

Department of Mathematics and Statistics

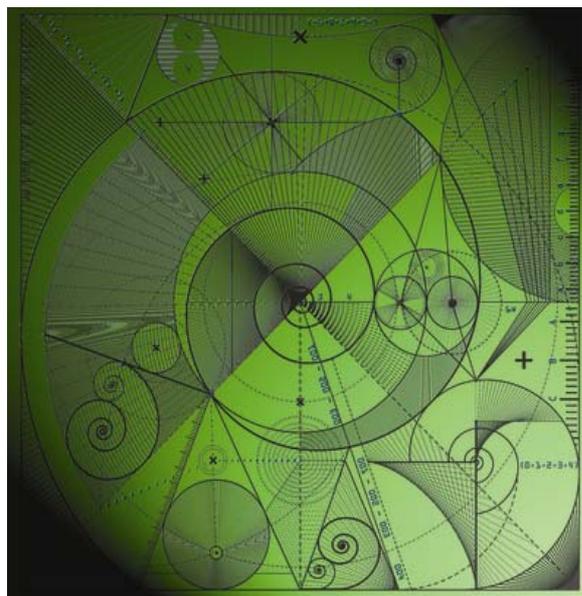
Preprint MPS-2013-01

9 January 2013

Scheduling satellite-based SAR acquisition
for sequential assimilation of water level
observations into flood modelling

by

Javier Garcia-Pintado, Jeff C. Neal, David C.
Mason, Sarah L. Dance and Paul D. Bates



Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling

Javier García-Pintado^{a,b,*}, Jeff C. Neal^c, David C. Mason^{a,b}, Sarah L. Dance^{a,b}, Paul D. Bates^c

^a*School of Mathematical and Physical Sciences, University of Reading, UK*

^b*National Centre for Earth Observation, University of Reading, Reading, UK*

^c*School of Geographical Sciences, University of Bristol, Bristol, UK*

Abstract

Satellite-based Synthetic Aperture Radar (SAR) has proved useful for obtaining information on flood extent, which, when intersected with a Digital Elevation Model (DEM) of the floodplain, provides water level observations that can be assimilated into a hydrodynamic model to decrease forecast uncertainty. With an increasing number of operational satellites with SAR capability, information on the relationship between satellite first visit and revisit time and forecast performance is required to optimise the operational scheduling of satellite imagery. By using an Ensemble Transform Kalman Filter (ETKF) and a synthetic analysis with the 2D hydrodynamic model LISFLOOD-FP based on a real flooding case affecting an urban area (summer 2007, Tewkesbury, Southwest UK), we evaluate the sensitivity of the forecast performance to visit parameters. We emulate a generic hydrologic-hydrodynamic modelling cascade by imposing a bias and spatiotemporal correlations to the inflow error ensemble into the hydrodynamic domain. First, in agreement with previous research, estimation and correction for this bias leads to a clear improvement in keeping the forecast on track. Second, imagery obtained early in the flood is shown to have a large influence on forecast statistics. Revisit interval is most influential for early observations. The results are promising for the future of remote sensing-based water level observations for real-time flood forecasting in complex scenarios.

Keywords: Data assimilation, Remote Sensing, Synthetic aperture radar, Flood forecasting, Urban flood, Parameter estimation

2010 MSC: 60G35

2010 MSC: 62M20

2010 MSC: 93E10

2010 MSC: 93E11

1. Introduction

Hydrodynamic simulation is a basic tool used by most real-time flood forecasting systems. Remote sensing has proved useful for obtaining water level observations (WLOs) during flood events.

*Corresponding author. Tel.: +44(0) 118 378 7722.

ESSC, Harry Pitt Building, 3 Earley Gate, University of Reading, Whiteknights Campus, RG6 6AH, Reading, United Kingdom

Email address: j.garcia-pintado@reading.ac.uk (Javier García-Pintado)

In the UK, as in many other places, a difficulty for flood observation is that standard gauges are typically sited only every ~ 20 km, so give little information on the spatial variations in the flood level, which may be particularly important in urban areas. Much more spatial information is contained in the flood extents captured in satellite Synthetic Aperture Radar (SAR) images. SAR is generally used for flood detection rather than visible-band sensors because of its all-weather day-night capability. Distributed water levels may be estimated indirectly along the flood extents in SAR images by intersecting the extents with a floodplain Digital Elevation Model (DEM) (Horritt et al., 2003; Lane et al., 2003; Raclot, 2006; Schumann et al., 2007). Consequently, a number of studies have focused on assimilating SAR-derived WLOs into hydrodynamic forecasting models (e.g., Neal et al., 2009; Hostache et al., 2010; Matgen et al., 2010; Giustarini et al., 2011). Specifically, Neal et al. (2009) analysed how dense a gauge network would need to be to match the performance of SAR-derived WLOs in a data assimilation context.

In the future, an alternative will be direct space-borne WLOs at high resolution using NASA/CNES's Surface Water and Ocean Topography (SWOT) mission, which will use K_a -band radar interferometry to measure surface water levels to 10 cm accuracy on rivers ~ 100 m wide. However, as SWOT is not scheduled for launch until 2020 and will not measure levels for floods less than 100 m wide, the water levels from SAR flood boundaries should continue to be an important source of data for assimilation into models, especially in the near future (Mason et al., 2012b).

Data assimilation is an iterative approach to the problem of estimating the state of a dynamical system using both current and past observations of the system together with a model for the system's time evolution. Within Data Assimilation (DA), the ensemble Kalman Filter (EnKF) is becoming a method of choice for large-scale data assimilation systems, along with variational methods, in a number of Earth science disciplines. For hydrodynamic experiments, e.g., Andreadis et al. (2007), Durand et al. (2008), and Biancamaria et al. (2011) successfully assimilated virtual observations of the proposed SWOT mission with simulations from the LISFLOOD-FP hydraulic model (Bates & De Roo, 2000). Specifically, the studies by Andreadis et al. (2007) and Biancamaria et al. (2011) were based on the square root implementation of the analysis scheme proposed by Evensen (2004). In variational techniques, Lai & Monnier (2009) used 4D-var to assimilate spatially distributed water levels into a shallow-water flood model. Alternatively, Matgen et al. (2010) and Giustarini et al. (2011) evaluated the performance of assimilation schemes based on the Particle Filter (PF), which does not require the Gaussian distribution of error assumed by the EnKF and variational methods. These two studies used SAR-derived WLOs, the former with synthetic and the latter with two real observations (ERS-2 and ENVISAT). However, their studies, both in a 19-km reach of the Alzette River, used the 1-D HEC-RAS hydrodynamic model within a single transect and one upstream boundary condition. With their model setup, the problem had a state vector length $n = 144$, and they used 64 particles to approach the PF problem. While Matgen et al. (2010) comment that their methodology can be extended to rivers with more complex geometry (which would need a 2-D model), they do not consider the issue of increase in dimensionality. As an example, the problem in the present study includes a number of distributed boundary conditions and affects rural and urban areas. To adequately represent the geometry, we consider $664 \times 408 = 270902$ pixels within a rectangular domain. Just considering flooded cells in the model, the maximum extent of the flooded area is about 15200 pixels. The state vector length is thus more than 100 times bigger than in these two studies. The feasibility of the ensemble Kalman filter with ensemble sizes much smaller than the state dimension has been demonstrated in operational numerical weather prediction (e.g., Houtekamer & Mitchell, 2005), and has some

theoretical justification (e.g., Furrer & Bengtsson, 2007). Conversely, as, discussed by Snyder et al. (2008), there are results showing that the standard particle filter must have an ensemble size exponentially large in the variance of the observation log likelihood or the filter will suffer from a “collapse”. Thus, despite current research to improve the PF efficiency for large dimensional problems, it remains unclear whether it will be a viable alternative in a near future for these operational flooding problems in areas with high human or economical risk.

Both EnKF and PF are Monte Carlo-based filters that require a number of ensembles of model runs to represent the forecast uncertainty. 2-D hydrodynamic models for simulating floods are expensive to run in ensemble mode with the result that, in operational cases, watershed scale hydrodynamic modelling is currently prohibitive, and thus the hydrodynamic model must either be restricted to a computationally feasible domain or use a lower resolution, which may not be adequate. In order to increase forecast lead times, within a modelling cascade, a low resolution hydrologic model can be used for obtaining the watershed response to rainfall, and this response can be used as input flow boundary conditions for the hydrodynamic model.

For ensemble simulations, the spread of the hydrologic model responses represents the hydrologic forecast uncertainty. Also, the ensemble mean will differ from the true watershed response. This difference will take the form of a time-correlated mean error, which will be considered a bias if it remains stationary during the time span of the simulations. The evolution of this mean error will be a function of the various errors inherent in the data (mostly rainfall) and the hydrologic model. This mean error in the input to the hydrodynamic domain tends to offset the benefit of the DA within the hydrodynamic model.

It has been shown that the persistence of DA improvement on hydrodynamic model simulations is limited if DA is just used for updating the state vector (water stage), as the errors in upstream boundary conditions can have a dominating effect on the flooded area within a short time after the assimilation step. To tackle this problem, some studies have proposed to estimate and correct the error in upstream inflows (Andreadis et al., 2007; Matgen et al., 2010), with different approaches. In general, DA can be used to estimate uncertain model parameters. From a DA point of view, input flow boundary conditions, as well as friction in land and channels, can be considered as parameters to be estimated; their difference, as mentioned above, being that inflow errors have a much higher variability in time than friction parameters can have.

Satellite-based SAR acquisitions have an undeniable cost to water authorities and risk management services. Although it is possible to obtain a sequence of SAR images of a flood using data from several different uncoordinated satellites, the only satellite constellation currently available to provide image sequences is the COSMO-SkyMed (CSK) constellation, which has sufficiently high resolution (up to 1m) to image flooding in urban areas. COSMO-SkyMed is an Italian Space Agency constellation with 4 satellites in sun-synchronous orbit with a 97.9° inclination. In a decision-making scenario (an imminently forecasted flood), for the CSK constellation, two visit parameters must be taken into account; the first visit time (the time of the first SAR image acquisition) and the revisit time (the time between two consecutive acquisitions over the same target). CSK offers 12 h and 24 h revisit times. Too late a first visit may miss important information for forecasting purposes, while one that is too early may provide little information or be prematurely scheduled without any further flood development, so incurring unnecessary costs.

We are interested in maximizing the capability of remotely-sensed WLOs to decrease the forecast uncertainty. Here we evaluate the influence of the time of the first visit and revisit time on the error characteristics of the flood event. We assume that the difference in time between the

acquisition of the CSK image and the time at which its WLOs are available to the user (the information age) is negligible. In practice the event sequence is not currently near real-time for high resolution SARs, though may become so in the near future. Operational considerations concerned with acquiring high resolution satellite SAR images of a developing flood and extracting WLOs in near real-time have been considered in Mason et al. (2012a,b). CSK is likely to be followed by other constellations with lower information ages (e.g., Sentinel-1). The aim of this paper is to be generic, so that the issue of information age should be an additional consideration for the particular satellite concerned.

This study builds upon previous analyses of remotely-sensed WLO DA. Our main goal is to evaluate the sensitivity of the forecasting and DA performance to a number of realistic hypothetical visit scenarios using satellite-based SAR WLOs. For this, we use a real flood in an urban area and real inflow measurements as base scenario, but employ a controlled identical twin experiment for the study. Firstly, we obtain a family of three curves that show mean forecast statistics (Root Mean Square Error) for the event as a function of visit times. Each curve represents a revisit time ($\Delta t_a = 12$ h, 24 h, and 48 h), and is built up by successively delaying the time of the first visit but keeping a common last visit time (at a late stage within the flood event). Secondly, for a selected revisit/DA time ($\Delta t_a = 24$ h) we simulate a budget-limited scenario, by successively delaying a fixed number of SAR overpasses. As a DA technique, we use an Ensemble Transform Kalman Filter (ETKF) and conduct parameter (inflow errors) estimation through augmentation of the state vector. We expand the discussion by highlighting related issues, such as the importance of inflow error estimation and the evolution of the correct spread, that should deserve further consideration in operational environments with sequential DA.

The rest of this paper is organised as follows: In Section 2, we describe the experimental design, the study domain, the hydrodynamic model, the generation of synthetic satellite observations, the ensemble filter, the generation of inflow boundary condition errors, and the applied verification methods. In Section 3, we present and discuss the results, describing the influence of updating the inflow boundary conditions during the assimilation process, the evolution of the ensemble during the sequential assimilation, and the sensitivity to first visit and revisit times. Conclusions are provided in Section 4.

2. Methods

2.1. Experimental Design

We use an identical twin experiment with a hydrodynamic model grounded in a real flood event. In this study, we assume that friction is known and constant (e.g., through prior model calibration), but that inflows are poorly known and their errors are estimated and corrected by the filter. For this, we choose pre-calibrated friction parameters for the floodplain and channels, and a set of measured inflow/stage boundary conditions to simulate a “true” event. Then, we obtain synthetic SAR-type WLOs from this “truth”, and for the same period we corrupt the inflow boundary conditions to generate an ensemble of inflows with added errors. As we assume that measured inflows are the truth, to generate the ensemble of inflows, we first impose a stationary mean error as a multiplicative bias on this truth. Then, the biased inflow time series are further corrupted by spatiotemporally-correlated errors to generate the ensemble of inflows into the study domain. This is described in Section 2.5. The inflow ensemble is used for generating an open-loop simulation, without DA, and for all the simulations assimilating the synthetic WLOs under various SAR visit scenarios.

Within ensemble Kalman filters and several contexts, it has been shown that as the size of the ensembles increases, correlations are estimated more accurately (e.g., Houtekamer & Mitchell, 2001). Note that ensemble Kalman filtering quantifies uncertainty only in the space spanned by the ensemble. If computational resources restrict the number of ensemble members m to be much smaller than the number of model variables n , this can be a severe limitation. Here, for our $\sim 1.5 \times 10^4$ effective state vector length (pixels within the flooded area), we arbitrarily set the ensemble size $m = 210$ as a relatively big one in comparison with that from typical operational applications with high computational demand, as is this case. The size of m was chosen to keep reasonable computing times given available computing resources. In this study, we do not conduct any test of the forecast-error covariance sensitivity to the ensemble size, and we do not use localization. Nevertheless, we investigate the ensemble reliability for the chosen size (see Section 3). We assume that the system can be represented on a discrete grid and, for the purposes of this study only, that the system model is “perfect”, i.e. it gives an exact description of the true behaviour of the system.

2.2. Study Domain and Hydrodynamic Model

This study focuses on the area of the lower Severn and Avon rivers in South West United Kingdom, over a $30.6 \times 49.8 \text{ km}^2$ (1524 km^2) domain. The case study is 1-in-150-year flood event that took place in July 2007 in the area. It resulted in substantial flooding of urban and rural areas, with about 1500 homes in Tewkesbury being flooded (Mason et al., 2010; Schumann et al., 2011). Tewkesbury lies at the confluence of the Severn, flowing from the Northwest, and the Avon, flowing in from the Northeast. Fig. 1 depicts the domain for the current study. The peak of the flood ($> 550 \text{ m}^3\text{s}^{-1}$ at Saxons Lode Us) occurred on July 22, and the river did not return to bank-full until July 31 ($\sim 350 \text{ m}^3\text{s}^{-1}$ at Saxons Lode Us).

We set up time-varying boundary conditions from real measurements of seven input flows and one downstream stage time series (see Fig. 1). The three boundary conditions with highest inflows were Bewdley (peak inflow $Q_p = 300 \text{ m}^3\text{s}^{-1}$) in the Severn, Evesham ($Q_p = 465 \text{ m}^3\text{s}^{-1}$) in the Avon, and Knightsford Bridge ($Q_p = 315 \text{ m}^3\text{s}^{-1}$) in the Teme. The Severn also had inflows from Kidder Callows ($Q_p = 33 \text{ m}^3\text{s}^{-1}$) and Hardford Hill ($Q_p = 36 \text{ m}^3\text{s}^{-1}$), prior to its junction with the Teme. The Avon also had a sharp, short duration, inflow from Hinton ($Q_p = 85 \text{ m}^3\text{s}^{-1}$), and from Besford Bridge ($Q_p = 315 \text{ m}^3\text{s}^{-1}$), downstream. The supplementary material includes plots for all hydrographs.

The simulation interval goes from 2007-07-19 13:00:00 UTC to 2007-08-01 19:15:00 UTC, and all time series have a 15' temporal resolution. The area is modelled by the flood inundation model LISFLOOD-FP, a coupled 1D/2D hydraulic model based on a raster grid (Bates & De Roo, 2000). It predicts water depth in each grid cell at each time step. After each assimilation step, the model is re-initialized with the updated state vector (water stage). LISFLOOD-FP has several formulations. Here, we apply the so-call “sub-grid” approach described by Neal et al. (2012). This scheme uses a computationally efficient finite difference numerical scheme adapted from the reach scale inundation model of Bates et al. (2010), and utilises gridded river network data, assuming a rectangular channel geometry. This is a scheme that considers the diffusion and local inertial term of the 1D shallow water equations as a means of increasing stable time steps. These 1D equations are solved for each face of a 2D grid cell to provide a 2D solution that is decoupled in x and y . Neal et al. (2012) analyse the scheme for an application in the River Niger Inland delta regarding the simulation of water surface elevation, inundation and wave propagation. LISFLOOD-FP is here applied to the domain with 75 m pixel resolution. Thus, the domain is $408 \times 664 = 270912$ pixels, but the maximum extent of the flooded area is about 15200 pixels. The source digital terrain model

(DTM) was the NEXTMap British digital terrain model dataset (5×5 m resolution), derived from airborne Interferometric Synthetic Aperture Radar (IFSAR), which was upscaled by explicitly removing channel depth, later parametrized into the sub-grid geometry. To describe the channel geometry, we used the power law relationship $d = \lambda w^\gamma$ between the channel width (w) and depth (d), where we used the parameters $\lambda = 0.30$, and $\gamma = 0.78$. For the main rivers, we estimated mean channel widths from field campaigns, and calibrated λ and γ using within bank water level dynamics measured by the available gauges, using the same method as Neal et al. (2012). Width values were $w = 20, 35, 50,$ and 60 m for the Teme, Avon, Severn upstream of its junction with the Avon, and Severn downstream from this junction, respectively. For smaller tributaries we kept the same λ and γ values, and assigned widths in 5–15 m on the basis of drainage areas obtained from the DTM. These seemed reasonable when cross-checked with field observations. Simulations are run with an initial time step of 20 s, which is internally adapted for every time step by the model to optimize calculations. The 15 minute input time series are linearly interpolated in time within the model to be adapted to the internal time step. The Manning friction parameters used for the channels and floodplain are $n_{ch} = 0.035$ and $n_{fp} = 0.06$, respectively.

2.3. Virtual Satellite Observations

Distributed water levels may be estimated indirectly along the flood extents in SAR images by intersecting the extents with the floodplain topography. The WLOs are produced as a continuous line along the boundary of a flood extent, and it is necessary to select a subset of level observations at individual points along the boundary for assimilation, because errors in adjacent levels along the flood extent will be strongly correlated (Mason et al., 2012b; Stephens et al., 2012). In theory, this is a tractable problem as correlated errors could be included into the assimilation procedure through a correct specification of the non-diagonal values in the observation error covariance matrix \mathbf{R} . However, in practice it is extremely difficult to correctly estimate these covariances, and if they are not correctly specified the filter will deviate from the optimal; e.g., the updated state will be excessively biased towards the observations if their error covariances are underestimated or assumed zero (Stewart et al., 2008). It is more straightforward to filter out observations so that just a subset with uncorrelated errors is assimilated. In the terminology of DA, this is commonly known as “thinning”.

In our synthetic experiment, we obtained, every 12 h, the gridded stages of the “true” simulation of the flood event. Note again that in real cases, satellite-based SAR samples only provide water elevations along the flood edge. Then, we used the thinning procedure described by Mason et al. (2012b), which itself is based on Ochotta et al. (2005). This uses a Moran’s I test to ensure that the samples remaining after thinning have no spatial autocorrelation at the 5% significance level, so that zero covariance can be assumed between observations. We assumed a standard deviation for the thinned set of WLOs of 0.25 m, with zero mean bias. This is the mean error in the estimate of water surface height at the final sampling points. These values are realistic as they were obtained from WLOs extracted from real high resolution SAR images of flood extents observed over this domain, and the absence of bias was validated with DEM-independent gauge-based WLOs (Mason et al., 2012b). The standard deviation includes components due to uncertainties both in the position of the SAR flood waterline and in the DEM. Height errors due to waterline position uncertainty were reduced by selecting waterline samples in flooded areas on low DEM slopes. The base $\Delta t = 12$ h WLO dataset was used to create the various first-visit-time and revisit-time ($\Delta t_a = 12$ h, 24 h, and 48 h) scenarios.

2.4. Ensemble Filter

Unknown parameters can be estimated as part of the data assimilation by using state space augmentation (Friedland, 1969). As the model state is augmented with model parameters, correlations develop between the parameters and the model variables. In data assimilation schemes using such an approach, the analysis updates an augmented state vector,

$$\mathbf{x} = \begin{bmatrix} \mathbf{z} \\ \boldsymbol{\beta} \end{bmatrix}, \quad (1)$$

where \mathbf{z} is the n_s -dimensional model state and $\boldsymbol{\beta}$ is a generic n_β -dimensional vector of parameters. Thus \mathbf{x} is the augmented n -dimensional state vector, with $n = n_s + n_\beta$. Here, we follow this approach with an ensemble representation, where our parameters are the inflow errors at the assimilation time. Then, in our case, after each assimilation step, the updated \mathbf{z} (an ensemble of water stage grids) evolves by integrating each member of the ensemble forward in time with the LISFLOOD-FP model, and, independently, the updated ensemble of inflow errors evolves in time according to our error forecast model (described in Section 2.5.2).

The Kalman filter equations (Kalman, 1960) to update the state vector in a linear system are:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}^f), \quad (2)$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f, \quad (3)$$

where the forecast (prior) and analysis (posterior) quantities are denoted by the superscripts f and a , respectively; $\mathbf{y} \in \mathfrak{R}^p$ is the vector of observations; \mathbf{H} is the $p \times n$ observation matrix (or ‘‘observation’’ or ‘‘forward’’ operator) mapping the state vector to the observation space; \mathbf{P} is the $n \times n$ state error covariance matrix; \mathbf{I} is the $n \times n$ identity matrix; and \mathbf{K} is the $n \times p$ Kalman gain matrix:

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1}, \quad (4)$$

where the superscript ‘‘T’’ denotes matrix transposition, and \mathbf{R} is the $p \times p$ observation error covariance matrix.

The Ensemble Kalman Filter (EnKF) was introduced by Evensen (1994). It makes it possible to apply the Kalman filter to high-dimensional discrete systems, when the explicit storage and manipulation of the system state error covariance is impossible or impractical. The EnKF methods may be characterized by the application of the analysis equations given by the Kalman filter to an ensemble of forecasts.

Let $\{\mathbf{x}_i\}$ ($i = 1, \dots, m$) be an m -member ensemble $\in \mathfrak{R}^n$. The ensemble mean is the vector defined by

$$\bar{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i. \quad (5)$$

The ensemble perturbation matrix for the augmented state is the $n \times m$ matrix defined by

$$\mathbf{X} = \frac{1}{\sqrt{m-1}} [\mathbf{x}_1 - \bar{\mathbf{x}} | \mathbf{x}_2 - \bar{\mathbf{x}} | \dots | \mathbf{x}_m - \bar{\mathbf{x}}]. \quad (6)$$

The ensemble covariance matrix \mathbf{P} is typically assumed to be carried by the ensemble by means of the relation

$$\mathbf{P} = \mathbf{X}\mathbf{X}^T = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \quad (7)$$

In addition to linear problems, the EnKF-based methods have been applied to nonlinear problems, as is our case. In general, the (possibly nonlinear) observation operator H maps state variables \mathbf{x} to observed variables \mathbf{y} :

$$\mathbf{y} = H(\mathbf{x}^t) + \boldsymbol{\epsilon}, \quad (8)$$

where the superscript t denotes the true state, and $\boldsymbol{\epsilon}$ results from a combination of instrument error and representation error. The covariance $\langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T \rangle$ is the abovementioned \mathbf{R} .

Within the EnKF methods, the Ensemble Square Root Filters (ESRF) is a family in which both the ensemble mean and the ensemble perturbations are updated explicitly, so that the analysis error covariance matches the theoretical value given by the Kalman filter (Tippett et al., 2003). The ETKF is a specific formulation within the ESRF. The ensemble mean is updated by using the analysis equation (2), and the ensemble perturbations are updated by an ensemble $m \times m$ transform matrix \mathbf{T} :

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{T}, \quad (9)$$

such that the analyzed covariance matrix obtained from \mathbf{X}^a using (9) matches the theoretical value (3):

$$\mathbf{X}^f \mathbf{T} (\mathbf{X}^f \mathbf{T})^T = (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{X}^f (\mathbf{X}^f)^T. \quad (10)$$

The forecast ensemble perturbation matrix, for a linear \mathbf{H} , is $\mathbf{Y}^f = \mathbf{H}\mathbf{X}^f$. By defining the matrix

$$\mathbf{D} = \mathbf{Y}^f (\mathbf{Y}^f)^T + \mathbf{R}, \quad (11)$$

and using (4), (10) can be rearranged as

$$\mathbf{T}\mathbf{T}^T = \mathbf{I} - (\mathbf{Y}^f)^T \mathbf{D}^{-1} \mathbf{Y}^f, \quad (12)$$

whose right-hand side may be rewritten using the Sherman-Morrison-Woodbury identity (Tippett et al., 2003, equation (15)) as

$$\mathbf{I} - (\mathbf{Y}^f)^T \mathbf{D}^{-1} \mathbf{Y}^f = \left(\mathbf{I} + (\mathbf{Y}^f)^T \mathbf{R}^{-1} \mathbf{Y}^f \right)^{-1}. \quad (13)$$

Since $(\mathbf{I} + (\mathbf{Y}^f)^T \mathbf{R}^{-1} \mathbf{Y}^f)$ is a symmetric positive definite matrix, whose eigenvectors are equivalent to the eigenvectors of $(\mathbf{Y}^f)^T \mathbf{R}^{-1} \mathbf{Y}^f$,

$$\mathbf{T}\mathbf{T}^T = \mathbf{C}(\mathbf{I} + \boldsymbol{\Gamma})^{-1} \mathbf{C}^T, \quad (14)$$

$$= \left(\mathbf{C}(\mathbf{I} + \boldsymbol{\Gamma})^{-1/2} \right) \left(\mathbf{C}(\mathbf{I} + \boldsymbol{\Gamma})^{-1/2} \right)^T, \quad (15)$$

where the columns in \mathbf{C} contain the orthonormal eigenvectors of $(\mathbf{Y}^f)^T \mathbf{R}^{-1} \mathbf{Y}^f$, and the diagonal matrix $\boldsymbol{\Gamma}$ contains the corresponding eigenvalues. This provides the solution

$$\mathbf{T} = \mathbf{C}(\mathbf{I} + \boldsymbol{\Gamma})^{-1/2}, \quad (16)$$

which is the ‘‘one-sided’’ solution given by Bishop et al. (2001). Clearly, any orthogonal matrix \mathbf{U} (i.e. $\mathbf{U}\mathbf{U}^T = \mathbf{I}$) can be attached to the right-hand side of (14) to provide an alternative solution. Specifically, being \mathbf{C} orthogonal, one solution is

$$\mathbf{T} = \mathbf{C}(\mathbf{I} + \boldsymbol{\Gamma})^{-1/2} \mathbf{C}^T, \quad (17)$$

which is called the “symmetric solution” by Ott et al. (2004) or the “spherical simplex” solution by Wang et al. (2004), and is also equivalent to the Local Ensemble Transform Kalman Filter (LETKF) solution given by Hunt et al. (2007) in the case without localization. The solution (17) is unbiased (Livings et al., 2008; Sakov & Oke, 2008), and is the solution adopted in this study.

The state space of our model is a water stage grid. We simply use a linear mapping \mathbf{H} from the state space into the SAR-derived WLOs by locating the inundated, and with “running water”, grid point closest to each individual observation. Thus, for each observation, the stage of this closest grid point is mapped, with a weight equal to 1, while remaining grid points have a weight equal to 0. \mathbf{H} is thus a sparse matrix containing 1s and 0s. The “running water” criterion refers to pixels whose water depth is above a threshold (1 mm in this study) considered as surface depression storage, below which water is not routed, and the pixel becomes hydraulically disconnected from the main flooded area.

2.5. Ensemble Generation

2.5.1. Perturbation to model inputs

The performance of most ensemble forecasts is influenced by the quality of the ensemble generation method, the forecast model, and also the analysis scheme. The perturbation of the forcing data to generate an ensemble of forecasted model state vectors is a key feature in the EnKF family. Here we assume that the model is free of structural errors and parameter uncertainty, so that all model errors arise from forcing data, i.e. input flow boundary conditions. At gauged points, errors in streamflows stem both from measurement errors in water level measurements and uncertainties in the rating curves (stage-flow relationships). It is acknowledged that errors in flow measurements are heteroscedastic (proportional to flow), and a number of approaches have been proposed to generate the error ensemble for the inflow boundary conditions into hydrodynamic models. On the other hand, errors attributed to missing lateral flow inputs through the domain boundary, not accounted for in the point flow boundary conditions, are not necessarily related to flow measurements.

For DA studies, several authors have perturbed the input forcing of a hydrologic model to obtain an ensemble of inflows into the hydrodynamic domain. In this way, Andreadis et al. (2007) used the VIC model with perturbed precipitation fields, and included a negative bias of 25% to the VIC simulated flows. Similarly, Matgen et al. (2010) and Giustarini et al. (2011) used the CLM hydrologic model, the former including a positive 25% bias to the CLM generated hydrographs, and the latter without adding any bias. Biancamaria et al. (2011) used Empirical Orthogonal Functions (EOF), following the methodology developed by Auclair et al. (2003), to perturb the most statistically significant modes of precipitation and temperature fields as input to the ISBA model, whose ensemble hydrograph output drove the hydrodynamic model LISFLOOD-FP. However, the statistics of the final inflow perturbations into the hydrodynamic model are not evident in these studies. For studies focused on DA within the hydrodynamic model it is useful to have a clear view of final inflow perturbations, as it is the errors in the hydrodynamic model and their value relative to the observation errors that determine the weight given to observations in the DA analysis.

In essence, from the point of view of the generation of the inflows for hydrodynamic models, and domains with a number of tributaries and boundary conditions, we could pose two general scenarios: a) input flows from real gauge observations, and b) input flows forecast by a hydrologic model. In both cases, the error evolution at each inflow will have some degree of temporal autocorrelation. On the other hand, scenario (a) should not show a significant correlation, if any,

between the errors at the various gauge locations, as errors in stage measurements and uncertainties in rating curves are normally independent between sites. In contrast, scenario (b) will generally introduce a, normally high, spatial correlation between the errors at the various inflows. The degree of this spatial correlation will be highly dependent on the perturbation of the forcing—chiefly precipitation fields—onto the hydrologic model, and the hydrologic model structure and parameters. For complex hydrodynamic domains, this distinction is key, as it will govern the development of the correlations in the state vector, and the general DA behaviour. Existing spatial correlation between boundary errors in different tributaries may well lead to one WLO (either from remote sensing or a standard gauge) at the head of one tributary to influence the error estimation at the others. This may be, especially for sparse observations (as is common for stage gauges), a positive DA outcome in linked hydrologic-hydrodynamic models as, in general, it will make the observations more influential in both correcting the hydrodynamic state and, possibly, correcting the hydrologic model errors. However, for scenario (a), if spatial correlations between inflow errors are erroneously assumed, or are developed as spurious in Monte Carlo-based methods (e.g., due to limited ensemble size), the DA updates could lead to biased error estimates.

In the current study, we evaluate a flood scenario with available real inflow measurements at the major tributaries. With this dataset, scenario (a) can be simulated by generating spatially-independent time-autocorrelated (and heteroscedastic) random errors to perturb measured inflows, and scenario (b) can be simulated by incorporating a spatiotemporal autocorrelation into the heteroscedastic errors.

With the number of operational gauges actually declining in the world (Vörösmarty et al., 2001), and considering that a linked hydrologic-hydrodynamic model should lead to increased flood forecast lead times, we choose scenario (b) for the remainder of this study. This approach has an advantage over selecting a specific hydrologic model in that it can be regarded as using a “generic” hydrologic model whose influence in generating inflow boundary conditions is explicitly modelled and known. This clarifies the analysis for our study.

Below, within Section 2.5.2, we detail how we simulate the inflow ensembles with random errors. As an example of the difference between the scenarios (a) and (b) when used with the described ETKF, Fig. 2 shows, for the study domain, the covariance between the inflow errors at Bewdley and water stage in the domain, after 5 forecast/DA steps, for both scenarios, for a revisit time $\Delta t_a = 24$ h. In both simulations the state vector is augmented with the inflow errors, which are updated every DA step. Clearly, for the case (b) where a spatial covariance is imposed between the errors at the various inflows, a positive cross-covariance develops not only between the state variable (stage) at the various tributaries, but also between inflow errors and stages at locations which are quite separate. In both cases the covariance is propagated in the flow direction downstream (for a sub-critical flow the covariance could also propagate upstream), while the covariance development is counterbalanced by the effect of the temporal decorrelation in the errors.

In real cases, the temporal characteristics of the errors in an ensemble of hydrologic model forecasts are not known. Here we simulate these errors by imposing a deterministic error, as a multiplicative bias (20%) to all flow measurements, and then by adding spatiotemporally correlated random errors to perturb the biased mean input flow to obtain the input inflow ensemble, which drives the hydrodynamic model. The imposed bias is reasonable given knowledge of real world discharge errors. It is noted that a bias is a particular case of error dynamics between the true hydrograph and the mean of the hydrograph ensemble that could be generated by a hydrologic model. In the real and general case, this mean error will not be stationary in time. Previous studies

have indicated that the improvement in forecasting skills due to assimilation of observations may have a short time span in hydrodynamic domains, as the inflow errors propagate downstream counterbalancing the improvement. So, the inflow errors estimated at the assimilation time can be used to correct boundary inflows until the next assimilation time, increasing the persistence of error reductions between times of observations (e.g., Andreadis et al., 2007; Matgen et al., 2010). In real cases, the assumptions one can make about the evolution of this error between consecutive assimilation steps, as well as the available information, should lead the design of the error forecast model used for correcting inflow errors. Thus, Andreadis et al. (2007) used an first order autoregressive model (for a 3-month case study), and Matgen et al. (2010) assumed stationarity of the estimated errors between assimilation steps (for a storm-flood study), so they used the estimated error to constantly correct all the inflows until the next assimilation time. Realistically, for storm-flood event durations (the focus of our interest), if no discharge data is available, satellite revisit times will make it very difficult to operationally implement more complex approaches than the latter. In any case, the deviations of the error forecast model from the real error dynamics will diminish the improvement due to the estimation and on-line correction of the inflow errors. In our synthetic case, we are imposing stationarity (bias) as “true” mean error dynamics. Accordingly, our error forecast model should not assume stationarity of the mean error between assimilation times, as the results would provide an overoptimistic correction of the inflow bias with respect to what can be realistically expected. Thus, to emulate this problem, our error forecast model is a decay toward 0 between consecutive assimilation times, as described below.

2.5.2. Simulation of heteroscedastic model errors

We now describe the procedure we follow to simulate spatiotemporally correlated heteroscedastic inflow errors, i.e. scenario (b) described above. This provides us with a time-evolution of the inflow errors, with which we augment the state vector (Section 2.4).

At a specific time, the covariance matrix $\Sigma \in \mathfrak{R}^{n \times n}$ can be decomposed as $\Sigma = \mathbf{D}^{1/2} \boldsymbol{\rho} \mathbf{D}^{1/2}$, where $\boldsymbol{\rho}$ is the $n \times n$ correlation matrix, and $\mathbf{D}^{1/2}$ is an $n \times n$ diagonal matrix containing the square root of the diagonal values of Σ (i.e. the standard deviations of the n state variables). By using this decomposition, it is possible to generate an $n \times m$ matrix \mathbf{A}' containing a pseudo-random sample of size m from $\mathcal{N}(\mathbf{0}, \Sigma)$, each sample in a column, through

$$\mathbf{A}' = \left(\mathbf{d}^{1/2} \mathbf{1}^T \right) \circ \left(\boldsymbol{\rho}^{1/2} \mathbf{A} \right), \quad (18)$$

where $\mathbf{d}^{1/2} \in \mathfrak{R}^n$ is a column vector containing the diagonal values of $\mathbf{D}^{1/2}$, $\mathbf{1} \in \mathfrak{R}^m$ is a column vector with all elements equal to 1, the symbol \circ represents the Schur (element-wise) product, and \mathbf{A} is an $n \times m$ matrix of pseudo-random numbers drawn from $\mathcal{N}(0, 1)$. Thus we can detach the process of spatiotemporal correlation (carried on by $\boldsymbol{\rho}$) from the process of variance scaling (expressed by \mathbf{d}), in order to generate the inflow errors. Heteroscedasticity may be embedded in \mathbf{d} . For this, we need to choose a model for the spatiotemporal correlation and a model for the heteroscedasticity.

We are interested in simulating errors in a number $n_{\partial\Omega}$ of spatially sparse inflow boundary conditions. A way of simulating an evolving multivariate time series of spatiotemporally correlated errors is through

$$\mathbf{q}_k = \alpha \mathbf{q}_{k-1} + \sqrt{1 - \alpha^2} \mathbf{w}_k, \quad (19)$$

where $\mathbf{q}_k \in \mathfrak{R}^{n_{\partial\Omega}}$ is the model error at time k , and $\alpha \in [0, 1]$ is a temporal autocorrelation coefficient

(Evensen, 2003). $\mathbf{w}_k \in \mathfrak{R}^{n_{\partial\Omega}}$ is a here white noise obtained by drawing samples from $\mathcal{N}(\mathbf{0}, \boldsymbol{\rho})$, where $\boldsymbol{\rho} \in \mathfrak{R}^{n_{\partial\Omega} \times n_{\partial\Omega}}$ is a distance-dependent correlation matrix (described below in Section 2.5.4).

Let $\{\mathbf{q}_{ki}\}$ ($i = 1, \dots, m$) be an m -member ensemble $\in \mathfrak{R}^{n_{\partial\Omega}}$ of inflow errors at time k . The ensemble matrix for the inflow errors is the $n_{\partial\Omega} \times m$ matrix defined by

$$\mathbf{Q}_k = [\mathbf{q}_{k1} | \mathbf{q}_{k2} | \dots | \mathbf{q}_{km}], \quad (20)$$

where each member of the ensemble \mathbf{Q}_k has evolved individually according to (19). This ensures that the diagonal of the covariance matrix of the ensemble \mathbf{Q}_k is made (approximately) of 1s as long as this is also true for \mathbf{Q}_{k-1} . In this way, we use the stochastic process defined by (19) to generate the spatiotemporally correlated errors in a normalized space, previous to the consideration of heteroscedasticity (i.e. \mathbf{Q}_k is analogous to $\boldsymbol{\rho}^{1/2} \mathbf{A}$ in (18)). After assimilation steps, errors are regenerated ($k = 1$). So $\mathbf{Q}_{k-1} \equiv \mathbf{Q}_0$ refers to the errors updated by the assimilation process. With this formulation, α being an scalar, we are assuming that the temporal autocorrelation dynamics of the errors are similar for all inflows.

Then, we account for heteroscedasticity in a later step. Let s_k be an arbitrary scalar, obtained through a measurement or a forecast, at time k , which is taken as the expected value of the scalar random variable of interest ($E[S_k] = s_k$). The variance $V[S_k] = \sigma_k^2$ can be assumed to be proportional in some form to s_k . For flow errors, we propose a general model to this proportion as

$$\sigma_k^2 = \sigma_0^2 \left(\frac{s_k}{s_0} \right)^h, \quad (21)$$

where σ_0^2 is a reference variance corresponding to a reference value s_0 , and h is a heteroscedasticity factor. If $h = 0$, errors are homoscedastic as σ_k^2 becomes independent of s_k . If $h = 1$, $\sigma_k^2 = \left(\frac{\sigma_0^2}{s_0} \right) s_k$, where the term in parentheses matches the so-called ‘‘hyper-parameter’’ proposed by Moradkhani et al. (2005). If $h = 2$, $\sigma_k = \left(\frac{\sigma_0}{s_0} \right) s_k$, where the term in parentheses is by definition a coefficient of variation (c.v.). In this study, we set $h = 2$, so heteroscedasticity is expressed by a common coefficient of variation, and set c.v.= 0.15 for the hydrographs of our virtual hydrologic model. This value is derived from historic observations for the rating curve calculation at Saxons Lode Us (c.v.~ 0.10), slightly increased to emulate the normally higher errors from hydrologic models. As a comparison, Clark et al. (2008) used a c.v.= 0.10 for a white noise perturbing flow observations to be assimilated into a hydrologic model.

Thus, analogous to (18), we obtain the heteroscedastic error ensemble ($n_{\partial\Omega} \times m$ matrix) \mathbf{Q}'_k as

$$\mathbf{Q}'_k = (\boldsymbol{\sigma}_k \mathbf{1}^\top) \circ \mathbf{Q}_k, \quad (22)$$

where $\boldsymbol{\sigma}_k \in \mathfrak{R}^{n_{\partial\Omega}}$ is the column vector of the standard deviations of the inflow errors, and $\mathbf{1} \in \mathfrak{R}^m$ is a column vector with all elements equal to 1. \mathbf{Q}'_k is the ensemble representation of the inflow errors, with which we augment the ensemble representation of the state vector \mathbf{z} at the time of assimilation. This is analogous to the generic parameter augmentation denoted by $\boldsymbol{\beta}$ in equation (1). It follows from (19) and (21) that the marginal distribution of the errors at each inflow is

$$q'_k \sim \mathcal{N}(0, \sigma_k^2). \quad (23)$$

Also, from (19) and (21), note that by using a constant time step between two arbitrary steps i and j , the covariance at a specific location between q'_i and q'_j is

$$\overline{q'_i q'_j{}^\top} = \alpha^{|i-j|} \sqrt{\left(\frac{s_i s_j}{s_0^2} \right)^h} \sigma_0^2. \quad (24)$$

After an assimilation step is conducted, the q'_0 ensemble at each inflow is the result from an updating together with the other variables in the state vector, and will generally deviate from both the mean and the variance given by (23). However, in time, both the mean and the variance of the newly simulated forecast errors will converge to these values, and this will occur faster for lower α values.

2.5.3. Determination of α

The factor α should be related to the real time step used and a specific time decorrelation length τ . The decay term in (24) can be also expressed as an exponential decay:

$$\alpha^{|i-j|} = e^{-\frac{\Delta t}{\tau}}, \quad (25)$$

which relates α and τ , and clarifies that, disregarding the heteroscedastic variance term in (24), the covariance in time between q'_i and q'_j is damped by a ratio e^{-1} over a time period $\Delta t_{ij} = \tau$ (see Evensen, 2003). For a specific time step k of length Δt_k , then

$$\alpha_k = e^{-\frac{\Delta t_k}{\tau}}, \quad (26)$$

which allows one to use (19) for any time step length by substituting α by the corresponding α_k , and, instead of (24), the error covariance, at each inflow, between any two time steps (i, j) is more generically expressed as

$$\overline{q'_i q'_j{}^T} = \sqrt{\left(\frac{s_i s_j}{s_0^2}\right)^h} \sigma_0^2 \prod_{k=i+1}^j \alpha_k. \quad (27)$$

2.5.4. Spatial correlation model for inflow errors

The spatial correlation matrix $\boldsymbol{\rho}$, for generation of the white noise \mathbf{w}_k in (19), can be created by any procedure which considers that correlation in inflow errors is dependent on the distance between the locations of the point inflow boundary conditions. Here we chose the Gaussian-decay correlation model

$$\rho_{ij} = e^{-\frac{1}{2}\left(\frac{d_{ij}}{\theta}\right)^2}, \quad (28)$$

where the subscripts i and j refer to any two boundary conditions, $\rho_{ij} \in [0, 1]$ is the corresponding spatial correlation and element in $\boldsymbol{\rho}$, d_{ij} is the distance between the corresponding locations, and θ is a spatial correlation coefficient.

2.5.5. Selection of τ and θ and inflow error estimation

As abovementioned (Section 2.5.1), the true dynamics of the mean error of measured or forecasted inflows are unknown in real cases. In this synthetic study we impose a deterministic stationary bias as a “true” mean error evolution, and we approach the DA problem as if we did not know about this error evolution to evaluate how it influences the forecast, and how DA is able to partially solve for it. To emulate errors from a “generic” hydrologic model, we first imposed a positive 20% bias on measured inflows. Then we perturbed the biased inflows with spatiotemporally correlated errors to generate the inflow ensemble. Generally, errors in precipitation inputs, and hydrologic model parameters and structures can generate a wide range of possible spatiotemporal correlations in the simulated hydrographs. Thus, two single values of τ and θ cannot embrace all possible

situations. Here, our parameters for the error forecast model were $\tau = 3$ days and $\theta = 62000$ m (e.g., the spatial correlation for the inflow errors between Bewdley and Evesham is 0.8). Despite being arbitrary, we chose these values as we believe they are representative of a relatively normal situation with a spatially distributed or semidistributed model, making use of continuous rainfall field inputs, and having undergone a certain degree of calibration with previous events. Fig. S1, in the supplementary material, shows a hypothetical example of the error forecast evolution, after one assimilation step, for two values of τ . In this study, as we have imposed a stationary bias in the true mean error, higher values of τ , will lead to better results, as they will exert a more persistent correction of the bias. So, the intentional mismatch between the error forecast model and the stationary bias serves to emulate the lack of knowledge of the mean error evolution in real cases. On the other hand, for a real case, the error forecast model should try to approach the real error dynamics; either by the parsimonious assumption of stationarity (e.g., Matgen et al., 2010), or by more complex models.

2.6. Verification Methods

To assess the strength and weaknesses of the forecasts, we use standard verification methods. The Root Mean Square Error (RMSE) is used as measurement of overall accuracy. The Brier Skill Score (BSS) is used to evaluate the forecast relative to a standard, which is chosen to represent an unskilled forecast. In our case, the unskilled forecast is the open loop simulation. The vectorized form of the BSS is

$$\text{BSS} = 1 - \frac{\overline{(\mathbf{f}_s - \mathbf{o})^2}}{\overline{(\mathbf{f}_r - \mathbf{o})^2}}, \quad (29)$$

where \mathbf{f}_s is the evaluated forecast state vector, \mathbf{f}_r , is the reference forecast (open loop) vector, \mathbf{o} is the actual outcome vector (here, the truth), and the overline denotes the average. The BSS $\in (-\infty, 1]$, where BSS= 0 indicates no skill when compared to the reference forecast, and BSS= 1 is a perfect score.

Finally, we use rank histograms for determining the reliability of ensemble forecasts and for diagnosis of errors in its mean and spread. A flat rank histogram is usually taken as a sign of reliability. A detailed interpretation of rank histograms for verifying ensemble forecasts is given by Hamill (2001).

3. Results and Discussion

3.1. Updating Inflow Boundary Conditions

Our results indicate that the improvement in forecasting skill due to assimilation of observations may have a short time span in hydrodynamic domains, as the inflow errors propagate downstream counterbalancing the improvement. This is in agreement with previous studies (e.g., Andreadis et al., 2007; Matgen et al., 2010; Giustarini et al., 2011). However, it is also important, in this context, to evaluate how the inflows are corrected at the boundary conditions themselves, as this is an indicator of the capability of the data assimilation scheme to obtain inflow time series that can be used as surrogate observations to feed back into an inverse hydrodynamic-hydrologic DA modelling cascade.

Fig. 3 compares the evolution of the inflow ensemble at the upstream boundary condition at Bewdley and the forecasted flood stage at a downstream location (Worcester) when inflow errors are not estimated and corrected by the assimilation against the case when they are corrected.

Simulations refer to a SAR assimilation revisit time $\Delta t_a = 24$ h. If inflows are not updated they are similar to an open loop without DA, so the DA-bias line overlies the input bias one (Fig. 3a). If inflow errors are also estimated and corrected according to the used error forecast model, each sequential assimilation pushes the inflows used by the model toward the truth (Fig. 3b). In both cases, the DA process does a good job in correcting the forecast toward the truth at Worcester. For each ensemble, this is clarified by the upper plots at Worcester, which show the evolution of the standard deviation (DA-SDev lines) and the mean bias (DA-bias lines) between the forecast and the truth. However, if the biases in the inflow (here mostly influenced by Bewdley at the North) are not corrected they have a control effect that, after any assimilation update, causes the forecast to drift away from the truth, leading to an early overestimation of the flood stage. A similar effect was shown by Matgen et al. (2010). The case with inflow updating keeps the forecast on track very close to the truth. Curves at the other inflows and sampled forecast locations show similar effects (see Figs. S2–S15; supplementary material). The speed at which the updated inflows drift away from the truth when they are updated is related to the lack of match between the used error forecast model (with $\tau = 3$) and the imposed stationary bias. As described in Section 2.5, in this case, higher τ values would result in a more persistent propagation of the errors estimated at the assimilation time, giving an improved mean inflow error estimation and correction in time with the forecast. In the remainder of this paper we use simulations with updating of the inflow errors, as this leads to a clear forecast improvement. However, we keep $\tau = 3$ to emulate the fact that any error forecast model that could be chosen for real cases (e.g., a stationary bias model as Matgen et al., 2010) will always fail to completely match the true (non-stationary) inflow error evolution. Here we assumed that friction parameters are known. In real cases, if friction in the channels and floodplain are considered to be uncertain, an attempt may be done to estimate them simultaneously by additional augmentation of the state vector. Generally, with additional parameters to be estimated, the filter would benefit from larger ensemble sizes. Estimation of friction, however, is beyond the scope of this study.

3.2. Ensemble Properties

The use of a finite ensemble size to approximate the error covariance matrix introduces sampling errors that are seen as spurious correlations. With each spurious update there is an associated reduction of ensemble variance. This ensemble collapse problem is present in all EnKF applications and can lead to filter divergence (Evensen, 2009). To the authors’ knowledge, there is no published study that evaluates the problem of ensemble collapse for hydrologic or hydrodynamic studies using sequential EnKF-based DA. Let us conduct a quick examination of the properties of the ensemble, taking as an example a simulation with $\Delta t_a = 24$ h revisit time, starting on the 20th of July, before the flood goes out of bank. Fig. 4 shows the evolution of the rank histograms evaluated with the forecasted ensemble at each assimilation time. To build the histograms, at least $5 \times m$ locations as evenly distributed as possible within the flooded area were taken at each visit time. These locations were used to sample from the truth and the forecasted ensemble, and the truth was ranked within the ensemble. In general, rank histograms in Fig. 4 appear different from uniform and the density concentrates around values lower than the rank mean as a result of the bias imposed on the simulated inflows. Clearly, the spread starts being too high, resulting from the coefficient of variation (c.v.= 0.15) we gave to the inflow errors in relation with the observation errors, representing the lower trust we have in the hydrologic model output. After that, as the event evolves, the variance decreases and the spread becomes more adequate. Nevertheless, the ensemble collapse is moderate, and the ensemble remains relatively stable along

with the sequential assimilation steps. This indicates that the ensemble size is enough, in general terms, for the case study. Here we use an ensemble size $m = 210$ for a state vector length of the order $\mathcal{O}(10^4)$. This is relatively high as compared with some available studies oriented to operational uses. For example, Houtekamer et al. (2009) used $m = 64$ for a Numerical Weather Predictions (NWP) model with a state vector length of the order $\mathcal{O}(10^7)$. Note, however, that these highly-dimensional operational NWP problems use methods, known as inflation and localization, for minimizing the impact of the spurious updates (e.g., Houtekamer & Mitchell, 2001; Hamill et al., 2001). Localization reduces the problems generated by reduced ensemble sizes by decreasing the weight, through several approaches, given to observations far from the estimated state variable (as the subspace in which the analysis is conducted is reduced). Inflation may be applied to either the background covariance or the analysis covariance during each assimilation cycle, and several multiplicative and additive inflation techniques have been proposed. For example, see the review in Hunt et al. (2007). In our study, despite ensemble collapse being slight, an attempt could have been made to further compensate it through inflation or localization. However, the spread is influenced by the assimilation interval and the start time of the assimilation (first visit). Thus we kept the same coefficient of variation as a general adequate value and did not apply any inflation/localization in order to make comparison among several visit scenarios straightforward.

3.3. Sensitivity to First Visit and Revisit Times

In a, normally budget-limited, operational context the decision to task a satellite to acquire SAR images needs to take into account the first visit and revisit times. Here we focus on these parameters. There are now a number of sensors acquiring SAR data (RADARSAT, TerraSAR-X, ALOS PALSAR, Cosmo-SkyMed, etc.), controlled by a number of different space agencies. Issues such as the difference in time between the acquisition of the SAR images and the time at which they are available to the user (the information age) should be considered for any current or future satellite mission. Note also that the assimilation strategy may well include satellite information from different satellites, not pertaining to the same constellation. These operational issues, as well as issues of data quality specific to each sensor are beyond the scope of this paper, which aims to be relatively generic. Fig. 5 shows the RMSE of the ensemble mean with respect to the synthetic truth, evaluated at specific locations. Each plot shows a family of three curves, which refer to the revisit/DA times $\Delta t_a = 12$ h, 24 h, and 48 h. For each curve, each point represents the RMSE for the assimilation scenario whose first visit is at the point time, and the RMSE is calculated between the mean of the ensemble and the truth over the entire time window. For example, for all $\Delta t_a = 24$ h curves in all Fig. 5 plots, the first point results from the simulation referred to by Fig. 3b and Fig. 3d. Note that the later a first visit occurs the less visit/DA steps are conducted, and all the statistics necessarily converge towards those from an open loop simulation. It is noted that satellite-based SAR acquisitions provide reliable water elevation only at locations when the flow is overbank. So, at the start of the rising limb, when flow is still within the river banks, it is highly unlikely to get valuable remote sensing observations. The RMSE shown at the two inflow boundary conditions, Bewdley and Evesham, indicates how state vector augmentation with boundary inflows is able to estimate and correct the inflow error. The forecast stage at Worcester is mostly controlled by the inflow at Bewdley, but also has some contributions from Kidder Callows and Hardford Hill. Just after the junction between the Teme and the Severn, Kempsey also includes the inflow from Knightsford Bridge. The forecast at Bredon depends on inflows from the Avon and its tributaries, i.e. Evesham, Hinton and Besford Bridge inflows.

Generally, for inflows and stage, the improvement due to the decrease of the revisit time is most clear when assimilation starts at an early stage of the flood event. After the peak stage is reached, from 22th July onwards, the curves have mostly converged. Also, for each Δt_a curve, the increase in the RMSE at the forecasted stages is very sharp just before the peak stage is reached, that is, when variation in stage is higher. This indicates that the early satellite overpasses on the rising limb, provided WLOs can be extracted from them, are the most useful.

If the observations are too early in relation to the arrival of flow, they do not provide useful information. However, an early overpass may simultaneously have observations of very low usefulness at downstream areas and much more valuable observations upstream if the flow is significant. For example, the assimilation of the 20th July overpass WLOs has a negligible benefit for the forecast at Mythe Bridge for the $\Delta t_a = 12$ h and 24 h revisit times. This is shown as the corresponding simulations with 21th as first visit have very similar RMSE values. Later, the 21th overpass has a very significant influence on the RMSE at Mythe Bridge, despite the flow being still very low at that point at the overpass time. But this is because at that time, upstream flow is much higher (e.g., around Bewdley). The benefit of assimilating upstream observations at that time is propagated downstream and has time to be highly influential at Mythe Bridge. On the other hand, at the same location, for the $\Delta t_a = 48$ h revisit time, a first visit at 20th July is very useful. But again this is not because of the observations around Mythe Bridge at that time, but because the benefit of observations being assimilated upstream has time to reach this point before the next first visit time is evaluated for this curve (22th July). The revisit scenario curves have generally converged after the 22th–23th July, which implies that the differences in the RMSE between the starting points for the three curves result from the improvement owing to the increased observation frequency during the rising limb. Interestingly, for the inflow at Evesham and the stage forecast at Bredon, when the first visit time is between the 22th and the 30th July the curves with $\Delta t_a = 48$ h, show better statistics than those with higher observation frequencies. While the differences are not big, these are related to the dynamics of the inflow at Evesham, which includes a secondary peak during the 26th–28th July. Minor modifications in the spread and the covariance create these differences.

The improvement in the forecast with respect to the open loop is indicated by Fig. 6, which shows the evolution of the Brier Skill Score (BSS) for the simulation with $\Delta t_a = 24$ h revisit time starting from the 20th of July. Fig. 6 corresponds to the simulation represented by the first point in the $\Delta t_a = 24$ h curves in Fig. 5, and also referred to by Fig. 3b and Fig. 3d). The BSS evolution is calculated against the open loop for forecast times $t+6$ h, $t+12$ h, and $t+24$ h, where t refers to revisit times. The BSS is calculated for the seven inflow boundary conditions (Fig. 6a), and for the forecast stages at the ten reference gauge locations (Fig. 6b). Note that after each assimilation step, the updated inflow boundary conditions evolve without using the hydrodynamic model, but according to the corrections made by the assimilation and the error forecast model to the inflows with imposed biases. Thus Fig. 6a mostly displays the inversion capability of the filter; but this capability is also influenced by the model structure, which develops the covariances between the inflows and the state vector (stage), used by the filter when SAR WLOs are assimilated.

For both inflows and forecasted stages, the BSS is very high throughout all the simulation. As expected, the BSS is generally better for the $t+6$ h forecast time, as the inflow corrections are partially lost for increasing forecast times, when the updated inflows drift back to the biased inflows, as shown by Fig. 3b. Still, the BSS for inflows has a very mild and stable decreasing trend along the simulations. This seems to match the decreasing trend in spread indicated by

the rank histograms. For the forecast stage, disregarding the 20th July forecast when flows are still very low and some observations are too early to be useful, the minimum values, given for the 25th and 26th July, are still very high ($\sim 0.85\text{--}0.90$). After that there is a recovery in the BSS. At the start of the simulation the inflow-stage covariance (e.g., Fig. 2b) fades faster downstream from the inflows than at a later time, and so this results in the updating for inflows being more influenced by the early local observations. As the flood event evolves, so do the covariances, and more SAR WLOs generally farther from the boundary conditions are affecting inflow updatings. So, some nonlinearities, and perhaps the development of spurious correlations, are likely affecting the updatings for the lower BSS values.

Finally, a budget-limited scenario is shown by Fig. 7, which considers 5 satellite overpasses with $\Delta t_a = 24$ h revisit time and successively delayed by one day. The RMSE is also calculated over the entire window. The RMSE patterns are very similar to those in Fig. 5, with generally increasing RMSE as the observations are delayed. This indicates that even considering a fixed number of satellite overpasses, the benefit of the DA still propagates downstream after all observations are completed. This is a similar effect to that shown by Biancamaria et al. (2011), which evaluated a number of synthetic SWOT orbit scenarios with partial coverage on an Arctic river, and indicated that those orbits that observed the upstream part of the river compared positively against those observing the downstream area, the reason being that the corrections propagate downstream. Still, the simulations with early observations have a higher RMSE than those with the same first visit time and $\Delta t_a = 24$ h in Fig. 5, as the cessation of observations moves the forecast back towards the open loop. The effect that observations that are too early do not provide very useful information is now well shown by the RMSE evolution at Mythe Bridge, where substitution of the 20th July overpass by the 25th July one (i.e. 1-day delay of the 5 visits) reduces the RMSE. One extra 1-day delay misses the rising limb and results in the largest increase in RMSE among the simulations.

As a summary, at the early stages of the flood event, the forecast within highly variable flood dynamics benefits from increased observation frequencies ($\Delta t_a = 12$ h revisit time). After the flood peak, as the event proceeds with smoother flood dynamics, the sensitivity to the revisit time is drastically reduced so it becomes adequate to decrease the observation frequency. This provides a longer time coverage of the event for the same cost.

Here we chose to use the ETKF, which has received a strong attention in recent theoretical and practical studies (e.g., Bishop et al., 2001; Hunt et al., 2007; Livings et al., 2008; Sakov & Oke, 2008). Very likely, results from SAR-based WLO assimilation are filter-dependent. A comparative study could shed light on the complexity of the implementation of the various available filters for data assimilation versus the efficiency for assimilating SAR-based WLOs in inundation problems.

4. Conclusions

This study focuses on the problem of scheduling satellite-based SAR acquisitions for sequential assimilation of SAR-derived WLOs into operational flood modelling. In particular, the interest is on areas of human or economic risk, thus involving urban areas which require detailed 2D flood modelling. For this study, we have used an ETKF and the 2D hydrodynamic model LISFLOOD-FP with a synthetic analysis based on a real flooding case around the Severn-Avon river junction in Southwest UK. We have touched on a number of related issues. Firstly, we have provided a clarification of the correlations originating from generic hydrologic-hydrodynamic modelling cascades. Secondly, our results indicate that the spread, in the case study, is relatively stable. Thus, localization and/or inflation techniques do not seem to be required for this study. However, this is

case-dependent. These techniques could be required in other scenarios for sequential ETKF-based DA in hydrodynamic modelling, if the best performance of the filter is wanted. In agreement with previous studies, we have shown that estimation/correction of the inflow errors leads to improved forecasts. Regarding the satellite visit parameters, for a standard budget-limited scenario, the operational scheduling of satellite SAR acquisitions should try to capture the early stages of the rising limb, possibly with the highest available observation frequency. After the flood peak, it becomes convenient to spread out the observations in time. This enables the forecast to be kept on track for a longer time and same cost for the less variable flood dynamics that occur during the falling limb period. This result should equally apply to airborne observations and data collection for offline model evaluation.

This study has assumed an error-free model and no parametric uncertainty. Errors in model parameters should be considered in future work. For example, friction parameters may even be variable over an event and are certainly non-stationary between events.

Acknowledgements

This work was supported by NERC through the DEMON (Developing Enhanced impact MOdels for integration with Next generation NWP and climate outputs) project, included in the NERC SRM (Storm Risk Mitigation) programme (NE/I005242/1). The authors thank three anonymous reviewers, which were a great help in improving this manuscript.

References

- Andreadis, K. M., Clark, E. A., Lettenmaier, D. P., & Alsdorf, D. E. (2007). Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophys. Res. Lett.*, *34*, L10403–.
- Auclair, F., Marsaleix, P., & Mey, P. D. (2003). Space-time structure and dynamics of the forecast error in a coastal circulation model of the gulf of lions. *Dyn. Atmos. Oceans*, *36*, 309–346.
- Bates, P. D., & De Roo, A. P. J. (2000). A simple raster-based model for flood inundation simulation. *J. Hydrol.*, *236*, 54–77.
- Bates, P. D., Horrit, M. S., & Fewtrell, T. J. (2010). A simple inertial formulation of the shallow water equations for efficient two dimensional flood inundation modelling. *J. Hydrol.*, *387*, 33–45.
- Biancamaria, S., Durand, M., Andreadis, K. M., Bates, P. D., Boone, A., Mognard, N. M., Rodríguez, E., Alsdorf, D. E., Lettenmaier, D. P., & Clark, E. A. (2011). Assimilation of virtual wide swath altimetry to improve arctic river modeling. *Remote Sens. Environ.*, *115*, 373–381.
- Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform kalman filter. part i: Theoretical aspects. *Mon. Weather Rev.*, *129*, 420–436.
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., & Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Adv. Water Resour.*, *31*, 1309–1324.
- Durand, M., Andreadis, K. M., Alsdorf, D. E., Lettenmaier, D. P., Moller, D., & Wilson, M. (2008). Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophys. Res. Lett.*, *35*, L20401–.
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, *99*, 10143–10162.
- Evensen, G. (2003). The ensemble kalman filter: theoretical formulation and practical implementation. *Ocean Dyn.*, *53*, 343–367.
- Evensen, G. (2004). Sampling strategies and square root analysis schemes for the enkf. *Ocean Dyn.*, *54*, 539–560. 10.1007/s10236-004-0099-2.
- Evensen, G. (2009). *Data Assimilation: The Ensemble Kalman Filter*. Springer.
- Friedland, B. (1969). Treatment of bias in recursive filtering. *IEEE Trans. Autom. Control*, *14*, 359–367.

- Furrer, R., & Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in kalman filter variants. *Journal of Multivariate Analysis*, *98*, 227–255.
- Giustarini, L., Matgen, P., Hostache, R., Montanari, M., Plaza, D., Pauwels, V. R. N., De Lannoy, G. J. M., De Keyser, R., Pfister, L., Hoffmann, L., & Savenije, H. H. G. (2011). Assimilating sar-derived water level data into a hydraulic model: a case study. *Hydrol. Earth Syst. Sci.*, *15*, 2349–2365.
- Hamill, T. M. (2001). Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Weather Rev.*, *129*, 550–560.
- Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Mon. Weather Rev.*, *129*, 2776–2790.
- Horritt, M. S., Mason, D. C., Cobby, D. M., Davenport, I. J., & Bates, P. D. (2003). Waterline mapping in flooded vegetation from airborne sar imagery. *Remote Sens. Environ.*, *85*, 271–281.
- Hostache, R., Lai, X., Monnier, J., & Puech, C. (2010). Assimilation of spatially distributed water levels into a shallow-water flood model. part ii: Use of a remote sensing image of mosel river. *J. Hydrol.*, *390*, 257–268.
- Houtekamer, P. L., & Mitchell, H. L. (2001). A sequential ensemble kalman filter for atmospheric data assimilation. *Mon. Weather Rev.*, *129*, 123–137.
- Houtekamer, P. L., & Mitchell, H. L. (2005). Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, *131*, 3269–3289.
- Houtekamer, P. L., Mitchell, H. L., & Deng, X. (2009). Model error representation in an operational ensemble kalman filter. *Mon. Weather Rev.*, *137*, 2126–2143.
- Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. *Physica D*, *230*, 112–126.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. ASME Ser. D: J. Basic Eng.*, *82*, 35–45.
- Lai, X., & Monnier, J. (2009). Assimilation of spatially distributed water levels into a shallow-water flood model. part i: Mathematical method and test case. *J. Hydrol.*, *377*, 1–11.
- Lane, S. N., James, T. D., Pritchard, H., & Saunders, M. (2003). Photogrammetric and laser altimetric reconstruction of water levels for extreme flood event analysis. *Photogramm. Rec.*, *18*, 293–307.
- Livingston, D. M., Dance, S. L., & Nichols, N. K. (2008). Unbiased ensemble square root filters. *Physica D*, *237*, 1021–1028.
- Mason, D. C., Davenport, I. J., Neal, J. C., Schumann, G. J.-P., & Bates, P. D. (2012a). Near real-time flood detection in urban areas and rural areas using high resolution synthetic aperture radar images. *IEEE Trans. Geosci. Remote Sensing*, *50*(8), 3041–3052.
- Mason, D. C., Schumann, G. J.-P., Neal, J. C., García-Pintado, J., & Bates, P. D. (2012b). Automatic near real-time selection of flood water levels from high resolution synthetic aperture radar images for assimilation into hydraulic models: a case study. *Remote Sens. Environ.*, *124*, 705–716.
- Mason, D. C., Speck, R., Deveraux, B., Schumann, G. J.-P., Neal, J. C., & Bates, P. D. (2010). Flood detection in urban areas using terrasars-x. *IEEE Trans. Geosci. Remote Sensing*, *48*, 882–894.
- Matgen, P., Montanari, M., Hostache, R., Pfister, L., Hoffmann, L., Plaza, D., Pauwels, V. R. N., De Lannoy, G. J. M., De Keyser, R., & Savenije, H. H. G. (2010). Towards the sequential assimilation of sar-derived water stages into hydraulic models using the particle filter: proof of concept. *Hydrol. Earth Syst. Sci.*, *14*, 1773–1785.
- Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state-parameter estimation of hydrological models using ensemble kalman filter. *Adv. Water Resour.*, *28*, 135–147.
- Neal, J. C., Schumann, G. J.-P., & Bates, P. D. (2012). A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas. *Water Resour. Res.*, *48*, W11506, 16pp.
- Neal, J. C., Schumann, G. J.-P., Bates, P. D., Buytaert, W., Matgen, P., & Pappenberger, F. (2009). A data assimilation approach to discharge estimation from space. *Hydrol. Process.*, *23*, 3641–3649.
- Ochotta, T., Gebhardt, C., Saupe, D., & Wergen, W. (2005). Adaptive thinning of atmospheric observations in data assimilation with vector quantization and filtering methods. *Quarterly Journal of the Royal Meteorological Society*, *131*, 3427–3437.
- Ott, E., Hunt, B., Szunyogh, I., Zimin, A., Kostelich, E., Corazza, M., Kalnay, E., Patil, D., & Yorke, J. (2004). A local ensemble kalman filter for atmospheric data assimilation. *Tellus Ser. A-Dyn. Meteorol. Oceanol.*, *56*.
- Raclot, D. (2006). Remote sensing of water levels on floodplains: a spatial approach guided by hydraulic functioning. *Int. J. Remote Sens.*, *27*, 2553–2574.
- Sakov, P., & Oke, P. R. (2008). Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Weather Rev.*, *136*, 1042–1053.
- Schumann, G. J.-P., Hostache, R., Puech, C., Hoffmann, L., Matgen, P., Pappenberger, F., & Pfister, L. (2007).

- High-resolution 3-d flood information from radar imagery for flood hazard management. *IEEE Trans. Geosci. Remote Sensing*, *45*, 1715–1725.
- Schumann, G. J.-P., Neal, J. C., Mason, D. C., & Bates, P. D. (2011). The accuracy of sequential aerial photography and sar data for observing urban flood dynamics, a case study of the uk summer 2007 floods. *Remote Sens. Environ.*, *115*, 2536–2546.
- Snyder, C., Bengtsson, T., Bickel, P., & Anderson, J. (2008). Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, *136*, 4629–4640.
- Stephens, E. M., Bates, P. D., Freer, J. E., & Mason, D. C. (2012). The impact of uncertainty in satellite data on the assessment of flood inundation models. *J. Hydrol.*, *414–415*, 162–173.
- Stewart, L. M., Dance, S. L., & Nichols, N. K. (2008). Correlated observation errors in data assimilation. *Int. J. Numer. Methods Fluids*, *56*, 1521–1527.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S. (2003). Ensemble square root filters*. *Mon. Weather Rev.*, *131*, 1485–1490.
- Vörösmarty, C., Askew, A., Grabs, W., Barry, R. G., Birkett, C., Dell, P., Goodison, B., Hall, A., Jenne, R., Kitaev, L., Landwehr, J., Keeler, M., Leavesley, G., Schaake, J., Strzepek, K., Sundarvel, S. S., Takeuchi, K., Webster, F., & Group, T. A. H. (2001). Global water data: A newly endangered species. *EOS Trans. AGU*, *82*, 54–58.
- Wang, X., Bishop, C. H., & Julier, S. J. (2004). Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Weather Rev.*, *132*, 1590–1605.

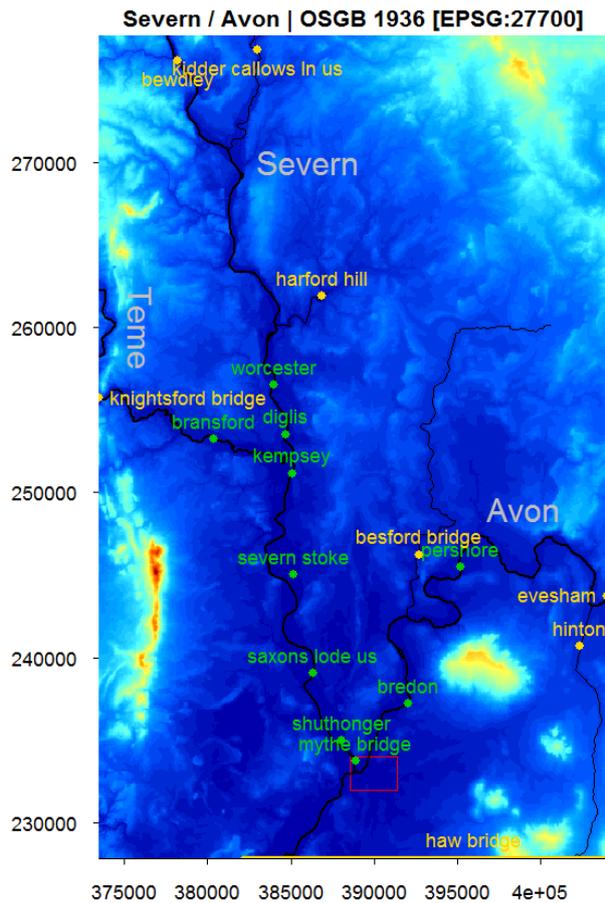


Figure 1: Study domain. OSGB 1936 British National Grid projection; coordinates in meters. Grey labels indicate major rivers (thick black lines). The red polygon surrounds the Tewkesbury urban area. Orange labels/dots refer to inflow boundary conditions, some of them on smaller tributaries (thin black lines). The orange line to the South indicates a time-varying stage boundary condition. Green labels/dots show locations with available stage observations for the event, from which we just use their locations as a reference in the current study. The background is the 75 m resolution DEM used for the model, based on upscaling the NEXTMAP British digital terrain model.

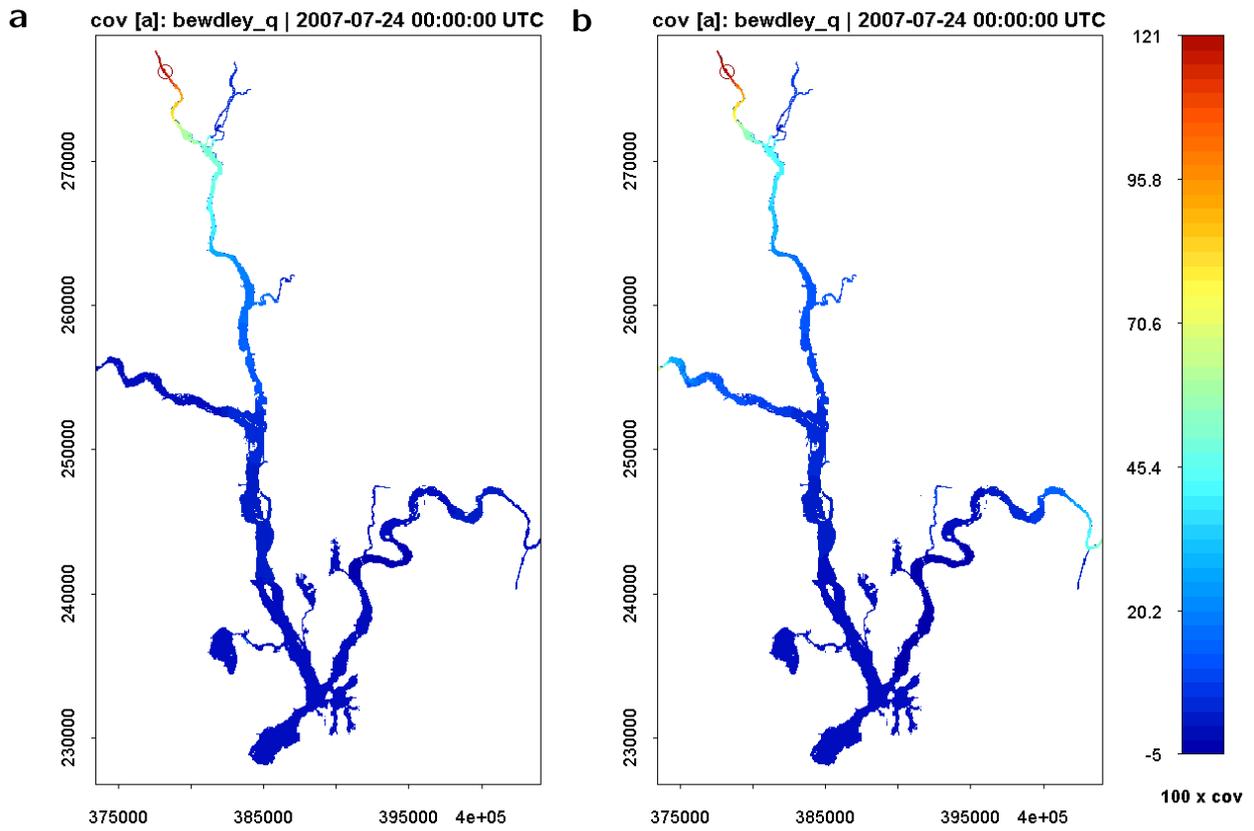


Figure 2: Spatial covariance developed after 5 daily ($\Delta t_a = 24$ h) SAR assimilation steps between errors at the inflow boundary condition at Bewdley (red circle), at the North of the Severn river, and water stage in the domain, for (a) no spatial covariance in the time-correlated inflow errors (representing measurement-driven hydrodynamic model), and (b) Gaussian-decay spatial covariance present (representing hydrologic model-driven hydrodynamic model). See details in Section 2.5. $\tau = 3$ days for both, and $\theta = 62000$ m for (b).

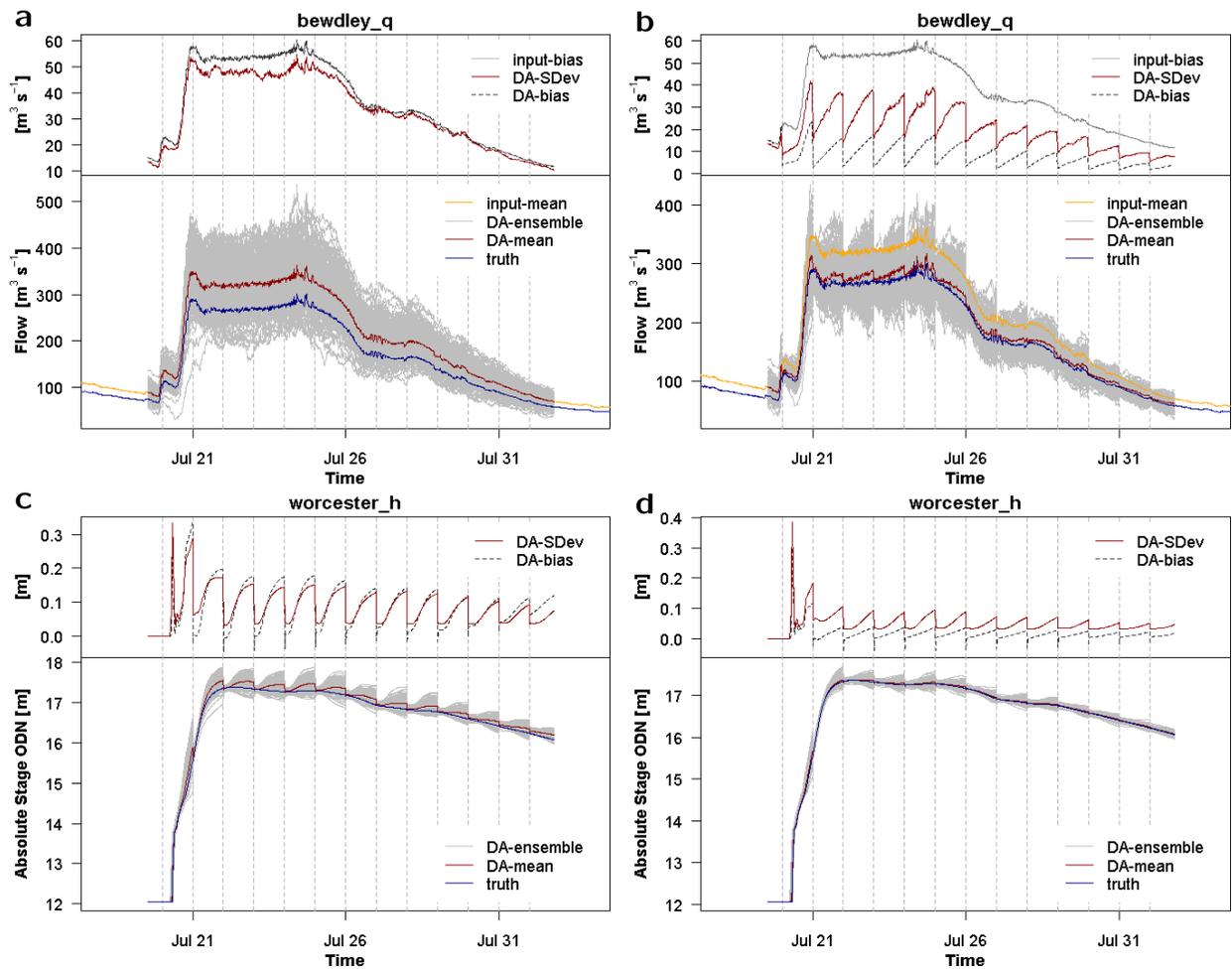


Figure 3: Evolution of the inflow at Bewdley (a), and corresponding forecast at Worcester (c), without attempting to estimate/correct the errors in the inflow boundary conditions. Inflow (b) and forecast (d) are as (a) and (c), respectively, but estimating and correcting the inflow errors by augmentation of the state vector. For each ensemble at Worcester, upper summary plots show the standard deviation of the ensemble (DA-SDev), and the bias between the mean of the ensemble and the truth (DA-bias). For the inflow at Bewdley, the input bias is also shown. Vertical lines indicate satellite overpass/DA times ($\Delta t_a = 24$ h).

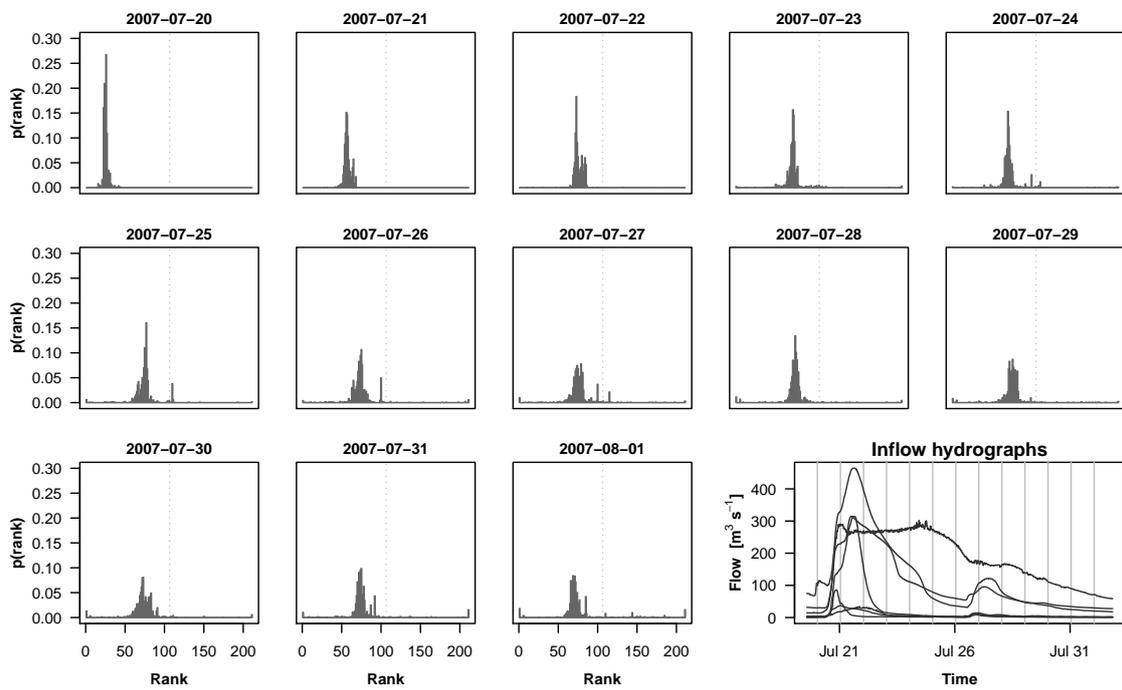


Figure 4: Evolution of the rank histogram evaluated for the forecast ensemble at each assimilation time for the $\Delta t_a = 24$ h revisit time simulation. The subplot at the lower-right corner is included as a reference indicating the corresponding assimilation times in relation with the various true inflow boundary conditions.

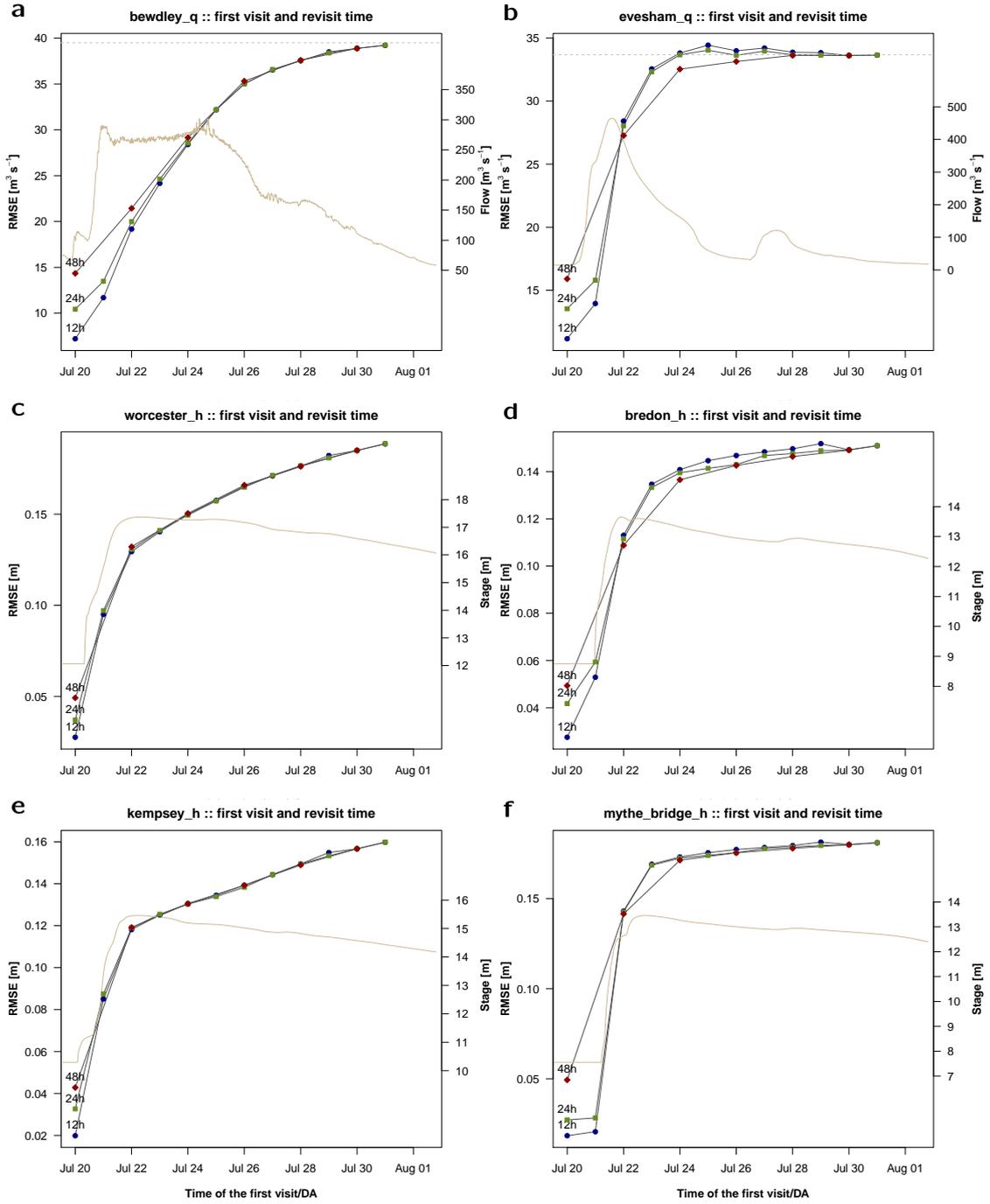


Figure 5: RMSE for inflows at the two boundary conditions with the highest inflow (Bewdley at the Severn, and Evesham at the Avon), and forecasted stage at four gauges: Worcester, at the river Severn; Kempsey, just after the junction between the Teme and the Severn; Bredon, in the Avon; and Mythe Bridge, in the Severn by Tewkesbury. True inflow/stage at the corresponding location is shown as a reference. Curves are calculated for revisit times $\Delta t_a = 12 \text{ h}$, 24 h , and 48 h . Each point in each curve denotes the first visit time and the corresponding RMSE over the entire window.

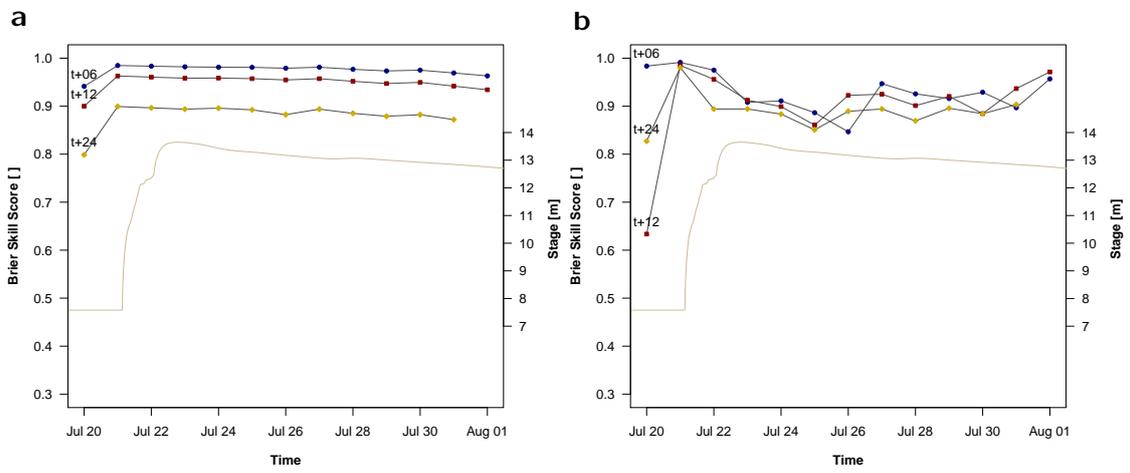


Figure 6: Brier Skill Scores (BSS) for the $\Delta t_a = 24$ h revisit time, starting the 20th of July, simulation evaluated against the open loop. At each t_a , the BSS is shown for forecast times $t+6$ h, $t+12$ h, and $t+24$ h. BSS is calculated at (a) the 7 inflow boundary conditions, and (b) at the forecasted stages at the 10 reference gauge locations. True stage at Saxons Lode US gauge is shown in both subplots as a reference.

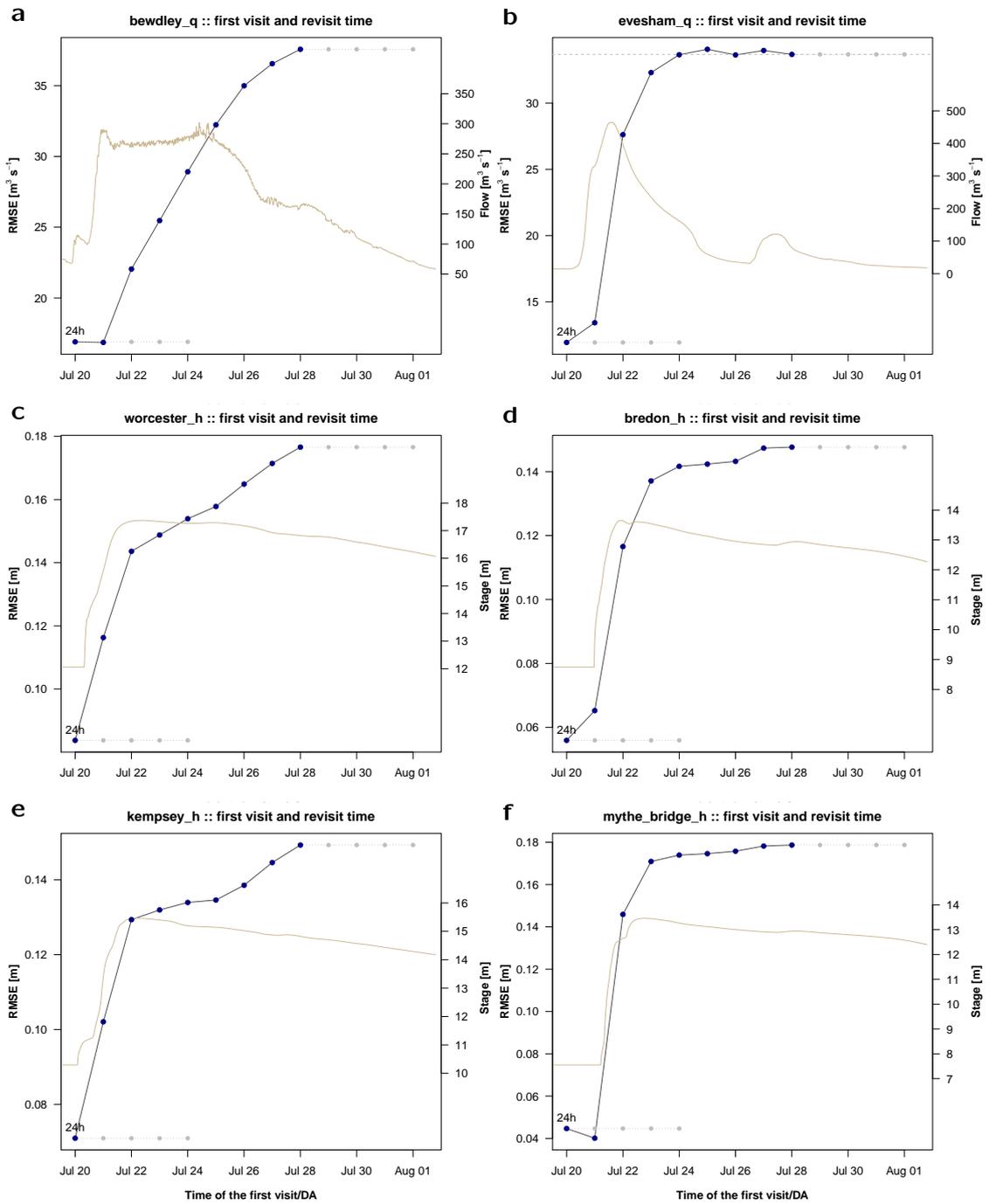


Figure 7: As Figure 5 for $\Delta t_a = 24$ h revisit time, but for five SAR overpasses successively delayed by one hour. Each blue point denotes the first visit time, and the corresponding RMSE over the entire window. As an example, for the first and last first visit time, all visits/DA times are shown as grey points.