**Hydrology and
Earth System
Sciences**

Adaptive correction of deterministic models to produce probabilistic forecasts

P. J. Smith[1]**, K. J. Beven**[1]**, A. H. Weerts**[2]**, and D. Leedal**[1]

[1]Lancaster Environment Centre, Lancaster University, UK
[2]Deltares, P.O. Box 177, 2600 MH Delft, The Netherlands

Correspondence to: P. J. Smith (p.j.smith@lancs.ac.uk)

Abstract. This paper considers the correction of deterministic forecasts given by a flood forecasting model. A stochastic correction based on the evolution of an adaptive, multiplicative, gain is presented. A number of models for the evolution of the gain are considered and the quality of the resulting probabilistic forecasts assessed. The techniques presented offer a computationally efficient method for providing probabilistic forecasts based on existing flood forecasting system output.

1 Introduction

The basis of many operational hydrological forecasting systems are process based models producing deterministic forecasts. Often significant resources have been invested in acquiring these models and users are familiar with their use and limitations. In many situations such models produce biased or inaccurate predictions of discharge or water level (Aronica et al., 1998; Pappenberger et al., 2007). This makes the issuing of accurate and reliable flood forecasts challenging.

Data assimilation (DA) has been used to address this challenge in two ways: assimilating observations to improve the process model predictions and assimilating observations to improve the representation of the prediction errors. Human forecasters widely practise both forms of DA. Manually altering the internal states of the model based on their interpretation of recent model forecast errors may act to improve future model predictions. The forecaster may use their knowledge of the recent prediction errors of the model in deciding when to issue flood warnings, thereby implicitly utilising the second type of DA. The effectiveness and consistency (across

forecasters) of such manual DA techniques is rarely reported formally (Seo et al., 2009).

These manual DA techniques can be formalised to produce deterministic assimilation schemes (e.g. Cole et al., 2009; Moore, 2007). The DA process can also be cast in a probabilistic framework with the aim of constructing the predictive distribution $P\left(y_{t+f} \mid \boldsymbol{y}_{1:T}\right)$ of the observation of some quantity of interest (e.g. water level or discharge); f time steps ahead given $\boldsymbol{y}_{1:T} = (y_1, \ldots, y_T)$ the observations of that quantity up to the current time t.

If the aim of the DA is to improve the predictions of a hydrological model $\mathcal{M}$, a common framework (e.g. Liu and Gupta, 2007) is to cast the model in state space form so that the hydrological states (indexed by time) $\boldsymbol{s}_t$ evolve according to

$$\boldsymbol{s}_{t+1} = \mathcal{M}\left(\boldsymbol{s}_t, \boldsymbol{u}_t, \varepsilon_t\right) \tag{1}$$

where the $\boldsymbol{u}_t$ are observed extraneous inputs (e.g. precipitation) and ε_t a stochastic noise. The model states are then related to the observed values by the observation function $\mathcal{H}$ and stochastic noise ζ_{t+1}:

$$y_{t+1} = \mathcal{H}\left(\boldsymbol{s}_{t+1}, \zeta_{t+1}\right). \tag{2}$$

The stochastic term ε_t may be additive, that is

$$\boldsymbol{s}_{t+1} = \mathcal{M}\left(\boldsymbol{s}_t, \boldsymbol{u}_t\right) + \varepsilon_t.$$

It may also act within $\mathcal{M}$ to represent a number of features such as noise on the forcing term $\boldsymbol{u}_t$ or time evolving model parameters (e.g. Rajaram and Georgakakos, 1989). By correcting the states of the model it may be hoped that predictions derived for unobserved sites (such as the internal

Table 1. Model considered for the evolution of the gain specified in terms of the state space form in Eq. (12) along with any parameter constraints.

Model	F_{11}	F_{12}	F_{22}	G_{11}	G_{22}	Constraints	Unknwon Parameters θ
RW	1	0	0	1	0	$q_\xi = 0$	$\left(\sigma^2, q_\eta\right)$
LLT	1	1	1	1	1		$\left(\sigma^2, q_\eta, q_\xi\right)$
DLLT	1	1	1	1	1	$q_\eta = q_\xi$	$\left(\sigma^2, q_\eta\right)$
RWD	1	1	1	1	0	$q_\xi = 0$	$\left(\sigma^2, q_\eta\right)$
IRW	1	1	1	0	1	$q_\eta = 0$	$\left(\sigma^2, q_\xi\right)$
AR	α	0	0	1	0	$q_\xi = 0$	$\left(\alpha, \sigma^2, q_\eta\right)$
SLLT	α	1	β	1	1		$\left(\alpha, \beta, \sigma^2, q_\eta, q_\xi\right)$
SRW	α	1	1	0	1	$q_\eta, = 0$	$\left(\alpha, \sigma^2, q_\xi\right)$
DT	1	1	β	1	1	$q_\eta = q_\xi$	$\left(\beta, \sigma^2, q_\eta\right)$

nodes of a hydraulic model) may also be improved. This of course cannot be validated until observations are taken at these points.

The operational usefulness of the predictive distribution constructed from the above state space formulation is dependent upon:

- an appropriate description of the distributions of ε_t and ζ_t;

- an adequate solution of the filtering problem inherent in producing the forecasts.

Addressing both of these topics introduces a number of barriers to the operational implementation of this technique.

If either $\mathcal{M}$ or $\mathcal{H}$ is non-linear, the solution to the filtering problem is not trivial. Approximate solutions to the filtering problem can be provided by a number of algorithms such as particle filters (e.g. Doucet et al., 2001; Moradkhani et al., 2005a; Weerts and El Serafy, 2006), non-linear extensions to the Kalman Filter (Rajaram and Georgakakos, 1989; DaRos and Borga, 1997; Evensen, 2003; Moradkhani et al., 2005b; Reichle et al., 2008) or variational techniques (Li and Navon, 2001; Madsen and Skotner, 2005; Seo et al., 2003). Combinations of these techniques may also be used (e.g. Shamir et al., 2010).

Particle filters, which approximate the desired distributions through Monte-Carlo sampling, can be considered the most flexible, although the computational burden can be large (Smith et al., 2008a) and implementation difficult when ε_t dominates the observation noise (Liu and Chen, 1998). The remaining techniques require less computational resource but introduce assumptions such as unbounded distributions that may require careful reparameterization of the hydrological model if the states are to remain hydrologically interpretable, e.g. volumes of water in the river channel must be greater or equal to zero.

All the techniques outlined above make multiple calls to the process model at each time step. The computational cost of this may be prohibitive for applications in real time when the lead times required for decisions about warnings are a constraint. This is particularly true if the implementation of the filtering algorithm is achieved by providing code that "wraps" the hydrological model and interacts by altering the initial state and parameter files (Weerts et al., 2010).

Regardless of the computational technique utilised great care should be taken in constructing the description of ε_t and ζ_t (Beven et al., 2008; Kirchner, 2006), particularly if there may be systematic biases, including phase errors, in the data or model (Reichle, 2008). The validation of these choices may require the re-analysis of a significant number of historic events, itself time consuming.

Using DA to improve the forecasts of the difference between the hydrological model and observed data can often be performed at minimal computational cost. If a suitable historic record of model output is maintained, the computational cost of setting up the DA may also be minimal. A wide variety of stochastic models have been proposed. These range from the classical auto-regressive moving average (ARMA) time series models of Box and Jenkins (1994) used operationally in the UK (Moore, 2007) to more complex semi-parametric methods (e.g. Krzysztofowicz and Maranzano, 2004; Maranzano and Krzysztofowicz, 2004).

To provide reliable forecasts (in the pragmatic and probabilistic sense), these formulations and others (e.g. Montanari and Brath, 2004; Weerts et al., 2011; Seo et al., 2006) have to attempt to capture the potentially complex evolution of the model residuals. These residuals may incorporate a systematic or temporally varying bias. Reliance on temporal correlation within the residuals must be tempered by the fact that the correlation may be non-stationary, often being low at key times such as during the rising limbs of hydrographs
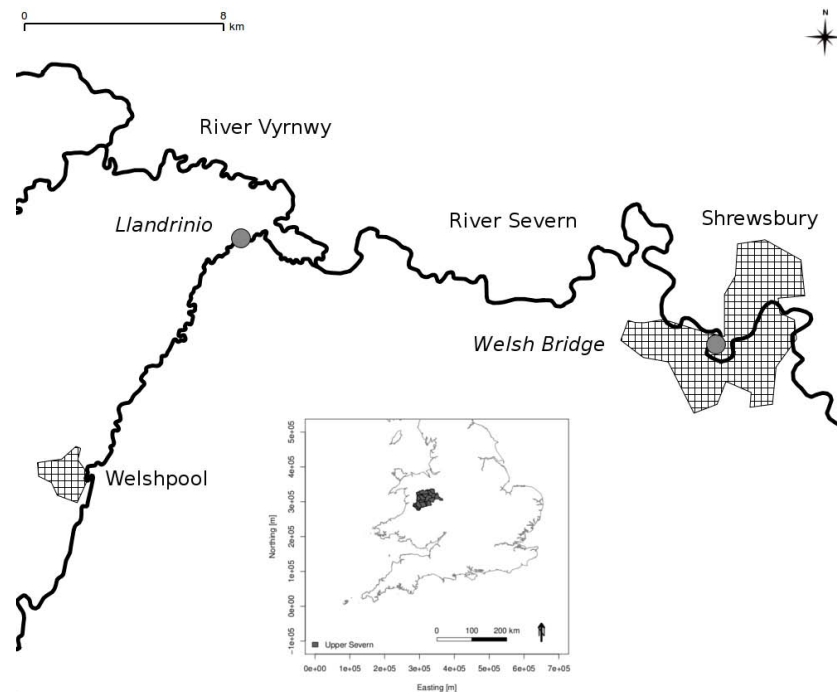
Fig. 1. Schematic map showing the upper River and Vyrnwy tributary which both flow from west to east. Gauged locations considered in the study (circles) and urban areas are shown (hatched). The insert shows the location of the catchment within the UK.

(Todini, 2008) and much higher during recession periods. Residuals in extreme situations, such as floods may also possess characteristics different to the majority of the data. Furthermore, each flood may reveal previously unknown shortcomings in the hydrological/hydraulic model(s) making their residuals difficult to predict. In such situations, it may be useful to utilise robust error models and predictive bounds (e.g. Rougier et al., 2009; Vernon et al., 2010).

This paper considers the use of DA to improve the forecasts of the difference between the output of a deterministic hydrological model and the observed data. It explores the use of a multiplicative gain to correct the deterministic forecasts. This gain is evolved stochastically. Forecasts of future observed values, that is future y_t, are expressed as probability distributions dependent upon both the gain and the deterministic forecast of the hydrological model. The approach presented has been utilised previously for operational flood forecasting (Lees et al., 1994). This paper extends previous work by considering a broader family of models for the evolution of the gain and two contrasting parameter estimation techniques.

Section 2 outlines the error model for providing probabilistic forecasts at a single observational site. In doing so a family of parsimonious models for the evolution of the gain are introduced. The use of the linear Kalman filter for DA and generation of predictive distributions is presented. Methods for estimating the parameters of the model are discussed in Sect. 3. Section 4 presents an example application using an operational flood forecasting model from the UK. The validity of the assumptions used in parameter estimation is investigated and the usefulness of the uncertainty representations illustrated.

2 Methodology

This section presents the stochastic error model utilised within this paper and the computation of the predictive distribution. The representation of the stochastic error model and its evolution is outlined in a state space framework giving a natural framework for computing the predictive distribution as a filtering problem.

2.1 Error model

Recall that $\mathbf{y}_{1:T} = (y_1, \ldots, y_T)$ is a vector of T observations indexed by time with corresponding deterministic hydrological/hydraulic model predictions $\mathbf{m}_{1:T}$. The observation y_t is then related to the prediction m_t by an adaptive gain g_t and noise term ϵ_t as outlined in Eq. (3).

$$y_t = m_t g_t + \epsilon_t \tag{3}$$

The gain g_t is a time varying correction for the bias in the model forecast. It is evolved stochastically according to the local level (Harvey, 1989) or generalised random walk (Jakeman and Young, 1984; Young et al., 1989) family of models outlined in this section. An alternative for the evolution of g_t
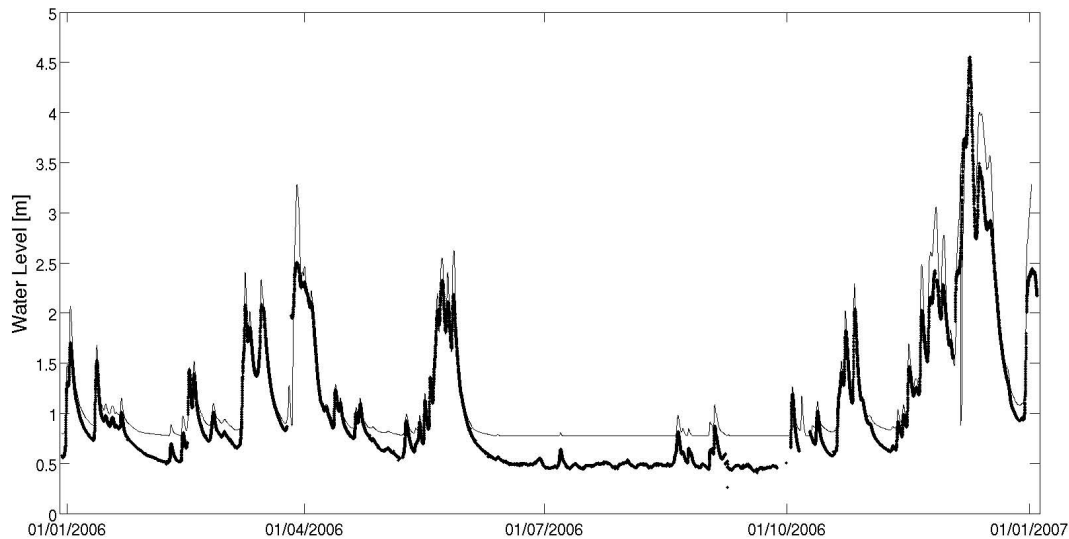
Fig. 2. Summary plots of the data available for Welsh bridge during the calibration period. Points represent observed data with the line being the concatenation of the output of the most recent hydraulic model initialisation. A bias in the prediction of low flows is clearly shown as are periods of missing data and poor model initialisation.

such as autoregressive moving average processes could also be considered.

The simplest local level model considered is a random walk where g_t is given as the sum of its previous value and the stochastic noise η_t. That is

$$g_t = g_{t-1} + \eta_t. \tag{4}$$

This is referred to as the random walk (RW) model. The local linear trend (LLT) model generalises this by introducing the slope d_t which follows a random walk driven by the stochastic noise ξ_t. Thus,

$$g_t = g_{t-1} + d_{t-1} + \eta_t \tag{5}$$
$$d_t = d_{t-1} + \xi_t. \tag{6}$$

In this paper it is assumed that ϵ_t, along with the stochastic noise terms η_t and ξ_t, are not correlated with each other or in time. Further, they are realisations of unimodal, symmetric, unbounded random variables that can be summarised by their first two moments which are defined using the parameters σ^2, q_η and q_ξ as

$$E[\epsilon_t] = E[\eta_t] = E[\xi_t] = 0$$
$$\text{Var}[\epsilon_t] = \sigma^2$$
$$\text{Var}[\eta_t] = q_\eta \sigma^2$$
$$\text{Var}[\xi_t] = q_\xi \sigma^2.$$

The validity of these assumptions can be assessed from the forecast residuals as shown in Sect. 4. Two methods for estimation of the parameters are presented in Sect. 3.

Three further local level models can be specified by placing constraints on the LLT model. If $q_\eta = q_\xi$, the trend in the

LLT model is deterministic, resulting in the deterministic local linear trend (DLLT) model. When q_ξ is zero, the slope is fixed and the evolution of the gain becomes a random walk with drift (RWD) model, that is

$$g_t = g_{t-1} + d + \eta_t.$$

Setting q_η to zero but allowing positive q_ξ results in an integrated random walk trend, referred to as the IRW model. This often results in a smoother adaptation of g_t compared to the RW model outlined in Eq. (4).

The models outlined above for g_t are parsimonious, the only unknown parameters other than σ^2 being the values of noise variance ratios q_η and q_ξ. A further level of complexity can be included by incorporating smoothing or damping parameters. Inclusion of such a parameter α in Eq. (4) results in a first order auto regressive (AR) model for the gain:

$$g_t = \alpha g_{t-1} + \eta_t. \tag{7}$$

Inclusion of the smoothing parameters (α and β) in the local linear trend model gives

$$g_t = \alpha g_{t-1} + d_{t-1} + \eta_t \tag{8}$$
$$d_t = \beta d_{t-1} + \xi_t. \tag{9}$$

This is referred to as the smoothed local linear trend (SLLT) model. Two special cases of this are the smoothed random walk (SRW) model; in which $\beta = 1$ and $q_\eta = 0$; and the damped trend (DT) model in which $q_\eta = q_\xi$ and $\alpha = 1$. More general and higher order representations are also possible such as doubly integrated random walks. Exploration of these is beyond the scope of this paper.

Table 2. Calibration results for Welsh bridge showing the log likelihood and RMSE (bracketed) for various forecast lead times (hours) and GRW models.

	2	6	9	12	18	24
RW	21 102.42 (0.06)	12 875.26 (0.15)	10 115.31 (0.21)	8958.92 (0.26)	8431.67 (0.31)	7986.98 (0.36)
AR	21 120.65 (0.06)	12 926.55 (0.15)	10 187.68 (0.21)	9035.47 (0.25)	8480.80 (0.30)	8024.50 (0.35)
LLT	21 195.19 (0.07)	12 033.98 (0.18)	8619.67 (0.26)	6278.50 (0.33)	3464.42 (0.47)	953.29 (0.63)
DLLT	21 100.06 (0.08)	10 418.60 (0.19)	6846.22 (0.27)	5215.57 (0.33)	716.47 (0.47)	−4174.00 (0.63)
RWD	20 939.18 (0.07)	12 033.98 (0.18)	8619.67 (0.26)	6278.50 (0.33)	3440.96 (0.47)	837.32 (0.63)
IRW	21 097.79 (0.08)	10 417.75 (0.19)	6846.24 (0.27)	5215.84 (0.33)	715.88 (0.47)	−4178.99 (0.63)
SRW	22 058.78 (0.06)	13 342.41 (0.15)	10 501.56 (0.21)	9430.35 (0.25)	8730.26 (0.29)	8190.68 (0.34)
DT	22 052.24 (0.06)	13 341.53 (0.15)	10 501.83 (0.21)	9430.56 (0.25)	8730.49 (0.29)	8190.91 (0.34)
SLLT	22 076.80 (0.06)	13 394.44 (0.14)	10 564.56 (0.20)	9491.22 (0.25)	8785.71 (0.29)	8251.69 (0.33)

2.2 State Space form

All the models outlined can be conveniently expressed in a state space form with state vector $x_t = \begin{bmatrix} g_t & d_t \end{bmatrix}'$ describing the gain (g_t) and its slope (d_t). The state vector evolves according the state transition matrix $\mathbf{F}$ and system noise matrix $\mathbf{G}$ as

$$x_t = \mathbf{F}x_{t-1} + \mathbf{G} \begin{bmatrix} \eta_t \\ \xi_t \end{bmatrix}. \tag{10}$$

The state vector is related to the observations by

$$y_t = h_t' x_t + \epsilon_t \tag{11}$$

where $h_t = \begin{bmatrix} m_t & 0 \end{bmatrix}'$. The values taken by $\mathbf{F}$ and $\mathbf{G}$ depend upon the local level model selected. Table 1 outlines the values taken in terms of the matrix forms given in Eq. (12) for the various models considered along with any other parameter constraints.

$$\mathbf{F} = \begin{bmatrix} F_{11} & F_{12} \\ 0 & F_{22} \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} G_{11} & 0 \\ 0 & G_{22} \end{bmatrix}. \tag{12}$$

Two methods for the estimation of the parameters are presented in Sect. 3. The following sub-section discusses the use of the Kalman filter to generate the expected value and covariance of the predictive distributions.

2.3 Prediction using the Kalman filter

The assumptions regarding the stochastic noise terms presented in Sect. 2.1 are the minimum required for application of the linear Kalman filter (Kalman, 1960). Suppose the distribution of x_t has similar properties to that of the stochastic noise terms with its expected value and variance given by $\hat{x}_t$ and $\sigma^2 \mathbf{P}_t$. (where $\mathbf{P}_t$ is a 2×2 matrix), respectively. The Kalman filter can be used to predict future states and assimilate the observed data as it becomes available.

The one step ahead predictions of the distribution of the states, conditional upon the data up to time t, are given by the expected value:

$$\hat{x}_{t+1|t} = \mathbf{F}\hat{x}_{t|t} \tag{13}$$

and variance $\sigma^2 \mathbf{P}_{t+1|t}$ where

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{G}\mathbf{Q}\mathbf{G}'. \tag{14}$$

The noise variance ratio matrix $\mathbf{Q}$ is constructed as

$$\mathbf{Q} = \begin{bmatrix} q_\eta & 0 \\ 0 & q_\xi \end{bmatrix}.$$

The f-step ahead prediction of the states given the information up to time t can be computed by repeated application of Eqs. (13) and (14).

The f-step ahead prediction error $v_{t+f|t}$ and prediction variance $\sigma^2 \psi_{t+f|t}$ can be computed from the forecast states using:

$$v_{t+f|t} = y_{t+f} - h_{t+f}' \hat{x}_{t+f|t} \tag{15}$$

$$\psi_{t+f|t} = 1 + h_{t+f}' \mathbf{P}_{t+f|t} h_{t+f} \tag{16}$$

Evaluation of these expressions requires knowledge of the future predictions of the flood forecasting model.

When a new observation becomes available it can be used to condition the distribution of the gain by updating the mean and covariance of the one step ahead prediction of the state distribution using Eqs. (17) to (19) (see for example Kalman, 1960, for a derivation).

$$k_{t+1} = \mathbf{P}_{t+1|t} h_{t+1} \psi_{t+1|t}^{-1} \tag{17}$$

$$\hat{x}_{t+1|t+1} = \hat{x}_{t+1|t} + k_{t+1} v_{t+1|t} \tag{18}$$

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|t} - k_{t+1} h_{t+1}' \mathbf{P}_{t+1|t} \tag{19}$$

To evaluate the above recursions some initial values for $\hat{x}_{0|0}$ and $\mathbf{P}_{0|0}$ are required. In this paper a representation based on taking a diffuse initial condition is used. The techniques presented in Durbin and Koopman (2001, Chapter 5) allow the conditional distributions of the states after two time steps; summarised by $\hat{x}_{2|2}$ and $\mathbf{P}_{2|2}\sigma^2$; to be computed analytically from the first two observation, the corresponding model predictions and θ. A suitable burn-in period denoted t_0 (see Sect. 4) is then used before commencing evaluation of the estimation criteria outlined in the following section.

Table 3. Fraction of observations at Welsh bridge bracketed by the estimated 95 % prediction intervals during calibration. Bracketed results are those for the SEFE calibration with the italicised and bold values corresponding to the empirical and theoretical symmetric bounds. Results are shown for all the time periods, high levels (> 2m) and periods where the hydrograph is rising for different combinations of lead time (hours) and GRW model.

	Period	2	6	9	12	18	24
RW	All	0.98 (*0.95*,**0.98**)	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.88*,**0.96**)	0.93 (*0.86*,**0.94**)	0.92 (*0.84*,**0.92**)	0.91 (*0.81*,**0.92**)	0.89 (*0.79*,**0.90**)	0.87 (*0.78*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.88*,**0.96**)	0.95 (*0.88*,**0.95**)	0.95 (*0.89*,**0.95**)	0.94 (*0.89*,**0.95**)	0.94 (*0.90*,**0.95**)
AR	All	0.98 (*0.95*,**0.98**)	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.88*,**0.96**)	0.93 (*0.87*,**0.94**)	0.92 (*0.86*,**0.92**)	0.91 (*0.84*,**0.92**)	0.89 (*0.82*,**0.90**)	0.87 (*0.80*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.88*,**0.96**)	0.95 (*0.89*,**0.95**)	0.95 (*0.90*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.90*,**0.95**)
LLT	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.96**)
	High	0.94 (*0.88*,**0.95**)	0.90 (*0.84*,**0.92**)	0.89 (*0.82*,**0.90**)	0.88 (*0.81*,**0.89**)	0.87 (*0.79*,**0.88**)	0.86 (*0.78*,**0.87**)
	Rising	0.96 (*0.88*,**0.97**)	0.94 (*0.88*,**0.95**)	0.94 (*0.89*,**0.94**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)
DLLT	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)
	High	0.94 (*0.89*,**0.95**)	0.92 (*0.83*,**0.92**)	0.90 (*0.81*,**0.90**)	0.89 (*0.80*,**0.89**)	0.88 (*0.80*,**0.88**)	0.87 (*0.78*,**0.87**)
	Rising	0.96 (*0.88*,**0.97**)	0.95 (*0.88*,**0.95**)	0.94 (*0.90*,**0.94**)	0.94 (*0.90*,**0.95**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)
RWD	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)
	High	0.93 (*0.87*,**0.95**)	0.90 (*0.84*,**0.92**)	0.89 (*0.82*,**0.90**)	0.88 (*0.81*,**0.89**)	0.87 (*0.79*,**0.88**)	0.86 (*0.78*,**0.87**)
	Rising	0.96 (*0.88*,**0.97**)	0.94 (*0.88*,**0.95**)	0.94 (*0.89*,**0.94**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)
IRW	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)
	High	0.94 (*0.89*,**0.95**)	0.92 (*0.83*,**0.92**)	0.90 (*0.81*,**0.90**)	0.89 (*0.80*,**0.89**)	0.88 (*0.80*,**0.88**)	0.87 (*0.78*,**0.87**)
	Rising	0.96 (*0.88*,**0.97**)	0.95 (*0.88*,**0.95**)	0.94 (*0.90*,**0.94**)	0.94 (*0.90*,**0.95**)	0.94 (*0.91*,**0.94**)	0.94 (*0.91*,**0.94**)
SRW	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.87*,**0.96**)	0.92 (*0.86*,**0.94**)	0.91 (*0.85*,**0.92**)	0.90 (*0.83*,**0.91**)	0.89 (*0.81*,**0.90**)	0.88 (*0.79*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.87*,**0.96**)	0.95 (*0.88*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.90*,**0.95**)	0.94 (*0.91*,**0.95**)
DT	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.87*,**0.96**)	0.92 (*0.86*,**0.94**)	0.91 (*0.85*,**0.92**)	0.90 (*0.83*,**0.91**)	0.89 (*0.81*,**0.90**)	0.88 (*0.79*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.87*,**0.96**)	0.95 (*0.88*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.90*,**0.95**)	0.94 (*0.91*,**0.95**)
SLLT	All	0.98 (*0.95*,**0.98**)	0.97 (*0.95*,**0.98**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)
	High	0.95 (*0.88*,**0.96**)	0.92 (*0.87*,**0.94**)	0.91 (*0.86*,**0.92**)	0.90 (*0.83*,**0.91**)	0.90 (*0.82*,**0.90**)	0.88 (*0.81*,**0.89**)
	Rising	0.97 (*0.88*,**0.97**)	0.95 (*0.88*,**0.96**)	0.95 (*0.88*,**0.95**)	0.95 (*0.90*,**0.95**)	0.95 (*0.91*,**0.95**)	0.95 (*0.91*,**0.95**)

3 Estimation

This section discusses the estimation of the unknown parameter vector θ defined for the models considered in Table 1. Two estimation techniques are outlined and there results contrasted in Sect. 4. The first technique is maximum likelihood estimation based upon the assumption that the prediction errors are independent realisations of Gaussian random variables. This introduces stronger assumptions about the stochastic noise terms than those introduced in Sect. 2.1. The second method which is based on minimising the sum of the squared expected forecast errors, is more heuristic. In both cases the validity of the error assumptions can be assessed. This is discussed along with the the construction of predictive error bounds.

3.1 Gaussian maximum likelihood

In Gaussian maximum likelihood (GML) estimation, the parameters θ are estimated by maximising the likelihood of the f-step ahead predictions when it is believed that $v_{t+f|t}$ is drawn independently from a zero mean Gaussian distribution with variance $\sigma^2 \psi_{t+f|t}$. Under these assumptions the log likelihood is

$$l(\theta) = K - \frac{1}{2} \sum_{t=t_0}^{T-f} \log\left(\sigma^2 \psi_{t+f|t}\right)$$
$$- \frac{1}{2\sigma^2} \sum_{t=t_0}^{T-f} v_{t+f|t}^2 \psi_{t+f|t}^{-1} \tag{20}$$

where K is a constant with respect to θ. The maximum likelihood estimate of σ^2 can be computed conditional upon the other parameters in θ as

$$\hat{\sigma}^2 = \frac{1}{T-f-t_0} \sum_{t=t_0}^{T-n} v_{t+f|t}^2 \psi_{t+f|t}^{-1}. \tag{21}$$

This allows σ^2 to be concentrated out of Eq. (20) leaving (Schweppe, 1965):

$$l\left(\theta \backslash \sigma^2\right) = K - \frac{1}{2} \sum_{t=t_0}^{T-f} \log\left(\hat{\sigma}^2 \psi_{t+f|t}\right) \tag{22}$$

Table 4. Precision of the forecasts (see text for definition) during calibration at Welsh bridge. Bracketed results are those for the SEFE calibration with the italicised and bold values corresponding to the empirical and theoretical symmetric bounds. Results are shown for all the time periods, high levels (> 2 m) and periods where the hydrograph is rising for different combinations of lead time (hours) and GRW model.

	Period	2	6	9	12	18	24
RW	All	0.18 (*0.04*,**0.29**)	0.33 (*0.10*,**0.54**)	0.41 (*0.15*,**0.66**)	0.45 (*0.19*,**0.69**)	0.48 (*0.23*,**0.73**)	0.50 (*0.26*,**0.76**)
	High	0.26 (*0.07*,**0.49**)	0.51 (*0.18*,**0.91**)	0.63 (*0.25*,**1.14**)	0.68 (*0.29*,**1.07**)	0.77 (*0.37*,**1.15**)	0.83 (*0.42*,**1.24**)
	Rising	0.19 (*0.04*,**0.31**)	0.35 (*0.11*,**0.57**)	0.43 (*0.16*,**0.71**)	0.47 (*0.20*,**0.73**)	0.51 (*0.25*,**0.77**)	0.54 (*0.27*,**0.81**)
AR	All	0.18 (*0.04*,**0.29**)	0.33 (*0.11*,**0.54**)	0.41 (*0.16*,**0.67**)	0.44 (*0.21*,**0.69**)	0.47 (*0.26*,**0.74**)	0.50 (*0.31*,**0.79**)
	High	0.26 (*0.07*,**0.49**)	0.50 (*0.19*,**0.91**)	0.61 (*0.27*,**1.14**)	0.67 (*0.32*,**1.07**)	0.76 (*0.41*,**1.17**)	0.83 (*0.49*,**1.27**)
	Rising	0.19 (*0.04*,**0.31**)	0.35 (*0.12*,**0.57**)	0.43 (*0.17*,**0.71**)	0.46 (*0.22*,**0.73**)	0.50 (*0.28*,**0.78**)	0.53 (*0.32*,**0.84**)
LLT	All	0.19 (*0.04*,**0.32**)	0.42 (*0.13*,**0.62**)	0.59 (*0.26*,**0.88**)	0.75 (*0.42*,**1.18**)	1.08 (*0.76*,**1.87**)	1.39 (*1.21*,**2.65**)
	High	0.24 (*0.06*,**0.54**)	0.63 (*0.22*,**1.05**)	0.87 (*0.40*,**1.35**)	1.11 (*0.61*,**1.70**)	1.67 (*1.14*,**2.79**)	2.41 (*2.05*,**4.50**)
	Rising	0.20 (*0.04*,**0.34**)	0.50 (*0.14*,**0.69**)	0.70 (*0.30*,**1.04**)	0.90 (*0.51*,**1.42**)	1.41 (*1.02*,**2.50**)	1.71 (*1.51*,**3.32**)
DLLT	All	0.19 (*0.03*,**0.31**)	0.46 (*0.17*,**0.68**)	0.63 (*0.29*,**0.96**)	0.77 (*0.44*,**1.24**)	1.07 (*0.80*,**1.97**)	1.42 (*1.20*,**2.69**)
	High	0.22 (*0.04*,**0.39**)	0.67 (*0.22*,**0.90**)	0.95 (*0.40*,**1.31**)	1.15 (*0.61*,**1.71**)	1.73 (*1.17*,**2.87**)	2.46 (*2.02*,**4.53**)
	Rising	0.20 (*0.04*,**0.32**)	0.53 (*0.20*,**0.82**)	0.74 (*0.36*,**1.17**)	0.91 (*0.54*,**1.51**)	1.31 (*1.08*,**2.66**)	1.63 (*1.51*,**3.38**)
RWD	All	0.19 (*0.04*,**0.31**)	0.42 (*0.13*,**0.62**)	0.59 (*0.26*,**0.88**)	0.75 (*0.42*,**1.18**)	1.08 (*0.76*,**1.87**)	1.41 (*1.21*,**2.65**)
	High	0.26 (*0.07*,**0.53**)	0.63 (*0.22*,**1.05**)	0.87 (*0.40*,**1.35**)	1.11 (*0.61*,**1.70**)	1.68 (*1.14*,**2.79**)	2.45 (*2.05*,**4.50**)
	Rising	0.20 (*0.04*,**0.33**)	0.50 (*0.14*,**0.69**)	0.70 (*0.30*,**1.04**)	0.90 (*0.51*,**1.42**)	1.41 (*1.02*,**2.50**)	1.73 (*1.51*,**3.32**)
IRW	All	0.19 (*0.04*,**0.30**)	0.46 (*0.17*,**0.68**)	0.63 (*0.29*,**0.96**)	0.77 (*0.44*,**1.24**)	1.07 (*0.80*,**1.97**)	1.42 (*1.20*,**2.69**)
	High	0.22 (*0.04*,**0.39**)	0.67 (*0.22*,**0.90**)	0.95 (*0.40*,**1.31**)	1.15 (*0.61*,**1.71**)	1.73 (*1.17*,**2.87**)	2.46 (*2.02*,**4.53**)
	Rising	0.20 (*0.04*,**0.32**)	0.53 (*0.20*,**0.82**)	0.74 (*0.36*,**1.17**)	0.91 (*0.54*,**1.51**)	1.31 (*1.08*,**2.66**)	1.63 (*1.51*,**3.38**)
SRW	All	0.17 (*0.03*,**0.30**)	0.34 (*0.09*,**0.55**)	0.44 (*0.13*,**0.68**)	0.49 (*0.22*,**0.74**)	0.54 (*0.30*,**0.87**)	0.57 (*0.36*,**0.98**)
	High	0.21 (*0.05*,**0.48**)	0.47 (*0.15*,**0.93**)	0.59 (*0.23*,**1.16**)	0.67 (*0.31*,**1.03**)	0.81 (*0.43*,**1.27**)	0.93 (*0.58*,**1.56**)
	Rising	0.17 (*0.04*,**0.32**)	0.37 (*0.09*,**0.59**)	0.48 (*0.14*,**0.72**)	0.55 (*0.25*,**0.82**)	0.62 (*0.35*,**1.02**)	0.64 (*0.42*,**1.12**)
DT	All	0.17 (*0.03*,**0.30**)	0.34 (*0.09*,**0.55**)	0.44 (*0.13*,**0.68**)	0.49 (*0.22*,**0.74**)	0.54 (*0.30*,**0.87**)	0.57 (*0.36*,**0.98**)
	High	0.21 (*0.05*,**0.49**)	0.47 (*0.15*,**0.93**)	0.59 (*0.23*,**1.16**)	0.67 (*0.31*,**1.03**)	0.81 (*0.43*,**1.27**)	0.93 (*0.58*,**1.56**)
	Rising	0.17 (*0.04*,**0.32**)	0.37 (*0.09*,**0.59**)	0.48 (*0.14*,**0.72**)	0.55 (*0.25*,**0.82**)	0.62 (*0.35*,**1.02**)	0.64 (*0.42*,**1.12**)
SLLT	All	0.17 (*0.03*,**0.30**)	0.33 (*0.09*,**0.55**)	0.44 (*0.14*,**0.68**)	0.49 (*0.22*,**0.73**)	0.55 (*0.30*,**0.85**)	0.59 (*0.38*,**0.95**)
	High	0.21 (*0.05*,**0.50**)	0.46 (*0.16*,**0.93**)	0.60 (*0.25*,**1.17**)	0.67 (*0.30*,**1.02**)	0.83 (*0.44*,**1.24**)	0.97 (*0.61*,**1.51**)
	Rising	0.17 (*0.03*,**0.32**)	0.36 (*0.10*,**0.58**)	0.48 (*0.15*,**0.73**)	0.54 (*0.24*,**0.80**)	0.64 (*0.35*,**0.98**)	0.66 (*0.43*,**1.07**)

which is dependent upon the remaining parameters (denoted $\theta \backslash \sigma^2$). This can be numerically optimised to give maximum likelihood parameter estimates of θ.

The uncertainty in the predictions can be expressed as percentile confidence intervals for the predictions constructed as:

$$\boldsymbol{h}'_{t+f} \hat{\boldsymbol{x}}_{t+f|t} \pm \kappa_p \hat{\sigma} \psi^{\frac{1}{2}}_{t+f|t} \qquad (23)$$

were κ_p is constant dependent upon p and can be computed from a standard normal distribution; for example $\kappa_{95} \approx 1.96$.

The normality of the forecast residuals and their correlation can be readily assessed using, for example, quantile and auto correlation plots (Box and Jenkins, 1994).

3.2 Minimising the sum of the squared expected forecast errors

The second estimation technique, referred to as SEFE for the remainder of this paper, is based on the appeal of minimising the sum of the squared expected forecasting error:

$$S_f = \sum_{t=t_0}^{T-f} v^2_{t+f|t}. \qquad (24)$$

This ensures that the expected value of the forecast is as close as possible (in terms of average squared error) to the observed data. The minimisation of S_f allows the estimation of all the parameters in θ except σ^2. A value for σ^2 can then be estimated using Eq. (21) if required. The error assumptions of the Kalman filter (Sect. 2.1) imply that each predictive distribution is uni-modal, symmetric and unbounded. Testing the symmetry of the forecast residuals, for example using Wilcoxon sign rank test (Wilcoxon, 1945), can indicate if this assumption is valid. Two methods for construction of predictive confidence intervals are considered. They make use of the theoretical symmetry of the forecast distribution and result in symmetric prediction intervals.

The symmetry of the forecast distribution implies that prediction confidence intervals can be expressed as

$$\boldsymbol{h}'_{t+f} \hat{\boldsymbol{x}}_{t+f|t} \pm \rho_p \psi^{\frac{1}{2}}_{t+f|t}. \qquad (25)$$
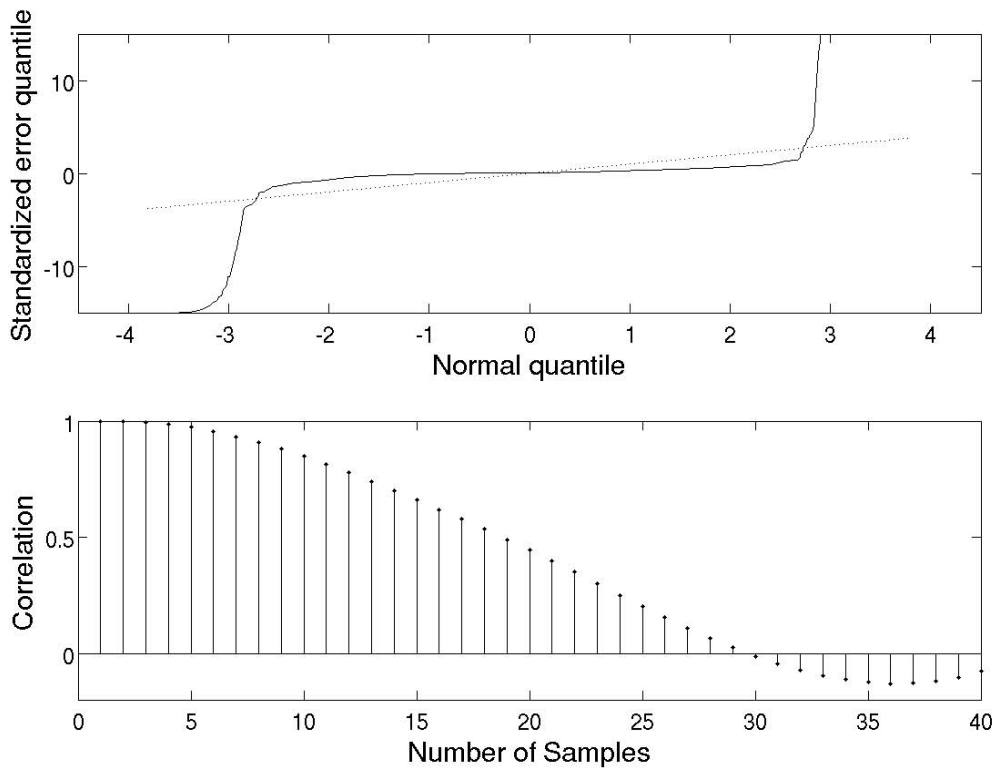
Fig. 3. Summary plots of the analysis of the 6 h forecast residuals at Welsh Bridge for the DT model calibrated using the Gaussian maximum likelihood methodology. The upper pane shows the quantile-quantile plot of the standardised residuals compared to a standard normal distribution. The lower pane shows the auto-correlation of the residuals.

The values of ρ_p can be estimated empirically as the pth percentile of $\left| v_{t+f|t} \, \psi_{t+f|t}^{-\frac{1}{2}} \right|$. Given the finite population of residuals, this empirical estimate of ρ_p may not be not robust at high values of p. The values of ρ_p can be adapted (for given p) as more data becomes available. Sequential tests for symmetry (e.g. Weed and Bradley, 1971) may be of use in such situations.

Pukelsheim (1994) gives theoretical results for the upper limits of ρ_p under the uni-model, symmetric and unbounded distributional assumptions. Specifically

$$
\Pr\left(\left| \frac{y_t - \mu_{t+f|t}}{\psi_{t+f|t}^{-\frac{1}{2}}} \right| \geq r \right) \leq \frac{4\sigma^2}{9r^2} \quad r > 1.63\sigma. \tag{26}
$$

The case $r = 3\sigma$ is the three sigma rule; that there is less than 5 % probability of a sample from univariate random variable random with the aforementioned properties being outside of 3 standard deviations from the mean. These upper limits can be used in two ways. Firstly, they allow for the estimation of conservative prediction confidence intervals, allowing for a more cautious view to be taken of the prediction uncertainty. The second use is as a means of analysing the suitability of the adaptive gain models considered by contrasting the sym-

metrical empirical estimates and the theoretic upper limits for a given p.

4   Upper Severn case study

To illustrate the effectiveness and limitations of the proposed methodology in an operational setting a case study based on the upper Severn catchment (UK) is presented. The upper Severn river network is situated on the border of England and Wales and shown in Fig. 1. The River Severn rises in the Cambrian mountains (741 mAOD) and flows to the northeast before meeting the Vyrnwy tributary at Crew Green. The valley is wide and flat in this confluence area, with a considerable extent of flood plain. The river then flows east to Shrewsbury. The lower boundary of the 2284 km² upper Severn catchment is defined by the gauge at Welsh bridge in Shrewsbury where the median annual flood is greater than 284 cumecs. Average annual rainfall can exceed 2500 mm in the head waters of the catchment. The catchment has seen seven significant flood events in the past twelve years.

The existing flood forecasting model consists of a number of simplified rainfall-runoff representations linked to a one dimensional ISIS hydrodynamic model (see http://www.halcrow.com/isis/default.asp). Results are shown for two sites seen in Fig. 1: Welsh bridge the lower boundary
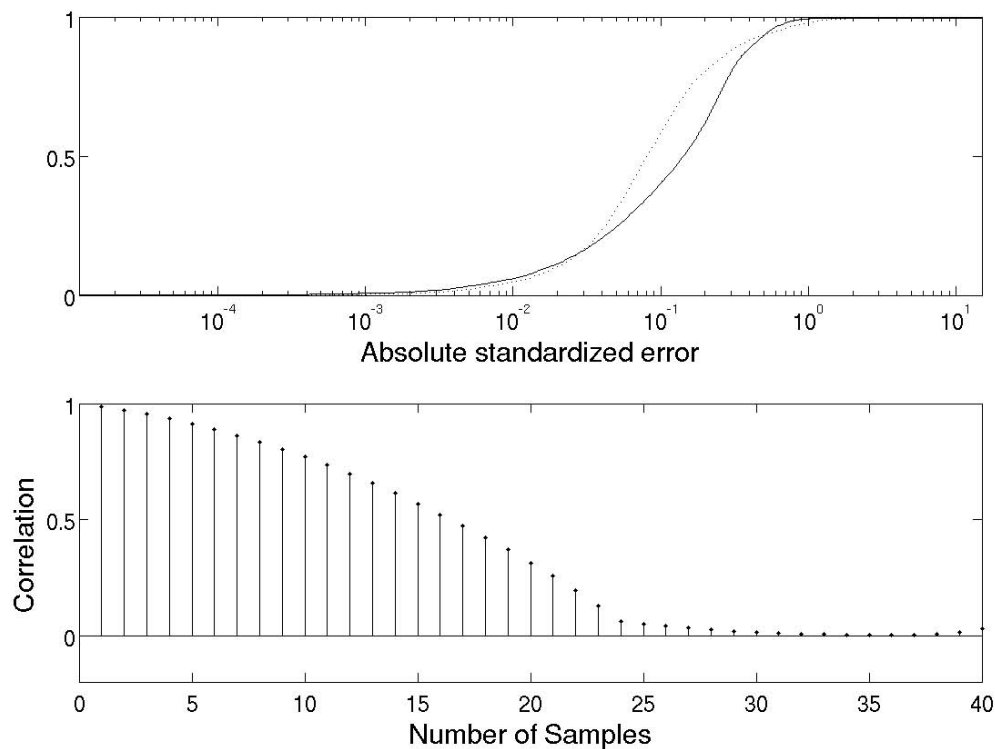
Fig. 4. Summary plots of the analysis of the 6 h forecast residuals for the RW model calibrated using the SEFE methodology. The upper pane shows the cumulative distributions of the absolute values of the residuals at Welsh Bridge. The solid line being positive residuals and the dotted line being negative residuals. The lower pane shows the auto-correlation of the residuals.

of the catchment and Llandrinio an internal node of the hydraulic model at which observations are available. The latter is just upstream of the junction with the major Vyrnwy tributary and may be affected by backwater effects at flood stages. At both sites water levels are observed every 15 min.

Operationally, the flood forecasting model is run with a 15 minute time step and evaluated twice daily. At midnight it is run in a continuous mode using observed inputs and output data assimilation to generate a set of "warm states" which are used to initialise the forecasts. The first set of forecasts issued at midnight (00:00) give up to 36 h lead time using forecast precipitation. The second set of forecasts is issued at midday (12:00). These forecasts are initialised by evolving the "warm states" using the observed meteorological variables between midnight and midday. Forecast precipitation is then used to evaluate the flood forecasting model giving forecasts of the hydrological variables with up to 36 h lead time. Further details can be found in Weerts et al. (2011).

The adaptive gain correction is initialised at the start of each forecast period. The first four hours of forecasts are used as a burn-in period for the adaptive gain. Then, in keeping with the operational system, forecasts are issued based on the most recent forecast run of the flood forecasting model for which the adaptive gain is burnt in. A single year of data (2006) which contains a significant flood event is used to identify and estimate the adaptive correction. Three years

(2007–2009) are used for validation purposes. The results for each of the sites are summarised below.

4.1 Welsh bridge

Figure 2 shows the hydraulic model predictions and observed data at Welsh bridge during the calibration period. There is a systematic over estimation of the water level during periods of low flow. Such a systematic bias may arise from the calibration of the hydraulic model. If the model was calibrated to discharge data and the representation of the gauged cross section poor at shallow depths, such a bias may result. Alternatively an artificially high low water level within the model, such as that seen here, can arise as a means of achieving acceptable representations of high water level periods.

The adaptive gain correction could be utilised to correct the forecasts in these low flow periods. The difference in relationship between the hydraulic model and observed data for the low flow periods and when this model is responding to rainfall suggests that a different calibration of the stochastic model for the adaptive gain may be required for each. For flooding purposes the response to rainfall is more important. Therefore, the calibration and validation criteria are only evaluated at Welsh bridge for the simulations of the hydraulic model which forecast a water level greater than 1 m.
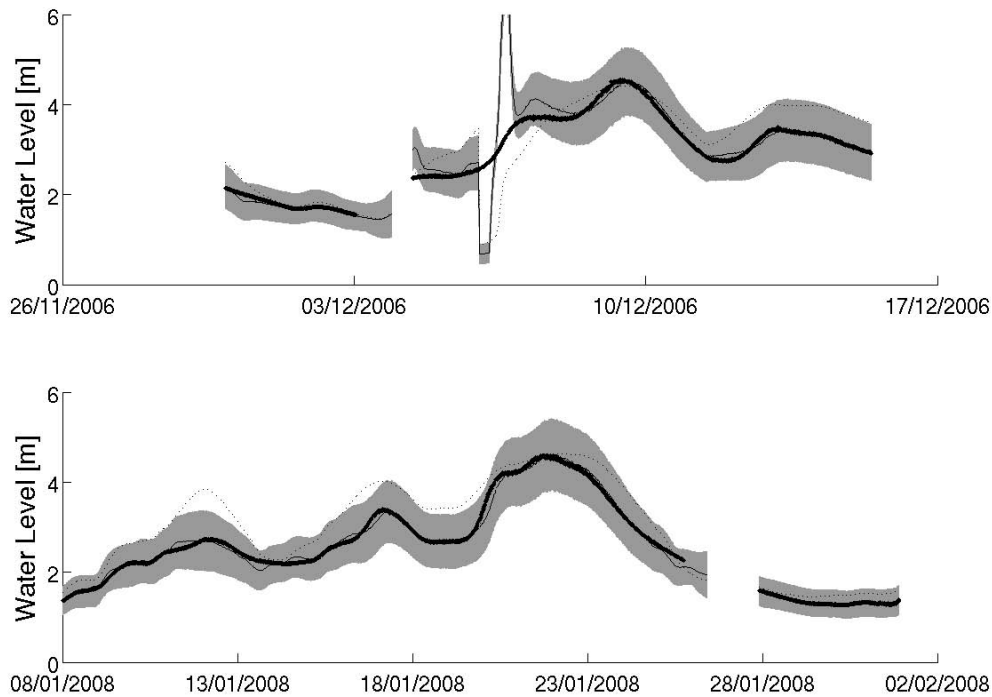
Fig. 5. Examples of 6 h ahead forecasts given at Welsh bridge for two large flood events during calibration (upper pane) and validation (lower pane) periods generated using the SLLT model calibrated using the Guassian methodology. The shaded area represents the 95 % prediction confidence interval with the solid line being the expected value of the predictions. Observed data points are also shown along with the deterministic model forecast (dotted line).
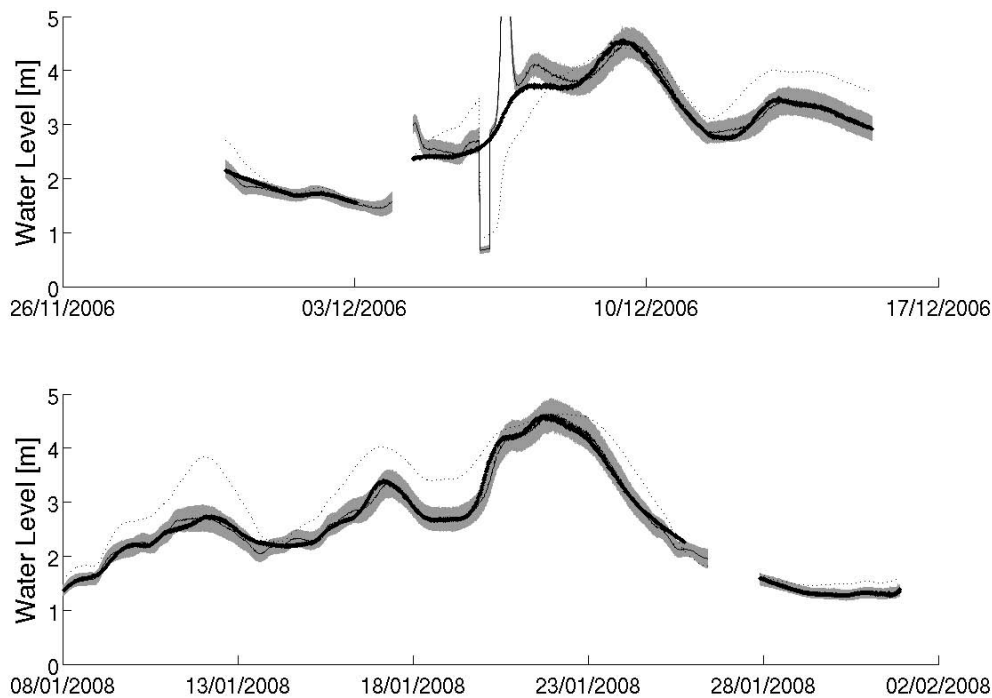

Fig. 6. Examples of 6 h ahead forecasts given at Welsh bridge for two large flood events during calibration (upper pane) and validation (lower pane) periods generated using the RW model calibrated using the SEFE methodology. The shaded area represents the 95 % prediction confidence interval with the solid line being the expected value of the predictions. Observed data points are also shown along with the deterministic model forecast (dotted line).

Table 5. Fraction of observations at Welsh bridge bracketed by the estimated 95 % prediction intervals during validation. Bracketed results are those for the SEFE calibration with the italicised and bold values corresponding to the empirical and theoretical symmetric bounds. Results are shown for all the time periods, high levels (> 2 m) and periods where the hydrograph is rising for different combinations of lead time (hours) and GRW model.

	Period	2	6	9	12	18	24
RW	All	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.96*,**0.99**)	0.99 (*0.96*,**0.99**)	0.99 (*0.95*,**0.99**)
	High	1.00 (*0.93*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.90*,**1.00**)	0.99 (*0.90*,**1.00**)	1.00 (*0.90*,**1.00**)
	Rising	1.00 (*0.93*,**1.00**)	1.00 (*0.93*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.95*,**1.00**)	0.99 (*0.94*,**1.00**)
AR	All	0.99 (*0.96*,**0.99**)	0.99 (*0.96*,**0.99**)	0.99 (*0.95*,**0.99**)	0.99 (*0.95*,**0.99**)	0.99 (*0.94*,**0.99**)	0.99 (*0.94*,**0.99**)
	High	1.00 (*0.93*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.91*,**1.00**)	0.99 (*0.89*,**1.00**)	0.99 (*0.90*,**1.00**)	0.99 (*0.91*,**1.00**)
	Rising	1.00 (*0.93*,**1.00**)	1.00 (*0.93*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.95*,**1.00**)	0.99 (*0.95*,**1.00**)
LLT	All	0.99 (*0.98*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.98*,**0.99**)
	High	0.99 (*0.96*,**1.00**)	0.99 (*0.94*,**1.00**)	0.99 (*0.93*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.94*,**1.00**)
	Rising	1.00 (*0.93*,**1.00**)	0.99 (*0.93*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.95*,**1.00**)	0.99 (*0.96*,**1.00**)	0.99 (*0.97*,**0.99**)
DLLT	All	0.99 (*0.98*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.98*,**0.99**)
	High	0.99 (*0.95*,**0.99**)	0.99 (*0.89*,**1.00**)	0.99 (*0.88*,**0.99**)	0.99 (*0.89*,**0.99**)	0.99 (*0.91*,**0.99**)	1.00 (*0.94*,**1.00**)
	Rising	1.00 (*0.93*,**1.00**)	1.00 (*0.92*,**1.00**)	1.00 (*0.93*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.95*,**1.00**)	1.00 (*0.97*,**0.99**)
RWD	All	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.98*,**0.99**)
	High	1.00 (*0.95*,**1.00**)	0.99 (*0.94*,**1.00**)	0.99 (*0.93*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.94*,**1.00**)
	Rising	1.00 (*0.93*,**1.00**)	0.99 (*0.93*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.95*,**1.00**)	0.99 (*0.96*,**1.00**)	0.99 (*0.97*,**0.99**)
IRW	All	0.99 (*0.98*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.98*,**0.99**)
	High	0.99 (*0.95*,**0.99**)	0.99 (*0.89*,**1.00**)	0.99 (*0.88*,**0.99**)	0.99 (*0.89*,**0.99**)	0.99 (*0.91*,**0.99**)	1.00 (*0.94*,**1.00**)
	Rising	1.00 (*0.93*,**1.00**)	1.00 (*0.92*,**1.00**)	1.00 (*0.93*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.95*,**1.00**)	1.00 (*0.97*,**0.99**)
SRW	All	0.99 (*0.97*,**0.99**)	0.99 (*0.96*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.96*,**0.99**)	0.99 (*0.95*,**0.99**)
	High	0.99 (*0.93*,**1.00**)	0.99 (*0.91*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.88*,**1.00**)	0.99 (*0.87*,**1.00**)	0.99 (*0.88*,**1.00**)
	Rising	1.00 (*0.94*,**1.00**)	1.00 (*0.92*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.94*,**1.00**)	0.99 (*0.94*,**1.00**)
DT	All	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.97*,**0.99**)	0.99 (*0.96*,**0.99**)	0.99 (*0.95*,**0.99**)
	High	0.99 (*0.93*,**1.00**)	0.99 (*0.91*,**1.00**)	0.99 (*0.92*,**1.00**)	0.99 (*0.88*,**1.00**)	0.99 (*0.87*,**1.00**)	0.99 (*0.88*,**1.00**)
	Rising	1.00 (*0.94*,**1.00**)	1.00 (*0.92*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.94*,**1.00**)	1.00 (*0.94*,**1.00**)	0.99 (*0.94*,**1.00**)
SLLT	All	0.99 (*0.97*,**0.99**)	0.99 (*0.96*,**0.99**)	0.99 (*0.95*,**0.99**)	0.99 (*0.94*,**0.99**)	0.99 (*0.94*,**0.99**)	0.99 (*0.94*,**0.99**)
	High	0.99 (*0.93*,**1.00**)	0.99 (*0.91*,**1.00**)	0.99 (*0.91*,**1.00**)	0.99 (*0.86*,**1.00**)	0.99 (*0.88*,**1.00**)	0.99 (*0.89*,**1.00**)
	Rising	1.00 (*0.92*,**1.00**)	1.00 (*0.92*,**1.00**)	1.00 (*0.93*,**1.00**)	1.00 (*0.93*,**1.00**)	0.99 (*0.94*,**1.00**)	0.99 (*0.95*,**1.00**)

Table 6. Calibration results for Llandrinio showing the log likelihood and RMSE (bracketed) for various forecast lead times (hours) and GRW models.

	2	6	9	12	18	24
RW	−523.75 (0.23)	−3879.11 (0.50)	−4576.16 (0.57)	−4868.47 (0.61)	−5135.70 (0.65)	−5446.65 (0.67)
AR	−466.31 (0.22)	−3783.83 (0.49)	−4486.50 (0.55)	−4773.20 (0.59)	−5020.97 (0.62)	−5307.53 (0.63)
LLT	−545.97 (0.27)	−7493.31 (0.82)	−10235.82 (1.26)	−12565.22 (1.72)	−16020.44 (2.62)	−18463.92 (3.53)
DLLT	−551.17 (0.30)	−12282.72 (0.83)	−10196.52 (1.26)	−12429.47 (1.72)	−15729.97 (2.63)	−18467.72 (3.53)
RWD	−1707.44 (0.31)	−8560.98 (0.97)	−11395.47 (1.47)	−13713.16 (1.98)	−17016.15 (3.02)	−19875.52 (4.04)
IRW	−545.97 (0.31)	−12294.39 (0.82)	−10192.66 (1.26)	−12425.94 (1.72)	−15726.36 (2.62)	−18463.99 (3.53)
SRW	1710.53 (0.20)	−307.89 (0.34)	−1115.54 (0.38)	−1515.08 (0.40)	−2109.19 (0.43)	−2588.85 (0.45)
DT	1709.51 (0.20)	−308.10 (0.34)	−1115.58 (0.38)	−1515.08 (0.40)	−2109.19 (0.43)	−2588.85 (0.45)
SLLT	1710.53 (0.20)	−307.89 (0.34)	−1115.36 (0.38)	−1513.65 (0.40)	−2092.73 (0.42)	−2553.96 (0.44)

Table 2 presents the log likelihood and root mean squared error (RMSE) summaries of the calibration results at various forecast lead times. For both calibration methodologies, the performance of the LLT model along with its special cases DLLT, RWD and IRW decays much more rapidly with lead time than the alternative models. This suggests that as with the SLLT, SRW and DT models there is the need to recognise the decay in information about the future values of the gain or its gradient through the use of damping parameters. The RMSE of the deterministic model prediction (based on always using the most recent model run) is 0.37. At lead times greater than 12 hours, none of the models achieves a significant decrease in the RMSE, though they still provide an estimate of the forecast uncertainty.
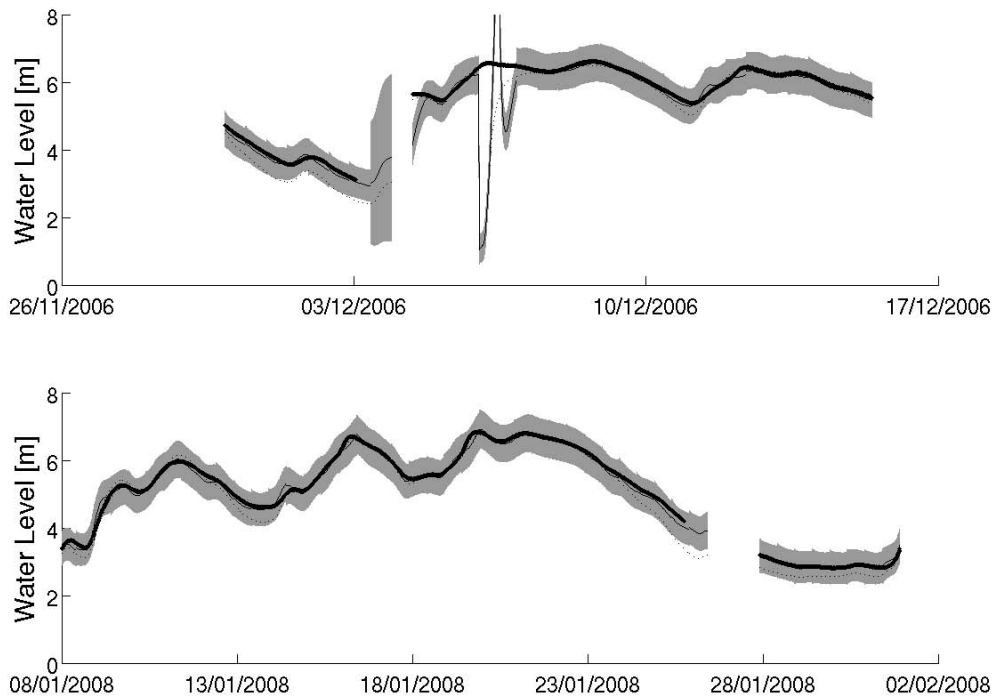
Fig. 7. Examples of 6 h ahead forecasts given at Llanrinio for two large flood events during calibration (upper pane) and validation (lower pane) periods generated using the SLLT model calibrated using the SEFE methodology with empirical error bounds. The shaded area represents the 95 % prediction confidence interval with the solid line being the expected value of the predictions. Observed data points are also shown along with the deterministic model forecast (dotted line).

Calibration with the GML methodology allows model selection to be performed using time series identification criteria, such as the AIC (Akaike, 1974) or BIC (Schwarz, 1978). This leads to the selection of the SLLT at all the lead times considered. The RMSE results of the SEFE calibration suggest that at shorter lead times, a more parsimonious model, such as the RW, may be suitable. Further analysis of the predictive performance of the models can be gained by examining the coverage and precision of the forecasts shown for the calibration period in Tables 3 and 4.

Table 3 shows the percentage of points falling within the 95 % prediction confidence interval for various lead times during calibration. Results are shown for three categories: the whole period, high water levels greater than 2 m and rising parts of the observed hydrograph where $y_{t+1} > y_t$. By definition the empirical bounds derived from the SEFE methodology bracket 95 % of all the (calibration) observations. Coverage of the high and rising flow categories for this methodology is less than 95 % for all models and lead times. Coverage of high flows decreases with increasing lead time for all the models. This indicates that the pooling of the residuals in computing the empirical bounds may not be appropriate. The values of κ_p, or indeed the model parameters, may be state or time dependent (Young, 2002; Young et al., 2001).

The theoretical prediction bounds of the SEFE methodology give in all cases percentages of coverage based on all the calibration data that is consistent with the error assumptions outlined in Sect. 2.1. A similar pattern to the empirical results for the coverage of the high and rising flows categories is shown.

The high percentages for coverage when using the Gaussian likelihood calibration at short forecast lead times support the suggestion that the assumptions used in deriving the likelihood may not be valid. However, at lead times up to six hours, there is less evidence of systematic differences in the percentage of observations covered in the three discharge categories. As with the SEFE calibration at longer lead times, the coverage of the high flow category deteriorates for all the models.

The results in Tables 4 summarise the precision of the of the forecasts through the mean of the forecasts standard deviation over each flow category. The empirically derived forecast error bounds are, especially at lead times up to six hours, significantly sharper than the theoretical upper limit and those resulting from the Gaussian calibration.

Figure 3 shows two summary plots of the standardised 6 h forecast residuals of the SLLT model calibrated using the Gaussian likelihood methodology. The upper pane shows a quantile-quantile plot of the standardised residual $v_{t+f|t} \hat{\sigma}^{-1} \psi_{t+f|t}^{-\frac{1}{2}}$ against a standard norm distribution. The heavier tails present in the distribution for the standardised residuals indicates that the Gaussian assumption may not be

Table 7. Fraction of observations at Llandrinio bracketed by the estimated 95 % prediction intervals during calibration. Bracketed results are those for the SEFE calibration with the italicised and bold values corresponding to the empirical and theoretical symmetric bounds. Results are shown for all the time period, high levels (> 4 m) and periods where the hydrograph is rising for different combinations of lead time (hours) and GRW model.

	Period	2	6	9	12	18	24
RW	All	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)
	High	0.96 (*0.94*,**0.96**)	0.94 (*0.90*,**0.94**)	0.93 (*0.89*,**0.94**)	0.93 (*0.89*,**0.93**)	0.92 (*0.89*,**0.92**)	0.91 (*0.88*,**0.91**)
	Rising	0.95 (*0.89*,**0.95**)	0.95 (*0.88*,**0.95**)	0.94 (*0.89*,**0.94**)	0.94 (*0.89*,**0.94**)	0.94 (*0.90*,**0.94**)	0.93 (*0.89*,**0.93**)
AR	All	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)
	High	0.96 (*0.95*,**0.96**)	0.94 (*0.92*,**0.94**)	0.93 (*0.91*,**0.94**)	0.93 (*0.90*,**0.93**)	0.92 (*0.91*,**0.92**)	0.91 (*0.90*,**0.91**)
	Rising	0.95 (*0.91*,**0.95**)	0.95 (*0.92*,**0.95**)	0.94 (*0.92*,**0.94**)	0.94 (*0.92*,**0.94**)	0.94 (*0.93*,**0.94**)	0.93 (*0.93*,**0.93**)
LLT	All	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)	0.95 (*0.95*,**0.95**)
	High	0.95 (*0.94*,**0.96**)	0.94 (*0.90*,**0.94**)	0.93 (*0.89*,**0.93**)	0.92 (*0.89*,**0.93**)	0.91 (*0.89*,**0.92**)	0.91 (*0.88*,**0.91**)
	Rising	0.95 (*0.91*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.91*,**0.94**)	0.94 (*0.92*,**0.94**)	0.94 (*0.93*,**0.94**)	0.93 (*0.91*,**0.93**)
DLLT	All	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)	0.95 (*0.95*,**0.95**)
	High	0.95 (*0.95*,**0.96**)	0.94 (*0.90*,**0.94**)	0.93 (*0.89*,**0.93**)	0.92 (*0.89*,**0.93**)	0.91 (*0.89*,**0.92**)	0.91 (*0.88*,**0.91**)
	Rising	0.95 (*0.91*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.91*,**0.94**)	0.94 (*0.92*,**0.94**)	0.94 (*0.93*,**0.94**)	0.93 (*0.91*,**0.93**)
RWD	All	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)	0.95 (*0.95*,**0.95**)
	High	0.95 (*0.93*,**0.95**)	0.93 (*0.90*,**0.94**)	0.93 (*0.89*,**0.93**)	0.92 (*0.89*,**0.93**)	0.91 (*0.89*,**0.92**)	0.90 (*0.88*,**0.91**)
	Rising	0.95 (*0.90*,**0.95**)	0.95 (*0.91*,**0.95**)	0.94 (*0.92*,**0.94**)	0.94 (*0.93*,**0.94**)	0.94 (*0.93*,**0.94**)	0.93 (*0.91*,**0.93**)
IRW	All	0.97 (*0.95*,**0.97**)	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)	0.95 (*0.95*,**0.95**)
	High	0.95 (*0.95*,**0.96**)	0.94 (*0.90*,**0.94**)	0.93 (*0.89*,**0.93**)	0.92 (*0.89*,**0.93**)	0.91 (*0.89*,**0.92**)	0.91 (*0.88*,**0.91**)
	Rising	0.95 (*0.91*,**0.95**)	0.95 (*0.90*,**0.95**)	0.94 (*0.91*,**0.94**)	0.94 (*0.92*,**0.94**)	0.94 (*0.93*,**0.94**)	0.93 (*0.91*,**0.93**)
SRW	All	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)	0.95 (*0.95*,**0.95**)
	High	0.95 (*0.95*,**0.96**)	0.92 (*0.88*,**0.94**)	0.89 (*0.87*,**0.93**)	0.88 (*0.86*,**0.92**)	0.87 (*0.86*,**0.91**)	0.87 (*0.86*,**0.91**)
	Rising	0.95 (*0.90*,**0.95**)	0.93 (*0.89*,**0.94**)	0.92 (*0.89*,**0.94**)	0.91 (*0.90*,**0.94**)	0.91 (*0.90*,**0.94**)	0.91 (*0.90*,**0.93**)
DT	All	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)	0.95 (*0.95*,**0.95**)
	High	0.95 (*0.95*,**0.96**)	0.92 (*0.88*,**0.94**)	0.89 (*0.87*,**0.93**)	0.88 (*0.86*,**0.92**)	0.87 (*0.86*,**0.91**)	0.87 (*0.86*,**0.91**)
	Rising	0.95 (*0.90*,**0.95**)	0.93 (*0.89*,**0.94**)	0.92 (*0.89*,**0.94**)	0.91 (*0.90*,**0.94**)	0.91 (*0.90*,**0.94**)	0.91 (*0.90*,**0.93**)
SLLT	All	0.97 (*0.95*,**0.97**)	0.96 (*0.95*,**0.97**)	0.96 (*0.95*,**0.96**)	0.96 (*0.95*,**0.96**)	0.95 (*0.95*,**0.95**)	0.95 (*0.95*,**0.95**)
	High	0.95 (*0.95*,**0.96**)	0.92 (*0.88*,**0.94**)	0.89 (*0.87*,**0.93**)	0.88 (*0.86*,**0.93**)	0.87 (*0.87*,**0.92**)	0.87 (*0.87*,**0.91**)
	Rising	0.95 (*0.90*,**0.95**)	0.93 (*0.89*,**0.94**)	0.92 (*0.90*,**0.94**)	0.91 (*0.90*,**0.94**)	0.91 (*0.91*,**0.94**)	0.91 (*0.91*,**0.93**)

valid. The residual distribution is approximately symmetric. Therefore, the use of the Kalman filter may be a valid approximation, even if the resulting prediction confidence intervals are not appropriately defined. The lower plot indicates that the residuals are correlated up to at least a lag of 28 samples, or 7 h, approximately the forecast lead time. This and the low correlation value suggest that the conditional independence assumption, while not strictly valid, may be a reasonable approximation.

The forecast residuals of the SEFE methodology can also be assessed. The lower pane of Fig. 4 shows the autocorrelation pattern in the residuals of the RW model fitted by the SEFE methodology for a 6 h leadtime. The correlation within the residuals lower and less persistent at longer lags than for the Gaussian calibration. The upper pane of Fig. 4 shows the cumulative distributions of the positive standardised residuals along with that of the absolute value of negative standardised residuals. It indicates a lack of symmetry in the residuals confirmed by a Wilcoxon sign rank test.

Table 5 shows the percentage of observations falling within the 95 % prediction confidence interval during the val-

idation period. The high percentages for the GML methodology further support the suggestion that the assumptions used in deriving the likelihood may not be valid. The results for the SEFE calibration indicate that the symmetrical empirical estimation of ρ_p derived during calibration perform well in the validation period. The predictive bounds derived using the theoretical upper limit appear to bracket around 99 % of the data during the validation period, suggesting they are unduly conservative.

Figure 5 shows the predictions made using the SLLT model calibrated using the Gaussian methodology for the largest flood events at Welsh bridge in the calibration and validation periods. The results are encouraging with the adaptive gain acting to correct the model forecast towards the observed values. During both the calibration and validation period adaptive gain methodology is able to correct for errors in both the timing and magnitude of the hydrograph peaks as well as the receding limb.

The calibration event (upper pane) highlights that the methodology is not able to correct for highly erroneous model forecasts. The validation results (lower pane) also

Table 8. Precision of the forecasts (see text for definition) during calibration at Llandrinio. Bracketed results are those for the SEFE calibration with the italicised and bold values corresponding to the empirical and theoretical symmetric bounds. Results are shown for all the time period, high levels (> 4 m) and periods where the hydrograph is rising for different combinations of lead time (hours) and GRW model.

	Period	2	6	9	12	18	24
RW	All	0.74 (*0.13*,**1.24**)	1.03 (*0.39*,**1.56**)	1.09 (*0.51*,**1.66**)	1.11 (*0.61*,**1.68**)	1.12 (*0.70*,**1.70**)	1.13 (*0.71*,**1.72**)
	High	1.00 (*0.27*,**2.53**)	1.51 (*0.59*,**2.38**)	1.68 (*0.81*,**2.62**)	1.77 (*0.99*,**2.76**)	1.87 (*1.22*,**2.93**)	1.96 (*1.27*,**3.09**)
	Rising	0.76 (*0.14*,**1.33**)	1.07 (*0.40*,**1.63**)	1.14 (*0.54*,**1.74**)	1.16 (*0.64*,**1.77**)	1.18 (*0.75*,**1.80**)	1.20 (*0.76*,**1.84**)
AR	All	0.73 (*0.13*,**1.24**)	1.02 (*0.43*,**1.56**)	1.08 (*0.53*,**1.66**)	1.10 (*0.60*,**1.69**)	1.11 (*0.70*,**1.72**)	1.12 (*0.74*,**1.74**)
	High	0.99 (*0.27*,**2.52**)	1.49 (*0.64*,**2.36**)	1.65 (*0.83*,**2.61**)	1.74 (*0.97*,**2.75**)	1.84 (*1.19*,**2.94**)	1.92 (*1.32*,**3.10**)
	Rising	0.75 (*0.14*,**1.33**)	1.06 (*0.44*,**1.63**)	1.13 (*0.56*,**1.74**)	1.15 (*0.63*,**1.78**)	1.17 (*0.74*,**1.82**)	1.19 (*0.79*,**1.86**)
LLT	All	0.86 (*0.11*,**1.72**)	1.83 (*0.78*,**3.02**)	2.64 (*1.24*,**4.43**)	3.31 (*1.67*,**5.75**)	4.55 (*2.37*,**8.15**)	6.77 (*3.09*,**11.21**)
	High	1.01 (*0.23*,**3.49**)	2.83 (*1.11*,**4.32**)	4.11 (*1.82*,**6.49**)	5.64 (*2.68*,**9.22**)	8.96 (*4.56*,**15.67**)	13.68 (*6.22*,**22.56**)
	Rising	0.90 (*0.12*,**1.84**)	2.01 (*0.85*,**3.31**)	3.18 (*1.52*,**5.40**)	4.20 (*2.19*,**7.52**)	5.99 (*3.30*,**11.35**)	9.35 (*4.35*,**15.76**)
DLLT	All	0.86 (*0.10*,**1.90**)	2.47 (*0.78*,**3.02**)	2.72 (*1.24*,**4.41**)	3.51 (*1.67*,**5.72**)	4.98 (*2.36*,**8.12**)	6.77 (*3.08*,**11.16**)
	High	1.01 (*0.19*,**3.60**)	4.55 (*1.11*,**4.32**)	4.14 (*1.82*,**6.48**)	5.79 (*2.68*,**9.20**)	9.67 (*4.55*,**15.62**)	13.69 (*6.20*,**22.46**)
	Rising	0.90 (*0.11*,**2.02**)	2.69 (*0.85*,**3.31**)	3.29 (*1.51*,**5.37**)	4.54 (*2.18*,**7.49**)	6.83 (*3.29*,**11.30**)	9.35 (*4.33*,**15.68**)
RWD	All	0.92 (*0.18*,**1.43**)	2.05 (*0.87*,**3.39**)	2.90 (*1.39*,**5.11**)	3.80 (*1.87*,**6.92**)	5.35 (*2.86*,**10.48**)	7.53 (*4.01*,**15.37**)
	High	1.22 (*0.37*,**2.93**)	3.29 (*1.31*,**5.10**)	4.77 (*2.10*,**7.69**)	6.62 (*3.02*,**11.15**)	10.66 (*5.49*,**20.09**)	15.47 (*8.05*,**30.83**)
	Rising	0.97 (*0.20*,**1.54**)	2.25 (*0.95*,**3.71**)	3.45 (*1.70*,**6.22**)	4.82 (*2.46*,**9.09**)	7.15 (*4.07*,**14.90**)	10.24 (*5.81*,**22.24**)
IRW	All	0.86 (*0.10*,**1.93**)	2.47 (*0.78*,**3.02**)	2.72 (*1.24*,**4.41**)	3.51 (*1.66*,**5.72**)	4.98 (*2.37*,**8.11**)	6.77 (*3.08*,**11.15**)
	High	1.01 (*0.19*,**3.58**)	4.55 (*1.11*,**4.32**)	4.14 (*1.82*,**6.48**)	5.79 (*2.67*,**9.19**)	9.66 (*4.55*,**15.60**)	13.68 (*6.20*,**22.44**)
	Rising	0.90 (*0.11*,**2.05**)	2.69 (*0.85*,**3.31**)	3.29 (*1.51*,**5.37**)	4.54 (*2.18*,**7.48**)	6.82 (*3.29*,**11.29**)	9.34 (*4.32*,**15.66**)
SRW	All	0.66 (*0.10*,**1.29**)	0.88 (*0.52*,**1.36**)	1.02 (*0.73*,**1.57**)	1.11 (*0.91*,**1.71**)	1.25 (*1.14*,**1.93**)	1.43 (*1.35*,**2.22**)
	High	0.71 (*0.20*,**2.51**)	0.97 (*0.60*,**1.57**)	1.10 (*0.84*,**1.80**)	1.23 (*1.07*,**2.02**)	1.60 (*1.55*,**2.63**)	1.97 (*1.96*,**3.23**)
	Rising	0.67 (*0.11*,**1.38**)	0.91 (*0.54*,**1.42**)	1.10 (*0.80*,**1.72**)	1.26 (*1.05*,**1.97**)	1.50 (*1.39*,**2.36**)	1.77 (*1.71*,**2.80**)
DT	All	0.66 (*0.10*,**1.29**)	0.88 (*0.52*,**1.36**)	1.02 (*0.73*,**1.57**)	1.11 (*0.91*,**1.71**)	1.25 (*1.14*,**1.93**)	1.43 (*1.35*,**2.22**)
	High	0.71 (*0.21*,**2.55**)	0.97 (*0.60*,**1.57**)	1.10 (*0.84*,**1.80**)	1.23 (*1.07*,**2.02**)	1.60 (*1.55*,**2.63**)	1.97 (*1.96*,**3.23**)
	Rising	0.67 (*0.11*,**1.38**)	0.91 (*0.54*,**1.42**)	1.10 (*0.80*,**1.72**)	1.26 (*1.05*,**1.97**)	1.50 (*1.39*,**2.36**)	1.77 (*1.71*,**2.80**)
SLLT	All	0.66 (*0.10*,**1.29**)	0.88 (*0.51*,**1.35**)	1.02 (*0.72*,**1.57**)	1.11 (*0.87*,**1.71**)	1.24 (*1.09*,**1.90**)	1.41 (*1.29*,**2.17**)
	High	0.71 (*0.21*,**2.60**)	0.97 (*0.59*,**1.55**)	1.10 (*0.83*,**1.80**)	1.23 (*1.04*,**2.02**)	1.59 (*1.50*,**2.62**)	1.94 (*1.90*,**3.17**)
	Rising	0.67 (*0.11*,**1.38**)	0.91 (*0.54*,**1.40**)	1.10 (*0.79*,**1.71**)	1.26 (*1.00*,**1.96**)	1.49 (*1.33*,**2.31**)	1.75 (*1.62*,**2.71**)

reveal one aspect of how the adaptive gain behaves in a less than ideal situation. During January there is a period where no observations are taken. During this time, the prediction interval can at first be seen to gradually widen, as no further observations are available to condition the correction of the current simulation of the flood forecasting model. Then, when the next run of the flood forecasting model becomes available the gain cannot be initialised, so no forecasts are issued. In such situations, the forecaster making operational decisions may wish to consider the past simulation of the flood forecasting model and the corresponding adaptive gain.

Figure 6 shows the predictions made for the same event using the RW model calibrated using the SEFE methodology with empirical prediction bounds. Visual comparison with Fig. 5 shows the forecasts to be more precise, but failing to capture the rising limb of the validation hydrograph.

4.2 Llandrinio

Tables 6, 7 and 8 summarise the performance in calibration of the adaptive gain models at Llandrinio. As at Welsh bridge, there is a bias in the baseflow of the hydraulic model. Performance is assessed only when the flood forecasting model

predicts a water level greater than 1.8 m. None of the models for the adaptive gain produce a RMSE lower than that of the uncorrected model (0.39) for lead times greater than 9 h. This time is shorter than that at Welsh bridge which is consistent with the location of the gauging site closer to the headwaters of the catchment. As for Welsh bridge the deterioration in performance is more marked for gain models, which include a gradient term but do not include a damping parameter.

The coverage probabilities during calibration (Table 7) are less dependant upon the water level than those at Welsh bridge. In keeping with the Welsh bridge results, the precision of forecasts issued using the SEFE methodology, with empirical error bounds, is higher than that of the gaussian calibration methodology. In validation (9) all the prediction error bounds appear conservative, bracketing a higher proportion of the observations than would be expected.

Figure 7 shows the 6 h lead item forecasts for two large calibration vents at Llandrinio. As at Welsh bridge the technique cannot correct for highly erroneous model outputs such as those around the 6 December 2006. An earlier period on the 3 December 2006 highlights the importance of

Table 9. Fraction of observations at Llandrinio bracketed by the estimated 95 % prediction intervals during validation. Bracketed results are those for the SEFE calibration with the italicised and bold values corresponding to the empirical and theoretical symmetric bounds. Results are shown for all the time period, high levels (> 4 m) and periods where the hydrograph is rising for different combinations of lead time (hours) and GRW model.

	Period	2	6	9	12	18	24
RW	All	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.97*,**0.99**)
	Rising	0.99 (*0.96*,**0.99**)	0.98 (*0.96*,**0.99**)	0.98 (*0.96*,**0.98**)	0.98 (*0.96*,**0.98**)	0.98 (*0.96*,**0.98**)	0.98 (*0.96*,**0.98**)
AR	All	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.97*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.99*,**0.99**)	0.99 (*0.98*,**0.99**)
	Rising	0.99 (*0.96*,**0.99**)	0.98 (*0.98*,**0.99**)	0.98 (*0.97*,**0.98**)	0.98 (*0.97*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
LLT	All	0.99 (*0.96*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)
	Rising	0.99 (*0.92*,**0.99**)	0.99 (*0.97*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
DLLT	All	0.99 (*0.97*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)
	Rising	0.99 (*0.94*,**0.99**)	0.99 (*0.97*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
RWD	All	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)
	Rising	0.99 (*0.97*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
IRW	All	0.99 (*0.97*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)
	Rising	0.99 (*0.95*,**0.99**)	0.99 (*0.97*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
SRW	All	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.97*,**0.99**)	0.98 (*0.96*,**0.99**)	0.97 (*0.96*,**0.99**)	0.96 (*0.95*,**0.99**)	0.95 (*0.95*,**0.99**)
	Rising	0.99 (*0.96*,**0.99**)	0.98 (*0.97*,**0.98**)	0.98 (*0.96*,**0.98**)	0.97 (*0.96*,**0.98**)	0.97 (*0.96*,**0.98**)	0.96 (*0.96*,**0.98**)
DT	All	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.97*,**0.99**)	0.98 (*0.96*,**0.99**)	0.97 (*0.96*,**0.99**)	0.96 (*0.95*,**0.99**)	0.95 (*0.95*,**0.99**)
	Rising	0.99 (*0.96*,**0.99**)	0.98 (*0.97*,**0.98**)	0.98 (*0.96*,**0.98**)	0.97 (*0.96*,**0.98**)	0.97 (*0.96*,**0.98**)	0.96 (*0.96*,**0.98**)
SLLT	All	0.99 (*0.97*,**0.99**)	0.99 (*0.98*,**0.99**)	0.99 (*0.98*,**0.99**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)	0.98 (*0.98*,**0.98**)
	High	0.99 (*0.99*,**0.99**)	0.99 (*0.97*,**0.99**)	0.98 (*0.97*,**0.99**)	0.97 (*0.96*,**0.99**)	0.96 (*0.97*,**0.99**)	0.96 (*0.97*,**0.99**)
	Rising	0.99 (*0.94*,**0.99**)	0.98 (*0.97*,**0.98**)	0.98 (*0.97*,**0.98**)	0.97 (*0.97*,**0.98**)	0.97 (*0.97*,**0.98**)	0.97 (*0.97*,**0.98**)

assimilating further observations after initialisation of the gain to improve the precision of the forecasts.

5 Conclusions

The results presented in this paper indicate that comparatively simple error models combined with real time data assimilation can provide probabilistic forecasts.

Results are shown for a medium size river basin. Two calibration methodologies are presented. Detailed residual analysis suggests that, though the error assumptions used in these derivations are not satisfied, they can produce probabilistic forecasts with conservative coverage and precision of the order of 10 % of the observed water level. The selection of which calibration methodology performs most adequately requires consideration on a case by case basis. Techniques for making this selection are presented.

The lead times at which the expected value of the forecasts can be seem to improve the RMSE of the uncorrected model predictions depends upon the time it takes for the catchment to respond at the gauged site. It remains to be seen if such a data assimilation methodology is able to perform adequately on small basins where the response to rainfall is more rapid and timing errors, due perhaps to errors in the rainfall forecast, are more severe (see Alfieri et al., 2011, for some initial results).

In an operational setting, only the linear Kalman filter evolving two states needs to be evaluated. Though not discussed this can be readily implemented in spreadsheet packages without recourse to more complex programming. This makes the addition of these data assimilation techniques to existing deterministic forecasting systems comparatively straightforward with a very low "cost of entry".

A number of future developments can be proposed. Firstly, analysis of the results suggests that recognition of state or time dependency within the model parameters (Young, 2002; Smith et al., 2008b) may be worthy of exploration. Recognising that the gain at different sites may be correlated opens up a further line for future research. If such correlations can be successfully exploited, it will both increase the robustness of the data assimilation scheme to missing observations and potentially allow forecast updating at sites where very limited

data are available (e.g. locations only observed during the calibration of a hydraulic model). The exploration of such strategies is the subject of on-going research.

References

Akaike, H.: New Look At Statistical-model Identification, IEEE T. Automat. Contr., AC19, 716–723, 1974.

Alfieri, L., Smith, P. J., Thielen-del Pozo, J., and Beven, K. J.: A staggered approach to flash flood forecasting – case study in the Cévennes region, Adv. Geosci., 29, 13–20, doi:10.5194/adgeo-29-13-2011, 2011.

Aronica, G., Hankin, B., and Beven, K.: Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, Adv. Water Res., 22, 349–365, 1998.

Beven, K., Smith, P., and Freer, J.: So just why would a modeller choose to be incoherent?, J. Hydrol., 354, 15–32, 2008.

Box, G. E. P. and Jenkins, G. M.: Time Series Analysis: Forecasting and Control, Prentice Hall PTR, 3rd Edn., 1994.

Cole, S. J., Robson, A., Bell, V., and Moore, R.: Model initialisation, data assimilation and probabilistic flood forecasting for distributed hydrological models., Geophys. Res. Abstr., 11, EGU2009-8048-3, 2009.

DaRos, D. and Borga, M.: Adaptive use of a conceptual model for real time flood forecasting, Nord. Hydrol., 28, 169–188, 1997.

Doucet, A., de Freitus, N., and Gordon, N.: An introduction to Sequential Monte Carlo Methods, in: Sequential Monte Carlo Methods in Practice, edited by: Doucet, A., de Freitus, N., and Gordon, N., Chap. 1, New York, Springer-Verlag, 3–13, 2001.

Durbin, J. and Koopman, S.: Time series analysis by state space methods, Oxford University Press, 2001.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implimentation, Ocean Dynam., 53, 343–367, 2003.

Harvey, A.: Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, 1989.

Jakeman, A. J. and Young, P. C.: Recursive Filtering and Smoothing Procedures For the Inversion of Ill-posed Causal Problems, Utilitas Mathematica, 25, 351–376, 1984.

Kalman, R.: A new approach to linear filtering and prediction problems, J. Basic Eng. T.-ASME, 82, 35–45, 1960.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, W03S04, doi:10.1029/2005WR004362, 2006.

Krzysztofowicz, R. and Maranzano, C. J.: Hydrologic uncertainty processor for probabilistic stage transition forecasting, J. Hydrol., 293, 57–73, 2004.

Lees, M. J., Young, P. C., Ferguson, S., Beven, K. J., and Burns, J.: An adaptive flood warning scheme for the River Nith at Dumfries, in: 2nd Inter-national Conference on River Flood Hydraulics, edited by: White, W. R. and Watts, J., Wiley, Chichester, 1994.

Li, Z. J. and Navon, I. M.: Optimality of variational data assimilation and its relationship with the Kalman filter and smoother, Q. J. Roy. Meteorol. Soc., 127, 661–683, 2001.

Liu, J. S. and Chen, R.: Sequential Monte Carlo methods for dynamic systems, J. Am. Stat. Assoc., 93, 1032–1044, 1998.

Liu, Y. Q. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resour. Res., 43, W07401, doi:10.1029/2006WR005756, 2007.

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.

Maranzano, C. J. and Krzysztofowicz, R.: Identification of likelihood and prior dependence structures for hydrologic uncertainty processor, J. Hydrol., 290, 1–21, 2004.

Montanari, A. and Brath, A.: A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, Water Resour. Res., 40, W01106, doi:10.1029/2003WR002540, 2004.

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. L., Gupta, H., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41, W05012, doi:10.1029/2004WR003604, 2005a.

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter, Adv. Water Resour., 28, 135–147, 2005b.

Pappenberger, F., Beven, K., Frodsham, K., Romanowicz, R., and Matgen, P.: Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach, J. Hydrol., 333, 275–287, 2007.

Pukelsheim, F.: The 3-sigma-rule, Am. Stat., 48, 88–91, 1994.

Rajaram, H. and Georgakakos, K. P.: Recursive parameter estimation of hydrologic models, Water Resour. Res., 25, 281–294, 1989.

Reichle, R. H.: Data assimilation methods in the Earth sciences, Adv. Water Resour., 31, 1411–1418, 2008.

Reichle, R. H., Crow, W. T., and Keppenne, C. L.: An adaptive ensemble Kalman filter for soil moisture data assimilation, Water Resour. Res., 44, W03423, doi:10.1029/2007WR006357, 2008.

Rougier, J., Sexton, D. M. H., Murphy, J. M., and Stainforth, D.: Analyzing the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles from Different but Related Experiments, Journal of Climate, 22, 3540–3557, 2009.

Schwarz, G.: Estimating Dimension of A Model, Ann. Stat., 6, 461–464, 1978.

Schweppe, F. C.: Evaluation of Likelihood Functions For Gaussian Signals, IEEE T. Inform. Theory, 11, 61–70, 1965.

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

Seo, D. J., Herr, H. D., and Schaake, J. C.: A statistical postprocessor for accounting of hydrologic uncertainty in short-range ensemble streamflow prediction, Hydrol. Earth Syst. Sci.

Discuss., 3, 1987–2035, doi:10.5194/hessd-3-1987-2006, 2006.

Seo, D. J., Cajina, L., Corby, R., and Howieson, T.: Automatic state updating for operational streamflow forecasting via variational data assimilation, J. Hydrol., 367, 255–275, 2009.

Shamir, E., Lee, B.-J., Bae, D.-H., and Georgakakos, K. P.: Flood Forecasting in Regulated Basins Using the Ensemble Extended Kalman Filter with the Storage Function Method, Journal of Hydrologic Engineering, 15, 1030–1044, 2010.

Smith, P. J., Beven, K. J., and Tawn, J. A.: Detection of structural inadequacy in process-based hydrological models: A particle-filtering approach, Water Resour. Res., 44, W01410, doi:10.1029/2006WR005205, 2008a.

Smith, P. J., Beven, K. J., Tych, W., Hughes, D., Coulson, G., and Blair, G.: The provision of site specific flood warnings using wireless sensor networks, in: Proceedings of FLOODRisk2008, 2008b.

Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, International Journal of River Basin Management, 6, 123–137, 2008.

Vernon, I., Goldstein, M., and Bower, R. G.: Galaxy Formation: a Bayesian Uncertainty Analysis, Bayesian Analysis, 5, 619–669, 2010.

Weed, H. D. and Bradley, R. A.: Sequential One-sample Grouped Signed Rank Tests For Symmetry – Basic Procedures, J. Am. Stat. Assoc., 66, 321–326, doi:10.2307/2283930, 1971.

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., El Serafy, G. Y., Hummel, S., Dhondia, J., and Gerritsen, H.: Application of generic data assimilation tools (DATools) for flood forecasting purposes, Comput. Geosci., 36, 453–463, 2010.

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.

Wilcoxon, F.: Individual Comparisons By Ranking Methods, Biometrics Bull., 1, 80–83, 1945.

Young, P. C.: Advances in real-time flood forecasting, Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences, 360, 1433–1450, 2002.

Young, P. C., Ng, C., and Armitage, P.: A Systems-approach To Recursive Economic-forecasting and Seasonal Adjustment, Comput. Math. Appl., 18, 481–501, 1989.

Young, P. C., McKenna, P., and Bruun, J.: Identification of nonlinear stochastic systems by state dependent parameter estimation, International Journal of Control, 74, 1837–1857, 2001.