University of

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

by

Javier Garcia-Pintado, Jeff C. Neal, David C.

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

Javier Garc a-Pintado[a,b,] , Je C. Neal[c], David C. Mason[a,b], Sarah L. Dance[a,b], Paul D. Bates[c]

[a]

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) succesfully 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 that 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 justi cation (e.g., Furrer & Bengtsson, 2007). Conversely, as, discussed by Snyder et al. (2008), there are results showing that the standard particle lter must have an ensemble size exponentially large in the variance of the observation log likelihood or the lter will su er from a \collapse". Thus, despite current research to improve the PF e ciency for large dimensional problems, it remains unclear whether it will be a viable alternative in a near future for these operational ooding problems in areas with high human or economical risk.

Both EnKF and PF are Monte Carlo-based lters that require a number of ensembles of model runs to represent the forecast uncertainty. 2-D hydrodynamic models for simulating oods are expensive to run in ensemble mode with the result that, in operational cases, watershed scale 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 (Δt_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 (Δt_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 spatiotemporallly-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 1.5×10^4 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 30.6×49.8 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 m^{3}s^{-1} at Saxons Lode Us) occurred on July 22, and the river did not return to bank-full until July 31 ($\approx$ 350 m^{3}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 m^{3}s^{-1}) in the Severn, Evesham (Q_p = 465 m^{3}s^{-1}) in the Avon, and Knightsford Bridge (Q_p = 315 m^{3}s^{-1}) in the Teme. The Severn also had inflows from Kidder Callows (Q_p = 33 m^{3}s^{-1}) and Hardford Hill (Q_p = 36 m^{3}s^{-1}

(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 $d =\ \ w$ between the channel width (w) and depth (d), where we used the parameters   = 0:30, and   = 0:78. For the main rivers, we estimated mean channel widths from eld campaigns, and calibrated   and   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   and   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 eld observations. Simulations

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,

$$\mathbf{x} = \begin{bmatrix} \mathbf{z} \\ \boldsymbol{\theta} \end{bmatrix}; \tag{1}$$

where $\mathbf{z}$ is the n_S-dimensional model state and $\boldsymbol{\theta}$ is a generic n_θ-dimensional vector of parameters. Thus $\mathbf{x}$ is the augmented n-dimensional state vector, with $n = n_S + n_\theta$. 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:

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

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

where the forecast (prior) and analysis (posterior) quantities are denoted by the superscripts f and a, respectively; $\mathbf{y} \in \Re^p$ 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:

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\mathbf{P}^f \mathbf{H}^{\mathrm{T}} + \mathbf{R} \right)^{-1}; \tag{4}$$

where the superscript "T" denotes matrix transposition, and $\mathbf{R}$ is the $p \times p$ observation error covariance matrix.

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 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 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. 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 becames 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 in ows. The degree of this spatial correlation will be highly dependent on the perturbation of the forcing | chie y precipitation elds | 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 di erent tributaries may well lead to one WLO (either from remote sensing or a standard gauge) at the head of one tributary to in uence 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 in uential in both correcting the hydrodynamic state and, possibly, correcting the hydrologic model errors. However, for scenario (a), if spatial correlations between in ow 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 ood scenario with available real in ow 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 in ows, 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 ood forecast lead times, we choose scenario (b) for the remainder of this study. This approach has an advantage over selecting a speci c hydrologic model in that it can be regarded as using a \generic" hydrologic model whose in uence in generating in ow boundary conditions is explicitly modelled and known. This clari es the analysis for our study.

Below, within Section 2.5.2, we detail how we simulate the in ow ensembles with random errors. As an example of the di erence between the scenarios at(b)enwithlet0(b)-2ased

(Evensen, 2003). $\mathbf{w}_k \in \Re^{n_@}$ is a here white noise obtained by drawing samples from $N(0, \ )$, where $\ \in \Re^{n_@ \ n_@}$ is a distance-dependent correlation matrix (described below in Section 2.5.4).

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

$$\mathbf{Q}_k = [\mathbf{q}_{k1} | \mathbf{q}_{k2} | \ldots | \mathbf{q}_{km}], \tag{20}$$

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 $\ ^{1/2}\mathbf{A}$ in (18)). After assimilation steps, errors are regenerated ($k = 1$). So $\mathbf{Q}_{k-1} \ \mathbf{Q}_0$ refers to the errors updated by the assimilation process. With this formulation, $\ $ 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_1

After an assimilation step is conducted, the q_0^0 ensemble at each in ow 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 values.

2.5.3. Determination of

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

$$ {}_{ji}\ {}_{jj} = e^{-\frac{t}{}}; \tag{25}$$

which relates and , and clari es that, disregarding the heteroscedastic variance term in (24), the covariance in time between q_i^0 and q_j^0 is damped by a ratio e^{-1} over a time period $t_{ij} =$ (see Evensen, 2003). For a speci c time step k of length t_k, then

$$ {}_k = e^{-\frac{t_k}{}}; \tag{26}$$

which allows one to use (19) for any time step length by subtituting by the corresponding ${}_k$, and, instead of (24), the error covariance, at each in ow, between any two time steps $(i; j)$ is more generically expressed as

$$\overline{q_i^0 q_j^{0\mathrm{T}}} = \sqrt{\frac{s_i s_j}{s_0^2}}^{\,h}\ {}^2_0 \prod_{k=i+1}^{j} {}_k. \tag{27}$$

2.5.4. Spatial correlation model for in ow errors

The spatial correlation matrix , for generation of the white noise w_k in (19), can be created by any procedure which considers that correlation in in ow errors is dependent on the distance between the locations of the point in ow boundary conditions. Here we chose the Gaussian-decay correlation model

$$ {}_{ij} = e^{-\frac{1}{2}\left(\frac{d_{ij}}{}\right)^2}; \tag{28}$$

where the subscripts i and j refer to any two boundary conditions, ${}_{ij} \in [0; 1]$ is the corresponding spatial correlation and element in , d_{ij} is the distance between the corresponding locations, and is a spatial correlation coe cient.

2.5.5. Selection of and and in ow error estimation

As abovementioned (Section 2.5.1), the true dynamics of the mean error of measured or forecasted in ows 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 in uences the forecast, and how DA is able to partially solve for it. To emulate errors from a "generic" hydrologic model, we rst imposed a positive 20% bias on measured in ows. Then we perturbed the biased in ows with spatiotemporally correlated errors to generate the in ow 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 and cannot embrace all possible situations. Here, our parameters for the error forecast model were = 3 days and = 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 . In this study, as we have imposed a stationary bias in the true mean error, higher values of , 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

$$\mathrm{BSS} = 1 - \frac{\overline{(\mathbf{f}_s - \mathbf{o})^2}}{\overline{(\mathbf{f}_r - \mathbf{o})^2}}; \tag{29}$$

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 dowstream 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 t_a = 24 h. If in ows 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 in ow errors are also estimated and corrected according to the used error forecast model, each sequential assimilation pushes the in ows 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 clari ed 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 in ow (here mostly in uenced by Bewdley at the North) are not corrected they have a control e ect that, after any assimilation update, causes the forecast to drift away from the truth, leading to an early overestimation of the ood stage. A similar e ect was shown by Matgen et al. (2010). The case with in ow updating keeps the forecast on track very close to the truth. Curves at the other in ows and sampled forecast locations show similar e ects (see Figs. S2{S15; supplementary material). The speed at which the updated in ows drift away from the truth when they are updated is related to the lack of match between the used error forecast model (with = 3) and the imposed stationary bias. As described in Section 2.5, in this case, higher values would result in a more persistent propagation of the errors estimated at the assimilation time, giving an improved mean in ow error estimation and correction in time with the forecast. In the remainder of this paper we use simulations with updating of the in ow errors, as this leads to a clear forecast improvement. However, we keep = 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) in ow error evolution. Here we assumed that friction parameters are known. In real cases, if friction in the channels and oodplain 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 lter would bene t from larger ensemble sizes. Estimation of friction, however, is beyond the scope of this study.

3.2. Ensemble Properties

The use of a nite 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 lter 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 t_a = 24 h revisit time, starting on the 20th of July, before the ood 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 m

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 $O(10^4$

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 t_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

the rank histograms. For the forecast stage, disregarding the 20th July forecast when ows are case-dependent. These techniques could be required in other scenarios for sequential ETKF-based

Furrer, R., & Bengtsson, T. (2007). Estimation of high-dimensional prior and posterior covariance matrices in kalman lter 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., P ster, L., Ho mann, 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 ltering of background error covariance estimates in an ensemble kalman lter. *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 ooded 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 ood 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 lter for atmospheric data assimilation. *Mon. Weather Rev.*, *129*, 123{137.

Houtekamer, P. L., & Mitchell, H. L. (2005). Ensemble kalman ltering. *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 lter. *Mon. Weather Rev.*, *137*, 2126{2143.

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). E cient data assimilation for spatiotemporal chaos: A local ensemble transform kalman lter. *Physica D*, *230*, 112{126.

Kalman, R. E. (1960). A new approach to linear ltering and prediction problems. *Trans. ASME Ser. D: J. Basic Eng.*, 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.

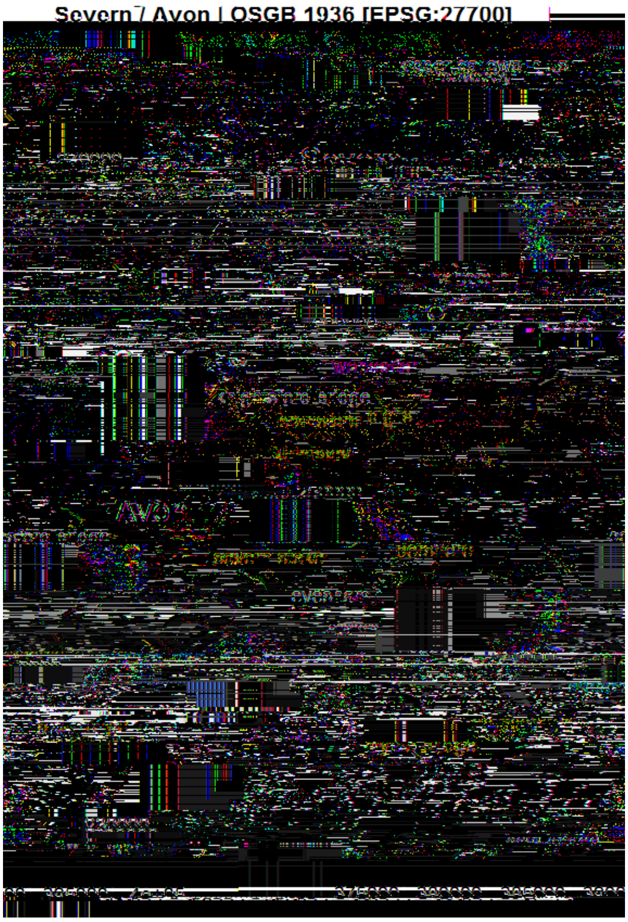
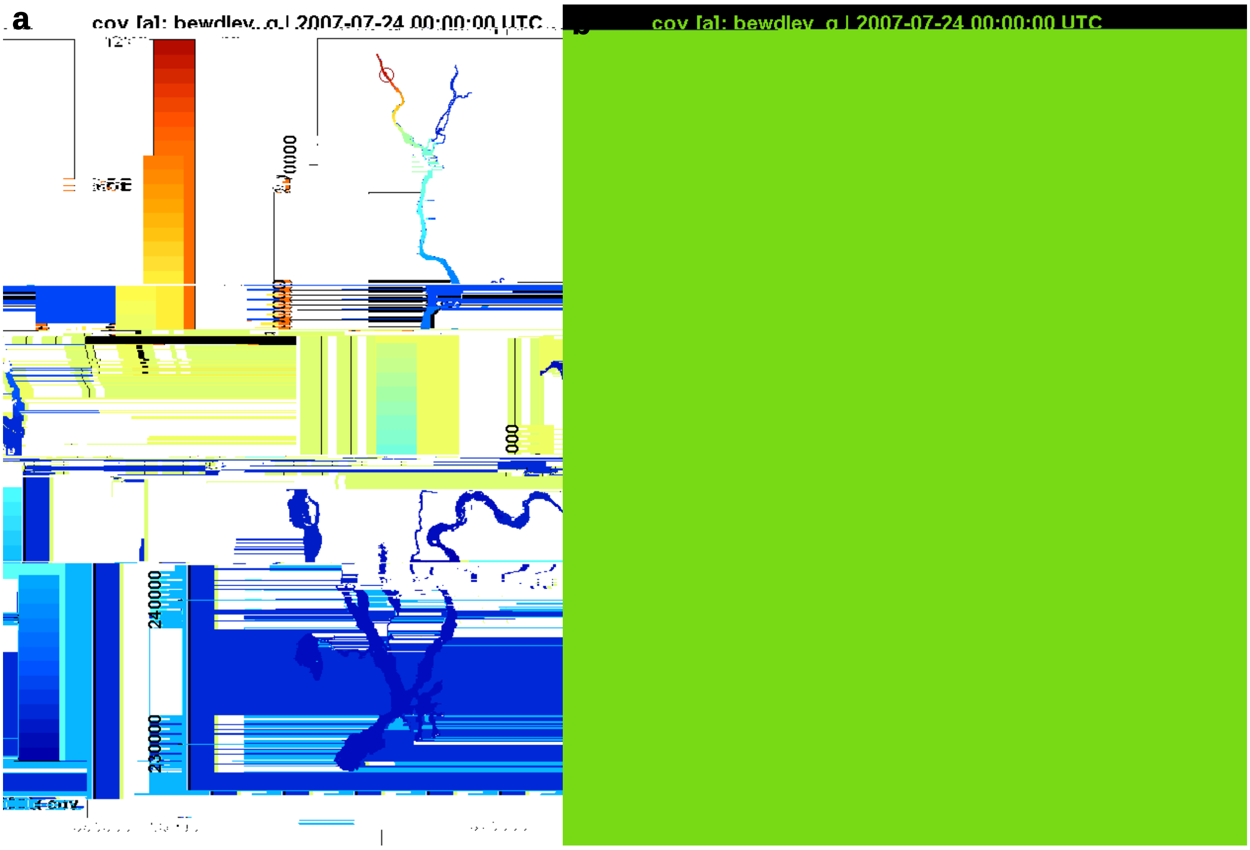


Figure 1: Study domain. OSGB 1936 British National Grid projection; coordinates in meters. Grey labels indicate major rivers (thick black lines). The red polygon surrounds the Tewkesbury urban area. Orange labels/dots refer to in ow boundary conditions, some of them on smaller tributaries (thin black 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.


a
cov [a]: bewdley_σ | 2007-07-24 00:00:00 UTC
240000
230000
cov
cov [a]: bewdley_σ | 2007-07-24 00:00:00 UTC

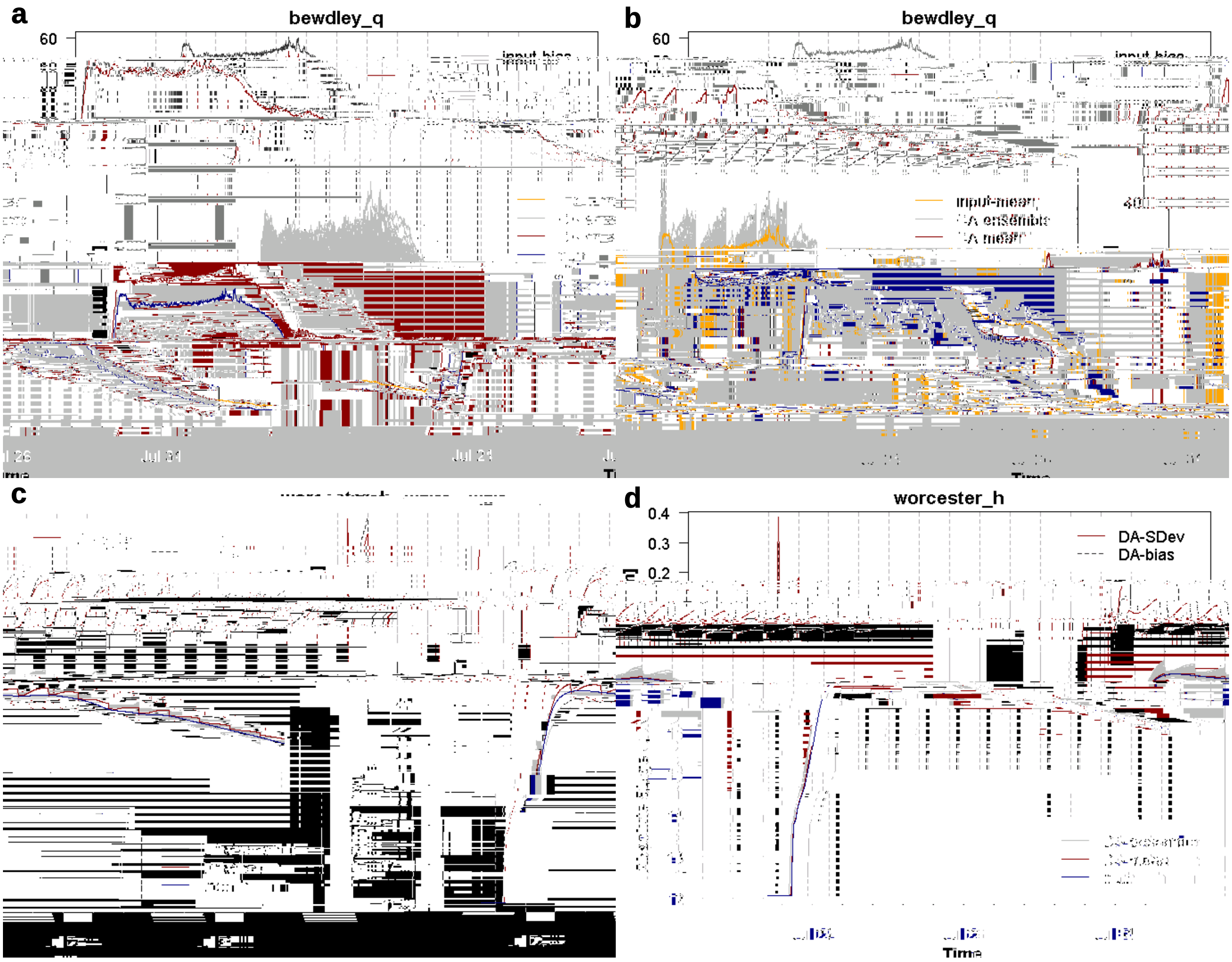


Figure 3: Evolution of the inflow at Bewdley (a), and corresponding forecast at Worcester (c), without attempting to estimate/correct the errors in the inflow boundary conditions. Inflow (b) and forecast (d) are as (a) and (c), respectively, but estimating and correcting the inflow errors by augmentation 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 (t_a = 24 h).

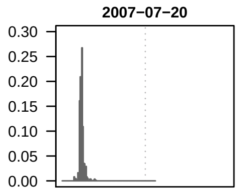


Figure 4: Evolution of the rank histogram evaluated for the forecast ensemble at each assimilation time for the t_a = 24 h revisit time simulation. The subplot at the lower-right corner is included as a reference indicating the corresponding assimilation times in relation with the various true in ow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

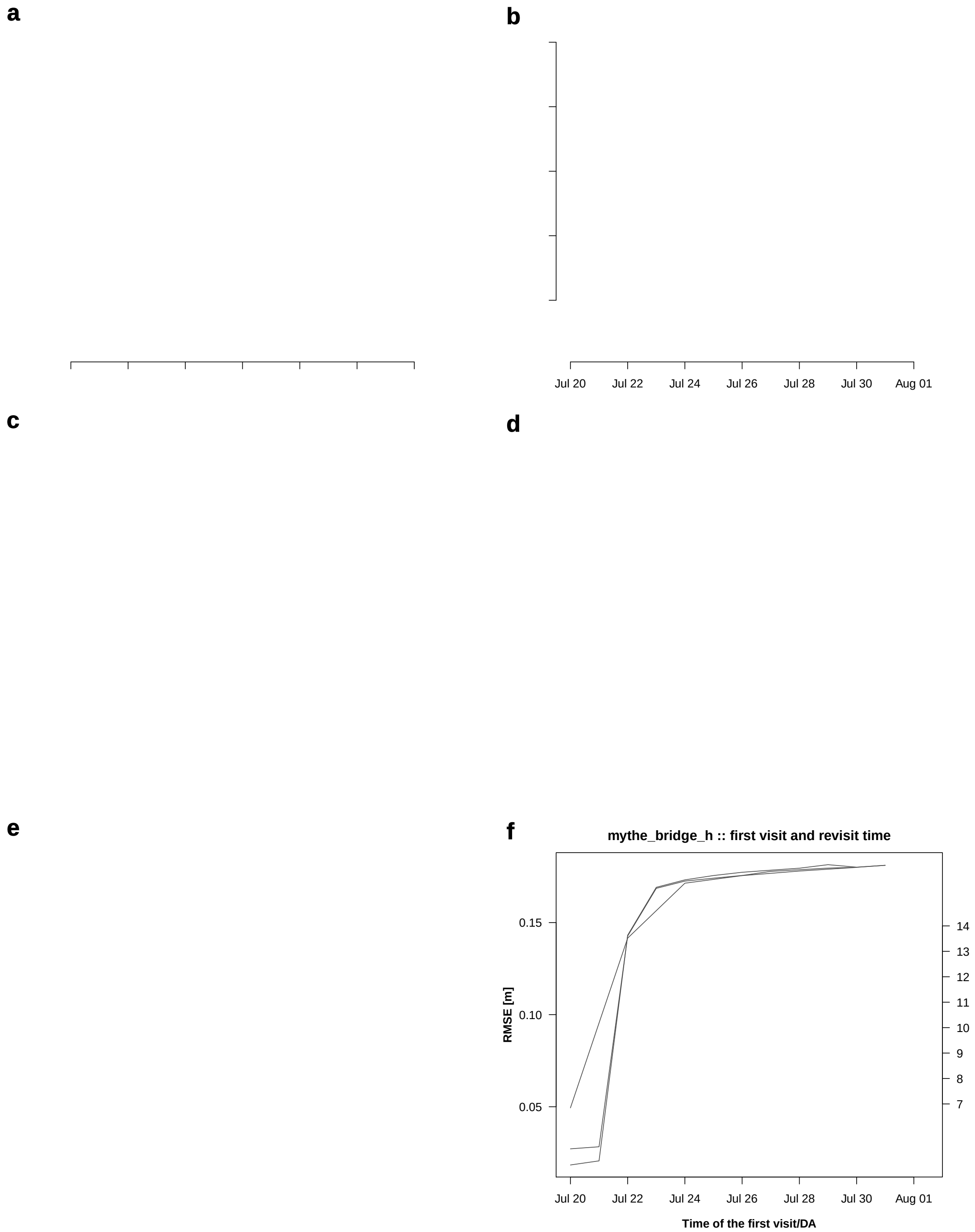


Figure 5: RMSE for in ows at the two boundary conditions with the highest in ow (Bewdley at the Severn, and Evesham at the Avon), and forecasted stage at four gauges: Worcester, at the river Severn; Kempsey, just after the junction between the Teme and the Severn; Bredon, in the Avon; and Mythe Bridge,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

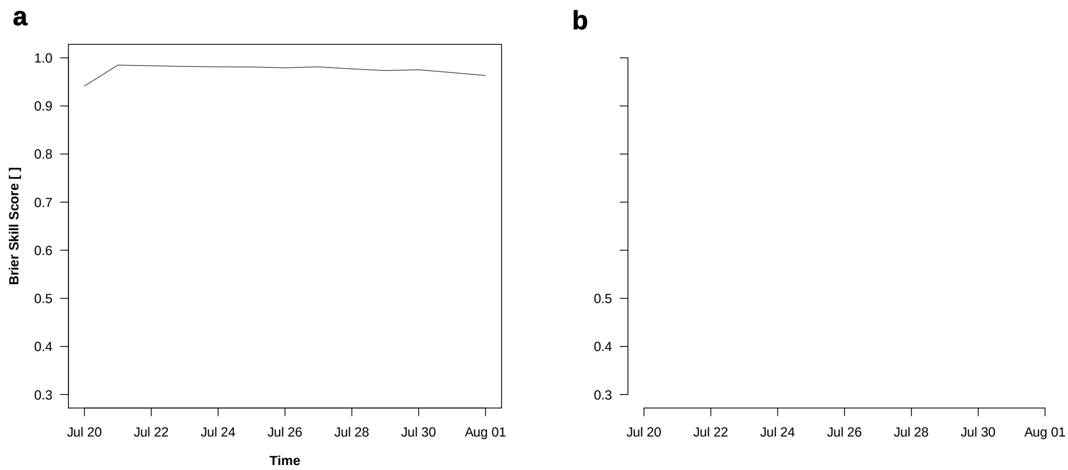


Figure 6: Brier Skill Scores (BSS) for the t_a

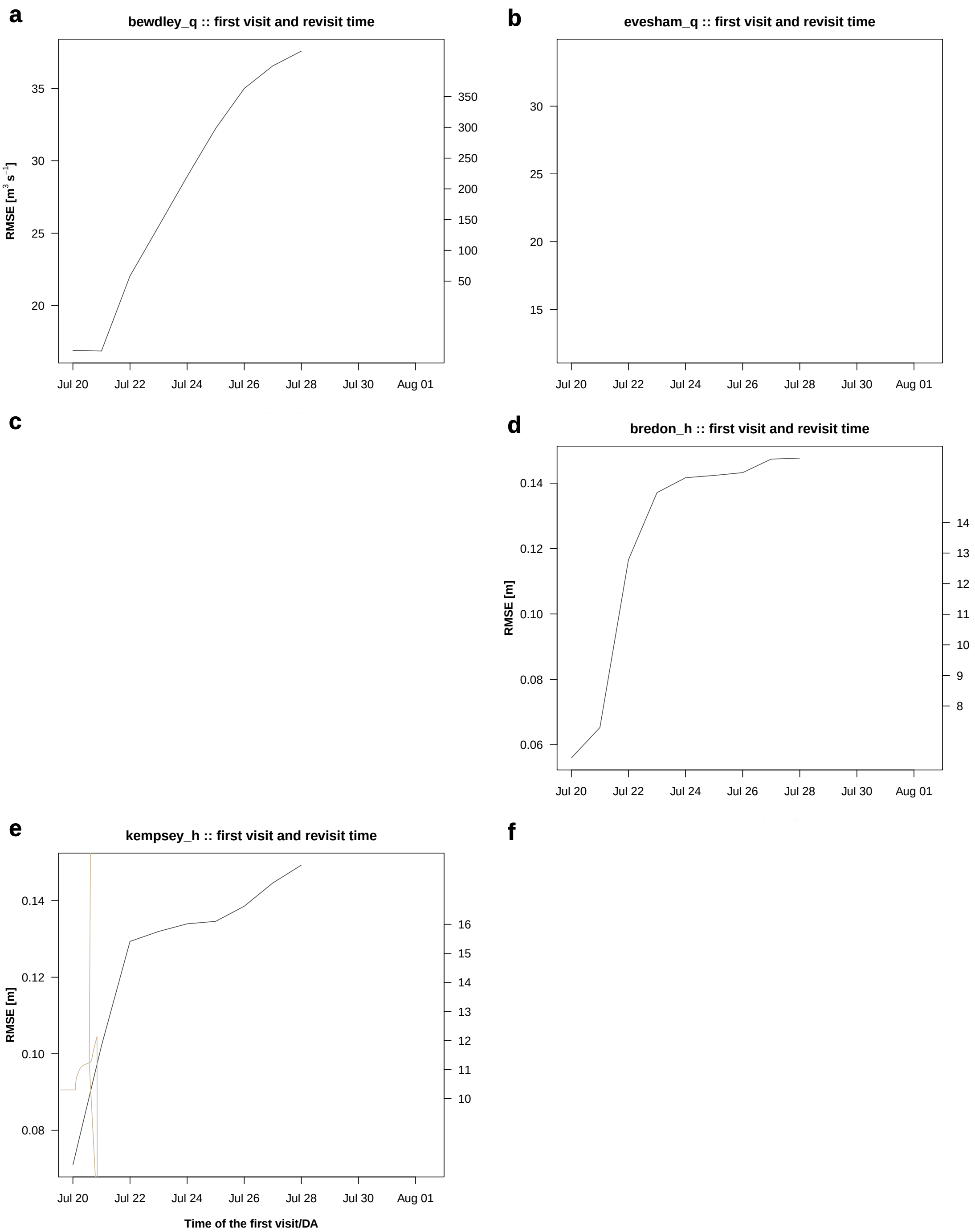


Figure 7: As Figure 5 for t_a = 24 h revisit time, but for ve SAR overpasses successively delayed by one hour. Each blue point denotes the rst visit time, and the corresponding RMSE over the entire window. As an example, for the rst and last rst visit time, all visits/DA times are shown as grey points.