



# On selection of the optimal data time interval for real-time hydrological forecasting

J. Liu<sup>1,2</sup> and D. Han<sup>2</sup>

<sup>1</sup>State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China

<sup>2</sup>Water and Environmental Management Research Centre, Department of Civil Engineering, University of Bristol, Bristol BS8 1TR, UK

Correspondence to: J. Liu (jia.liu@iwhr.com)

Received: 22 August 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 26 September 2012

Revised: 25 June 2013 – Accepted: 14 August 2013 – Published: 30 September 2013

**Abstract.** With the advancement in modern telemetry and communication technologies, hydrological data can be collected with an increasingly higher sampling rate. An important issue deserving attention from the hydrological community is which suitable time interval of the model input data should be chosen in hydrological forecasting. Such a problem has long been recognised in the control engineering community but is a largely ignored topic in operational applications of hydrological forecasting. In this study, the intrinsic properties of rainfall–runoff data with different time intervals are first investigated from the perspectives of the sampling theorem and the information loss using the discrete wavelet transform tool. It is found that rainfall signals with very high sampling rates may not always improve the accuracy of rainfall–runoff modelling due to the catchment low-pass-filtering effect. To further investigate the impact of a data time interval in real-time forecasting, a real-time forecasting system is constructed by incorporating the probability distributed model (PDM) with a real-time updating scheme, the autoregressive moving-average (ARMA) model. Case studies are then carried out on four UK catchments with different concentration times for real-time flow forecasting using data with different time intervals of 15, 30, 45, 60, 90 and 120 min. A positive relation is found between the forecast lead time and the optimal choice of the data time interval, which is also highly dependent on the catchment concentration time. Finally, based on the conclusions from the case studies, a hypothetical pattern is proposed in three-dimensional coordinates to describe the general impact of the data time interval and to provide implications of the selection

of the optimal time interval in real-time hydrological forecasting. Although nowadays most operational hydrological systems still have low data sampling rates (daily or hourly), the future is that higher sampling rates will become more widespread, and there is an urgent need for hydrologists both in academia and in the field to realise the significance of the data time interval issue. It is important that more case studies in different catchments with various hydrological forecasting systems are explored in the future to further verify and improve the proposed hypothetical pattern.

## 1 Introduction

Hydrological forecasting has always been a dominant and challenging field in operational hydrology. Owing to the spatial and temporal variability of catchment characteristics and rainfall patterns, together with a number of associated hydrological components which are highly nonlinear, time-dependent and spatially varying (Cluckie and Han, 2000; Tokar and Markus, 2000), rainfall–runoff transformation is always a complex hydrological process for modelling, and there still exist inevitably various uncertainties in the hydrological forecasting (Mantovan and Todini, 2006; Han et al., 2007). In the last decade, increased scientific interest has been shown in the topics of real-time updating schemes and data assimilation methods (Refsgaard, 1997; Madsen and Skotner, 2005; Moradkhani et al., 2005; Weerts and El Serafy, 2006; Komma et al., 2007; Clark et al., 2008), probabilistic forecasting with either physical or data-driven models

(Freer et al., 1996; Krzysztofowicz, 1999, 2002; Tamea et al., 2005; Mantovan and Todini, 2006; Chen and Yu, 2007), as well as numerical weather prediction (NWP), which provides precipitation forecasts as the input of the forecasting system and allows for an extension of the forecast lead time (Cloke and Pappenberger, 2009; Wood and Schaake, 2008; Lin et al., 2002, 2006, 2010). However, no matter how advanced these methods are, and how much they can improve the forecasting results, they all depend on a reliable hydrological forecasting model, which normally consists of a rainfall–runoff model (either lumped, distributed or semi-distributed) together with a real-time updating scheme. When constructing the forecasting system, there is an important issue that cannot be avoided, i.e. the selection of the time interval of the model input data (e.g. the rainfall, streamflow and evaporation, etc.) used to drive the system, which is, however, mostly ignored by the hydrological community in operational applications. A simple search using Google Scholar with the keywords “hydrological forecasting” reveals 84 500 articles covering scientific publications on hydrological forecasting in the last 30 yr and beyond. However, there is seldom the mention of “time interval”. Not only “data time interval” is seldom mentioned; also “model time interval” and “model time step” are rarely mentioned.

Many real-time hydrological forecasting systems tend to use data as they are originally measured. Take the National Flood Forecasting System (NFFS) (Werner et al., 2009) of the Environmental Agency (EA) as an example. The NFFS system provides a platform capable of using a standard set of EA-approved models to make real-time flood forecasts for eight regions in England and Wales. It has roughly been pointed by Sene et al. (2009) that for large, slowly responding rivers, an hourly time step is sufficient, while a shorter time step is required on flashier catchments. However, since rainfall–runoff data are measured at a minimal time interval of 15 min in the UK, the 15 min time step is normally adopted when forecasts are made by the NFFS system (Sene et al., 2010). Examples are taken from NFFS in the study of Weerts et al. (2011) to estimate the predictive uncertainty in flood forecasting. A variety of catchments with different sizes (30–1000 km<sup>2</sup>) and hydrological behaviours are selected from three EA regions: the North East, Midlands and Thames regions, which are respectively modelled by three rainfall–runoff models: the probability distributed model (PDM) (Moore, 2007), MCRM (Midlands Catchment Runoff Model) (Bailey and Dobson, 1981) and nested TCMs (Thames Catchment Models) models (Wilby et al., 1994). The observed rainfall–runoff data are directly used for calibration and forecasting with different lead times (0–48 h). Similarly, in the study of Schellekens et al. (2011), the whole Thames region is chosen to test the use of MOGREPS ensemble rainfall forecasts in the NFFS system. A configuration of 148 TCM models is employed with a deterministic 15 min time step to make forecasts for the range of 0–3 days.

A distinction should be made here between the sampling interval of the hydrological measurements and the time interval of the model input data. A sampling interval is determined by the instruments (e.g. the rain gauge and the flow gauge) and could be as frequent as the instrument can afford, while the time interval of the model input data refers to the time interval of the hydrological data (e.g. rainfall and streamflow) used to construct the forecasting system, or equivalently, the time step at which the forecasting system is operating. For an easier quotation, we use the term “data time interval” hereafter to refer to the interval of the model input data, and use “sampling interval” to represent the raw hydrological measurements.

Traditionally, hydrological measurements have been carried out manually with very low sampling rates (e.g. monthly, weekly and daily). However, modern telemetry systems have been developed to increase the hydrological sampling rate to per hour or minute or even in seconds (e.g. weather radar is able to measure rainfall once every one minute, while an optical rain gauge can measure in seconds). One burning question for hydrologists would be how to select the optimal time interval in the hydrological forecasting. If the modern telemetry system keeps advancing at its current pace, will our hydrological forecasting model be able to cope with this “data-rich” environment? Although continuous-time models can avoid the time interval issue and do not suffer from numerical problems of having a time step that is too fine (Young, 2011), they are not commonly used in operational hydrological forecasting, which is carried out by discrete digital computers. It is interesting to note that such a problem has already concerned modern control engineers. Unlike the “slow” sampling rate in hydrology, control engineers need to deal with very fast sampling, e.g. hundreds or even thousands of samples per second in rocket trajectory control. In the work of Åström (1969), a simple Gauss–Markov process was analysed using a parametric model, with the parameters completely describing the stochastic process. It was found that there was an optimal choice of the time interval and that the variance of parameter estimates would increase rapidly when the time interval increased or decreased from the optimal value. This is the first time the “optimal time interval” has been proposed with regard to the data used for model construction. Later on, Ljung (1987, 1999) pointed out that data with too large intervals would bear little information about the dynamics of the sampled signal, while too fast sampling would involve more measurement noise and lead to numerical problems in a model, as the model could only fit in high-frequency bands with poor performances for extra work. It was further proposed by Ljung that the optimal choices of the time interval would lie in the range of the time constant of the system, which measures how quickly the system responses. In the case of hydrological systems, the time constant may refer to the measurement of the catchment response in the rainfall–runoff transformation, i.e. the catchment concentration time. An overestimation of the optimal

time interval in either linear or nonlinear systems may lead to a dramatic increase of the variances for the model parameter estimates. It was also pointed out by Clark and Kavetski that for highly nonlinear systems, the optimal time interval should be even smaller than the system time constant (i.e. the catchment concentration time) due to the problem of solving the nonlinear ordinary differential equations numerically (Clark and Kavetski, 2010; Kavetski and Clark, 2010).

In the past decade, this issue has begun to gain attention from hydrologists in academia. A general consensus has been reached that the parameters, simulation results and process representations of the hydrological models are inherently and strongly time-step-dependent (e.g. Duan et al., 2006; Merz et al., 2009, 2011). More than a decade ago, Schaake et al. (1996) developed a simple water balance model and tested its sensibility for operating at time steps different from what it was calibrated at. It was recommended that in order to obtain the best performance, the model should be calibrated at the same time step as the one used for operation. Following this, a number of studies have been carried out on the time-step dependency of the hydrological model parameters (Finnerty et al., 1997; Tang et al., 2007; Littlewood and Croke, 2008; Wang et al., 2009; Cho et al., 2009; Kavetski et al., 2011), for its implications in regionalisation at ungauged catchments or climate impact analysis, etc. However, the impact of the model time step, or using the equivalent expression in this study, i.e. the time interval of the model input data, on the inference of the catchment structure, model parameters and, more generally, the catchment behaviour remains poorly understood (Kavetski et al., 2011). Furthermore, in operational applications, scientists are more interested in the improvement of the model accuracy and the appropriate use of the observational data rather than reducing the time dependency of the model parameters. Although the utilisation of hydrological observations at appropriately fine resolutions is advocated by Wagener et al. (2010), there still remains a lack of general guidance allowing for hydrologists to better cope with this time interval issue in operational applications. Moreover, most of the previous studies are focused on the use of hydrological models for simulation. With respect to hydrological forecasting, this issue becomes more complicated by the involvement of another temporal concept – the forecast lead time. In such cases, a suitable choice of the time interval of the model input data is much more important, and deserves further considerations.

In the area of hydrological forecasting, it has been firstly addressed in the work of Remesan et al. (2010) that when the artificial neural network (ANN) is used for real-time flood forecasting, there is an optimal time interval for the input data. It has been found that the 30 min interval can produce more accurate forecast results than the 15, 60 and 120 min intervals in the Brue catchment of the UK, and the significance of the time interval impact on the forecasting accuracy is more prominent for longer lead times than shorter ones. The results are quite meaningful to data-driven mod-

els, the performances of which are highly time-dependent, and improper data inputs could easily lead to overfitting (a serious weakness associated with such models). However, a question might arise: will this “optimal time interval” still exist for forecasting made by using the more widely applied conceptual rainfall–runoff models? The conceptual rainfall–runoff models encompass a broad spectrum of plausible descriptions of the physical rainfall–runoff processes, which are found to be both reliable and effective in various situations, especially for real-time hydrological forecasting (Bell et al., 2001). The purpose of this study is to explore the general impact of the data time interval on hydrological forecasting using the conceptual rainfall–runoff model and a real-time updating scheme.

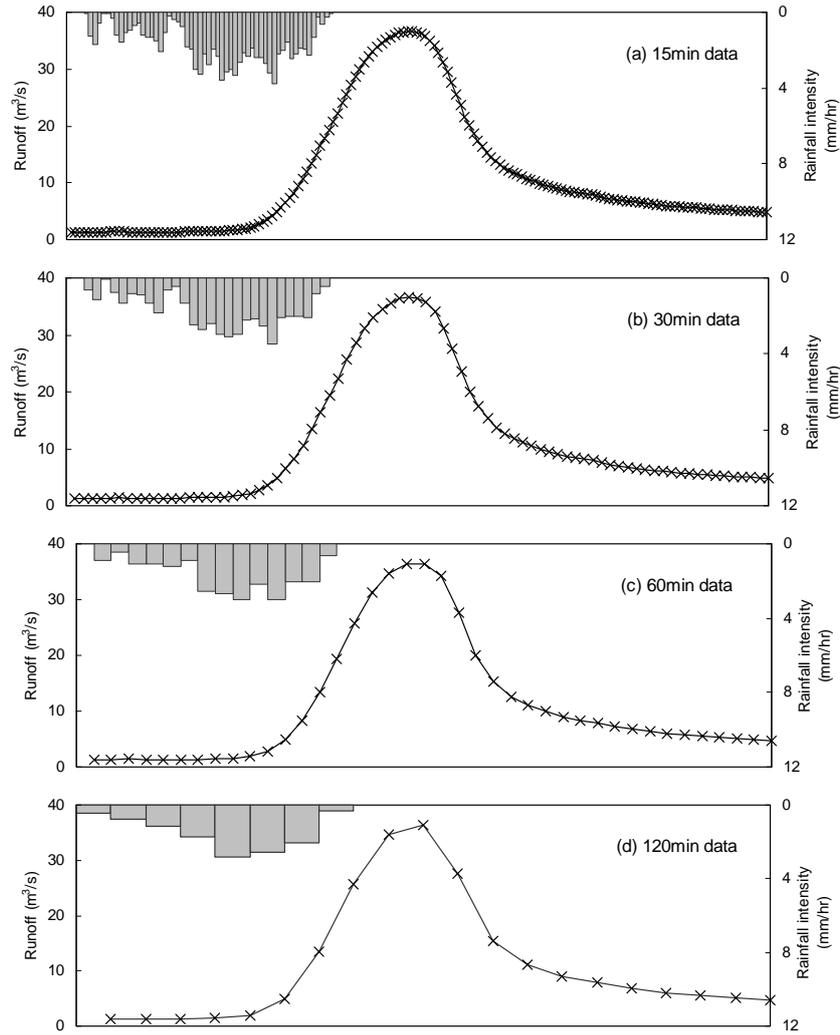
In this study, the intrinsic properties of observed rainfall–runoff data with different sampling intervals are firstly investigated from the perspective of the information loss using the spectral analysis tool of discrete wavelet transform. Then the impact of a data time interval on the accuracy of real-time hydrological forecasting is examined through case studies. A real-time forecasting system is constructed by incorporating the PDM with a real-time updating scheme, the autoregressive moving-average (ARMA) model. Rainfall–runoff data with different time intervals of 15, 30, 45, 60, 90 and 120 min are taken from four UK catchments with different drainage areas, and then used to drive the constructed forecasting system. Forecasts are made using different time-interval data with the forecast lead time ranging from 1 to 12 h. The underlying relation between the optimal data time interval and the forecast lead time is investigated, and the implications of the catchment concentration time on the selection of the optimal time interval is further discussed. Finally, based on the spectral analyses of the rainfall–runoff observations and the forecasting results from the case studies, a hypothetical pattern is proposed to describe the general impact of the data time interval and to provide hints on the selection of the optimal time interval in real-time hydrological forecasting.

## 2 Material and methods

### 2.1 Spectral analysis for rainfall–runoff data with different time intervals

#### 2.1.1 Sampling methods for rainfall–runoff observations

Sampling refers to the process of converting a signal (e.g. a function of continuous time or space) into a numerical sequence (a function of discrete time or space) (Shannon, 1949, 1998). Hydrological processes are continuous in the time domain; however, most of the observed hydrological measurements are only available at discrete time intervals. There are two methods by which a continuous time function can be represented in a discrete time domain. One is to use



**Fig. 1.** Observed rainfall–runoff data with different sampling intervals of (a) 15 min, (b) 30 min, (c) 60 min and (d) 120 min.

a sampling data function where the value of the continuous function  $Q(t)$  in the  $j$ th time interval  $Q_j$  is given simply by the instantaneous value of  $Q(t)$  at time  $j \Delta t$ :

$$Q_j = Q(t_j) = Q(j \Delta t). \tag{1}$$

The other way is to adopt a pulse data function, in which the value of the discrete time function  $Q_j$  is given by the area under the continuous function  $Q(t)$ :

$$Q_j = \int_{(j-1)\Delta t}^{j \Delta t} Q(t) dt. \tag{2}$$

The two principal variables of interest in hydrology, streamflow and rainfall, are measured as sampled data series and pulse data series, respectively (Chow et al., 1988). When the values of streamflow and rainfall are recorded by gauges at a given instant, the flow gauge value represents the flow rate at that instant (in dimensions of  $(L^3/T)$ ), while the rainfall

gauge value is the accumulated depth of rainfall (in dimension of  $(L)$  standing for the volume of rainfall  $(L^3)$ ) which has occurred up to that instant.

By representing a continuous signal using discrete measurements, the sampling process always results in some kind of information loss. Figure 1 shows the rainfall–runoff measurements with sampling intervals of 15, 30, 60 and 120 min. It can be seen that as the sampling interval increases from 15 to 120 min, the flow curve is less smooth and the shape of the columns representing the rainfall volume becomes more approximate. Slower sampling leads to subsets of the datasets produced by fast sampling, and hence is less informative (Ljung, 1991). However, from the theoretical aspect of the signal reconstruction, downsampled data can also result in a perfect reconstruction of the original signal by choosing an appropriate interpolation method and more importantly, complying with the sufficient condition of the Nyquist–Shannon sampling theorem (Nyquist, 1928, 2002).

The theorem, often known as the sampling theorem, provides a lower boundary of the sampling frequency and thus an upper boundary of the sampling interval as a sufficient condition for a perfect reconstruction of the original signal in the absence of observational noise:

$$f_s > 2B, \tag{3}$$

or equivalently,

$$B < f_s / 2, \tag{4}$$

where  $f_s$  is the sampling frequency and  $B$  is the one-sided baseband bandwidth of the band-limited original signal.  $f_s / 2$  is defined as the Nyquist frequency, which is a property of the sampled system. If the condition is not satisfied, then the part of information of the original signal with frequencies beyond the Nyquist interval  $(-f_s / 2, f_s / 2)$  will be lost during sampling, and the spectral density inside the interval can be distorted in the process of signal reconstruction, which will lead to the phenomenon of aliasing (Mitchell et al., 1988). On the other hand, if the frequency of the downsampled data is higher than twice the bandwidth of the original signal, then the information loss can be neglected, and the original signal can also be successfully reconstructed. This has, at least from the theoretical aspect of the signal reconstruction, refuted the intuition of many modellers, and demonstrated that choosing a relatively larger time interval will not always deteriorate the modelling results.

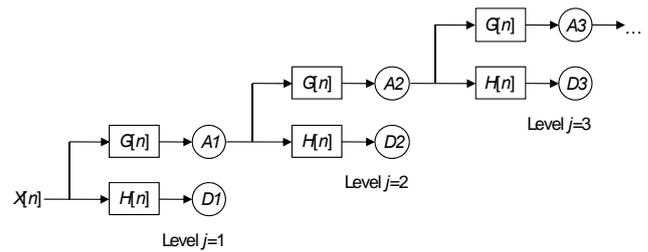
### 2.1.2 Discrete wavelet transform

In order to further investigate the information content of the observed rainfall–runoff data with different sampling intervals before they are used in hydrological forecasting, discrete wavelet transform (DWT) is applied to investigate the energy distribution of the observed rainfall–runoff data in different frequency domains. DWT is a powerful mathematical tool for the spectral analysis of discrete signals, which is more efficient than the Fourier transform in studying non-stationary time series (Meyer, 1993; Polikar, 1999). A most popular and efficient way to implement DWT is the Mallat decomposition algorithm (Mallat, 1989), the process of which is illustrated in Fig. 2.

The decomposition level  $j$  is associated with a frequency band  $\Delta F$  calculated based on the sampling frequency  $f_s$ :

$$2^{-j-1} f_s \leq \Delta F \leq 2^{-j} f_s. \tag{5}$$

The original signal is first decomposed into an approximation and an accompanying detail. The detail contains the high-frequency components of the signal within the frequency band  $\Delta F$ , while the approximation represents the low-frequency components below the band. Table 1 gives the corresponding frequency bands of the six wavelet decomposition levels for a 15 min data series obtained by using



**Fig. 2.** Process of the discrete wavelet decomposition following the Mallat decomposition algorithm.  $X[n]$  is a discrete signal with  $n$  samples, passing through a low-pass filter  $G$  and a high-pass filter  $H$  with impulse responses of  $G[n]$  and  $H[n]$ , respectively.  $A1, A2, A3$  and  $D1, D2, D3$  are the decomposed approximations and details on level  $j = 1, 2, 3, \dots$

Eq. (5). The decomposition process is iterated, with successive approximations being decomposed in turn so that the original signal is broken down into many lower-resolution components. In this study, the decomposition is carried out by decomposing 1 yr observed rainfall–runoff data into six decomposition levels using the Daubechies wavelet of order 10, 12 and 20 (Daubechies, 1990; Labat et al., 2000, 2004). Since we are only interested in the general frequency profile instead of the frequency distribution details, the six wavelet decomposition levels are sufficient for this purpose. The decomposition results are shown in Sect. 3.1 with explorations on the spectral characteristics of rainfall–runoff data with different time intervals.

## 2.2 Structure of the real-time hydrological forecasting system

Investigations on the impact of data time interval in this study are carried out based on a real-time hydrological forecasting system, which is constructed by integrating a conceptual rainfall–runoff model, the PDM, with a commonly used real-time updating scheme, the ARMA model.

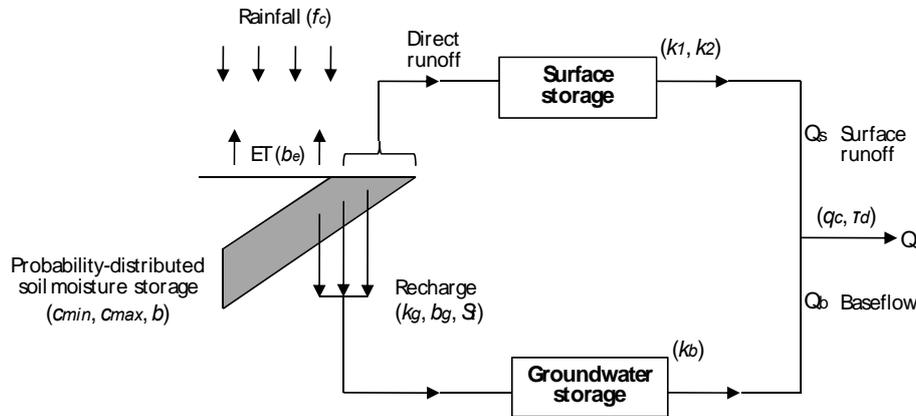
### 2.2.1 The rainfall–runoff model: PDM

The PDM (Moore, 2007) is a lumped conceptual rainfall–runoff model which has been widely applied in various catchments in the UK and abroad. It could be viewed as a representative of the conceptual saturation-excess models used for the runoff simulation in humid and semi-humid regions. The PDM model is developed based on the scheme of the Xinanjiang model with a soil moisture storage capacity that varies over the catchment and is described by a simple probability distribution curve. Inputs to the model are the rainfall and the potential evaporation. Actual evaporation depletes the soil moisture storage, the rate of which depends on the soil moisture deficit and the potential evaporation. Further loss as recharge to the groundwater is defined by assuming that the rate of draining depends linearly on the current

**Table 1.** Frequency bands of the wavelet decomposition levels for data with a sampling interval of 15 min.

Level $j$	$2^{-j-1}$	$2^{-j}$	Sampling rate $f_s$ (Hz)	Frequency band $\Delta F(\times 10^{-6}$ Hz)
1	0.25	0.5	1/900 s	[278, 556]
2	0.125	0.25	1/900 s	[139, 278]
3	0.0625	0.125	1/900 s	[69, 139]
4	0.03125	0.0625	1/900 s	[35, 69]
5	0.015625	0.03125	1/900 s	[17, 35]
6	0.007813	0.015625	1/900 s	[9, 17]

Note: for a certain decomposition level of the wavelet analysis, the detail contains the components of the original signal within the relevant frequency band, while the approximation represents the components below the frequency band.



**Fig. 3.** Conceptual structure of the PDM model. The 13 model parameters to be calibrated are in the brackets.

soil moisture conditions. The groundwater recharge is then routed through a “slow response system”, which can be best represented by a cubic form of a nonlinear storage model (Dooge, 1973). The direct flow, defined as the difference between the soil moisture storage at the beginning and the end of a time interval, is routed through “a fast response system”, which is often described by a cascade of two linear reservoirs (O’Connor, 1982). Further details of the model equations are presented by Moore (2007). Figure 3 shows the conceptual structure of the PDM model. Table 2 lists the 13 parameters to be calibrated in the model.

**2.2.2 The real-time updating scheme: ARMA**

Error prediction is now a well-established technique for forecast updating in real time (Box and Jenkins, 1970; Moore, 1982). The real-time updating scheme used in this study is wholly external to the deterministic model operation and thus can be easily applied in combination with any kind of rainfall–runoff models. A feature of errors from a conceptual rainfall–runoff model is that there is a tendency for errors to persist so that sequences of positive errors (underestimation) or negative errors (overestimation) are common (Moore, 2007). This dependent structure in the error sequence may be exploited by developing an error predictor which incorporates the error structure and allows for future

errors to be predicted. In this updating scheme, the structure of errors is analysed and the predictions of future errors are added to the deterministic model prediction to obtain the updated and improved forecasts of flow. With  $q_{t+l}$  representing the modelling result of the observed flow  $Q_{t+l}$  at a certain time  $t+l$ , directly obtained from the PDM model without incorporating the observed flow data, the error  $\eta_{t+l}$  is defined as  $Q_{t+l} - q_{t+l}$ . Let  $\eta_{t+l|t}$  denote a prediction of the error  $\eta_{t+l}$ , which is made  $l$  steps ahead from a forecast origin at time  $t$  using an error predictor. Then a real-time forecast,  $q_{t+l|t}$ , can be expressed as the sum of the predicted error and the original modelling result:

$$q_{t+l|t} = q_{t+l} + \eta_{t+l|t} \tag{6}$$

Among various forms of error predictors, the ARMA model is considered to be most appropriate and parsimonious (Moore, 2007). The equations of the error predictions  $\eta_{t+l|t}$  expressed by the ARMA model can be written as follows:

$$\eta_{t+l|t} = -\varphi_1\eta_{t+l-1|t} - \varphi_2\eta_{t+l-2|t} - \dots - \varphi_p\eta_{t+l-p|t} + \theta_1a_{t+l-1|t} + \theta_2a_{t+l-2|t} + \dots + \theta_qa_{t+l-q|t} \quad l = 1, 2, \dots, s \tag{7}$$

where  $\varphi_1, \varphi_2, \dots, \varphi_p$  and  $\theta_1, \theta_2, \dots, \theta_q$  are autoregressive and moving-average parameters respectively, with

$$a_{t+l-i|t} = \begin{cases} 0 & l-i > 0 \\ a_{t+l-i} & \text{otherwise,} \end{cases} \tag{8}$$

Table 2. Parameters in the PDM model (Moore, 2007).

Parameter	Unit	Suggested value	Description
$f_c$	none	1	Rainfall factor
$\tau_d$	h	0	Time delay
$c_{min}, c_{max}$	mm	0	Minimum and maximum soil moisture store capacity
$b$	none	0.5	Exponent of the soil moisture distribution
$b_e$	none	2.5	Exponent in the actual evaporation function
$k_g$	h mm <sup>bg-1</sup>	10 <sup>5</sup>	Groundwater recharge time constant
$b_g$	none	1.5	Exponent of the groundwater recharge function
$S_t$	mm	0	Soil tension storage capacity in the recharge function
$k_1, k_2$	h	1–20	Time constants of the surface routing
$k_b$	h mm <sup>2</sup>	5–100	Time constant of the groundwater storage routing
$q_c$	m <sup>3</sup> s <sup>-1</sup>	0	Constant flow representing returns/abstractions

Table 3. Definitions of the FEH catchment descriptors (Bayliss, 1999).

Catchment descriptors	Unit	Definitions
AREA	km <sup>2</sup>	Catchment area
ALBAR	m	Mean altitude of the catchment above sea level
LDP	km	Longest drainage path, defined by recording the greatest distance from a catchment node (50 m grid) to the catchment outlet
DPLBAR	km	Mean drainage path length, calculated as the mean distances between each catchment node and the outlet
DPSBAR	m km <sup>-1</sup>	Mean drainage path slope, calculated as the mean of all the inter-nodal slopes, which characterises the overall steepness of the catchment
ASPVAR	none	Invariability of the inter-nodal slope directions, where values near 0 indicate there is considerable variability in the aspect of catchment slopes. Values approaching 1 indicate that catchment slopes tend to face one particular direction.
SAAR	mm	Standard average annual rainfall (1961–1990)
QMED	m <sup>3</sup> s <sup>-1</sup>	Median annual maximum flood (1961–2008)
PROPWET	none	Proportion of time when SMD (soil moisture deficit) was equal to or below 6 mm during the period 1961–1990
SPRHOST	%	Standard percentage runoff derived by using the HOST (Hydrology Of Soil Types) soil classification
URBEXT	none	Extent of urban and suburban land cover (1990)

and  $a_{t+l-i}$  is the one-step ahead prediction error defined as

$$a_{t+l-i} \equiv a_{t+l-i|t+l-i-1} = \eta_{t+l-i} - \eta_{t+l-i|t+l-i-1} \quad (9)$$

$$= Q_{t+l-i} - q_{t+l-i|t+l-i-1}$$

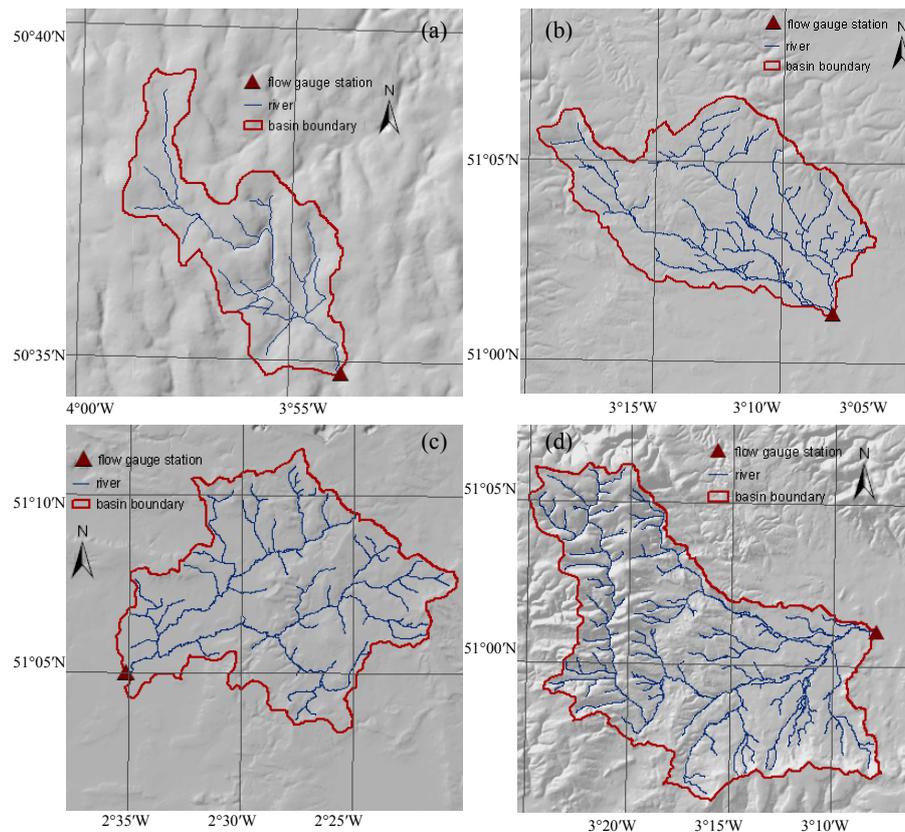
and

$$\eta_{t+l-i|t} = \eta_{t+l-i} = Q_{t+l-i} - q_{t+l-i} \text{ for } l-i \leq 0. \quad (10)$$

Equation (7) together with the related Eqs. (8–10) is used recursively to produce the error predictions of  $\eta_{t+1|t}, \eta_{t+2|t}, \dots, \eta_{t+l|t}$ , from the available values of  $a_t, a_{t-1}, \dots$  and  $\eta_t, \eta_{t-1}, \dots$ . Using this error prediction methodology, the PDM model original modelling results,  $q_{t+l}$ , can be updated using the error prediction  $\eta_{t+l|t}$ , to calculate the required real-time forecast,  $q_{t+l|t}$ , according to Eq. (6).

As for the number of parameters in the ARMA structure, i.e.  $\varphi_1, \varphi_2, \dots, \varphi_p$  and  $\theta_1, \theta_2, \dots, \theta_p$  in Eq. (7), a third-order autoregressive with dependence on three past model errors has

been proved to be an appropriate choice for UK conditions (Moore, 2007). Thus the ARMA structure containing three autoregressive parameters and one moving-average parameters (with  $p = 3$  and  $q = 1$ ) is chosen as the updating scheme in this study. It should be noted that for the forecasts made from an origin  $t$ , by calculating the error predictor  $\eta_{t+l|t}$ , the real-time observations of flow are assimilated into the forecasting results. For instance, with the structure of ARMA (3, 1), the observed flows at  $t, t - 1$  and  $t - 2$  are involved in the calculation of the error predictor  $\eta_{t+l|t}$ , which are then added to the original model prediction  $q_{t+l}$  to derive the updated result of  $q_{t+l|t}$  according to Eq. (6). For the obtaining of  $q_{t+l}$ , we assume the perfect knowledge of the future rainfall and potential evaporation, i.e. the observed catchment average rainfall and the MOSES (Met Office Surface Exchange Scheme) potential evaporation are used as the model inputs after the origin  $t$ .



**Fig. 4.** Locations and configurations of the four UK catchments in case studies with river networks and the flow gauging stations at the catchment outlets (a) Bellever, (b) Halsewater, (c) Brue and (d) Bishop\_Hull.

**Table 4.** Flow gauge locations and values of the FEH catchment descriptors for the four catchments in the case studies.

		(A)	(B)	(C)	(D)
		Bellever	Halsewater	Brue	Bishop_Hull
Flow gauge location	Latitude	50.582° N	51.022° N	51.075° N	51.019° N
	Longitude	3.898° W	3.133° W	2.587° W	3.134° W
Catchment descriptors	AREA (km <sup>2</sup> )	21.5	87.8	135.2	202.0
	ALBAR (m)	459	109	104	144
	LDP (km)	13.46	19.40	22.61	40.21
	DPLBAR (km)	6.28	9.57	13.62	17.75
	DPSBAR (m km <sup>-1</sup> )	94.9	85.7	71.1	98.0
	ASPVAR	0.25	0.30	0.16	0.17
	SAAR (mm)	2095	851	857	964
	QMED (m <sup>3</sup> s <sup>-1</sup> )	37.4	12.2	36.2	43.7
	PROPWET	0.46	0.35	0.37	0.36
	SPRHOST (%)	47.5	30.6	36.4	32.9
	URBEXT	0	0.006	0.007	0.007

An automatic calibration procedure utilising an efficient population-based evolutionary optimisation technique, i.e. the particle swarm optimisation (PSO) algorithm (Eberhart and Kennedy, 1995, 2001), is adopted to estimate the four

error-prediction parameters of the ARMA model and the 13 parameters of the PDM model.

## 2.3 Case study descriptions

### 2.3.1 Study catchments

Four catchments are selected from South West England, i.e. Bellever, Halsewater, Brue and Bishop\_Hull, to carry out the case studies based on the constructed real-time forecasting system. They have different sizes of the drainage areas, varying from 20 to 200 km<sup>2</sup>, as shown in Fig. 4. Except for the Bellever catchment, which is in the Devon area, with a maximum altitude of 604 m above sea level, the other three catchments are located in the area of North Wessex, with an average altitude of 119 m. All the catchments are predominantly rural areas with the main land-use types being the moorland, low-grade agriculture or woodland. Catchment descriptors are obtained from the Flood Estimation Handbook (FEH) (Bayliss, 1999) produced by the Centre of Ecology and Hydrology (CEH) in the UK. Meanings of the descriptors are explained in Table 3, and values of the respective descriptors for the four catchments are listed in Table 4. It can be seen from Table 4 that the average annual rainfall (SAAR) is 2095 mm for the Bellever catchment with the percentage runoff (SPRHOST) being 47.5 %, while the other three catchments have less rainfall (about 850–1000 mm per year) and slightly lower percentage runoff, i.e. 30.6 %, 36.4 % and 32.9 %. From the descriptors of the catchment size and configuration, it can be noticed that the increase of the catchment area (AREA) corresponds to an increase of LDP and DPLBAR, representing the longest and average length of the drainage path, which thus suggests an increasing travel time of the streamflow before it routes to the catchment outlet.

### 2.3.2 Data used in the case studies

Seven years of rainfall–runoff data are collected from the four catchments with a sampling interval of 15 min. The period is from October 1998 to September 2005 for Bellever, Halsewater and Bishop\_Hull. Data from the HYREX (Hydrological Radar Experiment) project funded by the NERC (Natural Environment Research Council) are used for the Brue catchment from September 1993 to May 2000. All the data are downsampled into other sequences with time intervals of 30, 45, 60, 90 and 120 min. Daily potential evaporation data are obtained from the MOSES and then disaggregated into the same time steps as the rainfall.

For each catchment, 1 yr data are selected for the validation of the calibrated parameters of the rainfall–runoff model and the real-time updating scheme. Another 1 yr is used for independent evaluation of the model performance (i.e. the forecast accuracy). The remaining data of nearly five years are used for calibration. It is widely accepted that the information quality of the calibration data (i.e. the representativeness of the catchment's hydrological responses) is of more importance in deciding the performance of the calibrated model than the data length (Gupta and Sorooshian, 1985a, b).

According to the conclusions of Liu and Han (2010), a calibration dataset of 12 months contains sufficient information of hydrological variability to result in a reliable and stable model in the Brue catchment. In order to reduce the burden of the calculation work and improve the efficiency of data utilisation, in this study the most appropriate 1 yr data are selected from the calibration datasets for calibration using an effective selection index called the information cost function (ICF). More detailed information of the ICF index can be found in the work of Liu and Han (2010).

The 5 yr calibration data are first split into a group of calibration scenarios with the fixed length of 12 months, using a moving window of one month. As a result, up to 50 calibration scenarios are resulted for each catchment. With the 1 yr validation data determined beforehand, the ICF index is used to select the most appropriate 10 calibration scenarios with sufficient hydrological information for calibration. Since the ICF index can only identify the relatively better calibration scenarios and not the absolutely best one, the 10 scenarios initially selected by ICF are used to carry out the calibration procedure. The best three calibrated models which have the best performances on the validation data are chosen to perform the real-time forecasting 1–12 h ahead using the testing data, an independent 1 yr dataset of the calibration and validation data. Finally, the forecasting performances of the best three models are averaged. This is to present more stable results, and in that way the uncertainty of calibration data, which is unavoidable in practice, can be involved in the final results.

It should be mentioned that the whole process described above is carried out for all the four catchments and repeated using data with different time intervals of 15, 30, 45, 60, 90 and 120 min. That is to say, for each data sequence with a certain time interval, the calibration and validation procedures are carried out to construct a forecasting system that is suitable to use with future data of the same time interval, and the real-time forecasting is made with the constructed system functioning at the same time step.

## 3 Results and discussion

### 3.1 Spectral differences of rainfall–runoff data with different time intervals

As mentioned in Sect. 2.1.2, the spectral analysis using discrete wavelet transform is carried out for 1 yr rainfall–runoff observations from the Brue catchment. The observed data are first obtained with a sampling interval of 15 min, and then downsampled to 30, 45, 60, 90 and 120 min data series according to different sampling methods of rainfall and streamflow described in Sect. 2.1.1. The decomposition results are shown in Table 5. A detailed data processing procedure can be found in the explanation below the table. Since the results from the Daubechies wavelet of order 10, 12 and 20

**Table 5.** Energy distributions in different frequency bands for 1 yr rainfall–runoff observations with different sampling intervals of 15, 30, 45, 60, 90 and 120 min.

Approx.	Frequency band ( $10^{-6}$ Hz)	Details						Total energy [0, 566]	
		[0, 9]	level 6 [9, 17]	level 5 [17, 35]	level 4 [35, 69]	level 3 [69, 139]	level 2 [139, 278]		level 1 [278, 566]
flow	15 min	356 196	18 422	6848	543	39	3	1	382 052
	30 min	356 176	18 375	6795	529	35	2	1	381 913
	45 min	356 115	18 298	6705	504	29	2	1	381 654
	60 min	356 026	18 175	6573	471	22	2	2	381 271
	90 min	355 880	17 892	6253	395	18	7	2	380 447
	120 min	355 718	17 477	5804	311	35	6	2	379 353
rain	15 min	112	56	71	80	86	78	43	526
	30 min	112	56	71	77	73	41	30	460
	45 min	112	56	70	71	57	33	16	415
	60 min	112	55	68	66	50	15	13	379
	90 min	112	55	63	50	29	8	10	327
	120 min	112	54	59	42	18	7	7	299

Note: the observed data are taken from the Brue catchment of the UK with a sampling interval of 15 min. Four-year rainfall–runoff data (19 September 1993 to 19 July 1997) are used to form 47 sets of 1 yr data with a moving window of one month. With the data collection methods of streamflow and rainfall described by Eqs. (1) and (2), each dataset is downsampled into another five series with time intervals of 30, 45, 60, 90 and 120 min. For a comparable analysis and a convenient calculation of the wavelet decomposition, extra data are added into the 30, 45, 60, 90 and 120 min data series in order to make them have the same amount of data as the 15 min series by either using a simple linear interpolation for the streamflow series or interpolating constant values for the rainfall. In such a case, all the data series have the same sampling interval of 15 min, and thus a sampling frequency of 1/900 Hz. The 47 pairs of data series of streamflow and rainfall with a certain time interval are used to perform the wavelet decomposition on 6 levels, which refer to the frequency bands of  $[278, 566] \times 10^{-6}$  Hz,  $[139, 278] \times 10^{-6}$  Hz,  $[69, 139] \times 10^{-6}$  Hz,  $[35, 69] \times 10^{-6}$  Hz,  $[17, 35] \times 10^{-6}$  Hz and  $[9, 17] \times 10^{-6}$  Hz from level 1 to level 6. Finally, the decomposition results are averaged for the 47 pairs of data series with a certain time interval, and the energy values are calculated for each frequency band based on the wavelet coefficients.

**Table 6.** Lead steps of the forecasting system with different time intervals.

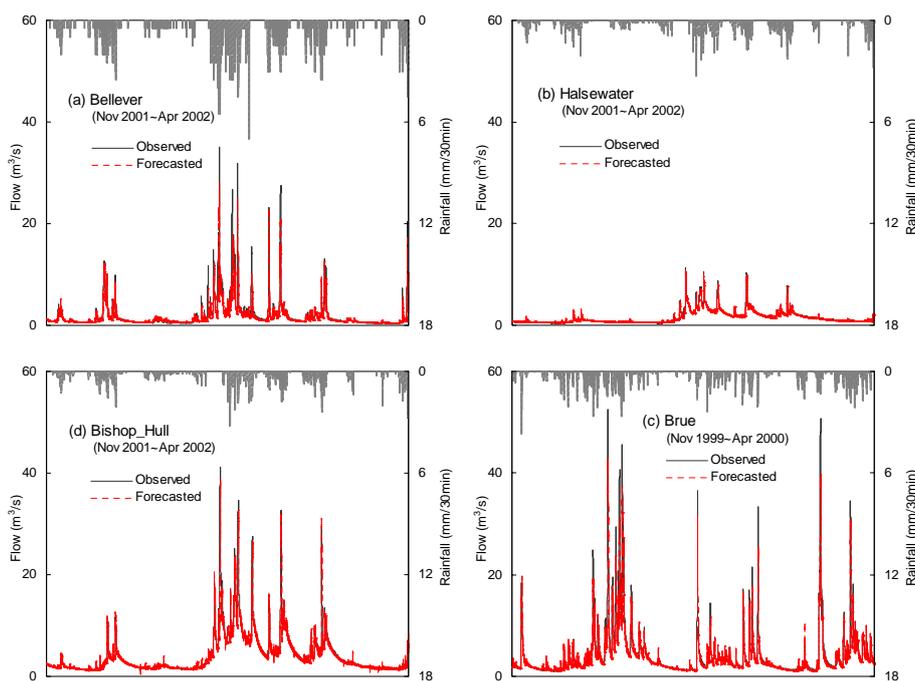
	1 h (60 min)	2 h (120 min)	3 h (180 min)	4 h (240 min)	5 h (300 min)	6 h (360 min)	9 h (540 min)	12 h (720 min)
15 min	4	8	12	16	20	24	36	48
30 min	2	4	6	8	10	12	18	24
45 min	–	–	4	–	–	8	12	16
60 min	1	2	3	4	5	6	9	12
90 min	–	–	2	–	–	4	6	8
120 min	–	1	–	2	–	3	–	6

have very similar patterns, only those from the Daubechies wavelet of order 20 are presented.

The energies shown in Table 5 can be interpreted as the magnitudes of information content of the 1 yr data in different frequency bands. For the flow, the majority of energy is distributed in the lower bands, with little energy in the higher bands of level 1 to level 4. As for the rainfall, the energy distribution is relatively balanced, with considerable amounts in all frequency bands. It has been pointed out by some studies (Bras, 1979; Bras and Rodriguez-Iturbe, 1976; Storm, 1989) that the catchment behaves like a low-pass filter to the climatic input data, e.g. rainfall and evapotranspiration, by absorbing their subtle time variability. To further investigate the energy variance in a certain frequency band caused by the flow and rainfall data with different time intervals, it can be noticed that the variances for the flow series are not ob-

vious, considering the relatively large amounts of the total energies; however, with respect to the rainfall, the variances are more outstanding, which are on an increasing trend from the approximation to lower decomposition levels. In the low-frequency band of  $[0, 35 \times 10^{-6}]$  Hz (approximation + level 6 + level 5), where most of the energy of the flow data is distributed, the energy variance of the rainfall series is less obvious compared to that in higher frequency bands. Moreover, the energies of approximations are exactly the same for all the rainfall series. The difference of energy distribution for the flow and rainfall data with different intervals might be caused by the intrinsic characteristics of the two different hydrological processes, and also the different sampling methods.

Similar to the low-pass-filtering function of the catchment, the high-frequency variances of the rainfall inputs can be



**Fig. 5.** Six-month length hydrographs of the observed rainfall–runoff data versus the best results of the 4 h ahead forecasts using the 30 min data. The NSE values calculated between the observed and the forecasted flows are 0.8932, 0.9544, 0.9167 and 0.9420 respectively for the four catchments in (a), (b), (c) and (d).

absorbed by the soil moisture reservoirs in the rainfall–runoff model (Oudin et al., 2004). As a consequence, the information content of the rainfall input in the higher frequency bands, which can show large variances with regard to a different time interval, is not likely to be transformed into the simulated flow through the rainfall–runoff model. As discussed above, since the energy variances of the rainfall data in low frequency bands are not obvious, the transformed flow will not have much difference when using rainfall data with different intervals as the input of the rainfall–runoff model. This indicates that the increased sampling rate for the rainfall data may not necessarily improve the performance of the rainfall–runoff model. This has been verified by many studies; for example, Schaake et al. (1996) found that using 6 h, 12 h, 1-day, 2-day and 4-day rainfall data could generate similar simulation results of flow as long as the model was calibrated and operated at the same time intervals. However, this conclusion may only be true for pure simulation using the rainfall–runoff model. The energy distribution from the wavelet analysis is not enough to provide a general pattern for the impact of time interval on real-time forecasting, where the existence of the real-time updating scheme involves the historical flow data as another system input together with the rainfall. In the following sections, the impact of data time interval in real-time hydrological forecasting is explored through the case studies.

### 3.2 Relation between the optimal time interval and the forecast lead time

For the four catchments in the case studies, the following analyses of data time interval are based on the performances of the real-time forecasting system on the respective 1 yr testing data with lead times of 1–6, 9 and 12 h. The lead steps of the forecasting system constructed from data with different time intervals are shown in Table 6. It can be noticed that for some forecast lead times such as 1, 2, 3, 4, 5 and 9 h, there are no integers for the time intervals of 45, 90 and 120 min. Therefore, forecasts are not made for such lead times by the forecasting systems constructed from the 45, 90 and 120 min data.

The forecasting results for the four catchments are shown in Table 7, evaluated by the root-mean-square error (RMSE). Example hydrographs resulted from the best calibration scenario for the 4 h ahead forecasts using the 30 min data are shown in Fig. 5. By examining each row of Table 7, it can be seen that for a forecasting system built on data with a certain time interval, the RMSE value increases with the increase of the forecast lead time. This is the common case in real-time forecasting. The forecasting error increases as the lead time extends. The trend is more obvious in Fig. 6, shown by the slopes of curves for the Nash–Sutcliffe efficiency (NSE) statistics (Nash and Sutcliffe, 1970). For each catchment in the subfigure of Fig. 6, all the six curves revealing the relationship between the performance of the forecasting system

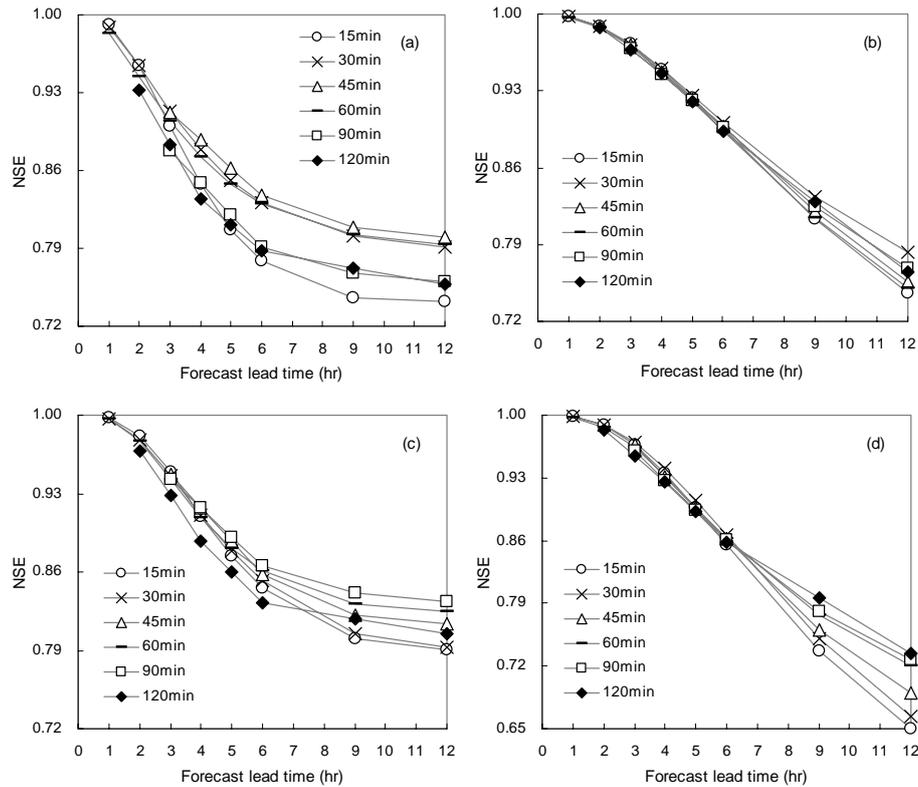
**Table 7.** Forecasting results shown in RMSE ( $\text{m}^3 \text{s}^{-1}$ ) for the four catchments using data with different time intervals. The lowest RMSE value for a certain lead time is highlighted to show the optimal data time interval.

	Forecast lead time (h)							
	1	2	3	4	5	6	9	12
(A) Bellever ( $21.5 \text{ km}^2$ )								
15 min	0.2099	0.4870	0.7232	0.8915	1.0039	1.0754	1.1530	1.1593
30 min	0.2359	0.4851	0.6699	0.7955	0.8801	0.9378	1.0200	1.0448
45 min	–	–	0.6754	–	–	0.9218	0.9981	1.0231
60 min	0.2938	0.5359	0.7073	0.8175	0.8907	0.9401	1.0143	1.0395
90 min	–	–	0.7942	–	–	1.0489	1.1083	1.1260
120 min	–	0.5833	–	0.9042	–	1.0189	–	1.0832
(B) Halsewater ( $87.8 \text{ km}^2$ )								
15 min	0.0555	0.1177	0.1859	0.2516	0.3111	0.3638	0.4863	0.5672
30 min	0.0570	0.1211	0.1883	0.2510	0.3061	0.3535	0.4585	0.5238
45 min	–	–	0.1896	–	–	0.3626	0.4827	0.5635
60 min	0.0605	0.1243	0.1930	0.2578	0.3161	0.3668	0.4842	0.5623
90 min	–	–	0.1905	–	–	0.3644	0.4802	0.5568
120 min	–	0.1246	–	0.2600	–	0.3668	–	0.5441
(C) Brue ( $135.2 \text{ km}^2$ )								
15 min	0.2641	0.7190	1.1873	1.5767	1.8617	2.0583	2.3422	2.3939
30 min	0.3172	0.7824	1.2185	1.5671	1.8263	2.0157	2.3155	2.3867
45 min	–	–	1.2121	–	–	1.9803	2.2150	2.2654
60 min	0.3208	0.8044	1.2549	1.5896	1.8114	1.9554	2.1539	2.1984
90 min	–	–	1.2584	–	–	1.9229	2.0934	2.1395
120 min	–	0.9306	–	1.7574	–	2.1469	–	2.3157
(D) Bishop_Hull ( $202.0 \text{ km}^2$ )								
15 min	0.1496	0.3849	0.6502	0.9099	1.1468	1.3567	1.8314	2.1099
30 min	0.1447	0.3677	0.6210	0.8713	1.1007	1.3048	1.7792	2.0676
45 min	–	–	0.7068	–	–	1.4327	1.8736	2.1137
60 min	0.1644	0.4088	0.6783	0.9291	1.1436	1.3213	1.6848	1.8869
90 min	–	–	0.7190	–	–	1.3865	1.7814	2.0101
120 min	–	0.4576	–	0.9703	–	1.3422	–	1.8412

and the forecast lead time are on decreasing trends. When comparing the differences between the six curves, it can be found that as the lead time increases, the distance of the curves becomes increasingly larger until finally the curves are clearly distinct from each other. This means that when the lead time is small, e.g. from 1 to 6 h, the variance of the forecasts resulted from data with different time intervals is also subtle, while with the increase of the lead time, e.g. when exceeding 6 hours, the variance becomes more and more significant. This proves that the choice of the time interval does have a considerable impact on the forecasting results, which is even more prominent with longer forecast lead times than shorter ones.

The lowest RMSE values for the forecasts made with a certain lead time are highlighted in Table 7 to show the optimal choice of the time interval for a certain lead time forecast. It can be seen that there is an increase of the optimal time

interval for all four catchments when the lead time increases from 1 to 12 h. To further illustrate this, the rankings of the time intervals according to the RMSE values are listed in Table 8. An interesting phenomenon is revealed by the rankings; that is, the optimal time interval is replaced gradually by longer ones as the increase of the forecast lead time. Using the Brue catchment as an example, the optimal interval is 15 min for 1 h to 3 h ahead forecasts, while it is replaced by 30 and 60 min for forecasting made with lead times of 4 and 5 h; later the 90 min interval replaces 60 min as the best one when the lead time is longer than 6 h. Moreover, it should be pointed out that for the 15 min interval, although it is the optimal choice for the 1, 2 and 3 h ahead forecasting, it appears to be at the bottom of the ranking when the lead time is 12 h. This again indicates the importance of selecting the appropriate time interval in real-time forecasting rather than



**Fig. 6.** Decreasing performance of the forecasting system constructed using data with different time intervals with the increase of the forecast lead time, for the catchments of Bellever, Halsewater, Brue and Bishop\_Hull, respectively in (a), (b), (c) and (d).

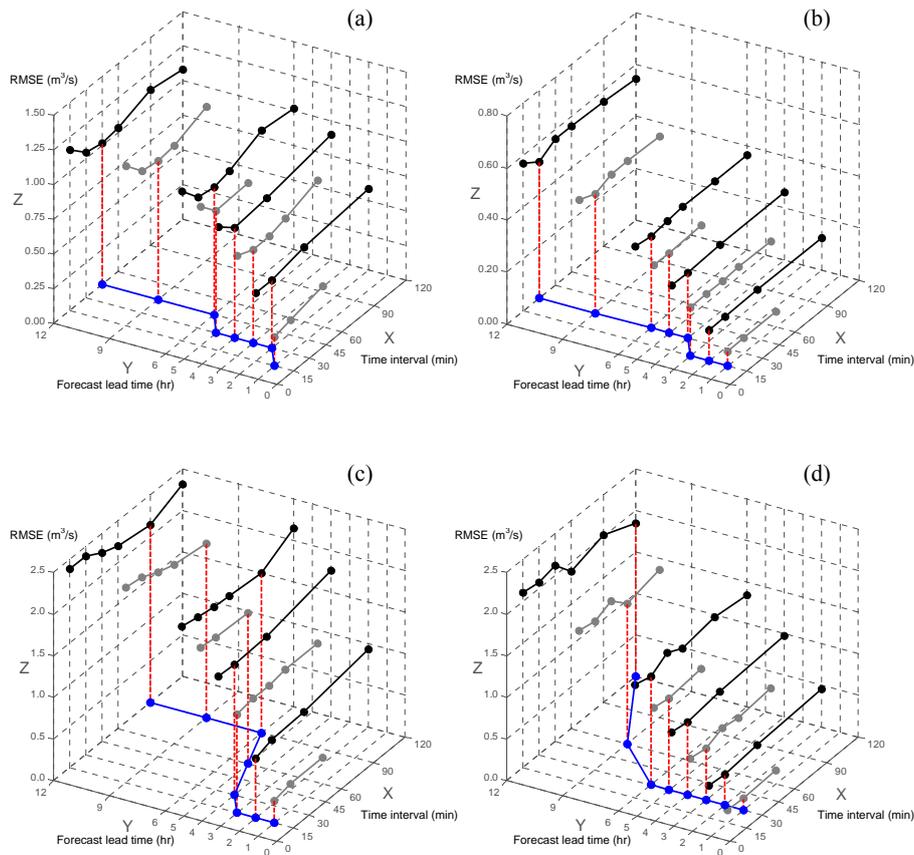
simply building the model using data as they are originally measured.

To make the patterns shown by Table 7 more obvious, the forecasting results are plotted in three-dimensional coordinates for the four catchments, as shown in Fig. 7. The *x* axis stands for the data time interval (15, 30, 45, 60, 90 and 120 min); the *y* axis represents the eight forecast lead times (1, 2, 3, 4, 5, 6, 9, 12 h); and the values on the vertical *z* axis show the forecasting results in RMSE. In each subfigure, each of the eight suspended curves indicates the variance of the forecast accuracy with respect to different time intervals for forecasts made with a certain lead time. The lowest points of the eight curves (representing the optimal time intervals which result in the least forecasting errors) are projected onto the *x*-*y* plane and then connected together. The projection curve (in blue) thus reveals the relationship between the optimal time interval and the forecast lead time. It can be seen that all the four projection curves for the four catchments are on increasing trends, which indicates the increase of the optimal time interval with the increase of the forecast lead time. This is consistent with the previous conclusions made by examining the rankings in Table 8. However, besides the positive relation between the optimal time interval and the forecast lead time, it can also be noted that the increasing patterns of the projection curves are quite dif-

ferent for the four catchments. The reason for these various increasing patterns will be fully discussed in the next section.

### 3.3 Implications of the catchment concentration time

By examining the increasing pattern of the projection curves in Fig. 7, it is interesting to note that the sharpest curve is generated by the catchment of Bishop\_Hull in Fig. 7d, which has the largest drainage area of 202.0 km<sup>2</sup> and the largest LDP (longest drainage path) value of 40.21 km (see Table 4). Further, when comparing the projection curves in Fig. 7a and c produced by Bellever and Brue, it can be found the two curves start at the same values of the optimal time interval (both at 15 min) when the lead time is within 2 h, and become stable after the lead time exceeds 6 h. However, the middle section of the curve in Fig. 7c, which rises to 90 min at the lead time of 6 h, is much steeper than that of the curve in Fig. 7a, which only increases to 45 min at the same lead time. This may also be explained by the concentration times of the two catchments. From Table 4, it can be seen that Brue has a larger area of 135.2 km<sup>2</sup> with higher indices of LDP and DPLBAR, suggesting a longer concentration time compared to Bellever, which has an area of only 21.5 km<sup>2</sup>. A further comparison of DPSBAR, which reflects the average steepness of the catchment, can also lead to the same conclusion. For the catchment of Brue, DPSBAR shows a lower value



**Fig. 7.** Forecasting results in three-dimensional coordinates for the four catchments, of Bellever, Halsewater, Brue and Bishop\_Hull, respectively in (a), (b), (c) and (d).

representing a lower catchment steepness, which thus indicates a longer concentration time compared to Bellever. All these facts imply that a longer concentration time might result in a sharper increase of the projection curve, while on the other hand, a flatter curve might be resulted from a catchment with a quicker response.

FEH provides a simple method to calculate the catchment concentration time using a generalised model of the catchment descriptors derived by regression analysis (Houghton-Carr, 1999), as shown by Eq. (11):

$$T_p = 4.270 \times \text{DPSBAR}^{-0.35} \times \text{PROPWET}^{-0.80} \times \text{DPLBAR}^{0.54} \times (1 + \text{URBEXT})^{-5.77} \quad (11)$$

where  $T_p$  stands for the time to peak of the instantaneous unit hydrograph. The advantage of this equation is that it is theoretically independent of the time interval, which enables an independent evaluation of catchment concentration time other than those normally derived by using rainfall-runoff data sampled with a certain time interval (Littlewood and Croke, 2008). The calculated results of the four catchments are 4.36, 6.81, 8.37 and 8.82 h, respectively for Bellever, Halsewater, Brue and Bishop\_Hull. This is in agreement with the above conclusion made by simply examining the values

of the catchment descriptors of LDP, DPLBAR and DPSBAR. It should be mentioned that results from the Eq. (11) also can only provide a provisional comparison of the concentration times, which may not be treated as the realistic response times of the catchment (Houghton-Carr, 1999).

By a further check of the correlations between the concentration times and the increasing patterns of the projection curves in Fig. 7, it can be noticed that the projection curve in Fig. 7b for the Halsewater catchment seems to be much flatter than the one in Fig. 7a produced by Bellever, although Halsewater has a larger catchment area together with larger values of LDP, DPLBAR and  $T_p$ , and a lower average steepness (DPSBAR), compared to Bellever. Nevertheless, this might be explained to some extent by another index, ASPVAR (see Tables 3 and 4), which reveals the invariability of the slope directions in every grid of the catchment. Halsewater has the largest ASPVAR value among the four catchments, which indicates the highest consistency of the directions of all catchment grids, and thus might result in a better concentration of the surface and subsurface flow.

**Table 8.** Rankings of different time intervals according to the forecasting results (RMSE) with a certain lead time (from the lowest to the highest RMSE). The optimal time intervals (ranked as the 1st) are highlighted.

	1 h	2 h	3 h	4 h	5 h	6 h	9 h	12 h
(A) Bellever (21.5 km <sup>2</sup> )								
1st	15	30	30	30	30	45	45	45
2nd	30	15	45	60	60	30	60	60
3rd	60	60	60	15	15	60	30	30
4th	–	120	15	120	–	90	90	120
5th	–	–	90	–	–	120	15	90
6th	–	–	–	–	–	15	–	15
optimal	15	30	30	30	30	45	45	45
(B) Halsewater (87.8 km <sup>2</sup> )								
1st	15	15	15	30	30	30	30	30
2nd	30	30	30	15	15	45	90	120
3rd	60	60	45	60	60	15	45	90
4th	–	120	60	120	–	90	60	60
5th	–	–	90	–	–	60	15	45
6th	–	–	–	–	–	120	–	15
optimal	15	15	15	30	30	30	30	30
(C) Brue (135.2 km <sup>2</sup> )								
1st	15	15	15	30	60	90	90	90
2nd	30	30	45	15	30	60	60	60
3rd	60	60	30	60	15	45	45	45
4th	–	120	60	120	–	30	30	120
5th	–	–	90	–	–	15	15	30
6th	–	–	–	–	–	120	–	15
optimal	15	15	15	30	60	90	90	90
(D) Bishop_Hull (202.0 km <sup>2</sup> )								
1st	30	30	30	30	30	30	60	120
2nd	15	15	15	15	60	60	30	60
3rd	60	60	60	60	15	120	90	90
4th	–	120	45	120	–	15	15	30
5th	–	–	90	–	–	90	45	15
6th	–	–	–	–	–	45	–	45
optimal	30	30	30	30	30	30	60	120

All the above analyses indicate that the concentration time might play a considerable role in influencing the selection of the optimal time interval when the forecast lead time is determined. It is known that a steeper, naturally wetter and more urbanised catchment tends to have a faster response, while the larger and longer the catchment is, the slower the response will be. However, besides the descriptors presented in Table 3 and used in Eq. (11), the concentration time is also affected by many other factors (Akan, 1989), e.g. the density of the watercourse, the consistency of the flow directions, the soil infiltration conditions, the rainfall intensity and the storm

path, etc. As a consequence, although it can be deduced from the above analyses that the longer the concentration time is, the steeper the projection curve tends to be, the various influencing factors make it difficult to figure out how exactly the selection of the optimal time interval is affected by the concentration time. And it is far too early to say that the concentration time is definitely the principal factor determining the increasing pattern of the optimal time interval with respect to the increase of the forecast lead time. More research with various catchments bearing different response characteristics is needed to explore the underlying factors and their functions in determining the optimal time interval.

### 3.4 A hypothetical pattern for the selection of the optimal time interval

Based on the analyses of the forecasting results in the case studies, it is interesting to note that the conclusions from the four catchments are consistent with the findings in modern control engineering as mentioned in the introduction part; that is, the best forecasts with a certain lead time are not always produced by the finest time interval or the largest one. Conversely, in hydrological forecasting, the optimal choice of the time interval is found to be increased with the extension of the forecast lead time. Following this, a generalised pattern for the selection of the optimal time interval in real-time hydrological forecasting can be proposed, as shown in Fig. 8, in similar three-dimensional coordinates as Fig. 7.

In Fig. 8, the *x*, *y* and *z* axes represent the time interval, the forecast lead time and the forecasting error in the three subfigures respectively. Two U-shape curves are used to describe the relations between the forecasting error and the time interval in two cases when forecasts are made with a small forecast lead time and a long lead time, as shown respectively in Fig. 8a and b. The lowest points of the two U-shape curves, (*X*<sub>1</sub>, *Z*<sub>1</sub>) and (*X*<sub>*N*</sub>, *Z*<sub>*N*</sub>) thus represent the optimal choices of the time interval in the two cases. As concluded from the case studies, with the increase of the forecast lead time, the forecasting accuracy is decreasing and the optimal time interval is on an increase. Consequently, the two lowest points have the relations of *Z*<sub>1</sub> < *Z*<sub>*N*</sub> and *X*<sub>1</sub> < *X*<sub>*N*</sub>. This trend is more obvious when the two curves are represented in the three-dimensional coordinates in Fig. 8c. A connection of the projections of the two lowest points on the *x*-*y* plane (i.e. the red dotted line) reveals a positive relation between the optimal time interval and the forecast lead time, which is that the longer the lead time is, the larger should the optimal time interval be. Although this projection curve is found to be positive by the case studies, the exact increasing pattern cannot be easily determined. The increasing rate can be either a constant or an accelerating/decelerating value, or even a non-monotonic rate, which is found to be highly related to the characteristics of the catchment, such as the concentration time discussed in Sect. 3.3.

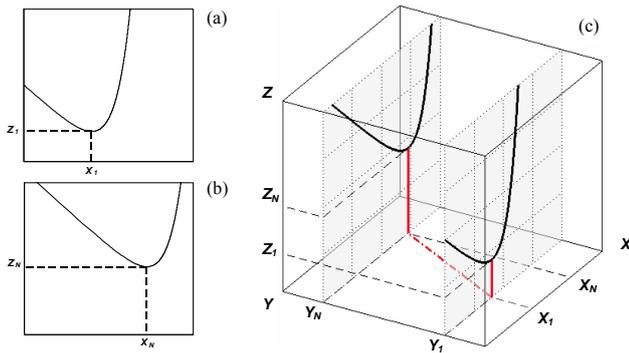
**Table 9.** Forecasting results shown in RMSE ( $\text{m}^3 \text{s}^{-1}$ ) for the four catchments using data with different time intervals: similar to Table 7 but without updating of the ARMA model. The lowest RMSE value for a certain lead time is highlighted to show the optimal data time interval.

	Forecast lead time (h)							
	1	2	3	4	5	6	9	12
(A) Bellever ( $21.5 \text{ km}^2$ )								
15 min	0.3304	0.5784	0.7568	0.8921	0.9897	1.0528	1.1896	1.3188
30 min	0.3529	0.5950	0.7604	0.8933	0.9830	1.0519	1.1723	1.2130
45 min	–	–	0.7679	–	–	1.0630	1.1730	1.2165
60 min	0.3538	0.6414	0.8478	0.9945	1.0962	1.1677	1.2809	1.2313
90 min	–	–	0.8688	–	–	1.1727	1.2980	1.3170
120 min	–	0.6447	–	0.9994	–	1.1895	–	1.3363
(B) Halsewater ( $87.8 \text{ km}^2$ )								
15 min	0.0821	0.1528	0.2160	0.2711	0.3183	0.3641	0.4581	0.5100
30 min	0.0830	0.1554	0.2172	0.2761	0.3258	0.3585	0.4462	0.4969
45 min	–	–	0.2205	–	–	0.3616	0.4500	0.5004
60 min	0.0844	0.1556	0.2209	0.2771	0.3306	0.3675	0.4571	0.4985
90 min	–	–	0.2241	–	–	0.3682	0.4612	0.5082
120 min	–	0.1582	–	0.2815	–	0.3723	–	0.5109
(C) Brue ( $135.2 \text{ km}^2$ )								
15 min	0.6278	1.1883	1.5938	2.0220	2.3034	2.4417	2.8441	3.0413
30 min	0.5948	1.1112	1.5266	1.8581	2.1142	2.3616	2.6827	2.7562
45 min	–	–	1.5295	–	–	2.2597	2.6521	2.7220
60 min	0.6204	1.1611	1.5926	1.9238	2.1723	2.3136	2.5642	2.6408
90 min	–	–	1.6541	–	–	2.5148	2.8087	2.8768
120 min	–	1.2532	–	2.1489	–	2.6775	–	2.8876
(D) Bishop_Hull ( $202.0 \text{ km}^2$ )								
15 min	0.3135	0.5994	0.8518	1.0685	1.2516	1.4048	1.7096	1.8420
30 min	0.3100	0.5935	0.8444	1.0606	1.2424	1.3942	1.6955	1.8247
45 min	–	–	0.8391	–	–	1.3891	1.6925	1.8221
60 min	0.2976	0.5687	0.8080	1.0133	1.1869	1.3319	1.6712	1.7982
90 min	–	–	0.8295	–	–	1.3664	1.6188	1.7399
120 min	–	0.5860	–	1.0438	–	1.3722	–	1.7704

It should be emphasised that this hypothetical pattern of the optimal time interval is especially proposed for real-time forecasting, rather than simulation with only the rainfall–runoff model. When making forecasts using either the data-driven model (e.g. ANN and TF model, etc.) or the physically based rainfall–runoff model together with a real-time updating scheme, an extrapolation is made based on the historical data and into the future. Too dense data will result in worse extrapolation further into the future, while too sparse data can also make the extrapolation fail in the very near future. In essence, this is why using different data time intervals makes a difference to the forecasting results. On the other hand, for the simulation mode using the rainfall–runoff model only, the simulated flow may not vary too much when different data time intervals are used. As analysed in Sect. 3.1, this is due to the low-pass-filtering function of the rainfall–runoff model, which filter out high-frequency variances of the rainfall data

with different time intervals as the input of the system. However, in real-time forecasting, besides rainfall, the flow observations are involved as another input of the modelling system through the updating scheme (e.g. the ARMA model). This can explain why the optimal time interval exists in real-time forecasting but not in the case of pure simulation, at least for those commonly used time intervals examined in this study. But it is believed that with an infinitely small or large time interval, modelling results in simulation mode can also deteriorate due to the involvement of measurement noises or the lack of efficient representation of the original signals of rainfall.

It should also be explained that the optimal time intervals found for the four catchments in the case studies and their relations with the forecast lead times are to some extent subject to the specific structure of the forecasting system (i.e. the rainfall–runoff model and the updating scheme).



**Fig. 8.** Hypothetical pattern for the general impact of data time interval on forecasting accuracy and the selection of the optimal time interval in real-time hydrological forecasting. The axes of  $X$ ,  $Y$  and  $Z$  respectively stand for the time interval of the model input data, the forecast lead time and the forecasting error.

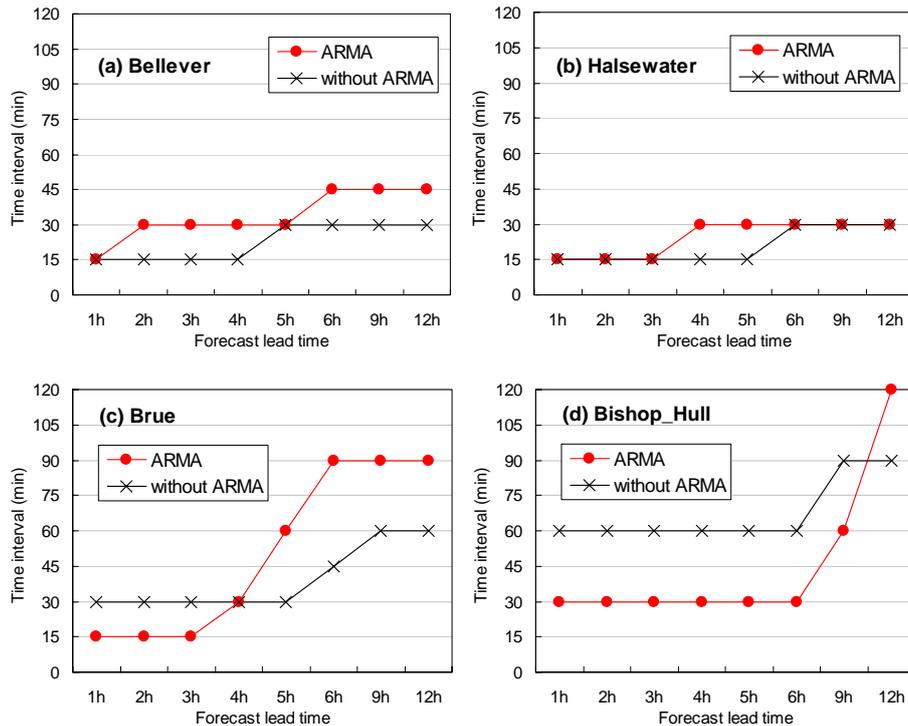
As discussed above, the updating scheme is the most likely reason for the existence of the optimal time interval and its positive relation with the forecast lead time. It is interesting to see how much the current results depend on the updating scheme, i.e. the ARMA model in this study. Table 9 lists the forecasting results of the four catchments using the same datasets and methodologies as described in the case studies, but without the real-time updating using the ARMA model. Instead, previous one-step errors are directly added to forecasts of the future flow. It can be seen in Table 9 that there is an overall increase of the forecasting errors compared to the results with the use of the ARMA model (shown by Table 7). For a certain forecast lead time, the optimal time intervals still exist, and they show an increasing trend with the increase of the forecast lead time. However, the optimal values are not identical to those in Table 7. For a better comparison, the optimal time intervals are plotted in Fig. 9 against the forecast lead time in the two circumstances with and without the updating of ARMA. It can be found for each study catchment in Fig. 9 that the increasing pattern of the curve without ARMA updating is not as sharp as the ARMA one. These modest patterns indicate that the impacts of the data time interval are alleviated when the ARMA model is not used. However, it is still true that the increasing pattern is sharper for larger catchments without the use of ARMA; for example, the optimal time intervals for the 12 h lead time increase to 90, 60, 30 and 30 min for the catchments of Bishop\_Hull, Brue, Halsewater and Bellever with a decreasing drainage area respectively. The results without the use of the ARMA model can be treated as a baseline to evaluate the impact of the data time interval when more complicated updating schemes are adopted in the forecasting system. Regarding the differences caused by using different updating schemes, it is important that more case studies are explored using different forecasting systems (i.e. combinations of different rainfall–runoff models with different updating schemes) in various

geographical and climatic regions. In such a case the proposed hypothetical pattern of the data time interval may be further verified and improved, or otherwise refuted; for example, is the pattern universal or only applicable to a certain types of forecasting systems and catchments?

Another question that deserves attention is the estimation of the catchment areal rainfall, which is regarded as one of most significant sources of uncertainty in hydrological forecasting (Sene et al., 2009). In this study, the simplest linear weighting method, i.e. Thiessen polygons, is used to average the rain gauge observations for the catchments. The four catchments have relatively small areas (20 to 200 km<sup>2</sup>) covered by dense rain gauges. Take the Brue catchment, for example; it has 49 tipping-bucket rain gauges over the drainage area of 135.2 km<sup>2</sup>. Although the other three catchments have lower rain gauge densities, they show very similar patterns of the forecasting results to the Brue. In this case, the estimation of the catchment areal rainfall is not a major problem in this study. However, for large catchments with sparse rain gauge networks, the averaging methods may bring uncertainties in revealing the true patterns between the optimal data time interval and the forecast lead time. The question becomes broader in the combination of the spatial information of radar with the point observations of rain gauges, especially when the radar estimates of rainfall are involved (Cole and Moore, 2008, 2009). Besides the estimation of the areal rainfall, the error measurement statistics adopted to evaluate the performance of the forecasting system may also bring uncertainties into the study. Hall (2001) addressed the error measurement issue and analysed 10 commonly used indices to evaluate the goodness of fit of a model to a set of observations, and it was found that not a single error measurement method was perfect. In this study, two square-error-based statistics, RMSE and NSE, were chosen considering their popularity in hydrological forecasting. However, other error evaluation methods are worth trying in order to attest to whether the hypothetical pattern is limited to a certain type of error measurement statistics.

#### 4 Conclusions

This paper explores the time interval of the model input data and its impact on the hydrological forecasting system, consisting of a conceptual rainfall–runoff model and a real-time updating scheme. Modern telemetry system is able to measure and record hydrological variables with increasingly higher sampling rates; for example, nowadays rainfall data can be collected once a second by the optical rain gauge. An important problem deserving of attention from hydrologists would be which suitable time interval should be used; that is, should our hydrological forecasting system use rainfall and flow data measured in seconds? Such a problem has long been recognised in control



**Fig. 9.** Comparisons on the data time interval patterns when the real-time hydrological forecasts are made in the two circumstances with/without the updating of the ARMA model.

engineering, but is a largely ignored topic in the area of operational hydrological forecasting.

In the beginning of this study, discrete wavelet transform is used to examine the spectral variances of rainfall–runoff data with different time intervals in the frequency domain. It is found that the rainfall signal has energy spread more widely than the flow, but due to the low-pass-filtering function of the rainfall–runoff model, the high-frequency variances of the rainfall signal with different time intervals are not likely to be transformed into the flow. This indicates that higher sampling rates may not always help to improve the results of rainfall–runoff modelling. To further investigate the impact of time interval in real-time hydrological forecasting, which becomes more complicated by involving the flow data as the system inputs by the updating scheme, case studies are carried out in four catchments with different drainage areas and concentration times. Main findings from the case studies can be concluded as follows: (1) the data time interval does have a considerable impact on the performance of the forecasting system, which is more prominent to longer lead times than shorter ones; (2) there exists an optimal time interval for forecasts made with a certain lead time, and the length of the optimal time interval is increasing with the increase of the forecast lead time; (3) the positive relation between the optimal time interval and the forecast lead time can show various patterns, which is found to be highly related to the catchment concentration time; and (4) the longer the concentration time

is, the more dramatically the optimal time interval tends to increase with the increase of the forecast lead time. Finally, according to the results of the case studies, a hypothetical pattern is proposed in three-dimensional coordinates for the selection of the optimal time interval in real-time hydrological forecasting.

It should be recognised that the data time interval issue addressed in this study has not yet been completely solved by the control engineers even though they have many decades of experience with this problem. According to Ljung (1999), it is impossible to deduce a mathematical expression for the identification of the optimal time interval for highly nonlinear systems. Only for very simple models, e.g. the Gauss–Markov process, may the problem be addressed by a deductive approach. Rainfall–runoff models together with the real-time updating schemes are highly nonlinear and diverse, which militate against a deductive derivation and favour an inductive approach. Although nowadays most operational hydrological forecasting systems still have low data sampling rates (daily or hourly), the future is that higher sampling rates will become more widespread and there is an urgent need for both academic and practising hydrologists to realise the significance of the data time interval issue. The main purpose of this study is to remind the hydrologists of this issue when they are dealing with a “data-rich environment” rather than providing a general mathematical formula for selecting the most appropriate data time interval. Therefore the best

way to identify the optimal time interval is to explore various forecasting systems with different catchments, and as a result hopefully a more general pattern applicable to a wide range of cases can be finally found.

*Acknowledgements.* The authors thank the editor and the reviewers for their insightful comments and valuable suggestions in improving the quality of the paper.

Edited by: D. Solomatine

## References

- Åström, K. J.: On the choice of sampling rates in parametric identification of time series, *Information Sciences*, 1, 273–278, 1969.
- Akan, A. O.: Time of concentration of overland flow, *Journal of Irrigation and Drainage Engineering*, ASCE, 112, 283–292, 1986.
- Akan, A. O.: Time of concentration formula for previous catchments, *Journal of Irrigation and Drainage Engineering*, ASCE, 115, 733–735, 1989.
- Bailey, R. and Dobson, C.: Forecasting for floods in the Severn catchment, *J. Inst. Water Engrs. Sci.*, 35, 168–178, 1981, 1981.
- Bayliss, A.: *Flood Estimation Handbook, Volume 5: Catchment descriptors*, Wallingford HydroSolutions, Wallingford, UK, ISBN: 0948540931, 1999.
- Bell, V. A., Carrington, D. S., and Moore, R. J.: Comparison of rainfall-runoff models for flood forecasting, Part 2: Calibration and evaluation of models, R&D Technical Report W242, Environment Agency, UK, 239 pp., 2001.
- Box, G. E. P. and Jenkins, G. M.: *Time series analysis forecasting and control*, Holden Day, San Francisco, 553 pp., 1970.
- Bras, R. L.: Sampling of interrelated random fields: the rainfall-runoff case, *Water Resour. Res.*, 15, 1767–1780, 1979.
- Bras, R. L. and Rodriguez-Iturbe, I.: Network design for the estimation of areal mean of rainfall events, *Water Resour. Res.*, 12, 1185–1195, 1976.
- Chen, S. T. and Yu, P. S.: Real-time probabilistic forecasting of flood stages, *J. Hydrol.*, 340, 63–77, 2007.
- Cho, J., Mostaghimi, S., Kang, M. S., and Chun, J. A.: Sensitivity to grid and time resolution of hydrology components of Dansat, *T Asabe*, 52, 1121–1128, 2009.
- Chow, V. T., Maidment, D. R., and Mays, L. W.: *Applied hydrology*, pp. 27–33, McGraw-Hill, New York, ISBN: 0-07-010810-2, 1988.
- Clark, M. P. and Kavetski, D.: Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, 46, W10510, doi:10.1029/2009WR008894, 2010.
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: use of streamflow observations to update states in a distributed hydrological model, *Adv. Water Res.*, 31, 1309–1324, 2008.
- Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, *J. Hydrol.*, 375, 613–626, 2009.
- Cluckie, I. D. and Han, D.: Fluvial flood forecasting, *Water Environ. J.*, 14, 270–276, 2000.
- Cole, S. J. and Moore, R. J.: Hydrological modelling using raingauge- and radar-based estimators of areal rainfall, *J. Hydrol.*, 358, 159–181, 2008.
- Cole, S. J. and Moore, R. J.: Distributed hydrological modelling using weather radar in gauged and ungauged basins, *Adv. Water Res.*, 32, 1107–1120, 2009.
- Daubechies, I.: The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Info. Theor.*, 36, 961–1005, 1990.
- Dooge, J. C. I.: *Linear theory of hydrologic systems*, Technical Bulletin 1468, US Department of Agriculture, Washington, 309 pp., 1973.
- Duan, Q., Schaake, J., Andreassian, V., Franks, S. W., G. Goteti, H. V. Gupta, Y. M. Gusev, F. Habets, A. Hall, L. Hay, T. Hogue, M. Huang, G. Leavesley, X. Liang, O.N. Nasonova, J. Noilhan, L. Oudin, S. Sorooshian, T. Wagener, and E. F. Wood: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, 320, 3–17, 2006.
- Eberhart, R. C. and Kennedy, J.: A new optimizer using particle swarm theory, in: *Proceedings of the 6th International Symposium on Micro Machine and Human Science*, 39–43, Nagoya, Japan, 1995.
- Eberhart, R. C. and Shi, Y.: Particle swarm optimization: developments, applications and resources. In: *Proceedings of the 2001 congress in evolutionary computation*, 81–86, Seoul, Korea, 2001.
- Finnerty, B. D., Smith, M. B., Seo, D. J., Koren, V., and Moglen, G. E.: Space-time scale sensitivity of the Sacramento model to radar-gage precipitation inputs, *J. Hydrol.*, 203, 21–38, 1997.
- Freer, J., Beven, K. J., and Ambroise, B.: Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach, *Water Resour. Res.*, 32, 2161–2173, 1996.
- Gupta, V. K. and Sorooshian, S.: The relationship between data and the precision of estimated parameters, *J. Hydrol.*, 85, 57–77, 1985a.
- Gupta, V. K. and Sorooshian, S.: The automatic calibration of conceptual catchment models using derivative-based optimization algorithms, *Water Resour. Res.*, 21, 473–486, 1985b.
- Hall, M. J.: How well does your model fit the data?, *J. Hydroinform.*, 3, 49–55, 2001.
- Han, D., Kwong, T., and Li, S.: Uncertainties in real-time flood forecasting with neural networks, *Hydrol. Proc.*, doi:10.1002/hyp.6184, 21, 223–228, 2007.
- Houghton-Carr, H.: *Flood Estimation Handbook, Volume 4: Re-statement and application of the Flood Studies Report rainfall-runoff method*, Wallingford HydroSolutions, Wallingford, UK, ISBN: 0948540931, 1999.
- Kavetski, D. and Clark, M. P.: Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, 46, W10511, doi:10.1029/2009WR008896, 2010.
- Kavetski, D., Fenicia, F., and Clark, M. P.: Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment, *Water Resour. Res.*, 47, W05501, doi:10.1029/2010WR009525, 2011.

- Komma, J., Blöschl, G., and Reszler, C.: Soil moisture updating by ensemble Kalman filtering in real-time flood forecasting, *J. Hydrol.*, 357, 228–242, 2007.
- Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, *Water Resour. Res.*, 35, 2739–2750, 1999.
- Krzysztofowicz, R.: Probabilistic flood forecast: bounds and approximations, *J. Hydrol.*, 268, 41–55, 2002.
- Labat, D., Ababou, R., and Mangin, A.: Rainfall-runoff relations for karstic springs, Part II: continuous wavelet and discrete orthogonal multiresolution analyses, *J. Hydrol.*, 238, 149–178, 2000.
- Lin, C. A., Wen, L., Béland, M., and Chaumont, D.: A coupled atmospheric-hydrological modeling study of the 1996 Ha! Ha! River basin flash flood in Quebec, Canada, *Geogr. Res. Lett.*, 29, 13/1–13/4, 2002.
- Lin, C. A., Wen, L., Lu, G., Wu, Z., Zhang, J., Yang, Y., Zhu, Y., and Tong, L.: Atmospheric-hydrological modelling of severe precipitation and floods in the Huaihe River Basin, China, *J. Hydrol.*, 330, 249–259, 2006.
- Lin, C. A., Wen, L., Lu, G., Wu, Z., Zhang, J., Yang, Y., Zhu, Y., and Tong, L.: Real-time forecast of the 2005 and 2007 summer severe floods in the Huaihe River Basin of China, *J. Hydrol.*, 381, 33–41, 2010.
- Littlewood, I. G. and Croke, B. F. W.: Data time-step dependency of conceptual rainfall-streamflow model parameters: An empirical study with implications for regionalization, *Hydrol. Sci. J.*, 53, 685–695, 2008.
- Liu, J. and Han, D.: Indices for calibration data selection of the rainfall-runoff model, *Water Resour. Res.*, 46, W04512, doi:10.1029/2009WR008668, 2010.
- Ljung, L.: *System Identification – Theory for the User*, Prentice-Hall, Englewood Cliffs, NJ, ISBN: 0-13-881640, 1987.
- Ljung, L., Wahlberg, B., and Söderström, T.: On sampling and reconstruction of continuous time stochastic systems, in: *Proceedings of the 9th IFAC/IFORS Symposium on System Identification and System Parameter Estimation*, Budapest, edited by: Bányász, C. and Keviczky, L., 1174–1178, Pergamon Press, Oxford, 1991.
- Ljung, L.: *System Identification – Theory for the User (2nd Edition)*, Prentice-Hall, Upper Saddle River, NJ, ISBN: 0-13-656695-2, 1999.
- Madsen, H. and Skotner, C.: Adaptive state updating in real-time river flow forecasting – a combined filtering and error forecasting procedure, *J. Hydrol.*, 308, 302–312, 2005.
- Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Patt. Anal. Mach. Intell.*, 11, 674–693, 1989.
- Mantovan, P. and Todini, E.: Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology, *J. Hydrol.*, 330, 368–381, 2006.
- Merz, R., Parajka, J., and Blöschl, G.: Scale effects in conceptual hydrological modeling, *Water Resour. Res.*, 45, W09405, doi:10.1029/2009WR007872, 2009.
- Merz, R., Parajka, J., and Blöschl, G.: Time stability of catchment model parameters: Implications for climate impact analyses, *Water Resour. Res.*, 47, W02531, doi:10.1029/2010WR009505, 2011.
- Meyer, Y.: *Wavelets: Algorithms and Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 1993.
- Mitchell, D. P. and Netravali, A. N.: Reconstruction filters in computer-graphics, *ACM SIGGRAPH Computer Graphics*, 22, 221–228, 1988.
- Moore, R. J.: Transfer functions, noise predictors and the forecasting of flood events in real-time. In: *Statistical analysis of rainfall and runoff*, edited by: Singh, V. P., 229–250, Water Resources Publication, Littleton, Colorado, USA, 1982.
- Moore, R. J.: The PDM rainfall-runoff model, *Hydrol. Earth Syst. Sci.*, 11, 483–499, doi:10.5194/hess-11-483-2007, 2007.
- Moradkhani, H., Hsu, K., Gupta, H. V., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: sequential data assimilation using the particle filter, *Water Resour. Res.*, 41, W05012, doi:10.029/2004WR003604, 2005.
- Nash, J. E. and Sutcliffe, J. V.: River flow forecasting using conceptual models, Part 1: a discussion of principles, *J. Hydrol.*, 10, 280–290, 1970.
- Nyquist, H.: Certain topics in telegraph transmission theory, *Trans. AIEE*, 47, 617–644, 1928.
- Nyquist, H.: Classic paper: Certain topics in telegraph transmission theory, *Proc. IEEE*, 90, 280–305, 2002.
- O'Connor, K. M.: Derivation of discretely coincident forms of continuous linear time-invariant models using the transfer function approach, *J. Hydrol.*, 59, 1–48, 1982.
- Oudin, L., Andréassian, V., Perrin, C., and Anctil, F.: Locating the sources of low-pass behaviour within rainfall-runoff models, *Water Resour. Res.*, 40, W11101, doi:10.1029/2004WR003291, 2004.
- Polikar, R.: The story of wavelets, in: *Physics and Modern Topics in Mechanical and Electrical Engineering*, edited by: Mastorakis, N., 192–197, World Scientific and Engineering Academic Society Press, Athens, Greece, 1999.
- Refsgaard, J. C.: Validation and intercomparison of different updating procedures for real-time forecasting, *Nordic Hydrol.*, 28, 65–84, 1997.
- Remesan, R., Ahmadi, A., Shamim, M. A., and Han, D.: Effect of data time interval on real-time flood forecasting, *J. Hydroinform.*, 12, 396–407, 2010.
- Schaake, J. C., Koren, V. I., Duan, Q. Y., Mitchell, K., and Chen, F.: Simple water balance model for estimating runoff at different spatial and temporal scales, *J. Geophys. Res.-Atmos.*, 101, 7461–7475, 1996.
- Schellekens, J., Weerts, A. H., Moore, R. J., Pierce, C. E., and Hildon, S.: The use of MOGREPS ensemble rainfall forecasts in operational flood forecasting systems across England and Wales, *Adv. Geosci.*, 29, 77–84, doi:10.5194/adgeo-29-77-2011, 2011.
- Sene, K., Weerts, A. H., Beven, K., Moore, R. J., Whitlow, C., and Young, P.: Risk-based Probabilistic Fluvial Flood Forecasting for Integrated Catchment Models, Phase 1, Science Report: SC080030/SR1, Environment Agency, UK, 2009.
- Sene, K., Weerts, A. H., Beven, K., Moore, R. J., Whitlow, C., Beckers, J., Minett, A., Winsemius, H. C., Verkade, J., Young, P., Leedall, D., Smith, P., Cole, S., Robson, A., Howard, P., Huband, M. and Breton, N.: Risk-based Probabilistic Fluvial Flood Forecasting for Integrated Catchment Models, Phase 2, Science Report: SC080030/SR2, Environment Agency, UK, 2010.
- Shannon, C. E.: Communication in the presence of noise, *Proc. Institute of Radio Engineers*, 37, 10–21, 1949.
- Shannon, C. E.: Classic paper: Communication in the presence of noise, *Proc. IEEE*, 86, 447–457, 1998.

- Storm, B., Høgh Jensen, K., and Refsgaard, J. C.: Estimation of catchment rainfall uncertainty and its influence on runoff prediction, *Nordic Hydrol.*, 19, 77–88, 1989.
- Tamea, S., Laio, F., and Ridolfi, L.: Probabilistic nonlinear prediction of river flows, *Water Resour. Res.*, 41, W09421, doi:10.1029/2005WR004136, 2005.
- Tang, Y., Reed, P., Wagener, T., and van Werkhoven, K.: Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, 11, 793–817, doi:10.5194/hess-11-793-2007, 2007.
- Tokar, A. S. and Markus, M.: Precipitation-runoff modelling using artificial neural network and conceptual models, *J. Hydrol. Eng.*, 5, 156–161, 2000.
- Wagener, T., Sivapalan, M., Troch, P. A., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N. B., and Wilson, J. S.: The future of hydrology: An evolving science for a changing world, *Water Resour. Res.*, 46, W05301, doi:10.1029/2009WR008906, 2010.
- Wang, Y., He, B., and Takase, K.: Effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge, *Hydrol. Sci. J.*, 54, 886–898, 2009.
- Weerts, A. H., and El Serafy, G. Y. H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water Resour. Res.*, 42, W09403, doi:10.1029/2005WR004093, 2006.
- Weerts, A. H., Winsemius, H. C. and Verkade, J. S.: Estimation of predictive hydrological uncertainty using quantile regression: Examples from the National Flood Forecasting System (England and Wales), *Hydrol. Earth Syst. Sci.*, 15, 255–265, doi:10.5194/hess-15-255-2011, 2011.
- Werner, M., Cranston, M., Harrison, T., Whitfield, D., and Schellekens, J.: Recent developments in operational flood forecasting in England, Wales and Scotland, *Meteorol. Appl.*, 16, 13–22, doi:10.1002/met.124, 2009.
- Wilby, R., Greenfield, B., and Glenny, C.: A coupled synoptic-hydrological model for climate change impact assessment, *J. Hydrol.*, 153, 256–290, doi:10.1016/0022-1694(94)90195-3, 1994.
- Wood, A. W. and Schaake, J. C.: Correcting errors in streamflow forecast ensemble mean and spread, *J. Hydrometeorol.*, 9, 132–148, 2008.
- Young, P. C.: *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*, 2nd Edition., 504 pp., Springer-Verlag, New York, ISBN: 978-3-642-21981-8, 2011.