Journal of Hydrology 397 (2011) 210–224

Contents lists available at ScienceDirect

Journal of Hydrology

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

A wavelet-based approach to assessing timing errors in hydrologic predictions

Yuqiong Liu [a,b,*], James Brown [a,c], Julie Demargne [a,c], Dong-Jun Seo [d]

[a] National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrologic Development, 1325 East-West Highway, Silver Spring, MD 20910, USA
[b] Riverside Technology, Inc., 2290 East Prospect Road, Fort Collins, CO 80525, USA
[c] University Corporation for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA
[d] Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX 76019, USA

ARTICLE INFO

Article history:
Received 24 May 2010
Received in revised form 21 October 2010
Accepted 26 November 2010
Available online 3 December 2010

This manuscript was handled by K. Georgakakos, Editor-in-Chief, with the assistance of Emmanouil N. Anagnostou, Associate Editor

Keywords:
Timing errors
Magnitude errors
Cross wavelet transform
Hydrologic evaluation
Forecast verification

SUMMARY

Streamflow predictions typically contain errors in both the timing and the magnitude of peak flows. These two types of error often originate from different sources (e.g. rainfall–runoff modeling vs. routing) and hence may have different implications and ramifications for both model diagnosis and decision support. Thus, where possible and relevant, they should be distinguished and separated in model evaluation and forecast verification applications. Distinct information on timing errors in hydrologic prediction could lead to more targeted model improvements in a diagnostic evaluation context, as well as better-informed decisions in many practical applications, such as flood prediction, water supply forecasting, river regulation, navigation, and engineering design. However, information on timing errors in hydrologic predictions is rarely evaluated or provided. In this paper, we discuss the importance of assessing and quantifying timing error in hydrologic predictions and present a new approach, which is based on the cross wavelet transform (XWT) technique. The XWT technique transforms the time series of predictions and corresponding observations into a two-dimensional time-scale space and provides information on scale- and time-dependent timing differences between the two time series. The results for synthetic timing errors (both constant and time-varying) indicate that the XWT-based approach can estimate timing errors in streamflow predictions with reasonable reliability. The approach is then employed to analyze the timing errors in real streamflow simulations for a number of headwater basins in the US state of Texas. The resulting timing error estimates were consistent with the physiographic and climatic characteristics of these basins. A simple post-factum timing adjustment based on these estimates led to considerably improved agreement between streamflow observations and simulations, further illustrating the potential for using the XWT-based approach for timing error estimation.

1. Introduction

Hydrologic predictions are subject to various types and sources of error. These include errors in the hydrometeorological forcing and initial soil moisture conditions, as well as erroneous or imperfect model structures and parameters (e.g., Huang and Liang, 2006; Liu and Gupta, 2007). Rigorous evaluation or verification of hydrologic predictions is necessary to quantify the reliability and skill of a prediction system, to assess the value of new improvements to a model, and to facilitate model inter-comparisons (e.g., Krause et al., 2005). More broadly, evaluation is central to various applications in hydrologic research, such as data analysis, model identification, parameter estimation, uncertainty analysis, sensitivity analysis, multi-model analysis, and Bayesian networks (Matott et al., 2009). It also constitutes an important component in verifying operational forecasts of hydrometeorological and hydrologic variables (e.g., Murphy and Winkler, 1992; Hersbach, 2000; Demargne et al., 2009).

Traditional evaluation strategies have received extensive criticism for their lack of 'diagnostic power' or inability to provide a comprehensive assessment of model performance (e.g., Teegavarapu and Elshorbagy, 2005; Gupta et al., 2008). A common criticism is that traditional approaches to hydrologic evaluation cannot distinguish between different models or parameter sets, a situation sometimes referred to as 'equifinality' in hydrology (e.g., Beven, 1993; Beven and Freer, 2001). While lack of data is often cited as a cause of difficulties in hydrologic evaluation, others suggest that more powerful forms of evaluation are required to enable appropriate performance assessment and to obtain diagnostic information for targeted model improvement (e.g., Briggs and Levine, 1997; Gupta et al., 2008). This has prompted considerable research into developing more powerful evaluation strategies, including using multi-criteria methods (e.g., Gupta et al., 1998;

* Corresponding author at: Hydrological Sciences Branch, NASA Goddard Space Flight Center, 8800 Greenbelt Road, Greenbelt, MD 20771, USA. Tel.: +1 301 713 0640x120.
E-mail address: Yuqiong.Liu@nasa.gov (Y. Liu).

0022-1694/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2010.11.040

Boyle et al., 2000; Willems, 2009), decomposition of conventional error measures into more meaningful components (e.g., Hersbach, 2000; Bradley et al., 2004; Gupta et al., 2009), development of improved benchmark models (e.g., Seibert, 2001; Schaefli and Gupta, 2007), development of new error measures (e.g., Teegavarapu and Elshorbagy, 2005; Jachner and van den Boogaart, 2007; Brown et al., 2010), development of mechanisms for integrating soft data and expert knowledge (e.g., Seibert and McDonnell, 2002), and development of widely applicable evaluation tools and software packages (e.g., Dawson et al., 2007; Jachner and van den Boogaart, 2007; Brown et al., 2010), among many others. However, despite the overwhelming research in developing new evaluation strategies or tools, "lumped" error measures, such as the root mean square error (RMSE) and the Nash–Sutcliffe Efficiency (Nash and Sutcliffe, 1970) remain as a fundamental basis for (or drawback of) most evaluation applications. In this context, "lumped" implies that no distinction is made between different types of error, and that the scale and time dependencies of the errors are typically not taken into account. This can lead to difficulties in hydrologic evaluation including the equifinality problem (e.g., Lane, 2007; Gupta et al., 2008).

The starting point for discussion in this paper concerns the need to distinguish between different types of error in hydrologic predictions and to account for the scale- and time-dependency of these errors in evaluation (e.g., Reusser et al., 2008). For example, in addition to commonly investigated magnitude errors, hydrologic predictions typically also contain errors in the predicted timing of a distinct "hydrologic feature", such as a high-frequency "peak flow" or a low-frequency "base flow". Such features must be distinct, whether in untransformed or transformed space, to make the diagnosis of a timing error meaningful. Thus, a timing error refers to a persistent signal in the lumped error that is separable from the magnitude error through analysis (e.g. in the frequency space). This is analogous to the location or position errors in spatial prediction of precipitation, due to the displacement of a distinct "meteorological feature" (e.g., the center of a hurricane; Gilleland et al., 2009). As will be discussed in detail in Section 2, information on timing errors in hydrologic predictions is not only useful for increasing the diagnostic power of model evaluation, but also important for many practical applications. One fundamental issue with traditional model evaluation and forecast verification involves "lumped" assessments of magnitude and timing errors, leading to underestimated predictive skills in most cases. For example, if a simulated hydrograph matches the observed hydrograph well in magnitude but with a small timing error, a lumped error statistic based on the RMSE or NSE would be misleadingly large.

The ability to reliably identify timing errors in hydrologic predictions is complicated by the scale- and time-dependency of timing errors and the need for long records of predictions and corresponding observations to evaluate those errors with reasonably small sampling uncertainty. In this paper, we propose wavelet analysis for identifying the scale- and time-dependent timing errors in hydrologic predictions, using, specifically, the cross wavelet transform technique, or XWT (Torrence and Compo, 1998). In hydrologic research and other geoscience studies, wavelet-based techniques have received increasing interest in recent years (see Kumar and Foufoula-Georgiou (1997) and references cited therein). This stems from their ability to provide spectral decompositions of a signal that are localized in both time and frequency, while avoiding the inherent limitations of the windowed Fourier transform (Torrence and Compo, 1998). The ability to provide time-dependent information is especially appealing given the increasing recognition of the non-stationarity of most hydrologic systems due to climate change and anthropogenic alterations (e.g., Milly et al., 2008; Wagener et al., 2010). A common application of wavelet analysis is for the detection of dominant patterns and inter-connections among different quantities that are tied to particular frequencies in time and space and, therefore, appear more readily when evaluated in the frequency domain. For example, wavelet analysis has been widely used to study the relationships between various hydrologic conditions (e.g., streamflow and ice conditions) and dominant patterns of climate variability such as the El Niño-Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO) (e.g., Torrence and Webster, 1999; Jevrejeva et al., 2003; Coulibaly and Burn, 2004; Grinsted et al., 2004; Labat et al., 2005; Zhang et al., 2007; among many others). In another area of application, Briggs and Levine (1997) used wavelet decomposition to develop multivariate closeness measures to improve forecast verification of meteorological variables. In hydrology, wavelet-based methods have also been used for the characterization of precipitation (e.g., Perica and Foufoula-Georgiou, 1996) and streamflow (e.g., Smith et al., 1998, 2004; Zoppou et al., 2002; Schaefli et al., 2007, Zolezzi et al., 2009), and for hydrologic model evaluation and parameter estimation (e.g., Lane, 2007; Schaefli and Zehe, 2009). The specific technique used in this study, namely the cross wavelet transform technique (XWT), analyzes the co-variation of two time series in time–frequency space. The XWT is less commonly used than univariate wavelet analysis, but has received some interest in the literature (e.g., Torrence and Compo, 1998; Grinsted et al., 2004; Marraun and Kurths, 2004). For example, XWT was used in Anctil et al. (2008) to assess the time-lags between the 150-cm air temperature and soil temperatures at different depths in the soil.

This paper evaluates the potential for using XWT to quantify the timing errors in hydrologic predictions. Our main goal is not to develop new statistical measures, but to call for the explicit assessment of timing errors in hydrologic model evaluation and forecast verification (wherever possible and relevant), and to propose an XWT-based approach for doing so. The importance of timing error assessment is discussed in Section 2, followed by a brief description of the test basins and the streamflow datasets used in this study in Section 3. Section 4 describes the XWT technique and its application for timing error estimation. Results from applying XWT to streamflow simulations with synthetic timing errors and to real streamflow simulations from a lumped hydrologic model for a number of headwater basins are discussed in Sections 5 and 6, respectively. A brief summary of the present study and discussions on the implications for hydrologic evaluation are included in Section 7.

2. The need for timing error assessment

As indicated above, hydrologic predictions typically contain errors in both magnitude and timing, which are different by nature and have different implications for both model evaluation and decision making. As pointed out in Liu et al. (2008), providing 'usable' scientific information with enhanced credibility and transparency is essential to bridging the gaps between science and decision making. For example, in an extreme flooding situation, distinct magnitude and timing uncertainty information regarding the flood forecasts would enable emergency managers and other forecast users to make better-informed decisions about flood protection and evacuation procedures, thus improving the use of resources and enhancing protection of life and property. Distinct information on timing errors could be used in many hydrological applications. For example, river flow timing, along with flow magnitude, duration, frequency, and rate of change are typically considered as the five fundamental attributes of flow regimes that have ecological significance (see Zolezzi et al., 2009 and the references cited therein). Other relevant applications include river navigation,

regulation, reservoir operation, engineering design, and water banking and marketing. For example, in order to estimate representative hydrographs for engineering design purposes, the timing variability among different observed hydrographs needs to be accounted for (or removed) to allow for the derivation of meaningful averaged hydrographs, via landmark registration or curve registration techniques (e.g., Ramsay and Li, 1998; Reilly et al., 2004).

In a model evaluation context (diagnostic or otherwise), assessment of timing errors in hydrologic predictions is critical, and may lead to appropriate bias corrections. For example, Meng et al. (2008) showed that a timing adjustment to simulated hydrographs could significantly improve the NSE (Nash and Sutcliffe, 1970). Abrahart et al. (2007) incorporated a timing error correction procedure in the calibration of a neural network-based rainfall–runoff model to obtain better predictions. Lane (2007) found that objective functions based on wavelet power difference and phase difference can help discriminate streamflow time series that have the same or similar NSE values, thus helping reduce equifinality in model predictions. The presence of timing errors can also have a negative impact on hydrologic data assimilation. For example, Seo et al. (2003, 2009) reported that the performance of a data assimilation procedure would become less satisfactory, or even unacceptable, when there are systematic timing errors in the model predictions. As pointed out by Ravela et al. (2007), most data assimilation approaches are designed for correcting amplitude errors only, and attempts to correct timing errors (or position errors to be more general) by adjusting amplitudes via data assimilation can adversely affect analyses and forecasts. In a diagnostic evaluation of the distributed Sacramento Soil Moisture Accounting Model (SAC-SMA, Koren et al. (2004)) of the National Weather Service (NWS), Yilmaz et al. (2008) found that the overall water balance error in streamflow simulation is strongly sensitive to changes in parameters that control the total amount of tension soil water and the partition between tension water and free water in the lower soil zone, while the timing error is mostly sensitive to changes in the upper zone free water capacity and the channel discharge routing parameter. Since timing errors in hydrologic predictions may point to distinct sources of uncertainty, their quantification may lead to targeted improvements in hydrologic models.

In meteorology, improved verification of spatial forecasts via feature-based or object-based techniques, as apposed to the traditional gridpoint-based verification techniques, have received increasing interest in recent years (see Gilleland et al. (2009) and the references cited therein). These new verification techniques aim to separate and evaluate the spatial displacement in the meteorological features or objects in the forecast, so that the residual magnitude errors can be quantified for a more meaningful analysis of forecast quality. An analogy for hydrologic evaluation would be to assess the timing errors in distinct hydrologic features (such as the crest of a hydrograph) and separate the timing errors from the prediction, in order to obtain a more meaningful quantification of magnitude errors. However, in hydrology, timing errors are rarely assessed or quantified, either in a model evaluation or real-time forecasting context. A few hydrologic studies have attempted to evaluate timing errors using averaged timing indices (e.g., Seo et al., 2003; Yilmaz et al., 2008) or qualitative measures (e.g., Jachner et al., 2007). This however ignores the temporal variations and scale dependence of hydrologic model performance, which are known to be important for diagnostic purposes (e.g., Wagener et al., 2003; Shamir et al., 2005a,b; Reusser et al., 2008). The inadequate research in evaluating timing errors in hydrologic predictions is partly due to insufficient recognition of the importance of timing error assessment itself, but also partly due to lack of widely applicable approaches that can estimate time- and scale-varying timing errors both reliably and in a reasonably automated way. As will be illustrated later in this paper, the XWT-based technique holds considerable potential for assessing time- and scale-dependent timing errors in hydrologic predictions.

3. Study basins and datasets

In this study, we focus on analyzing the timing errors in the crests or peaks of predicted streamflow hydrographs. Flow observations and simulations were obtained for 11 headwater basins from the US state of Texas (Fig. 1 and Table 1), which lie in the service area of the NWS West Gulf River Forecast Center (WGRFC). The study area encompasses a variety of hydro-climatic regimes ranging from semi-arid in the west to humid in the east, and subtropical along the coast. As shown in Table 1, the 11 study basins vary in size from less than 200 km^2 (QLAT2) to nearly 2000 km^2 (UVAT2). The mean annual precipitation ranges from 575 mm (UVAT2, a semi-arid basin) to 1126 mm (MTPT2, a costal basin); and the mean annual streamflow ranges from 1.17 m^3/s (SKMT2) to 6.40 m^3/s (MDST2). The shape and slope of the basins also vary greatly, leading to a wide range of time to peak (T_p) values (12–39 h). Additional information on the physiography of the area and the basins can be found in Kuzmin et al. (2008). Given that the model's skill in predicting timing of streamflow is intricately related to these and other hydro-climatic and physiographic characteristics of a basin, the variety of basins considered should help to rigorously evaluate the reliability of the XWT-based technique for timing error estimation.

Hourly streamflow observations were obtained from the U.S. Geological Survey (USGS) for the basin outlets, from June 04, 1998, or March 04, 1999, to January 31, 2007. The streamflow simulations for these basins were generated from the lumped SAC-

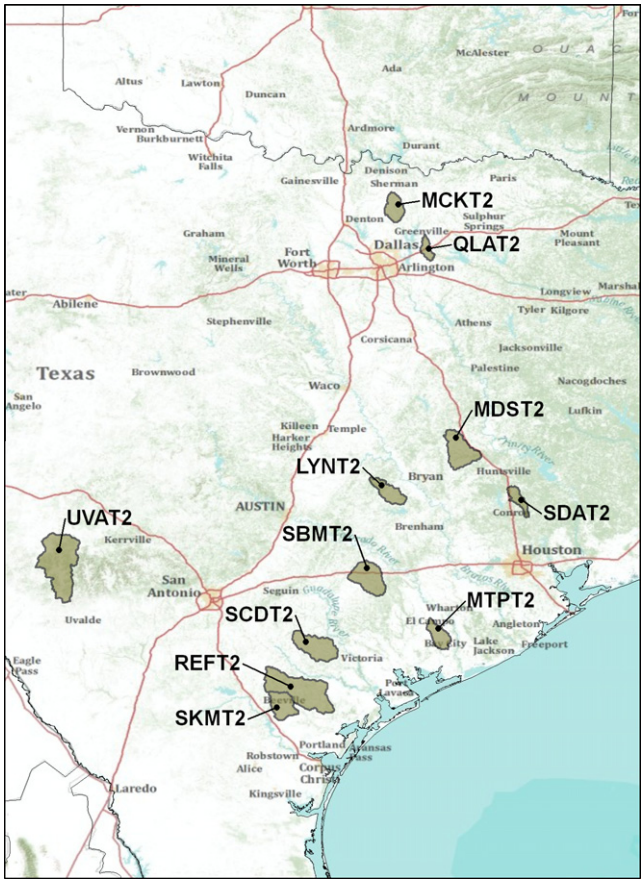

Fig. 1. The 11 study basins in Texas, US.

Table 1
The study basins.

	Basin ID	Basin name	USGS ID	Area (km^2)	Time to peak (hr)	Average annual discharge (m^3/s)	Average annual precipitation (mm)	Number of events analyzed ($Q_{peak} > 50\ m^3/s$)
1	LYNT2	Lyons – Davidson Creek	08110100	508	18	2.45	858	20
2	MCKT2	McKinney – East Fork Trinity	08058900	427	14	4.07	800	16
3	MDST2	Madisonville – Bedias Creek	08065800	870	21	6.40	933	41
4	MTPT2	Midfield – Tres Palacios	08162600	435	17	4.89	1126	39
5	QLAT2	Quinlan – South Fork Sabine	08017300	197	12	2.45	793	22
6	REFT2	Refugio – Mission River	08189500	1787	39	4.48	748	27
7	SBMT2	Sublime – Navidad River	08164300	896	26	5.01	934	39
8	SCDT2	Schroeder – Coleto Creek	08176900	932	14	2.97	845	30
9	SDAT2	Splendora – Caney Creek	08070500	285	17	2.73	1103	16
10	SKMT2	Skidmore – Aransas River	08189700	640	12	1.17	712	16
11	UVAT2	Laguna – Nueces River	08190000	1981	13	6.27	575	17

Note: The average annual discharge and precipitation values are calculated from hourly flow and precipitation records of approximately 10 years (1997–2006). Time to peak values are estimated from empirical unit hydrographs.

SMA model (Burnash et al., 1973), and a unit hydrograph model. The SAC-SMA is a conceptual rainfall–runoff model with spatially lumped parameters. The runoff amounts produced from the SAC-SMA model are translated into streamflow values at the basin outlet via an empirically estimated unit hydrograph model. In this study, both models were run at an hourly time step and the model parameters were obtained via calibration using the data records available for each basin (see Seo et al. (2009) for details).

Although the models were calibrated against a long record of observations, the hourly streamflow simulations still do not reproduce the observations satisfactorily. Fig. 2 shows the streamflow observations and simulations for two of the basins considered, namely LYNT2 and QLAT2. As expected, streamflow predictions in these basins contain both timing and magnitude errors, although the errors may vary significantly between events. Since providing timing error information is more critical for medium-to-high flow events, here we focus on only those events with a peak value exceeding a certain threshold when analyzing the timing errors. While the choice of the peak value threshold is arbitrary in this study, it helps to identify distinct hydrologic features (e.g., event peaks) for analyzing the timing errors. In practical applications, depending on the primary purposes of the analyses, it may be necessary to use a more meaningful threshold such as a climatological probability, or values corresponding to the minor, medium, or major flood levels of a river. To facilitate the analysis, a truncated time series of streamflow is created by removing the

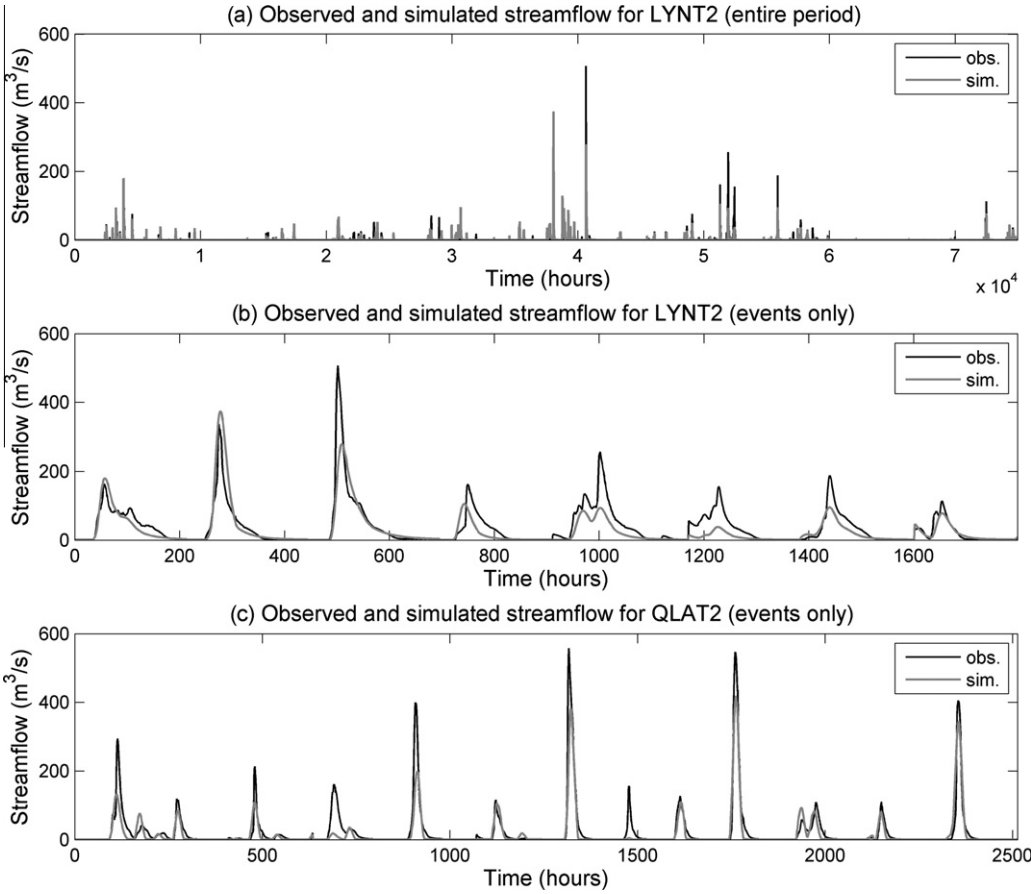

Fig. 2. Observed and simulated time series for LYNT2 and QLAT2. Note in (b) and (c) the time series are truncated to include only periods of large events (peak flow >=100 m^3/s).

extended low flow periods from the original time series using a threshold of 100 m^3/s for each of the test basins and is used in the synthetic experiments discussed in Section 5. Examples of truncated flow time series that include only event periods are shown in Fig. 2b and c for LYNT2 (8 events) and QLAT2 (13 events), respectively. The truncation reduces the length of the data record greatly (e.g., from 74,955 to 1798 data points for LYNT2), thus increasing the efficiency of wavelet analysis while making it more convenient to visually examine the results as shown in Section 5. In applying the XWT-based approach to the real streamflow simulations (Section 6), the original (un-truncated) time series are used to avoid the possible impact (if any) of the truncation on the accuracy of timing error estimation, due to the neighbor influences inherent in wavelet analysis (see relevant discussions in the following sections).

4. Using XWT for timing error estimation

Wavelet transform (WT) techniques have been widely used in geophysical research to analyze nonstationary time series at many different frequencies (e.g., Kumar and Foufoula-Georgiou, 1997; Smith et al., 1998). Timing errors in hydrologic predictions are typically both time- and scale-dependent, warranting a WT-like technique for localized analyses in both time and frequency domains. In this study, we use the continuous wavelet transform (CWT), which can be defined as the convolution of a discrete data sequence $X = \{x_{n'},\ n' = 1,\ldots,N\}$ with a scaled and translated version of a mother wavelet function $\psi(\eta)$ that depends on a non-dimensional time parameter η (Torrence and Compo, 1998):

$$W_n^X(s) = \sum_{n'=1}^{N-1} x_{n'} \psi^* \left[\frac{(n'-n)\delta t}{s}\right] \tag{1}$$

where s is the scale parameter and n is the location parameter; the asterisk indicates the complex conjugate of the wavelet function; δt is the time step of the analysis. W denotes the wavelet spectrum and the wavelet power is defined as $|W^2|$. The wavelet function ψ is normalized to have unit energy and is localized in both time and frequency spaces to allow for localized decompositions of x_n via Eq. (1) in both time and frequency domains.

Several factors need to be considered when choosing a mother wavelet for a particular application (Torrence and Compo, 1998). These include (1) orthogonal or nonorthogonal; (2) complex or real; (3) width, which determines the balance between time resolution and frequency resolution; and (4) shape, which should reflect the type of features found in the actual data sequence. In this study, we choose a mother wavelet that is nonorthogonal, to allow for smooth and continuous variations in wavelet power spectrum, and complex, to allow for the derivation of phase or timing errors from the real and imaginary parts of the wavelet spectrum (see Eq. (5) below). After some testing, the complex Morlet wavelet is chosen for the datasets considered in this study,

$$\psi(n') = \pi^{-1/4} e^{i\omega_0 n'} e^{-n'^2/2} \tag{2}$$

where ω_0 is the non-dimensional frequency, here set to 6 to meet the admissibility condition (i.e., having zero mean) and to allow a good balance between time and frequency localization (Torrence and Compo, 1998; Grinsted et al., 2004).

As noted above, the aim of this study is to estimate the timing difference between time-series of observed and simulated streamflows, by evaluating the cross wavelet phase relationship between the two time series (denoted here as X and Y for simulation and observation, respectively) using the XWT technique (e.g., Grinsted et al., 2004; Marraun and Kurths, 2004). Given the CWT of two time series ($W_n^X(s)$ and $W_n^Y(s)$), the cross wavelet spectrum ($W_n^{XY}(s)$) can be defined as

$$W_n^{XY}(s) = W_n^X(s) W_n^{Y^*}(s) \tag{3}$$

where $*$ denotes the complex conjugation. The cross wavelet spectrum $W_n^{XY}(s)$ can be used to calculate the cross wavelet power, $|W_n^{XY}(s)|$, which reveals common areas of high power in X and Y in the time–frequency domain. One could also calculate the wavelet coherence (which is analogous to a correlation coefficient in the time domain) to measure the cross correlation between the two time series in the time–frequency domain. Wavelet coherence is defined as the square of the smoothed cross power spectrum normalized by the individual smoothed power spectra of the two time series:

$$R_n^2(s) = \frac{|\langle s^{-1} W_n^{XY}(s)\rangle|^2}{\langle s^{-1}|W_n^X(s)|^2\rangle \cdot \langle s^{-1}|W_n^Y(s)|^2\rangle} \tag{4}$$

where $\langle\rangle$ denotes the smoothing operation in both time and frequency which is necessary to remove noise in the wavelet spectra (e.g., Torrence and Webster, 1999; Grinsted et al., 2004) and s^{-1} is a factor for conversion to unit energy. By definition, $R_n^2(s)$ takes values between 0 and 1, with 1 indicating highest coherence and 0 the lowest. The phase angle of the cross wavelet spectrum gives the phase difference between the two time series and can be computed from the real ($\Re$) and imaginary ($\Im$) parts of $W_n^{XY}(s)$ as follows:

$$\phi_n^{XY}(s) = \tan^{-1}\left[\frac{\Im(\langle s^{-1} W_n^{XY}(s)\rangle)}{\Re(\langle s^{-1} W_n^{XY}(s)\rangle)}\right] \tag{5}$$

In this study, the smoothing in Eqs. (4) and (5) follows the strategy presented in the appendix of Torrence and Webster (1999). For convenience, we convert the cross phase angle $\phi_n^{XY}(s)$ into a time lag as a means of evaluating the timing error in the streamflow simulations (as compared to the observations):

$$\Delta t_n^{XY}(s) = \phi_n^{XY}(s) * T/(2\pi) \tag{6}$$

where T is the equivalent Fourier period of the wavelet.

For illustration purposes, we apply XWT to the truncated time series of observations and simulation for LYNT2 (see Fig. 2b and discussion on the truncation in Section 3) using Eqs. (1)–(6) as described above. The results are presented in Fig. 3 for cross wavelet power, wavelet coherence, and timing errors, all as a function of both time and frequency. Note that in Fig. 3a–c, the shaded white areas at the beginning and end of the time window are known as the cone of influence (COI), where the results from the wavelet decomposition are considered unreliable, due to applying the wavelet transform to a non-cyclic time series of finite length. The size of the COI increases with scale and depends on the wavelength of the chosen wavelet. For a given scale, a narrow wavelet (e.g., the Mexican hat wavelet) will produce a much smaller COI than a wider wavelet (e.g., the Morlet wavelet). In this study, we follow the zero padding strategy used in Torrence and Compo (1998) (i.e., adding zeros to the beginning and end of the time series before applying the wavelet transform and remove them afterward) to reduce the impact of edge effects. Although zero padding could introduce discontinuities at the edges, the impact is less important in this study since hydrologic events typically start and end with low flow values.

Fig. 3a shows the pattern of areas with common high power (i.e., where both series simultaneously exhibit high variations) in the time–frequency space with time on the x-axis and Fourier period or scale on the y-axis. For the Morlet wavelet considered, the wavelet scale corresponds (approximately) to its Fourier period; and hence the two will be used interchangeably in this paper. The pattern of the power spectrum clearly follows the variations of individual events and is, therefore, closely related to amplitude information. It is expected that the phase angle computed from the cross wavelet transform may not be reliable when the two time

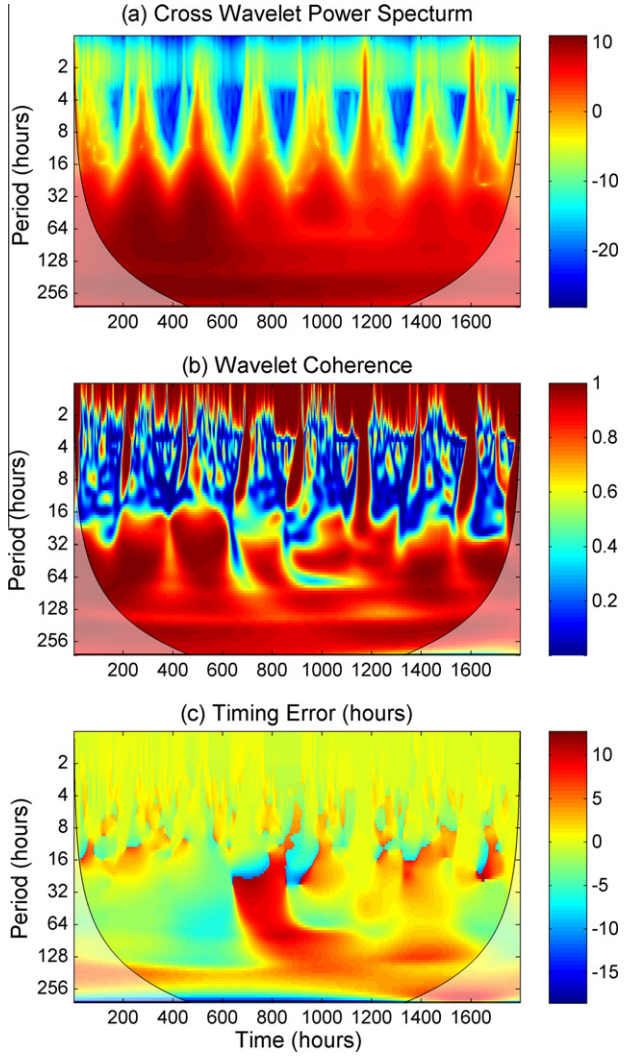

Fig. 3. (a) the cross wavelet power spectrum (on the log2 scale), (b) wavelet coherence, and (c) timing error between the observed and simulated streamflow time series for LYNT2, which are calculated based on Eqs. (3), (4), and (6), respectively. The shaded white areas outlined by black curves indicate the cone of influence (COI) regions. Positive timing errors in (c) indicate simulation leading observation.

series contain very different features or are completely out of phase. As such, the wavelet coherence distribution shown in Fig. 3b is useful for identifying areas in the time–frequency space from which the time error information should be extracted. In the case of LYNT2, we see that, for the observed and simulated time series to have relatively high wavelet coherence (e.g., $R_n^2(s) > 0.5$), the period (or scale) should be larger than approximately 20 h. However, larger scales do not necessarily lead to more reliable estimates of timing error, since the wavelet convolution tends to blur the distinction between neighboring events at large scales (see Fig. 3a and c and discussions in Sections 5 and 6). The timing error distribution in the time–frequency plane (hereinafter referred to as the timing error spectrum) in Fig. 3c clearly indicates the time- and scale-dependency of the timing errors. Visually comparing the observed and simulated time series in Fig. 2b indicates that the simulation tends to lag the observation for the first three events while leading the observation for the remaining five events, especially for the fourth event. This observation agrees well with the timing error spectrum for the scale band 20–150 h as shown in Fig. 3c (where positive timing errors indicate simulation leading observation), at least in a qualitative sense. In the subsequent discussion, we consider how to derive practical and quantitatively accurate timing error estimates from the time error spectra (as shown in Fig. 3c).

5. Application to synthetic timing errors

Before applying the XWT technique described in Section 4 to the other basins for further analysis, synthetic experiments are carried out to examine the reliability of XWT for analyzing the timing errors in streamflow predictions. In these experiments, synthetic streamflow simulations are generated by simply shifting the truncated observed time series to the right or left by a certain number of hours. As such, the "true" timing errors are known, providing a basis for quantitatively verifying the timing error estimates from XWT. The other important purpose of these synthetic experiments is to gain some insights with respect to the scale (or frequency) ranges and time periods over which the true timing errors can be properly extracted from the XWT-based timing error spectra. Cases of both a constant timing error and event-specific (or time-varying) timing errors are investigated. It is important to note that, since the same error is applied to an event as a whole (e.g., with no distinction between peak flows and base flows which often involve different scales), the prescribed timing errors are scale-invariant.

5.1. Constant timing error

In this synthetic experiment, the truncated time series of observed streamflow at LYNT2 (Fig. 2b) is shifted earlier or later by five hours to create two synthetic simulation series. As such, the two simulations should have a constant timing error of +5 or −5 h when compared to the untransformed series. Also, the wavelet coherence between the observation and the two synthetic simulations will be high, by construction, except at the smallest scales affected by the timing error and noises. The timing error results from applying XWT to the synthetic simulations and the observed time series are presented in Fig. 4. As indicated in Fig. 4a and b, the timing errors estimated by XWT are approximately equal to the true errors (i.e. +5 h for the first case and −5 h for the second case) for scales larger than 10 h, with small variations in the time–frequency space due to noise. This is further confirmed by the global timing errors obtained by averaging the timing error spectra over time, excluding areas inside the COI, in Fig. 4c, where it shows that the time-averaged errors are reasonably close to the true errors, except for very small and very large scales. A close examination of the time-averaged errors indicates that the scale band 10–150 h corresponds to a timing error range of +4.2 to +5.0 h for the first case and −4.2 to −5.0 h for the second case. Further narrowing down the scale band to 110–140 h would produce a timing error range of +4.9 to +5.0 (or −4.9 to −5.0) hours, as also can be observed in Fig. 4c. In fact, a close look at Fig. 4a and b also reveals an outstanding narrow frequency band in the same range (around 128 h) over which the true timing errors are most consistently manifest. These two scale bands (10–150 h and 110–140 h) are used to compute the scale-averaged timing errors (as a function of time) and the results are presented in Fig. 4d. The scale-averaged timing errors derived from XWT for the narrower scale band (110–140 h) are indeed almost constant over time, reproducing the true errors almost perfectly, except at the beginning and end of the time series, due to edge effects. The timing errors averaged over the larger scale band (10–150 h), although showing some small variations in time, are also reasonably close to the truth. Applying constant timing errors to the other basins produces very similar results (not shown here).

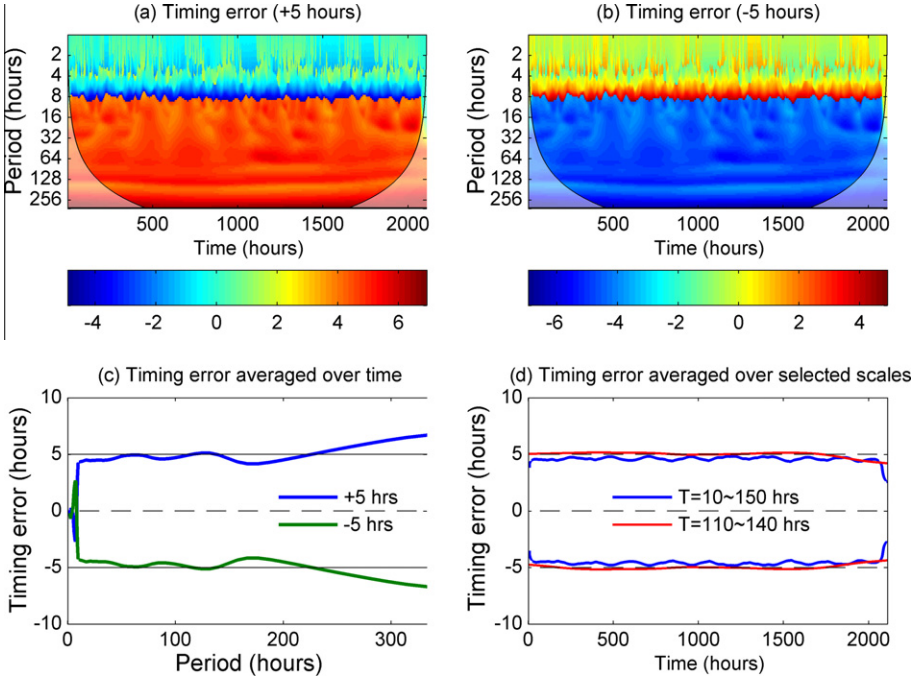

Fig. 4. Timing error spectra for synthetic simulations at LYNT2 with constant timing errors of +5 h (a) or −5 h (b). (c) is the time error averaged over the entire time period, as a function of scale; (d) is the time error averaged over selected ranges of scales, as a function of time. Positive timing errors indicate simulation leading observation.

5.2. Event-specific timing errors

In reality, timing errors in hydrologic predictions typically change with time (e.g., varying from one event to another). This requires capturing the temporal variations in timing error analysis. Since a wavelet is not completely localized in the time domain, the results from applying XWT may be impacted by the temporal resolution of the mother wavelet and the convolution process involved in wavelet transformations. In order to investigate this impact, additional synthetic time series are generated by applying a different timing error to each of the individual hydrologic events. Note that the prescribed timing error is still constant within the time period of a given event, but varies between events (i.e., the timing errors are event-specific).

Fig. 5 presents the results from applying XWT to synthetic simulation time series generated by applying event-specific timing errors to the truncated time series of observed streamflow for LYNT2 (Fig. 5a) and QLAT2 (Fig. 5b). The prescribed timing errors for the individual events for the two basins are listed in Table 2. In the case of LYNT2, the "true" timing error increases at a fixed interval of 3 h from 0 h for the first event to 21 h for the last; for QLAT2, the event-specific timing errors were randomly sampled from a discrete uniform distribution with hourly offsets between −25 h and +25 h. A maximum error of 25 h (+ or −) is chosen because it is likely to be the maximum timing error that might occur in the real streamflow simulations for all the basins considered in this work, according to visual inspections of the observed and simulated flow time series. As indicated by the timing error spectra in Fig. 5, although the same timing error is applied to the entire time period of a given event (i.e., the prescribed timing error is scale-invariant, in addition to being time-invariant, within the range of that specific event), the "true" timing error is only properly manifest within a certain range of scales. For a given event with a prescribed timing error, there appears to be a "critical" scale for the XWT-based approach to accurately reproduce the prescribed timing error. At scales smaller than the critical scale, the timing error estimates from XWT tend to be noisy and biased toward the opposite of the prescribed errors. It is also clearly indicated in Fig. 5a that the critical scale tends to increase as the prescribed timing errors increases, implying the need to use larger scales when analyzing large timing errors. For example, for the second to seventh events shown in Fig. 5a (the first and last events are not considered due to edge effects), the critical scales are approximately 6, 12, 18, 24, 30, and 36 h, respectively, which are double the prescribed timing errors of 3, 6, 9, 12, 15, and 18 h for these events. This indicates that for the XWT-based technique to reliably estimate timing errors in the synthetic simulations, the scales used must be at least twice as large as the expected timing error. Intuitively, this is understandable because the convolution window needs to be large enough to allow the identification of phase errors in both the rising and recession limbs of a hydrograph peak, so that the phase error of the entire peak can be properly determined; otherwise, if the scale is too small, erroneous timing error estimates would be produced, e.g., from mistakenly relating the recession limb of the observed hydrograph to the rising limb of the simulated hydrograph. This is also a result of comparing observed and synthetically simulated hydrographs that have the exactly same shape. In reality, when the shapes of the observed and simulated hydrographs are different, the "critical" scale is expected to be larger than twice of the timing errors (see relevant discussions in Section 6). However, at large scales the timing error estimates for small events (e.g., the 2nd–4th, 6th, 8th–9th, and 11th–12th events in Fig. 5b) tend to be influenced by nearby events that have much larger peak values (e.g., the 1st, 5th, 7th, 10th, and 13th events in Fig. 5b), as indicated by the blending of color streams corresponding to the prescribed timing errors of these large events into the time periods of smaller, neighboring events at large scales (e.g., >80 h). This is largely caused by the larger convolution window size at larger scales and the larger wavelet coefficients for signals of larger magnitudes from which the timing error estimates are derived. The influence of neighboring events can be fairly large, depending on the separation in time between the two neighboring events, and their difference in magnitude.

Since, in this case, the timing error varies between the events, it would be more sensible to calculate a timing error for each individ-

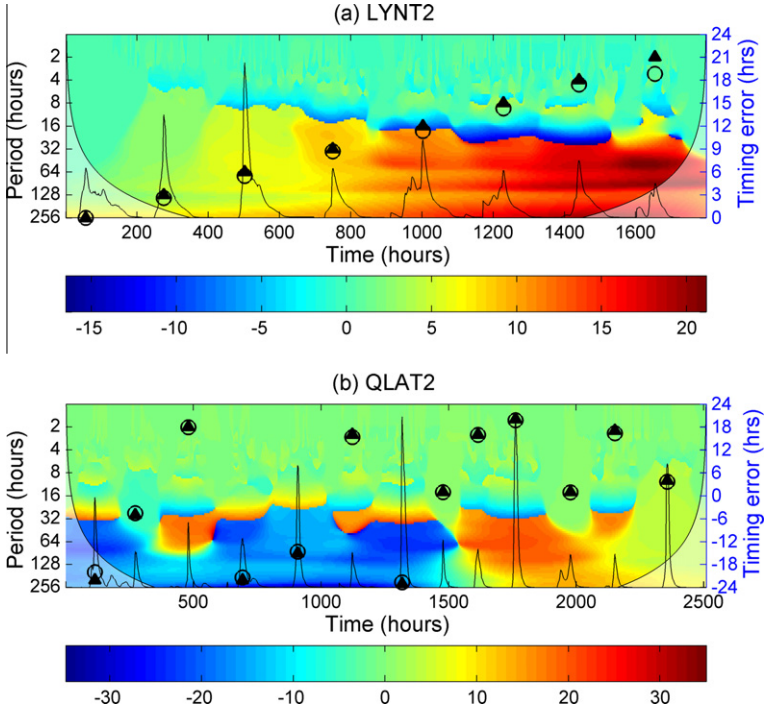

Fig. 5. Timing error spectra and the peak timing errors for the synthetic simulations for (a) LYNT2 and (b) QLAT2. Solid black triangles denote the prescribed timing errors, and open circles represent the estimated peak timing errors.

Table 2
The peak timing errors (TE) estimated from XWT for LYNT2 and QLAT2, as compared to the prescribed timing errors for individual events (Q_{peak} > 100 m^3/s).

Event	LYNT2			QLAT2		
	TE prescribed (h)	TE estimated (h)	Error in estimated TE (h)	TE prescribed (h)	TE estimated (h)	Error in estimated TE (h)
1	0	0.0	0.0	−22	−19.95	2.05
2	3	2.60	−0.40	−5	−4.56	0.44
3	6	5.47	−0.53	18	17.88	−0.12
4	9	8.68	−0.32	−22	−21.35	0.65
5	12	11.34	−0.66	−15	−14.60	0.40
6	15	14.29	−0.71	16	15.31	−0.69
7	18	17.41	−0.59	−23	−22.61	0.39
8	21	18.80	−2.20	1	0.92	−0.08
9	–	–	–	16	15.89	−0.11
10	–	–	–	20	19.70	−0.30
11	–	–	–	1	0.88	−0.12
12	–	–	–	17	16.32	−0.68
13	–	–	–	4	3.64	−0.36

ual event (instead of averaging the timing error spectrum over the entire time period to obtain a global timing error). To avoid the influence of neighboring events (i.e., the scale should be as small as possible) while considering the requirement of 'critical' scales, we focus on deriving the peak timing errors by averaging the timing error estimates within the scale range of $S_c \sim S_c + 20$ h (where S_c is the critical scale defined as twice of the absolute value of the prescribed timing error) and a time period of 20 h (~0.1 of the average event time period) around the peak time of the individual hydrographs. This averaging within the narrow scale and time ranges helps avoid erroneous timing error estimates, while also cancelling out some of the noise from wavelet transformation. The peak timing errors calculated in this manner for each individual event for LYNT2 and QLAT2, as compared to the prescribed (or "true") timing errors, are displayed in Fig. 5 and Table 2. Except for those events possibly influenced by edge effects (e.g., the last event for LYNT2 and the first event for QLAT2), the XWT-derived peak timing error estimates are fairly accurate for all the events, with a bias of less than one hour (i.e., the bias in the timing error estimates is within the range of the hourly time step of the data). It is worth noting that in practice, the "true" timing errors and hence the "critical" scales are typically unknown, requiring a different strategy to identifying suitable scales for deriving the timing errors (see Section 6 for more discussion on this issue).

To assess the uncertainty bounds of the bias in XWT-based timing error estimates, 1000 synthetic Monte Carlo simulations are produced with randomly generated event-specific timing errors for the basin LYNT2. Here, the original un-truncated (instead of the truncated) time series are used, to avoid or reduce the impact of edge effects, as well as neighbor influences on originally well separated events that otherwise could experience larger influences due to truncation. Again, the prescribed timing errors are randomly sampled from the uniform distribution of [−25, +25] hours. After applying XWT to each of the 1000 synthetic simulations, the event-specific peak timing errors are derived by averaging the timing error spectra over the scale range of $S_c \sim S_c + 20$ h and the time

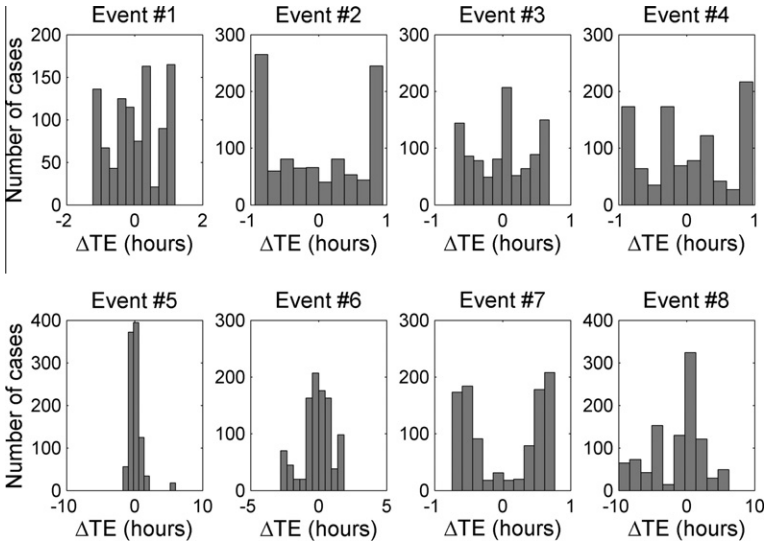

Fig. 6. Error in XWT-based peak timing error estimates for 1000 synthetic Monte Carlo simulations for the eight events of LYNT2. $\Delta TE = TE_{XWT} - TE_{true}$.

period of 20 h around the event peak times. The results of peak timing errors from the Monte Carlo experiments are presented as histograms of errors in the timing error estimates in Fig. 6, where it indicates that the timing error estimates for four of the events (#2, #3, #4, and #7) are accurate to within a range of −1 to +1 h (the time step of the data series). However, errors in the timing error estimates for the remaining four events, especially event #8, can be as large as 10 h. A close examination of these four events (see Fig. 2b) reveals that they are all double-peak or multi-peak events, indicating that the XWT-based timing error estimation may not be reliable for multi-peak and consecutive hydrologic events. The error in timing error estimation can be large when the two consecutive peaks have similar magnitudes (e.g., event #8) and would be small when the major peak is much larger than the second peak (e.g., events #1, #5, and #6). It is worth noting that the edge effects should be negligible here, as the use of the original (un-truncated) time series has created enough room for waves of 150 h or smaller to pass through without introducing possible edge effects on the eight events considered.

5.3. Summary of results from the synthetic experiments

The results from the synthetic experiment with constant prescribed timing errors (Section 5.1) indicate that the global timing error can be reliably estimated from XWT by averaging the timing error spectrum over an appropriate range of scales and the entire time period. However, when the timing error is event specific or varies with time (Section 5.2), as is typically the case for real-world hydrologic predictions, the influence of neighboring events becomes important, rendering it necessary to use a narrower scale band and an event-specific time period for averaging. The results indicate the presence of a "critical" scale, roughly equivalent to double the prescribed timing error, which must be equaled or exceeded for the XWT-based timing error estimates to be reasonable. A Monte Carlo experiment with 1000 synthetic simulations indicates that the timing error estimates obtained by averaging over a scale band of 20 h above the critical scale and a time period of 20 h around the event peaks are fairly reliable for single-peak events, with a small bias not exceeding the time step of the data series. However, for multi-peak or consecutive events, the XWT-based timing error estimates can be inaccurate, depending on the shape and magnitude of the peaks and the true timing errors associated with these peaks. This is intuitively sensible, as the presence of two or more consecutive peaks could be potentially confusing when relating a peak in the observation to a peak in the simulation to derive the phase relationship.

6. Application to hydrologic simulations

The synthetic experiments described in Section 5 help establish the confidence in using XWT for peak timing error estimation and provide useful insights into the factors that control its reliability. Here, the XWT-based technique is applied to the original, un-truncated, streamflow simulations (described in Section 3) to produce the timing error spectra for the 11 basins, which are then averaged over an appropriate scale range (to be discussed below) and a time period (20 h) around the peak time of the selected events to estimate their peak timing errors. In order to obtain more events for deriving the uncertainty bounds of the peak timing error for each basin, the threshold of the peak flow value for identifying the hydrologic events is set to 50 m^3/s (whereas 100 m^3/s is used in the synthetic experiments in Section 5 to limit the number of events for easier visual examination). It is important to note that the use of a smaller threshold increases the number of events in each basin that would reduce the sampling uncertainty in the peak timing error estimates, while the findings from the synthetic experiments should still be applicable. When averaging over time, a 20-h period around the peak time of each individual event is used (as in the synthetic experiment). This should allow proper derivation of timing error estimates for the event peaks while avoiding possible neighbor influences.

The estimation of appropriate scales is less straightforward than in the synthetic experiments. For these real simulations, the "true" timing errors (and thus the "critical" scales) are unknown, requiring other means of identifying the appropriate range of scales for extracting meaningful timing error estimates. In this study, the cross wavelet power spectrum is scanned to identify the scales that correspond to the maximum wavelet power at the peak time period of each individual event. These scales have the highest common energy among the simulated and observed streamflow hydrographs at their corresponding peak time periods. In other words, among all the scales analyzed, these constitute the dominant scales for both the observed and simulated event peaks (i.e., they contribute most significantly to the hydrographs during the peak time periods), making them potentially suitable scales for extracting the timing error estimates. For convenience, these are

subsequently referred to as the "characteristic" scale of the event peaks. It is expected that the "characteristic" scales (as defined here) would typically be smaller for high-frequency event peaks than for low-frequency base flows. In this study, the XWT-based timing error spectra is averaged over a scale band of 10 h around the characteristic scale and a time period of 20 h around the peak time to derive the peak timing error for each individual event. For all events, the characteristic scales are found to be much larger than the "critical scale" (double the timing error) suggested by the synthetic experiments, indicating that, when the observed and simulated hydrographs have different shapes, the scales appropriate for analyzing timing errors need to be larger than double of the timing error.

Fig. 7 (first row) shows the time error spectra for a few selected events for LYNT2 (one event), QLAT2 (two well separated events), and SBMT2 (three consecutive events).

It is obvious that there exists a considerable amount of variability in the timing errors in both time and frequency domains for all events considered. To derive reliable peak timing error estimates, the average cross wavelet power corresponding to the peak time period is calculated for each individual event and shown in the second row of Fig. 7. For all the events considered, there exists a "characteristic" scale that corresponds to the maximum cross wavelet power. For example, the characteristic scale is 62.4 h for the event of LYNT2, 55.6 and 37.1 h for the two events of QLAT2, and 124.8 h for the three events of SBMT2. When averaging the time error spectra over the 10-h scale band around the characteristic scales and the 20-h time period around the event peaks, the peak timing errors obtained for these events are 9.45, 9.02, 2.68, −12.74, −11.74, and −11.23 h, respectively. To check if these timing error estimates are reasonable, the hydrographs are timing-adjusted according to the corresponding peak timing error estimates and are compared to the observed hydrographs and the original simulated hydrographs as shown in the third row of Fig. 7. For example, for the event of LYNT2, since the estimated timing error is 9.45 h (i.e., simulation leading observation by 9 h), the simulated hydrograph is adjusted in timing by shifting it later (i.e., to the right) by 9 h. The events for QLAT2 and SBMT2 are also timing adjusted accordingly. As indicated in Fig. 7, after timing adjustment, the simulated events for LYNT2 and QLAT2 correspond much better with the observed events than the original simulated events. For the event of LYNT2, the RMSE is reduced from 31.4 to18.9 m^3/s and the correlation coefficient increased from 0.67 to 0.94; for the two events of QLAT2, the RMSE measures are reduced from 23.5 to 7.2 m^3/s and from 11.4 to 7.6 m^3/s, and the correlation coefficients increase from 0.59 to 0.98 and from 0.92 to 0.98. This indicates that the timing error estimates for these events are reasonable, while it also suggests that the presence of timing error can have a significant impact on the values of the evaluation measures.

For the three consecutive events considered for SBMT2, the results are less promising. The timing error estimate for the middle event, because of its much smaller magnitude, is heavily influenced by the two larger neighboring events. Visual inspection suggests that the original simulated hydrograph for the middle event corresponds with the observed hydrograph relatively well, with a very

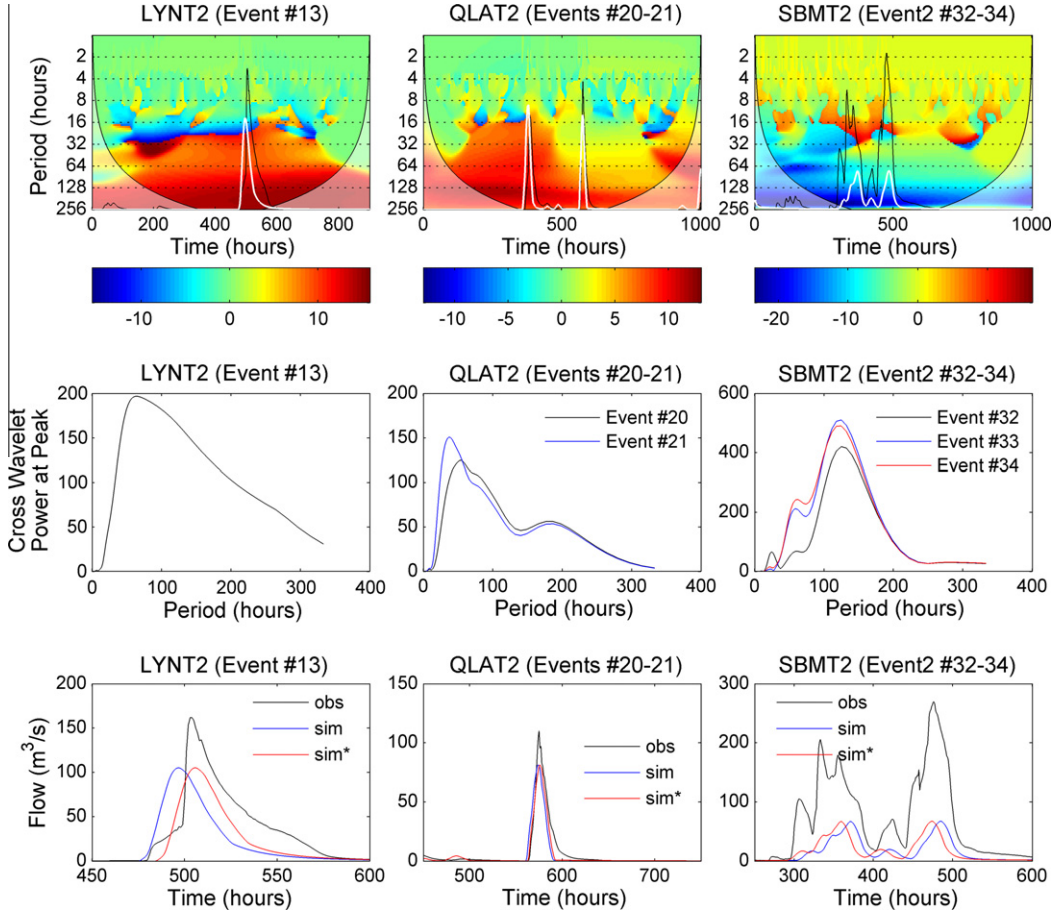

Fig. 7. Time error spectra (1st row), the cross wavelet power at event peaks (2nd row), and the observed and simulated hydrographs (3rd row) for selected events of LYNT2 (1st column), QLAT2 (2nd column) and SBMT2 (3rd column). In the first row, the observed and simulated flows are in black and grey, respectively. In the third row, sim* denotes the hydrographs obtained by shifting the original simulated hydrographs according to the estimated timing error.

small timing error (the bottom-right subplot in Fig. 7). However, since the simulations for the other two events are considerably lagging the observations, the estimated timing error for the middle event is heavily biased toward a negative timing error (−11.74 h). As a result, timing adjustment based on this erroneous timing error estimate leads to a degraded simulation with the correlation coefficient reduced from 0.97 to −0.57, and the RMSE increased from 37.0 to 40.6 m^3/s. For the other two events, because of their larger magnitudes, the neighbor influence is smaller and the timing adjustment does lead to some marginal improvement in the RMSE correlation coefficient measures. This is consistent with the observation from the synthetic experiments that the XWT-based timing error estimates may not be reliable for multi-peak or consecutive events.

Fig. 8 presents the peak timing errors estimated by the XWT-based approach for all 11 basins, arranged in order of increasing time to peak (T_p) value (see Table 1). The box-whisker plot shows the distribution of the peak timing errors for the 11 basins (see Table 1 for the number of events in each basin), with the boxes representing the central 50 percentile and the whiskers extending to the most extreme data points not considered as outliers (which are plotted individually as "+"). As indicated in Fig. 8, there are two distinct groups of basins, with the first group comprising six basins (QLAT2, SKMT2, UVAT2, MCKT2, SCDT2, and MTPT2), which exhibit relatively small timing errors, and the second group comprising the remaining five basins (SDAT2, LYNT2, MDST2, SBMT2, and REFT2), which exhibit relatively large timing errors. The first group of basins has a T_p value smaller than or equal to 17 h, while the second group has a T_p value larger than or equal to 17 h, indicating that the event-based timing errors generally increase with the time to peak value. For basins with similar T_p values, the timing errors tend to be larger for drier basins. For example, although MCKT2 and SCDT2 both have a T_p value of 14 h, the timing error is much larger for the drier basin SCDT2 (average annual discharge = 3.0 m^3/s) than for the wetter basin MCKT2 (average annual discharge = 4.1 m^3/s). This is also true for MTPT2, SDAT2, and LYNT2 which have similar T_p values of 17 or 18 h. However, the basin SBMT2, although has a relatively large T_p value (26 h), exhibits relatively small timing errors in the second group, probably due to the relatively high wetness of the basin (average annual discharge = 5.01 m^3/s and average annual precipitation = 934 mm). Given that the sample size (i.e., the number of events considered) is relatively small for each basin, the same results shown in Fig. 8 are presented using histograms in Fig. 9, to examine the timing error distribution of each individual basin for more detailed information that may not be readily observed in Fig. 8. For example, Fig. 9 shows that for the basin MCKT2, only two (out of 16 events considered) have an estimated timing error larger than −5 h (i.e., the time error estimates for all other events are between 0 and −5 h). As indicated by the histograms and the box plot, several basins tend to consistently exhibit either positive errors (e.g., QLAT2 and SDAT2) or negative timing errors (e.g., UVAT2, MCKT2, SBMT2, and REFT2), indicating that the hydrologic models may have a systematic deficiency in predicting streamflow in these basins. In the study area, large spatiotemporal variability exists in rainfall. As such, it is quite common that the assumptions behind unit hydrograph (Chow et al., 1988; Lee et al., 2008) are not well met. Also, the unit hydrographs used in this work were derived using all available data in one batch (Seo et al., 2009) and hence do not capture event-to-event variations. In general, these results suggest that the peak timing errors are not necessarily closely related to the size of a basin (contrary to what one would expect), but tend to depend on the average time for the events to peak and the amount of water in the basin. Using a climatological threshold, which presumably would be more meaningful given the hydro-climatic diversity among the basins, and a larger sample size (i.e., including more events from longer data records) might have led to timing error results that are more informative for model diagnosis purposes.

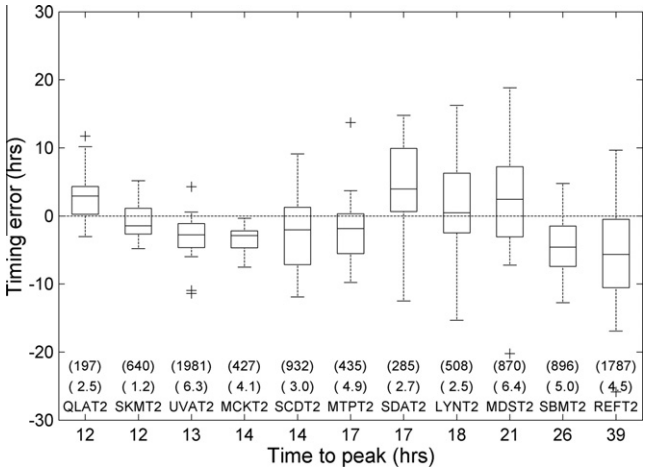

Fig. 8. Peak timing errors for all events in the 11 basins, arranged in order of increasing time to peak value. The numbers in the parentheses above the basin name give the basin size in km^2 (first row) and the average annual discharge in m^3/s (second row).

Since for the real simulations, the "true" timing errors are unknown, making it difficult to directly evaluate the accuracy of the timing error estimates in a strictly quantitative sense. It is, however, sensible to evaluate it in an indirect fashion by assessing the quality of the timing adjusted event hydrographs based on the peak timing errors estimated from the XWT-based approach, as shown in Fig. 7. If a timing error estimate is reasonable, the adjusted simulated hydrograph should line up better with the observed hydrograph than the original simulation. Here the streamflow simulations for all events of the 11 basins are adjusted in timing according to the event-based peak timing error estimates. The amount of adjustment for each individual event (283 events in total) is determined by rounding the XWT-derived timing error to the nearest integer hour (since the time step of the simulation is one hour). For example, given that the estimated timing error for the first event of LYNT2 is 0.79 h, the timing-adjusted simulation for this event is obtained by shifting the original simulation to the right by 1 h. The root mean squared error (RMSE) and correlation coefficient (CORR) values are then calculated for both the original simulation and the timing-adjusted simulation for each event. The results are presented in Fig. 10. A comparison between the correlation coefficient statistic before and after the timing adjustment in Fig. 10a indicates a large improvement due to the simple time shift based on the XWT-based timing error estimates, with the biggest improvement being as much as 0.8 (from 0.06 to 0.86) for the 12th event of SDAT2. Among the total 283 events, 194 events experience an improvement in correlation coefficient, while only 2 events experience a reduction in CORR that is larger than 0.1; the remaining events either have no change in CORR or exhibit a reduction of less than 0.1 in CORR. As far as the RMSE is concerned (Fig. 10b), the timing adjustment has led to a reduction in RMSE for 195 events, with the biggest reduction being as much as 80% for the 22nd event of REFT2 or 43.3 m^3/s for the 19th event of MDST2. Only one event out of the total 283 events suffers an increase in RMSE of larger than 5 m^3/s. A close examination of the observed hydrographs indicates that most of those events that experienced a poorer RMSE or CORR measure from timing adjustment are multi-peak or consecutive events, for which the timing error estimates from the XWT-based approach are known to be unreliable from the synthetic experiments. In this simple timing

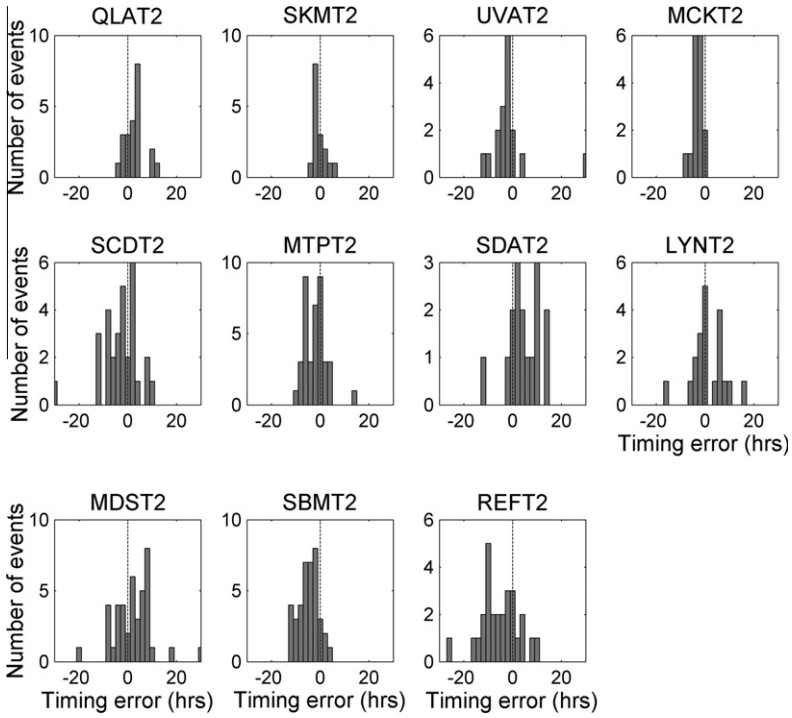

Fig. 9. Histograms of peak timing errors for all events in the 11 basins. The basins are arranged in order of increasing time to peak value.

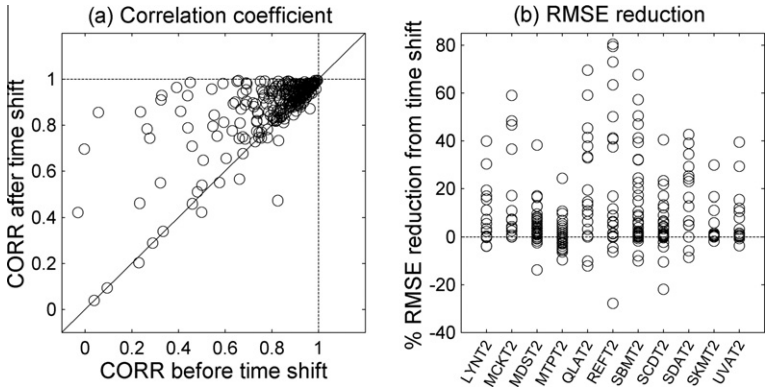

Fig. 10. (a) event-based correlation coefficient before and after timing adjustment and (b) percent reduction in event-based RMSE due to timing adjustment, for all events in the 11 basins.

adjustment process, rounding the timing error estimates to the nearest integer hours may have also introduced some degree of uncertainty, which could potentially be reduced if a temporal resolution finer than hourly were used. It is also important to note that, giving the different shapes of the observed and simulated hydrographs, removing timing error in the simulation may not always lead to an improved RMSE or CORR measure. In any case, evaluation based on the timing-adjusted simulation (instead of the original simulation) is expected to provide more effective information regarding the "true" magnitude error in the context of diagnostic model evaluation.

Fig. 11 presents the hydrographs of six selected events that receive large improvement from the timing adjustment, including the event mentioned above that receives the biggest increase in correlation coefficient (event #12, SDAT2). As indicated in Fig. 11, the XWT-derived timing error estimates for these events range from −25 h to +14 h, and the timing adjustments based on these timing error estimates have led to considerably improved agreement between the observations and the simulations. Together, the results in Figs. 10 and 11 indicate that the timing adjustment has effectively removed at least part of the timing error in the simulations, suggesting that the timing error estimates from XWT shown in Figs. 8 and 9 are mostly reasonable and reliable. These timing error estimates provide a useful reference for the potential uncertainty in the timing of the streamflow predictions in each of the 11 basins. However, a sufficiently large number of events need to be analyzed for each basin (e.g., by using longer data records) to derive statistically meaningful uncertainty bounds for timing predictions in each basin. Also, since the timing errors tend to be event specific, it would be difficult to implement the timing adjustment as discussed above in a real-time forecasting context to improve the forecast of a specific event, given that it is difficult to reliably estimate the timing errors in the forecast using the XWT-based approach with the observation unknown (see Section 7 for more discussion on potential ways of using timing error information in real-time forecasting).

The substantial changes in the RMSE and correlation coefficient measures from the timing adjustment suggest that the presence of

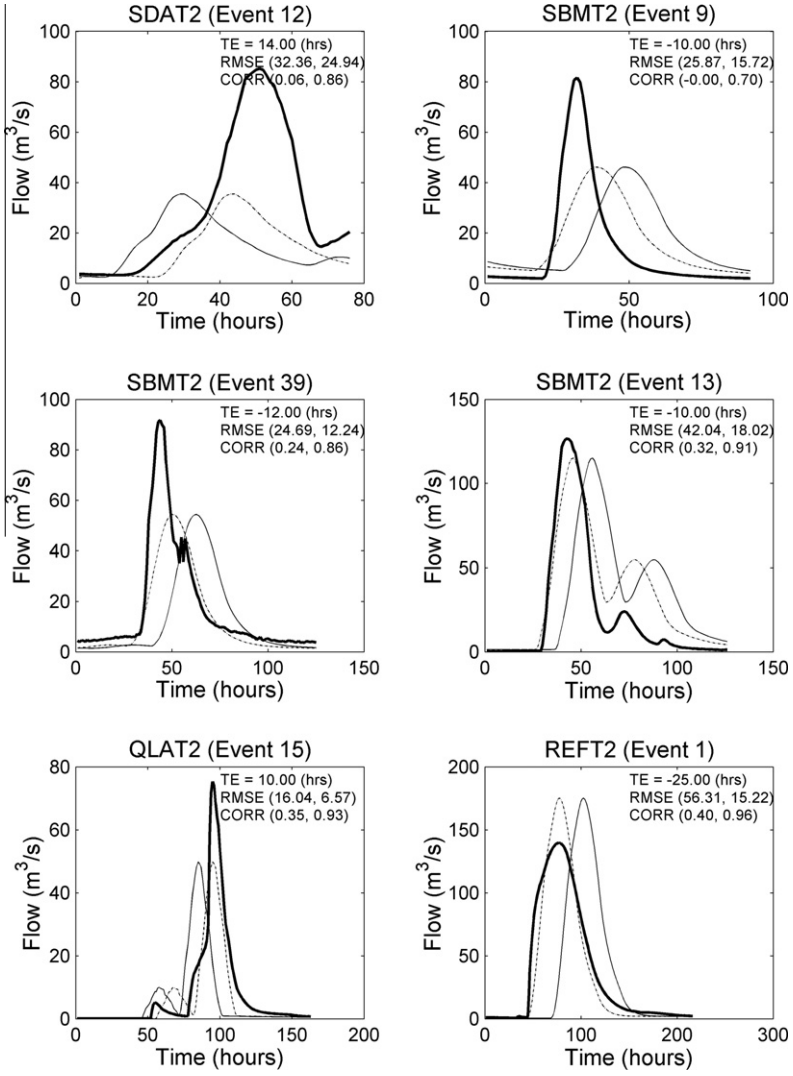

Fig. 11. Simulations before (light solid) and after (light dash) the timing adjustment, as compared to the observed hydrographs (dark) for selected events. TE = event-based peak timing error estimated from XWT; for RMSE and correlation coefficient, the first and second numbers in the parentheses give the values before and after the adjustment, respectively.

timing errors could have a significant impact on the evaluation of hydrologic predictions, indicating the need to separate timing errors from magnitude errors in hydrologic evaluation. Also, it is important to note that the improvement in the agreement between the observations and simulations would be much less appreciable if the RMSE and correlation coefficient measures were calculated for the entire time period of the data series, instead of each event period, suggesting the need to consider the temporal variations of timing errors when evaluating hydrologic predictions.

7. Discussion and future work

Streamflow predictions contain errors in both magnitude and timing. These different types of error have different implications and ramifications for both model diagnosis and decision support and thus should be distinguished and separated in model evaluation and forecast verification applications, where possible and relevant. In practice, however, this has rarely been the case, either in hydrologic research or operational forecasting, largely due to insufficient recognition of the importance of providing timing uncertainty information, as well as a lack of effective approaches to reliably assessing timing errors. In traditional hydrologic evaluation, error measures are typically used in a lumped fashion, with no distinction between errors in timing and magnitude and no consideration for the temporal variations of these errors. This has led to inefficiency in hydrologic evaluation and many other issues including the model equifinality problem.

In this paper, we discuss the importance of including timing error assessment as an essential and integral component of hydrologic evaluation, including forecast verification applications. The results from the experiments carried out in this study, with both synthetic and real streamflow simulations, indicate that the cross wavelet transform technique is an effective approach to quantifying timing errors in hydrologic predictions. The advantages of using an XWT-based approach include being able to provide localized timing error information in both time and frequency domains, and the possibility of readily applying the method to a long data record with minimum supervision (if a robust automatic peak identification method is available). The synthetic experiments with prescribed timing errors (including the Monte Carlo simulations) help examine the accuracy and reliability of this technique for timing error estimation, while providing useful insights into the specific conditions required for XWT-based timing error estimates to be reliable. Encouraging results from applications to the real streamflow simulations, including considerably improved agree-

ment between the observation and the simulation from a simple timing adjustment based on the estimated timing errors, further confirm the potential of the XWT-based method for estimating timing errors. As indicated by the application of XWT to both the synthetic cases and real streamflow simulations, the timing error estimates are reliable for well separated single-peak events, but may be unreliable or even erroneous for multi-peak or consecutive events, due to neighboring influences from the convolution involved in wavelet transformation, especially at large scales.

Further work should help establish the statistical significance of the XWT results for timing error estimation (e.g., Marraun and Kurths, 2004; Schaefli et al., 2007). Also, the choice of Mother wavelet warrants further investigation. The Morlet wavelet chosen in the study, although providing a good balance between temporal resolution and frequency resolution, does not have the best temporal resolution, which may have contributed to the difficulty in quantifying the timing errors of contiguous events. Hence, it would be worthwhile to explore the potential of a wavelet with a better time resolution, such as the complex order two Gaussian wavelet used in Lane (2007). One important purpose of timing error analysis is to be able to remove the timing error in the hydrologic prediction to produce a new prediction that is less contaminated by timing errors, and thus more amenable to traditional hydrologic evaluation, which typically focuses on magnitude errors. In this study, a simple approach is taken by shifting the hydrologic events in time according to the XWT-derived timing error estimates. This led to improved streamflow simulation in most cases. A more formal approach, however, would be to first remove the timing/phase errors in the wavelet domain and then use wavelet reconstruction techniques (Teti and Kritikos, 1992) to synthesize a new simulation that contains no or less timing error. It is worth noting here that, in most cases, the observed and modeled hydrographs have different shapes, rendering it impossible to completely separate timing and magnitude errors using the techniques discussed in this paper. In addition, historical data typically contain missing values which must be estimated via robust techniques before applying the XWT technique. Finally, considering the neighboring influences inherent in the XWT-based approach, only scale-invariant peak timing errors are investigated in the present study. However, since timing errors in hydrologic predictions can be scale-dependent, it would be desirable to also explore the scale-variant aspect of timing errors. For example, one could potentially expand the proposed XWT-based approach to assess timing errors associated with the different components of a hydrograph (e.g., quick flow and base flow) or the different segments of a hydrograph (e.g., rising limb, quick recession, and slow recession), to provide more specific timing error information to facilitate diagnostic model evaluation where the aim is to detect the exact sources of the timing errors. This is also an important consideration for real-time applications, where the forecast for the entire peak of an upcoming event may not be available, depending on the specific forecast creation time and the forecast lead time. Characterizing timing errors associated with the different hydrograph components (or segments) would require an effective strategy to decompose the hydrographs and a reliable solution to remove neighboring influences, which should constitute a potential topic for a follow-up study. For base flows, one could potentially consider filling sufficient low flow values between events to artificially separate them from each other. This strategy could also be potentially adopted to reduce the impact of neighboring events on the timing error estimation for multi-peak or consecutive events.

In this study, the XWT-based approach is only used to assess timing error in streamflow predictions. But the approach is expected to be applicable to other hydrologic or hydrometeorological variables such as precipitation, temperature, or soil moisture contents. There are many potential applications for reliable estimates of timing error. These could include diagnostic model evaluation, real-time verification, parameter sensitivity analysis, and model calibration, among many others. For instance, one could incorporate the XWT-based approach into a parameter sensitivity analysis study (e.g., Van Werkhoven et al., 2008) to identify those model parameters mainly responsible for the timing errors in the predictions, so that calibration efforts can focus on those specific parameters to more effectively correct the timing errors. As another example, timing error estimation can be used in studies aiming to inter-compare the performance of distributed and lumped hydrologic models, to more reliably assess the potential gain (if any) from distributed modeling in predicting the timing of hydrologic events. For real-time forecasting, there may be opportunities to distinguish between timing and amplitude errors in statistical post-processing techniques, leading to improved bias correction and uncertainty accounting; also, estimates of timing error based on appropriate historical analogs of forecast events and associated observations can lead to reasonable prediction of timing uncertainty in real-time flood forecasting, thus enhancing protection of life and property. With increasing development and uses of the XWT-based approach and other methods for quantifying timing errors in hydrologic predictions, it is expected that timing error assessment will become an essential and integral component of hydrologic evaluation.

Acknowledgement

This work is supported by the Advanced Hydrologic Prediction Service (AHPS) program of the National Weather Service (NWS) and the Climate Predictions Program for the Americas (CPPA) of the Climate Program Office (CPO) of the National Oceanic and Atmospheric Administration (NOAA). Core wavelet analysis routines were provided by A. Grinsted and are available at http://www.pol.ac.uk/home/research/waveletcoherence/. The authors would like to thank Robert Corby and Paul McKee of the West Gulf River Forecast Center of the US NWS for providing the basin map and streamflow datasets. Comments from Pedro Restrepo, Geoff Bonnin, Haksu Lee, and Gary Carter of the NWS Office of Hydrologic Development helped to improve an earlier version of this paper. The authors would also like to thank Alberto Viglione and an anonymous reviewer for their helpful comments.

References

Abrahart, R.J., Heppenstall, A.J., See, L.M., 2007. Timing error correction procedure applied to neural network rainfall–runoff modeling. Hydrological Sciences Journal 52 (3), 414–431.

Anctil, F., Pratte, A., Parent, L.E., Bolinder, M.A., 2008. Non-stationary temporal characterization of the temperature profile of a soil exposed to frost in southeastern Canada. Nonlinear Processes in Geophysics 15, 409–416.

Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modeling. Advances in Water Resources 16 (1), 41–51.

Beven, K.J., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. Journal of Hydrology 249, 11–29.

Boyle, D.P., Gupta, H.V., Sorooshian, S., 2000. Toward improved calibration of hydrological models: combining the strengths of manual and automatic methods. Water Resources Research 36(12), 3663–3674.

Bradley, A.A., Schwartz, S.S., Hashino, T., 2004. Distribution-oriented verification of ensemble streamflow predictions. Journal of Hydrometeorology 5 (3), 532–545.

Briggs, W.M., Levine, R.A., 1997. Wavelets and field forecast verification. Monthly Weather Review 125, 1329–1341.

Brown, J.D., Demargne, J., Seo, D.-J., Liu, Y., 2010. The Ensemble Verification System (EVS): a software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. Environmental Modelling & Software 25 (7), 854–872.

Burnash, R.J., Ferral, R.L., McGuire, R.A., 1973. A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers, US Department of Commerce National Weather Service and State of California Department of Water Resources.

Chow, V.T., Maidment, D.R., Mays, L.W., 1988. Applied Hydrology. McGraw-Hill, New York. pp. 530–537.

Coulibaly, P., Burn, D., 2004. Wavelet analysis of variability in annual Canadian streamflows. Water Resources Research 40, W03105. doi:10.1029/2003WR002667.

Dawson, C.W., Abrahart, R.J., See, L.M., 2007. HydroTest: a web-based toolbox of evaluation metrics for the standardized assessment of hydrological forecasts. Environmental Modelling & Software 22, 1034–1052.

Demargne, J., Mullusky, M., Werner, K., Adams, T., Lindsey, S., Schwein, N., Marosi, W., Welles, E., 2009. Application of forecast verification science to operational river forecasting in the US national weather service. Bulletin of the American Meteorological Society 90 (6), 779–784.

Gilleland, E., Ahijevych, D., Brown, B.G., Casati, B., Ebert, E., 2009. Intercomparison of spatial forecast verification methods. Weather and Forecasting 24, 1416–1430.

Grinsted, A., Moore, J.C., Jevrejeva, S., 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes in Geophysics 11, 561–566.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrological models: multiple and noncommensurable measures of information. Water Resources Research 34, 751–763.

Gupta, H.V., Wagener, T., Liu, Y., 2008. Reconciling theory with observations: towards a diagnostic approach to model evaluation. Hydrological Processes 22, 3802–3813.

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 modeling. Journal of Hydrology 377, 80–91.

Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting 15, 559–570.

Huang, M., Liang, X., 2006. On the assessment of the impact of reducing parameters and identification of parameter uncertainties for a hydrologic model with applications to ungauged basins. Journal of Hydrology 320, 37–61.

Jachner, S., van den Boogaart, K.G., 2007. Statistical methods for the qualitative assessment of dynamic models with time delay (R Package qualV). Journal of Statistical Software 22 (8).

Jevrejeva, S., Moore, J.C., Grinsted, A., 2003. Influence of the Arctic Oscillation and El Nino-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: the wavelet approach. Journal of Geophysical Research 108 (D21), 4677. doi:10.1029/2003JD003417.

Koren, V., Reed, S., Smith, M., Zhang, Z., Seo, D.-J., 2004. Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. Journal of Hydrology 291, 297–318.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. In: Krause, P., Bongartz, K., Flügel, W.-A. (Eds.), Proceedings of the 8th Workshop for Large Scale Hydrological Modelling-Oppurg 2004. Advances in Geosciences 5, 89–97.

Kumar, P., Foufoula-Georgiou, E., 1997. Wavelet analysis for geophysical applications. Reviews of Geophysics 35 (4), 385–412.

Kuzmin, V., Seo, D.-J., Koren, V., 2008. Fast and efficient optimization of hydrologic model parameters using a priori estimates and stepwise line search. Journal of Hydrology 353 (1–2), 109–128.

Labat, D., Ronchail, J., Guyot, J.L., 2005. Recent advances in wavelet analyses: Part 2 – Amazon, Parana, Orinoco and Congo discharges time scale variability. Journal of Hydrology 314, 289–311.

Lane, S.N., 2007. Assessment of rainfall–runoff models based upon wavelet analysis. Hydrological Processes 21, 586–607.

Lee, K.T., Chen, N.-C., Chung, Y.-R., 2008. Derivation of variable IUH corresponding to time-varying rainfall intensity during storms. Hydrological Sciences Journal 53 (2), 323–337.

Liu, Y., Gupta, H.V., 2007. Uncertainty in hydrological modeling: toward an integrated data assimilation framework. Water Resources Research 43, W07401. doi:10.1029/2006WR005756.

Liu, Y., Gupta, H.V., Springer, E., Wagener, T., 2008. Linking science with environmental decision making: Experiences from an integrated modeling approach to supporting sustainable water resources management. Environmental Modelling & Software 23, 846–858.

Marraun, D., Kurths, J., 2004. Cross wavelet analysis: significance testing and pitfalls. Nonlinear Processes in Geophysics 11, 505–514.

Matott, L.S., Babendreier, J.E., Purucker, S.T., 2009. Evaluating uncertainty in integrated environmental models: aA review of concepts and tools. Water Resources Research, 45, W06421. doi:10.1029/2008WR007301.

Meng, H., Green, T.R., Salas, J.D., Ahuja, L.R., 2008. Development of testing of a terrain-based hydrologic model for spatial Hortonian infiltration and run-of/on. Environmental Modelling & Software 23, 794–812.

Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D., Stouffer, R.J., 2008. Stationarity is dead: whither water management. Science 319, 573–574.

Murphy, A.H., Winkler, R.L., 1992. Diagnostic verification of probability forecasts. International Journal of Forecasting 7, 435–455.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I: a discussion of principles. Journal of Hydrology 10, 282–290.

Perica, S., Foufoula-Georgiou, E., 1996. Linkage of scaling and thermodynamic parameters of rainfall: results from midlatitude mesoscale convective systems. Journal of Geophysical Research – Atmospheres 101 (D3), 7431–7448.

Ramsay, J.O., Li, X., 1998. Curve registration. Journal of the Royal Statistical Society, Series B 60, 351–363.

Ravela, S., Emanuel, K., McLaughlin, D., 2007. Data assimilation by field alignment. Physica a 230, 127–145.

Reilly, C., Price, P., Gelman, A., Sandgathe, S.A., 2004. Using image and curve registration for measuring the goodness of fit of spatial and temporal predictions. Biometrics 60, 954–964.

Reusser, D.E., Blume, T., Schaefli, B., Zehe, E., 2008. Analysing the temporal dynamics of model performance for hydrological models. Hydrology and Earth System Sciences Discussions 5, 3169–3211.

Schaefli, B., Gupta, H.V., 2007. Do nash values have value? Hydrological Processes 21, 2075–2080.

Schaefli, B., Zehe, E., 2009. Hydrological model performance and parameter estimation in the wavelet-domain. Hydrology and Earth System Sciences Discussions 6, 2451–2498.

Schaefli, B., Maraun, D., Holschneider, M., 2007. What drives high flow events in the Swiss Apls? Recent developments in wavelet spectral analysis and their application to hydrology. Advances in Water Resources 30, 2511–2525.

Seibert, J., 2001. On the need for benchmarks in hydrological modeling. Hydrological Processes 15, 1063–1064. doi:10.1002/hyp.446.

Seibert, J., McDonnell, J.J., 2002. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. Water Resources Research 38 (11), 1241. doi:10.1029/2001WR000978.

Seo, D.-J., Koren, V., Cajina, N., 2003. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. Journal of Hydrometeorology 4, 627–641.

Seo, D.-J., Cajina, L., Corby, R., Howieson, T., 2009. Automatic state updating for operational streamflow forecasting via variational data assimilation. Journal of Hydrology 367, 255–275.

Shamir, E., Imam, B., Morin, E., Gupta, H.V., Sorooshian, S., 2005a. The role of hydrograph indices in parameter estimation of rainfall–runoff models. Hydrological Processes 19, 2187–2207.

Shamir, E., Imam, B., Gupta, H.V., Sorooshian, S., 2005b. Application of temporal streamflow descriptors in hydrologic model parameter estimation. Water Resources Research 41, W06021. doi:10.1029/2004WR003409.

Smith, L.C., Turcotte, D.L., Isacks, B.L., 1998. Stream flow characterization and feature detection using a discrete wavelet transform. Hydrological Processes 12, 233–249.

Smith, M.B., Koren, V.I., Zhang, Z., Reed, S.M., Pan, J.-J., Moreda, F., 2004. Runoff response to spatial variability in precipitation: an analysis of observed data. Journal of Hydrology 298, 267–286.

Teegavarapu, R., Elshorbagy, A., 2005. Fuzzy set based error measure for hydrologic model evaluation. Journal of Hydroinformatics 7 (3), 199–208.

Teti, J.G., Kritikos, H.N., 1992. SAR ocean image representation using wavelets. IEEE Transactions on Geoscience and Remote Sensing 30, 1089–1094.

Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 79 (1), 61–78.

Torrence, S., Webster, P.J., 1999. Interdecadal changes in the ENSO-monsoon system. Journal of Climate (12), 2679–2690.

Van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2008. Characterization of watershed model behavior across a hydroclimatic gradient. Water Resources Research 44. doi:10.1029/2007WR006271.

Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall–runoff modelling: dynamic identifiability analysis. Hydrological Processes 17 (2), 455–476.

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., Wilson, J.S., 2010. The future of hydrology: an evolving science for a changing world. Water Resources Research 46, W05301. doi:10.1029/2009WR008906.

Willems, P., 2009. A time series tool to support the multi-criteria performance evaluation of rainfall–runoff models. Environmental Modelling & Software 24, 311–321.

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 Resources Research 44, W09417. doi:10.1029/2007WR006716.

Zhang, Q., Xu, C.-Y., Jiang, T., Wu, Y., 2007. Possible influence of ENSO on annual maximum streamflow of the Yangtze River. China Journal of hydrology 333, 265–274.

Zolezzi, G., Bellin, A., Bruno, M.C., Maiolini, B., Siviglia, A., 2009. Assessing hydrological alterations at multiple temporal scales: Adige River, Italy. Water Resources Research 45, W12421. doi:10.1029/2008WR007266.

Zoppou, C., Nielsen, O.M., Zhang, L., 2002. Regionalization of daily stream flow in Australia using wavelets and k-means analysis. <http://wwwmaths.anu.edu.au/research.reports/mrr/02/003/> (accessed 26.01.10).