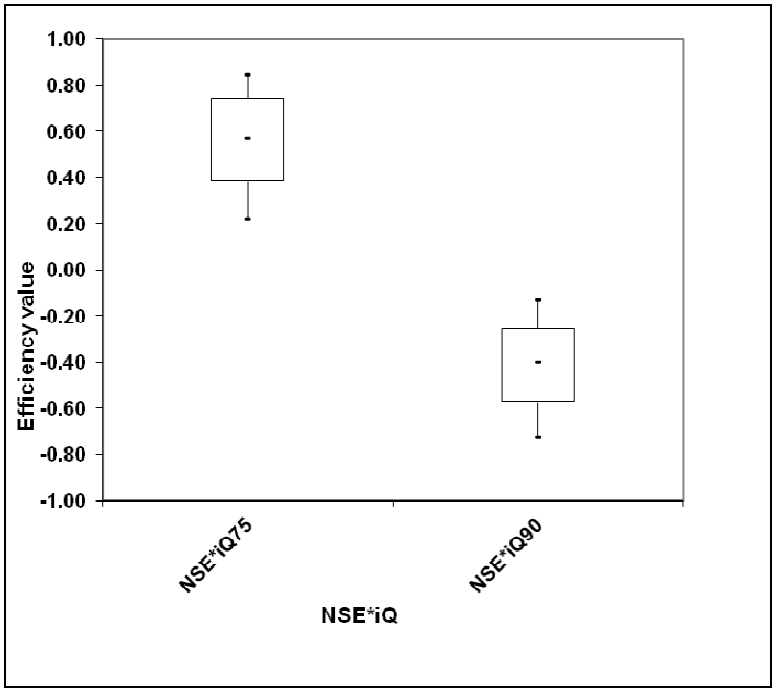


Figure 5.5: Model performance in simulation (i.e. without correction) mode at Q_{75} and Q_{90}

5.7.2. Influence of simulation error on model forecasts

Improving simulation efficiency can improve forecast quality. In this section, we analysed the scores which we calculated in the simulation mode and also in the forecast mode. The simulation bias (ratio of the mean observation and mean simulation at the two thresholds Q_{75} and Q_{90}) and the bias calculated from the contingency table at Q_{75} and Q_{90} are analysed to see the influence of simulation errors on model's forecasts.

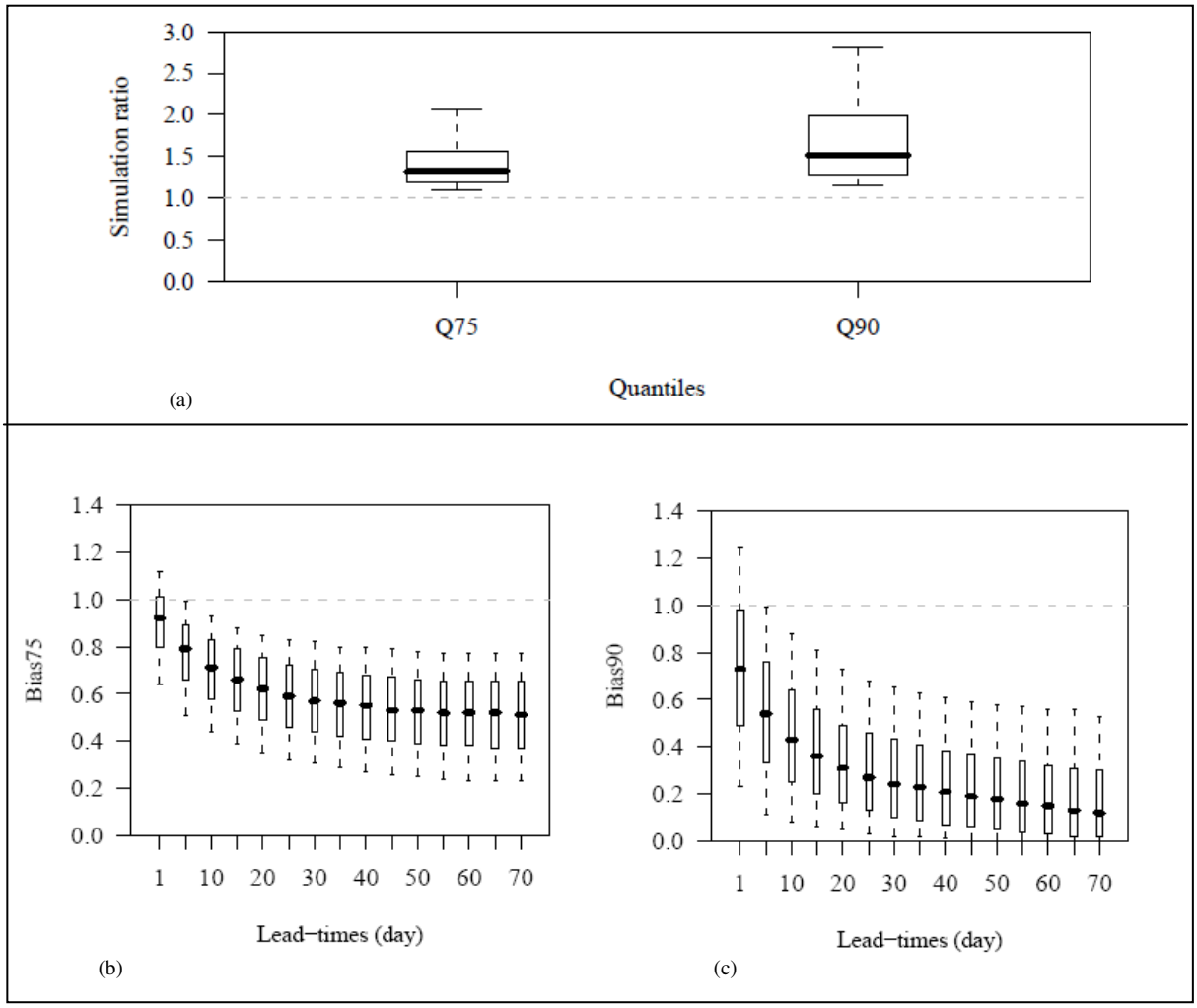


Figure 5.6: The influence of error in simulation mode (a) on model's forecasts (b and c). Boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles

In **Figure 5.6**, a value above 1 indicates overestimation and a value below 1 indicates underestimation. The box plots in figure (a) show that the model overestimates flows at the two selected thresholds and this overestimation leads to make more misses during forecasts. This makes under-forecasts. **Figure 5.6** (b and c) show the distribution of bias criteria calculated from the contingency table (see **Table 5.5**). This clearly shows that the forecasts are underestimated. This suggests that the biases in model simulation have an impact on the model's forecast quality. Some of the options to avoid the influence of biases will be discussed in the coming section.

5.7.3. How to improve model's forecast quality?

This section briefly explains the results of some of the methods used in our study to improve the forecast quality.

5.7.3.1. Model output bias correction

The results of the bias corrections used in this study (BC1, BC2) are compared with the non-corrected model (GR6J) forecasts. As the Brier skill score is the comparison of the performance of the model forecast with climatology, the analysis of all the tested bias or error corrections are carried out just by considering the Brier skill score.

Figure 5.7 shows that the improvement in the skill score for both thresholds is not significant using BC1 and BC2 compared to the non-corrected forecasts for the entire catchments. This might be due to the error associated with mean simulation with respect to the observation in the formulation of BC1. This simulation error will propagate into the members and hence this affects the quality of the forecast. In BC2, the forecast itself is used in association with observation to correct the members. Similar to the previous case, here also from the forecast used in the formulation, the error can transfer into the members and hence on quality of the forecast.

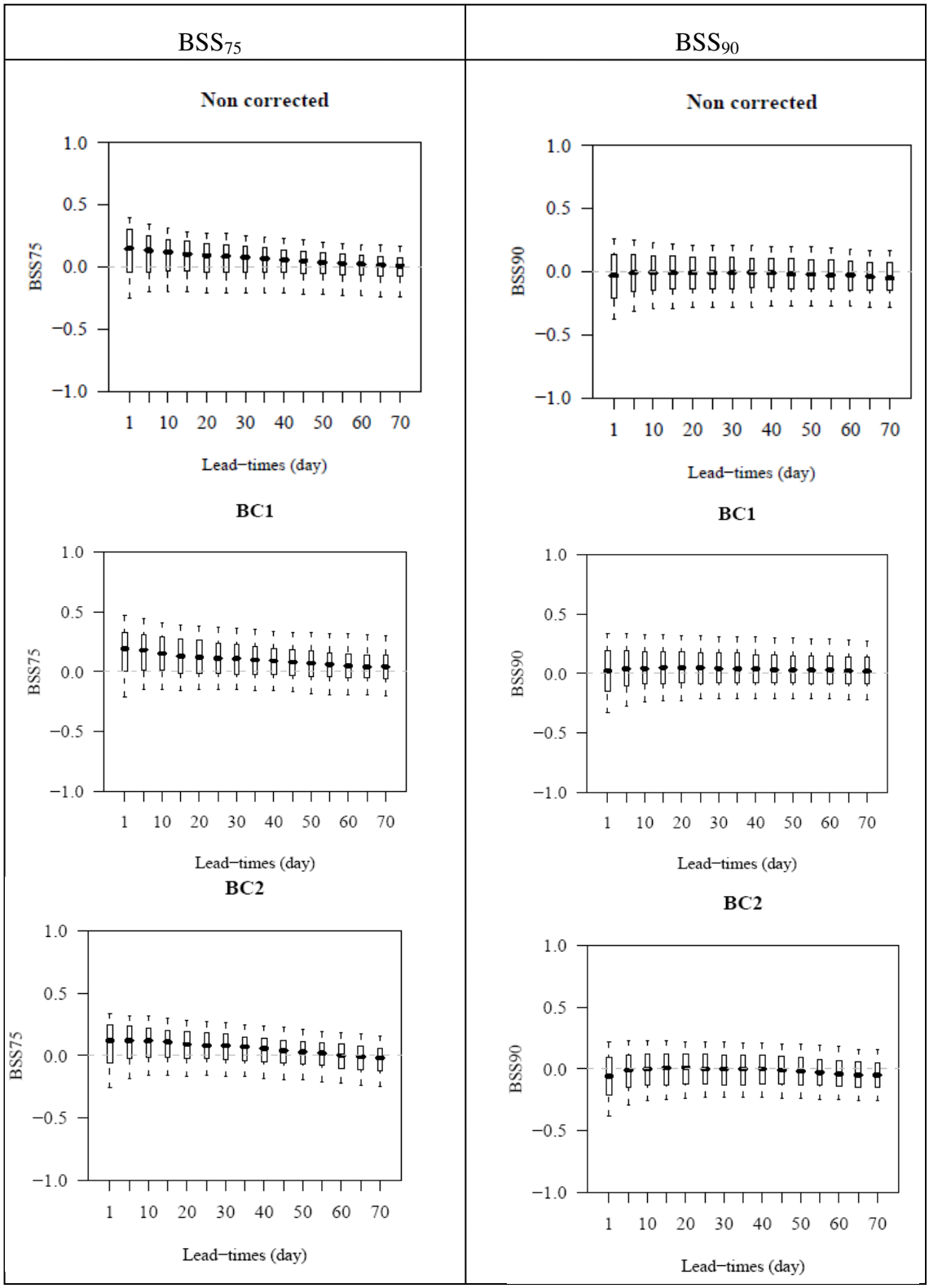


Figure 5.7: Comparison of model forecasts with and without bias corrections, BC1 & BC2 (the first column represents the results at threshold Q_{75} and the second column represents the results at Q_{90})

5.7.3.2. Model output updating with error corrections

Figure 5.8 shows that model updating with the error of the day of forecast can improve the model performance considerably. The EC1 correction yields more robust results than the EC2 correction. These two methods are superior to the previously mentioned bias corrections (BC1 and BC2). Compared to EC2, there is a possibility for huge improvement in the forecast skill with EC1 while changing the beta (β) value from 1.

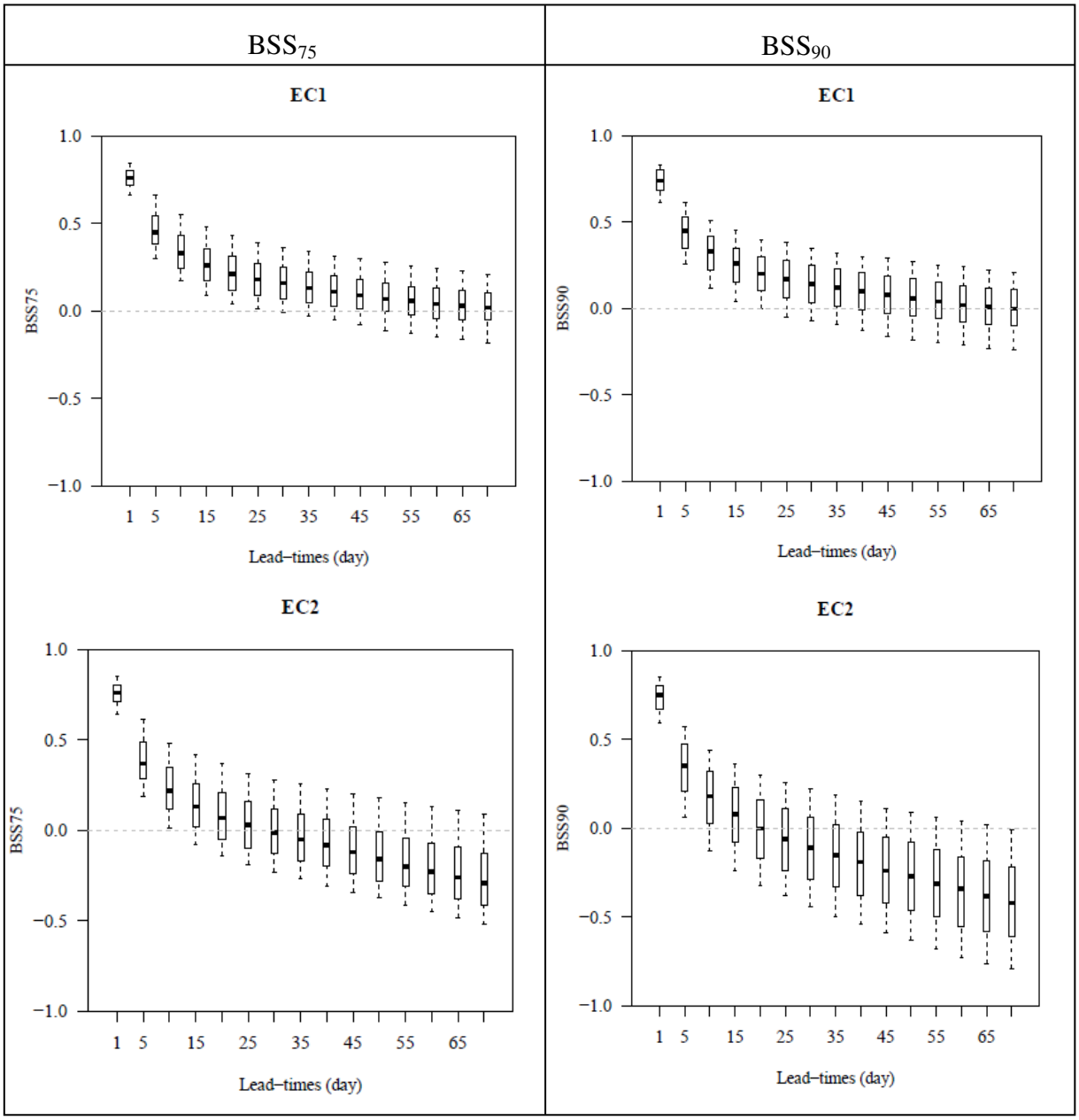


Figure 5.8: Forecast skill score with model output error corrections: EC1 and EC2

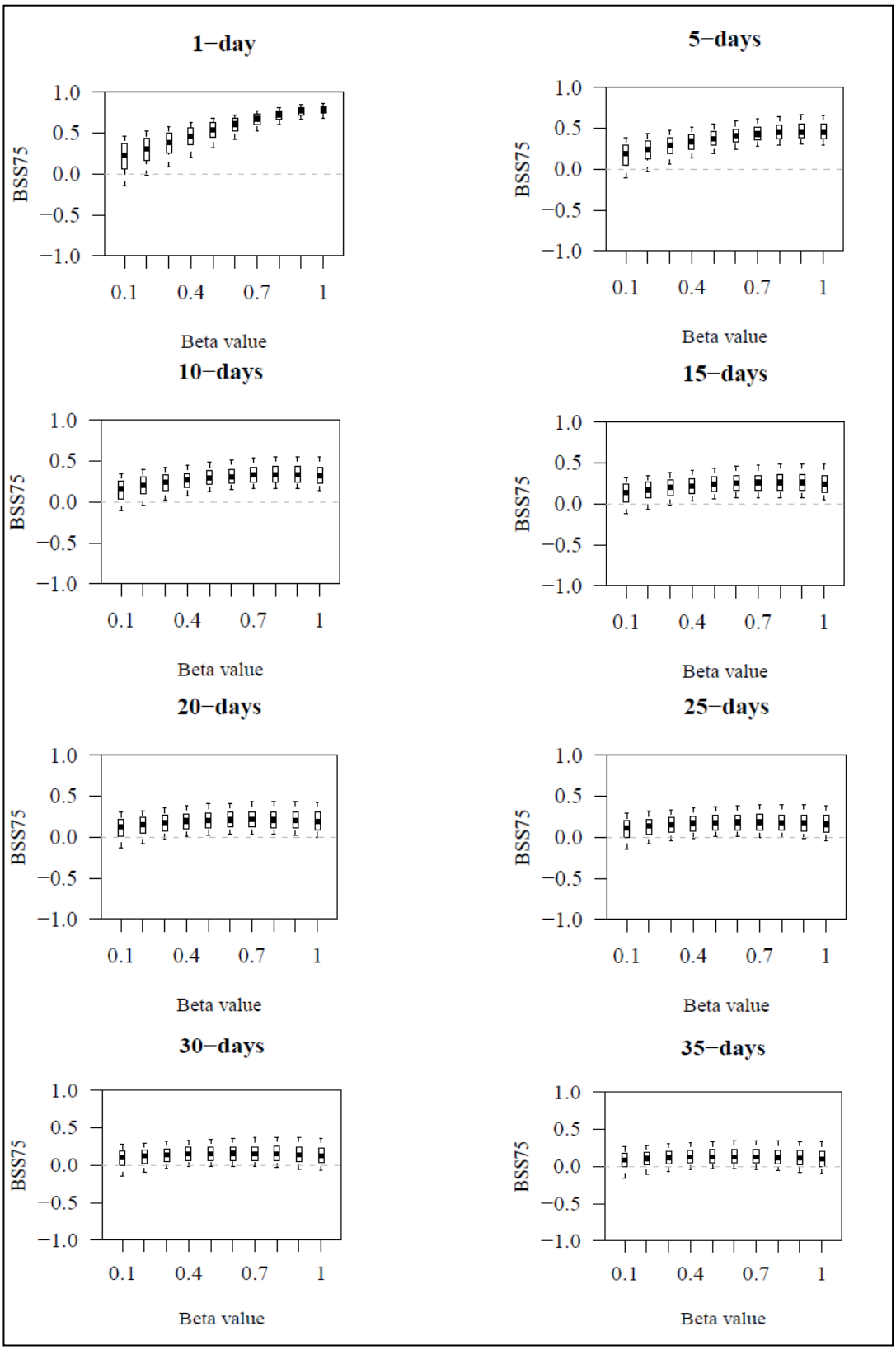


Figure 5.9: The performance of model forecasts with different beta values and different lead times

Therefore, the formulation of EC1 is further investigated. The tested β value ranges from 0.1 to 1. The corresponding model performance is presented in **Figure 5.9**. But the optimum β value (the beta value for which the forecasts has better BSS values) changes from one catchment to another with respect to the lead-time.

Hence, we identified β optimum ($Opt\beta$) for each catchment for each lead-time. As an example, for the first lead-time, the optimum betas (β_1, β_2, β_3,.... β_N) which correspond to the highest skill score for each catchment were identified. Then, we calculated the average ($Opt\beta_{1stlead}$) of all β values on the catchment set based on Eq. 5.9. This average is the optimum β for the first lead-time. This procedure is repeated for all lead-times (here it is 70 days).

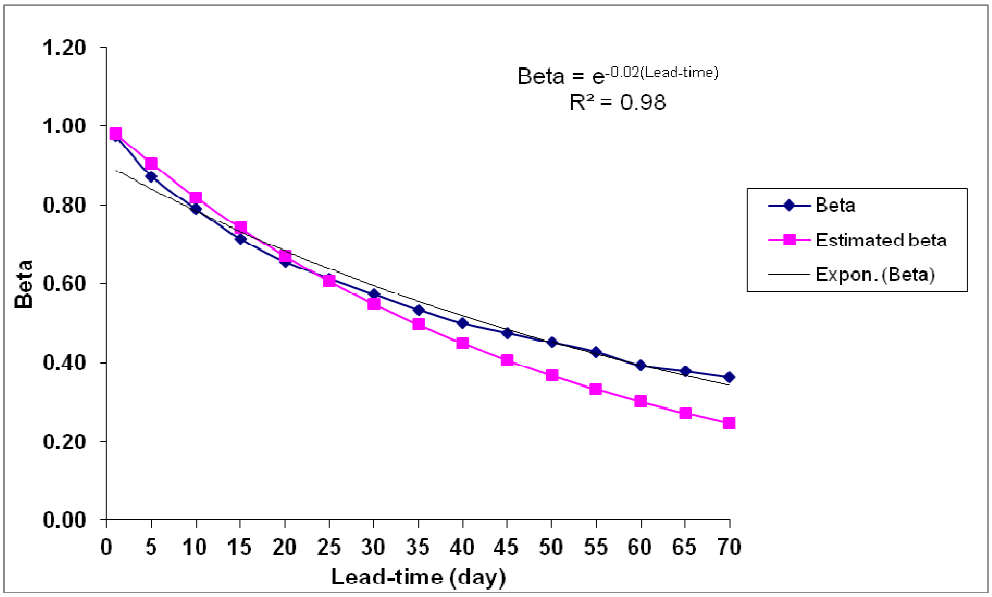


Figure 5.10: Dependency of beta value with forecast lead-time

Figure 5.10 shows the plot of dependency of optimum β values with forecast lead-time. The pink dots and line represent the estimated β values using the derived relation between the optimum β values (blue dots and line) and the lead-time. The derived exponential relation of the optimum β with lead-time, L with an R^2 value of 0.98 is represented in Eq. 5.10 and the resulting formulation of EC1 (the new formulation of EC1 is represented as EC1') is presented in Eq. 5.11.

$$Opt\beta_{1stlead} = \frac{1}{N}(\beta_1 + \beta_2 + \beta_3 + ... + \beta_N) \quad \textbf{Eq. 5.9}$$

$$\beta = \frac{1}{\exp(0.02L)} \quad \textbf{Eq. 5.10}$$

$$EC1' = \left(\frac{Q_j}{\hat{Q}_j}\right)^{\frac{1}{\exp(0.02L)}} \quad \textbf{Eq. 5.11}$$

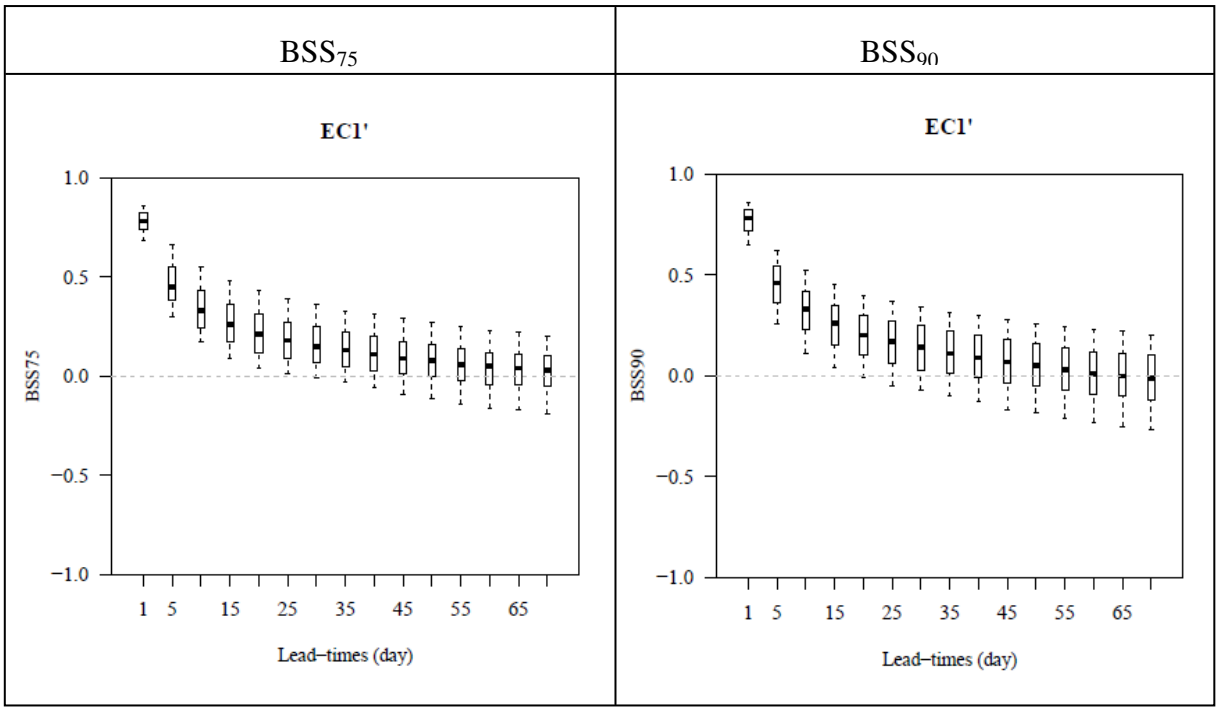


Figure 5.11: Forecast skill score with lead-time using model updating, EC1'

Figure 5.11 shows that there is a significant improvement in the skill score with the EC1'. There is a positive skill score till the 30th day in case of BSS_{75}, at the same time, 20 days in case of BSS_{90} while considering the full data set.

As discussed at the beginning of this Chapter, the absolute flow variability during the low-flow period is not so significant compared to the flood events. Therefore, the simple error

correction techniques can improve the forecast skill compared to the peak events. Our results are in agreement with this concept. The simple bias correction (EC1': correcting each day's forecast with the bias of the day of forecast) is superior to the other bias corrections and error corrections tested here.

5.7.3.3. Results of model state updating

This section briefly discusses the results of the updating of the model's state variables. Here we updated the levels of the routing stores as well as the production store (the soil moisture store) with the last model output information (last discharge value) at each time step. So that the new level of the stores for the next time step can make an improvement in the forecast quality. The updating of the routing stores shows better results compared to the updating of the production store. The results are illustrated in **Figure 5.12**. The improvement in the model forecast skill score with the updating of the levels of the two routing stores shows that the routing part of the model is more sensitive to the contribution of the low flows into the streams compared to the production module. The structure of the model itself says that the routing part of the model is mainly responsible to generate stream low flows. The present results also confirm that the changes in the routing part are highly responsible to make changes in the stream low flows.

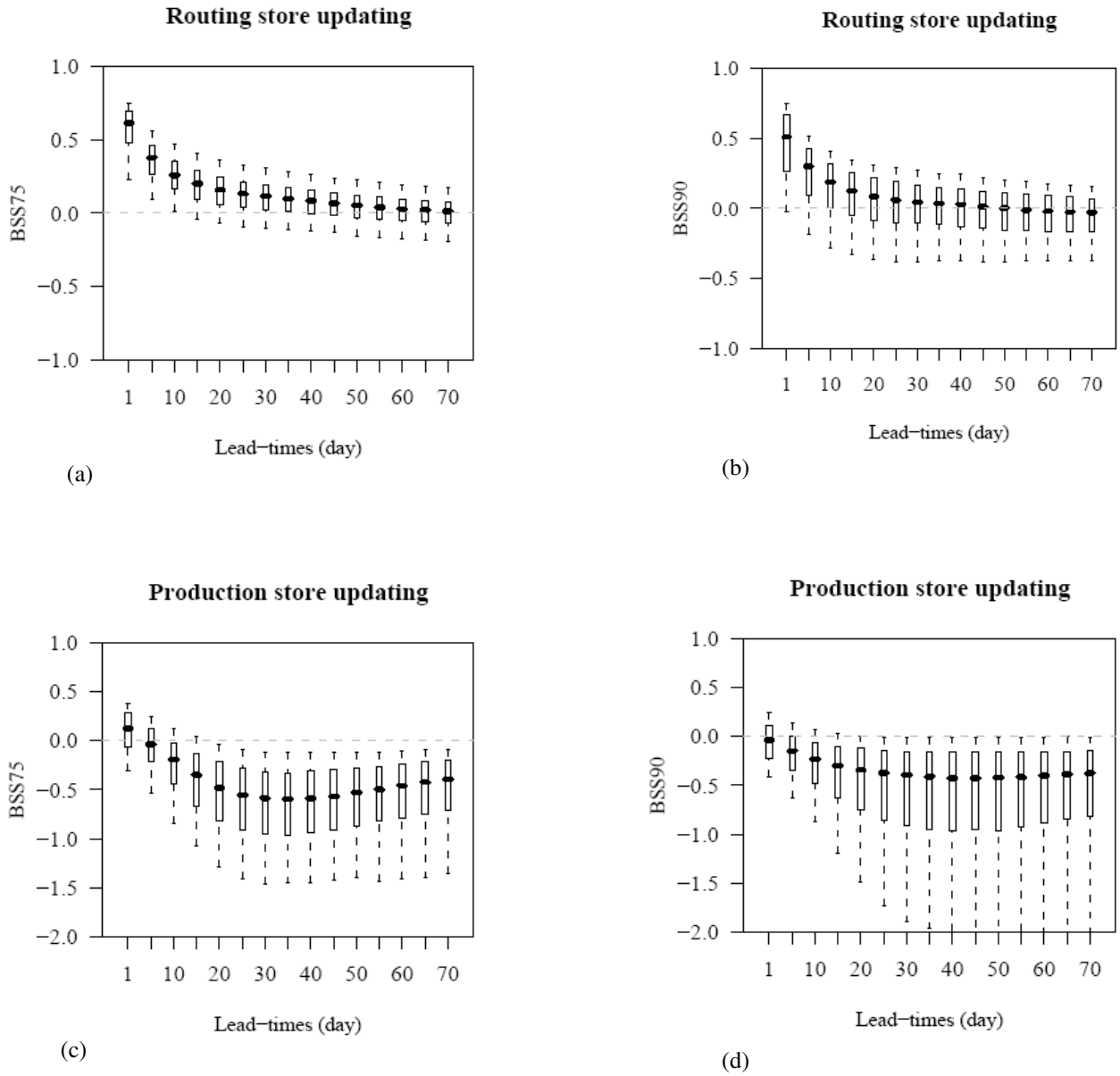


Figure 5.12: Illustration of the model's forecast skill score with the state updating ((a) & (b): updating of routing stores; (c) & (d): updating of the production store)

5.7.1. Maximum possible lead-time (MPLT)

The analysis of the results of bias corrections and updating techniques show that the error updating, EC1' is superior to other tested methods (see **Table 5.6**). Here we analysed the performance of this correction factor in terms of the other scores (POD and NSE^{*}_{iQ}) which we already presented in section 5.6. Here we presented the POD obtained with the EC1' method and the POD without correction (see **Figure 5.13**). A significant improvement in POD is visible with the integration of this correction factor. A similar trend is visible in **Figure 5.14**

with the NSE^*_{iQ} values. Thus the improvement in the POD (the score based on the contingency table), NSE^*_{iQ} (the deterministic score) and the BSS (the probabilistic score) with the integration of the EC1' highlights the applicability of this correction factor in improving the low-flow forecast skill with longer lead-times in our test catchments. Therefore, we selected this correction factor to improve the forecast skill in our tested catchments.

Table 5.6 Median values of the BSS for the different correction procedures

Lead-time	BSS_{75}			
	BC1	**BC2**	**EC1'**	**EC2**
1	0.19	0.12	0.78	0.75
5	0.18	0.12	0.45	0.35
10	0.15	0.12	0.33	0.18
15	0.13	0.11	0.26	0.08
20	0.12	0.09	0.21	0.00
25	0.11	0.08	0.18	-0.06
30	0.11	0.08	0.15	-0.11
35	0.10	0.07	0.13	-0.15
40	0.09	0.06	0.11	-0.19
45	0.08	0.04	0.09	-0.24
50	0.07	0.03	0.08	-0.27
55	0.06	0.02	0.06	-0.31
60	0.05	0.00	0.05	-0.34
65	0.04	-0.01	0.04	-0.38
70	0.04	-0.02	0.03	-0.42

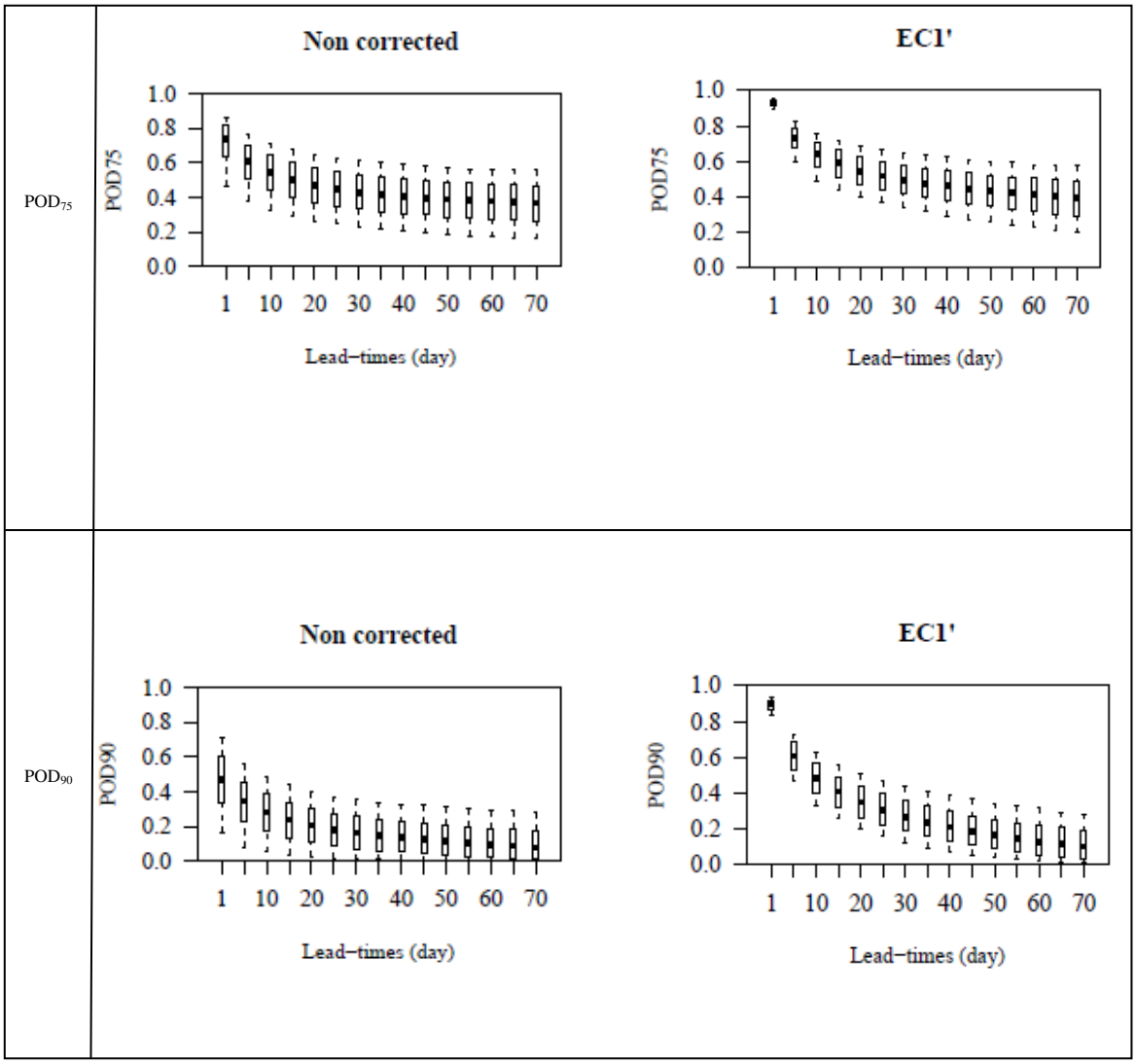


Figure 5.13: POD of forecasts with bias correction (EC1') and POD of forecasts without correction

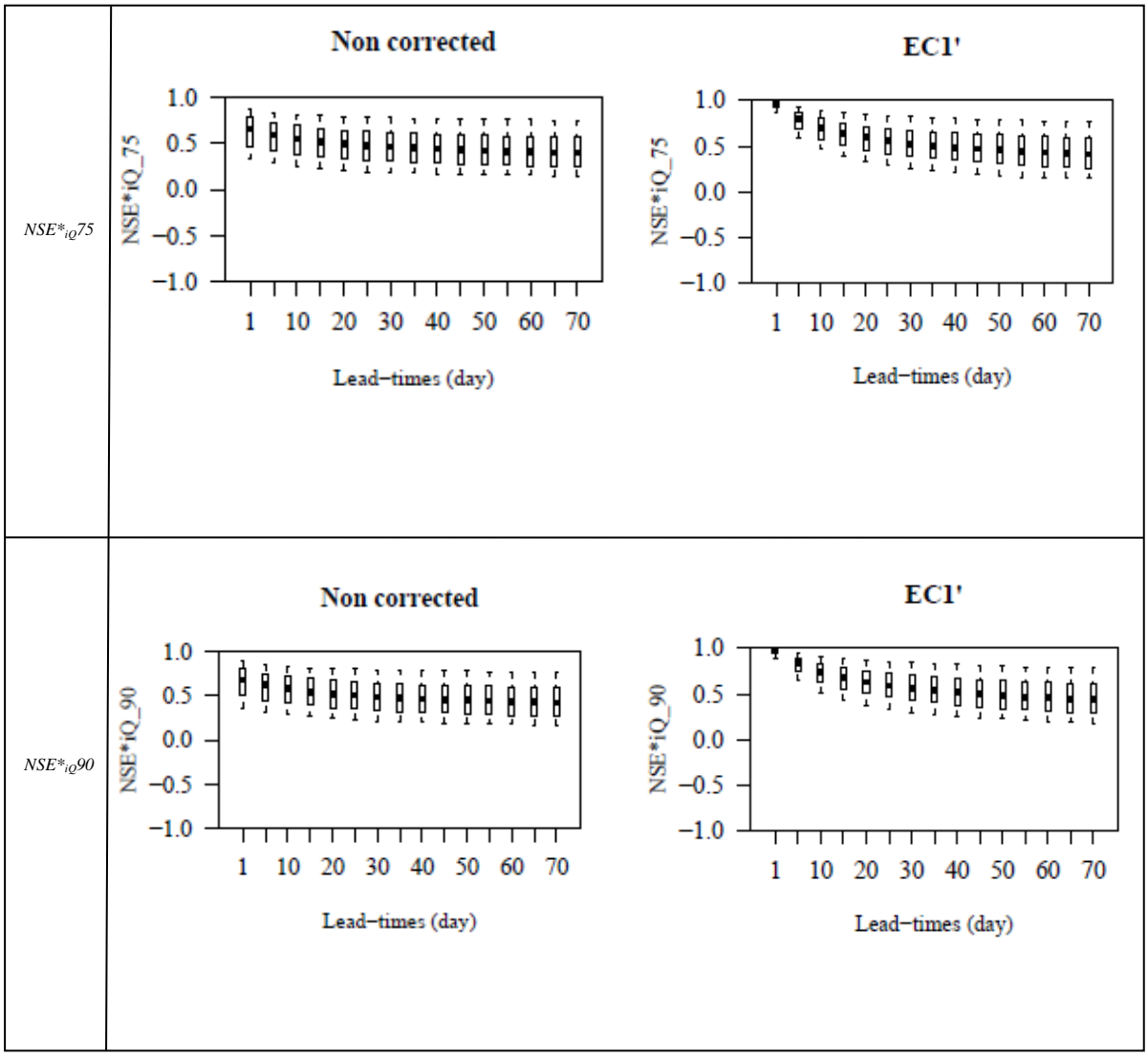


Figure 5.14: NSE^*_{iQ} of forecasts with bias corrected (EC1') and NSE^*_{iQ} of forecasts without correction

As in section 5.6.3, the maximum possible lead-time for the individual catchments is identified with respect to the BSS. In our tested catchments, we considered that a value of BSS greater than 0.2 indicates the significant gain in the model forecast compared to the reference forecast, i.e. “climatology”. Therefore, the BSS of 0.2 was the lower limit for each catchment to select the maximum possible lead-time.

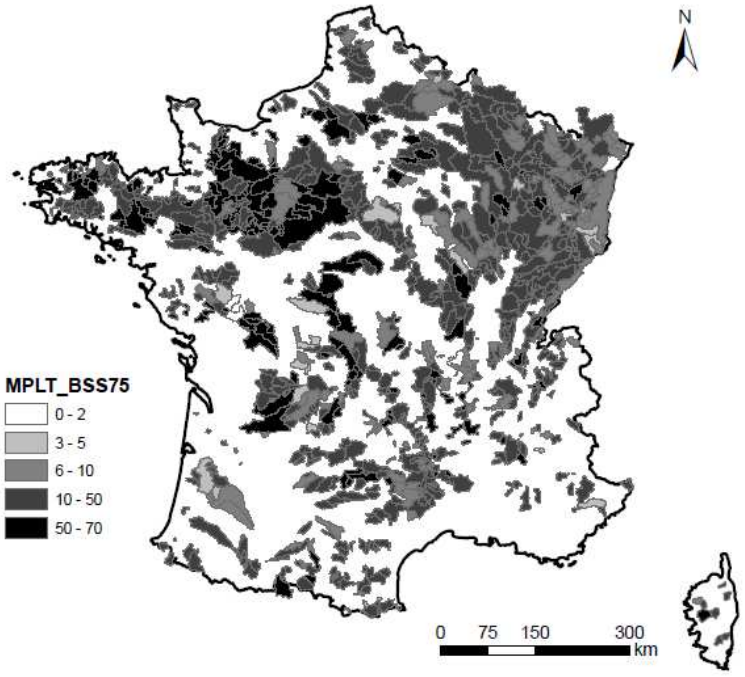


Figure 5.15: Spatial distribution of the maximum possible lead-time based on BSS_{75} for the data set

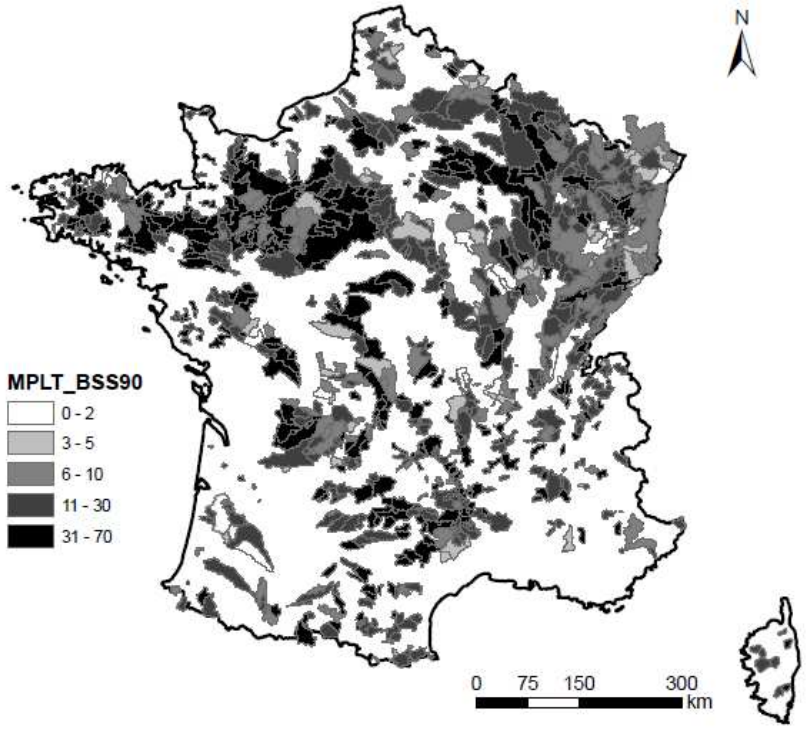


Figure 5.16: Spatial distribution of the maximum possible lead-time (in days) based on BSS_{90} for the data set

The spatial distributions of the maximum possible lead time (MPLT) for the full data set at the two thresholds (BSS_{75} and BSS_{90}) are illustrated in **Figure 5.15** and **Figure 5.16**. Most of the catchments in the North-Western part of France have longer forecast lead times. The MPLT for most of the catchments in the data set is less than 10 days. The spatial variability in the MPLT shows that this variability may depend on catchment characteristics such as catchment geology.

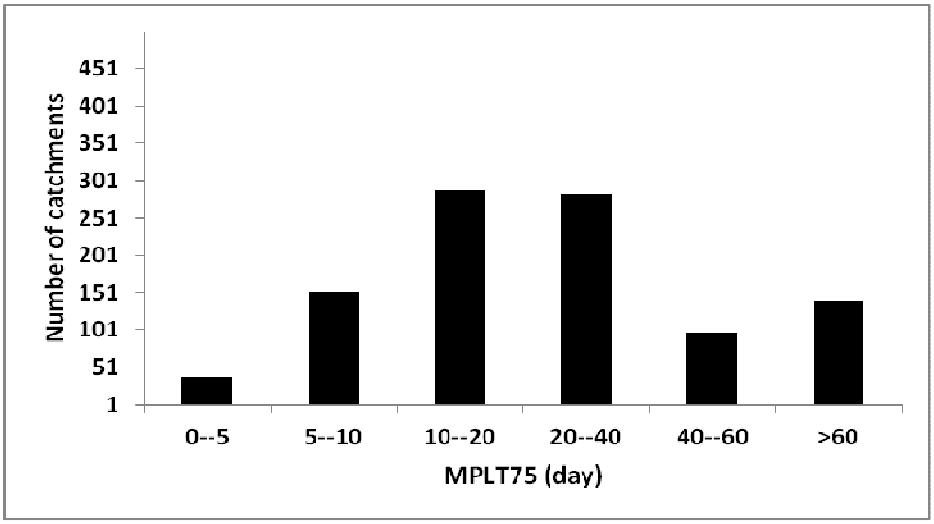


Figure 5.17: Distribution of the maximum possible lead time for the catchments on the data set

Figure 5.17 shows that 29% of our test catchments, forecasting with a lead-time of 10 to 20 days is highly possible. In 28% of data set, the MPLT reaches up to 40 days. 10% of catchments can attain a MPLT of 60 days. The possibility of long-term forcasting is identified for a limited number of catchments (>60 days), i.e., for 14% only. But, in 19% of our data set, we can issue only short-term forecasts (<10 days).

A brief overview of catchment geology and the corresponding lead-time for the catchments are presented in **Figure 5.18**. **Figure 5.18** illustrates the possible link of forecast lead-time with the catchment geology at a threshold of Q_{75}. Here we considered the major geology for each catchment as they are composed of several fractions of geology. From the illustration, it is not possible to derive a clear conclusion, even though we can interpret that catchments with geology of quaternary volcanics and detrital crystalline show longer lead-times. Test results also indicate that long-term forecasting is not possible in catchments with a geology of non-carbonates and flysch sediments. A more in-depth analysis would be needed to establish clearer links (e.g. only focusing on catchments where the model performs well to avoid introducing modelling noise in the analysis).

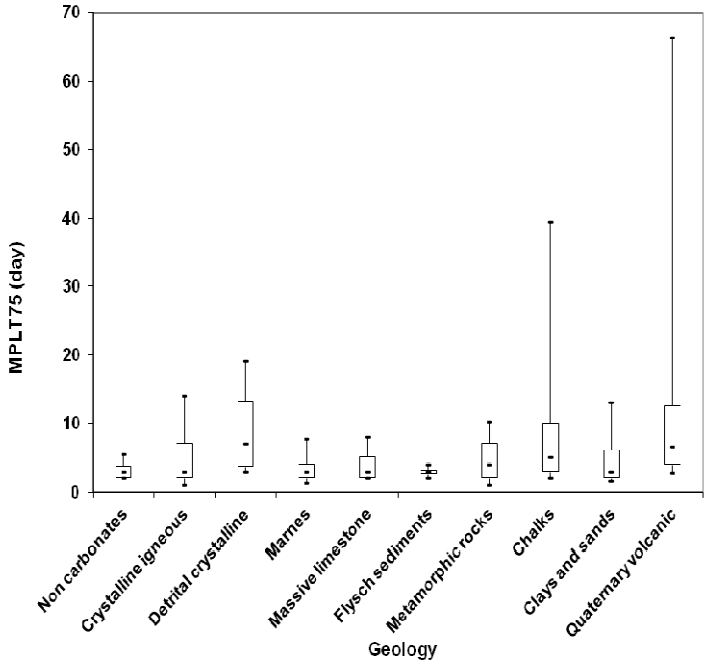


Figure 5.18: Possible link of lead-time with catchment geology

5.8. Conclusions

In this Chapter, we implemented an ensemble forecasting approach to issue long-term low-flow forecasts in the French river basins. The forecasts are issued using the rainfall scenarios from an historical archive as input to the GR6J model. As a first step, the model forecasts showed poor skills for both the low-flow thresholds (Q_{75} and Q_{90}) in terms of the Brier skill score. We tried to improve the skill score with the integration of bias corrections, model output updating and model state updating techniques. Among all the tested correction methods, the output error updating (EC1') showed significant improvement in the skill score with longer lead-time. Based on this tested methodology, we identified the maximum possible lead-time for the individual catchments in the data set for both the low-flow thresholds. The attempt to link forecast lead-time with catchment geology could provide only a general overview of our results. Further investigations are needed in this issue.

Chapter 6. Accounting for the influence of artificial reservoirs in low-flow forecasting

6.1.Introduction

Artificial reservoirs are intended to modify the natural hydrological processes and thereby they may affect the watershed behaviour by locally changing the watershed's response to precipitation in several ways. The construction of such reservoirs is one of the direct human influences on streamflow. Reservoirs lead to changes in streamflow downstream the reservoir and these changes on flow regime will depend on the storage capacity of the reservoir in comparison with stream runoff, its regulation purposes (e.g. irrigation diversions, hydro-electricity, flood control, low-flow augmentation, etc.), and its operating rules (Miquel and Roche, 1983; Williams and Wolman, 1984). The changes can be in the form of frequent occurrence of low flows downstream (Ye et al., 2003; Lajoie et al., 2007; Isik et al., 2008; Lopez-Moreno et al., 2009) which affects the downstream ecosystem (Hill et al., 1998), or reduced variability in the downstream flows, i.e., higher low flows and lower floods (see Lefèvre, 1974; Peters and Prowse, 2001; Batalla et al., 2004; Magilligan and Nislow, 2005; Maurel et al., 2008; Vernoux and Villion, 2008). Therefore, accounting for the influence of artificial reservoirs is highly important for the proper water management plans downstream the dam.

To account for the influence of reservoirs into the rainfall-runoff model, additional modules are required as the models are originally designed to simulate the natural rainfall-runoff transformation at the watershed scale. There are some studies carried out to integrate the reservoir information into the hydrological model to see the flow variations at the downstream by hydrological models with and without the reservoir information. Most of them are based on a distributed approach (Pociask-Karteczka et al., 2003; Hanasaki et al., 2006; Wu et al., 2007; Zhang et al., 2012). The distributed approach to account for artificial reservoirs will become difficult in case of a large number of reservoirs for which no detailed data are available. In chapter 4, we already discussed the simplicity in using lumped approach in hydrological modelling. Here also we continue with our lumped approach in modelling.

Moulin et al. (2005) conducted a study on the Seine River basin (France) to account for the influence of reservoirs into the lumped GR4J rainfall-runoff model. In their study, they integrated the influence into the lumped GR4J model. Their results show an improvement in the model performance in simulation in all the tested catchments. Following this work, Payan et al. (2008) also conducted a similar study on a set of catchments distributed worldwide, including some in the Seine and Loire basins. The introduction of the reservoir influence was

made by an additional store corresponding to the artificial reservoir, which was linked with the production as well as the routing modules of the model GR4J without any other change in its structure. Their results also indicate an improvement in the model simulation efficiency. Therefore, this chapter focuses on accounting for the influence of reservoirs into a lumped R-R model in order to evaluate the possibility to improve the model's forecast quality.

6.2.Scope of the study

Based on the previous literature, the integration of the information on reservoirs into the model improves the model's simulation efficiency. This improvement in simulation efficiency can improve the model forecast quality which was clearly discussed in chapter 5. Therefore, in this study, our objective is to evaluate to which extent the forecasting efficiency can be improved using this approach. In order to fulfil our objectives, we used a set of catchments in the Seine and Loire basins, which are directly influenced by large dams.

6.3. Test catchments

The data set presented in chapter 2 had been selected in such a way that catchments are lowly influenced by human activities. Here we selected a set of catchments in the Seine and Loire basins which are significantly influenced by reservoirs. The inhabitants demand, agricultural activities, navigation as well as the industrial demand highly motivated this basins' selection.

The Loire River basin, the longest river in France, has more than 11.5 million inhabitants. This basin is also extremely important for farming. Around 350,000 ha farm land in the basin are irrigated. This river is also used for navigation, recreation, generation of hydro and nuclear power from 38 dams and four power stations. One of the recent studies by Mathevet et al. (2007) in the Loire basins gives more information on the possibility of long-lead forecasting of streamflow in the Loire basin which is affected by several influences such as dams.

The Seine is the second longest river in France. Paris, the capital city of France is located at the centre of the drainage network of the Seine River basin with more than 10 million people including the neighbourhood (Billen et al., 2007). The Seine and its tributaries are used for fluvial transportation of agricultural production, timber, construction material etc. This river basin is an intensive agricultural area. There are also several industries in the downstream part of the basin. For more on the socio-economic issues and low flows in the Seine basin, please refer to Bousquet et al. (2003).

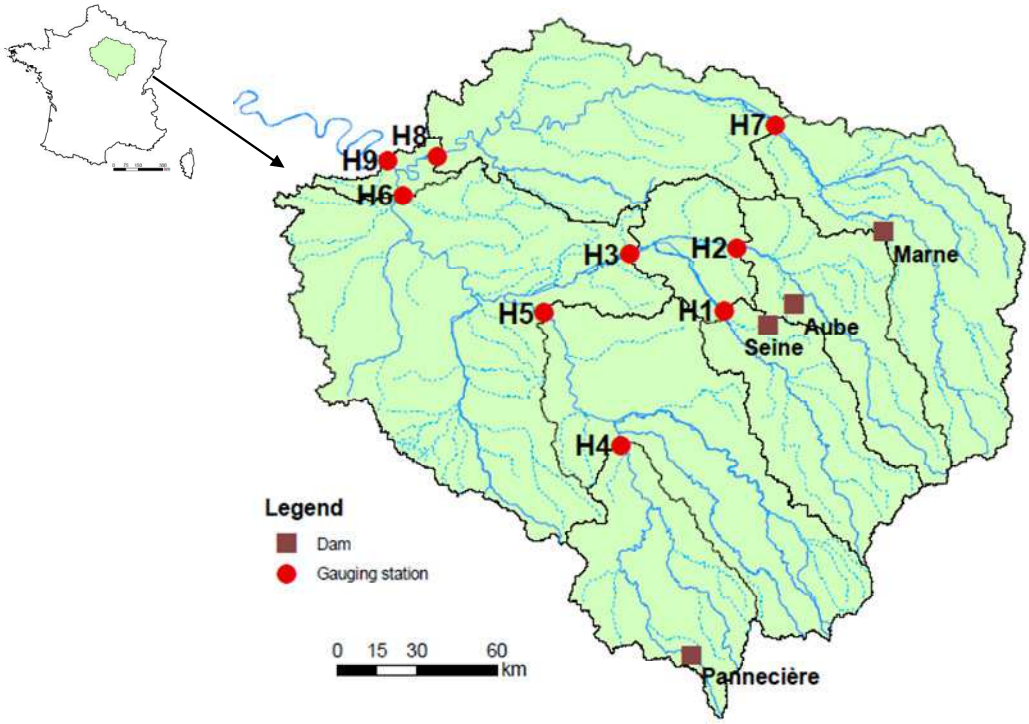


Figure 6.1: Location of test catchments and dams in the Seine river basin

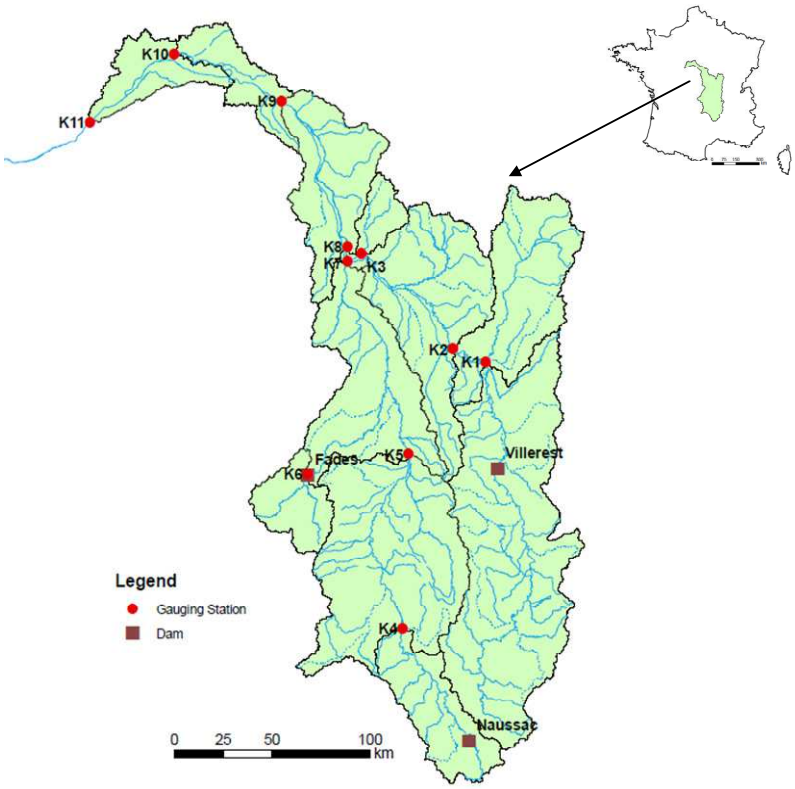


Figure 6.2: Location of catchments and dams in the Loire River basin

Because of these multipurpose activities, these two river basins can offer a clarifying case study with several catchments which are highly influenced by reservoirs.

Figure 6.1 and **Figure 6.2** show the location of the study catchments and dams in the Seine and Loire basins. In our test, we have a set of nine catchments and four dams in the Seine basin and eleven catchments and three dams in the Loire basin respectively. Their flow characteristics and the information on dams are presented in **Table 6.1** and **Table 6.2** for the Seine and Loire basins respectively.

The four dams considered here on the Seine basin are managed by Seine Grands Lacs, a Public Basin Authority (EPTB, www.grandslacsdeseine.fr). For more on the listed dams and their functions in the appropriate rivers in the Seine basin, please refer to Villion (1997). The dams on the Loire basin are managed by the Etablissement Public Loire (www.eptb-loire.fr) and/or by EDF.

6.1.Methodology

The methodology adopted to account for the influence of reservoirs into the lumped hydrological model is presented in the following subsection.

6.1.1. To account for the reservoirs into an R-R model

The methodology proposed by Payan et al. (2008) is used in this study. The reservoirs are grouped in a lumped mode in accordance with the lumped structure of the hydrological model. Below we give a brief description of the adopted methodology adapted from Payan et al. (2008).

Similar to the representation of the behaviour of watershed's hydrological processes in a lumped rainfall-runoff model, the artificial reservoirs over a watershed can be represented as a single store, whose content corresponds to the total stored volume actually observed in the reservoirs. The water depth (mm) in the store corresponds to the total stored volume divided by the catchment area at the downstream station. Thus the content of the store is represented in the same unit as the other internal variables of the model.

Table 6.1: Summary of data set in the Seine basin

Station #	HYDRO code	River	Gauging station	Catchment area (km²)	Concerned Reservoir	Capacity of reservoir (10^6 m^3)	P (mm/yr)	PE (mm/yr)	Q (mm/yr)	MAM7 (mm/yr)
H1	H0800010	Seine	Troyes	3410	Seine	205	853	662	294	68
H2	H1501010	Aube	Arcis	3590	Aube	170	829	660	296	65
H3	H1700010	Seine	Pont	9760	Seine, Aube	375	928	678	336	92
H4	H2221010	Yonne	Gurgy	3820	Pannecière	80	846	676	270	77
H5	H2721010	Yonne	Courlon	10700	Pannecière	80	777	665	223	75
H6	H4340010	Seine	Villeneuve St-Georges	30800	Seine, Aube, Pannecière	455	829	677	277	99
H7	H5201010	Marne	Charlons sur Marne	6280	Marne	350	960	668	361	74
H8	H5841010	Marne	Noisiel	12500	Marne	350	849	640	278	89
H9	H5920010	Seine	Paris	43800	Seine, Aube, Pannecière, Marne	805	814	676	224	67

Table 6.2: Summary of data set in the Loire basin

Station #	HYDRO code	River	Gauging station	Catchment area (km²)	Concerned Reservoir	Capacity of reservoir (10^6 m^3)	P (mm/yr)	PE (mm/yr)	Q (mm/yr)	MAM7 (mm/yr)
K1	K1180010	Loire	Digoin	9315	Villerest	106.2	892	633	310	39
K2	K1440010	Loire	Gilly-sur-Loire	13007	Villerest	106.2	907	634	371	41
K3	K1930010	Loire	Nevers	17570	Villerest	106.2	899	660	329	45
K4	K2330810	Allier	Vieille-Brioude	2269	Naussac	190	895	575	407	62
K5	K3030810	Allier	Saint-Yorre	8940	Naussac	190	923	619	354	62
K6	K3292020	Sioule	Saint-Priest-des-Champs	1300	Fades	69	1003	614	433	60
K7	K3650810	Allier	Cuffy	14310	Fades, Naussac	259	887	625	353	60
K8	K4000010	Loire	Cours-les-Barres	36610	Fades, Naussac, Villerest	365.2	898	660	296	61
K9	K4180010	Loire	Gien	35500	Fades, Naussac, Villerest	365.2	856	668	269	57
K10	K4350010	Loire	Orléans	36970	Fades, Naussac, Villerest	365.2	882	647	309	47
K11	K4470010	Loire	Blois	38320	Fades, Naussac, Villerest	365.2	857	645	310	65

The temporal evolution of the water volume stored in the reservoir is not natural and it depends on management decisions. Therefore the store that represents artificial reservoir will simply reflect the observations of the water storages. Therefore, the behaviour of this store is completely defined by observations and there is no parameterisation needed.

Let us consider the level of the store is V_i at time step i and V_{i+1} at time step $i+1$. If $V_{i+1} > V_i$, the stored volume has increased, a positive quantity of water $\Delta V = V_{i+1} - V_i$ has globally been taken in by the artificial reservoir from the natural system. Conversely, if $V_{i+1} < V_i$, the stored volume has decreased, a quantity of water has globally left the artificial reservoirs to feed the natural system. If $V_{i+1} = V_i$, there is no water exchange. Therefore, one can consider that the algebraic quantity ΔV represents the water exchanges between the lumped representation of the watershed and the lumped representation of the artificial reservoirs. ΔV is only defined by the observed values of the stored volumes, and therefore constitutes a new input to the model. **Figure 6.3** illustrates how artificial reservoirs are integrated into the GR6J model.

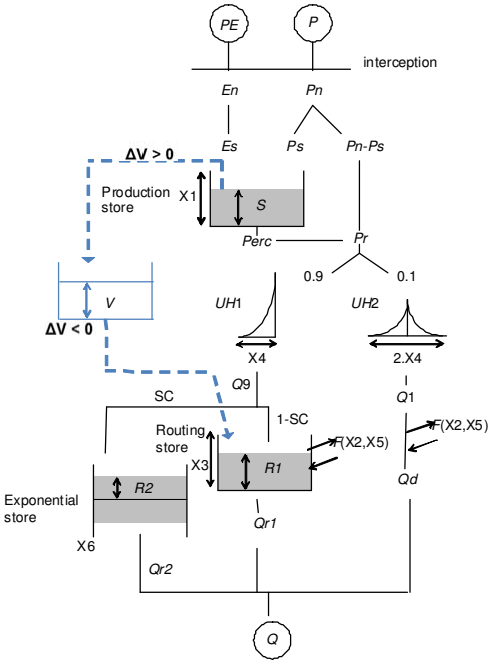


Figure 6.3: Illustration of how to account for the artificial reservoir volume variations (ΔV) in the GR6J model

Figure 6.3 shows that, if ΔV is positive, this quantity should be subtracted from the production store and if ΔV is negative, this quantity must be added to the routing module.

Here ΔV is added only to the non-linear power store. The trial on splitting ΔV into the routing stores did not make any significant change in the overall model performance.

The test was also conducted with the GR4J (which was already tested by Payan et al. (2008)) and GR5J models to make a comparative analysis in their performances with the influence of artificial reservoirs.

6.1.2. Model testing and evaluation

The model testing scheme was already discussed in chapter 3 and here we are not going into the details. The objective function for calibration was the square root transform (NSE^*_{sqrtQ}) and the criteria used to evaluate the model performance in low-flow simulation mode are the NSE^*_{lnQ} and NSE^*_{iQ} respectively. The model forecast quality was assessed based on the Brier Skill Score (BSS) at the two thresholds Q_{75} and Q_{90} which we discussed in chapter 5.

6.2. Results and discussion

This section illustrates some of the main results obtained in simulation as well as forecast modes. The results were grouped for the two basins and we conclude our results based on comparing the model performances on these two basins.

6.2.1. Can we really improve the simulation performance of models by accounting the reservoirs?

The model simulation performances on the Seine and Loire basins are discussed and are illustrated in **Figure 6.4** and **Figure 6.5** respectively. The overall performance values for the three models indicate the better performance of the GR6J model which supports our conclusion in Chapter 4. The models efficiencies are better in the Loire basin compared to the Seine, and the gain in efficiency by accounting for the dams is much lower in the case of the Loire basin. But in both cases, there is a strong improvement in NSE^*_{iQ} (which is our target criterion for low flows) compared to the NSE^*_{LnQ}.

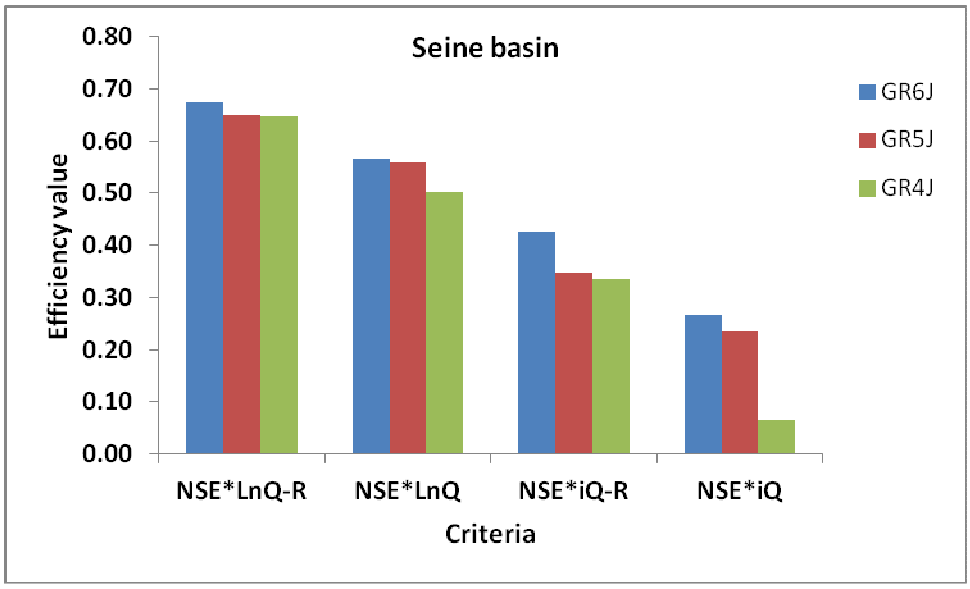


Figure 6.4: Illustration of model's efficiency values with (R) and without reservoirs

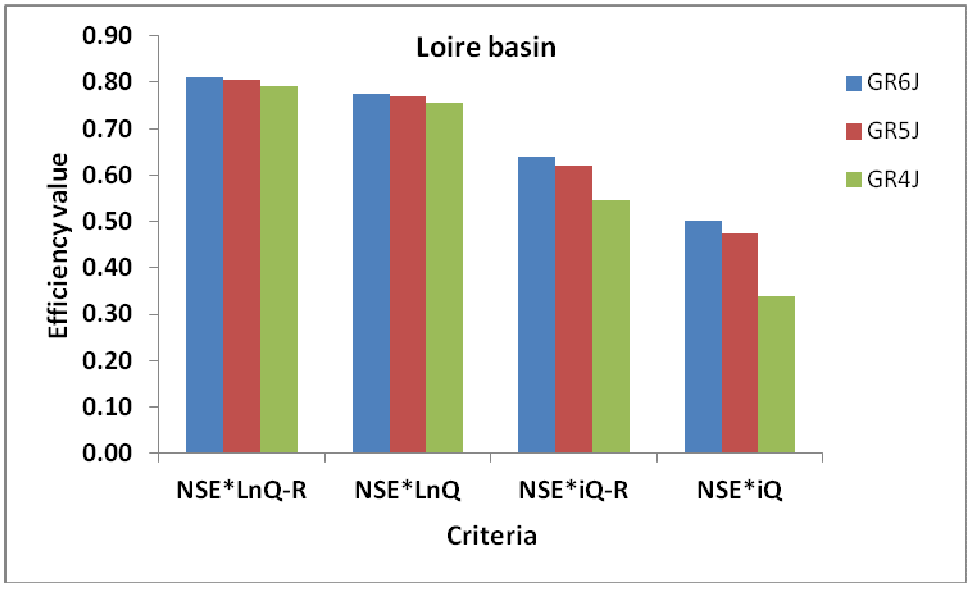


Figure 6.5: Illustration of model's efficiency values with and without reservoirs (R)

As GR6J has better overall efficiency values compared to GR4J and GR5J models, in the following sections we concentrate on the performances of GR6J in simulation as well as in forecast modes.

The analysis of the efficiency criteria for the individual catchments in the two basins using the GR6J model are shown in **Figure 6.6**. The points above the diagonal line indicates the improved performance by accounting for the dams. Most of the catchments show an improved performance, although model efficiency is lower on a few of them, especially on the NSE^*_{iQ}

criterion on the Loire basin. It means that the proposed model version to account for the influence is not optimal in all the cases. This corroborates the results by Payan et al. (2008) who showed that the introduction of dams with a "one-size-fits-all" solution was not optimal in all cases, since some factors (e.g. distance between dam and downstream gauging station) are not accounted for.

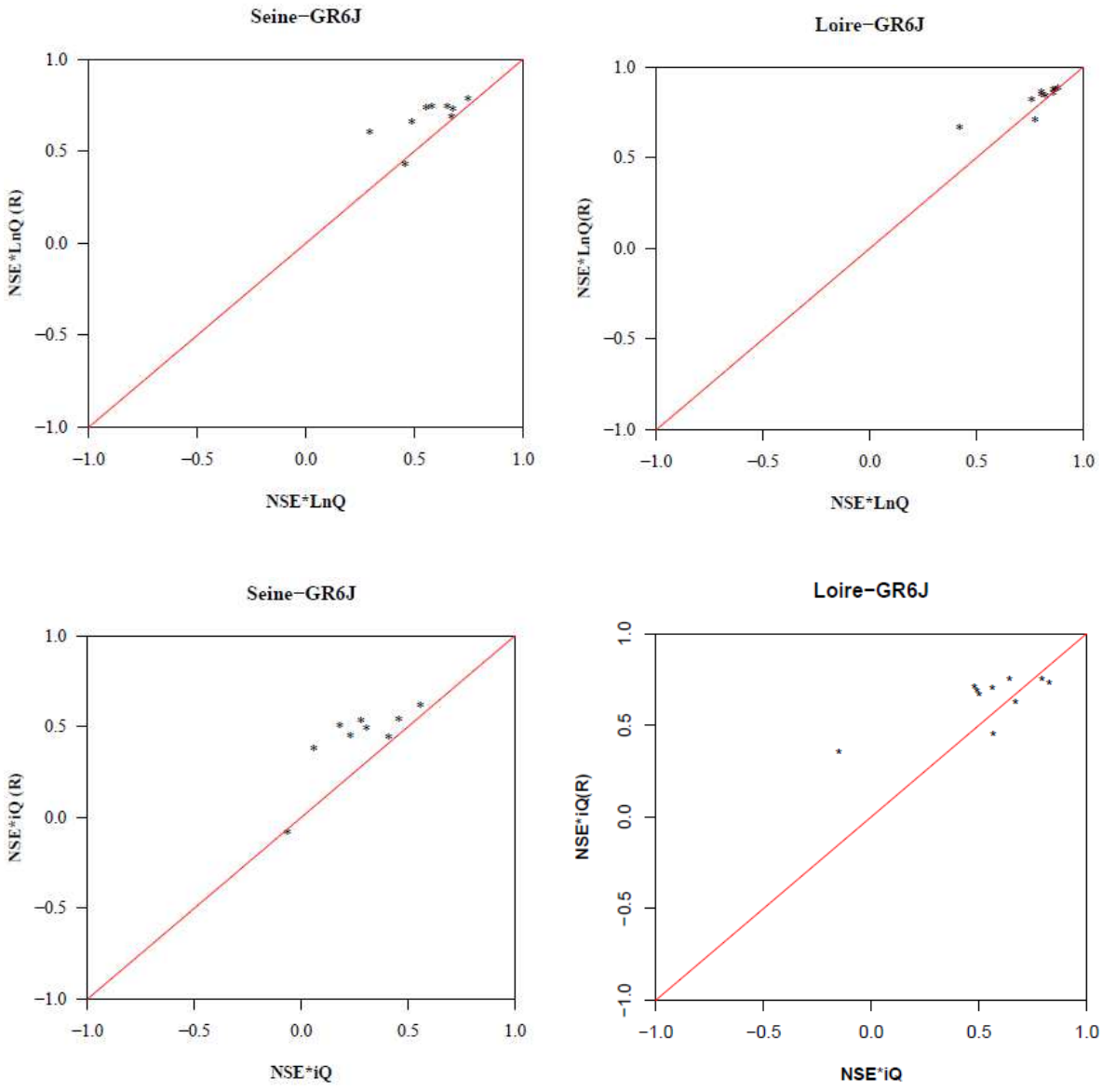


Figure 6.6: Comparison of model performances with and without the influence of reservoirs using the NSE^*_{LnQ} and NSE^*_{iQ} criterion

6.2.1.1. Hydrograph of GR6J simulations

Figure 6.7 illustrates the simulated and observed hydrographs of catchments H1 for a period of four years (2000-2003) which includes one of the driest periods (2003) in the last decade in France. The blue line represents the model simulation with the integration of artificial influences and the red line represents the simulation without influences. As we discussed in the previous section, the integration of influences caused an improvement in the sample catchment. The calculated NSE^*_{iQ} for the same time-period also is presented in **Figure 6.7**.

The next section discusses on the influence of simulation efficiency on model forecast quality.

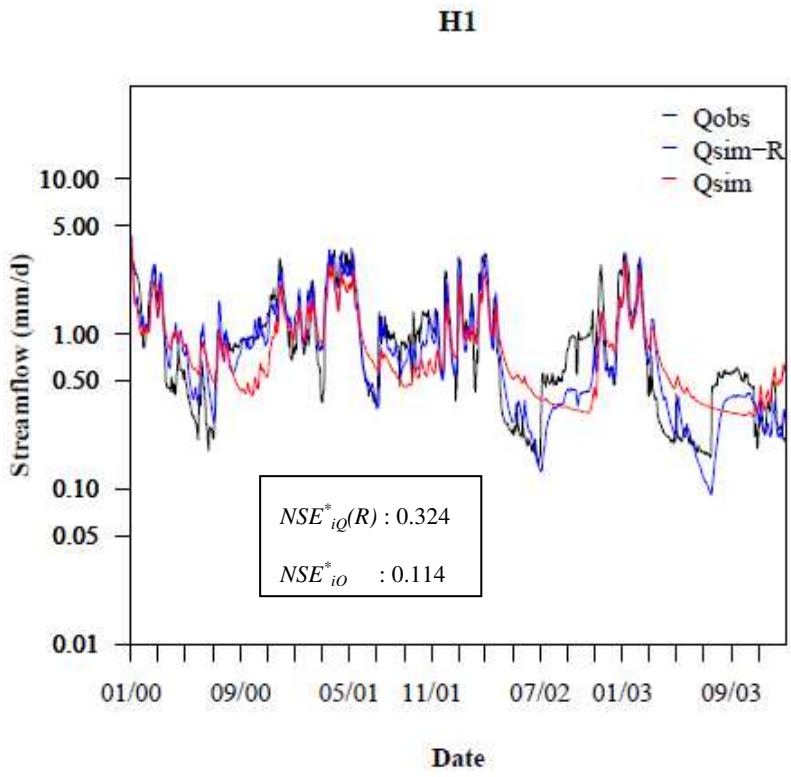


Figure 6.7: Illustration of the observed and simulated hydrographs (with and without R) of one of the test catchments for a period of four years (2000-2003) and its efficiency values

6.2.2. Can we improve the forecast quality?

In chapter 5, we discussed the link between simulation efficiency and the corresponding forecast quality of the model (see section 5.7.1). Here again we tried to link the same with the additional information on the influence of reservoirs into the model. We tested the quality of forecasts issued using the GR6J model with and without the integration of the influence of

reservoirs. The quality of forecasts was assessed with the BSS_{75} and BSS_{90} (Brier Skill Score at Q_{75} and Q_{90}).

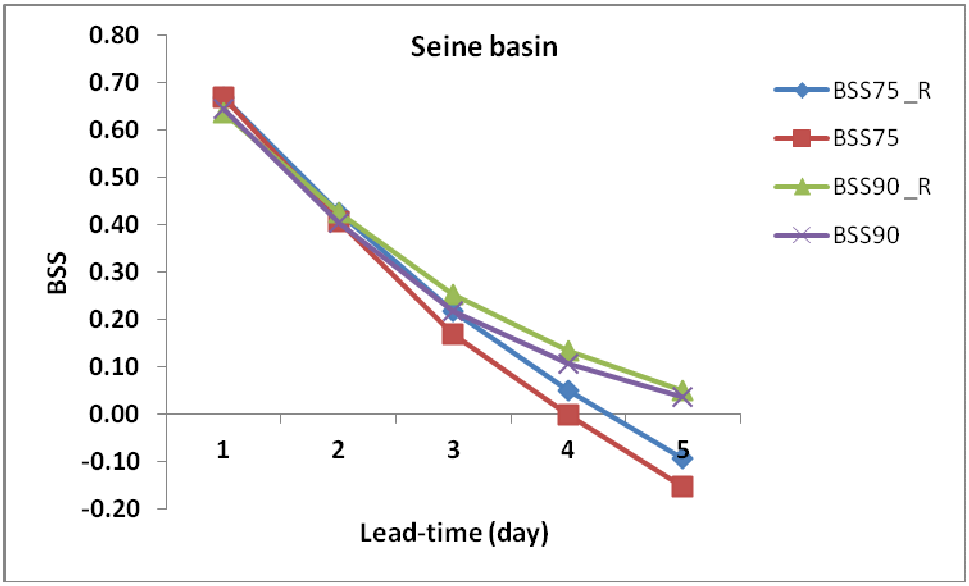


Figure 6.8: Illustration of forecast skill score of the model GR6J in the Seine basin

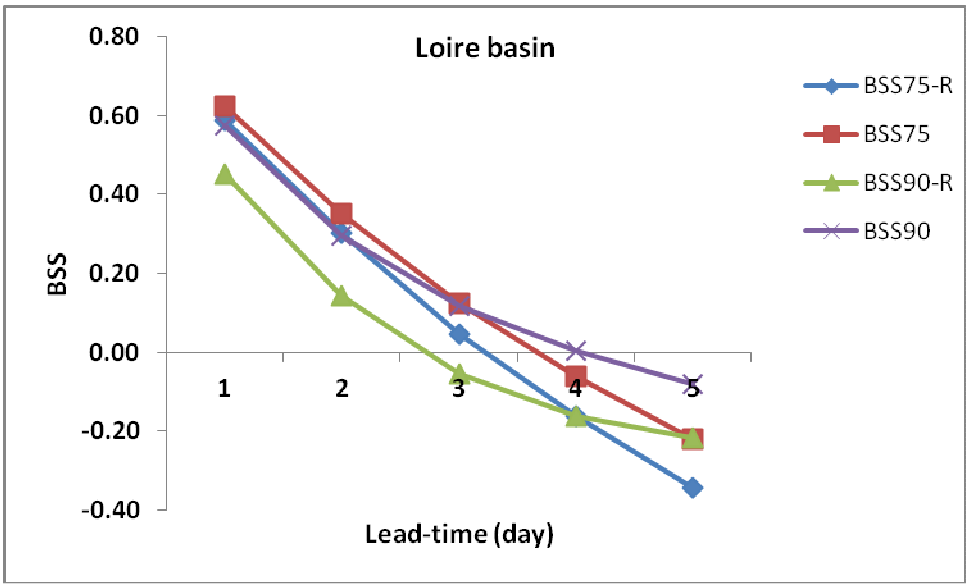


Figure 6.9: Illustration of forecast skill score of the model GR6J in the Loire basin

The mean values for the data set in the basins for each lead-time was used here to make the plots. **Figure 6.8** shows a slight upward shift in the curves of BSS_{75} and BSS_{90} when reservoirs are considered on the Seine basin, which indicates an improvement in forecast quality. Also, the positive skill values of BSS_{90} in both conditions (with and without R) indicates the better model simulations in the very low-flow conditions for the tested

catchments. This is an indication of better low-flow simulation with the additional information on reservoirs. But, **Figure 6.9** is an illustration of the negative impact of the reservoirs into the model's forecast quality. This shows that the proposed methodology is not suitable for this basin and further analysis is needed to improve the quality on these catchments.

6.3. Conclusion

The objective to account for the influence of reservoirs into the GR6J model provided satisfactory results in simulation mode. The proposed methodology is suitable in the Seine basin and it needs further modifications to cope with the Loire basin. Also, the small change in forecast quality compared to the change in simulation efficiency in the Seine basin itself indicates further steps are needed for accounting reservoirs so that we can have a significant improvement in the forecast quality. This study also suggests conducting further studies on this issue.

General conclusion

Synthesis

The thesis aimed at developing an ensemble approach for long-term low-flow forecasts on a large set of catchments distributed over France. In the first part, we discussed stream low flows and the low-flow generating factors. Since the evaluation of model simulations is still a matter of debate in operational hydrology, while the literature is very limited in the specific case of low-flow simulations, we focused chapter 3 on the analysis of different efficiency criteria, and on their ability to evaluate the efficiency of hydrological models to simulate low-flows. This analysis resulted in the proposition of a squared-error criterion, based on inverse transformed flows, which better evaluates the model simulations in low-flow conditions than the existing criteria.

In chapter 4, we investigated the suitability of existing rainfall-runoff (R-R) models for low-flow simulation. This led to the development of an improved model structure, better representing the low-flow conditions on average on our set of catchments. Chapter 4 briefly describes the step-by-step model development. It was mainly achieved by the structural sensitivity analysis of the tested models. Although the overall improvement in model efficiency is modest, it should contribute to better low-flow forecasts.

As a continuation, in chapter 5, we focused on the integration of this improved model version (GR6J) into an ensemble approach to issue long-term low-flow forecasts. The ensembles are made of the historic time series, which are injected into the model to produce ensemble forecasts. We evaluated how forecast uncertainty depended on forecast lead time. To reduce the uncertainty and to improve the forecast skill in long-term basis, we tested several straightforward updating techniques in the forecasting approach. The results suggest that simple bias correction factors depending on the forecast lead time can significantly improve the skill score. Evaluations were made using the Brier skill score, one of the most widely used probabilistic score. Based on this skill score, we identified the maximum possible lead times for the individual catchments.

In chapter 6, we described the application of low-flow forecasting on catchments influenced by large dams. In such cases, we needed to integrate the information on reservoirs into the forecast model, so that we could improve the quality of forecasts. In this chapter we clearly discussed the possibility of improvement in the quality with the consideration of reservoirs.

The results suggest that the adopted methodology to account for the reservoirs is not optimal in all the tested catchments. Further investigations are needed.

Perspectives

We noticed that low-flow studies are still too limited and that we will need more attention in this domain due to the possibly more frequent occurrence of stream low flows in the future and their high impact on the society. This study addressed some of the relevant issues in low-flow hydrology, in terms of improved simulation and long-lead forecasting. This study highlighted the necessity to continue the effort on low-flow hydrology and forecast, in particular in the following directions:

- The level of performance of the model GR6J in low-flow conditions has improved but remains quite modest. This may be due to the structural model errors, data quality or human influences. Further studies are needed to find ways to improve low flow simulation. Better accounting for human influences, especially abstractions, may bring significant improvements. Building comprehensive databases using the information from Water Agencies, ONEMA, etc., would be very useful in this perspective. More efficient methodologies may also be sought to better account for dams, and for links between surface and groundwater.
- Further investigations are also needed to better understand the specificities of low-flow forecasts. The basic updating or correction methods we evaluated in our exploratory tests significantly improved forecasting efficiency. More complex methods to update or correct models may provide better results and should be tested. The use of upstream-downstream relationship or regional coherence may also help improving forecasts. In terms of meteorological scenarios, investigating the usefulness of information provided by meteorological models would be useful. This may result in more reliable forecast at longer lead-times.
- Last, the GR6J simulation and forecasting model proposed here should be further tested and evaluated against other models aimed at making low-flow predictions. ONEMA and the Water & Biodiversity Direction from the Ministry of Ecology launched in 2010 a project to benchmark existing low-flow forecasting models. Several models developed by BRGM, Cegum (Univ. Lorraine), EDF-DTG, Météo-

France and Irstea are currently evaluated on French catchments. This may provide useful information for end-users to start implementing advanced operational tools to anticipate and better manage low-flow conditions.

References

[1] Ackroyd, E. A., W.C. Walton, and D.L. Hills, 1967. Ground-water contribution to streamflow and its relation to basin characteristics in Minnesota, Report of Investigations 6, University of Minnesota**:** 36p.

[2] Adler, M., and V. Ungureanu, 2006. Multi-model technique for low-flow forecasting. Climate Variability and Change- Hydrological Impacts (Proceedings of the Fifth FRIEND World Conference. Havana, Cuba, IAHS Publ. 308**:** 151-157.

[3] Anctil, F., C. Perrin, and V. Andreassian, 2003. ANN Output updating of lumped conceptual rainfall-runoff forecasting models. Journal of American Water Resources Association, 36(5): 1269-1279.

[4] Andréasian, V., A. Hall, N. Chahinian, and J. Schaake, 2006a. Introduction and synthesis: why should hydrologists work on a large number of basin data set? Large sample experiments for hydrological model parameterisation: results of the model parameter experiment - MOPEX. IAHS Publ, 307: 1-5.

[5] Andréasian, V., S. Bergström, N. Chahinian, Q. Duan, Y.M. Gusev, I. Littlewood, T. Mathevet, C. Michel, A. Montanari, G. Moretti, R. Moussa, O.N. Nasonova, K. O'Connor, E. Paquet, C. Perrin, A. Rousseau, J. Schaake, T. Wagener, and Z. Xie, 2006b. Catalogue of the models used in MOPEX 2004/2005, Large sample basin experiments for hydrological model parameterization: Results of the model parameter experiment-MOPEX. IAHS Publ, 307: 41-93.

[6] Andréassian, V., 2004. Waters and Forests: from historical controversy to scientific debate. Journal of Hydrology, 291: 1-27.

[7] Andréassian, V., C. Perrin, E. Parent, and A. Bardossy, 2010. Court of Miracles of Hydrology: can failure stories contribute to hydrological science? . Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 55(6): 849-856.

[8] Andréassian, V., E. Parent, and C. Michel, 2003. A distribution-free test to detect gradual changes in watershed behaviour. Water Resources Research, 39(9): 1252.

[9] Andréassian, V., C. Perrin, and C. Michel, 2004. Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. Journal of Hydrology, 286: 19-35.

[10] ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Management Committee, I. a. D. D., 1993. Criteria for Evaluation of Watershed Models. Journal of Irrigation and Drainage Engineering-Asce, 119(3): 429-442.

[11] Atger, F., 2001. Verification of intense precipitation forecasts from single models and ensemble prediction systems. Nonlinear Processes in Geophysics, 8: 401-417.

[12] Aubert, D., L. Cécile, and L. Oudin, 2003. Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model. Journal of Hydrology, 280: 145-161.

[13] Avalos Lingan, R. F., 1976. Essais methodologiques de prevision des debits d'etiages: Application au bassin de l'Oise a Sempigny. Sciences de la Terre. Montpellier, Universite des Sciences et Techniques du Languedoc. Thése de Doctorat**:** 145 p.

[14] Babovic, V., R. Canizares, H.R. Jensen, and A. Klinting, 2001. Neural networks as routine for error updating of numerical models. Journal of Hydraulic Engineering, 127: 181-193.

[15] Bartholmes, J. C., J. Thielen, M.H. Ramos, and S. Gentilini, 2009. The European flood alert system EFAS - Part 2: statistical skill assessment of probablistic and deterministic operational forecasts. Hydrological Earth System Sciences, 13: 141-153.

[16] Batalla, R. J., C.M. Gomez, and G.M. Kondolf, 2004. Reservoir-induced hydrological changes in the Ebro River basin (NE Spain). Journal of Hydrology, 290: 117-136.

[17] Bergström, S., and A. Forsman, 1973. Development of a conceptual deterministic rainfall-runoff model. Nordic Hydrology, 4: 147-170.

[18] Berthet, L., 2010. Prévision des crues au pas de temps horaire: pour une meilleure assimilation de l'information de débit dans un modèle hydrologique. Unité de recherche Hydrosystèmes et Bioprocédés, Cemagref. Antony, AgroParisTech. Thèse de Doctorat**:** 626p.

[19] Berthet, L., V. Andréassian, C. Perrin, and P. Javelle, 2009. How crucial is to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. Hydrological Earth System Sciences, 6: 1707-1736.

[20] Beven, K. J., and M.J. Kirby, 1979. A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 24(1): 43-69.

[21] Billen, G., J. Garnier, J.-M. Mouchel, and M. Silvestre, 2007. The Seine system: Introduction to a multidisciplinary approach of the functioning of a regional river system. Science of the Total Environment, 375: 1-12.

[22] Blenkinsop, S., and H.J. Fowler, 2007. Changes in European drought characteristics projected by the PRUDENCE regional climate models. International Journal of Climatology, 27: 1595-1610.

[23] Boé, J., L. Terray, E. Martin, and F. Habets, 2009. Projected changes in components of the hydrological cycle in French river basins during the 21st century. Water Resources Research, 45: W08426.

[24] Boucher, M. A., L. Perreault, and F. Anctil, 2009. Tools for the assessment of hydrological ensemble forecasts obtained by neural networks. Journal of Hydroinformatics, 11(3-4): 297-307.

[25] Bourgin, F., 2009. Calibration of a catchment-based land surface model in the Loire River basin (France) to assess hydrological impacts of climate change. Environmental Engneering, Technische Universitat Munchen. Master of Science**:** pp60.

[26] Bousquet, S., E. Gaume, and B. Lancelot, 2003. Évaluation des enjeux socio-économiques liés aux étiages de la Seine. La Houille Blanche, 3: 145-149.

[27] Box, G. E. P., and D.R. Cox, 1964. An analysis of transformations. J. of Roy. Stat. Soc. , 26: 211-252.

[28] Bradley, A. A., S.S. Schwartz, and T. Hashino, 2004. Distributions-Oriented Verification of Ensemble Streamflow Predictions. Journal of Hydrometeorology, 5: 532-545.

[29] Bravo, J. M., A.R. Paz, W. Collischonn, C.B. Uvo, O.C. Pedrollo, and S.C. Chou, 2009. Incorporating Forecasts of Rainfall in Two Hydrologic Models Used for Medium-Range Streamflow Forecasting. Journal of Hydrologic Engineering, 14(5): 435-445.

[30] Brier, G. W., 1950. Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78(1).

[31] Brown, J. D., and D.-J. Seo, 2010. A Nonparametric Postprocessor for Bias Correction of Hydrometeorological and Hydrologic Ensemble Forecasts. Journal of Hydrometeorology, 11: 642-665.

[32] Calanca, P., 2007. Climate change and drought occurrence in the Alpine region: How severe are becoming the extremes? G. Global and Planetary Change, 57: 151-160.

[33] Campolo, M., A. Soldati, et al., 1999. Forecasting river flow rate during low-flow periods using neural networks. Water Resources Research, 35(11): 3547-3552.

[34] Cavitte, J.-P., and J.-F. Moor, 2004. Les plans de gestion des étiages sur le bassin Adour-Garonne. Un exemple modèle de partenariat du monde agricole avec les organismes de bassin. La Houille Blanche, 1: 26-30.

[35] Céron, J. P., G. Tanguy, L. Franchistéguy, E. Martin, F. Regimbeau, and J.P. Vidal, 2010. Hydrological seasonal forecast over France: feasibility and prospects. Atmospheric Science Letters, 11: 78-82.

[36] Chiew, F. H. S., 2010. Lumped Conceptual Rainfall-Runoff Models and Simple Water Balance Methods: Overview and Applications in Ungauged and Data Limited Regions. Geography Compass, 4(3): 206-225.

[37] Chiew, F. H. S., and T.A. McMahon, 1993. Assessing the adequacy of catchment streamflow yield estimates. Australian Journal of Soil Research, 31(5): 665-680.

[38] Chiew, F. H. S., and T.A. McMahon, 2002. Global ENSO-streamflow teleconnection, streamflow forecasting and interannual variability. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 47(3): 505-522.

[39] Chiew, F. H. S., M.J. Stewardson, and T.A. McMahon, 1993. Comparison of six rainfall-runoff modelling approaches. Journal of Hydrology, 147: 1-36.

[40] Chiew, F. H. S., S.L. Zhou, and T.A. McMahon, 2003. Use of seasonal streamflow forecasts in water resources management. Journal of Hydrology, 270: 135-144.

[41] Chopart, S., and E. Sauquet, 2008. Usage des jaugeages volants en régionalisation des débits d'étiage. Revue des Sciences de l'Eau, 21(3): 267-281.

[42] Clarke, R. T., 2008. A critique of present procedures used to compare performance of rainfall-runoff models. Journal of Hydrology, 352: 379-387.

[43] Clausen, B., and K.R. Rasmussen, 1993. Low Flow Estimation and Hydrogeology in a Chalk Catchment. Nordic Hydrology, 24: 297-308.

[44] Cloke, H. L., and F. Pappenberger, 2009. Ensemble flood forecasting: A review. Journal of Hydrology, doi:10.1016/j.jhydrol.2009.06.005.

[45] Coeur, D., 2004. La connaissance historique des étiages extrêmes: quelles sources d'information? La Houille Blanche, 4: 84-86.

[46] Cunderlik, J. M. and S. P. Simonovic, 2005. Hydrological extremes in a southwestern Ontario river basin under future climate conditions. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 50(4): 631-654.

[47] Cuo, L., T.C. Pagano, Q.J. Wang, 2011. R review of quantitative precipitation forecasts and their use in short-to medium-range streamflow forecasting. Journal of Hydrometeorology, 12: 713-728.

[48] Dakova, S., Y. Uzunov, and D. Mandadjiev, 2000. Low flow - the river's ecosystem limiting factor. Ecological Engineering, 16: 167-174.

[49] Davison, B., and G. van der Kamp, 2008. Low flows in deterministic modeling: a brief review. Canadian Water Resources Journal, 33(2): 181-193.

[50] Dawson, C. W., R.J. Abrahart, and L.M. See, 2007. Hydro Test: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environmental modelling and software, 22(7): 1034-1052.

[51] de Vos, N. J., T.H.M. Rientjes, and H.V. Gupta, 2010. Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering. Hydrological Processes, 24(20): 2840-2850.

[52] de Wit, M. J. M., B. van den Hurk, P.M.M. Warmerdam, P.J.J.F. Torfs, E. Roulin, and W.P.A. van Deursen, 2007. Impact of climate change on low-flows in the river Meuse. Climatic Change, 82(3): 351-372.

[53] Demuth, S., 2005. Low flows and droughts-A European prospective. IAHS News letter 82, March 2005**:** 7-8.

[54] Desaint, B., P. Nogues, C. Perret, and R. Garçon, 2009. La prévision hydrométéorologique opérationnelle: l'expérience d'Electricité de France. La Houille Blanche, 5: 39-46.

[55] Despriée, J., P. Voinchet, H. Tissoux, J.-J. Bahain, C. Falguères, G. Courcimault, J. Dépont, M.-H. Moncel, S. Robinson, M. Arzarello, R. Sala, L. Marquer, E. Messager, S. Puaud, and S. Abdessadok, 2011. Lower and Middle Pleistocene human settlements recorded in fluvial deposits of the middle Loire River Basin, Centre Region, France. Quarternary Science Reviews, 30: 1474-1485.

[56] Diaz-Nieto, J., and R.L. Wilby, 2005. A comparison of statistical downscaling and climate change factor methods: Impacts on low flows in the river Thames, United Kingdom. Climate Change, 69: 245-268.

[57] Drogue, G., L. Pfister, T. Leviandier, A.E. Idrissi, J.-F. Iffly, P. Matgen, J. Humbert, and L. Hoffmann, 2004. Simulating the spatio-temporal variability of streamflow response to climate change scenarios in a mesoscale basin. Journal of Hydrology, 293: 255-269.

[58] Duband, D., 2010. Rétrospective hydro-pluviométrique des étiages rares depuis 140 ans, dans l'ouest de l'Europe (bassins Loire, Seine, Rhin, Rhône, Pô). La Houille Blanche, 4: 51-59.

[59] Duband, D., P. Schoeneich, and V.A. Stanescu, 2004. Exemple de l'étiage 1921 en Europe (Italie, France, Roumanie, Suisse,...): climatologie et hydrologie. La Houille Blanche, 5: 18-29.

[60] Edijatno, and C. Michel, 1989. Un modèle pluie-débit journalier à trois paramètres. La Houille Blanche, 2: 113-121.

[61] Edijatno, X. Y., Z. Makhlouf, and C. Michel, 1999. GR3J: a daily watershed model with three free parameters. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 44(2): 263-277.

[62] Efstratiadis, A., and D. Koutsoyiannis, 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. Journal of Hydrological Sciences, 55(1): 58-78.

[63] Eheart, J. W. and D. W. Tornil, 1999. Low-flow frequency exacerbation by irrigation withdrawals in the agricultural midwest under various climate change scenarios. Water Resources Research, 35(7): 2237-2246.

[64] Fenicia, F., H.H.G. Savenije, P. Matgen, and L. Pfister, 2006. Is the groundwater reservoir linear? Learning from data in hydrological modelling. Hydrological Earth System Sciences, 10: 139-150.

[65] Feyen, L., and R. Dankers, 2009. Impact of global warming on streamflow drought in Europe. Journal of Geophysical Research, 114: D17116.

[66] Fortin, V., and R. Turcotte, 2007. Le modéle hydrologique MOHYSE. Note de cours pour SCA7420, Université du Québec à Montréal : Département des sciences de la terre et de l'atmosphère.

[67] Furey, P. R. and V. K. Gupta, 2000. Space-time variability of low streamflows in river networks. Water Resources Research, 36(9): 2679-2690.

[68] Garçon, R., 1996. Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR. Bilan de l'année 1994-1995. La Houille Blanche, 5(71-76)

[69] Garçon, R., 1999. Modéle global Pluie-Débit pour la prévision et la prédétermination des crues (Lumped rainfall-runoff model for flood forecasting and design flood estimation, in French). La Houille Blanche, 7/8: 88-95.

[70] Garçon, R., C. Carré, and P. Lyaudet, 1999. Exemple de prévision et de simulation opérationnelle des débits d'étiage pour les besoins d'EDF. La Houille Blanche, 6: 37-42.

[71] Garnier, E., 2010. Bassesses extraordinaires et grandes chaleurs. 500 ans de sécheresses et de chaleurs en france et ans les pays limitrophes. La Houille Blanche, 4: 26-42.

[72] Garrick, M., C. Cunnane, and J.E. Nash, 1978. A criterion of efficiency for rainfall-runoff models. Journal of Hydrology, 36(3-4): 375-381.

[73] Gaume, E., 2004. Sécheresses et étiages: quelques réflexions. La Houille Blanche, 4: 77-83.

[74] Gazelle, F., 1979. Les répercussions des étiages de 1976 sur l'hydroélectricité languedocienne. La Houille Blanche, 1: 59-61.

[75] Girard, G., 1977. Etude de l'efficacité relative et du domaine d'application de différents modèles hydrologique pour prévoir les étiages. Paris, Bureau Central Hydrologique, Report**:** 115p.

[76] Guilbot, A., G. Bédiot, J.-M. Masson, and C. Ducastelle, 1976. Essai de prévision des étiages de l'Oise à Sempigny. La Houille Blanche, 6(7): 549-568.

[77] Giuntoli, I., and B. Renard, 2009. Identification des impacts hydrologiques du changement climatique: vers un réseau de référence pour la surveillance des étiages, Rapport d'étude. Convention ONEMA-Cemagref 2009**:** 86p.

[78] Goswami, M., K.M. O'Connor, K.P. Bhattarai, and A.Y. Shamseldin, 2005. Assessing the performance of eight real-time updating models and procedures for the Brosna River. Hydrology and Earth System Sciences, 9(4): 394-411.

[79] Gupta, H. V., and H. Kling, 2011. On typical range sensitivity and normalization of mean squared error and Nash-Sutcliffe efficiency type metrics. Water Resources Research, 47: W10601.

[80] Gupta, H. V., H. Kling, K.K. Yilmaz, and G.F. Martinez, 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. Journal of Hydrology, 377(1-2): 80-91.

[81] Gustard, A., A., Bullock, and J.M., Dixon, 1992. Low flow estimation in the United Kingdom, Institute of Hydrology, Report**:** 88.

[82] Hall, M. J., 2001. How well does your model fit the data? . Journal of Hydroinformatics, 3(1): 49-55.

[83] Hamlet, A. F., D. Huppert, and D.P. Lettenmaier, 2002. Economic value of long-lead streamflow forecasts for Columbia River Hydropower. Journal of Water Resources Planning and Management, 128(2): 91-101.

[84] Hanasaki, N., S. Kanae, and T. Oki, 2006. A reservoir operation scheme for global river routing models. Journal of Hydrology, 327: 22-41.

[85] Hashino, T., A.A. Bradley, and S.S. Schwartz, 2007. Evaluation of bias-correction methods for ensemble streamflow volume forecasts. Hydrological Earth System Sciences, 11: 939-950.

[86] Higgs, G., and G.E. Petts, 1988. Hydrological changes and river regulation in the UK. Regulated Rivers: Research & Management, 2: 349-368.

[87] Hill, N. M., P.A. Keddy,and I.C. Wisheu, 1998. A Hydrological Model for Predicting the Effects of Dams on the Shoreline Vegetation of Lakes and Reservoirs. Environmental Management, 22(5): 723-736.

[88] Hisdal, H., K. Stahl, L.M. Tallaksen, and S. Demuth, 2001. Have stremaflow droughts in Europe become more severe or frequent? International Journal of Climatology, 21: 317-333.

[89] Hoog, B. V. D. W., 1995. The effects of climate variability on discharge as dependent on catchment characteristics in the Upper Loire basin, France. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 40(5): 633-646.

[90] Houghton-Carr, H. A., 1999. Assessment criteria for simple conceptual daily rainfall-runoff models. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 44(2): 237-261.

[91] Hundecha, Y. and A. Bardossy, 2004. Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. Journal of Hydrology, 292(1-4): 281-295.

[92] Iacobellis, V., 2008. Probabilistic model for the estimation of T year flow duration curves. Water Resources Research, 44(2): w02413.

[93] Isik, S., E. Dogan, L. Kalin, M. Sasal, and N. Agiralioglu, 2008. Effects of anthropogenic activities on the Lower Sakarya River. Catena, 75(2): 172-181.

[94] Jacob-Rousseau, N., and L. Astrade, 2010. Sécheresses et pénuries d'eau dans la France méridionale entre 1815 et 1840. La Houille Blanche, 4: 43-50.

[95] Jain, S. K., and K.P. Sudheer, 2008. Fitting of Hydrologic Models: A Close Look at the Nash-Sutcliffe Index. Journal of Hydrologic Engineering, 10: 981-986.

[96] Jakeman, A. J., I.G. Littlewood, and P.G. Whitehead, 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology, 117: 275-300.

[97] Johnson, A. C., M.C. Acreman, M.J. Dunbar, S.W. Feist, A.M. Giacomello, R.E. Gozlan, S.A. Hinsley, A.T. Ibbotson, H.P. Jarvie, J.I. Jones, M.M. Longshaw, C.M. Stephen, J.N. Terry, N. Colin, R.N. Jonathan, A.P. Miles, W.R. Roger, S.S. Nick, A.S. Caroline, P.W. John, J. Richard, 2009. The British river of the future: How climate change and human activity might affect two contrasting river ecosystems in England. Science of The Total Environment, 407(17): 4787-4798.

[98] Johnson, R., 1998. The forest cycle and low river flows: a review of UK and international studies. Forest Ecology and Management, 109: 1-7.

[99] Karamouz, M. and S. Araghinejad, 2008. Drought mitigation through long-term operation of reservoirs: Case study. Journal of Irrigation and Drainage Engineering-Asce, 134(4): 471-478.

[100] Klemes, V., 1986. Operational testing of hydrological simulation models. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 31: 13-24.

[101] Krause, P., D.P. Boyle, and F. Baes, 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences, 5: 89-97.

[102] Laaha, G., and G. Blöschl, 2007. A national low flow estimation procedure for Austria. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 52(4): 625-644.

[103] Laio, F., and S. Tamea, 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrology and Earth System Sciences, 11(4): 1267-1277.

[104] Lajoie, F., A.A. Assani, A.G. Roy, and M. Mesfioui, 2007. Impacts of dams on monthly flow characteristics. The influence of watershed size and seasons. Journal of Hydrology, 334: 423-439.

[105] Lane, P. N. J., A.E. Best, K. Hickel, and L. Zhang, 2005. The response of flow duration curves to afforestation. Journal of Hydrology, 310(1-4): 253-265.

[106] Lang, C., 2007. Etiages et étirements : vers quelle modélisations ? : l'approche conceptuelle et l'analyse statistique en réponse à la diversité spatiale des écoulements en étiage des cours d'eau de l'Est français. Thèse de Doctorat, Université de Metz, CEGUM, 292 p.

[107] Lang, C., A. Freyermuth, E. Gille, and D. François, 2006. Le dispositif PRESAGES (PREvisions et Simulations pour l'Annonce et la Gestion des Etiages Sévères) : des outils pour évaluer et prévoir les étiages. Géocarrefour, 81(1), 15-24.

[108] Lang, C., E. Gille, D. Francois, G. Drogue, 2008. Improvement of a lumped rainfall-runoff structure and calibration procedure for predicting daily low flow discharges. Journal of Hydrology and Hydromechanics, 56(1): 59-71.

[109] Lang, C., E. Gille, and D. Francois, 2010. Estimation des paramètres d'un modèle mensuel pour l'évaluation des débits d'étiage en bassins non jaugés. IAHS Publ, 340: 620-628.

[110] Larras, J., 1972. Prévision et prédétermination des étiages et des crues, Bureau Central d'Etudes pour les Equipements d'Outre-Mer, Paris,p:9-17.

[111] Lefèvre, J., 1974. Le soutien des étiages en Loire à l'aide de réservoirs situés dans le haut bassin. Application au barrage de Naussac. La Houille Blanche, 4(5): 271-278.

[112] Lemarchand, C., R. Rosoux, and P. Berny, 2011. Ecotoxicology of the Eurasian Otter (Lutra lutra) along Loire river (France) and predictable trends due to global change. IUCN Otter Spec. Group Bull, 28(B)**:** 14 p.

[113] Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performance et du réalisme des modèles pluie-débit?. Thése de Doctorat, Cemagref (UR HBAN). Antony, Université Pierre et Marie Curie**:** 324p.

[114] Le Moine, N., V. Andréassian, C. Perrin, and C. Michel, 2007. How can rainfall-runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. Water Resources Research, 43: W06428.

[115] Letcher, R., F. Chiew, and A. Jakeman, Ed. 2004. An assessment of the value of seasonal forecasts in Australian farming systems. Biennial Meeting of the International Environmental Modelling and Software Society. Manno, Switzerland: 1511-1516.

[116] Linsley, R. K., 1982. Rainfall–runoff models-an overview. In: Proc. Int. Symp. on Rainfall-Runoff Modelling (ed. by V. P. Singh), 3–22. Water Resources Publications, Littleton, Colorado, USA.

[117] Longobardi, A. and P. Villani, 2008. Baseflow index regionalization analysis in a mediterranean area and data scarcity context: Role of the catchment permeability index. Journal of Hydrology, 355(1-4): 63-75.

[118] Lopez-Moreno, J. I., S.M. Vincent-Serrano, S. Begueria, J.M. Garcia-Ruiz, M.M. Portela, and A.B. Almeida, 2009. Dam effects on droughts magnitude and duration in a transboundary basin: The Lower River Tagus, Spain and Portugal. Water Resources Research, 45: W02405.

[119] Magilligan, F. J., and K.H. Nislow, 2005. Changes in hydrologic regime by dams. Geomorphology, 71: 61-78.

[120] Manoha, B., F. Hendrickx, A. Dupeyrat, C. Bertier, and S. Parey, 2008. Impact des évolutions climatiques sur les activités d'EDF (projet impec). La Houille Blanche, 02: 55-60.

[121] Mathevet, T., 2005. Quels modèles pluie-débit globaux au pas de temps horaire? Développements empiriques et comparison de modèles sur un large échantillon de bassins versants. Sciences de l'Eau, Thése de Doctorat, des Eaux et des Forêts, Paris**:** 463p.

[122] Mathevet, T., C. Michel, C. Perrin, and V. Andréassian, 2004. Experimental Design of a Lumped Hydrological Model Representing the Rainfall-Runoff Transformation at the Hourly Time-Step Hydrological Modeling. Conference Proceedings, Ohrid, FY Republic of Macedonia.

[123] Mathevet, T., C. Michel, V. Andréassian, and C. Perrin, 2006. A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins. Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment-MOPEX, IAHS Publication 307**:** 211-219.

[124] Mathevet, T., R. Garçon, C. Perret, M. Paquet, and J. Gailhard, 2007. Suivi et prévision de la ressource en eau à l'échelle d'un grand bassin versant: Exemple du bassin de la Loire. Grenoble, EDF-DTG, Report**:** 15p.

[125] Mathevet, T., C. Perret, R. Garçon, P. Periers, D. Goutx, J.M. Gibey, R. Oudin, H. Xhaard, J.L. Roy (2010). Modèles de prévision et prise de décision pour le soutien d'étiage de la Loire. La Houille Blanche, 5, 40-51.

[126] Maurel, F., T. Lepelletier, J. Moreau, C. Noeuveglise, and B. Lancelot, 2008. Bilan prospectif besoins-ressources en Seine-Normandie (partie hydrologique). La Houille Blanche, 5: 45-52.

[127] McCuen, R. H., Z. Knight, and A.G. Cutter, 2006. Evaluation of the Nash-Sutcliffe Efficiency Index. Journal of Hydrologic Engineering, 6(597-602).

[128] Merz, R., J. Parajka, and G. Blöschel, 2009. Scale effects in conceptual hydrologic modelling. Water Resources Research, 45: W09405. doi:09410.01029/02009WR007872.

[129] Michel, C., C. Perrin, and V. Andreassian, 2003. The exponential store: a correct formulation for rainfall-runoff modelling. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 48(1): 109-124.

[130] Michel, C., C. Perrin, and V. Andreassian, L. Oudin, and T. Mathevet, 2006. Has basin scale modelling advanced far beyond empiricism? IAHS Publ, 307: 108-116.

[131] Mignot, J. F., and J. Lefèvre, 1996. La problématique des étiages. La Houille Blanche, 6(7): 85-87.

[132] Mijuskovic-Svetinovic, T., and S. Maricic, 2008. Low flow analysis of the lower Drava River. Earth and Environmental Science, IOP Publishing. 4:doi:10.1088/1755-1307/4/1/012011.

[133] Mishra, A. K., and V.P. Singh, 2010. A review of drought concepts. Journal of Hydrology, 391: 202-216.

[134] Miquel, J., and P.A. Roche, 1983. La gestion d'un réservoir de soutien d'étiage peut-elle être optimale en cas de prévisions imparfaites? IAHS Publ, 147: 301-320.

[135] Moatar, F., A. Ducharne, D. Thiéry, V. Bustillo, É. Sauquet, and J.-P. Vidal, 2010. La Loire à l'épreuve du changement climatique. Géosciences. 12**:** 79-85.

[136] Moatar, F., M. Meybeck, and A. Poirel, 2009. Variabilité journalière de la qualité des rivières et son incidence sur la surveillance a long terme : exemple de la Loire moyenne. La Houille Blanche, 4: 91-99.

[137] Moore, R., V.A. Bell, and D.A. Jones, 2005. Forecasting for flood warning. Comptes Rendus Geoscience, 337: 203-217.

[138] Moore, R. J., 2007. The PDM rainfall-runoff model. Hydrological Earth System Sciences, 11(1): 483-499.

[139] Moreau, F., 2004. Gestion des étiages sévères: l'exemple de la Loire. La Houille Blanche, 4: 70-76.

[140] Moriasi, D. N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith, 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE, 50(3): 885-900.

[141] Moulin, L., and R. Thépot, 1999. La modernisation du système CRISTAL de gestion des crues et des étiages du bassin de la Loire. La Houille Blanche, 3(4): 77-81.

[142] Moulin, L., C. Perrin, C. Michel, and V. Andréassian, 2005. Prise en compte de barrages-réservoirs dans un modèle pluie-débit global: application au cas du bassin de la Seine amont. La Houille Blanche, 5: 79-87.

[143] Moussa, R., 2010. When monstrosity can be beautiful while normality can be ugly: assessing the performance of event-based flood models. Journal of Hydrological Sciences, 55(6): 1074-1084.

[144] Nash, J. E., and J.V. Sutcliffe, 1970. River flow forecasting through conceptual models Part I-A discussion of principles. Journal of Hydrology, 10: 282-290.

[145] Nathan, R. J., and T.A., McMohan, 1990. Evaluation of automated techniques for base flow and recession analyses. Water Resources Research, 26: 1465-1473.

[146] Ngigi, S. N., H.H.G. Savenije, and F.N. Gichuki, 2008. Hydrological Impacts of Flood Storage and Management on Irrigation Water Abstraction in Upper Ewaso Ng'iro River Basin, Kenya. Water Resources Management, 22: 1859-1879.

[147] Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andréassian, F. Anctil, and C. Loumagne, 2005a. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. Journal of Hydrology, 303(1-4): 290-306.

[148] Oudin, L., C. Michel, and F. Anctil, 2005b. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 1 - Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? Journal of Hydrology, 303(1-4): 275-289.

[149] Oudin, L., V. Andréassian, J. Lerat, and C. Michel, 2008. Has land cover a significant impact on mean annual streamflow? An international assessment using 1508 catchments. Journal of Hydrology, 357: 303-316.

[150] Oudin, L., V. Andréassian, T. Mathevet and C. Perrin, 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resources Research, 42: W07410.

[151] Pagano, T., D. L. Shrestha, Q.J. Wang, D. Robertson, and P. Hapuarachchi, 2012. Ensemble dressing for hydrological applications. Hydrological Processes, DOI: 10.1002/hyp.9313.

[152] Payan, J. L., C.Perrin, V. Andreassian, and C. Michel, 2008. How can man-made water reservoirs be accounted for in a lumped rainfall-runoff model? Water Resources Research, 44(3): w03420.

[153] Perrin, C., 2000. Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative. Thèse de Doctorat, INPG: Grenoble, Cemagref, Antony**:** 530p.

[154] Perrin, C., C. Michel and V. Andreassian, 2001. Long term forecasting for French rivers by continuous rainfall-runoff modeling. Meeting of the British Hydrological Society on Continuous river flow simulation., BHS Occasional Paper No.13, Wallingford, UK. pp.21-29.

[155] Perrin, C., C. Michel, and V. Andréassian, 2003. Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology, 279(1-4): 275-289.

[156] Perrin, C., V. Andréassian, and C. Michel, 2006. Simple benchmark models as a basis for model efficiency criteria. Archiv für Hydrobiologie. Large Rivers, 17(1-2): 221-244.

[157] Peters, D. L., and T.D. Prowse, 2001. Regulation effects on the lower Peace River, Canada. Hydrological Processes, 15: 3181-3194.

[158] Pociask-Karteczka, J., J. Czulak, J. Niedbala, and J. Niedbala, 2003. Model Assessing Changes of the Raba River Runoff Caused by the Dobczyce Reservoir (Poland). Polish Journal of Environmental Studies, 12(4): 485-488.

[159] Pointet, T., 2004. Les relations nappes-rivières, observations à partir des étiages de petites rivières. La Houille Blanche, 4: 87-91.

[160] Prudhomme, C., and E. Saquet, 2006. Modelling a regional drought index in France, Centre for Ecology & Hydrology, Natural Environmental Research Council, CEH Wallingford & Cemagref, Lyon**:** 52p.

[161] Pushpalatha, R., C. Perrin, N. Le Moine, and V. Andréassian, 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. Journal of Hydrology, 420-421: 171-182.

[162] Quintana-Segui, P., P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baiillon, C. Canellas, L. Franchisteguy, and S. Morel, 2008. Analysis of near-surface atmospheric variables:

validation of the SAFRAN analysis over France. Journal of Applied Meteorology and Climatology, 47(1): 92-107.

[163] Ramos, M. H., T. Mathevet, J. Thielen, and F. Pappenberger, 2010. Communicating uncertainty in hydro-meteorological forecasts: mission impossiible? Meteorological Applications, 17: 223-235.

[164] Randrianasolo, A., M.H. Ramos, G. Thirel, V. Andréassian, and E. Martin, 2010. Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. Atmospheric Science Letters, 11: 100-107.

[165] Refsgaard, J. C., 1997. Validation and Intercomparison of Different Updating Procedures for Real-Time Forecasting. Nordic Hydrology, 28: 65-84.

[166] Reilly, C. F., and C.N. Kroll, 2003. Estimation of 7-day, 10-year low-streamflow statistics using baseflow correlation. Water Resources Research, 39(9): 1236.

[167] Renard, B., M. Lang, P. Bois, A. Dupeyrat, O. Mestre, H. Niel, J. Gailhard, C. Laurent, L. Neppel, and E. Sauquet, 2006. Evolution des extrêmes hydrométriques en France à partir de données observées. La Houille Blanche, 6: 48-54.

[168] Reusser, D. E., T. Blume, B. Schaefli, and E. Zehe, 2009. Analysing the temporal dynamics of model performance for hydrological models. Hydrological Earth System Sciences, 13(7): 999-1018.

[169] Rifai, H. S., S.M. Brock, K.B. Ensor, P.B. Bedient, 2000. Determination of low-flow characteristics for Texas streams. Journal of Water Resources Planning and Management-Asce, 126(5): 310-319.

[170] Robinson, M., and C. Cosandey, 2002. Impact de la forêt sur les débits d'étiage. La Houille Blanche, 3: 59-63.

[171] Rutten, M., N. van de Giesen, J. Icke, and W. Uijttewaal, 2008. Seasonal forecasting of cooling water problems in the River Rhine. Hydrological Processes, 22: 1037-1045.

[172] Santhi, C., P. M. Allen, et al., 2008. Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. Journal of Hydrology, 351(1-2): 139-153.

[173] Sauquet, E., J. Lerat and C. Prudhomme, 2008. Seasonal forecasting of river flows-State -of-the-art and applications, La Houille Blanch/N-6**:** 77-84.

[174] Schaake, J., K. Franz, A. Bradley, and R. Buizza, 2006. The Hydrologic Ensemble Prediction Experiment (HEPEX). Hydrological Earth System Sciences Discussion, 3: 3321-3332.

[175] Schaake, J. C., T.M. Hamill, R. Buizza, and M. Clark, 2007. The Hydrological Ensemble Prediction Experiment. American Meteorological Society: 1541-1547.

[176] Schaefli, B., and H.V. Gupta, 2007. Do Nash values have value? Hydrological Processes, 21(15): 2075-2080.

[177] Schneider, W. J., 1957. Relation of geology to streamflow in the upper little Miami basin. The Ohio Journal of Science, 57(1): 11-14.

[178] Schwartz, S. S., 2007. Automated algorithms for heuristic base-flow separation. Journal of the American Water Resources Association, 43(6): 1583-1594.

[179] Seibert, J., 2001. On the need for benchmarks in hydrological modelling. Hydrological Processes, 15(6): 1063-1064.

[180] Seo, D.-J., H.D. Herr, and J.C. Schaake, 2006. A statistical post-processor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction. Hydrological Earth System Sciences Discussion, 3: 1987-2035.

[181] Sharda, V. N., P. Samraj, J.S. Samra, and V. Lakshmanan, 1998. Hydrological behaviour of first generation coppiced bluegum plantations in the Nilgiri sub-watersheds. Journal of Hydrology, 211: 50-60.

[182] Sikka, A. K., J.S. Samra, V.N. Sharda, P. Samraj, and V. Lakshmanan, 2003. Low flow and high flow responses to converting natural grassland into bluegum (*Eucalyptus globulus*) in Nilgiris watersheds of South India. Journal of Hydrology, 270: 12-26.

[183] Singla, P., J.-P. Céron, E. Martin, F. Regimbeau, M. Déqué, F. Habets, and J.-P. Vidal, 2011. Predictability of soil moisture and river flows over France for the spring season. Hydrological Earth System Sciences Discussion, 8: 7947-7966.

[184] Smakhtin, V. U., 2001. Low flow hydrology: a review. Journal of Hydrology, 240: 147-186.

[185] Smakhtin, V. Y. and M. Toulouse, 1998. Relationships between low-flow characteristics of South African streams. Water Sa, 24(2): 107-112.

[186] Smith, J. A., G.N. Day, and M.D. Kane, 1992. Nonparametric framework for long-range streamflow forecasting. Journal of Water Resources Planning and Management, 118(1): 334-345.

[187] Soubeyroux, J.-M., J.-P. Vidal, M. Baillon, M. Blanchard, J.-P. Céron, L. Franchistéguy, F. Régimbeau, E. Martin, and J.-C. Vincendon, 2010. Caractérisation et prévision des sécheresses et étiages en France à partir de la chaîne hydrométéorologique Safran-Isba-Modcou. La Houille Blanche, 5: 30-39.

[188] Staub, P., 2008. Prévision d'étiage par modèlisation hydrologique: Mise au point d'une méthode d'évaluation. Master 2 Professionnel, Univ. François Rabelais, Tours, CEMAGREF, Antony**:** 92p.

[189] Stravs, L., and M. Brilly, 2007. Development of a low-flow forecasting model using the M5 machine learning method. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 52(3): 466-477.

[190] Stromberg, J. C., V. B. Beauchamp, et al., 2007. Importance of low-flow and high-flow characteristics to restoration of riparian vegetation along rivers in and south-western United States. Freshwater Biology, 52(4): 651-679.

[191] Szilagyi, J. and M. B. Parlange, 1998. Baseflow separation based on analytical solutions of the Boussinesq equation. Journal of Hydrology, 204(1-4): 251-260.

[192] Szollosi-Nagy, A., 1976. Introductory remarks on the state space modelling of water resource systems. Int. Inst. for Applied Systems Analysis, RM-76-73: 81pp.

[193] Tague, C., and G.E. Grant, 2004. A geological framework for interpreting the low-flow regimes of Cascade streams, Willamette River Basin, Oregon. Water Resources Research, 40: W04303, doi:04310.01029/02003WR002629.

[194] Tallaksen, L. M., and H.A.J. van Lanen, 2003. Hydrological Drought: Processes and estimation methods for stream flow and groundwater, Developments in Water Science:48 FRIEND,1-417.

[195] Tangara, M., 2005. Nouvelle méthode de prévision de crue utilisant un modèle pluie-débit global. Thèse de doctorat, École pratique des hautes études de Paris, Cemagref, Antony**:** 374p.

[196] Thielen, J., J. Schaake, R. Hartman, and R. Buizza, 2008. Aims, challenges and progress of the Hydrological Ensemble Prediction Experiment (HEPEX) following the third HEPEX workshop held in Stresa 27 to 29 June 2007. Atmospheric Science Letters, 9: 29-35.

[197] Thiéry, D., 1978. Application à quatre bassins hydrologiques des méthodes de prévision des étiages par convolution. Ministère de l'environnement, Rapport 78SGN621HYD, BRGM, Orléans, 207 p.

[198] Thiéry, D., 2009. Presentation of the Gardénia software package developed by BRGM Note Technique, NT EAU 2009/02**:** 10 p.

[199] Thirel, G., E. Martin, J.-F. Mahfouf, S. Massart, S. Ricci, and F. Habets, 2010a. A past discharge assimilation system for ensemble streamflow forecasts over France - Part 1: Description and validation of the assimilation system. Hydrological Earth System Sciences, 14: 1623-1637.

[200] Thirel, G., E. Martin, J.-F. Mahfouf, S. Massart, S. Ricci, F. Regimbeau, and F. Habets, 2010b. A past discharge assimilation system for ensemble streamflow forecasts over France - Part 2: Impact on the ensemble streamflow forecasts. Hydrological Earth System Sciences, 14: 1639-1653.

[201] Thirel, G., F. Rousset-Regimbeau, E. Martin, and F. Habets, 2008. On the impact of short-range meteorological forecasts for ensemble streamflow predictions. American Meteorological Society, 9: 1301-1317.

[202] Tucci, C. E. M., R.T. Clarke, W. Collischonn, P.L.D. Dias, and G.S. de Oliveira, 2003. Long-term flow forecasts based on climate and hydrologic modeling: Uruguay River basin. Water Resources Research, 39(7): 1181.

[203] Valença, M., and T. Ludermir, 2004. Hydrological Forecasting and Updating Procedures for Neural Network, LNCS 3316. Verlag Berlin Heidelberg: 1304-1309.

[204] Vernoux, J.-F., and G. Villion, 2008. Outils d'aide à la gestion en étiage du lac-réservoir Marne prenant en compte les eaux souterraines. La Houille Blanche, 5: 30-37.

[205] Vidal, J.-P., E. Martin, L. Franchisteguy, M. Baillon, and J.-M. Soubeyroux, 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. International Journal of Climatology, 30: 1627-1644.

[206] Villion, G., 1997. Rôle des lacs-réservoirs amont : les grands lacs de Seine. La Houille Blanche, 8: 51-56.

[207] Vogel, R. M. and N. M. Fennessey, 1994. Flow-Duration Curves .1. New Interpretation and Confidence-Intervals. Journal of Water Resources Planning and Management-Asce, 120(4): 485-504.

[208] Wagener, T., H.S. Wheater, and H.V. Gupta, 2004. Rainfall-Runoff Modelling in Gauged and Ungauged Catchments, Imperial College Press, London.

[209] Wang, S. F., S.Z. Kang, L. Zhang, and F.S. Li, 2008. Modelling hydrological response to different land-use and climate change scenarios in the Zamu River basin of northwest China. Hydrological Processes, 22(14): 2502-2510.

[210] Weijs, S. V., G. Schoups, and N. van de Giesen, 2010. Why hydrological predictions should be evaluated using information theory. Hydrology and Earth System Sciences, 14(12): 2545-2558.

[211] Werner, K., D. Brandon, M. Clark, S. Gangopadhyay, 2005. Incorporating Medium-Range Numerical Weather Model Output into the Ensemble Streamflow Prediction System of the National Weather Service. Journal of Hydrometeorology, 6: 101-114.

[212] Whitfield, P. H., 2008. Improving the prediction of low flows in ungauged basins in Canada in the future. Canadian Water Resources Journal, 33(2): 207-214.

[213] Williams, G. P., and M.G. Wolman, 1984. Downstream effects of dams in alluvial rivers. US Geological Survey, Professional Paper: 1286.

[214] Willmot, C. J., 1984. On the evaluation of model performance in physical geography. In: Gaile, G.L., and C.J. Willmot (Eds.), Spatial Statistics and Models. D. Reidel, Dordrecht. pp. 443-460.

[215] Wipfler, L., H. van Lanen, F. Ludwig, L.M. Tallaksen, A.K. Fleig, S. Niemeyer, E. Sauquet, and M.-H. Ramos, 2009. 7th Framework Programme, Theme 6, Environment (Including Climate Change). FP7-ENV-2007-1 (ENV.2007.1.3.3.3.), Grant agreement no.: 211837**:** 46 p.

[216] WMO, 2008. Manual on Low-flow Estimation and Prediction, Operational Hydrology Report No. 50. A. Gustard, and S. Demuth. Geneva

[217] Woo, M., T. Long, and R. Thorne, 2009. Simulating monthly stream flow for the Upper Changjiang, China, under climatic change scenarios. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 54(3): 596-605.

[218] Wu, Y., and J. Chen, 2012. An Operation-Based Scheme for a Multiyear and Multipurpose Reservoir to Enhance Macroscale Hydrological Models. Journal of Hydrometeorology, 13: 270-283.

[219] Wu, Y., J. Chen, and B. Sivakumar, 2007. Numerical modelling of operation and hydrologic effects of Xinfengjiang reservoir in Southern China. MODSIM 2007 International Congress on Modelling and Simulation, Rocquencourt, France, Modelling and Simulation Society of Australia and New Zealand, 1561-1567.

[220] Xiong, L., and K.M. O'connor, 2002. Comparison of four updating models for real-time river flow forecasting. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 47(4): 621-639.

[221] Ye, B. S., D.Q. Yang, and D.L. Kane, 2003. Changes in Lena River streamflow hydrology: Human impacts versus natural variations. Water Resources Research, 39(7): 1200.

[222] Żelazinski, J., and M. Mierkiewics, 2009. Ensemble predictions in hydrology: necessity, preparation and using in water management decision processes. La Houille Blanche, 5: 35-38.

[223] Zhang, C., Y. Peng, J. Chu, and C.A. Shoemaker, 2012. Integrated hydrological modelling of small-and medium-sized water storages with application to the upper Fengman Reservoir Basin of China. Hydrological Earth System Sciences Discussion, 9: 4001-4043.

Appendices

A Selection of objective function for calibration

A.1 Test of a set of criteria

Here we analyzed the sensitivity of model performances in validation after using various objective functions for model calibration. The bounded formulation of Nash-Sutcliffe efficiency on natural flows, square root transformed flows, log-transformed flows and the inverse transformed flows were analysed to fix a general objective function (a criterion which would be general enough to give equal emphasis on high- as well as low-flow simulations) for calibration in our study.

The split sample testing scheme proposed by Klemes (1986) was used to evaluate model performance on a set of catchments which was already presented in Chapter 2. For each catchment, the period where rainfall, PE and flow data were available (at most 1970–2006) was split into two halves (P1 and P2) of similar length, alternatively used for model calibration and validation. It means that for each catchment and each tested model, two calibrations and two validations were systematically performed. The first year of each test period was used for model warm–up. To avoid initialization problems for catchments with long–term memories, five years of warm–up were considered in addition to the 1–year warm–up period: they were either the five years of observed data preceding the test period when available, or a mean year repeated five times otherwise. In this study, only performance in validation was considered to evaluate models.

Table A.1 shows the average performance of individual criterion with respect to the calibration test on the full catchment set. As expected, the analysis of results shows that the behaviour of criteria in validation depends on the objective function. Not surprisingly, the highest performance for each criterion in validation is obtained when the same criterion is used as objective function in calibration. But interestingly, the use of NSE^{*}_{sqrtQ} as objective function provided moderate performance for all the selected criteria during validation. This shows that the use of this criterion can give equal preference to the high- as well as low-flow conditions (For more on criteria analysis, refer to Pushpalatha et al. (2012)). This is useful when both the low- and high-flow simulations are of interest. This conclusion is in agreement with the previous study by Oudin et al. (2006). Hence in our all tests, we kept this criterion as the objective function for calibration.

Table A.1 Validation results with respect to the objective functions

	Criteria in validation			
Criteria in calibration	NSE^*_Q	NSE^*_{sqrtQ}	NSE^*_{lnQ}	NSE^*_{iQ}
NSE^*_Q	0.632	0.646	0.537	0.123
NSE^*_{sqrtQ}	0.621	0.675	0.617	0.230
NSE^*_{lnQ}	0.558	0.649	0.644	0.343
NSE^*_{iQ}	0.329	0.444	0.518	0.397

B A review of efficiency criteria suitable for evaluating low-flow simulations

Raji Pushpalatha, Charles Perrin, Nicolas Le Moine, Vazken Andréassian

(*Article published in Journal of Hydrology*)

Journal of Hydrology 420–421 (2012) 171–182

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A review of efficiency criteria suitable for evaluating low-flow simulations

Raji Pushpalatha [a], Charles Perrin [a,*], Nicolas Le Moine [b], Vazken Andréassian [a]

[a] Cemagref, UR HBAN, 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France
[b] Université Pierre et Marie Curie, UMR 7619 Sisyphe, 4 Place Jussieu, 75252 Paris Cedex 05, France

ARTICLE INFO

Article history:
Received 5 May 2011
Received in revised form 27 September 2011
Accepted 29 November 2011
Available online 8 December 2011
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Keywords:
Low flows
Simulation efficiency
Model evaluation
Benchmark model
Criteria

SUMMARY

Low flows are seasonal phenomena and an integral component of the flow regime of any river. Because of increased competition between water uses, the demand for forecasts of low-flow periods is rising. But how low-flow predictions should be evaluated? This article focuses on the criteria able to evaluate the efficiency of hydrological models in simulating low flows. Indeed, a variety of criteria have been proposed, but their suitability for the evaluation of low-flow simulations has not been systematically assessed.

Here a range of efficiency criteria advised for low flows is analysed. The analysis mainly concentrates on criteria computed on continuous simulations that include all model errors. The criteria were evaluated using two rainfall–runoff models and a set of 940 catchments located throughout France. In order to evaluate the capacity of each criterion to discriminate low-flow errors specifically, we looked for the part of the hydrograph that carries most of the weight in the criterion computation.

Contrary to what was expected, our analysis revealed that, in most of the existing criteria advised for low flows, high flows still make a significant contribution to the criterion's value. We therefore recommend using the Nash–Sutcliffe efficiency criterion calculated on inverse flow values, a valuable alternative to the classically used criteria, in that on average it allows focusing on the lowest 20% of flows over the study period.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

1.1. In the jungle of efficiency criteria

Hydrological modelling aims at understanding and interpreting catchment hydrological behaviour. It is also used to address a number of practical issues, ranging from flood estimation to water resources management and low-flow forecasting. Whatever model is applied, the model user needs appropriate and meaningful indicators informing on the actual capacity of the model.

However, the evaluation of goodness-of-fit is not as straightforward as it may seem at first glance. Of course, model performance can first be evaluated by the visual comparison of the observed and simulated flow hydrographs, but this remains extremely dependent on the evaluator's experience (Chiew and McMahon, 1993; Houghton-Carr, 1999). A more objective way to evaluate model performance is to use numerical criteria, but then the user may get lost in a jungle of potential criteria. Why is the choice so difficult? Several reasons can be put forward:

1. flows vary by several orders of magnitude that may not be equally useful for the modeller;
2. hydrological models often produce heteroscedastic errors, i.e. their variance is not independent of the flow value;
3. the range of target flows may vary significantly between evaluation periods;
4. the model may be used for different applications, which may require specific criteria.

For these reasons, a large variety of criteria have been proposed and used over the years in hydrological modelling, as shown for example by the lists of criteria given by the ASCE (1993), Dawson et al. (2007), Moriasi et al. (2007) and Reusser et al. (2009). Among these criteria, some are *absolute* criteria such as the widely used root mean square error, while others are *relative* criteria (i.e. normalized) such as the Nash and Sutcliffe (1970) efficiency criterion (*NSE*). In the latter, model errors are compared to the errors of a reference or benchmark model (Seibert, 2001; Perrin et al., 2006). This provides a useful quantification of model performance in that it indicates to which extent the model is better (or worse) than the benchmark. It also facilitates the comparison of performance between catchments.

However, the choice of a benchmark is difficult: different benchmarks are more or less demanding (and thus the comparison more or less informative) depending on the type of hydrological regime or the type of model application. This sometimes makes it difficult to interpret the relative criteria and a bit puzzling for an inexperi-

* Corresponding author. Tel.: +33 140966086; fax: +33 140966199.
E-mail address: charles.perrin@irstea.fr (C. Perrin).

0022-1694/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2011.11.055

enced end-user, who may simply want to know whether the model can be considered as "good", "acceptable" or "bad". Actually, there is no single criterion that can evaluate model performance in all cases (Jain and Sudheer, 2008). Many authors use several criteria simultaneously (see e.g. Efstratiadis and Koutsoyiannis (2010) for a review in the context of multi-objective calibration), but a discussion of these approaches is not within the scope of this article.

The merits and drawbacks of several efficiency criteria have already been discussed and debated in the literature, as well as the links between them. The Nash–Sutcliffe efficiency criterion has probably received the most attention (Garrick et al., 1978; Houghton-Carr, 1999; McCuen et al., 2006; Schaefli and Gupta, 2007; Clarke, 2008; Gupta et al., 2009; Moussa, 2010; Gupta and Kling, 2011). Like many other criteria based on the mean model square error, this criterion is known to put greater emphasis on high flows when calculated on a continuous simulation. Although it was shown to have several limitations (such as its sensitivity to the hydrological regime, sample size or outliers), it remains a valuable and popular means to evaluate models for high-flow simulation.

Comparatively little work has been carried out on the meaning and interpretation of the criteria used to evaluate models in low-flow conditions. The following section summarizes the existing studies.

1.2. Criteria used for the evaluation of low-flow simulation

Table 1 lists some of the studies discussing performance criteria able to judge low-flow simulations. Note that the various criteria formulations listed here depend on at least three factors:

1.2.1. Calculation period

Instead of calculating criteria only over the low-flow periods (which generally requires the subjective choice of a low-flow threshold), most of the existing criteria calculate model errors over the entire test period. Thus, they give some weight to the errors in low-flow as well as to the errors in high-flow conditions.

1.2.2. Target variable

The second major aspect in calculating criteria is the choice of a target variable. Some authors (e.g. Houghton-Carr, 1999) appeal to statistical measures classically used to characterize low flows, such as the base-flow index, a percentile of the flow duration curve (*FDC*) or some minimum accumulated flows over a continuous period (e.g. 7 days). By calculating the ratio between the simulated and observed values, a relative efficiency criterion is obtained. These criteria are very useful when studies focus on specific aspects of low flows. However, one may also continue calculating the sum of errors over the entire test period, provided that the appropriate transformation on flows is used (Box and Cox, 1964; Chiew et al., 1993). This transformation helps put more weight on low flows. The root square or the logarithms are among the most widely used transformations on low-flow values. For example, Smakhtin et al. (1998), Houghton-Carr (1999), Oudin et al. (2006), Jain and Sudheer (2008), and de Vos et al. (2010) used the sum of squared residuals calculated on the logarithms of flow values in order to reduce the biasing towards peak flows. Chiew et al. (1993) used the root squared transform to evaluate the model's performance in low-flow conditions. Krause et al. (2005) proposed using a relative variable as the ratio between simulation and observation, and calculated the distance of this variable from 1. Le Moine (2008) discussed a generalization of these transformations as a power law transformation with positive or negative exponents and proposed a family of squared criteria based on the power transformation of flows (of Box-Cox type), defined by:

$$RMSE(\lambda) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_i^{\lambda} - \widehat{Q}_i^{\lambda})^2} \quad (1)$$

where λ is the power of the flow transformation, n is the number of time steps and Q_i and $\widehat{Q}_i$ are the observed and simulated flows, respectively, at time step i. λ is not necessarily an integer and can take positive and negative values. When λ tends towards zero, the transformation tends towards the logarithm transformation. As noted by Le Moine (2008), low and high values of λ will tend to emphasize the model errors on the minimum and maximum flow values, respectively. For example, Chiew et al. (1993) had used a value of λ equal to 0.2 to give more emphasis on low flows.

1.2.3. Error normalization

Another aspect that differs between criteria is the type of error used and the way model error is normalized. Most of the criteria are based on the squared residuals, but absolute errors can also be considered. A power of these absolute errors may also be used (see e.g. Krause et al., 2005). In terms of model error normalization, most of the efficiency indexes use the form of the *NSE*. Willmot (1984) proposed the index of agreement as another way to normalize model square error, by dividing it by the potential error.

1.3. Are existing criteria appropriate to evaluate low-flow simulations?

Only a few authors have discussed the suitability of the variety of existing criteria to evaluate low-flow simulations. Oudin et al. (2006) compared several objective functions and concluded that the square root transformation provides an all-purpose efficiency measure, not specifically focusing on low flows. Analysing several efficiency indices, Krause et al. (2005) showed that some criteria are closely related while others show very different patterns. They advised using the relative efficiency index for the evaluation of low-flow simulations, noting that the logarithm transformation on flows provides a higher sensitivity to low flows, although they indicate that this criterion remains sensitive to high flows.

Actually, the impact of flow transformations significantly changes the way the hydrograph and model errors are considered in criteria, as shown by Le Moine (2008). This is illustrated in Fig. 1 in the case of natural, root square and logarithmic transformed flows. Fig. 1 shows the series of flow values and model errors as well as the cumulated quadratic error (to facilitate the comparison

Table 1
Studies that used criteria to evaluate low-flow simulation quality.

References	Low flow statistics	Residual errors	Standard deviation	Bias	Coeff. of determination	Nash efficiency (all transforms)	Coeff. of variation	Log MSE	Relative efficiency	Index of agreement
Chiew et al. (1993)						x				
Ye et al. (1998)	x	x				x				
Houghton–Carr (1999)	x					x				
Krause et al. (2005)					x	x			x	x
Oudin et al. (2006)		x				x				
Jain and Sudheer (2008)	x		x	x		x	x			
de Vos et al. (2010)								x		

between graphs, all series were normalized by the maximum value in the series, thus providing values ranging between 0 and 1). Fig. 1 illustrates the significant differences between transformations and the resulting error values: indeed, the part of the hydrograph that produces most of the errors strongly depends on the transformation chosen. Hence, a detailed analysis of power law transformations is needed and will be discussed in Section 3.6.

1.4. Scope of the article

The present work intends to complement previous studies by objectively identifying which criteria are best suited for the evaluation of low-flow simulations using actually observed data. This paper has two main objectives:

- First, we wish to discuss the links between the criteria used or proposed to evaluate low-flow simulations.
- Second, we wish to better understand which part of the hydrograph most influences the various criteria used for the evaluation of low-flow simulations.

To meet these objectives, we will evaluate various criteria, mainly pertaining to the family of criteria proposed by Le Moine (2008) in the form of the Nash–Sutcliffe efficiency index. They will be compared to other criteria proposed in the literature for low-flow evaluation on a large set of catchments showing various low-flow characteristics. We will restrict this analysis to (i) the case of nondimensional criteria (i.e. relative criteria in which the model error is normalized), and (ii) deterministic low-flow simulations. Although several authors recently questioned the relevance of deterministic predictions in an information theory perspective (Weijs et al., 2010), the criteria for the evaluation of probabilistic low-flow predictions are not within the scope of this study (see e.g. Laio and Tamea (2007), Bartholmes et al. (2009), Boucher et al. (2009), Ramos et al. (2010) or Randrianasolo et al. (2010) for discussions on this issue). Last, we will not discuss the issue of model calibration (i.e. the choice of objective function), but concentrate on the evaluation of simulations only.

The next section presents the data set and methodology used in this study. Then the results are detailed and discussed before the concluding remarks.

2. Data set, models and evaluation methodology

2.1. Data set

To carry out the evaluation, a set of 940 catchments throughout France was used to test the models in various hydro-meteorological conditions, including oceanic, Mediterranean and continental

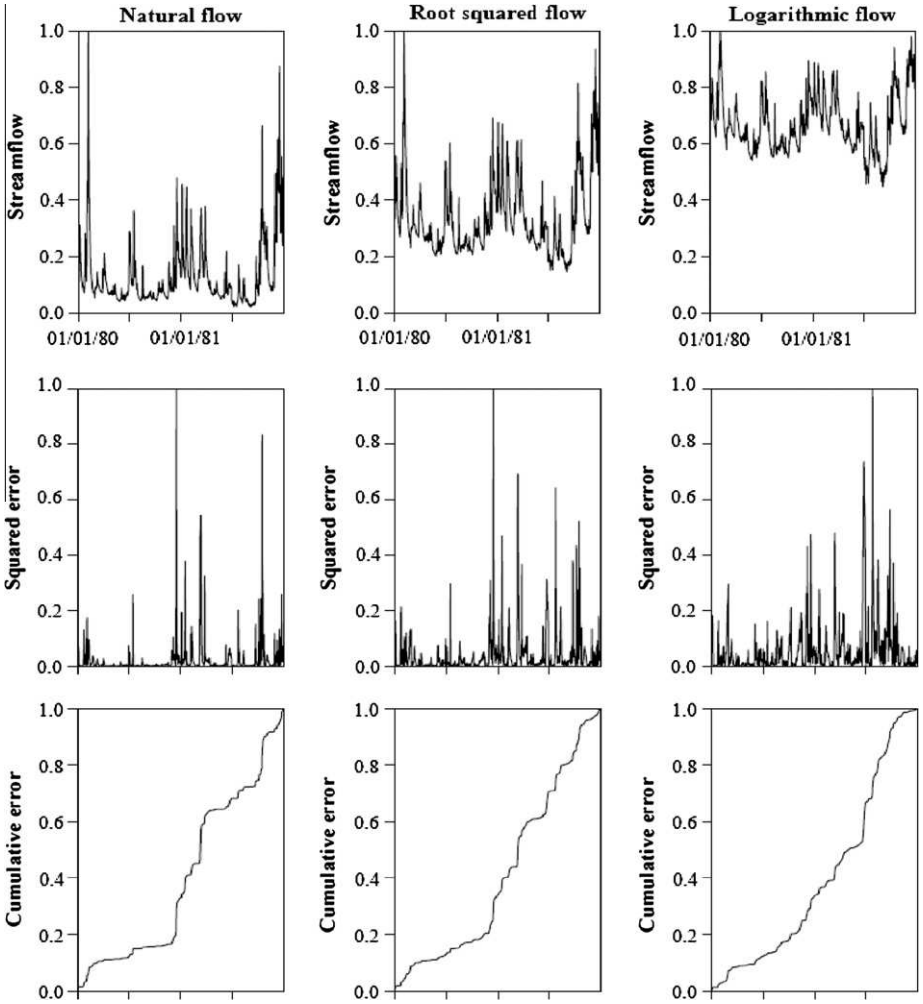


Fig. 1. Illustration of changes in the series of flows, model errors and cumulated squared errors using root square and logarithm transformations for the Moselle River at Custines (Hydro code A7010610; catchment area 6830 km^2; period: 01/01/1980–31/12/1981). All series are normalized between 0 and 1 by dividing by the maximum value over the period.

influences. This should provide more robust conclusions than testing the models on a single catchment (Andréassian et al., 2006a). Detailed information on this data set can be found in Le Moine et al. (2007) and Le Moine (2008). A few characteristics are given in Table 2. The catchments studied are all unregulated, although some of them may be influenced in the summer periods by human withdrawals.

Daily time series were used, as this time step is well adapted to the study of low flows. Continuous series of precipitation, potential evapotranspiration and streamflow were available for the 1970–2006 time period, providing good variability of meteorological conditions, with quite severe drought periods (e.g. the years 1976, 1989–1991, 2003 and 2005). Meteorological data come from the SAFRAN reanalysis of Météo–France (Quintana-Segui et al., 2008; Vidal et al., 2010). Daily potential evapotranspiration was estimated using the formulation proposed by Oudin et al. (2005) based on temperature and extra-terrestrial radiation.

To characterize the variability in low-flow characteristics in the catchment set, we calculated the ratio between Q_{90} (the flow exceeded 90% of the time) and Q_{50} (the median flow) for each catchment, to obtain a nondimensional indicator of the severity of low flows (Smakhtin, 2001). The less variable the flow regime, the higher the ratio value. As shown in Table 2 and Fig. 2, the catchments in this data set experience a large variety of low-flow conditions. Some catchments seem to have a very smoothed response with substantial low flows, while others seem much more responsive with high flow variability. A few catchments have an index value of zero, corresponding to a Q_{90} value of zero, which means that they are nonperennial catchments.

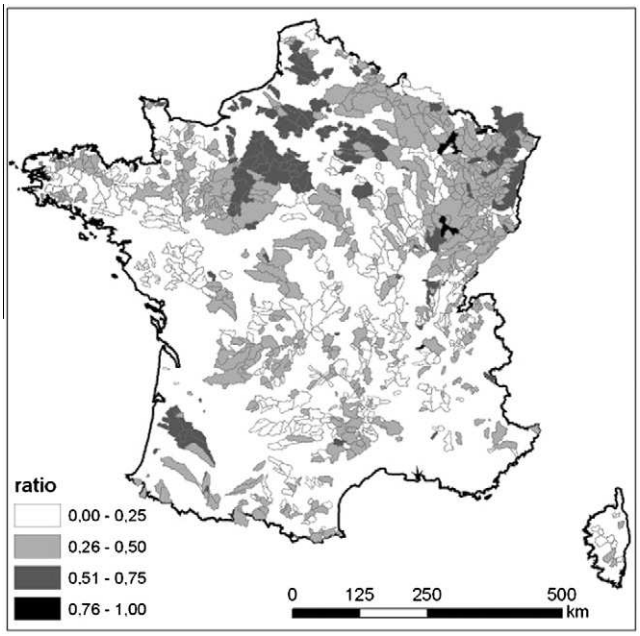


Fig. 2. Spatial distribution of the Q_{90}/Q_{50} index values over the set of 940 French catchments.

2.2. Hydrological models

The tests were carried out using two continuous lumped rainfall–runoff models: the four-parameter GR4J model (Perrin et al., 2003) and a six-parameter version of the MORDOR model (Garçon, 1999; Mathevet, 2005) called MORD hereafter to avoid confusion with the original model. These models are quite widely used in France for operational purposes. Schematic diagrams of model structures are shown in Fig. 3. For detailed descriptions of the models, please refer to Andréassian et al. (2006b). Here, models were used in the same conditions, i.e. they were fed exactly with the same rainfall and potential evapotranspiration inputs.

The lumped modelling approach adopted here is often thought unsuitable for quite large catchments like some of those used here (catchment area goes up to almost 10,000 km^2 in our catchment set). However, contrary to this preconceived idea, we noticed that the median Nash–Sutcliffe efficiency tends to increase with catchment size for the tested models on our catchment set (not shown here). This can be related to the results presented by Merz et al. (2009) who applied a hydrological model (with semi-distributed inputs but lumped parameters) to 269 Austrian catchments ranging in size from 10 to 130,000 km^2. They concluded that model efficiency tends to increase over the scale range of 10–10,000 km^2.

Note that the two models simulate the low flows in a quite different manner: in the GR4J model, low flows are generated by the percolation from the soil moisture store (S), outflow from the routing store (R) and the groundwater exchange function (F) (which simulates possible interactions with deep aquifers or surrounding catchments). However, in MORD, low flows are produced by two routing stores (L and N) (Fig. 3).

The models' parameters were calibrated using a two-step search procedure (a prior gross inspection of the parameter space to identify the most likely zone of convergence, followed by a local search algorithm). This approach proved efficient for these parsimonious models compared to more complex search algorithms (see Edijatno et al., 1999; Mathevet, 2005).

As an objective function for calibration, we used the root mean squared error calculated on root square flows, which was found by Oudin et al. (2006) to be a good compromise for an all-purpose model (not giving too much emphasis to low or high flows). Discussing the impact of the choice of the objective function on the quality of low-flow simulations is not within the scope of this article, and we will only examine the relevance of different criteria for model evaluation in validation mode.

2.3. Testing scheme

This section describes the method used to evaluate criteria. The models were run following the split sample testing scheme proposed by Klemes (1986), i.e. splitting the whole available record into two independent periods (P1 and P2) of equal length, calibrating the model on P1 and validating it on P2 and then exchanging the roles of P1 and P2. Hence the models could be evaluated in validation over the whole data record. The first year of each test period was used for model warm-up. In the following, the results shown correspond only to validation tests. To summarize the

Table 2
Characteristics of the 940 test catchments.

	Exceedance percentiles						
Catchment characteristics	1.00	0.95	0.75	0.50	0.25	0.05	0.00
Catchment area (km^2)	5.3	2075	391	162	78	28	9423
Mean annual precipitation (mm/y)	621	1548	1152	964	828	714	2143
Mean annual potential evapotranspiration (mm/y)	337	750	685	645	615	533	814
Mean annual runoff (mm/y)	50	1201	561	366	244	141	3689
Q_{90}/Q_{50} (–)	0.00	0.64	0.39	0.28	0.18	0.06	0.94

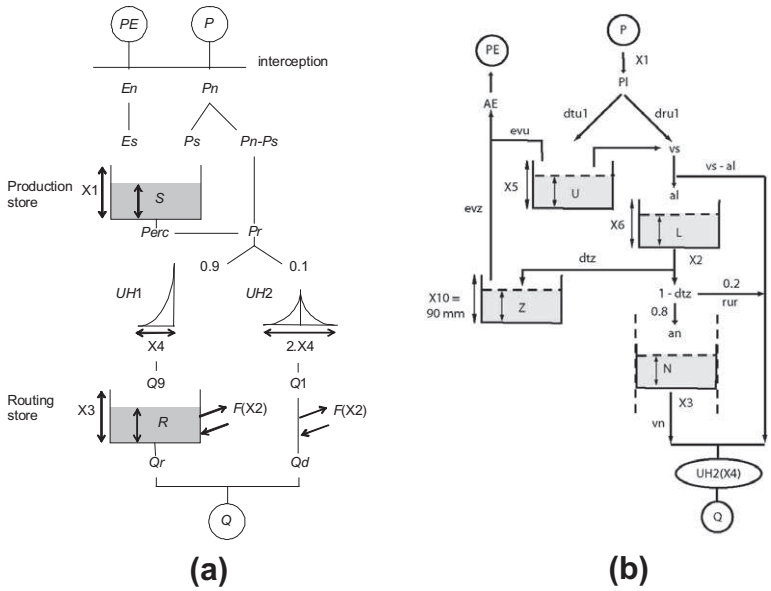


Fig. 3. Schematic diagrams of the structures of (a) the GR4J model and (b) the MORD model (PE: potential evapotranspiration; P: rainfall; Q: streamflow; X: parameters; other letters are internal model variables).

results on the whole set of catchments, we mainly used box plots based on the 940 × 2 results obtained in validation.

2.4. Criteria analysed

The analysis was carried out on the criteria detailed in Table 3:

- The first five criteria represent different Nash–Sutcliffe efficiencies calculated on flows (NSE_Q), root squared flows (NSE_{sqrtQ}), logarithmic transformed flows (NSE_{lnQ}), inverse flows (NSE_{iQ}) and relative flows (NSE_{rQ}). NSE_{iQ} is not commonly used in the literature and was suggested by Le Moine (2008). The comparison of these criteria will provide information on the effect of prior flow transformation.
- NSE_{uT} is the *NSE* calculated under a threshold taken equal to the 0.9 percentile of the observed flow duration curve of the test period (other thresholds could obviously be considered). Compared to the previous criteria, NSE_{uT} is calculated only on one-tenth of the time steps corresponding to the lowest observed flows and will provide information on the use of a threshold.
- IA_{rel} is the relative index of agreement proposed by Krause et al. (2005). Here, the normalization of the error is different from the one made in the *NSE*.
- B_{uT} is the bias under the same threshold as for NSE_{uT} (0.9 flow percentile). Contrary to the previous criteria that are all based on the squared error, B_{uT} is based on the cumulative error and will indicate the error of water balance in low-flow conditions.
- Last, $RLFD_{uT}$ is the flow deficit under the low-flow threshold (0.9 flow percentile). The formulation of $RLFD_{uT}$ is similar to that of B_{uT} and indicates the error of the water deficit in low-flow conditions based on the threshold.

As these criteria have no lower bounds, they may take strongly negative values in case of model failure. These individual values can severely impact the calculation of statistical moments of the set of criteria (e.g. the mean value). To avoid this bias, we used the mathematical transformation proposed by Mathevet et al. (2006), in which the range of variation of the transformed criterion is $]-1;1]$ instead of $]-\infty;1]$. The transformation is given by:

$$C^* = \frac{C}{2-C} \tag{2}$$

where C^* is the transformed criterion and C is the initial one. In addition to solving the problem of extreme negative values, the 0 and 1 values remain the same: $C^* = 0$ for $C = 0$ and $C^* = 1$ for $C = 1$, hence keeping the same interpretation relative to the benchmark. Note that for $C > 0$, C^* will be lower than C and the reverse for $C < 0$.

In the next sections, we will use the notations NSE^*_Q, NSE^*_{sqrtQ}, $NSE^*_{\ln Q}, \ldots, RLFD^*_{uT}$ for the bounded versions of the criteria listed in Table 3.

Note that for some criteria (e.g. $NSE^*_{\ln Q}$ and NSE^*_{iQ}), it is not mathematically possible to compute the transformed flow values when simulated or observed flows equal zero. However, these time steps should not be ignored during model evaluation, or else it could result in an overly optimistic assessment of model performance. To avoid this, one can add a small quantity to the flow value before computing its transformation. It can be, for example, a small fraction of the mean observed flow over the test period. In the following, ε will be set at one hundredth of the mean flow, but the choice of this value will be further discussed in Section 3.5. Note that the same comment applies to all criteria derived from Eq. (1) with negative λ values.

2.5. Approach for criterion analysis

The main objective of this study was to determine which criteria were actually the best adapted to the evaluation of low flows, by an objective analysis on a large set of catchments. To this aim, we identified the part of the hydrograph bearing most of the total model error. The procedure adopted is as follows:

1. Rank daily flow values over the study period by decreasing order. For each of them, the corresponding model squared error term *SE* can be computed.
2. Compute the empirical cumulative frequency, k/N, with k the rank of the flow value and N is the total number of time steps.
3. Compute the weighted empirical cumulative frequency by using the values of errors *SE* as weights w. The weighted cumulative frequency for the kth flow is therefore given by the ratio between the accumulated error for the k smallest flows and the total error (*SSE*) over the N time steps:

Table 3
Selection of evaluation criteria and their corresponding formulation and specific values (Q_i and $\widehat{Q}_i$ are the observed and simulated flows, respectively, n the total number of time steps, T a low-flow threshold (here $T = Q_{90}$), $\overline{Q}$ the mean of Q, and ε a small constant).

Criterion	Usual name	References	Mathematical formulation	Range of values	Value of perfect agreement	Target variable	Benchmark model
NSE_Q	NSE calculated on flows	Nash and Sutcliffe (1970)	$1-\frac{\sum_{i=1}^{n}(Q_i-\widehat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i-\overline{Q_i})^2}$	$]-\infty,1]$	1	Q_i	$\overline{Q_i}$
NSE_{sqrtQ}	NSE calculated on root squared transformed flows	Oudin et al. (2006)	$1-\frac{\sum_{i=1}^{n}\left(\sqrt{Q_i}-\sqrt{\widehat{Q}_i}\right)^2}{\sum_{i=1}^{n}\left(\sqrt{Q_i}-\sqrt{\overline{Q_i}}\right)^2}$	$]-\infty,1]$	1	$\sqrt{Q_i}$	$\overline{\sqrt{Q_i}}$
NSE_{lnQ}	NSE calculated on log transformed flows	Oudin et al. (2006)	$1-\frac{\sum_{i=1}^{n}(\ln(Q_i+\varepsilon)-\ln(\widehat{Q}_i+\varepsilon))^2}{\sum_{i=1}^{n}(\ln(Q_i+\varepsilon)-\overline{\ln(Q_i+\varepsilon)})^2}$	$]-\infty,1]$	1	$\ln(Q_i+\varepsilon))$	$\overline{\ln(Q_i+\varepsilon)})$
NSE_{iQ}	NSE calculated on inverse transformed flows	Le Moine (2008)	$1-\frac{\sum_{i=1}^{n}\left(\frac{1}{Q_i+\varepsilon}-\frac{1}{\widehat{Q}_i+\varepsilon}\right)^2}{\sum_{i=1}^{n}\left(\frac{1}{Q_i+\varepsilon}-\overline{\frac{1}{Q_i+\varepsilon}}\right)^2}$	$]-\infty,1]$	1	$\frac{1}{Q_i+\varepsilon}$	$\overline{\frac{1}{Q_i+\varepsilon}}$
NSE_{rQ}	NSE calculated on relative flows	Krause et al. (2005)	$1-\frac{\sum_{i=1}^{n}((Q_i-\widehat{Q}_i)/(Q_i+\varepsilon))^2}{\sum_{i=1}^{n}((Q_i-Q_i)/(\overline{Q_i}+\varepsilon))^2}$	$]-\infty,1]$	1	$\frac{Q_i}{Q_i}$	1
NSE_{uT}	NSE calculated under low-flow threshold		$1-\frac{\sum_{i\in[1,n]\mid Q_i<T}(Q_i-\widehat{Q}_i)^2}{\sum_{i\in[1,n]\mid Q_i<T}(Q_i-\overline{Q_i})^2}$	$]-\infty,1]$	1	$Q_i < Q_{90}$	$\overline{Q_i}(Q_i < Q_{90})$
IA_{rel}	Relative index of agreement	Krause et al. (2005)	$1-\frac{\sum_{i=1}^{n}((Q_i-\widehat{Q}_i)/Q_i)^2}{\sum_{i=1}^{n}((\mid Q_i-\overline{Q}_i\mid+\mid\widehat{Q}_i-\overline{Q}_i\mid)/\overline{Q}_i)2}$	$]-\infty,1]$	1	$\frac{Q_i}{Q_i}$	$\overline{Q_i}$
B_{uT}	Bias under low-flow threshold		$1-\frac{\sum_{i\in[1,n]\mid Q_i<T}\widehat{Q}_i}{\sum_{i\in[1,n]\mid Q_i<T}Q_i}$	$]-\infty,1]$	0	$Q_i < Q_{90}$	–
$RLFD_{uT}$	Ratio of Low-flow deficit		$1-\frac{\sum_{i\in[1,n]\mid Q_i<T}Max(0,T-\widehat{Q}_i)}{\sum_{i\in[1,n]\mid Q_i<T}(T-Q_i)}$	$]-\infty,1]$	0	$T < Q_i$	–

$$F_k = \frac{\sum_{i=1}^{k} w_i}{\sum_{i=1}^{N} w_i} = \frac{\sum_{i=1}^{k} SE_i}{\sum_{i=1}^{N} SE_i} = \frac{\sum_{i=1}^{k} SE_i}{SSE} \quad (3)$$

4. Summarize the weighted cumulative distribution as a box plot showing the flow percentiles of exceedance: 0.05, 0.25, 0.50, 0.75 and 0.95.
5. Transform back these flow percentiles into percentiles of the flow duration curve. The range between the 0.05 and 0.95 percentiles gives a 90% confidence interval of the range of observed flows contributing the most to the total model error.

An example of weighted and non weighted distributions and the results of steps 4 and 5 above are illustrated in Fig. 4. For the same example, Fig. 5 illustrates flow values in the hydrograph that contributes most of the model error. In this figure, the darker the observation, the larger the contribution of this time steps to the total model error.

This was done for each criterion and for each catchment. The resulting box plots can be compared between catchments. We did not analyse the results for each catchment individually, but for the whole catchment set to obtain more general conclusions. To this aim, we will draw a "master" box plot, in which each percentile shown will correspond to the median value of the same percentiles of all the box plots computed over the catchment set. An illustration of the derivation of the master box plot is presented in Fig. 6.

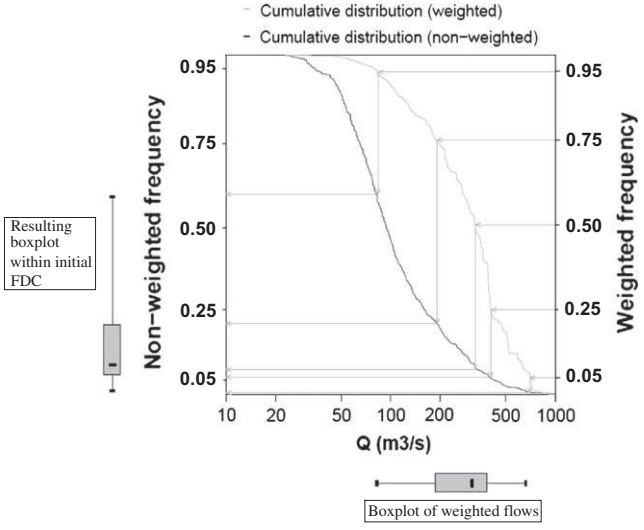


Fig. 4. Example of derivation of box plots from the distribution of weighted flows (here weights are the squared errors) to the flow duration curve (distribution of non weighted flows). The same example as in Fig. 1 is used here (period 01/01/1980–31/12/1981).

3. Results and discussion

This section first presents the distributions of the efficiency criteria over the 940 catchments and then analyses their interdependency and their behaviour in terms of dependence to low-flow values. This leads to a selection of criteria for the evaluation of low-flow simulation.

3.1. Overall efficiency distribution

Fig. 7 shows the box plots of the distributions of efficiency values for the entire catchment set (i.e. the 940 × 2 values obtained in validation) and for the nine criteria. The two models show quite similar behaviours.

The NSE^*_{uT} (NSE calculated under 0.9 exceedance flow percentile) exhibits very different behaviour from the other criteria. Its values mostly lie between −1 and 0. This indicates that on most catchments, the two models used are not better than the benchmark that provides the mean observed value under the threshold as the simulation. Although this result is a bit disappointing, it should be remembered that the variability of flow values in these conditions is limited, so that the benchmark is already a model that is difficult to equal (even if a constant flow is not perfect in terms of dynamics, it is already quite good). Besides, the benchmark is by construction perfect in terms of water balance, which means that no model can be better (see e.g. Perrin et al., 2006). Therefore, although they are clear in terms of the level of performance provided by the models, these mostly negative values make the crite-

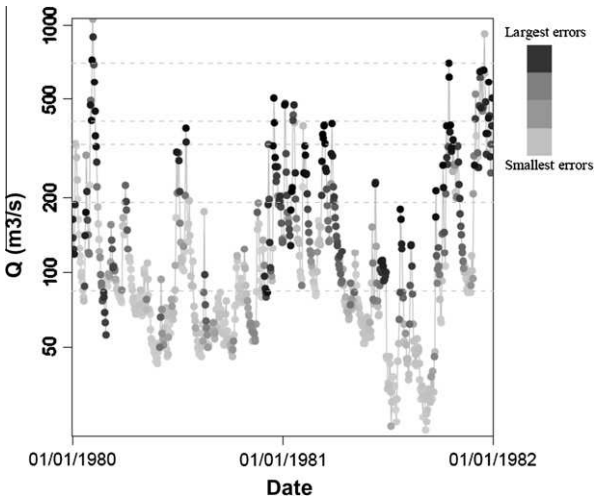


Fig. 5. Illustration of flow values that contribute the most of the total model error. The error considered here is the square error calculated on flows. The darker the dots, the larger the contribution to the total error. The same example hydrograph as in Fig. 1 is used here. Horizontal dashed lines correspond to the percentiles of the box plot shown in Fig. 4.

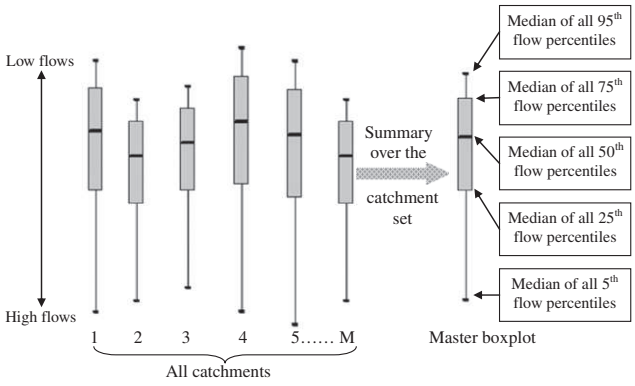


Fig. 6. Schematic representation of the derivation of master box plot which summarizes the behaviour of a given criterion over the entire catchment set (M stands for the total number of catchments).

ria difficult to interpret: the benchmark is probably too demanding in this case. The distributions of the other efficiency criteria show a much larger share of positive values (Fig. 7).

This variety of efficiency ratings obtained with the different criteria comes clearly from the differences in criteria formulations that put more or less emphasis on different parts of the hydrograph. This indicates that the tendency of modellers to declare that a model is "bad" or "good" based on the interpretation of relative efficiency criteria is often dangerous: it would be more rigorous to always refer to the benchmark used to say to what extent the model is "worse" or "better" than this benchmark.

3.2. Criteria interdependency

To analyse the possible interdependency between criteria, we drew scatter plots of pairs of criteria over the catchment set. The scatter diagrams for the possible combinations of criteria are presented in Tables 4 and 5 for the GR4J and MORD models, respectively (note that patterns in scatter diagrams were similar for the two models tested, which indicates that they are not model-specific). The scatter plots of NSE^*_Q vs. NSE^*_{sqrtQ}, NSE^*_{sqrtQ} vs. $NSE^*_{\ln Q}$, and NSE^*_{rQ} vs. IA^*_{rel} illustrate that there is a significant relation between these pairs of criteria. It means that they provide similar information on model efficiency. The relation between NSE^*_{rQ} and IA^*_{rel} could be expected since these criteria are based on the same model error and differ only in their normalization. The larger scatter between $NSE^*_{\ln Q}$ and NSE^*_Q shows that the log transformation substantially changes the information contained in the NSE_Q criterion. This is even more obvious when looking at the absence of relationships between NSE^*_Q, and NSE^*_{iQ} and NSE^*_{rQ} on the peak values. For the remaining pairs of criteria, the relationships are generally weak, indicating that they probably provide complementary information.

3.3. Analysis of model error terms

As explained in Section 2.5, we built master box plots over the entire set of catchments for the nine efficiency criteria to try to identify from which part of the flow duration curve most of the model error comes on average. The results are shown in Fig. 8. Here again, the results show similar patterns for the two simulation models, indicating that they are not highly model-dependent.

As could be expected for NSE^*_Q given the past analyses on this criterion, 90% of the model error accumulates below the 20th percentile of the *FDC*. This confirms the bias of NSE^*_Q towards high flows and strongly supports the suitability of NSE^*_Q for the evaluation of model performance on flood events.

The box plots for the NSE^*_{sqrtQ}, $NSE^*_{\ln Q}$, and NSE^*_{iQ} criteria show that the transformations with decreasing power (root square, logarithm, inverse) progressively put more emphasis on low flows. The inverse transformation seems the most appropriate if one wishes to concentrate on the very low flows, i.e. roughly the 20% lowest flows in the hydrograph. The box plots for NSE^*_{rQ} and IA^*_{rel} are similar. Ironically, $NSE^*_{\ln Q}$ did not show clear links with NSE^*_{rQ} and IA^*_{rel}, but these criteria seem to concentrate on similar parts of the hydrograph, with a stronger emphasis on low flows. NSE^*_{uT}, B^*_{uT} and $RLFD^*_{uT}$ stress the 10% lowest flows by construction.

3.4. Which criterion should be used for low-flow evaluation?

Our results tend to confirm previous comments made in the literature:

- NSE^*_Q should only be regarded as a criterion to evaluate the high-flow simulation efficiency and is of no use for the evaluation of low-flow simulations.
- NSE^*_{sqrtQ} provides more balanced information as the errors are more equally distributed on high- and low-flow parts, in agreement with the study by Oudin et al. (2006).
- $NSE^*_{\ln Q}$ is also intermediate like NSE^*_{rQ} and IA^*_{rel} but with greater emphasis on low flows. However, the sensitivity of $NSE^*_{\ln Q}$ to high flows noted by Krause et al. (2005) is confirmed here.

The *NSE* criterion on inverse flows (NSE^*_{iQ}) appears to be the most useful criterion to evaluate the very low flows among the criteria calculated over the total period, because it shows no sensitivity to high flows, contrary to $NSE^*_{\ln Q}$. This transformation is more relevant than the logarithmic transformation if one wishes to evaluate models in severe low-flow conditions. In comparison with NSE^*_{uT}, B^*_{uT} and $RLFD^*_{uT}$, it avoids the often arbitrary definition of a low-flow threshold: the most widely used low flows range from Q_{70} to Q_{99} (Smakhtin, 2001) but may depend on local conditions and regime type. In addition, it avoids obtaining the often negative values found with NSE^*_{uT}, which are due to an overly demanding benchmark model and are difficult to interpret.

Therefore, we would suggest that NSE^*_{iQ} is a good candidate (the most adequate of the criteria analysed here) to evaluate the low-flow simulation efficiency of a hydrological model. We wish to emphasize that even if at present $NSE^*_{\ln Q}$ is the most commonly preferred criterion to evaluate low flows, its value is influenced by a wider range of flows.

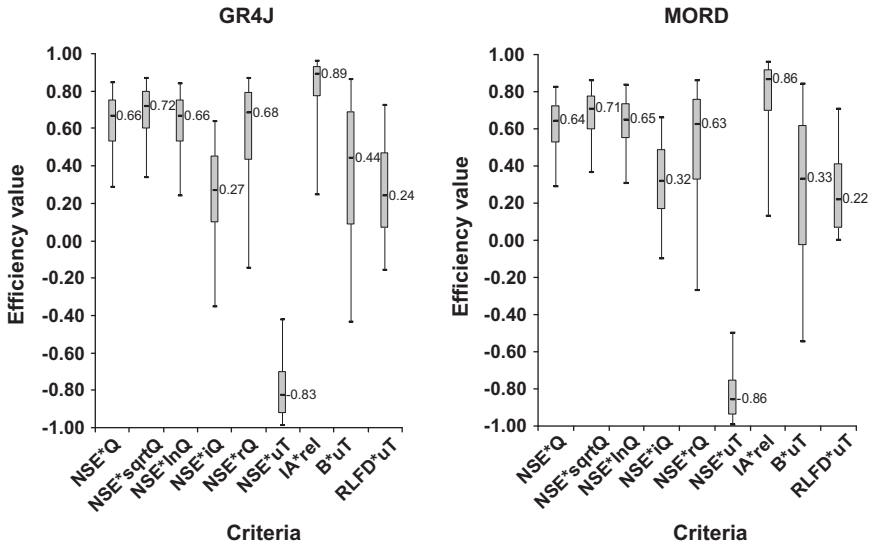
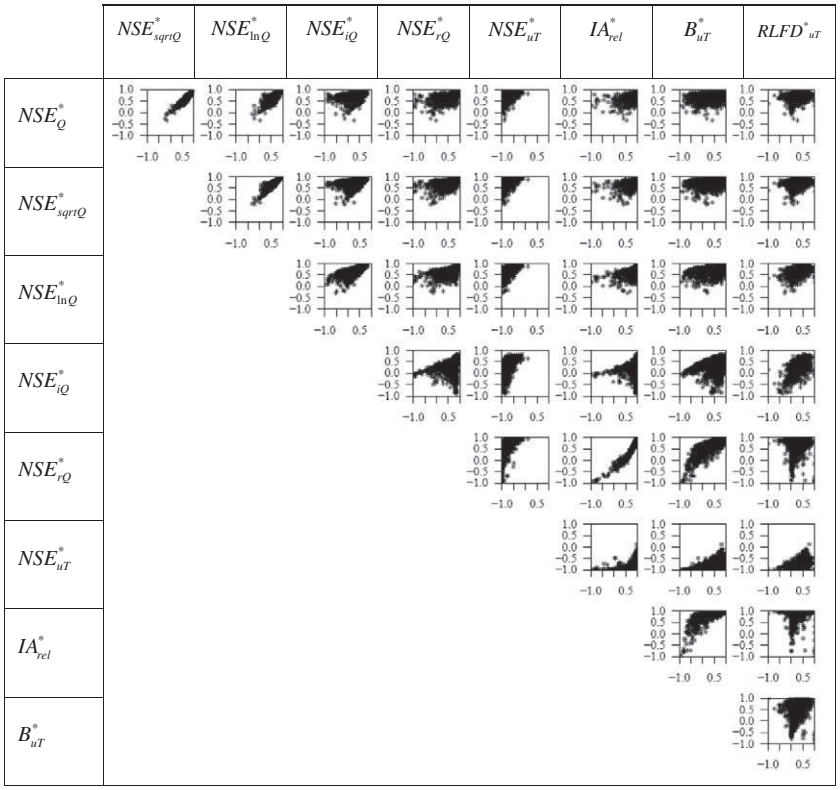


Fig. 7. Box plot of distribution of efficiency criteria obtained by the GR4J and MORD models over the entire catchment set in validation (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.05 and 0.95 percentiles).

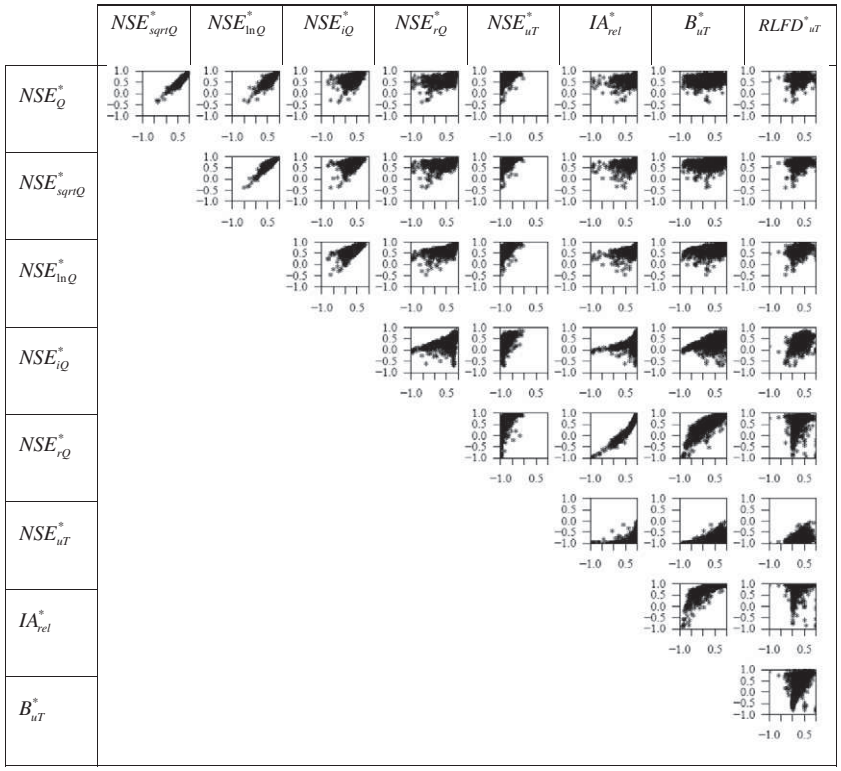
Table 4
Scatter plots of pairs of criteria on the 940-catchment set for the GR4J in validation.



3.5. What should be done with zero flows when calculating NSE^*_{iQ}?

As mentioned in Section 2.4, the calculation of the *NSE* on inverse flows raises a problem when the catchment is non-perennial or ephemeral, i.e. with some days with zero flows. To deal with zero-flow values, it was necessary to add a small constant, ε, to the flow values before calculating their transforms. However, the choice of ε may have an impact on the relative level of performance attributed to the model. As shown in Fig. 9, where values of ε ranging from one-tenth to one-hundredth of the mean flow were considered, the mean model efficiency on the catchment set varies significantly. ε should not be too large, or else it would tend to artificially enhance the relative level of model efficiency. Therefore the value of one-hundredth of the mean flow could be advised for ε, as it seems that this value corresponds to a plateau in the efficiency values, i.e. with low sensitivity of model efficiencies.

Table 5
Scatter plots of pairs of criteria on the 940-catchment set for the MORD in validation.



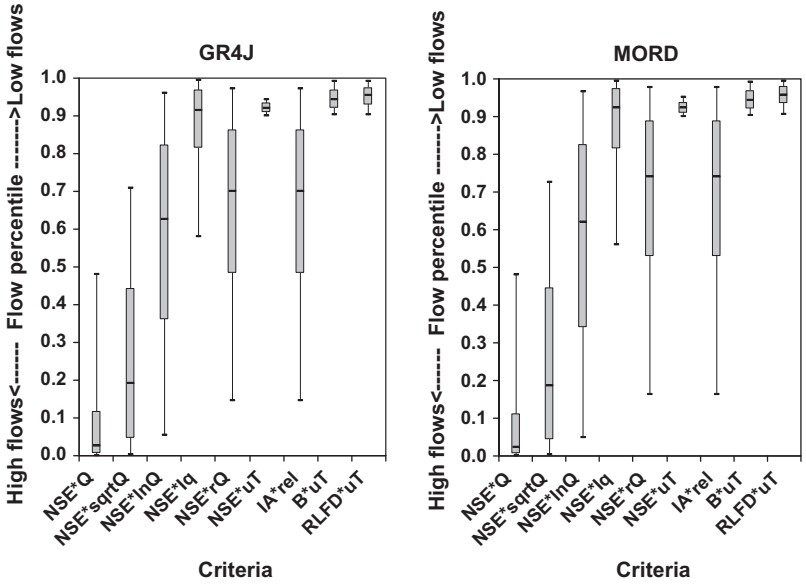


Fig. 8. Master box plot obtained over the entire catchment set for the GR4J and MORD models (each flow percentile of exceedance corresponds to the median of this percentile over all the box plots obtained on each catchment of the entire set).

3.6. Power transformation and criteria

The results in Section 3.3 clearly indicate that transformations on flows with power (λ) values lower than 0 give greater emphasis to low flows. A value of λ equal to -1 (i.e. the inverse transformation which we ultimately recommend) was shown to stress the approximately 20% of lowest flows on average on the catchment set, as shown in Fig. 8. However, the box plots in this figure indicate that this λ value does not emphasize the quality of model fit on exactly the same parts of the hydrographs for all catchments. Indeed, it may depend on regime characteristics, flow variability or model performance on low flows. We tried to quantify the potential variability of λ values from catchment to catchment which would emphasize model errors on the same parts of the

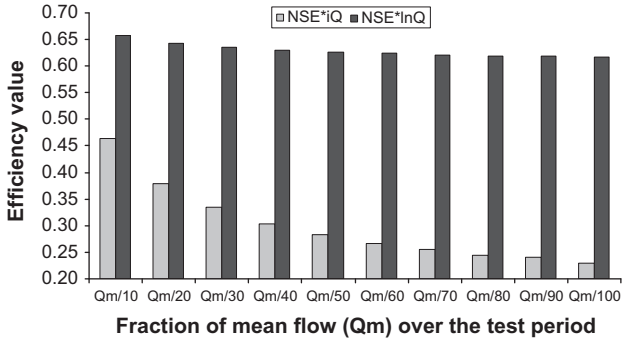


Fig. 9. Change in the mean value of NSE^*_{iQ} and $NSE^*_{\ln Q}$ obtained by the GR4J model in validation over the catchment set with different values of ε (fractions of Qm from Qm/10 to Qm/100).

hydrograph. For each catchment, we selected the λ value for which 80% of the largest errors were concentrated on the 20% lowest flows at most. Fig. 10 compares the λ values obtained on the two test periods for the two models. The range of λ values obtained for the two models is similar but also quite wide, which confirms that a single λ value will not emphasize the same flow range on different catchments. The scatter on the graphs also shows a quite strong dependency of λ values on the test period. A detailed analysis of the catchments where the λ value is very unstable shows that the criterion is very sensitive to a few time steps among the lowest flows in one period where the model makes large errors. Fig. 11 compares the λ values obtained by the two models. Interestingly, there is quite good agreement between them, though the scatter is quite wide for decreasing λ values. This is probably the indication that models behave similarly in low-flow conditions, but that their behaviour on extreme low flows is different. These results indicate a dependency on the characteristics of the test period and the model used.

The dependency of λ on flow characteristics is illustrated for the GR4J model in Fig. 12. It shows the relationship between λ and the Q_{90}/Q_{50} ratio that quantifies the variability of flows. Although the scatter is large, the value of λ seems to depend on this descriptor: the smoother the catchment response (i.e. the larger Q_{90}/Q_{50}), the lower the λ value.

All these results indicate that the interpretation of the Nash–Sutcliffe criterion based on inverse transformation does require experience and that one must be cautious when comparing values obtained on different periods and/or catchments; however, this was already the case with the original Nash–Sutcliffe criterion (see the comments by Perrin et al. (2006)).

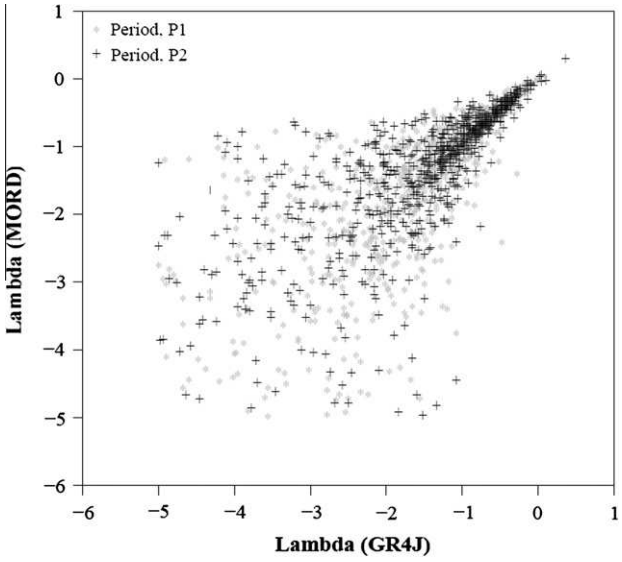


Fig. 11. Comparison of λ values obtained by the GR4J and MORD models.

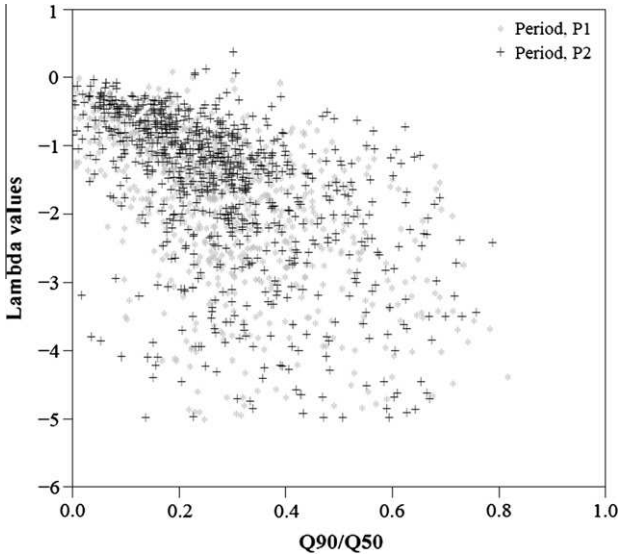


Fig. 12. Scatter plot of Q_{90}/Q_{50} vs. lambda values (λ) for the two test periods for the GR4J model.

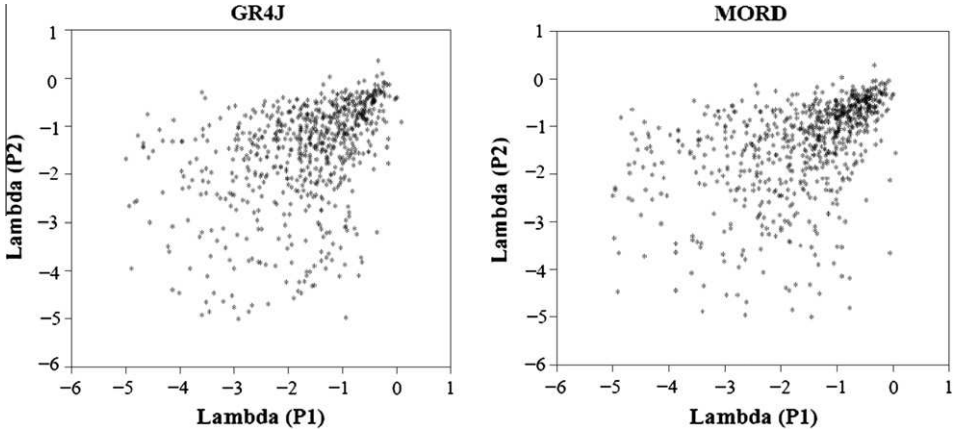


Fig. 10. Comparison of λ values obtained by GR4J (left) and MORD (right) models for the two test periods (P1 and P2).

4. Conclusions

This study aimed at identifying adequate criteria for evaluating the simulation of low flows using hydrological models. The behaviour of nine criteria was analysed on a set of 940 catchments. We used a master box plot representation to locate which part of the hydrograph contributes most of the error in these criteria, on average on the total catchment set. The detailed analysis of error distributions shows that the *NSE* calculated on inverse flows is better suited for the evaluation in very-low-flow conditions than the classically used *NSE* on logarithm flows, as it does not show sensitivity to high-flow values. It tends to focus on the 20% lowest flows on average, like the *NSE* on natural flows tends to focus on the 20% largest flows. Therefore, we recommend using this criterion calculated on inverse flow values for low-flow studies. However, it should be noted that the part of the flow duration curve that will cause most of the errors when using this transformation may differ from period to period and catchment to catchment, depending on model suitability and flow regime characteristics. This encourages to be cautious on the interpretation of this criterion when changing test conditions.

Acknowledgements

We are very grateful to Météo–France for providing meteorological data and SCHAPI for providing flow data. We also wish to thank DIM ASTREA of Région Ile-de-France and ONEMA for their financial support to conduct this study. Last, the two anonymous reviewers are thanked for their constructive comments on a previous version of the article, which helped to improve the manuscript.

References

ASCE, 1993. The ASCE task committee on definition of criteria for evaluation of watershed models of the watershed management committee, Irrigation and Drainage division, Criteria for evaluation of watershed models. J. Irrig. Drain. Eng. 119 (3), 429–442.

Andréassian, V., Hall, A., Chahinian, N., Schaake, J., 2006a. Introduction and Synthesis: Why Should Hydrologists Work on a Large Number of Basin Data Sets? Large Sample Experiments for Hydrological Model Parameterisation: Results of the Model Parameter Experiment–Mopex, vol. 307. IAHS Publication, pp. 1–5.

Andréassian, V., Bergstroem, S., Chahinian, N., Duan, Q., Gusev, Y.M., Littlewood, I., Mathevet, T., Michel, C., Montanari, A., Moretti, G., Moussa, R., Nasonova, O.N., O'connor, K., Paquet, E., Perrin, C., Rousseau, A., Schaake, J., Wagener, T., Xie, Z., 2006b. Catalogue of the models used in MOPEX 2004/2005, Large sample basin experiments for hydrological model parameterization: Results of the model parameter experiment–MOPEX, vol. 307. IAHS Publication, pp. 41–93.

Bartholmes, J.C., Thielen, J., Ramos, M.H., Gentilini, S., 2009. The european flood alert system EFAS – Part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrol. Earth Syst. Sci. 13, 141–153.

Boucher, M.A., Perreault, L., Anctil, F., 2009. Tools for the assessment of hydrological ensemble forecasts obtained by neural networks. J. Hydroinform. 11 (3–4), 297–307.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Stat. Soc. 26, 211–252.

Chiew, F.H.S., McMahon, T.A., 1993. Assessing the adequacy of catchment streamflow yield estimates. Aust. J. Soil Res. 31 (5), 665–680.

Chiew, F.H.S., Stewardson, M.J., McMahon, T.A., 1993. Comparison of six rainfall–runoff modelling approaches. J. Hydrol. 147, 1–36.

Clarke, R.T., 2008. A critique of present procedures used to compare performance of rainfall–runoff models. J. Hydrol. 352, 379–387.

Dawson, C.W., Abrahart, R.J., See, L.M., 2007. HydroTest: a web–based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. Environ. Model. Softw. 22 (7), 1034–1052.

de Vos, N.J., Rientjes, T.H.M., Gupta, H.V., 2010. Diagnostic evaluation of conceptual rainfall–runoff models using temporal clustering. Hydrol. Process. 24 (20), 2840–2850, DOI: 10.1002/hyp.7698.

Edijatno, Nascimento, N.O., Yang, X., Makhlouf, Z., Michel, C., 1999. GR3J: a daily watershed model with three free parameters. J. Hydrol. Sci. 44 (2), 263–277. doi:10.1080/02626669909492221.

Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi–objective calibration approaches in hydrological modelling: a review. J. Hydrol. Sci. 55 (1), 58–78.

Garçon, R., 1999. Modéle global Pluie–Débit pour la prévision et la prédétermination des crues (Lumped rainfall–runoff model for flood forecasting and design flood estimation, in French). La Houille Blanche 7 (8), 88–95.

Garrick, M., Cunnane, C., Nash, J.E., 1978. A criterion of efficiency for rainfall–runoff models. J. Hydrol. 36 (3–4), 375–381.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91.

Gupta, H.V., Kling, H., 2011. On typical range, sensitivity and normalization of mean squared error and Nash Sutcliffe efficiency type metrics. Water Resources Research 47, W10601.

Houghton–Carr, H.A., 1999. Assessment criteria for simple conceptual daily rainfall–runoff models. J. Hydrol. Sci. 44 (2), 237–261.

Jain, S.K., Sudheer, K.P., 2008. Fitting of hydrologic models: a close look at the Nash–Sutcliffe index. J. Hydrol. Eng. 13 (10), 981–986.

Klemes, V., 1986. Operational testing of hydrological simulation models. J. Hydrol. Sci. 31 (1), 13–24.

Krause, P., Boyle, D.P., Baese, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97.

Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrology and Earth System Sciences 11 (4), 1267–1277.

Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performance et du réalisme des modéles pluie–débit? PhD Thesis, Université Pierre et Marie Curie, Antony, 324 pp.

Le Moine, N., Andréassian, V., Perrin, C., Michel, C., 2007. How can rainfall–runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. Water Resour. Res. 43, W06428.

Mathevet, T., 2005. Quels modéles pluie–debit globaux au pas de temps horaire? Développements empiriques et comparison de modéles sur un large échantillon de bassins versants, PhD Thesis, des Eaux et des Foréts, Paris, 463 pp.

Mathevet, T., Michel, C., Andréassian, V., Perrin, C., 2006. A Bounded Version of the Nash–Sutcliffe Criterion for Better Model Assessment on Large Sets of Basins, Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment–MOPEX. vol. 307. IAHS Publication, pp. 211–219.

McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the Nash–Sutcliffe efficiency index. J. Hydrol. Eng. 11 (6), 597–602.

Merz, R., Parajka, J., Blöschl, G., 2009. Scale effects in conceptual hydrologic modelling. Water Resour. Res. 45, W09405. doi:10.1029/2009WR007872.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.

Moussa, R., 2010. When monstrosity can be beautiful while normality can be ugly: assessing the performance of event–based flood models. J. Hydrol. Sci. 55 (6), 1074–1084.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: Part I – a discussion of principles. J. Hydrol. 10, 282–290.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., 2006. Dynamic averaging of rainfall–runoff model simulations from complementary model parameterizations. Water Resour. Res. 42 (7), : W07410.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005. Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2––Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. J. Hydrol. 303 (1–4), 290–306.

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279 (1–4), 275–289.

Perrin, C., Andéassian, V., Michel, C., 2006. Simple benchmark models as a basis for model efficiency criteria. Archiv für Hydrobiologie. Large Rivers 17 (1–2), 221–244, Suppl. 161/1–2.

Quintana–Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., Morel, S., 2008. Analysis of near–surface atmospheric variables: validation of the SAFRAN analysis over France. J. Appl. Meteorol. Climatol. 47 (1), 92–107.

Ramos, M.H., Mathevet, T., Thielen, J., Pappenberger, F., 2010. Communicating uncertainty in hydro–meteorological forecasts: mission impossible? Meteorol. Appl. 17, 223–235.

Randrianasolo, A., Ramos, M.H., Thirel, G., Andréassian, V., Martin, E., 2010. Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models. Atmos. Sci. Lett. 11, 100–107.

Reusser, D.E., Blume, T., Schaefli, B., Zehe, E., 2009. Analysing the temporal dynamics of model performance for hydrological models. Hydrol. Earth Syst. Sci. 13 (7), 999–1018.

Schaefli, B., Gupta, H.V., 2007. Do Nash values have value? Hydrol. Process. 21 (15), 2075–2080.

Seibert, J., 2001. On the need for benchmarks in hydrological modelling. Hydrol. Process. 15 (6), 1063–1064.

Smakhtin, V.U., 2001. Low flow hydrology: a review. J. Hydrol. 240, 147–186.

Smakhtin, V.Y., Sami, K., Hughes, D.A., 1998. Evaluating the performance of a deterministic daily rainfall–runoff model in a low flow context. Hydrol. Process. 12 (5), 797–811.

Vidal, J.–P., Martin, E., Franchisteguy, L., Baillon, M., Soubeyroux, J.–M., 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. Int. J. Climatol. 30 (11), 1627–1644.

Weijs, S.V., Schoups, G., van de Giesen, N., 2010. Why hydrological predictions should be evaluated using information theory. Hydrol. Earth. Syst. Sci. 14 (12), 2545–2558.

Willmot, C.J., 1984. On the evaluation of model performance in physical geography. In: Gaile, G.L., Willmot, C.J. (Eds.), Spatial Statistics and Models. D. Reidel, Dordrecht, pp. 443–460.

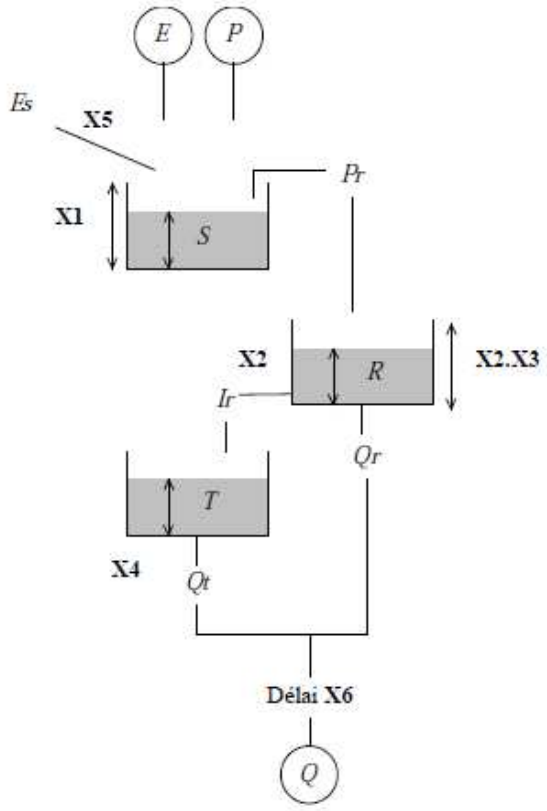
Ye, W., Jakeman, A.J., Young, P.C., 1998. Identification of improved rainfall–runoff models for an ephemeral low-yielding Australian catchment. Environ. Model. Softw. 13, 59–74.

C Description of the hydrological models tested in this study

This section presents a brief description of the hydrological models used in our study. The information on model structure, parameters and their original references are listed for each model. Some of the model structures presented is altered from their original references to transform them into a lumped one which makes the comparative analysis of model's performance simpler with the same input data.

C.1 Model Gardénia (GARD)

Structural layout



Reference of original model:Thiery (2009)

No of stores: 3

1. Surface store, S
2. Intermediary store, R
3. Underground store, T

No of parameters: 6
X1-Capacity of surface store
X2-Linear percolation constant
X3-Lateral outflow constant
X4-Outflow parameter of the underground store
X5-Correction coefficient of potential evapotranspiration
X6-Time lag

Mathematical formulation

$$S = S + P$$

$$\mathrm{Pr} = \max(0, S - X1); S = S - \mathrm{Pr}$$

$$Es = X5 \times E; S = S - Es$$

$$R = R + \mathrm{Pr}$$

$$Qr = \frac{R^2}{R + X2 \times X3}; R = R - Qr$$

$$Ir = R / X2; R = R - Ir$$

$$T = T + Ir$$

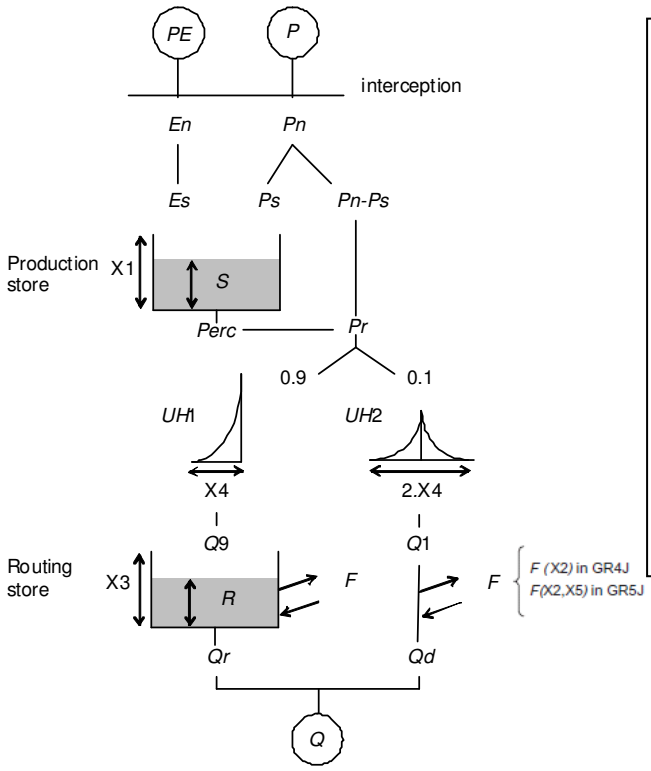
$$Qt = T / X4; T = T - Qt$$

$$Q = Qt + Qr$$

C.2 GR-series (GR4J & GR5J)

The structural layout of the GR4J and GR5J models are the same. The only difference is the additional 5th parameter in the GR5J model in the groundwater exchange function. Hence here we grouped these two models together to make the analysis simpler.

Structural layout



References:

1. GR4J: Perrin (2000)
2. GR5J: Le Moine (2008)

No of stores for the two models: 2

1. Soil moisture store, S
2. Routing store with power 5

No of parameters: 4 in GR4J & 5 in GR5J

X1- Capacity of production store
X2-Groundwater exchange function
X3-Capacity of routing store
X4-Time base of unit hydrograph
X5-Threshold for change in F sign

Mathematical formulation

if $P \geq E, Pn = P - E, En = 0$

if $P < E, En = E - P, Pn = 0$

$$Ps = \frac{Pn \times \left(1 - \left(\frac{S}{X1}\right)^2\right)}{1 + \frac{Pn}{X1}\left(1 + \frac{S}{X1}\right)}$$

$$Es = \frac{En \times \frac{S}{X1}\left(2 - \frac{S}{X1}\right)}{1 + \frac{En}{X1}\left(2 - \frac{S}{X1}\right)}$$

$$Perc = S - \left(S^{-4} + \left(\frac{9}{4}X1\right)^{-4}\right)^{\frac{1}{4}}; S = S - Perc$$

$$0 \leq j \leq X4, SH1(j) = \left(\frac{j}{X4}\right)^{\frac{5}{2}}; j > X4, SH1(j) = 1$$

$$0 \leq j \leq X4, SH2(j) = \frac{1}{2}\left(\frac{j}{X4}\right)^{\frac{5}{2}};$$

$$X4 < j \leq 2X4, SH2(j) = 1 - \frac{1}{2}\left(2 - \frac{j}{X4}\right)^{\frac{5}{2}}$$

$$j > 2X4, SH2(j) = 1$$

$$F = X2\left(\frac{R}{X3}\right)^{\frac{7}{2}} ------ (GR4J)$$

$$F = X2\left(\frac{R}{X3} - X5\right) ------ (GR5J)$$

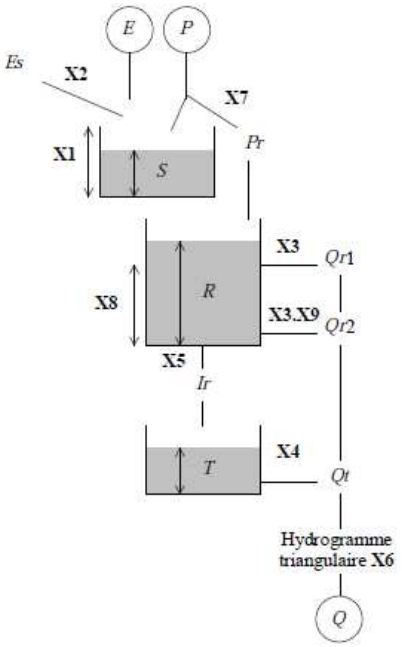
$$Qd = \max(0, Q1 + F); R = \max(\varepsilon, R + Q9 + F)$$

$$Qr = R - \left(R^{-4} + X3^{-4}\right)^{-\frac{1}{4}}; R = R - Qr$$

$$Q = Qr + Qd$$

C.3 HBV0

Structural layout



Reference: Bergström et Forsman (1973)

No of stores: 3

1. Soil store, S
2. Intermediary store, R
3. Groundwater store, T

No of parameters: 9

X1-Capacity of the soil reservoir
X2-Threshold of potential evapotranspiration
X3-Upper outflow constant of intermediary reservoir
X4-Outflow constant of the underground reservoir
X5-Percolation coefficient
X6-Time base of triangular hydrograph
X7-Exponent β
X8-Outflow threshold of the intermediary reservoir
X9-Lower outflow constant of the intermediary reservoir

Mathematical formulation

$\Pr = 0$

We realized a loop on each time step by decomposing the time step into five time steps.

Applying this rule, each of them a fifth of rain and a fifth of ETP.

$P5 = P/5; E5 = E/5$

$$\Pr i = P5 \times \left(\min(1, \frac{S}{X1}) \right)^{X7}; \Pr = \Pr + \Pr i$$

$$S = S + (P5 - \Pr i)$$

$$Esi = \min\left(S, E \times \frac{S}{X1} \right); S = S - Esi$$

$$R = R + \Pr$$

$$Qr1 = \max(0, (R - X8)/X3); R = R - Qr1$$

$$Qr2 = \frac{R}{(X3 \times X9)}; R = R - Qr2$$

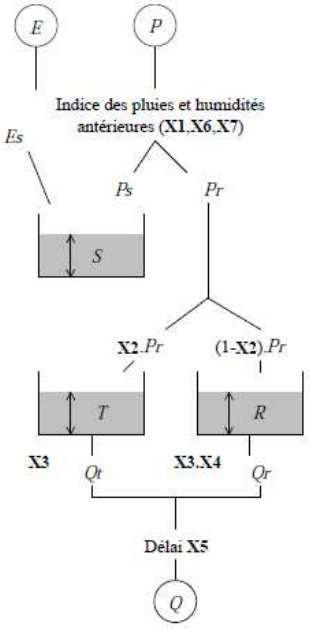
$$Ir = \min(S, X5); S = S - Ir$$

$$T = T + Ir; Qt = \frac{T}{X4}; T = T - Qt$$

$$Q = Qr1 + Qr2 + Qt$$

C.4 IHAC

Structural layout



Reference: Jakeman et al. (1990)

No of stores: 3

1. Surface store, S
2. Routing stores, T & R

No of parameters: 7

X1-Rainfall correction parameter
X2-Splitting coefficient of flow components
X3-Outflow constant of fast routing reservoir
X4-Outflow constant of slow routing reservoir
X5-Time lag
X6-PE modulation parameter
X7-Production parameter

Mathematical formulation

$$XS = S$$

$$E1 = \max(0., X7 - E / X6)$$

$$S = XS + \frac{P}{X1} - \frac{XS}{\exp(E1)}$$

$$\Pr = \frac{1}{2}(XS + S) \times P; T = T + X2 \times \Pr$$

$$R = R + (1 - X2) \times \Pr; Qt = \frac{T}{X3}$$

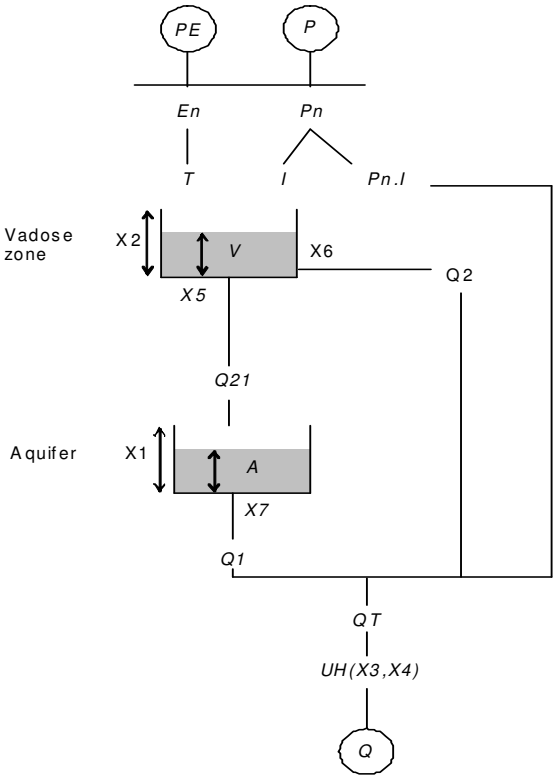
$$Qr = \frac{R}{X3 \times X4}; T = T - Qt$$

$$R = R - Qr$$

$$Q = Qt + Qr$$

C.5 MOHY

Structural layout



Reference: Fortin and Turcotte (2007)

No of stores: 2

1. Surface store (vadose zone), S
2. Underground store (aquifer), A

No of parameters: 7

X1-Capacity of the groundwater store
X2-Capacity of the surface reservoir
X3-Shape parameter of unit hydrograph
X4-Time base of unit hydrograph
X5-Outflow coefficient of vadose zone into the aquifer
X6-Outflow coefficient of vadose zone into the stream
X7-Outflow coefficient of aquifer into the stream

Mathematical formulation

if $P \geq E, Pn = P - E, En = 0$

if $P < E, En = E - P, Pn = 0$
$I - Infiltration; T - Transpiration$

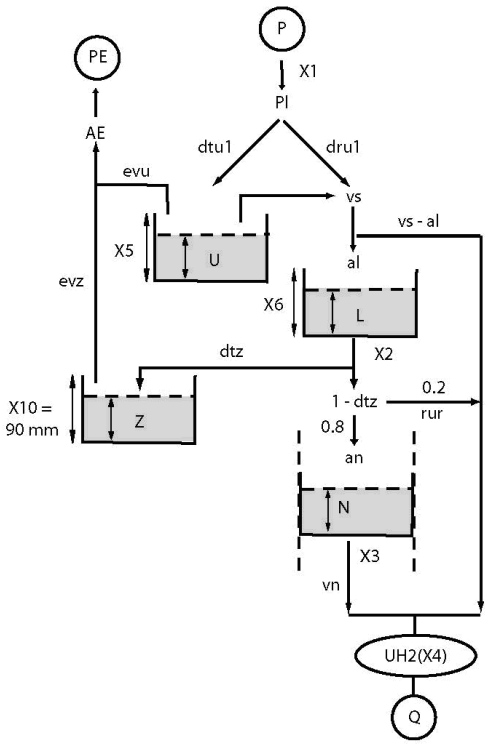
$Z = Z + I$
$I = \min(1, Z / X2); Z = Z + (1 - I) \times Pn$
$Z = Z - T$
$T = \min(Z / X2, En); Z = Z + \max(0, T)$

$Q2 = Z / X6; Z = Z - Q2$
$Q21 = Z / X5; Z = Z - Q21$
$A = A + Q21$
$Q1 = A / X7 / \max(X5, X6)$
$A = A - Q1$

$QT = Q1 + Q2 + I \times Pn$

C.6 MORD

Structural layout



Reference: Mathevet (2005)

No of stores: 4

1. Surface store, U
2. Evaporation store, Z
3. Intermediary store, L
4. Underground store, N

No of parameters: 6

X1-Correction coefficient of rainfall
X2-Outflow parameter for reservoir L
X3- Outflow parameter for reservoir N
X4-Time base of unit hydrograph
X5-Capacity of reservoir U
X6-Capacity of reservoir L
X10-Capacity of reservoir Z (here X10 is fixed and is 90mm)

Mathematical formulation

Correction for rainfall:

$Pl = P \times X_1$

Distribution of rainfall to U:

$dtr1 = Pl \times U / X5$

Evolution of the reservoir U:

$vs = dtr1 + \max(0, U - X5); U = \min(U + dtr1, X5); evu = \min(X5, E \times U / X5); U = U - evu$

Evolution of reservoir L:

$al = \min(X6 - L, vs \times (1 - L / X6); L = L + al; vl = L / X2; L = L - vl$

Evolution of reservoir Z:

$dtz = vl \times (1 - Z / 90); rur = 0.2 \times vl \times (Z / 90); an = 0.8 \times vl \times Z / 90; Z = Z + dtz$

$evz = \min(Z, (E - evu) \times Z / 90; Z = \min(90, Z - evz)$

Evolution of reservoir N:

$N = N + an; vn = \min(N, (N / X3)^3); N = N - vn$

Sum of different routing contributions:

$Qt = vs - al + rur + vn$

Unit hydrograph of GR4J (Perrin, 2000):

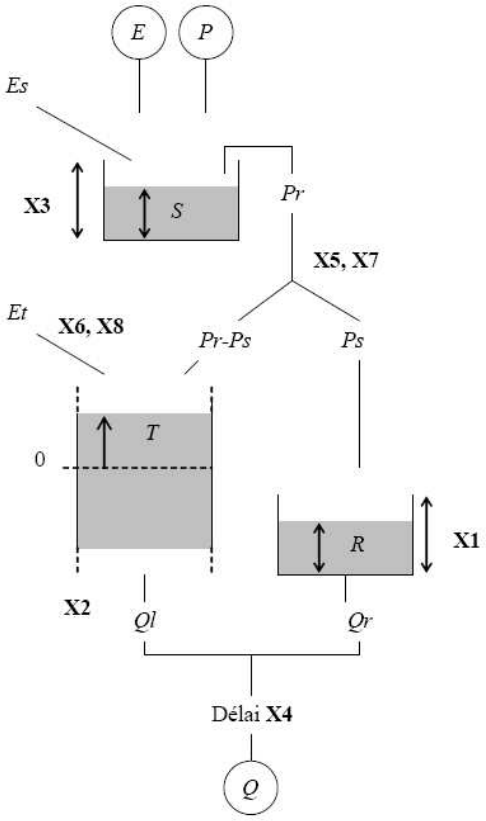
$0 \le j \le X4, SH2(j) = \frac{1}{2}(\frac{j}{X4})^{\frac{5}{4}}; X4 < j \le 2 \times X4, SH2(j) = 1 - \frac{1}{2}(2 - \frac{j}{X4})^{\frac{5}{4}}$

$j > 2 \times X4, SH2(j) = 1$

$Q = SH2(X4, Qt)$

C.7 TOPM

Structural layout



Reference: Beven and Kirkby (1979)

No of stores: 3

1. Interception store, S
2. Quadratic routing store, R
3. Underground store, T

No of parameters: 8

X1-Capacity of the quadratic reservoir
X2-Outflow parameter of the exponential reservoir
X3-Capacity of the interceptor reservoir
X4-Time lag
X5 & X6-Parameters of the topographic index curve
X7 & X8-Parameters controlling actual evapotranspiration

Mathematical formulation

$$S = S + P$$

$$Es = \min(S, E); S = S - Es; E' = E - Es$$

$$\Pr = \max(0, S - X3); S = S - \Pr$$

$$Ps = \frac{\Pr}{1 + \exp(X7 - \frac{T}{X5})}; T = T + \Pr - Ps$$

$$Es = \frac{E'}{1 + \exp(X6 - \frac{T}{X8})}; T = T + Es$$

$$R = R + Ps; Qr = \frac{R^2}{R + X1}; R = R - Qr$$

$$Ql = X2 \times \exp(\frac{T}{X2}); T = T - Ql$$

$$Q = Ql + Qr$$

D The developed model versions and their performances

D.1 Versions of model GR4J

As discussed in Chapter 4, several model versions were developed for the GR4J model structure. The modifications include structural, groundwater exchange function and changes in the percolation coefficient (see Appendix C). This section presents the developed versions of the GR4J model and their performances based on the selected evaluation criteria.

Table D.1 presents the derived versions of the model GR4J.

Table D.1 Versions of the model GR4J

Model version	Groundwater exchange function (*F*)		Additional routing stores						Parameters
	Eq.1	Others	Power-2 store	Power-5 store	Exponential store	Others	Added in parallel	Added in series	
N1	✓			✓			✓		5
N2	✓		✓				✓		5
N3		Formulation of Nascimento (1995)	✓				✓		5
N4		Eq.2	✓				✓		6
N5	✓		✓				✓		6
N6		Formulation of Nascimento (1995)	✓				✓		6
N7	✓		✓✓				✓	✓(New stores)	6
N8	-					2 linear stores	Optimised splitting coefficient	✓	6
N 9	✓					2 linear stores		✓	6
N10	-					3 linear stores	Optimised splitting coefficient	✓	8
N11	✓					3 linear stores		✓	8
N12	✓				✓		✓		5
N13	✓	Splitting coefficient applied to *F*							4
N14		Formulation of Nascimento (1995)							4
N15	✓	No percolation from the production store							4
N16	✓	Optimised coefficient of percolation							5
N17	✓					Power-6 store			4
N18	✓					Power-4 store			4

D.2 Test results of versions of GR4J

This section presents the performances of the different tested versions of the model GR4J. We used the NSE^*_Q, NSE^*_{lnQ} and NSE^*_{iQ} to evaluate the model's performances on high as well as on low-flow conditions.

Table D.2 Performances of GR4J model versions

Model acronym	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}	No of free Parameters
GR4J	0.621	0.617	0.23	4
N1	0.626	0.63	0.284	5
N2	0.618	0.635	0.287	5
N3	0.619	0.613	0.171	5
N4	0.623	0.646	0.319	6
N5	0.628	0.637	0.289	6
N6	0.632	0.64	0.263	6
N7	0.622	0.619	0.243	6
N8	0.501	0.496	0.24	6 (Lang, 2007)
N9	0.473	0.448	0.059	6 (Lang, 2007)
N10	0.507	0.505	0.29	8 (Lang, 2007)
N11	0.474	0.448	0.059	8 (Lang, 2007)
N12	0.628	0.638	0.328	5
N13	0.621	0.617	0.225	4
N14	0.62	0.582	0.063	4
N15	0.612	0.607	0.273	4
N16	0.624	0.625	0.264	5
N17	0.622	0.617	0.224	4
N18	0.617	0.613	0.231	4

D.3 Performance of versions of MORD & TOPM

As in the previous section, this section presents the results of the model versions of MORD and TOPM.

Table D.3 Performances of other model versions

Model acronym	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}	Number of free parameters
TOPM	0.574	0.584	0.216	8
TOPM9	0.603	0.606	0.216	9
MORD	0.603	0.616	0.302	6
MORD7	0.599	0.624	0.313	7

(where TOPM9 and MORD7 are the modified versions (with the integration of groundwater exchange functions) of TOPM and MORD models respectively. Here we are not going into the details of the model structures, as there is no structural modification in the new versions and only the integration of an exchange term, and it only changes the total number of parameters in both the models and structure remains the same. For more on the basic model structures, please see Appendix C.

E A downward structural sensitivity analysis of Hydrological models to improve low-flow simulation

Raji Pushpalatha, Charles Perrin, Nicolas Le Moine, Thibault Mathevet, Vazken Andréassian

(A*rticle published in Journal of Hydrology*)

Journal of Hydrology 411 (2011) 66–76

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

A downward structural sensitivity analysis of hydrological models to improve low-flow simulation

Raji Pushpalatha [a], Charles Perrin [a,*], Nicolas Le Moine [b], Thibault Mathevet [c], Vazken Andréassian [a]

[a] Cemagref, UR HBAN, 1, rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony Cedex, France
[b] Université Pierre et Marie Curie, UMR 7619 Sisyphe, 4 place Jussieu, 75252 Paris Cedex 05, France
[c] EDF–DTG, 21, avenue de l'Europe, BP 41, 38040 Grenoble Cedex 09, France

ARTICLE INFO

Article history:
Received 5 May 2011
Received in revised form 18 August 2011
Accepted 23 September 2011
Available online 29 September 2011

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Keywords:
Low flows
Simulation
Lumped model
Model efficiency
Uncertainty
Downward approach

SUMMARY

Better simulation and earlier prediction of river low flows are needed for improved water management. Here, a top–down structural analysis to improve a hydrological model in a low-flow simulation perspective is presented. Starting from a simple but efficient rainfall–runoff model (GR5J), we analyse the sensitivity of low-flow simulations to progressive modifications of the model's structure. These modifications correspond to the introduction of more complex routing schemes and/or the addition of simple representations of groundwater–surface water exchanges. In these tests, we wished to improve low-flow simulation while avoiding performance losses in high-flow conditions, i.e. keeping a general model.

In a typical downward modelling perspective, over 60 versions of the model were tested on a large set of French catchments corresponding to various low-flow conditions, and performance was evaluated using criteria emphasising errors in low-flow conditions. The results indicate that several best performing structures yielded quite similar levels of efficiency. The addition of a new flow component to the routing part of the model yielded the most significant improvement. In spite of the close performance of several model structures, we conclude by proposing a modified model version of GR5J with a single additional parameter.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Low flows and rainfall–runoff models

The occurrence of low flows is perhaps less spectacular than high flows, but low-flow consequences can be as costly, because they correspond to crucial periods in the functioning of both ecological and water management systems. For example, the cost of damage caused by the drought events in the years 1988–1989 in the United States was approximately US$40 billion, whereas the cost of the 1993 flood event was US$18–20 billion (Demuth, 2005). Thus, we consider that the simulation and advanced prediction of river low flows is an important challenge to improve low-flow management, both in the present climate and under the projected climate changes, which may well result in an increase in the occurrence of low-flow events (see e.g. Boé et al., 2009; Feyen and Dankers, 2009).

While a variety of lumped rainfall–runoff models are available to simulate streamflow irrespective of the flow conditions (see e.g. Singh and Frevert, 2002a,b), only a limited number of modelling studies focus on low-flow simulation. This study aims at identifying a generic model structure for improved low-flow simulation. Note that given the complexity of hydrological processes and the specificities of each catchment, some modellers have argued that model structures should be catchment-specific (e.g. Fenicia et al., 2008). However, we believe that before identifying catchment-specific models, the best possible general model that would include the representation of most of the dominant processes at work on catchments should be identified. This is the approach followed in this paper.

1.2. Specificities of the downward approach

To identify the general model structures that represent catchment behaviour, we followed a downward approach: a lumped representation of the catchment was used, in which only the main features of catchment hydrological behaviour are represented. This means that we did not attempt to build an explicit physical representation of the system but instead attempted to find the building blocks of the model that maximised modelling efficiency. The tests reported herein can be considered a structural sensitivity analysis. Some studies highlight the usefulness of sensitivity analysis for the improvement of hydrological models (see e.g. Andréassian et al., 2001; Oudin et al., 2006b; Tang et al., 2007; Bahremand and De Smedt, 2008; Ruelland et al., 2008). Other studies used sensitivity

* Corresponding author.
E-mail address: charles.perrin@cemagref.fr (C. Perrin).

0022-1694/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2011.09.034

analysis to better understand model behaviour with respect to inputs such as precipitation and potential evapotranspiration (Oudin et al., 2005a,b; Xu et al., 2006; Meselhe et al., 2009). Here we will focus on the sensitivity of low-flow simulation to the change in the components of the model structure responsible for low-flow simulation.

The main objective of this article is to analyse the extent to which a downward sensitivity analysis can help identify ways to improve low-flow simulation, while keeping the hydrological coherence in simulating the other parts of the flow regime. The downward search starts from a robust and parsimonious model structure. Then we will analyse how sensitive low-flow simulations are to the formulation of the model structure. This is done in trial-and-error mode, by testing many alternative model structures on a large set of catchments representing various physical and hydrometeorological conditions. The best candidate towards which our search converged is finally assessed in comparison with other model structures available in the literature.

1.3. A brief overview of low-flow modelling studies

The number of catchment modelling studies focusing on low-flow simulation using hydrological models is quite limited. One of the major problems with low-flow simulation is to account for surface water–groundwater interactions. During low-flow periods, water exchanges occur through the stream bed: the river may be fed by groundwater or, conversely, it may leak to feed the aquifer. Therefore, groundwater significantly influences low flows. A few studies that investigated these issues can be mentioned here. Fleckenstein et al. (2006) clearly mentioned the river–aquifer interactions and the significance of groundwater contribution during low-flow periods. Herron and Croke (2009) noted the improvement of lumped model predictions with the incorporation of groundwater exchange functions. The conclusions by Anderson et al. (2004) and Hughes (2004) also suggest that the model simulation efficiency can be improved by the addition of functions which represent the interaction between channel and aquifer flows. This is clearly shown in the study by Le Moine et al. (2007), who tested several options to account for inter-catchment groundwater flows using two rainfall–runoff models. Their results indicate that explicitly accounting for these groundwater fluxes significantly improves modelling efficiency.

Along with groundwater exchange functions, additional stores in the routing module can also enhance model performance, especially in the case of delayed flows (Wagener et al., 2004; Mathevet, 2005). Lang (2007) and Lang et al. (2008) analysed the performance of lumped models with respect to the addition of routing stores (to account for different water pathways underground) in an existing structure. Their study showed that some improvement can be achieved in the low-flow simulation, although they conclude that further work would be needed to improve lumped models for low-flow simulation. In a recent study, Kim et al. (2011) used the IHACRES-3S (3 Storage) model to evaluate the low-flow simulation together with the integration of base flow. The results showed a slight improvement in the model's performance, but they concluded that further studies are needed to obtain better low-flow simulation results. Last, Staudinger et al. (2011) analysed the sensitivity of recession simulation to various storage configurations on a snow dominated catchment in Norway within the FUSE framework. They conclude that the structural sensitivity is different in the winter and summer seasons, but that tests on a larger set of catchments are needed to get more general conclusions.

1.4. Scope of the paper

This article presents the end result of a long downward sensitivity analysis process that led to proposing an improved version of the GR4J catchment model (Perrin et al., 2003). Although our aim was to improve low-flow simulation specifically, we intended to find a generic solution, i.e. one that would improve low-flow representation without affecting the representation of high flows. This study builds on the previous studies by Mathevet (2005) and Le Moine (2008) who have already conducted tests to modify the existing model structures to improve modelling efficiency on a wide variety of catchments.

The next section discusses the data set and testing methodology. Then the results are presented and discussed before the concluding remarks.

2. Data set, models and methodology

This section presents the data set, models and testing methodology used for the analysis.

2.1. Data set

A set of 1000 catchments spread over France was used to test the model's generalisability (Andréassian et al., 2006). The database was built by Le Moine (2008). Continuous series of precipitation (P) and potential evapotranspiration (PE) were available for the 1970–2006 time period, providing a good variability of meteorological conditions, with quite severe drought periods (e.g. the years 1976, 1989–1991, 2003 and 2005). Meteorological data come from the SAFRAN reanalysis of Météo–France (Quintana-Segui et al., 2008; Vidal et al., 2010). Daily potential evapotranspiration was estimated using the formulation proposed by Oudin et al. (2005a) based on temperature and extra-terrestrial radiation. Streamflow (Q) data were extracted from the national HYDRO database. The length of the available flow record varies from one catchment to another, but at least 20 years of data were available on each selected catchment within the 1970–2006 period.

The variability of mean streamflow values can be expressed as a function of precipitation and PE. Fig. 1 plots the runoff coefficient (Q/P) as a function of the aridity index (P/PE) (see Mouelhi, 2003; Le Moine et al., 2007). It illustrates the variability of hydro-climatic conditions in the test catchments. As explained in detail by Le Moine et al. (2007), there are many catchments in this data set for which water losses are greater than PE (points lying below the line $y = 1 - 1/x$), which is an indication of leaky catchments. There are also catchments for which flow is greater than rainfall (points above the line $y = 1$), which mainly correspond to

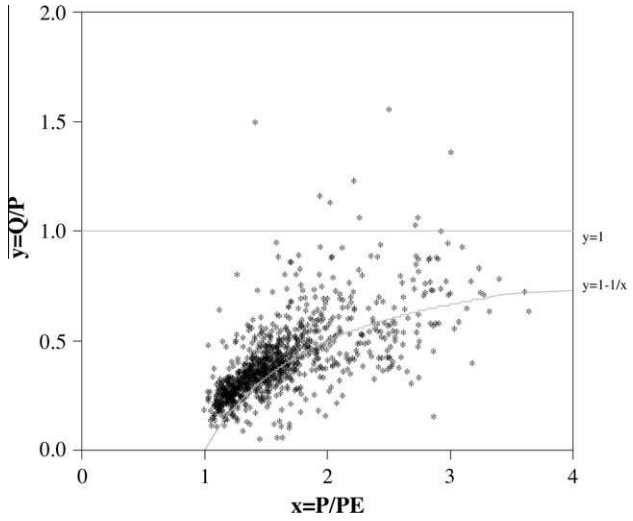


Fig. 1. Daily mean Q/P vs P/PE values for the 1000 catchments in the data set (P – rainfall; Q – streamflow; PE – potential evapotranspiration).

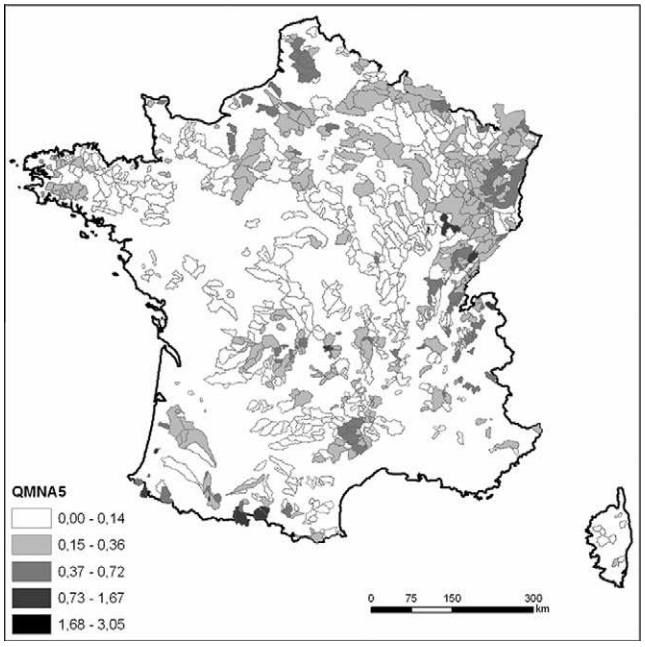


Fig. 2. Catchment location and illustration of the 5-year minimum monthly flow values.

catchments with karstic influences, i.e. those fed by inter-catchment groundwater flows from surrounding areas. Even though these catchments may prove more difficult to model, they were not discarded from the data set, as advocated by Andréassian et al. (2010).

Catchments with flow regulation structures (such as dams) were excluded from this data set. However, low flows may still be influenced by water withdrawals on some catchments: data on these influences were not available for this study. Given the size of our data set, the quality of flow data retrieved from the HYDRO database was trusted and not further checked in this study.

Catchment locations are shown in Fig. 2, along with the value of the minimum monthly flow of 5-year return period (called QMNA5 in France). QMNA5 is highly variable in this data set. It is influenced by various catchment characteristics, such as soil type, vegetation cover, geology and climatic conditions.

2.2. Tested models

The starting point of the present study was the GR4J rainfall–runoff model (Perrin et al., 2003), a lumped four-parameter model (see diagram in Fig. 3). It was already tested in various conditions with good results compared to other model structures. The water balance function that controls water balance in the GR4J model structure consists of a soil moisture accounting (SMA) reservoir (level S) and a conceptual water exchange function (F), expressed as:

$$F = X2 \cdot \left(\frac{R}{X3}\right)^{3.5} \quad (1)$$

in which $X2$ (mm) is the "groundwater" exchange coefficient and R and $X3$ (mm) are the water level and the capacity of the routing store, respectively. $X2$ can be positive or negative, meaning that the water exchange function can simulate imports or exports of water with the underground (i.e. connections with deep aquifers or surrounding catchments). Note that $X3$ is also used to parameterize the outflow from the routing store, which limits the interactions that would unavoidably exist between $X2$ and $X3$ if Eq. (1) was used alone. The routing part of the structure consists in two flow components routed by two unit hydrographs and a non-linear store. The latter is mainly responsible for low-flow simulations, along with leakage (percolation) from the SMA store. The groundwater exchange term F is added to the two flow components of the routing module.

Mathevet (2005) tested several modified versions of this model, especially by increasing the complexity of the routing part of the model and adding stores in parallel to the existing one. His tests, made at the hourly time step, showed limited sensitivity of model results, but the criteria he used focused more on high flows.

Following this work, Le Moine (2008) investigated the interactions between surface and groundwater and evaluated several modifications of the GR4J model to better account for these exchanges. These included different water exchange functions and the addition of a new store representing long-term memory. He proposed a five-parameter version of the model (GR5J) in which the groundwater exchange function has been modified (Fig. 3) to:

$$F = X2 \cdot \left(\frac{R}{X3} - X5\right) \quad (2)$$

where $X5$ is a dimensionless threshold parameter. It allows a change in the direction of the groundwater exchange within the year depending on the water level R in the routing store compared to this threshold. This model has shown significant performance improvement over the GR4J model, especially in low-flow conditions. It can be noted that the time-varying term F is only a very crude way to simulate groundwater–surface water connections. $X5$ can be seen as the external, quasi-stationary potential of the groundwater system and F is a "restoring flux" acting like a spring device with constant $X2$. Usually, $X2$ is negative: the more $R/X3$ departs from $X5$, the more intense the flux is, which tends to restore its value to $X5$.

Based on these previous results, the GR5J model's structure was used as a benchmark in our tests. In the subsequent sensitivity analysis, we will evaluate the extent to which modifications of the components used in the model to simulate low flows have an impact on model performance.

2.3. Model testing and assessment

The split sample testing scheme proposed by Klemes (1986) was used to evaluate model performance. For each catchment, the period where rainfall, PE and flow data were available (at most 1970–2006) was split into two halves (P1 and P2) of similar length, alternatively used for model calibration and validation. It means that for each catchment and each tested model, two calibrations and two validations were systematically performed. The first year of each test period was used for model warm-up. To avoid initialisation problems for catchments with long-term memories, five years of warm-up were considered in addition to the 1-year warm-up period: they were either the five years of observed data preceding the test period when available, or a mean year repeated five times otherwise. In this study, only performance in validation was considered to evaluate models.

The parameters were calibrated using a mean square model error calculated on root squared transformed flows as the objective function. This was found by Oudin et al. (2006a) to be a good compromise between high and low flows for model calibration. Here, as we focus on low flows, we could have chosen an objective function putting more weight on low flows. However, it would have been to the detriment of the simulation of high flows. So we preferred to keep this objective function to obtain a general model. This did not prevent us from assessing the model (in validation) over a wider range of criteria.

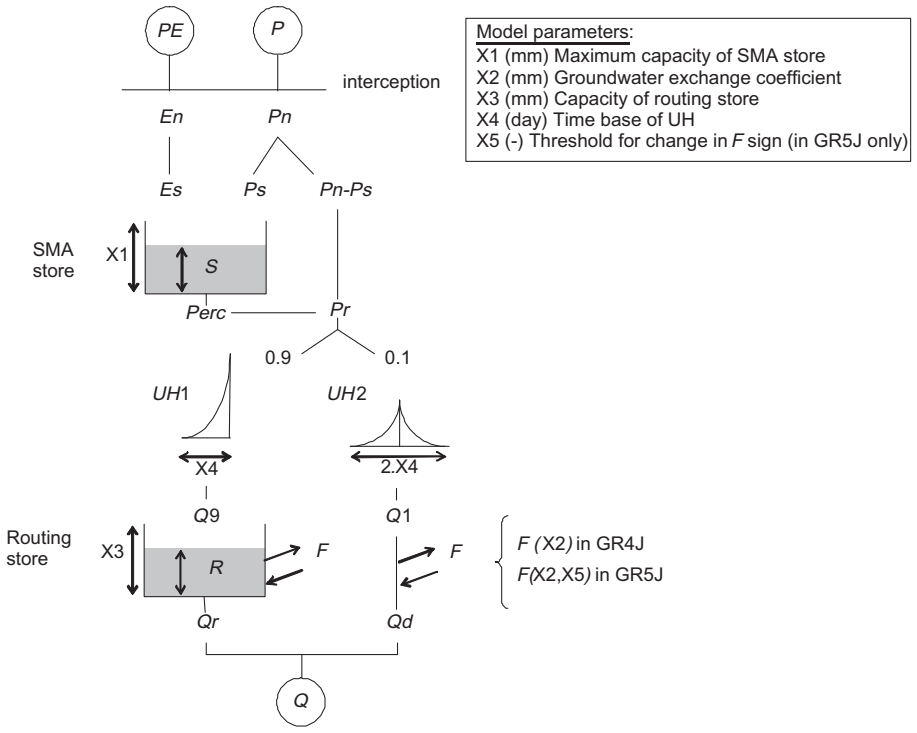


Fig. 3. Structure of the GR4J and GR5J models.

Several criteria based on the Nash and Sutcliffe (1970) efficiency index (*NSE*) were used to evaluate model performance in validation. *NSE* is given by:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{n}(Q_{obs,i} - \overline{Q_{obs}})^2} = 1 - \frac{E}{E_0} \quad (3)$$

where n is the number of time steps, $Q_{obs,i}$ and $Q_{sim,i}$ are the observed and simulated flows, respectively, at time step i. $\overline{Q_{obs}}$ is the mean of the observed flows over the selected period. E and E_0 are the mean squared error and the variance of observed flows respectively. The *NSE* index takes values over the range $]-\infty; 1]$, 1 indicating perfect simulation and 0 indicating a simulation equivalent to a constant flow equal to the mean observed flow.

As *NSE* has no lower bound, a bounded formulation of *NSE* was preferred (here noted NSE^*) to avoid the influence of strongly negative values while calculating the mean of the model performance over the test catchments (see Mathevet et al., 2006 for more details). NSE^* is derived from *NSE* using the following relationship:

$$NSE^* = \frac{1 - E/E_0}{1 + E/E_0} \quad (4)$$

NSE^* values vary over the range $]-1; 1]$. When $NSE = 1$ (i.e. $E = 0$), $NSE^* = 1$, and when $NSE = 0$ (i.e. $E = E_0$), $NSE^* = 0$, hence the interpretation of the two criteria is similar. Note that for $NSE > 0$, NSE^* values will be lower than *NSE* values and the reverse for $NSE < 0$.

The criterion on natural flows (NSE^*_Q) was used to check simulation consistency in high-flow conditions. The efficiency criteria calculated on logarithm transformed flows (NSE^*_{lnQ}) and inverse transformed flows (NSE^*_{iQ}) were used to put more weight on low-flow simulation. These prior transformations on flows are of the Box–Cox type. Pushpalatha et al. (submitted for publication) analyse the effect of such power transformations on *NSE* efficiency criteria and investigate which transformation seems more relevant to evaluate the efficiency in low-flow conditions. They found that an inverse transformation puts more weight on the 20% of lowest flows on average.

The overall performance of the tested models was computed over the 1000-catchment set, either using the mean value or distribution of performance criteria obtained in validation (i.e. a total of 2×1000 values).

Performance differences with the reference models were also quantified using the relative efficiency index initially suggested by Nash and Sutcliffe (1970) and more recently advocated by Seibert (2001) and Lerat (2009). This is a generalised form of the *NSE* criterion. It compares the performance of the tested model with respect to the performance of a benchmark model structure, by the following equation:

$$RE_{(sim/bench)} = 1 - \frac{\sum_{i=1}^{n}(Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^{n}(Q_{obs,i} - Q_{bench,i})^2} = 1 - \frac{E}{E_1} \quad (5)$$

where $Q_{bench,i}$ is the flow simulated by the benchmark, at time step i, and E_1 is the mean squared error of the benchmark model. Here GR5J was used as the benchmark in all the test cases. Like *NSE*, *RE* can be written under a bounded form (RE^*) using the same transformation as in Eq. (4) (here substituting E_0 by E_1 in Eq. (4)) and can be calculated on transformed flows, depending on the range of flows targeted in the analysis.

2.4. Structural sensitivity analysis

A sensitivity analysis of low-flow simulation to the formulation of the model structure was performed. We systematically evaluated various modifications of the GR5J model. Since it is difficult to detail all the tests, the following sections present the two main types of modifications that were performed, namely modifications of the groundwater exchange function and the routing component.

2.4.1. Sensitivity to changes in the groundwater exchange function

Groundwater (GW) is the main source for river flows during prolonged dry periods. Hence the recharge and release of groundwater is one of the important processes to consider for simulating low flows. Some authors considered only the flow towards the

stream using a specific groundwater reservoir (Davison and van der Kamp, 2008), but in the present analysis, we considered an exchange function that can account for both recharge and discharge from the groundwater reservoir, as in Le Moine (2008). During the course of this research, we evaluated the sensitivity of low-flow simulation to various formulations of the existing GW exchange functions.

2.4.2. Sensitivity to the addition of new stores

Stores and the empirical rules governing the transfer of water between them are the main components of rainfall–runoff models. Because of the complexity of the rainfall–runoff transformation, additional stores may improve model performance (Wagener et al., 2004; Mathevet, 2005). For low flows, this may provide additional components corresponding to different flow pathways.

In the GR5J model, a single routing store exists (R). A percolation from the soil moisture store also feeds flows during low-flow periods. As suggested by Mathevet (2005) and Le Moine (2008), we considered the parallel addition of new stores to the initial store, with various options to split effective rainfall into the different flow components. We also tested the serial addition of stores, as proposed by Lang (2007).

3. Results and discussion

In the following, we present the main results and discuss how sensitive low-flow simulations are to model formulation, following the modifications presented above. The selected versions of GR5J and their formulations are briefly presented in Table 1. As the number of modifications is almost infinite, we chose to present only a few of them to answer a number of simple questions that may arise when discussing the model's structure. Although these questions are sometimes interrelated, they are presented in sequence for the sake of clarity.

3.1. Can we design an improved model for low-flow simulation?

3.1.1. Can the existing groundwater exchange term in GR5J be improved?

We evaluated the sensitivity of low-flow simulation to various formulations of the groundwater exchange function. Starting from the GR5J model, several model versions that differ only by their groundwater exchange formulation were tested. In Table 1, three examples of modifications are provided:

- in M1, we gave seasonal dynamics to the exchanges by making them dependent on the SMA store and not on the routing store;
- in M2, we applied the splitting coefficient of flow components (0.1/0.9, as in Fig. 3) to the exchanges;
- in M3, we applied the formulation proposed by Nascimento (1995), i.e. making the exchanges a function of the level of the two stores, depending on the direction of the exchanges.

Fig. 4 shows the distribution of the performance of the selected versions, indicating significant sensitivity of the model's results to this function, which corroborates the findings of Le Moine et al. (2007). The existing exchange function in the base model appears to provide the best performance. This is in agreement with the results of Le Moine (2008) who had selected this function as the best performing among several other options.

3.1.2. Should the volumetric splitting between flow components be adapted to each catchment?

In GR4J and GR5J, 90% of the total effective rainfall is routed by the non-linear store (see Fig. 3). This volumetric proportion is fixed in the model for any catchment since Edijatno et al. (1999) showed that optimising it did not significantly improve the mean results. One factor to be considered when adding new stores to the initial store is the splitting coefficient of effective rainfall (SC in Fig. 5) between the stores. Mathevet (2005) and Le Moine (2008) conducted trials to divide effective rainfall between the existing store and an additional store. Their results tend to confirm that it is difficult to consider SC a free parameter. Here we simultaneously tested two model versions with two stores, one in which SC (version M4 in Table 1) was optimised on each catchment and the other in which SC was set at 0.4 (version M5).

The parameter analysis in M4 shows that SC values are very sensitive to the calibration conditions. The SC values obtained on the two test periods (P1 and P2) shows that this parameter is poorly defined and it will be difficult to relate it to catchment characteristics. The limited difference in model efficiency between the M4 and M5 versions (see Table 2) shows that SC can be set without significant efficiency loss. In the upcoming sections, we test versions considering only fixed splitting coefficients.

3.1.3. Should a new serial or parallel store be added?

Existing models propose a variety of conceptualisations for flow routing, using serial and/or parallel stores. Jakeman et al. (1990) discussed this issue in the IHACRES model, in which the routing

Table 1
Modified versions of the GR5J model and their main characteristics.

Model version	Characteristics of the groundwater exchange function		Characteristics of the additional routing stores					Number of routing stores	Number of free parameters
	Eq. (2)	Others	Power-2 store	Power-5 store	Exponential store	Added in parallel	Added in series		
M1	✓	Exchange dependent on SMA store						1	5
M2	✓	Splitting coefficient applied to F						1	5
M3		Formulation of Nascimento (1995)						1	4
M4	✓			✓		✓		2	7
M5	✓			✓		✓		2	6
M6	✓			✓			✓	2	6
M7	✓		✓			✓		2	6
M8	✓				✓	✓		2	6
M9	✓			✓	✓	✓		3	7
M10	✓				✓ ✓	✓		3	7
M11	✓		✓		✓	✓		3	7
M12	✓				✓	✓		2	6

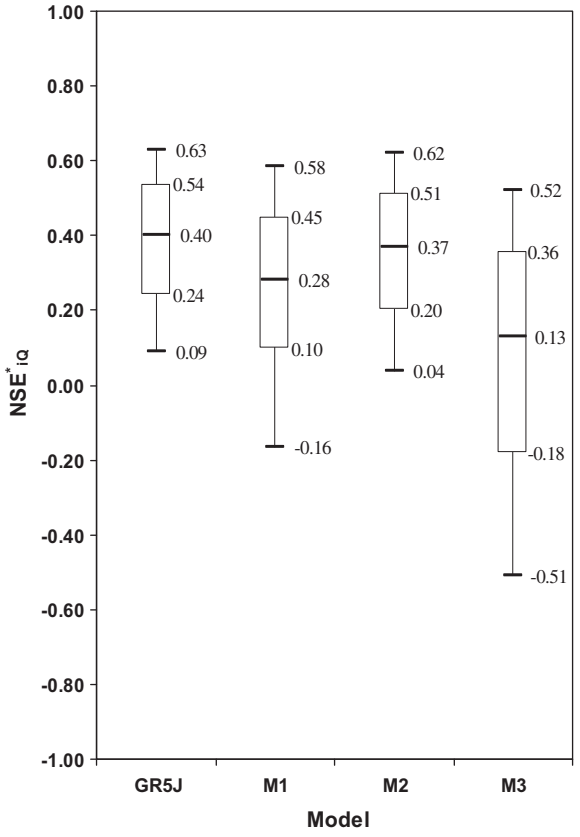


Fig. 4. Box plots of NSE^{*}_{iQ} values obtained in validation over the catchment set by GR5J and three model versions with modified groundwater exchange functions (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles).

module is made of linear stores. This model structure can be adapted to obtain several serial or parallel stores. Despite this flexibility, the authors indicate that in most cases, having two parallel stores is the most efficient configuration.

Here we analysed the sensitivity of the model's performance to the arrangement of routing stores, be they added to the existing parallel or serial stores. Two versions were tested, in which a new parallel store similar to the existing one was added (version M5, see Fig. 5) or a new serial store (version M6). Table 3 shows the mean performance of these two versions. The M5 version reaches higher efficiency values than M6. We also tried to add one more parallel routing store to obtain a third routed flow component (versions M9–M11 in Table 1). The results presented in Table 4 indicate that the improvements for low-flow simulation are not significant, which means that this additional complexity is not warranted by the data.

This confirms that the best compromise on average is to have two parallel stores. The series arrangement did not prove to be an efficient option. Therefore, following Jakeman et al. (1990), we suggest that the complexity of the routing part of the model should be increased by considering two independent flow components. This is a solution that provides more varied flow dynamics.

3.1.4. Does the formulation of the routing stores matter?

Here we tried to identify the best formulation of routing stores, i.e. the solution for which the model shows higher efficiency values. There is a variety of possible formulations of routing stores, ranging from linear to non-linear stores, e.g. power law or exponential stores (see Michel et al., 2003, for a good formulation of this store). In previous studies (see Edijatno and Michel, 1989; Edijatno et al., 1999), a power-5 non-linear routing store was identified as the most efficient. When adding a new parallel store, another formulation may be interesting to introduce a variety of behaviours in the flow components.

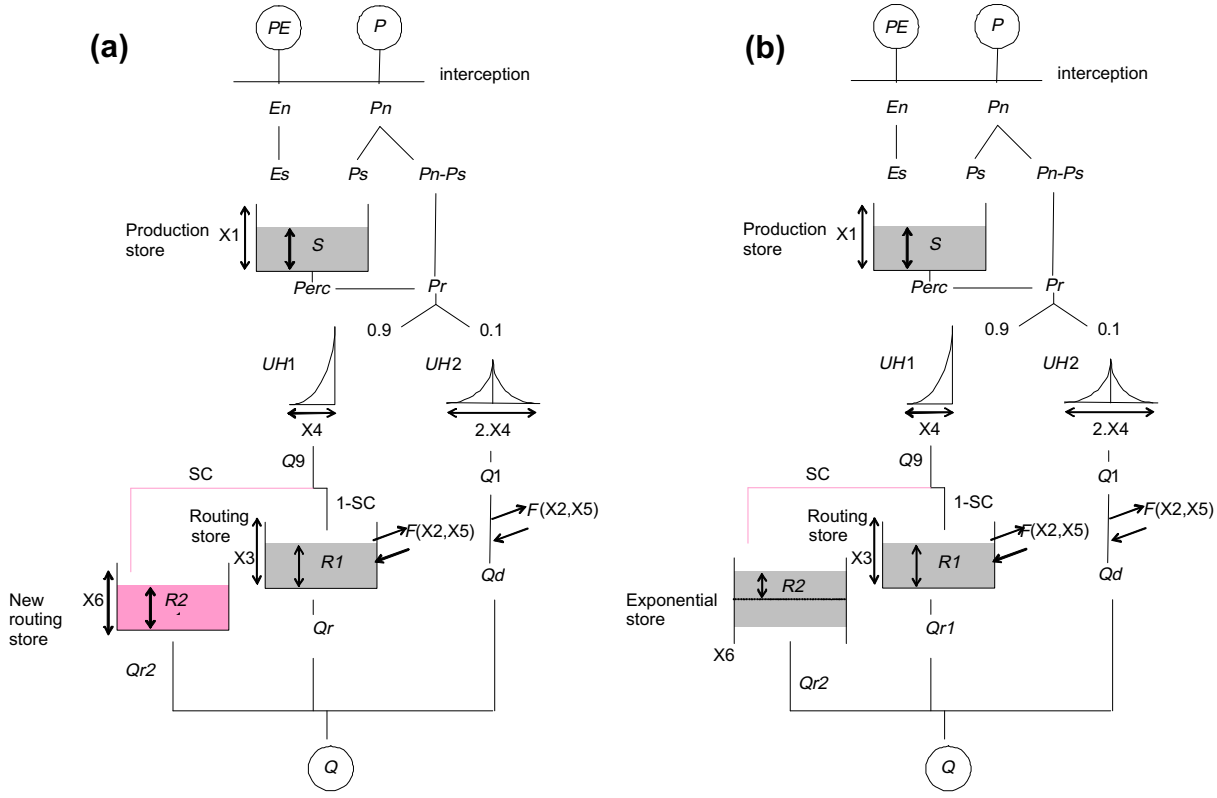


Fig. 5. Schematic representation of modified versions of the GR5J model (M5 (a) and M8 (b)).

Table 2
Mean model performance for versions M4 and M5.

	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}	Number of free parameters
M4	0.637	0.661	0.369	7
M5	0.634	0.659	0.365	6

Table 3
Mean efficiency values for versions M5 and M6.

	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
M5	0.634	0.659	0.365
M6	0.625	0.641	0.310

Table 4
Mean model performance for versions M8–M11 (multiple routing stores) and mean relative performance RE^* with reference to M8 over the catchment set.

	M8	M9	M10	M11
NSE^*_{iQ}	0.383	0.384	0.385	0.384
RE^* (%)	–	0.10	0.16	0.10

Various formulations were tested, among which we give the examples of versions M5, M7 and M8 in Table 1. Fig. 6 shows the corresponding distributions of efficiency values over the catchment set. The percentiles of the distribution of the model version M8 (with an additional exponential store) indicate better performance. The exponential store is known to be an efficient tool to simulate long recession spells (see Michel et al., 2003).

As suggested by Le Moine (2008) we also analysed the performance of M8 by removing the direct flow component (version M12). Indeed, the introduction of a new flow component may make this direct flow component unnecessary. However, the results are slightly lower than version M8 (Table 5), so we chose to keep this direct flow in the model. Note that this direct flow does not require specific free parameters.

Other versions were tested and several gave similar although slightly lower results. Thus, in all our tests, the M8 version was shown to be the most satisfactory and we chose to select it as a good candidate for providing improved low-flow simulation. We will call it GR6J hereafter (daily (J) version of the GR model with six free parameters).

3.2. Comparing the results of GR4J, GR5J and GR6J

This section quantifies the differences in the model's behaviour and performance between the GR4J, GR5J and GR6J versions in greater detail. Since GR5J was shown by Le Moine (2008) to yield better efficiency than GR4J, we mainly focus on the relative performance of GR5J and GR6J.

3.2.1. Relative performance of GR6J

The percentage improvement in the NSE^*_{iQ} values of GR6J are calculated in terms of relative efficiency values (RE^*). The RE^* values of the GR6J model are calculated with reference to the GR5J model. Table 6 shows the average relative performance for the three NSE^* criteria. The RE^* value based on NSE^*_{iQ} in Table 6 indicates a significant improvement in the simulation of low flows without losing efficiency on high flows. The significance of the improvement in performance is evaluated using the Student T-test at a 99% confidence level (T-values should be above 2.576). Although the differences may not seem large, remember that they were obtained on a large set of catchment, which makes them very significant (see also Mathevet (2005) for further discussion). When looking at the criterion on inverse flows, RE^* is positive on a majority of catchments, which means that the additional store improves this set of catchments.

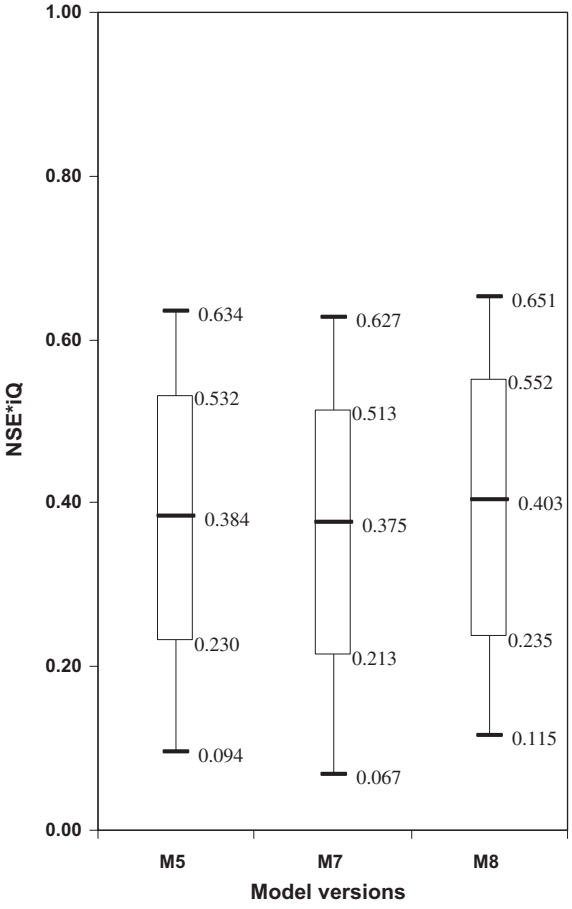


Fig. 6. Box plots of NSE^*_{iQ} values obtained in validation by model versions having different formulations of the additional routing store (boxes represent the 0.25 and 0.75 percentiles, with the median value inside, and the whiskers represent the 0.10 and 0.90 percentiles).

Table 5
Mean efficiency values of M8 vs M12.

	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
M8	0.634	0.662	0.383
M12	0.631	0.657	0.378

Table 6
Mean performance of GR5J and GR6J and relative performance of GR6J with reference to GR5J over the catchment set for various criteria (criteria on Q and iQ put more emphasis on floods and low flows, respectively), and significance of the improvement using the T-test (T-values should be greater than 2.576 at a 99% confidence level). The results were obtained in validation after calibration using another objective function.

Model	GR5J	GR6J	RE^* (%)
NSE^*_Q	0.629	0.634	0.83
NSE^*_{lnQ}	0.648	0.662	2.45
NSE^*_{iQ}	0.346	0.383	4.26

3.2.2. Illustration of the model's results

It is always difficult to select representative examples when working on a large catchment set. However, we wished to illustrate

the model's results on a few case studies, by providing simulated hydrographs. We selected three catchments (see Table 7) with different hydro-climatic conditions and considered their streamflow values for a period of 1 year. We chose the year 2003, which was one of the driest years over the past decade in France. Fig. 7 shows the observed flow series and the flow series simulated by the two models, GR5J and GR6J, for the three catchments. Note that the graphs use logarithmic scales to emphasise differences in low flows. In catchment A and B, the performance of the GR6J model is significantly better than GR5J's performance on the very low-flow. The GR5J model tends to underestimate these flows, especially in the case of catchment B. In catchment J, the two models give similar results and also similar dynamics in low-flow conditions, which indicates that the introduction of the new store is neutral on this catchment.

Table 7
Characteristics of sample catchments.

Catchment characteristics	A	B	J
Gauging station	Custines	Saint Michel	Drennec
River	Moselle River	Meuse River	Aber Wrac'h River
Catchment code	A7010610	B2220010	J3205710
Mean rainfall (mm/year)	1109	954	1087
Mean streamflow (mm/year)	530	378	593
Mean potential evapotranspiration (mm/year)	614	619	643
Catchment area (km^2)	6830	2540	24

3.2.3. Parameter stability and identifiability

Fig. 8 compares the stability of parameters of the GR5J and GR6J models obtained on the two calibration periods (P1 and P2). In general, there is a quite good agreement between periods, with the parameters showing good stability, which is a desirable property. However, the threshold values for groundwater exchange

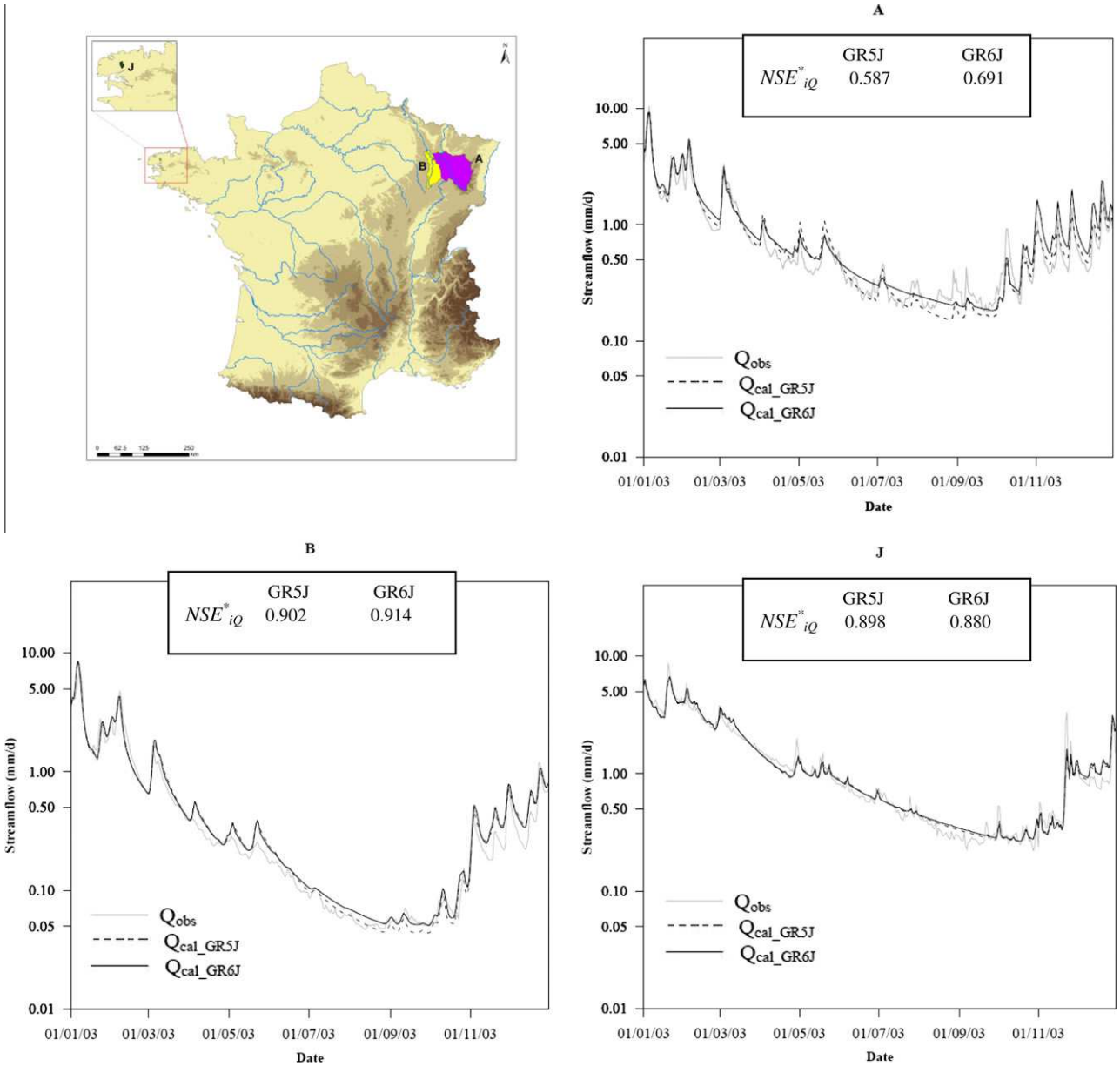


Fig. 7. Location of sample catchments and illustration of their hydrographs simulated by the GR5J and GR6J models, with corresponding NSE^*_{iQ} efficiency values.

(*X*5) and the sixth parameter of the GR6J model change significantly for a number of catchments, which may be due to a lower identifiability of these parameters for these catchments. The scatter seems a bit lower with the GR6J model for the *X*2 and *X*5 parameters. Interestingly, the reverse is observed for the capacity of the routing store (*X*3), for which the spread seems greater in

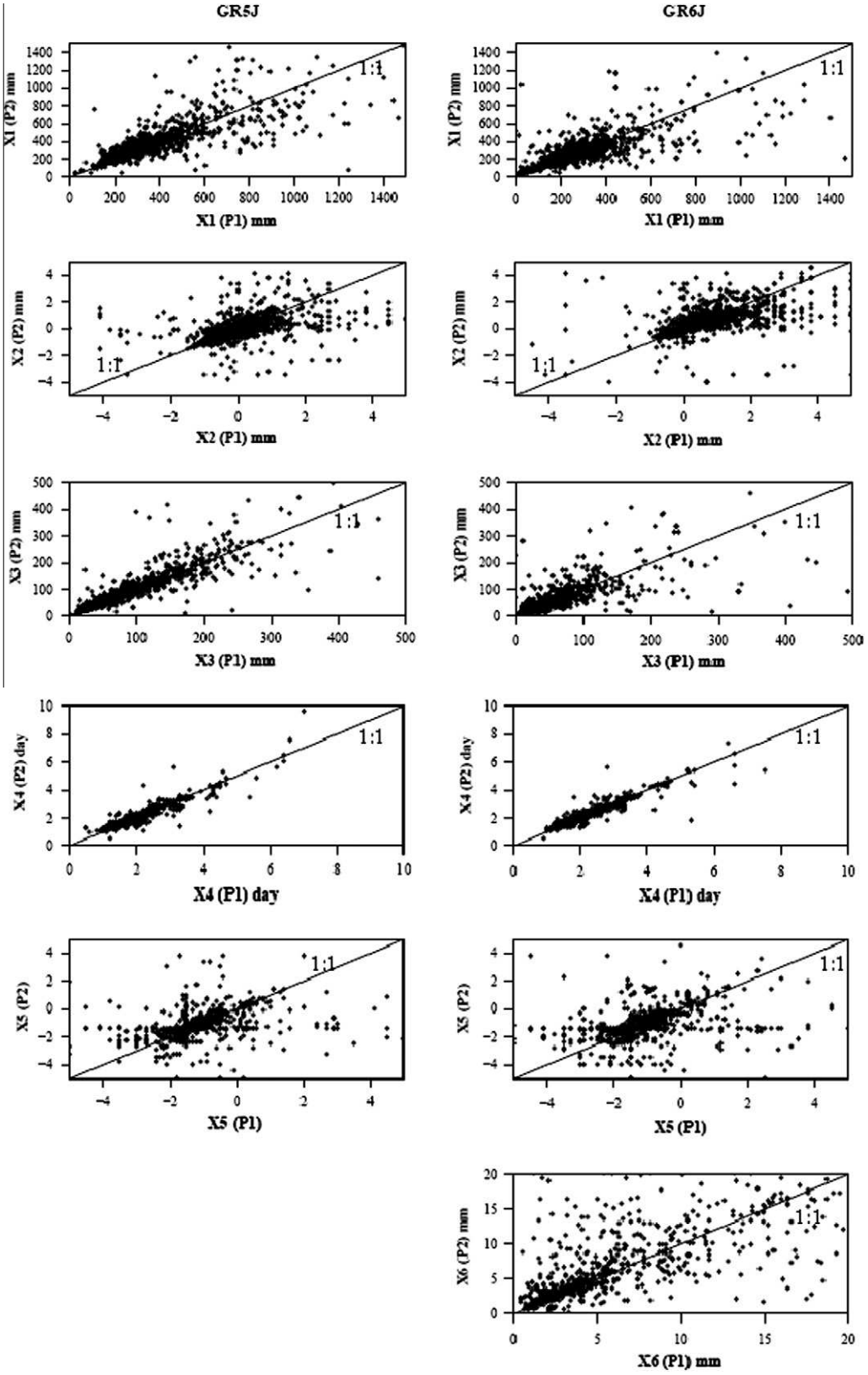


Fig. 8. Comparison of the parameter values obtained on the two calibration periods P1 and P2 for the GR5J and GR6J models.

Table 8
Average efficiency values for five lumped models compared to GR5J and GR6J for various criteria (criteria on Q and iQ put more emphasis on floods and low flows, respectively). Results were obtained in validation after calibration using another objective function.

Model acronym	Reference describing the original version	Number of free parameters	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
HBV0	Bergström and Forsman (1973)	9	0.546	0.559	0.156
IHAC	Jakeman et al. (1990)	6	0.528	0.556	0.196
MOHY	Fortin and Turcotte (2007)	7	0.493	0.554	0.229
MORD	Garçon (1999)	6	0.603	0.616	0.302
TOPM	Beven and Kirby (1979)	8	0.574	0.584	0.216
GR4J	Perrin et al. (2003)	4	0.621	0.617	0.230
GR5J	Le Moine (2008)	5	0.629	0.648	0.346
GR6J		6	0.634	0.662	0.383

the case of GR6J. This means that the introduction of the new routing store impacted the rest of the routing module, especially the initial routing store. The additional complexity in the model seems to be at the cost of a lower identifiability for some components. Note that in spite of the precautions taken for model initialisation, parameter optimisation may still have been hampered by unsuitable initial conditions on some groundwater dominated catchment, as discussed by Le Moine (2008).

3.2.4. GR6J vs existing models

To finalise the comparative assessment, the proposed GR6J version of the model was compared to independent lumped models. The selection of models is shown in Table 8, along with model's mean performance for three criteria. Note that, to be able to apply the models in exactly the same conditions (i.e. the same data, calibration procedure and testing scheme), we had to recode the models and sometimes slightly modify them. For this comparison, it was important to rely on model structures that were representative of various conceptualisations of low-flow simulation. More details on the modifications made are given by Perrin et al. (2003) and Mathevet (2005).

The average performance values calculated over the entire catchment set indicate that GR6J is suitable to simulate low flows on this data set, since it compares favourably well with the other models. While showing significant gains in low flows, it still remains efficient in high flows. This indicates that GR6J is a good candidate for various hydrological modelling applications, i.e. it is a generic model for end-users interested in the advantages of one-size-fits-all models (which we acknowledge may not be a generally shared opinion, see e.g. Savenije, 2009) or a good starting skeleton for hydrologists aiming at customised solutions.

4. Conclusions

Improving the low-flow simulation ability without impacting the high-flow simulation ability: this was one of our objectives in this study. We chose to proceed by trial and error, as recommended by Nash and Sutcliffe (1970) and Michel et al. (2006). Working on a large set of catchments proved to be a good way to prevent undue complexity in the proposed modified versions of the model. Here we started from the simple GR5J model and tested a number of modified versions, some having higher performance values compared to the initial model structures. The model's performance was not equally sensitive to all the tested modifications. Among the modifications that proved the most robust, the addition of an exponential routing store, in parallel to the existing routing store in the GR5J model, showed improvement in low flows on average, still remaining efficient in high-flow conditions. The complexity added by this modification (an additional free parameter) seems to be warranted by the model's results, as well as by the comparison to other existing models. In spite of this improvement, it is not possible to say that we improved the physical realism of this model, since the initial intention was not to explicitly represent the physical mechanisms. Instead this study focused more on identifying the main features of the rainfall–runoff transformation at the catchment scale and improving the model's predictive power. The improved model version provides a better representation of the catchment's hydrological behaviour.

Last, let us note that the level of performance in low-flow conditions seems to remain quite low. This may be for several reasons, including structural model errors, data quality or artificial influences. Nonetheless, it shows that specific research should be continued to improve the efficiency of hydrological models for low-flow simulation, a domain that was probably overlooked in the past.

Acknowledgements

We are very grateful to Météo–France for providing meteorological data and SCHAPI for providing flow data. We also wish to thank Domaine d'Intérêt Majeur Agrosciences, territoires, écologie, alimentation (DIM ASTREA) of Région Ile–de–France and Office National de l'Eau et des Milieux Aquatiques (ONEMA) for their financial support to conduct this study.

References

Anderson, M.L., Chen, Z.Q., Kavvas, M.L., 2004. Modeling low flows on the Cosumnes River. J. Hydrol. Eng. 9 (2), 126–134.
Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. J. Hydrol. 250, 206–223.
Andréassian, V., Hall, A., Chahinian, N., Schaake, J., 2006. Introduction and synthesis: why should hydrologists work on a large number of basin data sets? Large sample experiments for hydrological model parameterisation: results of the model parameter experiment – MOPEX. IAHS Publ. 307, 1–5.
Andréassian, V., Perrin, C., Parent, E., Bardossy, A., 2010. Editorial – The court of miracles of hydrology: can failure stories contribute to hydrological science? Hydrol. Sci. J. 55 (6), 849–856. doi:10.1080/02626667.2010.506050.
Bahremand, A., De Smedt, F., 2008. Distributed hydrological modeling and sensitivity analysis in Torysa Watershed, Slovakia. Water Resour. Manage 22, 393–408.
Bergström, S., Forsman, A., 1973. Development of a conceptual deterministic rainfall–runoff model. Nord. Hydrol. 4, 147–170.
Beven, K., Kirby, M.J., 1979. A physically based, variable contributing area model of basin hydrology. J. Hydrol. Sci. 24 (1), 43–69.
Boé, J., Terray, L., Martin, E., Habets, F., 2009. Projected changes in components of the hydrological cycle in French river basins during the 21st century. Water Resour. Res. 45, W08426.
Davison, B., van der Kamp, G., 2008. Low-flows in deterministic modelling: a brief review. Can. Water Resour. Assoc. 33 (2), 181–194.
Demuth, S., 2005. Low flows and droughts – a European prospective. IAHS Newsletter 82, March 2005, 7–8.
Edijatno, Michel, C., 1989. Un modèle pluie-débit journalier à trois paramètres. La Houille Blanche (2), 113–121.
Edijatno, Nascimento, N.O., Yang, X., Makhlouf, Z., Michel, C., 1999. GR3J: a daily watershed model with three free parameters. Hydrol. Sci. J. 44 (2), 263–277. doi:10.1080/02626669909492221.
Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L., 2008. Understanding catchment behavior through 664 stepwise model concept improvement. Water Resour. Res. 44, W01402. doi:10.1029/2006WR005563.
Feyen, L., Dankers, R., 2009. Impact of global warming on streamflow drought in Europe. J. Geophys. Res. 114, D17116.
Fleckenstein, J.H., Niswonger, R.G., Fogg, G.E., 2006. River–aquifer interactions, geologic heterogeneity, and low flow management. Ground Water 44 (6), 837–852.
Fortin, V., Turcotte, R., 2007. Le modéle hydrologique MOHYSE. Internal note, CEHQ, Québec.
Garçon, R., 1999. Modéle global Pluie-Débit pour la prévision et la prédétermination des crues (Lumped rainfall–runoff model for flood forecasting and design flood estimation, in French). La Houille Blanche 7 (8), 88–95.

Herron, N., Croke, B., 2009. Including the influence of groundwater exchanges in a lumped rainfall–runoff model. Math. Comput. Simul. 79, 2689–2700.
Hughes, D.A., 2004. Incorporating groundwater recharge and discharge functions into an existing monthly rainfall–runoff model. J. Hydrol. Sci. 49 (2), 297–311.
Jakeman, A.J., Littlewood, I.G., Whitehead, P.G., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. J. Hydrol. 117, 275–300.
Kim, H.S., Croke, B.F.W., Jakeman, A.J., Chiew, F.H.S., 2011. An assessment of modelling capacity to identify the impacts of climate variability on catchment hydrology. Math. Comput. Simul. 81, 1419–1429.
Klemes, V., 1986. Operational testing of hydrological simulation models. J. Hydrol. Sci. 31, 13–24.
Lang, C., 2007. Etiages et tarissements: vers quelles modélisations? L'approche conceptuelle et l'analyse statistique en réponse à la diversité spatiale des écoulements en étiage des cours d'eau de l'Est français. Ph.D. thesis. Univerité Paul Verlaine (Metz), 292 pp.
Lang, C., Gille, E., Francois, D., Drogue, G., 2008. Improvement of a lumped rainfall–runoff structure and calibration procedure for predicting daily low flow discharges. J. Hydrol. Hydromech. 56 (1), 59–71.
Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performance et du réalisme des modéles pluie-débit? Ph.D. thesis. Université Pierre et Marie Curie (Paris), Cemagref (Antony), 324 pp.
Le Moine, N., Andréassian, V., Perrin, C., Michel, C., 2007. How can rainfall–runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments. Water Resour. Res. 43, W06428.
Lerat, J., 2009. Quels apports hydrologiques pour les modèles hydrauliques? Vers un modèle intégré de simulation des crues. Ph.D. thesis. Université Pierre et Marie Curie (Paris), Cemagref (Antony, Montpellier), 300 pp.
Mathevet, T., 2005. Quels modéles pluie-debit globaux au pas de temps horaire? Développements empiriques et comparison de modéles sur un large échantillon de bassins versants. Ph.D. thesis. ENGREF (Paris), Cemagref (Antony), France, 463 pp.
Mathevet, T., Michel, C., Andréassian, V., Perrin, C., 2006. A bounded version of the Nash–Sutcliffe criterion for better model assessment on large sets of basins, large sample basin experiments for hydrological model parameterization: results of the model parameter experiment – MOPEX. IAHS Publ. 307, 211–219.
Meselhe, E.A., Habib, E.H., Oche, O.C., Gautam, S., 2009. Sensitivity of conceptual and physically based hydrologic models to temporal and spatial rainfall sampling. J. Hydrol. Eng. 14 (7), 711–720.
Michel, C., Perrin, C., Andreassian, V., 2003. The exponential store: a correct formulation for rainfall-runoff modelling. J. Hydrol. Sci. 48 (1), 109–124.
Michel, C., Perrin, C., Andréassian, V., Oudin, L., Mathevet, T., 2006. Has basin scale modelling advanced far beyond empiricism? IAHS Publ. 307, 108–116.
Mouelhi, S., 2003. Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. Ph.D. thesis. ENGREF, Cemagref Antony, France, 323 pp.
Nascimento, N.O., 1995. Appréciation à l'aide d'un modèle empirique des effets d'action anthropiques sur la relation pluie-débit à l'échelle du bassin versant. Ph.D. thesis. CERGRENE/ENPC, Paris, 550 pp.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I – A discussion of principles. J. Hydrol. 10, 282–290.
Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005a. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. J. Hydrol. 303 (1–4), 290–306.
Oudin, L., Michel, C., Anctil, F., 2005b. Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 1 – Can rainfall–runoff models effectively handle detailed potential evapotranspiration inputs? J. Hydrol. 303 (1–4), 275–289.
Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., 2006a. Dynamic averaging of rainfall–runoff model simulations from complementary model parameterizations. Water Resour. Res. 42, W07410.
Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., Michel, C., 2006b. Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. J. Hydrol. 320, 62–83. doi:10.1016/j.jhydrol.2005.07.016.
Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279 (1–4), 275–289.
Pushpalatha, R., Perrin, C., Le Moine, N., Andréassian, V., submitted for publication. Efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol.
Quintana-Segui, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., Morel, S., 2008. Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France. J. Appl. Meteorol. Climatol. 47 (1), 92–107.
Ruelland, D., Ardoin-Bardin, S., Billen, G., Servat, E., 2008. Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpretation on a large basin in West Africa. J. Hydrol. 361, 96–117.
Savenije, H.H.G., 2009. The art of hydrology. Hydrol. Earth Syst. Sci. 13, 157–161.
Seibert, J., 2001. On the need for benchmarks in hydrological modelling. Hydrol. Process. 15 (6), 1063–1064.
Singh, V.P., Frevert, D.K. (Eds.), 2002a. Mathematical Models of Large Watershed Hydrology. Water Resources Publications, Highlands Ranch, Colorado, p. 14 pp.
Singh, V.P., Frevert, D.K. (Eds.), 2002b. Mathematical Models of Small Watershed Hydrology and Applications. Water Resources Publications, Highlands Ranch, Colorado, p. 72 pp.
Staudinger, M., Stahl, K., Seibert, J., Clark, M.P., Tallaksen, L.M., 2011. Comparison of hydrological model structures based on recession and low flow simulations. Hydrol. Earth Syst. Sci. Discuss. 8, 6833–6866. doi:10.5194/hessd-8-6833-2011.
Tang, Y., Reed, P., Wagener, T., van Werkhoven, K., 2007. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrological Earth System Sciences 11, 793–817.
Vidal, J.-P., Martin, E., Franchisteguy, L., Baillon, M., Soubeyroux, J.-M., 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. Int. J. Climatol. 30, 1627–1644.
Wagener, T., Wheater, H.S., Gupta, H.V., 2004. Rainfall–Runoff Modelling in Gauged and Ungauged Catchments. Imperial College Press, London.
Xu, C., Tunemar, L., Chen, Y.D., Singh, V.P., 2006. Evaluation of seasonal and spatial variations of lumped water balance model sensitivity to precipitation data errors. J. Hydrol. 324, 80–93.

F Model's sensitivity to temporal variations of potential evapotranspiration

F.1 Introduction

Chapter 4, Appendix C and Appendix D discussed the model's structural sensitivity. Similar to the structural sensitivity, model results are also sensitive to the input data, rainfall and potential evapotranspiration (PE). Similar to the rainfall data, PE plays an important role in the long-term watershed balance and hence the imperfect information on PE can influence the calibration of model parameters and thereby it can influence model simulations. In a study, Andréasian et al. (2004) tested the model sensitivity (models GR4J and TOPM) to different Penman PE. The results indicate that the improved information on PE to the R-R model did not improve its efficiency. This might be due to the ability of calibration process to compensate for biased input data. At the end of their study, they highlighted the necessity to find alternate options to estimate PE other than the Penman PE to the R-R model. In another study, Oudin et al. (2005a) also concluded the insensitivity of models (GR4J, IHAC, HBV0 and TOPM) by testing them with two Penman PE data (mean PE and temporally varying PE) on a large catchment set. In a succeeding article, Oudin et al. (2005b) proposed a formulation to estimate the PE using temperature and radiation to provide best streamflow simulations.

At catchment scale, evaporation will play a key role in drying up soil which eventually will decrease the volume of water reaching the river. Hence much attention is needed while selecting the PE data to simulate stream low flows. Therefore, in the present study, we focused on the extend of model's PE sensitivity (mean PE and temporally varying PE) on low-flow simulations using Oudin's PE formulation. The data set discussed in chapter 2 is used to test the models. The PE model is as follows:

$$PE = R_e(T_a + 5)/100 \times \lambda\rho \qquad \textbf{Eq. F.1}$$

where R_e is the extraterrestrial radiation ($MJm^{-2}day^{-1}$); T_a is the mean daily air temperature ($^{\circ}C$); λ is the latent heat flux ($MJKg^{-1}$); ρ is the density of water (Kgm^{-3}).

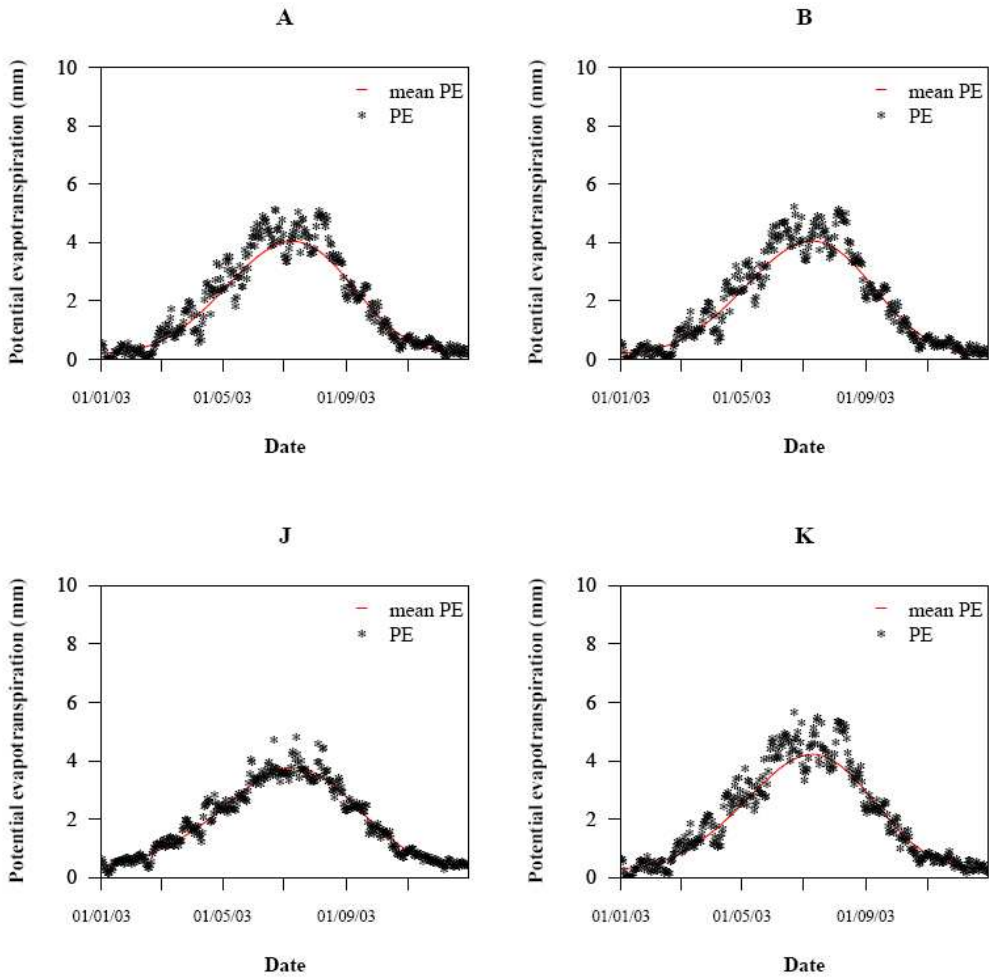


Figure F.1: Daily mean PE and temporally varying PE (the inter-annual mean PE was calculated for a period of 1970 to 2006)

Temporal variation of PE in the sample catchments (for more details on the sample catchments, see chapter 2) for a period of one year (2003, one of the driest periods in France) is presented in **Figure F.1**. We can see the variability in PE demand in the four catchments and in this study, we are interested to analyse the influence of temporal variation of PE on stream low-flow simulation efficiency of R-R models. As in **Figure F.1**, we tested the models with two sets of PE data, one with daily mean PE and another with temporally varying PE.

F.2 Tested models

The GR-series (GR4J, GR5J and GR6J) which we already discussed in chapter 4 are used to analyse the sensitivity with PE data due to their higher performance in both high- as well as low-flow simulations. The model testing and validation approach is similar to the one used in

these previous chapters. The performances of these models are evaluated using the NSE^*_Q, NSE^*_{lnQ} and NSE^*_{iQ} criteria in order to assess the models in both high- and low-flow conditions.

F.3 Results and discussion

This section presents brief description of the sensitivity of model performances with different PE data (mean PE and temporally varying PE). The mean efficiency values over the full data set for the tested models are presented in **Table F.1**. The percentage values presented in **Table F.1** are considerable as they are calculated on a large set of catchments. Also any improvement in simulation can make an improvement in model's forecast quality, hence we need to consider these percentage values listed in **Table F.1**. The results suggest that the temporally varying PE can improve the model's low-flow simulation efficiency compared to the mean PE.

Table F.1: Mean efficiency values in validation

Model	NSE^*_Q		NSE^*_{lnQ}		NSE^*_{iQ}		Difference (%)		
	With PE	With mean PE	With PE	With mean PE	With PE	With mean PE	NSE^*_Q	NSE^*_{lnQ}	NSE^*_{iQ}
GR4J	0.621	0.611	0.617	0.604	0.230	0.224	1.0	1.3	0.6
GR5J	0.629	0.617	0.648	0.634	0.346	0.346	1.2	1.4	0.0
GR6J	0.634	0.622	0.662	0.643	0.383	0.367	1.2	1.9	1.6

The results also reveal that the sensitivity changes from one model to another with respect to their structural changes and groundwater exchange functions. Compared to the other tested models, the mean percentage increment in the performance level of GR6J (see the highlighted values in **Table F.1**) over the full data set is 1.9% and 1.6% in case of NSE^*_{lnQ} and NSE^*_{iQ} with the temporally varying PE data. This shows its sensitivity to the PE data compared to the other two models. This might be due to the additional exponential store in the structure compared to the other GR6 models tested in this study. The GR6J simulations with mean PE and temporally varying PE for the sample catchments are presented below:

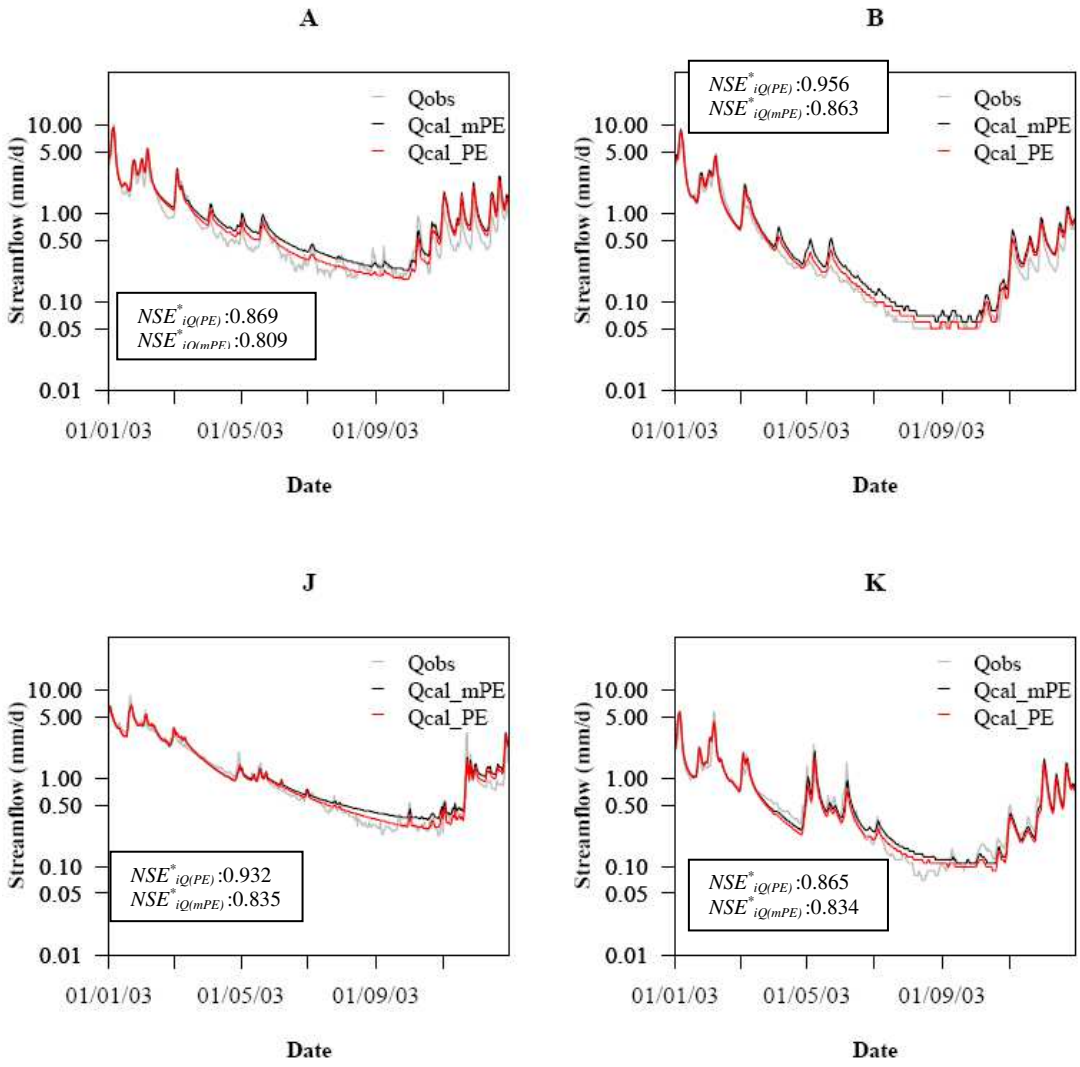


Figure F.2: Illustration of the GR6J model simulations with mean PE (mPE) and temporally varying PE (PE)

Figure F.2 shows comparatively slightly better simulations in all the four sample catchments with the temporally varying PE data (the red simulations in the plot) especially in low-flow conditions. Based on this result, we used the temporally varying PE data as input to all hydrological models which we used in this thesis.

F.4 Conclusions

Here, we analysed the sensitivity of low-flow simulations to PE data. Use of PE data instead of mean PE shows an improvement in the overall efficiency value on low-flow simulations. But the level of sensitivity depends on model structure. However, the model GR6J is more sensitive to the PE data on low-flow simulation compared to the other tested models.