Journal of Hydrology 575 (2019) 1308–1327

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Hydrological modelling at multiple sub-daily time steps: Model improvement via flux-matching

Andrea Ficchì*, Charles Perrin, Vazken Andréassian

Irstea, UR HYCAR, 1 rue Pierre-Gilles de Gennes, CS 10030, 92761 Antony, France

ARTICLE INFO

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Saman Razavi, Associate Editor

Keywords:
Hydrological modelling
Temporal flux consistency
Parameter identifiability
Interception
Model efficiency
Flood events

ABSTRACT

Consistent hydrological models for multiple time steps are needed to respond to the operational requirement of using a model for different objectives and with varied data sets. In this paper, we propose a rationale for multi-time step model development based on the temporal consistency of the fluxes modelled. The new methodology is applied for the temporal downscaling of a lumped rainfall-runoff model (GR4J) from daily to sub-hourly time steps, within a coherent modelling framework. The modelling tests are performed at eight time steps from daily to 6-min, for a set of 240 French catchments, for which 2400 flood events were selected. To identify an improved multi-time step model structure, a two-step approach is followed. First, we propose a model diagnosis to evaluate the internal behaviour of the baseline model across time steps, whereby a significant inconsistency was found in the simulated water fluxes, especially interception. Second, we recommend a prognosis to improve the scale invariance of the modelled water balance by matching the model fluxes across time scales. The new model structure retained for sub-daily time steps is derived from the daily baseline model by refining the representation of the interception process, so that the flux-matching condition is satisfied. A complementary modification of the groundwater exchange function is also resumed following a previous study. The structural changes are motivated by the improvement of model flux consistency across time steps, while also improving model performance. This new dual paradigm for model identification, based on both flux coherence and output accuracy, also results in more robust model parameters across time steps.

1. Introduction

1.1. Multi-time–step hydrological modelling: motivation and challenges

Hydrological models are routinely used to meet various objectives, such as flood forecasting and drought anticipation, which require streamflow estimates over a range of spatial and temporal scales (Blöschl and Sivapalan, 1995). For the temporal resolution, many flood forecasting models currently operate at sub-daily time steps (e.g. Marty et al., 2013; Wetterhall et al., 2011). The requirement for short time steps, hourly or even sub-hourly, stems from the need to adequately represent the characteristics of a flood (Obled et al., 2009). In fact, the model time step operates as a filter for the high-frequency dynamics of streamflow and for all the hydrological processes represented, as the sampling interval of observations does. Thus, ideally, the end-users should be able to choose the model's temporal resolution according to the level of detail required for a specific application and the characteristic response time of the basin and flood events studied.

However, most hydrological models were traditionally designed at a fixed time step and most operational systems still choose the model to be used by considering the objectives and 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 a few minutes), new techniques are needed to attain hydrological models able to work efficiently and consistently at different resolutions, up to sub-hourly. The transposability of model equations, state variables and parameters across time scales is a recognized challenge for all types of hydrological models, from empirical to physically-based (Blöschl and Sivapalan, 1995; Morel-Seytoux, 1988). The physically-based models rely on equations which are often appropriate only at fine spatio-temporal scales, generally smaller than the available data time interval. 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). For example, any rate-limited model for infiltration (e.g. Horton's model) is a meaningful description of the hydrologic response to a storm only when the temporal pattern of precipitation is sufficiently fine (< 1 h, in general), because of the sub-hourly characteristic time scales of infiltration and

* Corresponding author at: University of Reading, Department of Geography and Environmental Science, Whiteknights, RG6 6AB Reading, UK.
E-mail address: a.ficchi@reading.ac.uk (A. Ficchì).

https://doi.org/10.1016/j.jhydrol.2019.05.084
Received 10 July 2018; Received in revised form 16 April 2019; Accepted 27 May 2019
Available online 01 June 2019
0022-1694/

its highly nonlinear nature (Kandel et al., 2005; Morel-Seytoux, 1988). The empirical and conceptual rainfall-runoff models are often developed by following a top-down (or downward) approach (Klemeš, 1983; Sivapalan et al., 2003) seeking for progressively more complex model structures at finer time scales (e.g. Atkinson et al., 2002; Farmer et al., 2003; Gupta et al., 2012; Jothityangkoon et al., 2001). The potential benefits of refining the model resolution may be jeopardised by the structural deficiencies that may emerge with the need for increasing model complexity. Therefore, a trade-off may exist between model parsimony (or robustness) and the potential increase in model accuracy from refined resolutions. Various strategies have been proposed to counterbalance these factors, while the choice of the most parsimonious model structure possible is traditionally advocated (e.g. Gupta et al., 2012; Obled et al., 2009; Wagener et al., 2001). From the traditional top-down perspective, "model complexity is systematically increased in response to demonstrated deficiencies in model predictions until acceptable accuracy is achieved" (Atkinson et al., 2002). Other authors have suggested alternative pathways that can justify increased model complexity to provide more "realistic" representations of hydrological processes with consequent increased predictive power: either by extracting more information from available streamflow data, for example through different hydrological signatures and "prior constraints" from expert knowledge (e.g. Clark et al., 2011; Euser et al., 2013; Gharari et al., 2014; Hrachowitz et al., 2014) or by using auxiliary data sources to help identify the processes to model (e.g. Fenicia et al., 2008a; Son and Sivapalan, 2007; Winsemius et al., 2008). Along the lines of these previous studies, multi-time–step models can benefit from extracting the information available from multi-resolution data. In this study, we will propose an alternative model-development approach based on prior constraints on the transposability of the modelled fluxes across model time steps.

1.2. The need for adapting model structures across time steps

As illustrated by Morel-Seytoux (1988) for different hydrological processes, the 'successful' scaling of models requires an adaptation of their physical description, implying 'enlightened simplification' and multiple integrations (in space, time and in terms of expectations and processes). In catchment hydrology, meaningful model conceptualisations should change across spatio-temporal scales (Klemeš, 1983), as "different causal processes operate at different scales" (Blöschl, 2006). Some model components needed at one level of scale might become unnecessary (or ill suited) to operate at other levels, while other functions might emerge or need to be enhanced. For this reason, a multi-scale consistent modelling paradigm should be advocated in hydrology as a comprehensive representation of physical processes at different scale levels, where equations can be changed while verifying their coherence (consistency) across scales. This unifying modelling paradigm across scales would be analogous to the "flexible modelling approach" proposed (e.g. Fenicia et al., 2011) for a synthesis across hydrological processes. Some authors adopting the flexible modelling approach have focused on different aspects of model "consistency" to identify whether a model structure is suitable for a certain catchment based on multiple hydrological signatures (e.g. Euser et al., 2013; Hrachowitz et al., 2014). In our study we will use a different definition of consistency to identify a coherent model functioning across temporal resolutions.

In the field of conceptual and empirical models, an increasing literature has focused on the impact of time steps on model parameters and performance (e.g. Bastola and Murphy, 2013; Cullmann et al., 2006; Finnerty et al., 1997; Ishidaira et al., 2003; Kavetski et al., 2011; Littlewood and Croke, 2008; Littlewood et al., 2011; Melsen et al., 2016; Ostrowski et al., 2010; Wang et al., 2009). In the field of physically based models, rare cases of consistent multi-time–step models can be found, such as the AFFDEF model (Moretti and Montanari, 2007). Another approach is to use a probability-based model (Kandel et al., 2005) consisting of a single model structure working at different resolutions by means of a statistical distribution of the fluxes within the time step supported by the available data.

There is still no 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). The main hydrological processes expected to show emerging importance in modelling passing from daily to sub-hourly time scales are interception and infiltration-excess overland-flow (e.g. Blöschl and Sivapalan, 1995; Fenicia et al., 2008b; Gerrits et al., 2010; Savenije, 2004). Interception is the first active process in the rainfall-runoff transformation and refers to the storage of part of the rainfall above the ground surface, mostly in vegetation in a natural environment, and the evaporation from this storage that generally occurs during the rainfall event and shortly after. Its main effect is filtering the sub-daily rainfall variability and smoothing the peaks up to a certain threshold given by the interception capacity (i.e. storage). The magnitude of interception loss is dependent on many factors as climate, seasonality and vegetation, but in general, in temperate climates, the annual interception loss may represent a significant portion (> 10%) of the total evaporation and precipitation (Calder, 1990; Gerrits, 2010; Savenije, 2004). The infiltration-excess process also acts typically at sub-hourly time scales and is highly nonlinear (e.g. Kandel et al., 2005). Its main effect is, similarly to interception, smoothing out the catchment response to rainfall up to a certain threshold given by the infiltration rate capacity.

Since some hydrological processes are masked by the time step, it seems intuitive that the model conceptualization requires a structure of increasing complexity as the time step decreases. This has been confirmed in the literature by several authors, for example by Ye et al. (1997), Jothityangkoon et al. (2001), Atkinson et al. (2002) and Kavetski et al. (2011). Most of these studies focused on the range from annual to daily time scales, while there are fewer studies dealing with changes in model structures from (sub-)daily to (sub-)hourly scales (Jeong et al., 2010; Kavetski et al., 2011).

The GR chain of models is an example of models developed specifically for different time steps, from annual (GR1A) to monthly (GR2M), daily (GR4J, GR5J) and hourly (GR4H, GR5H), with some structural similarities and increasing complexity for decreasing time steps (Mouelhi et al., 2006). On the other hand, when passing from the daily to hourly time step, an increase of the complexity of the GR model structure has unexpectedly not been validated in previous studies (Le Moine, 2008; Mathevet, 2005). For example, Mathevet (2005) tested the addition of an interception store in the hourly model, which was judged not necessary for general model performance criteria focusing on regime and average accuracy (over a multi-year simulation period). However, there is an acknowledged need for further investigations to evaluate the optimality of the GR4 model structure at multiple sub-daily time steps (Ficchì et al., 2016; Le Moine, 2008; Mathevet, 2005). In particular, Ficchì et al. (2016) identified a degradation of performance for a consistent number of catchments as the time step decreases from daily to sub-hourly, and a general worsening of the underestimation of flood volumes at sub-daily time steps. This problem suggests a possible limitation of the model structure causing overestimated water losses and/or underestimated gains at sub-daily time steps compared to the daily model.

1.3. The need for flux-matching across spatio-temporal scales

Any physical model (e.g. hydrological, atmospheric or land-surface) can be transferred across scales and "the search for an appropriate level of conceptualization can proceed either upward or downward along the hierarchy of scales" (Klemeš, 1983). These two scaling routes are generally referred to as upscaling or downscaling, and involve, respectively, the problem of aggregating or disaggregating models and measurements, including inputs and descriptors (Blöschl and Sivapalan, 1995; Klemeš, 1983; Raupach and Finnigan, 1995). As the temporal and spatial resolutions change, the mass and energy conservation principles should be verified, by averaging land–air fluxes from the finer to the coarser scales. This truism means that the simulations of a consistent hydrological model at different time steps (or spatial discretisations) should lead to mass conservation of water fluxes, i.e. flux matching, to be physically consistent and potentially meaningful (Raupach and Finnigan, 1995), also when its parameters are different (e.g. recalibrated). The flux-matching condition across scales should be used as a constraint for any physically based or conceptual model, as already recognised by some authors, focusing on flux-matching tests across spatial scales, especially in the climate and land-surface modelling community (Rakovec et al., 2016; Raupach and Finnigan, 1995; Samaniego et al., 2010; Samaniego et al., 2017; Wood, 1997). This flux-matching condition across scales has been proved to be consistent with the fact that process parameterizations (i.e. equations) often depend on the resolution and parameters are re-calibrated at different scales.

In contrast, in the hydrological literature, few studies have evaluated the consistency of the internal functioning of a model at multiple time steps. To our knowledge, only Haddeland et al. (2006) analysed the coherence of simulated moisture fluxes at different time steps on the Variable Infiltration Capacity (VIC) model. These authors showed that the VIC model simulates inconsistent fluxes, when runs are performed at different sub-daily time steps (1-, 3-hourly and daily), with the same parameters for different model modes (i.e. energy balance closure assumptions) and time steps. They showed that the differences in fluxes are mainly a result of the parameterization of canopy evaporation from intercepted water at different time steps. The possibility of intercepting the whole current precipitation at the daily time step leads to increased intercepted volumes at daily rather than at sub-daily time steps. They proposed a scheme to reconcile these flux differences, based on simple parameter adjustments *to match simulated moisture fluxes* from runs with shorter time steps to the daily water balance simulation results. Their study constitutes a first promising result promoting the use of flux-matching criteria across time steps for hydrological model identification. Consistent models must be able to reproduce the same cumulated fluxes over time (e.g. over each validation period) when run at different resolutions, which is a necessary but not sufficient condition for any type of model to be hydrologically meaningful.

1.4. Summary of the research questions and scope of the article

Despite the recognised importance of having consistent hydrological models for multiple time steps, there is a lack of research on the development of coherent models across time steps, especially from sub-hourly to daily. Further research should consider the use of flux-matching criteria across time scales to constrain not only model parameterisation, as done by Haddeland et al. (2006), but also model structures. Flux-matching criteria across time steps could be combined with usual criteria on model performance to find new routes for model development, similarly to what was suggested by other authors using hydrological signatures (e.g. Euser et al., 2013; Hrachowitz et al., 2014). The present study explores this direction to develop a general methodology for the identification of a multi-time–step hydrological model based on the temporal consistency of the simulated fluxes. Our tests are presented for the downscaling of a lumped rainfall-runoff model (GR4J, Perrin et al., 2003) from daily to sub-hourly time steps, which is particularly important for flood forecasting applications. The use of a large sample of 240 catchments will provide general conclusions (Andréassian et al., 2009). Based on the modelling results, we will answer two specific research questions: (i) Does flux-matching improve parameter consistency? and (ii) does flux-matching improve model performance?

In Section 2, we present the hydro-climatic data, the catchments and flood events set, the baseline GR4 rainfall-runoff model and the basic calibration-evaluation procedure. In Section 3, we propose a new methodology for model identification by considering both model accuracy and flux-consistency across time steps, and we present the application to the GR4 model. Section 4 presents and discusses the results of the model diagnosis showing the symptoms of the baseline model at different time steps. In Section 5 the proposed structural modifications are presented, with detailed results for the new model structure retained in terms of both model performance and internal coherence across time steps. Section 6 provides concluding remarks and a discussion on the limitations of this work and the perspectives for further research.

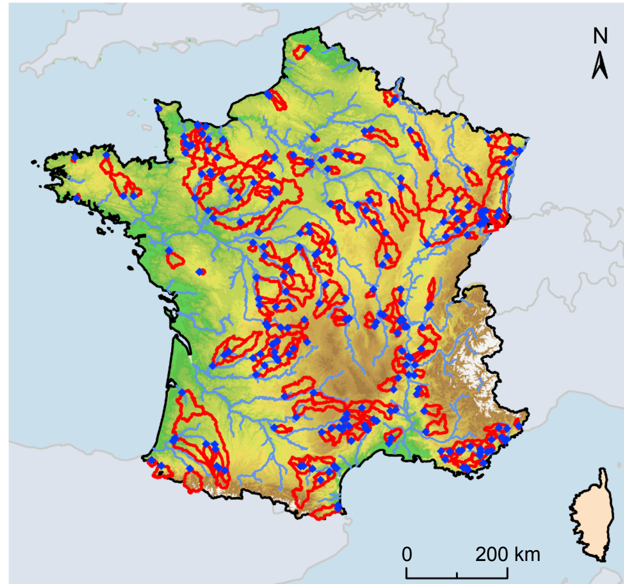

Fig. 1. Set of 240 French catchments (red lines: catchment divides; blue dots: gauging stations at catchment outlet; light blue lines: rivers), from Ficchì et al. (2016). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2. Materials and methods

2.1. Data, catchment and flood event sets

We used the database of 240 French catchments (Fig. 1) set up and described by Ficchì et al. (2016). Here we briefly review the main features of this data set (see Ficchì et al. (2016) for further details). The input data for the GR4 lumped model are precipitation and potential evapotranspiration (estimated from temperature data), spatially aggregated at the catchment scale. We used the SAFRAN countrywide climate reanalysis (Vidal et al., 2010) at the daily time step and 8-km resolution, to calculate the daily precipitation and temperature over each catchment. The daily data were downscaled to sub-daily time scales while respecting the mass and energy conservation principles, by ensuring the consistency between input amounts at different time steps. The daily precipitation amounts were distributed at sub-daily time steps (up to 6 min) using ground observations at the 6-min time step for 1405 automatic rain gauges in France. First, the 6-min rain-gauge data were spatially aggregated by catchment (using the Thiessen polygon method); then the sub-daily temporal pattern, obtained as the distribution function of the areal 6-min rainfall estimate over each day, was used to disaggregate the catchment daily precipitation estimated from the SAFRAN reanalysis at higher spatial-resolution (see Ficchì et al. (2016) for further details). Potential evapotranspiration (E) was estimated using the temperature-based formula proposed by Oudin et al. (2005) at the daily time step and then distributed over each day using a parabolic empirical formula, reaching its maximum from 12:00 p.m. to 1:00 p.m. (UTC), with null values between 7:00 p.m. and 6:00 a.m.

According to the availability of the shortest-length data set (rain

Table 1
Summary of the distribution of several morphological and hydro-climatic characteristics of the 240 test catchments.

Catchment characteristics	Definition/reference	Min	1st Quartile	Median	3rd Quartile	Max
Catchment area [km^2]	–	4	163	356	770	8790
Annual precipitation (P) [mm]	Mean inter-annual P	651	820	940	1110	2108
Annual potential evapotranspiration (E) [mm]	Mean inter-annual E	594	665	712	763	1129
Runoff coefficient [–]	Ratio of mean annual streamflow to P	0.04	0.26	0.34	0.46	1.69
Aridity index [–]	Ratio of mean annual E to P	0.29	0.64	0.76	0.91	1.53

gauges), we selected the 8-year period from 01/08/2005 to 31/07/2013 to test the model. For calibration and evaluation, we used streamflow observations from the *Banque Hydro* database (Leleu et al., 2014). Streamflow time series at different fixed time steps (from 6 min to 1 day) were generated by linear interpolation and temporal integration of original hydrometric data at variable (sub-hourly) sampling times.

The 240 unimpaired catchments were selected to obtain no or limited snow influence (< 10% solid precipitation), because we wished to use GR4 without its snow module. Some criteria on data quality were also tested to satisfy a minimum rain gauge density (for the 6-min network) and a low gap fraction in the streamflow time series. The 240 test catchments (Fig. 1) represent the variety of hydroclimatic regimes in metropolitan France, ranging from oceanic, Mediterranean to continental. The morphological and hydro-climatic characteristics of the 240 catchments are varied (see Table 1 here, and Table 2 in Ficchì et al. (2016)).

Since short time steps are mostly useful for hydrological events with fast dynamics, i.e. floods, the model was evaluated over the whole time series and over flood events only. Ten flood events were selected for each catchment over the 8-year test period (2005–2013) using an automated procedure (Ficchì et al., 2016), providing a set of 2400 events. The characteristics of the flood events are also varied, with durations from less than 1 day to several days and response times from 1 h to more than 3 days, as summarised by Ficchì et al. (2016) (Table 3 therein).

2.2. The baseline rainfall-runoff model (GR4) and a modified version (GR5)

2.2.1. Previous tests of the GR4 model at sub-daily time steps

The baseline model chosen in this study is the GR4 model (Perrin et al., 2003), a lumped bucket-type rainfall-runoff model with four free parameters, known as GR4J in its daily original version (where J stands for *journalier*, i.e. daily). Previous work on the development of this model and adaptation at the hourly time step (Le Moine, 2008; Mathevet, 2005) mainly focused on model performance while keeping a parsimonious structure: on the basis of model accuracy at the hourly time step, these studies concluded that an increase of the complexity of the daily model was not necessary and it was sufficient to adapt some fixed and calibrated parameters. Also, other studies have reported the use of the GR4 hourly model (GR4H) with satisfactory results when compared to other hourly models (de Boer-Euser et al., 2017; Van Esse et al., 2013). The analogous adaptation of the model at other sub-daily time steps used in our previous study (Ficchì et al., 2016) confirmed a good level of performance from 6-min to daily time steps on our catchment set. However, we also suggested the need for some refinements of the model structure at shorter time steps, especially for solving the worsening of the flood volume underestimation.

Table 2
Median values of the three water–balance-related parameters of GR4, calibrated at the daily and 6-min time steps, and median relative changes at time steps from 6 min to 12 h with respect to the daily value, over the 240 catchments.

GR4 (water-balance-related) parameter	Median parameter (*normalised at 1 h ref.*)		Median relative change ($\Delta_{\text{rel.}}$) of parameter calibrated at time step Δt, with respect to the daily reference; $\Delta_{rel} = 100\cdot[x_i(\Delta t) - x_i(1d)]/\lvert x_i(1d)\rvert$ [%] where $x_i(\Delta t)$ is the *i-th* parameter calibrated at t.s. Δt						
	1 d	6 m	6 m	12 m	30 m	1 h	3 h	6 h	12 h
x_1 [mm]	264	232	−10	−10	−9	−8	−7	−7	−5
x_2 [mm/h]	−0.58	−1.07	−59	−57	−53	−46	−41	−33	−24
x_3 [mm]	114	106	−6	−6	−5	−7	−6	−5	−2

2.2.2. GR4 baseline model structure and parameters

GR4 is composed of two main parts: (i) a production part, modelling the water balance, made up of an interception function, a soil-moisture accounting store and a groundwater exchange function; and (ii) a routing part, composed of unit hydrographs and a non-linear store (see Fig. 2). A comprehensive description of the equations of the GR4 model can be found in Perrin et al. (2003), and an overview is provided hereafter, given the importance of defining the fluxes that will be considered in the model diagnosis. The model has four free parameters to be optimised (see Fig. 2) and fixed parameters, whose values were set by Perrin et al. (2003).

2.2.3. GR4 baseline model equations

Here we present the GR4 model, with the equations for the fluxes determining the water balance (other fluxes representing transfers within the catchment are reported in Appendix A). Note that, in the paper, all quantities marked with capital letters refer to amounts cumulated over a time step ([mm]), e.g. P and E, are the precipitation and potential evapotranspiration amounts, while small letters, e.g. p and e, are the corresponding rates ([mm/time]).

The first operation is the determination of an interception loss, i.e. the amount of evaporation from intercepted water (E_i), by a neutralisation function (Eq. (1)) of the precipitation (P) by the potential evapotranspiration (E). P and E are consequently reduced to a net rainfall (P_n) and a net evapotranspiration amount (E_n), respectively:

$$E_i = \min(P, E) \tag{1}$$

$$P_n = P - E_i \tag{2}$$

$$E_n = E - E_i \tag{3}$$

The actual evaporation from the production store (AE_s) is a function of the net evapotranspiration (E_n):

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

where S [mm] is the level in the production store, and x_1 [mm] is its capacity. The amount of water infiltrating in the soil-moisture accounting store (P_s) is analogously determined as a function of the net rainfall (see Eq. (A.1), Appendix A). The production store level S is updated ($S + P_s - AE_s$), and then a percolation leakage Perc is removed from the store (Eq. (A.2)). The part of the net rainfall not infiltrating in the production store ($P_n - P_s$) reaches the routing part of the model directly, after being added to the percolation (*Perc*).

The routing function consists of linear routing with two unit hydrographs (*UH*) and a non-linear routing store. The total amount of

Table 3
Median values of the three water–balance-related parameters of the GR4-I model, calibrated at daily and 6-min time steps, their coefficient of variation and the median relative changes at time steps from 6 min to 12 h with respect to the daily value, over the 240 catchments.

GR4-I (water-balance-related) parameter	Median parameter (*converted to 1-h ref. values*)		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. $\Delta_{rel} = 100\bullet[x_i(\Delta t) - x_i(1d)]/\|x_i(1d)\|$ [%] 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

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 a non-linear routing store, and 10% is routed only by another unit hydrograph (UH2). Both UHs 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.

A groundwater exchange loss (or gain) is released from (or added to) both flow components, representing fluxes to (from) the outside of the basin. The potential half exchange, F, is calculated as:

$$F = x_2\left(\frac{R}{x_3}\right)^{3.5}\bullet\Delta t \quad (5)$$

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 (one time-step ahead). 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. Actual groundwater gains (F_G) in case x_2 is positive are double F (Eq. (6)), while actual losses F_L in case x_2 is negative are limited by the water available in the routing store and in the flow components from the unit hydrographs (Eq. (7)).

$$F_G = 2F \quad (6)$$

$$F_L = -[\min(|F|, R + Q_9\Delta t) + \min(|F|, Q_1\Delta t)] \quad (7)$$

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 groundwater exchange component F. The outflow from the routing store gives the first flow component Q_r (Eq. (A.3)). The output of the second unit hydrograph UH2 provides the direct flow component Q_d, after being subject to the same exchange flux: $Q_d = \max(0; Q1 + F)$. The total streamflow is finally given by the sum of the two flow components ($Q_r + Q_d$).

2.2.4. GR5 model variant

Le Moine (2008) proposed a modified model version, called the GR5 model (also GR5J or GR5H, respectively, at the daily and hourly time step), with a complexified exchange function. The GR5 version proved particularly useful for improving the simulation of low flows. The modified exchange function includes an additional free parameter (x_5 [−]), representing the level of the routing store for which the exchanged

Free parameter	Unit	Definition
x_1	mm	Maximum capacity of the production store
x_2	mm / time step	Groundwater exchange coefficient (positive for gains; negative for losses)
x_3	mm	Reference capacity of the routing store one time step ahead
x_4	time step	Time base of the unit-hydrograph

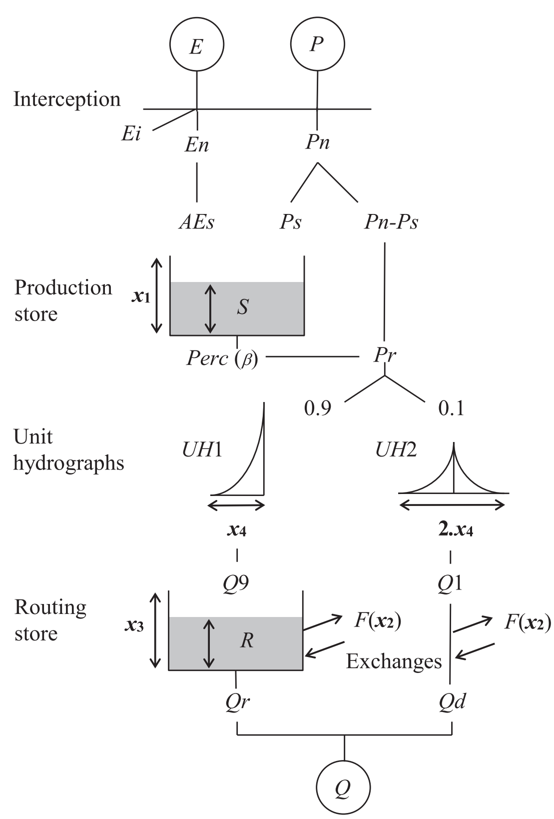

Fig. 2. Schematic representation of the GR4 baseline model structure (Perrin et al., 2003).

flux changes sign. As proposed by Le Moine (2008), the linear exchange function (replacing the power function in Eq. (5)) is:

$$F = x_2\left(\frac{R}{x_3} - x_5\right)\cdot\Delta t \tag{8}$$

where R is the level in the routing store [mm], x_3 its reference capacity [mm], x_5 [–] the threshold level for which the exchange function changes sign (with $x_5 \in [0, 1]$) and x_2 the water exchange coefficient [mm/time step].

2.2.5. Baseline multi-time–step adaptation

The baseline adaptation of the GR4 model from daily to sub-daily time steps is done by changing the parameters that theoretically depend on the time step and the UH exponent value. The UH exponent (see Eqs. (10), (13)–(14) in Perrin et al., 2003) was lowered from 2.5 to 1.25 passing from the daily to hourly time step, as proposed by Mathevet (2005). Ficchì et al. (2016) summarised all the time-step dependencies of the fixed and calibrated parameters of the GR4 model to be considered to run the model at any sub-daily time step (see Table A.1). The theoretical relationships of the fixed parameters with the time step are included directly in the model equations, while the corresponding relationships for the free parameters are verified after calibrating the model, to control artificial bias in applying the model at different time steps.

An implementation of the GR4 model at daily and hourly time steps is freely available in the airGR R package (Coron et al., 2017a,b). The GR4 daily model version by Coron et al. (2017a) was used as the baseline version tested here, after implementing code adaptations to run the model at multiple time steps. We will use a set of eight time steps ensuring a sufficient continuity in the temporal sub-daily scale: 6, 12, 30 min, 1, 3, 6, 12 h and 1 day.

2.3. Calibration and evaluation of model accuracy

The available 8-year period was split into two 4-year sub-periods (2005–2009 and 2009–2013) to apply a calibration-validation procedure. A 2-year warm-up period before each test sub-period (01/08/2003–31/07/2005, for which SAFRAN data were available, and 01/08/2007–31/07/2009) was used and judged sufficiently long to initialise the model states. A split-sample test (Klemeš, 1986) was applied by calibrating the model on each sub-period and validating on the other one. As the objective function, we used the Kling–Gupta efficiency (KGE) criterion (Gupta et al., 2009), calculated on the modelled streamflow over the whole calibration sub-period. The KGE criterion is calculated as:

$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2} \tag{9}$$

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; 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 evaluation was performed on both the whole validation sub-periods and on the set of selected flood events for each catchment. For significance of the statistics calculated on the floods, the criteria were calculated on the series of ten events together for each catchment. The periods of the shortest selected events were adjusted to last at least 3 days, thus leading to a validation period over floods of at least 30 days for each catchment.

The differences in performance between modelling tests were evaluated by means of:

(i) the KGE criterion and its components, evaluated both on the whole time series and on flood events only;

(ii) certain criteria on the flow duration curve (FDC): ratios of extreme quantiles of simulated and observed flows, and the percent bias in the slope of the mid-segment of the FDC [%] between 0.2 and 0.7 flow exceedance probabilities (Yilmaz et al., 2008).

The Friedman rank test (Friedman, 1937) was used to detect statistically significant changes in mean ranks of performance across alternative modelling options over the catchment set (at significance level 0.05).

3. A new approach for model diagnosis and improvement

3.1. Proposed approach based on flux matching at multiple time steps

The rationale underlying the proposed approach is that a consistent rainfall-runoff model should simulate accurate and coherent outputs and the same internal fluxes at multiple time steps (see Section 1.3). We propose a two-step approach for empirical identification and improvement of a consistent multi-time step model, which complements the evaluation of model performance with the consistency of fluxes and parameters across scales. This approach can be interpreted as a model diagnosis, i.e. "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" (Yilmaz et al., 2008). This type of process can identify model structure inadequacies to be addressed by targeted changes in a trial-and-error iterative procedure.

To this aim, we applied the diagnostic and prognostic approach summarised in Fig. 3. The model is calibrated and run at different time steps using inputs and simulating internal fluxes and outputs at the same time step. For each model run, the model time step is fixed over the whole (8-year) simulation period, and equal to the resolution of forcing data. Then two evaluations are made: (i) model fluxes are cumulated in time and their consistency across simulation time steps is evaluated, (ii) the accuracy of the outputs, i.e. streamflow, is evaluated by comparing simulated streamflow and observations at the same time step as the model equations.

Based on the outcomes of the model diagnosis (Section 4), targeted structural changes can be proposed to improve both model consistency across time scales and output accuracy, or either one of these two desired features (Section 5). A structural change is rejected if it does not improve any of these two aspects or leads to one of them degrading. Note that the proposed changes in structure may target one (or more) specific time step(s) and may be a function of the time step (see Fig. 3), because different equations might be needed as the time step changes. Following our research questions (see Section 1.4), this approach will allow to evaluate whether the modified model structures targeting fluxes consistency provide improvements in: (i) the coherence of model parameters and (ii) model performance across time steps.

3.2. Application to the GR4 model

The flux-matching condition across time steps will be verified for the water-balance fluxes simulated by the model and cumulated over time. Following the equations in Section 2.2, the catchment water-balance equation of GR4 can be written as:

$$\boldsymbol{P} + \boldsymbol{F_G} = \boldsymbol{Q} + \boldsymbol{E_i} + \boldsymbol{AE_s} - \boldsymbol{F_L} + \Delta S \tag{10}$$

where the terms marked in bold are fluxes cumulated over the analysis period, i.e. the 8-year validation period (2005–2013) used in our study. The variation of the amount of water in the production store (ΔS) may contribute to the water balance equation but is negligible when cumulated over long periods (e.g. annual). The (total) net outflux to the atmosphere and groundwater is the sum of the internal fluxes of GR4 ($\boldsymbol{E_i} + \boldsymbol{AE_s} + |\boldsymbol{F_L}| - \boldsymbol{F_G}$), defined as:

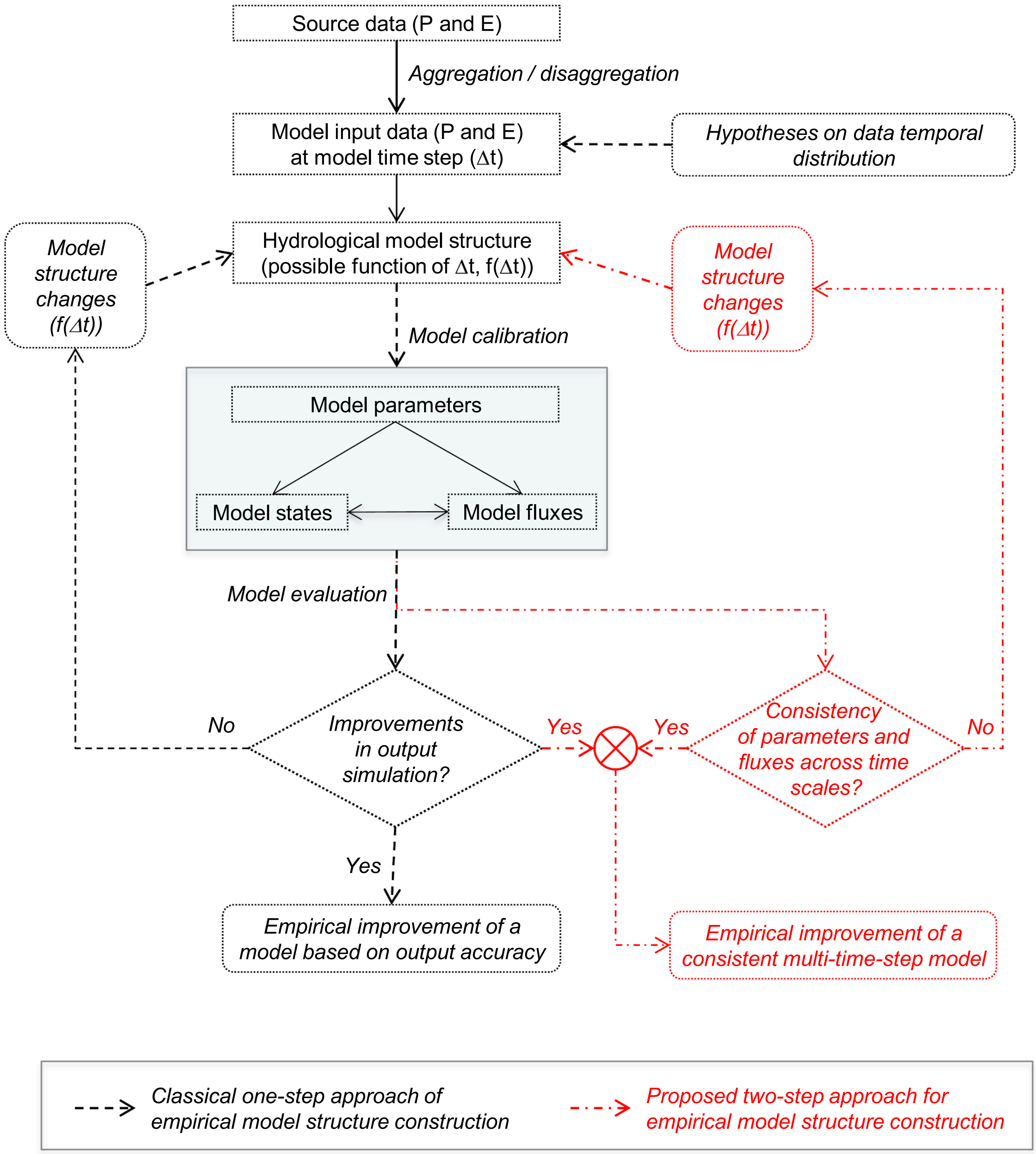

Fig. 3. Schematic representation of the proposed methodology for model diagnosis and identification of improvements at multiple time steps.

a) *Interception loss* ($\boldsymbol{E_i}$[mm]), representing the evaporation from intercepted water, which occurs at a potential rate during a rainfall event and shortly after (e.g. Gerrits, 2010). In the GR4 baseline model, the interception loss is modelled as a neutralisation function between precipitation, P, and potential evapotranspiration, E, at each time step, as in Eqs. (1–3).
b) *Actual evaporation from the production store* ($\boldsymbol{AE_s}$ [mm]), corresponding to the amount of water evaporated from the soil-moisture accounting store under the effect of the net evapotranspiration E_n, calculated as in Eq. (4).
c) *Exchanged fluxes or inter-catchment groundwater flows* ($\boldsymbol{F_L}$ and $\boldsymbol{F_G}$ [mm]), i.e. water losses ($\boldsymbol{F_L}$, negative) or gains ($\boldsymbol{F_G}$, positive) for the catchment, according to the sign of the calibrated x_2 parameter (Eqs. (5–7)).

These fluxes are difficult to measure at the catchment scale and are obtained by calibration on only the most easily measurable flux, i.e. streamflow at the catchment outlet. We suspect that accurate simulations of the model output at different time steps can be obtained by inconsistent internal model functioning, because of possible compensations between the simulated fluxes (i.e. equifinality). However, the cumulated fluxes over long periods (longer than 1 day) should be consistent for multiple modelling time steps to comply with the mass conservation equation. We analyse the consistency of the cumulated fluxes obtained by simulations at the tested time steps (6, 12, 30 min, 1, 3, 6, 12 h and 1 day), over the whole validation period of 8 years and over the selected flood events. Since a reference is needed to evaluate the flux consistency across multiple time steps, we would ideally need observations of these fluxes at the catchment scale. However, such

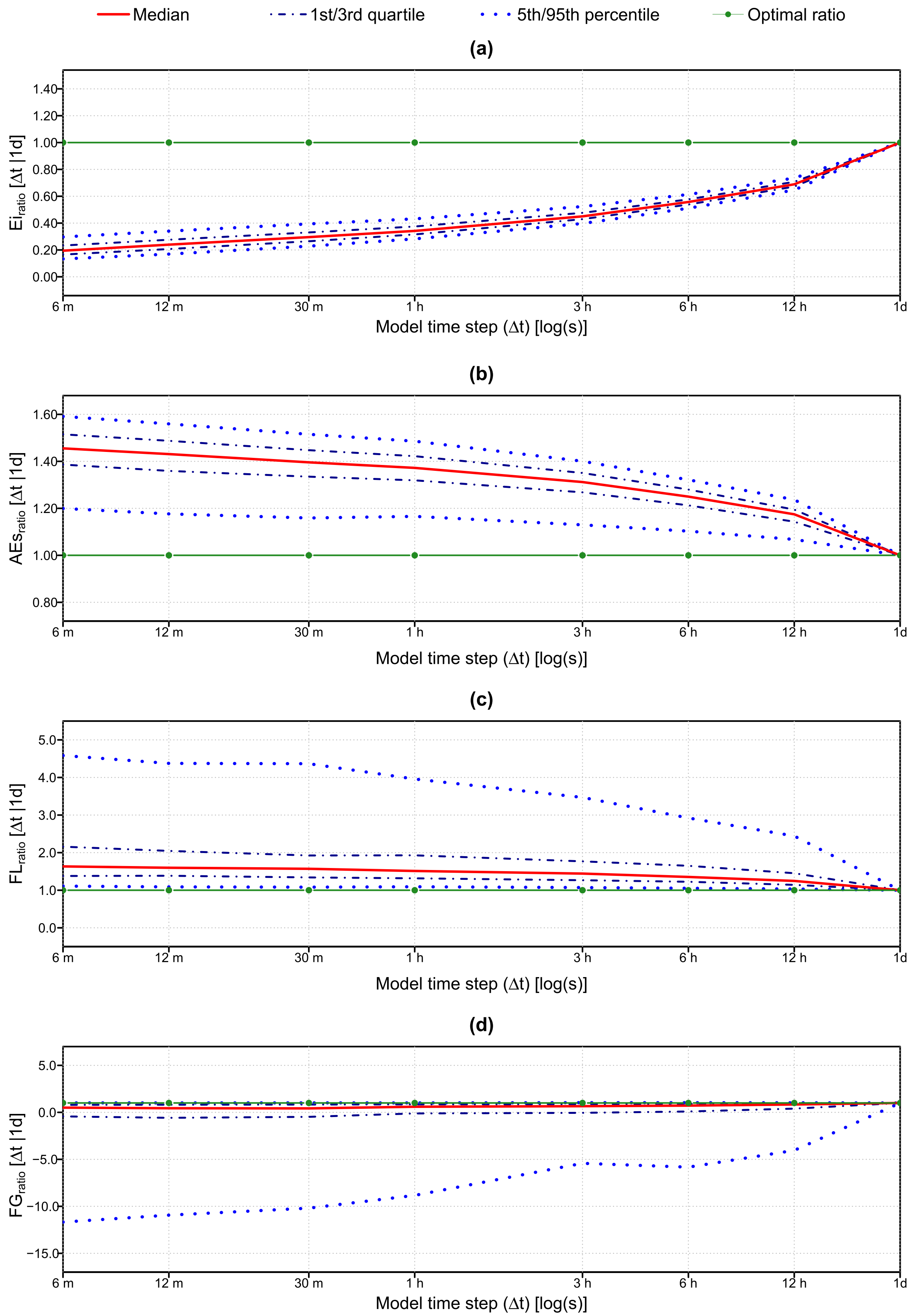

Fig. 4. Summary of the cumulative flux ratios of the GR4 baseline model at different time steps (with daily reference) over the 8-year validation period and the 240-catchment set, for (a) interception loss, *Ei*; (b) actual evaporation from production store, AE_s; (c) groundwater losses, F_L; (d) groundwater gains, F_G.

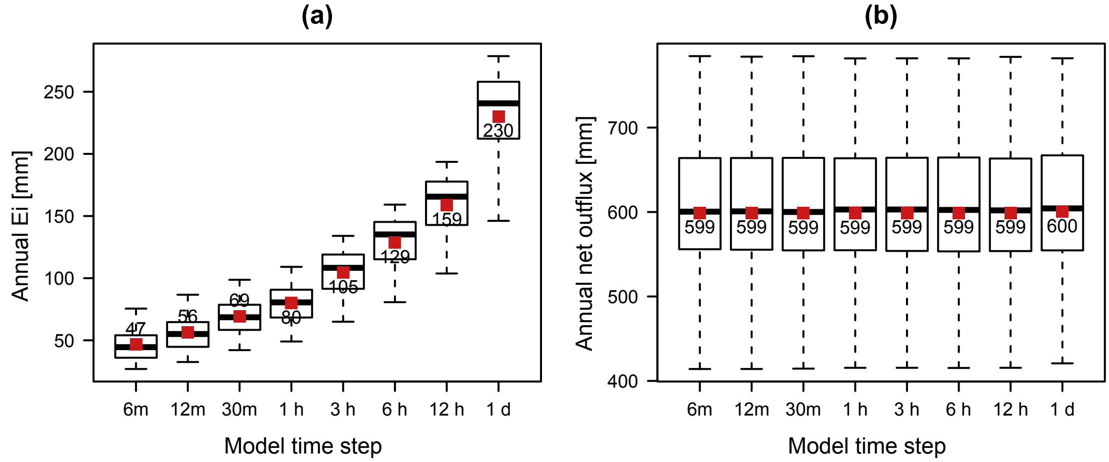

Fig. 5. Distribution over the catchment set of annual average fluxes, over the 8-year validation period, from the GR4 model simulations at different time steps: (a) interception loss, *Ei*; (b) net outflux ($E_i + AE_s + |F_L| - F_G$). The boxplots report the median value, interquartile range, and the whiskers represent the 5th and 95th percentiles; the red dots refer to mean values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

measurements are not available over our catchment set. Therefore, we considered the daily model as the reference in terms of simulated fluxes, because it was shown that the catchment water balance is better reproduced by the daily GR4 model than the sub-daily versions of the model (Ficchì et al., 2016). For each time step Δt tested, from 6 min to 1 day, we therefore calculated the ratio of the cumulated flux simulated at time step Δt normalised on the corresponding reference flux simulated at the daily time step. These ratios, called the cumulative flux ratios, are non-dimensional indexes, evaluated for each catchment, by the following equations:

$$Ei_{ratio}[\Delta t|1\ d] = \frac{\boldsymbol{E_i}[\Delta t]}{\boldsymbol{E_i}[1\ d]} \quad [--] \tag{11}$$

$$AE_{s\ ratio}[\Delta t|1\ d] = \frac{\boldsymbol{AE_s}[\Delta t]}{\boldsymbol{AE_s}[1\ d]} \quad [--] \tag{12}$$

$$FL_{ratio}[\Delta t|1\ d] = \frac{\boldsymbol{F_{L/G}}[\Delta t]}{\boldsymbol{F_L}[1\ d]} \quad [--] \tag{13}$$

$$FG_{ratio}[\Delta t|1\ d] = \frac{\boldsymbol{F_{L/G}}[\Delta t]}{\boldsymbol{F_G}[1\ d]} \quad [--] \tag{14}$$

where $\boldsymbol{E_i}[\Delta t]$, $\boldsymbol{AE_s}[\Delta t]$, and $\boldsymbol{F_{L/G}}[\Delta t]$ represent the cumulated interception loss, actual evaporation and inter-catchment exchanges (losses and gains) [mm] of the GR4 model at the sub-daily time step Δt with $\Delta t \in (6, 12, 30\text{min}, 1, 3, 6, 12\text{h})$; $\boldsymbol{E_i}[1d]$, $\boldsymbol{AE_s}[1d]$, $\boldsymbol{F_L}[1d]$ and $\boldsymbol{F_G}[1d]$ represent the cumulated interception loss, actual evaporation and groundwater exchanges (losses, $\boldsymbol{F_L}$, and gains, $\boldsymbol{F_G}$) [mm] of the GR4 model at the daily time step ($1d$). The optimal value of the cumulative flux ratios is 1, which corresponds to consistent modelled fluxes at different time steps.

The computation of the cumulative flux ratios for the exchanged fluxes, i.e. $FL_{ratio}[\Delta t|1d]$ and $FG_{ratio}[\Delta t|1d]$, is admissible on our catchment set, because the cumulated fluxes $\boldsymbol{F_L}[1d]$and$\boldsymbol{F_G}[1d]$ are not null for all catchments. Note that since the exchanged fluxes can be either positive or negative, the correspondent cumulative flux ratios were divided into the two cases of Eqs. (13) and (14) corresponding to:

- $\boldsymbol{F_L}[1d] < 0$, i.e. catchments presenting losses at the daily time step (197 catchments of our set).
- $\boldsymbol{F_G}[1d] > 0$, i.e. catchments presenting gains at the daily time step (43 catchments).

This allows one to univocally interpret the values of the ratios, because of the possible sign change of the numerator ($\boldsymbol{F_{L/G}}[\Delta t]$) with respect to the denominator. For example, for the interpretation of $FG_{ratio}[\Delta t|1d]$ (Eq. (14)), three cases are possible:

(i) $\text{FG}_{ratio}[\Delta t|1d] \geq 1$, i.e. gains are larger at the sub-daily time step (Δt) than at the daily or they are equal ($\text{FG}_{ratio} = 1$);
(ii) $0 < \text{FG}_{ratio}[\Delta t|1d] < 1$, i.e. gains are larger at the daily time step than at the sub-daily;
(iii) $\text{FG}_{ratio}[\Delta t|1d] < 0$, i.e. the exchanges are gains at the daily time step but become losses at the sub-daily time step *x*.

Three similar cases exist for the interpretation of Eq. (13).

4. Results and discussion on model diagnosis: the GR4 symptoms at sub-daily time steps

4.1. The undesirable model symptoms (1): flux inconsistencies

The outcomes of the model diagnosis are presented in two parts: the flux-matching condition is evaluated first over the entire 8-year validation period (Section 4.1.1) and then over the flood events selected (Section 4.1.2).

4.1.1. Flux-matching analysis over the whole validation period

Fig. 4(a–d) shows a summary of how the cumulative flux ratios evolve over the 8-year validation period as the time step changes, over the catchment set, highlighting the following clear trends:

a) The interception loss steadily decreases from daily to shorter time steps, reaching a median $I_{ratio}[6\text{min}|1\text{d}]$ of 19%, i.e. on average, the cumulated interception loss simulated by the 6-min model is less than one-fifth of the corresponding amount simulated by the daily model. The dispersion around the average ratios is small, indicating that the interception losses significantly decrease at shorter time steps for all 240 catchments. The decrease rate is the highest when passing from 1 d to 12 h. These results show that the neutralisation function (Eq. (1)) leads to increasingly less intercepted water as the time step decreases (see also Fig. 5(a)).
b) The actual evaporation from the production store increases as the time step decreases, reaching a median $AE_{sratio}[6\text{min}|1\text{d}]$ of about 145%. The trend in actual evaporation is opposite the trend for interception, with the highest increase rate when passing from 1 d to 12 h, suggesting a compensation between reduced interception loss

and increased actual evaporation.

c) The loss to groundwater (negative exchange F at the daily reference) increases as the time step decreases, reaching a median $\mathrm{FL}_{ratio}[6\mathrm{min}|1\mathrm{d}]$ of about 165%. The trend in groundwater losses is opposite the trend for interception, analogously to actual evaporation. So it seems that the model compensates the reduced interception loss at sub-daily time steps also by increasing groundwater loss.
d) The gain from groundwater (positive exchange F at the daily reference) decreases at shorter time steps, on average, reaching a median $\mathrm{FG}_{ratio}[6\mathrm{min}|1\mathrm{d}]$ of about 50%. Moreover, for 15 out of the 43 catchments with gains at the daily time step ($F_G[1d] > 0$), the gains become zero and even change sign as the time step Δt decreases ($F[\Delta t] < 0$). This means that for these catchments the groundwater gains at the daily time step become losses at shorter time steps, implying a sign change for the water exchange coefficient.

The large relative changes in cumulated fluxes must be further specified in absolute terms (i.e. mm), without normalisation by the daily reference, to outline the magnitude of the changes. Also, the possible changes in total net outflux must be evaluated ($\boldsymbol{E_i + AE_s + |F_L| - F_G}$). Fig. 5 reports the distribution of the annual interception loss and total net outflux for all the model time steps tested, over the catchment set. It shows that the strong decrease of annual interception loss at shorter model time steps is counterbalanced perfectly, providing a stable net outflux (around 600 mm/y, on average). This is obtained via the compensations between interception and other fluxes, i.e. actual evaporation and groundwater exchanges (Fig. 4(a–d)). In fact, in absolute terms, when passing from daily to 6-min time step, the median values over the catchment set for annual actual evaporation and groundwater losses increase from 311 and 72 mm/y, respectively, to 455 and 122 mm/y, counterbalancing the decrease in interception (241 to 45 mm/y) shown in Fig. 5(a).

The absolute changes of annual interception loss shown in Fig. 5(a) are very large (also compared to the total net outflux, see Fig. 5(b)). Although we do not have measurements of these fluxes for a physical validation of the model, it is still interesting to discuss roughly the order of magnitude of the annual interception loss simulated at different time steps, compared to annual precipitation and total evaporation (Table 1). From the literature it is expected that, in temperate climates, the annual interception loss may represent a significant portion of the total evaporation, and amount to up to 15–50% of precipitation (Calder, 1990; Gerrits, 2010; Savenije, 2004). Thus, the simulated annual interception loss at the daily time step is more consistent with this expectation (see Table 1 and Fig. 5(a)), accounting for a substantial proportion of both total evaporation ($E_i + AE_s$) and precipitation (on average about 40% and 25%, respectively). This proportion significantly decreases at shorter time steps, dropping unrealistically to less than 10% and 5% of total evaporation and precipitation, respectively, at the 6-min time step. This rough comparison tends to support our choice of using the daily modelled fluxes as a reference, which should be further validated if more data were available. It is also interesting to analyse further the dispersion of the changes of all cumulated fluxes around the average values over the catchment set. Fig. 6 shows the comparison between the annual 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 for all fluxes. It is clear that the current structure of the hourly GR4 model is not consistent with its daily counterpart for the whole catchment set.

Still the GR4 model manages to obtain a coherent water balance at the outlet for all catchments, except one outlier (Fig. 6(d)), given the compensations between interception loss, actual evaporation and groundwater exchanges. The outlier catchment is the Espérelle River at Roque-Sainte-Marguerite, in southern France, which is characterised by karst with significant losses to groundwater, determinant for its water balance.

At shorter time steps, the model components after the interception function receive more water and the actual evaporation from the production store and the groundwater losses are used to get rid of the excess water. This compensative mechanism might be more difficult to balance for karst leaking catchments, as the outlier suggests.

These internal compensations can be attributed (at least partially) to the inconsistent interception component (Eq. (1)), since it is the first operation in the chain of interlinked functions of the model.

4.1.2. Flux-matching analysis over flood events

The same trends of changes in fluxes simulated at different time steps are found on flood events. Fig. 7 shows a summary of the evolution of the daily average fluxes across time steps over the selected floods. In contrast to the case of the whole-period evaluation, over flood events, the changes in fluxes do not counterbalance, and the total net outflux increases as the time step decreases. This increase in net outflux can be attributed to the increasing losses from the underground exchanges, while interception loss and actual evapotranspiration counterbalance almost perfectly. Note that while, on average, the absolute changes of daily cumulated fluxes might seem small, they correspond to significant relative changes. For example, by passing from the daily simulation to the 6-min simulation, the daily net outflux increases by about 30%.

This increase of total net outflux at shorter time steps is counterproductive for model accuracy over flood events, given the general significant decrease of the ratio of means with shorter model time steps, already discussed in Ficchì et al. (2016). Thus, the internal compensations in the model can ensure a coherent water balance at all time steps only on average over the long term, but they lead to increasingly biased simulations over flood events.

4.2. The undesirable model symptoms (2): parameter inconsistencies

While in the GR4 structure the interception loss is independent of any model parameter, the other fluxes considered depend on three of the four calibrated parameters in GR4. Consequently, while the changes in interception loss across time steps are due to structural inadequacy, the changes in the other fluxes are driven by the calibration process. The actual evaporation from the production store depends on the capacity of this reservoir (x_1), as in Eq. (4), while the exchanges are related to the groundwater exchange coefficient (x_2) and to the capacity of the routing store (x_3, see Eq. (5)). Thus, the inconsistencies of model fluxes across time steps foretell spurious impacts on these three calibrated parameters.

Table 2 reports statistics describing the changes in the three water–balance-related parameters at the time steps tested. The median values of the parameters at the two extreme time steps, i.e. 1 day and 6 min, show large relative differences, especially for the x_2 parameter. This trend is confirmed at all time steps by the median values of the relative changes (see Δ_{rel} equation in Table 2) between the daily reference parameters and the corresponding calibrated values. On average, all three water-balance parameters decrease at sub-daily time steps with respect to the daily reference values, with a maximum relative change for the water exchange coefficient, x_2 (59% from daily to 6 min).

These changes were not desirable, but logically expected from the related flux inconsistencies. However, only the changes in the water exchange coefficient (x_2) can be directly explained by its role in the flux equation (Eq. (5)), while for the two production and routing store capacities (x_1 and x_3), their role is filtered via the store-filling rates and their contribution to other functions. The filling rates are found to be stable across time steps for both stores (not shown). Thus, only the most significant change of x_2 is linked to the spurious compensation of the reduced interception loss at shorter time steps. On the other hand, the

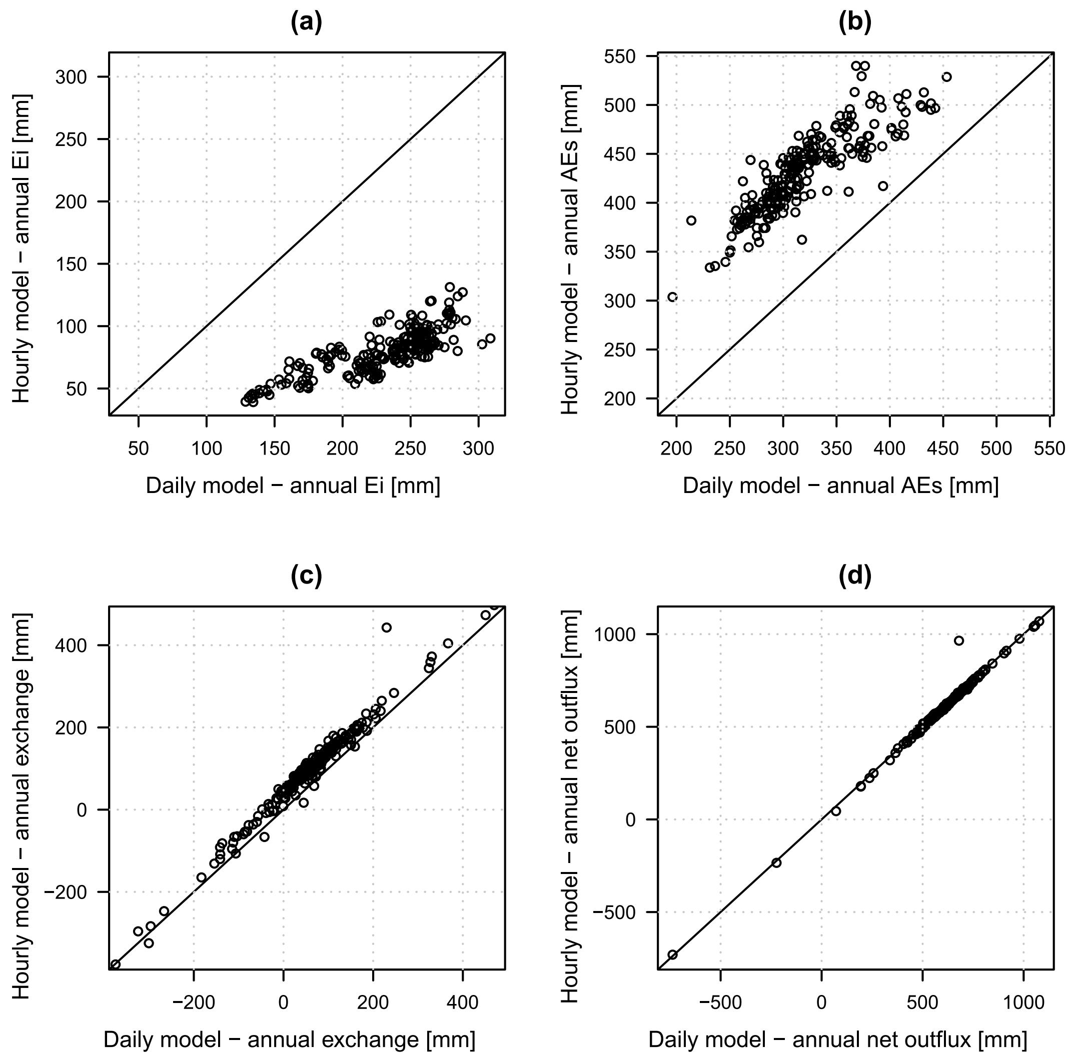

Fig. 6. Annual average cumulated fluxes from daily and hourly GR4 model simulations over the 8-year validation period for the 240 catchments (each point represents one catchment): (a) interception loss, Ei; (b) actual evaporation from production store, AE_s; (c) groundwater exchange losses (negative values represent gains, i.e. $-F_{L/G}$); (d) net outflux ($E_i + AE_s + |F_L| - F_G$).

increase in actual evaporation depends on the increasing amounts of net evaporation capacity at shorter time steps, as a direct consequence of reduced interception losses. These findings are in line with Savenije (2004), who suggested that "an error introduced in modelling interception automatically introduces errors in the calibration of the subsequent processes."

5. Results and discussion on model prognosis: the new GR5-I structure

5.1. A new structure with an interception store for sub-daily time steps

From the model diagnosis presented above, we concluded that it was a priority to stabilise the interception loss at sub-daily time steps to make all the model fluxes more consistent across time steps. This could be done using an interception store, i.e. extending the memory of the function used for interception to durations longer than the model time step. The neutralisation function, initially introduced for the daily GR4, allows intercepting water only when both precipitation and potential evapotranspiration are non-null, and this leads to temporally inconsistent fluxes, as proven empirically in Section 4. We therefore tested the insertion of an interception store at the top of the GR4 model structure, replacing the neutralisation function, as represented in Fig. 8. To distinguish this model structure from the baseline, we named it "GR4-I", or "GR5-I", the latter indicating the inclusion of a fifth free parameter to be calibrated (i.e. either the capacity of the interception store or another parameter used in the exchange function).

The interception store has a capacity of a few millimetres (I_{max} [mm]), which allows one to temporarily store an amount of intercepted rainfall (I [mm]) for a few hours or days. The stock of water in the store is reduced by evaporation (E_i [mm]) at potential rate (e [mm/time]), unless the effective water supply rate ($p + I/\Delta t$) is limiting, and overflows producing throughfall (P_{th} [mm], at rate $p_{th} \leq p$ [mm/time]) when the capacity I_{max} is exceeded. Depending on the order of the interactions of the two climatic inputs (P and E) with the interception store, two alternative implementations of the interception component are possible. We tested both options and chose the implementation with the evaporation calculated prior to throughfall, at the beginning of each fixed time step. While model performance was equivalent (not shown), the chosen implementation has two advantages: (i) its behaviour is more intuitive, leading to a monotonous relationship between the store capacity and the time step (see Fig. 9), and (ii) it is in line with what is traditionally done in other 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). The chosen implementation of the interception store is governed by three subsequent equations at each time step Δt:

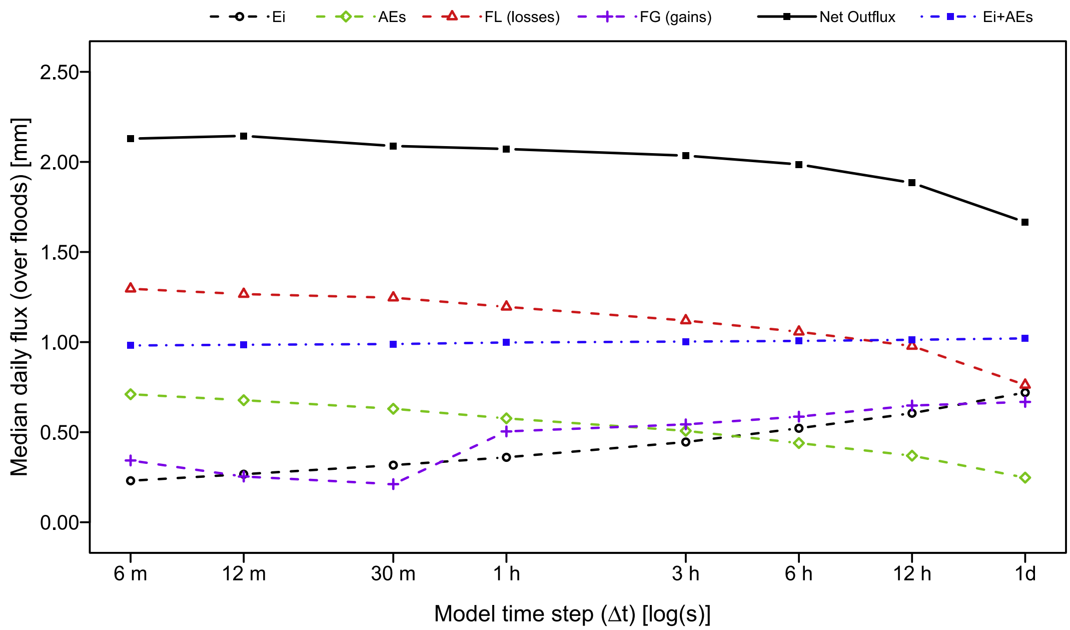

Fig. 7. Evolution across time steps of the daily average fluxes modelled by the GR4 model (median values) over the 2400 flood events: interception loss (E_i); actual evaporation from production store (AE_s); groundwater exchange losses (FL) and gains (FG); net outflux ($E_i + AE_s + |F_L| - F_G$); total losses to atmosphere ($E_i + AE_s$).

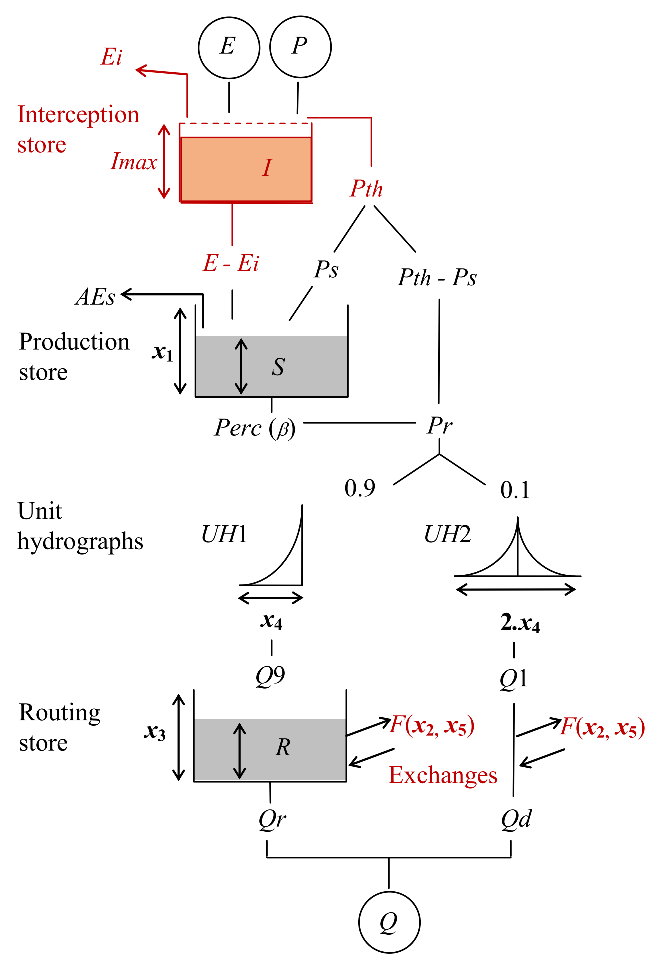

Fig. 8. Schematic representation of the GR4-I and GR5-I model structure, modified from the GR4 baseline model by insertion of an interception store of maximum capacity *Imax*. The components changed with respect to the baseline model are highlighted in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

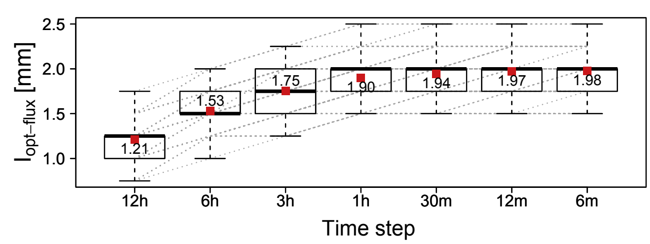

Fig. 9. Distribution over the catchment set of the capacity of the interception store of the GR4-I model at different time steps fixed via flux matching with the daily reference. The boxplots report the median value, interquartile range, and the whiskers represent the 10th and 90th percentiles; the red dots refer to mean values. The dotted grey lines behind the boxplots report the trajectories of the capacity for all 240 catchments (it should be remembered that the search step was 0.25 mm, so the lines overlap on a finite number of possible trajectories). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$e_i = \min\left(e,\ p + \frac{I_0}{\Delta t}\right): \Leftrightarrow E_i = \min(E,\ P + I_0) \tag{15}$$

$$p_{th} = max\left[0,\ p - \frac{(I_{max} - I_0)}{\Delta t} - e_i\right]:$$
$$\Leftrightarrow P_{th} = max\,[0,\ P - (I_{max} - I_0) - E_i] \tag{16}$$

$$I = I_0 + (p - e_i - p_{th})\Delta t: \Leftrightarrow I = I_0 + P - E_i - P_{th} \tag{17}$$

where P and E are the rainfall and potential evapotranspiration amounts over the time step, derived from the corresponding rates (p and e); $\boldsymbol{I_{max}}$ is the interception store capacity and $\boldsymbol{I_0}$ its water content at the beginning of the time step; $\boldsymbol{E_i}$ and $\boldsymbol{P_{th}}$ are the actual evaporation and throughfall from the store, with the latter representing the new net rainfall $\boldsymbol{P_n}$. Since part of the potential evapotranspiration energy available ($\boldsymbol{E}$) is used to evaporate water from the interception store ($\boldsymbol{E_i}$), only the difference ($\boldsymbol{E - E_i}$) remains available to evaporate water from the soil-moisture accounting store. The total actual evaporation loss will be the sum of $\boldsymbol{E_i}$ and $\boldsymbol{AE_s}$, where $\boldsymbol{AE_s}$ is the amount of water evaporated from the production store (Eq. (4)), calculated as a function of the new net evaporation capacity ($\boldsymbol{E_n = E - E_i}$).

5.2. Fixing the interception capacity by flux matching

The interception store capacity (I_{max}) can be either fixed or calibrated, according to criteria based on model performance or internal flux consistency. We evaluated three alternatives with calibration-validation tests:

(i) I_{max} fixed at the same value for all catchments. We tested values of I_{max} between 0.1 and 15 mm, i.e. including the typical range of interception capacities from the literature, ranging from 1.7 to 3.8 mm for France: lower than 2 mm for broad-leaved tree forests and larger than 3 mm for coniferous forests (see Aussenac, 1968);

(ii) I_{max} calibrated. I_{max} was calibrated for each catchment within a continuous domain (from 0 to 15 mm) based on modelled streamflow accuracy (see Eq. (9)), bringing the number of calibrated parameters to five. This model version is called here GR5-I* (the asterisk is used to distinguish it from the five-parameter model version finally retained);

(iii) I_{max} fixed via flux matching across time steps for each catchment. We searched for the interception store capacities ensuring the consistency of the interception fluxes at sub-daily time steps ($I_{opt\text{-}flux}$) with the reference given by the daily neutralisation function. The flux-matching criterion to optimise is the cumulative flux ratio for interception (see Eq. (11)) normalised on the daily reference given by neutralisation of P and E. An iterative search algorithm was set up to maximise the coherence of the interception fluxes at each time step with the daily reference. We explored the interception capacity space between 0 and 15 mm by a search step of 0.25 mm (deemed sufficient), calculating the corresponding interception loss. This procedure identified the interception store capacity ensuring coherent interception losses across model time steps.

For each test, the GR4's four original free parameters were recalibrated.

At the hourly time step, the GR4-I model with I_{max} fixed at the same value for all catchments (alternative (i)) markedly improves model accuracy with respect to the baseline model, especially over flood events (Table 4). The median KGE over floods is maximum with the 2.5-mm interception capacity. However, the performance scores of GR4-I with capacities from 2 to 10 mm are not significantly different (Friedman test). With an interception store of increasing capacity, we observed a decrease in the absolute value of the water exchange coefficient. This confirms the observed compensative feedback mechanism between interception loss and exchanges observed for GR4.

The Friedman test indicates that, at the hourly and daily time steps, there is no significant difference between the scores over the whole series and over flood events for GR5-I* with calibrated interception capacity (alternative (ii)) and GR4-I with any fixed capacity between 2.25 and 10 mm. Therefore, the additional complexity of a free interception capacity is not justified in terms of model performance. This may be explained by a reduction of the robustness of the model with an increased number of free parameters and the compensations between exchanges and interception. At the daily time step, the differences in average performance between GR4-I (/GR5-I*) and the baseline model are negligible ($< 10^{-2}$) and do not warrant the addition of the interception store in the daily GR4.

The distributions of the optimal interception capacities obtained via flux matching (alternative (iii)) for the GR4-I model at different time steps (Fig. 9) show the stabilisation of the parameter (at 2 mm, on average) as the time step decreases. The relationship between interception capacity and time step is monotonous not only on average but also for each catchment (see lines behind boxplots). This model version, which provides flexible interception capacity for flux matching, while keeping a similar level of performance to the other versions (see Table 4) and not introducing any extra free parameter, was retained for the tests detailed in the next section.

5.3. Flux consistency at different time steps for the GR4-I model

5.3.1. Flux-matching analysis over the whole validation period

Fig. 10 reports a summary of the evolution of the internal fluxes of the GR4-I model cumulated over the entire 8-year validation period as the time step decreases from 1 day to 6 min. The fluxes are normalised to the chosen reference given by the fluxes of the daily GR4 model. Fig. 10 shows that the model fluxes were stabilised across time steps by introducing the interception store (see Fig. 4 for comparison with the original model). In fact, the median value and the interquartile range of the cumulative flux ratios are centred around the optimal ratio (1) at all time steps, which is a substantial improvement over the baseline model case. The cumulated interception flux does not change more than 1.5% at all time steps. Therefore, all other modelled fluxes deviate on average by less than 5% from the corresponding reference daily model fluxes. Fig. 10 shows that, for interception loss and actual evaporation, the extreme quantiles of the cumulative flux ratios also do not deviate much from 1 (5% at most). Only for the exchange fluxes do a few large relative changes (> 20%) remain. However, these large relative deviations are associated with small absolute deviations (in mm) with respect to the whole annual water balance (< 1% of the total net outflux, on average), because the reference exchange flux can be small. Moreover, for the GR4-I model even the largest relative deviations in the exchange fluxes are much lower than those observed for the GR4 baseline model (see Fig. 4 for comparison).

For an overview of the dispersion across the 240 catchments, Fig. 11 shows the comparison of the cumulated fluxes between the hourly and daily models. It shows very good consistency of the simulated fluxes as the time step changes for almost all the catchments (see Fig. 6 for a comparison with the original model). The new version of the model

Table 4
Median values over the 240 catchments of performance criteria for different hourly model structures derived by insertion of an interception store of fixed, calibrated or flux-matching capacity in the GR4 baseline model and in the GR5 variant with linear exchange.

Hourly model	Median criteria on the whole series (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 baseline	0.820	0.989	1.009	0.897	0.972	1.39	0.952	0.727	0.969	0.904	0.832
GR4-I with fixed 2.5-mm I_{max} capacity	0.829	0.995	1.005	0.907	0.981	5.18	0.880	0.747	0.988	0.923	0.836
GR5-I* with calibrated I_{max} capacity	0.829	0.996	0.997	0.911	0.980	7.50	0.837	0.752	0.976	0.934	0.834
GR4-I with flux-matching ($I_{opt\text{-}flux}$) capacity	0.827	0.991	1.008	0.905	0.979	4.53	0.887	0.742	0.979	0.920	0.836
GR5 variant of baseline with linear exchange (Eq. (8))	0.826	0.993	1.002	0.903	0.979	−2.35	1.023	0.737	1.018	0.924	0.829
GR5-I with interception store ($I_{opt\text{-}flux}$) and linear exchange (Eq.(8))	0.831	0.998	1.002	0.909	0.991	−2.03	1.017	0.747	1.032	0.936	0.835

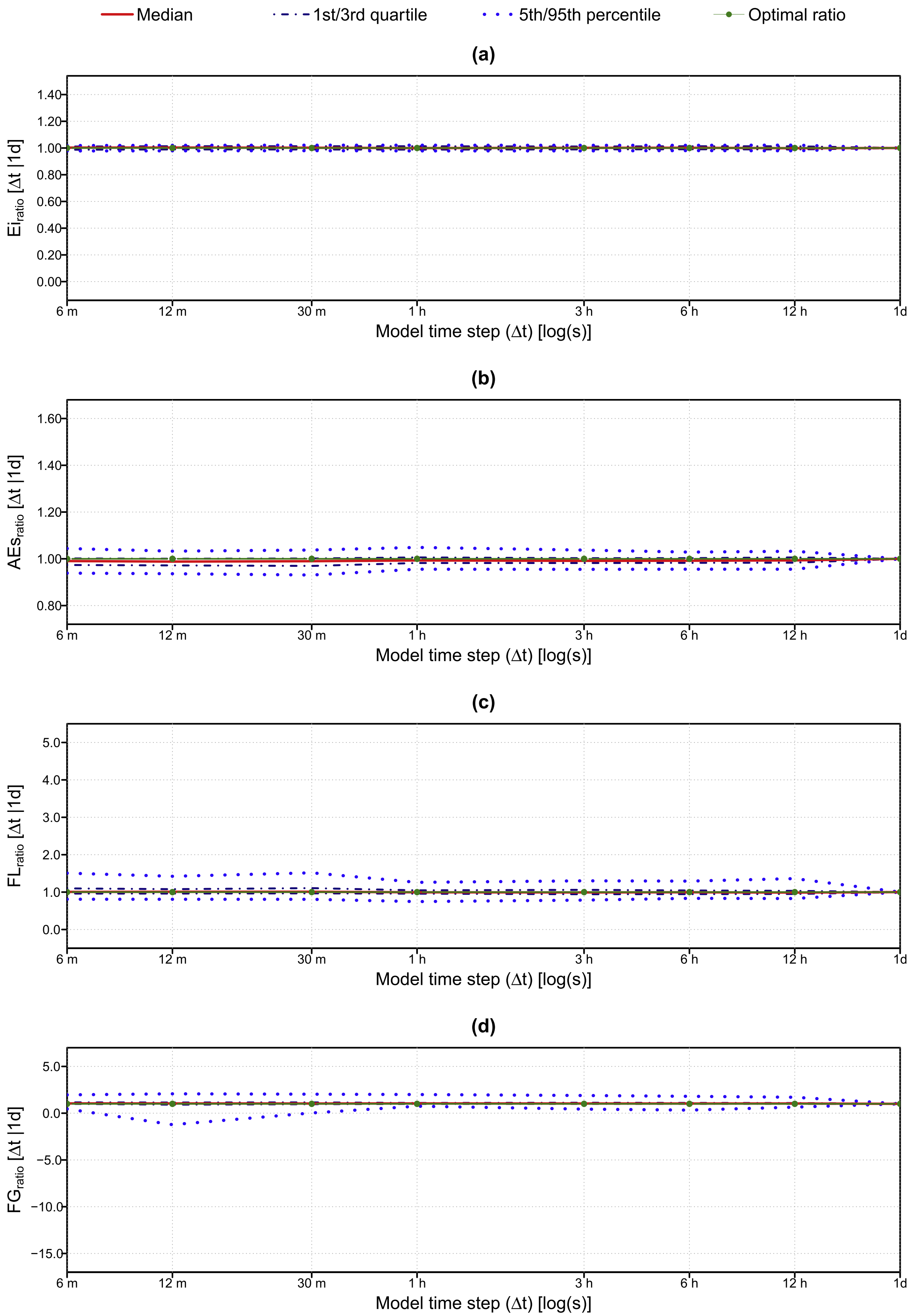

Fig. 10. Summary of the cumulative flux ratios of the GR4-I model at different time steps (with daily reference) over the whole validation period and the 240-catchment set: (a) interception loss, *Ei*; (b) actual evaporation from the production store, AE_s; (c) groundwater losses, F_L; (d) groundwater gains, F_G.

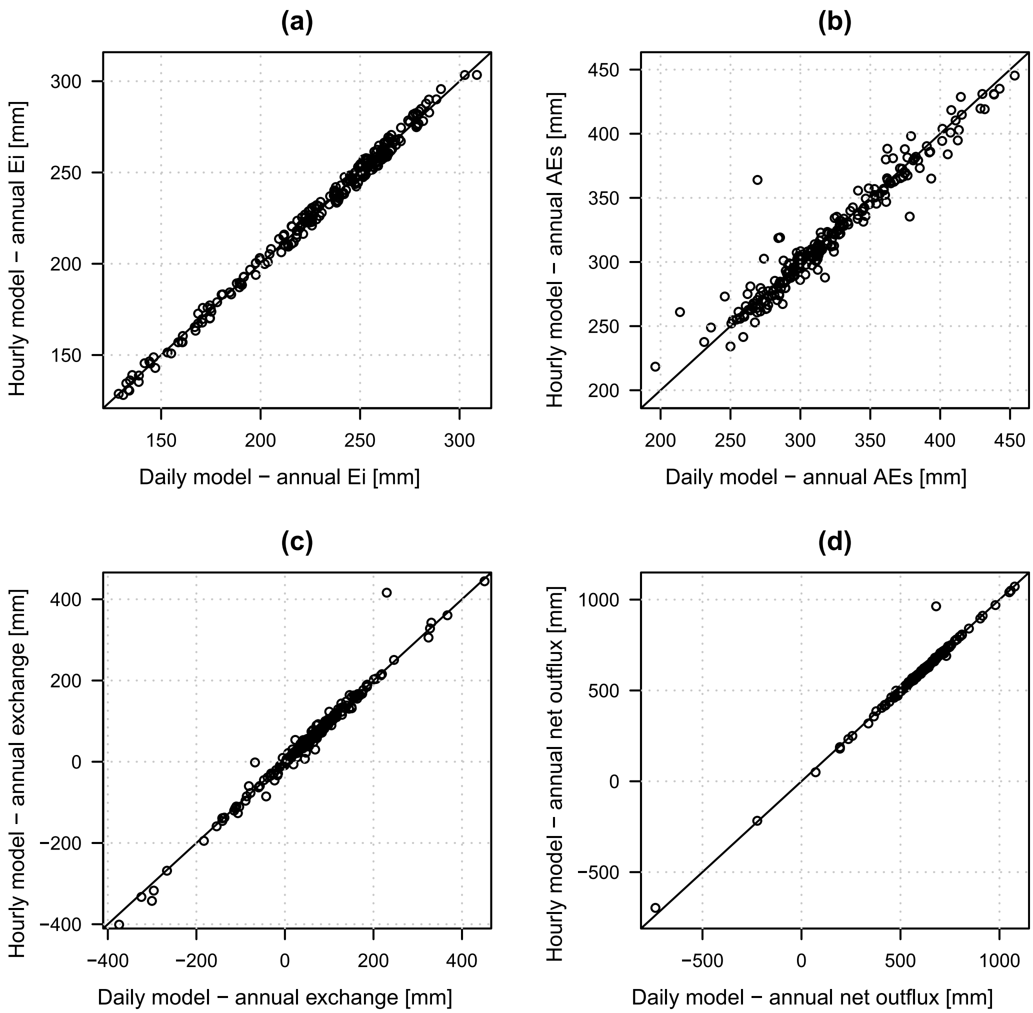

Fig. 11. Annual average cumulated fluxes from daily and hourly GR4-I model simulations over the 240 catchments (each dot represents one catchment): (a) interception loss, *Ei*; (b) actual evaporation from the production store, AE_s; (c) groundwater exchange losses (negative values represent gains, i.e. $-F_{L/G}$); (d) net outflux ($\boldsymbol{E_i + AE_s + |F_L| - F_G}$).

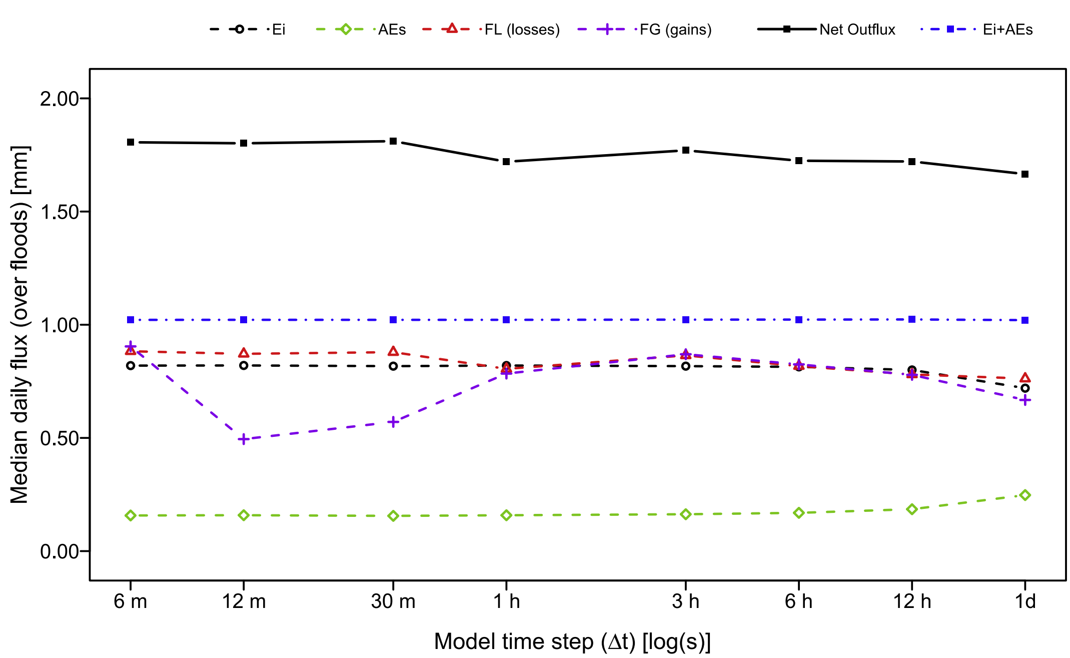

Fig. 12. Evolution across time steps of the daily average fluxes modelled by the GR4-I model (median values) over the 2400 flood events: interception loss (E_i); actual evaporation from production store (AE_s); groundwater exchange losses (*FL*) and gains (*FG*); net outflux ($E_i + AE_s + |F_L| - F_G$); total losses to atmosphere ($E_i + AE_s$).

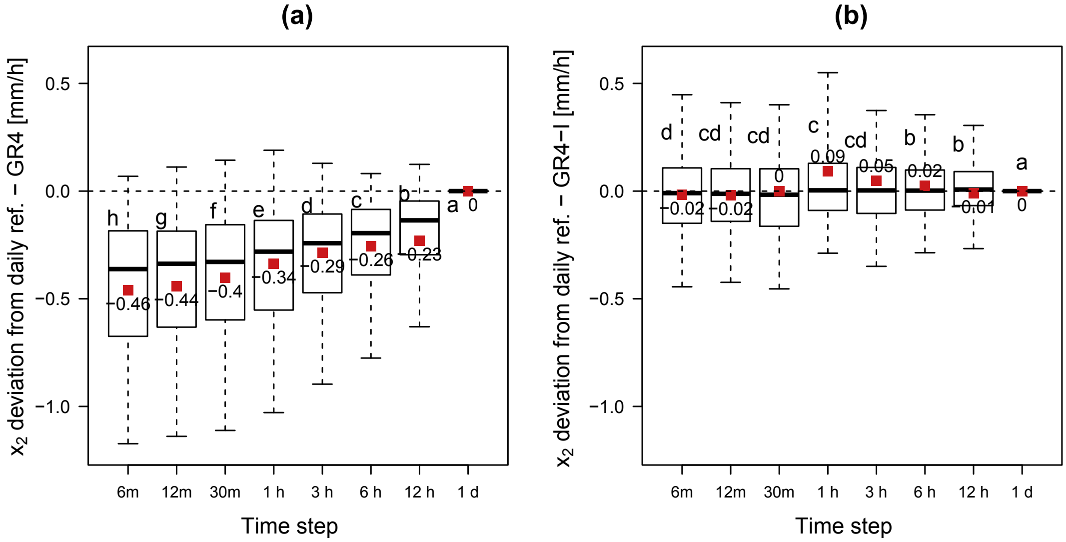

Fig. 13. Distribution over the 240 catchments of the deviations of the water exchange parameter (x_2) between the daily model reference (GR4) and the adapted model at time steps from 6 min to 12 h: (a) GR4 baseline; (b) GR4-I with the interception store of catchment-dependent capacity fixed via flux matching ($I_{opt\text{-}flux}$). The boxplots report the median value and interquartile range, the whiskers the 10th and 90th percentiles, and the red dots the mean values. The letters above each boxplot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letters are not significantly different).

with the interception store (GR4-I) stabilised the fluxes across time steps. Also, the problem of groundwater exchanges changing sign across model time steps is almost completely solved, as shown in Fig. 11(c).

5.3.2. Flux-matching analysis over flood events

Fig. 12 provides the median values of the daily cumulated fluxes over the 2400 flood events simulated at different time steps. The average change in interception loss over floods is at most 0.1 mm/day and it is perfectly compensated by the actual evaporation changes (see Fig. 12). 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. In the new GR4-I model the groundwater losses increase slightly by passing from daily to shorter time steps (while the gains show a slight oscillation), although they no longer need to compensate spurious changes of interception loss volumes. However, most of the groundwater loss changes across time steps has been reduced (they increased by about 90% in the baseline model, while only by 16% here).

5.4. Does flux coherence improve parameter consistency?

We expect that the spurious time-step dependency detected for the water exchange coefficient (x_2) of GR4 will be reduced with the new model version, given the considerable improvement in the temporal consistency of the fluxes of the GR4-I model. Fig. 13 shows the distributions of the deviations of the x_2 parameters of the GR4 baseline model (left panel) and the GR4-I model (right panel) calibrated at the sub-daily time steps tested from the corresponding value obtained by calibration at the daily time step. For a consistent comparison, the parameters at all the different time steps are converted to their corresponding hourly value ([mm/h]) by applying the theoretical relationship derived from the integration of the model's governing equations (see Table A.1, Appendix A).

Fig. 13 shows that the water exchange coefficient parameter is stabilised after the insertion of the interception store. The decreasing trend observed for GR4 (left panel) has disappeared for GR4-I (right panel). For the new model version, the average deviation of the parameter is around zero at all time steps. A few large deviations at different time steps are still found over the catchment set, but no trend is detected. This stabilisation of the parameter corroborates the results presented by Kavetski et al. (2011), who showed that some parameters (common to different model structures of increasing complexity) are highly scale-dependent in simpler models but become progressively more stable in more complex models. Table 3 reports the statistics that provide a better understanding of the positive effect of the interception store on the temporal coherence of all the water–balance-related parameters. The median values of the parameters at the two extreme time steps, 1 day and 6 min, are much more stable than what was observed for GR4 (see Table 2). Note that the relative changes of the parameters at different time steps are now much lower than the relative standard deviation of the parameters over the catchment set.

5.5. Does flux coherence improve model performance?

Table 4 reports the median values of the criteria of model performance over the entire validation period and over flood events for the hourly simulation of the baseline model (GR4) and the modified versions in which the interception store has fixed, calibrated or flux-matching capacity. The results are analogous at all sub-daily time steps. All the new model versions with the interception store contribute a general improvement in the KGE criterion and its components (relative variability, ratio of means and correlation), evaluated both on the whole series and on flood events. This improvement is more significant over flood events, especially in terms of reduced balance bias, which is improved by 2–3% starting from a score of 90% with the baseline model. However, a significant decrease of performance over low flows is detected, with a worsened underestimation, detected by the flow quantile ratios (Q20 is more underestimated by the new model versions GR4-I/GR5-I* than by the baseline GR4). This also implies an increasing slope bias of the FDC. This problem could be possibly solved by modifying the exchange function, which should also be improved because of its slight residual temporal inconsistency (Fig. 12). The general improvement of model performance in terms of regime and floods may be explained by the increased freedom of the new model (with refined interception) in adjusting soil moisture and in better describing antecedent wetness conditions, as discussed also by Fenicia et al. (2008b).

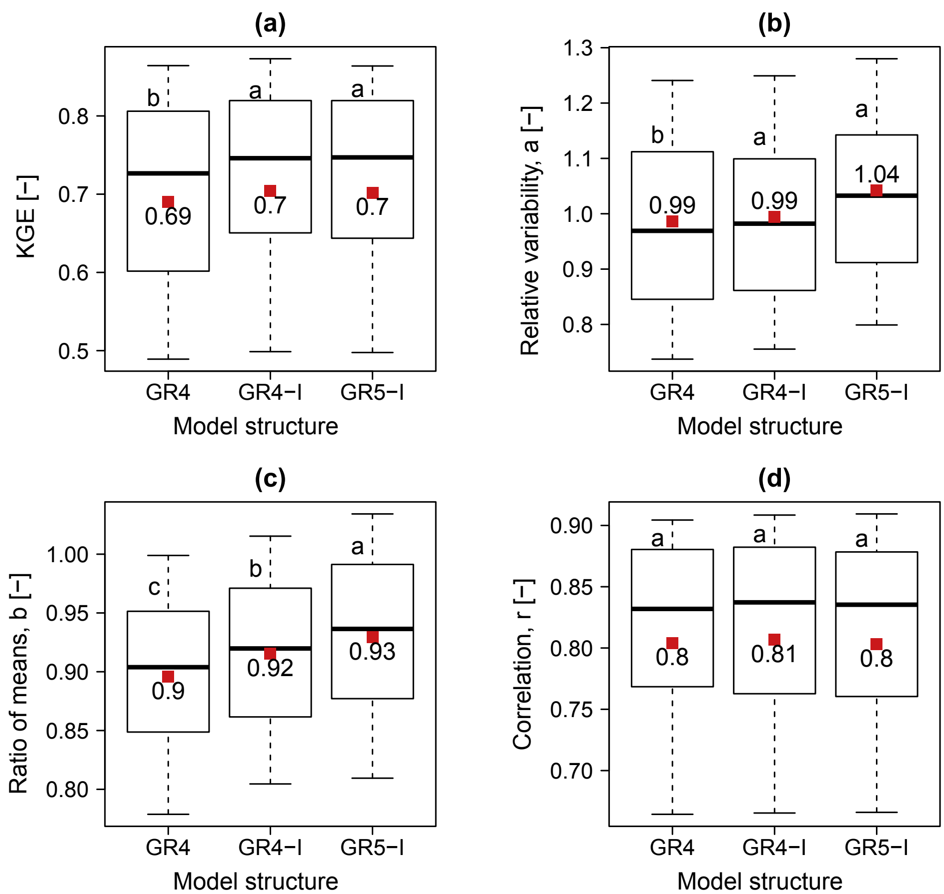

Fig. 14. Distribution of the performance criteria over the 2400 flood events for three models at the hourly time step: GR4, GR4-I (with interception store) and GR5-I (with interception store and linear exchange function): (a) *KGE*, (b) relative variability, *a*; (c) ratio of means, *b*; (d) correlation, *r*. The boxplots report median and interquartile range, the whiskers the 10th and 90th percentiles, and the red dots the mean values. The letters above each boxplot specify the ranking (alphabetical order) and the significant differences detected by the Friedman test at significance level 0.05 (distributions with the same letters are not significantly different).

5.6. A complementary structural modification for improving low-flows: GR5-I with a refined exchange function

We tested several modifications to the exchange function, in combination with the interception store with catchment-dependent capacity fixed via flux matching. Among the alternative functions tested, the linear exchange function (Eq. (8)) proposed by Le Moine (2008) proved to be the best option. This function has an additional free parameter leading to a model with five free parameters, called GR5-I. This model version solves the problem of performance degradation on low flows and provides further improvement over flood events (bias reduction), as summarised in Table 4. Fig. 14 shows the distribution of the performance scores over flood events for the GR4 baseline, the GR4-I and GR5-I models at the hourly time step. The improvements provided by both the GR4-I and GR5-I structures are statistically significant for KGE, relative variability and ratio-of-means criteria. The interquartile range of the KGE score distribution is particularly improved with the new models. The GR5-I model provides an additional significant improvement in the ratio-of-means with respect to GR4-I. These findings lead us to retain the GR5-I model with the interception store and linear exchange function at sub-daily time steps. This model ensures substantial improvements of the temporal consistency of the fluxes modelled (as shown for its parent structure GR4-I) while also significantly improving output accuracy in all conditions (regime, low and high flows).

6. Conclusions and perspectives

In this paper, we aimed to understand which structural improvements were needed to adapt the daily GR4 model for multiple sub-daily time steps, based on both model internal coherence and performance. For the sake of generality, the rainfall-runoff model simulations were run and evaluated at eight time steps, from 6 min to 1 day, over a large and varied set of 240 catchments and 2400 flood events. Simulations showed significant deficiencies in the coherence of the baseline model across time steps on the whole catchment set. The change of the GR4 model time step results in large and monotonous changes of the internal fluxes. The interception loss, i.e. the first flux calculated in the model, is significantly reduced at shorter time steps and this affects the subsequent fluxes by strong compensation effects. While over the long term the coherence of the net water balance is ensured by this chain of compensation effects, the net outflux is inconsistent over flood events and worsens the simulation bias at shorter time steps. These differences in model fluxes also translate into spurious changes of the calibrated model parameters across time steps.

The addition of a refined interception component at sub-daily time steps provides consistent model flux estimates and parameters across time steps. Significant benefits of the new modified structure with the interception store are also found on model performance, particularly over floods. Our modelling experiments show that fulfilling a flux-matching condition across time scales is important for reducing spurious parameter compensations, while also improving model performance across time steps. These results suggest that the consistency of internal fluxes should be considered a prerequisite for the identification of any multi-time–step model but can also help to detect structural deficiencies of the model at fixed time steps. The main general conclusions from our study that should be independent of the model/setting used are:

(i) a flux-matching condition can help better identify model structural deficiencies and improvements at multiple time steps;
(ii) the interception is essential for flood modelling at sub-daily time steps and its adequate representation may improve model performance and consistency when using continuous simulation models, confirming conclusions of previous studies (e.g. Fenicia et al.,

2008b);

(iii) in models with multiple free calibrated parameters, related to the water balance, a mis-representation of the interception leads to uncontrolled compensations between internal fluxes and spurious dependencies of parameters on time step. This corroborates results previously found also with the continuous state-space version of the same model (Santos et al., 2018);

(iv) to solve this issue, we included a parametric interception component, commonly used in lumped hydrological modelling (Fenicia et al., 2008b; Kandel et al., 2005), and we showed that for its parameterization, calibration is not needed if a flux-matching condition is used.

The operational value of this study is related to the flood forecasting application of the GR4 model. Most of the French flood forecasting services routinely use a slightly modified hourly version of the discrete GR4 model, called GRP (Berthet et al., 2009), based on a simplified version of GR4 better adapted to the forecasting mode. The current hourly version of the GRP model uses the GR4 neutralisation function for interception, which is subject to the limitations shown in this paper. The replacement of this function by an interception store should be validated for the GRP model version.

In this paper, we used the fluxes simulated at the daily time step as a reference to evaluate flux-matching, also because no measurement of the internal fluxes was available for this large-sample study. In future studies using this model or others, it would be useful to compare the internal fluxes modelled at different time steps to flux observations that may be available on small experimental catchments. This would potentially provide a better reference for the assessment of the flux 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.).

Recently, a state-space representation of the GR4 model was proposed by Santos et al. (2018), showing flow simulations that were very similar to the discrete GR4 model at the daily and hourly time steps. The formulation of the fluxes is equivalent in the discrete and state-space formulations, and the only difference in model structure stands in the unit hydrographs, replaced by a Nash cascade in Santos et al. (2018). Also in the state-space model, the interception component was directly written in its discrete form as a neutralisation function, since this is not governed by a dynamical component. Santos et al. (2018) also found that the water-balance parameters of GR4 are affected by spurious time-step dependencies and they argue that these are due to temporal inconsistencies in the interception calculation, referring to our work. Therefore, also in the state-space GR4 model, the neutralisation function should be replaced by the interception store proposed here, which is expected to lead to positive impacts in terms of both flux/parameter consistency and model accuracy.

Last, another perspective is to apply the framework for model identification based on flux matching to semi-distributed models across different spatio-temporal scales. The impact of the combined refinement of temporal and spatial resolutions on model fluxes should be investigated to detect possible inconsistencies across scales, which could guide the improvement of the semi-distributed model at multiple spatial scales. For example, the semi-distributed GR-SD model, based on the GR5 baseline model variant, might benefit from further research in this sense.

Declaration of Competing Interest

None.

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 Agnes Ducharne, Hubert Savenije, Marco Borga, Nicolas Massei, Isabella Zin, Lionel Berthet and Nicolas Le Moine for their useful comments and the discussion on the PhD thesis presented by the first author, which also helped to improve the present paper. The authors thank the Associate Editor Saman Razavi, as well as Shervan Gharari and the other anonymous reviewer for the constructive comments which helped to improve the manuscript.

Appendix A. Theoretical relationships of the model parameters at different time steps and model equations

Table A.1 reports the theoretical time-step dependencies of the fixed and free model parameters of the GR4 model (first five lines) and the new GR4-I and GR5-I models (all lines).

Here we report the equations of the fluxes representing transfers within the catchment to complement the model presentation in Section 2. The discrete model equations derive from the integration of the continuous equations over a time step, apart from the equations representing the interception loss and the exchange fluxes, which are directly written in their discrete form (see Section 2.2.2). In the following equations, the *water fluxes are integrated over* the model time step and are expressed in millimeters.

Table A.1
Theoretical time-step dependency of all fixed and free (calibrated) parameters of the GR4 model and GR5-I models (adapted from Table C.1 in Ficchì et al., 2016).

Model parameter	Parameter definition	Theoretical transformation from Δt_1[s] to Δt_2[s]	Source of the time-step dependency
β [-]	Percolation coefficient (fixed, $\beta = 9/4$ at the daily time step)	$\beta_{(\Delta t_2)} = \beta_{(\Delta t_1)}\left(\frac{\Delta t_1}{\Delta t_2}\right)^{\frac{1}{4}}$	Integration of the percolation rate from the production store, which is a power 5 function
$x_1\,[mm]$	Capacity of the production store (calibrated)	–	No time-step dependency
$x_2\left[\frac{mm}{timestep}\right]$	Water exchange coefficient (calibrated)	$x_{2(\Delta t_2)} = x_{2(\Delta t_1)}\left(\frac{\Delta t_1}{\Delta t_2}\right)^{-\frac{1}{8}}$	Integration of the exchange flux formulation (dependent on the routing store level)
$x_3\,[mm]$	Capacity of the routing store (calibrated)	$x_{3(\Delta t_2)} = x_{3(\Delta t_1)}\left(\frac{\Delta t_1}{\Delta t_2}\right)^{\frac{1}{4}}$	Integration of the emptying function of the routing store that is a power 5 function
$x_4\,[timestep]$	Time base parameter (calibrated)	$x_{4\ (\Delta t_2)} = \left[x_{4\ (\Delta t_1)}\left(\frac{\Delta t_1}{\Delta t_2}\right)\right]$	Value expressed in time-step units and rounded to the nearest integer
$x_5\,[-]$	Threshold for exchange function sign change (calibrated)	–	No time-step dependency
$I_{max}\,[-]$	Interception store capacity (fixed by flux-matching criteria with daily GR4 reference)	$I_{max(\Delta t)} = f(P_{daily}, E_{daily})$	Value determined by matching the interception flux with the daily neutralisation (function of P and E)

The water transfers in the production part of the model include the infiltration in the soil-moisture accounting store (P_s) and the percolation ($Perc$):

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

$$Perc = S\left\{1-\left[1+\left(\frac{S}{\beta_{\Delta t}\bullet x_1}\right)^4\right]^{-\frac{1}{4}}\right\} \tag{A.2}$$

The discrete Eq. (A.2) is obtained from the integration over the time step of an instantaneous leaking function, which leads to a time–step-dependent percolation constant $\beta_{\Delta t}$ (see Table A.1).

In the routing part, the outflow from the store is:

$$Q_r = R\left\{1-\left[1+\left(\frac{R}{x_3}\right)^4\right]^{-\frac{1}{4}}\right\} \tag{A.3}$$

Similar to the percolation equation, the discrete Eq. (A.3) is obtained from the integration over the time step of an instantaneous leaking function, which leads to a theoretical time-step dependency of the routing store reference capacity x_3 (see Table A.1).

References

Andréassian, V., et al., 2009. Crash tests for a standardized evaluation of hydrological models. Hydrol. Earth Syst. Sci. 13 (10), 1757–1764.

Atkinson, S.E., Woods, R.A., Sivapalan, M., 2002. Climate and landscape controls on water balance model complexity over changing timescales. Water Resour. Res. 38 (12), 501–5017.

Aussenac, G., 1968. Interception des précipitations par le couvert forestier. Annales des sciences forestières 25 (3), 135–156. https://doi.org/10.1051/forest/19680302.

Bastola, S., Murphy, C., 2013. Sensitivity of the performance of a conceptual rainfall-runoff model to the temporal sampling of calibration data. Hydrol. Res. 44 (3), 484–494. https://doi.org/10.2166/nh.2012.06.

Berthet, L., Andréassian, V., Perrin, C., Javelle, P., 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. Hydrol. Earth Syst. Sci. 13 (6), 819–831. https://doi.org/10.5194/hess-13-819-2009.

Blöschl, G., 2006. Hydrologic synthesis: across processes, places, and scales. Water Resour. Res. 42 (3). https://doi.org/10.1029/2005wr004319.

Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrol. Process. 9 (3–4), 251–290. https://doi.org/10.1002/hyp.3360090305.

Calder, I.R., 1990. Evaporation in the Uplands. John Wiley & Sons, Chichester, UD, pp. 148.

Clark, M.P., Kavetski, D., Fenicia, F., 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resour. Res. 47 (9), W09301. https://doi.org/10.1029/2010WR009827.

Coron, L., Perrin, C., Michel, C., 2017a. airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling. R package version 0.7.2.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., Andréassian, V., 2017b. The suite of lumped GR hydrological models in an R package. Environ. Modell. Software 94, 166–171. https://doi.org/10.1016/j.envsoft.2017.05.002.

Cullmann, J., Mishra, V., Peters, R., 2006. Flow analysis with WaSiM-ETH – model parameter sensitivity at different scales. Adv. Geosci. 9, 73–77.

de Boer-Euser, T., et al., 2017. Looking beyond general metrics for model comparison – lessons from an international model intercomparison study. Hydrol. Earth Syst. Sci. 21 (1), 423–440. https://doi.org/10.5194/hess-21-423-2017.

Euser, T., et al., 2013. A framework to assess the realism of model structures using hydrological signatures. Hydrol. Earth Syst. Sci. 17 (5), 1893–1912. https://doi.org/10.5194/hess-17-1893-2013.

Farmer, D., Sivapalan, M., Jothityangkoon, C., 2003. Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: downward approach to water balance analysis. Water Resour. Res. 39 (2), SWC11–SWC121.

Fenicia, F., McDonnell, J.J., Savenije, H.H.G., 2008a. Learning from model improvement: on the contribution of complementary data to process understanding. Water Resour. Res. 44 (6). https://doi.org/10.1029/2007wr006386.

Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L., 2008b. Understanding catchment behavior through stepwise model concept improvement. Water Resour. Res. 44 (1), W01402. https://doi.org/10.1029/2006WR005563.

Fenicia, F., Kavetski, D., Savenije, H.H.G., 2011. Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. Water Resour. Res. 47 (11). https://doi.org/10.1029/2010WR010174.

Ficchì, A., Perrin, C., Andréassian, V., 2016. Impact of temporal resolution of inputs on hydrological model performance: an analysis based on 2400 flood events. J. Hydrol. 538, 454–470. https://doi.org/10.1016/j.jhydrol.2016.04.016.

Finnerty, B.D., Smith, M.B., Seo, D.J., Koren, V., Moglen, G.E., 1997. Space-time scale sensitivity of the Sacramento model to radar-gage precipitation inputs. J. Hydrol. 203 (1–4), 21–38. https://doi.org/10.1016/S0022-1694(97)00083-8.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. 32 (200), 675–701. https://doi.org/10.2307/2279372.

Gerrits, A.M.J., 2010. The Role of Interception in the Hydrological Cycle. PhD Thesis. TU Delft, Delft, the Netherlands.

Gerrits, A.M.J., Pfister, L., Savenije, H.H.G., 2010. Spatial and temporal variability of canopy and forest floor interception in a beech forest. Hydrol. Process. 24 (21), 3011–3025. https://doi.org/10.1002/hyp.7712.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., Savenije, H.H.G., 2014. Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. Hydrol. Earth Syst. Sci. 18 (12), 4839–4859. https://doi.org/10.5194/hess-18-4839-2014.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Gupta, H.V., Clark, M.P., Vrugt, J.A., Abramowitz, G., Ye, M., 2012. Towards a comprehensive assessment of model structural adequacy. Water Resour. Res. 48 (8). https://doi.org/10.1029/2011wr011044.

Haddeland, I., Lettenmaier, D.P., Skaugen, T., 2006. Reconciling simulated moisture fluxes resulting from alternate hydrologic model time steps and energy budget closure assumptions. J. Hydrometeorol. 7 (3), 355–370. https://doi.org/10.1175/JHM496.1.

Hrachowitz, M., et al., 2014. Process consistency in models: the importance of system signatures, expert knowledge, and process complexity. Water Resour. Res. 50 (9), 7445–7469. https://doi.org/10.1002/2014wr015484.

Ishidaira, H., et al., 2003. Effect of Spatial and Temporal Resolution of Precipitation Data on the Accuracy of Long-Term Runoff Simulation. IAHS-AISH Publication, pp. 186–193.

Jeong, J., et al., 2010. Development and integration of sub-hourly rainfall-runoff modeling capability within a watershed model. Water Resour. Manage. 24 (15), 4505–4527. https://doi.org/10.1007/s11269-010-9670-4.

Jothityangkoon, C., Sivapalan, M., Farmer, D.L., 2001. Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development. J. Hydrol. 254 (1–4), 174–198. https://doi.org/10.1016/S0022-1694(01)00496-6.

Kandel, D.D., Western, A.W., Grayson, R.B., 2005. Scaling from process timescales to daily time steps: a distribution function approach. Water Resour. Res. 41 (2), 1–16. https://doi.org/10.1029/2004WR003380.

Kavetski, D., Fenicia, F., Clark, M.P., 2011. Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: insights from an experimental catchment. Water Resour. Res. 47 (5). https://doi.org/10.1029/2010WR009525.

Kirchner, J.W., Feng, X., Neal, C., Robson, A.J., 2004. The fine structure of water-quality dynamics: the (high-frequency) wave of the future. Hydrol. Process. 18 (7), 1353–1359.

Klemeš, V., 1983. Conceptualization and scale in hydrology. J. Hydrol. 65 (1–3), 1–23.

Klemeš, V., 1986. Operational testing of hydrological simulation models. Hydrol. Sci. J./Journal des Sciences Hydrologiques 31 (1), 13–24.

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, pp. 348.

Leleu, I., et al., 2014. La refonte du système d'information national pour la gestion et la mise à disposition des données hydrométriques. La Houille Blanche 1, 25–32.

Littlewood, I.G., Croke, B.F.W., 2008. Data time-step dependency of conceptual rainfall-streamflow model parameters: an empirical study with implications for regionalisation. Hydrol. Sci. J. 53 (4), 685–695. https://doi.org/10.1623/hysj.53.4.685.
Littlewood, I.G., Croke, B.F.W., Young, P.C., 2011. Discussion of "effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge". Hydrol. Sci. J. 56 (3), 521–524. https://doi.org/10.1080/02626667.2011.565769.
Marty, R., Zin, I., Obled, C., 2013. Sensitivity of hydrological ensemble forecasts to different sources and temporal resolutions of probabilistic quantitative precipitation forecasts: flash flood case studies in the Cévennes-Vivarais region (Southern France). Hydrol. Process. 27 (1), 33–44. https://doi.org/10.1002/hyp.9543.
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, pp. 463.
Melsen, L., et al., 2016. 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. https://doi.org/10.5194/hess-20-2207-2016.
Morel-Seytoux, H.J., 1988. Soil aquifer stream interactions – a reductionist attempt toward physical-stochastic integration. J. Hydrol. 102 (1–4), 355–379. https://doi.org/10.1016/0022-1694(88)90107-2.
Moretti, G., Montanari, A., 2007. AFFDEF: a spatially distributed grid based rainfall-runoff model for continuous time simulations of river discharge. Environ. Modell. Software 22 (6), 823–836. https://doi.org/10.1016/j.envsoft.2006.02.012.
Mouelhi, S., Michel, C., Perrin, C., Andréassian, V., 2006. Stepwise development of a two-parameter monthly water balance model. J. Hydrol. 318 (1–4), 200–214. https://doi.org/10.1016/j.jhydrol.2005.06.014.
Obled, C., Zin, I., Hingray, B., 2009. Choix des pas de temps et d'espace pour des modélisations parcimonieuses en hydrologie des crues. La Houille Blanche 5, 81–87. https://doi.org/10.1051/lhb/2009059.
Ostrowski, M., Bach, M., Gamerith, V., de Simone, S., 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., et al., 2005. Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 – towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. J. Hydrol. 303 (1–4), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026.
Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279 (1–4), 275–289. https://doi.org/10.1016/S0022-1694(03)00225-7.
Rakovec, O., et al., 2016. Multiscale and multivariate evaluation of water fluxes and states over European River Basins. J. Hydrometeorol. 17 (1), 287–307. https://doi.org/10.1175/jhm-d-15-0054.1.
Raupach, M.R., Finnigan, J.J., 1995. Scale issues in boundary-layer meteorology: surface energy balances in heterogeneous terrain. Hydrol. Process. 9 (5–6), 589–612. https://doi.org/10.1002/hyp.3360090509.
Samaniego, L., et al., 2017. Toward seamless hydrologic predictions across spatial scales. Hydrol. Earth Syst. Sci. 21 (9), 4323–4346. https://doi.org/10.5194/hess-21-4323-2017.
Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res. 46 (5), W05523. https://doi.org/10.1029/2008WR007327.
Santos, L., Thirel, G., Perrin, C., 2018. Continuous state-space representation of a bucket-type rainfall-runoff model: a case study with the GR4 model using state-space GR4 (version 1.0). Geosci. Model Dev. 11 (4), 1591–1605. https://doi.org/10.5194/gmd-11-1591-2018.
Savenije, H.H.G., 2004. The importance of interception and why we should delete the term evapotranspiration from our vocabulary. Hydrol. Process. 18 (8), 1507–1511. https://doi.org/10.1002/hyp.5563.
Singh, V.P., Frevert, D.K., 2005. Watershed Models. Taylor & Francis.
Sivapalan, M., Blöschl, G., Zhang, L., Vertessy, R., 2003. Downward approach to hydrological prediction. Hydrol. Process. 17 (11), 2101–2111. https://doi.org/10.1002/hyp.1425.
Son, K., Sivapalan, M., 2007. Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data. Water Resour. Res. 43 (1). https://doi.org/10.1029/2006wr005032.
Van Esse, W.R., et al., 2013. The influence of conceptual model structure on model performance: a comparative study for 237 French catchments. Hydrol. Earth Syst. Sci. 17 (10), 4227–4239. https://doi.org/10.5194/hess-17-4227-2013.
Vidal, J.P., Martin, E., Franchistéguy, L., Baillon, M., Soubeyroux, J.M., 2010. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. Int. J. Climatol. 30 (11), 1627–1644. https://doi.org/10.1002/joc.2003.
Wagener, T., et al., 2001. A framework for development and application of hydrological models. Hydrol. Earth Syst. Sci. 5 (1), 13–26. https://doi.org/10.5194/hess-5-13-2001.
Wang, Y., He, B., Takase, K., 2009. Effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge. Hydrol. Sci. J. 54 (5), 886–898. https://doi.org/10.1623/hysj.54.5.886.
Wetterhall, F., He, Y., Cloke, H., Pappenberger, F., 2011. Effects of temporal resolution of input precipitation on the performance of hydrological forecasting. Adv. Geosci. 29, 21–25. https://doi.org/10.5194/adgeo-29-21-2011.
Winsemius, H.C., Savenije, H.H.G., Bastiaanssen, W.G.M., 2008. Constraining model parameters on remotely sensed evaporation: justification for distribution in ungauged basins? Hydrol. Earth Syst. Sci. 12 (6), 1403–1413. https://doi.org/10.5194/hess-12-1403-2008.
Wood, E.F., 1997. Effects of soil moisture aggregation on surface evaporative fluxes. J. Hydrol. 190 (3), 397–412. https://doi.org/10.1016/S0022-1694(96)03135-6.
Ye, W., Bates, B.C., Viney, N.R., Sivapalan, M., Jakeman, A.J., 1997. Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments. Water Resour. Res. 33 (1), 153–166. https://doi.org/10.1029/96wr02840.
Yilmaz, K.K., Gupta, H.V., Wagener, T., 2008. A process-based diagnostic approach to model evaluation: application to the NWS distributed hydrologic model. Water Resour. Res. 44 (9), W09417.