

Error-correction across gauged and ungauged locations: A data assimilation-inspired approach to post-processing river discharge forecasts

Gwyneth Matthews^{1,2}, Hannah L Cloke^{1,3}, Sarah L Dance^{1,4,5}, and Christel Prudhomme²

¹Department of Meteorology, University of Reading, Reading, United Kingdom

²European Centre for Medium-range Weather Forecasts, Reading, United Kingdom

³Department of Geography and Environmental Science, University of Reading, Reading, United Kingdom

⁴Department of Mathematics and Statistics, University of Reading, Reading, United Kingdom

⁵National Centre for Earth Observation (NCEO), Reading, United Kingdom

Correspondence: Gwyneth Matthews (g.r.matthews@pgr.reading.ac.uk)

Abstract. Forecasting river discharge is essential for disaster risk reduction and water resource management, but forecasts of the future river state often contain errors. Post-processing reduces forecast errors but is usually only applied at the locations of river gauges, leaving the majority of the river network uncorrected. Here, we present a data-assimilation-inspired method for error-correcting ensemble simulations across gauged and ungauged locations in a post-processing step. Our new method employs state augmentation within the framework of the ~~Localised-Local~~ Ensemble Transform Kalman Filter (LETKF) ~~to estimate~~. Using the LETKF, an error vector representing the forecast residual is estimated for each ensemble member. The LETKF uses ensemble error covariances to spread observational information from gauged to ungauged locations in a dynamic and computationally efficient manner. To improve the efficiency of the LETKF we define new localisation, covariance inflation, and initial ensemble generation techniques that can be easily transferred between modelling systems and river catchments. We implement and evaluate our new error-correction method for the entire Rhine-Meuse catchment using forecasts from the Copernicus Emergency Management Service's European Flood Awareness System (EFAS). The resulting river discharge ensembles are error-corrected at every grid box but remain spatially and temporally consistent. ~~The skill is evaluated at 89 proxy-ungauged locations~~ A spatial cross-validation strategy is used to assess the ability of the method to spread the correction along the river network to ungauged locations. The skill of the ensemble mean is improved at almost all locations including stations both up- and downstream of the assimilated observations. Whilst the ensemble spread is improved at short lead-times, at longer lead-times the ensemble spread is too large leading to an underconfident ensemble. In summary, our method successfully propagates error information along the river network, enabling error correction at ungauged locations. This technique can be used for improved post-event analysis and can be developed further to post-process operational forecasts providing more accurate knowledge about the future states of rivers.

1 Introduction

River discharge forecasts are essential tools for taking effective preparatory actions for disaster mitigation and water resource planning (UNDRR, 2015)(UNDRR, 2015). However, despite the increased sophistication of forecasting systems over the past few decades, river discharge forecasts still contain uncertainty (Boelee et al., 2019). The uncertainty is introduced at several stages of the forecasting system including the meteorological forcings, the initial conditions, and the hydrological model structure and parameters (Valdez et al., 2022)(Valdez et al., 2022). Ensemble river discharge forecasts typically aim to account for the meteorological uncertainty by forcing a hydrological model with many meteorological forcings ~~either from multiple numerical weather prediction (NWP) systems or from an ensemble weather forecast created using multiple sets of initial conditions~~ (Cloke and Pappenberger, 2009; Wu et al., 2020). However, ensemble forecasts can still contain errors. Different methods for correcting these errors have been developed including pre-processing of the meteorological forcings, calibration of the hydrological model, improving the initial conditions using data assimilation, and post-processing of the river discharge forecast (Bourdin et al., 2012). Of these approaches post-processing is often considered the most computationally efficient and its ability to correct for multiple sources of errors simultaneously is appealing.

In meteorological forecasting, post-processing at non-observed locations is common (see Vannitsem et al., 2021). However, hydrological forecasting ~~also~~ requires consideration of the spatial heterogeneity introduced by the river network (e.g., Li et al., 2017; Woldemeskel et al., 2018; Ye et al., 2014; Xu et al., 2019; Liu et al., 2022; Lee and Ahn, 2024) ~~and the application of making hydrological post-processing methods at ungauged locations is still a difficult challenge. The lack of gauged locations along river networks is a particular problem as is the lack of agreed data sharing practices for the areas that are gauged (Lavers et al., 2019; Hannah et al., 2011), which means that global river gauge network is sparse (Krabbenhoft et al., 2022), and even in regions where gauges exist, river discharge data are often not widely shared (Lavers et al., 2019; Hannah et al., 2011).~~ Therefore, the development of post-processing techniques for ungauged locations is essential. However, current techniques are generally too computationally expensive for operational river flow forecasting applications (Emerton et al., 2016). For example, defining a joint distribution between the river discharge at multiple locations would allow forecasts to be conditioned on observations available at specific locations (Engeland and Steinsland, 2014)(Engeland and Steinsland, 2014). However, for large-scale distributed systems and multiple lead-times the size of the joint distribution quickly becomes too large. Alternatively, error-correction can be performed at a gauged location and the results interpolated to ungauged locations. One such method used to interpolate error-correction parameters is top-kriging (Pugliese et al., 2018; Skøien et al., 2021) (Pugliese et al., 2018; Skøien et al., 2021). Top-kriging takes into account the river network but the relationship between errors at different locations is assumed static regardless of the hydrometeorological situation (Skøien et al., 2016, 2006). Another option is to use a river routing model to propagate error-corrected river discharge forecasts between gauged locations ~~using a river routing model~~ (Bennett et al., 2022). Whilst this approach maintains spatial consistency between locations, the additional run of the model could be computationally expensive for an operational application.

The aim of this paper is to present and evaluate a novel technique ~~to spread for~~ spreading observation information from gauged to ungauged locations in a computationally efficient and temporally varying manner. The new method is based on

55 data assimilation techniques, ~~commonly used to improve the initial conditions of forecasts (Valdez et al., 2022), but applied as post-processing so that additional, computationally expensive executions of the hydrological model are not required.~~ Data assimilation is a mathematical technique that combines modelled predictions and observations to produce an improved modelled state relative to the true state of the system (Nichols, 2003, 2010). Data assimilation is often used to improve the initial conditions of forecasts (Valdez et al., 2022). However, in this paper we modify the techniques to apply them in a post-processing environment such that additional, computationally expensive, executions of the hydrological model are not required. The error correction method proposed in this study is based on state augmentation (Dee, 2005) and the Local Ensemble Transform Kalman Filter (LETKF, Hunt et al., 2007) ~~and state augmentation (Dee, 2005).~~ State augmentation is a technique that allows the estimation of the state and parameters/biases of a system simultaneously, and is often used for online bias-estimation in data assimilation (Ridler et al., 2018; Gharamti and Hoteit, 2014; Smith et al., 2013, 2009; Martin et al., 2002). The LETKF is part of the Kalman Filter family of methods and uses an ensemble of model states to estimate the state error covariances. Due to ~~its~~ ~~their~~ computational efficiency and ability to handle non-linear dynamics without an adjoint model, ensemble Kalman Filters are common data assimilation methods ~~for hydrological applications in hydrological research~~ (Rouzies et al., 2024; Li et al., 2023; Mason et al., 2020; Ridler et al., 2018; Khaki et al., 2017; Xie and Zhang, 2010; Clark et al., 2008). ~~State augmentation is a technique that allows the estimation of the state of a system and the parameters of the model used to simulate that system simultaneously (Ridler et al., 2018; Gharamti and Hoteit, 2014; Smith et al., 2013, 2009; Martin et al., 2002)~~

Whilst many studies have shown the benefits of data assimilation for hydrological forecasting (Tanguy et al., 2025; Valdez et al., 2022; Pi, , the process is rare in operational systems (Pechlivanidis et al., 2025), particularly in large-scale systems (Wu et al., 2020). This limited uptake is partly due to data latency issues (WMO, 2024), time constraints, and the potential impact on the interpretation of the forecasts (e.g., thresholds based on model climatology may no longer be consistent; Emerton et al., 2016). Additionally, the benefit of data assimilation at longer lead-times is uncertain (e.g., Valdez et al., 2022). In this paper, we leverage key advantages of data assimilation—such as the ability to propagate observational information to ungauged locations—within a post-processing framework that is more readily integrated into operational systems.

The proposed method aims to improve the skill of the ensemble mean and the reliability of the ensemble spread by adjusting each ensemble member, as will be discussed in more detail in Section 2. However, it is equally, if not more, important that the ensembles are spatially and temporally consistent in order to aid with decision making (Bennett et al., 2022). This is particularly important for large scale systems that provide forecasts across administrative boundaries, such as the Copernicus Emergency Management Service’s (CEMS) European Flood Awareness System (EFAS) used in this study (Matthews et al., 2025). The specific research questions to be addressed in this study are therefore,

1. Can data assimilation techniques be used in a post-processing environment to ~~spread-propagate~~ observation information to ungauged locations ~~in a spatiotemporally consistent manner?~~
2. Are the resulting ensemble predictions of river discharge more skillful than the raw ensemble?

This paper is organised as follows. In Section 2 we define the errors which we aim to correct and introduce some terminology and notation. In Section 3 we ~~formulate~~ ~~describe~~ the data assimilation techniques used within this study. In Section 4 we outline

the proposed error-correction method and detail how the ensemble is corrected. Section 5 provides some additional components of the method that improve the efficacy of the method but ~~which~~ can be adjusted to suit the data availability of any system and/or domain. Section 6 outlines the strategy used to evaluate the efficacy of the proposed method ~~and~~. Section 7 presents the results, first ~~assessing the ability of the method to spread observational information to ungauged locations~~ ~~investigating the impact of assimilating the observations~~, and then assessing the skill of the error-corrected ensembles. In Section 8 we discuss key features of the proposed method and their impact on the error-corrected ensembles. In Section 9 we conclude that the proposed method successfully improves the skill of the ensemble ~~and maintains spatiotemporal consistency~~ ~~mean~~, and highlight priorities for future developments.

Please note that throughout the paper ‘hindcast ensemble’ refers to the ensembles of river discharge that we are error-correcting. These ensembles are past operational EFAS forecasts (see Section 6.1); however, when we perform the error-correction we use observations that are available within the forecast (hindcast) period ~~which~~. This would not be possible in an operational system as these timesteps would be in the future. Therefore, we refer to these ensembles as hindcasts for clarity.

2 Ensemble error-correction framework

Here, we define the errors which we aim to correct and provide some notation that is used throughout the paper. ~~Where possible we follow the standard data assimilation notation provided in Ide et al. (1997).~~ Let the true state of the system at time k be defined as $\mathbf{x}_k^{true} \in \mathbb{R}^n$, where each element represents the true river discharge in one of the n grid boxes in the domain of interest. Hydrological forecasts ~~, including the EFAS forecasts used in this study (Section 6.1),~~ generally estimate the true state of the system ~~by using~~ a modelled state, denoted $\mathbf{x}_k$, where the lack of superscript ‘true’ indicates it is a modelled estimate. ~~Hydrological~~ ~~In this study, the hydrological~~ ensemble forecasts consist of N potential realizations of future river discharge, referred to as ensemble members. We define the ensemble river discharge hindcasts ~~used in this study~~ as

$$\left\{ \mathbf{x}_k : \mathbf{x}_k^{(i)}, \text{ for } i = 1, 2, \dots, N \text{ and } k = 0, 1, \dots, L \right\}. \quad (1)$$

where the superscript (i) indicates the i -th member of the ensemble, N is the ensemble size, the timestep k refers to the lead-time of the hindcast, and L is the maximum lead-time. The ensemble mean is defined as

$$\bar{\mathbf{x}}_k = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_k^{(i)} \in \mathbb{R}^n. \quad (2)$$

The ensemble perturbation matrix is defined as

$$\mathbf{X}_k = \begin{pmatrix} \mathbf{x}_k^{(1)} - \bar{\mathbf{x}}_k & \mathbf{x}_k^{(2)} - \bar{\mathbf{x}}_k & \dots & \mathbf{x}_k^{(N)} - \bar{\mathbf{x}}_k \end{pmatrix} \in \mathbb{R}^{n \times N} \quad (3)$$

where the i -th column represents the i -th ensemble member’s departure from the ensemble mean at lead-time k . The perturbation matrix contains information about the spread of the ensemble and the spatial structure of the deviations ~~from the mean~~ of each ensemble member ~~from the mean~~. From the definition of the perturbation matrix, the ensemble covariance matrix is

defined as

$$\mathbf{P}_k = \frac{1}{N-1} \mathbf{X}_k \mathbf{X}_k^T \in \mathbb{R}^{n \times n}. \quad (4)$$

120 where the superscript T indicates the matrix transpose.

~~Hydrological ensembles may still contain errors, so a post-processing step is usually necessary within a hydrological forecasting system. In this paper, we propose a method to spread an error-correction from gauged locations to every grid box in the system domain. We assume that there is an additive relationship between each hindcast ensemble member and an error vector such that~~

$$125 \quad \underline{\mathbf{x}_k^{true} = \mathbf{x}_k^{(i)} + \mathbf{b}_k^{(i)true} \in \mathbb{R}^n}$$

~~where $\mathbf{b}_k^{(i)true}$ is the error of the i -th domain. The proposed method estimates an additive error vector for each hindcast ensemble member with respect to the true state at each timestep. Each element of the error vector is represents the error associated with a single grid box. The proposed method estimates the additive error vector, $\mathbf{b}_k^{(i)}$ (where the lack of the superscript $true$ indicates it is an estimate) for each hindcast ensemble member at each timestep in the domain. Collectively,~~

130 ~~these error vectors form an ensemble defined as,~~

$$\left\{ \mathbf{b}_k^{(i)} \in \mathbb{R}^n \text{ for } i = 1, 2, \dots, N \right\} \quad (5)$$

where N is the same ensemble size as the river discharge hindcast, n is the number of grid-boxes in the hindcast domain, and k such that the resulting hindcast distribution, $\mathbf{x}_k^{new}$, is defined by

$$\underline{\mathbf{x}_k^{new} \sim \mathcal{N}(\bar{\mathbf{x}}_k + \bar{\mathbf{b}}_k, \mathbf{P}_k + \mathbf{\Gamma}_k)}$$

135 ~~where $\bar{\mathbf{x}}_k \in \mathbb{R}^n$ and $\mathbf{P}_k \in \mathbb{R}^{n \times n}$ are the ensemble mean is the timestep. The error ensemble mean, $\bar{\mathbf{b}}_k$, and the ensemble covariance matrix of the raw ensemble, respectively, $\bar{\mathbf{b}}_k \in \mathbb{R}^n$ is the ensemble mean of the estimated error vectors, and $\mathbf{\Gamma}_k$ perturbation matrix, $\mathbf{B}_k$, are calculated by substituting $\mathbf{b}_k^{(i)}$ in place of $\mathbf{x}_k^{(i)}$ in Eqs. (2) and (3), respectively. We assume there is an additive spread correction matrix. Defining an error-corrected ensemble in terms of mean bias and spread correction parameters is a common post-processing technique used, for example, in the Ensemble Model Output Statistics (EMOS Gneiting et al., 2005; Skoien et al., 2021) method relationship between each hindcast ensemble member and the corresponding error vector such that the i -th error-corrected ensemble member, $\mathbf{x}_k^{new,(i)}$, is defined as~~

$$\mathbf{x}_k^{new,(i)} = \mathbf{x}_k^{(i)} + \mathbf{b}_k^{(i)} \in \mathbb{R}^n. \quad (6)$$

The estimation of the error ensemble at each timestep is described in Section 4.

To aid with the estimation of the error vectors, we assume that at each timestep the system is observed at p_k river discharge gauges ~~such that we have a vector of river discharge observations, $\mathbf{y}_k \in \mathbb{R}^{p_k}$. We assume the observation vector, $\mathbf{y}_k \in \mathbb{R}^{p_k}$, is related to the true state of the system as~~

$$\mathbf{y}_k = \mathbf{H}_k(\mathbf{x}_k^{true}) + \boldsymbol{\epsilon}_k \quad (7)$$

where $\epsilon_k \in \mathbb{R}^{p_k}$ is a vector of unbiased Gaussian noise with covariance matrix $\mathbf{R}_k \in \mathbb{R}^{p_k \times p_k}$, such that $\epsilon_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$, and $\mathbf{H}_k \in \mathbb{R}^{p_k \times n}$ is the linear observation operator which. The observation operator maps the variables from the state space to observation space. The observation operator used in this study selects the grid boxes from within the modelled drainage network of the hydrological model that represent the location that correspond to the locations of the river gauges.

3 Data Assimilation

As discussed in Section 1, the proposed method is based on common data assimilation techniques: state augmentation and the Local Ensemble Transform Kalman Filter (LETKF). In this section we present the formulations, we provide an overview of these techniques used in this study and introduce the necessary equations. In Section 4, we adapt and apply these methods in a non-standard way due to their application in a post-processing environment.

3.1 State augmentation

In the proposed method, an ensemble of augmented states is defined between the ensemble river discharge hindcast (see Section 2) and an ensemble of additive error vectors. We define this ensemble of error vectors at time k as

$$\left\{ \bar{\mathbf{b}}_k^{(i)} \in \mathbb{R}^n \text{ for } i=1, 2, \dots, N \right\}$$

where N is State augmentation is a technique used for online bias-correction in data assimilation that allows the simultaneous estimation of the same ensemble size as the river discharge hindcast and n is the number of grid-boxes in the hindcast domain. The error ensemble mean, $\bar{\mathbf{b}}$, and the ensemble perturbation matrix, $\mathbf{B}$, are calculated by substituting $\bar{\mathbf{b}}_k^{(i)}$ in place of $\mathbf{x}_k^{(i)}$ in Eqs. (2) and (3), respectively. The generation of the initial error ensemble for timestep $k=1$ is described in Section 5.3.

The ensemble of augmented states is then defined such that the system state and biases. An augmented state is defined by appending the biases to the state vector, allowing both to be updated by the data assimilation method. In this study, the i -th ensemble member member of the augmented ensemble is defined as

$$\mathbf{w}_k^{(i)} = \begin{pmatrix} \mathbf{x}_k^{(i)} \\ \mathbf{b}_k^{(i)} \end{pmatrix} \in \mathbb{R}^{2n}. \quad (8)$$

where $\mathbf{x}^{(i)} \in \mathbb{R}^n$ and $\mathbf{b}^{(i)} \in \mathbb{R}^n$ are the i -th hindcast and error ensemble members, respectively. The augmented ensemble mean and perturbation matrix are given by

$$\bar{\mathbf{w}}_k = \begin{pmatrix} \bar{\mathbf{x}}_k \\ \bar{\mathbf{b}}_k \end{pmatrix} \in \mathbb{R}^{2n} \quad \text{and} \quad \mathbf{W}_k = \begin{pmatrix} \mathbf{X}_k \\ \mathbf{B}_k \end{pmatrix} \in \mathbb{R}^{2n \times N} \quad (9)$$

where $\bar{\mathbf{x}}$ and $\bar{\mathbf{b}}$ are the ensemble means of the hindcast and error ensembles, respectively, and $\mathbf{X}$ and $\mathbf{B}$ are the perturbation matrices of the hindcast and error ensembles, respectively.

175 ~~The next step of~~ In this study, state augmentation is used within an LETKF (described in Section 3.2) and it is therefore necessary to define the evolution of the augmented states between timesteps. The ~~hindcast is evolved by the evolution of the hindcast~~ and error ensembles determines the evolution of the augmented states. The hindcasts used in this study were generated using the LISFLOOD hydrological model ~~(the hydrological model used to create the EFAS forecasts; Van Der Knijff et al., 2010). For the~~, which is used in the EFAS operational system (Van Der Knijff et al., 2010). As the true evolution of the error vectors at all
180 grid-boxes is unknown, we assume a simple persistence model, ~~we adopt the common assumption that the error is constant between timesteps (Martin, 2001),~~ such that

$$\mathbf{b}_k^{(i)} = \mathbf{b}_{k-1}^{(i)}$$

~~Based on these independent evolution equations and~~ $\mathbf{b}_k^{(i)} = \mathbf{b}_{k-1}^{(i)}$. This is a common assumption used in state augmentation (Pauwels et al., 2020; Ridler et al., 2018; Rasmussen et al., 2016; Martin, 2001). Based on the independent evolution of
185 the hindcast and error ensembles, and the additive relationship between ~~the hindcast ensemble members and the error ensemble members (see their members~~ (Eq. (6)), we define the propagation of the augmented ensemble members as

$$\mathbf{w}_k^{(i)} = \begin{pmatrix} \mathbf{M}_{k-1} & \mathbf{I}_{k-1} \\ \mathbf{0}_{k-1} & \mathbf{I}_{k-1} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{k-1}^{(i)} \\ \mathbf{b}_{k-1}^{(i)} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_k^{(i)} + \mathbf{b}_{k-1}^{(i)} \\ \mathbf{b}_{k-1}^{(i)} \end{pmatrix}. \quad (10)$$

where $\mathbf{M}_{k-1} \in \mathbb{R}^{n \times n}$ is a linear evolution operator representing the LISFLOOD hydrological model ~~, $\mathbf{I}_{k-1}^{n \times n}$ and $\mathbf{I}_{k-1} \in \mathbb{R}^{n \times n}$~~ is the identity matrix ~~acting on the error component of the augmented state, and $\mathbf{x}_k^{(i)} \in \mathbb{R}^n$ is the i -th member of the precomputed~~
190 ~~the hindcast ensemble~~. Since we use precomputed hindcast ensembles the propagation of the hindcast ensemble members requires no additional computation ~~and the~~. The full non-linear LISFLOOD hydrological model is also used without the need to define a linear approximation.

3.2 Local Ensemble Transform Kalman Filter (LETKF)

The Local Ensemble Transform Kalman Filter ~~(LETKF; Hunt et al. (2007))~~ (LETKF; Hunt et al., 2007) updates the mean state
195 and the ~~square root of the covariance perturbation~~ matrix of an ensemble ~~(i.e., the perturbation matrix)~~ by combining the modelled and observed data. As a sequential data assimilation method, the LETKF consists of a *propagation step* (also known as a *forecast step*) and an *update step* (also known as an *analysis step*) that are ~~iterated. We cycled.~~ In this method, we use the LETKF to update the ensemble of error vectors at each ~~timestep but hindcast~~ timestep for which observations are available. However, we modify the propagation step to use precomputed hindcasts. The propagation step evolves the augmented states
200 forward in time from time $k-1$ to k , as described in Eq. (10). Rather than evolve the hindcast ensemble explicitly (which would require the hydrological model) ~~we instead substitute, we use~~ the precomputed hindcast ~~ensemble for at~~ timestep k ~~into the propagated augmented state.~~

The update step of the LETKF calculates the ~~optimal estimate of the~~ state of the system at timestep k by combining the modelled augmented states and observations, ~~both~~. Both data are weighted by their respective uncertainties, represented by

205 their covariance matrices. As the LETKF is a well documented method we only provide the key update equations and. For more detailed derivations, we direct the reader to Hunt et al. (2007) and Livings et al. (2008) for more detailed derivations.

Livings et al. (2008). To apply the LETKF to the augmented ensemble we extend the definition of the observation operator, $\mathbf{H} \in \mathbb{R}^{p \times n}$, given in Eq. (7) such that, we create a *model-observation ensemble* with an ensemble mean, $\bar{\mathbf{y}}_k^x$, defined as

$$\bar{\mathbf{y}}_k^x = (\mathbf{H}_k \ 0) \bar{\mathbf{w}}_k = \mathbf{H}_k^{(i)} \bar{\mathbf{x}}_k + \mathbf{H}_k^{(i)} \bar{\mathbf{b}}_{k-1} \quad (11)$$

210 where $\hat{\mathbf{H}} \in \mathbb{R}^{p \times 2n}$ is the augmented observation operator. As discussed in Section 2, the observation operator maps state variables from state space to observation space by extracting data for the appropriate grid-boxes. $\mathbf{H}_k \in \mathbb{R}^{p \times 2n}$ is the observation operator defined in Eq. (7). The LETKF can then update the augmented ensemble mean, $\bar{\mathbf{w}}_k$, such that,

$$\bar{\mathbf{w}}_k^a = \bar{\mathbf{w}}_k^f + \begin{pmatrix} \mathbf{K}_{\mathbf{x}_k} \\ \mathbf{K}_{\mathbf{b}_k} \end{pmatrix} (\mathbf{y}_k - \bar{\mathbf{y}}_k^{x,f}), \quad (12)$$

where the ~~subscripts~~ ^{superscripts} f and a indicate the state before and after the update step, respectively; $\mathbf{K}_{\mathbf{x}_k} \in \mathbb{R}^{n \times p}$ and $\mathbf{K}_{\mathbf{b}_k} \in \mathbb{R}^{n \times p}$ are the components of the Kalman gain matrix acting on the hindcast ensemble and the error ensemble respectively; and $\mathbf{y}_k \in \mathbb{R}^p$ is the observation vector defined in Eq. (7). The difference between the observations and the model state in observation space (i.e., $\mathbf{y}_k - \hat{\mathbf{H}}_k \bar{\mathbf{w}}_k^f$) is called the innovation vector. The Kalman gain matrix ~~weights the prior modelled state and the observations based on their respective uncertainties and~~ determines the impact of the innovation vector in the update step. ~~The respective uncertainties of the prior modelled state and the observations determine their weight within the LETKF.~~ Large observation uncertainties reduce the Kalman gain, while large uncertainties in the prior state increase the Kalman gain. Both the hindcast and the error components of the Kalman gain are functions of the covariance matrix of the augmented ensemble (see Eqs. (8) and (9) in Bell et al., 2004) (see Appendix A; Bell et al., 2004). The covariance matrix describes the state error covariances between grid-boxes allowing the Kalman gain to spread the observation information to ungauged locations. To update the error component specifically, it is the cross-covariances between the error component and the hindcast component that control the spread of the observation information to ungauged locations (see Eq. (9) in Bell et al., 2004). This ability to spread the observational information is key to the error-correction method presented in this study.

The LETKF updates the augmented ensemble perturbation matrix, $\mathbf{W}_k$, such that,

$$\mathbf{W}_k^a = \mathbf{W}_k^f \mathbf{T}_k \quad (13)$$

where $\mathbf{T}_k \in \mathbb{R}^{N \times N}$ is the square root transform matrix (Livings et al., 2008). The square root transform matrix is derived using the Kalman gain matrix which gives the weighting between the modelled state and the observations (Livings et al., 2008). Using an eigenvector decomposition, the square root transform matrix rescales and rotates the ensemble members such that the updated perturbation matrix represents the uncertainty in the updated ensemble mean. The square root transform matrix allows the covariance matrix of the ensemble to be updated without the need for the covariances to be explicitly calculated which can be computationally expensive (Bishop et al., 2001; Hunt et al., 2007). These update equations are used to update the error component only as will be discussed in Section 4.1.

4 Spatially consistent error-correction method for river discharge

In this section, we describe how we use the data assimilation techniques discussed in Section 3, ~~namely state augmentation and the LETKF, to correct to post-process~~ the hindcasts across the domain, including at ungauged locations (Fig. 1). The correction is applied in a post-processing environment, avoiding the need for additional executions of the hydrological model which can be computationally expensive. ~~The proposed method consists of two steps: 1) updating~~ In Section 3.1, we describe how the error ensemble (defined in is updated at every timestep. In Section 3.1), and 2) ~~adjusting the hindcast ensemble members using the~~, we describe how updated error ensemble (Fig. 1) is used to error-correct the hindcast ensemble. Specific experimental design choices are discussed in Section 5.

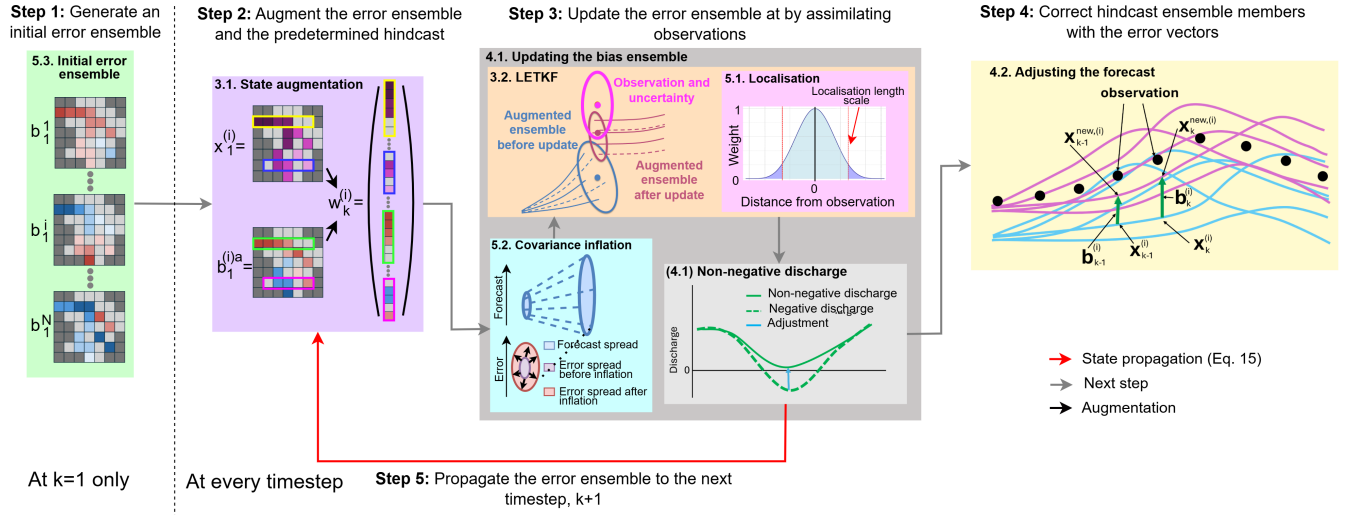


Figure 1. Schematic of the new error-correction method for gauged and ungauged locations. Coloured boxes indicate different components of the method. An initial error ensemble is created for timestep $k=1$ (green box). Then, the error ensemble is augmented to the hindcast ensemble (purple box). At each timestep the covariance of the augmented ensemble is inflated (cyan box) before being updated using the LETKF which uses localisation to improve the results of the update (collectively the orange box). The updated error ensemble is adjusted to ensure non-negative discharge values (light grey box) before being used to error-correct the hindcast (yellow box). The non-negative error ensemble is propagated to the next timestep (red arrow). More details are provided for each component in the section indicated in the top left corner of the corresponding box.

4.1 Updating the error ensemble

At each timestep the error ensemble is updated to estimate the optimal set of error vectors to correct the hindcast at that timestep. The update is performed using the LETKF defined in Section 3.2. ~~Equations (A2) and (13) are the Kalman update equations for the augmented state.~~ Using the definition of the augmented state (see in Eq. (8)) the update equations for the

error ensemble only are,

$$\bar{\mathbf{b}}_k^a = \bar{\mathbf{b}}_k^f + \mathbf{K}_{\mathbf{b}_k} (\mathbf{y}_k - \bar{\mathbf{y}}_k^{\mathbf{x}} \bar{\mathbf{b}}_k^f) \quad (14)$$

250 and

$$\mathbf{B}_k^a = \mathbf{B}_k^f \mathbf{T}_k. \quad (15)$$

As the hindcast component is not explicitly evolved (see Section 3.2), we assume that the raw hindcast is a good approximation for the hindcast analysis state ~~were the component if the component were to be~~ updated. This allows the substitution of the precomputed hindcast in place of the propagated state at the next timestep. Thus, the updated mean of the augmented
255 ensemble can be defined as

$$\bar{\mathbf{w}}_k^a = \begin{pmatrix} \bar{\mathbf{x}}_k \\ \bar{\mathbf{b}}_k^a \end{pmatrix} \in \mathbb{R}^{2n}. \quad (16)$$

where $\bar{\mathbf{x}}_k$ is the ensemble mean of the raw hindcast ensemble and $\bar{\mathbf{b}}_k^a$ is the updated error ensemble mean (Eq. (14)). The perturbation matrix of the updated augmented ensemble follows a similar pattern such that

$$\mathbf{W}_k^a = \begin{pmatrix} \mathbf{X}_k \\ \mathbf{B}_k^a \end{pmatrix}. \quad (17)$$

260 where $\mathbf{X}_k$ is the ensemble perturbation matrix of the raw hindcast ensemble and $\mathbf{B}_k^a$ is the updated error ensemble perturbation matrix (Eq. (15)). The assumptions made in EqEqs. (16) and Eq-(17) make our system **suboptimal. However, we provide sub-optimal** from a data assimilation perspective but are necessary to avoid rerunning the hydrological model. Importantly, we aim to estimate the error of the precomputed model output at each lead time. Therefore, while the lack of state evolution makes the hindcast component update sub-optimal, the update of the error ensemble remains mathematically consistent. In this study, we
265 **provide** proof-of-concept ~~in this study~~ that the resulting error ensemble improves the skill of the hindcast (see Section 7.1.7.1).

The Kalman filter is not constrained to enforce non-negativity of the analysis state, and therefore, could lead to negative discharge values for some grid boxes if the cross-covariances are incorrectly defined. We enforce non-negativity by further adjusting the error ensemble members after the LETKF update step (Fig 1). ~~The adjustment is done separately for each grid box and each ensemble member only if they result in a negative river discharge as follows:-~~

$$270 \quad \underline{\text{if } \mathbf{x}_k^{(i)}[j] + \mathbf{b}_k^{(i)a}[j] < 0; \quad \text{then } \hat{\mathbf{b}}_k^{(i)a}[j] = -\mathbf{x}_k^{(i)}[j] + \|\zeta_k\|}$$

~~where $\hat{\mathbf{b}}_k^{(i)}$ is the adjusted error ensemble member that results in non-negative discharge, j indicates the j -th grid box, i indicates the i -th ensemble member, $\|\cdot\|$ indicates the modulus, and ζ_k a random noise value.~~ For any ensemble member and grid box where the sum of the hindcast discharge and the updated error is negative, we modify the error value so that the total becomes a small positive value, mitigating the potential for instabilities caused by zero-values. This small positive value is

275 sampled from a Gaussian distribution with ~~mean 0 and a~~ mean of zero and a standard deviation equal to 10% of the standard deviation of the updated error ensemble at the grid-box of interest.

The updated positive-definite augmented states are propagated to the next timestep as defined in Eq. (10). The updated positive-definite augmented states are also used to error-correct the hindcast (Section 4.2).

4.2 Adjusting the forecast

280 After the error component of the augmented state has been updated using Eqs. (14) and (15), and non-negativity has been enforced (Section 4.1), the error ensemble members are added to the respective hindcast ensemble members such that

$$\mathbf{x}_k^{new,(i)} = \mathbf{x}_k^{(i)} + \hat{\mathbf{b}}_k^{(i)a} \quad (18)$$

where $\mathbf{x}_k^{new,(i)}$ and $\mathbf{x}_k^{(i)}$ are the i -th ensemble members of the error-corrected and raw hindcast ensembles, respectively, and $\hat{\mathbf{b}}_k^{(i)}$ is the error vector associated with the i -th error ensemble member where the caret indicates a non-negativity check has
285 been applied. Consequently, the error-corrected hindcast ensemble mean and perturbation matrix are given by

$$\bar{\mathbf{x}}_k^{new} = \bar{\mathbf{x}}_k + \bar{\mathbf{b}}_k^a \quad (19)$$

and

$$\mathbf{X}_k^{new} = \mathbf{X}_k + \hat{\mathbf{B}}_k^a. \quad (20)$$

~~The-This update results in an additive spread correction matrix defined in Eq. (??) is the result of calculating the covariance~~
290 ~~matrix of the error-corrected hindcast ensemble as in Eq. (3). The form of the additive spread correction matrix is-~~ Γ_k , with the form

$$\Gamma_k = \mathbf{X}_k \hat{\mathbf{B}}_k^{aT} + \hat{\mathbf{B}}_k^a \mathbf{X}_k^T + \hat{\mathbf{B}}_k^a \hat{\mathbf{B}}_k^{aT} \quad (21)$$

where $\mathbf{X}_k$ and $\hat{\mathbf{B}}_k^a$ are the perturbation matrices of the raw hindcast and error ensembles, respectively, and the superscript T indicates the matrix transpose (Section 5.2 in Martin, 2001).

295 5 Experimental implementation

In Section 4 we presented a new method of spreading observation information to ungauged locations in a post-processing environment based on common data assimilation techniques. In this section, we describe three key components of the method—localisation, covariance inflation, and the generation of the initial error ensemble—which are crucial for its performance but can be implemented in various ways.

300 5.1 Localisation

Localisation is used to reduce the effect of spurious correlations which can arise due to sampling errors caused by the small ensemble size (Hamill et al., 2001; Hunt et al., 2007). The LETKF uses observation localisation which reduces the impact of

observations by multiplying the inverse of the observation-error covariance matrix by a *localisation matrix*, $\rho \in \mathbb{R}^{p_k \times p_k}$, such that

$$\mathbf{R}^{-1} = \rho \circ \mathbf{R}_{nl}^{-1} \quad (22)$$

where $\mathbf{R} \in \mathbb{R}^{p_k \times p_k}$ is the localised observation-error covariance matrix used in the LETKF (Section 3.2), $\mathbf{R}_{nl} \in \mathbb{R}^{p_k \times p_k}$ is the non-localised observation-error covariance matrix, and the symbol $\circ$ indicates the Schur product (also known as the Hadamard product) which is an element-wise matrix multiplication (Golub and Van Loan, 2013). We assume that $\mathbf{R}_{nl}$ and, by definition, $\mathbf{R}$ are diagonal matrices. In this study we use distance-based localisation so the impact of the multiplication described in Eq. (22) is to increase the effective uncertainty of distant observations and thus decrease their impact on the analysis state. The impact of the localisation on the spatial extent of the analysis increments is demonstrated in Section 7.1.

The localisation matrix is defined using the Gaspari-Cohn function which has a parameter called the *localisation length scale* (Eq. 4.10 in Gaspari and Cohn (1999)). (Appendix ??; Gaspari and Cohn, 1999). The Gaspari-Cohn function smoothly decreases the weights assigned to an observation as the distance from the observation location increases, starting from a value of 1 at the observation location and reaching 0 for distances greater than twice the localization length scale (pink box, Fig. 1). In this study, the distance is calculated along the river network which has been shown to improve the analysis for fluvial applications (García-Pintado et al., 2015; El Gharamti et al., 2021; Khaniya et al., 2022)(García-Pintado et al., 2015; El Gharamti et al., 2021; Khaniya et al., 2022). The distance between a grid-box and the location of an observation is calculated using the local drainage direction map and the channel length used in the hydrological model (Choulga et al., 2023). As the distance is defined along the river network, observations cannot impact grid-boxes in a different drainage basin.

~~The localisation length scale is often a tuned parameter but the tuning process can be time and resource intensive. We~~ Sensitivity experiments conducted during the development of this method found that the optimal length scale varied by location, lead-time, and tuning metric of choice, but overall, the differences were small for length scales from 65 km to 786 km (not shown). Therefore, we propose instead for the localisation length scale to be defined as the maximum distance between any grid point and its closest observation. This 1) ensures that all grid boxes are updated in the update step of the LETKF reducing the potential for discontinuities in the analysis state, 2) can adapt to changes in the availability of observations, and 3) can be applied to different domains and hydrological model configurations without requiring a tuning experiment.

5.2 Covariance inflation

Small ensemble sizes can cause underestimation of the ensemble spread ~~which in turn reduces~~ reducing the impact of the observations on the analysis (Furrer and Bengtsson, 2007). ~~In addition to the issues caused by the small ensemble size, we also make the simplified assumption that the~~ Additionally, we assume the error ensemble is constant between timesteps (Eq. (??)) ~~which~~ which, while simplifying implementation, could introduce model errors into the ensemble (Evensen et al., 2022). Covariance inflation is an approach often used to ~~To~~ ameliorate these issues, ~~although an inflation method that is optimal for all situations has yet to be identified (Due et al., 2020; Scheffler et al., 2022).~~

335 We aim to inflate the ensemble perturbation matrix such that at time $k + 1$ the spread better represents the true uncertainty of the mean error prior to the update step. Due to the unusual approach of using predefined ensembles, we propose a new method to inflate the covariance of the error ensemble. Various covariance inflation techniques are often used (Duc et al., 2020; Scheffler et al., 2022). We take inspiration from the ‘relaxation to prior perturbations’ technique (RTPP, Zhang et al., 2004; Kotsuki et al., 2017) which blends the analysis perturbation matrix with the perturbation matrix prior to the analysis step. This results in both additive and multiplicative inflation that is proportional to the impact of the assimilation of observations (Whitaker and Hamill, 2012). Rather than accounting for errors introduced in the update step we want to account for errors introduced primarily in the propagation step. We therefore adapt the RTPP method to blend the propagated analysis perturbation matrix, $\mathbf{W}_k^a$ defined in Eq. (17), and an alternative estimate of the perturbation matrix at the next timesteps, $\mathbf{W}_{k+1}^{est}$, such that

$$\mathbf{W}_{k+1}^{inf} = (1 - \alpha) \begin{pmatrix} \mathbf{M}_k & \mathbf{I}_k^{n \times n} \\ \mathbf{0}_k & \mathbf{I}_k^{n \times n} \end{pmatrix} \mathbf{W}_k^a + \alpha \mathbf{W}_{k+1}^{est}$$

345 implement a heuristic covariance inflation method inspired by the *relaxation-to-prior perturbations* technique (Zhang et al., 2004; Kotsuki et al., 2017). However, as we are working within a post-processing context, we adapt the method for use with predefined ensembles (i.e., without evolving the inflated perturbations between timesteps).

We blend the prior perturbation matrix at $k + 1$ with an estimated perturbation matrix $\mathbf{W}_{k+1}^{est}$ similar to the use of a climatological covariance matrix in Valler et al. (2019). The resulting perturbation matrix is given by

$$350 \quad \mathbf{W}_{k+1}^{inf} = (1 - \alpha) \begin{pmatrix} \mathbf{M}_k & \mathbf{I}_k \\ \mathbf{0}_k & \mathbf{I}_k \end{pmatrix} \mathbf{W}_k^a + \alpha \mathbf{W}_{k+1}^{est} \quad (23)$$

where α is an inflation parameter to be defined (and the definition of the matrices $\mathbf{M}_k$ and $\mathbf{I}_k^{n \times n}$ are evolution matrices introduced. $\mathbf{I}_k$ are given in Section 3.1 and α is an inflation parameter that must be defined. The estimation of $\mathbf{W}_{k+1}^{est}$ could use an alternative model to evolve the analysis perturbation matrix forward or be a climatological matrix (Valler et al., 2019). In this study we give $\mathbf{W}_{k+1}^{est}$ the form

$$355 \quad \text{as} \quad \mathbf{W}_{k+1}^{est} = \begin{pmatrix} \mathbf{X}_{k+1}^{est} + \mathbf{B}_{k+1}^{est} \\ \mathbf{B}_{k+1}^{est} \end{pmatrix}. \quad (24)$$

where $\mathbf{X}_{k+1}^{est}$ and $\mathbf{B}_{k+1}^{est}$ can be estimated separately. We assume the hindcast covariance matrix is correct such that $\mathbf{X}_{k+1}^{est} = \mathbf{X}_{k+1}$, and estimate the error covariance matrix at time $k + 1$ as proportional to the state covariance matrix (Dee, 2005; Martin et al., 2002). We assume that the constant of proportionality is 1 such that $\mathbf{B}_{k+1}^{est} = \mathbf{X}_{k+1}$. When substituted into Eq. (23), this form of $\mathbf{W}_{k+1}^{est}$ maintains consistency between the error terms in the hindcast and error components of the augmented state ensemble.

The inflation parameter, α_k , controls the weighting of propagated analysis perturbation matrix and. During development, it was found that the estimated matrices must have spatial structures consistent with the river network and be forecast and lead-time-dependent. For simplicity, and as the raw hindcast perturbations satisfy these requirements, we set both $\mathbf{X}_{k+1}^{est}$ and $\mathbf{B}_{k+1}^{est}$

equal to the raw hindcast perturbation matrix (Dee, 2005; Martin et al., 2002). In future studies, the ~~estimated perturbation~~
 365 ~~matrix~~. Here, estimated perturbation matrices could be defined using alternative models to evolve the analysis perturbation
 matrix between timesteps or be climatological matrices (Valler et al., 2019).

The inflation parameter α_k ~~determines how much of the uncertainty at time $k+1$ is due to uncertainty at time k and how~~
~~much is not captured by the propagated matrix~~. We assume that the change in the variance of the hindcast ensemble between
 timesteps is an indication of how much the spread of the error ensemble would change if the propagation model of the error
 370 ~~was correct~~. Therefore controls the weighting between the prior and estimated matrices. To account for changing uncertainty
 across lead-times and forecasts, we define ~~the inflation parameter as the fractional change in the hindcast variance~~, α_k using a
 smoothed estimate of the relative change in hindcast ensemble variance

$$\alpha_k = \frac{1}{k} \sum_{l=k-2}^{l=k} \max \left\{ \frac{|Tr(\mathbf{P}_l) - Tr(\mathbf{P}_{l+1})|}{Tr(\mathbf{P}_l)}, 1 \right\} \quad (25)$$

where k is the ~~the current timesteps and $Tr(\mathbf{P}_l)$~~ current timestep and $Tr(\mathbf{P}_l)$ is the trace of the raw hindcast covariance matrix
 375 at ~~timesteps~~ timestep l . A maximum value of 1 is set to avoid instabilities, particularly at short lead-times where the change in
 variance between timesteps can be large. The average over the past three timesteps is taken to ensure that ~~alpha~~ α is smoothly
 changing between timesteps, again to avoid instabilities. ~~An inflation value of 1 suggests the uncertainty of the modelled~~
~~state has changed so much between timesteps that the uncertainty at the previous timestep is no longer relevant. An inflation~~
~~values of 0 implies the uncertainty at the previous timesteps should be trusted~~. This approach of estimating α was selected
 380 after sensitivity testing (not shown for brevity), which indicated that the inflation factor must be both lead-time dependent and
 forecast dependent. While α is not spatially varying, it is applied to perturbation matrices with spatial structures consistent
 with the river network, ensuring physically plausible ensemble perturbations.

5.3 ~~Initial~~ Initialising the error ensemble for the first timestep

We must define an initial error ensemble to perform the state augmentation at the first timestep. ~~Due to the application to a~~
 385 ~~In a forecast~~ post-processing environment there is no “warm-up” period in which a state of equilibrium can be reached, and
 therefore the initial error ensemble must be physically plausible. Here, the initial error ensemble is defined using three sets of
 river discharge data: in-situ observations, $\mathbf{y}_k \in \mathbb{R}^{p_k}$, simulations created by forcing a hydrological model with meteorological
 observations, $\mathbf{s}_k \in \mathbb{R}^n$, and the ensemble mean and ensemble perturbation matrix of a single lead-time from a previous hindcast;
 ~~$\bar{\mathbf{x}}_k \in \mathbb{R}^n$ and $\mathbf{X}_k \in \mathbb{R}^{n \times N}$~~ . A single ensemble is generated for the full EFAS domain and then the elements associated with the
 390 domain of interest (in this study the Rhine-Meuse catchment) are extracted.

~~We define the initial error ensemble in two steps: 1) the error mean is estimated based on the errors due to the hydrological~~
~~model, and 2) ensemble perturbations are estimated based on the perturbations of hindcast members. The~~ The estimation has
 two main steps: estimating the mean ~~of the initial error ensemble is estimated error~~ and generating the perturbations around
 that mean. The ensemble mean is intended to capture biases in the hydrological model at the initial time. It is computed as
 395 follows:

1. **Calculate the errors at gauged locations:** For each river gauge location, we calculate the average relative error of the simulation compared to the observations between observed and simulated river discharge over the past d days (here 10 days) at all p_k stations with available observations, $\delta \in \mathbb{R}^{p_k}$ is calculated as

$$\delta[j] = \frac{\sum_{k=-d}^{k=1} \frac{y_k[j] - s_k[j]}{s_k[j]}}{10}$$

where $\delta[j]$. To limit the influence of outliers or representation errors, $y_k[j]$ and $s_k[j]$ are the relative error, the observation and the simulated value at station j at time k , respectively. If the value of $\delta[j]$ is greater than 1 (or less than -1) then $\delta[j]$ is set to 1 (or -1). This reduces the impact of representation errors due to the mapping of stations, for example, these errors are capped at ± 100

2. **Interpolate the errors to ungauged locations:** Inverse distance weighted interpolation is used to estimate the average relative error at ungauged locations ensuring closer stations have a greater influence (Lu and Wong, 2008). The Euclidean distance, denoted d_{gj} , is calculated between a grid-box, g , and each of the closest G stations (here 100 stations). Using inverse distance weighting, we interpolate the errors from gauged to ungauged locations. The value at each grid-box is a weighted average of relative errors from the 100 stations, and the average relative error weighted accordingly. The Euclidean distance is used here to allow the method to be applied to all catchments. Therefore, the inverse distance weighted formula used to calculate the relative error at grid-box g , denoted $\Delta[g]$, is

$$\Delta[g] = \frac{\sum_{j=1}^{100} \delta[j] / \sqrt{d_{gj}}}{\sum_{j=1}^{100} 1 / \sqrt{d_{gj}}}$$

nearest stations, with closer stations given more influence (Lu and Wong, 2008). All available stations, including those outside the catchment of interest, are used in this calculation to capture spatial variability.

3. **Impose the river network structure:** The mean of the initial error ensemble is calculated by multiplying the field of estimated relative errors, Δ , with the simulation at time $t = -1$, s_{-1} , such that at grid-box g the initial error ensemble mean, $\bar{\mathbf{b}}_1^f[g]$ is calculated as

$$\bar{\mathbf{b}}_1^f[g] = \Delta[g] \times s_{-1}$$

where the superscript f indicates the ensemble has not been updated by the LETKF (Section 3.2) interpolated error field is then multiplied by the simulated river discharge values at each grid point. This enforces the spatial pattern structure of the river network by ensuring the value of the initial error mean is, ensuring errors are proportional to the magnitude of the discharge in the size of the river.

The perturbations from the ensemble mean are then defined as follows. Since the form of the covariance matrix of the errors Since the true error covariance is unknown, we use a common technique of scaling the system state covariance matrix (e.g. Martin et al., 2002). We use the ensemble members assume a reasonable estimate can be derived from a previous river discharge ensemble forecast as follows:

1. **Calculate the ensemble statistics:** We calculate the ensemble mean and perturbation matrix from the second lead-time of the hindcast from a hindcast issued two days prior. This choice avoids unrealistically low spread often seen at the first lead-time was selected as the spread of the ensemble at due to a lead-time of one day can often be very narrow due to the use of a single set of initial conditions (see 6.1). The steps to define the initial error perturbations are:-

430 2. **Calculate the ensemble statistics:** The ensemble mean and perturbation matrix of the the second lead-time of the hindcast from two days prior are calculated. The valid-time for these ensemble values is, $t = 0$ so they are denoted $\bar{\mathbf{x}}_0$ and $\mathbf{X}_0$, respectively.

3. **Inflate the covariance matrix:** The spread of the hindcast ensemble should account for the uncertainty due to the meteorological forcings. To ensure the variance is correct we scale the perturbation matrix by perturbation matrix is adjusted by calculating the error of the ensemble mean compared to the hindcast ensemble mean at each grid-box relative to a simulation forced by meteorological observations. A vector This provides a set of scaling factors, $\mathbf{f}$ is defined such that

$$\mathbf{f} = \left(\frac{s_0[1] - \bar{x}_0[1]}{\bar{x}_0[1]}, \frac{s_0[2] - \bar{x}_0[2]}{\bar{x}_0[2]}, \dots, \frac{s_0[n] - \bar{x}_0[n]}{\bar{x}_0[n]} \right)^T$$

440 where $s_0[n]$ is the simulation at the n -th gridbox. These scaling factors can be very small so we set a requirement that the scaled standard deviation at each grid-box must not go below used to inflate the perturbation matrix. To avoid underestimating uncertainty, we impose a minimum threshold of on the resulting standard deviation of 10% of the simulated value, $s_0[j]/10$. In practice this is done in two steps: i) scale the perturbation matrix by the vector $\mathbf{f}$ such that

$$\mathbf{B}_\pm^f = \mathbf{f} \mathbf{X}_0$$

445 and then ii) reinflate the spread where necessary. By scaling the perturbation matrix in this way we relate the spatial variability of the ensemble spread to the error due to the meteorological forcings.

local simulated river discharge.

The error mean, $\bar{\mathbf{b}}_1^f$, and the error perturbation matrix, $\mathbf{B}_1^f$ are used to define the error ensemble at timesteps $k = 1$ where they are

450 The resulting error ensemble mean and perturbations define the initial ensemble, which is then updated using the LETKF with state augmentation, as described in Section 4.1.

6 Evaluation strategy

6.1 European Flood Awareness System (EFAS)

The hindcasts used in this study were produced by the European Flood Awareness System (EFAS) as operational forecasts (Barnard et al., 2020). EFAS is part of the Early Warning component of the European Commission's Copernicus Emergency

Management Service (CEMS), and aims to provide complementary forecast information to hydro-meteorological services throughout Europe (Matthews et al., 2025). EFAS streamflow forecasts are produced by forcing a calibrated hydrological model, LISFLOOD (De Roo et al., 2000; Van Der Knijff et al., 2010; Arnal et al., 2019) (De Roo et al., 2000; Van Der Knijff et al., 2010; , with the output from meteorological numerical weather prediction (NWP) systems. Whilst the operational EFAS system is a multi-model system with four sets of meteorological forcings, in this study we focus only on the medium-range river discharge forecasts generated with meteorological forcings from the 51-member medium-range ensemble from the European Center for Medium-range Weather Forecasts (ECMWF) due to its large ensemble size. The meteorological forcings are interpolated to the EFAS grid. A single set of initial hydrological conditions are used for all ensemble members often leading to small ensemble spreads at short lead-times. The spread then increases as the different meteorological forcings propagate through the system. No data assimilation is performed in the generation of the initial hydrological conditions. Instead, the LISFLOOD hydrological model is forced with meteorological observations (and meteorological forecasts when observations are not available) to generate the initial conditions (Smith et al., 2016).

As an operational system, EFAS is constantly evolving. For the evaluation presented here we use EFAS version 4 (operational from 14 October 2020 to 20 September 2023) aggregated to daily timesteps with a maximum lead-time of 15 days. The ensembles have 51 members and predict the average river discharge for each timestep for each grid-box within the domain (see 6.2). The hindcasts have a spatial resolution of $5\text{km} \times 5\text{km}$ with a ETRS89 Lambert Azimuthal Equal Area Coordinate Reference System. Hindcast from the 00 UTC daily cycle are used resulting in a total of 365 hindcasts used in the evaluation.

6.2 Rhine-Meuse catchment

The Rhine-Meuse catchment has a drainage area $195,300\text{ km}^2$, a channel length of about $38,370\text{ km}$ in EFAS, and consists of 7812 grid-boxes. It is the 5th largest catchment in EFAS. The Rhine river originates in the Swiss Alps, flows through the Central Uplands and the North European Plain, before finally discharging into the North Sea. The Meuse river originates from the Langres Plateau in France, flows through the Ardennes Massif and the low-lying plains of the Netherlands, before merging with the Rhine and entering the North Sea. The catchment consists of rivers of different sizes, topologies, and levels of human influence, making it an ideal test catchment to see how the method deals with changes along the river network.

6.3 Observations

The Rhine-Meuse catchment has a dense river gauging station network. The main set of observations used in this study are daily river discharge observations from 89 stations across the Rhine-Meuse catchment for the time period from 21 December 2020 to 15 January 2022. The minimum value across the stations is $0.516\text{ m}^3\text{s}^{-1}$ and the maximum value is $7662.917\text{ m}^3\text{s}^{-1}$. These observations were assimilated as part of the error-correction method to update error ensemble and used in the evaluation of the corrected forecasts (Section 6.4 describes the cross-validation approach used). Whilst the error-correction method can adapt to missing observations, these 89 stations were selected as they have no missing data for the time period of interest allowing this analysis to focus on the spread of observational information to ungauged locations. The maximum distance between any grid-box and the closest of the 89 stations is 262 km which is set as our localisation length scale (cut-off distance

is therefore 524 km; see Section 5.1). In addition to these stations, all available observations from across Europe, were used to generate the initial error ensembles (total 505 stations). All river discharge observations were provided by local and national authorities and collated by the CEMS Hydrological Data Collection Centre (see <https://confluence.ecmwf.int/display/CEMS/EFAS+contributors>).

The construction of the non-localised observation error covariance matrix, $\mathbf{R}_k^{nl}$, is a key component of all data assimilation methods. The matrix describes the uncertainty associated with each observation and (defined in Eq. (7)). This uncertainty arises due to instrument uncertainty, observation processing, observation operator error and scale mismatch between the observations and the model resolution (Janjić et al., 2018). The matrix also describes the correlation between errors of different observations (Stewart et al., 2013; Fowler et al., 2018)(Stewart et al., 2013; Fowler et al., 2018). In this study, we assume that the observation errors from different gauge stations are uncorrelated such that $\mathbf{R}_k^{nl}$ is a diagonal matrix with all off-diagonal elements set to 0. We also assume that the standard Observation errors are also assumed to be uncorrelated with the prior errors which is a standard assumption in data assimilation (Janjić et al., 2018). We estimate the standard deviation of the observation errors is as 10% of the observation magnitude (Refsgaard et al., 2006; McMillan et al., 2018, 2012) (Refsgaard et al., 2006; McMillan et al., 2018, 2012).

In the leave-one-out verification experiments (see Section 6.4) we use the observations from the non-assimilated station as validation data and assume they are the truth with no errors.

6.4 Experiments

We use three experimental schemes to investigate the effect that the error-correction scheme has on the ensemble hindcasts.

1. **Single station experiments:** Only observations from one of the 89 station are assimilated when estimating the error-vector. All available observations are used in the generation of the initial error ensemble. These experiments allow the impact of an observation to be identified and allow the effects of localisation to be explored.
2. **All station experiments:** Observations from all stations are assimilated when estimating the error-vector and used in the generation of the initial error ensemble. These experiments allow the complete method to be assessed and for any spatiotemporal inconsistencies to be identified.
3. **Leave-one-out experiments:** Observations are withheld from one of the 89 stations and are not assimilated when estimating the error-vector nor used in the generation of the initial error ensemble. This cross-validation framework allows the skill of the adjusted hindcasts to be assessed at the locations of stations as if they were ungauged locations.

Each experiment scheme is applied to all hindcasts from 1 January 2021 to 31 December 2021. However, for brevity, for the single station and all station experiments we only discuss two hindcasts: 7 July 2021 and 8 October 2021. These dates represent high and normal flow conditions, respectively, allowing the ability of the method to be assessed for different circumstances.

6.5 Evaluation metrics

The following metrics are used to investigate the skill of the error-corrected hindcast ensemble mean and the reliability of the ensemble spread.

For the ensemble mean, the three components of the modified Kling-Gupta Efficiency: correlation, mean bias, and variability bias are used to assess different types of errors within the ensemble mean (Kling et al. (2012); Gupta et al. (2009)) (Kling et al., 2012; Gupta et al., 2009). Pearson's correlation coefficient measures the linear relationship between the simulated timeseries and the observations indicating timing errors (score range $[-1, 1]$). The mean bias given by the ratio between the mean of the simulated timeseries and mean of the observations indicates whether the flow is consistently over or underestimated (score range $(-\infty, +\infty)$). The variability bias given by the ratio between the coefficient of variation of the simulation and the coefficient of variation of the observations indicates whether the variability in the flow is consistently over or underestimated (score range $(-\infty, +\infty)$). All three components have a perfect score of 1. Additionally, to investigate whether the magnitude of the error of the forecast mean is decreased by the proposed method we use the Normalised Mean Absolute Error (NMAE Hodson, 2022; Jackson et al., 2019). The metric is normalised by dividing by the mean of the observations for that station. Normalising the metric makes the scores at different stations comparable. The NMAE has a perfect score of 0.

To analyse the reliability of the spread of the ensemble forecast we use the rank histogram (Harrison et al., 1995; Anderson, 1996; Hamill and Colucci, 1997; Talagrand, 1999). To generate the histogram the rank of the observation relative to the sorted ensemble values is calculated for each hindcast. The frequencies with which the observation has a rank from 1 to $M + 1$ are plotted as a histogram. The shape of the histogram provides information about the reliability of the ensemble spread and bias of the ensemble (Hamill, 2001).

7 Results

7.1 Impact of assimilating observations

In this section ~~we investigate the spatial and lead-time dependent impact of assimilating the observations. To assess the spatial impact of these observations, we analyze the~~, we discuss the efficacy of the proposed error-correction method. In Section 7.1, we discuss how observation information is propagated along the river network and, in particular, we explore how the method reacts to different flow scenarios, both spatially and across different lead-times. In Section 7.2, we evaluate the skill of the resulting error-corrected ensembles in terms of their means and distributions.

7.1 How is observation information propagated along the river network?

7.1.1 Spatial propagation of the observation information

Here, we investigate how the observation information is propagated spatially from gauged locations to ungauged locations. We investigate the analysis increments of the mean — the difference between the ensemble mean before and after the update step

(term 2 in Eq. (14)) — ~~across the domain at a single lead-time for single station experiments and all station~~ ~~for single station~~ ~~and all station~~ experiments (Fig. 2). ~~We focus on single station~~ Specifically, we focus on the single station experiments for the
 550 Bonn station on the Rhine (~~left panels~~) and the Uckange station on the Moselle (~~middle panels~~) ~~for~~. To investigate the impact
 of different flow scenarios, we study the hindcasts generated on 8 October 2021 (upper panels) and 7 July 2021 (lower panels),
 which represent normal and high flow scenarios, respectively.

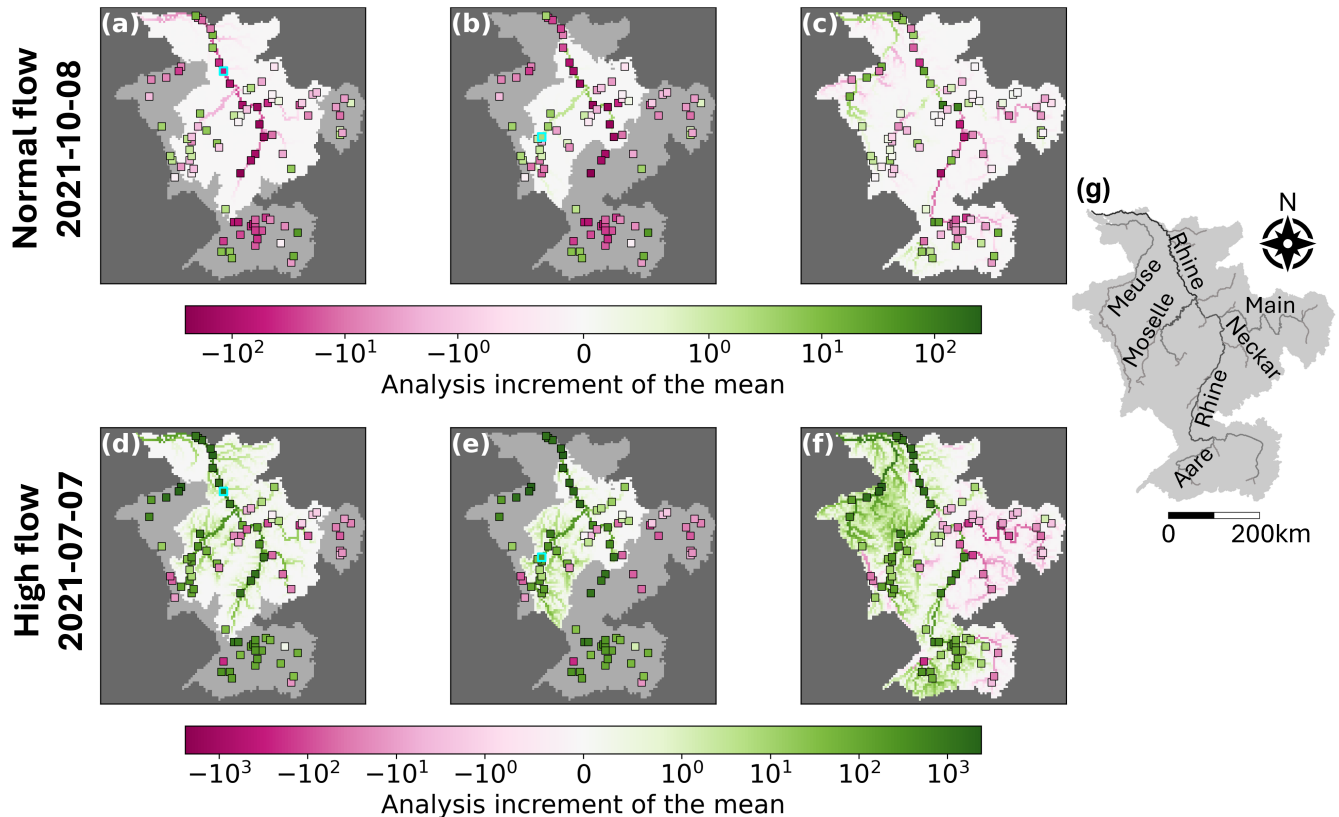


Figure 2. Analysis increments of the mean for a lead-time of 9 days for single station (a, b, d, and e) and all station (c and f) experiments for the hindcasts generated on 8 October 2021 (upper panels) and 7 July 2021 (lower panels). Assimilated stations for the single station experiments (cyan outline) are the Bonn station on the Rhine (a and d) and the Uckange station on the Moselle (b and e). The shaded region of the catchment is outside the localisation length of the assimilated station. Markers show the innovation at all stations. Catchment area: $195,300 \text{ km}^2$. Panel g shows the Rhine-Meuse catchment and highlights the rivers discussed within this section.

For 8 October 2021, the assimilation of an observation at Bonn results in the largest analysis increments near the observation location, with the increments diminishing to zero at distances greater than 524 km due to localization (background color of
 550 Fig. 2a). In Fig. 2, the shaded regions show the parts of the catchment that are outside the localization region for the assimilated observation. The number of grid-boxes within the localization regions of the Bonn and Uckange stations differ because the distance is calculated along the river network and the channel length within each grid box is not constant (4662 grid boxes and

2451 grid boxes, respectively). Increasing (decreasing) the localisation length scale results in a more (less) gradual dampening of the analysis increments and more (fewer) grid-boxes being impacted by a single observation (not shown). The number of grid-boxes in the localization regions of the Bonn and Uckange stations differ (4662 grid boxes and 2451 grid boxes, respectively) because the distance is calculated along the river network and the channel length within each grid box is not constant. Square markers indicate the innovation — the difference between the observation and the error-corrected ensemble mean prior to the update step (Fig. 2). Ideally, the analysis increment (background colour in Fig. 2) should reflect similar spatial behavior to the innovations within the localisation region. This would imply the ensemble is being adjusted towards the observations at each station.

For the October experiment, the innovation at Bonn is negative and results in negative analysis increments across the domain (Fig. 2a). For the Uckange experiment, the innovation is positive and the analysis increments are also all positive (Fig. 2b) indicating positive ensemble covariances. For both experiments, the analysis increments match the sign of the innovations at neighbouring stations (Figs. 2a and 2b), but at greater distances this is not the case. For example, the innovations along the Rhine in the Uckange experiment are negative whilst the analysis increments are positive.

The localisation implemented in this study allows the assimilated observations to influence the error ensemble both up- and downstream, although the influence is dampened at longer distances. We here discuss whether this choice of implementation is useful by studying the spatial patterns of the innovations. Focusing first on the Bonn experiment for October (Fig. 2a), we see that downstream (north) of the assimilated observations the innovations can be both positive and negative. Upstream of the assimilated observation the innovations are negative, matching the innovation at the Bonn station. The assimilation of the observation at Bonn is therefore primarily beneficial upstream, with some benefit also seen at specific locations downstream. For the Uckange experiment (Fig. 2b), the pattern is reversed with downstream innovations showing more consistency with the innovation at the assimilated location. The inconsistent spatial patterns could be because, in the LETKF, we update the errors rather than the river discharge directly. The errors are dependent not only on the observed hydrological conditions but also the model structure and configuration. The spatial structure of the errors may therefore extend both up- and downstream. For example, if the drainage area within an upstream grid-box is overestimated due to the hydrological model spatial resolution, all grid-boxes downstream will be impacted by that overestimation. The benefit, in terms of consistency between the innovations and analysis increments, that is seen both up and downstream suggests that the localisation implementation is appropriate. However, we note that there may be additional factors other than distance, that could be included in the localisation to better modulate the observation influence (e.g., river confluences, regulation, or river size).

In the July experiments, we see that the innovations both up and downstream of the assimilated observations are positive, matching the innovations at the Bonn and Uckange stations, respectively. For the July experiment, the innovations are spatially homogeneous for greater distances along the river network (Figs. 2a and 2b).d and 2e). This indicates a greater spatial correlation length, likely due to the low-pressure system which covered large parts of the west of the catchment during this period (Mohr et al., 2023). The different correlation scales suggest that an adaptive localisation length scale may be beneficial.

The spatial heterogeneity for the October experiments suggests that assimilating a single observation cannot correct the entire domain. However, when all observation are assimilated the analysis increments vary across the domain, demonstrating

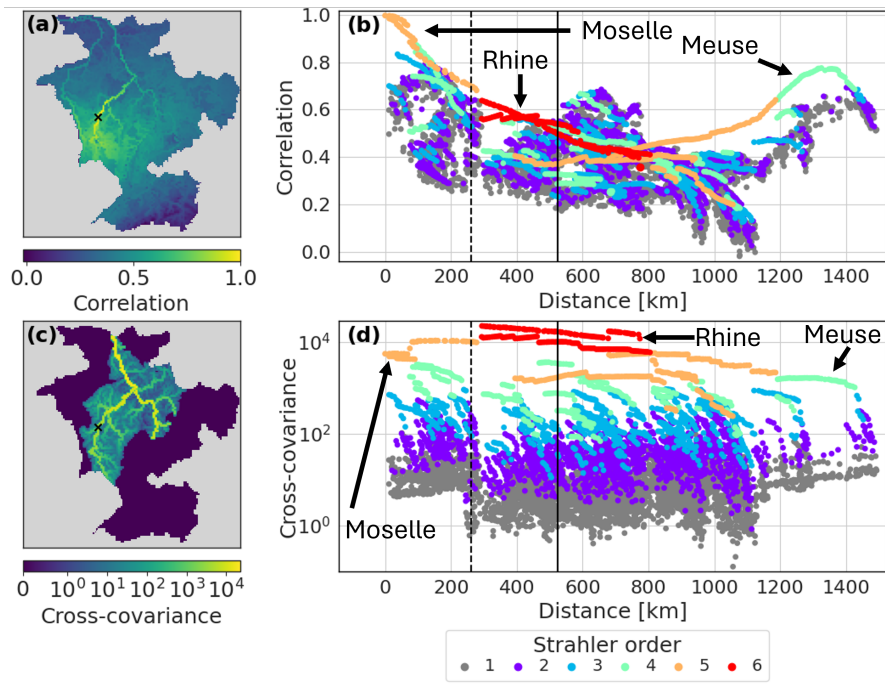


Figure 3. Ensemble correlations (upper panels) and cross-covariances (lower panels) between the error ensemble and the hindcast component of the augmented state averaged across all all-station experiments. (a) Map of the correlation between the Uckange station and all other grid-boxes and (c) the same for the cross-covariances. (b) Scatter plot of the correlation between the Uckange station and all other grid-boxes and (d) the same for the cross-covariances. Grid-boxes on rivers discussed in the text are broadly indicated by the arrows. Dashed black line shows the localisation length scale and the solid black line shows the effective cut-off point beyond which the observation has no impact.

the method’s ability to adapt to the errors on different stretches of the river. In both all-station experiments (Figs. 2c and 2f), the analysis increments vary smoothly along the river network, which suggests the error-corrected ensemble will also change gradually. This is important because it ensures the hindcasts remain spatially consistent, with no abrupt transitions between adjacent grid boxes.

In general, for the July experiment, small rivers exhibit larger increments than in the October experiment. This indicates the assimilated observations have a greater impact across more of the domain. For October, the assimilation of an observation at Bonn results in the largest analysis increments near the observation location, with the increments diminishing to zero at distances greater than 524 km due to localization (Fig. 2a). Interestingly, in the Uckange experiment, the largest increments occur not near the station, but along the Rhine near the confluence with the Moselle (Fig. 2b). In both experiments, the increments tend to be larger along bigger rivers, with smaller rivers showing smaller increments. This occurs due to large ensemble covariances between the location of the assimilated observation and locations along the bigger rivers (Fig. 3).

The spreading of observational information along the river network is dictated by the cross-covariances ~~between the error component and the hindcast component~~ of the augmented ensemble prior to the update step (Section 3.2). The magnitude of

the cross-covariance between two locations depends on the correlation ~~at the two locations~~, the variance of the augmented ensemble ~~between the locations and the ensemble variance~~ at both locations, ~~and the dampening enforced by localisation~~. The correlation between the location of the Uckange station and ~~a any~~ grid-box is highest along the same river stretch (the Moselle) and decreases at longer distances from the station (Fig. 3a). Nearby grid boxes that are not on the same river stretch have lower correlations in general. ~~Comparing (Figs. 3a and 3d indicates that along the river localisation is appropriate for this application, as with other hydrological data assimilation systems (e.g., El Gharamti et al., 2021), as it results in the impact of an observation being restricted to locations with higher, physically plausible, correlations.~~

Ensemble correlations (upper panels) and cross-covariances (lower panels) between the error component and the hindcast component of the augmented state prior to the update step for all station experiments. Map of the correlation (a) and localised cross-covariance (d) averaged across all lead-times and forecasts between the Uckange station (shown by black cross) and all grid-boxes. Scatter plot of the correlation (b), localised cross-covariance (e), and non-localised cross-covariance (f) averaged across all lead-times and forecasts between the Uckange station and all grid-boxes against distance from the Uckange station. Grid-boxes on rivers discussed in the text are broadly indicated by the arrows. Dashed black line shows the localisation length scale and the solid black line shows the effective cut-off point beyond which the observation has no impact (twice the localisation length scale; see Section 5.1). Correlation (c) and localised cross-covariance (g) between the Uckange station on the Moselle and the Bonn station on the Rhine for all forecast and for each lead-time of the hindcast (365 values per lead-time, one for each forecast).

The magnitude of the cross-covariances are similar ~~b~~. Downstream from the Uckange station the correlation is highest along the Moselle and ~~for some parts of the Rhine despite lower correlations and the application of localisation~~ the downstream along the Rhine. On the other hand, the correlation upstream is more uniform across the grid-boxes (Fig. 3b ~~and 3e~~). Whilst the correlation initially decreases with increasing distance, the magnitude of the non-localised cross-covariances is primarily dependent on the size of the river (note the horizontal bands of Strahler orders (a measure of stream size where larger orders indicate larger rivers) in Fig. 3f). Localisation enforces a dependence on distance such that smaller rivers near the station have a larger localised cross-covariance than large rivers very far from the station there are regions in the south of the catchment with which the correlation is small, in general there is a correlation of around 0.3 even with distance locations (Fig. 3e). ~~However, some grid-boxes on the Rhine (Strahler order of 6) still have larger cross-covariances than smaller rivers that are closer b~~. This is likely spurious correlation and exemplifies the need for localisation. The correlations begins to rise again at longer distances due to grid-boxes that are geographically close to the station but the distance along the river network is large, such as the Meuse (Fig. 3e). ~~The impact of this can be seen in Fig. 2b where the analysis increments along the Rhine are larger than those along parts of the Moselle. b~~. Note that the similarity between the localisation length scale (dashed line) and the distance between the Uckange station and grid-boxes on the Rhine (change from a Strahler order of 5 to 6) is coincidental but does suggest that the method for defining the localisation length scale (see Section 5.1) is capable of capturing the order of magnitude of the relevant spatial scales for the Rhine catchment.

In Fig. 2, the square markers indicate the innovation — the difference between the observation and the error-corrected ensemble mean prior to the update step. Ideally, the analysis increment (background colour in Fig. 2) should reflect similar

behavior to the innovations within the localisation region, implying that the ensemble is being adjusted towards the observations at each station. For 8 October, at Bonn the innovation is negative and results in negative analysis increments across the domain. Despite lower correlations, the magnitude of the cross-covariances are larger along the Rhine than for grid-boxes closer to the Uckange station on the Moselle (Fig. 2a). For the Uckange station the innovation is positive and the analysis increments are also all positive (Fig. 2b) indicating positive ensemble covariances (Fig. 3d). For both of the 8 October experiments the analysis increments match the sign of the innovation vectors for neighbouring stations (Figs. 2a and 2b). At greater distances the analysis increments do not follow the same behaviour as the innovations. For example, the innovations along the Rhine in the Uckange experiment are negative whilst the analysis increments are positive (Fig. 2b) suggesting spurious ensemble covariances between the Uckange station and locations on the Rhine. In the experiments for 7 July, the analysis increments show a similar behaviour to the innovations for a much greater distance along the river network (Figs. 2d and 2e). In contrast to the 8 October experiments (Fig. 2e), the innovations for the 7 July experiments are more spatially homogeneous indicating a greater spatial correlation length. This is likely due to the low pressure system which covered large parts of the West of the catchment during the hindcast period of the 7 July experiments (Mohr et al., 2023). In Figure 3b, the average correlation (Fig. 3c). Whilst the correlation is dependent on distance, the magnitude of the cross-covariances is primarily dependent on the size of the river (note the horizontal bands of Strahler orders (a measure of stream size where larger orders indicate larger rivers Strahler, 1957) in Fig. 3d). Larger cross-covariances can lead to larger analysis increments as can be seen to begin to increase again for larger distances. This is due to grid boxes on different rivers (here, the Meuse and the Moselle; see rivers names in Fig. 2) being close geographically but far apart along the river network. Geographically close locations may be impacted by the same weather systems even if the water drains into different rivers as happened for the 7 July period.

Whilst the spatial heterogeneity for 8 October suggests that the assimilation of an observation from a single station cannot correct the entire domain (Figs. 2a and 2b), when all observations are assimilated where the analysis increments vary across the domain (Fig. 2e). This heterogeneity of the analysis shows the ability of the method to vary the correction across the domain adapting to the errors in different stretches of the river. The analysis increments are smoothly changing along a river stretch therefore, the changes to the error-corrected ensemble will also be smoothly changing spatially. Along the Rhine are larger than those along parts of the Moselle. Localisation enforces a dependence on distance such that observations have less impact on large rivers very far from the station but this may not outweigh the larger cross-covariances.

Another difference between the 8 October and the 7 July experiments is that for the 7 July hindcast small rivers exhibit larger increments, indicating a greater impact from the assimilated observation (Figs. 2e and 2f). Two factors contribute to this increased influence. First, the increased spatial correlation length means the observation is more informative for longer distances. However, the correlation between, for example, the locations of the Uckange and Bonn stations is higher at a lead-time of 9 days for the 8 October than for the 7 July (Fig. 3e). Therefore, the second factor, larger ensemble variances in the 7 July period compared to the 8 October period, is likely the more dominant component (Figs. 4b, 4c, 4e, and 4f). The increase in spread increases the cross-covariances (Fig. 3g) and allows the observation to have more influence.

Ensemble trajectories for a single station experiment for the hindcast generated on 7 July 2021 for the assimilated station (Uckange station on the Moselle; panels a-e) and a non-assimilated station (Bonn station on the Rhine; panels d-f). The plots

show the trajectory of all members and the ensemble mean of the raw hindcast ensemble (left panels), the hindcast component of the augmented state (middle panels), and the error ensemble members (right panels; different y-axis scale). Markers show the river discharge observations (a, b, d, and e), and the error of the raw hindcast mean (c and f).

7.1.2 Lead-time dependence of the analysis increments

Here, we investigate how the impact of assimilating observations changes over different lead-times. Figure 4 shows the trajectories of the three intermediate ensembles used in the LETKF for the 7 July hindcast for a single-station-single-station experiment where observations are assimilated at the Uckange station: the raw hindcast (left columns), the hindcast component of the augmented ensemble (middle columns), and the error component of the augmented ensemble (right columns). The evolution of the augmented ensemble are discussed in Sections 3.1-5. It should be noted that none of these ensembles are the final error-corrected ensemble but intermediate ensembles used in the LETKF. The lower panels show the trajectories at the Bonn stations for which no observations are assimilated during this experiment. By plotting the raw hindcast trajectories and the observations we can visualise the errors to be estimated. We can see that for both stations the error of the hindcast mean is negative (observations are smaller than the hindcast mean) for lead-times up to 8 days, and positive at longer lead-times. Whilst this behaviour is similar for the Bonn station, the magnitude of the error is different by a factor of 10 at most lead-times.

The middle column of Fig. 4 shows the hindcast component of the augmented ensemble. We can see that using To propagate this component between timesteps without rerunning a hydrological model, we assume that the raw hindcast as an is a reasonable approximation of the analysis state is not optimal (discussed in Section 4.1). As expected, this assumption results in a sub-optimal ensemble mean estimate. For example, at lead-times greater than lead times beyond 10 days at the Uckange station the update takes the ensemble further away. Uckange, the update moves the ensemble mean further from the observations (Fig. 4b). This occurs also at the Bonn station, and a similar effect is seen at Bonn (Fig. 4e). This is not unexpected as our approximation assumes. Also by using the precomputed ensemble, the assimilated observations do not update the ensemble perturbations; although, the perturbations do change between lead-times as the precomputed ensemble is lead-time dependent. This assumption does ensure the analysis hindcast component is always physically plausible (e.g., the river discharge is always positive), and provides a reasonable estimate of the uncertainty as the raw hindcast is more accurate than the hindcast corrected with the error ensemble members from the previous time step (Section 4.1). However, this assumption is necessary to propagate the hindcast to the next time step without the use of a hydrological model (Section 3.1) generated using the output from an ensemble NWP. Additionally, at each timestep we aim to correct the raw hindcast, therefore this assumption provides consistency between the hindcast component and the error component of the augmented state.

It is the error ensemble that is most important to for our application (Figs. 4c and 4f). Despite the non-optimal formation of the analysis augmented state, the error ensembles are updated beneficially, with the analysis error ensemble mean moving closer to the error of the raw hindcast mean at each lead-time for the assimilated location (Fig. 4c) and the non-assimilated location (Fig. 4f). At short lead-times the updates to the error ensemble at the Bonn station do not appear to be beneficial (Fig. 4f). However, as this experiment only assimilates observations from one station this discussion should be considered a demonstration of how the method updates proxy-ungauged locations rather than an evaluation of the error-corrected ensemble

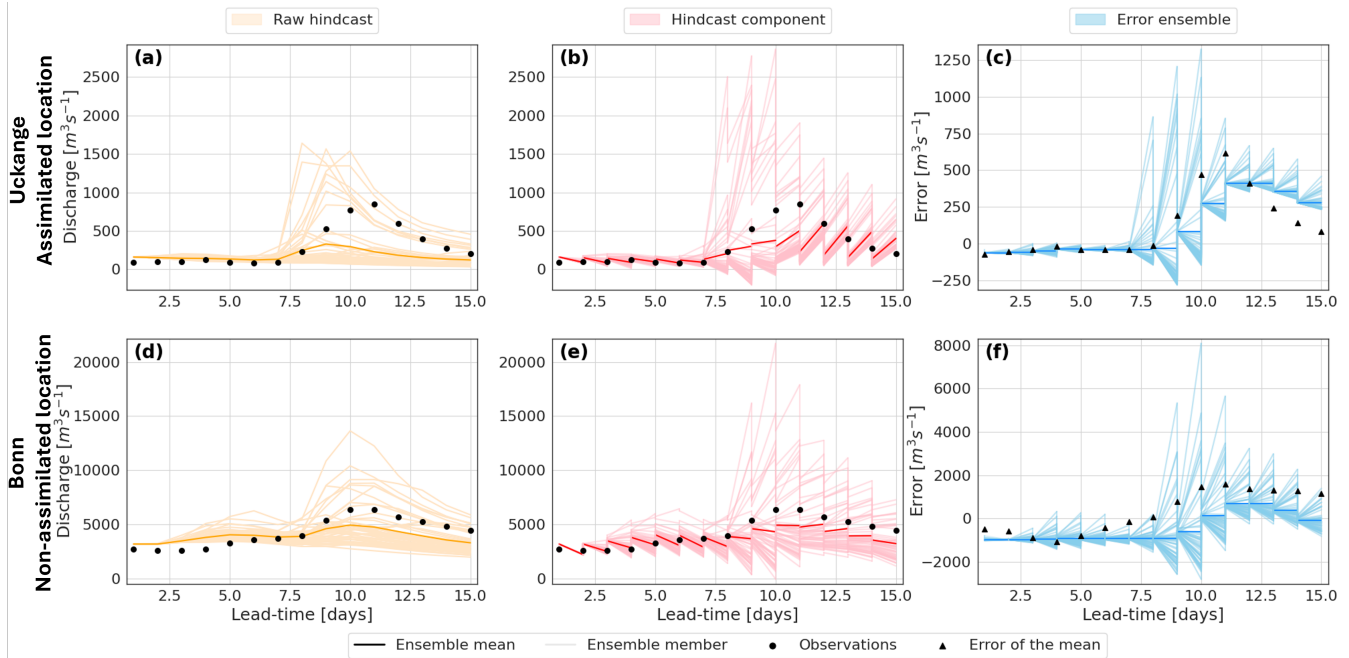


Figure 4. Ensemble trajectories for a single station experiment for the hindcast generated on 7 July 2021 for the location of the assimilated observation (Uckange station on the Moselle; panels a-c) and a location where an observations is not assimilated (Bonn station on the Rhine; panels d-f). The plots show the trajectory of all members and the ensemble mean of the raw hindcast ensemble (left panels), the hindcast component of the augmented state (middle panels), and the error ensemble members (right panels; different y-axis scale). Markers show the river discharge observations (a, b, d, and e), and the error of the raw hindcast mean (c and f).

710 (which is provided in Section 7.2). First we note, that the updates at the assimilated location do not result in the error ensemble mean (dark blue line) matching the error of the mean (markers). This is expected and is due to the consideration of the observational uncertainty within the LETKF. This ensures spatial consistency across assimilated and non-assimilated locations, and combines whilst combining the modelled and observed data to estimate the true state of the system across the domain.

The error-ensemble is narrow after the update step and it is the covariance inflation that increases the spread between 715 timesteps. The spread of the hindcast is due to meteorological forcings, predominantly precipitation. Therefore, in general, the hindcast spread is larger for longer lead-times as the precipitation forecasts become more uncertain and this uncertainty is propagated along the river network, and higher river discharge values (when precipitation is above 0 mm). Since the covariance inflation technique presented here results in the blending of the hindcast perturbation matrix with the error-ensemble from the previous timestep, this behaviour in the hindcast spread is transferred to the error-ensemble. As demonstrated in Figs. 4c and 720 4f, this can result in the error ensemble spread being large for the rising limb of an event and smaller for the falling limb. This can result in the error not being updated sufficiently and the spread of the analysis state being too narrow, as seen after the peak in Fig. 4c and discussed later along with Fig. 5b.

7.2 Ensemble skill ~~How skillful are the error-corrected ensemble hindcasts?~~

In this section ~~we look at~~, we investigate whether the updated ensemble is more skillful than the raw hindcast ensemble. Using
725 leave-one-out experiments we evaluate the ensemble mean and ensemble spread at proxy-ungauged locations (~~Section 6.4~~). The hydrographs in Fig. 5 show the raw and error-corrected ensembles for three proxy-ungauged locations from the leave-one-out experiments. The hydrographs are used to illustrate the method's ability to correct the ensemble and some of the limitations.

7.2.1 Skill of the ensemble mean

To investigate the ~~impact on different types of errors in skill~~ of the ensemble mean we calculate the correlation, mean bias,
730 variability bias and the ~~NMAE-N-RMSE~~ for each station and each lead-time (~~Section ??~~). Figure 6 compares the skill of the ensemble ~~mean means~~ of the raw and the error-corrected ensembles focusing ~~on the overall change in skill (a, d, g, and i), the~~
~~the~~ spatial dependency of the skill (~~b, c, h, and k a-d~~), and the lead-time dependency of the skill (~~e, f, i, and l e-h~~).

The error-corrected ensemble means show a stronger correlation with observations than the raw hindcast ensemble means, with an average increase from 0.82 to 0.92 ~~; and an overall shift towards the perfect value of 1 (Fig. 6a(not shown))~~. Figure
735 5a shows an example of how the error-corrected ensemble can better capture the dynamics of the ~~observations improving the~~
~~river discharge resulting in an increased~~ correlation. It can be seen that the resulting ensemble is temporally consistent ~~and~~
(i.e., does not have improbable changes between timesteps. ~~However, at four stations the correlation worsens~~). ~~The correlation~~
~~is worsened~~ compared to the raw hindcast ensemble ~~at four stations~~ (Fig. 6**b**). ~~The a~~. Focusing ~~on the~~ two most southern of
these ~~worsened stations, are near to confluences with larger rivers which have different correlation patterns in the raw hindcast~~
740 ~~to those of the~~ stations, we see that the correlation values of the raw ensembles at nearby stations are very different compared
~~to the correlation at the~~ two stations of interest (note the much lighter colours for nearby stations; Fig. 6a). The ensemble
covariances are not capturing this change in regime correctly so the observational information is not being advantageously
spread between these rivers. The remaining two stations ~~that have degraded correlation~~ are the most upstream stations on
their rivers. At these ~~; stations~~ the updates made to the error-corrected ensemble are dependent on observations assimilated
745 downstream. The assimilated observations are therefore providing information about a past state of the river upstream which
could be the cause of the decreased correlation (a measure of timing errors) at these upstream stations. Whilst most upstream
stations are improved by the error-correction method, stations which have much smaller upstream areas ~~that than~~ their closest
downstream station tend to be improved less than those that have a similar upstream area, particularly if the distance to the
neighboring station is large.

750 ~~The error-corrected ensemble generally has a lower mean bias than the raw hindcast ensemble, with the average mean bias~~
~~shifting from 1.027 (overestimation) to 1.004 (less overestimation). However, there is a slight shift towards underestimation~~
~~(Fig. 6d)~~. Just over half of the stations (47) show improvement in ~~the~~ mean bias averaged across all lead times (Fig. 6**b**), but
no clear spatial pattern emerges, as most rivers have a mix of improved and worsened stations. This spatial heterogeneity is
also seen in the raw hindcast ensemble, with stations on the same river stretch often showing different biases. For example,
755 stations on the Neckar ~~; and~~ upstream of the Meuse show stations ~~; that~~ are under- and overestimated, as well as stations with

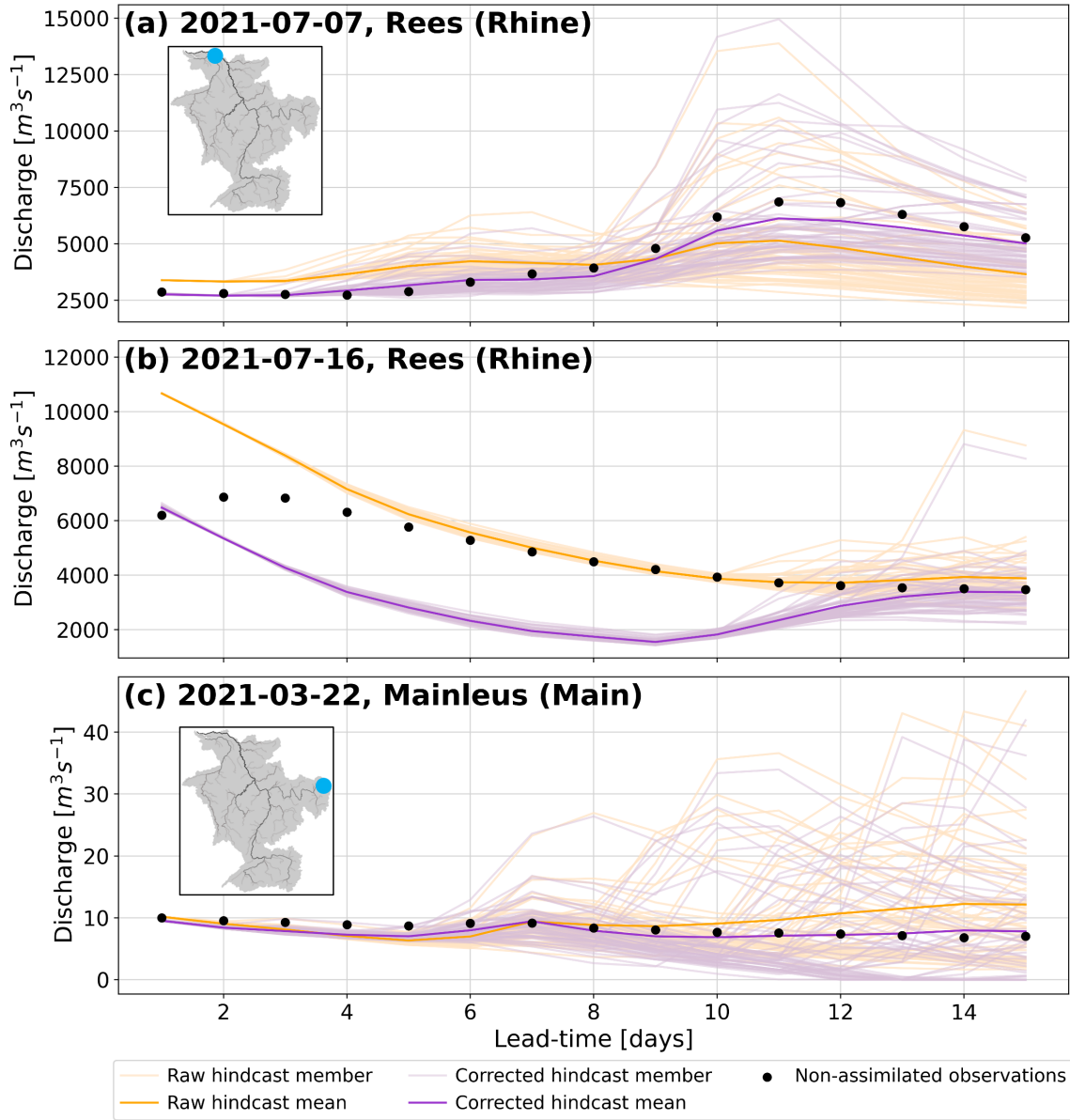


Figure 5. Raw and error-corrected hydrographs for proxy-ungauged locations in leave-one-out experiments at the Rees station on the Rhine (upstream area: $159,320 km^3$) and the Mainleus station on the Main (upstream area: $1,164 km^3$). Catchment illustrations indicate the location of the station (see Fig. 2 for **river names** river names and scale).

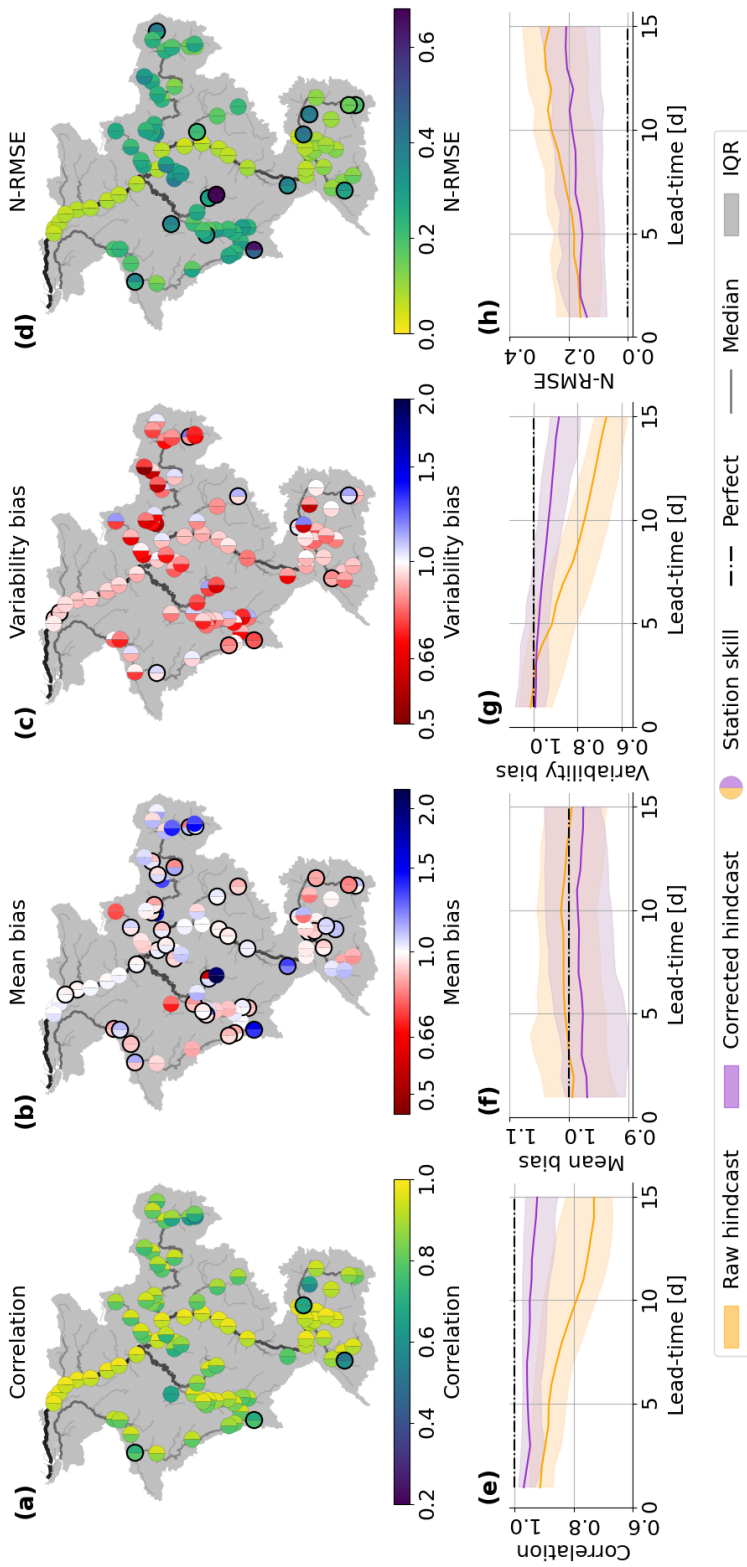


Figure 6. Skill of the ensemble means in terms of the correlation (a, b, ee), mean bias (db, e, f), variability bias (gc, h, i, g) and normalised root mean absolute-square error (NMAEN-RMSE; j, k, l). The histograms show the distributions of the metrics pooled over all stations and lead-times for the raw (orange) and error-corrected (purple) ensembles (a, d, g, jh). A perfect score for the metric is shown by the dashed black line. Catchment The catchment maps show the metric averaged across all lead-times at all 90-89 stations (b, e, h, ka-d). The left (right) half of the marker shows markers show the skill for the raw (error-corrected) ensemble. Black outlines indicate stations for which the updated ensemble has worse skill than the raw hindcast ensemble. Line plots show the distribution of the metric pooled over all 90-89 stations for each lead-time for the raw (orange) and error-corrected (purple) ensembles (e, f, i, le-h). The solid line shows the median value of the metric and the shaded region shows the interquartile range (IQR) of the metric. A perfect score for the metric is shown by the dashed black line.

very little bias. The heterogeneity suggests local factors, which are not fully captured in the ~~modelling-system-hydrological~~ ~~model~~, considerably influence flow bias. Stations showing the most improvement tend to have similar mean bias values to their neighboring stations in the raw hindcast ensemble, such as on the middle stretch of the Meuse, where four stations with similar biases show improvement (Fig. 6eb). Spatial patterns of errors that are related to domain-wide model structure rather than local factors, such as weirs, are more likely to be portrayed by the ensemble covariances allowing observational information to be more helpfully spread along the river network.

The raw hindcast ensemble mean generally underestimates flow variability, with a variability bias below 1 (red in Fig. 6gc). The error-corrected ensemble improves this, ~~increasing the mean variability bias from 0.82 to 0.95, although although~~ ~~there is an increase in~~ the frequency of overestimation of ~~flow variability is also increased (Fig. 6g)-~~ the flow variability. Stations where the error-corrected ensemble overestimates the variability are often the most upstream station on their rivers (e.g., Plochigen station on the Neckar) or are ~~much~~ closer to downstream ~~than upstream~~ neighbours (e.g., Chooz station on the Meuse). This suggests the hindcast covariances between downstream stations and upstream locations are too large, causing excessive adjustment at upstream locations. Ten stations show worsened variability bias, including two stations downstream on the Rhine (Fig. 6h). ~~The cause of the worsening of these two stations is~~ For the two stations on the Rhine, the degradation ~~is caused by the forecasts of the~~ ~~adjustment for the~~ falling limb of a flood peak in July (Fig. 5b). Here, the hindcast uncertainty was very small at short lead-times, causing the analysis to ignore observations and the error ensemble to remain relatively unchanged, despite changes in the error ~~behavior following the peak-~~ ~~behaviour after the peak~~ (also shown in Fig. 4f).

Overall, the error-corrected ensemble reduces the ~~absolute error, with the average NMAE decreasing from 0.33 to 0.23~~ ~~(Fig. 6j)-~~ ~~The 7 stations with worsened NMAE are typically (5/7) the most upstream on their N-RMSE~~ but there are 14 stations where the skill is reduced. Typically, these stations are on the upstream reaches of their respective rivers (Fig. 6kd; see discussion ~~about on~~ correlation). Interestingly, the ~~NMAE-N-RMSE~~ does not follow the same spatial pattern as the mean bias. The decrease in ~~absolute errors~~ ~~N-RMSE~~, despite an increase in mean bias, suggests that the error-corrected ensembles consistently underestimate flow, while the raw hindcast ensemble fluctuates more between under- and overestimation, which can compensate for each other in the mean bias metric.

The raw and error-corrected ensemble means both decrease in skill in terms of correlation, variability bias, and ~~NMAE-N-~~ ~~RMSE~~ with increasing lead-times. The raw hindcast ensemble loses skill more quickly in particular for lead-times longer than 5 days (Figs. 6e, 6i, and 6le-h). The uncertainty in the observations is not lead-time dependent. However, Fig. 3d shows that the ensemble covariances do change across lead-times, increasing for longer lead-times. The ~~reduction in skill as greater gain in~~ ~~skill for longer~~ lead-times ~~increase suggest that the ensemble covariances are not able to spread the observational information~~ ~~along the river network as accurately is~~ likely due to larger covariances allowing the observations to have more influence (e.g., in Fig. 5b). However, the decrease in skill of the error-corrected ensemble means at longer lead-times ~~-This is likely due to an~~ ~~over-estimation of the variance~~ suggests that the ensemble covariances are not as accurate at longer lead-times.

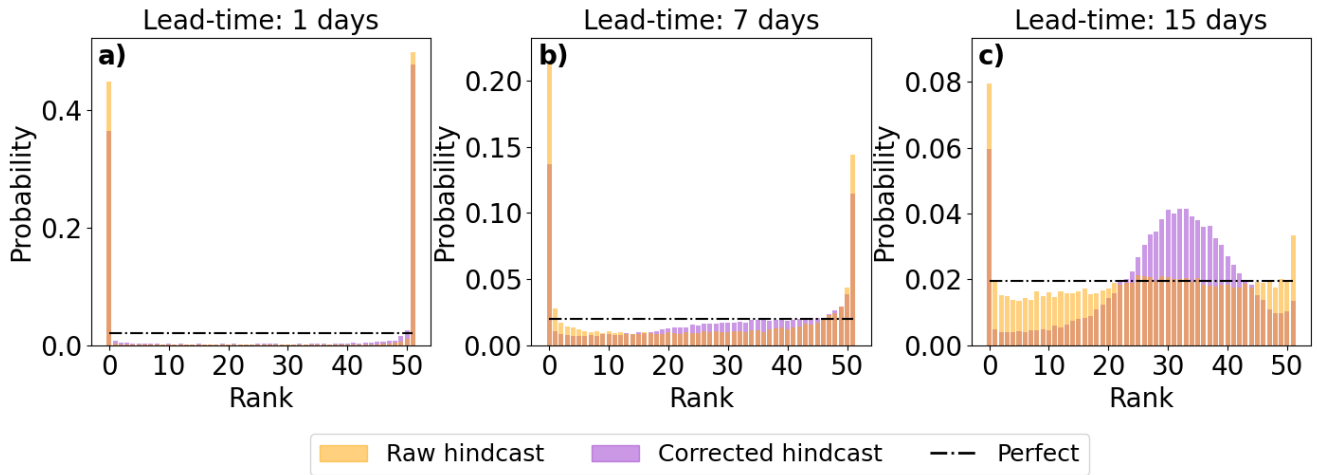


Figure 7. Reliability of the ensemble. Histograms show the rank of the ensemble pooled over all forecasts and stations for lead-times of 1 day (a), 7 days (b), and 15 days (c).

7.2.2 Skill of the ensemble distribution

The reliability of the ensemble distribution is assessed using rank histograms at different lead times (Figs. 7a, 7b, and 7c). At short lead times, the raw hindcast ensemble is underdispersed, likely due to the use of a single set of initial conditions (Fig. 7a). Although the error-corrected ensemble shows slight improvement, it remains overconfident with minimal correction to the spread. Both the raw and error-corrected ensembles generally appear unbiased, with observations falling both above and below the ensemble predictions at similar frequencies. However, some bias may be masked by the narrow ensemble spread and as it is known that some stations are biased (Fig. 6b), likely contributing to the peaks at ranks 0 and 51 in the rank histograms.

As the lead-time increases, the spread of both ensembles becomes more reliable, and fewer observations fall outside the ensemble (Fig. 7b). However, even at a 15-day lead time, both ensembles show a tendency to overestimate observations, leading to a peak at rank 0, mostly due to a few stations consistently overestimating flow (Fig. 6b). Up to 7-day lead times, the rank histograms for both raw and error-corrected ensembles show similar shapes. Beyond 7 days, the raw hindcast ensemble's histogram flattens, suggesting a reliable ensemble, while the error-corrected ensemble shows a peak around ranks 25-35, suggesting overdispersion (Fig. 7c). The left-skewness of the histograms is likely due to the inherent skewness in river discharge distributions. The LETKF update step seeks to minimise the difference between the ensemble mean and the true state of the system. The ensemble mean is often larger than the ensemble median leading to the observations falling in ranks above 25 if the adjustment method is successful and at minimising the error of the mean (Figs. 5a and 5c).

As discussed in Section 4.1, the Kalman filter is not restricted to ensure positive discharge and there is therefore a need to adjusted-adjust the error ensemble before correction of the hindcast. Enforcing non-negative discharge was necessary, for example, for the Mainleus station on the Main for the hindcast generated on the 22 March 2021 (Fig. 5c). Whilst the ensemble mean is error-corrected at most lead-times, several members indicate river discharge values of $0 \text{ m}^3 \text{ s}^{-1}$. The river discharge

is below $10 \text{ m}^3 \text{ s}^{-1}$ but a zero flow is unlikely in reality. This suggests the ensemble spread is not sufficiently corrected even though the ensemble mean is improved as is also suggested by Fig. 7c.

810 8 Discussion

In general, the proposed data-assimilation-inspired method successfully spreads observational information along the river network improving the skill of the ensemble mean at ungauged locations. Locations downstream from assimilated observations are improved most although locations upstream are usually improved as well, even if they are far from neighbouring stations. This is likely due to two reasons: 1) constant biases in the river discharge estimates that are propagated downstream and
815 hence can be accounted for when a downstream observation is assimilated, and 2) the daily aggregation of the river discharge extending the time period for which a downstream observations provides relevant information. If the error patterns of the ensemble mean at a location differ from those at nearby stations the method struggles to spread the observational information correctly. At shorter lead-times the reliability of the ensemble is slightly improved due to the decrease in the error of the ensemble mean. However, at longer lead-times the ensemble spread is often too large leading to an under-confident forecast.

820 ~~Despite the method's ability to correct upstream, it could be beneficial to assimilate observations from as far upstream as possible. These observations do not necessarily need to be traditional in-situ observations but could come from Earth Observation (EO; Durand et al., 2023), crowdsourced or community observations (Le Coz et al., 2016; Etter et al., 2020), or camera-based sensors (Vandaele et al., 2021). The key requirement is that an observation operator can be defined. Observation operators map the state of the system from state space to observations space. In our study the observation operator selects~~
825 ~~the grid-point that represents the location of the station on the modelled~~ The ability of the method to correct the forecasts typically depends on the consistency of the error vectors between nearby locations. The localisation method implemented here depends only on the distance from the station along the river network. The ~~mapping of the station locations from the physical river network to the modelled river network is not trivial and several studies have attempted to automate this step (Isikdogan et al., 2017; Li et al., 2018). If this mapping is incorrect then representation errors can be introduced~~
830 ~~(Janjić et al., 2018). For example, if a station on a bypass channel is incorrectly located on the main channel. Observations from the station will undoubtedly provide erroneous information in the update step~~ method does successfully correct the forecast both up- and downstream; however, if the station is on a different river or if there is a confluence between the station and the grid-box of interest, the errors are often not consistent for as long a distance along the river network. Therefore, it could be beneficial to investigate the impact of including information about the river stretch into the localisation length scale.
835 Additionally, the errors were found to be more consistent when the catchment was impacted by large-scale weather systems. It may therefore be useful to incorporate information about the meteorological situation into the localisation function as well.

The covariance inflation method used here maintains consistency between the spread of the error ensemble and the spread of the hindcast (Section 5.2). This successfully stops the error ensemble from collapsing such that the observations are not ignored. However, in situations where the uncertainty of the hindcast ensemble is over- or under-estimated the covariance
840 inflation does not correct the error ensemble covariances correctly. This can lead to the observations being ignored as for short

lead-times in Fig. 5b, ~~and could also~~. Additionally, if the hindcast perturbations do not provide an accurate estimate of the true error ensemble perturbations, this method may introduce errors which could be the cause for the slight degradation in skill of the ensemble mean with lead-time shown in Fig. 6e, ~~6f~~e, 6g, and 6h. Correcting the spread of the hindcast before using it in the inflation of the error covariances could solve this issue (~~Section 5.2~~). Covariance inflation techniques that use the innovation
845 statistics could be used to first adjust the hindcast ensemble (e.g., Kotsuki et al., 2017). Alternatively, a lower threshold for the variance of the ensemble could be set - say 10% of the ensemble mean similarly to the observation error covariance matrix or the root mean square-error of the initial conditions. However, caution is needed not to artificially inflate the covariances too much such that the analysis increments become too large, ~~in particular at short lead-times when the correlation is small (Fig. 3)~~.

850 As discussed in Section 7.2.2, the resulting ensemble must be adjusted in some cases to avoid negative discharge values (Section 4.1). This does in some cases lead to ensemble members close to $0 \text{ m}^3 \text{ s}^{-1}$ when a zero flow value is unlikely (e.g., Fig. 5c). This occurs due to the analysis increment being larger in magnitude than the value of some of the raw ensemble members. In general, this is due to the skewed distribution of discharge (Bogner et al., 2012). Future work could look into applying anamorphosis, or normalising transformations, to make the ensemble distribution more Gaussian-like (~~Nguyen et al., 2023~~).
855 ~~The~~ (Nguyen et al., 2023; Bogner et al., 2012). This was not done in this proof-of-concept study for simplicity and to facilitate the interpretation of the errors. The results also showed that the covariances between grid boxes on larger rivers and the station locations tend to be large even when the correlation is small. This is due to larger rivers having larger variances which is partially due to their larger river discharge magnitudes. Localisation ~~does enforce~~ enforces a distance dependence on the covariance magnitudes ~~but~~. However, transforming the river discharge values to be comparable across the domain may ~~also~~
860 ~~help minimise the impact of overestimated ensemble spread~~ help regulate the covariances based on river size. A transformation between river discharge and specific discharge (river discharge divided by upstream area) could be used to ensure that the ensemble covariances more accurately represent the true relationship between locations.

In this study, the initial estimate of the error ensemble mean is defined using the observations and the simulation forced with meteorological observations from the 10 days before the forecast. The average difference between the observations and
865 simulations is calculated at gauged locations and interpolated to every grid-box using inverse distance weighted interpolation. The aim is to provide a physically plausible first guess of the errors which is then updated at each timestep. By taking the average over a 10-day period, we aim to capture the biases of the hydrological model but also to allow for seasonal/dynamic variation in this bias. However, the choice of 10 days has not been tuned, and may be more applicable to larger catchments with slowly changing errors than for smaller catchments (Matthews et al., 2022). Further research into the accuracy of the initial
870 error ensemble, and how it influences the skill of the error-corrected ensemble, is needed. It should be noted that this component of the method is an implementation choice and can be adjusted depending on system configuration and data availability. The only requirement is that the initial error ensemble is physically plausible as there is no warm-up period within this application (Kim et al., 2018).

We assume that the errors are sufficiently slowly changing such that a persistence model can be used to propagate the errors
875 between timesteps. It should be noted that the LETKF updates the errors at each timestep so the analysis errors used to correct

the hindcasts are not constant for all timesteps. However, the assumption that the errors are slowly changing is likely only true for larger rivers that respond more slowly. Future studies could investigate the use of a simple time-varying evolution model. The model would need to be simple enough that the calculations do not add too much computational time to the method. Additionally, the error values at every grid-box would need to be evolved; therefore, the evolution model should either rely
880 only on the model output or must be spatially interpretable if using observations. For example, a model dependent on the hindcast river discharge magnitude could be used to evolve the errors between timesteps.

The leave-one-out approach used in this study allows the corrected ensembles to be assessed at proxy-ungauged locations. However, only one station is omitted at a time. Future work could use a block cross-validation strategy whereby multiple stations are omitted simultaneously (Roberts et al., 2017). This would allow the impact of the density of stations and their
885 specific locations along the river network to be investigated more thoroughly (Rakovec et al., 2012). The impact of not having any observations along a river stretch could also be more thoroughly investigated. One benefit of this method is that the assimilated observations do not necessarily need to be traditional in-situ observations but could come from Earth Observation (EO; Durand et al., 2023), crowdsourced or community observations (Le Coz et al., 2016; Etter et al., 2020), or camera based sensors (Vandaele et al., 2021). The key requirement is that an observation operator can be defined. Observation op-
890 erators map the state of the system from state space to observations space. In our study the observation operator selects the grid-point that represents the location of the station on the modelled river network. The mapping of the station locations from the physical river network to the modelled river network is not trivial and several studies have attempted to automate this step (Isikdogan et al., 2017; Li et al., 2018). If this mapping is incorrect then representation errors can be introduced (Janjić et al., 2018). For example, if a station on a bypass channel is incorrectly located on the main channel, observations from
895 the station will undoubtedly provide erroneous information in the update step.

The code developed for this study is designed to allow for research flexibility rather than operational efficiency. However, the error-adjustment of a single forecast took on average 8.5 minutes for the whole of the Rhine-Meuse catchment - a large catchment. This suggests that, with proper parallelization, the method could be operationalized and applied to all gauged catchments in Europe. Before that, though, the method needs to be evaluated on additional catchments. The Rhine was selected
900 because it is highly gauged, but this also means that the raw ensemble's skill is relatively high due to the effectiveness of the hydrological model calibration process. This could influence the method's performance in two ways: 1) the error ensemble may evolve more linearly than in less calibrated catchments, and 2) the hindcast ensemble's covariance may better represent the covariances between the estimated errors. The next step should be applying this method to a catchment with lower skill than the Rhine.

The method presented in this study spreads observation information along the river network but cannot yet be used as a
905 post-processing method because observations from the hindcast period (the future) are assimilated. We envisage the method being developed further to make it applicable operationally as a hydrological forecast post-processing method. Nevertheless, it may still be useful in certain situations, such as post-event analysis. After a flood event an assessment is often performed estimating the severity of the event as well as potential causes and mitigating factors. However, in-situ river gauges only
910 present a snapshot of the event at specific locations and are often damaged during flood events, resulting in missing or incorrect

data. EO estimations of river discharge could fill in some of the gaps but this would depend on the satellite's orbit and its availability at the right time (Douben, 2006). Reanalysis is another option, but it requires additional hydrological model runs and may contain errors due to the structure of the hydrological model or errors in the meteorological observations. The method proposed here could offer a domain-wide estimate of observations without requiring additional model runs or a "warm-up" period typically needed in hydrological simulations to stabilize antecedent water storage within the catchment.

9 Conclusion

We present and evaluate a data-assimilation-inspired method for spreading observation information from gauged to ungauged locations in a post-processing environment. This method enables the error-correction of an ensemble simulation at all grid boxes. The method utilises state augmentation within an LETKF framework to estimate an ensemble of error vectors. The error vectors are then used to correct each hindcast ensemble member separately.

Overall, the method successfully reduces the errors of the ensemble mean at ungauged locations in leave-one-out experiments. The adjusted ensemble mean has a higher correlation with the observed river discharge and is more able to capture the variability of the river discharge at a point. Whilst the magnitude of the errors are reduced the ensemble spread is not adjusted sufficiently resulting in an under-confident ensemble spread at longer lead-times. The adjusted ensembles are spatially and temporally consistent with the river discharge predictions showing smooth evolution both between grid-boxes on the same river and between lead-times. The method is most limited in its applications to locations further upstream than the assimilated observations and for hindcasts where the variance of the ensemble is incorrectly small which most often happens at shorter lead-times. These limitations can be minimised by further investigation into the localisation approach, for example having a different localisation length upstream and downstream from the assimilated observation, and the covariance inflation approach, which may involve applying a spread-correction to the hindcast ensemble as well as the error-ensemble.

Our method of spreading observation information could be used to improve post-event analysis. However, as the computational requirements and processing time are both small the method could also be developed further to allow for its application to the post-processing of operational forecasts. The prediction of river discharge at ungauged locations is a crucial challenge for hydrological research and once successfully achieved will allow for better preparedness for floods.

Code and data availability. The code used in this study for the error-correction method, evaluation of river discharge forecasts, and generation of the figures presented in this manuscript is available upon request. The river discharge forecast used in this study are from the Copernicus Emergency Management Service (CEMS)'s European Flood Awareness System (EFAS) and are available to download from <https://ewds.climate.copernicus.eu/datasets/efas-forecast>. The local drainage direction and channel length data is available from <https://data.jrc.ec.europa.eu/dataset/f572c443-7466-4adf-87aa-c0847a169f23>.

940 Appendix A: Kalman gain matrix decomposition

The Kalman gain matrix has the following form for timestep k :

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^\top \left(\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^\top + \mathbf{R}_k \right)^{-1}$$

where $\mathbf{P}_k^f$ is the prior ensemble covariance matrix, $\mathbf{R}_k$ is the error-covariance matrix of the observations, and $\mathbf{H}_k$ is the observation operator (Livings, 2005; Hunt et al., 2007; Kalman, 1960). Substituting the definitions of the perturbation matrix
945 and the observation operator for the augmented state given in Eqs. (9) and (11) gives:

$$\mathbf{K}_k = \mathbf{W}_k^f \mathbf{W}_k^{fT} \mathbf{H}_k^\top \left(\mathbf{H}_k \mathbf{W}_k^f \mathbf{W}_k^{fT} \mathbf{H}_k^\top + \mathbf{R}_k \right)^{-1}.$$

This can then be decomposed into the hindcast and error components as

$$\mathbf{K}_k = \begin{pmatrix} \mathbf{K}_{\mathbf{x}_k} \\ \mathbf{K}_{\mathbf{b}_k} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_k^f + \mathbf{B}_k^f \\ \mathbf{B}_k^f \end{pmatrix} (\mathbf{H}_k \mathbf{X}_k^f + \mathbf{H}_k \mathbf{B}_k^f)^T ((\mathbf{H}_k \mathbf{X}_k^f + \mathbf{H}_k \mathbf{B}_k^f)(\mathbf{H}_k \mathbf{X}_k^f + \mathbf{H}_k \mathbf{B}_k^f)^T + \mathbf{R}_k)^{-1} \quad (\text{A1})$$

The analysis of the ensemble mean of the augmented states is therefore given by

$$\bar{\mathbf{w}}_k^a = \begin{pmatrix} \bar{\mathbf{x}}_k^f + \bar{\mathbf{b}}_k^f \\ \bar{\mathbf{b}}_k^f \end{pmatrix} + \begin{pmatrix} \mathbf{K}_{\mathbf{x}_k} \\ \mathbf{K}_{\mathbf{b}_k} \end{pmatrix} (\mathbf{y}_k - \bar{\mathbf{y}}_k^{\mathbf{x}}), \quad (\text{A2})$$

where $\bar{\mathbf{x}}_k^f$ and $\bar{\mathbf{b}}_k^f$ are the ensembles mean of the raw hindcast and the prior error ensemble at timestep k , $\mathbf{y}_k$ is the observation vector, and $\bar{\mathbf{y}}_k^{\mathbf{x}}$ is the model-observation ensemble mean.

Appendix B: Gaspari-Cohn function

The Gaspari-Cohn function is correlation function commonly used in data assimilation to define the localisation weights
955 (Gaspari and Cohn, 1999). It has the following form:

$$\rho(r) = \begin{cases} 1 - \frac{5}{3}r^2 + \frac{5}{8}r^3 + \frac{1}{2}r^4 - \frac{1}{4}r^5, & 0 \leq r \leq 1 \\ -\frac{2}{3}r^{-1} + 4 - 5r + \frac{5}{3}r^2 + \frac{5}{8}r^3 - \frac{1}{2}r^4 + \frac{1}{12}r^5, & 1 < r \leq 2 \\ 0, & r > 2 \end{cases}$$

where $r = d/c$ where d is the physical distance between two points, and c is the localisation length scale. The function has a value of 1 when $r = 0$ and a value of 0 when $r > 2$.

Author contributions. GM, HC, SD, and CP developed the method and designed the study. GM developed the software, performed the
960 forecast error-correction and the evaluation, and drafted the manuscript. All the co-authors contributed to the editing of the manuscript, and to the discussion and interpretation of the results.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. We gratefully acknowledge the financial support provided by the following: Gwyneth Matthews is supported in part by the EPSRC DTP 2018-19 University of Reading (grant number: EP/R513301/1) and ECMWF. Hannah L Cloke is supported by the
965 UKRI Natural Environment Research Council, Evolution of Global Flood Risk (EVOFLOOD; grant number: NE/S015590/1). Sarah L Dance is supported in part by the UKRI Natural Environment Research Council National Centre for Earth Observation (grant numbers: NE/X019063/1, NE/W004984/1, NE/Y006216/1).

References

- Anderson, J. L.: A method for producing and evaluating probabilistic forecasts from ensemble model integrations, *Journal of climate*, 9,
970 1518–1530, 1996.
- Arnal, L., Asp, S., Baugh, C., De Roo, A., Disperati, J., Dottori, F., Garcia, R., GarciaPadilla, M., Gelati, E., Gomes, G., et al.: EFAS upgrade
for the extended model domain—technical documentation, Tech. rep., JRC Technical Reports, EUR 29323 EN, Publications Office of the
European . . . , 2019.
- Barnard, C., Blick, M., Wetterhall, F., Mazzetti, C., Decremer, D., Jurlina, T., Baugh, C., Harrigan, S., Battino, P., and Prudhomme, C.: River
975 discharge and related forecasted data from the European Flood Awareness System, v4.0, <https://doi.org/10.24381/cds.9f696a7a>, [Dataset],
2020.
- Bell, M. J., Martin, M. J., and Nichols, N. K.: Assimilation of data into an ocean model with systematic errors near the equator, *Quarterly
Journal of the Royal Meteorological Society*, 130, 873–893, <https://doi.org/https://doi.org/10.1256/qj.02.109>, 2004.
- Bennett, A., Stein, A., Cheng, Y., Nijssen, B., and McGuire, M.: A process-conditioned and spatially consistent method for reducing system-
980 atic biases in modeled streamflow, *Journal of Hydrometeorology*, 23, 769–783, 2022.
- Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects,
Monthly weather review, 129, 420–436, 2001.
- Boelee, L., Lumbroso, D. M., Samuels, P. G., and Cloke, H. L.: Estimation of uncertainty in flood forecasts—A comparison of methods,
Journal of Flood Risk Management, 12, e12 516, 2019.
- 985 Bogner, K., Pappenberger, F., and Cloke, H. L.: The normal quantile transformation and its application in a flood forecasting system, *Hy-
drology and Earth System Sciences*, 16, 1085–1094, 2012.
- Bourdin, D. R., Fleming, S. W., and Stull, R. B.: Streamflow modelling: a primer on applications, approaches and challenges, *Atmosphere-
Ocean*, 50, 507–536, 2012.
- Choulga, M., Moschini, F., Mazzetti, C., Grimaldi, S., Disperati, J., Beck, H., Salamon, P., and Prudhomme, C.: LISFLOOD static and
990 parameter maps for Europe, <https://doi.org/http://data.europa.eu/89h/f572c443-7466-4adf-87aa-c0847a169f23>, [Dataset], 2023.
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assim-
ilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, *Advances in
Water Resources*, 31, 1309–1324, <https://doi.org/https://doi.org/10.1016/j.advwatres.2008.06.005>, 2008.
- Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, *Journal of hydrology*, 375, 613–626, 2009.
- 995 De Roo, A., Wesseling, C., and Van Deursen, W.: Physically based river basin modelling within a GIS: the LISFLOOD model, *Hydrological
Processes*, 14, 1981–1992, 2000.
- Dee, D. P.: Bias and data assimilation, *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied
meteorology and physical oceanography*, 131, 3323–3343, 2005.
- Douben, K.-J.: Characteristics of river floods and flooding: a global overview, 1985–2003, *Irrigation and Drainage: The journal of the
1000 International Commission on Irrigation and Drainage*, 55, S9–S21, 2006.
- Duc, L., Saito, K., and Hotta, D.: Analysis and design of covariance inflation methods using inflation functions. Part 1: Theoretical framework,
Quarterly Journal of the Royal Meteorological Society, 146, 3638–3660, 2020.

- Durand, M., Gleason, C. J., Pavelsky, T. M., Prata de Moraes Frasson, R., Turmon, M., David, C. H., Altenau, E. H., Tebaldi, N., Larnier, K., Monnier, J., et al.: A framework for estimating global river discharge from the Surface Water and Ocean Topography satellite mission, *Water Resources Research*, 59, e2021WR031 614, 2023.
- El Gharamti, M., McCreight, J. L., Noh, S. J., Hoar, T. J., RafieeiNasab, A., and Johnson, B. K.: Ensemble streamflow data assimilation using WRF-Hydro and DART: novel localization and inflation techniques applied to Hurricane Florence flooding, *Hydrology and Earth System Sciences*, 25, 5315–5336, <https://doi.org/10.5194/hess-25-5315-2021>, 2021.
- Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., et al.: Continental and global scale flood forecasting systems, *Wiley Interdisciplinary Reviews: Water*, 3, 391–418, 2016.
- Engeland, K. and Steinsland, I.: Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times, *Water Resources Research*, 50, 182–197, <https://doi.org/https://doi.org/10.1002/2012WR012757>, 2014.
- Etter, S., Strobl, B., van Meerveld, I., and Seibert, J.: Quality and timing of crowd-based water level class observations, *Hydrological Processes*, 34, 4365–4378, 2020.
- Evensen, G., Vossepoel, F. C., and Van Leeuwen, P. J.: Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem, Springer Nature, 2022.
- Fowler, A. M., Dance, S. L., and Waller, J. A.: On the interaction of observation and prior error correlations in data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 144, 48–62, <https://doi.org/https://doi.org/10.1002/qj.3183>, 2018.
- Furrer, R. and Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants, *Journal of Multivariate Analysis*, 98, 227–255, 2007.
- García-Pintado, J., Mason, D. C., Dance, S. L., Cloke, H. L., Neal, J. C., Freer, J., and Bates, P. D.: Satellite-supported flood forecasting in river networks: A real case study, *Journal of Hydrology*, 523, 706–724, 2015.
- Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757, 1999.
- Gharamti, M. and Hoteit, I.: Complex step-based low-rank extended Kalman filtering for state-parameter estimation in subsurface transport models, *Journal of Hydrology*, 509, 588–600, 2014.
- ~~Gneiting, T., Raftery, A. E., Westveld III, A. H., and Goldman, T.: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation, *Monthly weather review*, 133, 1098–1118, 2005.~~
- Golub, G. H. and Van Loan, C. F.: Matrix computations, JHU press, 2013.
- Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *Journal of hydrology*, 377, 80–91, 2009.
- Hamill, T. M.: Interpretation of rank histograms for verifying ensemble forecasts, *Monthly Weather Review*, 129, 550–560, 2001.
- Hamill, T. M. and Colucci, S. J.: Verification of Eta–RSM short-range ensemble forecasts, *Monthly Weather Review*, 125, 1312–1327, 1997.
- Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Monthly Weather Review*, 129, 2776–2790, 2001.
- Hannah, D., Demuth, S., van Lanen, H., Looser, U., Prudhomme, C., Rees, R., Stahl, K., and Tallaksen, L.: Large-scale river flow archives: importance, current status and future needs, *Hydrological Processes*, 25, 1191–1200, 2011.

- 1040 Harrison, M., Richardson, D., Robertson, K., and Woodcock, A.: Medium-range ensembles using both the ECMWF T63 and unified models—An initial report, UK Meteorological Office Tech. Rep, 153, 25, 1995.
- Hodson, T. O.: Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not, *Geoscientific Model Development*, 15, 5481–5487, <https://doi.org/10.5194/gmd-15-5481-2022>, 2022.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman
1045 filter, *Physica D: Nonlinear Phenomena*, 230, 112–126, 2007.
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation: Operational, sequential and variational (gtspecial issue) data assimilation in meteorology and oceanography: Theory and practice), *Journal of the Meteorological Society of Japan. Ser. II*, 75, 181–189, 1997.
- Isikdogan, F., Bovik, A., and Passalacqua, P.: RivaMap: An automated river analysis and mapping engine, *Remote Sensing of Environment*,
1050 202, 88–97, 2017.
- Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., and Ames, D. P.: Introductory overview: Error metrics for hydrologic modelling - A review of common practices and an open source library to facilitate use and adoption, *Environmental Modelling & Software*, 119, 32–48, <https://doi.org/https://doi.org/10.1016/j.envsoft.2019.05.001>, 2019.
- Janjić, T., Bormann, N., Bocquet, M., Carton, J., Cohn, S. E., Dance, S. L., Losa, S. N., Nichols, N. K., Potthast, R., Waller, J. A., et al.: On
1055 the representation error in data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 144, 1257–1278, 2018.
- Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems, *Journal of Basic Engineering*, 82, 35–45, <https://doi.org/10.1115/1.3662552>, 1960.
- Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., Van Dijk, A. I., Schumacher, M., and Pattiaratchi, C.: Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model, *Advances in Water Resources*, 107, 301–316, 2017.
- 1060 Khaniya, M., Tachikawa, Y., Ichikawa, Y., and Yorozu, K.: Impact of assimilating dam outflow measurements to update distributed hydrological model states: Localization for improving ensemble Kalman filter performance, *Journal of Hydrology*, 608, 127 651, <https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127651>, 2022.
- Kim, K. B., Kwon, H.-H., and Han, D.: Exploration of warm-up period in conceptual hydrological modelling, *Journal of Hydrology*, 556, 194–210, 2018.
- 1065 Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios, *Journal of Hydrology*, 424, 264–277, 2012.
- Kotsuki, S., Ota, Y., and Miyoshi, T.: Adaptive covariance relaxation methods for ensemble data assimilation: Experiments in the real atmosphere, *Quarterly Journal of the Royal Meteorological Society*, 143, 2001–2015, 2017.
- 1070 Krabbenhoft, C. A., Allen, G. H., Lin, P., Godsey, S. E., Allen, D. C., Burrows, R. M., DelVecchia, A. G., Fritz, K. M., Shanafield, M., Burgin, A. J., et al.: Assessing placement bias of the global river gauge network, *Nature Sustainability*, 5, 586–592, 2022.
- Lavers, D. A., Harrigan, S., Andersson, E., Richardson, D. S., Prudhomme, C., and Pappenberger, F.: A vision for improving global flood forecasting, *Environmental Research Letters*, 14, 121 002, <https://doi.org/10.1088/1748-9326/ab52b2>, 2019.
- Le Coz, J., Patalano, A., Collins, D., Guillén, N. F., García, C. M., Smart, G. M., Bind, J., Chiaverini, A., Le Boursicaud, R., Dramais, G.,
1075 et al.: Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand, *Journal of Hydrology*, 541, 766–777, 2016.

- Lee, D.-G. and Ahn, K.-H.: Improving medium-range streamflow forecasts over South Korea with a dual-encoder transformer model, *Journal of Environmental Management*, 368, 122–134, 2024.
- 1080 Li, J., Li, T., Liu, S., and Shi, H.: An Efficient Method for Mapping High-Resolution Global River Discharge Based on the Algorithms of Drainage Network Extraction, *Water*, 10, <https://doi.org/10.3390/w10040533>, 2018.
- Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., and Di, Z.: A review on statistical postprocessing methods for hydrometeorological ensemble forecasting, *Wiley Interdisciplinary Reviews: Water*, 4, e1246, 2017.
- Li, Y., Cong, Z., and Yang, D.: Remotely sensed soil moisture assimilation in the distributed hydrological model based on the error subspace transform Kalman filter, *Remote Sensing*, 15, 1852, 2023.
- 1085 Liu, S., Wang, J., Wang, H., and Wu, Y.: Post-processing of hydrological model simulations using the convolutional neural network and support vector regression, *Hydrology Research*, 53, 605–621, 2022.
- [Livings, D.: Aspects of the ensemble Kalman filter, Reading University Masters Thesis, 2005.](#)
- Livings, D. M., Dance, S. L., and Nichols, N. K.: Unbiased ensemble square root filters, *Physica D: Nonlinear Phenomena*, 237, 1021–1028, 2008.
- 1090 Lu, G. Y. and Wong, D. W.: An adaptive inverse-distance weighting spatial interpolation technique, *Computers & Geosciences*, 34, 1044–1055, <https://doi.org/10.1016/j.cageo.2007.07.010>, 2008.
- Martin, M., Bell, M., and Nichols, N. K.: Estimation of systematic error in an equatorial ocean model using data assimilation, *International Journal for Numerical Methods in Fluids*, 40, 435–444, 2002.
- 1095 Martin, M. J.: Data assimilation in ocean circulation models with systematic errors., Ph.D. thesis, University of Reading, 2001.
- Mason, D., Garcia Pintado, J., Cloke, H. L., Dance, S., and Munoz-Sabater, J.: Assimilating high resolution remotely sensed soil moisture into a distributed hydrologic model to improve runoff prediction, *ECMWF Technical Memorandum*, 2020.
- [Matthews, G., Barnard, C., Cloke, H., Dance, S. L., Jurlina, T., Mazzetti, C., and Prudhomme, C.: Evaluating the impact of post-processing medium-range ensemble streamflow forecasts from the European Flood Awareness System, *Hydrology and Earth System Sciences*, 26, 2939–2968, 2022.](#)
- 1100 Matthews, G., Baugh, C., Barnard, C., De Wiart, C. C., Colanese, J., Decremier, D., Grimaldi, S., Hansford, E., Mazzetti, C., O'Regan, K., Pappenberger, F., Ramos, A., Salamon, P., Tasev, D., and Prudhomme, C.: Chapter 14 - On the operational implementation of the European Flood Awareness System (EFAS), in: *Flood Forecasting (Second Edition)*, edited by Adams, T. E., Gangodagamage, C., and Pagano, T. C., pp. 251–298, Academic Press, second edition edn., ISBN 978-0-443-14009-9, <https://doi.org/10.1016/B978-0-443-14009-9.00005-5>, 2025.
- 1105 McMillan, H., Krueger, T., and Freer, J.: Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality, *Hydrological Processes*, 26, 4078–4111, 2012.
- McMillan, H. K., Westerberg, I. K., and Krueger, T.: Hydrological data uncertainty and its implications, *WIREs Water*, 5, e1319, <https://doi.org/10.1002/wat2.1319>, 2018.
- 1110 Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., Ehmele, F., Feldmann, H., Franca, M. J., Gattke, C., Hundhausen, M., Knippertz, P., K  pfer, K., M  hr, B., Pinto, J. G., Quinting, J., Sch  fer, A. M., Scheibel, M., Seidel, F., and Wisotzky, C.: A

- multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe – Part 1: Event description and analysis, *Natural Hazards and Earth System Sciences*, 23, 525–551, <https://doi.org/10.5194/nhess-23-525-2023>, 2023.
- 1115 Nguyen, T. H., Ricci, S., Piacentini, A., Simon, E., Rodriguez-Suquet, R., and Luque, S. P.: Gaussian anamorphosis for ensemble kalman filter analysis of SAR-derived wet surface ratio observations, *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- Nichols, N. K.: Data Assimilation: Aims and Basic Concepts, in: *Data Assimilation for the Earth System*, edited by Swinbank, R., Shutyaev, V., and Lahoz, W. A., pp. 9–20, Springer Netherlands, Dordrecht, ISBN 978-94-010-0029-1, 2003.
- 1120 Nichols, N. K.: Mathematical concepts of data assimilation, *Data assimilation: making sense of observations*, pp. 13–39, 2010.
- Pauwels, V. R., Hendricks Franssen, H.-J., and De Lannoy, G. J.: Evaluation of State and Bias Estimates for Assimilation of SMOS Retrievals Into Conceptual Rainfall-Runoff Models, *Frontiers in Water*, 2, 4, <https://doi.org/10.3389/frwa.2020.00004>, 2020.
- Pechlivanidis, I. G., Du, Y., Bennett, J., Boucher, M.-A., Chang, A. Y., Crochemore, L., Dasgupta, A., Di Baldassarre, G., Luterbacher, J., Pappenberger, F., et al.: Enhancing research-to-operations in hydrological forecasting: innovations across scales and horizons, *Bulletin of the American Meteorological Society*, 106, E894–E919, 2025.
- 1125 Piazzì, G., Thirel, G., Perrin, C., and Delaigue, O.: Sequential data assimilation for streamflow forecasting: Assessing the sensitivity to uncertainties and updated variables of a conceptual hydrological model at basin scale, *Water Resources Research*, 57, 2021.
- Pugliese, A., Persiano, S., Bagli, S., Mazzoli, P., Parajka, J., Arheimer, B., Capell, R., Montanari, A., Blöschl, G., and Castellarin, A.: A geostatistical data-assimilation technique for enhancing macro-scale rainfall–runoff simulations, *Hydrology and Earth System Sciences*, 22, 4633–4648, 2018.
- 1130 Rakovec, O., Weerts, A., Hazenberg, P., Torfs, P., and Uijlenhoet, R.: State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy, *Hydrology and Earth System Sciences*, 16, 3435–3449, 2012.
- Rasmussen, J., Madsen, H., Jensen, K. H., and Refsgaard, J. C.: Data assimilation in integrated hydrological modelling in the presence of observation bias, *Hydrology and earth system sciences*, 20, 2103–2118, <https://doi.org/10.5194/hess-20-2103-2016>, 2016.
- 1135 Refsgaard, J., Van der Keur, P., Nilsson, B., Müller-Wohlfeil, D.-I., and Brown, J.: Uncertainties in river basin data at various support scales—Example from Odense Pilot River Basin, *Hydrology and Earth System Sciences Discussions*, 3, 1943–1985, 2006.
- Ridler, M.-E., Zhang, D., Madsen, H., Kidmose, J., Refsgaard, J. C., and Jensen, K. H.: Bias-aware data assimilation in integrated hydrological modelling, *Hydrology Research*, 49, 989–1004, 2018.
- 1140 Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Aroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, *Ecography*, 40, 913–929, 2017.
- Rouzies, E., Lauvernet, C., and Vidard, A.: Comparison of different ensemble assimilation methods in a modular hydrological model dedicated to water quality management, *Hydrology and Earth System Sciences Discussions*, 2024, 1–33, 2024.
- 1145 Scheffler, G., Carrassi, A., Ruiz, J., and Pulido, M.: Dynamical effects of inflation in ensemble-based data assimilation under the presence of model error, *Quarterly Journal of the Royal Meteorological Society*, 148, 2368–2383, <https://doi.org/10.1002/qj.4307>, 2022.
- Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging-geostatistics on stream networks, *Hydrology and Earth System Sciences*, 10, 277–287, 2006.
- 1150 Skøien, J. O., Bogner, K., Salamon, P., Smith, P., and Pappenberger, F.: Regionalization of post-processed ensemble runoff forecasts, *Proceedings of the International Association of Hydrological Sciences*, 373, 109–114, 2016.

- Skøien, J. O., Bogner, K., Salamon, P., and Wetterhall, F.: On the Implementation of Postprocessing of Runoff Forecast Ensembles, *Journal of Hydrometeorology*, 22, 2731 – 2749, <https://doi.org/10.1175/JHM-D-21-0008.1>, 2021.
- 1155 Smith, P., Thornhill, G., Dance, S., Lawless, A., Mason, D., and Nichols, N.: Data assimilation for state and parameter estimation: application to morphodynamic modelling, *Quarterly Journal of the Royal Meteorological Society*, 139, 314–327, 2013.
- Smith, P., Pappenberger, F., Wetterhall, F., Del Pozo, J. T., Krzeminski, B., Salamon, P., Muraro, D., Kalas, M., and Baugh, C.: On the operational implementation of the European Flood Awareness System (EFAS), in: *Flood forecasting*, pp. 313–348, Elsevier, 2016.
- Smith, P. J., Dance, S. L., Baines, M. J., Nichols, N. K., and Scott, T. R.: Variational data assimilation for parameter estimation: application
1160 to a simple morphodynamic model, *Ocean Dynamics*, 59, 697–708, 2009.
- Stewart, L. M., Dance, S. L., and Nichols, N. K.: Data assimilation with correlated observation errors: experiments with a 1-D shallow water model, *Tellus A: Dynamic Meteorology and Oceanography*, 65, 19546, <https://doi.org/10.3402/tellusa.v65i0.19546>, 2013.
- Strahler, A. N.: Quantitative analysis of watershed geomorphology, *Eos, Transactions American Geophysical Union*, 38, 913–920, 1957.
- Talagrand, O.: Evaluation of probabilistic prediction systems, in: *Workshop Proceedings" Workshop on Predictability"*, 20-22 October 1997,
1165 ECMWF, Reading, UK, 1999.
- [Tanguy, M., Eastman, M., Chevuturi, A., Magee, E., Cooper, E., Johnson, R. H., Facer-Childs, K., and Hannaford, J.: Optimising ensemble streamflow predictions with bias correction and data assimilation techniques, *Hydrology and Earth System Sciences*, 29, 1587–1614, 2025.](#)
- 1170 UNDRR: Sendai Framework for Disaster Risk Reduction 2015–2030, Tech. rep., United Nations Office for Disaster Risk Reduction, Geneva, <https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030>, 2015.
- Valdez, E. S., Anctil, F., and Ramos, M.-H.: Choosing between post-processing precipitation forecasts or chaining several uncertainty quantification tools in hydrological forecasting systems, *Hydrology and Earth System Sciences*, 26, 197–220, 2022.
- 1175 Valler, V., Franke, J., and Brönnimann, S.: Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation, *Climate of the Past*, 15, 1427–1441, 2019.
- Van Der Knijff, J., Younis, J., and De Roo, A.: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, *International Journal of Geographical Information Science*, 24, 189–212, 2010.
- 1180 Vandaele, R., Dance, S. L., and Ojha, V.: Deep learning for automated river-level monitoring through river-camera images: an approach based on water segmentation and transfer learning, *Hydrology and Earth System Sciences*, 25, 4435–4453, 2021.
- Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., Lerch, S., Roberts, N., Theis, S., Atencia, A., Bouallāgue, Z. B., Bhend, J., Dabernig, M., Cruz, L. D., Hieta, L., Mestre, O., Moret, L., PlenkoviÄž, I. O., Schmeits, M., Taillardat, M., den Bergh, J. V., Schaeybroeck, B. V., Whan, K., and Ylhäisi, J.: Statistical Postprocessing for Weather Forecasts: Review, Challenges,
1185 and Avenues in a Big Data World, *Bulletin of the American Meteorological Society*, 102, E681 – E699, <https://doi.org/10.1175/BAMS-D-19-0308.1>, 2021.
- ~~Whitaker, J. S. and Hamill, T. M.: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation, *Monthly Weather Review*, 140, 3078 – 3089~~

- WMO: Hydromet Gap Report 2024, Tech. rep., ~~2012~~-World Meteorological Organization and Alliance for Hydromet Development, <https://library.wmo.int/idurl/4/68926>, accessed: 2025-01-10, 2024.
- 1190 Woldemeskel, F., McInerney, D., Lerat, J., Thyer, M., Kavetski, D., Shin, D., Tuteja, N., and Kuczera, G.: Evaluating post-processing approaches for monthly and seasonal streamflow forecasts, *Hydrology and Earth System Sciences*, 22, 6257–6278, 2018.
- Wu, W., Emerton, R., Duan, Q., Wood, A. W., Wetterhall, F., and Robertson, D. E.: Ensemble flood forecasting: Current status and future opportunities, *Wiley Interdisciplinary Reviews: Water*, 7, e1432, 2020.
- 1195 Xie, X. and Zhang, D.: Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter, *Advances in Water Resources*, 33, 678–690, <https://doi.org/https://doi.org/10.1016/j.advwatres.2010.03.012>, 2010.
- Xu, J., Anctil, F., and Boucher, M.-A.: Hydrological post-processing of streamflow forecasts issued from multimodel ensemble prediction systems, *Journal of Hydrology*, 578, 124 002, 2019.
- Ye, A., Duan, Q., Yuan, X., Wood, E. F., and Schaake, J.: Hydrologic post-processing of MOPEX streamflow simulations, *Journal of hydrology*, 508, 147–156, 2014.
- 1200 Zhang, F., Snyder, C., and Sun, J.: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter, *Monthly Weather Review*, 132, 1238 – 1253, [https://doi.org/10.1175/1520-0493\(2004\)132<1238:IOIEAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2), 2004.