An adaptive hydrological model for multiple time-steps : diagnostics and improvements based on fluxes consistency

Andrea Ficchí

THESE DE DOCTORAT DE
L'UNIVERSITÉ PIERRE ET MARIE CURIE

Ecole doctorale 398 Géosciences, ressources naturelles et environnement

Présentée par

M. Andrea Ficchí

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

An adaptive hydrological model for multiple time-steps: Diagnostics and improvements based on fluxes consistency

soutenue le 27 février 2017

devant le jury composé de :

Mme Agnès Ducharne	CNRS, Paris	Présidente
M. Marco Borga	Université de Padoue (Italie)	Rapporteur
M. Nicolas Masséi	Université de Rouen	Rapporteur
M. Lionel Berthet	DREAL Centre-Val de Loire, Orléans	Examinateur
M. Hubert Savenije	Université de Delft (Pays-Bas)	Examinateur
Mme Isabella Zin	Obs. des Sciences de l'Univers, Grenoble	Examinatrice
M. Vazken Andréassian	Irstea, Antony	Directeur de thèse
M. Charles Perrin	Irstea, Antony	Co-encadrant

“L'acqua che tocchi de' fiumi è l'ultima di quella che andò
e la prima di quella che viene.
Cosí il tempo presente.”

Leonardo da Vinci

Acknowledgments

I want to express here my gratitude to the people who contributed to make my PhD journey pleasant and successful.

First, I would like to warmly thank my two supervisors, Charles Perrin and Vazken Andréassian, for assigning me this interesting research project and for the careful guidance and expert advice during the three years spent at Irstea. Thank you for the inspiring discussions and suggestions, your teachings in hydrology and beyond, and for the trust and freedom conferred on my creativity. Your contagious enthusiasm and your positive attitude helped me to keep high motivation and optimism from the sources until the end (the sea?) of this adventure (apparently also the hydrological poems had an effect!). Also, I am grateful for your quick and effective reviewing of the thesis manuscript (fortunately, your reading flow rate was appropriate to the length of my chapters!).

I would like to thank the members of my jury for the discussions at the PhD defence, and for the interest and the positive feedback on my work. I was honoured by the two rapporteurs, Marco Borga and Nicolas Massei, for agreeing to evaluate my dissertation, and I want to thank them for the constructive and relevant remarks on my work and on the manuscript. Thanks to Agnès Ducharne, serving as Chair of the committee, and to Hubert Savenije, Isabella Zin and Lionel Berthet for sharing relevant comments and appropriate suggestions to improve the manuscript.

I would also like to thank the members of my PhD guidance committees, Nicolas Le Moine, Charles Obled, Federico Garavaglia, Pierre Javelle, Rémi Lamblin, and Flora Branger who participated to the two mid-term meetings on my research work. They shared their independent point of view and ideas on the possible directions to follow in this research and helped me to choose some good paths. Special thanks are owed to Nicolas for agreeing to represent my doctoral school (ED GRNE, UPMC) for these committees, and for the other inspiring discussions we had on other occasions. Many thanks also to Maria-Helena Ramos and Guillaume Thirel for participating in the discussions of the guidance committees, as permanent members of the team, and for the suggestions on other occasions. Thanks also to Bernard Cappelaere for his comments and exchanges by email for my mid-term committee.

This PhD research work was financed by Irstea and SCHAPI (Service central d'hydrométéorologie et d'appui à la prévision des inondations). I owe thanks to the people at SCHAPI who believed in this project and agreed to support it, and particularly to Bruno Janet, Rémi Lamblin and Léa Garandeau, who significantly contributed to this.

SCHAPI and Météo-France provided the hydro-climatic data used in this thesis and they need to be thanked for the essential steps of data collection and provision. I contacted also some French hydrological services of DREAL (regional environment agencies) and SPCs (flood forecasting services), to retrieve additional information on the streamflow data and I received many kind replies by managers and technicians. I owe my thanks to all of them.

Working in the Catchment Hydrology team at Irstea (Antony) was an unforgettable experience and it is difficult to leave such a great team! The list of colleagues I had the chance to meet in these three years is long. I would like to mention all "*les hydros*", but here I will do my best in remembering the names of all of those with whom I spent more time… Carine, Philippe, Louise, Laure, Alban, François, Florent, Pierre N., Olivier, Julie, Pierre B., Laurent, Carina, Angelica, Léonard, Cédric, Manon, Morgane, Jose, Helena, Guillaume, Charles, Vazken, David, Arnaud, Gaia, Julien, Damien, Sylvia, Audrey, Tito, Louise II, Cuan, and Jeff. My regards also go to all the nice people staying for a shorter time, as the *stagiaires*, for participating to the very nice ambiance, the sport activities, the socio-hydrological events at *Chez Mamane*, and the *NASH* celebrations for the new hydrological year. Special thanks to Alban and Julien for the nice scientific collaboration, to Olivier for sharing his multiple expertise in R, GIS, and statistics, and to the "*anciens*", particularly François and Florent, for their good example and their support in the beginning. A friendly thought to all the PhD mates for the good moments, the coffee breaks, and the mutual support. I am already missing my "*super colocs*", Philippe, Carine, Louise, and Manon (what a good company… sharing sweets and magic pears in the end!). Now I wish a successful continuation to the next PhD adventurers!

I cannot forget all who shared many nice moments at Irstea playing Ultimate Frisbee and Football! In particular, thanks to the many people from my research unit HBAN. Thanks to the "*MPG*" players for the funny moments in real and simulated games. Thanks to the administrative and informatics team at Irstea, in particular to Roger, Sylvain, Nathalie, Laurence and Elizabeth, for their kind and expert support.

Thanks to Dovy Tristani, François Baudin and Loïc Labrousse for the kind support, the opportunity of participating to the "stage de terrain", and the training sessions of the Ecole Doctorale "Géosciences, Ressources Naturelles et Environnement".

Thanks also to David Dorchies, Pierre-Olivier Malaterre, Luciano Raso, and Francesca Pianosi for the inspiring collaboration during my MSc thesis which prompted me to embark on this journey.

Thanks to my friends, "*les cocos*", and my Parisian family of "*Moulinois(es)*" who made these years so lovely and rich of good times, music, cooking, and laughs. Heartfelt thanks to Anne, Diana, John, Thomas, Justine, Ulvika, and Ankur.

Finally, last but not least, I would like to thank my family, especially my parents, and Massimo and Mari, for the continuous encouragement, love and great support over the years. Special thanks to my uncle Donato, for the support in the last days before the PhD defence. My highest thoughts go to my aunt, Maria, who passed away too soon, for her smile and positive approach to life that I will keep with me.

Thank you all! Merci beaucoup! Grazie mille!

Abstract

This thesis aims at exploring the question of temporal scaling in lumped conceptual hydrological modelling. The main objectives of the thesis are to: (i) study the effects of varying the modelling time step on the performance, parameters and structure of hydrological models; (ii) develop a hydrological model operating at different time steps, from daily to sub-hourly, through a unified, robust and coherent modelling framework at different time scales. Our starting point is the chain of conceptual rainfall-runoff models called *'GR'*, developed at Irstea, and in particular the daily *'GR4J'* lumped model. The GR4J model will be the baseline model to be effectively *downscaled* up to sub-hourly time steps following a *top-down* approach. An hourly adaptation of this model had already been proposed in previous research studies, but some questions on the optimality of the structure at sub-daily time steps were still open. This thesis builds on these previous studies on the hourly model and responds to the operational expectations of improving and adapting the model at multiple sub-daily and sub-hourly time steps, which is particularly interesting for flood forecasting applications. For our modelling tests, we built a database of *240 unregulated catchments* in metropolitan France, at multiple time steps, from 6-minute to 1 day, using fine time step hydro-climatic datasets available: (i) 6-min rain gauges and higher spatial-density daily reanalysis data for precipitation; (ii) daily temperature data for potential evapotranspiration (making assumptions on sub-daily patterns); (iii) sub-hourly variable time step streamflow data. We investigated the impact of the inputs temporal distribution on model outputs and performance in a flood simulation perspective based on 2400 selected flood events. Then our model evaluation focused on the consistency of model internal fluxes at different time steps, in order to ensure obtaining a satisfactory model performance by a coherent model functioning at multiple time steps. Our model diagnosis led us to identify and test a significant improvement of the model structure at sub-daily time steps based on the complexification of the interception component of the model. Thus, we propose a new version of the model at multiple sub-daily time steps, with the addition of an interception store without extra free parameters. Our tests also confirm the suitability at multiple time steps of a modified groundwater exchange function proposed earlier, leading to overall improved model accuracy and coherence.

Résumé

Cette thèse vise à explorer la question du changement d'échelle temporelle en modélisation hydrologique conceptuelle globale. Les principaux objectifs de la thèse sont : (i) étudier les effets du changement du pas de temps sur les performances, les paramètres et la structure des modèles hydrologiques ; (ii) mettre au point un modèle pluie-débit applicable à différents pas de temps, allant du journalier à l'infra-horaire, au travers d'un cadre de modélisation unifié, robuste et cohérent à différents pas de temps. Notre point de départ est la chaîne de modèles 'GR', développée à Irstea, et plus particulièrement le modèle global journalier GR4J. Ce modèle a été choisi comme le modèle de référence à adapter à d'autres résolutions plus fines, jusqu'à des pas de temps infra-horaires, en suivant une approche descendante. Une adaptation horaire de ce modèle avait déjà été proposée dans des travaux antérieurs, mais certaines questions sur l'optimalité de la structure à des pas de temps infra-journaliers restent ouvertes. Cette thèse s'appuie sur ces études antérieures sur le modèle horaire et est une réponse aux attentes opérationnelles d'améliorer et d'adapter le modèle à des pas de temps infra-journalier et infra-horaires, ce qui est particulièrement intéressant pour la prévision des crues. Pour nos tests de modélisation, nous avons construit une base de données de 240 bassins versants non influencés en France métropolitaine, à différents pas de temps allant de 6 minutes à 1 jour, en utilisant des jeux de données disponibles à pas de temps fin : (i) les données pluviométriques à 6 minutes et la réanalyse des lames d'eau journalières à plus haute résolution spatiale ; (ii) les données de température journalière pour le calcul de l'évapotranspiration potentielle ; (iii) les données hydrométriques à pas de temps variable. Nous avons étudié l'impact de la distribution temporelle des entrées sur les sorties et les performances du modèle en se focalisant sur la simulation de crue, sur la base de 2400 événements. Ensuite, notre évaluation du modèle a porté sur l'analyse de la cohérence des flux internes du modèle à différents pas de temps, afin d'assurer une performance satisfaisante à travers un fonctionnement du modèle cohérent. Notre diagnostic du modèle nous a permis d'identifier et de tester une amélioration de la structure du modèle à différents pas de temps infra-journaliers basée sur la complexification de la composante d'interception du modèle. Ainsi, nous proposons une nouvelle version du modèle, fonctionnant à différents pas de temps infra-journaliers, avec l'ajout d'un réservoir d'interception sans paramètre libre additionnel. Nos tests confirment aussi l'adéquation à pas de temps multiples d'une fonction d'échanges souterrains proposée précédemment, conduisant à une meilleure qualité et cohérence globale du modèle.

Contents

Introduction *1*

1 State-of-the-art of rainfall-runoff modelling and scale issues in hydrology: A critical overview *9*

1.1 Catchment hydrology **11**

1.2 Rainfall-runoff modelling **14**

1.2.1 Hydrological modelling framework 14
1.2.2 Hydrological modelling limitations and temporal averaging effects 19
1.2.3 Objectives of modelling: simulation and forecast over a wide range of time scales 21
1.2.4 Types of models: towards a scale-dependent process-based classification 22
1.2.5 Desirable properties of rainfall-runoff models 30
1.2.6 The fixed and flexible modelling paradigms and our model choice 32

1.3 The space and time scale issues in hydrology **33**

1.3.1 General concepts and typical scales of hydrological processes 33
1.3.2 The importance of spatial and temporal rainfall distribution 36
1.3.3 Catchment characteristic time scales 38

1.4 State-of-the-art on the temporal scale issue in rainfall-runoff modelling **39**

1.4.1 Relation between process, observation and modelling scale 39
1.4.2 Effects of time steps on model parameters and performance 40
1.4.3 Model structures 45

1.5 Synthesis and resulting research objectives **52**

2 Material and methods *55*

2.1 Introduction **57**

2.2 Building-up the hydro-climatic database **57**

2.2.1 Precipitation data 57
2.2.2 Streamflow data 64
2.2.3 Evapotranspiration data 66

2.3 Catchment selection criteria **67**

2.3.1 Criteria on catchment characteristics 68
2.3.2 Criteria on rainfall data 69
2.3.3 Criteria on flow data 73
2.3.4 Additional catchments recovered for particular reasons 75

2.4 Summary of the characteristics of our catchment sample **77**

2.4.1 Morphological and hydro-climatic characteristics 77
2.4.2 Hydrological regimes 82
2.4.3 Operational interest of our catchment sample 85
2.4.4 Synthesis of the catchment characteristics 86

2.5 Flood event set 88

2.6 Model and calibration-evaluation procedure 91

2.6.1 The GR4 rainfall-runoff model 91

2.6.2 Calibration and evaluation procedure 96

2.7 Synthesis 100

3 Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events 103

3.1 Introduction 109

3.1.1 Importance of short time-step data 109

3.1.2 Modelling at short time steps to evaluate at larger time steps 110

3.1.3 Scope and structure of the article 111

3.2 Testing approach and evaluation methodology 112

3.2.1 Testing approach for model evaluation at different time steps 112

3.2.2 Calibration and evaluation procedure 114

3.3 Results and discussion 115

3.3.1 Performance with time step over the whole catchment set 115

3.3.2 Behavioural catchment clusters for the evolution of model performance 121

3.3.3 Understanding the causes underlying the behavioural catchment clusters 124

3.4 Conclusions and perspectives 128

3.4.1 Summary 128

3.4.2 Limits of this study and further research needs 129

3.5 Acknowledgments 131

4 Model diagnostics at multiple time steps 133

4.1 General scheme of our model diagnostics framework 136

4.1.1 Model diagnosis scope 136

4.1.2 General remarks on the evaluation procedure 137

4.2 Hypotheses on the model inputs temporal distribution 138

4.2.1 Precipitation temporal distribution 139

4.2.2 Potential evapotranspiration temporal distribution 143

4.3 Consistency of the model internal fluxes at different time steps 146

4.3.1 Motivation 146

4.3.2 Analysis of the consistency of interception, evapotranspiration and groundwater exchanges 147

4.3.3 Implications and formal proof of the temporal inconsistency of the neutralisation function for interception 158

4.4 Consistency of model parameters and states at different time steps 161

4.4.1 Model parameters 161

4.4.2 Model states 165

4.5 Synthesis 170

5 Towards a consistent multi-time step model 173

5.1 General scheme of our model structure modifications 177

5.2 Refinement of the interception component 178

5.2.1 A new structure with a bucket-style interception store 179

5.2.2 Results of the GR4-I model with an interception store of fixed capacity 181

5.2.3 Results of the GR5-I model with interception store of free capacity 186

5.2.4 Fixing the interception store capacity by seeking the fluxes coherence at different time steps 188

5.2.5 Results of the GR4-I2 model with interception store of catchment-dependent capacity based on fluxes consistency 191

5.3 Attempts at improving model performance by modification of the exchange function 206

5.3.1 Tested variants of the exchange function 206

5.3.2 Synthesis of the results with different exchange functions 208

5.4 Attempts at taking into account rainfall intensities for improving the production function 212

5.4.1 Infiltration-excess runoff based on precipitation intensity 212

5.4.2 Precipitation intensity-based correction 215

5.5 Synthesis 217

Conclusions 219

Main contributions 221

Perspectives for further research 223

References 225

Appendices 241

A - Complements to literature review: Time stepping schemes 243

B - Complements to Chapter 2: Analysis of the quality of the 6-minute precipitation measurements 247

C - List of the 240 selected catchments 251

D - Flood event selection procedure 257

E – GOUE index 259

F - Consistency of the GR4 model parameters at different time steps 261

G - Formulation of the GR5-I model (complements to Chapter 5) 263

Introduction

Context

Hydrological models are used for several practical and essential purposes, such as flood forecasting, droughts anticipation and decision making for water planning and management. These varied objectives require hydrological estimates (river flows) at a wide range of spatial and temporal scales (Blöschl and Sivapalan, 1995). For this reason, the possibility of using a model with different spatial and temporal resolutions corresponds to an important operational demand.

Until now, the question of refining the spatial resolution of hydrological models has drawn more attention than the temporal resolution. For instance, a recent review by Melsen et al. (2016a) has shown the systematic lack of attention payed to the temporal resolution in hydrological modelling. While many model developers continuously increased the spatial resolution of hydrological models, the temporal resolution did not keep pace with this and with the *characteristic velocity* (Blöschl et al., 1995) of hydrological processes. Presently, it is frequent to find semi-distributed and distributed hydrological models with a refined spatial mesh (e.g. 100-km^2), similar to the typical space scale of hydro-meteorological processes, but with a daily time step, which is not sufficiently fine for describing many processes (Blöschl and Sivapalan, 1995). This asymmetry of attention is surprising, given the strong inter-play between space and time, a long-lasting debate in hydrology and all physical sciences. It seems urgent to rebalance this asymmetry, spending more resources (especially time) on temporal issues in hydrology!

Nowadays, most hydrological models are designed at a fixed time step and most operational systems choose the model to be used in function of their objectives and of the data sampling rates (typically daily or hourly). Given the increasing data availability from different sources (e.g. automatic rain gauges, radar, satellite) and higher sampling rates (up to 5 min), new techniques to get hydrological models able to work at different scales are needed.

In the flood forecasting context, the ability of running a model at different time steps corresponds to an actual operational demand. Flood forecasting services wish to have models able to run at different time steps, ranging from a few minutes (e.g. for flash floods anticipation) to daily (e.g. for floods warning in large river basins), for example to adapt to different data sources. Also, it is useful for a same basin to change the model time step when needed, as for example refining it in extreme conditions. For example, in France, some of the flood forecasting centres, especially those located in the Mediterranean regions, would need a model able to run at (up to) sub-hourly time steps, to predict intense fast-response events (flash floods). Improved predictions are needed to effectively alert populations and improve preparedness to these extreme events, which frequently cause fatalities (almost every year) and billions of euros in losses in the Mediterranean part of France (see, for instance, Antoine et al., 2001; Delrieu et al., 2005; Vinet, 2007; Boudou et al., 2016).

Currently, most of the French flood forecasting services routinely use an hourly rainfall-runoff model (called *GRP*) to forecast river flows (Furusho et al., 2016). This model has been

adapted to the hourly time step (Berthet, 2010), and to the forecasting mode, from the daily *GR4J* simulation model (Perrin et al., 2003), but it seems necessary to further explore the temporal *adaptability* of this model to shorter time steps (Furusho et al., 2016). It is also expected that some essential model improvements of the accuracy of flood peaks and volumes estimations could be achieved by considering the temporal variability of model inputs at shorter time scales.

The relevance of this research is evident: the *GRP* model is currently used as part of an early warning system which informs decision makers and citizens of the upcoming flood risks in France. The model outputs are used to update a real-time information system and publish flood watch maps. These maps are currently published twice a day on the web-site www.vigicrues.gouv.fr and concern 22,000 kilometers of watercourses in France (Escudier et al., 2016), covering 75% of the population living in flood zones[1].

State-of-the-art and scientific questions

The question of temporal scaling in hydrological modelling implies multiple scientific and technical questions, such as:

- the relationship between processes, observation and modelling scale, as discussed, for instance, by Blöschl and Sivapalan (1995) and Obled et al. (2009);
- the dependency of model parameters on time step (e.g. Littlewood and Croke, 2008; Wang et al., 2009; Kavetski et al., 2011);
- the dependency of model structures on time step (e.g. Atkinson et al., 2002; Mouelhi et al., 2006b) and the linked problem of the coherence of the internal model functioning at different time steps (Haddeland et al., 2006).

These issues are highly inter-connected, and their comprehensive analysis is a complex task.

An ideal link exists between the processes and observation scale, as pointed out by Blöschl and Sivapalan (1995), reminding the well-known filtering effects of sampling, formalized in the signal processing field by the Nyquist-Shannon sampling theorem (Shannon, 1949). Obled et al. (2009) highlight the implications of sampling for hydrological modelling, suggesting that a "*maximum acceptable time step*" should be defined for each model application in function of the catchment response time. This analysis should be considered an essential first step for hydrological modelling, but instead it is generally ignored in the literature.

An increasing number of papers investigate the impact of calibration data time step on model parameters. As discussed by Littlewood et al. (2011), the results of different research groups share some conclusions, while presenting also intriguing differences in the messages

[1] Source: Ministère de l'Environnement, de l'Énergie et de la Mer (http://www.developpement-durable.gouv.fr/)

delivered. The use of different models and calibration methods, on a small number of catchments, seems to be the actual limitation for seeking general conclusions. For example, some previous studies (e.g. Littlewood and Croke, 2008; Wang et al., 2009; Kavetski et al., 2011) showed varying parameters-dependencies on time step (from daily to sub-hourly) depending on which models, parameters, and catchments are considered. While some parameters (in specific models and catchments) are time-step dependent and seem to reach *stable values* as the time step decreases, other parameters reveal either a low sensitivity or unclear trends.

The dependency of model structure on time steps stems from the commonly recognized fact that the formulation of hydrological processes at one scale is not immediately transferable at the adjoining levels of scale (Klemeš, 1983). Nevertheless, in hydrological modelling, it is a common practice to use a same model structure for different time steps. The construction of a *time-step adaptive structure* has so far been neglected in the literature of conceptual and empirical hydrological models. Still, some authors showed that the structure of conceptual hydrological models requires additional complexity at shorter time steps (see, for instance, Atkinson et al., 2002; Mouelhi et al., 2006b; Kavetski et al., 2011). However, there is not yet a systematic knowledge on the *level of complexity* and on which model *components* are needed at different time steps, especially at sub-daily time scales.

There is a lack in detailed evaluations of model functioning at multiple time steps, for example for assessing the internal coherence of a model. To our knowledge, only one example of analysis of the coherence of the simulated moisture fluxes in a hydrological model run at different time steps can be found in the work by Haddeland et al. (2006) on the Variable Infiltration Capacity model. This kind of evaluation seems to be necessary for the identification of a multi-time step hydrological model.

Objectives

The main objective of this research work will be the **development of a coherent multi-time step model** that could be run over a continuous range of time steps, at least from daily to sub-hourly, by simple adaptation of its parameters and structural components, while maintaining an internal coherence of its functioning. Also, we will try to improve the performance of the rainfall-runoff models developed at different time steps with special focus on flood simulation. The results of this work should be transferred in the operational flood forecasting model (*GRP*) developed at Irstea and used in the flood forecasting centres in France.

The starting point of this work is the GR (*Génie Rural*) models chain, a set of rainfall-runoff models empirically developed since the 1980s for different fixed time steps (at Irstea, UR HBAN, France). In particular, we will start on the basis of the daily GR4J rainfall-runoff model (Perrin et al., 2003), which concentrated most of the efforts in the development process. Previous attempts were made by Mouelhi (2003) to find consistent model structures from the daily to the yearly time steps. Mouelhi (2003) searched for the optimal model structures at these larger time steps, finding some similarities in the structures identified at the

monthly and daily time step. Our work will be in the continuation of this *harmonization* of model structures across different time steps, but we will move in the other direction of the temporal axis, from daily to sub-hourly time steps. There were previous works done at Irstea to adapt the simulation model to sub-daily time steps: Mathevet (2005) and Le Moine (2008) successively proposed improved versions at the hourly time-step, partly based on explicit relationships of model parameters with time step. Their attempts made to make modifications or include additional complexity in the hourly model structure brought some improvements, but some questions on the optimality of the structure at sub-daily time steps are still open. Here we will build on these works by focusing on a wider range of time steps.

The first step will be to propose an original unified modelling protocol at different time steps to evaluate *model performance* in a consistent way as the time step changes. In this context, one research question is whether modelling at shorter time steps could improve model performance at larger time scales. This will allow understanding the benefits of refining the data and modelling time steps for model accuracy also when the interest is on water volumes simulations (and not only on actual peaks or timing of short-duration events). By working on a large catchment set, we will be able to investigate the links between the optimal model time step and catchment and flood events characteristics.

Then, the *adaptability* of the GR simulation models across different time steps will be evaluated in terms of coherence of parameters and structural components. We will consider the *temporal coherence* (of model parameters, fluxes and states) as an essential principle for building a multi-time step consistent model. This coherence could also help to better understand the interactions between representation of hydrological processes and time scale.

Structure of the thesis

This thesis consists of five main chapters. Some complements are provided in the Appendixes.

Chapter 1 introduces the context of our work, providing a review of scale issues in hydrology, with particular emphasis on modelling. In particular, we summarize several previous works that investigated the impact of time step on model performance, parameters and structure. This review outlines the current limits for the transferability of hydrological models across time steps. Also, we introduce our choice of a lumped modelling approach and we present the GR4 rainfall-runoff model that we will use and try to improve at different time steps.

Chapter 2 sets the methodological basis of our study. It presents the source data used (for precipitation, potential evapotranspiration and streamflow), and the treatments done to build a large catchment data base at multiple time steps. The characteristics of the 240 catchments and 2400 flood events selected are presented. Also, the model used is further presented as well as the calibration and evaluation framework.

Chapter 3 is an article published in *Journal of Hydrology*, where we propose our model testing approach to analyse the impact of time step on the simulation model performance. It

shows that the impact of the modelling and data time steps is significant also when performance is evaluated at larger time scales. We detected different behaviours of model performance as the time step changes, which are found to be dependent on flood characteristics. The analysis of this chapter highlights one specific limit of the current adaptation of the GR4 model at sub-daily time steps in the capacity of controlling the water balance over floods.

Chapter 4 presents a thorough diagnosis of the GR4 model at different time steps, mainly based on the analysis of the internal behaviour of the model as the time step changes, in terms of model parameters, fluxes and states. This analysis allows detecting some clear structural inadequacies of the model that emerge by analysing the temporal consistency of the internal fluxes of the model. In particular, a defect in the conceptualization of the interception component in the model is shown to impact all the water balance model components and to jeopardize model performance (on flood events) at sub-daily time steps.

Chapter 5 presents our process of model improvement at sub-daily time steps based on the performance and the internal coherence of the model. We show the need of a refinement of the production part of the model at sub-daily time steps by taking into account what can be considered the interception process. Other structural modifications are also tested, especially for the exchange function. These tests lead to validate and retain a new improved structure of the model at sub-daily time steps.

The last chapter (Conclusions) provides a global synthesis of the work, highlighting the main contributions and the implications of the results for flood forecasting applications. Finally, some ideas for future studies are suggested.

1
State-of-the-art of rainfall-runoff modelling and scale issues in hydrology: A critical overview

In order to contextualize this research work, the first part of this chapter (Sections 1.1 and 1.2) introduces rainfall-runoff modelling in the context of the hydrological science. We introduce the object of study, i.e. the catchment and its water balance, and we briefly explain what a rainfall-runoff model is. An overview of the general framework of development of hydrological models is provided, highlighting the underlying *time concepts*. We briefly illustrate some limitations of rainfall-runoff modelling, particularly with respect to temporal discretization and data-averaging effects. Then we provide a critical review of the traditional classification of existing models, focusing more on the combination of spatial and temporal scales than on model conceptualization or development. Furthermore, we discuss the properties that a rainfall-runoff model should possess, and, finally, we support our choice of a spatially lumped approach and a *fixed model structure paradigm* at different time steps.

In the second part of this chapter (Section 1.3 and 1.4), we provide a literature review on scale issues in hydrology, and particularly on temporal scaling in rainfall-runoff modelling. After presenting some general concepts on scaling and the typical scales in hydrology, we highlight the multiple inter-related aspects of this research work. The state of the art of the time scale issues in rainfall-runoff modelling is presented in details with particular emphasis devoted to:

(i) the relation between processes, observations and model scales;
(ii) the dependency of hydrological model parameters and performance on the time step;
(iii) the relation between model structures and time step.

Some other questions, as the numerical time stepping schemes and the multi-time scale assessment of modelling, are also briefly outlined along our review. Since the discussion on the time stepping schemes requires more details, but is not essential to understand and develop this research work, it will be further presented in Appendix A. At the end of this chapter, following the perspectives enlightened by our overview, the objectives of this research work will be presented and formulated in some detailed research directions.

1.1 Catchment hydrology

Hydrology is the science that studies the water cycle, i.e. the movement of masses of water in the hydrosphere, i.e. the part of the Earth system where water is (from about 15 km up into the atmosphere to about 1 km down into the lithosphere). A schematic representation of the water cycle is given in Figure 1.1, from Chow et al. (1988). We can observe that, at a global scale, hydrology focuses on the exchanges of water between the atmosphere, the land surface, the watercourses, the subsurface and the ocean. However, in order to contextualise our work, this classic definition of hydrology is too large as it encompasses all the branches and characterizations of the discipline that can focus on one or more compartments of the hydrosphere, across a wide range of spatial and temporal scales. Continental hydrology is usually further defined as the study of the continental part of the water cycle (excluding ocean and seas) from a quantitative point of view. The major physical processes considered in continental hydrology include: precipitation, interception, evaporation (from intercepted water

and soil), transpiration, infiltration, percolation, runoff (from overland, sub-surface and groundwater flow) and open-channel flow (in watercourses).

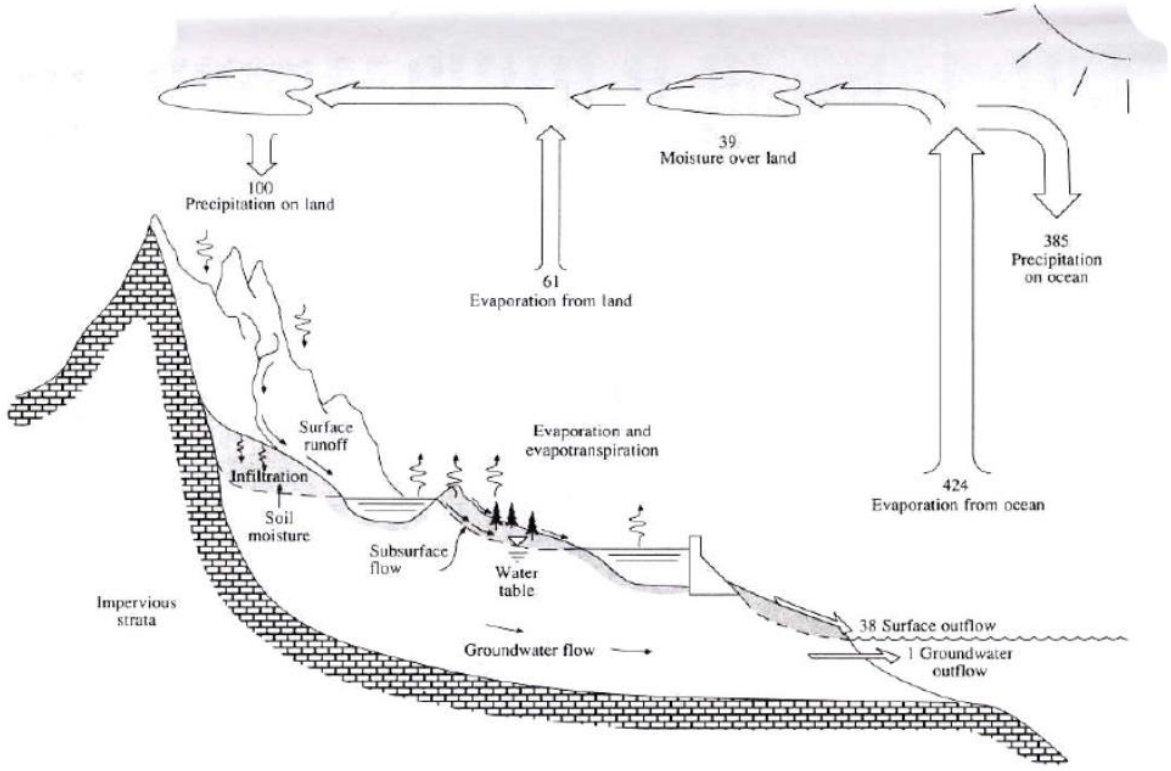


FIGURE 1.1 – The water cycle, with the main hydrologic processes, and a general indication of the annual average water fluxes given in units relative to a value of 100 for the rate of precipitation on land (from Chow et al., 1988).

In the field of continental hydrology, we can further specify the *catchment hydrology* that deals with the integration of hydrological processes at the *catchment scale* (Singh and Woolhiser, 2002). The ***catchment***, also called *drainage basin (or watershed in US English)*, is defined at any point or cross-section of a river, as the entire topographical and geological area drained by the river and its tributaries upstream of this section. It may be as small as an agricultural parcel, smaller than 1 km^2 (e.g. for a small mountain creek), or as large as hundreds of thousands or a few millions of km^2 (for the largest river basins in the world, e.g. those of Amazon, Congo, Nile, Mississippi, Ob and Paranà Rivers).

Despite the definition given above, a catchment is usually defined only on a topographical basis, by the so-called drainage divide line that lies along the topographical ridges, as represented in Figure 1.2 (from Le Moine, 2008). This allows including in the definition the entire area that contributes to the streamflow by surface runoff (also called overland flow), and by infiltration in those aquifers that are drained by the river. This area contributes to the streamflow also with flow components from subsurface or groundwater. However, the area defined by the drainage divide line is not always perfectly superposing with the whole geological contributing area that could be identified following the sub-surface modes of supply of the watercourse streamflow, as discussed by Le Moine (2008). This difference can

be taken into account by enlarging/reducing the system, considering the hydrogeological basin, or by considering the topographical catchment as an open system exchanging with the groundwater body, in the case where it exists, so keeping a general approach and a unique definition of catchment. The latter way is the choice followed in the present research work.

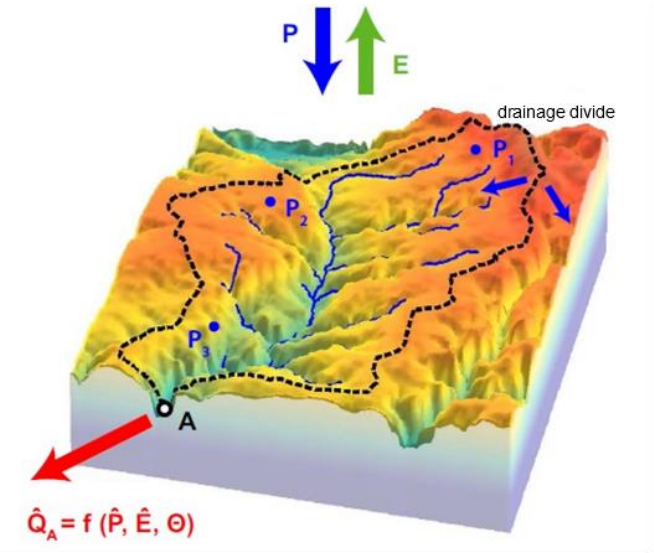


FIGURE 1.2 – Definition of the catchment (from Le Moine, 2008) at the outlet A. The streamflow Q at the river section in A is "*forced*" by the areal precipitation P, and the evapotranspiration E, occurring during a time interval (day, hour, etc.). The main forcing variable is the areal precipitation P (ratio between total volume of rainfall fallen over the catchment in a time interval and basin surface) that is estimated by interpolating some point measurements P_1, P_2, …, P_n and hence is subject to sampling and measurement uncertainties. The catchment response, i.e. the functional relationship *f( )*, depends on the catchment characteristics and conditions, i.e. the parameters Θ.

Like in other environmental sciences, the object of catchment hydrology is a complex system with highly interrelated, spatially distributed and time varying phenomena involved. To deal with this complexity, there is a variety of approaches that can be summarized in two general complementary strategies:

(1) the **reductionist approach** (also called *mechanistic*): it is based on the reduction of the system studied into a number of interconnected elementary units of smaller spatial dimensions. In catchment hydrology this corresponds to the a-priori prescription of splitting a catchment area into sufficiently small grid elements where some elementary physical laws can be applied.

(2) the **lumped approach** (also called *systemic* or *holistic*): it analyses the system at the catchment scale. Given this large spatial scale, the functional relationship describing the rainfall-runoff transformation must be defined with some *conceptual* or *empirical* relationships, still with the possibility of representing the hydrological processes integrated in space. This second approach is the one followed in our work.

1.2 Rainfall-runoff modelling

Rainfall-runoff (RR) models are mathematical models that aim at transforming the precipitation fallen in a certain time interval on a certain area, here the catchment, into the streamflow at the outlet during the same time interval. As input forcing of the system, in addition to rainfall, other variables can be used to reproduce the output flows, such as temperature or evapotranspiration. A *mathematical model*, as defined by Eykhoff (1974), is a representation of the essential aspects of an existing system (or one to be constructed) which presents knowledge of that system in usable form. This is a good definition also for rainfall-runoff models which describe the behaviour of hydrological systems in a simplified form that can be used for practical purposes by generating streamflow time series.

The 'utilitarian' view of hydrological models has clearly driven their historical development that "*has been largely pioneered by practitioners who needed solutions to real problems*" as noted by Linsley (1982). From the 19th century and the beginning of the 20th, hydrologists (often engineers) proposed a variety of empirical formulae, starting from the so-called "rational formula" by Mulvaney who proposed the first known hydrological model in 1850. This simple first model, linking the peak flow to rainfall intensity on a catchment, has been largely used for a variety of engineering tasks, as sewer design. Only in the second part of the 20th century, models have been developed for describing the mechanisms of flow in greater details, as for hydrologic research on the flow processes (as cited by Linsley, 1982).

Today, many different hydrological models exist (maybe a few hundred or more) with a wide variety of levels of sophistication. Many of them are currently used and continue to be improved for research and operational purposes, in a long-term development process. A review of the existing hydrological models and their historical advances goes beyond our scope. For this, one can refer for example to Todini (1988) and Singh and Woolhiser (2002).

1.2.1 Hydrological modelling framework

The development of hydrological models has traditionally followed a framework methodology involving the following steps, as highlighted by Blöschl and Sivapalan (1995) and frequently recalled in the literature like in Refsgaard and Henriksen (2004):

(a) collecting and analysing data;

(b) "*developing a conceptual model (in the researcher's mind) which describes the important hydrological characteristics of a catchment*"; for a sake of clarity, we prefer calling it a '*meta-model'* or '*perceptual model*' (since the adjective 'conceptual' is often used for indicating a type of model structures, as reported in Section 1.2.4);

(c) translating the meta-model into a mathematical model, generally using difference or differential equations (solved analytically or by numerical schemes);

(d) "*calibrating* the mathematical model to fit a part of the historical data by adjusting various coefficients", i.e. determining the values of the model free parameters by fitting the model outputs with the observed time series;

(e) "*validating* the model against the remaining historical data set".

Variants and reiterations of this general scheme are sometimes followed. For example, the development of the so called '*empirical*' models can rely on a trial-and-error modelling approach as acknowledged by some authors (e.g. Bergström, 1991; Perrin et al., 2003).

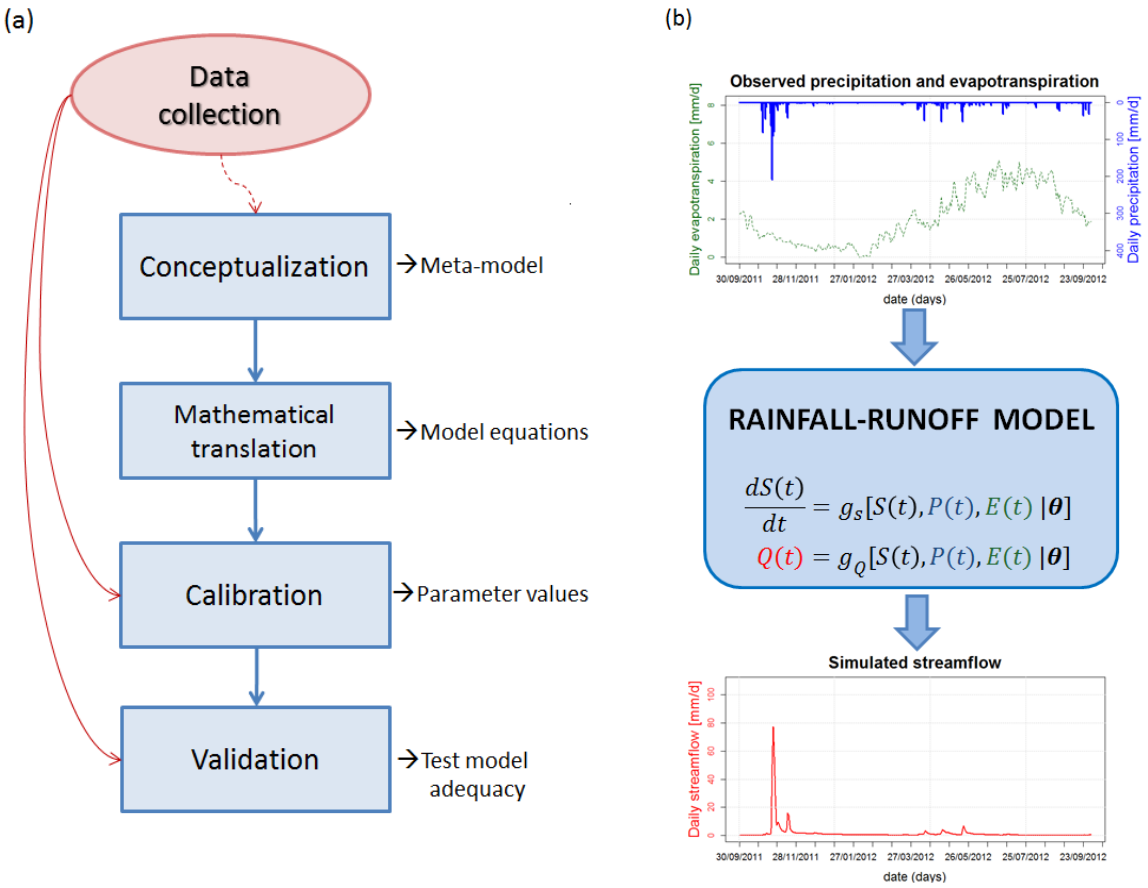


FIGURE 1.3 – (a) The rainfall-runoff modelling process with the phases of model development and their interplay. Note that the conceptualization phase has a dashed connection with the upstream phase of data collection since the degree of this interplay depends on the type of model. (b) Scheme of functioning of a spatially lumped rainfall-runoff model with (from top to bottom): (i) input trajectories of catchment daily rainfall (observed data from the French catchment of the Hérault river at Laroque, 916 km2) and estimated potential evapotranspiration (calculated using temperature data), (ii) essential mathematical representation of a lumped hydrological model (in the equations $\boldsymbol{g}(\cdot)$ are input-output fluxes, S are the model state variables and $\boldsymbol{\theta}$ are the model parameters), and (iii) output trajectory of simulated streamflow at daily time step (by using the GR4J model, see Perrin et al., 2003).

Figure 1.3(a) gives a schematic representation of the rainfall-runoff modelling process, with the phases of model development. Figure 1.3(b) shows a basic representation of the transformation operated by a rainfall-runoff model on the main variables, i.e. *precipitation*

(P), potential evapotranspiration (E), and streamflow (Q). Note that Figure 1.3(b) is more representative for the *lumped (systemic)* approach than the reductionist one, because for the latter several input trajectories should be used (one for each grid element of the spatial discretization) and the fluxes equations should be based on partial differential equations.

1.2.1.1 Time concepts in modelling and linkages across scales

Along the different phases of the modelling process, the choice of a spatial and temporal discretization must be made several times. This choice should take into account the dominant hydrological processes (those considered important for a specified application), the data available, the model developed/used and the objectives of modelling. Some *linkages across scales* may be necessary along the modelling process and they should be explicit. For example, for the temporal domain, we agree with Melsen et al. (2016a) in identifying several different "*time concepts which in practice are often mixed up and misinterpreted*". We distinguish ***eight different time concepts***, extending the six proposed by Melsen et al. (2016a):

(i) the process time scale;
(ii) the source data time interval (spacing between data samples, also called sampling interval);
(iii) the model input resolution;
(iv) the numerical resolution (model time step);
(v) the model output resolution;
(vi) the calibration time interval;
(vii) the validation time interval;
(viii) the interpretation time interval.

In these concepts, as stated by Melsen et al. (2016a), one may note the distinction between the term *'scale'*, that is a *continuous* variable, while *'resolution'* and *'time interval'* indicate *discrete* variables. The resolution refers to a model property, while the time interval is independent of the used model. In the following of this work we will use the term '*scale*' in a more general and common sense and also the term '*time step*' will refer to model resolution and data time intervals. These different concepts are summarized in Figure 1.4(a), while examples of linkages across these time concepts are given in Figure 1.4(b) and 1.4(c). The first causal link between process time scale and source data time interval can be considered as ideal, because observing processes at the scale they occur "*is not always feasible*" (Blöschl and Sivapalan, 1995). However, a deterministic link exists between the time scale of processes and the data resolution needed for their perfect reconstruction. This link is explained by the well-known fundamental criterion of signal processing and sampling, the *Nyquist-Shannon sampling theorem* (Shannon, 1949), also called the *Nyquist criterion*. For further details on the transposition of the *Nyquist criterion* to hydrology, we refer the reader to Section 1.4.1, while here we remind it in the two examples in Figure 1.4. First (Figure 1.4(b)), a thunderstorm event with steady spatial location, short time duration (e.g. < 30 min) and high temporal variability (approximated highest frequency of 10 min^{-1}) should be observed at short source data time intervals (< 5 min) to detect the highest frequency and intensity of the

rainfall peaks. This may be important also given that such intense events are likely to generate an infiltration-excess response, depending on rainfall intensity, with characteristic time scales smaller than half an hour (see also Section 1.3). Conversely, a steady frontal rain event is characterized by longer duration (> 1 day) and much less temporal variability, and it is likely to activate slower mechanisms of runoff response as saturation-excess (Figure 1.4(c)). In this case, daily rainfall measurements may be sufficient to properly observe the event and model the catchment response. Note that for non-stationary spatial phenomena, as rainfall showers (i.e. intermittent events, both in time, space or intensity) one should design the necessary source data time interval by analyzing not only their temporal variability but also their *characteristic velocity* in space (see Section 1.2.4.3).

Following the review on scale issues by Blöschl and Sivapalan (1995), Figure 1.4(a) suggests that the *linkages across scales can be performed either in a deterministic or stochastic way*. A deterministic approach for scaling can be followed when information on the actual spatial or temporal pattern is available; otherwise a stochastic approach is often chosen by means of distribution functions. Moreover, linkages across scales can go in both directions through the hierarchies of scales. The term ***upscaling*** denotes transferring information from a given scale to a larger scale (e.g. from 5 min to 1 hour, Figure 1.4(b)), while ***downscaling*** refers to transferring information to a smaller scale (e.g. from daily to hourly, Figure 1.4(c)). For scaling from source data to input data, upscaling is usually trivial and based on simple aggregation (Blöschl and Sivapalan, 1995). Conversely, downscaling involves disaggregating and singling out from average values (source data at larger scale) to "point" values (model input data at smaller scale). In this case, "*the stochastic approach has a particular appeal as the detailed pattern is rarely known and distribution functions can be derived more readily*" (Blöschl and Sivapalan, 1995). For further details and examples of linkages across scales from a modelling perspective we remind to the review by Blöschl and Sivapalan (1995).

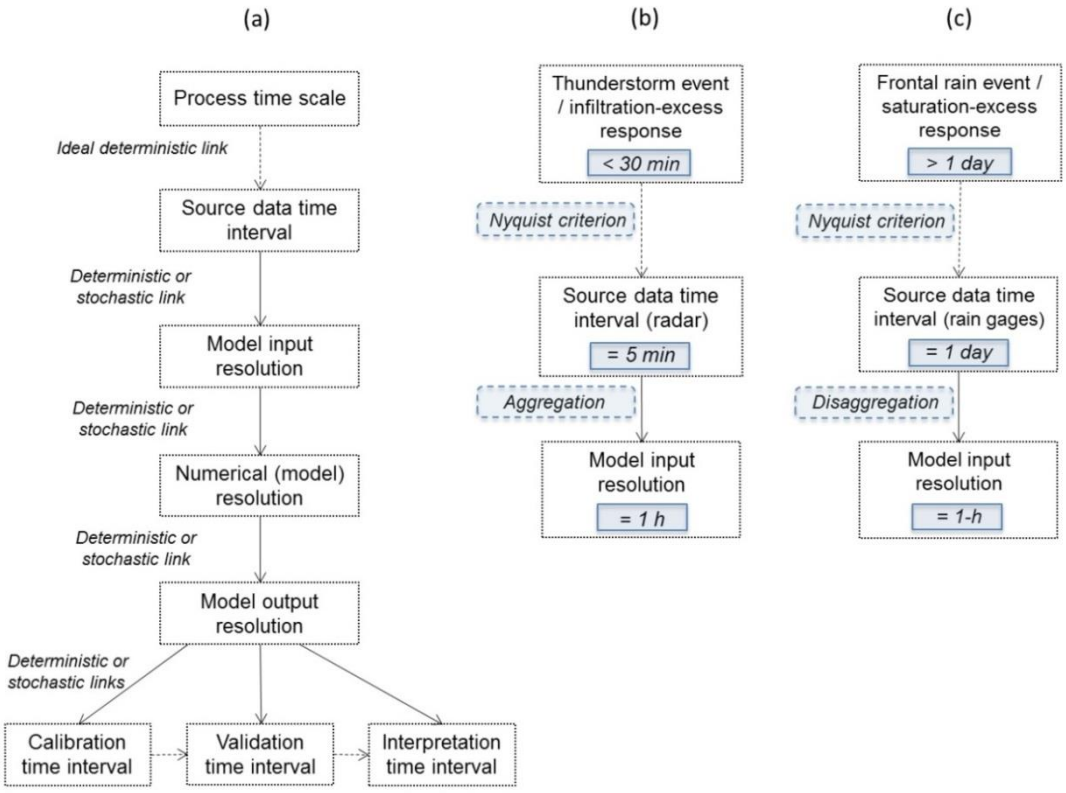


FIGURE 1.4 – (a) Scheme of the eight time concepts of the modelling process. (b) and (c) Examples of linkages across process time scale, source data time interval and model input resolution.

1.2.1.2 Time issues in the calibration-validation procedure

We think that the commonly used validation framework, in which the validation time interval is considered always as equal as the calibration time interval, should be extended to try *corroborating* models also on the basis of their *'scaling' capability*. For example, one may calibrate a model at hourly time intervals while validating at daily time intervals after aggregation of model outputs. This would extend the methods discussed by Klemeš (1986b) who proposed a systematic framework to validate hydrological models (i.e. test their "*operational adequacy*"). This framework includes the commonly used '*split-sample test'* and four other levels of test: *simple and differential split-sample tests* and *simple and differential proxy-catchment tests*. Andréassian et al. (2009) discussed the use of this validation framework, highlighting its utility, although they observe that only the first level, i.e. the *split-sample test*, is in standard use in evaluating hydrological models today. They suggest that a possible reason for the limited use of the full Klemeš testing scheme could be that it is excessively demanding (see Le Moine, 2008). However, we agree with Andréassian et al. (2009) that modellers should not fear demanding validation tests for at least three reasons:

(i) finding model limits by advanced tests helps finding ways for model improvement;

(ii) as some authors suggested, models can only be *corroborated* or *refuted* (Popper, 1959) or just *evaluated in relative terms* (e.g. Oreskes et al., 1994);

(iii) analysing model failures will make it possible to define its real limits of application and lead to a safer use of models, as *crash-tests* for automobiles (Andréassian et al., 2009), typically for applications on catchments with changing conditions (see Thirel et al., 2015).

1.2.2 Hydrological modelling limitations and temporal averaging effects

When the system of interest is complex, as it is the case of hydrological systems, the conceptualization phase needs to start from a *simplification* of the real system. This simplification is inevitable and justified because the modelling process is affected at different levels by some *limitations* that concern for example:

- the scientific understanding of complex phenomena (as the groundwater processes);
- the measurement techniques, that have their measurement errors and their spatial and temporal sampling scale;
- the spatial and temporal model resolutions that smooth out an excessive detail in description of physical laws (Morel-Seytoux, 1988);
- the computers calculation capabilities that limit the choice of a too fine spatial and temporal discretization.

Here we want to focus on the impact of the *temporal averaging of data* that is a key constraint of rainfall-runoff modelling addressed by this research work. As represented in Figure 1.3, the source data, usually collected at a fixed discrete time interval, are used along the different phases of model development. In the conceptualization phase, data are used to different extent according to the type of model, e.g. they should not be used for *fully-distributed physically-based* models (at least in principle), but they are generally used for *empirical* models (see *Section 1.2.4*). Rainfall can be measured continuously (e.g. by weighing-type rain gauges) or more generally at variable time steps (e.g. by tipping bucket rain gauges). Then rainfall data are usually aggregated over a certain time interval of accumulation (usually daily, hourly or sometimes some minutes). Similarly streamflow is usually first sampled at variable time steps and then averaged over fixed time steps. Evapotranspiration is usually calculated from averaged temperature data. For this reason, all models are affected by data-averaging and -sampling time scale effects, which *smooth* or *smear* the processes that rainfall-runoff models must reproduce, as it can be seen in Figure 1.5.

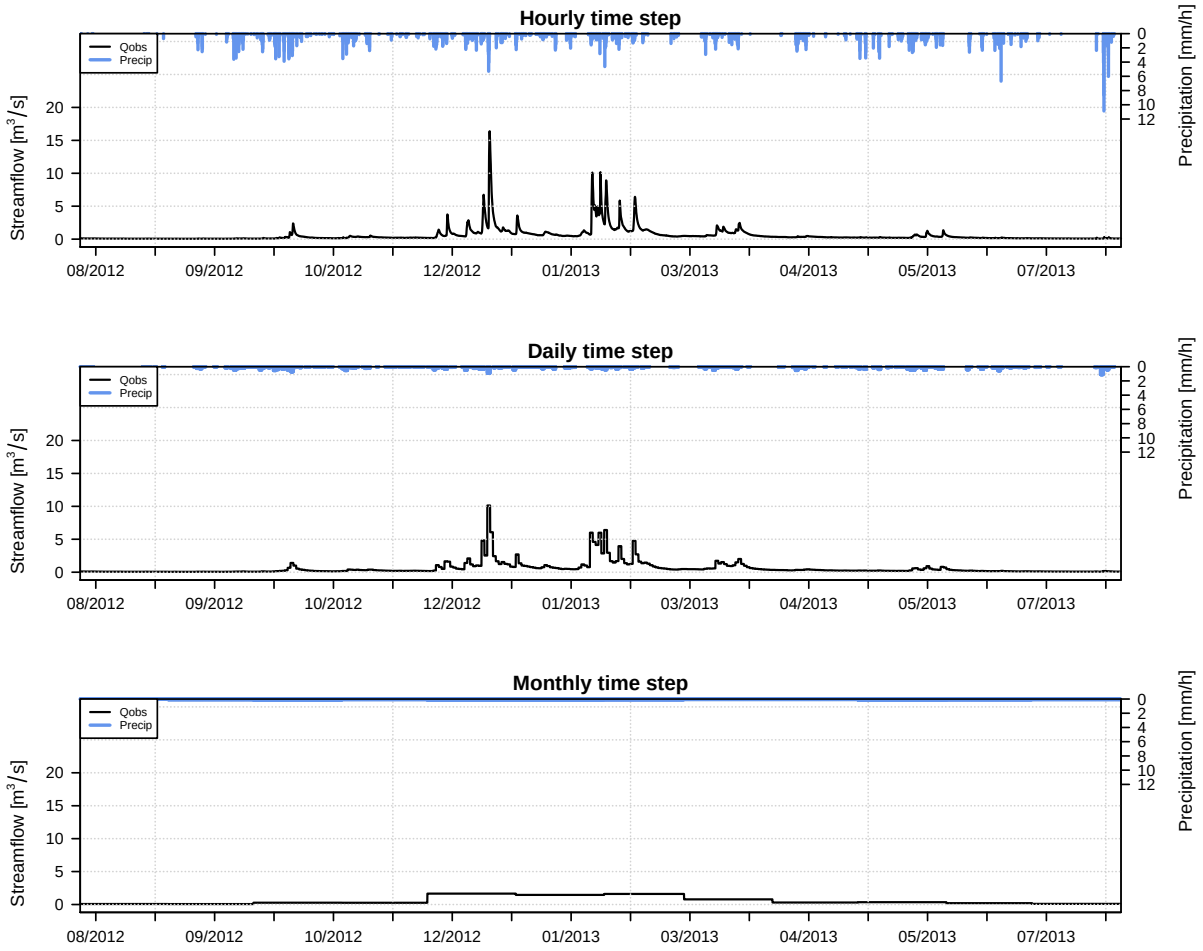


FIGURE 1.5 – Data-averaging effects on streamflow and precipitation data: example for the French experimental catchment of the Orgeval River at Boissy-le-Châtel (105 km²) studied by Irstea. The source data at sub-hourly time interval (variable sub-hourly step for streamflow and 6-min step for precipitation) are averaged at hourly (top panel), daily (middle) and monthly time intervals (bottom).

Figure 1.5 shows the progressive information loss about temporal dynamics and extreme values of precipitation and streamflows 1-year series as the time step increases. For the catchment considered here, located in the Parisian basin under temperate oceanic climate, it seems that there is less difference in the streamflow dynamics between daily and hourly time steps, rather than between daily and monthly time steps, although the time steps ratio is almost the same in the two cases. The ratios of the maximum streamflow peaks averaged at different time steps appear to be of about 62% and 17% respectively for daily/hourly and monthly/daily time steps. For precipitation data the ratios between maximum values over the year are similar for daily/hourly and monthly/daily scaling (note that the peak of monthly average is 0.15 mm/h). However, this observation cannot be generalized since the temporal variability of streamflows is obviously catchment-dependent. The difference in extreme values of streamflow at daily and hourly time step should be larger for smaller catchments and for regions subject to higher rainfall intensities. For example, Ostrowski et al. (2010) analysed the relationship between the time step and the extreme values of rainfall and

streamflow on two Austrian catchments of about 20 km^2 and showed ratios of about 33% for average maximum flow at daily and hourly time steps (see Figure 9 in Ostrowski et al., 2010).

As discussed by Kavetski et al. (2011), the "*act of averaging the rainfall forcing over a time step* Δt *introduces 'smearing' errors into the forcing data, which necessarily translates into* Δt*-dependent errors in the model predictions and calibrated parameters*". The '*smearing effects*' affect also the observed runoff data that is to be fitted in the calibration and validation procedure. These effects will strongly impact the identification of fast and non-linear processes and the calibration of their associated model parameters. Moreover, they showed that these effects have interplay with the effects of numerical schemes used to translate continuous differential equations into discrete-time equations (see Appendix A).

1.2.3 Objectives of modelling: simulation and forecast over a wide range of time scales

The knowledge and anticipation of the spatial and temporal distribution of the water in a catchment is essential for various human activities. Hydrological models are routinely used as a basis to find solutions for different problems like floods and droughts, erosion and sediment transports, water pollution, or to provide information for the management of water resources. For all these purposes, rainfall-runoff models are required to produce hydrological estimates at a range of ***time coverages (extents)***, typically from hours to hundreds of years, and with a wide range of ***time resolutions (time steps)***, from minutes to years. Note the distinction between these two concepts: the *temporal horizon* covered by the model outputs *(coverage or extent)* and the spacing between data in the time series *(resolution, time step or time interval).* To indicate these two time concepts the unique term '*time scale'* is generally used in the literature (e.g. Blöschl and Sivapalan, 1995).

Figure 1.6 shows some examples of application of hydrological models and their typically required *time coverages* (horizontal axis in the Figure) and *model time steps* (vertical axis). This scheme has been adapted from Blöschl and Sivapalan (1995) (cf. their Figure 1) where the model time steps were not represented and focus was only on the time coverage.

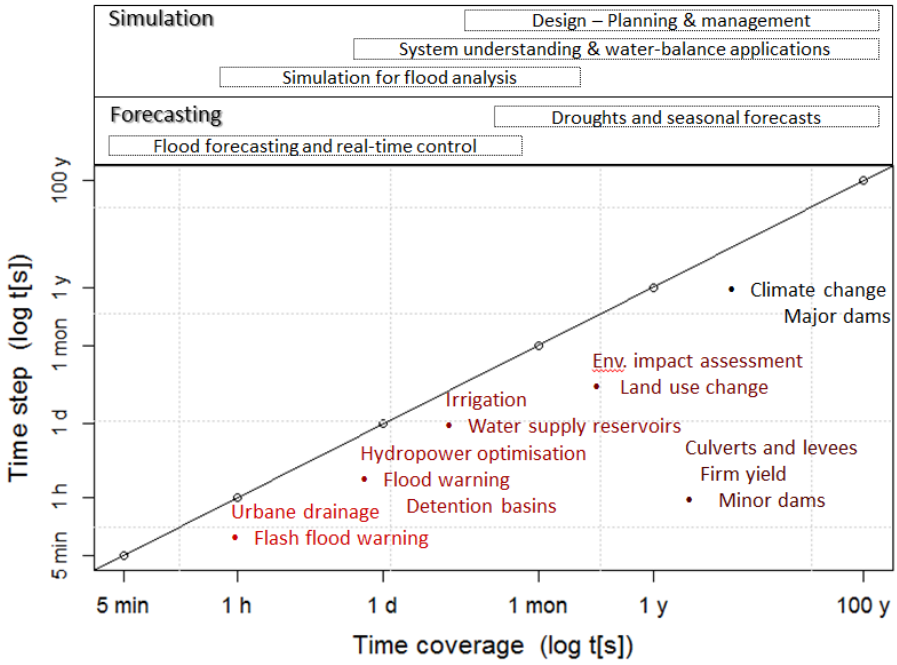


FIGURE 1.6 – Typical time coverage and time step of hydrological models required for different purposes (original figure inspired from Blöschl and Sivapalan (1995)). Model applications are classified in simulation or forecasting modes and in distinct categories of problems. Typical time coverages range from hourly to 100-years periods, while typical corresponding time steps range from 5-minutes to 1 year.

To meet these different objectives, a rainfall-runoff model can be used in two different modes:

(1) In the *simulation mode*, given the initial state of the system and the trajectory of the input variables over a period T_s, called the *simulation horizon*, the model simulates the trajectory of states and output variables over the same period T_s. In simulation, models are typically run over *long horizons*, i.e. from one month to several years, and with a range of time steps going from hourly to yearly intervals.

(2) In the *forecasting mode*, given the records of input, state and output variables up to the instant of forecast, t_f (when the forecast is issued), the model predicts the states and output values for a time horizon h following t_f, where h is called *lead time*. In forecasting, the model is generally run over *shorter periods (lead times)*, i.e. from one hour to some months (or years), and with a range of time steps going *from sub-hourly to monthly intervals*.

1.2.4 Types of models: towards a scale-dependent process-based classification

In order to respond to the various objectives presented above, different types of hydrological models have been developed in the last century. The choice of a particular type of model is

dependent on the objective of modelling, on data availability and on the modeller's understanding and conceptualization of the system. A classification of the types of models available is needed, for better orientation in the wide variety of hydrological modelling options, and to specify our modelling approach. At least ***five significant levels of classifications*** of models may be distinguished, according to the different hypotheses made by the modeller on the spatial and temporal scales (both resolution and coverage) and the processes conceptualization. These five levels have been already recognized by many authors but often the focus is put on each level individually. Here we propose an original comprehensive synthesis of this classification. Moreover we criticize the traditional classification based on processes conceptualization, going towards a new scale-dependent classification that gives more importance to the assumptions on spatial and temporal scales, especially the temporal one. The latter has long been disregarded in the hydrological and land surface modelling literature, where more importance has been given to the spatial dimension, as already noted by some authors (e.g. Archfield et al., 2015; Melsen et al., 2016a).

1.2.4.1 A four-level scale-dependent classification

1) A first level of model classification can be done on the basis of the *spatial resolution*, i.e. the assumption to represent the continuous spatial domain of the real system into the discrete one of the model. In this way, models can be denoted as:

   - ***Spatially lumped***, treating the catchment as a single unit, by a systemic approach, that implicitly accounts for only the mean variability;
   - ***Semi-distributed***, subdividing the catchment into sub-catchments, each one treated as a whole single unit (systemic approach);
   - ***Distributed***, subdividing the catchment into grid elements (i.e. *reductionist* approach); the degree of refinement of the discretization may be very different within this class. We would further distinguish the ***fully-distributed*** models, for which grid elements are so small as to be consistent with the point-scale measurements (see Section 1.2.4.3).

2) A second level of classification relies on the *spatial coverage (or extent).* This is less common in the literature, though different model denominations according to coverage can be found as:

   - ***Macroscale*** or ***large-scale,*** for applications of modelling either on large basins (as hundreds of thousands of km^2) or regional (continental) scale or global scale (Kauffeldt et al., 2016);
   - ***Mesoscale*** *(upper and lower)* for applications on small catchments or medium-size river basins, from 1 km^2 to 100'000 km^2 (Schultz, 1994);
   - ***Microscale*** or ***plot-scale***, for spatial coverages smaller than 1 km^2 (Schultz, 1994).

3) A classification is based on the model time step, i.e. the numerical *time resolution* used to represent the continuous temporal domain of the system into the discrete one of the

model. Thus models are denoted as ***inter-annual***, ***annual***, ***seasonal***, ***monthly***, ***weekly***, ***daily***, ***hourly*** or ***sub-hourly*** *models.* These mentioned resolutions are the most used, but also any other time resolution can be used.

4) As for the *temporal coverage* (and the initialization mode), one can find:

- ***Continuous time models***, considering a long '*continuous'* temporal span (generally at least one year); they are suitable for simulations on *long time coverages* thus allowing to by-pass the initial condition determination problem thanks to a *warm-up* period (see Kitanidis and Bras, 1980); these models may use a wide range of time resolutions.

- ***Event-based models*** that are used for single storm events and require the *initial condition to be specified* by the modeller, assuming that it can be guessed easily from a short past period (less than one year); in event-based models, only short temporal resolutions are used.

1.2.4.2 The traditional conceptualization-based classification

According to the assumptions made by the modeller in the processes conceptualization phase, a commonly used model classification consists of three classes (see for example Wheater et al., 1993):

(i) ***Purely empirical*** or ***black-box models***, also called *metric* or *data-based models*: they are based on inference from the data used in the modelling process and limited to providing accurate output predictions. The model parameters do not correspond to measurable physical variables, so they are usually derived by automatic calibration. Some of the most popular examples are: the linear ARMAX models (Auto-Regressive Moving Average with eXogenous inputs) initially developed by Box and Jenkins (1976); the Artificial Neural Networks (Dawson and Wilby, 2001); Transfer Functions (e.g. Young, 2011); and Data-Based Mechanistic (DBM) models (e.g. Young and Beven, 1994).

(ii) ***Conceptual or bucket-style models***, also called ***soil-moisture accounting*** models: they attempt at describing the main hydrological processes occurring in the system following a *systemic* approach. These models are often written in state-space form where state variables *S(t)* are used to describe the evolution of the catchment system conditions, by a set of ordinary differential equations as:

$$\frac{dS(t)}{dt} = g_s[S(t), P(t), E(t)\,|\theta] \tag{1.1}$$

$$Q(t) = g_Q[S(t), P(t), E(t)\,|\theta] \tag{1.2}$$

where g_s and g_Q are functions describing the internal fluxes and input-output relationships and are derived by a simplified representation of the hydrological processes using components as *reservoirs* (e.g. for soil moisture accounting) and *unit hydrographs* (for routing); $P(t)$ and $E(t)$ are the time-dependent input forcing,

respectively precipitation and evapotranspiration; $Q(t)$ is the system output, i.e. the streamflow; θ are the parameters, which are not directly measurable and are generally derived by calibration (automatic or manual). The discrete equations of conceptual models are usually derived by integration of the differential equations (Eqs. (1.1) and (1.2)) or sometimes directly formulated by difference equations. The integration may be analytical or approximated numerically, and in the latter case numerical artefacts may be introduced in the model (see Appendix A). Some examples of popular conceptual models are: Sacramento (SAC-SMA, Burnash et al., 1973), TOPMODEL (Beven and Kirkby, 1979), Xinanjiang (Zhao et al., 1980), PDM (Moore and Clarke, 1981), ARNO (Todini, 1996), SMAR (Tan and Oconnor, 1996), HBV (Lindstrom et al., 1997), TOPKAPI (Ciarapica and Todini, 2002), and GR4J (Perrin et al., 2003). Despite the number of conceptual models, many of them share the same concepts and are very similar (Moore et al., 2005).

(iii) ***Physically-based models***: they attempt at describing all the processes occurring in the catchments by applying physical laws at a *distributed* scale (*reductionist* approach). These models are based on partial differential equations, as the Saint-Venant equations for river flows and the Boussinesq or Richards equations for flows in saturated and unsaturated soils. These equations are applied on a fine discretization grid to take into account the spatial variability of the basin. Each grid cell is characterized by a set of parameters having a physical principle and theoretically derivable by field measurements (Beven, 1989). However, many difficulties arise in the measurement of parameters, because of the technical and economic difficulties of measuring such a number of parameters at the required small scale. So the parameters cannot be all derived by measurements and calibration is still required. The discrete time equations of these models derive from the integration of partial differential equations, so they are often impacted by the effects of numerical schemes (see Appendix A). The time step of model functioning is generally fine, whatever the data time step is, because numerical errors are larger as the model time step increases. The frequent use of larger resolutions of source data limits the appropriate physical representation of the hydrological processes (see Section 1.3.2). A good way of scaling up in time a physically-based model does not exist, thus aggregation of models results is necessary (Singh and Frevert (2005), page 524). Some examples of physically based models are: SHE (Abbott et al., 1986), IHDM (Beven et al., 1987) and SWAT (Arnold et al., 1998).

Table 1.1 summarizes the characteristics of the model classes of this traditional classification, highlighting the main differences, with particular emphasis to the time scale issue.

Modelling approach	**Mathematical form**	**Spatial resolution**	**Temporal coverage**	**Time-steps (numerical and data resolution)**	**Effects of data-averaging in…**
Purely empirical (Black-box)	Input-output form (0 state variables)	Lumped	Continuous and single-event models	From sub-hourly to annual and multi-annual	Model forcing; Conceptualization; Calibration and validation.
Conceptual and hybrid empirical-conceptual	*State-space form (dozen state variables)*	*Lumped or semi-distributed* or distributed	*Continuous* and single-event models	*From sub-hourly to annual*	*Model forcing; Conceptualization; Calibration and validation.*
Physically based	State-space form (hundreds or thousands of variables)	Distributed and fully distributed	Continuous and single-event models	From sub-hourly to daily	Model forcing; Calibration and validation.

TABLE 1.1 – Summary of the characteristics of the models classes of the traditional classification based on the processes conceptualization. The classes in *italic* font are the ones considered in this work.

Limits of the traditional conceptualization-based classification

In the last three decades, more and more authors rejected this conceptualization-based classification, especially questioning the realism of the physically-based concept (e.g. Grayson et al., 1992; Beven, 2002; Kirchner, 2006; Beven and Cloke, 2012; Montanari and Koutsoyiannis, 2012), but also the overlapping and uncertain contours between the three classes (Andréassian, 2005). The classes of empirical and conceptual models may overlap leading to combinations of conceptual and empirical models called *hybrid metric-conceptual* (or *empirical-conceptual*) models (Wheater et al., 1993). In our opinion this is more frequent than it is recognized, because observations are generally used to corroborate a hypothesized conceptual model structure and improve it by iterative experiments in an empirical paradigm (a sort of trial-and-error procedure). A classic example of this hybrid type of model is the IHACRES model, acronym of Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow data (Ye et al., 1997). Another emblematic case is the GR models chain developed at Irstea (see also Section 1.4.3.2 for details). At the daily and monthly time steps, the GR models GR4J (Perrin et al., 2003) and GR2M (Mouelhi et al., 2006b) are hybrid models, as their building elements belong to the conceptual models class (e.g. soil moisture reservoir, unit hydrographs) but were selected and assembled by using observations. At the annual and inter-annual time steps, the GR model is a simple input-output equation with a basic memory function (Mouelhi et al., 2006a), which may be classified more easily as purely empirical.

Also the classes of physically-based and conceptual models may often overlap (Andréassian, 2005). This is due to the difficulty of pursuing a fully-distributed representation of physical processes in space and time, as already mentioned. Without a fully-distributed approach a so-called *physically based* model is reduced to an excessively detailed and sophisticated *conceptual* model, with a semblance of physics (see also next Section 1.2.4.3). Moreover, all

the types of models are impacted by the effects of data temporal averaging, as discussed by Kavetski et al. (2011): the *time scale effects* of data-averaging and sparse sampling are *inherent to any model forced with* (and/or calibrated to) *averaged data*, whether the model is based on conceptual or physical principles (e.g. Clark et al., 2008) or transfer functions (e.g. Young and Garnier, 2006). These effects undermine the physical basis of any model.

1.2.4.3 The need of a scale-dependent process-based classification

In our view, the main limitation of the long-established classification based on processes conceptualization is that it is assumed to be *scale independent* and in general the three classes (*empirical*, *conceptual*, and *physically-based*) overlap for the commonly used temporal and spatial scales. In particular, the concept of physically-based model is not realistic at all scales, because at coarser spatial and temporal resolutions it overlaps with the *conceptual* model category (Andréassian, 2005). Recently some authors stated that there are no purely physically-based models for mesoscale or large-scale catchments (see, for instance, Beven and Cloke, 2012; Montanari and Koutsoyiannis, 2012). Nowadays, it is not feasible to model these large systems at resolutions consistent with the point measurements, which would be necessary to apply small-scale physics. This limitation is due to the lack of data for these large extents, particularly at fine spatio-temporal resolutions, and the inadequacy of computational capabilities for high-resolution, long-term continuous simulations at large spatial coverages (e.g. Singh and Woolhiser, 2002). Today, the concept of physically-based models could be *'realistic'* only for micro-scale spatial coverages (< 1 km^2), where a combination of fine spatio-temporal modelling scales could be possible (i.e. spatial resolution < 1 m and temporal resolution < 5 min).

Outside this set of fine spatio-temporal scales, models can be defined more accurately as '***process-based***', a term that is gaining increasing popularity in the literature (e.g. Montanari and Koutsoyiannis, 2012). As a matter of fact, hydrological processes may be represented by deterministic or stochastic equations even at large spatial and temporal resolutions, depending on the inherent *velocity* and *frequency* of the processes considered. This is easy to understand for slow processes as groundwater and base flow. Also, the integration of a process over space could be straightforward in cases where the inputs and catchment spatial variability is small. In these cases, it is likely that the distribution over time of inputs and catchment response could be more important than subdividing the catchment into fine spatial grid elements. Along similar lines, Singh and Woolhiser (2002) highlighted the great dependency of spatial variability on a range of scales and on location-dependent properties, by resuming the field examples examined by Seyfried and Wilcox (1995), as for example the spatial variability of shrub canopy affecting infiltration and surface runoff. Singh and Woolhiser (2002) well resumed this concept:

"*Depending on the scale, the sources of variability can be stochastic or deterministic or both. It is not possible to describe watersheds in terms of a single deterministic length scale, independent of scale and watershed characteristics*."

In other words, models should describe the relevant spatial and temporal variability of inputs and catchment characteristics that is appropriate for the coverages considered and the relevant processes. In this sense, we think that the integration of smaller-scale variabilities to larger modelling scales could lead to a *process-based model* even for spatially lumped representations on mesoscale catchments. Also, we think that refining the temporal resolution (not only the numerical but also the input data one) should be considered at least as important as refining the spatial resolution, and this could lead to more appropriate process-based representations of the catchment functioning. Some processes could be well modelled at some large time intervals as the year and the day which are '*physically based*' (Klemeš, 1983), for example the evaporation from soil. Conversely, other processes could need higher temporal resolutions to account for the threshold behaviors depending on rainfall intensities, as the infiltration-excess runoff.

These observations lead us towards a new classification of models that we would term a '*scale-dependent process-based classification*'. It is schematically represented in Figure 1.7, which is inspired from Melsen et al. (2016a) (see their Figure 1, resumed also by our modified Figure 1.9). Melsen et al. (2016a) resumed a long-established inventory of the time and space scales of hydrological processes (following Blöschl and Sivapalan, 1995). Therein, the concept of '***characteristic velocity***' of hydro-meteorological processes was represented. It is defined as the ratio of characteristic length and time scales of processes and is called also '*process scale'* (Blöschl and Sivapalan, 1995; Melsen et al., 2016a). Blöschl and Sivapalan (1995) suggested that the characteristic velocity is roughly constant across a range of scales and stated that "*for atmospheric processes this characteristic velocity is of the order of 10 m/s, for channel flow it is 1 m/s and for subsurface stormflow it is less than, say, 0.1 m/s*."

In our opinion, an efficient process-based model must follow the direction of the characteristic velocities (Figure 1.7) by using a pair of spatial and temporal resolutions (for inputs and model functioning) whose ratio should be equal to and not greater than a maximum velocity. We name this ratio the '***Efficient Process-based Model Speed*** *(**EPMS**)*. By analyzing Figure 1.7, one may intuitively note that this *model speed* should not be greater than the *velocity of the most rapid process* playing a role within the catchment; otherwise the process evolution over space cannot be tracked by the model for a lack of spatial information. On the other hand, the use of a lower '*model speed'* may be associated either to a lack of temporal information to describe the temporal evolution of the process or to a redundant spatial representation of the process in the model (i.e. not efficient model speed). The EPMS is dependent on the inputs and catchment characteristics and conditions. For example, for a small catchment (e.g. 1-km^2) subject to a convective storm event with a maximum characteristic velocity in the order of 1 m/s (as that of the channel flow), a process-based model should have temporal and spatial resolutions with such a ratio (see corresponding line in Figure 1.7). In this case, the choice of a spatially lumped approach would be deterministically linked by the EPMS concept to a necessary temporal resolution (approximately 17 min for the 1-km^2 catchment).

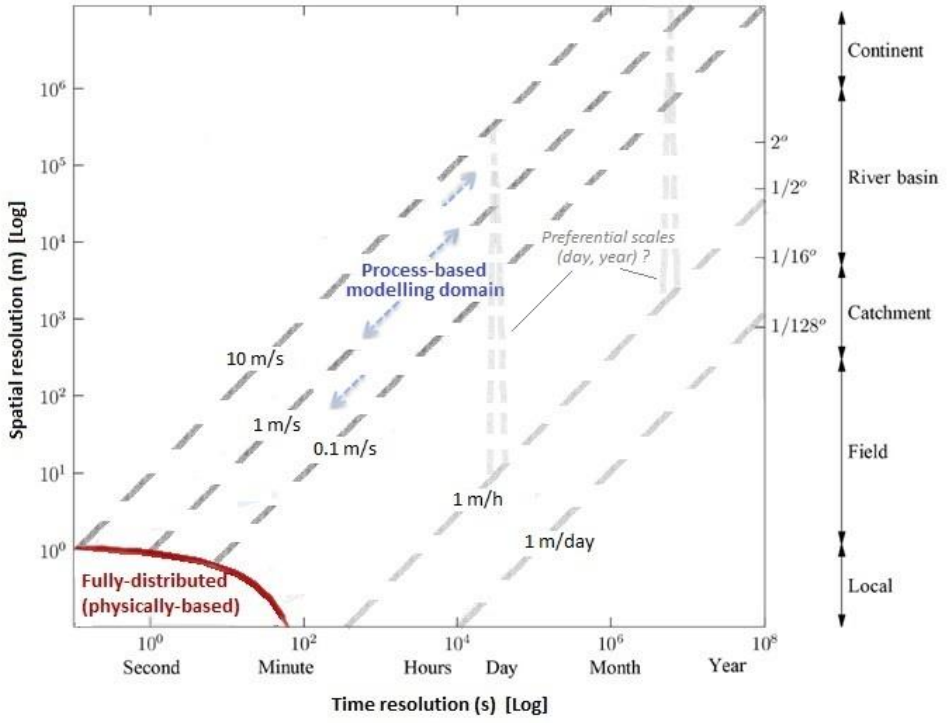


FIGURE 1.7 – Scheme of the fundamental basis of a '*scale-dependent process-based*' model classification, with a rough indication of the *fully-distributed* and *process-based* modelling domains in the time-space resolutions plan (on a log-log scale). The grey dashed lines represent the *characteristic velocities* of some hydrological processes. Original figure adapted from the scheme of characteristic process-scales lately presented by Melsen et al. (2016a) (which in turn was adapted from Blöschl and Sivapalan, 1995).

This scheme in mind, we agree with Beven and Cloke (2012) in saying that changing scales is just a necessary but not sufficient step for representing hydrological processes, as they well stated: "*Changing the scale of implementation of hydrological models does not, in itself, resolve the issues arising from fundamental lack of knowledge*."

Thus the scheme that we have proposed here could just help to determine the combination of spatial and temporal resolutions that a model should have to accurately and efficiently represent the involved processes (EPMS) and to aspire to be called 'process-based'. Then, a model with such a combination of resolutions needs also a proper representation of the hydrological processes that can be either conceptual (bucket-style) or physical (if the resolution is sufficiently fine). Any other model with neither a proper ratio of resolutions nor a sound representation of processes cannot be called process-based, making no significant difference between *a self-styled 'physically-based' model or a conceptual/empirical* (or *metric-conceptual*) model, given the overlaps already discussed.

As for the fully-distributed modelling paradigm, we confined it in the bottom-left corner of the time-space plan in Figure 1.7, for the necessity of the simultaneous refinement of the temporal and spatial resolutions, as already discussed above. A spatially distributed model, say at 1-m resolution and at daily time step, applied in a 1-km^2 catchment with a maximum characteristic velocity of 1 m/s (*EPMS*), has no chance of being called *process-based*, since

its model speed is 1 m/d (far slower than the EPMS). The physically-based notion applied to this model would be even more groundless. On the other hand, a spatially lumped model applied to the same 1-km^2 catchment could be *process-based* if the temporal resolution is chosen so as to obtain the EPMS ratio (around 17 min) and the processes are described with *realism* (for some notions of realism see Section 1.2.5).

1.2.4.4 Our approach: spatially lumped, mesoscale, continuous, multi time-step

By following the levels of classifications detailed above, we may now categorize our modelling approach that will focus on spatially lumped, mesoscale and continuous time models at different temporal resolutions from daily to sub-hourly. Our choice of a spatially lumped approach is motivated by the three reasons listed by Perrin et al. (2003):

(i) the importance of trying "*to determine how a catchment works as a whole*", given the actual limitations of knowledge on processes;

(ii) the fact that lumped models are usually used as "*building blocks" of distributed models*, and so their development must precede their gathering in interconnected distributed systems;

(iii) the fact that "*the practical superiority of distributed or semi-distributed approaches over lumped ones for streamflow simulation has not been clearly demonstrated yet*".

These reasons are also supported by our concept of *Efficient Process-based Model Speed* (explained in the previous Section), according to which the spatial distribution, in itself, is not sufficient to develop a process-based representation of the rainfall-runoff relationship, but processes should be modelled at a suitable combination of spatio-temporal resolutions consistently with the catchment characteristics.

If one wants to still categorize our approach by the traditional classification on model conceptualization, we may say to follow a hybrid combination of the so-called empirical and conceptual paradigms, by combining the use of model structures and components typical of the conceptual models with an empirical approach for model development based on data (see Michel et al., 2006).

1.2.5 Desirable properties of rainfall-runoff models

Some relevant properties are generally required for rainfall-runoff models including: *reliability*, *robustness*, *realism* (*the 3 R's*) and *transposability* (across space and time scales), as resumed by Gupta et al. (2014). These properties are interdependent concepts. The reliability is probably the most general concept, hooked on all the other mentioned properties, and also others. To be reliable, a model should be robust, realistic and transposable, but also accurate, general and consistent. So all these properties should be verified by an extensive model diagnostics based on a large-sample approach, in which traditional aggregate performance indices are not sufficient. As cited by Gupta et al. (2014), to be demonstrably

robust, models must pass effective crash testing procedures proposed for instance by Klemeš (1986b) and Andréassian et al. (2009).

As for *realism*, it is important to note here that even if proving (and even defining) the "realism" of a rainfall-runoff model is a very difficult task, especially for an empirical or conceptual model, it is still possible to find out proofs of hydrological likelihood (i.e. a possible definition of *realism*) of a model structure (e.g. Le Moine et al., 2007). We think that this attitude vis-à-vis of models must be used to corroborate their conceptualization, because it is important that models "*work well for the right reasons"* (Klemeš, 1986a; Kirchner, 2006). To ensure at least a minimum level of model credibility, we remind the importance of standard calibration-validation schemes (Klemeš, 1986b). Finally, we refer the reader to Section 1.3 for further discussion on the question of *transposability* across space and time scales, which is central to this work.

Model property	Definition	Modes of evaluation	Comments
Accuracy	"*Ability to closely reproduce the historic hydrograph throughout the range of flows*" (Linsley, 1982)	- Statistical measures of similarity between observed and simulated hydrographs as the *Nash-Sutcliffe efficiency* (NSE) (Nash and Sutcliffe, 1970a) - Hydrological "signatures" (e.g. Yilmaz et al., 2008)	It is the first and best basis for model selection and improvement.
Applicability	Good matching between model requirements and actual available data and computing capabilities	- Comparison of required data and calculation time of the model with the resources available	The initial choice of models complexity should take into account the available resources right from the phase of conceptualization (Bergström, 1991).
Ease-of-use	Ease of a proper use of the model for the final end-users purposes	- End-users understanding and efficiency of use of the model	Ease-of-use should not be placed before accuracy, as stated by Linsley (1982).
Generality	Ability of the model to provide accurate results on many catchments and over long time coverages	-Extensive testing on large-samples of catchments (Andréassian et al., 2009)	Nowadays, the application of models on large data-sets is possible, thanks to the continuous developments of computing capabilities and increasing data availability (Gupta et al., 2014).

TABLE 1.2 – Summary and comments on the four essential desirable properties of rainfall-runoff models promoted by Linsley (1982).

Table 1.2 provides an overview of the other four fundamental characteristics promoted by Linsley (1982): *accuracy*, *applicability*, *ease of use* and *generality*. The choice to prioritize one feature or the other may be justified in light of the modelling objectives. In this research work, we will not explicitly prioritize one feature on the others, but we will extensively consider model accuracy, measured by statistical metrics of similarity between observed and

simulated hydrographs, and also generality, by conducting the tests and evaluations over a large sample of catchments.

1.2.6 The fixed and flexible modelling paradigms and our model choice

The approach of applying a single model on many catchments is necessary to seek generality, as discussed above, and corresponds to the so-called '*fixed modelling'* paradigm. This approach is opposed to the '*flexible modelling*' paradigm (e.g. Fenicia et al., 2011; Van Esse et al., 2013). Here it is interesting to briefly present and discuss the contrast of these two paradigms and introduce the approach followed in this work. In our view, both approaches are valuable and important and present different advantages.

The '*fixed modelling'* paradigm is based on the choice of building a general model with a fixed structure that is validated on many different catchments and over long time series, as promoted by some authors (e.g. Perrin et al., 2001; Le Moine et al., 2007). We agree with Linsley (1982) in saying that in general "*the fundamental processes of hydrology are the same in all catchments*", even if the dominant processes may change . Some processes may vary and become more dominant with the local characteristics of climate (e.g. presence of snow), landscape (e.g. vegetation) and soil (e.g. lithology). Anyway, we believe that a general rainfall-runoff model may have a fixed structure but enhance (or deactivate) a process thanks to the pre-setting or calibration of some of its parameters. Despite the difficulties that may emerge in the calibration to well represent the *uniqueness* of individual catchments (Beven, 2000), we think that building a general fixed model structure has many merits, as increasing the *credibility* of the model, and so the *confidence* of the end-users and "*lead to meaningful generalizations*" (Klemeš, 1986b). On the other hand, the *flexible approach* is based on searching the model structure that best 'fits' each specific catchment of interest, among a set of *multiple working hypotheses* (Clark et al., 2011). It is a promising approach because it should reduce more easily structural uncertainty (Van Esse et al., 2013), and this is to be verified, because of its great interest in research applications. However, further work is needed for improving its applicability and ease-of-use, above all for applications out of the research community. The necessity of customizing a model structure for each specific case of application is time consuming and less practical for operational purposes.

For our research work, we have chosen to base it on the long-established, fixed modelling paradigm that seems to us a more direct approach for answering our research questions and operational expectations. In particular, we have decided to start from a suite of parsimonious models widely used in the hydrological community, choosing the GR models chain at different time steps, including the *GR4J* and *GR4H* models (Perrin et al., 2003; Mathevet, 2005; Le Moine, 2008). This choice is motivated by the need of continuing investigating the transferability of these models (structure and parameters) across temporal resolutions, which stems from a specific *operational demand* from the 'SCHAPI' (the French national service for flood forecasting). At the same time, this is a *current general challenge in hydrological modelling* (e.g. Singh and Woolhiser, 2002). The GR models have been widely used for

research purposes on large sets of catchments around the world with generally reliable results (e.g. Perrin et al., 2001; Perrin et al., 2003; Le Moine et al., 2007; Van Esse et al., 2013) and their modified version called *GRP* is currently used for flood forecasting in France (Berthet et al., 2009). Thus the choice of these models is a good starting point for such a specific research work on the time scale issue and reveals a dual interest (both of research and operational). From an epistemological point of view, a parallel can be drawn between our research on model structures across different temporal resolutions and the flexible framework for hydrological modelling promoted by Fenicia et al. (2011). In fact, in both cases the objective is to "*generalize and systematize the currently fragmented field of conceptual models*" (in our case, at different time steps; in the flexible framework, at different locations).

1.3 The space and time scale issues in hydrology

The spatial and temporal scaling issue in hydrological modelling has been recognized as one of the most rewarding challenges for models development and advance in hydrological science by many authors, as, for instance: Klemeš (1983), Blöschl and Sivapalan (1995), Singh and Woolhiser (2002), Blöschl (2006) and Merz et al. (2009). As highlighted by Blöschl (2006), *hydrological synthesis is needed across scales*, to find out the general characteristics of processes as a function of space and time scales for the same site or an ensemble of sites.

1.3.1 General concepts and typical scales of hydrological processes

The *term 'scale'* is normally defined as the sampling interval size at which hydrologic observations are made (*observation scale*) or as the grid size used for numerical computations (*model scale*) or as the characteristic time or length of a process (*process scale*). Thus, the *scale (grid size)* will correspond to the length in the spatial domain and to the duration in the time domain (Singh and Woolhiser, 2002).

Many authors explored different aspects of the spatial and temporal scaling issues of hydrological processes. Here we summarize some of the main contributions with a deeper focus on the temporal scale issue that is the central theme of our work. However, the problem of taking into account the temporal variability in hydrological modelling presents some analogies and is linked with the problem of the spatial variability (as we have highlighted in Section 1.2.4.3). So we report also some findings of studies investigating the spatial representation that could be transferred also to time.

- Klemeš (1983) discusses some basic features of space and time scale to bear in mind when dealing with scaling issue in hydrology. He argues that *scale should be viewed as a quality* and not just a purely quantitative size. "*In nature, scales of things are not arbitrary*" and meaningful conceptualization of physical processes is possible only at some discrete scales. In nature we can find preferred nodes on the spectrum of scales, as in physics we have different discrete entities from the quanta and subatomic particles to

the macroscopic bodies. A qualitative feature of scales is that different forces tend to dominate at different levels of scale and this limits a mechanical extrapolation of the mathematical relationships expressing physical laws from one level of scale to another. Another important feature, often disregarded in scientific research, is the intrinsic relation between the time and space scales. An indifferent attitude to this interconnection of space and time scales makes the understanding and conceptualization of natural processes more difficult. This is in line with the scheme we have proposed in Section 1.2.4.3 for defining process-based modelling. Klemeš (1983) highlights that "*given the time scale, the choice of the spatial scale determines the kind of relationships that one can hope to identify*" and vice versa. It is necessary to identify the time lags between rainfall and runoff of the given spatial scale and compare with the time interval selected to understand whether "*the dynamic effects of the spatial unit will be swallowed by integration over a single time interval*". This is particularly important when defining the scientific expectations of hydrological modelling at the "basin scale". However, he argued that "*the arbitrary use of spatial and temporal scales is not, per se, the most disturbing aspect of our approaches to conceptualization of hydrologic processes."* This is well explained by his metaphor:

"*Nobody can be blamed for not immediately knowing the correct way through a complex labyrinth. What he can be blamed for is an insistence on a preconceived idea of the correct route and unwillingness to check it out*."

- Morel-Seytoux (1988) suggests that "*the passage from a smaller to a larger scale requires: enlightened simplification, integration in many senses that is*: *(1) in time; (2) in space; (3) in an expectation sense; and (4) in a process sense, and finally enlightened coupling*." The author focuses on a simplified but physical description of hydrological processes and suggests that temporal patterns of rainfall, temporal fluctuations in river stage and other spatial and temporal variability features of the hydrological processes significantly affect the catchment response. For example, an illustration of the importance of temporal patterns of rainfall is given (see paragraph below).

- Blöschl and Sivapalan (1995) provide a review for scale issues in hydrology and a framework for scaling hydrological models. They analyse the operation of '*scaling'* that is defined as *transferring information across scales*. In a *model-oriented approach*, the information to be scaled consists of state variables, model parameters and inputs, as well as the model conceptualization itself (i.e. the model structure). *Upscaling* typically consists of two steps, distributing and aggregating, while *downscaling* involves disaggregation and singling out. Blöschl and Sivapalan (1995) give some examples of deterministic and stochastic approaches used for scaling in hydrology. For example, to link model conceptualizations across spatial scales, a common useful way is the use of sub-grid parameterizations to represent the effects of smaller scale processes than the modelling spatial scale, i.e. an example of top-down approach (as discussed also in Beven (2009), p. 9). In a more *holistic (hydrologic) approach*, scaling may be based on the concept of similarity, i.e. determining *scale factors* to link systems across scales. To this

end, similarity analysis, dimensional analysis and functional normalization can be used to determine empirical relationships between different variables at different scales.

Following Blöschl and Sivapalan (1995) and Melsen et al. (2016a) we report a classification of hydro-meteorological *processes* according to their ***characteristic scale*** (time and length) by a diagram of typical process scales based both on data and heuristic considerations (Figure 1.8).

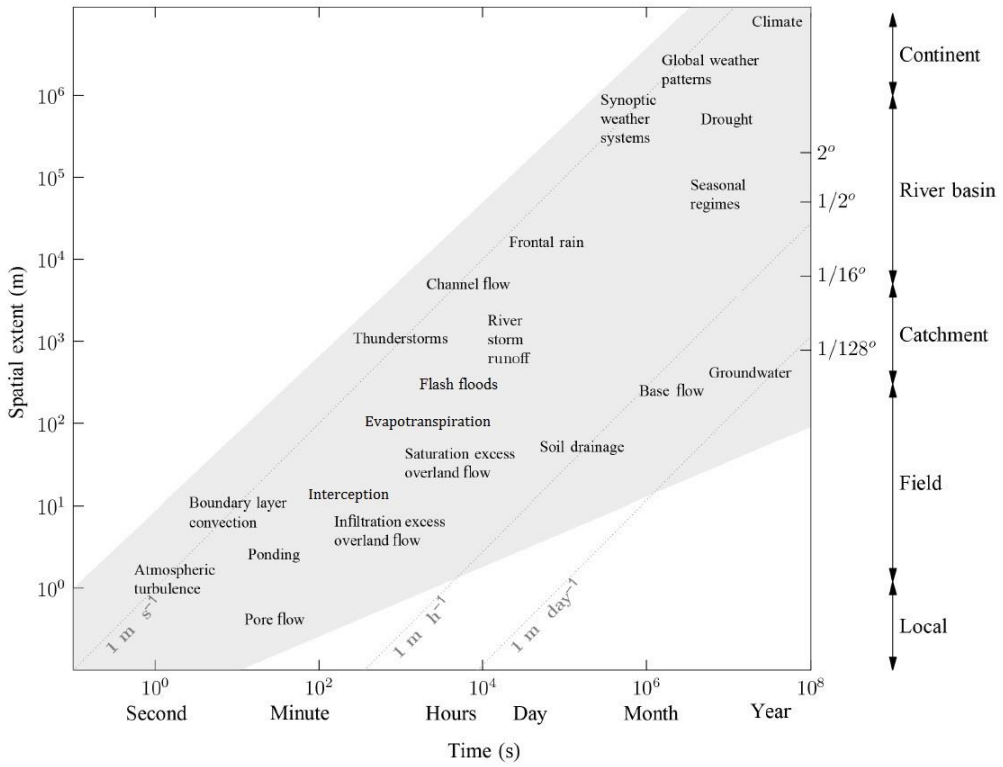


FIGURE 1.8 – Typical scales of several hydro-climatic processes in the spatio-temporal domain (seconds – meters, log-log scale). Figure modified from Melsen et al. (2016a), which in turn is based on Blöschl and Sivapalan (1995).

This diagram was based on previous ones used in the atmospheric sciences and has been lately resumed by Melsen et al. (2016a). Figure 1.8 reports our adaptation of this latter version. With respect to the original figure, our main change is the *separation of the **interception** process from the **evapotranspiration** one*, based on knowledge gained from other previous works (see Calder, 1990; Savenije, 2004; Gerrits et al., 2010). Indeed, these two processes are distinct and, even though they are bound and driven by common variables (mainly solar radiation) they are generally active on different scales. Following a well-established terminology, *interception* as a process refers to the storage of part of the rainfall above the ground surface, mostly in vegetation (canopy and forest floor) in a natural environment, and the evaporation from this storage that generally occurs during the rainfall event and shortly after the end of the event (e.g. Gerrits, 2010). From our analysis of the literature cited above, we have placed the interception process at higher spatio-temporal resolutions than evapotranspiration. One should note that defining the scale of hydro-

meteorological processes is a hard task because these processes are '*active*' (variable) at a *wide range of scales* and interact among them by stochastic linkages. For example, one may think at the time variability of the *interception*: this process is highly *seasonal* (because of seasonality of potential evaporation, precipitation and storage capacity of vegetation), but presents also *sub-daily and sub-hourly patterns* (due respectively to daily temperature pattern, and sub-hourly variabilities of precipitation and wind-speed). In general, Figure 1.8 shows approximately the highest frequency of each process (in time and space), as we have done for interception. The shaded grey region shows the approximate whole domain of activity of hydrological process, while the straight lines crossing this region represent the *characteristic velocities* of processes, suggesting a positive correlation between spatial and temporal scales (see also Section 1.2.4.3).

1.3.2 The importance of spatial and temporal rainfall distribution

Some authors have shown the importance of rainfall variability in space and time for runoff simulations. For spatial variability the literature is abundant, but there is no consensus on the impact of spatial resolution of rainfall on the performance of hydrological models (see for instance Lobligeois, 2014, and references therein). On the contrary for the impact of temporal variability, hydrological literature is more limited, although it is in rapid development. However, it seems that there is more consensus on the significant impact of temporal resolution of rainfall on runoff simulations. It has been recognized by many authors that runoff generation is highly affected by sub-hourly dynamics of precipitation, as for example by Morel-Seytoux (1988), Woolhiser and Goodrich (1988), Krajewski et al. (1991), Kandel et al. (2005), and Paschalis et al. (2014). Precipitation controls the high-frequency catchment response despite the catchment low-pass behavior (Oudin et al., 2004), which smoothes out high frequencies.

This importance of rainfall for the high-frequency response may be explained by the *sub-hourly characteristic time scales of* ***infiltration*** and its highly *nonlinear* nature (Kandel et al., 2005). Some studies have shown that (infiltration-excess) surface runoff is modelled better by using fine time step rather than daily time step models and, particularly, peak rates of rainfall are recognized as the most important controls for rainfall-runoff modelling (e.g. Yu et al., 1998; Socolofsky et al., 2001; Kandel et al., 2004; Kandel et al., 2005). All these studies indicate that sub-daily variability of rainfall intensity is important and quite short time scales should be represented in models to capture this variation. This may be explained by the *rate-limited behaviour* of infiltration-excess mechanism. Here we briefly summarize some main findings about this widely investigated subject.

- Morel-Seytoux (1988) discusses how the *temporal pattern of rainfall* impacts the *runoff coefficient* of a soil by illustrating the case of a rain with a total depth of 3 cm over one hour and three possible rainfall patterns, as reported in Table 1.3. The *assumption of rainfall rate uniformity over one hour is unrealistic* and has a negative impact on models results, as it can be deduced from Table 1.3. As the rainfall intensity increases the runoff coefficient increases too. This example let the author suggest that infiltration is at least as

sensitive to intensities as to total depth of rainfall and an improvement in forecasting of the overland flow of a catchment will come from an improvement in the definition of the rainfall pattern rather than from a refinement of simple infiltration formulae.

Rainfall pattern	**Rate 3 cm h^{-1}, duration 1h**	**Rate 6 cm h^{-1}, duration 0.5 h**	**Rate 12 cm h^{-1}, duration 0.25 h**
Homogeneous soil	0	5	26
Heterogeneous soil	9	17	31

TABLE 1.3 – Influence of rainfall temporal pattern and of soil heterogeneity in hydraulic conductivity on the runoff coefficient (%). *Source:* Morel-Seytoux (1988).

- Some other authors investigate the importance of the time distribution of rainfall for runoff response and agree with the suggestion made by Morel-Seytoux (1988). For example, Koch and Kekhia (1987), using a physically based model and three soil types, show that the required complexity of the storm temporal distribution is dependent on the soil type and the constant intensity approximation is less adequate for more permeable soils. Also Woolhiser and Goodrich (1988) point out the significance of the rainfall intensity-infiltration interaction by using a simple physically based model, concluding that *the **constant intensity rainfall pattern** cannot be recommended*, especially for rapid catchments (response time of the same order or smaller than the duration of rainfall excess).

- Some authors investigate at the same time the influence of both spatial and temporal rainfall patterns on the catchments hydrological response. Krajewski et al. (1991) investigate the sensitivity of a distributed model performance with respect to rainfall spatial and temporal sampling density for a very small catchment (7.5 km^2). Their results indicate a greater sensitivity of basin response with respect to the temporal resolution than to the spatial resolution of the rainfall inputs. Menabde and Sivapalan (2001) investigate the relationships between the various time and space scales of variability in rainfall and runoff process within a similarity framework. They analyse the spatial scaling properties of peak flows and show the importance of the ratio between two controlling variables, namely the **storm duration** and the **concentration time of catchments**. This conclusion is supported also by other works as those by Robinson and Sivapalan (1997) and Gabellani et al. (2007). Nicótina et al. (2008) investigate the spatial scales at which rainfall spatial variability influence the flood response by coarse-graining rainfall input fields from 100 m- to 50 km-resolution for some Italian basins of different size. They found that the catchments response is less influenced by spatial variability of rainfall for small basin areas (up to about 3500 km^2), rather than in larger basins where the travel time in the channels is a more important part of the total residence time. Paschalis et al. (2014) show that the flood response is strongly affected by the temporal correlation of rainfall and to a lesser extent by its spatial variability.

1.3.3 Catchment characteristic time scales

Several definitions of the catchment characteristic time scales can be found in the literature, describing in different ways the time lag between a rainfall event and the resulting hydrograph. The primary importance in hydrological modelling of these time characteristics is evident. This can be argued easily as almost all hydrological models contain at least one time-dimensioned parameter (Morin et al., 2002), and routing components (as Unit Hydrographs) are used to simulate the time lag between the rainfall event and the resulting streamflow peak. Some of the most frequently used definitions of catchment time scales are: the *time of concentration*, the *lag time*, the *time to peak* and the *time base* (see definitions in Table 1.4). For a detailed review of the estimation methods of these time parameters one can see Gericke and Smithers (2014) (particularly their Table A3).

Catchment time scale	Definition	Methods of estimation	Comments
Time of concentration	Time required for a drop of water falling on the most remote part of a drainage basin to reach its outlet (Singh, 1992, pp. 451-452)	Empirical formulae (e.g. see formulae of Giandotti, Kirpich, Passini, Turazza, Ventura, etc.)	Used in the rational formula (Singh, 1992, pp. 595)
Lag time	Time between the centroid (centre of mass) of rainfall excess and the centroid of the direct runoff hydrograph	As in its definition or using the peak of the runoff hydrograph instead of the centre of mass (Morin et al., 2002)	Examples of its use are provided by Caroni et al. (1986) and Simas and Hawkins (1998) who evaluated its variation
Time to peak	Time between the beginning of direct runoff and the peak of direct runoff hydrograph (Singh (1992), pp. 451-452)	Algorithms aimed at defining the *unit hydrograph (UH)*, e.g.: the DPFT method by Duband et al. (1993), or the IHACRES method (Schreider et al., 1996) Time lag that maximizes the *cross-correlation between rainfall and streamflow* (Dong et al., 2005)	Many hydrological applications as peak discharge estimation. Definition of a maximum acceptable time step for rainfall-runoff models (Obled et al., 2009)
Time base	Duration of the direct runoff hydrograph (Singh, 1992)	As in its definition.	Unit hydrographs

TABLE 1.4 – Summary of four common catchment characteristic time parameters: definitions and estimation methods.

Other characteristic time scale parameters were suggested in the literature as the *time to equilibrium*, *time of travel*, *rise-time*, *volume/peak ratio*, *critical-lag time*, *infiltration-opportunity time and time to ponding* (Singh, 1992). The definition of all these parameters goes beyond the scope of this work. However, we note that different methods can be developed to determine a time parameter describing the timing of the catchment response and this can be done even just by data analysis. Some data-driven methods are based on *similarity of rainfall and runoff patterns* (Morin et al., 2001; Morin et al., 2002) or *spectral analysis* (e.g. Tessier et al., 1996). It can be noted that most of the definitions of catchment

characteristic time scales are related to the fast response, which is only one component of the outlet flow. Very little is made on long characteristic time scales that relate to slower flow components which are typically related to storages in models. According to us, this aspect is as important as for fast response and should deserve more research in the future.

As for the relationship between the catchment response time and the hydrological model time steps the literature is less abundant. On this topic, the work by Obled et al. (2009) (see Section 1.4.1*)* uses the definition of '***time to peak***' to define a *maximum acceptable time step* that can be used for modelling the rainfall-runoff relationship. However, we have not found published works dealing with the analysis of the relationship between the time-dimensioned parameters of rainfall-runoff models and the catchment time scales time parameters from neither a conceptual nor a physical point of view. This would be important for improving the regionalization of model parameters, as highlighted by Littlewood and Croke (2008).

1.4 State-of-the-art on the temporal scale issue in rainfall-runoff modelling

The question of temporal scaling in hydrological modelling involves different aspects, as the relationship between process, observation and model scale or the dependency of model parameters and structures on time step, that are discussed in the following sections.

1.4.1 Relation between process, observation and modelling scale

As discussed by Blöschl and Sivapalan (1995), processes should be ideally observed and modelled at the scale they occur. However, this is not always possible because some hydrological processes have characteristic time scales of minutes (often shorter than the data available resolution) and they simultaneously operate at a range of scales. Available data for rainfall-runoff modelling are discrete: rainfall is measured with total accumulations over fixed time durations, usually hourly or daily, and streamflows are punctually measured at some instants with a fixed or variable sampling interval. Data sampling in time has a filtering effect on processes as we have already highlighted in Section 1.2.2. As Blöschl and Sivapalan (1995) observe, "*processes larger than the data coverage appear as trends in the data, whereas processes smaller than the resolution (spacing) appear as noise*". This is a well-known problem in the field of signal processing, where the *Nyquist-Shannon theorem* (Shannon, 1949) defines the highest frequency f_n of a oscillatory continuous process that can be detected from a discretized data set of *spacing d*, as: $f_n = \frac{1}{2d}$.

The ***Nyquist criterion*** may be *transposed to hydrological modelling* after some small adjustments to consider that: (i) average (or cumulated) inputs/outputs over a time step are usually considered in hydrological modelling, instead of 'punctual' sampling (Chow et al., 1988); (ii) hydrological systems are not oscillating, so their response is damped: at the catchment scale, an impulsive input of rainfall never generates an impulsive decrease of streamflow rate (Obled et al., 2009).

Thus, in hydrology, by following a lumped modelling approach, an adapted observation and modelling scale can be fixed by the dynamics of the system's output, i.e. the outlet streamflow, that integrates all the different processes and time scales. As highlighted by Obled et al. (2009), the observation and modelling time step should be upper-bounded to "*properly reproduce the most rapid dynamics of the system*". Thus, by following the notions of *impulse response* and *time to peak*, Obled et al. (2009) define a *maximum acceptable time step* (MATS) to represent the peak response of a catchment as:

$$MATS = \frac{T_p}{3} \tag{1.3}$$

where T_p is the time to peak of the catchment (for its definition one can see the *Section 1.3.3*). The coefficient $\frac{1}{3}$ derives from the fact that the sampling technique used for streamflows is averaging over the time step Δt. Obled et al. (2009) empirically show that a time step $\Delta t \leq \frac{T_p}{3}$ is sufficient to represent a peak response without denaturing its form by testing this on a flood hydrograph case study (see also Figure 1 in Obled et al., 2009). This recommendation is coherent with the one by Maniak (1997) that also indicates a *time step upper bound of one third to one fifth of the time to peak of a discharge event*.

The chosen time step (MATS) is used by Obled et al. (2009) to define the maximum acceptable spatial resolution which allows to consider the precipitation inputs as homogeneous over each discretization grid element. To this end, they suggest the use of some known results in the field of geo-statistical analysis of precipitation, in particular on the *"decorrelation distance" (a),* i.e. distance for which the correlation decreases to *1/e* of its maximum (Rubel, 1996). It has been shown that the *decorrelation distance* increases with the time step Δt over which rainfall is cumulated and this relationship can be described by an exponential model representing the correlation in function of the distance h and the time step of accumulation Δt. Obled et al. (2009) suggest that the maximum spatial 'step' Δx for which one can consider the rainfall uniform over the domain $S = (\Delta x)^2$ is in the order of a at most. So, referring to typical values of a from the literature, between 20 and 25 km for hourly time step, and expressing the function $a(\Delta t)$ as a power of Δt with exponent between 0.3 and 0.5, a maximum spatial resolution can be defined as:

$$S_{km^2} \approx [a(\Delta t)]^2 = N_h(\Delta t) \cdot 400 \; to \; 625 \; km^2 \tag{1.4}$$

where $N_h(\Delta t)$ is the number of hours of the time step Δt (adimensional). This means that for catchments with a spatial resolution in the order of S_{km^2} one can reasonably use a lumped model, with a single input of rainfall for the whole catchment.

1.4.2 Effects of time steps on model parameters and performance

The effects of the temporal sampling of calibration data on model parameters and corresponding model performance have been investigated by an increasing number of authors, such as: Finnerty et al. (1997), Ishidaira et al. (2003), Cullmann et al. (2006), Tang et al.

(2007), Littlewood and Croke (2008), Wang et al. (2009), Cho et al. (2009), Ostrowski et al. (2010), Littlewood et al. (2011), Kavetski et al. (2011), Bastola and Murphy (2013), and Melsen et al. (2016b). This important area of research for rainfall-runoff modelling for both operational and research aspects has been overlooked for a long time in the literature, but is now rapidly developing.

In general, all the studies cited above show that a great part of parameters of hydrological models depend upon the calibration data temporal resolution and highlight the importance of accounting for this dependency in order to reduce the uncertainty of simulations, to improve parameters identifiability and regionalisation. There are parameters that are time step independent and can be viewed as intrinsic properties of the basin, while other parameters depend on the time step of calibration data. The explicit definition of parameters-time step dependency is important to effectively transfer a model at different time steps, which is often necessary in an operational context, because longer time series are more frequently available at coarser time steps (e.g. daily). So it is useful to define a method to transfer the information derived from larger time steps data to shorter time steps (e.g. Nalbantis, 1995). Further research is required on this topic by analysing the effects of changing time steps over a larger range of time scales and testing different models on a wider range of catchments. Here below we summarize some main findings that one can find in the literature about this topic.

- Mathevet (2005) and Le Moine (2008) investigated the relationship between the parameters of the lumped GR4J model at *daily time step* and those of the derived *hourly* model version, GR4H (see also Section 1.4.3.2 for more details about these models). The two models share the same structure and functions, but the *temporal scaling* generates a transformation of some fixed and free parameters. First, these transformations were derived *a-priori* from the *integration at different time steps of the model governing equations*, expressing the fluxes of the model (Le Moine, 2008, pp. 172-173). This can be done for the fixed parameter (not calibrated) describing the *percolation* of the production reservoir and the *four free parameters* to be optimized. The four free parameters are:

    (i) the capacity of the production reservoir, θ_1[mm], that is theoretically independent from the time step;

    (ii) the water exchange coefficient, θ_2 [mm], whose value is derived from the hypothesis of slow dynamics of this process (i.e. daily fluxes uniformly distributed over the hourly steps).

    (iii) the capacity of the routing reservoir, θ_3 [mm], whose value at different time steps is derived from the integration of the reservoir emptying function;

    (iv) the base time of the UH, θ_4 [time], that is directly expressed in time step units.

The temporal transformations were empirically verified for the optimized-parameters obtained at the two different time steps over a large sample of 1040 French catchments (Le Moine, 2008, p. 179). Le Moine (2008) found a quite good coherence between the calibrated parameters of the hourly-model and the theoretical values obtained from the daily calibrated

values, as shown in Figure 1.9 (from Le Moine, 2008), especially for the two reservoirs capacities. The base time of the unit-hydrograph systematically deviates from the theoretical relationship for values lower than 24h. This is explained in Le Moine (2008) by the mathematical form of the UH: by construction at the daily time step the UH form does not change for different values of θ_4 lower than 24 h.

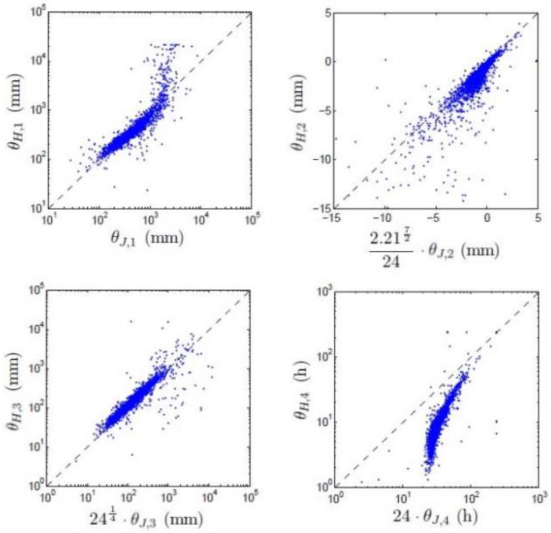


FIGURE 1.9 – Coherence between the calibrated values of the parameters of the hourly model GR4H (vertical axis) and their values derived from the theoretical relationships linking them with the parameters of the daily model GR4J (reported from Le Moine, 2008, p. 180, with a different notation).

- Littlewood and Croke (2008) investigated the data time-step dependency of the *6 parameters* of a conceptual, lumped, unit-hydrograph-based model, *IHACRES*, for *two catchments* in Wales (10.6 and 298 km^2). The motivation of their work was to search for time-step independent parameters that could improve regionalisation. They showed that all the parameters change substantially when calibrated on data sets at different time steps from 1 h to 1 day. As the time step decreases, the parameters reach *stable values*, especially for the smaller basin. For the larger catchment this trend seems less clear, probably for the larger spatial variability and uncertainty in areal averaged rainfalls. The model performance criteria, the *Nash-Sutcliffe efficiency*, NSE, (Nash and Sutcliffe, 1970a) and *average relative parameter error*, were relatively good and did not change substantially over the range of time steps. This study presents some limitations, some of which recognized by the authors as the fact of using only one model on only two catchments and the need to extend the time steps range to sub-hourly data. We found also two other limitations that presumably affect their results. First, too short calibration periods are used for the two catchments (7 and 27 months). Second, the NSE measures, calculated at each model time step, have been used also to compare the performance at different time steps, although these scale-dependent measures are not directly comparable. In fact, the time step strongly affects the statistical moments and correlation and so the NSE measures too. For comparisons at different time steps, these statistics should be calculated based on aggregation of the output series at a common larger time step.

- Wang et al. (2009) applied a conceptual model with *17 free parameters* to a 0.21-km^2 forested experimental catchment in Japan. The model consists of reservoirs representing interception, evapotranspiration, infiltration, percolation processes. By employing *10-minute, hourly and daily data* they evaluate (i) daily discharge predictions on a 5-years period by using hourly and daily rainfall data and (ii) hourly discharge predictions on 10 storm events by using 10-min and hourly data. In both cases results obtained with higher resolution data show better predictions especially on the peak discharge and recession part of the hydrographs and better efficiency in terms of mean relative error and NSE. We note that in contrast to Littlewood and Croke (2008), in these comparisons the error statistics were calculated at the same time step of aggregation (respectively 1 day and 1 hour in tests (i) and (ii) respectively) and this is consistent with our observation above.

  About the parameters-time step dependency, they show that parameters describing slow flow processes reveal a low sensitivity, remaining roughly constant across the time steps used, while quick flow parameters reveal higher dependence on the resolution. Finally, Wang et al. (2009) found that the rapid response parameters are roughly proportional to the square root of the time interval and analysed the link of this relationship with the rainfall depth-duration relationship by using the Sherman equation. The work by Wang et al. (2009) however presents some limitations, as outlined by Littlewood et al. (2011) in a discussion paper that we briefly summarize here. The main critical argument is that in a model with such a number of free parameters (17) there is likely strong correlation between at least some of the calibrated parameters, leading to problems of identifiability (Jakeman and Hornberger, 1993). Thus, in the analysis by Wang et al. (2009) it is not clear what is the impact of fixing some model parameters in contrast to what done by Littlewood and Croke (2008) who calibrated all five parameters of their model at the same time. This could be one reason of the different conclusions derived by Wang et al. (2009) compared to Littlewood and Croke (2008) who found that all parameters change with Δt, including non-time dimensioned and slow flow response parameters.

- Ostrowski et al. (2010) employed a soil moisture accounting model, dominantly based on infiltration-excess runoff, for two small experimental catchments in Austria (about 20 km^2) to analyse parameters-time step dependencies. By using data at different time steps, *from 5-minute to 1 day*, over a period of 3 months, they found a strong impact of the time step on some parameters. They showed that an irregular response surface for larger time steps becomes smoother for small time steps, facilitating efficient automatic calibration. They proved the existence of non-linear and linear dependencies for some parameters as hydraulic conductivity and maximum infiltration rates. Other non-time dimensioned parameters, as wilting point and field capacities, resulted to be slightly or not dependent on time step. For the hydraulic conductivity parameter k_f they tested successfully the functional relationship proportional to the square root of the time step suggested by Wang et al. (2009). Moreover they showed also the similarity of this power function with the relationship between rainfall maximum intensity and time steps.

- Kavetski et al. (2011) investigated the time step dependencies of models parameters and their uncertainties by examining also their relationship with model complexity (see next Section 1.4.3 for the latter). They used *4 model structures* including *increasing complexities*, i.e. more model components and parameters (from 5 to 9), in a *flexible model framework* (Clark et al., 2008; Fenicia et al., 2008). Also they used both the fixed-step explicit and implicit Euler time stepping schemes (see Appendix A). The parameters of the four tested model structures were inferred from calibration data with *eight different time steps from 30 min to 3 days* using two inference schemes for parameter uncertainty analysis, i.e. *standard least squares (SLS)* and *weighted heteroscedastic least squares (WLS)*. Their results show that fixed-step explicit Euler time stepping scheme and SLS inference scheme tend to introduce strong *spurious time scale dependencies*, especially for quick-flow processes parameters. As for the parameter uncertainty across time scales, no general trends are identified but a time scale at which some parameters become identifiable is observed, indicating the *typical time scale of the processes they represent*. For example some quick-flow processes parameters are largely non-identifiable for large time steps ($\Delta t \geq 6h$), but converge to stable values when the time step decreases. Conversely, base flow parameters stabilize for much larger time scales.

- Bastola and Murphy (2013) evaluated the sensitivity of model parameters and performance to the time step using rainfall and runoff data at *3-, 6-, 12- and 24-hourly time steps*. They employed a lumped conceptual model, HYMOD, with *5 parameters*, on two Irish catchments (1803 and 2452 km^2). First, they showed a loss in performance (measured in terms of NSE and volume error) by using models calibrated with daily data for simulations at shorter time steps. This supports the argument that recalibration or scaling of model parameters is desirable if the model time step is changed. Then they assumed a-priori a linear scaling relationship between parameters and temporal scale. This relationship was used in a multi-time step calibration method to estimate model parameters for all time steps simultaneously (by maximizing the average performance for all model time steps). Compared to unscaled parameters, the loss in performance corresponding to the use of scaled parameters is shown to be significantly smaller. This indicates the potential of such scaling techniques in situations where sub-daily data may not be available. An alternative approach could have been calibrating parameters for each considered time step and subsequently estimating the scaling relationship from the median parameter values of parameter sets at the different time steps. Further work could compare this approach to the one followed by Bastola and Murphy (2013).

- Melsen et al. (2016b) investigated to what extent the parameters of a model are transferable across temporal and spatial resolutions. They stated that "*the degree to which parameters are transferable across temporal and spatial resolutions is an indicator of how well spatial and temporal variability is represented in the models*". They used the Variable Infiltration Capacity (VIC) model with 7 parameters for the Thur basin in Switzerland (1703 km^2), run with *four spatial resolutions (1-km^2, 25-km^2, 100-km^2, lumped)* and evaluated for *three temporal resolutions (1 h, 1 day and 1 month)*, both applied with uniform and distributed forcing. Three objective functions were used for

model evaluation: (i) the Kling-Gupta efficiency (KGE) (Gupta et al., 2009) on the streamflow; (ii) the Nash-Sutcliffe efficiency (NSE) on the streamflow; (iii) the Nash-Sutcliffe efficiency (NSE) on the logarithm of the streamflow. The transferability of the parameters across resolutions was evaluated by sampling the parameters space and selecting the best parameter sets, called 'behavioural', for each resolution. Thus the 'transferability' across resolutions was defined as the percentage of agreement in these behavioural sets for each spatio-temporal resolution. They found that the parameters of the VIC model are largely transferable across spatial scales but hardly over temporal scales, especially when passing from daily to monthly scales. Their results show that "*with monthly data it is impossible to determine the optimal parameter set for the hourly or daily time step*". Model performance was significantly affected by both temporal and spatial resolutions. For the uniform forcing the spatially lumped model outperformed the higher-resolution models, while for the distributed forcing the 100-km^2 resolution model outperformed the others. The most significant changes of performance were found when passing from daily to monthly time steps. However their performance comparisons at different time steps are affected by the same limitations than those of the work of Littlewood and Croke (2008) (see discussion above). Still, Melsen et al. (2016b) recognized these limitations saying that: "*the monthly model results are simply an aggregation from the hourly model results, which might imply that the higher score on the monthly time step is the result of errors which compensate for each other, and that the model performance scores for the monthly time step are based on a considerable lower number of points*". Finally, we stress the interest of the change in performance that was found with respect to the different objective functions, in particular at monthly time steps.

1.4.3 Model structures

Rainfall-runoff models have been developed historically to perform simulation at fixed time-steps and many authors have developed specific models for different time-steps. This delineates today a fragmented field composed by a plethora of models at different time steps. Some authors use a same model structure for different time steps, without changing the model components (i.e. the equations) and their parameters in function of the time step. This is the case of some conceptual models: for example one can see the application of IHACRES at 4-hourly time step by Schreider and Jakeman (2001) or the application of HBV at hourly time step by Kobold and Brilly (2006). However, it is evident that this choice is generally suboptimal and leads to lower performance at the time steps for which the structure is less adapted. The construction of a *time-step adaptive structure* has not yet been presented in the literature of conceptual and empirical hydrological models. In the field of physically based models, rare cases of transposable models can be found as the AFFDEF model (Moretti and Montanari, 2007). Another approach is to use a probability-based model (Kandel et al., 2005) that consists of a same model structure working at different time resolutions by means of a statistical distribution of the fluxes within the time step supported by the available data. Kandel et al. (2005) developed a temporal scaling method to scale between process time scales (e.g. minutes) and typical measurement time scales (e.g. daily) using a *probability*

distribution function of the rainfall intensity within a day. This approach is analogous to the one followed in some models to represent the spatial variability of processes (e.g. the PDM and ARNO models).

1.4.3.1 Changing model complexity across time scales

There is still not a systematic knowledge on the *level of complexity* and on which model *components* are needed at different time steps, especially at sub-daily time scales (Kirchner et al., 2004). However, this question is essential in hydrological modelling, because the choice of a time-step has a direct impact on how to represent the input-output relationship. The model time step operates as a filter respect to processes dynamics, as the sampling interval of observations does, implying higher information content in higher resolution data. Since some hydrological processes are *masked* by the time step, it seems intuitive that the *model conceptualization requires a model structure of increasing complexity as the time decreases.* This is generally accepted and has been confirmed in the literature by different authors as for example by: Schaake et al. (1996), Ye et al. (1997), Jothityangkoon et al. (2001), Atkinson et al. (2002), Farmer et al. (2003) and Kavetski et al. (2011). Some key studies are summarized here below.

- Ye et al. (1997) evaluated the performance of *three conceptual rainfall-runoff models at the daily and monthly time step* in three low-yielding, ephemeral catchments in Australia. Their objective was to understand which level of complexity is required in conceptual models for predicting runoff in dry catchments. The models used are a simple conceptual model (GSFB; 8 parameters), a hybrid metric/conceptual model (IHACRES; 6 parameters), and a complex conceptual model, the Large Scale Catchment Model (LASCAM; 22 parameters). They showed that in these dry catchments passing from the monthly to the daily time step, a slightly more complex model (IHACRES or LASCAM) performs better than a simple model that is more adequate for monthly time steps. Their study also shows that a *better accuracy is obtained by calibrating models on daily data and integrating the outputs up to monthly rather than calibrating directly a model on a monthly time step.*

- Jothityangkoon et al. (2001) presented the formulation of a *water-balance model at annual, monthly and daily time scales* for a large semi-arid catchment in Australia by following a *downward approach* (Klemeš, 1983). Their first objective was to investigate the impact of some process controls (spatial variability of soil depths, rainfall and vegetation cover) on runoff variability at different time scales. As the time step decreases they found a gradual increase in the required model complexity. At the annual scale a simple bucket model was found adequate, including a saturation excess overland flow and multiple stores, connected in parallel, to reflect the spatial distribution of soil depths and rainfall. At the monthly scale, additional processes are required as the subsurface runoff, to introduce a delay mechanism in the runoff generation process, and a more detailed description of evapotranspiration. At the daily time scale, they found important to include non-linearity in the storage-discharge relationship relating to subsurface runoff, a deep groundwater store, and finally stream-flow routing in the river network.

- Atkinson et al. (2002) examined the choice of the *appropriate model complexity* in a downward approach for prediction of streamflow responses in 4 New Zealand catchments at *annual, monthly and daily time steps*. They showed that the simple single Manabe bucket water balance model (1 parameter) produces acceptable predictions of annual runoff, while at monthly and daily time steps gives progressively less accurate results. The authors defined a sufficient accuracy as predictions with a *good timing*, expressed by a *correlation coefficient > 0.8* and *good magnitude, i.e. ratio between mean annual observed and predicted runoff close to 1*. They increased progressively the complexity of the model by adding a subsurface runoff component, or *delayed flow*, expressed through a non-linear storage-discharge relationship (3 more parameters) that allows obtaining a reasonable accuracy for the monthly time scale. This is still not sufficient for the daily time scale, especially during low flows, because of the presence of a persistent base flow component throughout the year that is not captured by the model. This is solved by a revised daily model including a base flow discharge as a *linear function* of the bucket storage (1 more parameter) that improves the prediction of low-flows, though it changes slightly the performance with respect to the timing and magnitude of the predictions. By means of a sensitivity analysis, they showed that with decreasing time scales the model becomes increasingly more sensitive to the estimated parameters, with increasingly poorer confidence in the predictions. A qualitative relationship of the model complexity expressed as function of a measure of climatic wetness (*dryness index*) and the time-step is suggested, i.e. *required model complexity increases with decreasing time scale and increasing dryness index*.

- Kavetski et al. (2011), in their work that we have already summarized for the parameters issue (see Section 1.4.2), showed that parameters related to *catchment's quick response* to rainfall become progressively *better identified as the data resolution is refined*, thus *supporting additional model complexity* (i.e. transfer functions, non-linear reservoirs). The motivation is also intuitive: since their study catchment has a sub-hourly characteristic time scale, the averaging of observed data above hourly scales smears this quick-response feature of the catchment. This makes the quick-flow components of the model appear progressively non-identifiable, or redundant, as the time step increases. Moreover, they showed the effect of the model structure on the time scale dependencies of parameters. For example, they noted that some parameters that are common to different model structures of increasing complexity are highly scale dependent in simpler models but become progressively more stable in more complex model structures. In other words, the presence of structural errors in exceedingly simple models may be compensated by spurious time scale dependencies of the parameters.

As for the general approach that can be followed to investigate the model complexity dependency on the time step, we remind that *linking model conceptualizations across scales can follow either an upward or a downward route*, according to the direction along the hierarchy of scales (Klemeš, 1983). Some examples of application of the downward approach to derive model structures at different time scales can be found in Sivapalan et al. (2003), Jothityangkoon et al. (2001) and Eder et al. (2003). The essence of the downward analysis is

that model complexity evolves from the simplest form at the largest time step to its final quasi-distributed framework with multiple storages and process interactions in response to inadequacies in signature prediction at finer time steps (Farmer et al., 2003).

1.4.3.2 The emblematic case of the GR conceptual models chain

In the field of lumped conceptual models, a clear example of the *progressive complexification* of the models structure *as the time step becomes finer* is the GR models chain developed at Irstea by different authors (e.g. Michel, 1983; Edijatno et al., 1999; Perrin, 2000; Perrin et al., 2003; Mathevet, 2005; Mouelhi et al., 2006b; Mouelhi et al., 2006a; Le Moine et al., 2007; Le Moine, 2008; Pushpalatha et al., 2011). These models have been developed for different specific time-steps, up to date: *annual (GR1A), monthly (GR2M), daily (GR4J, GR5J, GR6J) and hourly (GR4H, GR5H).* They are the result of more than 30 years of progressive improvements, starting from a simple lumped model at the daily time step. Thus the development of the models for the other time steps has followed both the *upward and downward routes*. The actual daily model structures at four, five or six parameters (GR4J, GR5J, GR6J) derive from a first 1-parameter structure (Michel, 1983) and the progressive improvements by Edijatno (1991), Edijatno et al. (1999), Perrin (2000), Mathevet (2005), Le Moine (2008) and Pushpalatha et al. (2011). This long development of the model has followed a *progressive increase in the complexity of the structure* that has been *accepted only if justified by a significant improvement in the performances* proved *over large catchment samples* (consisting of hundreds of catchments).

Figure 1.10 shows four of the selected models for the multi-annual, annual, monthly and daily time steps: the Turc equation (Mouelhi, 2003), the GR1A (Mouelhi et al., 2006a), GR2M (Mouelhi et al., 2006b) and GR4J (Perrin et al., 2003) models. It clearly shows the *complexification* of the structures as the time step is shortened. Table 1.5 summarizes the different complexities for these four models, two other daily versions (GR5J and GR6J) and the hourly models (GR4H and GR5H). Note that the hourly models share almost the same structures than their corresponding daily versions (GR4J and GR5J), thus one may refer to the daily structure represented in Figure 1.10. Some comments about this progressive *complexification* follow:

- At the multi-annual time step the streamflow is only function of the inputs, P and E, at the same time step (see Turc equation).

- At the annual time step it is introduced a dependency on the previous inputs taking into account the ***antecedent annual rainfall***. This is a more simple way than adding a soil-moisture accounting storage or a transfer component and it is resulted sufficient at the annual step (Mouelhi et al., 2006a).

- The necessity of a ***soil-moisture (or production) reservoir*** and a ***transfer component*** emerges at the monthly time step with also the need of an additional linear function expressing the ***underground exchanges*** of the basin. The production reservoir evolution is described by a power law of its storage (derived by integration over the

time step of its quadratic differential equations). At the monthly time scale, the transfer function involves just a ***quadratic routing reservoir***.

- Then, at the daily time step, while the soil moisture part rests unchanged, the ***transfer function*** of the model becomes *more sophisticated* including two ***unit hydrographs***. The water that reaches the routing functions is partitioned into *two distinct fluxes*. While the most part (90%) is routed by a UH and then by the non-linear routing store, a part (10%) bypasses the routing reservoir and is routed directly by a UH (Perrin et al., 2003). A more complex *groundwater exchange function* is applied to both components and is described by a power law of the routing reservoir level.

- The daily model has been *adapted to the hourly time step* by Mathevet (2005) and Le Moine (2008) following a *top-down approach.* The hourly structure was developed by empirical analysis (based on a trial-and-error process and evaluation of model performance), starting from the daily structure (GR4J), and testing many possible modifications on the model using a sample of 313 French catchments. Mathevet (2005) questioned the original daily structure at the hourly time step by: (1) *changing the fixed parameters of the models*, (2) *complexifying the routing function of the model*, and (3) *complexifying the production function of the model*. Only the first strategy succeeded, by *changing the parameters that theoretically depend on the time step* (see Section 1.4.2) and the *UH exponent value.* The UH exponent was lowered, leading to a more crushed form, as it is reported in Figure 1.11 from Mathevet (2005). The *complexification strategies* (2 and 3) failed. These strategies included many variations, as for example: changing the emptying laws of the two reservoirs; adding some new reservoirs in series or parallel; changing the form of the UH, or adding a dependency between the exponent of the UH, its base time and the production reservoir level. All these tests showed that no significant improvement was gained and *the daily structure complexity was judged sufficient also for the hourly time step,* which may be a counterintuitive finding. Moreover, Mathevet (2005) even tried to simplify the structure, finding that the use of only one UH with two branches does not modify the performance, simplifying the structure. Le Moine (2008) confirmed these conclusions, by also testing different addition points for the water exchanges fluxes and more options for the UH form. Le Moine (2008) (see p. 200) suggests two possible reasons for this *complexification-failing*: (i) indeed, "*the representation of the rainfall-runoff transformation at the hourly time step does not need to be more complex than at the daily time step, the entire 'added value' would come from the richer information in the hourly rainfall signal"*; (ii) the *identifiability* of the parameters would be more difficult at the hourly time step, because of the "*simultaneous identification of parameters simulating a wide range of frequencies*". We think that both reasons are plausible and deserve further researches.

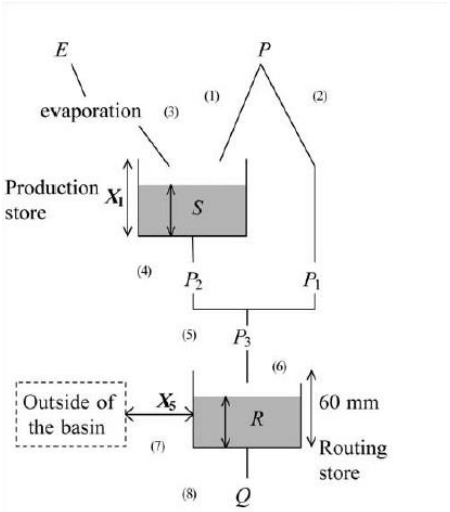
Multi-annual time step *(Turc eq., in Mouelhi, 2003)*

$$Q = P\left\{1 - \frac{1}{\left[1 + \left(\frac{P}{E}\right)^2\right]^{0.5}}\right\}$$

Annual time step *(GR1A, Mouelhi et al., 2006a))*

$$Q_k = P_k\left\{1 - \frac{1}{\left[1 + \left(\frac{0.7P_k + 0.3P_{k-1}}{X1.E_k}\right)^2\right]^{0.5}}\right\}$$

Monthly time step *(GR2M, Mouelhi et al., 2006b)*



(1) $S_1 = \dfrac{S + X_1\varphi}{1 + \varphi\dfrac{S}{X_1}}$ with $\varphi = \tanh\left(\dfrac{P}{X_1}\right)$

(2) $P_1 = P + S - S_1$

(3) $S_2 = \dfrac{S_1(1-\psi)}{1 + \psi\left(1 - \dfrac{S_1}{X_1}\right)}$ with $\psi = \tanh\left(\dfrac{E}{X_1}\right)$

(4) $S = \dfrac{S_2}{\left[1 + \left(\dfrac{S_2}{X_1}\right)^3\right]^{1/3}}$ $\quad P_2 = S_2 - S$

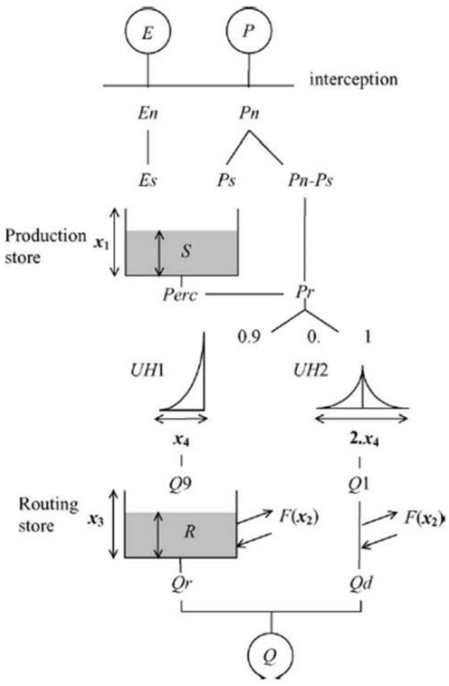
(5) $P_3 = P_1 + P_2$

(6) $R_1 = R + P_3$

(7) $R_2 = X_5.R_1$

(8) $Q = \dfrac{R_2^2}{R_2 + 60}$ $\quad R = R_2 - Q$

Daily time step *(GR4J, Perrin et al., 2003)*



If $P \geq E$, then $P_n = P - E$ and $E_n = 0$ (1)

otherwise $P_n = 0$ and $E_n = E - P$ (2)

$$P_s = \frac{x_1\left(1 - \left(\frac{S}{x_1}\right)^2\right)\tanh\left(\frac{P_n}{x_1}\right)}{1 + \frac{S}{x_1}\tanh\left(\frac{P_n}{x_1}\right)} \quad (3)$$

$$E_s = \frac{S\left(2 - \frac{S}{x_1}\right)\tanh\left(\frac{E_n}{x_1}\right)}{1 + \left(1 - \frac{S}{x_1}\right)\tanh\left(\frac{E_n}{x_1}\right)} \quad (4)$$

$S = S - E_s + P_s$ (5)

$$\text{Perc} = S\left\{1 - \left[1 + \left(\frac{4}{9}\frac{S}{x_1}\right)^4\right]^{-1/4}\right\} \quad (6)$$

$S = S - \text{Perc}$ (7)

$P_r = \text{Perc} + (P_n - P_s)$ (8)

For $t \leq 0$, $\text{SH1}(t) = 0$ (9)

For $0 < t < x_4$, $\text{SH1}(t) = \left(\frac{t}{x_4}\right)^{5/2}$ (10)

For $t \geq x_4$, $\text{SH1}(t) = 1$ (11)

For $t \leq 0$, $\text{SH2}(t) = 0$ (12)

For $0 < t \leq x_4$, $\text{SH2}(t) = \frac{1}{2}\left(\frac{t}{x_4}\right)^{5/2}$ (13)

For $x_4 < t < 2x_4$, $\text{SH2}(t) = 1 - \frac{1}{2}\left(2 - \frac{t}{x_4}\right)^{5/2}$ (14)

For $t \geq 2x_4$, $\text{SH2}(t) = 1$ (15)

$\text{UH1}(j) = \text{SH1}(j) - \text{SH1}(j-1)$ (16)

$\text{UH2}(j) = \text{SH2}(j) - \text{SH2}(j-1)$ (17)

$$F = x_2\left(\frac{R}{x_3}\right)^{7/2} \quad (18)$$

$R = \max(0; R + Q9 + F)$ (19)

$$Q_r = R\left\{1 - \left[1 + \left(\frac{R}{x_3}\right)^4\right]^{-1/4}\right\} \quad (20)$$

$R = R - Q_r$ (21)

$Q_d = \max(0; Q1 + F)$ (22)

$Q = Q_r + Q_d$ (23)

FIGURE 1.10 – GR models chain at multiannual, annual, monthly and daily time steps.

Time-step	Multiannual	Annual	Monthly	Daily			Hourly	
Model	GR0P	GR1A	GR2M	GR4J	GR5J	GR6J	GR4H	GR5H
Number of free parameters	0	1	2	4	5	6	4	5
Number of reservoirs	0	0	2	2	2	3	2	2
Underground exchanges	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Unit hydrographs (n.)	No	No	No	Yes (2)	Yes (1)	Yes (2)	Yes (1)	Yes (1)

TABLE 1.5 – Differences between the structures of the GR models at different time-steps, from multiannual to hourly time step. Table inspired from Mathevet (2005), modified from his table 6.1 (page 205).

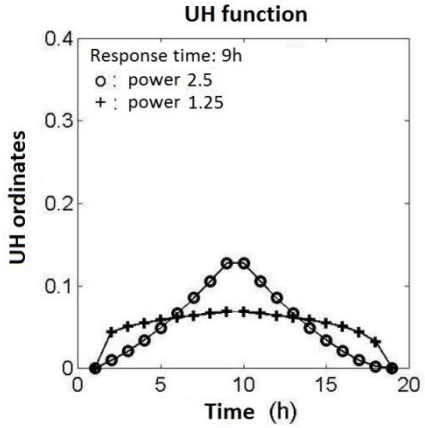


FIGURE 1.11 – UH form for two values of the exponent of its power law: power=2.5, used in the daily model GR4J, and power=1.25 selected for the hourly model GR4H (from Mathevet (2005)).

To conclude, although the GR models have structural affinities at different time steps, sharing some identical functions, especially for monthly, daily and hourly time steps, an exhaustive explicit link between them has not been clearly established and some questions are still open. For example the parameters dependency on time steps should be tested on a *more continuous temporal domain*. Moreover these models at different time steps should be further validated in a *multi-criteria framework* (flood, regime, low flows) and considering also their *inner consistency* (Mouelhi, 2003; Mathevet, 2005; Le Moine, 2008). Thus, we will start from these similar models at different time steps and we will try to validate their current state on a wider range of time steps, especially ***from the daily time step downwards***, because of the primary operational interest of these time scales. By testing a larger set of time steps we will be able to derive a continuous relationship between the time steps and the model parameters and structures. The time step could be used as a sort of fixed-input of the model that may change continuously some other parameters and components (e.g. some function forms, as the UH function) and activate (or switch off) some components as the time step changes. In this direction, new investigations must be done, for example on the *consistency of the state variables and internal fluxes* of the models as the time step changes, topic that has not yet been investigated in the literature.

1.5 Synthesis and resulting research objectives

In this chapter, we have focused our review on rainfall-runoff modelling and provided a synthesis of the general modelling framework, distinguishing *eight different time concepts* that are not sufficiently well recognized in the literature. We have thoroughly discussed the classification of different types of models, by focusing more on the underlying *hypotheses on temporal and spatial scales* than what is generally done. We have argued that this focus is essential for a meaningful and *process-based* conceptualization. Then we have discussed the general principles of the spatial and temporal scaling issues in hydrology, focusing in more details on the following aspects involved in the temporal scaling of conceptual hydrological models: (i) the relationship between process and modelling scale; (ii) the dependency of model parameters and performance on time step; (iii) the impact of temporal resolution on model structures. In each of these aspects, some advances have been made, but more research is needed to improve our qualitative and quantitative understanding on these issues. In this study, we will try to consider all these aspects together with their interplay, by testing *different model structures at different time steps* (starting from the cited GR4J and GR4H models) on a *large set of catchments*. The key issues that will deserve attention in our work can be summarized in the following points.

The importance of the time-distribution of rainfall for streamflows simulation

Absolutely central to the time scale issue in rainfall-runoff modelling is the importance of the time-distribution of rainfall for runoff simulations. Nowadays, we can better explore the crucial significance of the ***rainfall sub-daily variability*** in conceptual hydrological modelling, thanks to the increased computer power and *increased availability of observations at sub-hourly time steps* (from different sources, e.g. rain gauges and radars). We think that the research on the temporal variability in hydrological modelling is more urgent than the spatial variability for different reasons, as for example: (1) for data availability, because, typically, space resolutions are still poorer than time resolutions in hydrology (Blöschl and Sivapalan, 1995); (2) for a large sensitivity of runoff response to precipitation temporal variability, that is possibly more important than spatial variability as highlighted by some authors (e.g. Krajewski et al., 1991; Ishidaira et al., 2003; Paschalis et al., 2014); (3) the fact that, despite its importance, this temporal scaling issue in hydrological modelling has been the object of a smaller number of case studies respect to the spatial scaling issue. Given these considerations, we will investigate the ***impact of model time step on model performance*** by using rainfall data at short sub-daily time steps. Our model assessment will be based on a ***multi time scale approach***, by evaluating model outputs at different time steps, larger than the model functioning one. This approach, often neglected in the literature, seems to us a valuable and demanding validation test for model *consistency* and *temporal transposability*.

Model parameters dependency on temporal resolution

Some authors have explored the issue of model parameters-time scale dependency, providing some insights on the ***impact of calibration data time step*** on the estimated parameters of conceptual models. A general consensus exists that *most model parameters are time scale*

dependent (Littlewood and Croke, 2008; Wang et al., 2009; Littlewood et al., 2011; Bastola and Murphy, 2013; Littlewood and Croke, 2013). However, despite all these studies, as Kavetski et al. (2011) note, *there is still insufficient quantitative understanding about the precise underlying causes of these dependencies, their mathematical representations and physical interpretation.* Current treatments and analyses of parameters scaling are largely heuristic and empirical (e.g. Littlewood and Croke, 2008; Wang et al., 2009; Bastola and Murphy, 2013) and only sometimes based on a sound mathematical basis (integration of the model equations), as in Le Moine (2008). In this context, one objective of our research work will be to further investigate the models parameters dependency on the calibration data time step and try to explicit and understand the conceptual basis of this dependency. This work could help to improve the physical interpretation of calibrated model structures and elucidate their connections to catchment attributes, facilitating their regionalization to ungauged catchments (e.g. Wagener and Wheater, 2006; Bárdossy and Singh, 2008).

Model structures

Some works have shown that the ***complexity in model structure*** increases as the time step decreases, as for instance Atkinson et al. (2002) and Farmer et al. (2003). The GR models chain is an example of set of specific models developed for different time steps, with some similarities and increasing complexity passing from the monthly to the daily time step (Mouelhi, 2003). However, in contrast to expectations, a complexification of the GR model structure has not been proofed to be necessary at the hourly time step (Mathevet, 2005; Le Moine, 2008).

In our work, we will continue the development of the GR models at different time steps, and search for the actual limits in the temporal transferability of these model structures. For this, we will start from the simple GR4J daily model (Perrin et al., 2003) and follow a downward scaling approach for model development at multiple time steps. In this direction, new investigations should be done for analysing the ***internal coherence of the model***. This will imply the analysis of the impact of time step not only on model parameters and performance, but also on the ***consistency of the state variables and internal fluxes*** of the models as the time step changes. The time step will be an input of the model that will adapt its structure and parameters as the time step changes. The explicit statement of the time step size in the model equations will allow to easily applying the same model to data with different time steps, as promoted by Kavetski and Clark (2011).

2
Material and methods

2.1 Introduction

In this chapter we present the hydro-climatic data set used for this thesis. In order to develop robust and general models, it is necessary to work on a wide range of hydro-climatic conditions and a variety of hydrologic environments. Thus, we built a large catchment data base, as promoted by previous authors, as for example by Perrin et al. (2001) and Andréassian et al. (2009) among others. This scientific approach is frequently adopted in the "Catchment-Hydrology" team at Irstea (Antony), where this work was carried out. One may see some previous PhD theses carried out at Irstea, such as Le Moine (2008) and Lobligeois (2014).

In this chapter, we present in details our catchment selection procedure and the characteristics of the large and varied set of 240 mesoscale French catchments selected.

Finally, we briefly present the GR4 simulation model, the calibration-evaluation procedure and the main evaluation criteria that will be used in this thesis.

2.2 Building-up the hydro-climatic database

2.2.1 Precipitation data

The precipitation data series available for our work consist of the following products at different time steps.

i. The **catchment precipitation time-series at the daily time step** were constructed at Irstea for **3701 French catchments** by Bourgin et al. (2010) starting from the SAFRAN reanalysis developed by Météo-France (Vidal et al., 2010). This is a reanalysis of surface observations combined with data from analysis of meteorological models estimates of climate variables (liquid and solid precipitation, temperature, humidity, wind and solar radiation). The SAFRAN data, originally available at 8-km-resolution, were aggregated at the catchment scale for the French metropolitan territory. This database is constantly updated and currently covers the period 1/8/1958 - 31/7/2013, without gaps. The catchment rainfall series provide the cumulated rainfall for each day *i*, between 6h UTC of the same day and 6h UTC of the day *i+1*.

ii. Rain gauge data of **precipitation at 6-minute time step** (Météo-France) are available for 1622 automatic rain gauges in Metropolitan France (1405 stations) and French Overseas Departments and Territories (217). The time-series cover the period 2005-2013 (starting and ending at different dates depending on the station), with gaps. The 6-minute rainfall data for a given time step t_i corresponds to the cumulated rainfall between instants t_i-Δt (Δt=6 minutes) and t_i (for example, the data assigned to 0:06 h corresponds to the rainfall cumulated between 0h and 0:06 h UTC).

iii. The **catchment precipitation time-series at the hourly time step** were constructed by Lobligeois (2014) for 1133 French catchments from the COMEPHORE reanalysis developed by Météo-France (Guéguen et al., 2011). This is a reanalysis at 1-km spatial resolution of surface observations from *rain gauges* at hourly and daily time steps combined with data from *radar* measures at 5-minute time step. The time-series cover the period 1997-2006, without gaps. The hourly rainfall data for a given time step t_i corresponds to the cumulated rainfall between instants t_i-Δt (Δt=1h) and t_i

Because of our interest in testing a wide range of model time steps, up to sub-hourly, we decided to use only products (i) and (ii) to construct the rainfall time-series that we will use in our modelling tests. Note also that product (iii) is available for a more limited number of catchments than (i), while product (ii) is made of punctual observations over the whole Metropolitan French territory.

Then, by using product (ii), the maximum time coverage of our modelling tests will be 8 years (from 2005 to 2013). The hourly rainfall time-series (iii) will be used, in combination with the catchment rainfall time-series at the daily time step (i), to construct rainfall series for some years before 2005 used as a warm-up period. Product (iii) is used also for a critical assessment of the quality of product (ii) by analysing the consistency of the two products (see Appendix B). Since the catchment daily precipitation time-series (i) are available for a longer period and without gaps, the constraints on data availability will concern only 6-minute rainfall data and streamflow data.

2.2.1.1 Analysis of the quality of the 6-minute data

Here we briefly present the analysis of the 6-minute precipitation database available for the French Metropolitan territory.
Figure 2.1 shows the geographical distribution of the **1405 rain gauges in Metropolitan France** and the corresponding Thiessen polygons. The spatial distribution of these rain gauges is not homogeneous over the French territory. The network is very dense in the Paris region and in areas with higher spatial variability of precipitation fields (Cevennes, Vosges, Var, Alpes-Maritimes, and Rhône-Alpes regions) and sparser elsewhere.

The average area of the Thiessen polygons (also called *average rain gauges area*, see section 2.3.2) of the 1405 rain gauges at 6 minutes time step on the French metropolitan territory is about 487 km^2, i.e. there is ***one rain gauge every 487 km^2 on average***.
This density is approximately the same than for the hourly rain gauges network (459 km^2), and more than 3 times less than the daily network (136 km^2), according to the analyses by Lobligeois (2014) on the latter two networks. Even the range of the Thiessen polygons areas is almost the same than for the hourly network, being:

- 16 - 1650 km^2 for the 6 minutes pluviometers network,
- 6 - 1275 km^2 for the hourly pluviometers network (Lobligeois, 2014);
- 2 - 440 km^2 for the daily pluviometers network (Lobligeois, 2014).

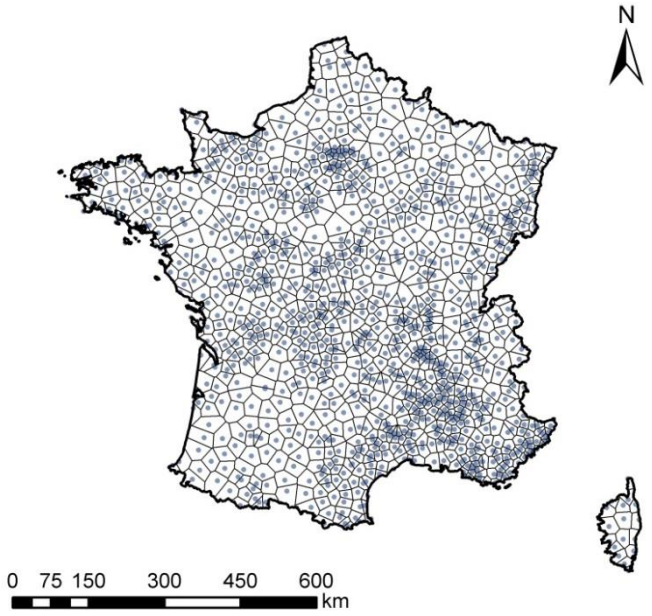


FIGURE 2.1 - Network of 1405 rain gauges stations at 6 minutes sampling intervals in Metropolitan France and correspondent Thiessen polygons.

For most rain gauges, the 6-minute time series are **8-year long**, starting in 2005 (from June to August) and ending in 2013 (August). More than 80% of the 1405 rainfall time series begin before the 1st July 2005 and terminate after the 13th August 2013. So, for our modelling work we decided to consider the period **1/8/2005 - 31/7/2013**, including 8 years (i.e. 2922 days or 701280 intervals of 6 minutes), to have a common test period for all the basins for practical reasons. This 8-year test period is judged sufficiently long to allow applying a model calibration-validation procedure, in line with previous studies such as Perrin et al. (2008) and Merz et al. (2009). We can add some previous years to the test period 2005-2013, as a warm-up period, in order to set the initial conditions for the model simulations in calibration and validation periods. To obtain the rainfall data for the years before 2005, for sub-daily time steps, we decided to either disaggregate the daily data using a uniform temporal distribution over the day or use the hourly data from the COMEPHORE reanalysis when this is available.

The quality of the 6-minute data was checked by cumulating and comparing these data with the rain gauges measurements at the hourly time step reconstructed from the COMEPHORE reanalysis by Météo-France for the corresponding pixels (see point (iii) in Section 2.2.1). This analysis of data coherence was performed at 257 rain gauges operating over the period July 2005 – December 2006 that is common between the two databases we have. This analysis confirmed a good consistency of the two products, as it is shown in Appendix B, providing some summary statistics about this comparison.

2.2.1.2 Constitution of catchment rainfall series at time-intervals from sub-hourly to daily

Given the two rainfall databases available at daily and 6-minute sampling times (section 2.2.1, points (i) and (ii)) we can derive catchment rainfall time-series at all time steps from 6 minutes upwards.

For sub-daily intervals, **we temporally disaggregate the daily precipitation series** (see Section 2.2.1, point (i)) **using the information of the sub-hourly distribution contained in the data of the 6-minute resolution rain gauges** (see Section 2.2.1, point (ii)). One may wonder why we did not use directly the 6-minute data to calculate the catchment sub-daily rainfall estimates instead of disaggregating the daily precipitation series. There are two reasons for this choice: firstly, the 6-minute rain gauges network is much less dense than the daily network (at least 3 times less dense, as already discussed in Section 2.2.1.1). Furthermore, the spatial correlation of the precipitation at 6-minute time steps decreases with the distance more quickly than the spatial correlation of daily precipitation. The dependence of the correlation function of rainfall on the time scale has been discussed by a number of authors (e.g. Zawadzki, 1973; Obled et al., 2009). A key measure in this context is the so-called *decorrelation distance*, i.e. distance for which the correlation decreases to *1/e* of its maximum (Rubel, 1996). Considering the accumulation time of precipitation, typical decorrelation distances are 10 km for 15 minutes (Zawadzki, 1973), 50 km for hourly (Rubel, 1996), 200 km for 12-hourly and daily (Rubel, 1996) and 300 km for monthly (Berndtsson, 1987) accumulated values of precipitation. For these reasons the spatial interpolation of 6-minute rainfall data is more hazardous than that of daily data. Thus, to avoid problems of water balance, we take as daily accumulated precipitation the values given by the SAFRAN reanalysis and **use only the sub-daily temporal distribution given by the 6-minute network**. This choice is analogous than that of Le Moine (2008).

The temporal disaggregation of the catchment daily rainfall is operated by calculating a distribution function (*DF*) of the 6-minute rainfall data over each day, as described in more details below. As for sub-daily time intervals, we decided to consider only intervals multiple of 6 minutes, otherwise we should distribute the precipitation making arbitrary assumptions about the sub-hourly distribution that are not supported by information in the data and then adding uncertainty. The procedure followed for the temporal distribution at sub-hourly intervals involves the following steps:

1. We determine the *N* rain gauges of the 6-minute network that influence each basin *b*, by **intersection of the Thiessen polygons** (Figure 2.1) **with the catchment area**. For each Thiessen polygon *k* we calculate its percentage of coverage p_k on the basin.

2. For each basin *b* we calculate the **areal rainfall estimates** $P_b^{6-\min}$ **at 6-minute time steps** t_i as an average of the 6-minute rain gauges data weighted by the percentages of coverage of the correspondent Thiessen polygons on the basin, as:

$$P_b^{6-\min}(t_i) = \sum_{k=1}^{N} P_k\,(t_i) \cdot p_k(t_i)$$

where $P_k(t_i)$ is the cumulated rainfall between time step t_i-*6'* and t_i and $p_k(t_i)$ is the weight of the station k, that depends on the time step, since for some time steps one or more station can be not operational (see discussion below).

3. From $P_b^{6-\min}$ we calculate the **discrete distribution $DF_{b,d}$ of the catchment rainfall series at 6-minute time steps for each day** *d*, from 6h UTC of day *d* to 6h UTC of the next day *d+1*, that is then used to disaggregate the daily rainfall up to 6-minute time resolution, as:

$$P_b(t_i) = DF_{b,d}(t_i) \cdot P_b^{Safran}(d)$$

where t_i is a sub-daily time step on day *d*.

Note that if one (or more) station(s) *k* is (/are) not operational on a certain time interval, its (/their) weight p_k is uniformly redistributed among the other stations influencing the basin. If no data of the 6-minute stations influencing a basin is available for a certain time interval Δt, we search for the nearest station *k'* of the 6-minute network that is in operation on that time interval Δt and we use its measures with a weight $p_{k'}(\Delta t)=1$ (only in this case, otherwise its weight is always zero). These cases are limited by considering only basins located in an area with a high density of the pluviometers network (see criteria explained in Section 2.3.2). In fact, for our final catchment sample of 240 basins, this absence of operating pluviometers happens for 133 basins for an average time interval (of 'absence' of pluviometers) of less than 5 days and a maximum of 45 days over the total study period of 2922 days.

2.2.1.3 Analysis of the goodness of areal rainfall estimates derived from 6-minute rain gauges at the daily time step

In order to provide a first evaluation of the goodness of the areal rainfall estimates derived from spatial aggregation of the sole 6-minute data, we compared the average areal rainfall time series at 6-minute time steps ($P_b^{6-\min}$, see point 2 of previous section), cumulated at the daily time step, with the daily rainfall series derived from the daily SAFRAN reanalysis (P_b^{Safran}, see point (i) in Section 2.2.1). Note that this comparison aims at assessing the coherence of two independent estimates of catchment daily rainfall, the first being based on the 6-minute rain gauges data only, and the second on the daily reanalysis. The objective of this analysis is to validate the goodness of the estimates derived from the 6-minute rain gauges only and so the quality of the original 6-minute data set. In fact, we can assume that the daily reanalysis is the reference areal rainfall at the daily time step, since it is the best estimate available for the two reasons already discussed in Section 2.2.1.2 (higher density of the daily rain gauges network and higher spatial correlation at the daily time step respect to 6-minute step).

To quantify the goodness of areal rainfall estimates, we used two indexes proposed by Andréassian et al. (2001): the *GORE* and the *BALANCE* indexes.

The ***Goodness of Rainfall Estimation* (*GORE*) index** is the transposition of the Nash-Sutcliffe criterion (Nash and Sutcliffe, 1970b) in the rainfall domain, and compares the sum of the squared errors in the rainfall estimates P_b^E (in our case $P_b^{6-\min}$) for a basin b to the temporal variance of the reference areal precipitation P_b^{ref} (in our case P_b^{Safran}). It is computed with the square roots of the rainfall data (to reduce the weight of extreme events), as:

$$GORE = 1 - \frac{\sum_{i=1}^{n}\left(\sqrt{P_{b,i}^{E}} - \sqrt{P_{b,i}^{ref}}\right)^2}{\sum_{i=1}^{n}\left(\sqrt{P_{b,i}^{ref}} - \overline{\sqrt{P_b^{ref}}}\right)^2}$$

where n is the number of time steps of the period of analysis and $\overline{\sqrt{P_b^{ref}}}$ is the mean of the square root of the reference precipitation over the study period. The GORE index can vary between $-\infty$ and 1. Its maximum value is reached when the estimated rainfall equals the reference rainfall; otherwise the index decreases as the estimates become poorer.

The ***BALANCE* index** compares the sum of estimated rainfall to the sum of reference rainfall over the analysis period, and is defined as:

$$BALANCE = \frac{\sum_{i=1}^{n} P_{b,i}^{E}}{\sum_{i=1}^{n} P_{b,i}^{ref}}$$

The *BALANCE* index is greater than 1 in case of rainfall overestimation, and smaller than 1 in case of underestimation.

These two easy-to-interpret indexes describe the quality of the rainfall time distribution and of the total depth, which are very important aspects in hydrological modelling.

In Figure 2.2, we report the empirical cumulative distribution of GORE index values for the 240 catchments of our final sample (see Sections 2.3 for the selection process). The daily temporal distribution of areal rainfall estimates derived from the 6-minute data corresponds very well to that of the reference data (derived from the SAFRAN reanalysis data). In fact, the median value of GORE index is 0.94 and the 10th percentile is 0.90. Only for two basins of the sample the GORE index is below 0.82, being the worst value 0.43 and the second worst 0.78.

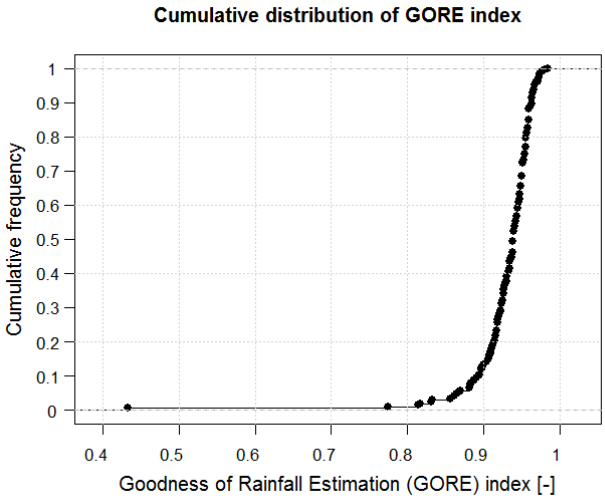


FIGURE 2.2 – Distribution of the Goodness of Rainfall Estimation (GORE) index for 240 catchments of our sample.

In Figure 2.3 we report the empirical cumulative distribution of BALANCE index values for the 240 catchments of our final sample.

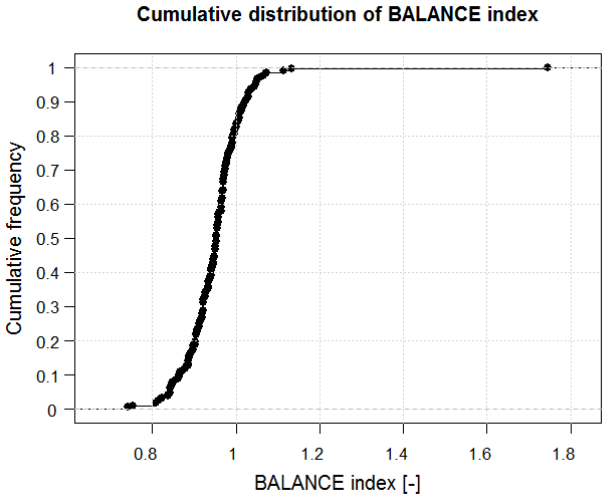


FIGURE 2.3 - Distribution of the BALANCE index for 240 catchments of our sample.

The areal rainfall estimates derived from the 6-minute data tend to underestimate the daily catchment rainfall with respect to the reference data, being the 80% of the BALANCE index values below 0.99. However this underestimation is limited: the minimum value of BALANCE index is 0.74 and the 10^{th} percentile is 0.87. The worst value of BALANCE index is 1.75, which corresponds to an overestimation of 75% of the total cumulated rainfall over the whole study period. This value concerns the basin of our sample with the worst value of GORE index too (0.43). This basin (i.e. the Tarnon River at Florac, code 'O3064010') seems

to be suffering from the problem of having rain gauges not operating for a great portion of the study period, since 2 of the 3 pluviometers influencing the catchments are inoperative for 25% of the time.

For the most part of our catchment sample, the GORE and BALANCE indexes validate the goodness of the daily areal rainfall estimates derived from the 6-minute data, as for 99% of our sample the indexes fall in a confidence band of 20% around the optimal value of 1. Only for a few basins there are bad outliers for the indexes of quality of rainfall estimates. However we decided to keep in our sample also these outliers, to evaluate the impact of the quality of rainfall estimates on model performance. Moreover, the uncertainties derived from the spatial aggregation of 6-minute rainfall data will concern the sub-daily temporal distribution of rainfall and not the accumulation of daily rainfall, due to our procedure of disaggregation of the daily precipitation series.

2.2.2 Streamflow data

Streamflow measures at variable sampling time were available at **1859 hydrometric stations in France** from the "***Banque-Hydro"* data base**, under HYDRO2 (SCHAPI, 2014), thanks to the extraction performed at Irstea (UR HHLY) in December 2013. These data cover our test period 2005-2013 and correspond to a half of the 3701 French catchments for which daily rainfall data are available. For these reasons, and given the time required to operate the extraction under HYDRO2 of flow time series at variable time steps, we considered that this data set was sufficient and it was not necessary to extract a large amount of additional data for other stations (only some additional data for few stations have been extracted to complement this data base).

As for the **time standard of these time-series**, we discovered that each gauging station has its specific time reference history, since we found in the metadata ('comments' or 'events' field) that for some of the 1859 stations a change of time standard reference occurred during our study period. However, the hydrometric services have not always kept track of the time standard followed and eventual changes occurred during the time series, in the metadata of the centralized Banque-Hydro archive. So we asked directly the time standards information to the individual services that manage the hydrometric stations of our data base: 21 DREALS, 2 SPC, one research centre (Irstea) and one regional Parc. The collected elements (although presenting a certain degree of uncertainty) allowed us to **reprocess the data** extracted from the *Banque Hydro* to put them in a **uniform time standard reference**, the **Coordinated Universal Time (UTC)**. This was judged necessary to avoid potential biases due to the time standard changes occurred during our study period at a great part of stations. These changes involve two or even three time standards (*winter time*, *summer time* and *UTC*) and so lead to systematic biases in the streamflow series caused by one hour or two of artificial time lag in the temporal reference.

2.2.2.1 *Statistical characterization of the temporal resolution of streamflow series at variable time steps*

Given our interest to work at multiple temporal scales, including sub-hourly scales, here we analyse the temporal resolution of the streamflow series at variable time steps to see whether they have sufficient information content. The question arises because the hydrometric gauging stations are configurable and each hydrometric service operating a set of gauging stations in France may set up a different configuration. The time step can be lowered up to the minute but it is usually larger to avoid not affecting the autonomy of field devices (*personal communication*). In particular, the variability of the sampling time intervals depends on the variability of the river flows, and instruments are configured to operate more frequent measure recordings when the flow varies more quickly. For this reason, our expectation is to have time steps as small as possible and in any case at least less than one hour for the part of the streamflow series with the fastest variations (i.e. the high-flows periods, which generally cover a small portion of the series). Figure 2.4 shows the cumulative distribution function of the 10% quantile of the variable sampling intervals of flow series for the 240 catchments of our sample. Given that the median of this distribution is 28 minutes and the 90th percentile is 60 minutes (i.e. for 90% of our stations the 10%-quantile of the variable sampling intervals is less than 1 hour), the temporal resolution of our data set satisfies our expectation. Note also that the 10%-quantile of the time steps of streamflow series is greater than 2 hours only for two catchments in the sample.

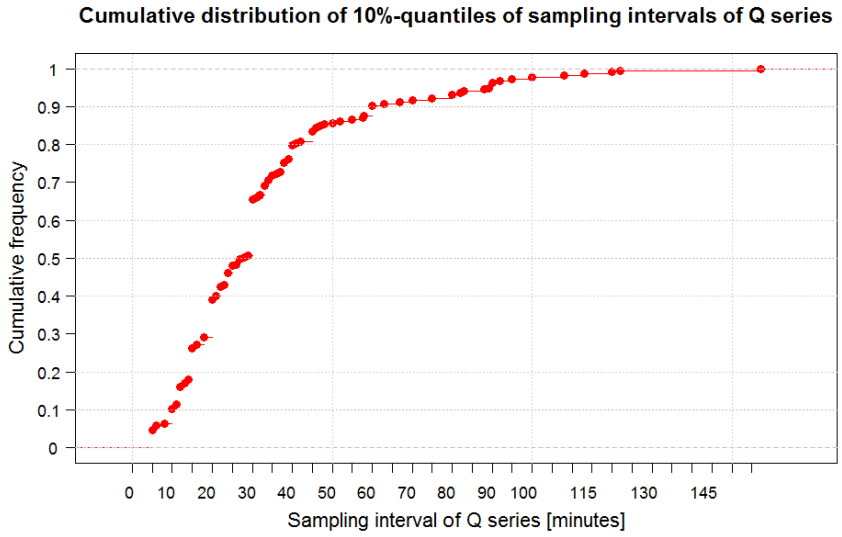


FIGURE 2.4 - Distribution of the 10% quantile of the variable sampling intervals of flow series for the 240 gauging stations corresponding to the catchments selected in our final sample.

We made this analysis on the temporal resolution of streamflow data *a posteriori*, after selecting our catchment sample, and since the result is satisfactory we decided not to further discard catchments by fixing a limit on the distribution of the sampling intervals of the streamflow series.

2.2.2.2 Constitution of streamflow series at time-intervals from sub-hourly to daily

The streamflow data at variable time steps have been aggregated at different time steps proceeding as done in previous works by Le Moine (2008) and Lobligeois (2014). Streamflow time-series at different sampling times are generated by **linear interpolation of original data at variable time step and integration on every desired time interval**. If a variable time step data necessary to calculate a new fixed time interval data is considered as a gap, the aggregated data is not calculated, it is considered a gap too. Data for which the **quality of measurements** (indicated by a quality code in the *Banque-Hydro* database) is unknown (quality code 'I' or 'S') are considered as **data gaps**. Also, all the data corresponding to initial measurements after a suspension of the station operation (quality code ending with ';1') are considered as data gaps, to avoid interpolating between two non-continuous measurements.

As done by Lobligeois (2014) and unlike Le Moine (2008), we have retained all the data whether they are good (quality code '9') or reconstituted (quality code '8') or estimated (quality code '5'). This treatment avoids cutting the hydrographs above a threshold, due to data judged unreliable by the hydrometric services for the highest gauging points of the rating curves. Moreover, we applied two **additional criteria**, in order to avoid interpolating data over long time intervals, by **defining as gaps**:

- Data corresponding to an **interval between two measurements larger than 15 days**;
- Data corresponding to an **interval between two measurements larger than 5 days and a difference of two subsequent flow values larger than 0.05 m^3/s.** This is to avoid interpolating over too long periods when streamflow significantly varies, while avoiding considering as gaps the low flows or dry periods (streamflow close to 0 and without significant variations) for which flow measures registration is sometimes suspended

2.2.3 Evapotranspiration data

Meteorological data necessary for the calculation of potential evapotranspiration (*PE*, also noted E_P) are available at daily time intervals from the SAFRAN database by Météo-France. However, this information is not available at sub-daily time intervals. So we decided to calculate the potential evapotranspiration (*PE*) at daily time intervals and then disaggregate these data at the sub-daily scale by some empirical assumptions, as described below.

The *PE* formula that we used is the one proposed by Oudin et al. (2005) that depends only on data of daily air temperature, the latitude of the catchment and the Julian day in the year:

$$E_P = \frac{R_e}{\lambda \cdot \rho}\left(\frac{T_a + 5}{100}\right)$$

where E_P is the daily evapotranspiration (mm/d), T_a is the air temperature (°C), R_e is the extra-terrestrial radiation (MJ/m^2/d), λ the latent heat of vaporization (MJ/kg) and ρ the water density (kg/m^3).

For the temporal disaggregation of daily PE, we followed the method used in Lobligeois (2014), where daily *PE* is distributed at hourly time steps according to a parabola. This parabola is null between 19h and 6h UTC and reaches its maximum from 12h to 13h UTC (see Figure 2.5). This is clearly an approximation, but it should not generate significant errors in streamflow simulations, given the insensitivity of rainfall-runoff models to high-frequency *PE* variations as shown by Oudin et al. (2004).

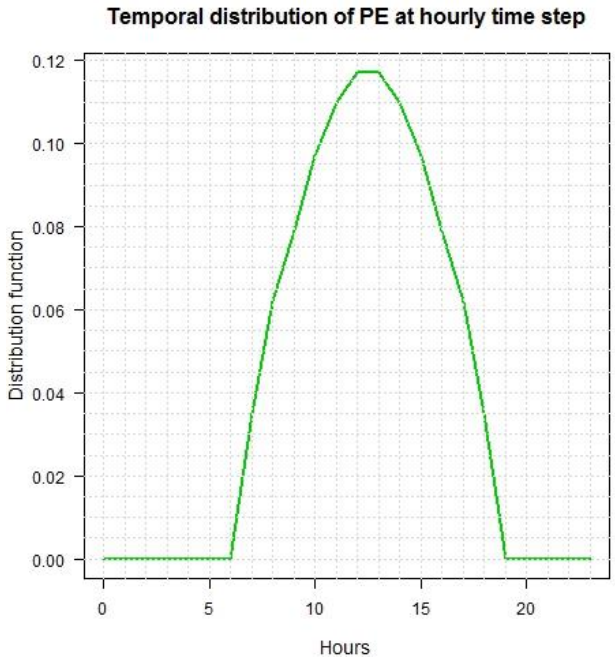


FIGURE 2.5 - PE temporal distribution at hourly time step used to disaggregate daily PE (Lobligeois, 2014).

2.3 Catchment selection criteria

Our scope is to construct a sample of unregulated catchments as large and varied as possible, aiming at the same time to contain the number of basins below 250 to keep the computation time reasonable. Catchments are selected to have no or low human and snow influences, since the model used is designed for natural catchments and is applied without a snow module. Some criteria on data quality are also required, as a minimum rain gauges density (for the 6-min network) and a limit to streamflow data gaps.

We aim to keep basins with different hydrological behaviours including basins with fast response times and slower ones. We desire also that our sample contains some catchments followed by the French flood forecasting services (SPC, French acronym for "*Services de Prévision des Crues")* and low-flows monitoring stations. So, we verified this feature at the end of our selection process (see Section 2.4).

In this section, we present the steps followed to select our test catchments, according to criteria on catchments characteristics and on rainfall and flow data.

2.3.1 Criteria on catchment characteristics

The selection has been performed starting from a list of **3701 catchments in metropolitan France** (Figure 2.6) having a hydrometric station at the outlet indexed in the Banque Hydro data-base (SCHAPI, 2014). It is a consistent part of the 4437 catchments set constructed at Irstea by Bourgin et al. (2010), by defining the basin contours from the hydrometric stations position and a digital terrain model (DTM) at 100-m-resolution, from which the flow directions were defined. The choice to start from the 3701 catchments subset is due to the availability of catchment rainfall time-series at the daily time step only for this subset.

FIGURE 2.6 - The contours of the 3701 catchments in metropolitan France used as starting point for our catchment set selection.

The following selection criteria were applied:

- We kept **3421 catchments** having a surface larger than 5 km^2 and smaller than 10000 km^2. This range of surfaces is judged sufficiently large to ensure the interest of our modelling work with different time steps, from sub-hourly to daily. We decided to exclude basins with a surface smaller than 5 km^2 because of the too large uncertainty that we would have on the precipitations for these basins (we remember that the daily rainfall data were originally available at 8-km-resolution from the SAFRAN data-base).
- We kept **3245 catchments** whose hydrometric station is not coded as virtual and has a finality of general hydrometric record or flood warning or low-flow monitoring or experimental basin. In this way we excluded stations designed for dam management and water police.
- We kept **2829 catchments** indexed in the Banque Hydro data base as basins with low or no influence.
- Another selection criterion followed for the scope of keeping only natural catchments, without human influence, was to consider an upper limit for the estimated storage capacities upstream each station. For this scope we used the database of dams storages (in Mm^3) inventoried by Payan, 2007. When the sum of all these storage capacities

represents more than a 10-mm equivalent water depth, the station is discarded (278 stations). This led to a number of **2551 catchments**.

- To remove the basins with a too large snow influence, we kept only the ones with a percentage of solid precipitation lower than (or equal to) 10% (**1977 catchments**).
- A last criterion was to discard the basins containing some key-words in their names as "channel", "source", "resurgence", and "pond" and this lead us to select **1961 catchments**.

2.3.2 Criteria on rainfall data

The quality of precipitation data is critically important for hydrological models development and calibration, as highlighted in the MOPEX project (Duan et al., 2006). Oudin et al. (2004) showed that the quality of streamflow simulation is highly dependent on the quality of rainfall data (contrary to evapotranspiration). When using precipitation data in lumped hydrological models, uncertainty comes from the raw data errors but also from the spatial aggregation at the catchment scale.

2.3.2.1 Selection on data gaps frequency

A first essential requirement is that the rainfall time series does not contain data gaps (missing data). The daily catchment rainfall time series are without gaps. However the 6-minute series contain gaps. Therefore a first selection is operated on the 1405 rain gauges stations of the 6-minute network based on the gaps percentages in these series. Our scope is to limit as much as possible the frequency of gaps in the series to ensure that the sub-daily temporal distribution of rainfall is mostly based on 6-minute observations and not arbitrary assumptions. To do so, we kept in consideration only the stations with less than 5% of gaps in the observations series for more than 90% of the days (1366 stations). Furthermore we selected only the stations with series including at least 5 years of data with less than 10% of gaps (**1128 gauge stations**).

2.3.2.2 Selection on rain gauge density

Further criteria were imposed to ensure a minimum density of the 6-minute rain gauges to limit the errors introduced by the disaggregation of daily catchment rainfall using the measurements of the 6-minute network. Otherwise these errors can be great given the lower spatial correlation of rainfall at sub-hourly time steps (see discussion in Section 2.2.1.2).

Previous studies for the MOPEX project have established a **minimum density requirement for rain gauges based on basin size** (Duan et al., 2006). Schaake et al. (2000) made a practical estimate of gauge density requirements to obtain highly accurate mean areal precipitation (MAP) estimates. They developed an equation by using observations from a very dense gauge network (45 km^2 per gauge station) across a number of basins of different basin size in the United States. The time step used to develop this equation is one-fourth of the basin lag time. The required number of gauges for a basin of area A (km^2) is:

$$N = 0.6\,A^{0.3} \tag{2.1}$$

The exponent 0.3 implies that the required number of gauges doubles as the basin size increases by a factor of 10. The number of gauges given by Equation (2.1) should give MAP estimates for each time step that are accurate with a maximum MAP error of 20 percent for 80 percent of the time during thunderstorm rainfall events (Schaake et al., 2000). The equation is applicable for basins between 200 and 20000 km^2. Below 200 km^2, Equation (2.1) may underestimate gauge requirements because the spatial correlation function for precipitation tends to fall more quickly with distance for shorter distances. This means that a minimum number of gauges, about 3, are needed for basins smaller than 200 km^2, to filter the noise associated with this "nugget" of the decorrelation function. In our catchment sample selection, we decided to verify these density criteria for the 6-minute rain gauges, reducing to 2 the minimum number of gauges for basins smaller than 25 km^2, because of the difficulty of having 3 gauges influencing such a basin area. To verify these gauges density criteria, we identified the pluviometers located in a square region of side 100 km around each basin and constructed the Thiessen polygons of these rain gauges (using ArcGIS). We counted the **number N' of pluviometers influencing each catchment** (with corresponding Thiessen polygon intersecting the catchment) and checked whether N' verifies Equation (2.1) or is at least 3 for basins smaller than 200 km^2 or at least 2 for basins smaller than 25 km^2. In this way, **1157 catchments** are selected. Moreover, for the basins larger than 200 km^2 (682 basins) we applied an additional criterion requiring at least one rain gauge located within the catchment area (discarding 212 basins without any rain gauge). We did not apply this criterion for smaller basins (730 basins), in order to take in the sample a sufficient number of these basins that are interesting for our work at finer time steps. Note also that even if no rain gauge is located within the catchment boundaries for about 85% of these small catchments, a large number of them present a high density of rain gauges close to their boundaries (see Figure 2.7).

Furthermore, we used **additional criteria** to limit the impact of the rain gauge density on the quality of the MAP estimates. To this end, we used a measure of the density of the 6-minute rain gauges network for each basin *b*, the **average rain gauges area** $\rho_{P,b}$, defined as the average area of influence of the rain gauges. In practice it is calculated as the average area of the Thiessen polygons intersecting the catchments weighted by the percentage of coverage of each polygon in the basin:

$$\rho_{P,b} = \sum_{k=1}^{N} A_k^{Th} \cdot \frac{A_k^{Int.}}{A_b} \tag{2.2}$$

where N is the number of pluviometers, A_k^{Th} the area of the Thiessen polygon of pluviometer k, $A_k^{Int.}$ the area of the intersection of the Thiessen polygon with the basin, and A_b the area of basin b. Equation (2.2) can be viewed as a natural definition of the rain gauge network density, and it has been already used in literature, for example by Lobligeois (2014). However, this formula makes no difference if a rain gauge is within the basin or not and this is not good for our scope of selection of catchments. We do not want to discard a catchment for a low density of the rain gauges just outside the basin (with large values of $\rho_{P,b}$) but to keep a catchment that has a high density inside its surface even if lower outside. To this end,

we modified Equation (2.2) taking into account only the area of Thiessen polygons intersecting the basin in case of pluviometers *k* inside the basin:

$$\rho'_{P,b} = \sum_{j=1}^{N_{in}} \frac{A_k^{Int.2}}{A_b} + \sum_{k=1}^{N_{out}} A_k^{Th} \cdot \frac{A_k^{Int.}}{A_b} \tag{2.3}$$

where N_{in} is the number of pluviometers inside the basin and N_{out} is the number of pluviometers outside ($N_{in} + N_{out} = N$).

Some statistics of the number of rain gauges influencing the selected catchments and their 'average rain gauge area' will be reported later in this chapter (see Table 2.1, Section 2.4.4). Note that the additional criteria on rain gauge density are useful, since there is a weak positive correlation between the number of rain gauges influencing each catchment and the 'average rain gauge area', as defined above (Pearson correlation coefficient r=0.28).

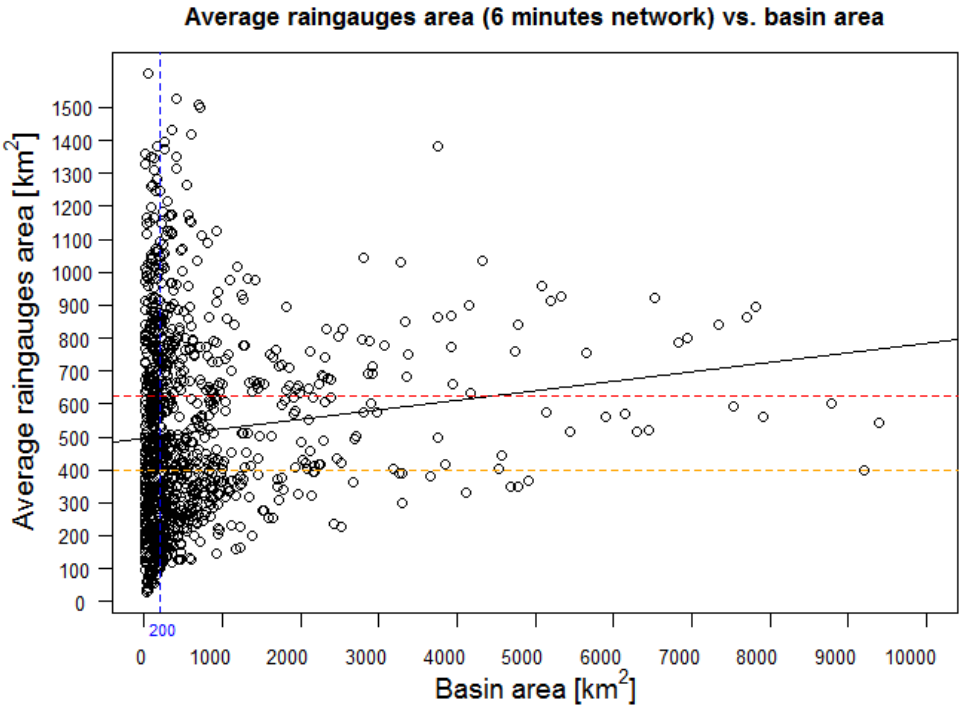


FIGURE 2.7 - Values of average rain gauges area and basin area for the 1157 catchments sample, with regression line. The vertical line (at 200 km²) represents the limit of basin area chosen to fix two classes of criteria for the minimum average rain gauges area (400 km² and 225 km²).

Figure 2.7 reports the values of average rain gauges area obtained by Equation (2.3) in relation to basin area for the sample catchments selected so far (339). The range of average rain gauges area is (30, 1603) km^2, while the range of basin area is (5, 9389) km^2. As example, Figure 2.8 shows the rain gauges influencing two basins of our sample with the

Thiessen polygons used to calculate the value of the average rain gauges area by Eq. (2.3). The first basin is smaller and presents a denser rain gauge network than the second, as shown by the values of average rain gauges area.

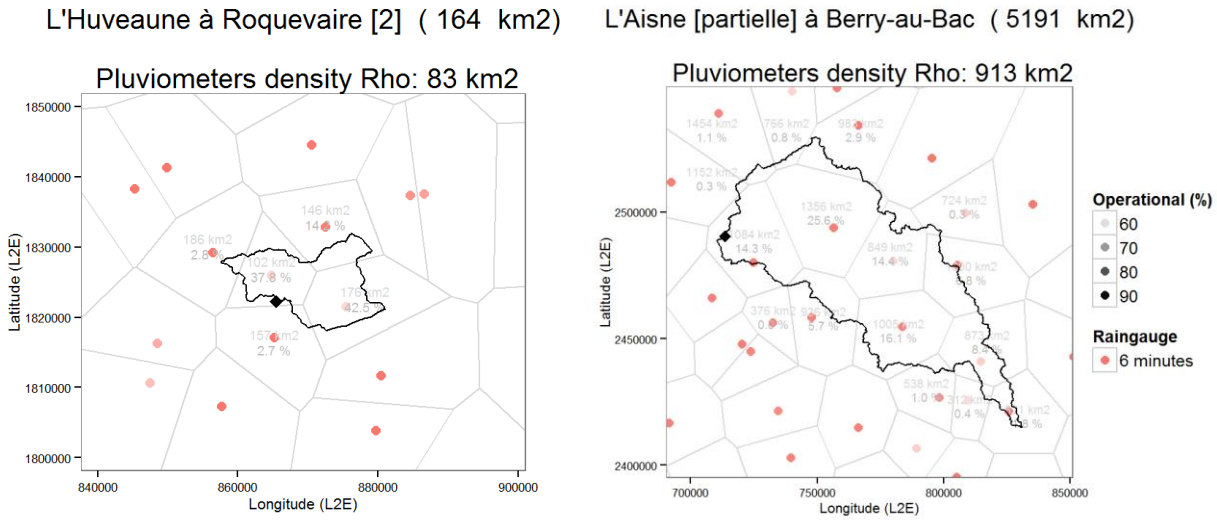


FIGURE 2.8 - Average rain gauges area for the 6-minutes rain gauge network on two catchments of different size (Y4414030, 164 km² and H6321011, 5191 km²). The surface of the Thiessen polygons (km²) and their percentages of coverage on the basin are shown in each polygon that intersects the catchment.

We defined two different **limits for the average rain gauges area** depending on the catchment surface (A_b):

$$\begin{cases} \rho'_{P,b} \leq 400\ km^2 \ \ if\ A_b < 200\ km^2 \\ \rho'_{P,b} \leq 625\ km^2\ if\ A_b \geq 200\ km^2 \end{cases} \tag{2.4}$$

These limit values were chosen empirically, considering the following ideas: (a) given the trend of growth of the average rain gauges area with the basin size (see Figure 2.7) we tried to be not too selective with respect to basins larger than 200 km²; (b) for smaller basins we fixed a more restrictive criterion to ensure a high density in the proximity of the basin to compensate the fact that we have not required the presence of at least one rain gauge within the catchment; (c) we also verified to keep Thiessen polygons with area values in the range of circle area with radius equal to the typical decorrelation distances of rainfall for sub-hourly time steps, i.e. from 10 km for 15 minutes (Zawadzki, 1973) to 50 km for hourly (Rubel, 1996); this allows us to exclude basins with an excessive portion farther than these distances from the nearest rain gauge; (d) we aimed also at being consistent with the maximum spatial resolution proposed by Obled et al. (2009), considering hourly time steps (see Section 1.4.1).

By applying the selection criteria of Equation (2.4), we kept the **584 catchments** represented in Figure 2.9. The selected catchments are fairly well distributed in France, although there are obviously regions more represented than others, corresponding to the regions where the

density of the pluviometers network is higher (as it can be seen comparing Figure 2.9 with Figure 2.1).

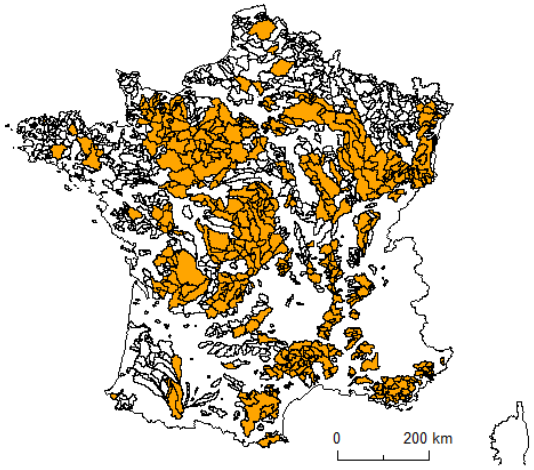


FIGURE 2.9 - The 584 French catchments (orange) selected by the criteria on rainfall data from the 1961 catchments (hollow and orange) selected by basins characteristics criteria.

2.3.3 Criteria on flow data

As for the flow data availability, we kept only the hydrometric stations for which **less than 35% of daily flow data** are **missing over each of the two periods 01/8/2005 – 31/7/2009 and 01/8/2009 - 31/7/2013** (being the two periods of equal length). The limit of 35% of gaps has been fixed by sensitivity analysis on the resulting sample, so as to keep a certain variety of catchments. Moreover this value can be considered over the limit below which the split-sample calibration-validation approach, advocated by Klemeš (1986b), would be difficult to apply. To have a measure of this problem, one can see for example Merz et al. (2009) that analysed the impact of the number of years available for calibration on model performances for 269 catchments in Austria. They show that model performance increases with the calibration period length, and using more than 3 years of calibration, one tends to sample sufficiently diverse hydrological conditions and the likelihood that one particular catchment performs poorly in the predictive mode is much lower than with shorter periods. In our case, by allowing at most 35% of data gaps on each calibration period of 4 years, the actual length of the data series available on each period will be at least almost 3 years. Note also that applying this criterion to daily time-series is more restrictive than to smaller time steps because the gaps frequency decreases with the sampling time, given the construction procedure of the flow series by interpolation of the variable time steps data (see Section 2.2.2.2). By applying the criteria on flow data availability explained above, we selected the **223 catchments** shown in Figure 2.10.

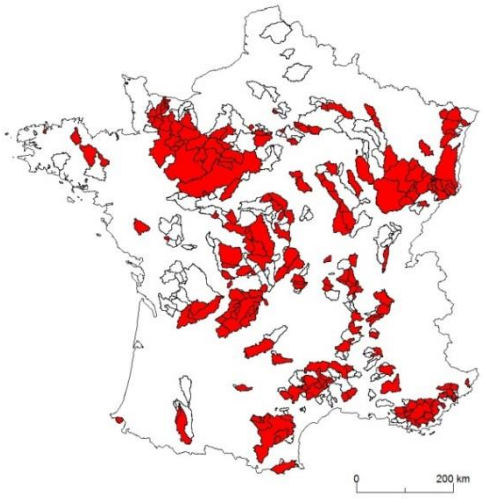


FIGURE 2.10 - Catchment sample of 223 French catchments (red) selected by criteria on flow data from the 584 catchments (hollow and red) selected so far, by basins characteristics and rainfall data criteria.

Among the 584 catchments selected so far: 261 were discarded because the source time series of streamflow (at variable time step) are not available; 56 were discarded because the source time series do not cover at all the period 2005-2013; 44 were discarded because their streamflow time series have more than 35% of data gaps over at least one of the two sub-periods used for the split-sample calibration-validation procedure. The limit of 35% of gaps will be sufficient to have enough data to calibrate and validate our rainfall-runoff models, also given the fact that the gaps frequency is less than 15% for about 90% of the catchments over the two periods (see Figure 2.11). The mean value of data gaps frequency is about 5% and the median value is less than 3% over the two periods. Thus for most of the catchments we will have almost 4 years of data availability (for each calibration-validation period) that can be considered a time period sufficiently long to have enough climate variability.

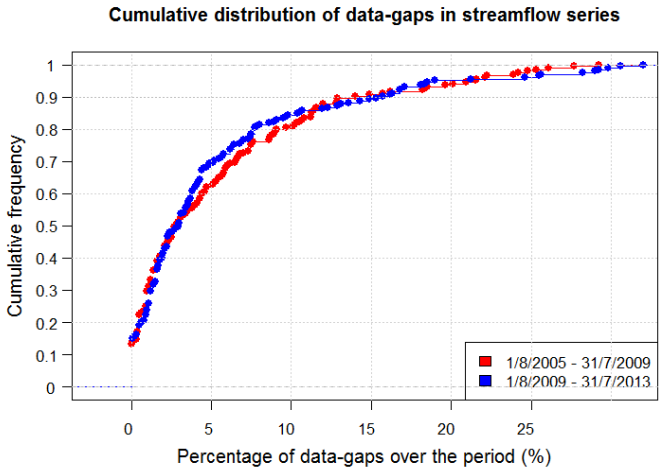


FIGURE 2.11 - Cumulative distribution of data-gaps in streamflow series for the 223 catchments of the sample selected by criteria on flow data, over the two periods 01/08/2005 – 31/07/2009 and 01/08/2009 - 31/07/2013.

2.3.4 Additional catchments recovered for particular reasons

At the end of the process of catchment sample selection explained above, we added some catchments for specific interests, by allowing a slight relaxation of the selection criteria especially for criteria on rainfall data.

Firstly, we added **two very small catchments (*S<5 km*^{2})**, for which the sub-hourly time steps are likely of the same order as the catchment response time. We selected them from the catchments with surface between 3 and 5 km^2, with the same criteria than the ones described above for the other catchments, except for the *minimum number of pluviometers of the 6-minutes network influencing the catchments*, that we *reduced to one*, while requiring the *distance of the pluviometers from the catchment centroid to be less than 2 km*. These catchments are: (i) the Morcille brook at Villié-Morgon, 4.12 km^2 (code 'U4506010'), that is an experimental basin followed by Irstea, for which we have also hourly rainfall measures at one rain gauge located within the catchment; (ii) the Sienne river at Saint-Sever-Calvados, 3.54 km^2 ('I7001040'). We remember that on these very small basins, the uncertainty on the daily precipitations is larger because they are derived from original data at 8-km-resolution from the SAFRAN data-base. For this reason, we have limited the number of these small basins in our sample.

We added also another **small experimental basin of 46 km^2**, the Yzeron river at Craponne ('V3015010'), located in the Rhône-Alpes region in eastern France.

We decided to include also **some additional basins of the southwest and northern parts of France**, because these regions were under-represented for the lower density of rain gauges of the 6-minutes network. So we recover some basins that were discarded for our criteria on rain gauges density, but that almost respect these criteria. To this scope we relaxed the criteria on *maximum average rain gauges area, 900 km^2 instead of 625 km^2 for basins with a surface*

larger than 200 km^2, and *minimum number of gauges, 2 instead of 3, for basins smaller than 200 km^2*. In this way, we recovered: **5 catchments** of the **southwest** of France, 'O5292510' (1603 km^2), 'Q0224020' (154 km^2), 'Q3120010' (7707 km^2), 'Q3464010' (843 km^2) and 'S2242510' (1678 km^2), and **4 catchments** of the **north**, 'B5572010' (368 km^2), 'E4035710' (392 km^2), 'G2011010' (307 km^2) and 'G2203010' (317 km^2).

Also an additional set of **5 catchments from the Adour basin**, in the Atlantiques Pyrenees region, was included in the sample to respond to the request to work on some basins *followed by the flood forecasting service of the Adour basin* (*SPC GAD, Gironde-Adour-Dordogne*). These catchments are: (i) the Adour river at Tarbes ('Q0120060'); (ii) the Arros river at Tournay ('Q0522530'); (iii) the Lees river at Lannux ('Q1094010'); (iv) the Gave d'Oloron river at Oloron-Sainte-Marie ('Q7002910'); (v) the Nive river at Ossees ('Q9102510'). The criteria of catchment selection described above were not applied for these catchments, being originally discarded for:

- dams influence on the streamflows (only one catchment shows an influence significantly greater than the limit in terms of stored volume, i.e. the basin 'Q7002910' with a total storage capacity of 32-mm equivalent water depth; the other catchments are below the limit of 10-mm or just around, 11-mm in one case);
- density of the 6-minute rain gauges network (3 catchments: 'Q0522530', 'Q1094010', 'Q9102510', that however present an acceptable density of rain gauges in their vicinity, with $\rho'_{P,b} < 900\ km^2$ for the worst case);
- snow influence (2 basins, 'Q012060' and 'Q7002910', present a percentage of solid precipitation significantly larger than our limit of 10%, i.e. around 25%).

Note that the different characteristics of these additional basins could negatively impact the results, but it could be interesting to analyse this impact, for example in term of loss of performance of the models due to the reduced density of the 6-minute rain gauges network, or the lack of description of the melting snow process, or dams operation.

Figure 2.12 shows the **final sample of 240 catchments** (223 catchments selected so far + 17 additional catchments) that will be used in our modelling tests. The list of the selected catchments is reported in Appendix C.

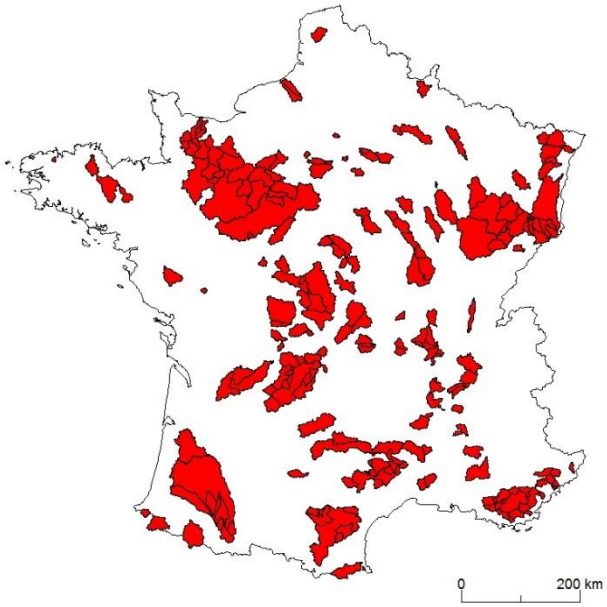


FIGURE 2.12 - Final catchment sample of 240 French catchments (red), including basins selected by criteria on catchment characteristics, rainfall and flow data, and basins recovered for specific reasons (basins smaller than 5 km², in the southwest part of France, and followed by the flood forecasting service of the Adour basin).

2.4 Summary of the characteristics of our catchment sample

2.4.1 Morphological and hydro-climatic characteristics

Our final catchment sample includes **240 lowland mesoscale French catchments** (Figure 2.12), whose surfaces vary from 3.5 km^2 to 3296 km^2. The selected basins are geographically well distributed over the whole French metropolitan territory and consequently represent a wide variety of morphological and hydro-climatic characteristics. Figure 2.13 and Figure 2.14 show respectively the statistical and geographical distribution of twelve physiographic, morphological and hydro-climatic catchment characteristics (definitions of some characteristics and comments are provided here below).

i. **Surface** (S [km^2]): our catchment sample mainly presents small catchments (Figure 2.13): the catchment surfaces vary from 3.5 km^2 to 8790 km^2, with a median value of 356 km^2 (Figure 2.13), and about 93% of the basins are smaller than 1000 km^2. Thus our sample concerns just small rivers, upstream of the larger French rivers.
ii. **Hydraulic length** (HL [km]): it is an index indirectly related to the water course length and is calculated as the average hydraulic length of all the pixels in the raster Digital Elevation Model (DEM) falling in the basin. The average hydraulic length of each pixel *i* is calculated as the distance from the outlet of the basin, following the flow direction grid from pixel *i*; in our catchment sample HL ranges from about 2 to 173 km.

iii. **Drainage density** (DD [km^2]): for the Drainage Density we used an index that represents the total surface of the "source basins" as proposed by Le Moine (2008); the higher the density of the river network, the lower the value of this index and vice versa. The value of DD for a certain basin *b* was obtained, following Lobligeois (2014), as the geometric mean value of the surfaces of the "source basins" that intersect the basin *b*. The lower values of DD in our sample are in the Mediterranean area and in the Pyrenees region (where there is the most part of the basins with DD<0.1 km^2; see Figure 2.14), where the surface drainage network consists of a lot of small water courses. High values of DD (DD>1 km^2) are concentrated especially in the north of France, but the highest value (DD=110 km^2) is for a basin in the Mediterranean area characterized by karst geology (the *Mas Pomier* creek at Nant in the Regional Natural Park of "*Grands Causses*", i.e. the Park of large limestone plateaus, from French). In fact, a large surface of the "source basins" means that water tends to infiltrate in the soil instead of flowing in surface, as observed also by Lobligeois (2014).

iv. **Average altitude** (z [m]): it is calculated as the average altitude of the DEM pixels falling in the basin; almost all our catchments are lowland basins (z<1000 m) with a range of average altitude between 70 m and 1308 m above the sea level and a median of 362 m; the catchments with the highest altitudes are in the Pyrenees, Alps and Cévennes regions.

v. **Average slope** (tan(β), [-]): it is the average of the slopes of the DEM pixels falling in the basin, where the slope of each pixel *i* is calculated as:

$$\tan(\beta) = \frac{(z_i - z_{up})}{\sqrt{(x_i - x_{up})^2 + (y_i - y_{up})^2}}$$

where β is the slope in degrees, (x_i, y_i, z_i) the projected coordinates *x, y* and altitude *z* of the considered pixel and (x_{up}, y_{up}, z_{up}) the projected coordinates *x, y* and altitude *z* of the upstream pixel, identified by the steepest slope around pixel *i*. The average slope is correlated with the average altitude as logically expected (Figure 2.14).

vi. **Topographic index** ($\log(n_i)/\alpha_i$, [-]): it is the topographic index of TOPMODEL (Beven and Kirkby, 1979) calculated as by Ducharne (2009), i.e. rescaled to be independent from the DEM resolution, as:

$$I_{topo} = \log(\frac{n_i}{\alpha_i})$$

where n_i is the number of pixels upstream pixel *i* and α_i is the slope (tan(β)) of pixel *i*. From its definition the topographic index is correlated to the basin surface (Figure 2.14).

vii. **Average annual streamflow** (Q, [mm/y]): it is the annual mean value of stream-flows observed at the basin outlet, converted into water sheet (mm) on the catchment area to be comparable between different basins; it is calculated over the whole period of daily data availability from the data-base constructed at Irstea by Bourgin et al. (2010). For our catchment sample it ranges from 36 to 1524 mm/y (Figures 2.13 and 2.14). Together with the average annual rainfall and evapotranspiration, it can be used to

characterize the hydrological regimes and annual water balance of our catchment sample (see also Figure 2.15).

viii. **Average annual rainfall** (P, [mm/y]): calculated as the mean annual precipitation over the whole period of daily data availability; for our catchment sample, P ranges from about 650 to 2108 mm/y (Figures 2.13 and 2.14).

ix. **Average annual potential evapotranspiration** (PE, [mm/y]): calculated as the mean annual potential evapotranspiration over the whole period of daily data availability; for our catchment sample PE ranges from about 600 to 1130 mm/y (Figures 2.13 and 2.14).

x. **Autocorrelation of streamflows at time-lag 24 hours** ($\rho(Q_t, Q_{t+24h})$, [-]): it is the correlation coefficient between the daily streamflows (Q) at time steps t and $t+24h$. A high value of autocorrelation indicates a slow response of the catchment to precipitation (that is more smoothed out), while on the contrary a fast response is characterized by lower values of the autocorrelation. In our sample the catchments with the faster response are in the Mediterranean area (Figure 2.14).

xi. **Daily precipitation intensity coefficient** (P_{99}/P_m, [-]): it is calculated as the ratio between the 99th percentile of the daily rainfall (P_{99}) and the inter-annual daily mean value of rainfall (P_m). A high value of P_{99}/P_m indicates that the catchment is exposed to intense convective precipitation events. The catchments in our sample with the highest values of the daily rainfall intensity coefficient are concentrated above all in the Cévennes region and more generally distributed in the Mediterranean area.

xii. **Base Flow Index** (BFI, [-]): it is defined as the ratio between the base flow component and the total flow at the outlet; thus the base flow index (BFI) ranges between 0 and 1. The BFI here is calculated following the method described by Gustard et al. (1992). A high BFI value (BFI>0.65) indicates the presence of a large base flow component in the catchment, generally reflecting the importance of the underground flow. The highest values of BFI in our sample are mostly located in the north of France that is well-known for the geologic dominance of *chalk* formations. However, some basins with high BFI are present also in the Mediterranean area. Note also that high BFI values are highly correlated to high DD values (see comments in point (iii) here above and Figure 2.14).

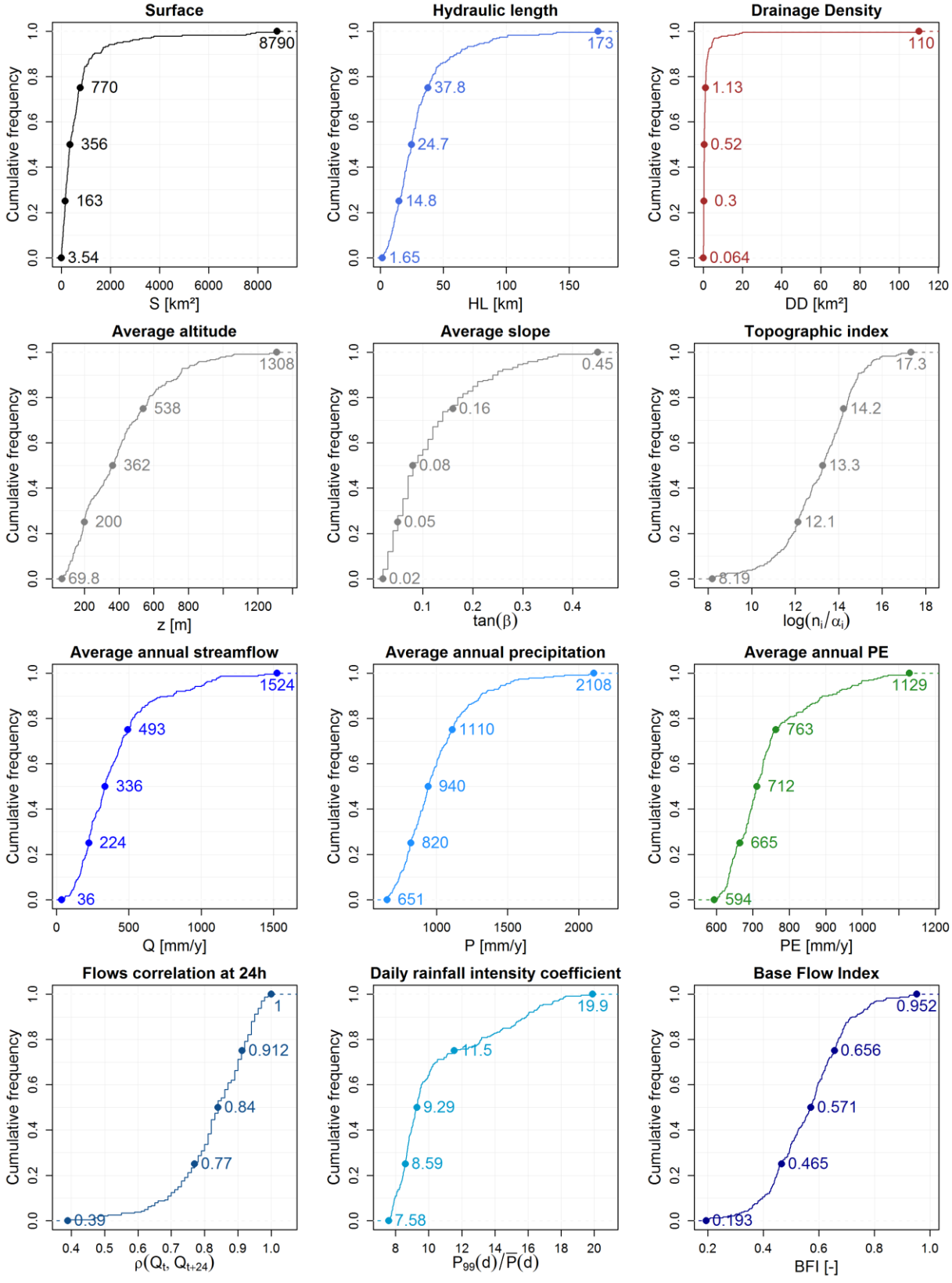


FIGURE 2.13 - Distribution of the morphological and hydro-climatic characteristics of our catchment sample.

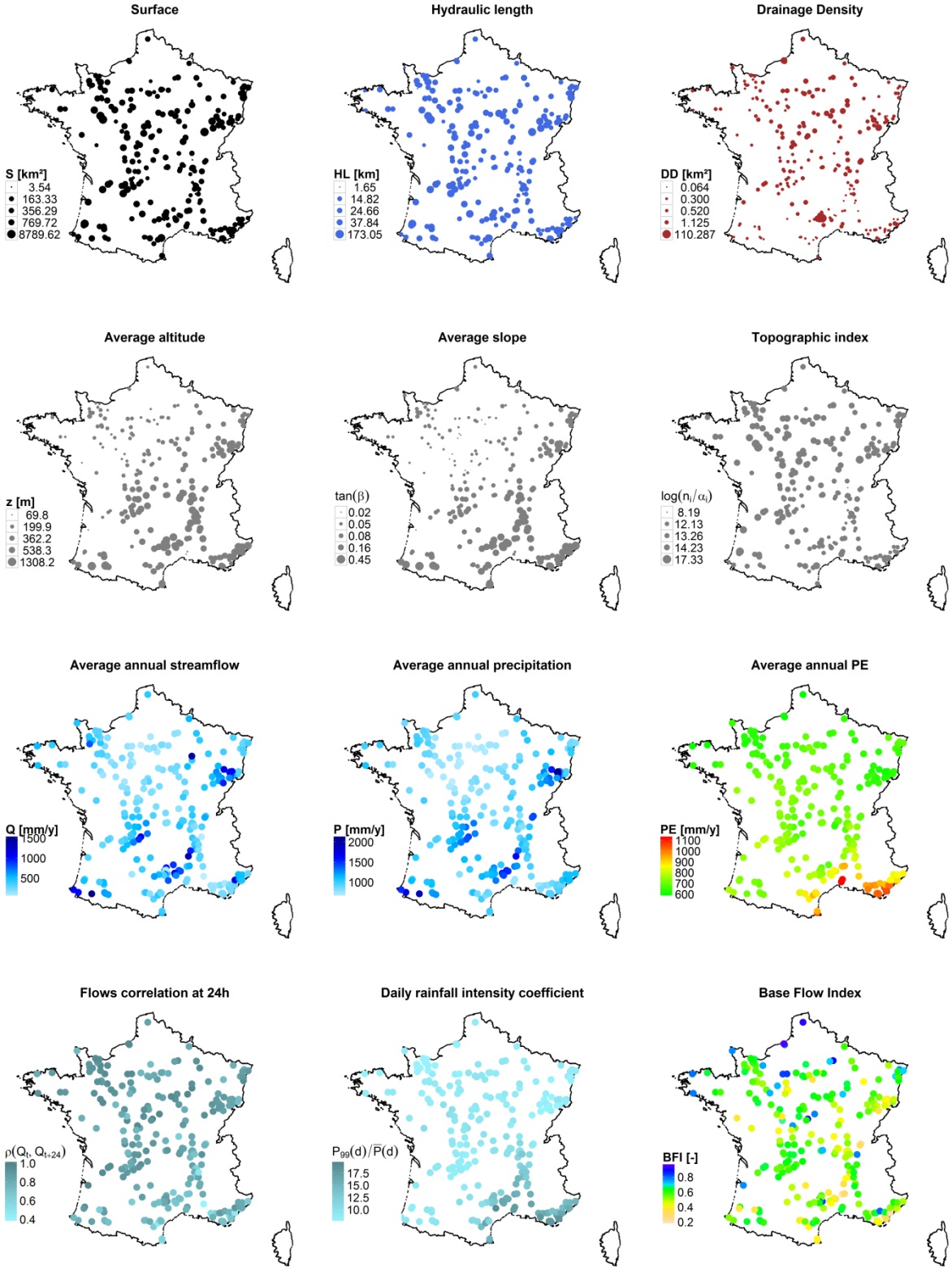


FIGURE 2.14 - Spatial distribution of the morphological and hydro-climatic characteristics of our catchment sample.

Figure 2.15 shows the **annual hydrologic balance**, between precipitation P, potential evapotranspiration E_P and streamflow Q for the 240 catchments of our sample (red points) in the non-dimensional graph Q/P=f(P/E$_P$). The two dashed curves in Figure 2.15 define the *realistic physical region* that should normally contain all basins without groundwater contribution to or from its neighbours. Normally for these basins the flow should never exceed precipitation (i.e. $Q/P \leq 1$). On the other hand, since the potential evaporation is the upper bound of the actual evaporation, the flow should always be greater than the difference P - E_p (i.e. $Q/P \geq 1 - E_P/P$). Only a dozen of our 240 catchments are outside these limits, and these mostly correspond to basins with high values of the base flow index (BFI>0.6; Figure 2.14). Note also that the annual water balance of our catchment sample is largely representative of the wide variety of regimes of the French basins (grey points, Figure 2.15).

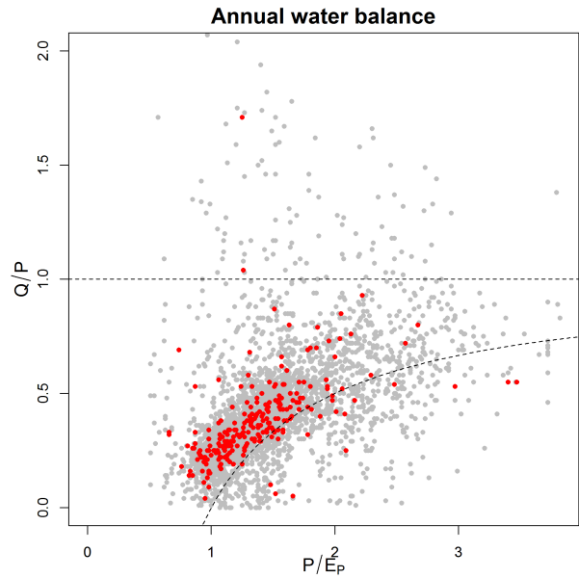


FIGURE 2.15 - Representation of the annual hydrologic balance of the 240 catchments of our sample (red points) and of 4437 French catchments (grey points) for which hydro-climatic data are available.

2.4.2 Hydrological regimes

Here we present a classification of the hydrological regimes that can be found in our catchment sample, following the classification of river flow regimes made by Sauquet et al. (2008), who identified 12 regime groups for France according to the **distribution of the normalized inter-annual monthly runoff time series** (ratio between mean monthly and annual flows, Q_m/Q_y). As discussed by Sauquet et al. (2008), twelve reference hydrographs were constructed by hierarchical cluster analysis based on similarity criteria (Ward's minimum variance method and Euclidian distance) to construct a flow regime map of France. Figure 2.16 shows the reference hydrographs constructed by Sauquet et al. (2008) for the 9 classes of regimes present in our catchment sample. Figure 2.17 reports the geographical distribution of the 9 groups of catchments for each hydrological regime in Figure 2.16. Note that obviously snowmelt-fed regimes (groups 10-12 in Sauquet et al., 2008) are not represented in our sample, as we discarded the catchments with a solid precipitation fraction

greater than 10%. However, there are some basins in the transition (or *pluvio-nival*) regime and the nivo-pluvial regime (Groups 8 and 9) distributed in the mountainous areas of Alps, Massif Central and Pyrenees (16 basins). These basins include for example the two catchments not respecting the criterion on solid precipitation fraction recovered in the sample for a specific reason (see Section 2.3.4). For these catchments, high flows are observed in spring, which reflects the fact that the seasonal variation of streamflow is influenced as much by precipitation as snowmelt timing.

The great part of our catchment sample is characterized by pluvial river flow regimes (Groups 1 to 7), representing the dominant regimes in metropolitan France, with the oceanic pluvial regimes that reigns outside the Mediterranean area. These groups mainly differ by the contrast between the maximum and the minimum of monthly streamflow and small nuances in their timing. Nearly uniform flows through most of the year (*Group 1*) are found in 9 basins of our sample, mostly coinciding with catchments where large aquifers moderate flows (BFI>0.65) as in the northern part of France and in the southern karst region of the Regional Park of "*Grands Causses*" (see also section 2.4.1, point (iii)). All our basins of Brittany and almost all of Normandy regions are allocated in *Group 5*, which describes the strong homogeneity of these regions in terms of river flow regime. This runoff pattern includes 35 basins (not only in north-west part of France) characterized by very low flow in summer, reflecting the lack of deep groundwater storages in these basins. A similar but more smoothed behaviour characterizes 86 basins of our sample belonging to *Groups 4 and 2*, which present only one maximum observed flow in winter (similarly to Group 5) and low flows in summer (less extreme than Group 5). These basins are quite spread over France, but include almost all the basins of the Eastern part of France (Alsace and Lorraine regions). Similar to this flows regime are also *Groups 3 and 6*, including 73 basins of our sample, most of them in the centre of France, presenting a less monotonous trend of high flows (increasing of flows also in spring). Finally, 21 basins are allocated to *Group 7* that is representative of Mediterranean river flow regimes, particularly in the Cevennes region, where small rivers basins experience hot and dry summers and intense rainy events in autumn and winter. Their runoff pattern therefore exhibits severe low flows in summer, with possible totally dry periods, and maximum annual high flows in November, with possible flash floods due to intense but short rainy events. These basins present also a local maximum flow in spring (in April), reflecting the influence of snowmelt in the Alps and Massif Central areas.

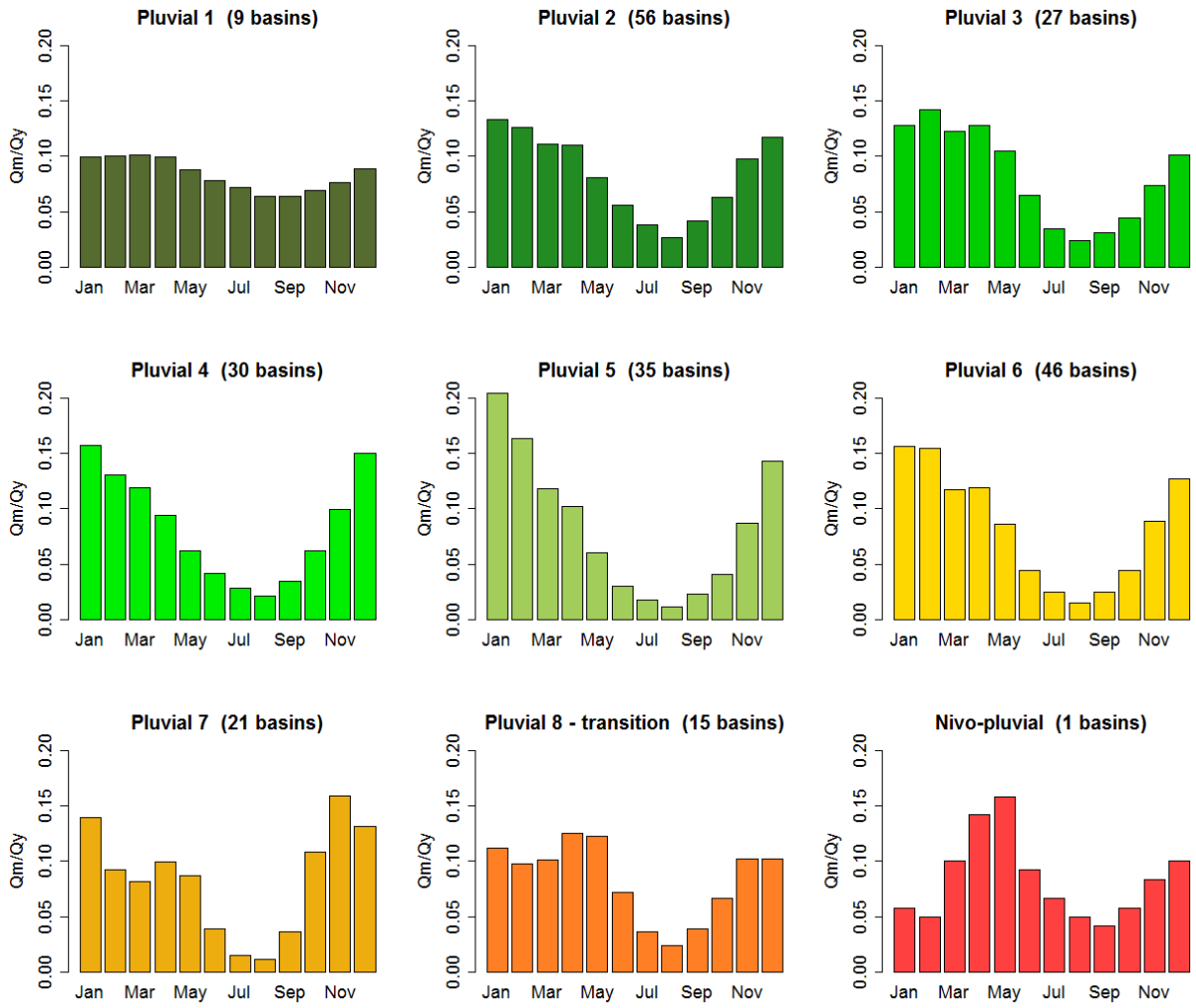


FIGURE 2.16 - Hydrological regimes of our catchment sample (with absolute frequency) based on the reference hydrographs for France defined by Sauquet et al. (2008).

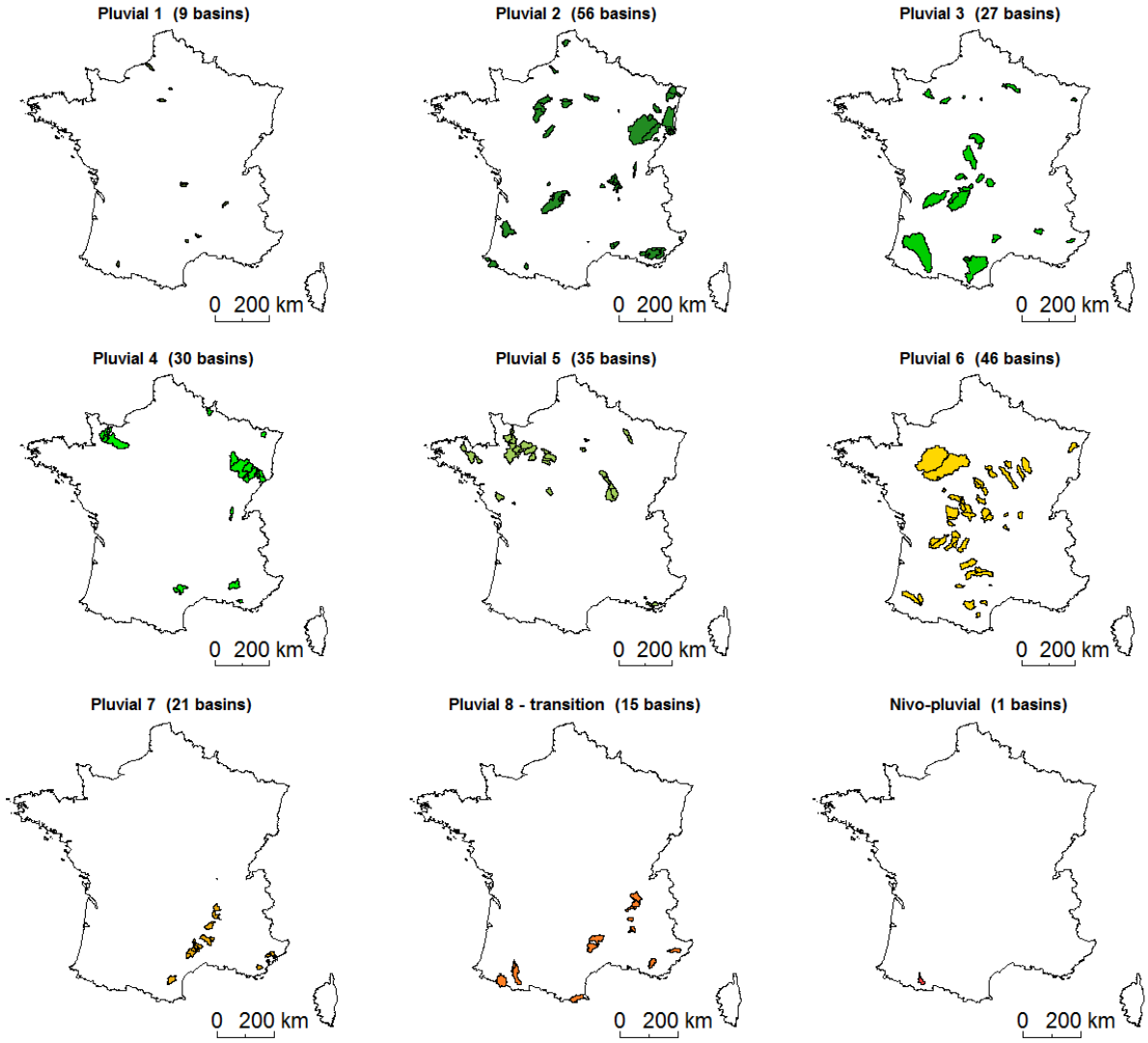


FIGURE 2.17 - Geographical distribution of the hydrological regimes of our catchment sample (with absolute frequency) based on the reference hydrographs for France defined by Sauquet et al. (2008).

2.4.3 Operational interest of our catchment sample

Figure 2.18 and Figure 2.19 show the catchments of our sample followed respectively for flood warning by the French flood forecasting services, (S.P.C., *Service de Prévision des Crues*), and for low-flows monitoring, by the Regional Directions of the Environment, Planning and Housing (DREAL, *Directions régionales de l'environnement, de l'aménagement et du logement*).

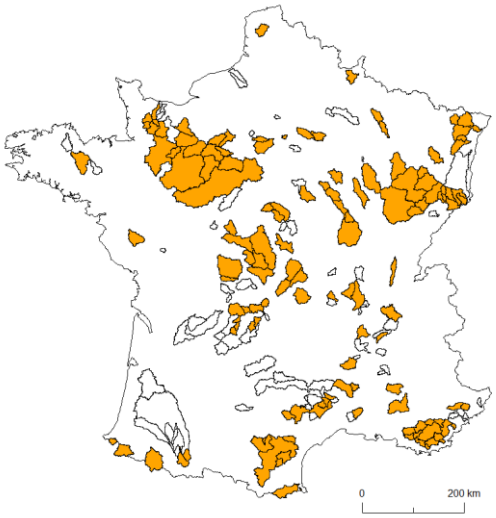


FIGURE 2.18 - Catchments of our sample (hollow) followed by the SPC, i.e. French flood forecasting services (134 catchments coloured in orange).

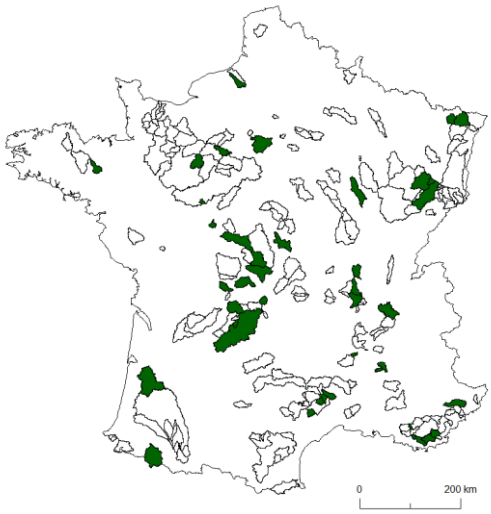


FIGURE 2.19 - Catchments of our sample (hollow) followed for low-flows monitoring by DREAL (42 catchments coloured in green).

2.4.4 Synthesis of the catchment characteristics

The test catchments are geographically well distributed over the entire area of metropolitan France and represent the variety of oceanic and Mediterranean pluvial regimes existing in continental France.

Table 2.1 summarizes the distribution of ten morphological and hydro-climatic characteristics of the 240 catchments. These characteristics were chosen to investigate their relationship with model behaviour at different time steps (see the analyses of Chapter 3). Typically, smaller and steeper catchments are expected to present faster responses to precipitation and to benefit

more from a refined temporal resolution of rainfall. The daily precipitation intensity coefficient (P_{99}/P_m ratio) and the flow auto-correlation at 24 h characterize the temporal variability and dynamics of rainfall and streamflow, respectively. The highest values of the daily precipitation intensity coefficient are concentrated mainly above the Cévennes region and in the south-eastern Mediterranean area, indicating the occurrence of intense convective precipitation events. In the same area, the lowest values of the flow auto-correlation are found, corresponding to catchments presenting the fastest response to precipitation with frequent flash floods (Delrieu et al., 2005; Berne et al., 2009; Saulnier and Le Lay, 2009; Javelle et al., 2010; Braud et al., 2014).

The catchment set includes as many catchments as possible respecting the selection criteria discussed above, with the objective to get general results, i.e. not too much dependent on specific catchment types or climate conditions. No a priori removal of "outlier" catchments (e.g. karstic catchments or groundwater-dominated catchments) was done, as advised by Andréassian et al. (2009). As a consequence, two catchments (the Laine river at Soulaines−Dhuy and the Siagne river at Callian) which are notoriously karstic, have mean annual runoff coefficients larger than 1 (Table 2.1).

The ten characteristics in Table 2.1 may present strong linear correlations. The highest Pearson correlation coefficients ($r>0.75$) are found between catchment area, hydraulic length and topographic index, and between average altitude and slope. A quite large negative correlation ($r=-0.62$) was found between the daily precipitation intensity coefficient and the flow auto-correlation at 24 h, reflecting that catchments with the fastest flow dynamics are also subject to the most intense precipitation events. Despite these correlations, all descriptors were kept in subsequent analyses, since linearly correlated descriptors may not be similarly relevant given the non-linearity of hydrological processes.

#	Catchment characteristics	Definition / reference	Min	1st Quartile	Median	3rd Quartile	Max
MH1	Catchment area [km^2]	-	4	163	356	770	8790
MH2	Hydraulic length [km]	Average hydraulic length of all the catchment pixels	2	15	25	38	173
MH3	Average altitude [m]	-	70	200	362	538	1308
MH4	Average slope [-]	-	0.02	0.05	0.08	0.16	0.45
MH5	Topographic index [-]	As calculated in Ducharne (2009)	8.20	12.10	13.30	14.20	17.30
MH6	Runoff coefficient [-]	Ratio of mean annual runoff to precipitation	0.04	0.26	0.34	0.46	1.69
MH7	Aridity index [-]	Ratio of mean annual potential evapotranspiration to precipitation	0.29	0.64	0.76	0.91	1.53
MH8	Base Flow Index [-]	As calculated by Gustard et al. (1992) using daily streamflow data	0.19	0.47	0.57	0.66	0.95
MH9	Flow auto-correlation at 24 h [-]	-	0.39	0.77	0.84	0.91	1.00
MH10	Daily precipitation intensity coefficient, P_{99}/P_m [-]	Ratio of the 99th percentile to the mean daily precipitation	7.58	8.59	9.29	11.50	19.90
RGD1	Number of rain gauges influencing the catchment [-]	Rain gauges of the 6-min network with Thiessen polygons intersecting the catchment	1	3	5	7	32
RGD2	Average rain gauge area [km^2]	Average area of the Thiessen polygons weighted by coverage percentage over the catchment	30	208	302	393	1126

TABLE 2.1 – Summary of the distribution of ten morphological and hydro-climatic (MH) characteristics and two rain gauge density (RGD) indicators over the 240 test catchments.

2.5 Flood event set

An automated procedure of flood event selection adapted from Lobligeois (2014) (see Appendix D for greater detail) was used to select ten flood events for each catchment over the 8-year test period (2005–2013), providing a set of 2400 flood events over the 240 catchments. Table 2.2 presents the distribution of some average flood descriptors for the 240 catchments. These flood characteristics were chosen because they are conceptually related to the temporal dynamics of flood events and therefore are expected to be relevant for our analyses. One

descriptor, the Goodness of Uniform Estimates (GOUE) Index, was proposed in this study and is further described in Appendix E. It is applied to both precipitation and streamflow series, to evaluate their respective degree of temporal variability at a fine resolution (6 min) compared to a reference larger time scale (1 day).

As for the catchment characteristics, some of these flood descriptors are correlated. The highest Pearson correlation coefficients ($r>0.75$) are found between: flood and storm durations ($r=0.88$); the GOUE Index on streamflows and the 24-h flow shape coefficient ($r=-0.83$); rainfall-runoff lag time and duration of storms ($r=0.79$) and floods ($r=0.76$); the GOUE index on streamflows and the flood duration ($r=0.77$); and the 24-h flow shape coefficient and mean flow gradient ($r=0.76$). No strong correlations of this order ($r>0.75$) were found between the flood descriptors in Table 2.2 and the catchment morphological and hydro-climatic characteristics (Table 2.1). However, the daily flow auto-correlation is correlated with the 24-h flow shape coefficient as well as the mean flow gradient ($r=-0.64$) and with the GOUE Index on streamflows ($r=0.64$). The expected correlations between rainfall-runoff lag time and catchment area and hydraulic length were partially confirmed (respectively with $r=0.61$ and $r=0.68$), while the slope is less correlated with lag time ($r=-0.40$).

#	Flood characteristics	Definition	Min	1st Quartile	Median	3rd Quartile	Max
FE1	Flood duration [h]	Number of hourly time steps between the start and end of the event	20	65	111	232	518
FE2	Storm duration [h]	Number of hourly time steps with non-zero precipitation between the start and peak of the event	5	21	33	55	155
FE3	Amount of precipitation for event [mm]	Cumulated precipitation over the total event duration	18	45	62	86	204
FE4	24-h flow shape coefficient [-]	Ratio between the maximum instantaneous flow and the maximum of mean flows on 24-h duration around the peak	1.01	1.10	1.22	1.45	4.60
FE5	Mean flow gradient [mm/h^2]	Ratio between the flow difference between peak flow and base flow and the rising limb duration	0.0000	0.0045	0.0125	0.0304	0.35
FE6	Rainfall-runoff lag time [h]	Time shift maximizing the cross-correlation of rainfall-runoff series over each event at 6-min resolution	1.25	7.56	13.10	24.06	92.00
FE7	GOUE(P) [-] of daily precipitation at 6-min reference [-]	Nash-Sutcliffe criterion between daily rainfall uniformly disaggregated at 6-min time step and the 6-min observed series (see Appendix E)	−0.04	0.11	0.16	0.22	0.47
FE8	GOUE(Q) [-] of daily streamflow at 6-min reference [-]	Nash-Sutcliffe criterion between daily streamflow uniformly disaggregated at 6-min time step and the 6-min observed series (see Appendix E)	−0.30	0.49	0.71	0.87	0.99

TABLE 2.2 – Summary of the distribution of the flood event (FE) characteristics over the 2400 selected events averaged by catchments (by the median values for each catchment).

2.6 Model and calibration-evaluation procedure

2.6.1 The GR4 rainfall-runoff model

The model used in this study is the GR4 model (Perrin et al., 2003), a four-parameter lumped rainfall-runoff model. In its daily version, it is popularly known as GR4J. In this thesis, since we will use and adapt this model at multiple time steps, we have removed the 'J' letter from the original name, because it stood for daily (from French 'journalier').

This model was chosen because of the expertise gained from previous works on its development at daily and hourly time steps and for its operational relevance for the French flood forecasting services (e.g. Berthet et al., 2009). A detailed description of the structure of the GR4 model can be found in Perrin et al. (2003) and an overview of the model equations is provided hereafter. The GR4 model is composed of a production function, similar to soil moisture accounting functions, followed by routing and water exchange functions, as schematically represented in Figure 2.20. The soil moisture accounting store receives part of the net rainfall (P_s) and satisfies part of the net potential evapotranspiration (E_s) determined by zero-capacity interception storage. The routing function consists of a linear routing with two unit hydrographs (UH) and a non-linear routing store. The groundwater exchange term makes it possible to import (export) water from (to) the outside of the basin. The four free parameters to be optimized are:

(i) the maximum capacity of the production store (x_1, [mm]);

(ii) the water exchange coefficient (x_2, [mm/time step]), which can be either positive for water gains or negative for water losses;

(iii) the maximum capacity of the routing store at one time step ahead (i.e. capacity after removing the output of the store) (x_3, [mm]);

(iv) the time base of the unit hydrograph (x_4, [time step]).

The model also has fixed parameters, the value of which is justified by Perrin et al. (2003).

As discussed in Chapter 1, there were previous works to adapt the GR4 simulation model to sub-daily time steps: Mathevet (2005) and later Le Moine (2008) successively proposed improved model versions at the hourly time-step. Le Moine (2008) made more explicit the relationships of model parameters with time step. Following his work, we identified the theoretical time-step dependencies of the fixed and free model parameters (see Table 2.3). Thus, the GR4 model can be already tested at sub-daily time steps by simply adapting its time-step dependent parameters. This will be the starting point for our modelling tests at multiple time steps.

The current hourly version of the model, called GR4H (in the literature), proved already adequate with respect to model performance, as showed by Mathevet (2005) and Le Moine

(2008). Also, other works report the use of the GR4H model with satisfactory results when compared to other hourly models (Van Esse et al., 2013; de Boer-Euser et al., 2017).

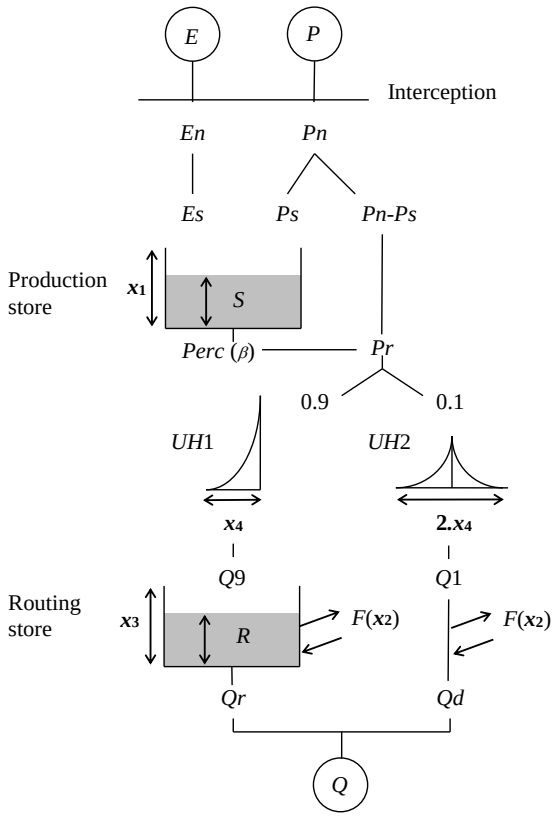


FIGURE 2.20 – Schematic representation of the GR4 model structure (Perrin et al., 2003).

GR4 model parameter	Theoretical transformation from $\Delta t_1[s]$ to $\Delta t_2[s]$	Source of the time-step dependency
β [-]	$\beta_{(\Delta t_2)} = \beta_{(\Delta t_1)} \left(\frac{\Delta t_1}{\Delta t_2}\right)^{\frac{1}{4}}$	Integration of the percolation rate from the production store that is a power 5 function
x_2 [mm/time step]	$x_{2\,(\Delta t_2)} = x_{2\,(\Delta t_1)} \left(\frac{\Delta t_1}{\Delta t_2}\right)^{-\frac{1}{8}}$	Integration of the exchange flux formulation (dependent on the routing reservoir level)
x_3 [mm]	$x_{3\,(\Delta t_2)} = x_{3\,(\Delta t_1)} \left(\frac{\Delta t_1}{\Delta t_2}\right)^{\frac{1}{4}}$	Integration of the emptying function of the routing reservoir that is a power 5 function
x_4 [time step]	$x_{4\,(\Delta t_2)} = \left\lfloor x_{4\,(\Delta t_1)} \left(\frac{\Delta t_1}{\Delta t_2}\right) \right\rceil$	Value expressed in time-step units and rounded to the nearest integer

TABLE 2.3 – Theoretical time-step dependency of one fixed and three free time step-dependent parameters of the GR4 model.

Model equations

Here we present the discrete equations of the GR4 model used as baseline model for multiple time steps. All equations derive from the integration of the continuous equations over a time step Δt, apart from the equations representing the interception loss and the exchange fluxes, which are directly written in their discrete form. In the following equations, the water fluxes are integrated over the model time step and are expressed in mm on the integration time step.

The first operation is the determination of an *interception loss*, i.e. evaporation from intercepted water (E_i), by a *neutralisation* function of the precipitation P by the potential evapotranspiration E, which are consequently reduced to a net rainfall (P_n) and a net evapotranspiration amount (E_n):

$$E_i = \min(P, E) \tag{2.5}$$

$$P_n = P - E_i \tag{2.6}$$

$$E_n = E - E_i \tag{2.7}$$

In case the precipitation P is greater than the potential evapotranspiration E, the net evapotranspiration is null, while if E is greater than P, the net rainfall is null.

The production store is filled by a part P_s of the net rainfall, representing the part of rainfall infiltrating in the soil moisture accounting store. The remaining part of the net rainfall ($P_n - P_s$) bypasses the production store and reaches directly the routing part of the model. The amount of water P_s is determined as:

$$P_s = \frac{x_1 \left(1 - \left(\frac{S}{x_1}\right)^2\right) \tanh\left(\frac{P_n}{x_1}\right)}{1 + \frac{S}{x_1} \tanh\left(\frac{P_n}{x_1}\right)} \tag{2.8}$$

where S [mm] is the level in the production store at the beginning of the time step and x_1 is the maximum capacity of the production store.

Under the effect of the net evapotranspiration amount, the production store is emptied of a quantity of water E_s, i.e. the actual evaporation from the store, determined as:

$$E_s = \frac{S \left(2 - \frac{S}{x_1}\right) \tanh\left(\frac{E_n}{x_1}\right)}{1 + \left(1 - \frac{S}{x_1}\right) \tanh\left(\frac{E_n}{x_1}\right)} \tag{2.9}$$

Given the inputs to the store (net rainfall and net evapotranspiration) and a certain store capacity (x_1), the actual evaporaton E_s increases with the level S in the store, while the amount of infiltrating water P_s decreases.

The total amount of water evaporated from the model on the time step is therefore given by the sum of E_i and E_s.

The discrete Equations (2.8) and (2.9) are derived from the integration over the time step of the differential equation $\frac{dS}{dt} = [1 - \left(\frac{S}{x_1}\right)^2]P_n - \frac{S}{x_1}\left(2 - \frac{S}{x_1}\right)E_n$, as reported by Edijatno (1991) (see pp. 329-332).

The production store level S is updated by adding P_s and removing E_s. Then a percolation leakage $Perc$ is removed from the store. It is calculated as:

$$Perc = S\left\{1 - \left[1 + \left(\frac{S}{\beta_{\Delta t} \cdot x_1}\right)^4\right]^{-\frac{1}{4}}\right\} \tag{2.10}$$

The discrete Equation (2.10) is obtained from the integration over the time step of an instantaneous leaking function of power 5, which leads to a time-step dependent percolation constant $\beta_{\Delta t}$ (see Table 2.3).

The percolation $Perc$ reaches the routing part of the model, where it is added to the part of the net rainfall that has bypassed the production store ($P_n - P_s$).

The total amount of routed water, i.e. $P_r = Perc + (P_n - P_s)$, is divided into two flow components by a fixed ratio: 90% of P_r is routed by a unit hydrograph (UH1) and then a non-linear routing store, and 10% of P_r is routed only by another unit hydrograph (UH2). Both unit hydrographs depend on the same time parameter x_4, expressed in time-step units, with a time base equal to x_4 for UH1 and $2x_4$ for UH2. The ordinates of the unit hydrographs are derived from the corresponding *S-curves (SH1* and *SH2)* defining the cumulative distribution of the inputs with time, as detailed by Perrin et al. (2003). These curves are calculated over a number of time steps equal to the base time of the unit hydrographs, as:

$$For\ 0 \le t \le x_4\text{:}\ SH1(t) = \left(\frac{t}{x_4}\right)^{5/2} \tag{2.11}$$

$$For\ 0 < t \le x_4\text{:}\ SH2(t) = \frac{1}{2}\left(\frac{t}{x_4}\right)^{5/2}, \tag{2.12}$$

$$\text{and } for\ x_4 < t \le 2x_4\text{:}\ SH2(t) = 1 - \frac{1}{2}\left(2 - \frac{t}{x_4}\right)^{5/2} \tag{2.13}$$

The ordinates of the unit hydrographs are then calculated as:

$$UH1(j) = SH1(j) - SH1(j-1) \tag{2.14}$$

$$UH2(j) = SH2(j) - SH2(j-1) \tag{2.15}$$

where *j* is an integer between 1 and the maximum number of ordinates, *n* and *m*, for *UH1* and *UH2*, respectively, i.e. the smallest integers exceeding x_4 and $2x_4$. The ordinates are used in the convolution product of effective rainfall to calculate the outputs of the unit hydrographs.

A *groundwater exchange* loss (or gain) is then released (or added) from (to) both flow components. The potential half exchange, *F*, is calculated as:

$$F = x_2 \left(\frac{R}{x_3}\right)^{3.5} \cdot \Delta t \tag{2.16}$$

where *R* [mm] is the level in the routing store at the beginning of the time step, x_2 [mm/time step] is the water exchange coefficient, and x_3 [mm] is the reference capacity of the routing store. In absolute terms, the higher the level in the routing store and the higher the water exchange coefficient x_2, the larger the exchange F. The actual exchange losses are limited by the water available in the routing store and by the flows components coming from the unit hydrographs.

Then the water content of the routing store *R* is updated by adding the output of the unit hydrograph UH1 ($Q9$) and removing (or adding) the exchange component *F*. The outflow from the routing store gives the first flow component Q_r, and is calculated as:

$$Q_r = R\left\{1 - \left[1 + \left(\frac{R}{x_3}\right)^4\right]^{-\frac{1}{4}}\right\} \tag{2.17}$$

Similarly to the percolation equation, the discrete Equation (2.17) is obtained from the integration over the time step of an instantaneous leaking function of *R* at power 5, which leads to a theoretical time step dependency of the routing store reference capacity x_3 (see Table 2.3).

The output of the second unit hydrograph UH2 provides the direct flow component Q_d, after being subject to the same groundwater exchange component (loss or gain) *F*, as:

$$Q_d = \max(0; Q1 + F) \tag{2.18}$$

The total streamflow is finally given by the sum of the two flow components ($Q_r + Q_d$).

Baseline model implementation and considered time steps

An implementation of the GR4 model at daily and hourly time steps is freely available in the "*airGR*" R package developed at Irstea by Coron et al. (2017). This implementation is used as the baseline version of the model tested in this thesis, with some code adaptations to run the model at multiple time steps.

In order to achieve the objectives of this thesis, we will use a set of model time steps ensuring a sufficient continuity in the temporal sub-daily scale. The time steps that will be used are the following: 6, 12, 30 min, 1, 3, 6, 12 h and 1 day.

2.6.2 Calibration and evaluation procedure

The available 8-year period was split into two test sub-periods (01/08/2005–31/07/2009 and 01/08/2009–31/07/2013). Eight years were judged sufficient to apply a common calibration-validation procedure, following the split-sample test (Klemeš, 1986b) on the two 4-year sub-periods. Some previous studies with different conceptual hydrological models empirically indicated that this is a period length sufficient to obtain robust parameters (e.g. Yapo et al., 1996; Perrin et al., 2007; Merz et al., 2009).

For our data set, this choice is also supported by two aspects: (i) the stability of the climatic conditions between the two sub-periods, and (ii) the significant event-based variability within each period.

First, there is no strong change in the climatic conditions between the two periods, over the catchment set, as shown in Figure 2.21. On average, over the catchment set, the mean annual precipitation amount varies between the two 4-year sub-periods of about 5% (median), while the maximum relative variation is about 33% (see Figure 2.21(a)). The strongest relative variations (>20%) of mean annual precipitation between the two periods are found in catchments located in Southern France, where higher interannual variability of precipitation is expected. For both periods, the mean annual precipitation is about 940 mm on average over the catchment set (median), which is an evidence of a stationary climate over this 8-year period. For the potential evapotranspiration, the median relative variation of the mean annual values between the two sub-periods is lower than 1%, while the maximum variation is about 3% (see Figure 2.21(b)). The mean annual potential evapotranspiration is about 700 mm on average over the catchment set for both periods. On average, the mean annual streamflow varies slightly more between the two periods, with a median relative variation of about 13% and a maximum of about 92% (see Figure 2.21(c)). Still, the average hydrological regime is stationary between the two periods (no general increase/decrease is observed) and only a few catchments are affected by large absolute deviations of streamflow between the two periods (i.e. catchments with low runoff coefficients, located in Southern France, which therefore are more sensitive to changes in rainfall). The median absolute deviation (with sign) of annual streamflow is only about +13 mm by passing from the first to the second 4-year period (compared to an annual value of about 300 mm). Regarding the precipitation distribution, the daily precipitation intensity coefficient (P_{99}/P_m, see Section 2.4.1) varies on average of only 8% between the two sub-periods (see Figure 2.21(d)).

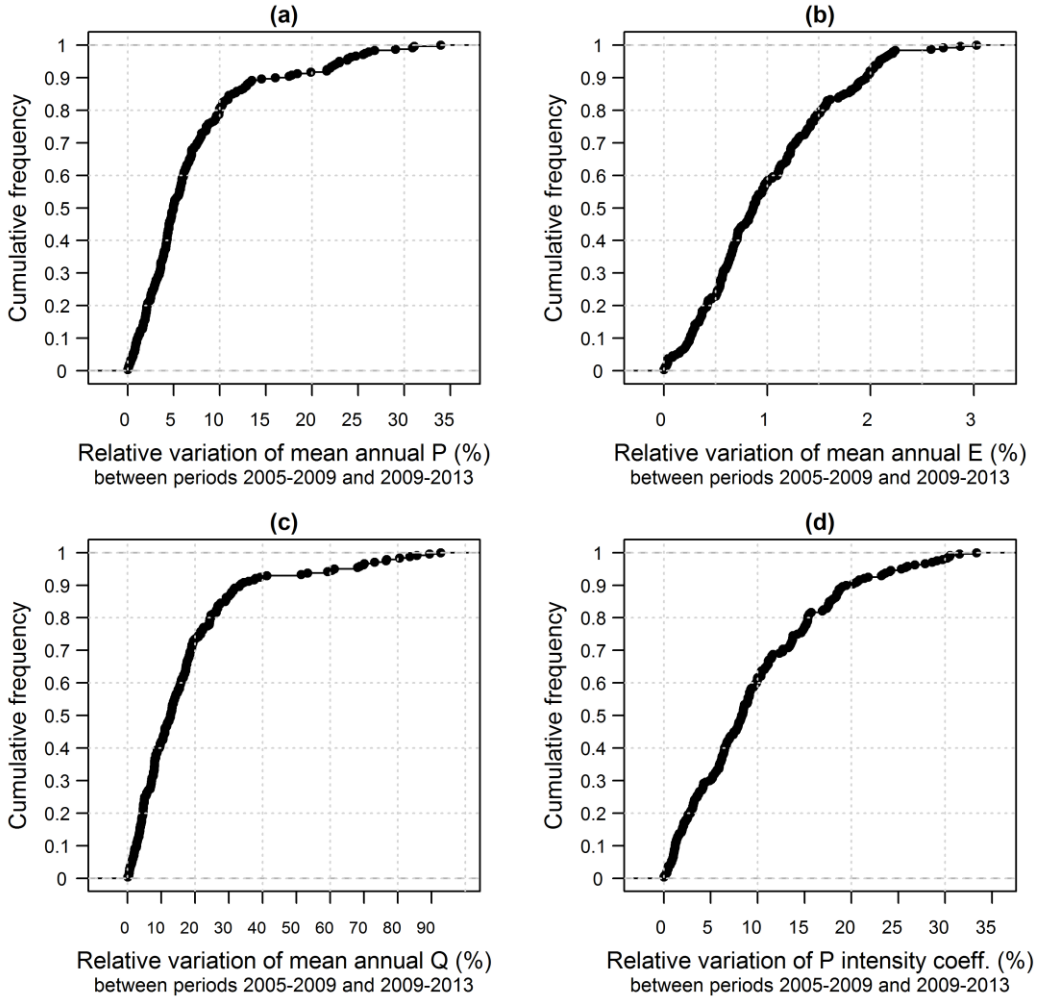


FIGURE 2.21 – Cumulative distribution over the catchment set of the relative variation between the two 4-year periods (2005-2009 and 2009-2013) of: (a) mean annual precipitation, (b) potential evapotranspiration, (c) mean annual streamflow, and (d) daily rainfall intensity coefficient coefficient (P_{99}/P_m). The variation is expressed in absolute terms and then normalized to the mean value over the whole 8-year period.

Second, within each 4-year sub-period, the information content in the data-set is sufficiently varied, including dry and wet conditions, and a representative variability of flood events characteristics, as shown in Table 2.4. The statistics of the flood event characteristics over the two periods show that the variability of the events is similar across the two 4-year periods. Overall, the statistics provided in Figures 2.21 and Table 2.4 show that the climatic regimes of the two periods are not significantly contrasting and each 4-year sub-period includes varied events. Thus, the periods can be considered sufficiently long to lead to robust parameter sets.

#	Flood characteristics	Period I (01/08/2005 – 31/07/2009)				Period II (01/08/2009 – 31/07/2013)			
		Min	**Median**	**Max**	**Mean width of variation range by catchment**	**Min**	**Median**	**Max**	**Mean width of variation range by catchment**
FE1	Flood duration [h]	18	103	550	141	16	122	521	187
FE2	Storm duration [h]	5	30	175	50	4	37	152	57
FE3	Amount of precipitation for event [mm]	8	59	218	60	16	66	349	86
FE4	24-h flow shape coefficient [-]	1.01	1.22	3.46	0.64	1.01	1.21	5.92	0.66
FE5	Mean flow gradient [mm/h^2]	0.00	0.01	0.42	0.06	0.00	0.01	0.29	0.08
FE6	Rainfall-runoff lag time [h]	0.30	13.10	120.20	13.68	0.20	13.90	160.20	17.41
FE7	GOUE(P) [-] of daily precipitation at 6-min reference [-]	−0.04	0.16	0.47	0.22	−0.08	0.16	0.42	0.24
FE8	GOUE(Q) [-] of daily streamflow at 6-min reference [-]	−0.22	0.70	0.99	0.43	−0.42	0.74	0.99	0.43

TABLE 2.4 – Summary of the distribution of the flood event (FE) characteristics over the 2400 selected events divided by the two 4-year periods 2005-2009 (1149 events) and 2009-2013 (1251 events) and averaged by catchment (by the median values for each catchment over each period). For each period, the "mean width of variation range by catchment" represents the average absolute value of the range amplitude of the characteristic over the catchment set.

After some first tests and analyses (not shown), we decided to use a warm-up period of two years before each calibration and simulation sub-period (i.e. 01/08/2003 to 31/07/2005 and 01/08/2007 to 31/07/2009; see section 2.2.1.1), which was judged sufficiently long to initialize the model states.

2.6.2.1 Model calibration

The model is calibrated on each 4-year sub-period, using the **Kling–Gupta efficiency (KGE) criterion** (Gupta et al., 2009) as objective function, calculated on all the time steps of the sub-period. The KGE criterion is calculated as:

$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2} \quad (2.19)$$

where:

- r is the linear correlation coefficient between simulated and observed flows;

- *a* is the ratio of the standard deviation of simulated flows to the standard deviation of observed flows, i.e. a measure of relative variability between simulated and observed values;

- *b* is the ratio of the mean of simulated flows to the mean of observed flows, i.e. an index of the overall bias in water balance.

The 'ideal' value for KGE and its three components (*a,b,r*) is 1. Since in general we aim at reproducing the overall volume of flow (ideally $\beta=1$), the variability (ideally $\alpha=1$) and the timing and shape of the hydrograph (ideally $r=1$), maximising the KGE criterion is a good compromise solution because it minimizes the Euclidian distance from the ideal point ($\alpha=1, \beta=1, r=1$).

The model calibration is performed by a two-step optimization procedure proposed by Mathevet (2005):

(i) A pre-filtering search in the parameter space is performed using three quantiles for each of the four parameters. The quantiles of the parameters (used as possible starting values) are given by parameter sets libraries resulting from previous works on the GR4 model development on large catchment sets (Perrin et al., 2008).

(ii) Starting from the best solution provided by the pre-filtering search, an iterative local optimization is performed using the "*pas-à-pas"* (*step-by-step*) method developed at Irstea. This procedure proved to be effective and robust for different conceptual rainfall-runoff models, including the GR4 model (Mathevet, 2005; Perrin et al., 2008).

2.6.2.2 Model validation

The model validation (or evaluation) is performed by various criteria calculated either over the whole validation sub-period or only over the flood events of the validation sub-period.

For the **evaluation over the whole period**, the KGE criterion and its three components (variability ratio *a*, ratio of means *b* and correlation *r*) are routinely used.

As a complement, other criteria are used at some moments of our model diagnostics and identification process, in order to evaluate the model in a multi-criteria perspective (flood, average regime and low-flows). These complementary criteria are based on the *Flow duration curve* (FDC) to evaluate the spread of simulated flows with respect to the observed flows frequency distribution and to detect any trend of over or under-estimation of extremes. In particular, the criteria used are:

a. *Ratio of lower quantiles* (e.g. 20th percentiles) of simulated and observed streamflows ($\frac{Q_{20}^{sim}}{Q_{20}^{obs}}$), characterizing low flows; we used the 20th percentiles and not lower quantiles, to avoid constraining our evaluation to reproduce very low flows (which is not the objective of short time steps modelling), and because of the large relative errors in low-flow measurements;

b. *Ratio of the higher quantiles* (e.g. ninety-ninth percentiles) of simulated and observed streamflows: $(\frac{Q_{99}^{sim}}{Q_{99}^{obs}})$, characterizing high flows.

c. The *percent bias in the slope of the mid-segment of the FDC* [%], that is related to the vertical soil moisture redistribution as discussed by Yilmaz et al. (2008). The slope of the mid-segment of the FDC is stepper for catchments with a more "flashy" response (due to small storage capacity and larger percentage of overland flow), while is flatter for catchments having slower and more sustained groundwater flow response. The percent bias in the *slope of the mid-segment of the FDC, between 0.2 and 0.7 flow exceedance probabilities*, can be defined as:

$$SlopeBias_{FDC(0.2-0.7)} = 100\left(\frac{\log\left(FDC_{sim}(0.2)\right) - \log\left(FDC_{sim}(0.7)\right)}{\log\left(FDC_{obs}(0.2)\right) - \log\left(FDC_{obs}(0.7)\right)} - 1\right) [\%] \quad (2.20)$$

where $FDC_{sim/obs}(\mathrm{p})$ represents the simulated/observed flow with exceedance probability p. Being a bias, the ideal value of the criterion $SlopeBias_{FDC(0.2-0.7)}$ is 0%. A positive percentage indicates that the simulated catchment response is more rapid than the correspondent observed catchment response, while a negative value indicates the opposite tendency.

For the **evaluation over the flood events**, we mainly used the KGE and its components. In order to ensure the significance of these aggregated statistics (*KGE, a, b and r*) over short time series, these criteria are calculated on the ten flood events taken altogether for each catchment (and not on each single event individually). Moreover, for events with durations shorter than three days, the event start and end are adjusted to have an evaluation period lasting at least three days.

In Chapter 3, a complementary assessment is carried out on each flood event individually to evaluate the proper identification of flood peak magnitude and timing. This was based on three event-based criteria to evaluate the peak flows, the timing and the volume errors (see Chapter 3).

2.7 Synthesis

In order to meet the research objectives of this thesis, we have built up a large data base of French catchments, for which high-resolution hydro-climatic data is available, over a period of at least 8 years. We have presented the data available and the treatments done to prepare the time series of catchment precipitation, potential evapotranspiration and streamflow at different time steps, needed for our modelling tests.

The catchment selection procedure was reported in detail. This procedure led to a set of 240 test catchments presenting a wide range of morphological and hydro-climatic characteristics.

Finally, we have presented the GR4 simulation model (general naming of the GR4J model at different time steps) and the calibration-evaluation procedure that will be used in this thesis. The GR4 model, developed at Irstea (Antony), has already been adapted from the daily to the hourly time step in some previous works. A first simple way of adapting the model at different time steps has been reported, based on expliciting its time-step dependent parameters. Thus, the GR4 model represents the baseline model version that we will test and try to improve at a varied set of sub-daily resolutions.

3

Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events

This chapter corresponds to an article published in *Journal of Hydrology*. :

Ficchì, A., Perrin, C., Andréassian, V. (2016). Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events, Journal of Hydrology, Vol. 538, 454–470, http://dx.doi.org/10.1016/j.jhydrol.2016.04.016.

Notes

In this chapter, we have removed some parts of the original paper dealing with:

- the presentation of the hydro-climatic data, and the catchment and flood event sets (Section 2 of the original paper);
- the GR4 model (Section 3.1 of the original paper).

These parts of the original paper were removed to avoid repetition, because they were already presented in the previous chapter of the thesis (2 - Material and methods).

About the GR4 model, here we start from the assumption that the same model structure can be maintained whatever the time step is, from daily to sub-hourly. This hypothesis is supported by previous work on the model (Mathevet, 2005; Le Moine, 2008) showing that the optimal structure was rather stable passing from daily to hourly time steps and no additional complexity seemed necessary, on the basis of model performance. As suggested by Le Moine (2008) (p. 200), there can be different reasons for this, including the hypothesis that the representation of the rainfall-runoff relationship at the catchment scale would not need to be more complex as the time step decreases from daily to hourly. In support of this, there would be the possibility that the greater variability of the streamflow at shorter time steps would be completely explained by the rainfall information at higher resolution. This hypothesis is to be verified by considering also other possible problems that could affect the results at different time steps (e.g. identifiability of parameters and internal coherence of the model). This chapter addresses this issue more thoroughly from the point of view of the model performance. We remind that to limit any bias in applying the model at different sub-daily time steps, we identified the theoretical time-step dependencies of the fixed and free model parameters following Mathevet (2005) and Le Moine (2008) (see Chapter 2, Table 2.3). Moreover, we have done a first analysis of the consistency of the parameters calibrated at different time steps finding an overall good coherence. A detailed discussion of the impact of the time step on parameter values is not within the scope of this paper. However, we present in Appendix F a summary of the consistency of the parameter values at different time steps, for the GR4 model used in this chapter.

Abstract

Hydro-climatic data at short time steps are considered essential to model the rainfall–runoff relationship, especially for short-duration hydrological events, typically flash floods. Also, using fine time step information may be beneficial when using or analysing model outputs at larger aggregated time scales. However, the actual gain in prediction efficiency using short time-step data is not well understood or quantified. In this paper, we investigate the extent to which the performance of hydrological modelling is improved by short time-step data, using a large set of 240 French catchments, for which 2400 flood events were selected. Six-minute rain gauge data were available and the GR4 rainfall-runoff model was run with precipitation inputs at eight different time steps ranging from 6 minutes to 1 day. Then model outputs were aggregated at seven different reference time scales ranging from sub-hourly to daily for a comparative evaluation of simulations at different target time steps. Three classes of model performance behaviour were found for the 240 test catchments: (i) significant improvement of performance with shorter time steps; (ii) performance insensitivity to the modelling time step; (iii) performance degradation as the time step becomes shorter. The differences between these groups were analysed based on a number of catchment and event characteristics. A statistical test highlighted the most influential explanatory variables for model performance evolution at different time steps, including flow auto-correlation, flood and storm duration, flood hydrograph peakedness, rainfall-runoff lag time and precipitation temporal variability.

3.1 Introduction

3.1.1 Importance of short time-step data

The transformation of rainfall into streamflow includes a large number of processes, with various dynamics and characteristic time scales on the order of 1 minute to hundreds of years (Blöschl and Sivapalan, 1995). The proper description and simulation of these processes may require short time steps for at least three reasons: (i) because of the short duration of the modelled runoff events (e.g. flash floods); (ii) because of the considerable intra-storm variability that controls some runoff processes and (iii) for numerical reasons especially related to the integration of differential equations in the model structure. This raises the issue of the appropriate time step of data used as input to hydrological models (typically precipitation).

Until the 1990s, hydrologists had to rely mostly on data at the daily step at best, e.g. ground accumulated rainfall amounts recorded once a day by observers. This could cause limitations in the applicability of rainfall-runoff models needing shorter time steps, which had to be run with data disaggregated over the shorter time steps either uniformly or by mass curves (Blöschl and Sivapalan, 1995) or by more sophisticated stochastic generators (Creutin and Obled, 1980). However, over the last two decades, the availability of hourly and even sub-hourly data tremendously increased in many countries, especially with the implementation of automatic rain gauge networks and meteorological radars (e.g. Creutin and Borga, 2003; Berne and Krajewski, 2013). This boosted the development of hydrological models running at short time steps to make use of these available data (e.g. Hughes, 1993; Moretti and Montanari, 2007; Chu and Steinman, 2009; Jeong et al., 2010).

One idea underlying these developments is that data at short time steps contain more information and therefore should contribute to better modelling of the rainfall-runoff relationship. This is supported by several studies that showed that runoff generation is highly affected by sub-hourly dynamics of precipitation, particularly where the infiltration-excess overland flow mechanism dominates the rainfall-runoff response (e.g. Koch and Kekhia, 1987; Morel-Seytoux, 1988; Woolhiser and Goodrich, 1988; Krajewski et al., 1991; Kandel et al., 2005; Paschalis et al., 2014). The precipitation controls the high-frequency catchment response, contrary to evapotranspiration whose variations are much more smoothed (Oudin et al., 2006). The temporal distribution of rainfall affects not only the runoff temporal distribution, i.e. flood shape, but also the peak discharge value (Gabellani et al., 2007) and the runoff volume (Viglione et al., 2010). This is due to the nonlinear nature of infiltration (and runoff) processes, with characteristic time scales of a few minutes (Blöschl and Sivapalan, 1995; Kandel et al., 2005). Woolhiser and Goodrich (1988) point out the significance of the rainfall intensity–infiltration interaction using a simple physically based model, concluding that the constant intensity rainfall pattern on an hourly scale cannot be recommended, especially for rapid catchments. Various studies have shown that infiltration excess surface runoff is modelled better using a sub-daily time step rather than daily time-step models, and

peak rates of rainfall are recognized as the most important controls for rainfall-runoff modelling (e.g. Yu et al., 1998; Socolofsky et al., 2001; Kandel et al., 2004; Kandel et al., 2005).

3.1.2 Modelling at short time steps to evaluate at larger time steps

Given the importance of sub-daily variability of rainfall for runoff modelling, one may be more confident in a model running with short time-step data than a model running with larger time-step data, even when the target model assessment time step is large. However, in the literature there is a limited number of case studies, where a rainfall-runoff model is run at a short time step and its outputs are used or evaluated at a larger time step (e.g. Hughes, 1993; Finnerty et al., 1997; Schreider and Jakeman, 2001; Kannan et al., 2007; Jeong et al., 2010; Yang et al., 2016). Typically (sub-)hourly or daily time steps are used for running the model and then performance assessment is based on daily or monthly aggregated outputs, respectively. Hughes (1993) discussed the advantages of using fine sub-daily time steps up to 5 min by applying a variable time-step model structure to two semi-arid catchments and a total of six storm events. The results suggested that the simulated runoff volume may be improved as the time step decreases up to 1 h for one catchment and even up to 5 min for the other catchment with higher rainfall intensities and a faster response. Finnerty et al. (1997) analysed the sensitivity of the SAC-SMA model to the spatial and temporal discretization of rainfall inputs while holding the parameters constant. They showed that the runoff volumes cumulated over a 9-month period significantly changed when the time step decreases from 6 h to 1 h. The surface runoff resulted in being the most temporally sensitive model component, which is attributed to the varied averaging of high-intensity short-duration precipitation events that affect surface runoff. Jeong et al. (2010) developed a sub-hourly version of the SWAT model and tested it on a small catchment. They showed the improvement in model performance when sub-daily predicted streamflows (at 15 min and 1 h) are aggregated to daily averages compared to daily simulation results. Similar results were shown also by Yang et al. (2016), using hourly and daily rainfall observations as inputs of the SWAT model for daily streamflow simulation on one medium-sized catchment, while, for the same model, contrasting results were found by Kannan et al. (2007) for the ranking of sub-daily and daily inputs options on one small catchment.

Despite these overall encouraging findings, the common modelling practice is still to choose the model and input data time step equal to the evaluation time step. The assessment of aggregated outputs using shorter time-step data is rarely reported, even among the increasing number of studies examining the time scale dependencies of rainfall-runoff model parameters (e.g. Littlewood and Croke, 2008; Wang et al., 2009; Ostrowski et al., 2010; Littlewood et al., 2011; Bastola and Murphy, 2013; Littlewood and Croke, 2013). In these studies, simulation outputs and performance scores at different time steps are sometimes compared at one aggregation time scale, daily or hourly (e.g. Wang et al., 2009) and sometimes without a preliminary aggregation (e.g. Littlewood and Croke, 2008). However, the rankings of model performance at different time steps may depend on the evaluation time scales chosen for

comparative analysis. A comparison across a wide range of evaluation time scales could help find general tendencies or specific behaviours emerging at certain time scales.

The case of artificial reservoirs is an example of an application of hydrological models that could benefit from time steps shorter than the operation model time step. For large flood-control or water-supply reservoirs, their management may only require the forecast of daily inflows, but these daily inflow forecasts may be obtained using an hourly model and aggregating the outputs to obtain the daily inflow. This could provide a better description of flood events, which contribute most of the flow volume. Although this approach is still rare in the reservoir operation literature, examples of its advantages can be found. For example, Schreider and Jakeman (2001) applied the IHACRES model at a 4-hourly time step for ten catchments in the Upper Murray Basin, feeding Hume and Dartmouth lakes, two of Australia's four largest reservoirs. They showed that long-term daily forecasts of streamflows used for reservoirs' operational management can be obtained by aggregating the 4-hourly step simulations with the same or higher accuracy than by daily model simulations.

Although this approach of using shorter time steps intuitively makes sense, there are several reasons that may limit the added value of short time-step data when looking at results at larger aggregated time scales. First the model input data, especially rainfall, may have a lower signal-to-noise ratio at shorter time steps due to the greater difficulty validating data and the greater uncertainty in areal averaged rainfalls (Yu et al., 1997; Obled et al., 2009). Second, catchments behave like low-pass filters, which may smooth out the short-term variability of input and limit the sensitivity of outputs to this additional information. This may be especially true when the characteristic time of studied events is far longer than the time step used (e.g. Obled et al., 2009). The model structure itself may also be less appropriate to catch the greater complexity of processes at shorter time steps, as already expressed for example by Hughes (1993). Last, the averaging effect of output data aggregation may also limit the usefulness of using fine time-step input data. Hence, it is useful to investigate the influence of the time step on modelling results, since there may be a compromise between the expected advantages obtained by refining the inputs and the model time step, and the possible limits affecting model efficiency at shorter time steps more than at the larger evaluation time step.

3.1.3 Scope and structure of the article

The literature review has shown that, despite the general knowledge of the importance of sub-daily variability of rainfall for flood volume modelling, the advantages of using rainfall data at fine temporal resolution for flow simulation are still not well quantified. There is a need for further investigations to evaluate the usefulness of short time-step information for hydrological model simulations, by comparing different model time-step outputs at common aggregated time scales, using a large set of catchments. A parallel can be made between these investigations on the temporal discretization issue and the studies conducted to investigate the impact of refined spatial discretization of catchments on modelling results, which have received more attention in the literature (see e.g. Obled et al., 1994; Das et al., 2008; Lobligeois et al., 2014). Recently, Melsen et al. (2016a) noted the surprising contrast between

the trend of increasing spatial resolutions and the stagnation of time steps used for calibration and validation of the VIC model (based on 192 studies of last two decades), and they advocated the need of a simultaneous refinement of temporal resolutions.

The present research aims at better understanding the usefulness of fine time-step hydro-climatic data for hydrological model performance on a large set of 240 catchments and 2400 flood events. This will allow going beyond the specific conclusions of previous single-catchment tests by investigating the relationship of model performance behaviour at different time steps with a number of catchment and flood event characteristics. Simulations made at different sub-daily time steps (up to 6 min) will be aggregated and compared at selected evaluation time scales going from sub-hourly to daily. This choice is essential to provide a rigorous evaluation framework and to ensure consistency in the comparison. This analysis has the advantage of quantifying the possible improvements in hydrological simulations evaluated at certain aggregated time scales using shorter time-step information.

In Section 3.2, the general testing framework and evaluation methodology followed for comparing simulations at different model time steps are presented. In Section 3.3, the results of model performance at different time steps on the whole catchment set are presented and discussed. The different performance behaviours and the possible explanatory variables behind these differences are investigated. Section 3.4 provides some concluding remarks and discussion on the limitations of this work and the perspectives for further research.

3.2 Testing approach and evaluation methodology

3.2.1 Testing approach for model evaluation at different time steps

The main objective of this study is to analyse the usefulness of using short time step information and modelling when the analysis is performed at larger (aggregated) time steps. To this aim, we applied the testing approach summarized in Fig. 3.1.

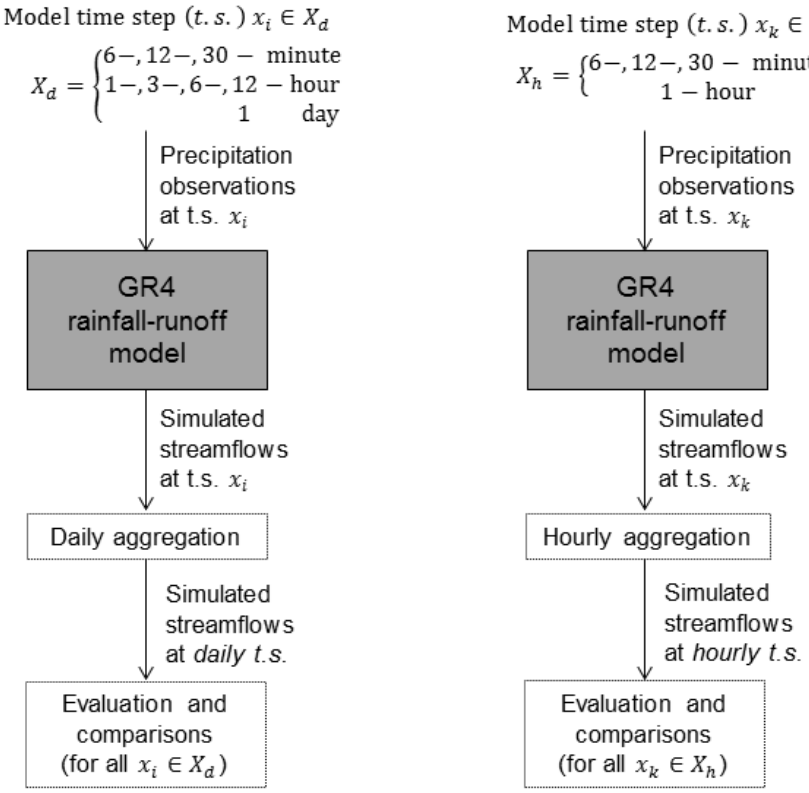


FIGURE 3.1 – Testing approach followed for model evaluation and comparisons at different time steps. Left panel: daily evaluation for simulation time steps from 6 min to 1 day; right panel: hourly evaluation for simulation time steps from 6 min to 1 h.

The GR4 model is run at a short time step using inputs and simulating outputs at the same time step. Then model outputs are aggregated at a larger time step, which is called the model evaluation reference time scale. The model was run at eight time steps: 6, 12, 30 min, 1, 3, 6, 12 h and 1 day (d). So, all time steps are multiples of shorter time steps (except in the case of 30 and 12 min). For an exhaustive comparative evaluation, output series at a certain model time step were aggregated at all the larger time steps: 12, 30 min, 1, 3, 6, 12 h and 1 day. This methodology aims to answer the question: "When evaluated at a certain reference time scale, can model performance benefit from shorter time-step information and modelling?". To test the significance of the differences between model performance at different time steps, statistical tests were performed. Since the evaluation statistics of models at different time steps are paired (i.e. the same basins are used) and not normally distributed, we used the Friedman rank test, also known as the Friedman two-way analysis of variance (Friedman, 1937), to determine whether two or more paired samples have been selected from populations having equal medians.

3.2.2 Calibration and evaluation procedure

The available 8-year data period was split into two test sub-periods (01/08/2005–31/07/2009 and 01/08/2009–31/07/2013). Eight years were judged sufficient to applying a common calibration-validation procedure (e.g. as shown at a daily time step by Merz et al., 2009), following the split-sample test (Klemeš, 1986b) on the two 4-year sub-periods. The models were calibrated on each sub-period, using the Kling–Gupta efficiency (KGE) criterion (Gupta et al., 2009) as objective function, calculated on all the time steps of the sub-period before aggregation at the evaluation reference time scale. The KGE criterion is calculated as:

$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2} \tag{3.1}$$

where r is the linear correlation coefficient between simulated and observed flows; a is the ratio of the standard deviation of simulated flows to the standard deviation of observed flows, i.e. a measure of relative variability between simulated and observed values; b is the ratio of the mean of simulated flows to the mean of observed flows, i.e. an index of the overall bias in water balance. The 'ideal' value for KGE and its three components (a, b, r) is 1.

The model evaluation is presented here only on the flood events of the validation sub-period after temporal aggregation at each reference time scale. The overall evaluation was performed with the KGE criterion and its three components (variability ratio a, ratio of means b and correlation r). To ensure the significance of these aggregated statistics (KGE, a, b and r), these criteria were calculated on the ten events taken altogether for each catchment (and not on each single event individually). Moreover, for events with durations shorter than three days, the event start and end were adjusted to have an evaluation period lasting at least three days.

For the comparison of simulations at different model time steps, we used the relative KGE Index (R_{KGE}) as formulated by Lerat et al. (2012) to compare two modelling options, here called alternatives r, i.e. simulation at the reference time step r, and x, i.e. simulation at the shorter time step $x<r$:

$$R_{KGE}(x|r) = \frac{m(q_{obs}, \hat{q}_r) - m(q_{obs}, \hat{q}_x)}{m(q_{obs}, \hat{q}_r) + m(q_{obs}, \hat{q}_x)} \tag{3.2}$$

where $m = 1 - KGE$ is a metric measuring the discrepancies between simulated and observed flows, with m=0 for a perfect simulation; q_{obs} is the observed streamflow at the reference time step r; $\hat{q}_r$ is the simulated flow using the model at the reference time step r and $\hat{q}_x$ is the flow simulated at time step $x<r$ and then aggregated at time step r. R_{KGE} ranges from −1 to 1, with positive values corresponding to better performance at time step x with respect to the reference, and vice-versa for negative values. The interpretation of this index is further detailed in Table 3.1 adapted from Lerat et al. (2012). The bounded range of the index is an advantage for calculating mean results over large sets of catchments (e.g. Mathevet et al., 2006).

$R_{KGE}(x\|r)$	$m(q_{obs}, \hat{q}_x)/\, m(q_{obs}, \hat{q}_r)$	Interpretation
1	0	Simulated flow (q_{sim}) at time step (t.s.) x is perfect according to metric m
0.5	1/3	Error metric m for q_{sim} at t.s. x is 3 times smaller than for reference r
0.33	1/2	Error metric m for q_{sim} at t.s. x is twice as small as for reference r
0.05	9/10	Error metric m for q_{sim} at t.s. x is 10% smaller than for reference r
0	1	Error metric m for q_{sim} at t.s. x is equal to that of reference r
−0.05	10/9	Error metric m for q_{sim} at t.s. x is 10% larger than for reference r
−0.33	2	Error metric m for q_{sim} at t.s. x is twice as large as for reference r
−0.5	3	Error metric m for q_{sim} at t.s. x is 3 times larger than for reference r
−1	+∞	q_{sim} at t.s. r is perfect according to metric m

TABLE 3.1 – Summary for interpretation of the relative KGE index, R_{KGE} , between simulations at the reference time step r and the shorter time step x using the error metric $\boldsymbol{m(q,\hat{q}) = 1 - KGE(q,\hat{q})}$ (Lerat et al., 2012).

A complementary assessment was carried out on each flood event individually to evaluate the proper identification of flood peak magnitude and timing. This was based on three event-based criteria, namely $\Delta Q_P, \Delta t_P$, and VE, defined as:

$$\Delta Q_P = \frac{Q_P^{sim} - Q_P^{obs}}{Q_P^{obs}} \tag{3.3}$$

$$\Delta t_P = t\left(Q_P^{obs}\right) - t\left(Q_P^{sim}\right) \tag{3.4}$$

$$VE = 1 - \frac{\sum_{i=1}^{n} \left| q_{sim,i} - q_{obs,i} \right|}{\sum_{i=1}^{n} q_{obs,i}} \tag{3.5}$$

where ΔQ_P [-] is the 'peak flow error' that has an ideal value at zero and is positive for peaks overestimation and negative for underestimation; Q_P^{sim} and Q_P^{obs} are the simulated and observed peak flow amplitudes, respectively; Δt_P [time steps] is the 'time to peak error' that has an ideal value at zero; $t\left(Q_P^{obs}\right)$ and $t\left(Q_P^{sim}\right)$ are the time steps of the observed and simulated peaks, respectively; VE [-] is the 'volumetric efficiency', proposed by Criss and Winston (2008), that has an ideal value at 1. For an unbiased model, VE represents the fraction of water delivered at proper time along a time horizon (i.e. the flood event period of *n* time steps, in our case), so it is a measure of the overall goodness of simulated streamflows timing.

3.3 Results and discussion

3.3.1 Performance with time step over the whole catchment set

In this section, we present a summary of the results of our testing approach for the whole catchment set by evaluating over the selected flood events the outputs obtained at different model time steps and then aggregated at the reference evaluation time scales. Figure 3.2

shows a summary of how model performance evolved when evaluated at a daily reference scale by the relative KGE Index (R_{KGE}) as the model time step varies from daily to 6 min (from right to left on the horizontal axis). The median R_{KGE} Index is positive at all the sub-daily time steps (Fig. 3.2) and reaches its maximum at 6 h with R_{KGE}=0.05, meaning that the error metric (1-KGE) is 10% smaller than for the daily reference (see Table 3.1).

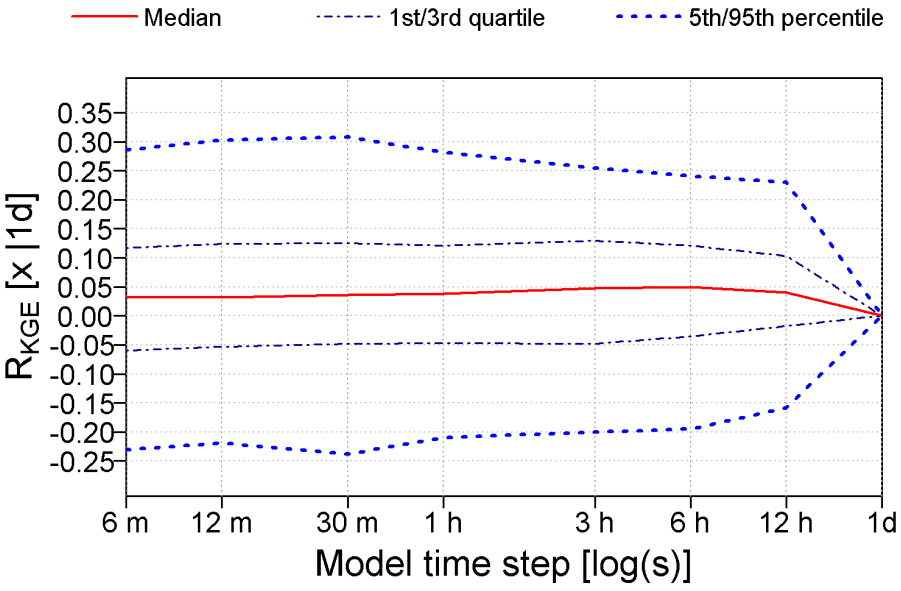


FIGURE 3.2 – Summary of the relative KGE index normalized with daily reference over the whole catchment sample as the model time step changes from 6 min to 1 day.

Figure 3.2 suggests that the model performance at sub-daily time steps evaluated on a daily basis is higher than at the daily time step on average over the entire catchment set and this improvement seems to stabilize from 12-h or 6-h time steps. The Friedman test applied on the KGE indices at different time steps confirms the significance of performance differences between the median KGE values (p-value < 2.2e-16) and the trends in performance seen in Fig. 3.2: (i) the performance greatly improves passing from 1 day to 12 h, and the performance upper bound is reached at the 6-h time step, for which the median KGE is significantly higher than at all other time steps, except the 3-h time step; (ii) a trend reversal (KGE degradation) is found going from the 1-h to the 6-min time step, which results in a statistically non-significant difference between KGE from the 6-min and 1-day model time steps (at significance level 0.05).

The distinctive KGE components (variability ratio a, ratio of means b and correlation r) were used to evaluate whether the differences in performance at different time steps can be related more to the variability, balance or timing of streamflow. Figure 3.3 (a-b-c) shows a summary of the statistics of how these three components evolved as evaluated at the daily reference scale over the time steps tested. Note that the optimal value for the three components is 1. On average over the whole sample, the relative variability and the correlation at the daily time step are improved by going towards shorter time steps, while on the contrary the capacity of the model to reproduce the water balance (ratio of means b) has slightly deteriorated. It

therefore seems that reducing the time step has the counter-effect of limiting the capacity of the model to maintain a good water balance simulation in high-flow conditions. However, note that: (i) the median bias over floods was already significant at the daily time step (b=0.935) and worsens of 0.036 points passing to 6-min time step; (ii) when evaluated on the whole time series, the median ratio of means (b) is stable at 1 for all time steps (not shown).

Table 3.2 reports a summary of the evaluation by the three event-based criteria ($\Delta Q_P, \Delta t_P$, and VE), calculated on each flood event after aggregation of the simulated hydrographs at the daily reference time scale. All the event-based performance criteria improved as the model time step becomes shorter. The largest improvement is obtained when passing from 1 day to 12 h. The negative median peak flow errors show that, on average, simulated peaks are under-estimated by the model at all time steps, but are closer to observed peaks at shorter time steps. The overall timing of the hydrographs is better reproduced at sub-daily time steps with respect to daily, as reflected by the VE values. The flood peak timing is improved at shorter time steps on a significant part of the 2400 events (e.g. passing from daily to 6-h or shorter time steps, Δt_P becomes null for more than 300 events). This improvement concerns the cases of advanced ($\Delta t_P \geq 1d$) or delayed ($\Delta t_P \leq 1d$) simulated peaks, that are reduced at sub-daily time steps, and is more significant considering only events with sub-daily lag times (e.g. Δt_P is null respectively for 60% and 79% of the 1079 events with lag time shorter than 12h at daily and 6-h model time step). This expected result confirms previous findings in the hydrological literature, illustrating that daily models inevitably fail to correctly reproduce hydrograph timing due to the aggregation effect at the daily scale (e.g. Asadzadeh et al., 2016).

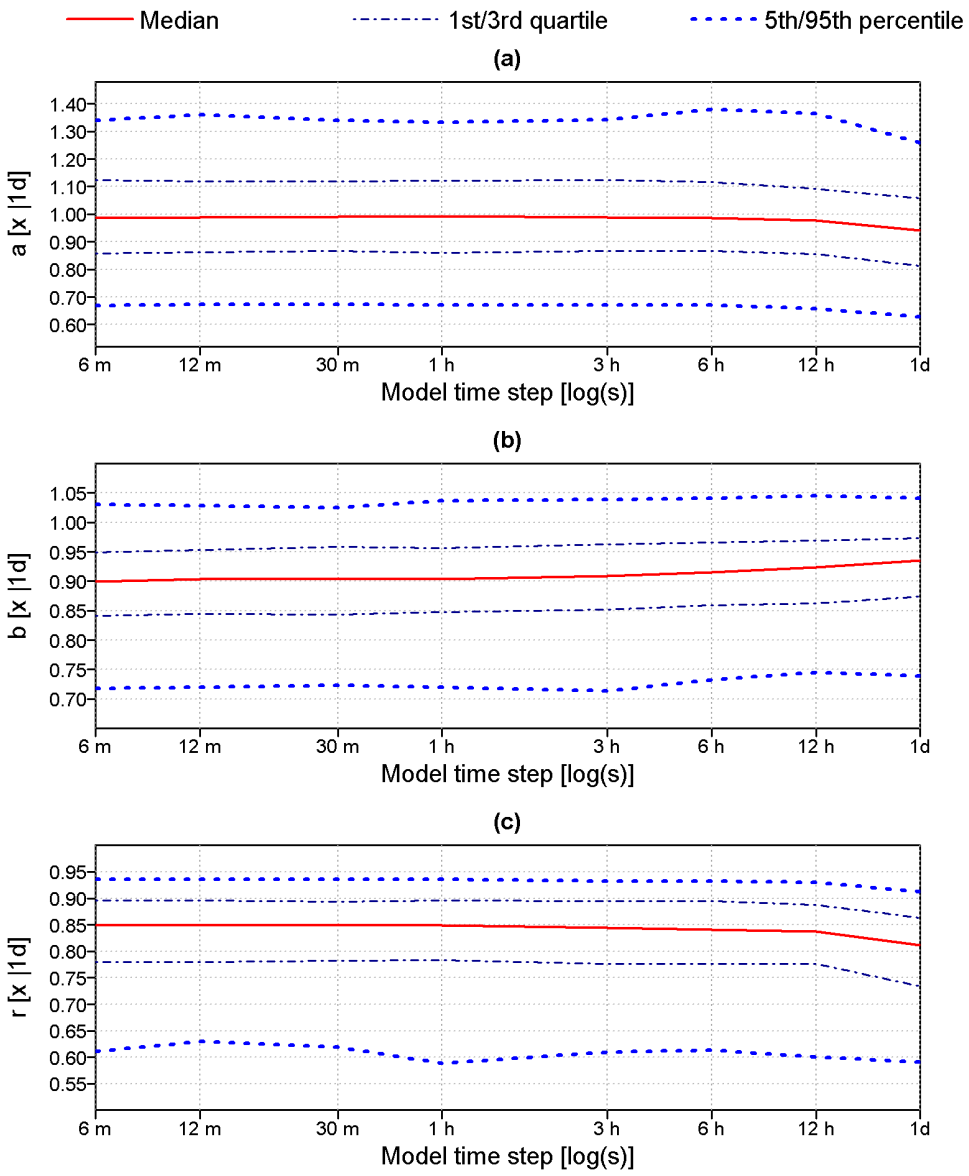


FIGURE 3.3 – Summary of the KGE components evaluated on flood events at the daily reference scale over the whole catchment sample as the model time step changes from 6 min to 1 day. (a) Relative variability, *a*; (b) ratio of means, *b*; (c) correlation, *r*.

	Model time step							
	6-min	**12-min**	**30-min**	**1-h**	**3-h**	**6-h**	**12-h**	**1-d**
Median peak flow error, ΔQ_P [%]	*-10.5* %	-10.6 %	-10.9 %	-10.6 %	-10.7 %	-11.2 %	-12.6 %	-17.1 %
Median volumetric efficiency, VE [-]	0.759	0.760	0.759	0.759	0.761	*0.763*	0.756	0.737
Number of events with null time to peak error, $\Delta t_P = 0\ d$ (% on the whole set)	1760 (73.3%)	*1764 (73.5%)*	1758 (73.3%)	1760 (73.3%)	1747 (72.8%)	1748 (72.8%)	1680 (70.0%)	1443 (60.1%)
Number of events with positive time to peak error $\Delta t_P \geq 1\ d$ (% on the whole set)	351 (14.6%)	*348 (14.5%)*	352 (14.7%)	353 (14.7%)	357 (14.9%)	362 (15.1%)	387 (16.1%)	480 (20.0%)
Number of events with negative time to peak error $\Delta t_P \leq -1\ d$ (% on the whole set)	289 (12.0%)	288 (12.0%)	290 (12.1%)	*287 (12.0%)*	296 (12.3%)	290 (12.1%)	333 (13.9%)	477 (19.9%)

TABLE 3.2 – Summary of model performance over the 2400 flood events set, evaluated at the daily reference time scale, by the three event-based criteria, $\Delta Q_P, \Delta t_P$, and VE (best values are highlighted in *italics*).

The same simulation results were evaluated at all the reference time scales from 12 h to 12 min. Figure 3.4 shows that the distribution of the relative KGE index has analogous behaviour at all the reference time scales. The R_{KGE} distribution is symmetrical around zero, which means that the performance improved by using shorter model time steps for approximately one-half of the catchments and that it deteriorated for the other half. This result may appear a bit paradoxical since one could expect that a shorter time step systematically yields performance improvement by providing additional information. However, this can be explained by the limits of the added value of short time-step data discussed in the introduction (see Section 3.1.2).

The median R_{KGE} index is quite stable at all model time steps for all the reference time scales, being slightly positive just for the 12-h reference time scale (as it was for 1 day) and slightly negative at all other reference scales. This is in line with the results found at the daily reference time scale for which KGE median values of the 6-h time-step model were significantly higher than other model time steps. Hence these results indicate that, for reference time steps shorter than or equal to 6 h, on average and on our catchment set, there is no added value of using shorter model time steps (when evaluation is based on outputs aggregated at the reference time scales).

As the reference time scale decreases, the interquartile range is narrower around zero, meaning that the differences in performance become smaller. This could be expected, since

the time scales are increasingly closer (note the logarithmic scale of the x-axis in Fig. 3.4). For the extreme behaviours, the largest jumps in performance are found passing from the reference time scale to a model time step equal to one-half or one-third of the reference. After that, there is a slower trend, or a plateau of performance is reached.

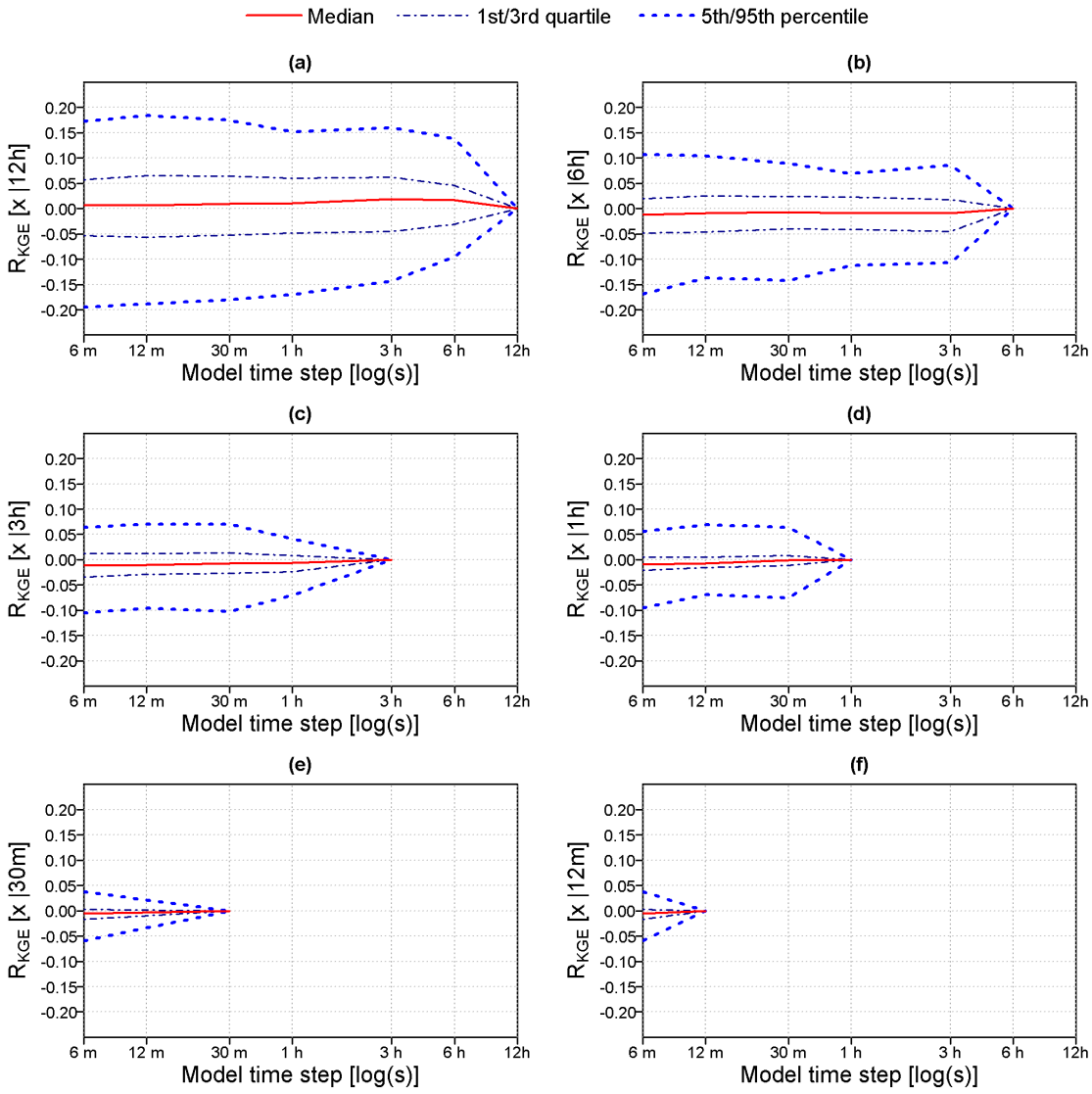


FIGURE 3.4 – Summary over the whole catchment set of the relative KGE Index normalized to different reference time scales: (a) 12 h; (b) 6 h; (c) 3 h; (d) 1 h; (e) 30 min; (f) 12 min.

Like the daily evaluation, the Friedman test applied on the KGE indices evaluated at all the different reference time scales confirmed the presence of significant differences between the median KGE values at different model time steps (with a *p*-value < 2.2e-16). The ranking of the results from different model time steps is coherent with that found at the daily evaluation time step.

The evaluation by the event-based criteria at all the reference time scales leads to similar conclusions. Consistently with the daily evaluation, from 12-h to 3-h reference time scales, the identification of flood peaks and timing is improved at shorter time steps. For example, for the 6-h reference, the median error $\Delta \boldsymbol{Q_P}$ is reduced from -0.17 at 6-h time step to -0.13 at 1-h or shorter, $\boldsymbol{VE}$ is stable, and $\Delta \boldsymbol{t_P}$ becomes null for more than 430 events at 1-h time step or shorter. The differences in performance are smaller as the reference scale decreases, and vanish from 1-h or shorter. For the hourly reference, median values of $\Delta \boldsymbol{Q_P}$ and $\boldsymbol{VE}$ are stable (at -0.14 and 0.73, respectively) for all shorter model time steps.

3.3.2 Behavioural catchment clusters for the evolution of model performance

For more detailed analysis, the daily reference was chosen because of the greater differences between performance at the daily reference scale and at shorter time steps. We distinguished three types of performance behaviours with time steps: (i) improvement, (ii) insensitivity and (iii) degradation of the KGE Index as the time step becomes shorter. We partitioned the set of catchments into distinct groups using clustering analysis techniques. The optimal number of clusters was confirmed to be three, based on the optimum average silhouette width and the Calinski-Harabasz criteria (Caliński and Harabasz, 1974). To partition the catchment behaviours, we used the K-means algorithm, a well-known iterative algorithm solving the clustering problem by minimizing a sum of squared Euclidean distances from the centroid of the groups. The result of this clustering method is shown in Fig. 3.5 in terms of the changes with time step in the relative KGE Index at the daily reference scale. A statistical summary of the three catchment clusters is reported in Table 3.3. It shows that half of the catchments are not statistically sensitive to the time step (not at all or only slightly) and that a similar proportion (about one-fourth each) is subject to either improvement or degradation, with a larger number of improvements.

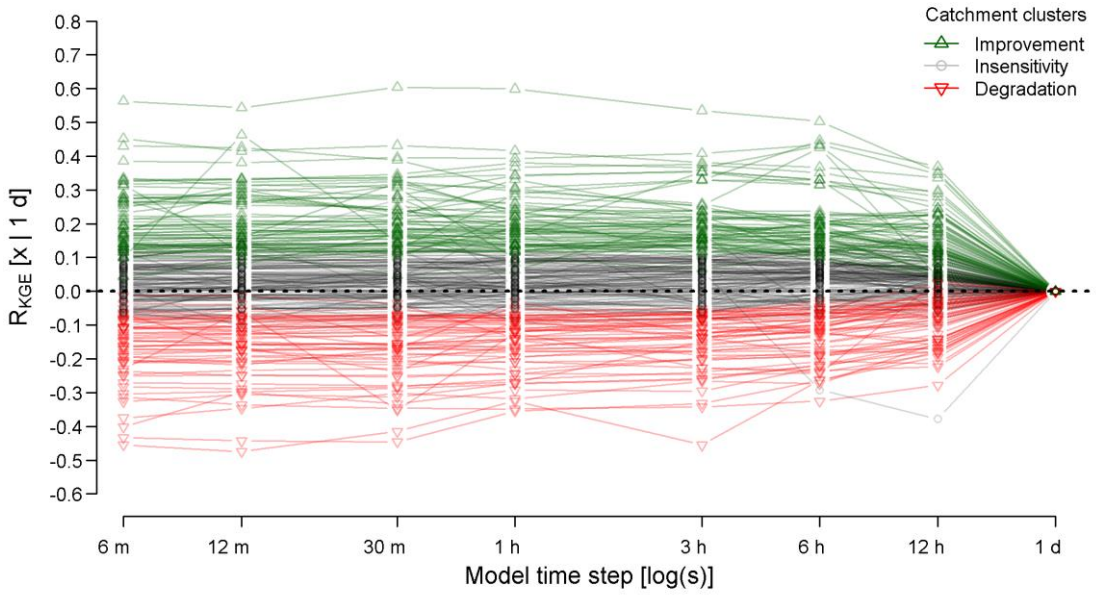


FIGURE 3.5 – Partitioning of model performance with time steps over the 240 catchments (one line per catchment) in three groups using the *k-means* algorithm. The three clusters correspond to (i) improvement, (ii) insensitivity or (iii) degradation of the KGE Index as the time step becomes finer.

	Improvement cluster	**Insensitivity cluster**	**Degradation cluster**
Min, median, max relative KGE index values of 1-h model (with daily ref.)	0.06	−0.09	−0.36
	0.19	0.03	−0.14
	0.60	0.20	−0.03
Number of catchments (% on the whole set)	65 (27%)	122 (51%)	53 (22%)
Median KGE of daily model (evaluated at daily ref.)	0.708	0.700	0.795
Median KGE of 12-h model (evaluated at daily ref.)	0.785	0.712	0.762
Median KGE of 6-h model (evaluated at daily ref.)	0.815	0.720	0.742
Median KGE of 3-h model (evaluated at daily ref.)	0.804	0.713	0.720
Median KGE of 1-h model (evaluated at daily ref.)	0.803	0.716	0.735
Median KGE of 30-min model (evaluated at daily ref.)	0.804	0.716	0.697
Median KGE of 12-min model (evaluated at daily ref.)	0.799	0.722	0.692
Median KGE of 6-min model (evaluated at daily ref.)	0.798	0.718	0.698

TABLE 3.3 – Statistical summary of the three behavioural catchment clusters of model performance evolution, with median KGE values of simulations at different time steps, evaluated on flood events at the daily reference scale.

Figure 3.6(a–d) shows a summary of the distribution of the KGE Index and its three components (*a*, *b*, *r*) over the three catchment clusters for the two extreme model time steps (daily and 6-min) evaluated at the daily reference. Figure 3.6(a) shows that, on average, the

improvement catchment cluster presents lower KGE values for the daily reference model than the degradation cluster. On these catchments, the margin of progress is therefore greater, and the large time step may be an actual limitation for the model to reach higher performance. For the improvement cluster, the KGE increase is related to an improvement in the relative variability and correlation (i.e. at the 6-min time step, *a* and *r* values are closer to 1), while the average ratio of means (*b*) slightly deteriorates. For the degradation cluster, the drop in performance is mainly explained by a counterproductive augmentation of the relative variability, which was already too high ($a>1$), and a degradation of the bias (ratio of means *b*, which was already too low), while the correlation is fairly stable. The Kruskal-Wallis non-parametric test (Kruskal and Wallis, 1952) was used to evaluate whether the mean ranks of the KGE Index and its components are significantly different between the catchment clusters. The test confirmed the significance in differences (at the 0.05 significance level) between the three groups of all the performance indices at the reference and finest time scales (daily and 6-min), except the daily bias. The most significant differences are in the relative variability component.

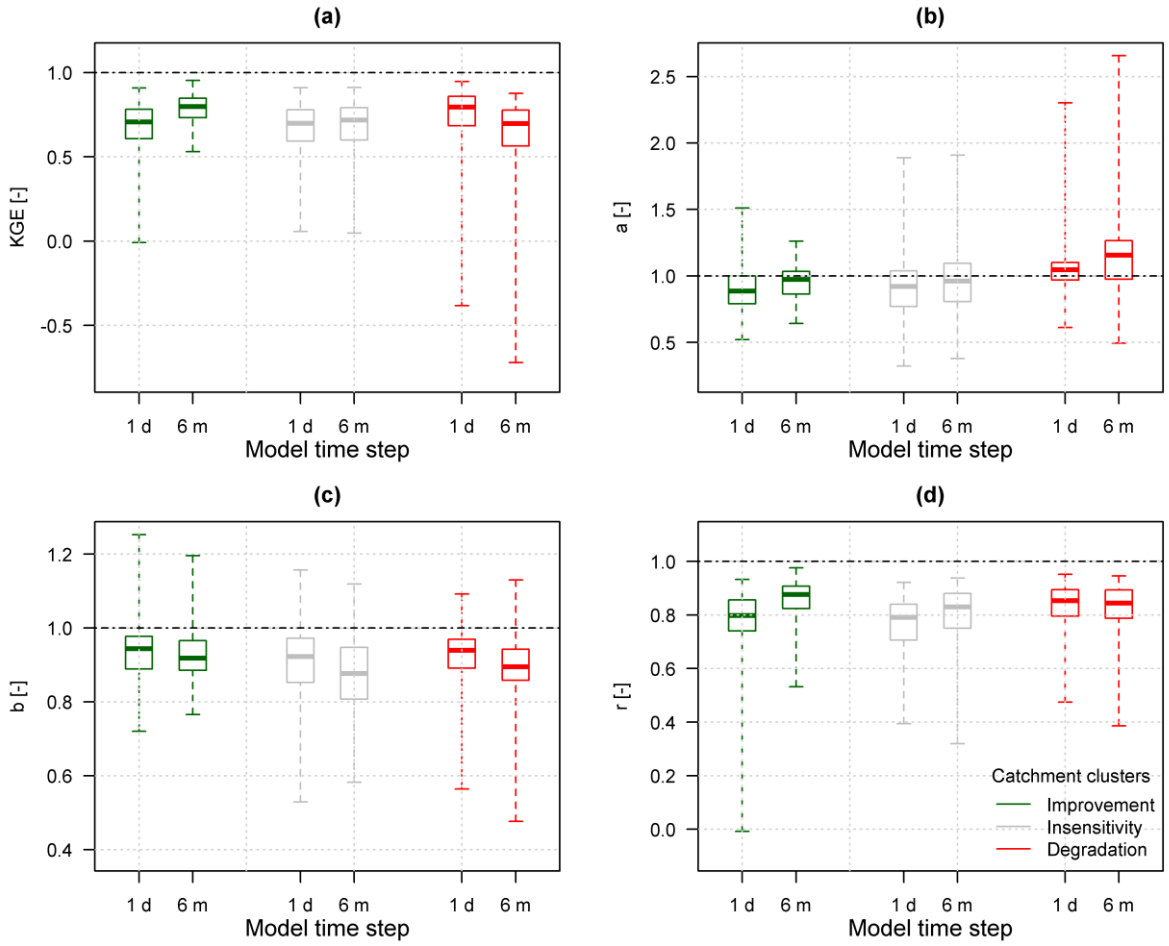


FIGURE 3.6 – Distribution over the three catchment clusters of the performance scores of the daily and 6-min models evaluated at the daily reference scale: (a) KGE; (b) relative variability, *a*; (c) ratio of means, *b*; and (d) correlation, *r*. The box plots report the median value and interquartile range, and the whiskers represent the minimum and maximum values. The optimal score (1) is highlighted by a horizontal dashed line.

3.3.3 Understanding the causes underlying the behavioural catchment clusters

To better understand the differences between clusters, we first searched for possible regional trends. The geographical distribution of clusters is shown in Fig. 3.7. Except for a cluster of degrading catchments in northwestern France, no strong regional trend could be identified. This may indicate that local physical characteristics have a greater influence than regional climate-based influences. The comparison of Fig. 3.7(c) and Fig. 2.1 (with the network of 6-min rain gauges, see Chapter 2) also shows that the zones with lower benefits (or degradation) are not systematically the zones with lower density of rain gauges (see for example the south-east Mediterranean arc).

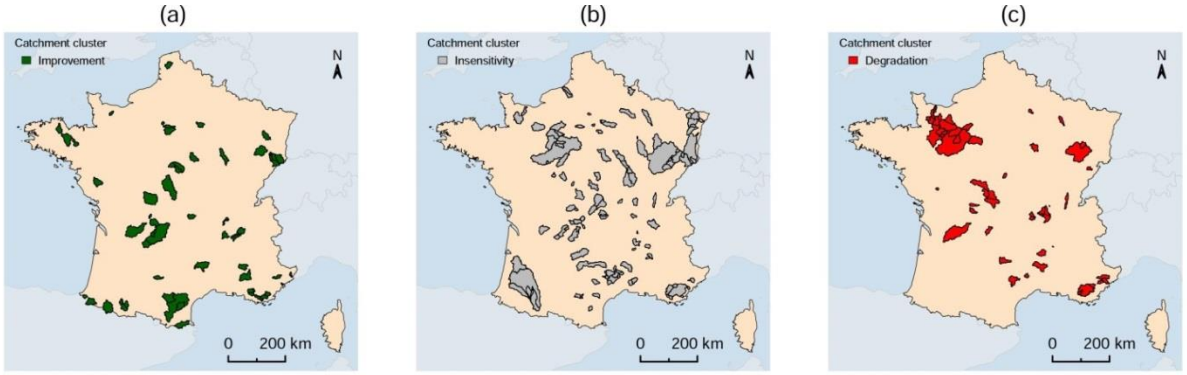


FIGURE 3.7 – Geographical distribution of the three behavioural catchment clusters corresponding to (a) improvement, (b) insensitivity and (c) degradation of the KGE Index as the time step becomes finer.

Therefore, the different behavioural catchment clusters were analysed in relation to 20 explanatory variables, i.e. the morphological and hydro-climatic (MH) characteristics, the rain gauges density (RGD) indicators (see Chapter 2, Table 2.1) and the flood event-based (FE) characteristics (see Chapter 2, Table 2.2). The Kruskal-Wallis test was performed to decide whether the distributions of these variables are significantly different for the three catchment clusters. The rejection of the null hypothesis ($p<0.05$) means that the explanatory variable considered in the test is likely to have an imprint on the evolution of model performance with the time step. Table 3.4 reports the results of Kruskall-Wallis test on the explanatory variables' distributions and their median values over the three catchment clusters. The characteristics that turn out to have significant differences among catchment clusters (at significance level 0.05) are mostly flood event characteristics. Only the event's amount of precipitation does not appear among the most influential FE characteristics. None of the morphological or hydro-climatic descriptors showed significant differences between clusters, except the streamflow auto-correlation, but this is highly dependent on the dynamics of flood events. The smaller catchments could be expected to present faster responses to precipitations and to benefit more from fine temporal resolution of rainfall but this was not the case on our catchment set. None of the 6-min rain gauges density indicators showed significant differences between catchment clusters of model performance. This surprising result could be related to other factors impacting the quality of rainfall data that we did not consider in our analysis (e.g. measurement errors of 6-min rain gauges, uncertainty on daily amount, rain gauges spatial distribution, etc.). We remind also that the spatial resolution of the daily data used is 64 km^2, so the uncertainty of daily precipitation is larger for smaller catchments. This could partly explain the non-significant difference of catchment area and rain gauges density indicators between clusters.

For the influential characteristics, the clusters' ranking is almost completely monotonic (i.e. the same along the entire cumulative distribution), as shown in Fig. 3.8(a-h). As expected, the model performance tends to benefit more from finer time steps for rapid basins with a shorter response time and higher streamflows and rainfall temporal variability, i.e. a lower GOUE Index, a lower daily flow auto-correlation, and a higher flow shape coefficient and gradient. Also, for catchments with shorter flood and storm duration the performance significantly

benefits from finer time steps. Conversely, performance tends to be insensitive or deteriorate at shorter time steps for slower catchments with highly auto-correlated streamflows and longer flood events. Note that the GOUE Index calculated on streamflows proves more significant than the one calculated on rainfall, as shown by the different orders of magnitude of the p-value in Table 3.4 and the distance of cluster distributions in Fig. 3.8(c) and (h). This is in line with the fact that the rainfall temporal variability is not sufficiently smoothed out by the model for the degradation cluster's catchments, as shown by the over-estimated flow variability discussed in the previous section.

Note also that some of these significant characteristics are correlated, as discussed in Chapter 2, but we chose to keep all of them in the analyses. This choice is motivated by the fact that setting a discriminating threshold of correlation would be an excessively arbitrary choice, because some characteristics prove more significant than others even if highly correlated with each other. This is proved by the different orders of magnitude of p-values for the correlated flood-based characteristics. Moreover, some characteristics, such as catchment area and hydraulic length, do not show an imprint in the evolution of model performance, even if quite well correlated with the rainfall-runoff lag time ($r=0.61$ and $r=0.68$).

As complementary analysis, we performed the statistical test by considering different classes of catchments, in order to search for possible patterns masked by considering all the catchments together. We split the total set of 240 catchments into sub-sets of 120 catchments, by considering the median value of each characteristic that did not show significant differences between clusters (MH & RGD). On each sub-set we performed the Kruskal-Wallis test to decide whether the distributions of the variables (MH & RGD) not used for sub-setting were significantly different between catchment clusters. Moreover, we checked whether the clusters' ranking was the same along the cumulative distributions of the characteristics subject to testing (as in Fig. 3.8). Only the daily precipitation intensity coefficient emerged as influential for some conditions (at significance level 0.05): (i) for catchments with low altitudes ($z<363$m); (ii) for catchments with low runoff coefficient ($RC<0.34$). This analysis, still non-exhaustive, shows that other masked general patterns are not easy to detect by considering only sub-sets of particular catchments.

#	Explanatory variable	Median by cluster			*p-value*
		(i) Improvement	(ii) Insensitivity	(iii) Degradation	
MH9	*Flow auto-correlation at 24 h [-]*	*0.80*	*0.85*	*0.88*	*9.4×10^{-6}*
FE1	*Flood duration [h]*	*79.00*	*116.25*	*136.50*	*9.7×10^{-6}*
FE8	*GOUE Index of daily streamflow (at 6-min ref.) [-]*	*0.59*	*0.73*	*0.80*	*9.3×10^{-5}*
FE4	*24-h flow shape coefficient [-]*	*1.38*	*1.20*	*1.17*	*1.3×10^{-4}*
FE2	*Storm duration [h]*	*26.50*	*35.75*	*40.00*	*8.1×10^{-4}*
FE5	*Mean flow gradient [mm/h²]*	*0.018*	*0.009*	*0.012*	*6.7×10^{-3}*
FE6	*Rainfall-runoff lag time [h]*	*9.55*	*14.13*	*15.40*	*9.3×10^{-3}*
FE7	*GOUE Index of daily precipitation (at 6-min ref.) [-]*	*0.13*	*0.18*	*0.17*	*2.0×10^{-2}*
RGD2	Average rain gauge area [km²]	289.05	316.91	270.23	2.1×10^{-1}
MH8	Base Flow Index [-]	0.58	0.57	0.55	2.9×10^{-1}
MH6	Runoff coefficient [-]	0.34	0.33	0.38	3.6×10^{-1}
MH10	Daily precipitation intensity coefficient [-]	9.29	9.22	9.19	4.2×10^{-1}
MH4	Average slope [-]	0.09	0.08	0.07	4.5×10^{-1}
FE3	Event amount of precipitation [mm]	54.14	63.02	65.55	4.6×10^{-1}
RGD1	Number of rain gauges influencing the catchment [-]	5	5	5	5.1×10^{-1}
MH5	Topographic Index [-]	13.02	13.38	13.24	8.3×10^{-1}
MH2	Hydraulic length [km]	24.27	24.66	24.91	8.6×10^{-1}
MH7	Aridity Index [-]	0.78	0.75	0.76	9.2×10^{-1}
MH3	Average altitude [m]	371.38	365.04	325.19	9.3×10^{-1}
MH1	Catchment area [km²]	381.56	332.43	358.51	9.8×10^{-1}

TABLE 3.4 – Results of Kruskal-Wallis tests on the differences of potential explanatory characteristics (flood event-based (FE), morphological and hydro-climatic (MH) characteristics and rain gauge density (RGD) indicators) for the three behavioural catchment clusters ordered by *p*-value. The first eight characteristics highlighted in italics show significantly different medians among catchment clusters at a significance level 0.05.

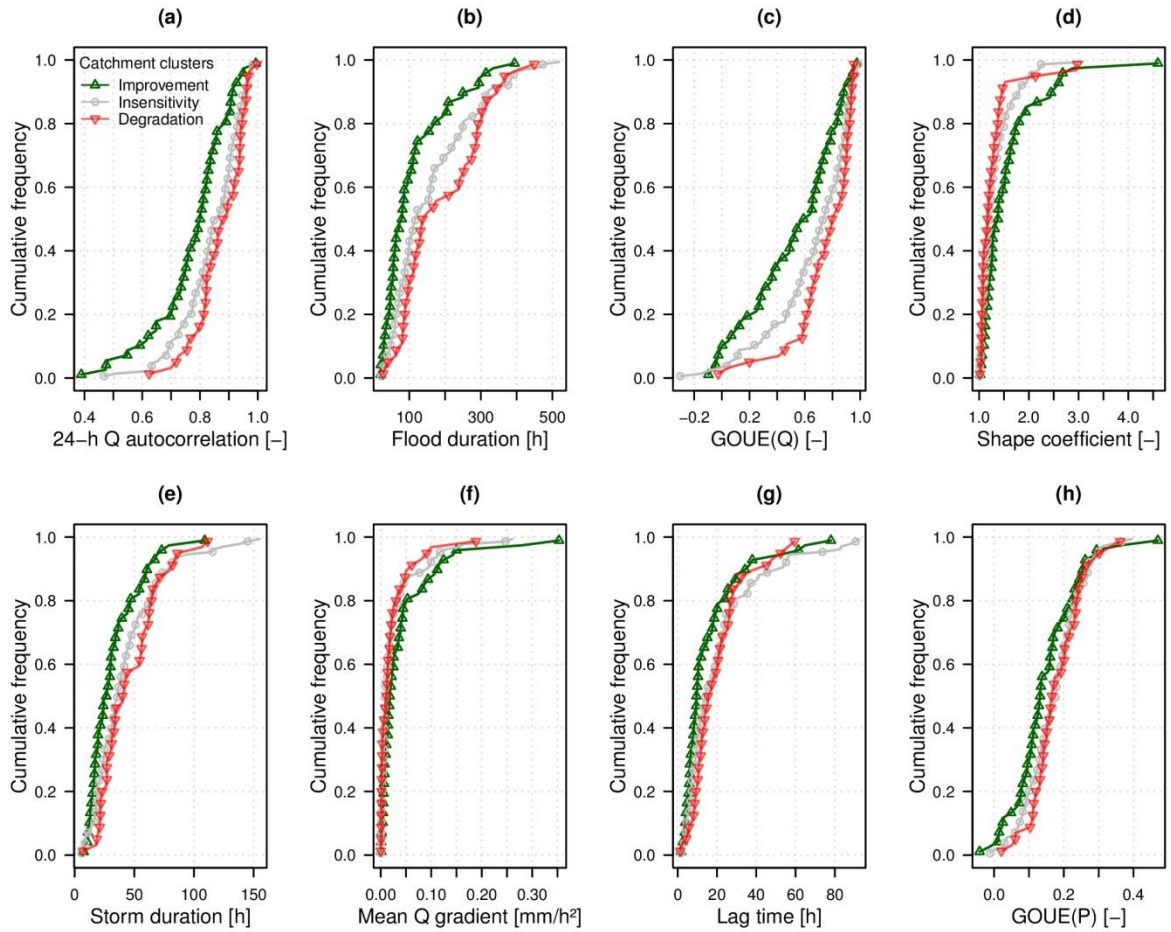


FIGURE 3.8 – Cumulative distributions of the eight characteristics presenting significant differences among the three behavioural catchment clusters: (a) flow auto-correlation at 24 h, (b) flood duration, (c) GOUE Index of daily streamflow (at 6-min reference), (d) 24-h flow shape coefficient, (e) storm duration, (f) mean flow gradient, (g) rainfall-runoff lag time, (h) GOUE Index of daily rainfall (at 6-min reference).

3.4 Conclusions and perspectives

3.4.1 Summary

This research aimed at understanding the potential hydrological benefit of short time step hydro-climatic data to simulate streamflow, over a large and varied set of 240 catchments and 2400 flood events. A lumped rainfall-runoff model was run at eight time steps from 6 min to 1 day. The modelling results at different time steps were aggregated and compared at different evaluation time scales ranging from sub-hourly to daily. The statistical analysis of the simulation results on the whole catchment set showed significant differences in model performance at different time steps.

On average, the use of shorter time steps significantly improved streamflow simulations when evaluating on a daily or 12-h reference time scale, with the greatest improvements obtained

for a time step of 6 h. At shorter reference time scales (3 h or shorter), the differences in performance are smaller and there is a trend reversal, i.e. a slight average degradation of performance at finer time steps on our catchment set. At all the evaluation reference time scales, three types of behaviours of performance evolution with time step can be found: (i) improvement, (ii) insensitivity and (iii) degradation of the KGE Index as the time step becomes shorter. At all reference time scales, the largest jumps in performance are found when changing to a model time step equal to one-half or one-third of the reference. However, as the reference time scale decreases, performance varies less.

An assessment of the ability of the model to correctly reproduce peaks magnitude and timing over the 2400 flood events showed a marked improvement passing from daily to sub-daily time steps for both aspects. In particular, the results highlight that a daily model time step provides poorer results in representing flood peaks and timing, even when the simulation is evaluated at a daily scale.

The three different classes of model performance behaviour were further analysed at the daily reference time scale, for which the largest performance differences were found. The KGE components of relative variability and correlation are significantly different among catchment clusters at the two extreme model time steps (daily and 6 min), and are improved or degraded at shorter time steps depending on the catchment clusters. In contrast, the daily bias is not significantly different among clusters and is slightly degraded on average over all three clusters.

The relationship of the different behaviours with a number of catchment and flood event descriptors was investigated by statistical tests. Some characteristics showed significant differences between behavioural catchment clusters: the 24-h flow auto-correlation, the flood and storm duration, the rainfall-runoff lag time, the flow gradient and shape coefficient and the GOUE Index of temporal variability in rainfall and streamflow series. On average, model performance is significantly improved at finer time steps for the basins presenting shorter response times and flood durations, lower streamflow auto-correlation, and shorter and highly variable storm events, i.e. low GOUE(P) values. On the other hand, for the catchments characterized by higher flow auto-correlation and longer but less intense events, model performance is degraded at shorter time steps, on average, mainly because of the excessive variability of the simulated flows.

3.4.2 Limits of this study and further research needs

The positive and original aspect of this work is the use of a large number of catchments and an evaluation methodology based on different reference time scales, for the sake of seeking general conclusions. The results confirmed that working on different catchments and time scales is necessary, since different model performance dependencies on time step were found over different catchments and across the range of reference time scales. However, the results may still be affected by certain limitations.

First, the results may be dependent on the single model structure used in this study across different catchments and time steps. The GR4 model is a simple but robust model, which has already demonstrated good performance across a wide variety of catchments at daily and hourly time steps (e.g. Perrin et al., 2003; Le Moine et al., 2007; Van Esse et al., 2013). However, we know that some of our colleagues would prefer to adapt the model structure (e.g. Fenicia et al., 2011; Van Esse et al., 2013). Also, the same model structure was applied at all time steps by simply adapting certain time step-dependent parameters, with average good performance at all time steps. This is in line with previous studies on the GR4 model at daily and hourly time steps (e.g. Le Moine, 2008). The coherent and good performance across time steps confirms that the GR4 structure is flexible enough to run properly at daily and sub-daily time steps as well. However, since the GR4 model was originally developed (and more extensively tested) at a daily time step, some refinements of the model structure may still be necessary at sub-daily time steps. Specifically, it was noted that the bias (evaluated on flood events) tended to slightly degrade from daily to shorter time steps, i.e. the model capacity of controlling the water balance is slightly degraded at sub-daily time steps. Thus the proposed model testing approach across time scales proves to be a valuable tool to diagnose model deficiencies. The performance of GR4 model at sub-daily time steps would rise if a solution was brought to the water balance degradation issue (and this would reduce the number of surprising cases where overall performance degraded with shorter time steps). Thus, the actual benefits in using refined resolution input data may be underestimated by the use of the current fixed structure of GR4 at sub-daily time steps. More research on large data sets using other models is still needed to improve our quantitative understanding of the benefits of using shorter time step data.

The results may also partially depend on the calibration and evaluation methodology used. In this study, a simple calibration function based on the KGE Index calculated on streamflows at the model time step resolution was used. The choice of the objective function may be partially responsible for the bias degradation on flood events. Indeed the simulations at shorter time steps may be more constrained to follow the fine temporal dynamics than to reproduce the long-term water balance. Hence some adaptation of the objective function may be needed when changing model time step and this could help solve some cases of degradation when evaluating at longer time steps.

Last, another limitation of this study is the application of only a lumped approach for spatial discretization. This does not allow considering the spatial heterogeneity in precipitation over the catchment, which is greater as the time step decreases, and the interactions between spatial and temporal discretization. Therefore, further research will explore the extension of the testing framework presented at multiple time steps with a semi-distributed modelling approach, in line for example with the study reported by Lobligeois et al. (2014), thus investigating the effect of spatial and temporal discretization simultaneously.

3.5 Acknowledgments

The authors thank Météo-France and SCHAPI for providing the climatic and hydrological data, respectively, used in this study. SCHAPI is also acknowledged for providing financial support to the first author. We would like to thank two anonymous reviewers and Bernard Cappelaere for their useful comments that helped to improve the manuscript. Linda Northrup is also thanked for her correction of the manuscript.

4
Model diagnostics at multiple time steps

In Chapter 3, we have shown that the GR4 rainfall-runoff model may be run at multiple time steps simply by adaptation of some of its parameters, with good average accuracy over a large catchment set. However, this is only a necessary starting point, and some improvements at shorter time steps seem to be needed. In fact, some problems have been noticed, such as the degradation of performance affecting a significant number of catchments as the time step decreases. In particular, we have highlighted a worsening of the underestimation of flood volumes when using shorter modelling time steps. In order to investigate the possible causes of these undesirable problems, we proceeded to a model diagnosis by specific evaluations of the model input-state-output response at different time steps.

First, we show that the temporal variability of precipitation inputs has a significant impact on model performance over flood events, contrary to potential evapotranspiration. At first glance, this impact may seem just an expected general improvement. However, by analysing the model bias, a counter-intuitive degradation of the capacity of reproducing the water balance is still observed by using higher resolution precipitation patterns, instead of a constant-intensity pattern over larger durations. Since the precipitation volume is constant and only its temporal distribution changes, this result clearly indicates the presence of structural problems in the way the precipitation is processed in the model.

Then our model diagnosis is based on targeted evaluations of the consistency of the inner fluxes of the model, especially in terms of volumes of water losses and gains that contribute to reproduce the water balance. Our analyses show how neglecting what can be considered as the interception process in the GR4 model (or badly representing it) has a cascade of consequences at different time steps. The automated calibration of the model alleviates only partially the lack of interception losses by the compensation of other fluxes, which hides the problems especially on long-term evaluations. However, this cascade of consequences results in a worsening of flood simulations and in particular in an underestimation of the flood volumes. Also, the calibrated parameters are impacted by spurious time-step dependencies, because of the lack of interception losses. Following the results of the evaluations presented in this chapter, some possible ways of improvement of the GR4 model structure will be suggested, and they will be tested and analysed in Chapter 5.

4.1 General scheme of our model diagnostics framework

4.1.1 Model diagnosis scope

Our model diagnostics will deal with the performance and internal consistency of the model at daily and sub-daily time steps. We designed a set of tests and analyses with the aim of understanding the problems that affect the current model structure. This approach can be interpreted as a *model diagnosis*, following Yilmaz et al. (2008):

"*Model diagnosis is a process by which we make inferences about the possible causes of an observed undesirable symptom via targeted evaluations of the input-state-output response of the system model."*

The results of a model diagnosis should highlight "*inadequacies in model performance*" and point "*toward the specific aspects of the model structure and/or parametrization that are causing the problem(s)*" (Yilmaz et al., 2008). Our targeted analyses are summarized in Figure 4.1 and will provide the basis for the identification and improvement of a multi time-step model, ensuring the consistency in model performance and internal functioning at multiple time steps. The following targeted evaluations will be presented in the present chapter:

(i) the impact of the inputs temporal distribution on model performance (Section 4.2);

(ii) the consistency of model internal fluxes at different time steps (Section 4.3);

(iii) the consistency of model parameters and states at different time steps (Section 4.4);

First, for the model inputs, we analyse the impact of two opposite assumptions for the temporal distribution of precipitation and potential evapotranspiration: a distributed temporal pattern and a uniform pattern (see Section 4.2). Second, the analysis of the consistency of model internal fluxes simulated at different time steps is proved to be a key question of this model diagnosis (see Section 4.3). The model fluxes considered are interception loss, actual evapotranspiration, and groundwater exchanges. Although we do not have measures of these fluxes, we can expect or require the model to satisfy some physical principles, as the continuity equation in time of the volumes of cumulated fluxes. In the end of the Chapter, the analysis of model parameters and states consistency at different time steps is proposed to complete the diagnosis (see Section 4.4) and to better explain the impacts of time step on the model fluxes.

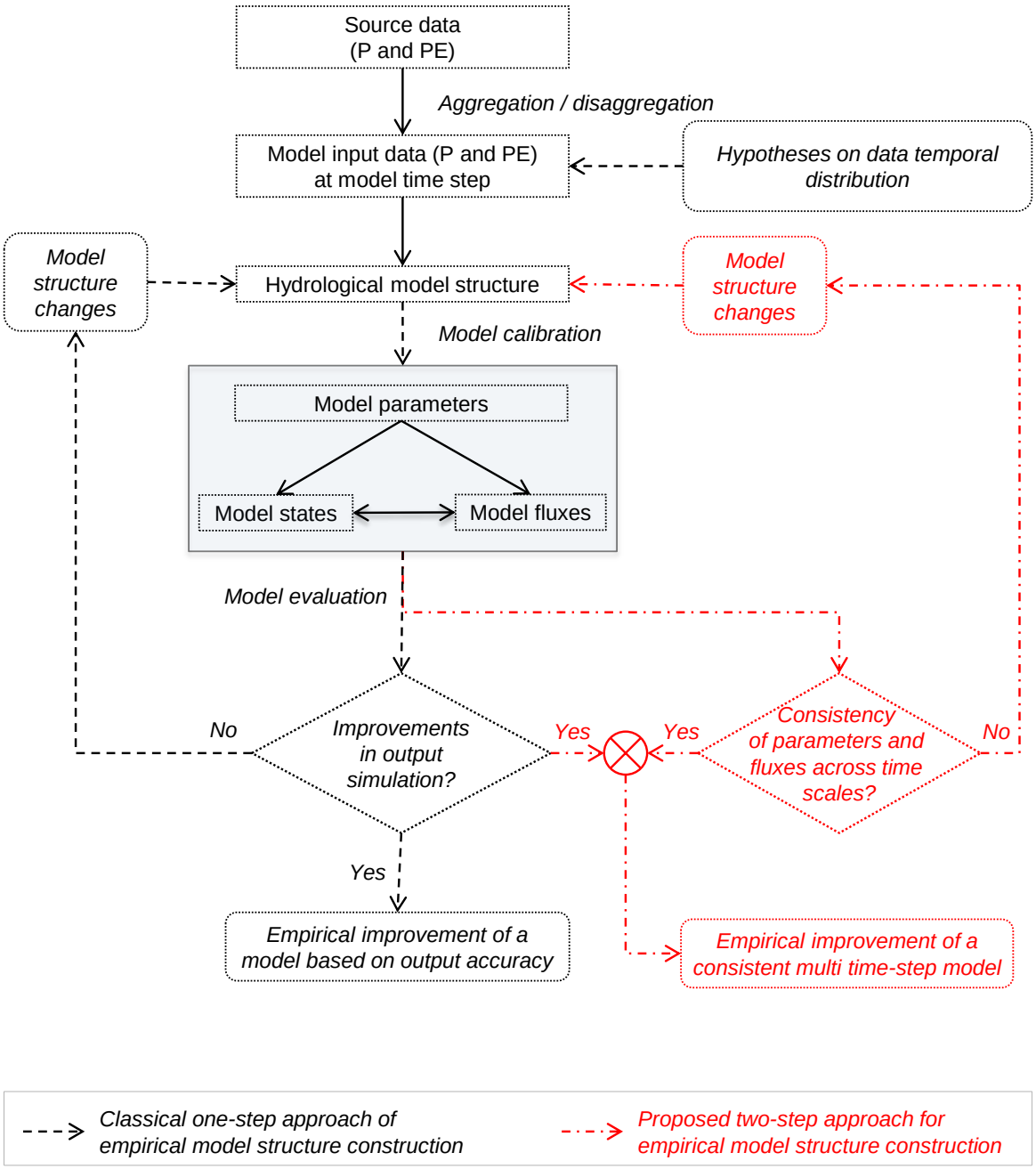


FIGURE 4.1 – Scheme of our process of model diagnosis and improvement at multiple time steps.

4.1.2 General remarks on the evaluation procedure

In all the tests presented in the following sections, the calibration-validation procedure is the same as the one presented in Chapter 3. We remind that the available 8-year period was split into two 4-year test sub-periods (2005–2009 and 2009–2013) and the split-sample test (Klemeš, 1986b) was applied. The model is calibrated on each sub-period, using the KGE criterion (Gupta et al., 2009) as objective function, calculated on the streamflow at all the time

steps of the calibration sub-periods. The evaluation was performed on both the whole validation sub-periods and on the set of selected flood events. The differences in performance between different modelling hypotheses were evaluated by means of the following criteria:

(i) the KGE criterion and its components (relative variability, ratio of means and correlation), evaluated both on the whole time series and on flood events only;

(ii) the criteria on the Flow Duration Curve (FDC), i.e. ratios of extreme quantiles of simulated and observed flows, and bias in the slope of the FDC.

A non-parametric statistical hypothesis test, i.e. the *Friedman test* (Friedman, 1937), was used to detect significant changes in mean ranks of performance across the alternative tested modelling options over the catchments set (at significance level 0.05). The post-hoc analysis of the Friedman test (Conover, 1999) is used to rank the different alternatives (by multiple pairwise comparisons) and the ranking is showed as a complement to box plots or to average values reported in tables. Since the Friedman test is based on ranking the variables (the criteria values in our case), it must be applied according to the boundaries and optimal values of the different criteria:

(i) for criteria with optimal value at the upper limit of their domain (e.g. KGE and r) the test is directly applied on the original score values;

(ii) for criteria with optimal values not equal to the domain upper boundary (e.g. relative variability, ratio of means, ratio of quantiles of the FDC and slope bias), the test is applied on the transformed values so as to make the optimal value infinity (e.g. taking the inverse of the absolute value of the criterion centered at zero at its original optimum).

Note also that when a difference in ranks between alternative options is judged “significant” by the statistical test, it should be further discussed according to the absolute difference in average values of criteria. In fact, the difference in ranks does not mean that the differences in the criteria distributions are significant from the point of view of a “meaningful” hydrological interpretation. To this end, an arbitrary level for the numeric differences between criteria should be considered. For example we consider negligible an absolute difference of average values of KGE (and its components) and FDC quantiles ratios smaller than 0.01.

4.2 Hypotheses on the model inputs temporal distribution

In this section, we present the results of some modelling tests using different hypotheses of the temporal distributions of model inputs, i.e. precipitation and evapotranspiration. Note that the volumes of the model inputs are consistent across the following tests, i.e. the daily precipitation and evapotranspiration do not change but their distribution over the day does.

4.2.1 Precipitation temporal distribution

Some tests were performed for assessing the impact of the precipitation constant-intensity hypothesis over different durations on model performance. This is interesting because the temporal variability of rainfall is more uncertain than the total amount, and the catchment low-pass filtering behaviour smooths out this variability (see the statistical analysis on our data sets in Chapter 2). In Chapter 3 we showed that the streamflow variability simulated by the GR4 model generally increases as the model time step decreases. This increase in simulated variability, together with a balance bias degradation, is responsible for the model performance degradation on flood events detected for a significant part of our catchment set (about 20%). This finding means that for a part of our catchment set, the GR4 model does not smooth out sufficiently the precipitation input variability. For this reason, we can expect that for some catchments, a precipitation constant-intensity pattern could even improve the model performance at one short time step, e.g. one hour, with respect to the use of the observed precipitation temporal pattern. We tested this hypothesis, by running the GR4 model at the hourly time step by the use of larger time steps (i.e. 3-, 6-, 12-h and 1-day) for the source data of precipitation and a hypothetic uniform distribution over this time span. The use of this constant-intensity pattern was compared with the use of the hourly precipitation observations by evaluating model performance over the whole validation period and on flood events only.

4.2.1.1 Evaluation over the whole validation period

The evaluation over the whole validation period indicates that the precipitation constant-intensity hypothesis has a very slight impact on the average model performance at the hourly time step with respect to the use of observed hourly precipitation (see Table 4.1).

The statistics in Table 4.1 summarize the distributions over the catchment set of the performance scores (KGE, its components, and the FDC-based criteria) calculated over the whole time series. In general, this set of criteria shows limited changes in the average model performance across the different tested options. However the Friedman test and its post-hoc analysis detect the following significant changes in the performance mean ranks of the aggregated KGE statistics (with a *p-value* of 10^{-15}):

(i) mean scores of the hourly model using source data at 6 h time step are higher than in other cases;

(ii) mean scores of the hourly model using 3 h and 12 h source data are higher than with hourly data;

(iii) mean scores with daily data are lower than all other time steps.

These differences in ranks are not impressive on the average values of KGE and not even on other quantiles (not shown) because the average differences for the different tests are in the order of 0.01 of KGE values, and so can be considered as negligible. For example, for more than two-thirds of the basins, the KGE values increase using source data at 6-h time step with respect to hourly data, but the mean absolute increase in KGE is only 0.007. Note also that

opposing trends are observed, for example for 40% of the catchments a slight unexpected improvement is observed passing from 1-h to 1-d constant-intensity duration with a mean increase in KGE of 0.02. Conversely, the opposite trend for the rest of the catchments (60%) is associated to a mean increase of 0.03. The most evident changes among the KGE components are found in the correlation which degrades relatively more passing from the sub-daily to daily constant-intensity duration.

Statistics of model performance criteria [-] evaluated over the 8-years validation period (at the hourly resolution)	**Precipitation constant-intensity duration [time step]**				
	1-h	**3-h**	**6-h**	**12-h**	**1-d**
Median Kling-Gupta Efficiency, *KGE* [-]	0.820	0.822	0.823	0.822	0.813
Friedman test on the KGE distributions: significant differences (for different letters) and ranking (alphabetical order)	c	b	a	b	d
Median relative variability, *a* [-]	0.989	0.988	0.988	0.989	0.988
Median ratio of means, *b* [-]	1.009	1.004	1.004	1.005	1.007
Median linear correlation, *r* [-]	0.897	0.899	0.900	0.895	0.884
Median 99th quantiles ratio, $Q99_{sim}/Q99_{obs}$ [-]	0.972	0.973	0.977	0.984	0.980
Percent bias in the slope of the mid-segment of the flow duration curve (0.2 – 0.7 flow exceedance probabilities) [%]	1.394	1.600	2.830	4.167	6.854

TABLE 4.1 – Summary of the distributions over the catchment set of the performance criteria evaluated over the whole validation period for the hourly model with the precipitation constant-intensity hypothesis over different durations (1, 3, 6, 12 h and 1 d). The ideal value is 1 for all the criteria except for the percent bias in the slope of the flow duration curve (FDC) that is 0%. The results of the Friedman test on the KGE scores are shown for a significance level of 0.05.

Among the indexes calculated on the FDC, the ratios of the 99th quantiles of simulated and observed flows show an interesting unexpected trend, being improved with larger source data resolutions (12 h and 1 d), on average. For example, by using hourly source data, the high flows with an exceedance probability of 1% are underestimated on average of about 3%, while the average underestimation is reduced to about 2% by using constant-intensity precipitation over larger durations (12 h and 1 d). This means that, counterintuitively, the high-flows volumes and peaks (linked to the largest precipitation events) seem to be better simulated by neglecting the actual temporal distribution of precipitation over the day. On the other hand, the percent bias in the slope of the mid-segment of the FDC ($SlopeBias_{FDC(0.2-0.7)}$) shows an expected trend with marked differences. On average, the bias $SlopeBias_{FDC(0.2-0.7)}$ is positive for all the alternative tests, indicating that the simulated response is more rapid than the observed catchment response. A constant improvement is detected by using finer temporal resolutions for precipitation data, with a reduction of the bias of more than 5% passing from 1 d to 1 h constant-intensity durations.

4.2.1.2 Evaluation over the flood events only

Figure 4.2 shows the comparison of the results of the hourly model simulations evaluated on flood events, by using the hourly observed precipitation and the uniform temporal patterns over different durations. The box plots in Figure 4.2(a) show that, on average, the model using the observed hourly precipitation markedly outperforms the model using a uniform precipitation pattern over time intervals larger than 3 h. In fact, the median KGE value decreases from 0.729 using hourly observations to 0.645 using daily uniform precipitation (i.e. a decrease of more than 11%). The Friedman test (with its post-hoc analysis) confirms that this trend is significant, with a p-value $< 10^{-44}$, and specifies also that all the options with sub-daily patterns (1, 3, 6, 12 h) outperform the one with daily uniform precipitation. Overall, this is explained by a coherent average ranking for the variability and correlation components Figure 4.2(b, d). Conversely, the water balance criterion (ratio of means) follows an opposite and counterintuitive trend being improved by the uniform precipitation pattern over larger durations, coherently with the analogous trend of the 99^{th} flow quantiles ratio observed in the previous section. The ratio of means (Figure 4.2(c)) shows a general underestimation of the flood volumes (of 10% on average) by the hourly model with hourly precipitation data. This underestimation is partially (but significantly) reduced by the constant-intensity precipitation pattern over the day.

Despite the general positive trend highlighted above for the aggregated KGE statistics, for a part of our catchment set (about 15%), the richer sub-daily temporal information in the rainfall signal steadily degrades the global performance of the GR4 model. For the 36 catchments for which model performance degrades passing from the use of daily uniform patterns to hourly observations, the mean degradation in KGE is 0.06 that is not negligible. These results confirm the findings presented in Chapter 3, specifying that while for a major part of the catchments, shorter time step information is effective for flood simulation, a degradation of the model performance may be found as the data time step decreases even when using the same model and evaluation time steps (e.g. 1 hour). This degradation is mainly due to:

(i) a *general decrease of the ratio of means* (i.e. water balance bias degradation) for the majority of the cases;

(ii) an *excessive increase of the relative variability* for a minor part of the catchment set.

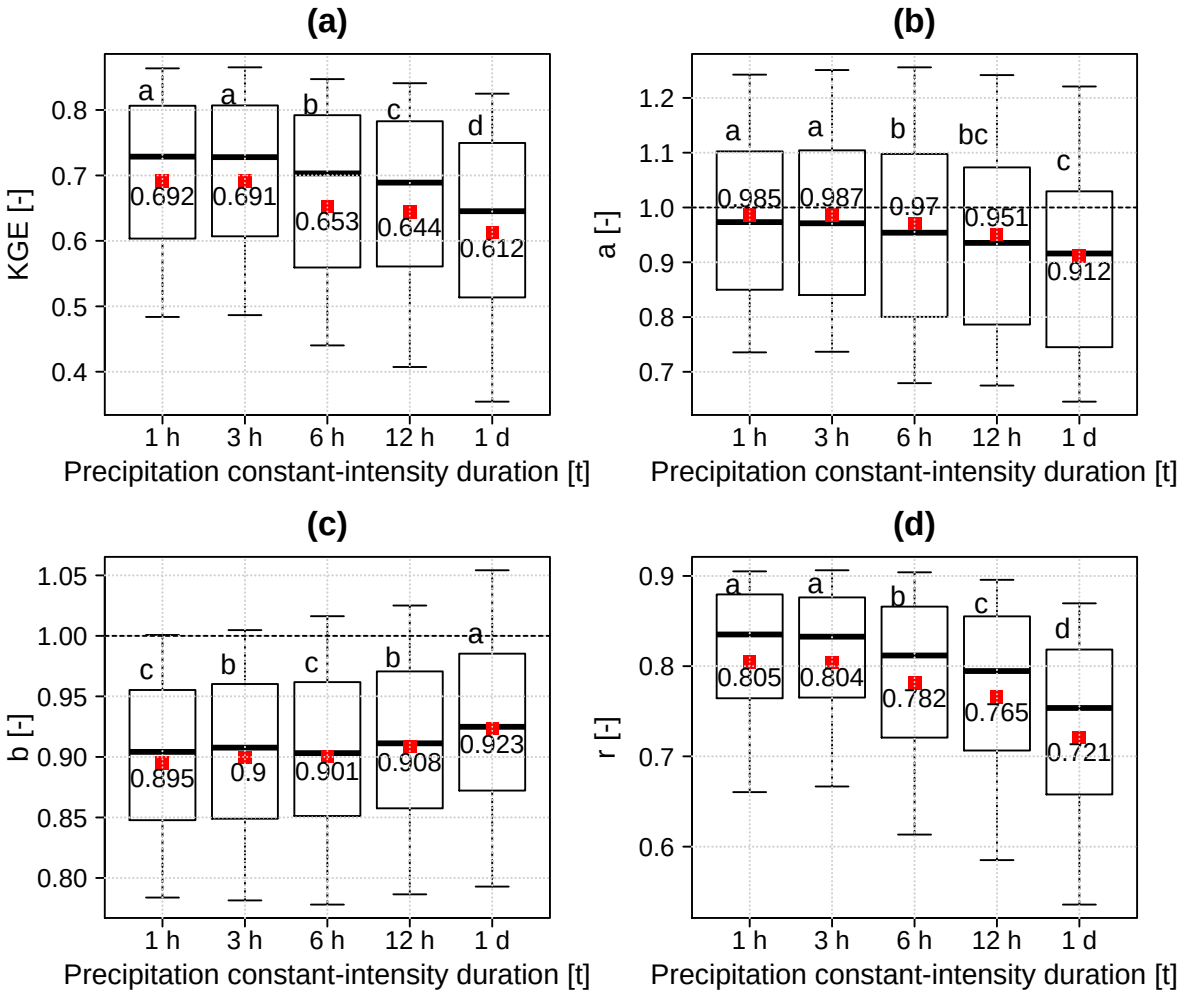


FIGURE 4.2 – Distribution over the catchment set of the performance criteria over the selected flood events for the hourly model with the precipitation constant-intensity hypothesis over different durations (1, 3, 6, 12 h and 1 d): (a) *KGE*, (b) relative variability, *a*; (c) ratio of means, *b*; and (d) correlation, *r*. The box plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red points refer to mean values. The letters above each box plot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letter are not significantly different).

The trend of the model balance bias is particularly interesting for the GR4 model diagnostics for two reasons:

(i) While the problem of the relative variability increase as the time step decreases could be due to errors in precipitation and streamflow data for some catchments (at least partially), the negative trend in model balance bias cannot be a problem of data errors**.** In fact we remind that the precipitation and streamflow volumes are constant across the different tests presented here, where only the temporal patterns change. This means that **the balance bias degradation is due to inadequacies in the rainfall-runoff model at sub-daily time steps or to more problems in the calibration phase at shorter time steps**.

(ii) The general decrease of the ratio of means (*b*) using finer time step information has a negative impact on model performance for most of the 240 basins. For example *b* degrades passing from daily to hourly constant-intensity patterns for about 65 % of the catchments. This happens also when the overall performance (i.e. KGE) improves (as for analogous tests in Chapter 3). On the other hand, the increase of relative variability is a counter-productive issue only for a minor part of the catchment set (for about 30 % of the basins passing from daily to hourly precipitation constant patterns).

4.2.2 Potential evapotranspiration temporal distribution

We remind that the potential evapotranspiration (PE) at the daily time step was calculated using the temperature-based formula proposed by Oudin et al. (2005) and daily temperature data from the SAFRAN reanalysis. Hourly temperature observations were not available, so the original source data time step for PE is daily. Thus, in order to disaggregate PE at the sub-daily time steps some assumptions have to be made. To evaluate the impact of the hypothesized temporal distribution of potential evapotranspiration over the day, we tested two different temporal patterns:

(i) ***"Uniform pattern"***: daily values of PE uniformly disaggregated over sub-daily time steps (see Figure 4.3(a)); this is the simplest hypothesis to disaggregate daily data at shorter time steps, in the condition of lack of data;

(ii) ***"(Quasi-)sinusoidal pattern"***: daily values of PE disaggregated over the day using the shape of the crest section of a sine-wave between 6:00 and 19:00 (with maximum from 12:00 noon to 13:00) and null values in the rest of the day/night (Figure 4.3(b)); this is a more physically-based hypothesis, that is typically used for describing diurnal variations of temperature, radiation and degree-day factors (e.g. Tobin et al., 2013).

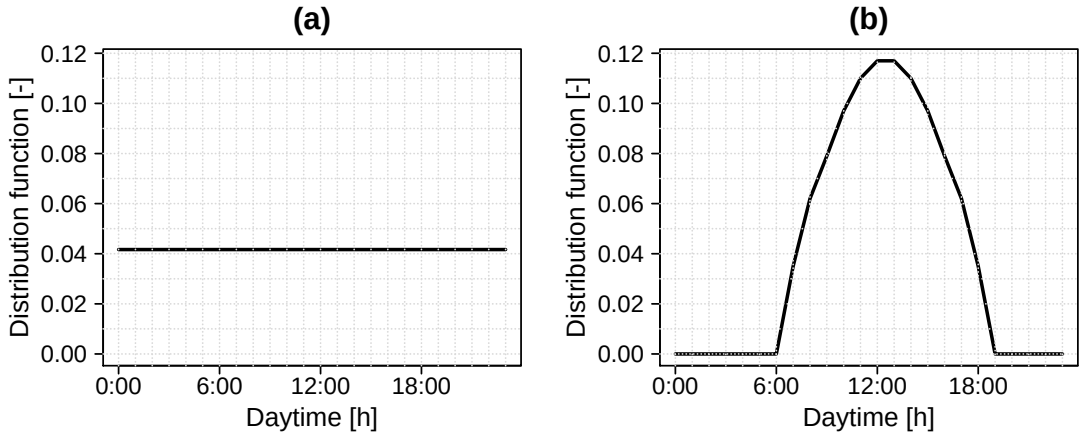


FIGURE 4.3 – Hypotheses of sub-daily temporal patterns for potential evapotranspiration (disaggregation of a daily unit value): (a) uniform pattern; (b) (quasi)-sinusoidal pattern.

The results of model performance with these two different hypotheses were analysed for different sub-daily time steps (from 6 min to 12 h) on the whole time period of model

validation and on flood events periods only. However, only the hourly model results are presented here, because these are representative of all the sub-daily time steps tests.

4.2.2.1 Evaluation over the whole validation period

If we evaluate the GR4 hourly model simulations by the KGE criterion and its components on the whole time series, we note that the two PE temporal patterns (uniform and sinusoidal) lead to very similar performance scores, as shown in Table 4.2. The median KGE on the catchment set is 0.82 for both the uniform and sinusoidal sub-daily distribution. The entire distributions of KGE and its components (*a*, *b* and *r*) remain stable too. The Friedman test detects a slightly significant difference between the ranks of KGE scores of the two testing hypotheses, with a p-value < 10^{-3} (just slightly lower than the significance level). However, the differences between the two modelling options in terms of KGE scores are negligible. For example for 60% of the catchments the uniform pattern slightly outperforms the sinusoidal, with a mean difference in KGE of only 0.003 points.

Also by evaluating the two tested PE temporal patterns by the indexes based on the Flow Duration Curve (FDC), the conclusions are similar. No significant difference is detected by the statistical test and the distributions of the scores are very similar for the two options. For example, Table 4.2 reports the median values of the ratio of the 99th quantiles of simulated and observed streamflow and of the bias in the slope of the FDC.

Statistics of model performance criteria [-] evaluated over the 8-years validation period (at the hourly resolution)	**PE temporal pattern**	
	Uniform	**Sinusoidal**
Median Kling-Gupta Efficiency, *KGE* [-]	0.822	0.820
Friedman test on the KGE distributions: significant differences (for different letters) and ranking (alphabetical order)	a	b
Median relative variability, *a* [-]	0.986	0.989
Median ratio of means, *b* [-]	1.008	1.009
Median linear correlation, *r* [-]	0.899	0.897
Median 99th quantiles ratio, $Q99_{sim}/Q99_{obs}$ [-]	0.972	0.972
Percent bias in the slope of the mid-segment of the flow duration curve (0.2 – 0.7 flow exceedance probabilities) [%]	2.286	1.394

TABLE 4.2 – Summary of the distributions over the catchment set of the performance criteria evaluated over the whole validation period for the hourly model with the uniform and sinusoidal temporal patterns for potential evapotranspiration. The ideal value is 1 for all the criteria except for the percent bias in the slope of the flow duration curve (FDC) that is 0%. The results of the Friedman test on the KGE scores are shown for a significance level of 0.05.

4.2.2.2 Evaluation over the flood events only

The two PE temporal distributions lead to very similar results also when model performance is evaluated on flood events only, as reported in Figure 4.4. The absolute difference in mean KGE and components values on the 240 catchments with the uniform and the sinusoidal

distributions are smaller than 0.003 points, so the statistical differences in mean ranks detected by the statistical test are considered negligible.

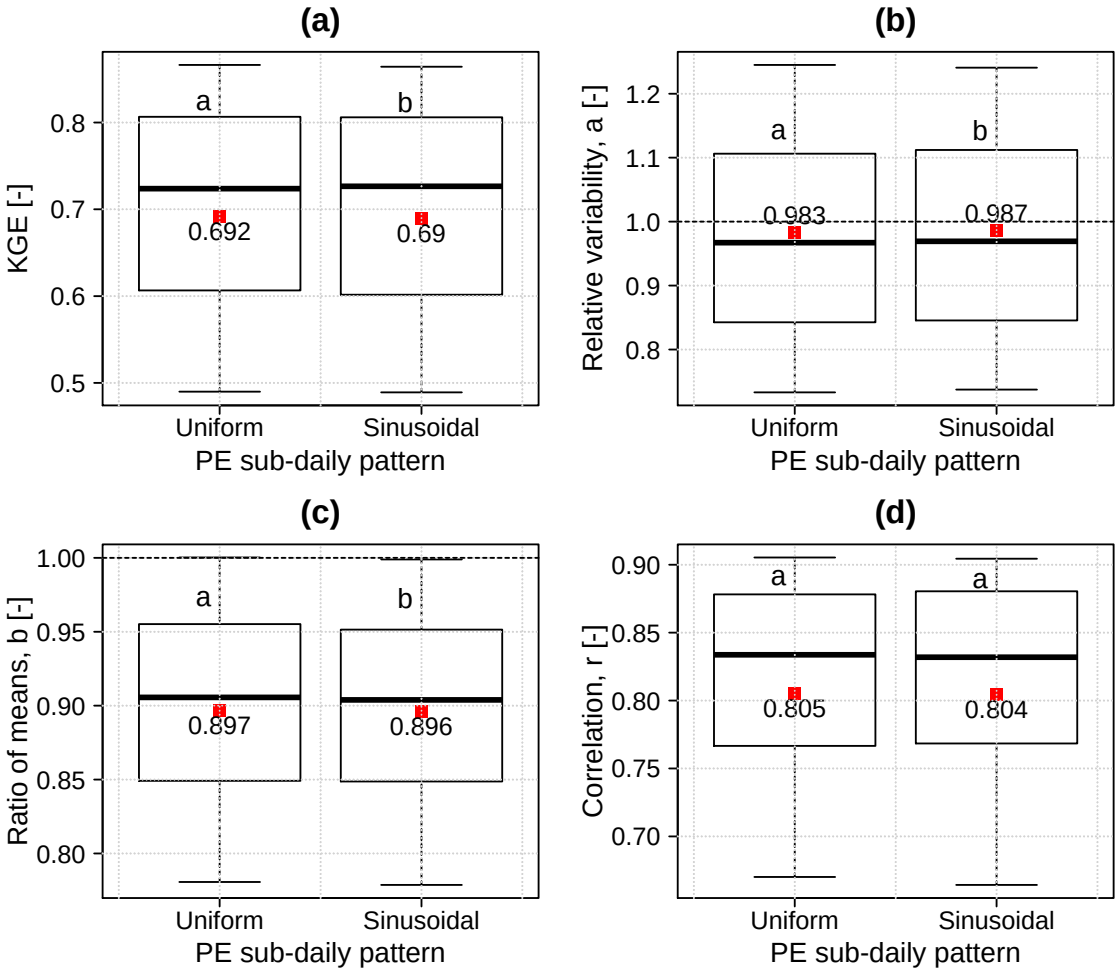


FIGURE 4.4 – Distribution over the catchment set of the performance criteria over the selected flood events for the hourly model with uniform and sinusoidal temporal patterns for potential evapotranspiration: (a) *KGE*, (b) relative variability, *a*; (c) ratio of means, *b*; and (d) correlation, *r*. The box plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red points refer to mean values. The letters above each box plot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letter are not significantly different).

4.2.2.3 Conclusions and our choice of a sub-daily pattern

The almost identical performance scores for the sub-daily simulations with the two PE temporal patterns represent a useful outcome for at least two reasons:

(i) it led us to validate the choice of a quasi-sinusoidal temporal pattern for PE because, for a same level of performance, we prefer to adopt a more physically-based assumption (e.g. see Fig. 2 in Tobin et al., 2013);

(ii) the non-influence of sub-daily patterns of PE confirms that the high frequencies of potential evaporation are smoothed out by the catchment low-pass filtering behaviour, contrary to precipitation (Oudin et al., 2006); so we expect that the lack of data for sub-daily estimates of PE does not affect our modelling results.

4.3 Consistency of the model internal fluxes at different time steps

4.3.1 Motivation

In general, the consistency of the model internal fluxes in time should be considered a prerequisite for the identification of a multi-time step model, as suggested in Section 4.1.1 (see Figure 4.1). A consistent rainfall-runoff model at different time steps should simulate the same cumulated internal fluxes, in addition to the same accurate outputs. One could perhaps argue that accurate simulations of the only output (streamflow at the outlet) at different time steps are also achievable by inconsistent internal model functioning at different time steps, thanks to compensations among the internal fluxes. However, we think that a model displaying temporal inconsistencies of fluxes would be sub-optimal for different reasons:

(i) Inconsistent model fluxes and compensations in the model are likely to be associated to inconsistent calibrated parameters at different time steps.

(ii) If accurate streamflow simulations are achieved by inconsistent model fluxes at different time steps, this would be evidence of large structural uncertainties.

(iii) Since it is preferable to get "*the right answer for the right reasons" (Kirchner, 2006)*, a lumped hydrological model at multiple time steps should respect the physical law of the mass conservation, in terms of volumes consistency. In other words, by decreasing the model time step we can expect that the dynamics of each flux are better described (thanks to a finer representation of inputs temporal variability) but the volumes cumulated in time should be coherent with the ones obtained by simulations at larger time steps.

Also, previous evaluations of the model indicated the need of an analysis of the consistency of model fluxes at different time steps. As we have shown in Chapter 3 and Section 4.2.1, a problem of balance bias degradation for the GR4 model simulations over flood events emerges as the data and modelling time step becomes shorter than daily. This problem is not linked to input data because of the construction of the tests (i.e. the consistency of cumulated inputs at multiple time steps, as already discussed), so it is due to the rainfall-runoff model (structure and/or parameters). In particular, the decrease in the ratio-of-means criterion over flood events, by using shorter time steps for input data and modelling (see Figure 4.2(c) and 3.8(c)), indicates that:

(i) On average, in high-flow conditions, the catchment water balance is currently better reproduced by the daily GR4 model rather than the sub-daily versions of the model.

(ii) Since for all time steps we used consistent inputs (same cumulated volumes of P and PE), we deduce that, over flood events, the simulated water losses are overestimated and/or the gains underestimated at sub-daily time steps with respect to the daily model.

For these reasons, it is essential to analyse the changes in the model internal fluxes across different time steps not only for building a consistent multi-time step model but also to understand the causes of the *observed undesirable balance bias degradation at shorter time steps over flood events*.

4.3.2 Analysis of the consistency of interception, evapotranspiration and groundwater exchanges

We focus here on the internal fluxes of the GR4 rainfall-runoff model corresponding to catchment water losses and gains, i.e. fluxes going outside of the basin before reaching the outlet or gains from the outside of the basin. We do not focus on water fluxes that are just transfers in time (e.g. the percolation from the production reservoir or the unit hydrographs outputs). We consider here the integrated fluxes over time (in mm over the integration time step), obtained by multiplying the model fluxes (mm/time step) by the time step. Then the fluxes of the GR4 model considered here are:

a) **Interception loss** (E_i or I [mm]), representing the evaporation from intercepted water (see Section 1.3.1, p. 35, for a broader definition of the process and its scale), that occurs mainly during a rainfall event (and shortly after). In the GR4 baseline model, the interception loss is modelled at the basin scale as the difference between input precipitation, P, and net rainfall, P_n (see Section 2.6.1, Eqs. (2.5-2.7)). This formulation allows evaporation of intercepted water only for the time steps with rainfall ($P>0$) and thus neglects the possible "*memory*" of the process. We remind here that the net rainfall P_n is calculated as the difference between precipitation P and potential evapotranspiration E (limited to positive values or zero), and then the interception loss can be formulated as:

$$E_i = \min(P,\ E) \qquad \text{[mm]} \qquad (4.1)$$

where: E_i is the interception loss "*computed as if there were an interception storage of zero capacity*" (Perrin et al., 2003) and zero resistance to potential evapotranspiration energy; P and E are the precipitation and potential evapotranspiration cumulated over the current time step Δt [mm]. Accordingly, the net rainfall and net evaporation capacity are respectively: $P_n = P - E_i$ and $E_n = E - E_i$.

b) **Actual evapotranspiration** (*AE or* E_s [mm]) from the production (soil moisture accounting) store, corresponding to the amount of water evaporated from the store

under the effect of the net evapotranspiration amount (see Section 2.6.1, Eq. (2.9)). We remind here below the discrete equation to calculate *AE*:

$$AE = \frac{S\left(2 - \frac{S}{x_1}\right)\tanh\left(\frac{E_n}{x_1}\right)}{1 + \left(1 - \frac{S}{x_1}\right)\tanh\left(\frac{E_n}{x_1}\right)} \quad \text{[mm]} \quad (4.2)$$

where: S [mm] is the production store water content (at the beginning of the time step); x_1 [mm] is the maximum capacity of the SMA store; E_n is the net evapotranspiration capacity (i.e. $E - E_i$) [mm].

c) **Exchanged fluxes** or **inter-catchment groundwater flows** (F_L and F_G [mm]) that can be losses (F_L) or gains (F_G), according to the sign of the calibrated x_2 parameter (and *F*). The half potential exchange, *F*, is calculated as in Eq. (2.9) (see Section 2.6.1) that is also reported below (see Eq. (4.3)). Actual gains are as twice as *F* (see Eq. (4.4)). Actual losses F_L are limited by the water available in the routing store and by the flows components coming from the unit hydrographs (see Eq. (4.5)):

$$F = x_2\left(\frac{R}{x_3}\right)^{3.5} \cdot \Delta t \quad \text{[mm]} \quad (4.3)$$

$$F_G = 2F \quad \text{[mm]} \quad (4.4)$$

$$F_L = -[\min(|F|,\ R + Q_9\Delta t) + \min(|F|,\ Q_1\Delta t)] \quad \text{[mm]} \quad (4.5)$$

where: R [mm] is the level in the routing store at the beginning of the time step; x_3 [mm] is its one-time-step-ahead capacity; x_2 [mm/time-step] is the groundwater exchange coefficient; Q_9 [mm/time-step] is the water input into the routing store, i.e. the output of the first unit hydrograph (UH1); and Q_1 [mm/time-step] is the direct flow component coming from the second unit hydrograph (UH2).

In the following we analyse the accumulation in time of these different fluxes obtained by simulations at different time steps, by cumulating each flux over the whole validation period of 8 years (including the two 4-year validation sub-periods) and over the selected flood events. These fluxes are not measurable at the catchment scale and are obtained by calibration on the measurable flux that is the outlet streamflow. However, their accumulation over long periods (larger than a day) should be consistent for multiple modelling time steps to comply with the simple mass balance equation.

To evaluate the consistency of the simulated fluxes at multiple sub-daily and daily time steps, we considered the daily model fluxes as the reference, in terms of volumes, since we showed that the latter (daily fluxes) allow obtaining a better water balance simulation.

For each tested model time step *x*, from 6 min to 1 day, we calculated the ratio of each cumulated flux simulated by the model at time step *x* normalized on the corresponding reference flux simulated at daily time step. We called these ratios the ***cumulative flux ratios***. They are non-dimensional indexes, evaluated for each catchment, by the following equations:

$$I_{ratio}[x|1\ d] = \frac{E_i[x]}{E_i[1\ d]} \quad [-] \quad (4.6)$$

$$AE_{ratio}[x|1\ d] = \frac{AE[x]}{AE[1\ d]} \quad [-] \quad (4.7)$$

$$\mathrm{FL}_{ratio}[x|1\ d] = \frac{F\ \ [x]}{F_L[1\ d]} \quad [-] \quad (4.8)$$

$$\mathrm{FG}_{ratio}[x|1\ d] = \frac{F\ \ [x]}{F_G[1\ d]} \quad [-] \quad (4.9)$$

where: $E_i[x]$, $AE[x]$, and $F\ \ [x]$ represent the cumulated interception loss, actual evaporation and exchanges fluxes (losses or gains) [mm] of the GR4 model at the sub-daily time step x with $x \in (6, 12, 30\ \mathrm{min}, 1, 3, 6, 12\ \mathrm{h})$; $E_i[1\ d]$, $AE[1\ d]$, $F_L[1\ d]$ and $F_G[1\ d]$ represent the cumulated interception loss, actual evaporation and exchanges fluxes (losses, F_L, and gains, F_G) [mm] of the GR4 model at the daily time step d. The "optimal" value of the cumulative flux ratios is 1, which corresponds to consistent modelled fluxes at different time steps.

Note that:

(i) The calculation of the cumulative flux ratios for the exchanged fluxes (*F*), i.e. $FL_{ratio}[x|1\ d]$ and $FG_{ratio}[x|1\ d]$, is admissible on our catchment set, because $F_L[1\ d]$ and $F_G[1\ d]$ are not null for all the catchments (the minimum absolute value of $F_L[1\ d]$ and $F_G[1\ d]$ is about 4 mm).

(ii) Since the exchanged fluxes can be either positive or negative we calculated the index $F_{ratio}[x|1\ d]$ by Eqs. (4.8) and (4.9) corresponding to the two cases:

(A) $F\ \ [1\ d]<0$, i.e. catchments presenting losses at daily reference (*197 catchments* of our sample).

(B) $F\ \ [1\ d]>0$, i.e. catchments presenting gains at daily reference (*43 catchments*).

This allows to univocally interpret the values of the ratios, because of the possible change in sign in the numerator ($F\ \ [x]$) with respect to the denominator. This means that for $\mathrm{FG}_{ratio}[x|1\ d]$ we can have the following cases:

- (B1) $\mathrm{FG}_{ratio}[x|1\ d] > 1$, i.e. gains are larger at sub-daily time step *x*;
- (B2) $\mathrm{FG}_{ratio}[x|1\ d] < 1$, i.e. gains are larger at daily time step;
- (B3) $\mathrm{FG}_{ratio}[x|1\ d] < 0$, i.e. the exchanges are gains at daily time step but become losses at sub-daily time step *x*.

Three similar cases could potentially be found for the case (A).

4.3.2.1 Evaluation over the whole validation period

Figure 4.5 (a-d) shows a statistical summary of the evolution of the cumulative flux ratios calculated over the eight-year validation period as the time step changes (for the catchment set). The evolution of the flux ratios is presented by some quantiles over the catchment

sample as the time step decreases from 1 day to 6 min (from right to left on the x-axis). The following clear trends in the fluxes evolution are detected:

a) The *interception loss* steadily decreases as the time step decreases reaching a median $I_{ratio}[6\,min|1\,d]$ of about 19 %. In other words, on average, the cumulated interception loss simulated by the 6-min model is only less than one fifth of the same flux simulated by the daily model. Note that the dispersion around the average cumulative ratios is relatively small (the 5th and 95th quantiles are close to less than 10% around the median ratio). It means that the interceptive losses significantly and constantly decrease at shorter time steps for all the 240 catchments. The decrease rate is the highest (-31%) when passing from 1 d to 12 h.

b) The *actual evaporation from the production reservoir* monotonically increases as the time step becomes finer, reaching a median $AE_{ratio}[6\,min|1\,d]$ of about 145 %. The increase rate is the highest (+18%) when passing from 1 d to 12 h. The trend in actual evaporation cumulative ratios is opposite to the one for interception. This may be explained by the direct compensation of a reduced interception loss (see also discussion below on absolute differences in fluxes and Table 4.3).

c) The *negative groundwater exchanged fluxes at daily reference (losses)* get stronger as the time step becomes shorter, reaching a median increase, $\mathrm{FL}_{ratio}[6\,min|1\,d]$, of about 165 %. As it was noted for actual evaporation, the trend in negative exchanged fluxes is opposite to the one for interception. So, it seems that the model compensates the reduced interception loss at sub-daily time steps also by increasing exchange losses. In opposition to this general trend, the losses are reduced at some sub-daily time steps (with respect to daily) only for 6 out of the 197 catchments with groundwater losses at daily time step.

d) The *positive groundwater exchanged fluxes at daily reference (gains)* decrease with time step, on average, reaching a median decrease at 6 min of about 50 % ($\mathrm{FG}_{ratio}[6\,min|1\,d] = 0.5$). Moreover, for 15 out of the 43 catchments with groundwater gains at daily time step (35% of the set), the gains become zero and even change sign as the time step *x* decreases (F[*x*]<0). This means that these catchments gain water at the daily time step, but as the time step decreases their water gains become losses that get larger as the time step becomes finer.

These large changes in cumulated fluxes, over the whole series, must be further specified in absolute terms (i.e. in mm) for each time step, without the normalization by the daily reference, in order to allow a comparison of the magnitude of changes for the different fluxes. Table 4.3 reports the median values of annual averages of cumulated fluxes for all the tested time steps and the median value of the net cumulated losses (the sum of interception loss, actual evaporation and exchanges for each catchment). It shows that the changes in the different fluxes over the whole series counterbalance each other perfectly, by providing a constant average net (total) loss across the different time steps.

The absolute changes of cumulated fluxes shown in Table 4.3 are very large. Unfortunately, we do not have any measurements for these fluxes to provide a physical validation of the internal behaviour of the model. In future works, it would be interesting to compare the internal fluxes modelled at different time steps to observations in a small experimental catchment (with measurements of the fluxes of interest), to have a reference for the assessment of the fluxes consistency (even considering the potential problems to overcome, e.g. correspondence between modelled and observed variables, representativeness of the measurements at the basin scale, etc.). However, here it is interesting to discuss at least the order of magnitude of the annual interception loss simulated at different time steps, compared to the mean annual precipitation and to the total evaporation. In fact, from the literature, it is expected that, at least in temperate climates (as in France), the annual interception loss may represent a significant portion of the total evaporation, and amount up to 15-50% of precipitation (Calder, 1990; Savenije, 2004; Gerrits, 2010). Thus, one may note that the simulated annual interception loss at the daily time step is more consistent with this expectation (see Table 4.3). In fact, at the daily time step the interception loss represents an important part of total evaporation (about 40%) but significantly decreases at shorter time steps. Moreover, while the simulated interception flux at the daily time step represents on average the 25% of precipitation, it drops down to less than 5% at the shortest 6-min time step (we remind that the median value of mean inter-annual precipitation over the catchment set is 940 mm/y, see Chapter 2). This rough comparison provides a first validation of our choice of using the daily modelled fluxes as a reference.

Annual cumulated flux statistics [mm/y] over the 8-years validation period	**Model time step**							
	6-m	**12-m**	**30-m**	**1-h**	**3-h**	**6-h**	**12-h**	**1-d**
Median annual interception loss, I [mm/y]	45	55	69	81	108	135	166	241
Median annual actual evaporation from production reservoir, AE [mm/y]	455	449	437	430	409	391	367	311
Median annual groundwater losses, from basins losing water at daily time step, F_L [mm/y]	122	122	117	113	108	101	92	72
Median annual groundwater gains, from basins gaining water at daily time step, F_G [mm/y]	29	30	32	53	57	53	60	78
Median annual net losses (I+AE+F_L-F_G) [mm/y]	601	601	600	603	603	603	602	604

TABLE 4.3 – Summary of the annual averages of the cumulated internal fluxes modelled by the GR4 model at different time steps.

It is also interesting to analyse the dispersion over the catchment set of the absolute changes of cumulated fluxes around the averages provided in Table 4.3. To provide an example, Figure 4.6 shows the comparison between the annual averages of cumulated fluxes simulated at the daily time step and the same fluxes obtained by the hourly simulation. For a consistent

multi-time step model the points should be aligned along the 1:1 line. Thus, it is evident that the current structure of the hourly GR4 model is not consistent with its daily original version for the whole catchment set. Still the model manages to get a coherent water balance at the outlet (see the equality of total ‘net losses’ in Figure 4.6(d)) thanks to the compensations between interception loss, actual evaporation and groundwater exchanges.

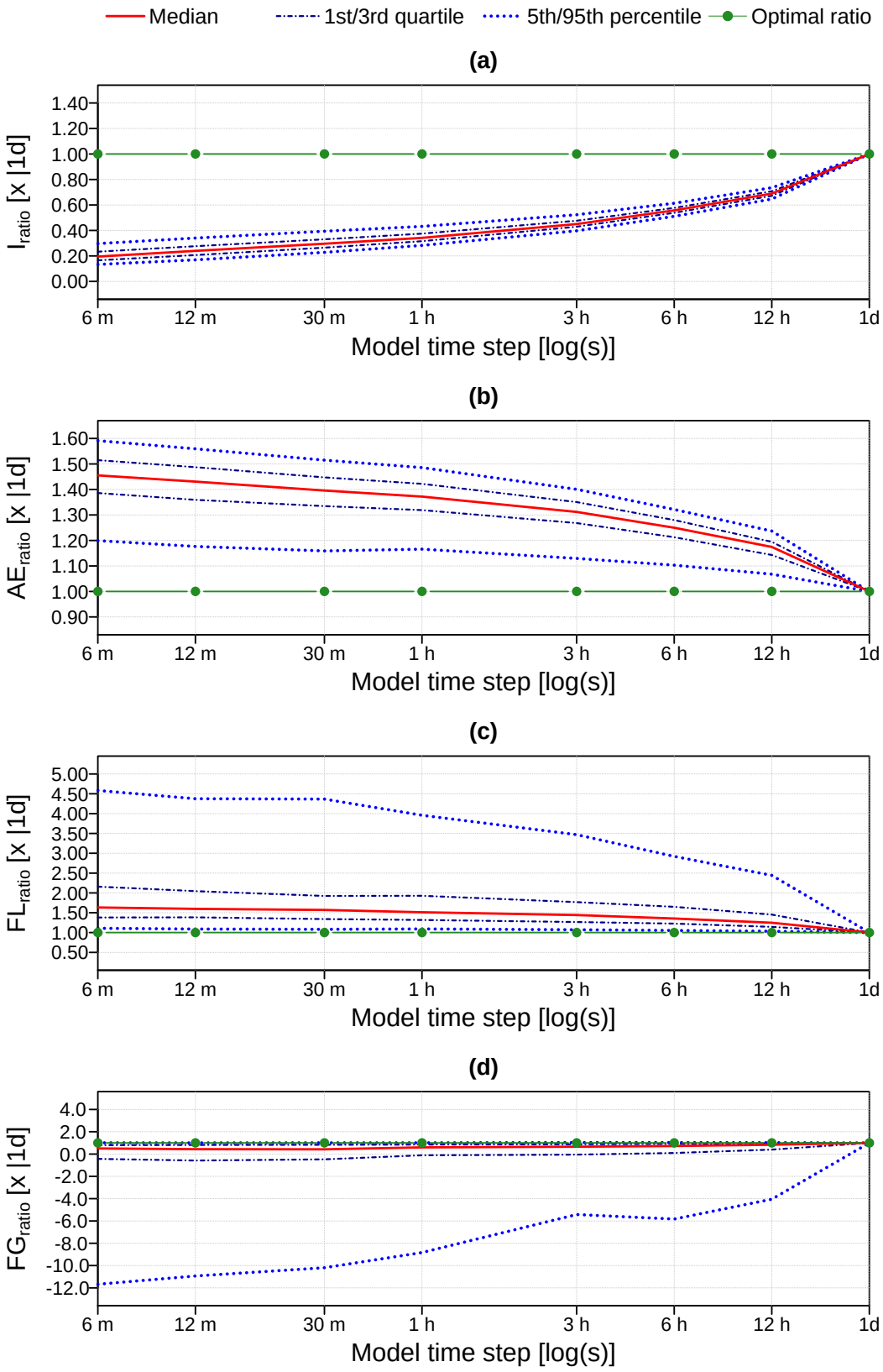


FIGURE 4.5 – Summary of the cumulative flux ratios at different time steps (with daily reference) over the whole validation period and the 240-catchment set: (a) interception loss, *I*; (b) actual evaporation from production reservoir, *AE*; (c) groundwater losses, *FL*; (d) groundwater gains, *FG*.

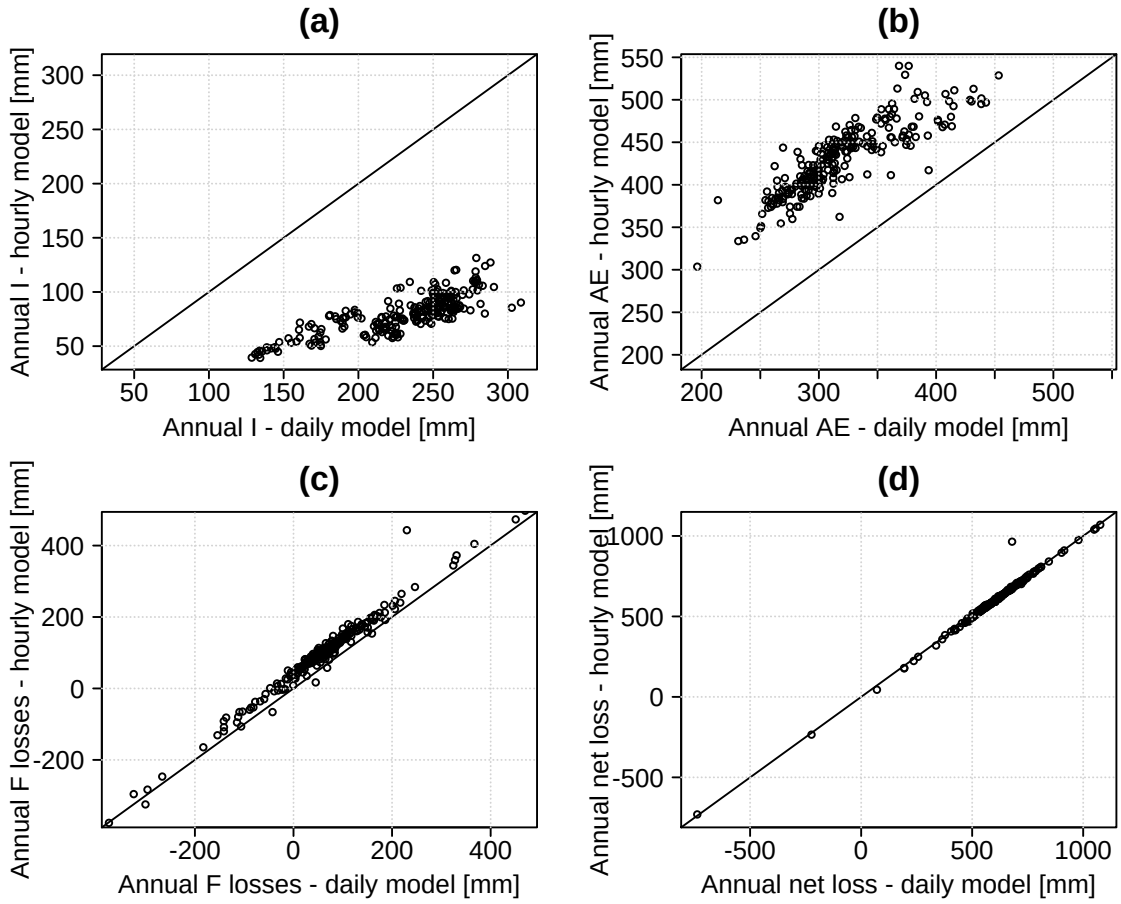


FIGURE 4.6 – Annual average cumulated fluxes from daily and hourly GR4 model simulations over the 240-catchment set: (a) interception loss, I; (b) actual evaporation from the production reservoir, AE; (c) actual groundwater losses, (negative values represent gains), F; (d) net losses (=I+AE+F).

4.3.2.2 Evaluation over the flood events only

The same analysis of the internal fluxes volumetric consistency at different time steps was performed also by evaluating on the selected flood events only. Figure 4.7(a-d) shows a statistical summary of the evolution of the cumulative flux ratios as the time step changes, by evaluating over the set of 2400 selected flood events.

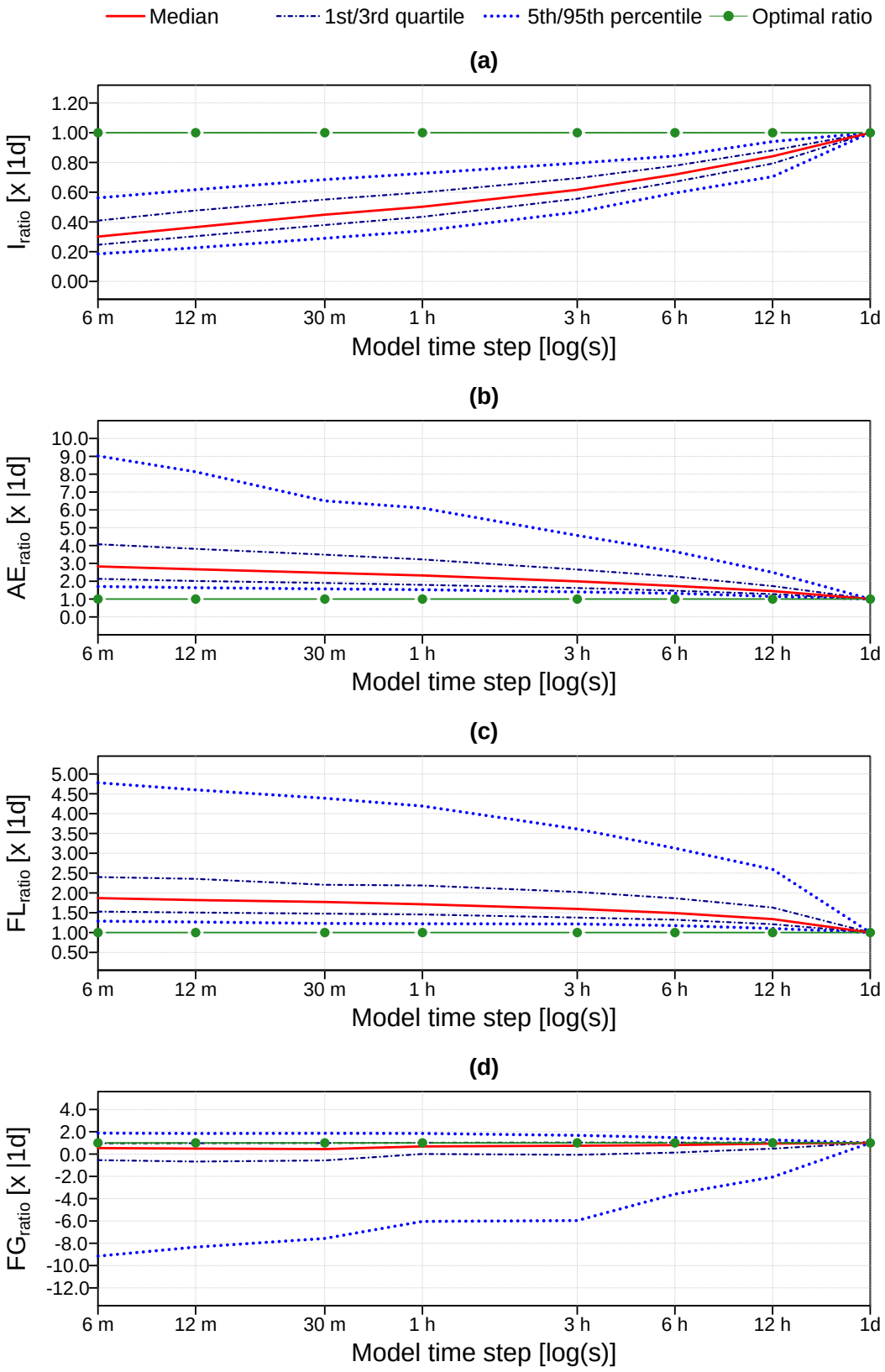


FIGURE 4.7 – Summary of the cumulative flux ratios at different time steps (with daily reference) over the 2400 flood events for the 240-catchment set: (a) interception loss, *I*; (b) actual evaporation from the production reservoir, *AE*; (c) groundwater losses, *FL*; (d) groundwater gains, *FG*.

By comparing the statistics of flux ratios in Figures 4.7 and 4.5, it can be noted that the same trends of changes in fluxes simulated at different time steps are found when evaluating either on flood events only or over the whole 8-years validation period. In more detailed and relative terms, it seems that:

a) The *interception loss* decreases with time step also on flood events, but at a slightly lower rate than on the whole validation period, reaching a median $I_{ratio}[6\ min|1\ d]$ of about 30 %.

b) The *actual evaporation from the production reservoir* monotonically increases at shorter time steps also on flood events, and at a higher rate than on the whole series, reaching a median $AE_{ratio}[6\ min|1\ d]$ of about 285 %.

c) The *negative groundwater exchanged fluxes at daily reference (losses)* get stronger as the time step becomes shorter at around the same rate than on the whole series, reaching a median increase, $FL_{ratio}[6\ min|1\ d]$, of about 190 %.

d) The *positive groundwater exchanged fluxes at daily reference (gains)* decrease, on average at the same rate than on the whole series, reaching a median decrease at 6 min of about 50 % ($FG_{ratio}[6\ min|1\ d] = 0.54$).

Also for flood events Table 4.4 reports a summary of the changes in model fluxes at different time steps in absolute terms (mm), averaged on a daily basis that is the most appropriate scale to evaluate the average fluxes over the selected floods (since the average flood-duration is of a few days). It shows the median values of the daily average cumulated fluxes and the net cumulated losses (the sum of interception loss, actual evaporation and exchanges over floods for each catchment). Table 4.4 shows that, **in contrast to the case of the whole series validation, the changes in the different fluxes over flood events do not counterbalance each other, but the total net losses increase as the time step decreases**. This increase in total net losses is mainly due to the increasing losses from the underground exchanges while interception loss and actual evapotranspiration counterbalance almost perfectly. For example, by passing from the daily simulation to the 6-min one, the average daily total net losses increase of about 30% (while the losses from the surface compartments, interception and evaporation, change of only 4%).

Daily cumulated flux statistics [mm/d] over the 2400 selected flood events	**Model time step**							
	6-m	**12-m**	**30-m**	**1-h**	**3-h**	**6-h**	**12-h**	**1-d**
Median daily interception loss, I [mm/d]	0.23	0.27	0.32	0.36	0.45	0.52	0.61	0.72
Median daily actual evaporation from production reservoir, AE [mm/d]	0.71	0.68	0.63	0.58	0.51	0.44	0.37	0.25
Median daily groundwater losses, from basins losing water at daily time step, F_L [mm/d]	1.30	1.27	1.25	1.20	1.12	1.06	0.98	0.76
Median daily groundwater gains, from basins gaining water at daily time step, F_G [mm/d]	0.34	0.25	0.21	0.50	0.54	0.59	0.65	0.67
Median daily net losses ($I+AE+F_L-F_G$) [mm/d]	2.13	2.15	2.09	2.07	2.04	1.99	1.88	1.67
Median daily losses from interception + evaporation ($I+AE$) [mm/d]	0.98	0.99	0.99	1.00	1.00	1.01	1.01	1.02

TABLE 4.4 – Summary of the daily average of the cumulated internal fluxes modelled by the GR4 model at different time steps over the 2400 selected flood events.

The scatter-plots in Figure 4.8 show the comparison of the daily average fluxes calculated from hourly and daily simulations and confirm the trends discussed above for the whole catchment set. It shows that the net losses are generally inconsistent and greater at hourly time step than at daily (points lying above the identity line for positive losses). This result is analogous at all time steps: the net losses are greater at all sub-daily time steps with respect to the daily model for at least 93% of the 240 catchments (for 222 catchments at 12-h and for 226 at 6-min time step).

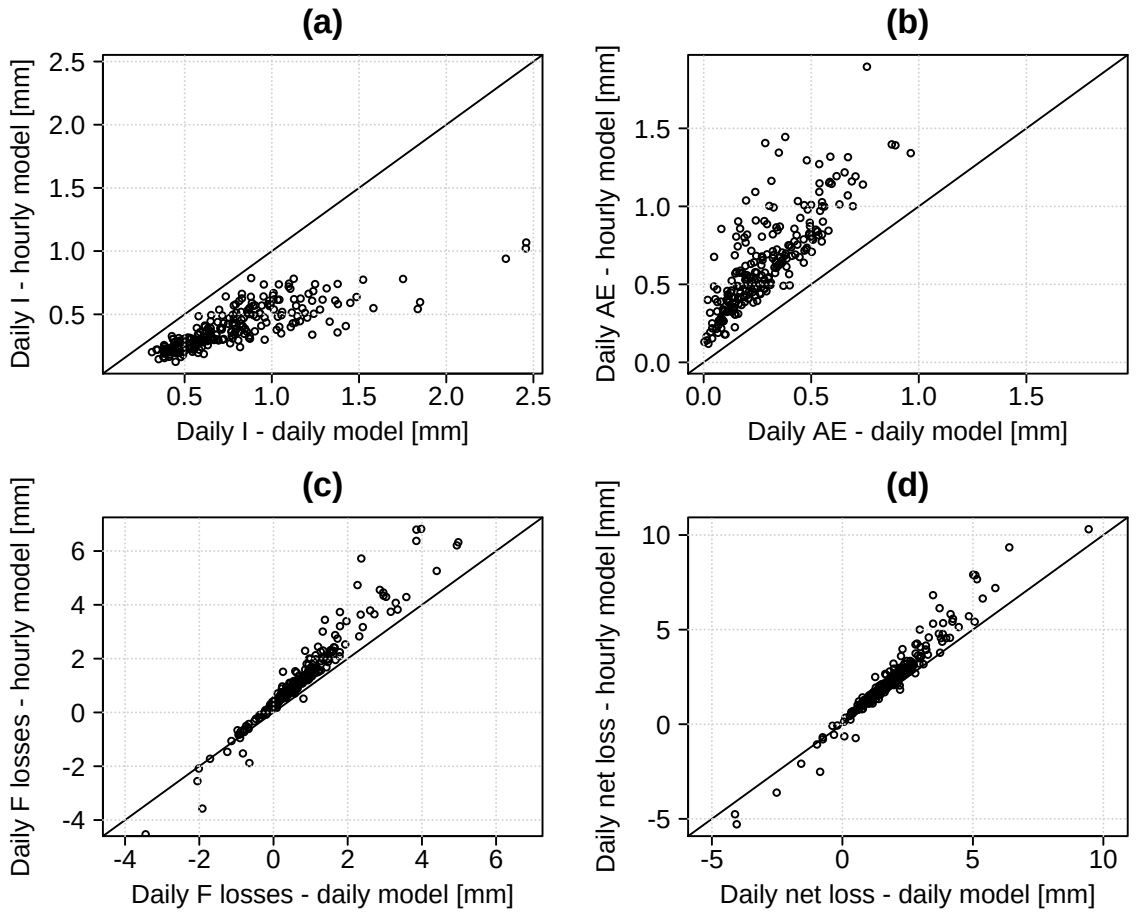


FIGURE 4.8 – Daily average cumulated fluxes from daily and hourly GR4 model simulations over the 2400 selected flood events: (a) interception loss, *I*; (b) actual evaporation from the production reservoir, *AE*; (c) actual groundwater losses, (negative values represent gains), *F*; (d) net losses (=*I*+*AE*+*F*).

4.3.3 Implications and formal proof of the temporal inconsistency of the neutralisation function for interception

The problem of fluxes inconsistency reported in the previous sections suggests that some of the GR4 model components should be changed as the time step decreases in order to ensure a consistent water balance. In particular, *the current simple representation of interception as a P-E neutralisation function* (see Eq. (4.1)), *corresponding to the assumption of zero-capacity interception storage*, *is not consistent at sub-daily time steps*, since it leads to increasingly reduced interception losses at shorter time steps. In order to ensure coherent fluxes, it seems necessary to change the simple neutralisation function governing the interception flux to intercept more water at shorter time steps. Our analyses indicate that, at sub-daily time steps, the process of interception would be represented in a more consistent way if its 'memory' would be extended up to a daily time interval. This could be done by adding an interception reservoir in the model at sub-daily time steps. These results are in line with previous findings in the literature, as analogous empirical analysis of the impact of time step on the modelled

interception loss by Haddeland et al. (2006), and also with the broad definition of the interception as a process (e.g. Gerrits, 2010).

Haddeland et al. (2006) analysed the sensitivity of the simulated internal fluxes of the Variable Infiltration Capacity (VIC) model to model time step and data temporal resolution (decreasing from daily to hourly). Similarly to our results, Haddeland et al. (2006) observed that, also for the VIC model, the changes in the cumulated model fluxes and outputs at different time steps are explained by the scheme used to parametrize canopy evaporation. The possibility of intercepting the whole current precipitation at the daily time step leads to increased intercepted volumes at daily time steps rather than at sub-daily time steps.

Gerrits (2010), reviewing the definitions of interception in the literature, argued that interception should be defined by considering it as a process, governed by the sum of the change of an interception storage and the evaporation from this stock. She reported that the time scale of the interception process is in the order of one day. After a daily time span, *for most climates*, the change of interception storage approaches zero and the interception loss becomes upper-limited by the potential evaporation, i.e. our assumption of zero-capacity interception store.

In the following, we propose a proof to add a formal argument for the inadequacy of the neutralisation function at sub-daily time steps. In a retrospective way, this proof lays the foundation of our empirical findings on the temporal inconsistency of the assumption of zero-capacity interception.

Proof of the temporal inconsistency of zero-capacity interception

A simple formal proof of the general inconsistency of the neutralisation function (Eq. (4.1)), corresponding to zero-capacity interception) at multiple time steps is given below based on classical deductive logic. It is proved by contradiction that the neutralisation function may provide consistent results when applied on data at two different time steps ($\Delta t_{fine} < \Delta t_{large}$) if and only if (*iif*) a particular condition is verified. The condition is that the precipitation and potential evapotranspiration values at the shorter time step (Δt_{fine}) are ranked in the same way (i.e. either $P \leq E$ or $P \geq E$) for all the duration of the larger time step (Δt_{large}). If this condition is not verified the neutralisation calculated on data at the fine resolution and cumulated over the larger time step Δt_{large} is smaller than the neutralisation calculated directly on data at the Δt_{large} resolution.

For facilitating the understanding of the notation, the case of hourly (*h*) and daily (*d*) time steps is assumed. However, the same proof is valid for any other couple of time steps (with notation *h* for the shorter time step and *d* for the larger one).

Proof of the temporal inconsistency of the neutralisation function for interception

Hypotheses: Let $H = \{h_1, h_2, \dots, h_{24}\}$ be the set of the 24 hours of a day.
The hourly precipitation (P_h) and potential evapotranspiration (E_h) are consistent with the daily values (P_d and E_d) in terms of daily accumulations:

$$\begin{cases} \sum_{i=1}^{24} P_{h_i} = P_d \\ \sum_{i=1}^{24} E_{h_i} = E_d \end{cases}$$

Claim: $\left[\min(P_d, E_d) = \sum_{i=1}^{24} \min\left(P_{h_i}, E_{h_i}\right)\right] \underset{iif}{\Leftrightarrow} (P_h \geq E_h \ \forall h \in H) \mid (P_h \leq E_h \ \forall h \in H)$
Otherwise, if the condition on the right-side is not true, $\min(P_d, E_d) > \sum_{i=1}^{24} \min\left(P_{h_i}, E_{h_i}\right)$.

Proof: First, the *simple conditional statement* (*if condition*) can be easily proved directly from the hypotheses. Let the estimations of interception loss at daily and hourly time step follow the notation I_d and I_{h,cum_d} respectively. From the neutralisation function (Eq. (4.1)), we have: $I_d = \min(P_d, E_d)$ and $I_{h,cum_d} = \sum_{i=1}^{24} \min\left(P_{h_i}, E_{h_i}\right)$.
I_d and I_{h,cum_d} will be equal in the two following alternative cases, as stated by the claim:
1. If $P_h \geq E_h \ \forall h \ \rightarrow I_{h,cum_d} = \sum_{h=1}^{24} \min(P_h, E_h) = \sum_{h=1}^{24} E_h \ = E_d = \min(P_d, E_d) = I_d$
2. If $P_h \leq E_h \ \forall h \ \rightarrow I_{h,cum_d} = \sum_{h=1}^{24} P_h = P_d = \min(P_d, E_d) \ = I_d$

Second, the *biconditional* statement (*iif*) in the claim can be proved by contradiction.
The claim is negated assuming the following assertion (i.e. its negation, noted as *p*):

$$p \begin{cases} \left[\min(P_d, E_d) = \sum_{i \in H} \min\left(P_{h_i}, E_{h_i}\right)\right], \ and \\ H \text{ can be partitioned into two subsets } H_1 \text{ and } H_2 \text{ with } \left(P_{h_j} > E_{h_j} \ \forall h_j \in H_1\right) \text{ and } \left(P_{h_k} < E_{h_k} \ \forall h_k \in H_2\right) \end{cases}$$

The possible existence of a third subset H_3 with $\left(P_{h_l} = E_{h_l} \ \forall h_l \in H_3\right)$ is not considered since it would not change the proof.
By considering this partition of the H set, it is true that:
$\sum_{i \in H} \min\left(P_{h_i}, E_{h_i}\right) = \sum_{j \in H_1} \min\left(P_{h_j}, E_{h_j}\right) + \sum_{k \in H_2} \min\left(P_{h_k}, E_{h_k}\right) = \sum_{j \in H_1} E_{h_j} + \sum_{k \in H_2} P_{h_k}$.
Then, following the negation of the claim to be proved (*p*) we would have:
$\min(P_d, E_d) = \sum_{i=1}^{24} \min\left(P_{h_i}, E_{h_i}\right) \rightarrow \min(P_d, E_d) = \sum_{j \in H_1} E_{h_j} + \sum_{k \in H_2} P_{h_k}$ (Assertion *q*).
However, the sum on the right-side of the last assertion cannot be equal to neither P_d neither E_d, but must be smaller than $\min(P_d, E_d)$. This comes from the definitions of the sets H_1 and H_2, which lead to:
$\sum_{k \in H_2} P_{h_k} < (E_d - \sum_{j \in H_1} E_{h_j}) \rightarrow \sum_{j \in H_1} E_{h_j} + \sum_{k \in H_2} P_{h_k} < E_d$, and
$\sum_{j \in H_1} E_{h_j} < (P_d - \sum_{k \in H_2} P_{h_k}) \rightarrow \sum_{j \in H_1} E_{h_j} + \sum_{k \in H_2} P_{h_k} < P_d$.
So, when the condition $(P_h \geq E_h \ \forall h \in H) \mid (P_h \leq E_h \ \forall h \in H)$ is negated we have:
$\min(P_d, E_d) > \sum_{i=1}^{24} \min\left(P_{h_i}, E_{h_i}\right)$, as stated in the end of the claim.
The negation of the claim implies two contradictory assertions (*p* and *q*). Since these cannot both be true, the assumption that the claim is false must be wrong. Then the claim is true.

Q.E.D. (end of proof)

Note that on a normal rainy day (when interception is an active process) the precipitation and potential evapotranspiration rarely respect the condition for the temporal consistency of neutralisation proved above. This condition should be ensured in two particular types of climate: (i) hot temperatures and very low-intensity precipitation (P[mm]<E[mm]), or (ii) cold temperatures and high-intensity precipitation (P[mm]>E[mm]). Conversely, in a temperate

climate as in France, it is more likely that precipitation and potential evapotranspiration at sub-daily time steps do not rank the same all along the day. It is sufficient to see some statistics of average hourly values of precipitation and evapotranspiration in a temperate region (as in France) and think at the intermittency of storm events to understand it. Moreover, empirical proof of this was provided by our analyses of the inconsistency of the interception loss in the GR4 model simulations at different time steps on our catchment set.

4.4 Consistency of model parameters and states at different time steps

4.4.1 Model parameters

Three of the four free parameters of the GR4 model affect the formulation of two internal fluxes considered in the analysis of the previous section, i.e. actual evaporation and groundwater exchanges. Among the three internal fluxes considered, only the interception loss is independent of all parameters. The actual evaporation from the production reservoir depends on the capacity of this reservoir (x_1), as stated in Equation (4) in Perrin et al. (2003). The groundwater exchanges are directly related to the groundwater exchanges coefficient (x_2) and the capacity of the routing reservoir (x_3), as in Equation (18) in Perrin et al. (2003). So we can logically expect that the inconsistencies in model internal fluxes at different time steps are likely to imply inconsistent parameter values (of x_1, x_2 and x_3) across time steps.

As we have highlighted in Chapter 2, three of the four free parameters of the model are theoretically time-step dependent (x_2, x_3, and x_4), and their theoretical relationships can be derived from the integration of the model governing equations (see also Le Moine, 2008, pp. 172-173). So, the comparison of the calibrated parameters at different time steps has to be done after normalization at a same reference time step by using these theoretical relationships (see Table 2.3 in Chapter 2). We have shown a brief statistical summary of the changes in the distributions of the four parameters of the GR4 baseline model in Appendix F (Figure F.1). Here we further analyse this issue by searching for possible links between the fluxes inconsistencies and the deviations of the parameters from their theoretical time-step dependencies.

To provide a first overview of the impact of time step on parameters, Figure 4.9 shows the comparison of the parameters calibrated at the hourly and daily time steps on our catchment set. The daily parameter values are normalized at an hourly reference by using their theoretical relationships (see Table 2.3, Chapter 2). For a consistent multi-time step model, the points should be aligned along the 1:1 line. In general a good but not perfect coherence may be observed in the four graphs. At first glance, the most marked deviation from the identity line regards the time base of the unit hydrograph (x_4). This deviation is due to the unit hydrographs discrete construction, which leads to increasingly precise estimations of the x_4 parameter as the time step decreases (as further discussed in Appendix F). The other parameters seem to be relatively more consistent, even if an asymmetric dispersion around the

identity line seems to appear for x_1 and x_2, for which a trend of decrease of the parameters at shorter time steps seems to be plausible. However, a thorough comparison cannot be done by means of Figure 4.9, because the ranges of the parameter values are too large compared to their average deviations at different time steps. To complement the analysis, we must further analyse the deviations of the parameters from their theoretical time step dependency by taking the daily model parameters as a reference (see Table 4.5 and Figure 4.10).

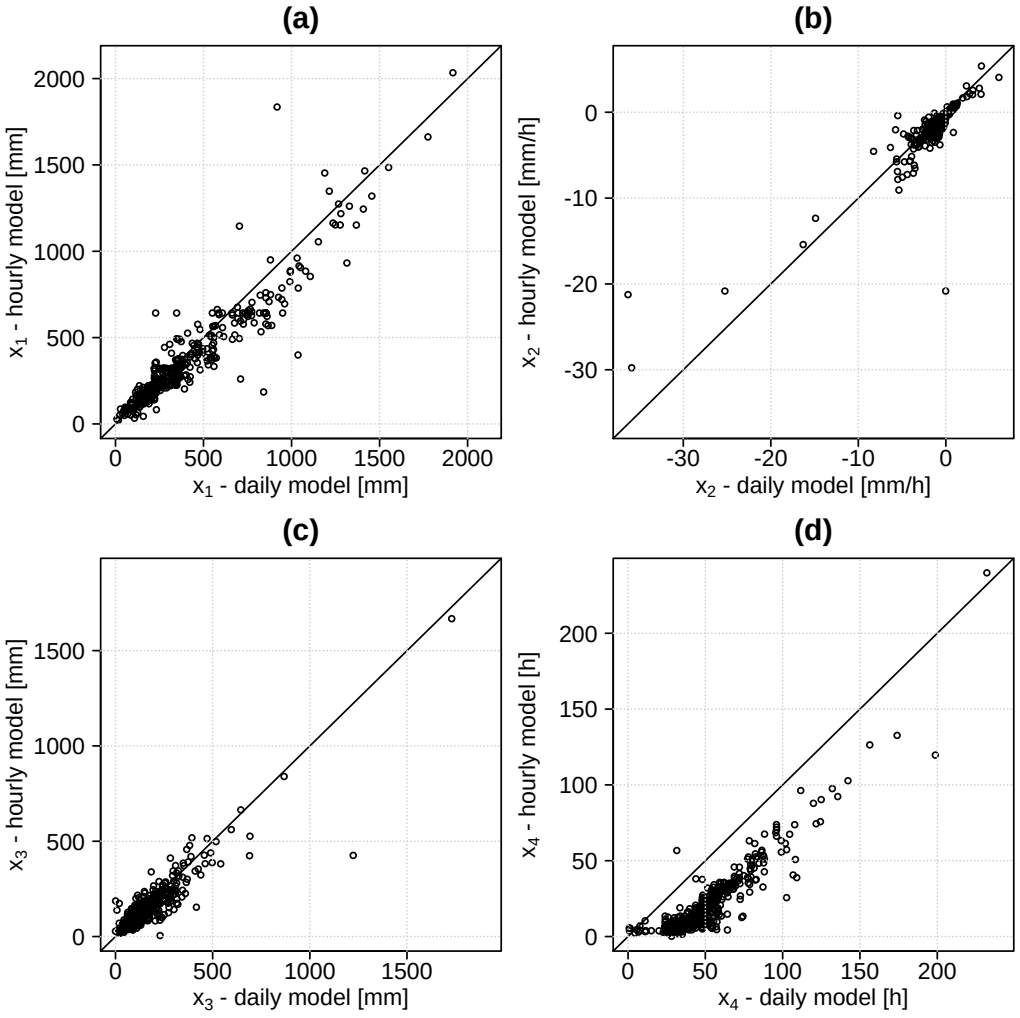


FIGURE 4.9 – Comparison of the GR4 model parameters calibrated at the hourly and daily time steps over the 240-catchment set (the daily model parameters are normalized at the hourly reference by their theoretical relationships): (a) capacity of the production store (x_1); (b) water exchange coefficient (x_2); (c) capacity of the routing store at one time step ahead (x_3); (d) time base of the unit hydrograph (x_4).

Table 4.5 reports some statistics useful to better understand the changes in parameters at the eight tested time steps. The median values of the parameters at the two extreme time steps, i.e. 1 day and 6 minutes, show large relative differences (compared to the parameter average value) especially for the x_2 and x_4 parameters. This trend is confirmed at all the tested time steps by the median values of the relative changes (see Δ_{rel} equation in Table 4.3) between the daily reference parameters and the corresponding calibrated values. Note that these relative

changes at different time steps are comparable to the relative standard deviation (modulus of the coefficient of variation) of the parameters over the catchment set.

Figure 4.10 shows the distributions of the deviations of the parameters calibrated at the eight time steps (from 6-minute to 1-day) from the corresponding value obtained by calibration at the daily time step (all parameters are normalized at the hourly reference). The optimal deviation for a consistent multi-time step model should be around zero, as the daily reference, at least for the parameters governing the water balance (above all x_1 and x_2, and x_3 to a minor extent). We report this complementary analysis also for the time base parameter, x_4, governing the linear routing of the unit hydrographs. However, the impact of time step on x_4 was already clear in Figure 4.9 and it has already been discussed (see also Le Moine, 2008) that it is normal that its values stabilize at shorter time steps, going towards more precise estimations of the catchment response time.

Parameter	**Units**	**Median parameter (*normalized at 1 h ref.*)**		**Relative standard deviation $\left\|\frac{\sigma}{\mu}\right\|$ [%]**		**Median relative change ($\Delta_{rel.}$) of parameter calibrated at time step Δt, with respect to the daily reference** $\Delta_{rel} = 100 \cdot [x_i(\Delta t) - x_i(1d)]/\|x_i(1d)\|$ [%] where $x_i(\Delta t)$ is the *i-th* parameter calibrated at t.s. Δt						
		1 d	**6 m**	**1 d**	**6 m**	**6 m**	**12 m**	**30 m**	**1 h**	**3 h**	**6 h**	**12 h**
x_1	[mm]	264	232	181	91	-10	-10	-9	-8	-7	-7	-5
x_2	[mm/h]	-0.58	-1.07	282	176	-59	-57	-53	-46	-41	-33	-24
x_3	[mm]	114	106	153	154	-6	-6	-5	-7	-6	-5	-2
x_4	[h]	48	12	50	123	-72	-71	-70	-68	-62	-53	-35

TABLE 4.5 – Median values of the GR4 parameters calibrated at daily and 6-min time step, their coefficient of variation, and the median relative changes in parameter values at seven time steps from 6-min to 12-h with respect to the daily value over the 240-catchment set.

This additional analysis allows disclosing some clear trends in the parameters dependency with time step also for x_1 and x_2, and to a smaller extent for x_3:

(i) On average, the capacity of the production store, x_1, decreases at sub-daily time steps with respect to the daily reference values. This decrease is significant in terms of ranks distributions (see results of the Friedman test in Figure 4.10) and is also important in relative terms (see Table 4.5).

(ii) The water exchange coefficient, x_2, steadily decreases with time step. The decrease is clearly significant in terms of distributions (see also Friedman test results) for all the time steps. In relative terms, the changes of x_2 with time step are the most impressive among the first three parameters, since they are the closest to the order of magnitude of the coefficient of variation (see Table 4.5).

(iii) On average, the capacity of the routing store at one time step ahead, x_3, slightly decreases (see Table 4.5). This change is not always significant for all the tested

time steps (see results of the Friedman test). The distributions of the parameter deviations from the daily values are quite symmetric around 0 (Figure 4.10) with increasing dispersion.

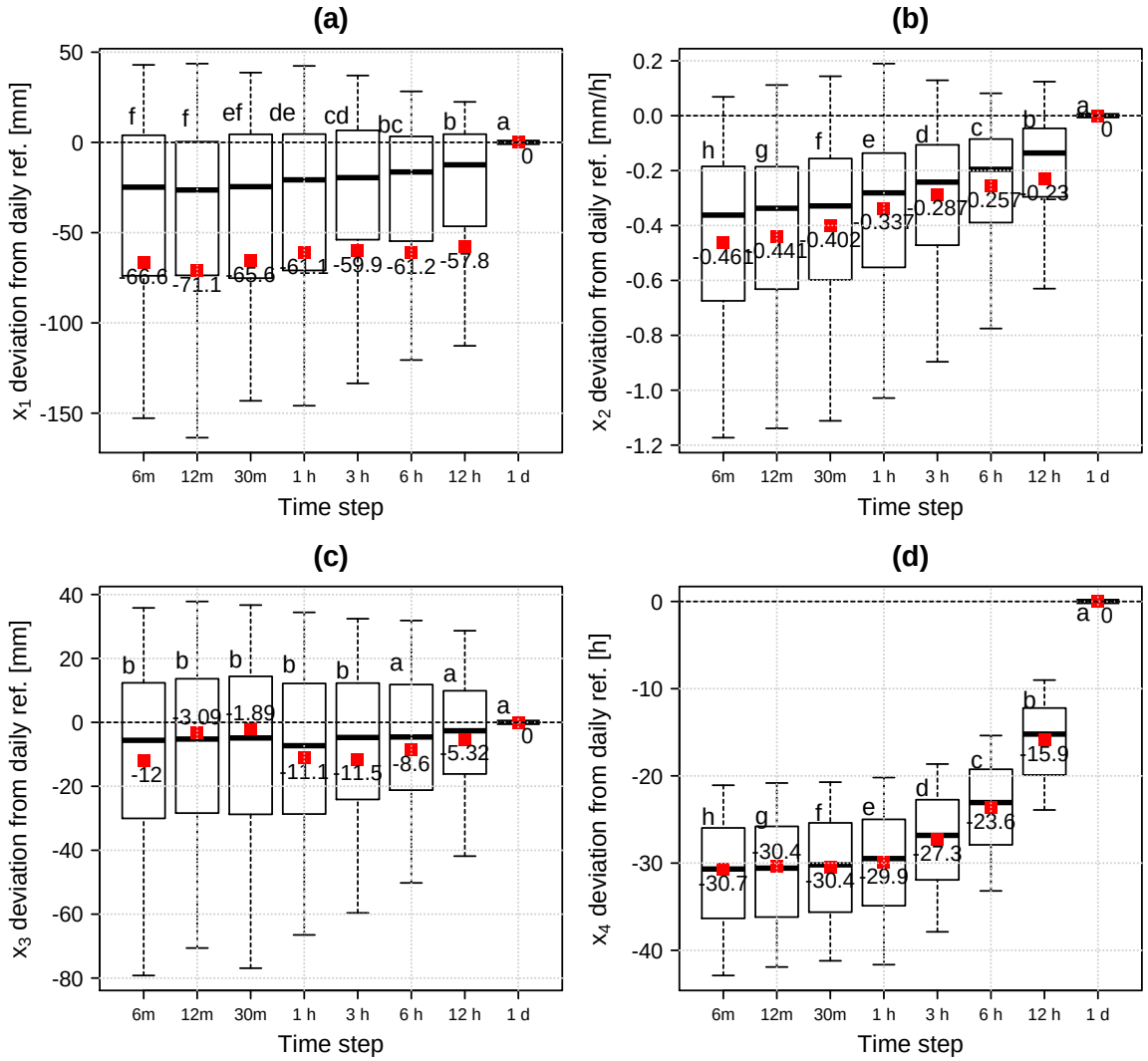


FIGURE 4.10 – Distributions of the deviations of the GR4 model parameters between the daily model reference and the eight time steps from 6 min to 1 day over the 240-catchment set (the parameters are normalized at the hourly reference by their theoretical relationships): (a) capacity of the production store (x_1); (b) water exchange coefficient (x_2); (c) capacity of the routing store at one time step ahead (x_3); (d) time base of the unit hydrograph (x_4). The box plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red points refer to mean values. The letters above each box plot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letter are not significantly different).

The presence of these trends highlighted above for the x_1, x_2 and x_3 parameter is coherent with our expectations following the model fluxes inconsistencies.

The impact of time step on the water exchange coefficient x_2 is the largest among the three parameters (after the logically-expected one on x_4) and it is straightforward to interpret. In fact, the GR4 model groundwater exchanges (see Eq. (4.4)-(4.6)) have the sign of x_2 and their absolute value increases with the absolute value of x_2. So, the observed changes in the water exchange coefficient correspond well to the observed changes in model fluxes at shorter time steps, i.e. the water losses increase (larger negative x_2 values) and the gains decrease (reduction of positive x_2 values).

On the other hand, the links between the changes in the x_1 and x_3 parameters and the corresponding governed fluxes (actual evaporation, *AE*, and exchanges, *F*, respectively) are less straightforward. Indeed, the fluxes depend on the evolution of the corresponding state variables, i.e. production and routing store levels respectively (see Eq. (4.2) and (4.3)). Moreover, the increase in the AE fluxes at shorter time steps is at least partially explained by the net evapotranspiration capacity ($E_n = E - I$) that increases as the interception loss decreases. So we further discuss the possible links between fluxes changes and production and routing store levels in the following section where we present the impact of time step on these state variables.

4.4.2 Model states

The GR4 model states include:

- the water content in the production store, i.e.: S [mm], linked to the non-linear functions of actual evaporation (see Eq. (4.2)), infiltration and percolation (Perrin et al., 2003);
- the level of the routing store, i.e. R [mm], linked to the exchanged fluxes (see Eq. (4.3)) and to its outflow, i.e. the main permanent component of the simulated streamflow;
- the unit hydrographs inputs, i.e. the rainfall excess that is being routed by the two unit hydrographs.

The linearity of the unit hydrographs (UH) ensures the consistency of the state-response function of the UH to the rainfall excess at different time steps. So here we focus only on the two first model states, i.e. the production and routing stores levels, which in addition are directly linked to the internal fluxes impacted by the model time step (*AE* and *F*).

Table 4.6 reports a summary of the distributions of the filling rates of the two stores (i.e. the ratios between the amount of water in the store and its capacity, i.e. $\frac{S}{x_1}$ for the production store and $\frac{R}{x_3}$ for the routing store) for the GR4 model at different time steps, over the whole catchment set. The statistics of the two stores levels are calculated over the 8-year simulation period at the same resolutions as the model time steps (from 6-min to 1-day).

Note that:

- The maximum capacity of the production store (x_1) does not depend theoretically on the time step (as already discussed in the previous section). So, any possible change of the filling rates of this store at multiple time steps would depend directly on the actual changes in the store inputs (P_n and E_n) and the concerned fluxes (*AE*, and the store filling and percolation).

- The filling rate of the routing store is calculated as the ratio of the stock and the actual maximum capacity (x_3) calibrated at each model time step without the normalization at a common reference time step (given by integration of the governing equation of the store) This choice is motivated by the importance of analysing the actual values of the filling rates, which are the ones directly impacting the exchange (see Eq. (4.4)).

Median of model stores filling rates (FR) [-] time-series statistics over the catchment set	**Model time step**							
	6 m	**12 m**	**30 m**	**1 h**	**3 h**	**6 h**	**12 h**	**1 d**
Production store mean FR	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56
Production store FR 10th quantile	0.31	0.30	0.30	0.30	0.30	0.30	0.30	0.29
Production store FR 90th quantile	0.78	0.78	0.79	0.78	0.78	0.79	0.79	0.78
Production store FR coefficient of variation	0.32	0.33	0.32	0.33	0.33	0.33	0.34	0.34
Routing store mean FR	0.13	0.16	0.20	0.24	0.31	0.37	0.43	0.51
Routing store FR 10th quantile	0.10	0.11	0.14	0.17	0.22	0.26	0.31	0.36
Routing store FR 90th quantile	0.17	0.20	0.26	0.31	0.40	0.48	0.56	0.66
Routing store FR coefficient of variation	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22

TABLE 4.6 – Summary of the distribution of the states of the production and routing stores of the GR4 model at different time steps over the whole validation period and the selected flood events for the 240-catchment set.

The results in Table 4.6 show that:

(i) For the production soil moisture accounting (SMA) store: the filling rates do not change on average when changing the model time step. The median value (for the 240 catchments) of the mean filling rate over the 8-years period is stable at about 56% across all model time steps. Also, the extreme quantiles (10th and 90th) and the coefficient of variation of the production store filling rate do not significantly change. This is coherent with the expected slow dynamics of this store, which must lead to a constant mean filling rate (for different model time steps).

(ii) For the routing store: the filling rates significantly decrease with the model time step. The median value steadily decreases from 51% at the daily time step to 13% at the 6-min time step. The same decreasing trend is observed on the extreme quantiles. It is interesting to note that the coefficient of variation is constant (around 22%) across time steps.

Some interesting interpretations of these results can be deduced by linking the statistics of the model stores filling rates at different time steps with the changes observed in parameters and fluxes.

4.4.2.1 Discussion on the changes in the production store filling rates

Figure 4.11 shows the coherence of the hourly and daily filling rates of the production reservoir for each catchment (mean and 90th quantile). Note that the slight dispersion around the 1:1 line (emerging for the 90th quantile) is likely to be explained by the fact that the statistics are calculated at each model resolution. For a perfectly fair comparison, the hourly simulations could have been preliminarily aggregated at the daily time step (before statistics calculation).

Note that even if the maximum capacity of the production store decreases on average at shorter time steps (see Table 4.5 and Figure 4.10), its filling rate is stable (Figure 4.11).

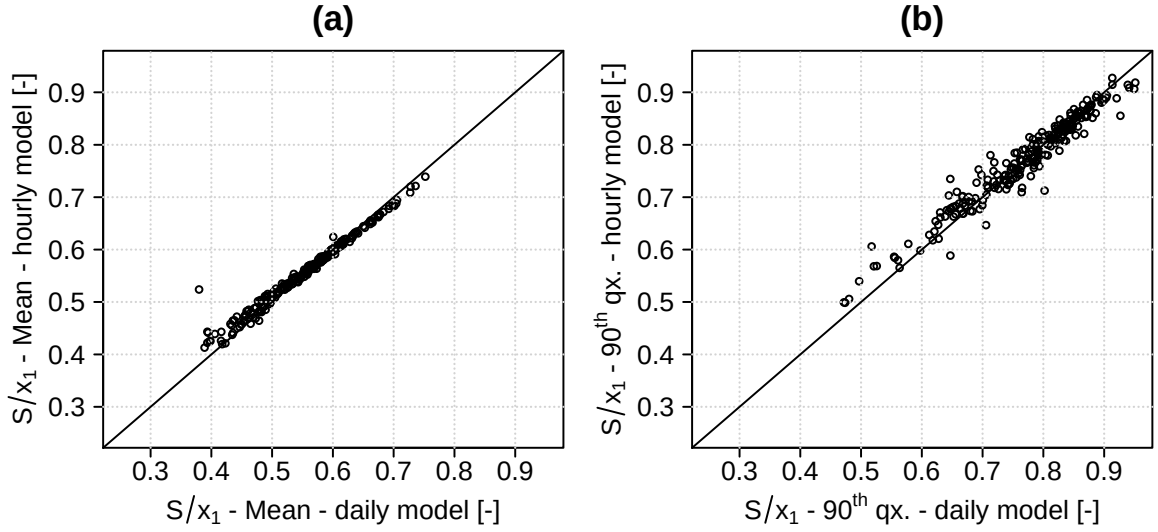


FIGURE 4.11 – Coherence of the production store filling rates between daily and hourly GR4 simulations over the catchment set: (a) mean values; (b) 90th quantiles of filling rates.

We have showed that by decreasing the time step of the GR4 model, the interception loss decreases (due to inputs temporal distribution) and, as a consequence, the actual evaporation from the production store (*AE*) increases (see Table 4.3). Since the level of the production store does not change, the *AE* changes can be explained by the differences of net precipitation and net potential evaporation inputs (due to interception loss changes). In fact, in Equation (4.2) we may note that the *AE* flux increases with the filling rate of the production store and

with the net evapotranspiration capacity ($E_n = E - E_i$). Since the filling rates are stable as the time step changes, the observed increase in AE fluxes at shorter time steps is essentially due to the increase of the net evapotranspiration capacity (E_n). So, the impact of the time step on the production part of the model is only due to the structural inadequacy of the interception component! This induces a feed-back on the actual evaporation from the production store, that is only due to the different inputs (E_n and P_n) and not to other structural problems.

4.4.2.2 Discussion on the changes in the routing store filling rates

The level of the routing store is linked to two internal fluxes of the GR4 model: (1) the exchange fluxes (F) and (2) the outflow of the store (Q_R).

(1) **Is there a link between the decreasing filling rates and the exchange?**

The exchange fluxes (F) depend on the routing store filling rate as we have reported in Eq. (4.3) (from Perrin et al., 2003). One may easily note from this equation that the absolute value of the exchange ($|F|$) increases with the filling rate of the routing store $\left(\frac{R}{x_3}\right)$. Thus, the decreasing trend observed in the routing store levels at shorter time steps is in opposition to the observed increase of magnitude of the exchange. This means that the decrease of the routing store filling rates is not caused by the "need" of increasing the water losses by the exchange component to compensate the reduced interception loss.

(2) **Is there a link between the decreasing filling rates and the outflow?**

The outflow of the routing reservoir is governed by a differential non-linear emptying function, such as:

$$\frac{dR}{dt} = -hR^{\alpha} \tag{4.10}$$

where h and α are two parameters that can be fixed or calibrated. In the GR4 model, h depends on the calibrated parameter x_3 (see here below), while α is fixed at 5.

As showed by previous authors who contributed to the GR4 models chain development (e.g. Le Moine, 2008, Appendix C, pp. 275-276), the integration of this equation on a time step gives the outflow of the reservoir given by Equation (20) in Perrin et al. (2003), here reported:

$$Q_R = R_t - R_{t+\Delta t} = R_t \left(1 - \left[1 + \left(\frac{R_t}{x_3} \right)^{\alpha - 1} \right]^{-\frac{1}{\alpha - 1}} \right) \quad \text{[mm]} \tag{4.11}$$

where we have:

$$x_3 = \left[\frac{1}{(\alpha - 1) \cdot h \cdot \Delta t} \right] \quad \text{[mm]} \tag{4.12}$$

which is used to transform the integrated equation over a time step and obtain an easy-to-interpret parameter to be calibrated (x_3), representing the 'reference' store-capacity one time step ahead (that cannot be exceeded at the end of the time step).

From Equation (4.12), as Le Moine (2008) observes, it is possible to obtain the relationship to transform the parameter x_3 from one time step Δt into the corresponding parameter ${x_3}'$ at time step $\Delta t'$, i.e.:

$$\frac{{x_3}'}{x_3} = \left(\frac{\Delta t}{\Delta t'}\right)^{\frac{1}{\alpha-1}} \tag{4.13}$$

For example, by passing from the time step $\Delta t = 24\ h$ to $\Delta t' = 1\ h$, the ratio of maximum 'reference' capacities is $24^{\frac{1}{\alpha-1}}$ (around 2.21, for $\alpha = 5$). Since the ${x_3}'$ capacity increases at shorter time steps and the average stocks should not change, the mean filling rates should be reduced as the time step decreases by the same factor of Equation (4.13).

So we found that the theoretical transformation described above explains well the observed trends reported in Table 4.6. In fact, the decrease in the routing store filling rates at shorter time steps corresponds to the transformation described above, not only for the mean filling rates but also for their 10th and 90th quantiles. One may easily check this relationship by verifying the transformation (Eq. (4.13)) applied to the statistics of Table 4.6. As a complement, Figure 4.12 shows the coherence of the hourly and daily filling rates of the routing store for each catchment (mean and 90th quantile) after applying the transformation of Equation (4.13), over the whole catchment set. These results are satisfactory, since they should allow inferring the model states (e.g. initial conditions of reservoirs before a storm event) at one short time step (e.g. 1 h) from the state at a larger time step (e.g. 1 day).

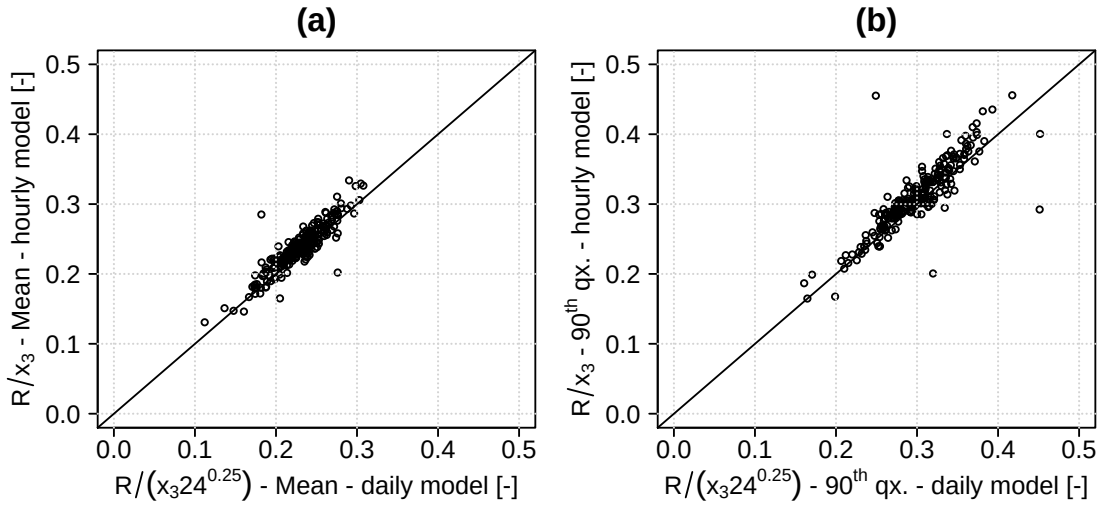


FIGURE 4.12 – Coherence of the routing store filling rates between daily and hourly GR4 simulations over the catchment set, by applying the theoretical relationship between time-step and routing store capacity: (a) mean values; (b) 90th quantiles of filling rates.

4.5 Synthesis

In this chapter, we have set up a diagnostic framework of the internal consistency of the GR4 model at multiple time steps. This analysis was based on targeted evaluations of the input-state-output response of the GR4 model. Particular attention was devoted to:

(i) the impact of inputs temporal distribution on model performance;

(ii) the consistency of the internal fluxes of the model across different time steps;

(iii) the impact of time step on calibrated parameters and model states.

Eventually, this manifold diagnosis led us to point towards serious inadequacies of the current model structure at sub-daily time steps and to understand possible ways of improvement.

Some problems in the performance of the GR4 model at sub-daily time steps were already presented in Chapter 3. In particular, we showed that for a significant part of the catchment set (about 20%) model performance degrades using shorter model time steps on flood events. It was shown that this degradation is mainly due to the undesired worsening of the *ratio of means* criterion (bias) by using shorter time steps for the majority of cases (even when average global model performance improves).

In our model diagnosis, we started analysing the impact of the temporal distribution of the precipitation and evapotranspiration inputs over the day, showing that the sub-daily distribution of precipitation inputs is more important than for evapotranspiration.

Despite the general improvement of simulations of flood events by using precipitation data at shorter time steps (i.e. shorter constant-intensity durations), some contrasting results were found. Above all, the water balance criterion follows an opposite trend being improved as the constant-intensity duration increases. This is counterintuitive because the use of larger data time steps corresponds to a loss of information (*P* temporal variability), while the volumes are consistent. Nevertheless, this problem is in line with what was found in Chapter 3, i.e. general significant decrease of the *ratio of means* over flood events with shorter model time steps. These results indicate that the capacity of the GR4 model to reproduce the water balance is jeopardized by simply using shorter temporal resolutions for precipitation inputs (even using the same model time step!). These results suggest some structural inadequacies that are strictly linked to the use of precipitation in the GR4 model equations at shorter time steps.

As for the potential evapotranspiration, we tested two different temporal patterns at sub-daily time steps: the uniform and "*sinusoidal*" patterns. The two patterns lead to very similar results (with no significant change) by evaluating either over the whole validation period or over flood events only. Thus, the sinusoidal temporal pattern is to be preferred to the uniform one, because it has a sounder physically-based basis, and is likely to lead to more consistent representations of the moisture fluxes.

The problem of water balance degradation at shorter time steps was linked to the rainfall-runoff model functioning, by a detailed analysis of the internal modelled fluxes inconsistency.

In fact, the interception loss, actual evapotranspiration and inter-catchment groundwater flows markedly change as the time step becomes finer. **The implicit assumption in the model structure of zero-capacity interception storage turned out to be not admissible at sub-daily time steps**. This was proved both empirically and formally: by our simulations of the GR4 model leading to large changes of interception loss volumes at different time steps (based on this assumption) and by a mathematical proof.

The large under-estimation of interception loss volumes at shorter time steps affects the other fluxes by strong effects of compensation. Since **the model components after the interception function receive more water at shorter time steps**, the other functions for water balance-closure, i.e. the actual evaporation from the production store and the losses by groundwater exchange, **are used to get rid of more water outside of the catchment**. These compensations seem to be able to ensure the good water balance at all time steps only by evaluating on average over the whole series, but **they lead to increasingly biased simulations over flood events**.

In the end of our model diagnosis, we evaluated the impact of the time step on model parameters and states. These impacts were analysed with particular attention to their links with the previously examined temporal inconsistencies of model fluxes.

For the GR4 model parameters, our analysis disclosed some clear trends of the values of the water-balance parameters (the capacity of production store x_1 and the water exchange coefficient x_2) across different time steps that differently from x_4 were neither *a-priori* expected nor obvious by looking at the distribution over the catchment set. However, the inconsistencies of model fluxes foretell possible *spurious* impacts on parameters that are eventually confirmed by our analyses (for x_1 and x_2). On average, the capacity of the production store, x_1, and the water exchange coefficient, x_2, decrease at sub-daily time steps with respect to the daily reference values. These impacts are directly related to the fluxes inconsistencies and to the model states. In fact, while the production store capacity (x_1) decreases the average filling rate of the store is constant across time steps. This confirms that the increase in actual evaporation directly depends on the different amounts of net precipitation and evaporation capacity at different time steps. On the other hand, the water exchange coefficient (x_2) decrease is directly explained by its role in the exchange fluxes equation. The latter equation depends on x_2 and x_3 (via the routing store filling rate), but only the x_2 changes in order to contribute compensating the reduced interception loss at shorter time steps.
As for the filling rates of the routing store, it is shown that they decrease with time step, but coherently to the integration of their outflow differential equation. Our analysis of the GR4 model states has an important practical implication: **the model states representing the levels of the production and routing stores are transferrable across different model time steps**.

In the next chapter, we will present our attempts performed to solve the problem of internal fluxes consistency. Our primary structural modification will be adding an interception store,

to favour the capacity of the model of evaporating intercepted water at sub-daily time steps. After solving the problem of the interception loss inconsistency, we should still focus on the actual role of the exchanges at different time steps. In fact, it has been shown that they seem determinant for the bias over flood events, while actual evaporation and interception loss counterbalance each other. Other exchange functions will be tested to see whether a reduction of the role of exchanges in high-flows conditions would further improve the water balance.

5
Towards a consistent multi-time step model

A modeller's Holy Grail: improving simulations accuracy by ensuring the fluxes consistency at multiple time steps

In Chapter 4 we have shown that the GR4 rainfall-runoff model is affected by problems of internal inconsistencies when run at multiple sub-daily time steps. It has been proved that these internal fluxes inconsistencies depend on the model structure. In particular, a structural inadequacy has been detected in the neutralisation function which represents the interception process at daily and sub-daily time steps. The GR4 model, with the neutralisation function, intercepts less and less water as the time step decreases from daily to 6 min. This is likely to be the cause of a chain of compensation effects on the model internal fluxes and parameters, which worsens the simulation bias over flood events at shorter time steps.

To solve this problem, a change of the interception component is necessary. This structural modification is tested by adding a simple interception store as the first model component in the production part of the GR4 model. We propose to fix its capacity parameter, in order to ensure the interception flux consistency. The results of this new model structure are evaluated by criteria on model performance and on the temporal consistency of the modelled fluxes. The two aspects are evaluated over the whole simulation period and flood events only. This analysis validates the insertion of the interception store in the GR4 model at sub-daily time steps, because it leads to significant benefits for model fluxes temporal consistency and for model performance over floods.

However, a decrease of performance over low flows is detected. It is argued that this problem could be solved by testing alternative exchange functions. The use of the exchange function proposed by Le Moine (2008) in combination with the model with the interception store proposed here solves the problem on low flows and provides further improvement to model performance over flood events (reduction of the bias).

Finally, other complementary tests are performed to take into account the precipitation intensity in the model. Our tested implementations of an infiltration-excess runoff component and a precipitation correction factor are briefly presented. They did not provide any significant improvement to the current version of the model. However, these tests are not exhaustive and some of these issues may deserve further research in the future.

5.1 General scheme of our model structure modifications

In this chapter, we present the structural modifications of the GR4 model that have been tested at daily and sub-daily time steps with the aims of:

(i) Improving model **performance** and particularly reduce the simulations bias (underestimation) on flood events that was detected at daily and even more at sub-daily time steps;

(ii) Improving the internal **fluxes coherence** at multiple time steps and particularly stabilize the interception loss and the other interdependent model fluxes (actual evaporation from production reservoir and inter-catchment groundwater exchanges).

We apply the testing protocol for model calibration and validation used in the previous tests of the GR4 model at multiple time steps (see Chapter 3 and 4). Calibration is performed using the KGE criterion on streamflows at all the time steps of the time series as objective function. As explained in Chapter 4, we propose a two-step approach for empirical model identification of a consistent multi-time step model, which complements the validation of model performance with the validation of the consistency of model fluxes and parameters across time scales. In our approach, **a structural modification is accepted if and only if most of the criteria of model performance are improved (or not degraded) and the fluxes consistency criterion is improved (or not degraded) too**.

In the validation process, model performance will be evaluated using different criteria on streamflow over the whole validation period and over flood events, i.e. KGE, relative variability, ratio of means, correlation and FDC-based criteria (extreme quantiles of flow ratios and slope bias in the mid-segment of the FDC). Note that since our evaluation of model performance is based on a multi-criteria approach, the problem of model identification is not an easy task. A compromise choice must be made by the modeller if significant improvements in some criteria are detected at the price of a degradation of other criteria.

For the validation of the temporal consistency of fluxes (at multiple time steps), we will use the multiplicative bias indexes defined in Chapter 4 and called ***cumulative flux ratios*** (i.e. ratios of cumulated fluxes simulated at different time steps) on the whole validation time series and on the flood events only.

Regarding the models naming, the **baseline model structure**, simply called **GR4**, refers to the structure described in Perrin et al. (2003) and run at multiple time steps by simple adaptation of its fixed time-dependent parameters (i.e. the model used in the previous chapters). We remind that the number in the model name stands for the number of free parameters (4). Analogously, the model versions presented in this chapter will be called by a code (i.e. an acronym) starting with *GRX* where *X* is the number of free parameters of the new version. The final part of the code is composed of letters and numbers, with letters

abbreviating a particular structural modification and numbers standing either for the ordinal number of the implemented structure and/or for a new fixed parameter value, as for instance: '*GR4-I1*' stands for GR4 baseline model with a first implementation of an interception model component.

For sake of clarity, sometimes we will refer to the "*order*" of a new derived structure that can be defined as the number of model components involved by the structural modification with respect to the baseline GR4 model. For example, a *first-order* derived model will concern the insertion (or modification) of only one particular model component (function) with respect to the baseline model (e.g. the addition of only an interception model component or the sole modification of the groundwater exchange function lead to first-order derived models). The combination of two first-order model adaptations in one new structure generates a *second-order* derived structure, and so on.

To summarize all the modifications, the following four points recall the *first-order modifications* of the GR4 baseline model that we have tested at multiple time steps and that will be presented in this chapter:

I. Insertion of an **interception** storage component (see Section 5.2);

II. Modification of the **groundwater exchange function** (see Section 5.3);

III. Insertion of a function accounting for **infiltration-excess capacity** (see Section 5.4.1);

IV. Insertion of a **precipitation correction** factor (see Section 5.4.2).

The largest part of our work has been devoted to the first point (interception), and will be thoroughly presented in this chapter. Our investigations on the other three types of structural modifications have been less exhaustive and will be presented in fewer details. Only some of all the possible higher-order modifications (given by the combination of the first-order ones) have been tested and will be presented in this chapter, such as the combination of the insertion of an interception store with a modified exchange function saturating on high-flows (see Section 5.3).

5.2 Refinement of the interception component

The need of a refinement of the interception component of the model is motivated by the results of our model diagnosis (Chapter 4): the neutralisation function currently used in the GR4 model proved inadequate to consistently represent the interception process across daily and sub-daily time steps, as demonstrated by both an empirical analysis and a formal proof. This inadequacy is the cause of large undesired time-step effects on model internal fluxes and parameters, which are presumably related to a degradation of the model simulations over flood events (i.e. increasing bias as the time step decreases).

Given the observed compensations between model fluxes, it seems necessary to stabilize in priority the interception loss at different time steps in order to make all the model fluxes more

consistent. This could be done by extending the memory of the function used for representing interception to durations longer than the model time step by using a store.

Thus, we will test the insertion of an interception store in the GR4 model at daily and sub-daily time steps, to replace the temporally inconsistent neutralisation function. A similar complexification of the interception component of the GR4 model was already tested in a previous study on the development of the hourly version of the model, without detecting any significant improvement of model performance (Mathevet, 2005; pp. 220-221). Mathevet (2005) tested the insertion of an interception store of either fixed or free capacity in the GR4 hourly model but he did not report in his PhD thesis the tested function for interception and detailed results. According to his evaluation framework, he claimed that the insertion of an interception store would not lead to any significant improvement in model performance. However, unlike us, Mathevet (2005) did not evaluate neither the model fluxes consistency nor the simulation performance on flood events. For this reason, we may expect that the refinement of the interception component deserves more attention and that we could find different conclusions than Mathevet (2005), thanks to our different evaluation framework. Also, Mathevet (2005) (p. 223) recognizes that his efforts in the hourly model development focused more on the routing part of the model than on the production part. So, he argued that further work should deepen the question of optimality of the production functions (interception, evaporation and infiltration).

5.2.1 A new structure with a bucket-style interception store

For the complexification of the interception component, we consider the insertion of a simple bucket-style store at the top of the current GR4 model structure (see Figure 5.1). The **interception store** has a **capacity of a few millimetres** (I_{max}), which allows to temporarily store an amount of intercepted rainfall (I) for a few hours or days. The stock of water in the store overflows and produces **throughfall** (P_{th}) when the capacity I_{max} is exceeded, while is reduced by **evaporation** (E_i) **at the potential rate**, unless the effective water supply rate ($P+I/\Delta t$) is limiting. This simple type of interception model component has been already used in rainfall-runoff modelling at sub-daily time steps by other authors (e.g. Kandel et al., 2005).

The interception store capacity (I_{max}) could be either fixed or calibrated, according to criteria based on model performance and internal fluxes consistency. We will evaluate both possibilities by empirical modelling tests. As for denoting the new versions of the model, we name as **GR4-I** the model obtained by adding an interception store of fixed capacity and as **GR5-I** the model with the interception store of free capacity (which would become the 5th free parameter of the model).

Given the fixed time-step numerical implementation of the model, the order of the interactions of the two climatic inputs (P and E_p) with the interception store could impact the modelling results, especially at larger time steps. Two possible implementations of the interception model component represented in Figure 5.1 are possible, depending on the priority given at the beginning of each fixed time step to either throughfall or evaporation

from the store. In the first implementation (named I1), the throughfall (P_{th}) is calculated prior to evaporation (E_i), while in the second (named I2), the evaporation is prior to throughfall (see Table 5.1). In the following, the derived model structures of these two implementations are denoted respectively as GR4-I1 (or GR5-I1) and GR4-I2 (or GR5-I2).

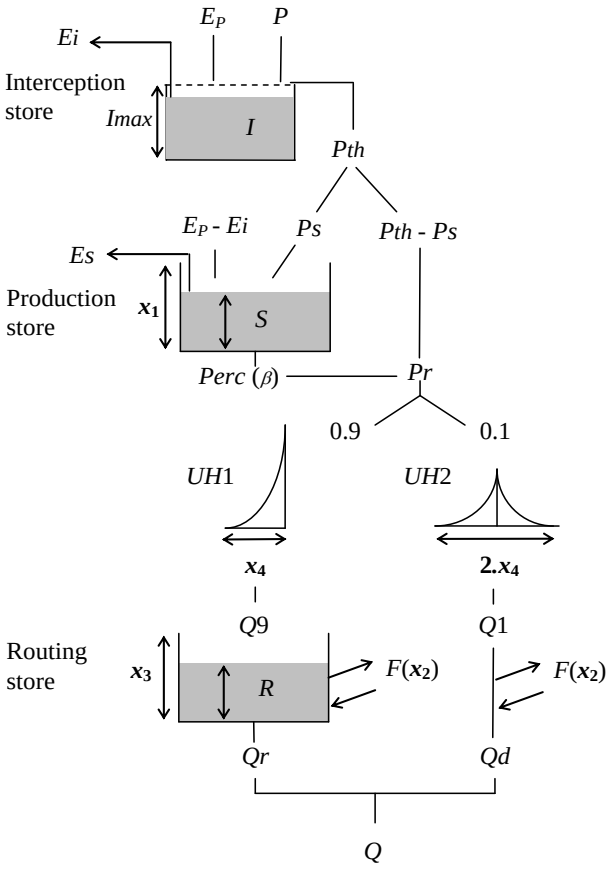


FIGURE 5.1 – Schematic representation of the GR4-I and GR5-I model structure, modified from the GR4 baseline model by Perrin et al. (2003) by insertion of an interception store (of maximum capacity I_{max}). The store capacity may be either fixed (in GR4-I) or calibrated as a fifth free parameter (in GR5-I).

Since part of the potential evapotranspiration energy available (E_p) is used to evaporate water from the interception store, only the difference ($E_P - E_i$) remains available to evaporate water from the production soil-moisture accounting store. In other words, with respect to the GR4 baseline model and the corresponding notation established by model equations in Perrin et al. (2003) (see also Section 2.6.1), in the GR4-I and GR5-I model structures (for both -I1 and -I2 versions) the net evaporation capacity E_n is equal to $E_P - E_i$. In the same way, the rainfall intensities are smoothed by the interception store, and the throughfall P_{th} is the new net rainfall P_n.

The total actual evaporation loss (E_A) will be the sum of E_i and E_s, where E_s is the amount of water evaporated from the production store (see Eq. (2.9) in Section 2.6.1), calculated here as a function of the new net evaporation capacity:

$$E_s = \frac{S\left(2 - \frac{S}{x_1}\right)\tanh\left(\frac{E_P - E_i}{x_1}\right)}{1 + \left(1 - \frac{S}{x_1}\right)\tanh\left(\frac{E_P - E_i}{x_1}\right)} \quad (5.1)$$

Interception store with throughfall prior to evaporation (I1) – Equations over a time step Δt	**Interception store with evaporation prior to throughfall (I2) – Equations over a time step Δt**
1 – The throughfall (P_{th}) rate is calculated as: $P_{th} = max\left[0, P - \frac{(I_{max} - I_0)}{\Delta t}\right]$ 2 – The actual evaporation (E_i) rate is calculated as: $E_i = \min(E_P, P + \frac{I_0}{\Delta t} - P_{th})$ 3 – The *I* store water content is then updated as: $I = I_0 + (P - P_{th} - E_i)\Delta t$	1 – The actual evaporation (E_i) rate is calculated as: $E_i = \min(E_P, P + \frac{I_0}{\Delta t})$ 2 - The throughfall (P_{th}) is calculated as: $P_{th} = max\left[0, P - \frac{(I_{max} - I_0)}{\Delta t} - E_i\right]$ 3 – The *I* store water content is then updated as: $I = I_0 + (P - E_i - P_{th})\Delta t$

TABLE 5.1 – Equations of the two implementations of the interception store depending on the priority given to either evaporation or throughfall (model names with either I1 or I2). P and E_P are the rainfall and potential evapotranspiration rates over the time step Δt, I_{max} is the interception store capacity; I_0 is the initial water content in the interception store (at the beginning of the time step Δt); E_i is the actual evaporation rate from the *I* store; P_{th} is the throughfall rate from the *I* store.

5.2.2 Results of the GR4-I model with an interception store of fixed capacity

The interception store capacity I_{max} was initially fixed, with different values (ranging from 0 to 15 mm), to test the sensitivity of model results to this parameter. This choice was made to avoid possible interactions with the other production parameters of the model (i.e. capacity of production store and water exchange coefficient) in the calibration step. We have chosen to test a limited number of fixed values (less than 20) to limit the number of models runs at short time steps and so the calculation time. The chosen set of possible fixed values, named '*F*', consists in the following eighteen values:

- 0.1 mm;
- all values between 0.25 and 3 mm with a step of 0.25 mm;
- 4, 5, 7.5, 10 and 15 mm.

Note that the neutralisation function of the GR4 baseline model is equivalent to an interception store of null capacity, and so the tested capacity values are nineteen.

The range of I_{max} values was chosen by extending of a few millimetres the typical range of interception storage capacities reported in the literature for different land cover types (e.g. Gerrits, 2010). The choice of considering more values between 0.25 and 3 mm was based also on our first analysis on the fluxes consistency at multiple time steps. In fact, this analysis indicated that the capacities between 0.25 and 3 mm were more effective to stabilize the internal fluxes at multiple time steps, as reported in details later (see Section 5.2.4).

For each I_{max} value in the *F* set, we recalibrated the GR4-I models, and so the four free parameters that are in common with the GR4 baseline model are changed in the calibration phase. To summarize the results of the GR4-I1 and GR4-I2 models (19 modelling tests for each version and time step), we selected for each catchment the fixed value of interception store capacity (among the nineteen) that leads to the maximum KGE value in calibration mode. We named this fixed catchment-dependent capacity as *'F-opt'*, to indicate that it consists of the optimal I_{max} values within the *F* set.

In this way, the selected results provide a first estimation of the upper-bound of performance that can be achieved by adding an interception store of fixed capacity (within the limited *F* set). A summary of the model performance at hourly and daily time steps is provided respectively in Table 5.2 and Table 5.3 for the two new model versions (GR4-I1 and -I2) and for the baseline model without interception store (GR4).

The results at the hourly time step (Table 5.2) show that, on average, the GR4-I1 and -I2 models lead to a marked improvement with respect to the GR4 baseline model for both evaluations over the whole 8-year period and over flood events. The average improvement in KGE is of about 1 point (i.e. 10^{-2}) over the whole period and 3 points over flood events with both GR4-I1 and -I2 implementations. It is important to note the positive pronounced trend in the ratio of means (*b*) in flood conditions. At the hourly time step, the median ratio of means increases from about 90% with the GR4 model to 93% with the GR4-I models with interception store. In other words, the average flood volumes underestimation is significantly reduced by introducing the interception store in the GR4 model: *the flood volumes are underestimated on average by 7% with the new GR4-I model, while they were underestimated by 10% with the GR4 baseline model.* This confirms that **the introduction of the interception store is especially advantageous for improving the simulation of flood events with respect to the baseline model**, as it was suggested in Chapter 4.

On the other hand, at the daily time step (Table 5.3) the differences in average performance between GR4 and GR4-I1/-I2 are smaller, and over the whole validation period they are negligible ($<10^{-2}$). A significant improvement (1.5 points of KGE) is still detected by evaluating over flood events, but is clearly less pronounced than for the hourly time step.

Hourly model	Median I_{max} store capacity [mm]	Median criteria on the whole series (validation)				Median criteria on flood events (validation)			
		KGE [-]	a [-]	b [-]	r [-]	KGE [-]	a [-]	b [-]	r [-]
GR4	0	0.820	0.989	1.009	0.897	0.727	0.969	0.904	0.832
GR4-I1 (*F-opt*)	7.5	0.832	0.998	0.999	0.912	0.759	0.975	0.929	0.839
GR4-I2 (*F-opt*)	7.5	0.833	0.999	0.999	0.912	0.759	0.972	0.928	0.841

TABLE 5.2 – Summary of the median performance criteria of the **hourly** GR4 (baseline), GR4-I1 and -I2 models, with fixed capacity I_{max}, over the catchment set. The criteria presented are the KGE and its components over the whole validation period and over the selected flood events. The results are presented for the I_{max} capacity (*F-opt*) which provides the maximum KGE value in calibration mode, selected for each catchment among the set of 19 values between 0 and 15 mm.

Daily model	Median I_{max} store capacity [mm]	Median criteria on the whole series (validation)				Median criteria on flood events (validation)			
		KGE [-]	a [-]	b [-]	r [-]	KGE [-]	a [-]	b [-]	r [-]
GR4	0	0.836	0.992	1.005	0.906	0.718	0.934	0.935	0.800
GR4-I1 (*F-opt*)	5.0	0.836	0.998	1.000	0.915	0.735	0.938	0.950	0.807
GR4-I2 (*F-opt*)	5.0	0.833	0.997	1.001	0.914	0.731	0.936	0.947	0.811

TABLE 5.3 – Summary of the median performance criteria of the **daily** GR4 (baseline), GR4-I1 and -I2 models, with fixed capacity I_{max}, over the catchment set. The criteria presented are the KGE and its components over the whole validation period and over the selected flood events. The results are presented for the I_{max} capacity (*F-opt*) which provides the maximum KGE value in calibration mode, selected for each catchment among the set of 19 values between 0 and 15 mm.

Note that the two implementations of GR4-I1 and -I2 lead to very similar model performance scores at the hourly and daily time steps by both evaluations over the whole time series and over flood events. The differences in the median values of KGE and its components between these two implementations are negligible, being in the order of 10^{-3}. Thus, with respect to model performance, a choice between the two model versions (GR4-I1 and -I2) would be arbitrary. The same conclusion is inferred by looking at the performance scores at all the other sub-daily time steps, from 6 min to 12 h.

In Figure 5.2, we provide a more detailed summary of model performance scores over the catchment set for the evaluation over flood events. The statistics of performance of the GR4-I2 model are presented for a representative sub-set of interception capacities ranging from 0 to 15 mm and for the sub-set of optimal capacities selected in the set *F (F-opt)*. The Friedman

test shows the significant improvement in the whole distribution of the KGE and ratio-of-means criteria, when passing from zero-interception storage to any interception store capacity greater or equal to 1 mm. On the contrary, on average, the relative variability and the correlation are more stable across the different tests reported in Figure 5.2. The differences in performance between the GR4-I model with the catchment-dependent interception capacities (*F-opt*) and with the fixed capacity between 5 and 10 mm (for all catchments) are not significant, except for the correlation (see results of the test in Figure 5.2). Also, the maximum median KGE over floods for a single fixed capacity (among the 19) for all catchments is obtained with the interception capacity of 2.5 mm. However, the performance scores of the models with capacities from 2 to 10 mm are not significantly different according to the Friedman test (note that the test was applied for all the 19 tested capacities and not only the ones reported in the figure).

The average performance scores reported in Table 5.2 and Table 5.3 are associated to large median interception store capacities, i.e. 7.5 and 5 mm respectively at the hourly and daily time steps. It is interesting to note that these values are larger than the typical values reported in the literature (i.e. observations), even if an accurate physical interpretation of our simple interception model is not our objective and is made difficult by the representativeness of the measurements at the basin scale. One may refer for example to the review of canopy interception by Gerrits (2010) (see her Table 1.1 and references therein). The values of typical water storage capacities reported for France range from 1.7 to 3.8 mm, with values lower than 2 mm for broad-leaved trees forests and values larger than 3 mm for coniferous (see Aussenac, 1968).

So, the GR4 model benefits more in terms of model performance from higher values of interception storage capacity than what has been reported in the literature (even if the comparison is rough). This could be linked to the specific internal functioning of the GR4 model and in particular to what happens inside the model when the interception capacity increases. As highlighted in Chapter 4, there is an internal feedback mechanism that compensates for interception loss and exchange fluxes. This is verified by looking at the changes of the parameters of the GR4-I model for different interception store capacities (see Figure 5.3). In fact, when adding an interception store of increasing capacity, we observe a consequent decrease (in absolute value) in the water exchange coefficient, as shown in Figure 5.3(b). At the hourly time step, the median water exchange coefficient (x_2) passes from -0.98 mm/h with the GR4 baseline model to -0.38 mm/h with both the GR4-I2 and GR4-I1 models with the interception capacities selected on the basis of performance (*F-opt*). This is a relative change from the baseline GR4 parameter of about 61%. At the daily time step, the analogous relative change is of about 42% (the median x_2 passes from -0.87 to -0.51 mm/d). The second maximum relative change is in the production store capacity (about 11% between GR4 and GR4-I2-F-opt). On average, the two other parameters are more stable when adding the interception store. So the need of higher values of interception capacities in the GR4-I model (to maximise model performance) is especially linked to this mechanism of compensations of the interception and exchange components of the model. This issue will be further discussed in terms of model fluxes in Section 5.2.5.

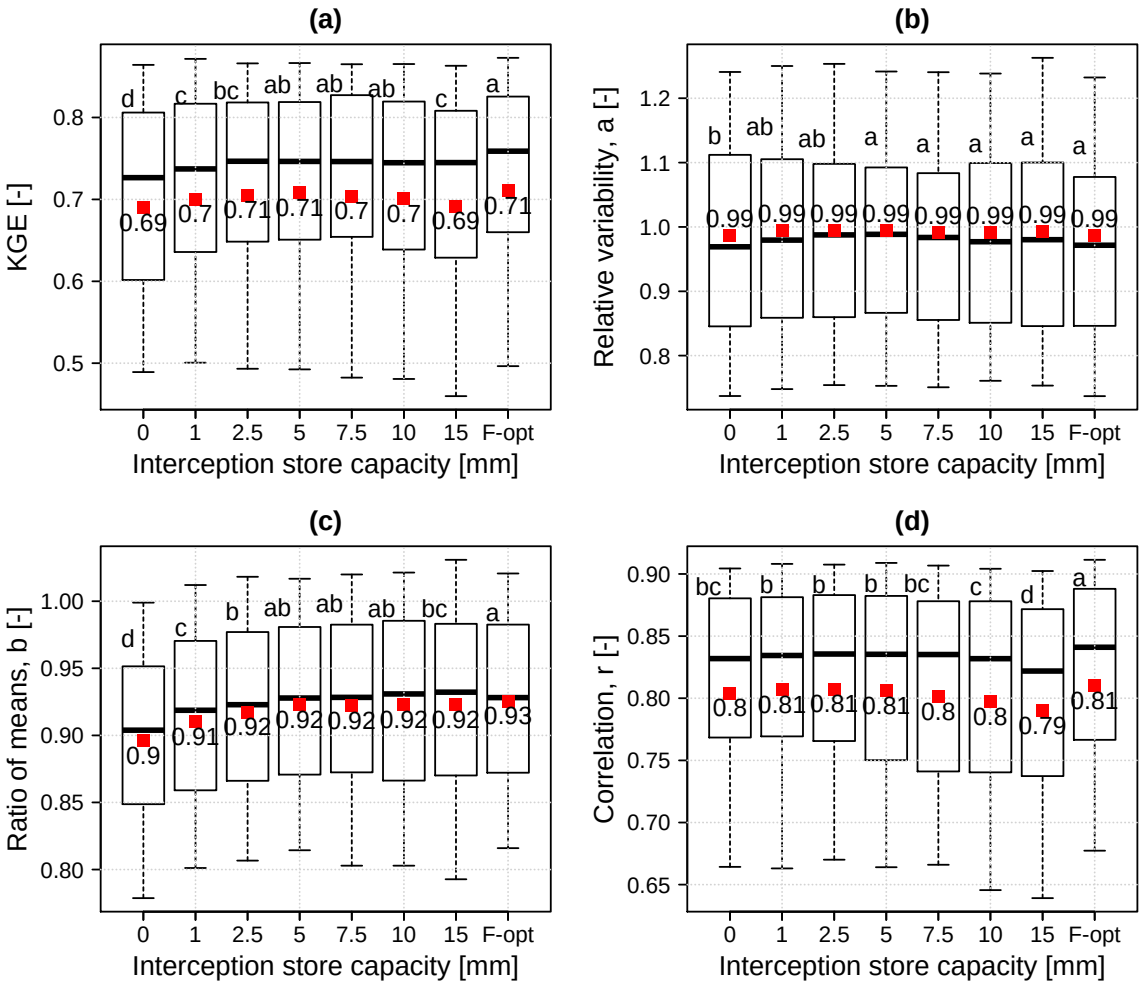


FIGURE 5.2 – Distribution over the catchment set of the performance criteria over the selected flood events for the GR4-I2 hourly model with different interception store capacities: (a) *KGE*, (b) relative variability, *a*; (c) ratio of means, *b*; and (d) correlation, *r.* The model with null interception store capacity corresponds to the GR4 baseline model, while the "F-opt" capacity corresponds to the catchment-specific capacity selected in the *F* set of 19 capacities between 0 and 15 mm. The box-plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red points refer to mean values. The letters above each box plot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letter are not significantly different).

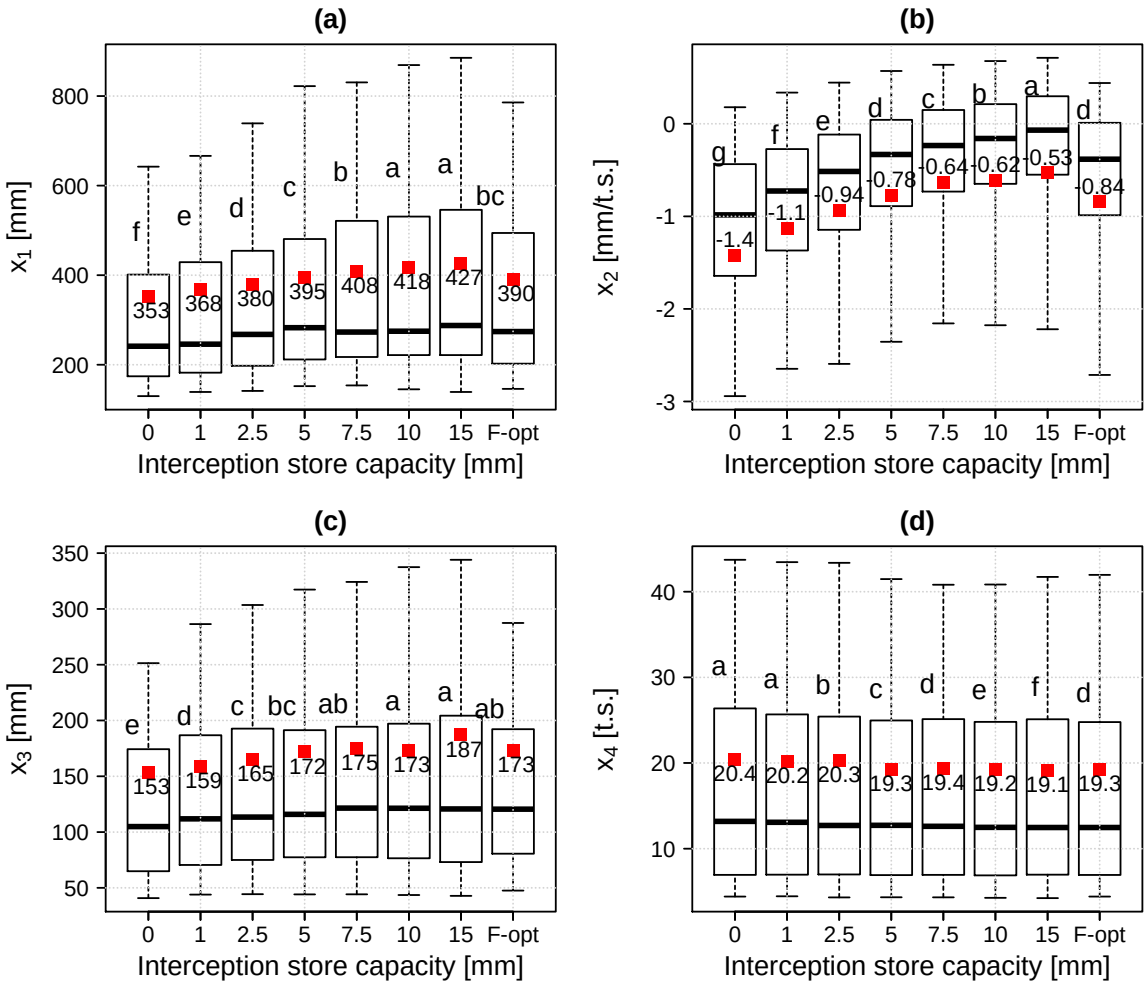


FIGURE 5.3 – Distribution over the catchment set of the calibrated values of the four free parameters of the GR4-I2 hourly model with different interception store capacities: (a) capacity of the production store (x_1); (b) water exchange coefficient (x_2); (c) capacity of the routing store at one time step ahead (x_3); (d) time base of the unit hydrograph (x_4). The model with null interception store capacity corresponds to the GR4 baseline model, while the "F-opt" capacity corresponds to the catchment-dependent capacity selected in the *F* set of 19 capacities between 0 and 15 mm. The box plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red points refer to mean values. The letters above each box plot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letter are not significantly different).

5.2.3 Results of the GR5-I model with interception store of free capacity

The findings reported in the previous section led us to question the possible benefits of calibrating the interception store capacity for each catchment by automated calibration, within a continuous domain, instead of choosing a fixed value within a limited set. So we tested the GR5-I model (Figure 5.1) with interception store with free capacity, bringing the number of calibrated parameters of the model from four to five.

Our scope here is to evaluate whether the performance of the GR5-I model is improved with respect to the GR4-I model, thanks to the increased degrees of freedom in the model, or if the increased number of free parameters would jeopardize model calibration and so performance and robustness. We remind that this analysis based on performance improvement is only the first step in our process of model identification at multiple time steps that will be followed by an analysis of the impact of the interception store on model fluxes consistency at different time steps (see next Section 5.2.5).

Table 5.4 and Table 5.5 report the average performance statistics for the two versions of the GR5-I model at the hourly and daily time steps. The two model versions (GR5-I1 and -I2) lead to the same average scores. Note that the median calibrated values of the interception capacity are very similar for the two model versions at the hourly time step, while they are more different at the daily time step. This is due to the fact that the different order of operations in the two implementations of the interception store has an impact only at larger time steps, as logically expected (because inputs are aggregated).

Hourly model	**Median I_{max} store capacity [mm]**	**Median criteria on the whole series (validation)**				**Median criteria on flood events (validation)**			
		KGE [-]	**a [-]**	**b [-]**	**r [-]**	**KGE [-]**	**a [-]**	**b [-]**	**r [-]**
GR5-I1	5.79	0.828	0.995	0.996	0.912	0.753	0.977	0.933	0.833
GR5-I2	5.96	0.829	0.996	0.997	0.911	0.752	0.976	0.934	0.834

TABLE 5.4 – Summary of the median performance criteria of the **hourly** GR5-I1 and -I2 models, with free (calibrated) capacity I_{max}, over the catchment set. The criteria presented are the KGE and its components over the whole validation period and over the selected flood events.

Daily model	**Median I_{max} store capacity [mm]**	**Median criteria on the whole series (validation)**				**Median criteria on flood events (validation)**			
		KGE [-]	**a [-]**	**b [-]**	**r [-]**	**KGE [-]**	**a [-]**	**b [-]**	**r [-]**
GR5-I1	6.45	0.830	1.001	1.000	0.912	0.725	0.935	0.946	0.807
GR5-I2	5.08	0.832	1.001	1.002	0.912	0.724	0.926	0.944	0.807

TABLE 5.5 – Summary of the median performance criteria of the **daily** GR5-I1 and -I2 models, with free (calibrated) capacity I_{max}, over the catchment set. The criteria presented are the KGE and its components over the whole validation period and over the selected flood events.

We compared the performance statistics of the GR5-I models with the statistics of the GR4 and GR4-I models with fixed interception store capacities reported in the previous section (Table 5.2 and Table 5.3; Figure 5.2). The average scores of the GR5-I models are slightly lower than those of the GR4-I models (with capacity *F-opt*) at both time steps and they are even lower than the scores of the GR4 baseline model at the daily time step. However, in general these differences in average performance are not significant (less than 10^{-2} in KGE and components values). Also, the Friedman test was applied to detect any significant

difference between the distributions of the performance scores of the GR5-I model and those of all the nineteen GR4-I models (for all tested fixed capacities). The test stated that there is no significant difference (at a 0.05 significance level) between the KGE scores over the whole series and over flood events for the GR5-I models and the GR4-I model with any fixed capacity between 2.25 and 10 mm (using the same capacity for all catchments). This means that there is no added value to be gained by calibrating the interception capacity together with the other four free parameters of the GR4 model.

In other words, the added complexity of a free interception store capacity is not necessary and may also be counterproductive in terms of model performance. The interception store capacity should be fixed independently from the other parameters and a search step of about 0.25 mm should be sufficient in terms of model performance. This finding could be explained by a reduction of the robustness of the model with an increased number of free parameters. The difficulty of calibrating the interception store capacity in the GR5-I model is revealed also by the interaction between this parameter and the water exchange coefficient, already discussed in the previous section (see Figure 5.3). By means of this interaction, the same levels of model performance are achieved for different parameters sets and this proves the presence of *equifinality* problems in the GR5-I model calibration. In this context, a possible way out of these problems is our second step of empirical model identification based on the temporal consistency of the internal fluxes.

5.2.4 Fixing the interception store capacity by seeking the fluxes coherence at different time steps

5.2.4.1 The choice of a reference for the interception flux

As demonstrated in the previous sections, the insertion of an interception store in the GR4 model may lead to significant improvements in model performance at the hourly time step, for both evaluations over the whole 8-year period and over flood events. At the daily time step, only a slight improvement over flood events was detected for the optimal fixed interception capacities (*F-opt*). No improvement of model performance was found by calibrating the interception parameter, with even a slight degradation at the daily time step. So in terms of model performance, our tests lead us to ***accept the insertion of an interception store in the GR4 model at hourly but not at the daily time step***.

Moreover we have shown that there is a problem of identification of the interception store capacity due to interactions with the exchange coefficient, if model identification is based on the sole basis of model performance. For this reason, we expect that the improvement of the temporal consistency of the model fluxes could help identifying this new component of the model at sub-daily time steps.

The choice of not accepting the insertion of the interception store at the daily time step, because of model performance issues, brings us to the possibility of considering as a reference the daily interception flux calculated by the neutralisation function. Moreover, the fact that the insertion of an interception store seems necessary only at sub-daily time steps is in line

also with the definition of the interception process in the literature. As argued by Savenije (2004), evaporation from interception may be defined "*as the fast feedback to the atmosphere (within a time span of about one day) of the rainfall that does not reach the root-zone or the drainage system*". Also Gerrits (2010) recognizes that "*the time scale of the interception process is in the order of one day*". The broad definition of interception and its characteristic time scale support our choice of using the daily neutralisation as the reference flux for the interception loss calculated by the interception store at sub-daily time steps.

Thus we searched for the interception store capacities $I_{opt\text{-}flux}$ that could ensure the coherence of the interception fluxes at different sub-daily time steps (from 6 min to 12 h) with the reference given by the daily neutralisation function. The criterion of fluxes consistency that we chose to optimize is the ***cumulative flux ratio*** (see Eq. (4.6) in Chapter 4). We remind that it is defined as the ratio of the cumulated flux simulated by the model at time step *x* normalized on the daily reference flux cumulated over the 8-year simulation period. So the cumulative flux ratio is equivalent to a ratio-of-means criterion (ratio of simulated flux on daily reference).

5.2.4.2 Determination of the interception store capacities ensuring the temporal consistency of the flux

A simple iterative search algorithm was set up in order to maximise the coherence of the interception fluxes at each time step with the daily reference flux. The iterative procedure is based on exploring the interception capacity space between 0 and 15 mm and calculating the corresponding interception loss at each model time step *x*. A search step of 1 mm is chosen for a first screening of the parameter space to detect a provisional (rough) optimal interception parameter for the temporal consistency of the fluxes (between time step *x* and the daily reference). Then the search is reiterated in a range of 2 mm around this first optimal value and the search step is reduced to 0.25 mm in order to find a refined solution. The effectiveness of the search step size is finally verified by checking that the cumulative flux ratio is close to 1 (optimal value) with a maximum tolerance distance of 5 % (arbitrary threshold). Applied to each catchment and for each time step, this procedure led us to identify the interception store capacity that is optimal for ensuring an estimation of the cumulated interception loss that is coherent across model time steps.

The average values (over the catchment set) of the resulting optimal interception capacities for the fluxes temporal consistency are reported in Table 5.6 for the two tested implementations (I1 and I2) of the interception store.

For the two implementations, the maximum relative deviation around the optimal cumulative flux ratios is about 3 % at any time step, and **the median relative difference in cumulated interception fluxes over the catchment set is lower than 0.5 % at all time steps**. These satisfactory results confirm that the discrete search step of 0.25 mm is fine enough for our scope. We remind that for the GR4 baseline model (without interception store) the *minimum* relative difference of the cumulated interception fluxes at different time steps was at least

30 % (and largely greater than this for time steps shorter than 12 h), as shown in Chapter 4 (see Figure 4.5).

Implementation of the interception store	**Median interception store capacity [mm] at time step *x* for the maximisation of the coherence of interception fluxes with daily reference**								**Max relative deviation from optimal cumulative flux ratio [%]**
	1 d	12 h	6 h	3 h	1 h	30 m	12 m	6 m	
I1	0	2.50	2.25	2.00	2.00	2.00	2.00	2.00	3.3
I2	0	1.25	1.50	1.75	2.00	2.00	2.00	2.00	3.1

TABLE 5.6 – Median values of the optimal interception capacities at different time steps over the catchment set. The criterion optimized for the temporal consistency of the simulated interception fluxes is the cumulative flux ratio of the simulated fluxes at time step *x* normalized with the daily reference.

The results in Table 5.6 show that, on average, for the two implementations, the interception capacity for ensuring the consistency of fluxes with the daily reference stabilizes at a value of 2 mm when the time step decreases from daily to 6 min. However, a different trend emerges for the two implementations for time steps larger than 3 h. The first implementation (I1) needs larger interception capacities at larger sub-daily time steps (e.g. 2.5 mm at 12 h), and presents a not monotonous trend (increasing between 1 d and 12 h and then decreasing at shorter time steps). This may be explained by the fact that the interception store with throughfall prior to evaporation (see equations in Table 5.1) needs larger capacities to avoid being bypassed by the precipitation at larger time steps, because the throughfall is calculated before evaporation. On the contrary, for the second implementation (I2), the trend of the optimal capacities for different time steps is monotonous, increasing from the daily zero-storage to the 2-mm capacity at the 6-min scale. For this reason, **we choose the second implementation for the interception store (I2)**, since we think that its behaviour is more intuitive for this monotonous relationship between the storage capacity and the time steps. This implementation is also in line with what is classically done in conceptual models (e.g. Kandel et al., 2005) and with the physical description of the interception process that is usually proposed (see, for instance, Aussenac, 1968, Section 3.2). In particular, it is recognized that evaporation phenomena mostly occur at the beginning of the interception process, because the surface temperature of the leaves may be higher than the air temperature. Then, with the continuation of the rainfall event, these phenomena rapidly reduce. Conversely, throughfall increases with time. At the beginning of a rainfall event, most of the water drops can be intercepted by the vegetation. Then intercepted water accumulates over the plant surfaces until the maximum storage capacity is reached, i.e. until the surface tension forces are exceeded by the force of gravity.

The distribution of the optimal interception capacities of the GR4-I2 model at different time steps is reported in Figure 5.4. It shows the stabilization of the parameter (to values around 2 mm) as the time step decreases. The relationship between interception capacity and time step was verified to be monotonous not only on average but also for each catchment (not shown).

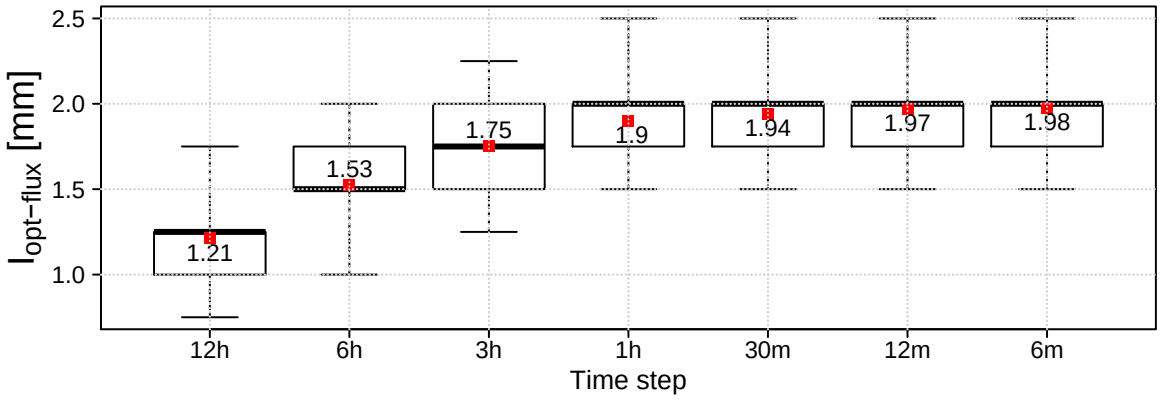


FIGURE 5.4 – Distribution over the catchment set of the optimal capacity of the interception store of the GR4-I2 model for ensuring the temporal consistency of the simulated interception flux with the daily reference flux. The box plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red points refer to mean values.

5.2.5 Results of the GR4-I2 model with interception store of catchment-dependent capacity based on fluxes consistency

After the determination of the interception capacities ensuring the temporal consistency of the interception flux, we have run the GR4-I2 model with these catchment-dependent capacities. In the following, we present the evaluation of this model from the point of view of the temporal consistency of the simulated fluxes corresponding to catchment water losses and gains: interception loss, actual evapotranspiration from the production store and groundwater exchanges. The motivations for analysing these specific internal fluxes of the model were introduced in Chapter 4. The reader can refer to Section 4.3.2 for the definition of these fluxes in the GR4 baseline model and for the definition of the *cumulative flux ratios* that will be used also in this section. Note that only the equations governing the interception fluxes have been directly changed in the GR4-I2 model with respect to the GR4 baseline model. So the changes in the other fluxes will indirectly result from this change (and the possible changes in optimized parameter values).

Here our scope is to analyse the effectiveness of the GR4-I2 model in order to reduce the large undesired changes of fluxes simulated by the GR4 baseline model at different time steps. We will analyse whether the stabilization of the interception flux as the time step changes in the GR4-I2 model leads to a stabilization of all the other model fluxes. In case this happens, we will evaluate whether the improved fluxes consistency leads to improved water balance simulations over floods.

5.2.5.1 Evaluation of the temporal consistency of fluxes over the whole validation period

Figure 5.5 reports a summary of the evolution of the internal fluxes of the GR4-I2 model cumulated over the whole 8-year validation period as the time step decreases from 1 day to 6

min (over the catchment set). The changes are normalized to the chosen reference given by the fluxes of the daily GR4 model (GR4-I2 model with zero-capacity interception storage).

Figure 5.5 shows that **the model fluxes have been stabilized across different time steps thanks to the introduction of the interception store**. In fact, the median value and the interquartile range of the cumulative flux ratios are centred around the optimal ratio (1) for all the fluxes at all time steps, unlike the baseline model case. The cumulated interception flux does not change of more than 1.5 % at all time steps. As a consequence, also all the other modelled fluxes deviate on average by less than 5% from the correspondent reference daily model fluxes. So, these results show that the insertion of an interception store in the GR4 model at sub-daily time steps can provide large improvements in terms of temporal consistency of the modelled fluxes over a long period. Figure 5.5 shows that, for interception loss and actual evaporation, the extreme quantiles (5^{th} and 95^{th}) of the cumulative flux ratios do not deviate much from 1 (about 5 % at most). Only for the exchange fluxes there are still some important relative changes (>20%). However, these large relative deviations can be associated to small absolute deviations (in mm) with respect to the whole annual water balance of the catchment, because the reference exchange flux can be small. Moreover, for the GR4-I2 model even the largest relative deviations in the exchange fluxes are much lower than the ones observed for the GR4 baseline model (see Figure 4.5). Note that the *y*-axis scales of the graphics in Figure 5.5 and Figure 4.5 are different. We chose this to allow detecting the relative changes of the actual fluxes of the GR4-I2 and GR4 models which are different (by more than a factor of 10 in general). If the same *y*-scales were used, the relative changes in the fluxes of the GR4-I2 model would be not visible because they would be crushed in a too narrow strip around the optimal ratio.

To summarize the changes in fluxes in absolute terms (mm), Table 5.7 provides the median values of the annual cumulated fluxes and the total net cumulated losses simulated at the eight tested time steps. The fluxes appear to be stabilized also by looking at their average absolute values simulated by the model at different time steps. Figure 5.6 shows the comparison of the cumulated fluxes between the hourly and daily models for all the 240 catchments. It shows a very good temporal consistency of the simulated fluxes over the whole catchment set, with almost all the catchments presenting approximately the same fluxes as the time step changes from daily to hourly. So the problem of large changes in cumulated fluxes is solved in general over the catchment set by the new version of the model with interception store (GR4-I2).

Also the problem of cases of groundwater exchange fluxes changing sign as the model time step changes is almost completely solved, as shown in Figure 5.6(c). There were 15 cases with the GR4 model, and now with the GR4-I2 model only 3 catchments still present this problem but with very small exchange fluxes (absolute cumulated annual values lower than 20 mm).

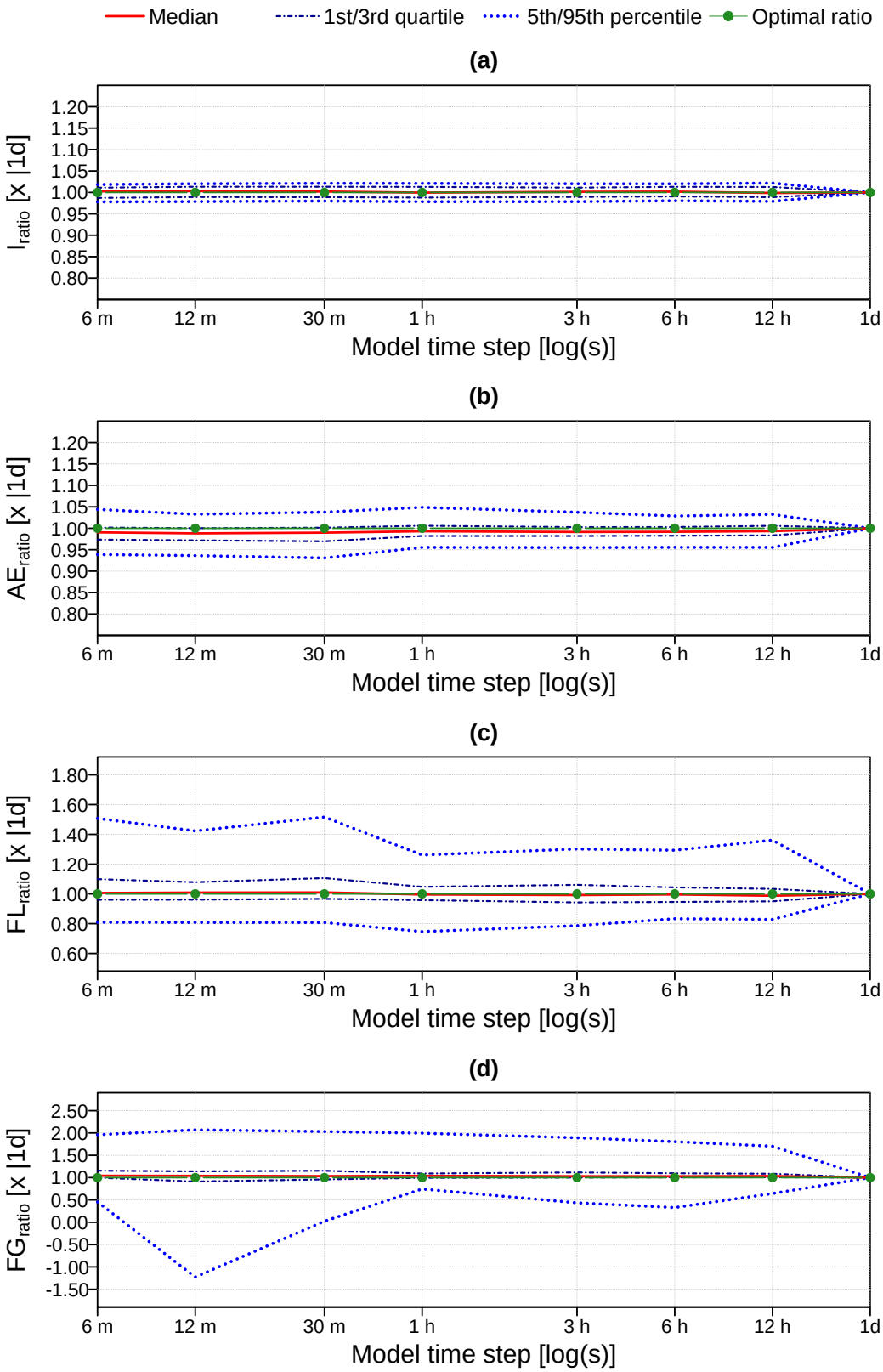


FIGURE 5.5 – Summary of the cumulative flux ratios of the GR4-I2 model at different time steps (with daily reference) over the whole validation period and the 240-catchment set: (a) interception loss, *I*; (b) actual evaporation from the production reservoir, *AE*; (c) groundwater losses, *FL*; (d) groundwater gains, *FG*.

In this general view of good consistency of modelled fluxes, only one very critical outlier is detected by looking at the scatterplots in Figure 5.6(c-d). It is the only point presenting a difference of daily and hourly fluxes of more than 67 mm/y (186 mm/y). This catchment is the Espérelle brook at Roque-Sainte-Marguerite (602 km^2; *Banque-Hydro* code: "O3395010"), that is characterized by a typical karst geology (limestone plateaus). The hydro-climatic characteristics of this catchment reveal that its significant water losses are determinant for the water balance: with inter-annual mean values of precipitation and potential evapotranspiration of 1200 mm/y and 725 mm/y respectively, the mean annual flow is only 55 mm/y! The drainage density indicator of this catchment is very high (2.6 km^2), indicating that water tends to infiltrate in the soil. The model calibration well detects this characteristic, by assigning one of the largest negative values to the exchange coefficient (about -21 mm/h). So, in this case, it is not surprising that a refined interception component does not manage to solve the problem of inconsistent values of exchange fluxes across time steps.

This particular case indicates that another problem of structural inconsistency (not driven by a compensation of the interception loss) seems to jeopardize the coherence of the exchange flux across time steps when the groundwater exchange flux is necessarily very high. This seems to be due to a structural problem located in the exchange function itself. So it seems that this problem should be solved by other specific structural modifications of the exchange function.

Annual cumulated flux statistics [mm/y] over the 8-years validation period	**Model time step**							
	6-m	**12-m**	**30-m**	**1-h**	**3-h**	**6-h**	**12-h**	**1-d**
Median annual interception loss, I [mm/y]	242	243	241	240	242	242	242	241
Median annual actual evaporation from production reservoir, AE [mm/y]	306	305	305	308	307	308	308	311
Median annual groundwater losses, from basins losing water at daily time step, F_L [mm/y]	77	74	77	72	75	74	73	72
Median annual groundwater gains, from basins gaining water at daily time step, F_G [mm/y]	70	60	64	77	79	80	78	78
Median annual net losses (I+AE+F_L-F_G) [mm/y]	602	602	604	603	603	602	602	604

TABLE 5.7 – Summary of the annual averages of the cumulated internal fluxes modelled by the GR4-I2 model at different time steps.

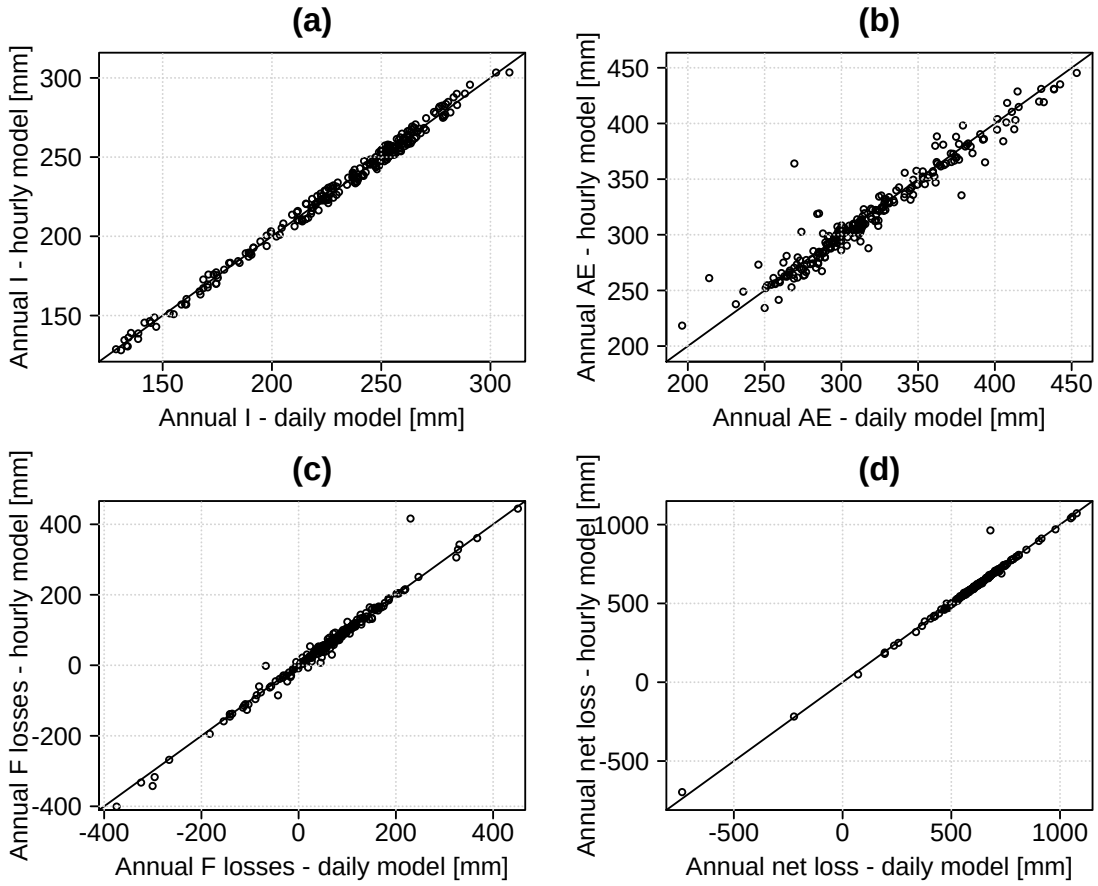


FIGURE 5.6 – Annual average cumulated fluxes from daily and hourly GR4-I2 model simulations over the 240-catchment set: (a) interception loss, *I*; (b) actual evaporation from the production reservoir, *AE*; (c) actual groundwater losses, (negative values represent gains), *F*; (d) net losses (=*I*+*AE*+*F*).

5.2.5.2 Evaluation of the temporal consistency of fluxes over the flood events only

Figure 5.7 shows the relative changes of the fluxes cumulated over the selected flood events provided by the simulations of the GR4-I2 model at the eight tested time steps. As in the case of the evaluation over the whole validation period, also over the flood events, there is a large improvement of the fluxes consistency with respect to the baseline model. In fact, the median cumulative flux ratios at 6-min time step are 1.14, 0.64, 1.16 and 1.34 for interception loss, actual evaporation, groundwater losses and gains respectively. These values are much closer to one than for the case of the GR4 baseline model, for which they were about 0.3, 2.85, 1.9, and 0.54 respectively (see Figure 4.7).

The fact that the interception flux over flood events is not stabilized could be a surprise at first, given the methodology followed to build the GR4-I2 model. However, this is likely to be due to the fact that we have not taken into account the initial and final conditions of the interception store (whether it is empty or not) at the beginning and at the end of each flood event. In fact, the interception flux is calculated over the same actual duration of the flood events for each time step. However, in this specific evaluation, the use of a same temporal window is not the most correct choice to assess a temporal coherence of the flux, because the

interception store used at sub-daily time steps could provide more water to evaporate, coming from the time steps before the considered storm event. Also, it could store some water that will be evaporated the day after the end of the selected flood event period. We decided not to further explore the issue, because the change in the interception flux is relatively low and is not important in absolute terms. In fact, by looking at the average daily fluxes of interception loss and actual evaporation from the production reservoir (see Table 5.8), one may observe that the average change in interception loss over floods is at most 0.1 mm/d and it is perfectly compensated by the actual evaporation changes (see last line in Table 5.8). Since this compensation is confined to the evaporative losses, this means that we have solved the problematic effects chain affecting the groundwater losses that was detected in the GR4 model.

Still, the groundwater losses significantly increase on average of about 16 % (by passing from daily to hourly or sub-hourly time steps) also in the new GR4-I2 model, although they no longer have to compensate for spurious changes of interception loss volumes. The largest part of the changes of groundwater losses across time steps has been reduced (they increased by about 90 % in the baseline model) by adding the interception store. The residual problem of inconsistency of the exchange fluxes in the GR4-I2 model is likely due to a structural inconsistency of the exchange function itself. An indication of the same kind is also found by looking at the evolution of the groundwater gains (Figure 5.7(d)). In fact, on average, we remind that in the GR4 model they decreased with time step to compensate for the reduction of interception losses. Now, in the GR4-I2 model, they increase without the need to compensate for other fluxes changes.

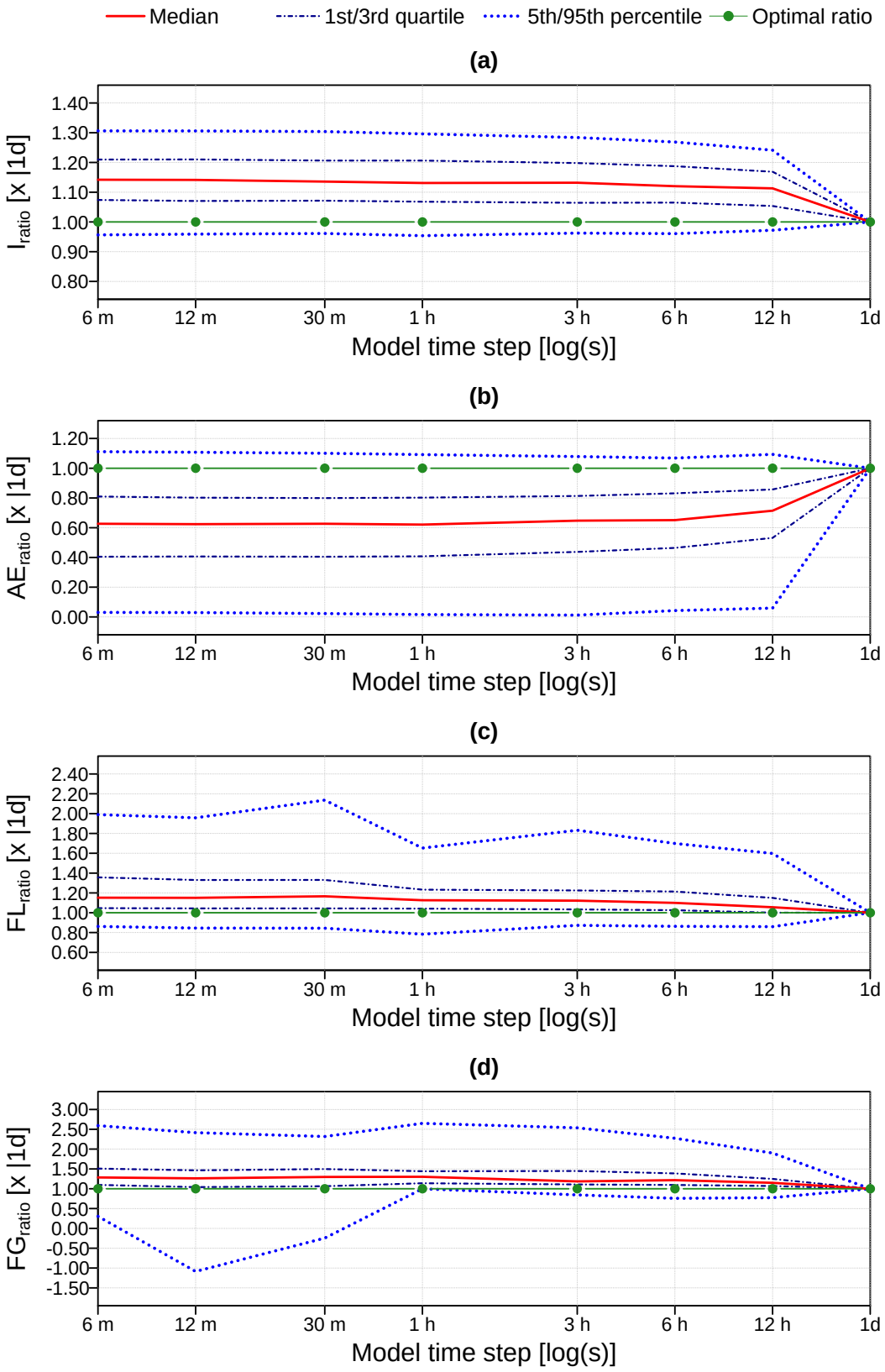


FIGURE 5.7 – Summary of the cumulative flux ratios at different time steps (with daily reference) over the 2400 flood events for the 240-catchment set: (a) interception loss, *I*; (b) actual evaporation from the production reservoir, *AE*; (c) groundwater losses, *FL*; (d) groundwater gains, *FG*.

Table 5.8 shows that the median daily net losses still present an increase at shorter time steps in the GR4-I2 model of about 8 % (it is almost 4 times less than in the baseline model). This increase in total net losses is now essentially due to the underground exchange function itself, given the trends observed in the modelled fluxes at different time steps and the perfect compensation of interception loss and actual evaporation from the production reservoir (Table 5.8). The scatter-plots in Figure 5.8 present the comparison of the daily average exchange fluxes and the total net losses calculated from hourly and daily simulations, confirming the trend observed in the average values. If the simulation bias is still worse at shorter time steps than at daily, this increase of exchange losses is likely to be a potentially undesired structural inadequacy of the model. In the next section we will evaluate this possible residual impact on the GR4-I2 model performance, especially on the bias of flood events simulations at shorter time steps.

Daily cumulated flux statistics [mm/d] over the 2400 selected flood events	**Model time step**							
	6-m	**12-m**	**30-m**	**1-h**	**3-h**	**6-h**	**12-h**	**1-d**
Median daily interception loss, I [mm/d]	0.82	0.82	0.82	0.82	0.82	0.81	0.80	0.72
Median daily actual evaporation from production reservoir, AE [mm/d]	0.16	0.16	0.16	0.16	0.16	0.17	0.19	0.25
Median daily groundwater losses, from basins losing water at daily time step, F_L [mm/d]	0.88	0.87	0.88	0.81	0.87	0.82	0.78	0.76
Median daily groundwater gains, from basins gaining water at daily time step, F_G [mm/d]	0.90	0.50	0.57	0.78	0.87	0.83	0.78	0.67
Median daily net losses (I+AE+F_L-F_G) [mm/d]	1.81	1.80	1.81	1.72	1.77	1.72	1.72	1.67
Median daily losses from interception + evaporation (I+AE) [mm/d]	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02

TABLE 5.8 – Summary of the daily average of the cumulated internal fluxes modelled by the GR4-I2 model at different time steps over the 2400 selected flood events.

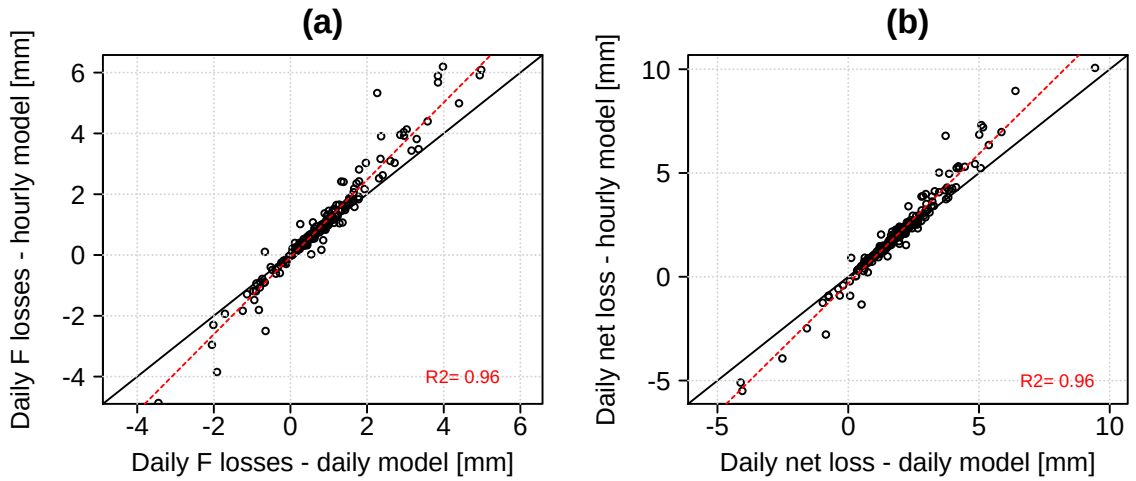


FIGURE 5.8 – Daily average cumulated fluxes from daily and hourly GR4-I2 model simulations over the 2400 selected flood events (with regression line): (a) actual groundwater losses, (negative values represent gains), *F*; (d) net losses (=*I*+*AE*+*F*).

5.2.5.3 Model parameters consistency of the GR4-I2 model

In Chapter 4, we have shown that, for the GR4 baseline model, the inconsistencies in model internal fluxes at different time steps result in inconsistent parameter values. In particular, the clearest impact of the fluxes inconsistencies on the parameters of the GR4 model is detected on the water exchange coefficient, x_2, which steadily decreases with time step (going towards larger negative values). Here, we expect that the spurious time-step dependency of x_2 is reduced, given the large improvement in the temporal consistency of the fluxes of the GR4-I2 model. Figure 5.9 shows the distributions of the deviations of the x_2 parameters of the GR4 baseline model (left panel) and the GR4-I2 model (right panel) calibrated at the eight time steps from the corresponding value obtained by calibration at the daily time step. For a consistent comparison, the parameters at all different time steps are converted to their corresponding hourly value ([mm/h]) by applying the theoretical relationship derived from the integration of the model governing equations, as already discussed in Chapters 3 and 4 (see also Le Moine, 2008, pp. 172-173).

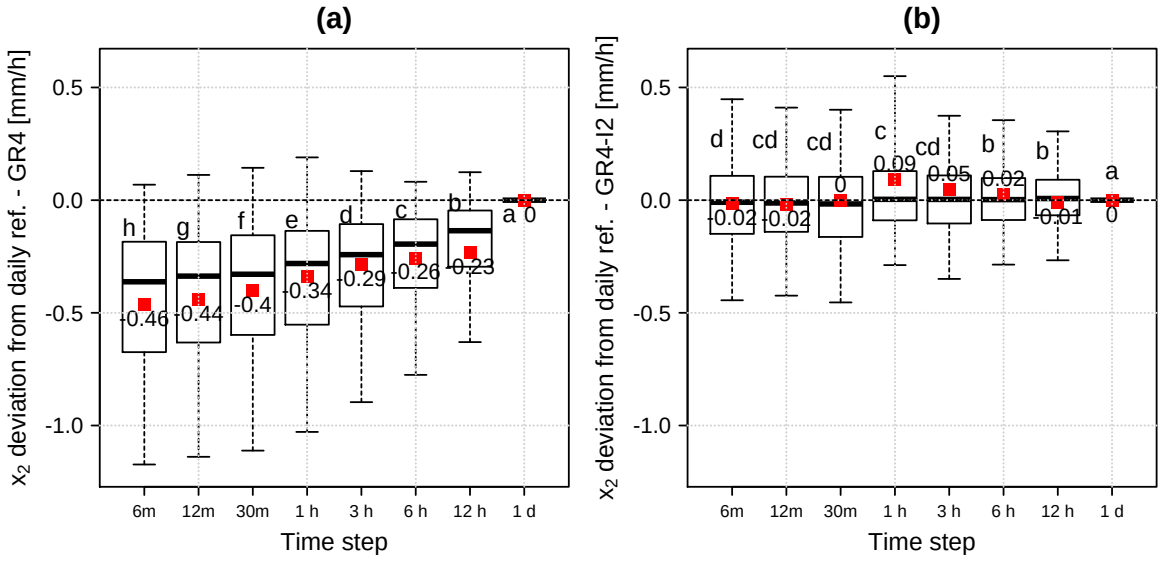


FIGURE 5.9 – Distributions of the deviations of the water exchange coefficient (x_2) parameter between the daily model reference and the eight time steps from 6 min to 1 day over the 240-catchment set (the parameters are converted to their corresponding hourly values by their theoretical relationships): (a) GR4 baseline model; (b) GR4-I2 with interception store of catchment-dependent capacity to ensure the flux temporal consistency ($I_{opt\text{-}flux}$). The box plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red points refer to mean values. The letters above each box plot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letter are not significantly different).

Figure 5.9 shows that the water exchange coefficient parameter is stabilized thanks to the insertion of the interception store. The decreasing trend that was observed for GR4 (left panel) has disappeared for GR4-I2 (right panel). For the new model version, the average deviation of the parameter is centred at around zero (its optimal value) at all time steps. Some cases of large relative deviations of the parameter at different time steps are still found over the catchment set, but no particular trends are detected.

Table 5.9 reports some statistics useful to better understand the positive effect of the interception store on the temporal coherence of all the GR4-I2 parameters. The median values of the parameters at the two extreme time steps, i.e. 1 day and 6 minutes, are much more stable than what was observed for GR4 (see Table 4.5). Note that the relative changes of the first three parameters at different time steps are now much lower than the relative standard deviation of the parameters over the catchment set.

The spurious trends that were observed in the GR4 parameters dependency with time step (especially for x_1 and x_2) have been drastically reduced after refining the interception component. This finding corroborates the results presented by Kavetski et al. (2011), who showed that some parameters that are common to different model structures of increasing complexity are highly scale dependent in simpler models but become progressively more stable in more complex model structures.

Parameter	Units	Median parameter (*converted to 1-h ref. values*)		Relative standard deviation $\left\|\frac{\sigma}{\mu}\right\|$ [%]		Median relative change (Δ_{rel}) of parameter calibrated at time step Δt, with respect to the daily reference $\Delta_{rel} = 100 \cdot [x_i(\Delta t) - x_i(1d)]/\|x_i(1d)\|$ [%] where $x_i(\Delta t)$ is the *i-th* parameter calibrated at t.s. Δt						
		1 d	6 m	1 d	6 m	6 m	12 m	30 m	1 h	3 h	6 h	12 h
x_1	[mm]	264	244	181	90	-3	-4	-4	-4	-2	-2	-2
x_2	[mm/h]	-0.58	-0.63	282	227	-2	-3	-3	-3	+1	+1	+3
x_3	[mm]	114	122	153	164	+4	+4	+5	+5	+3	+3	+2
x_4	[h]	48	12	50	110	-72	-72	-71	-71	-63	-63	-36

TABLE 5.9 – Median values of the GR4-I2 parameters calibrated at daily and 6-min time step, their coefficient of variation, and the median relative changes in parameter values at seven time steps from 6-min to 12-h with respect to the daily value over the 240-catchment set.

5.2.5.4 Evaluation of model performance of the GR4-I2 model

Table 5.10 reports the median values of the criteria of model performance over the whole validation period and over flood events for the hourly simulation of the GR4-I2 model with the capacity $I_{opt\text{-}flux}$ (ensuring the temporal consistency of the model fluxes). For comparison, the same statistics are reported for the GR4 baseline model and for the GR5-I2 model.

Hourly model	Median criteria on the whole series (validation)				Median FDC-based criteria (validation)			Median criteria on flood events (validation)			
	KGE [-]	a [-]	b [-]	r [-]	$\frac{Q99,sim}{Q99,obs}$ [-]	Slope-bias FDC [%]	$\frac{Q20,sim}{Q20,obs}$ [-]	KGE [-]	a [-]	b [-]	r [-]
GR4	0.820	0.989	1.009	0.897	0.972	1.394	0.952	0.727	0.969	0.904	0.832
GR4-I2 ($I_{opt\text{-}flux}$)	0.827	0.991	1.008	0.905	0.979	4.526	0.887	0.742	0.979	0.920	0.836
GR5-I2	0.829	0.996	0.997	0.911	0.980	7.499	0.837	0.752	0.976	0.934	0.834

TABLE 5.10 – Summary of the median performance criteria of the hourly GR4 (baseline), GR4-I2 and GR5-I2 models. The interception capacity in the GR4-I2 model is the one fixed to ensure the flux temporal consistency ($I_{opt\text{-}flux}$), while in the GR5-I2 model it is calibrated using the KGE criterion on the whole series as objective function.

Comparison with the baseline model

The GR4-I2 model with $I_{opt\text{-}flux}$ capacity outperforms the baseline model for the criteria over the whole period (i.e. regime) and more significantly over flood events, as confirmed also by the Friedman test (see also Figure 5.11 that is reported later for comparison with also another

structure). On the other hand, a decrease of performance in low-flow conditions is detected by a marked degradation of the ratio of the lower quantiles of the flow duration curve (Table 5.10). Low flows appear to be more under-estimated by the new GR4-I2 model with interception store than they were by the baseline model. This consequently increases the bias in the slope of the mid-segment of the flow duration curve.

Comparison with the GR5-I2 model

First, we remind that the GR5-I2 model (with the calibrated interception capacity) is a sub-optimal choice with respect to GR4-I2 (with the $I_{opt\text{-}flux}$ capacity) because GR5-I2 does not ensure the model fluxes temporal consistency as well as GR4-I2 does. However, it is interesting to see if some performance benefits could be provided by calibrating the interception store (GR5-I2), and understand the reasons of the performance changes.

With respect to the GR5-I2 model, the GR4-I2 model with $I_{opt\text{-}flux}$ capacity shows a slight decrease of performance over flood events (not significant according to the Friedman test). However, the GR5-I2 model degrades more the simulation of lows flows than GR4-I2 does with respect to GR4. One may note that the performance changes between GR5-I2 and GR4-I2 are essentially related to the following differences in the interception and exchange components:

(i) the average interception store capacity is about 6 mm for GR5-I2 and 2 mm for GR4-I2;

(ii) the change of interception capacity has a direct impact on the exchange flux and its related parameter (see Section 5.2.2; Figure 5.3(b)): when the interception capacity increases, the absolute value of the exchange coefficient decreases.

Note that the changes in the exchange have a direct impact on the simulation bias, while the changes in the interception flux are almost perfectly compensated by the feedback of the actual evaporation from the production store (as we proved by our previous analyses).

So a logical deduction from the observations above could explain the observed changes of performance. The further improvement of the floods volume simulation of GR5-I2 with respect to GR4-I2 should be linked to the reduced importance of the exchange due to the reduction of the water exchange coefficient. In fact we remind that for most of the catchments (about 80 % of the set) the exchange is negative (losses). So the reduction of the absolute value of the exchange coefficient in GR5-I2 is likely to be the cause of the improvement of the ratio-of-means (*b*) criterion over floods. For this reason, we think that further tests should attempt at recovering the gains in performance of the GR5-I2 model over flood events by changing the exchange function shape to reduce the exchange flux in high-flow conditions with respect to regime.

Decision on the acceptance of the GR4-I2 model structure

The GR4-I2 model structure with $I_{opt\text{-}flux}$ capacity (ensuring the fluxes temporal consistency) is accepted as an improvement of the current GR4 model at sub-daily time

steps. In fact, it significantly improves the model adaptability at different time steps thanks to the internal fluxes stabilization and its positive consequences (e.g. improved transferability of parameters). This model seems to be better, in this provisional version (as it has been proposed so far), for applications of the model focusing on regime or high-flow conditions rather than low flows. In these conditions, our modification of the model significantly improves the fluxes consistency and provides significant benefits in performance with respect to the baseline model. Further tests should focus on possible improvements in low-flow simulations. These could be achieved in different ways, such as:

(i) modifying the interception function itself for low incoming precipitation rates;

(ii) modifying the exchange function shape with particular attention to improving low-flow simulations;

(iii) testing the insertion of a precipitation correction factor.

Here we decided to terminate our tests on the refinement of the interception function, because its introduction has already proved many benefits and it seems now more urgent to work on other model components. In the rest of this chapter, we will focus more on the exchange function providing evidence of its importance for improving model performance on low flows.

Further research for improving the interception component

We report here some ideas on further possible ways of improvement of our simple interception model that could help improving low-flow simulation. These ideas are inspired by the well-established Rutter model (Rutter et al., 1971), one of the most often used conceptual models for interception at sub-daily time scales (Valente et al., 1997; Gerrits, 2010). The two following characteristics of the Rutter model could be combined with our bucket interception store:

(i) a slow leaking function from the interception store, that allows draining a small portion of intercepted water to the production reservoir;

(ii) the addition of a free throughfall branch, i.e. throughfall which does not touch the interception store at all (i.e. a fixed portion of the precipitation *P*).

These modifications should reduce the evaporation from the interception store when it is filled by low precipitation rates, but they could not be effective for improving low-flow simulation because of the compensative feedback of the evaporation from the production store. Thus, we suspect that these changes would be of second order with respect to the modification of the exchange function to attempt reducing the impact on low flows.

5.2.5.5 Relationship between interception store capacity and catchment characteristics

We performed a first rough analysis of the possible relationship between the interception parameter and some catchment characteristics. A proper correlation analysis was not possible

because the interception capacity values have a very low dispersion, being highly concentrated around a median value and with only a few values around. For example, at the hourly time step, there are only five capacity values for the 240 catchments, ranging from 1.5 to 2.5 mm and their standard deviation is 0.22 mm. This low dispersion is likely to be (at least partially) due to the choice of using a search step of 0.25 mm for screening the parameter space in our optimization procedure. However, the impact of this discrete search step on the cumulated flux is not large, as already discussed. So we think that a first analysis of the parameter dependency on catchment characteristics is still possible.

The analysis was performed at the hourly time step, for which the capacities are stabilized at their values corresponding to short sub-hourly time steps. The indicators of catchment characteristics that we considered are the following twelve: surface, altitude, topographic index, mean annual precipitation and potential evapotranspiration, aridity index, runoff coefficient, daily precipitation intensity index, GOUE index of precipitation temporal variability (at 6-min), percentage of broad-leaved, coniferous and mixed forest covers (see Chapter 2).

By grouping the 240 catchments in five classes according to their interception capacity parameter, we searched for possible clear monotonous relationships between the capacity values and the catchment characteristics. This trend was visually identified by means of box-plots if most of the quantiles of the distribution of the catchment characteristics changed in the same way (increasing or decreasing) by increasing interception capacity.

Among the 12 indicators considered, the clearest trend was found for the aridity index, as shown in Figure 5.10: the most arid the catchment is, the lower the interception capacity. This trend corresponds to the logical expectation of a decrease of the interception capacity for increasingly arid environments, where the vegetation cover should be sparser. This information could be useful to rapidly prescribe first approximate capacity values of the interception store of the GR4-I2 model in function of the aridity index of the catchment. However, our advice is to favor a more accurate method for ensuring the temporal consistency of the interception flux.

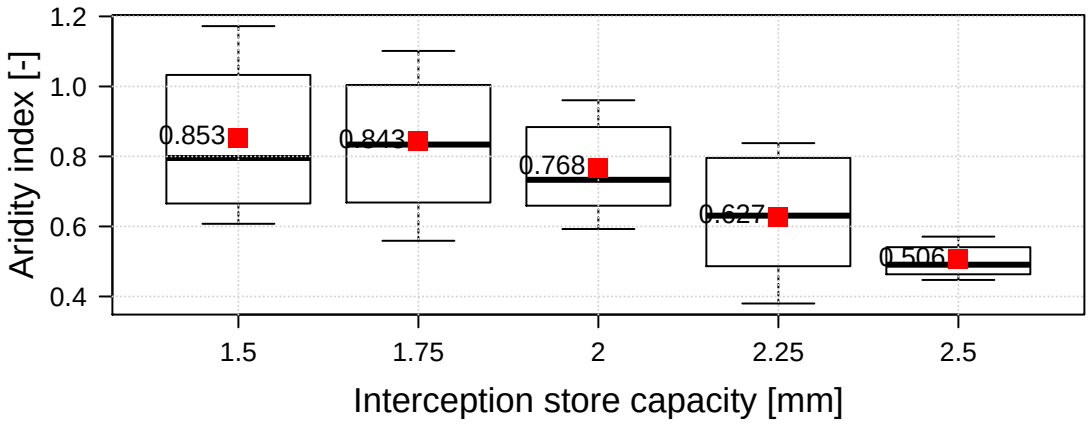


FIGURE 5.10 – Aridity index of the 240 catchments grouped by the interception store capacity of the GR4-I2 model at hourly time step. The values of the store capacity are those fixed to ensure the temporal consistency of the interception flux at different time steps.

Two other variables turned out to be seemingly correlated with the interception capacity parameter, but to a lower extent: the surface, and the GOUE index of precipitation temporal variability. Smaller catchments and catchments with higher precipitation variability are associated to higher interception capacities. The effect of the catchment size is likely to be due to the spatial averaging of this parameter, which should increasingly smooth out possible localized high values of storage capacities as the catchment surface increases.

Some limits of our analysis could *'hide'* the possible links of the model parameter with some expected explanatory variables, like the forest cover. In addition to the low number of pre-defined values of capacity values, another limit is the fact that we defined the capacities by the internal coherence of the model (with the daily modelled flux) and not by a physical approach.

5.3 Attempts at improving model performance by modification of the exchange function

As shown in our previous analyses of model fluxes across time steps, the underground losses are generally too large during flood events, at daily and even more at sub-daily time steps, for both the GR4-I2 and GR4 models. In particular, in the GR4-I2 model, the exchange flux is not subject to a compensation of interception loss changes, but it is still increasing significantly at shorter time steps and this worsens the flood volumes simulation bias. We attempted solving this problem by searching ways of reducing the groundwater losses during floods. To this end, we tested some variants of the ground-water exchange function designed to either saturate (i.e. come to a limit) or reduce the increase of exchanged fluxes in high-flow conditions or change its dependency on the routing store level.

5.3.1 Tested variants of the exchange function

The common thread of our tested modifications is that they are supposed to change the distribution of the exchange flux along different streamflow levels.

We did not question the **dependency of the exchange function on the routing reservoir level** itself. This starting point relies on previous works on the development of the GR4 model (Nascimento, 1995; Perrin, 2000; Le Moine, 2008) showing that the current solution outperforms several alternatives (such as the dependency from the production reservoir level). Moreover, since our scope is to change the distribution of the exchange for different levels of streamflows, the dependency on the routing store level is the most logical one. Note that we re-tested some previously tested functions for the exchange (Perrin, 2000; Le Moine, 2008), as changing the exponent of the power function, because our evaluation framework is different from the previous works cited. In fact, the added value of our evaluation stems from the attention paid to the internal fluxes of the model and to the flood events simulation.

The following four classes of exchange function were tested:

1. As a reference comparison, a simplified version of the baseline model **without exchange function at all (*F=0*)** has been tested. This model version has been called ***GR3-F0*** (for zero-exchange flux, *F*) and has only **3 free parameters**, since the exchange coefficient is not used. Of course, this model is not expected to improve model performance, given the importance of the exchange flux to allow a correct catchment water balance. In fact, in regime conditions, the current exchange function of the GR4 model seems essential and effective (as shown by the analysis of the fluxes over the whole series). The aims of this test are: (i) to show the possible lower-bound of model performance when the exchange component is neglected, and (ii) especially to show the impact of a null exchange flux on the bias over flood events.

2. We tested the exchange function of the model developed by Le Moine (2008) starting from the GR4 baseline model. This exchange function was identified by Le Moine (2008)

as a valid alternative to the GR4 model's one, particularly for improving the simulation of low flows. This originated a well-established model version derived from the GR4 baseline model at the daily and hourly time step that has been called GR5 (5 free parameters). This version of the model is used as the elementary block of the semi-distributed GR-SD model (e.g. Lobligeois et al., 2014). The exchange function of this model is a **linear function of the routing reservoir level**. Its formulation includes an **additional free parameter** (x_5 [-]), representing the **level of the routing store for which the exchanged flux changes sign**. As proposed by Le Moine (2008), this linear exchange function is:

$$F = x_2 \left(\frac{R}{x_3} - x_5 \right) \tag{5.2}$$

where: R is the level in the routing store [mm], x_3 its reference capacity (one time step ahead) [mm], x_5 the threshold level for which the exchange function changes in sign (with $x_5 \in [0, 1]$), and x_2 the water exchange coefficient [mm/t.s.].

In the following we will call this structure ***GR5-FL*** (where *L* stands for linear function for exchange, *F*) to better distinguish it from other models names. This linear exchange function has been tested also with the interception store of the GR4-I2 formulation, leading to a *second-order modification* called ***GR5-I2-FL.***

3. The first new exchange function variant that we propose involves the only **modification of the exponent *(p)* of the power function** (the current exponent is 3.5) of the level of the routing reservoir. The *p*-power exchange function is:

$$F = x_2 \left(\frac{R}{x_3} \right)^p \tag{5.3}$$

where: R is the level in the routing store [mm], x_3 its reference capacity [mm], x_2 the water exchange coefficient [mm/t.s.], and p a positive exponent. The sensitivity of model performance to different values of p has been tested. We tested values of p lower and greater than 3.5 (ranging from 2 to 5). Our expectation is that the exchanges in high-flow conditions are reduced by increasing the exponent of the power function, because F is calculated by raising a term that is always lower than 1 (i.e. $\frac{R}{x_3}$) to power p. However, also values of p lower than 3.5 have been tested because of the interaction of the change in p values with the calibrated values of x_2. This structure is called ***GR4-Fp-X***, where *X* is the exponent of the power function.

4. One may note that by raising the exponent of the power function, exchanges are overall reduced whatever the routing reservoir level is. On the contrary, our previous analyses on the modelled fluxes seem to indicate that the exchanges should be reduced only in high-flows conditions. For this reason, we tested some exchange functions saturating as the routing reservoir filling rate approaches 1, in order to reduce the losses only when the routing reservoir is fuller.

We tested several functions based on **hyperbolic tangent** and sigmoid functions of the routing reservoir level (or of its power functions). Some preliminary tests with slightly different arguments of these functions (i.e. with different power functions of the routing reservoir level, as argument) led us to choose the following formulation:

$$F = x_2 tanh\left(5\frac{R}{x_3}\right)^4 \tag{5.4}$$

The model derived from the baseline model with the exchange defined by using the hyperbolic tangent function of the routing reservoir level is called ***GR4-Fth*.** This function has been tested also with the interception store of the GR4-I2 formulation, building a *second-order modification* called ***GR4-I2-Fth.***

Our tests for the improvement of the exchange function are obviously not exhaustive. For example, other variants of the exchange functions above and other combinations with the GR4-I2 model structure with interception store could have been tested. However, as already mentioned, our study follows other previous works for the empirical development of the model that have tested other tens of variants of the exchange functions.

5.3.2 Synthesis of the results with different exchange functions

Table 5.11 shows the average performance scores over the catchment set for seven representative implementations of the model variants described in the previous section at the hourly time step. The performance scores of the GR4 baseline and the GR4-I2 models are also reported to ease comparison. The model performance is evaluated by different criteria calculated over the whole time series, the FDC and the flood events.

The average values of the criteria over the catchment set (see Table 5.11) show that:

1. The GR3-F0 model without exchange function at all presents the most significant loss of performance for all aspects (regime, low and high flows). In particular, it leads to a large overestimation of the mean long-term streamflow, and of the low flows. This proves that the exchange function of the GR4 model is essential to provide accurate water-balance simulations. This finding corroborates the results presented by Le Moine (2008) who showed that it is better to include the exchange function in the model, also rather than other surrogate solutions (as input correction factors). Also, it is important to note that even without exchange losses at all, the flood volumes underestimation is significantly degraded with respect to GR4.

2. The GR5-FL model, with the linear exchange function proposed by Le Moine (2008), provides, in general, better average scores (for the different criteria) than all other tested options except the ones derived from GR4-I2. The use of a linear exchange function (with a fifth additional free parameter) provides a marked improvement with respect to GR4 for both low flows and flood events simulation. The average scores of GR5-FL over the whole series and over flood events are slightly lower than for the GR4-I2-Fth model, but

GR5-FL is better than the latter for low flows. For this reason, the combination of the insertion of an interception store (GR4-I2) with the use of the linear exchange function of GR5-FL seems promising, because the two modifications are advantageous for different aspects. Indeed, this expectation is met: **the GR5-I2-FL model (with interception store and linear exchange function with change in sign) provides the best average scores over all conditions (regime, low and high flows).** However, this gain of performance is achieved by increasing the complexity of the model by adding a fifth free parameter.

3. The performance of the GR4-Fp models (exchange function with different exponents of the power function) is not satisfactory. For exponents lower than 3.5, even if the flood bias is improved, a degradation of model performance is observed over the whole validation period and more significantly over low flows. For higher exponents, performance is only improved over low flows but is significantly degraded over the whole series and flood events.

4. The GR4-Fth model derived from the baseline by saturating the exchange flux in high-flows (by the hyperbolic tangent function) performs well in general, but highly underestimates low flows. The floods volume bias is reduced, but less than by the GR5-FL and GR4-I2 models which provide also better results in general.

Figure 5.11 shows the distribution of the performance scores over flood events (at the hourly time step) for the GR4 baseline model, the GR4-I2 and GR5-I2-FL model with interception store and linear exchange function. It shows that:

(i) The performance improvements provided by both the GR4-I2 and GR5-I2-FL structures are statistically significant for the KGE, relative variability and ratio-of-means criteria over floods.

(ii) The GR5-I2-FL model leads to an additional significant improvement in the ratio-of-means criterion over floods also with respect to the GR4-I2 model.

Hourly model	Median criteria on the whole series (validation)				Median FDC-based criteria (validation)			Median criteria on flood events (validation)			
	KGE [-]	a [-]	b [-]	r [-]	$\frac{Q99,sim}{Q99,obs}$ [-]	Slope-bias FDC [%]	$\frac{Q20,sim}{Q20,obs}$ [-]	KGE [-]	a [-]	b [-]	r [-]
GR4	0.820	0.989	1.009	0.897	0.972	1.39	0.952	0.727	0.969	0.904	0.832
GR4-I2 ($I_{opt\text{-}flux}$)	0.827	0.991	1.008	0.905	0.979	4.526	0.887	0.742	0.979	0.920	0.836
GR3-F0	0.658	0.960	1.177	0.857	0.972	-21.26	1.671	0.668	0.868	0.866	0.795
GR4-Fp-2	0.815	0.997	0.997	0.894	0.976	11.01	0.795	0.707	0.955	0.923	0.813
GR4-Fp-5	0.809	0.977	1.019	0.882	0.973	-1.37	1.006	0.711	0.953	0.869	0.829
GR4-Fth	0.825	0.991	1.004	0.900	0.975	8.92	0.820	0.725	0.957	0.914	0.828
GR4-I2-Fth	0.830	0.989	1.007	0.904	0.977	4.46	0.894	0.743	0.977	0.921	0.834
GR5-FL	0.826	0.993	1.002	0.903	0.979	-2.35	1.023	0.737	1.018	0.924	0.829
GR5-I2-FL	0.831	0.998	1.002	0.909	0.991	-2.03	1.017	0.747	1.032	0.936	0.835

TABLE 5.11 – Summary of the median performance criteria of different hourly model structures derived by modification of the exchange function in the GR4 baseline model and in the new GR4-I2 modified version with interception store.

Note that all the model solutions presented above that are derived from GR4 only by changing the exchange function (without interception store) are a sub-optimal choice because of the modelled fluxes inconsistency across time steps. In fact, in all these models (GR3-F0, GR4-Fp-X, GR4-Fth, GR5-FL), the interception process is modelled by the same neutralisation function as in the GR4 baseline model. So the interception flux is affected by the same temporal inconsistencies that affect also all the sub-sequent model fluxes, as in the GR4 baseline model. However, we tested these models to show the changes in performance given only by the first-order modification of the exchange (independently from interception).

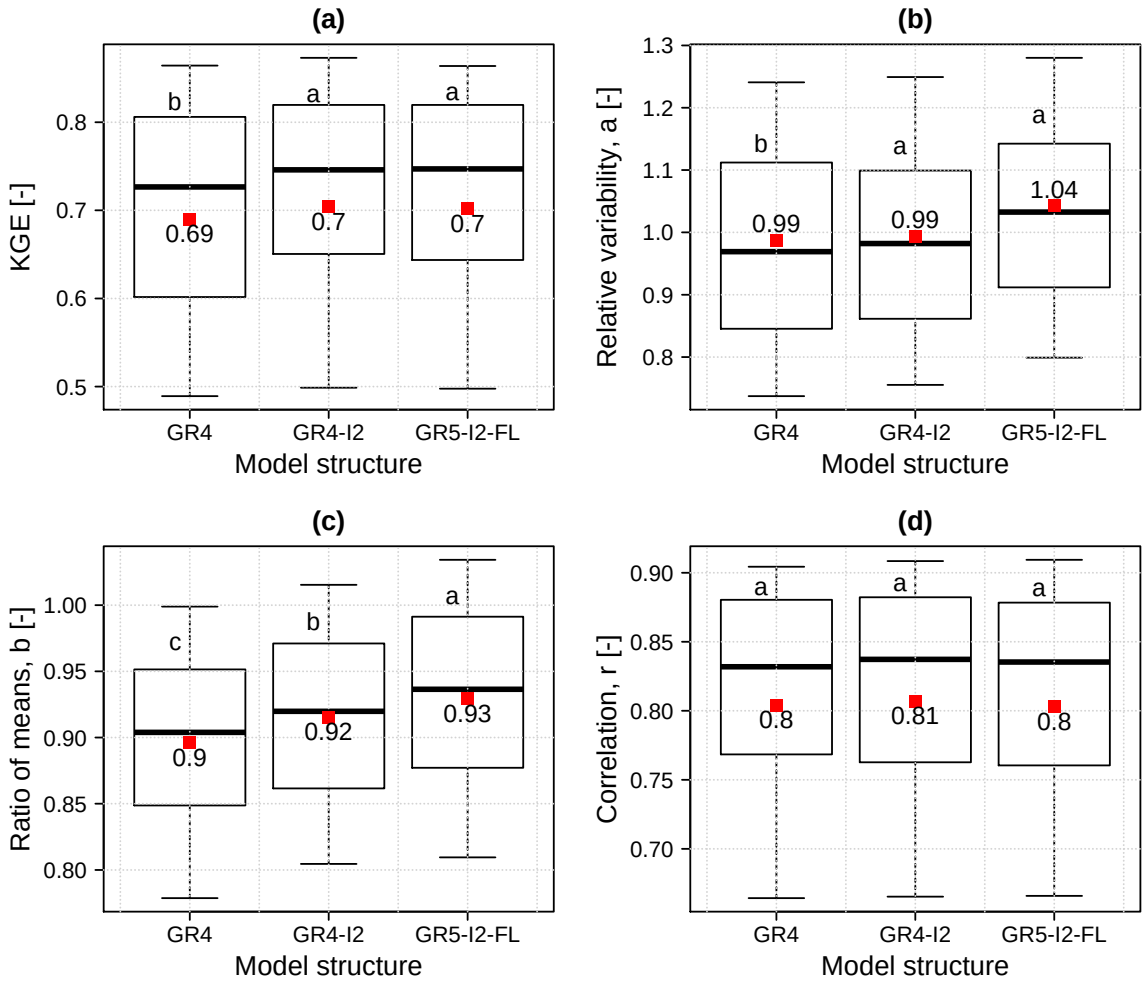


FIGURE 5.11 – Distribution over the catchment set of the performance criteria over the selected flood events for three models run at the hourly time step: GR4, GR4-I2 (with interception store) and GR5-I2-FL (with interception store and linear exchange function): (a) *KGE*, (b) relative variability, *a*; (c) ratio of means, *b*; and (d) correlation, *r.* The box plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red points refer to mean values. The letters above each box plot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letter are not significantly different).

Finally, the findings of this section lead us to **retain the GR5-I2-FL model with the interception store and the linear exchange function at sub-daily time steps.** For sake of brevity, the model could be renamed as *GR5-I* (with 5 parameters and interception store). This model ensures large improvements of the temporal consistency of the modelled fluxes (as shown for its parent model *GR4-I2*) while also significantly improving model performance in all conditions (regime, low and high flows). The *GR5-I* model equations are further presented in Appendix G.

5.4 Attempts at taking into account rainfall intensities for improving the production function

As we have highlighted in Chapter 1, it is generally recognized in the literature that the rainfall intensity is a key driver for runoff generation. This is particularly important in rainfall-runoff modelling at short time steps, as hourly and sub-hourly, for the possibility of considering the actual variability of rainfall intensities.

In the GR4 and GR4-I2 models, rainfall intensity is considered in the interception function and in the infiltration to the production reservoir. However, the latter function is not '*rate-limited*', i.e. there is no threshold limiting infiltration for high values of precipitation. We identified two other different ways to take into account rainfall intensity.

First, a *rate-limited* function can be introduced to represent the infiltration-excess runoff process. Several authors argued that this is a key mechanism of runoff generation that acts at short time durations (particularly at hourly and sub-hourly scales) as already reported in Chapter 1. A second way of considering the precipitation intensity can be the use of a correction factor depending on the precipitation intensity. A few implementations of these two additional components of the model were tested and they are briefly discussed in the following sections.

5.4.1 Infiltration-excess runoff based on precipitation intensity

We designed and tested some mathematical functions for representing an infiltration-excess runoff mechanism, by a simplification of the perceptual model of the so-called Horton overland flow. Following this concept, we decided to test the introduction of a rate-limited function in the model that acts immediately after interception to generate a direct runoff component for high values of precipitation. The function is activated only for precipitation intensities higher than a threshold, with the aim of increasing runoff in these conditions. Conversely, low precipitation rates may still entirely infiltrate in the production store. We tested the addition of this function in both the GR4 baseline model and its derived version with interception store (GR4-I2). In the GR4 baseline model, the infiltration-excess runoff component acts on the net precipitation (*Pn*, given by the neutralisation function). Analogously, its insertion in the GR4-I2 model is located on the throughfall (*Pth*) branch, after interception, as shown in Figure 5.12.

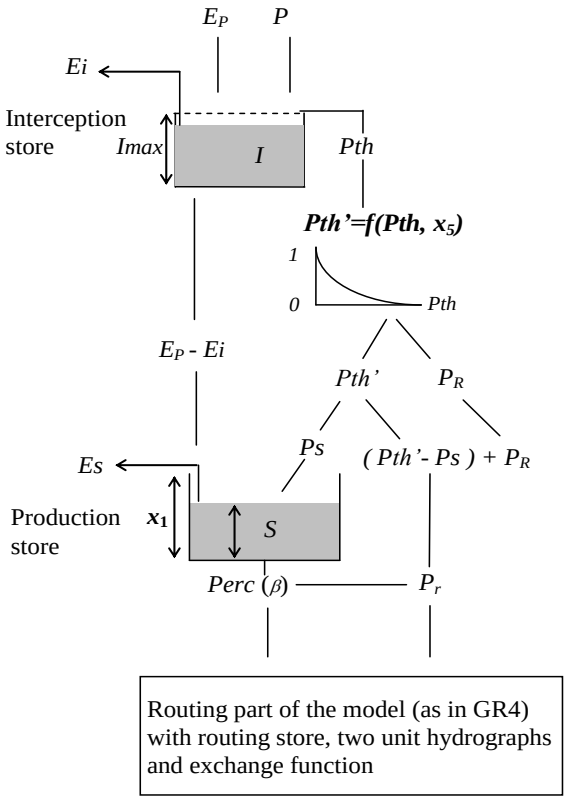


FIGURE 5.12 – Schematic representation of the production part of the GR5-I-PR model structure, derived from the GR4-I2 model (baseline model by Perrin et al. (2003) with interception store), by addition of an infiltration-excess runoff function, *f(Pth, x_5)*, depending on the intensity of net precipitation (i.e. throughfall, *Pth*).

This infiltration-excess runoff function takes as input the net precipitation (throughfall) and separates it in two components:

- a part of the throughfall *Pth'* may potentially infiltrate in the production store; the part *Ps* (of *Pth')* that actually fills the store is determined as in the GR4 baseline model (in function of the store filling rate).
- the remaining part, named P_R, contributes to the direct runoff component by-passing the production store.

The following two alternative functions were tested to operate this separation:

$$P_R = P_{th} \cdot \left(1 - e^{-x_5 \cdot {P_{th}}^{\gamma}}\right) \tag{5.5}$$

$$P_R = P_{th} \cdot \left(1 - \frac{1}{\left(1 + (P_{th} x_5)^{\gamma}\right)^{\frac{1}{\gamma}}}\right) \tag{5.6}$$

where, for both the alternative functions: γ is a positive fixed parameter [-], and x_5 [-] is a free parameter representing the speed of the yield rate of the infiltration-excess runoff (P_R/Pth) in approaching the unity. The variation range of the free parameter has been fixed to $[0, +10]$. After some first sensitivity tests, we have fixed a value of 3.0 and 5.0 for γ in equations (5.5) and (5.6) respectively. The main difference between the two functions is that one (5.6) increases more gradually as the precipitation intensity increases than the other.

The two alternative functions above (Eq. (5.5) or (5.6)) are inserted in the production part of the model, as represented in Figure 5.12. As for the new model names, when the infiltration-excess function is added in the GR4 baseline model, the new model structures are called *GR5-PRexp* and *GR5-PRpf* respectively for the exponential function (Eq. (5.5)) and the power function (Eq. (5.6)). Analogously, when inserted in the GR4-I2 model, the new structures codes are *GR5-I2-PRexp* and *GR5-I2-PRpf*.

Table 5.12 reports the average performance scores of the new model structures obtained by adding the infiltration-excess runoff component to either the GR4 baseline model or the GR4-I2 model with interception store. It shows that for all the criteria the new model structures lead to similar levels of average performance as their *parent model structure* (GR4 or GR4-I2) without any significant change. So it seems that this complexification of the model is not effective for improving the simulation of flood events, as it could be expected. The best performance scores over floods is provided by the GR5-I2-PRpf model but the changes in KGE and components are lower than 10^{-2} and can be considered negligible.

Note that these tests intended to be only a first exploration of this issue and do not provide a comprehensive view. Future work on the model development should further investigate this issue by testing other possible implementations.

Hourly model	Median criteria on the whole series (validation)				Median FDC-based criteria (validation)			Median criteria on flood events (validation)			
	KGE [-]	a [-]	b [-]	r [-]	$\frac{Q99,sim}{Q99,obs}$ [-]	Slope-bias FDC [%]	$\frac{Q20,sim}{Q20,obs}$ [-]	KGE [-]	a [-]	b [-]	r [-]
GR4	0.820	0.989	1.009	0.897	0.972	1.39	0.952	0.727	0.969	0.904	0.832
GR4-I2 ($I_{opt\text{-}flux}$)	0.827	0.991	1.008	0.905	0.979	4.53	0.887	0.742	0.979	0.920	0.836
GR5-PRpf	0.822	0.989	1.010	0.897	0.971	-0.11	0.976	0.728	0.974	0.910	0.838
GR5-PRexp	0.820	0.986	1.008	0.897	0.970	0.71	0.963	0.729	0.974	0.904	0.835
GR5-I2-PRpf	0.826	0.994	1.007	0.906	0.984	2.12	0.930	0.749	0.985	0.926	0.841
GR5-I2-PRexp	0.828	0.995	1.007	0.904	0.983	4.62	0.909	0.744	0.979	0.921	0.831

TABLE 5.12 – Summary of the median performance criteria of different hourly model structures derived by insertion of an infiltration-excess runoff function in the GR4 baseline model and in the GR4-I2 model with interception store.

5.4.2 Precipitation intensity-based correction

The introduction of a correction factor for precipitation has been considered only for reference, as done by Perrin (2000) at the daily time step. In fact, we agree with Perrin (2000) in stating that the use of such correction factors is tricky and ambiguous because it is halfway between data processing and hydrological modelling.

The correction factor that we tested is a multiplicative coefficient applied to precipitation. We defined it as a function depending on precipitation intensity that decreases as the precipitation increases. The tested function was defined by the following equation:

$$P_C = P \cdot (1 + x_5 e^{-\gamma P}) \tag{5.7}$$

where: x_5 is a free parameter determining the correction factor magnitude for small values of P [-]; and γ is a fixed parameter determining the 'rapidity' of the correction factor in approaching the unity as the precipitation increases [-]. The free parameter's domain is fixed to $[-1, +10]$. The fixed parameter γ has been fixed to 1.5, after first sensitivity tests (by calibrating it and fixing it at its median value).

The precipitation correction factor has been added to the baseline model either with exchange (in GR4) or without exchange function (in GR3-F0) to analyze the contribution of only the correction function without its interaction with the exchange. This function did not bring any

significant improvement with respect to the baseline version of the model. When added to the model without exchange it provides slightly lower average scores than the GR4 baseline model. We do not report the statistics of its performance scores since these tests did not lead us to any other specific conclusion.

5.5 Synthesis

In this chapter, we have presented the empirical tests performed for improving the model performance and the internal coherence of the GR4 model at sub-daily time steps. As indicated by our previous analysis (see Chapters 3 and 4), a modification of the GR4 model structure at sub-daily time steps is needed to improve the consistency of the model fluxes at multiple time steps. This is expected to possibly improve also simulations accuracy, especially over flood events.

To this end, we have tested the insertion of an interception store in the production part of the GR4 model, as our previous analysis indicated the need of increasing the capacity of the model to evaporate intercepted water at sub-daily time steps. This structural modification has been implemented by a simple bucket-style interception store with either a fixed or calibrated maximum capacity. Two possible implementations of the interception component on a discrete time steps were tested (named GR4-I1 and GR4-I2), depending on the priority given to either throughfall or evaporation calculation. A first evaluation of this structural modification was based on model performance. It validated the benefit of adding the interception store in the GR4 model at sub-daily time steps. In fact, both implementations of the interception store with either a fixed or calibrated capacity led to a significant increase of model performance. **The flood volumes bias is significantly reduced by the insertion of the interception store in the baseline model**. No significant improvement of model performance was found by calibrating the interception parameter with respect to the use of a fixed value (in a range between 2 and 10 mm). Moreover, some difficulties in the calibration of this parameter were detected because of its interactions with the exchange function (and its calibrated parameter).

Then, we evaluated the benefit of the insertion of the interception store for improving the coherence of the internal fluxes at different time steps. For each catchment, we calculated the "*optimal*" interception capacity that allows ensuring the coherence of the cumulated interception flux at all time steps using the daily reference flux of the GR4 model. The relationship between this optimal interception capacity and the time step was shown to be clearer for one of the two implementations of the interception store (GR4-I2, with evaporation prior to throughfall at each time step). So, this implementation was preferred and tested at time steps ranging from 6-min to 1 day, with the catchment-dependent capacity ensuring the fluxes consistency. This test proved that **the stabilization of the interception flux as the time step changes leads to a significant improvement of the coherence of all the other model fluxes**. The consistency of the model fluxes leads to a consistency in model parameters. Also, model performance is significantly improved over regime and high-flow conditions. On the other hand, a decrease of performance for low-flows simulations is detected. To improve this aspect, we have focused on the modification of the exchange function.

Thus, we have presented the tests performed with a few different exchange functions. These tests also aimed at further improving model performance over flood events. Some new

functions are proposed to reduce the exchanged fluxes in high-flow conditions, but, contrary to our expectation, this did not lead to any other significant improvements for the floods bias. Also, we tested another existing alternative of the exchange function of GR4, i.e. the linear exchange function proposed by Le Moine (2008), with an additional free parameter permitting a seasonal change of sign in the exchange flux. The combination of this exchange function with the GR4-I2 structure proposed here is proved to be the most effective solution for further increasing model performance with respect to the baseline model, also on low flows. Then, this model structure, with the interception store and the linear exchange function, provides significant improvements of model performance at sub-daily time steps with respect to GR4 in a multi-criteria perspective (regime, floods and low flows).

Thus, **we recommend the new model structure** (renamed here as *GR5-I,* with an interception store and five free parameters; see Appendix G), **for the improvement of both model performance and internal coherence at sub-daily time steps.** This is a positive and promising conclusion because it means that ***ensuring the improvement of the temporal consistency of the model fluxes (with the interception store) represents an added value for model performance in any condition***.

In the end, we have briefly presented some attempts of introducing in the model a function taking into account the impact of rainfall intensity on flow production. These tests based on the introduction of either an infiltration-excess runoff component or a precipitation correction factor did not provide any significant improvement of model performance. However, our implementations of these functions are not exhaustive, and further work on the model development could focus on other variants or combinations of these structural modifications.

Conclusions

This PhD thesis focused on issues related to temporal scaling in hydrological modelling. This issue is linked to essential operational demands, as for example the need of adaptive models for multiple time steps, in flood forecasting systems. It involves also multiple inter-related scientific questions such as: the relationship between processes, the observation and modelling scales, and the dependency of model parameters and structures on time step. These issues, long overlooked in the hydrological literature, have only received attention very recently.

In this PhD thesis, we have explored the *adaptability* of the GR rainfall-runoff models for different daily and sub-daily time steps. The main objective was to identify an adaptive multi-time step model starting from the well-known GR4 model, initially developed at the daily time step, and popularly known as GR4J (see Perrin et al., 2003). To this end, we have first carried out a diagnosis to highlight the current structural inadequacies of the model. Then, we have based the modification of the model on the improvement of its internal consistency at different time steps.

Main contributions

All along this thesis, we have evaluated our modelling tests over a large set of 240 catchments and 2400 observed flood events, in order to obtain as general results as possible and to ensure a robust model diagnosis over a wide variety of hydro-climatic conditions.

This thesis brings a *methodological innovation* with the proposal of a new approach for empirical model identification based on two validation steps:

(i) the improvement of model performance;
(ii) the evaluation of fluxes consistency across time scales.

Impact of time step on model performance

The evaluation of model performance at multiple time steps was based on two levels: (a) the evaluation of the outputs aggregated at larger time scales, to allow comparisons of criteria across different model time steps; (b) the evaluation of the outputs at the model time step, for assessing performance at the model (short) temporal scales. The distinction of these two levels is necessary when dealing with a multi-time step model.

The use of refined modelling time steps (hourly and sub-hourly) is obviously needed to estimate timing and intensity of short-duration events (flash floods), but one open question is whether it could contribute also to improve the simulation of events, with different dynamics and durations, in terms of volumes. We tried to answer this question by using the GR4 model at different sub-daily time steps. Our results indicate that, on average, the use of a shorter model time step (up to 3 h) may significantly improve model performance even when simulations are evaluated at a larger (daily) time scale. Thanks to our extensive tests on a large data set, we were able to analyse the links between model performance behaviour and a number of catchment and flood events characteristics. This led us to detect that improvements

of model performance at shorter time steps should be rather expected for fast-reactive catchments, subject to short-duration and highly-variable flood and storm events.

This analysis was useful also to detect the presence of model deficiencies at sub-daily time steps, such as the degradation of the model bias at shorter time steps. This kind of analysis could be applied to other models for the same diagnostic purpose.

Diagnosis of the internal coherence of the model

We have proposed a model diagnostic procedure based on some specific evaluations of the model input-state-output response at different time steps. First, the impact of inputs temporal distribution on model performance has been evaluated. Despite the general positive impact of higher-resolution precipitation inputs on model performance, our results confirmed the presence of a structural inadequacy in the GR4 model at shorter time steps, affecting its capacity to reproduce adequately the water balance.

A major outcome of our diagnosis was given by the analysis of the fluxes consistency at multiple time steps. The quest of the internal coherence of the model across time steps led us to understand and localize the structural inadequacies at multiple time steps. In particular, the evaluation of the consistency of the water gains and losses as the model time step changes proved to be essential. For the GR4 model, it was proved that the current representation of the interception (i.e. fast-feedback evaporation term) as a simple neutralisation function generates multiple internal inconsistencies in the model at sub-daily time steps. The automated calibration of the model seeks to compensate the underestimation of interception loss at shorter time steps by modifying other fluxes (evaporation from production reservoir and exchange). In this way, the problems are hidden but the capacity of reproducing the water balance in high-flow conditions is jeopardized. Moreover, the inconsistency of interception fluxes due to this structural inadequacy leads to spurious time-step dependencies of the calibrated parameters.

Improvements of the GR models at multiple sub-daily time steps

Finally, we have shown that it is possible to solve the problems of model internal coherence by targeted structural changes following the indications provided by the fluxes diagnostics at multiple time steps. In particular, in the case of the GR models, the insertion of an interception store of fixed capacity proved to be effective to stabilize all the model fluxes across time steps. This turned out to make the calibrated parameters more consistent across time steps. This finding indicates that the analysis of parameters time-step dependencies must always be preceded by an evaluation of the coherence of the model fluxes across time steps, to avoid confusing spurious parameters trends with something hydrologically meaningful.

Last but not least, the structural modifications made to improve the internal coherence of the model lead also to improve model performance at sub-daily time steps. The simulation bias over flood events is significantly reduced thanks to the insertion of the interception store, something that nobody would have suspected initially.

Thus, we have finally retained this new improved model structure with interception store (GR4-I) that is best suited for coherent simulations at multiple time steps. This structure includes an interception store (component 'activated' at sub-daily time steps) with a fixed (i.e. not optimised) capacity, which must be computed for a given time step from climate inputs of the catchment, so that the interception flux equals the daily neutralisation flux. The use of the linear exchange function with an additional parameter proposed in the previous work by Le Moine (2008) is also suggested as a complementary change that fits better with the introduction of the interception store in the model, leading to a new model with 5 free parameters (GR5-I).

Perspectives for further research

Here below, we briefly discuss some of the most interesting perspectives of this PhD thesis.

Further improvements of the model

Regarding the transferability of the model across time steps, our work brought large improvements in the coherence of the model fluxes, solving almost all the identified inconsistencies. However, a residual problem of (relatively slight) inconsistency of the exchange fluxes in the GR4-I model remains. This could be due to a residual structural inconsistency of the exchange function. This hypothesis should be further tested in future works on the model development. Moreover, linked to this, a new improved exchange function could be identified by starting from a differential equation formulation allowing an analytical integration over fixed time steps (as it is currently done for the other model equations that are sequentially integrated over the time step).

Finally, further work on model development could evaluate the impact of the currently used "*operator splitting*" (or *sequential flux)* approach (Clark and Kavetski, 2010) for the integration of the model governing equations of the GR models on a fixed time step. It would be interesting to analyse the differences between this approach and a possible global solution of the whole state-space system (or a numerical approximation), in terms of model performance and consistency of parameters across time steps. Work on this issue by Santos et al. (2017) is in progress (at Irstea) and may provide further insights on the structural behaviour of the model.

Operational flood forecasting applications

The structural improvements that we have introduced in the GR simulation model at sub-daily time steps will be directly transferable in the GRP forecasting model (Berthet, 2010) that is currently used (at the hourly time step) in the flood forecasting centres in France. In fact, the production part of the GRP model is composed of the same components as the GR4 simulation model initially used in this work (with the neutralisation function for interception). So, the GRP model is affected by the same problem of inconsistency of the interception flux across time steps. To solve this, we suggest using the interception store proposed in this thesis (in place of the neutralisation function) which should allow effectively (and coherently)

changing the time step of the GRP model. We think that the effects of the current representation of interception by a neutralisation function in the GRP model are qualitatively the same as what has been shown for the GR4 model. Thus, based on our findings in the simulation context, the insertion of an interception store in the GRP model is likely to be a good solution for improving:

(i) the adaptability of the model across different sub-daily time steps;
(ii) the model's ability to reproduce the flood volumes.

The structural modification proposed for the forecasting model (GRP) could be tested on our large database, using different time steps (from 5 minutes to 1 day). Last, in the forecasting context, it will be interesting to analyse the effects of data assimilation techniques when the model time step changes.

Combined refinement of temporal and spatial resolutions

Further research on the temporal issues should be carried out not only with a lumped approach, but also with a semi-distributed (or distributed) one. A spatially (semi-)distributed approach could perhaps be preferred at shorter time steps because the spatial variability of precipitation increases as the time step decreases. However, we consider that the improvement of the lumped model presented in this work was a necessary prior step to testing a semi-distributed version of this model. Nevertheless, an exploration of the simultaneous refinement of the temporal and spatial model resolutions was conducted in parallel to this thesis. I have collaborated in the co-supervision of a MSc thesis work (Goullet, 2016) leading to encouraging results. We have tested a semi-distributed version of the GR models, called GR-SD (Lobligeois, 2014), with different combinations of temporal and spatial resolutions, ranging respectively from 3 to 24 h and from 250 km^2 to spatially lumped. These modelling tests were performed on three catchments (extracted from our 240 catchment-set) presenting varied characteristics. The results (reported in Goullet, 2016) showed that the simultaneous refinement of spatial and temporal resolutions could lead to a synergistic effect for improving model performance, especially in Mediterranean catchments subject to highly-variable storms. The positive effects of a combined refinement of spatio-temporal resolutions may enhance the benefits of modelling at high temporal resolutions obtained in this thesis.

References

Abbott, M. B., J. C. Bathurst, J. A. Cunge, P. E. O'Connell and J. Rasmussen (1986). "An introduction to the European Hydrological System - Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system." Journal of Hydrology **87**(1-2): 45-59.

Andréassian, V. (2005). Trois énigmes de modélisation hydrologique. (Three riddles in hydrological modeling.). HDR Thesis (Mémoire d'Habilitation à Diriger les Recherches), UPMC, Paris. 154 pp.

Andréassian, V., C. Perrin, L. Berthet, N. Le Moine, J. Lerat, C. Loumagne, L. Oudin, T. Mathevet, M. H. Ramos and A. Valéry (2009). "Crash tests for a standardized evaluation of hydrological models." Hydrology and Earth System Sciences **13**(10): 1757-1764.

Andréassian, V., C. Perrin, C. Michel, I. Usart-Sanchez and J. Lavabre (2001). "Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models." Journal of Hydrology **250**(1–4): 206-223.

Antoine, J. M., B. Desailly and F. Geselle (2001). "Casualty-causing flood: From the Roussillon region to the Cevennes country." Annales de Geographie(622): 597-623.

Archfield, S. A., M. Clark, B. Arheimer, L. E. Hay, H. McMillan, J. E. Kiang, J. Seibert, K. Hakala, A. Bock, T. Wagener, W. H. Farmer, V. Andréassian, S. Attinger, A. Viglione, R. Knight, S. Markstrom and T. Over (2015). "Accelerating advances in continental domain hydrologic modeling." Water Resources Research **51**(12): 10078-10091.

Arnold, J. G., R. Srinivasan, R. S. Muttiah and J. R. Williams (1998). "Large area hydrologic modeling and assessment part I: Model development." Journal of the American Water Resources Association **34**(1): 73-89.

Asadzadeh, M., L. Leon, W. Yang and D. Bosch (2016). "One-day offset in daily hydrologic modeling: An exploration of the issue in automatic model calibration." Journal of Hydrology **534**: 164-177.

Atkinson, S. E., R. A. Woods and M. Sivapalan (2002). "Climate and landscape controls on water balance model complexity over changing timescales." Water Resources Research **38**(12): 501-5017.

Aussenac, G. (1968). "Interception des précipitations par le couvert forestier." Annales des sciences forestières **25**(3): 135-156.

Bárdossy, A. and S. K. Singh (2008). "Robust estimation of hydrological model parameters." Hydrology and Earth System Sciences **12**(6): 1273-1283.

Bastola, S. and C. Murphy (2013). "Sensitivity of the performance of a conceptual rainfall-runoff model to the temporal sampling of calibration data." Hydrology Research **44**(3): 484-494.

Bergström, S. (1991). "Principles and confidence in hydrological modelling." Nordic Hydrology **22**(2): 123-136.

Berndtsson, R. (1987). "On the use of cross-correlation analysis in studies of patterns of rainfall variability." Journal of Hydrology **93**(1–2): 113-134.

Berne, A., G. Delrieu and B. Boudevillain (2009). "Variability of the spatial structure of intense Mediterranean precipitation." Advances in Water Resources **32**(7): 1031-1042.

Berne, A. and W. F. Krajewski (2013). "Radar for hydrology: Unfulfilled promise or unrecognized potential?" Advances in Water Resources **51**: 357-366.

Berthet, L. (2010). "Prévision des crues au pas de temps horaire: pour une meilleure assimilation de l'information de débit dans un modèle hydrologique. (Flood forecasting at the hourly time step: for a better assimilation of flow information in a hydrological model.) PhD Thesis, INPG, Cemagref (Irstea) Antony, France, 604 pp.".

Berthet, L., V. Andréassian, C. Perrin and P. Javelle (2009). "How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments." Hydrology and Earth System Sciences **13**(6): 819-831.

Beven, K. (1989). "Changing ideas in hydrology - The case of physically-based models." Journal of Hydrology **105**(1-2): 157-172.

Beven, K. (1997). "TOPMODEL: a critique." Hydrological Processes **11**(9): 1069-1085.

Beven, K. (2002). "Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system." Hydrological Processes **16**(2): 189-206.

Beven, K. (2009). "Environmental modelling: an uncertain future?, CRC Press.".

Beven, K., A. Calver and E. M. Morris (1987). "The Institute of Hydrology Distributed Model." Report - UK Institute of Hydrology **98**.

Beven, K. J. (2000). "Uniqueness of place and process representations in hydrological modelling." Hydrology and Earth System Sciences **4**(2): 203-213.

Beven, K. J. and H. L. Cloke (2012). "Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water" by Eric F. Wood et al." Water Resources Research **48**(1): n/a-n/a.

Beven, K. J. and M. J. Kirkby (1979). "A physically based, variable contributing area model of basin hydrology." Hydrol Sci Bull Sci Hydrol **24**(1): 43-69.

Blöschl, G. (2006). "Hydrologic synthesis: Across processes, places, and scales." Water Resources Research **42**(3). W03s02.

Blöschl, G., R. B. Grayson and M. Sivapalan (1995). "On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling." Hydrological Processes **9**(3-4): 313-330.

Blöschl, G. and M. Sivapalan (1995). "Scale issues in hydrological modelling: a review." Hydrological Processes **9**(3-4): 251-290.

Boudou, M., M. Lang, F. Vinet and D. Cœur (2016). "Comparative hazard analysis of processes leading to remarkable flash floods (France, 1930–1999)." Journal of Hydrology **541, Part A**: 533-552.

Bourgin, P. Y., F. Lobligeois, J. Peschard, V. Andréassian, N. Le Moine, L. Coron, C. Perrin, M. H. Ramos and A. Khalifa (2010). "Description des caractéristiques morphologiques, climatiques et hydrologiques de 4436 bassins versants français. Guide d'utilisation de la base de données hydro-climatique." Cemagref (Irstea), Antony, France. 37 p.

Box, G. E. P. and G. M. Jenkins (1976). Time series analysis: forecasting and control. San Francisco, CA, Holden-Day.

Braud, I., P. A. Ayral, C. Bouvier, F. Branger, G. Delrieu, J. Le Coz, G. Nord, J. P. Vandervaere, S. Anquetin, M. Adamovic, J. Andrieu, C. Batiot, B. Boudevillain, P. Brunet, J. Carreau, A. Confoland, J. F. Didon-Lescot, J. M. Domergue, J. Douvinet, G. Dramais, R. Freydier, S. Gérard, J. Huza, E. Leblois, O. Le Bourgeois, R. Le Boursicaud, P. Marchand, P. Martin, L. Nottale, N. Patris, B. Renard, J. L. Seidel, J. D. Taupin, O. Vannier, B. Vincendon and A. Wijbrans (2014). "Multi-scale hydrometeorological observation and modelling for flash flood understanding." Hydrology and Earth System Sciences **18**(9): 3733-3761.

Burnash, R. J. C., R. L. Ferral, R. A. McGuire and R. A. McGuire (1973). A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers. Sacramento, California, U.S. Department of Commerce, National Weather Service, and State of California, Department of Water Resources.

Calder, I. R. (1990). Evaporation in the uplands. Chichester, UD, John Wiley & Sons.

Caliński, T. and J. Harabasz (1974). "A dendrite method for cluster analysis." Communications in Statistics **3**(1): 1-27.

Caroni, E., R. Rosso and F. Siccardi (1986). Nonlinearity and Time-variance of the Hydrologic Response of a Small Mountain Creek. Scale Problems in Hydrology. V. K. Gupta, I. Rodríguez-Iturbe and E. F. Wood, Springer Netherlands. **6:** 19-37.

Cho, J., S. Mostaghimi, M. S. Kang and J. A. Chun (2009). "Sensitivity to grid and time resolution of hydrology components of dansat." Transactions of the ASABE **52**(4): 1121-1128.

Chow, V. T., D. R. Maidment and L. W. Mays (1988). Applied Hydrology, McGraw-Hill.

Chu, X. and A. Steinman (2009). "Event and Continuous Hydrologic Modeling with HEC-HMS." Journal of Irrigation and Drainage Engineering **135**(1): 119-124.

Ciarapica, L. and E. Todini (2002). "TOPKAPI: A model for the representation of the rainfall-runoff process at different scales." Hydrological Processes **16**(2): 207-229.

Clark, M. P. and D. Kavetski (2010). "Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes." Water Resources Research **46**(10): W10510.

Clark, M. P., D. Kavetski and F. Fenicia (2011). "Pursuing the method of multiple working hypotheses for hydrological modeling." Water Resources Research **47**(9): W09301.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener and L. E. Hay (2008). "Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models." Water Resources Research **44**. W00b02.

Conover, W. J. (1999). Practical nonparametric statistics. New York, Wiley.

Coron, L., G. Thirel, O. Delaigue, C. Perrin and V. Andréassian (2017). "The Suite of Lumped GR Hydrological Models in an R Package." Environmental Modelling and Software **94**: 166–171.

Creutin, J. D. and M. Borga (2003). "Radar hydrology modifies the monitoring of flash-flood hazard." Hydrological Processes **17**(7): 1453-1456.

Creutin, J. D. and C. Obled (1980). "Modelling spatial and temporal characteristics of rainfall as input to a flood forecasting model." IAHS-AISH Publication(129): 41-49.

Criss, R. E. and W. E. Winston (2008). "Do Nash values have value? Discussion and alternate proposals." Hydrological Processes **22**(14): 2723-2725.

Cullmann, J., V. Mishra and R. Peters (2006). "Flow analysis with WaSiM-ETH - Model parameter sensitivity at different scales." Advances in Geosciences **9**: 73-77.

Das, T., A. Bárdossy, E. Zehe and Y. He (2008). "Comparison of conceptual model performance using different representations of spatial variability." Journal of Hydrology **356**(1–2): 106-118.

Dawson, C. W. and R. L. Wilby (2001). "Hydrological modelling using artificial neural networks." Progress in Physical Geography **25**(1): 80-108.

de Boer-Euser, T., L. Bouaziz, J. De Niel, C. Brauer, B. Dewals, G. Drogue, F. Fenicia, B. Grelier, J. Nossent, F. Pereira, H. Savenije, G. Thirel and P. Willems (2017). "Looking beyond general metrics for model comparison – lessons from an international model intercomparison study." Hydrol. Earth Syst. Sci. **21**(1): 423-440.

Delrieu, G., V. Ducrocq, E. Gaume, J. Nicol, O. Payrastre, E. Yates, P. E. Kirstetter, H. Andrieu, P. A. Ayral, C. Bouvier, J. D. Creutin, M. Livet, S. Anquetin, M. Lang, L. Neppel, C. Obled, J. Parent-Du-Châtelet, G. M. Saulnier, A. Walpersdorf and W. Wobrock (2005). "The catastrophic flash-flood event of 8-9 September 2002 in the Gard Region, France: A first case study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory." Journal of Hydrometeorology **6**(1): 34-52.

Dong, X., C. M. Dohmen-Janssen and M. J. Booij (2005). "Appropriate spatial sampling of rainfall for flow simulation." Hydrological Sciences Journal **50**(2): 279-298.

Duan, Q., J. Schaake, V. Andréassian, S. Franks, G. Goteti, H. V. Gupta, Y. M. Gusev, F. Habets, A. Hall, L. Hay, T. Hogue, M. Huang, G. Leavesley, X. Liang, O. N. Nasonova, J. Noilhan, L. Oudin, S. Sorooshian, T. Wagener and E. F. Wood (2006). "Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops." Journal of Hydrology **320**(1-2): 3-17.

Duband, D., C. Obled and J. Y. Rodriguez (1993). "Unit hydrograph revisited: an alternate iterative approach to UH and effective precipitation identification." Journal of Hydrology **150**(1): 115-149.

Ducharne, A. (2009). "Reducing scale dependence in TOPMODEL using a dimensionless topographic index." Hydrology and Earth System Sciences **13**(12): 2399-2412.

Eder, G., M. Sivapalan and H. P. Nachtnebel (2003). "Modelling water balances in an Alpine catchment through exploitation of emergent properties over changing time scales." Hydrological Processes **17**(11): 2125-2149.

Edijatno (1991). "Mise au point d'un modèle élémentaire pluie-débit au pas de temps journalier. (Development of an elementary rainfall-runoff model at a daily time-step.), PhD Thesis, Université Louis Pasteur/ENGEES, Strasbourg, Cemagref (Irstea) Antony, France, 242 pp.".

Edijatno, N. D. De Oliveira Nascimento, X. Yang, Z. Makhlouf and C. Michel (1999). "GR3J: A daily watershed model with three free parameters." Hydrological Sciences Journal **44**(2): 263-277.

Escudier, A., P. A. Hans, C. Astier and J. L. Souldadié (2016). "Flood forecasting: Crisis management and sharing flooded areas information." Houille Blanche **2016-October**(5): 5-10.

Eykhoff, P. (1974). "System Identification. Parameter and State Estimation, J. Wiley, London."

Farmer, D., M. Sivapalan and C. Jothityangkoon (2003). "Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis." Water Resources Research **39**(2): SWC11-SWC121.

Fenicia, F., D. Kavetski and H. H. G. Savenije (2011). "Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development." Water Resources Research **47**(11). W11510.

Fenicia, F., H. H. G. Savenije, P. Matgen and L. Pfister (2008). "Understanding catchment behavior through stepwise model concept improvement." Water Resources Research **44**(1). W01402.

Finnerty, B. D., M. B. Smith, D. J. Seo, V. Koren and G. E. Moglen (1997). "Space-time scale sensitivity of the Sacramento model to radar-gage precipitation inputs." Journal of Hydrology **203**(1-4): 21-38.

Friedman, M. (1937). "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance." Journal of the American Statistical Association **32**(200): 675-701.

Furusho, C., C. Perrin, J. Viatgé, R. Lamblin and V. Andréassian (2016). "Collaborative work between operational forecasters and scientists for better flood forecasts." Houille Blanche **2016-August**(4): 5-10.

Gabellani, S., G. Boni, L. Ferraris, J. von Hardenberg and A. Provenzale (2007). "Propagation of uncertainty from rainfall to runoff: A case study with a stochastic rainfall generator." Advances in Water Resources **30**(10): 2061-2071.

Gericke, O. J. and J. C. Smithers (2014). "Review of methods used to estimate catchment response time for the purpose of peak discharge estimation." Hydrological Sciences Journal **59**(11): 1935-1971.

Gerrits, A. M. J. (2010). The role of interception in the hydrological cycle. PhD, TU Delft, Delft, the Netherlands.

Gerrits, A. M. J., L. Pfister and H. H. G. Savenije (2010). "Spatial and temporal variability of canopy and forest floor interception in a beech forest." Hydrological Processes **24**(21): 3011-3025.

Goullet, J. (2016). A la recherche des résolutions spatiales et temporelles caractéristiques du comportement hydrologique d'un bassin versant. (Searching for characteristic spatio-temporal resolutions for modelling the hydrological behavior of a catchment.) Master Thesis, UPMC, Paris, France. 52 pp.

Grayson, R. B., I. D. Moore and T. A. McMahon (1992). "Physically based hydrologic modeling: 2. Is the concept realistic?" Water Resources Research **28**(10): 2659-2666.

Guéguen, C., G. L'Henaff, O. Laurantin and J.-M. Soubeyroux (2011). "Description des traitements en vue d'une réanalyse des lames d'eau ("Description of the treatments for a reanalysis of water slides")." Météo-France. 1-28.

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.

Gupta, H. V., C. Perrin, G. Blöschl, A. Montanari, R. Kumar, M. Clark and V. Andréassian (2014). "Large-sample hydrology: A need to balance depth with breadth." Hydrology and Earth System Sciences **18**(2): 463-477.

Gustard, A., A. Bullock and J. M. Dixon (1992). "Low flow estimation in the United Kingdom." Report 108, UK Institute of Hydrology. 88 p.

Haddeland, I., D. P. Lettenmaier and T. Skaugen (2006). "Reconciling Simulated Moisture Fluxes Resulting from Alternate Hydrologic Model Time Steps and Energy Budget Closure Assumptions." Journal of Hydrometeorology **7**(3): 355-370.

Hughes, D. A. (1993). "Variable time intervals in deterministic hydrological models." Journal of Hydrology **143**(3–4): 217-232.

Ishidaira, H., K. Takeuchi, Z. Xu, T. Ao, J. Magome and M. Kudo (2003). "Effect of spatial and temporal resolution of precipitation data on the accuracy of long-term runoff simulation." IAHS-AISH Publication(282): 186-193.

Jakeman, A. J. and G. M. Hornberger (1993). "How much complexity is warranted in a rainfall-runoff model?" Water Resources Research **29**(8): 2637-2649.

Javelle, P., C. Fouchier, P. Arnaud and J. Lavabre (2010). "Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations." Journal of Hydrology **394**(1–2): 267-274.

Jeong, J., N. Kannan, J. Arnold, R. Glick, L. Gosselink and R. Srinivasan (2010). "Development and Integration of Sub-hourly Rainfall-Runoff Modeling Capability Within a Watershed Model." Water Resources Management **24**(15): 4505-4527.

Jothityangkoon, C., M. Sivapalan and D. L. Farmer (2001). "Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development." Journal of Hydrology **254**(1–4): 174-198.

Kandel, D. D., A. W. Western and R. B. Grayson (2005). "Scaling from process timescales to daily time steps: A distribution function approach." Water Resources Research **41**(2): 1-16.

Kandel, D. D., A. W. Western, R. B. Grayson and H. N. Turral (2004). "Process parameterization and temporal scaling in surface runoff and erosion modelling." Hydrological Processes **18**(8): 1423-1446.

Kannan, N., S. M. White, F. Worrall and M. J. Whelan (2007). "Sensitivity analysis and identification of the best evapotranspiration and runoff options for hydrological modelling in SWAT-2000." Journal of Hydrology **332**(3–4): 456-466.

Kauffeldt, A., F. Wetterhall, F. Pappenberger, P. Salamon and J. Thielen (2016). "Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level." Environmental Modelling & Software **75**: 68-76.

Kavetski, D. and M. P. Clark (2010). "Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction." Water Resources Research **46**(10). W10511.

Kavetski, D. and M. P. Clark (2011). "Numerical troubles in conceptual hydrology: Approximations, absurdities and impact on hypothesis testing." Hydrological Processes **25**(4): 661-670.

Kavetski, D., F. Fenicia and M. P. Clark (2011). "Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment." Water Resources Research **47**(5). W05501.

Kirchner, J. W. (2006). "Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology." Water Resources Research **42**(3). W03s04.

Kirchner, J. W., X. Feng, C. Neal and A. J. Robson (2004). "The fine structure of water-quality dynamics: The (high-frequency) wave of the future." Hydrological Processes **18**(7): 1353-1359.

Kitanidis, P. K. and R. L. Bras (1980). "Real-time forecasting with a conceptual hydrologic model. 1. Analysis of uncertainty." Water Resources Research **16**(6): 1025-1033.

Klemeš, V. (1983). "Conceptualization and scale in hydrology." Journal of Hydrology **65**(1-3): 1-23.

Klemeš, V. (1986a). "Dilettantism in hydrology: Transition or destiny?" Water Resources Research **22**(9S): 177S-188S.

Klemeš, V. (1986b). "Operational testing of hydrological simulation models." Hydrological Sciences Journal/Journal des Sciences Hydrologiques **31**(1): 13-24.

Kobold, M. and M. Brilly (2006). "The use of HBV model for flash flood forecasting." Natural Hazards and Earth System Science **6**(3): 407-417.

Koch, R. W. and M. F. Kekhia (1987). Effect of rainfall intensity distribution on excess precipitation. Engineering Hydrology Conference, A Symposium; August 3-7, 1987, Williamsburg Hilton National Conference Center, Williamsburg, Virginia, United States, ASCE.

Krajewski, W. F., V. Lakshmi, K. P. Georgakakos and S. C. Jain (1991). "A Monte Carlo study of rainfall sampling effect on a distributed catchment model." Water Resources Research **27**(1): 119-128.

Kruskal, W. H. and W. A. Wallis (1952). "Use of Ranks in One-Criterion Variance Analysis." Journal of the American Statistical Association **47**(260): 583-621.

Le Moine, N. (2008). Le bassin versant de surface vu par le souterrain : une voie d'amélioration des performances et du réalisme des modèles pluie-débit? (The surface watershed seen by the underground: a way to improve performance and realism of rainfall-runoff models?) PhD Thesis, University Pierre et Marie Curie, Cemagref (Irstea), Paris, France. 348 pp.

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**(6). W06428.

Lerat, J., V. Andréassian, C. Perrin, J. Vaze, J. M. Perraud, P. Ribstein and C. Loumagne (2012). "Do internal flow measurements improve the calibration of rainfall-runoff models?" Water Resources Research **48**(2). W02511.

Lindstrom, G., B. Johansson, M. Persson, M. Gardelin and S. Bergstrom (1997). "Development and test of the distributed HBV-96 hydrological model." Journal of Hydrology **201**(1-4): 272-288.

Linsley, R. K. (1982). Rainfall-runoff models: An overview. Proceedings of the International Symposium on Rainfall-Runoff Modelling, Water Resour. Publ., Littleton, Colorado.

Littlewood, I. G. and B. F. W. Croke (2008). "Data time-step dependency of conceptual rainfall-streamflow model parameters: An empirical study with implications for regionalisation." Hydrological Sciences Journal **53**(4): 685-695.

Littlewood, I. G. and B. F. W. Croke (2013). "Effects of data time-step on the accuracy of calibrated rainfall-streamflow model parameters: Practical aspects of uncertainty reduction." Hydrology Research **44**(3): 430-440.

Littlewood, I. G., B. F. W. Croke and P. C. Young (2011). "Discussion of "effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge"." Hydrological Sciences Journal **56**(3): 521-524.

Lobligeois, F. (2014). Mieux connaître la distribution spatiale des pluies améliore-t-il la modélisation des crues ? Diagnostic sur 181 bassins versants français. ("Better understanding the spatial distribution of rainfall improves flood modeling? Diagnosis on 181 French catchments.") PhD Thesis, Antony, France. 310 pp.

Lobligeois, F., V. Andréassian, C. Perrin, P. Tabary and C. Loumagne (2014). "When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events." Hydrology and Earth System Sciences **18**(2): 575-594.

Maniak, U. (1997). Hydrologie und Wasserwirtschaft (Hydrology and Water Management). Heidelberg, Germany, Springer-Verlag.

Mathevet, T. (2005). Quels modèles pluie-débit globaux pour le pas de temps horaire? Développement empirique et comparaison de modèles sur un large échantillon de bassins versants. (Which lumped rainfall-runoff models for the hourly time-step? Empirical development and comparison of models on a large sample of catchments.) PhD Thesis, ENGREF, Cemagref (Irstea), Paris, France. 463 pp.

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." IAHS-AISH Publication(307): 211-219.

Melsen, A. L., J. A. Teuling, J. J. F. P. Torfs, R. Uijlenhoet, N. Mizukami and P. M. Clark (2016a). "HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models." Hydrology and Earth System Sciences **20**(3): 1069-1079.

Melsen, L., A. Teuling, P. Torfs, M. Zappa, N. Mizukami, M. Clark and R. Uijlenhoet (2016b). "Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-Alpine basin." Hydrol. Earth Syst. Sci. **20**(6): 2207-2226.

Menabde, M. and M. Sivapalan (2001). "Linking space–time variability of river runoff and rainfall fields: a dynamic approach." Advances in Water Resources **24**(9–10): 1001-1014.

Merz, R., J. Parajka and G. Blöschl (2009). "Scale effects in conceptual hydrological modeling." Water Resources Research **45**(9). W09405.

Michel, C. (1983). "Que peut-on faire en hydrologie avec modèle conceptuel à un seul paramètre?" La Houille Blanche(1): 39-44.

Michel, C., C. Perrin, V. Andreassian, L. Oudin, T. Mathevet, V. Andréassian, A. Hall, N. Chahinian and J. Schaake (2006). "Has basin-scale modelling advanced beyond empiricism?" IAHS-AISH publication(307): 108-116.

Montanari, A. and D. Koutsoyiannis (2012). "A blueprint for process-based modeling of uncertain hydrological systems." Water Resources Research **48**(9). W09555.

Moore, R. J., V. A. Bell and D. A. Jones (2005). "Forecasting for flood warning." Comptes Rendus - Geoscience **337**(1-2): 203-217.

Moore, R. J. and R. T. Clarke (1981). "A distribution function approach to rainfall runoff modeling." Water Resources Research **17**(5): 1367-1382.

Morel-Seytoux, H. J. (1988). "Soil aquifer stream interactions - a reductionist attempt toward physical-stochastic integration." Journal of Hydrology **102**(1-4): 355-379.

Moretti, G. and A. Montanari (2007). "AFFDEF: A spatially distributed grid based rainfall-runoff model for continuous time simulations of river discharge." Environmental Modelling & Software **22**(6): 823-836.

Morin, E., Y. Enzel, U. Shamir and R. Garti (2001). "The characteristic time scale for basin hydrological response using radar data." Journal of Hydrology **252**(1-4): 85-99.

Morin, E., K. P. Georgakakos, U. Shamir, R. Garti and Y. Enzel (2002). "Objective, observations-based, automatic estimation of the catchment response timescale." Water Resources Research **38**(10): 30.

Mouelhi, S. (2003). Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. (Towards a coherent chain of lumped rainfall-runoff models at pluri-annual, annual, monthly and daily time-steps.) PhD Thesis, ENGREF, Antony, France. 323 pp.

Mouelhi, S., C. Michel, C. Perrin and V. Andréassian (2006a). "Linking stream flow to rainfall at the annual time step: The Manabe bucket model revisited." Journal of Hydrology **328**(1–2): 283-296.

Mouelhi, S., C. Michel, C. Perrin and V. Andréassian (2006b). "Stepwise development of a two-parameter monthly water balance model." Journal of Hydrology **318**(1–4): 200-214.

Nalbantis, I. (1995). "Use of multiple-time-step information in rainfall-runoff modeling." Journal of Hydrology **165**(1-4): 135-159.

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. (Estimation of the anthropogenic effects on the rainfall-runoff relationship at the catchment scale by using an empirical model.) PhD Thesis, ENPC, Paris, France. 390 pp.

Nash, J. E. and J. V. Sutcliffe (1970a). "River flow forecasting through conceptual models part I - A discussion of principles." Journal of Hydrology **10**(3): 282-290.

Nash, J. E. and J. V. Sutcliffe (1970b). "River flow forecasting through conceptual models part I — A discussion of principles." Journal of Hydrology **10**(3): 282-290.

Nicótina, L., E. Alessi Celegon, A. Rinaldo and M. Marani (2008). "On the impact of rainfall patterns on the hydrologic response." Water Resources Research **44**(12). W12401.

Obled, C., J. Wendling and K. Beven (1994). "The sensitivity of hydrological models to spatial rainfall patterns: an evaluation using observed data." Journal of Hydrology **159**(1-4): 305-333.

Obled, C., I. Zin and B. Hingray (2009). "Choix des pas de temps et d'espace pour des modélisations parcimonieuses en hydrologie des crues." La Houille Blanche(5): 81-87.

Oreskes, N., K. Shrader-Frechette and K. Belitz (1994). "Verification, validation, and confirmation of numerical models in the earth sciences." Science **263**(5147): 641-646.

Ostrowski, M., M. Bach, V. Gamerith and S. de Simone (2010). "Analysis of the Time-step Dependency of Parameters In Conceptual Hydrological Models." Department for Engineering Hydrology and Water Resources Management, Darmstadt University, Darmstadt, Germany

Oudin, L., V. Andréassian, C. Perrin and F. Anctil (2004). "Locating the sources of low-pass behavior within rainfall-runoff models." Water Resources Research **40**(11): W1110101-W1110114.

Oudin, L., F. Hervieu, C. Michel, C. Perrin, V. Andreassian, F. Anctil and C. Loumagne (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." Journal of Hydrology **303**(1-4): 290-306.

Oudin, L., C. Perrin, T. Mathevet, V. Andréassian and C. Michel (2006). "Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models." Journal of Hydrology **320**(1-2): 62-83.

Paschalis, A., S. Fatichi, P. Molnar, S. Rimkus and P. Burlando (2014). "On the effects of small scale space-time variability of rainfall on basin flood response." Journal of Hydrology **514**: 313-327.

Perrin, C. (2000). Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative. (Towards an improvement of a lumped rainfall-runoff model through a comparative approach.). PhD Thesis, Antony, France. 287 pp.

Perrin, C., V. Andréassian, C. R. Serna, T. Mathevet and N. Le Moine (2008). "Discrete parameterization of hydrological models: Evaluating the use of parameter sets libraries over 900 catchments." Water Resources Research **44**(8). W08447.

Perrin, C., C. Michel and V. Andréassian (2001). "Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments." Journal of Hydrology **242**(3–4): 275-301.

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

Perrin, C., L. Oudin, V. Andreassian, C. Rojas-Serna, C. Michel and T. Mathevet (2007). "Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models." Hydrological Sciences Journal **52**(1): 131-151.

Popper, K. (1959). The logic of scientific discovery. London, Routldege.

Pushpalatha, R., C. Perrin, N. Le Moine, T. Mathevet and V. Andréassian (2011). "A downward structural sensitivity analysis of hydrological models to improve low-flow simulation." Journal of Hydrology **411**(1-2): 66-76.

Refsgaard, J. C. and H. J. Henriksen (2004). "Modelling guidelines - Terminology and guiding principles." Advances in Water Resources **27**(1): 71-82.

Robinson, J. S. and M. Sivapalan (1997). "An investigation into the physical causes of scaling and heterogeneity of regional flood frequency." Water Resources Research **33**(5): 1045-1059.

Rubel, F. (1996). "Scale dependent statistical precipitation analysis." Proc.of the Int. Conf. on Water Resour. & Environ. Res., 29-31 October 1996, Kyoto, 317-324.

Rutter, A. J., K. A. Kershaw, P. C. Robins and A. J. Morton (1971). "A predictive model of rainfall interception in forests, 1. Derivation of the model from observations in a plantation of Corsican pine." Agricultural Meteorology **9**(C): 367-384.

Santos, L., G. Thirel and C. Perrin (2017). A state-space representation of the GR4J rainfall-runoff model. Geophysical Research Abstracts, EGU General Assembly 2017, Vienna, Austria, Copernicus Publications.

Saulnier, G. M. and M. Le Lay (2009). "Sensitivity of flash-flood simulations on the volume, the intensity, and the localization of rainfall in the Cévennes-Vivarais region (France)." Water Resources Research **45**(10). W10425.

Sauquet, E., L. Gottschalk and I. Krasovskaia (2008). "Estimating mean monthly runoff at ungauged locations: An application to France." Hydrology Research **39**(5-6): 403-423.

Savenije, H. H. G. (2004). "The importance of interception and why we should delete the term evapotranspiration from our vocabulary." Hydrological Processes **18**(8): 1507-1511.

Schaake, J., Q. Duan, M. Smith and V. Koren (2000). "Criteria to select basins for hydrologic model development and testing." 15th Conference on Hydrology, 10-14 January 2000, Amer. Mereor. Soc., Long Beach, CA, Paper P1.8.: 149-152.

Schaake, J. C., V. I. Koren, Q. Y. Duan, K. Mitchell and F. Chen (1996). "Simple water balance model for estimating runoff at different spatial and temporal scales." Journal of Geophysical Research: Atmospheres **101**(D3): 7461-7475.

Schreider, S. Y. and A. J. Jakeman (2001). "Streamflow modelling on a subdaily time step in the Upper Murray Basin." Mathematical and Computer Modelling **33**(6-7): 659-668.

Schreider, S. Y., A. J. Jakeman and A. B. Pittock (1996). "Modelling rainfall-runoff from large catchment to basin scale: The Goulburn Valley, Victoria." Hydrological Processes **10**(6): 863-876.

Schultz, G. A. (1994). "Meso-scale modelling of runoff and water balances using remote sensing and other GIS data." Hydrological Sciences Journal **39**(2): 121-142.

Seyfried, M. S. and B. P. Wilcox (1995). "Scale and the Nature of Spatial Variability: Field Examples Having Implications for Hydrologic Modeling." Water Resources Research **31**(1): 173-184.

Shannon, C. E. (1949). "Communication in the Presence of Noise." Proceedings of the IRE **37**(1): 10-21.

Simas, M. J. and R. H. Hawkins (1998). Lag time characteristics for small watersheds in the U.S. International Water Resources Engineering Conference - Proceedings.

Singh, V. P. (1992). Elementary Hydrology, Prentice Hall.

Singh, V. P. and D. K. Frevert (2005). Watershed Models, Taylor & Francis.

Singh, V. P. and D. A. Woolhiser (2002). "Mathematical modeling of watershed hydrology." Journal of Hydrologic Engineering **7**(4): 270-292.

Sivapalan, M., G. Blöschl, L. Zhang and R. Vertessy (2003). "Downward approach to hydrological prediction." Hydrological Processes **17**(11): 2101-2111.

Socolofsky, S., E. E. Adams and D. Entekhabi (2001). "Disaggregation of daily rainfall for continuous watershed modeling." Journal of Hydrologic Engineering **6**(4): 300-309.

Tan, B. Q. and K. M. Oconnor (1996). "Application of an empirical infiltration equation in the SMAR conceptual model." Journal of Hydrology **185**(1-4): 275-295.

Tang, Y., P. Reed, T. Wagener and K. Van Werkhoven (2007). "Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation." Hydrology and Earth System Sciences **11**(2): 793-817.

Tessier, Y., S. Lovejoy, P. Hubert, D. Schertzer and S. Pecknold (1996). "Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions." Journal of Geophysical Research: Atmospheres **101**(21): 26427-26440.

Thirel, G., V. Andréassian, C. Perrin, J. N. Audouy, L. Berthet, P. Edwards, N. Folton, C. Furusho, A. Kuentz, J. Lerat, G. Lindström, E. Martin, T. Mathevet, R. Merz, J. Parajka, D. Ruelland and J. Vaze (2015). "Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments." Hydrological Sciences Journal **60**(7-8): 1184-1199.

Tobin, C., B. Schaefli, L. Nicótina, S. Simoni, G. Barrenetxea, R. Smith, M. Parlange and A. Rinaldo (2013). "Improving the degree-day method for sub-daily melt simulations with physically-based diurnal variations." Advances in Water Resources **55**: 149-164.

Todini, E. (1988). "Rainfall-runoff modeling - Past, present and future." Journal of Hydrology **100**(1-3): 341-352.

Todini, E. (1996). "The ARNO rainfall-runoff model." Journal of Hydrology **175**(1-4): 339-382.

Valente, F., J. S. David and J. H. C. Gash (1997). "Modelling interception loss for two sparse eucalypt and pine forests in central Portugal using reformulated Rutter and Gash analytical models." Journal of Hydrology **190**(1-2): 141-162.

Van Esse, W. R., C. Perrin, M. J. Booij, D. C. M. Augustijn, F. Fenicia, D. Kavetski and F. Lobligeois (2013). "The influence of conceptual model structure on model performance: A comparative study for 237 French catchments." Hydrology and Earth System Sciences **17**(10): 4227-4239.

Vidal, J. P., E. Martin, L. Franchistéguy, 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**(11): 1627-1644.

Viglione, A., G. B. Chirico, J. Komma, R. Woods, M. Borga and G. Blöschl (2010). "Quantifying space-time dynamics of flood event types." Journal of Hydrology **394**(1-2): 213-229.

Vinet, F. (2007). "Flood risk management in French Mediterranean basins." WIT Transactions on Ecology and the Environment **104**: 261-270.

Wagener, T. and H. S. Wheater (2006). "Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty." Journal of Hydrology **320**(1-2): 132-154.

Wang, Y., B. He and K. Takase (2009). "Effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge." Hydrological Sciences Journal **54**(5): 886-898.

Wheater, H., A. Jakeman and K. Beven (1993). "Progress and directions in rainfall-runoff modelling. In Jakeman, A.J., Beck, M.B., and McAleer, M.J. (eds.) Modelling change in environmental systems. John Wiley & sons, Chichester, pp. 101-132. ."

Wood, E. F., D. P. Lettenmaier and V. G. Zartarian (1992). "A land-surface hydrology parameterization with subgrid variability for general circulation models." Journal of Geophysical Research **97**(D3): 2717-2728.

Woolhiser, D. A. and D. C. Goodrich (1988). "Effect of storm rainfall intensity patterns on surface runoff." Journal of Hydrology **102**(1-4): 335-354.

Yang, X., Q. Liu, Y. He, X. Luo and X. Zhang (2016). "Comparison of daily and sub-daily SWAT models for daily streamflow simulation in the Upper Huai River Basin of China." Stochastic Environmental Research and Risk Assessment **30**(3): 959-972.

Yapo, P. O., H. V. Gupta and S. Sorooshian (1996). "Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data." Journal of Hydrology **181**(1): 23-48.

Ye, W., B. C. Bates, N. R. Viney, M. Sivapalan and A. J. Jakeman (1997). "Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments." Water Resources Research **33**(1): 153-166.

Yilmaz, K. K., H. V. Gupta and T. Wagener (2008). "A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model." Water Resources Research **44**(9). W09417.

Young, C. B. (2011). "Recursive Estimation and Time-Series Analysis, 2nd ed., XVIII, 502p., Springer-Verlag, New York.".

Young, P. C. and K. J. Beven (1994). "Data-based mechanistic modeling and the rainfall-flow nonlinearity." Environmetrics **5**(3): 335-363.

Young, P. C. and H. Garnier (2006). "Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems." Environmental Modelling & Software **21**(8): 1055-1072.

Yu, B., U. Cakurs and C. W. Rose (1998). "An assessment of methods for estimating runoff rates at the plot scale." Transactions of the American Society of Agricultural Engineers **41**(3): 653-661.

Yu, B., C. A. A. Ciesiolka, C. W. Rose and K. J. Coughlan (1997). "A note on sampling errors in the rainfall and runoff data collected using tipping bucket technology." Transactions of the American Society of Agricultural Engineers **40**(5): 1305-1309.

Zawadzki, I. I. (1973). "Statistical properties of precipitation patterns." J. Appl. Meteor. **12**: 459-472.

Zhao, R.-J., Z. Yi-Lin, L.-R. Fang, X.-R. Liu and S. Zhang Quan (1980). "The Xinanjiang model." Hydrological forecasting. Proc. Oxford symposium, April 1980, (International Association of Hydrological Sciences, Washington DC; IAHS-AISH Publication 129): 351-356.

Appendices

A - Complements to literature review: Time stepping schemes

One important issue in the context of temporal scaling in rainfall-runoff modelling is the accuracy and stability of numerical schemes. Often analytical solutions for the systems of differential equations classically used in hydrological models do not exist for most fluxes formulations. Thus, numerical approximations techniques must be used (e.g. see Clark and Kavetski, 2010). These approximations may be needed at *two different levels*: for the whole system of interconnected fluxes and for each flux separately.

At a first level we may distinguish between:

I. *Operator-splitting* (OS), or *"sequential flux", approach* (Clark and Kavetski, 2010) which decomposes unwieldy systems of differential equations into simpler sub-problems, which can be treated individually using an analytical solution or a numerical approximation for each of them (see the second level, for the possible numerical approximations of each flux);

II. *Global solutions of the whole state-space system* of differential equations.

At a second level, different classes of numerical methods may be applied, for each individual flux, if we have adopted the OS approach, or for the global system, otherwise. One may distinguish between:

i. *Explicit* vs. *implicit* formulations;

ii. *Fixed-step* vs. *adaptive* implementations.

The most commonly used scheme in current conceptual hydrological models is the *explicit Euler method (EE).* It approximates the analytical solution using the fluxes at the beginning of each time step (explicit formulation). It is commonly used, for its algorithmic simplicity and computational speed, but it is known to be highly unreliable unless numerical error control is used. Conversely, *implicit Euler method (IE)* uses the fluxes at the end of the time step (implicit formulation). It requires more expensive iterative solutions, but it is generally more robust and facilitates model calibration thanks to derived smooth response surfaces. Usually both these methods are used with a fixed-step implementation.

As discussed by Clark and Kavetski (2010), it is recognized that the fixed-step size time stepping schemes can produce inaccurate (in both explicit or implicit formulations) or unstable (in the explicit case) solutions. To overcome these limitations the adaptive approximation scheme reduces the step size Δt until the error is below a user-prescribed tolerance τ. However, the adaptive integration with error control remains rare in conceptual hydrological models. A synthesis of the main numerical schemes used in hydrological modelling is reported in Table A.1, which highlights their advantages and disadvantages.

In the GR4J rainfall-runoff model (Perrin et al., 2003), it was chosen to adopt the *operator splitting technique with analytical integration of each individual model flux in a predetermined sequence*. Clark and Kavetski (2010) and Kavetski and Clark (2011) suggest that the use of OS is probably suboptimal for the additional "splitting" errors with respect to the whole state-space system (delays or over/under estimation of the flux given by the original governing equations). However, to our knowledge, the impact of the operator splitting in hydrological modelling has not been thoroughly investigated yet. The authors cited above only tested the impact of the numerical approximations of the second level of the classification above, but they did not analyse the case of the operator splitting with analytical integration of each flux.

Still, we agree with Clark and Kavetski (2010) in saying that it is not recommended to use the OS scheme without formulating the governing differential equations. They state that governing equations are seldom formulated when OS is used. However, in the case of the GR4J model, the governing equations have been formulated for all fluxes (see Le Moine, 2008), except the groundwater exchange. The main limitation of the OS approach without formulation of governing equations is that "*the model conceptualization is conflated with the numerical implementation*". This may reveal a methodological weakness "*because modelling errors arising from the physical conceptualization and numerical implementation become difficult to separate, diagnose, and resolve*" (Clark and Kavetski, 2010).

Clark and Kavetski (2010) evaluate *8 different time stepping schemes* in terms of numerical reliability and computational efficiency by using 6 different conceptual rainfall-runoff models on 13 catchments. They show that the numerical errors of fixed-step explicit time stepping schemes *routinely dwarf the structural errors* of the model conceptualization. These numerical errors have significant implications for hypothesis testing, sensitivity analysis, parameter estimation (and interpretation) and operational prediction. For example, the numerical artefacts of unreliable time-stepping schemes can compensate for structural weaknesses in the model, as highlighted also by Kavetski and Clark (2010) and Kavetski et al. (2011), leading a model to obtain better results for the wrong reasons. From the range of 8 time stepping schemes evaluated by Clark and Kavetski (2010), the fixed-step IE method and the adaptive explicit Heun method emerge as good practical choices.

Also Kavetski et al. (2011) show the importance of using robust numerical approximations. The use of the unreliable fixed-step EE time stepping scheme is shown to introduce artificial time scale dependencies in model parameters, not detected by using the implicit Euler scheme. The EE scheme is unreliable also in the context of hypothesis testing, since it produces changes in performances between validation and calibration that are not observed when using a robust numerical approximation as the IE scheme. Another effect of numerical approximations noted by Kavetski et al. (2011) is that since the EE scheme evaluates the outflows before the inflows are added to the stores, it introduces a delay into the response dynamics, which unreliably interacts with the explicit routing component of the models.

Numerical method	Advantages	Disadvantages	Example of models implementing it
Explicit Euler method (EE)	Algorithmic simplicity; computational speed.	Only conditionally stable; unreliable (unless using numerical control error); inconsistencies at different time steps; less robust for larger step sizes	The most part of conceptual hydrological models
Implicit Euler method (IE)	Unconditionally stable; robust even for large step sizes; smooth solution with respect to input forcing and parameters (important in gradient-based parameter optimization).	Need of expensive iterative solutions (e.g. Newton-Raphson root solver); algorithmic complexity; computational cost.	Models in SUPERFLEX framework (Kavetski et al., 2011)
Adaptive sub-stepping schemes	Stability and accuracy.	Computational cost; roughness of the resulting objective function (Kavetski and Clark, 2010)	Rarely used (e.g. Clark and Kavetski, 2010)
Operator splitting with analytical integration of most fluxes	Algorithmic simplicity; computational speed; analytical solutions may become possible for individual fluxes	Potential "splitting" errors and inconsistencies at different time steps (to be verified case by case); less robust for larger step sizes	GR4J (Perrin et al., 2003) TOPMODEL (Beven, 1997) VIC (Wood et al., 1992)

TABLE A.1 – Summary of approximation methods generally used to solve numerically continuous differential equations in hydrological modelling with main advantages/disadvantages and examples of models implementing these methods.

B - Complements to Chapter 2: Analysis of the quality of the 6-minute precipitation measurements

This appendix is a complement to Chapter 2 (Section 2.2.1 “Rainfall data”). It provides some results of our analysis of the coherence between 6-minute rain gauges measurements and hourly data reconstructed at the same stations (at Irstea) by using the COMEPHORE reanalysis by Météo-France. This analysis of data coherence was performed at 257 rain gauges stations of the 6-min network operating over the period July 2005 – December 2006, which is common between the two databases we have.

Figure B.1 shows that the mean absolute deviations between the hourly cumulus from 6-min rain gauges data and from the hourly reanalysis are negligible (always smaller than 0.01 mm).

Figure B.2 shows that the maximum absolute deviations are also small on average (3 mm). Note that the presence of rare and not significant errors can be due to the fact that the reference hourly data reconstructed from the COMEPHORE reanalysis are combined with daily surface measurements and 5-minute radar observations.

Figure B.3 shows that the null-error frequency is very high, being at least 95% for all the stations.

This analysis confirms a very good consistency of the two products, which is an evidence of good quality for the precipitation data used in this work.

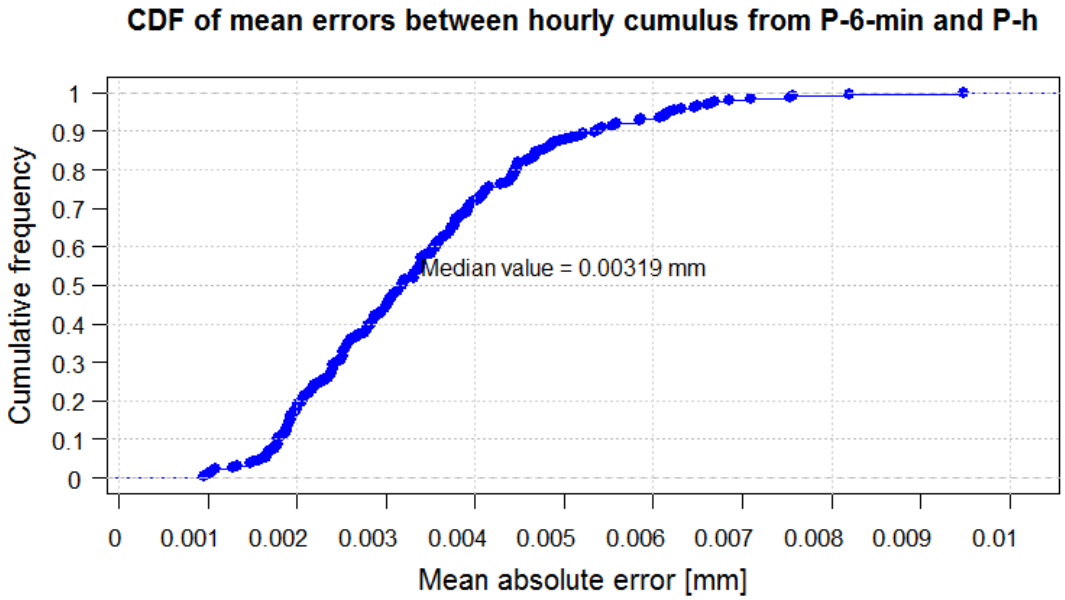


FIGURE B.1 – Cumulative distribution of mean absolute error between hourly cumulus from the 6-minute rain gauges and hourly values reconstructed at the same stations at Irstea by using the COMEPHORE reanalysis by Météo-France (257 stations with enough data over the common period July 2005 – December 2006).

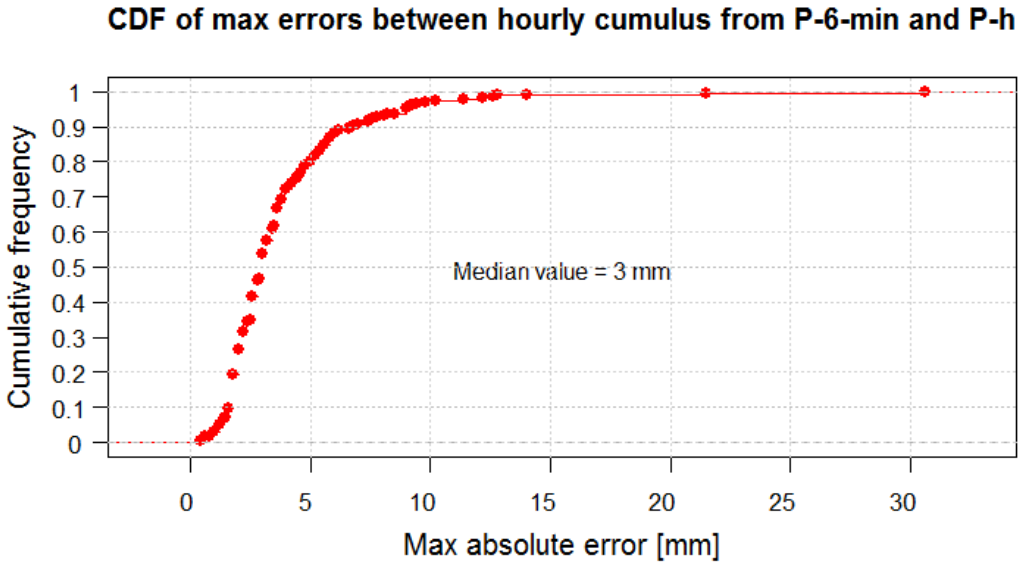


FIGURE B.2 – Cumulative distribution of maximum absolute error between hourly cumulus from 6-minute rain gauges and hourly values reconstructed at the same stations at Irstea by using the COMEPHORE reanalysis by Météo-France (257 stations with enough data over the common period July 2005 – December 2006).

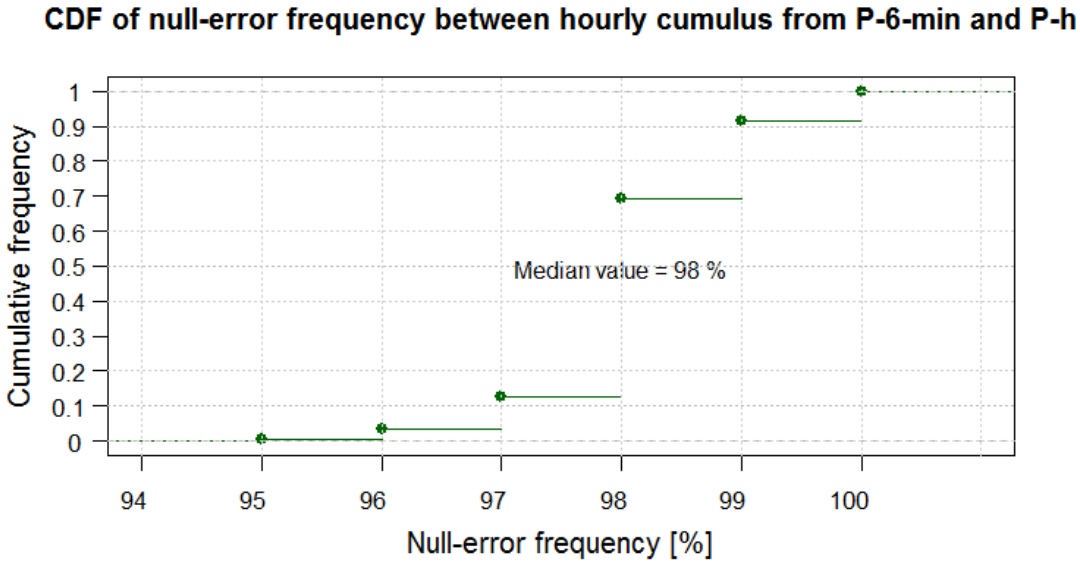


FIGURE B.3 – Cumulative distribution of null-error frequency (number of time steps with a null error/total n. of time steps) between hourly cumulus from 6-minute rain gauges station and hourly values reconstructed at the same stations at Irstea by using the COMEPHORE reanalysis by Météo-France (257 stations with enough data over the common period July 2005 – December 2006).

C - List of the 240 selected catchments

Station name	Code	S [km^2]	Z [m]
L'Ill à Altkirch	A1050310	238	282
L'Ill à Didenheim	A1080330	668	242
L'Ill à Osthouse	A2240310	3296	150
La Mossig à Soultz-les-Bains	A2842010	167	169
La Bruche à Holtzheim 2	A2860110	676	144
La Moder à Schweighouse-sur-Moder aval	A3301010	622	144
La Zorn à Waltenheim-sur-Zorn	A3472010	684	147
La Moder à Drusenheim	A3501010	787	118
La Mortagne à Gerbéviller	A6731220	498	234
L'Eichel à Oermingen	A9352050	280	214
La Meuse à Domrémy-la-Pucelle	B1150010	1035	266
La Seine à Plaines-Saint-Lange	H0100020	686	180
L'Aube à Bar-sur-Aube	H1201010	1298	167
La Laine à Soulaines-Dhuys	H1333010	22	137
Le Serein à Bierre-lès-Semur	H2322010	267	312
Le Serein à Dissangis	H2332020	645	189
Le Serein à Chablis Pont de la déviation	H2342020	1119	125
Le Serein à Beaumont	H2342030	1352	85
L'Armance à Chessy-les-Prés	H2473010	477	110
L'Ouanne à Charny	H3122010	559	133
L'Ouanne à Gy-les-Nonains	H3122020	877	99
Le ru d'Ancoeur à Blandy	H3923010	188	70
L'Orge à Épinay-sur-Orge Le Breuil	H4232040	641	38
L'Yvette à Villebon-sur-Yvette	H4243010	231	54
L'Orge à Morsang-sur-Orge	H4252010	934	38
Le Réveillon à Férolles-Attilly La Jonchère	H4333410	57	73
Le Grand Morin à Meilleray	H5702010	350	125
L'Orgeval à Boissy-le-Châtel Le Theil, ultrasons	H5723011	105	79
Le Grand Morin à Pommeuse	H5732010	769	62
L'Aire à Beausite Amblaincourt	H6102010	283	222
L'Aire à Varennes-en-Argonne	H6122010	629	154
La Vesle à Saint-Brice-Courcelles	H6412010	735	71
L'Ardres à Faverolles-et-Coëmy	H6423020	145	84
L'Ysieux à Luzarches Bertinval	H7833540	55	37
Le Lieutel à Neauphle-le-Vieux	H7913620	77	64
L'Eure à Saint-Luperce	H9021010	315	155
L'Avre à Acon	H9202010	477	119

L'Avre à Muzy	H9222010	872	78
La Dives au Mesnil-Mauger	I2051040	616	16
L'Orne à la Courbe	I3121010	935	141
Le Noireau à Cahan Les Planches - CD 911	I3462010	525	54
L'Orne à Grimbosq barrage du Val de Viard	I3531010	2263	16
La Seulles à Juvigny-sur-Seulles	I4022010	134	56
La Seulles à Tierceville	I4032010	256	8
La Souleuvre à Carville	I5053010	116	76
La Vire à Malloué	I5101010	478	63
La Vire à Tessy-sur-Vire	I5121020	640	39
La Vire à Saint-Lô Moulin des Rondelles	I5221010	883	15
L'Aure à Maisons Pont-Fatu	I5321510	136	17
La Drôme à Sully	I5352010	241	22
Le Petit Douet à Héauville	I6964010	12	21
Le Trieux à Saint-Clet Moulin-de-Châteaulin	J1721720	416	12
L'Horn à Mespaul Kertanguy	J3014330	38	40
Le Moros à Concarneau Pont D 22	J4514010	21	20
L'Oust à Pleugriffet La Tertraie	J8202310	931	42
L'Yvel à Loyat Pont D 129	J8363110	301	35
La Coise à Larajasse Le Nézel	K0663310	61	583
L'Anzon à Débats-Rivière-d'Orpra Cotes	K0744010	180	410
L'Aix à Saint-Germain-Laval	K0813020	197	378
Le Gand à Neaux	K0974010	86	364
Le Rhins à Saint-Cyr-de-Favières Pont Mordon	K0983010	435	293
L'Arroux à Dracy-Saint-Loup Surmoulin	K1251810	772	291
L'Arroux à Étang-sur-Arroux Pont du Tacot	K1321810	1792	268
La Vouzance à Saint-Léger-sur-Vouzance	K1414010	131	242
Le Bedat à Saint-Laure	K2773120	348	297
L'Ambène à Ennezat	K2774020	109	308
La Morge à Maringues Côte Rouge	K2783010	668	292
La Bouble à Chareil-Cintrat	K3373010	563	246
La Vauvise à Saint-Bouize La Grange	K4073110	389	151
La Cisse à Nazelles-Négron	K4853000	799	53
Le Cher à Chambonchard La Caborne	K5090900	524	321
L'Aumance à Hérisson	K5383020	877	199
L'Ouatier à Moulins-sur-Yèvre Maubranche	K5554580	161	136
L'Auron à Bourges L'Ormediot	K5653010	575	129
L'Arnon à Méreau Alnay	K6192420	2177	98
La Grande Sauldre à Brinon-sur-Sauldre 2	K6332520	617	130
La Petite Sauldre à Ménétréol-sur-Sauldre	K6373020	332	151
La Sauldre à Salbris Valaudran	K6402520	1220	103
L'Indre à Ardentes	K7202610	688	157
La Ringoire à Déols	K7207510	96	140
L'Indre à Saint-Cyran-du-Jambot	K7312610	1707	82
L'Indrois à Genillé	K7433030	402	78

L'Échandon à Saint-Branchs	K7514010	128	66
La Vienne à Eymoutiers	L0050630	375	408
La Combade à Roziers-Saint-Georges	L0093020	170	314
La Briance à Condat-sur-Vienne Chambon Veyrinas	L0563010	605	218
L'Aurence à Isle	L0614020	86	235
La Creuse à Felletin	L4010710	190	507
La Petite Creuse à Fresselines Puy Rageaud	L4411710	853	218
La Bouzanne à Velles Forges	L4653010	438	124
La Gartempe à Saint-Victor-en-Marche	L5001810	78	434
La Gartempe à Folles Bessines	L5101810	568	297
Le Vincou à Bellac 2	L5223020	286	178
La Brame à Oradour-Saint-Genest	L5323010	232	171
L'Anglin à Mérigny	L5741910	1634	77
La Sarthe à Saint-Céneri-le-Gérei Moulin du Désert	M0050620	909	124
L'Orne Saosnoise à Montbizot Moulin Neuf Cidrerie	M0243010	502	52
L'Huisne à Nogent-le-Rotrou Pont de bois	M0361510	833	102
L'Huisne à Montfort-le-Gesnois La Pécardière	M0421510	1936	57
La Sarthe à Saint-Denis-d'Anjou Beffes	M0680610	7523	21
L'Ozanne à Trizay-lès-Bonneval Prémoteux	M1034020	267	126
Le Loir à Saint-Maur-sur-le-Loir	M1041610	1080	118
L'Yerre à Saint-Hilaire-sur-Yerre Bêchereau	M1114011	293	101
L'Escotais à Saint-Paterne-Racan	M1354020	68	66
Le Loir à Durtal	M1531610	7918	22
L'Aisne à Javron-les-Chapelles Les Chapelles	M3014010	144	145
La Mayenne à Madré	M3020910	329	129
La Mayenne à Ambrières-les-Vallées Cigné	M3060910	832	102
La Mayenne à l'Huisserie Bonne	M3340910	2908	45
L'Ouine au Breuil-Bernard Les Alleuds	M7005610	63	163
La Maine à Remouillé	M7453010	595	19
Le Tarnon à Florac	O3064010	132	554
Le Tarn à Mostuéjouls La Muse	O3141010	945	405
Le ruisseau du Mas de Pomier Durzon à Nant Le Mas du Pré	O3335010	110	535
La Dourbie à Millau Massebiau 3	O3394030	654	373
Le ruisseau de l'Espérelle	O3395010	602	408
Le Cernon à Sainte-Eulalie-de-Cernon terrain de Football	O3414010	8	592
Le ruisseau de Balastière	O3416610	18	532
L'Anrou	O3575510	12	769
Le Dourdou à Vabres-l'Abbaye Le Poujol	O3594020	684	302
Le Rancé à Saint-Sernin-sur-Rance	O3754010	289	305
Le Tescou à Saint-Nauphary	O4984320	284	99
Le Dourdou à Conques	O7874010	545	236
Le Célé à Orniac Les Amis du Célé	O8133520	1246	142
La Triouzoune à Saint-Angel	P0924010	79	631
La Loyre à Voutezac Pont de l'Aumonerie	P3234010	104	153

La Loyre à Saint-Viance Pont de Burg	P3274010	254	105
La Corrèze à Corrèze Pont de Neupont	P3352510	164	478
La Corrèze à Tulle Pont des soldats	P3502510	354	224
La Corrèze à Brive-la-Gaillarde Le Prieur	P3922510	954	103
La Vézère à Montignac Le Pertuis	P4161020	3174	74
La Vézère à Campagne	P4271010	3657	56
L'Auvézère à Benayes	P6202510	22	361
L'Auvézère à Lubersac	P6222510	113	291
L'Auvézère à Tourtoirac	P6362510	665	133
L'Auvézère au Change Aubarède	P6382510	884	99
La Lizonne à Saint-Séverin Le Marchais	P8284010	624	51
La Dronne à Bonnes	P8312520	1912	37
La Dronne à Coutras Coutras aval	P8462520	2798	6
L'Adour à Hères Ju Belloc	Q0360010	1093	144
L'Adour à Cahuzac-sur-Adour	Q0450010	1279	120
La Nivelle à Saint-Pée-sur-Nivelle	S5144010	142	31
Le Coney à Fontenoy-le-Château	U0124010	316	258
Le Breuchin à la Proiselière-et-Langle	U0415010	123	343
La Lanterne à Fleurey-lès-Faverney	U0474010	1028	209
La Colombine à Frotey-lès-Vesoul	U0525010	143	220
La Saône à Ray-sur-Saône	U0610010	3761	195
L'Ognon à Chassey-lès-Montbozon Bonnal	U1044010	852	249
L'Ognon à Beaumotte-Aubertans	U1054010	1259	229
L'Ognon à Pin	U1074020	1699	203
La Saône à Auxonne	U1120010	8790	178
Le Saint-Nicolas à Rougemont-le-Château	U2305210	9	474
La Bourbeuse à Froidefontaine	U2324210	300	329
L'Allan à Fesches-le-Châtel	U2334010	705	321
La Savoureuse à Belfort	U2345030	144	358
La Savoureuse à Vieux-Charmont	U2345040	235	318
Le Rhome à Lachapelle-sous-Chaux Lachapelle Bis	U2345420	19	408
La Rosemontoise à Rougegoutte	U2345830	23	444
L'Allan à Courcelles-lès-Montbéliard	U2354010	1109	311
Le Cusancin à Cusance	U2425250	154	311
La Grosne à Jalogny Cluny	U3214010	334	243
L'Azergues à Châtillon	U4624010	337	211
La Turdine à l' Arbresle Gobelette	U4636610	169	231
L'Azergues à Lozanne	U4644010	798	198
La Bourbre à Tignieu-Jameyzieu	V1774010	696	204
Le Suran à Neuville-sur-Ain La Planche	V2814020	331	272
Le Suran à Pont-d'Ain	V2814030	359	241
L'Yzeron à Francheville Taffignon	V3015020	127	190
Le Rival à Beaufort	V3424310	467	280
Le Ternay à Savas Ternay	V3517010	25	520
La Cance à Sarras	V3524010	382	148

La Galaure à Saint-Uze	V3614010	228	164
Le Doux à Colombier-le-Vieux	V3724010	380	244
L'Embroye à Toulaud	V4025010	7	360
La Véore à Beaumont-lès-Valence Laye	V4034020	195	128
La Glueyre à Gluiras Tisoneche	V4145210	72	430
Le Roubion à Soyans	V4414010	192	276
Le Jabron à Souspierre	V4455010	79	256
L'Ardèche à Ucel	V5014030	478	207
La Cèze à Tharaux	V5454010	665	109
La Cèze à Montclus	V5464015	833	89
L'Ouvèze à Entrechaux Pont Saint Michel	V6042010	475	241
Le Gardon de Saint-Jean à Corbès Roc Courbe	V7135010	262	147
Le Gardon de Saint-Jean à Saint-Jean-du-Gard Bâtiment communal	V7135017	153	185
L'Herbasse à Clérieux Pont de l'Herbasse	W3534020	190	142
Le Coulon à Oppède La Garrigue	X3484020	885	109
La Massane à Argelès-sur-Mer Mas d'en Tourens	Y0115410	16	101
Le Tech à Argelès-sur-Mer Pont d'Elne	Y0284060	722	12
Le Fresquel à Villepinte	Y1314010	211	129
Le Lampy à Raissac-sur-Lampy	Y1345010	58	139
Le Fresquel à Carcassonne Pont Rouge	Y1364010	936	94
L'Orbieu à Saint-Martin-des-Puits	Y1524010	172	187
L'Orbieu à Luc-sur-Orbieu	Y1564010	590	38
La Cesse à Mirepeisset	Y1605050	251	35
L'Aude à Moussan Moussoulens - écluse	Y1612020	4914	12
L'Arre au Vigan La Terrisse	Y2015010	155	198
La Vis à Saint-Laurent-le-Minier	Y2035010	306	156
L'Hérault à Laroque	Y2102010	916	139
La Lergue à Lodève	Y2214010	181	160
Le Vistre à Bernis	Y3514020	280	18
Le Vistre lit mineur au Cailar	Y3534010	496	2
L'Arc à Pourrières	Y4002010	50	252
L'Arc à Meyreuil Pont de Bayeux	Y4022010	297	174
L'Huveaune à Roquevaire 2	Y4414030	165	167
Le Gapeau à Solliès-Pont	Y4604020	184	81
Le Réal Martin à la Crau Decapris	Y4615020	284	31
Le Gapeau à Hyères Sainte-Eulalie	Y4624010	536	12
Le Cauron à Bras Pont de l'Avocade	Y5005210	146	254
L'Argens à Châteauvert	Y5032010	505	183
La Bresque à Salernes Les Vingalières	Y5115020	166	212
L'Argens aux Arcs	Y5202010	1651	42
L'Aille à Vidauban Le Baou	Y5215020	229	46
La Nartuby à Trans-en-Provence	Y5235010	194	151
L'Argens à Roquebrune-sur-Argens	Y5312010	2514	8
Le Reyran à Fréjus Sainte-Brigitte	Y5325010	73	23
La Giscle à Cogolin	Y5424010	66	7

La Siagne à Callian Ajustadoux	Y5514040	166	235
La Brague à Biot Plan Saint-Jean	Y5605210	42	19
Le Loup à Tourrettes-sur-Loup Les Vallettes	Y5615010	206	133
Le Loup à Villeneuve-Loubet Moulin du Loup	Y5615030	289	7
L'Estéron à Sigale Pont du Coude	Y6434005	262	377
L'Estéron au Broc La Clave	Y6434010	443	140
La Bévéra à Sospel Pont D 2204	Y6635010	96	343
La Sienne à Saint-Sever-Calvados La Croix du Rocher	I7001040 *	4	264
L'Adour à Tarbes 2	Q0120060 *	388	314
L'Arros à Tournay	Q0522530 *	239	260
Le Larcis à Lannux	Q1094010 *	426	92
Le Gave d'Oloron à Oloron-Sainte-Marie Oloron-SNCF	Q7002910 *	1101	211
La Nive à Ossès	Q9102510 *	597	110
La Morcille à Villié-Morgon Pont des Versauds	U4506010 *	4	316
L'Aveyron à Laguépie 1	O5292510 *	1604	163
L'Échez à Bordères-sur-l'Échez	Q0224020 *	154	286
Le Luy à Saint-Pandelon	Q3464010 *	1144	6
L'Eyre à Salles	S2242510 *	1678	14
La Sormonne à Belval	B5572010 *	369	146
L'Aa à Wizernes	E4035710 *	392	19
La Béthune à Saint-Aubin-le-Cauf	G2011010 *	307	13
L'Eaulne à Martin-Église	G2203010 *	317	4
L'Adour à Saint-Vincent-de-Paul	Q3120010 *	7707	6
L'Yzeron à Craponne	V3015010 *	46	243

TABLE C.1 – List of the 240 selected catchments with correspondent code (“Banque-Hydro” data base), surface (S [km2]) and mean altitude (Z [m]). In this list, 17 catchments (reported at the end of the list and identified by ’*’ after the station code), were selected after relaxing some selection criteria (see Section 2.3.4).

D - Flood event selection procedure

The flood event selection procedure includes the following steps:

i. Select the current maximum discharge (Q_{max}) in the hourly streamflow time series.

ii. Within a 15-day period before (after) the peak flow, the beginning (end) of the event is placed at the time step at which the streamflow is lower than a threshold discharge Q_0:

$$Q_0 = \max\left(Q_{max} \cdot 0.25,\ Q_{min} + 0.05(Q_{max} - Q_{min})\right) \qquad (D.1)$$

where Q_{min} is the minimum discharge observed over the 15-day period before (after) the peak flow.

iii. If the precipitation is not null at the beginning of the event defined so far, then the beginning of the event is shifted to the first of the preceding time steps at which precipitation is null.

iv. The current event selected is discarded if one of the following criteria is verified:

- if the duration of the rising (falling) limb is shorter than 3 h. This allows discarding high flows within a rising/falling limb.
- if the event period overlaps with a previously defined flood event period. This allows one to consider independent flood events.
- if the precipitation amount, cumulated before the discharge peak, is less than 2.5 mm, the current event is discarded. This allows one to automatically remove implausible values in the precipitation–flow pair data.
- if the flow series over the selected period is almost constant, according to two empirically derived conditions: (1) maximum absolute relative difference of two successive hourly flows less than 10% and (2) ratio of peak flow on mean flow less than 1.3. These empirical conditions allow discarding events that are not actual independent floods with recognizable rising/declining limbs, but are part of recession flows.
- if the hourly flow series around the peak presents either a gap rate greater than 10% or an overly long interpolated piece (linearly interpolated from variable time-step data). The latter is verified by empirically derived conditions on the maximum allowed duration of subsequent equal flow differences around the peak ($d^h_{\Delta Q\ cost}$): (1) $d^h_{\Delta Q\ cost}$ less than 24 h; (2) $d^h_{\Delta Q\ cost}$ less than 50% of the duration of rising and falling limbs. This allows one to keep only floods with a minimum quality of observed flow data to evaluate streamflow simulations at sub-daily time steps.

The restrictive selection criteria of point (iv) were adjusted empirically by preliminary visual analysis of the hydrographs in order to avoid subjectivity in the selection of a representative and criticized set of floods. We believe that this automated procedure should be preferred to manual selection by hydrograph visualization and expert judgement only.

E – GOUE index

To quantify the degradation of the information on precipitation and streamflow temporal dynamics when passing from a short time step (e.g. 6 min) to a larger time step (equal to *s* short time steps, e.g. 1 h), we built the Goodness of Uniform Estimates (GOUE) derived from the Goodness of Rainfall Estimation (GORE) Index proposed by Andréassian et al. (2001). The GOUE Index is based on the Nash-Sutcliffe criterion (Nash and Sutcliffe, 1970b) formulation applied to rainfall or streamflow, considering the data at the shortest temporal resolution available as the reference observation. The GOUE Index is computed on rainfall and streamflow for a given catchment as in Equations (E.1) and (E.2):

$$GOUE(P) = 1 - \frac{\sum_{i=1}^{n}(P_i^{Ls} - P_i^{\mathrm{f}})^2}{\sum_{i=1}^{n}(P_i^{\mathrm{f}} - \overline{P^{\mathrm{f}}})^2} \qquad \text{(E.1)}$$

$$GOUE(Q) = 1 - \frac{\sum_{i=1}^{n}(Q_i^{Ls} - Q_i^{\mathrm{f}})^2}{\sum_{i=1}^{n}(Q_i^{\mathrm{f}} - \overline{Q^{\mathrm{f}}})^2} \qquad \text{(E.2)}$$

where *n* is the number of time steps of the period; P_i^{f} and Q_i^{f} [mm] are, respectively, the reference catchment precipitation and streamflow at the short time step *i* (6 min); P_i^{Ls} and Q_i^{Ls} are the precipitation and streamflow estimate at time step *i* obtained by uniformly disaggregating the value at a large time step (i.e. 1/*s* of that precipitation or streamflow); and $\overline{P^{\mathrm{f}}}$ and $\overline{Q^{\mathrm{f}}}$ are the mean of the reference precipitation and streamflow. The GOUE Index varies between $-\infty$ and 1.

GOUE Index values are expected to decrease as *s* increases and to depend on the characteristics of precipitation on the basin. The GOUE Index should decrease more quickly for basins subject to intense convective storm events than for basins with mostly stratiform events.

F - Consistency of the GR4 model parameters at different time steps

As discussed in Chapter 2, the capacity of the production store, x_1, is the GR4 model's only free parameter that is independent of time step. For the other three calibrated parameters, the time-step dependency can be formulated by theoretical relationships derived from the model governing equations (see Table 2.3).

In this study, after calibrating the free parameters at each model time step, we verified the consistency of parameter values obtained at the different time steps tested (according to the theoretical relationships in Table 2.3), as done in previous studies (Le Moine, 2008).

Figure F.1 reports a statistical summary of the comparison of the calibrated parameter values at different time steps for the 240 catchments used in this study, after normalization at the hourly reference time step thanks to the theoretical relationships given in Table 2.3. The parameters' consistency at different time steps is verified, with no clear deviations from their theoretical relationships, except for the time base of the unit hydrograph (x_4), because of its discrete construction. In fact, for example at the daily time step the UH form does not change for different values of x_4 lower than 24 h, since x_4 is rounded to the nearest integer value. Thus, as the time step decreases the time base parameter is identified with greater precision. No other strong time-scale dependencies can be identified, but only a slight decrease of the water exchange coefficient (x_2). Kavetski et al. (2011) feared that strong artificial time-scale dependencies in parameter estimates are introduced when unreliable time stepping schemes are used. However, in the present case, where the operator splitting technique is used the results in Fig. F.1 show that the GR4 model parameters are fairly stable.

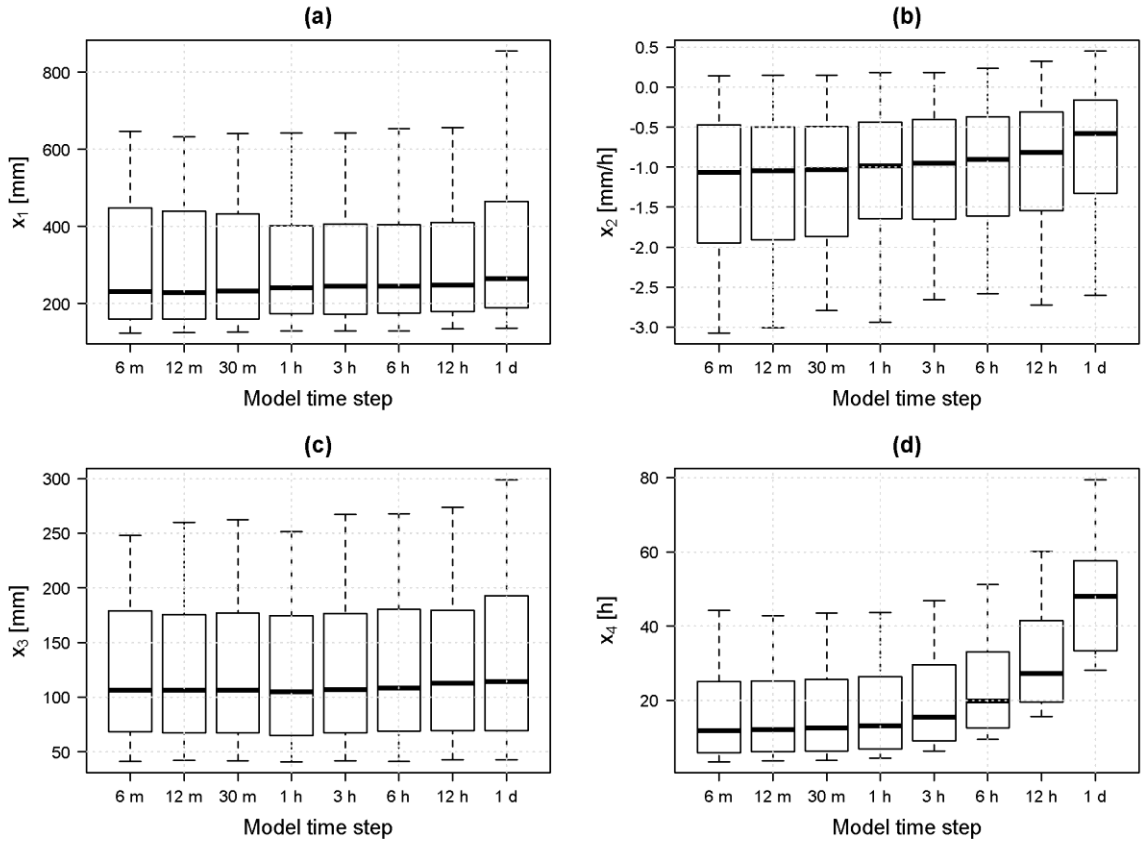


FIGURE F.1 – GR4 model parameters calibrated at eight time steps from 6 min to 1 day: (a) maximum capacity of the production store (x_1); (b) water exchange coefficient (x_2); (c) maximum capacity of the routing store at one time step ahead (x_3); (d) time base of the unit hydrograph (x_4). The box plots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles.

G - Formulation of the GR5-I model (complements to Chapter 5)

Here we summarize the **formulation of the *GR5-I* model structure retained in this thesis** as the new recommended model version **for multiple sub-daily time steps** (see Figure G.1), based on both model performance and temporal consistency of the fluxes. This structure differs from the GR4 baseline model described in Chapter 2 for the insertion of an interception store of fixed capacity, determined to maximize the coherence of the interception fluxes across time steps (see Chapter 5), and for the use of the linear exchange function proposed by Le Moine (2008), with an additional free parameter.

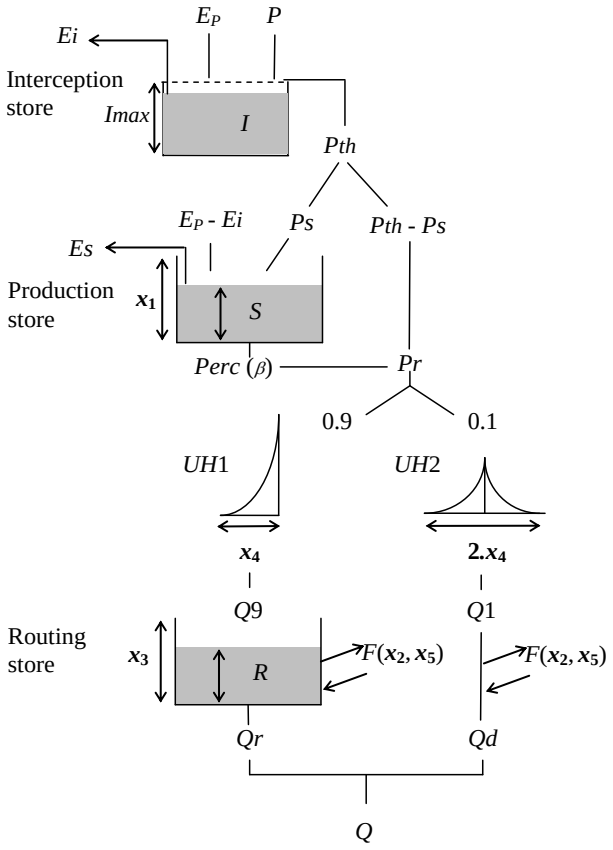


FIGURE G.13 – Schematic representation of the GR5-I model structure, modified from the GR4 baseline model by Perrin et al. (2003) by insertion of an interception store of maximum capacity *Imax* and modification of the exchange function, as a linear function with a fifth free parameter. The interception store capacity is fixed by optimizing the consistency of the interception loss at different time steps.

Note:

The equations of the model are here **presented over a generic sub-daily time step Δt**. We remind that for the daily time step, a zero-capacity interception store is recommended, i.e. using the neutralisation function of GR4J. Thus, the GR5-I model at the daily time step coincides with the GR5J model proposed by Le Moine (2008).

Model equations

Inputs

The inputs of the model are the precipitation P and the potential evapotranspiration E over each time step Δt.

Parameters

The interception store capacity I_{max} can be fixed prior to running the model, simply by processing the two time-series of the inputs (P and E) to ensure the consistency of the interception losses at different time steps (as discussed in Chapter 5). To this end, the calculation of the interception loss can be iterated with the equations summarized here below for different values of the capacity I_{max}, until getting to the same value than the interception loss obtained at a daily time step by a zero-capacity interception store.

The five free parameters to be calibrated are:

- the maximum capacity of the production store (x_1, [mm]);
- the water exchange coefficient (x_2, [mm/t.s.]);
- the routing store reference capacity (x_3, [mm]);
- the time base of the unit hydrograph (x_4, [t.s.]);
- the threshold level for which the exchange function changes in sign ($x_5[-] \in [0, 1]$).

Before starting the simulation, the ordinates of the two unit hydrographs, which will be used to spread effective rainfall at each time step, can be computed based on the x_4 parameter only. These ordinates are calculated from the corresponding *S-curves* (cumulative distribution of the input with time), as explained by Perrin et al. (2003), denoted as *SH1* and *SH2*. These curves are calculated over a number of time steps equal to the base time of the unit hydrographs, as:

$$For\ 0 \leq t \leq x_4 \colon\ SH1(t) = \left(\frac{t}{x_4}\right)^{5/2}$$

$$For\ t > x_4 \colon\ SH1(t) = 1$$

$For\ 0 < t \leq x_4\colon\ SH2(t) = \frac{1}{2}\left(\frac{t}{x_4}\right)^{5/2}$

$For\ x_4 < t \leq 2x_4\colon\ SH2(t) = 1 - \frac{1}{2}\left(2 - \frac{t}{x_4}\right)^{5/2}$

$For\ t > 2x_4\colon\ SH2(t) = 1$

The ordinates of the unit hydrographs are then calculated as:

$UH1(j) = SH1(j) - SH1(j-1)$

$UH2(j) = SH2(j) - SH2(j-1)$

where *j* is an integer between 1 and the maximum number of ordinates, *n* and *m*, respectively for *UH1* and *UH2*, i.e. the smallest integers exceeding x_4 and $2x_4$.

Initial conditions

At the beginning of each time step, the model states are:

- I_0, i.e. the initial water content in the interception store [mm];
- S , i.e. the initial water content in the production store [mm];
- R , i.e. the initial water content in the routing store [mm];

Interception

The first operation is the determination of an *interception loss* E_i, i.e. evaporation from intercepted water (*I*), by means of a simple bucket interception store. The interception loss E_i is calculated as:

$E_i = \min(E_P, P + I_0)$

Thus, the potential evapotranspiration is reduced to a net evapotranspiration capacity (E_n):

$E_n = E_P - E_i$

The precipitation is reduced to throughfall or net rainfall (P_{th}) that is calculated as:

$P_{th} = max[0, P - (I_{max} - I_0) - E_i]$

The interception store water content is then updated as:

$I = I_0 + (P - E_i - P_n)$

and will be used as initial condition for the next time step.

Production store

Then, in case P_s is greater than zero, the production store is filled by a part P_s of the throughfall calculated as:

$$P_s = \frac{x_1\left(1-\left(\frac{S}{x_1}\right)^2\right)\tanh\left(\frac{P_{th}}{x_1}\right)}{1+\frac{S}{x_1}\tanh\left(\frac{P_{th}}{x_1}\right)}$$

In case E_s is greater than zero, the evaporation from the production store E_s is calculated as:

$$E_s = \frac{S\left(2-\frac{S}{x_1}\right)\tanh\left(\frac{E_P-E_i}{x_1}\right)}{1+\left(1-\frac{S}{x_1}\right)\tanh\left(\frac{E_P-E_i}{x_1}\right)}$$

The production store level S is updated by adding P_s or removing E_s.

Then a percolation leakage $Perc$ is calculated as:

$$Perc = S\left\{1-\left[1+\left(\frac{S}{\beta_{\Delta t}\cdot x_1}\right)^4\right]^{-\frac{1}{4}}\right\}$$

where the percolation constant $\beta_{\Delta t}$ is a function of the model time step Δt (in seconds):

$$\beta_{\Delta t} = 5.25\left(\frac{3600}{\Delta t}\right)^{\frac{1}{4}}$$

The production store level S is updated by removing *Perc* and will serve as initial condition for the next time step.

The percolation $Perc$ reaches the routing part of the model, where it is added to the part of the net rainfall that has bypassed the production store, i.e. $P_r = Perc + (P_n - P_s)$.

The routed water amount P_r is divided into two flow components by a fixed ratio: 90% is routed by a unit hydrograph (*UH*1) and then a non-linear routing store, and 10% only by another unit hydrograph (*UH*2).

Unit hydrographs

At the current time step *k*, the outputs *Q*9 and *Q*1 of the unit hydrographs *UH*1 and *UH*2 correspond to the convolution of the previous rainfalls by the distribution key corresponding to the UH ordinates. They are calculated by:

$$Q9(k) = \frac{9}{10}.\sum_{j=1}^{n} UH\ 1(j).\ \Pr(k - j + 1)$$

$$Q1(k) = \frac{1}{10}.\sum_{j=1}^{m} UH\ \ 2(j).\ \Pr(\ k\ -\ j\ +\ 1)$$

Groundwater exchange term

A groundwater exchange loss (or gain) is then released (or added) from (to) both flow components. The potential half exchange, *F*, is calculated as proposed by Le Moine (2008):

$$F = x_2\left(\frac{R}{x_3} - x_5\right)$$

If $F > 0$ the actual exchange gains (F_G) are twice as *F*.

If $F < 0$, the actual exchange losses (F_L) are limited by the water available in the routing store and by the flow components coming from the unit hydrographs:

$$F_L = -[\min(|F|,\ \ R + Q_9) + \min(|F|,\ Q_1)]$$

Routing store

Then the water content of the routing store *R* is updated by adding the output of the unit hydrograph UH1 ($Q9$) and removing (or adding) the exchange component *F*.

The outflow from the routing store gives the first flow component Q_r, and is calculated as:

$$Q_r = R\left\{1 - \left[1 + \left(\frac{R}{x_3}\right)^4\right]^{-\frac{1}{4}}\right\}$$

The content of the routing store is updated by removing Q_r and will serve as initial condition for the next time step.

Pseudo-direct flow component

The output of the second unit hydrograph UH2 ($Q1$) provides the direct flow component Q_d, after being subject to the same groundwater exchange component (loss or gain) *F*, as:

$$Q_d = \max(0; Q1 + F)$$

Simulated flow

Finally, the total streamflow at the outlet is given by the sum of the two flow components:

$$Q = Q_r + Q_d$$