AI-Empowered Latent Four-dimensional Variational Data Assimilation for River Discharge Forecasting

Kun Wang, Gabriele Bertoli, Sibo Cheng, Kai Schröter, Enrica Caporali, Matthew D. Piggott, Yanghua Wang, and Rossella Arcucci

***Abstract*—Flood forecasting and warning, as a proactive strategy to mitigate potential adverse consequences, have attracted significant attention. However, river discharge forecasting, a crucial component of flood forecasting, presents challenges due to the high dimensionality of its parameters, often suffering from issues, such as low forecasting accuracy and high computational costs. 4-D variational (4D-Var) data assimilation, as a technique that integrates information from various sources to improve forecasting accuracy, has the advantage of incorporating the time dimension, which enables it to provide more precise predictions, making it well suited for river discharge forecasting. However, traditional hydrological forecasting models and 4D-Var methods are highly time-consuming, and 4D-Var requires access to tangent linear and adjoint models in order to evaluate the cost function, which limits their practical application in river discharge forecasting. Therefore, this article proposes an AI-empowered latent 4D-Var methodology to address these two issues: eliminating the reliance on the tangent linear model and adjoint model, and lowering the computational cost associated with traditional methods. The method first uses a convolutional autoencoder to compress the state field into a latent space, then employs a long short-term memory (LSTM) network as the surrogate model for the forward model in this latent space, and finally minimizes the cost function of 4D-Var directly in the latent space without the need for the tangent linear or adjoint models. We test the proposed AI-empowered latent 4D-Var on real datasets, utilizing data from the European Flood Awareness System (EFAS) as the state field and LamaH-CE as the observational data. Our method outperforms the EFAS historical simulation and the two baseline models, Voronoi-based LSTM and latent 3-D variational data assimilation, across five evaluation metrics. Furthermore, the proposed method completes 500 iterations for the predictions of the next day river discharge in just 100 s, demonstrating significant computational efficiency. In practice, we tested our method on a flood event occurred in June 2013, where it provided more accurate and timely forecasting compared to other methods, particularly the EFAS historical simulation. Moreover, this method is not only applicable to flood forecasting but also holds significant potential in various other fields.**

***Index Terms*—Latent data assimilation, 4DVar, flood forecasting, neural network, data learning, river discharge.**

Received 20 June 2025; revised 19 August 2025; accepted 7 September 2025. Date of publication 17 September 2025; date of current version 3 October 2025. The work of Kun Wang was supported by Resource Geophysics Academy at Imperial College London. The work of Gabriele Bertoli was supported by Italian National Recovery and Resilience Plan (PNRR), European Union – NextGenerationEU, under Ministerial Decree 351/2022. *(Corresponding author: Rossella Arcucci.)*

Kun Wang is with Resource Geophysics Academy, Imperial College London, SW7 2BP London, U.K., and also with the Department of Earth Science and Engineering, Imperial College London, SW7 2AZ London, U.K..

Gabriele Bertoli and Enrica Caporali are with the Department of Civil and Environmental Engineering, University of Florence, 50121 Firenze, Italy.

Sibo Cheng is with CEREA, ENPC, and EDF R&D, Institut Polytechnique de Paris, 91120 Palaiseau, France.

Kai Schröter is with the Leichtweiß-Institute for Hydraulic Engineering and Water Resources, Division Hydrology and River Basin Management, Technische Universität Braunschweig, 38106 Braunschweig, Germany.

Matthew D. Piggott is with the Department of Earth Science and Engineering, Imperial College London, SW7 2AZ London, U.K..

Yanghua Wang is with Resource Geophysics Academy, Imperial College London, SW7 2BP London, U.K..

Rossella Arcucci is with the Department of Earth Science and Engineering, Imperial College London, SW7 2AZ London, U.K., and also with Data Science Institute, Imperial College London, SW7 2AZ London, U.K. (e-mail: r.arcucci@imperial.ac.uk).

Digital Object Identifier 10.1109/JSTARS.2025.3611136

I. Introduction

RIVER floods are among the most impacting natural hazards, causing significant loss of life, displacement of people, and economic damage worldwide [1], with expected increasing impacts in the future [2], [3]. Flood forecasting and warning systems help reduce the vulnerability of people and assets to potential flood events by enabling authorities to act in advance and to enact necessary protection measures. Flood forecasting models aim to forecast river discharge, usually expressed in m^3/s, and are generally categorized into hydrological (process-based), data-driven, and hybrid models. Hydrological models aim at simulating complex rainfall–runoff processes, predicting the expected river discharge for given conditions (rainfall, soil moisture, groundwater state, etc). Examples of these models are HEC-HMS [4], which is suitable for catchment-scale modelling, or soil and water assessment tool for long-term predictions [5]. Hydrological modeling is often coupled with hydraulic modelling, to seamlessly simulate how floodwaters move across the landscape. One example of a model that integrates both runoff and hydraulic modelling is the LISFLOOD model, developed by the Joint Research Centre of the European Commission [6], dedicated to large-scale flood forecasting. It is operatively used by European Flood Awareness System (EFAS) [7] and Global Flood Awareness System [8] systems for providing large-scale flood forecasting data. Data-driven flood forecasting techniques usually leverage statistical and machine learning technologies, combined with historical observations, to forecast river discharge [9], [10], [11] without relying on physical laws, while hybrid models combine multiple technologies and modelling techniques together, e.g., blending machine learning tools with hydrological modelling [12]. Among the current limitations of the abovementioned techniques, there are the time and the resources required for running the models, especially at large spatial scales, and the uncertainties involved in the modelling (e.g., model state uncertainties and parameter uncertainties),

© 2025 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

affecting both the availability and the reliability of the predictions [13], [14]. Enhancing flood forecasting capabilities, in terms of efficiency and accuracy, would represent a strong contribution to the effectiveness of flood early warning systems, acknowledged to be seminal instruments for saving lives during floods [15] and reducing losses [16].

Data Assimilation (DA) is a method that enhances the predictive accuracy of models by integrating observational data, potentially from multiple sources, into the dynamic processes of the forecasting model. This method can account for both the spatial and temporal distribution of the data, as well as potential errors. DA is widely applied in river discharge and flood forecasting to improve model performance [17], [18], [19], [20]. 4-D variational (4-D Var) DA obtains the analysis field—an optimal estimate of the system state—by minimizing a cost function, incorporating both the forward model and the temporal distribution of the observations. This method yields more accurate predictions than model simulations and is particularly effective for solving high-dimensional, nonlinear time series problems, such as flood and river discharge forecasting [21], [22], [23]. Traditional Four-dimensional Variational Data Assimilation (4D-Var) relies on the tangent linear model and the adjoint model to solve the cost function [24]. In practical applications, these two models are approximated numerically by discretising the continuous models or by directly constructing discrete models, leading to increased errors and computational costs of 4D-Var [25]. Currently, the authors in[25] and [26] considered the use of neural networks (NNs) to provide the tangent linear model and adjoint model. These methods reduce computational cost and leverage the nonlinear mapping capabilities of Neural Networks (NNs) for potentially improved accuracy [27]. However, these methods still rely on approximations, which can introduce inaccuracies. In addition, a method [28], [29] has been proposed to minimize the cost function without the need for the tangent linear model and adjoint model. This approach leverages the long short-term memory (LSTM) memory mechanism to learn how to adjust the direction of the state update based on the current gradient information of the cost function. As a result, the optimization process not only updates the state according to the current gradient but also incorporates historical information, thereby improving the optimization efficiency. However, a drawback of this approach is that the LSTM model needs to be retrained whenever a new observation operator is introduced, which is inconvenient in practical applications.

In addition, timeliness is crucial in river discharge forecasting, meaning that efficient predictive models are essential [30]. Currently, NN-based surrogate models are used to replace traditional physics-based hydrological models, such as LISFLOOD and MIKE SHE [31], to achieve faster computational speeds [32], [33], [34]. Among these surrogate models, Long Short-term Memory (LSTM) is particularly effective in addressing long-term dependence issues in time series prediction, while utilizing relatively fewer trainable parameters, which makes it widely used in river discharge forecasting [35], [36], [37]. Furthermore, substantial high-dimensional input data for river discharge forecasting models lead to significant computational load and inefficiency [38], [39], [40]. Moreover, when applying 4D-Var to improve the accuracy of river discharge predictions, it is crucial to explore strategies to reduce the computational load, as the iterative process required for optimizing the cost function is computationally intensive. To achieve this, latent DA [41], [42], [43] has been proposed to compress the state field into a latent space using compression models, effectively reducing its dimensionality and, consequently, lowering the computational burden. In particular, NN-based compression models, such as autoencoder [44] and convolutional autoencoder (CAE) [45], excel at extracting nonlinear features, making them a more advantageous choice for practical applications [46], [47] compared with linear transformation-based compression models, such as proper orthogonal decomposition and principal component analysis [48], [49].

To address the following two key issues: reliance on the tangent linear and adjoint models for solving the cost function of 4D-Var, and the high computational costs of traditional 4D-Var methods and forecasting models, this article proposes an AI-empowered latent 4D-Var method to achieve both accurate and efficient river discharge forecasting. The method obtains the analysis field by calculating the derivatives of the cost function without relying on the tangent linear model and adjoint model, thereby improving both accuracy and efficiency. In addition, it leverages NNs as both the compression model and the forward model, enabling 4D-Var to be performed in the latent space, which significantly enhances computational efficiency. To evaluate the performance of our method, we compare it with EFAS historical simulations and establish two baseline models: the Voronoi-based LSTM [50], [51], and the AI-empowered latent 3-D variational (3D-Var) DA. The proposed method is tested on real world datasets, with state field data sourced from the EFAS historical simulation dataset [52] and observation data obtained from the LamaH-CE dataset [53]. In addition, to evaluate the method's effectiveness in predicting real flood events, we select the flood that occurred in June 2013 as a test case.

In summary, this study introduces the AI-empowered latent 4D-Var method that integrates 4D-Var with machine learning, effectively overcoming the limitations of existing 4D-Var methods for river discharge forecasting. The key contributions include the following.

1) The method allows for 4D-Var operations to be performed in the latent space by leveraging NNs as both the compression model and the forward model. This significantly improves computational efficiency compared to traditional methods.
2) The effectiveness of the proposed method is rigorously evaluated against existing models. It is compared with the EFAS historical simulation and two baseline models: a Voronoi-based LSTM model and an AI-empowered latent 3D-Var model, showcasing its superior performance.
3) The method is tested using real datasets, including state field data from the EFAS historical simulation and observation data from the LamaH-CE dataset, including real flood events, demonstrating its practical applicability and robustness in real-world scenarios.
4) In addition, we achieve next-day river discharge forecasting in just 100 s after 500 iterations of cost function

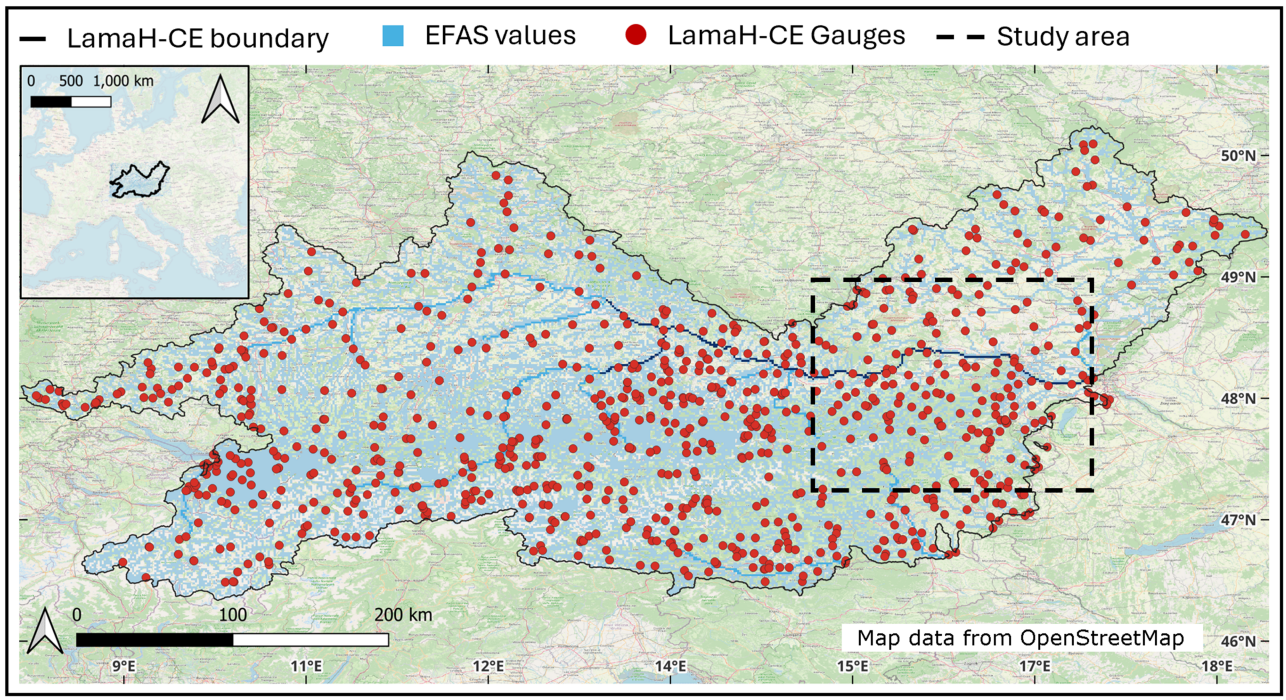


Fig. 1. Map showing the LamaH-CE boundary and hydrometric gauges, indicated by red dots. Mean EFAS river flow simulation values are illustrated in shades of blue, with darker shades representing higher river discharges. The dashed frame outlines the subset focused on in this study. In the top-left corner is shown the location of the dataset with respect to the European continent. The background map is from OpenStreetMap.

minimization, highlighting the method's efficiency in delivering timely flood forecasting.

The rest of this article is organized as follows. Section II introduces the datasets utilized in this work, along with the preprocessing steps applied to the data. Section III provides a detailed description of the proposed method and the baseline models. Section IV presents the numerical results, accompanied by an analysis and discussion of the findings, and one test case. Finally, Section V summarizes the overall findings and outlines plans for future work.

II. Data

A. Study Area and Dataset

The study area covers approximately 30 000 km^2 in Austria, Europe, including the capital Wien and a portion of its upstream catchments. We select this area due to the availability of river discharge observations, from the LamaH-CE large-scale hydrological dataset for central Europe, combined with the possibility to obtain predictions of environmental variables (including river discharge) from EFAS historical simulation dataset. LamaH-CE dataset contains historical observations of river discharges from 859 hydrometric gauges, across nine countries in central Europe, at hourly resolution, covering the period from 1981 to 2017. EFAS is an early warning system designed to provide pan-European flood forecasting and flood alerts. It is jointly developed by the European Commission and the European Centre for medium-range weather forecasts, with its first version officially been released in 2018. EFAS historical simulations are available from 1991 to present days, they have 6 h time resolution and 1 arc-min of spatial resolution, thus being particularly suitable for larger rivers and large spatial-scales. EFAS river discharge forecastings are based on the LISFLOOD hydrological model, integrating meteorological forecasts, historical, and real time data. The map in Fig. 1 shows the LamaH-CE dataset boundary and its gauges, represented with red dots. EFAS simulations are reported as their mean value over the whole time period—for illustration purposes only—in blue shades, where darker blue is for higher river discharges. The dashed frame highlights the subset area analyzed in the present study.

B. Data Preprocessing

After selecting the study area and datasets, we proceeded preprocessing both datasets, to prepare them for being analyzed. In this experiment, we utilize data from 1992 to 2017, training our model on data from 1992 to 2012 and testing it on data from 2013 to 2017. Consequently, within the study area, there are 58 gauge stations in the LamaH-CE dataset whose operational periods align with this time window. Given the data source typology difference between the two datasets, the next step consisted in associating the EFAS raster pixels to the corresponding river gauge. The resolution of the EFAS grid (1 arc-min) is designed for large-scale modeling but does not accurately capture narrow river networks and dense tributaries, introducing uncertainties when identifying the appropriate EFAS cell for a specific gauge, as these cells may not overlap precisely with river gauges. To address this challenge, an automatic cell selection algorithm, first introduced by Wang et al. [54], was applied to achieve the most accurate association of EFAS pixels with LamaH-CE river

TABLE I
PREPROCESSED DATA IN TWO DATASETS

Names	EFAS	LamaH-CE
Type	state field	observation
Variable	river discharge	
Time coverage	1992–2017	
Temporal resolution	6-h	
Spatial resolution	1 arc-min × 1 arc-min	\
Number of gauge	\	58
Dimension of data	128×128	58×1

gauges. This method selects the EFAS cells with the highest flow values within a watershed domain, which are located at the catchment outlet and thus comparable to the associated river gauges. However, some peculiar watershed shapes (e.g., very narrow outlets) can affect the automatic selection accuracy, necessitating manual checks to reduce uncertainties. LamaH-CE offers hourly resolution observations, while EFAS only has 6 h average river discharge simulation values. Even though it is always preferable to avoid reducing the resolution of the available data, we opted for this to allow a perfect alignment between EFAS and LamaH-CE river discharge values without introducing further uncertainties. Thus, hourly observations from LamaH-CE dataset are averaged over 6-h. LamaH-CE gauges observations are eventually aggregated inside a unique matrix. The preprocessed data are presented in Table I.

Then, to ensure that the data maintain a uniform numerical scale, avoid large disparities between different features, and mitigate the impact of noise and outliers—thereby improving the training effectiveness and stability of the network—it is generally necessary to normalize data before training NNs. In this experiment, due to the substantial differences in river discharge between main streams and tributaries, we designed a customized normalization approach using logarithmic scaling, as shown in the following:

$$\mathbf{x}_{\text{norm}} = \frac{\log(\mathbf{x}+1)}{\log(\mathbf{x}_{\max}+1)} \tag{1}$$

where $\mathbf{x}$ represents the state field, $\mathbf{x}_{\max}$ represents the maximum value of the state field, $\mathbf{x}_{\text{norm}}$ represents the normalized state field. Following this process, both the state field and observation data are normalized and prepared for use in subsequent experiment.

III. METHODOLOGY

In this section, we detail our methodology, including the design of the NN-based compression and forward models, as well as the proposed AI-empowered latent 4D-Var method. In addition, we introduce two baseline models to facilitate performance comparison with our proposed method.

A. Compression Model

In the experiment, the state field, derived from river discharge data in the EFAS historical simulation dataset and originally sized 128×128, is compressed using an NN-based compression model. This model reduces the dimensionality of the state field into a latent space of size 128×1, enabling more efficient processing in the latent space. The structure of the compression model is a Convolutional Autoencoder (CAE), composed of an encoder $\mathcal{E}$ and a decoder $\mathcal{D}$, as expressed in the formula given as follows:

$$\begin{aligned} \tilde{\mathbf{x}}_t &= \mathcal{E}(\mathbf{x}_t) \\ \hat{\mathbf{x}}_t &= \mathcal{D}(\tilde{\mathbf{x}}_t) \\ \mathcal{L} &= \frac{1}{n}\sum_{t=1}^{n} \|\hat{\mathbf{x}}_t - \mathbf{x}_t\|^2 \end{aligned} \tag{2}$$

where $\mathbf{x}_t$ represents the state field at the t_{th} time step, whereas $\tilde{\mathbf{x}}_t$ denotes the compressed latent representation of the state field $\mathbf{x}_t$ at the same time step, which can be reconstructed from the latent space using the decoder, $\hat{\mathbf{x}}_t$ represents the reconstructed state field at the t_{th} time step, and $\mathcal{L}$ represents the loss function of the training process.

CAE is an NN structure designed to extract spatial features from input data using convolutional layers in the encoder for dimensionality reduction, and subsequently reconstruct compressed input through convolutional layers in the decoder [45]. This architecture is effective for compressing high-dimensional data while preserving essential features for reconstruction. The CAE trains its model parameters by minimizing the distance between the original input and its reconstruction, effectively learning to compress and reconstruct the data. To enhance the extraction of spatial features from the state field, both the encoder and decoder incorporate residual blocks, which improve feature learning and overall model performance [55]. Each encoder and decoder consists of one fully connected layer and four residual blocks, with each residual block comprising two 2-D convolutional layers, two BatchNorm layers, and one ReLU activation function, as illustrated in Fig. 2(a). The detailed structural information of the CAE is presented in Table II.

TABLE II
DETAILED STRUCTURAL INFORMATION OF CAE

Types	Layers	Input	Output
Encoder	Residual block	$1 \times 128 \times 128$	$64 \times 64 \times 64$
	Residual block	$64 \times 64 \times 64$	$128 \times 32 \times 32$
	Residual block	$128 \times 32 \times 32$	$256 \times 16 \times 16$
	Residual block	$256 \times 16 \times 16$	$512 \times 8 \times 8$
	Fully connected layer	$512 \times 8 \times 8$	128×1
Decoder	Fully connected layer	128×1	$512 \times 8 \times 8$
	Residual block	$512 \times 8 \times 8$	$256 \times 16 \times 16$
	Residual block	$256 \times 16 \times 16$	$128 \times 32 \times 32$
	Residual block	$128 \times 32 \times 32$	$64 \times 64 \times 64$
	Residual block	$64 \times 64 \times 64$	$1 \times 128 \times 128$

After compressing the state field through the encoder, the representation of the state field in the latent space is obtained, which will be used to train the forward model. Furthermore, DA will also be conducted in the latent space.

B. Forward Model in Latent Space

Once the state field is represented in the latent space, a forward model is developed to forecast the future states in that space, enhancing computational efficiency. In this experiment, LSTM is selected as the forward model to forecast the state field, with its structure, as illustrated in Fig. 2(b). LSTM is a variant of

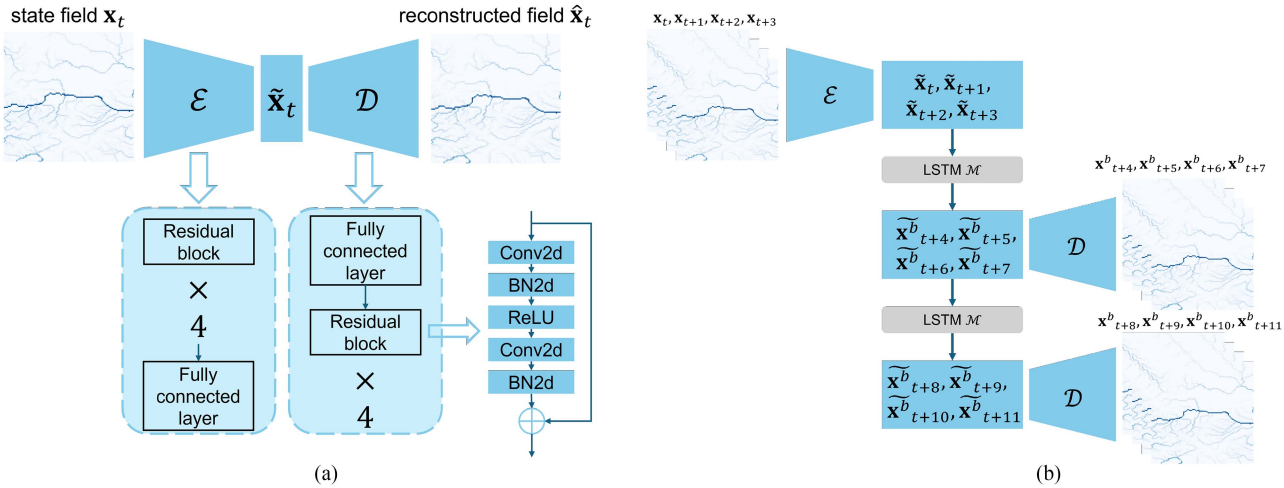


Fig. 2. Structure of the NN-based models. (a) Structure of the compression model, where the input is the state field at the t_{th} time step, and the output is the reconstructed field at the same time step. The encoder and decoder consist of four residual blocks and one fully connected layer. Each residual block contains two Conv2D layers, two BatchNorm layers, and one ReLU activation function. (b) Structure of the forward model in latent space. The input consists of the compressed state field for four time steps in latent space, and the output is the background field in the latent space for the subsequent four time steps. The output then passes through the decoder, where the background field for the following four time steps is reconstructed. (a) Structure of the compression model. (b) Structure of the forward model.

recurrent NN designed to effectively manage and predict the sequential data. It features three specialized gate structures: the input gate, the forget gate, and the output gate. These gates allow LSTM to selectively remember and forget information, making it particularly adept at addressing long-term dependence issues [56].

In designing the LSTM, a 4×4 structure is used to achieve fast prediction cycles and stable forward forecasting. The input to the LSTM $\mathcal{M}$ consists of the state field in the latent space over four time steps, denoted as $\tilde{\mathbf{x}}_{t:t+3}$. The output, referred to as the background field in DA, represents the forecasted state field in the latent space for the next four time steps, denoted as $\widetilde{\mathbf{x}^b}_{t+4:t+7}$. This relationship can be expressed mathematically, as shown in the following:

$$\widetilde{\mathbf{x}^b}_{t+4:t+7} = \mathcal{M}(\tilde{\mathbf{x}}_{t:t+3}). \tag{3}$$

In designing the LSTM structure, the model consists of one LSTM layer, two fully connected layers, one LayerNorm layer, and one ReLU activation function. Detailed structural information of the LSTM is presented in Table III.

The LSTM parameters are trained by minimizing the distance between the predicted background field and the ground truth over four time steps in latent space. Despite this training, the initial predictions of the state field remain preliminary and exhibit moderate accuracy; thus, integrating DA techniques with observations is crucial for improving forecasting accuracy.

C. AI-Empowered Latent 4D-Var

In this experiment, we propose the AI-empowered latent 4D-Var to enhance the forecasting accuracy of the designed LSTM and address the practical limitations of traditional 4D-Var, particularly the reliance on a tangent linear model and an adjoint model, along with the substantial computational load and approximation errors. The structure is illustrated in Fig. 3.

TABLE III
DETAILED STRUCTURAL INFORMATION OF LSTM

Layers	Types	Parameters
LSTM	Input dim	128
	Hidden dim	128
	Output dim	128
	Number of layers	4
	Length of sequence	4
	Length of prediction	4
Fully connected layer	Input dim	128
	Output dim	128
Activative function	ReLU	default
Normalization	LayerNorm	default
Fully connected layer	Input dim	128
	Output dim	128

In this method, the four time steps of the background field $\mathbf{x}^b_{t:t+3}$ predicted by the forward model $\mathcal{M}$ are compressed into the latent space, resulting in the representation of the background field in the latent space $\widetilde{\mathbf{x}^b}_{t:t+3}$. Then, eight time steps of observations in vector space $\tilde{\mathbf{y}}_{t:t+7}$ are used to assimilate the background field, and by minimizing the cost function of AI-powered latent 4D-Var $\mathcal{J}$, the analysis field in the latent space $\widetilde{\mathbf{x}^a}_{t:t+3}$ is obtained. The cost function of AI-empowered latent 4D-Var is expressed mathematically in the following:

$$\begin{aligned}\mathcal{J}(\tilde{\mathbf{x}}_{t:t+3}) = &\left\|\tilde{\mathbf{x}}_{t:t+3} - \widetilde{\mathbf{x}^b}_{t:t+3}\right\|^2_{\tilde{\mathbf{B}}^{-1}} \\ &+ \|\tilde{\mathbf{y}}_{t:t+3} - \mathcal{H}_o(\tilde{\mathbf{x}}_{t:t+3})\|^2_{\mathbf{R}^{-1}} \\ &+ \|\tilde{\mathbf{y}}_{t+4:t+7} - \mathcal{H}_o(\mathcal{M}(\tilde{\mathbf{x}}_{t:t+3}))\|^2_{\mathbf{R}^{-1}}\end{aligned} \tag{4}$$

where $\tilde{\mathbf{y}}_t$ represents the observation in vector space at the t_{th} time step, $\mathcal{H}_o$ represents the observation operator, $\tilde{\mathbf{B}}$ represents the background error covariance matrix in the latent space,

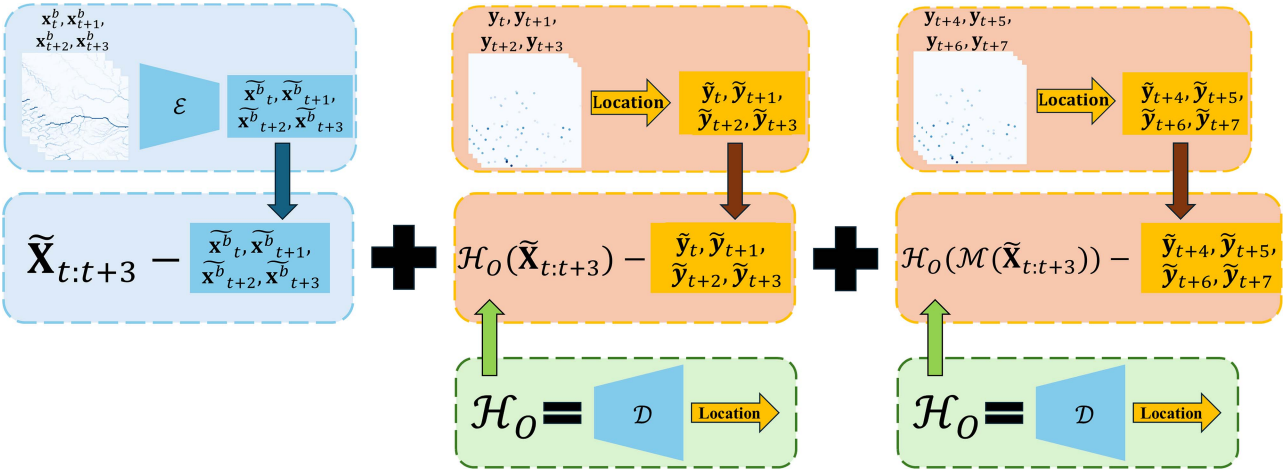


Fig. 3. Structure of the AI-empowered latent 4D-Var. The input consists of the compressed state field for four time steps in latent space and the observations for eight time steps. The blue block represents the background component, while the orange blocks represent the observation component. The green block represents the observation operator, which is composed of a decoder followed by the observation location matrix. This setup integrates both the state field and observation data to improve the prediction model's performance.

and $\mathbf{R}$ represents the observation error covariance matrix. The observation operator $\mathcal{H}_o$ consists of two parts: the decoder $\mathcal{D}$ and the observation location matrix $\mathcal{L}$. The observation location matrix $\mathcal{L}$ selects the location with observations in full space $\mathbf{y}_t$ and arranges them into a vector, resulting in the observations in vector space $\tilde{\mathbf{y}}_t$. The background error covariance matrix is defined as the error covariance matrix of the background field in the latent space, while the observation error covariance matrix is established as the self-covariance matrix of the observations in vector space, where the diagonal elements represent the variances of the corresponding elements, indicating that each observation is independent [57]. Since the AI-empowered latent 4D-Var utilizes four time steps of the background field and eight time steps of observations, performing DA every four time steps, the analysis state obtained from the assimilation is used as input for the subsequent forecasting. Consequently, the size of $\tilde{\mathbf{B}}$ is 512×512, and the size of $\mathbf{R}$ is 232×232. If 4D-Var were performed in full space, then the size of the background error covariance matrix $\mathbf{B}$ would be 65536×65536. Consequently, minimizing the cost function in this scenario would be significantly more time-consuming compared to the AI-empowered latent 4D-Var.

AI-empowered latent 4D-Var employs PyTorch to compute the derivatives of the cost function directly, facilitating its minimization without relying on a tangent linear model or an adjoint model, which are required in traditional 4D-Var methods. Furthermore, PyTorch's capability to utilize graphics processing units (GPUs) enhances computational efficiency significantly. In this implementation, the AdamW algorithm [58] is utilized to solve the cost function for the AI-empowered latent 4D-Var. The detailed algorithm of AI-empowered latent 4D-Var is presented in Algorithm 1, where t represents the number of time steps, k represents the number of the iterations, $\tilde{\mathbf{x}}_t^a$ represents the analysis field in the latent space at the t_{th} time step, and $\mathbf{x}_t^a$ represents the analysis field in the full space at the same time step.

Algorithm 1: AI-empowered latent 4D-Var.

Inputs: $\mathbf{x}^b_{t-3:t}, \mathbf{y}_{t-3:t+4}, \mathcal{E}, \mathcal{D}, \mathcal{L}, \mathcal{H}_o, \mathcal{M}$
Parameters: $t, k, \tilde{\mathbf{B}}, \mathbf{R}$
$k = 0, \tilde{\mathbf{x}^b}_{t-3:t} = \mathcal{E}(\mathbf{x}^b_{t-3:t}), \tilde{\mathbf{y}}_{t-3:t+4} = \mathcal{L}(\mathbf{y}_{t-3:t+4})$
if $\frac{t}{4} == 0$ and $t > 0$ **then**
  while $k < k_{max}$ **do**
    $\mathcal{J}(\tilde{\mathbf{x}}_{t-3:t}) = \|\tilde{\mathbf{x}}_{t-3:t} - \tilde{\mathbf{x}^b}_{t-3:t}\|_{\tilde{\mathbf{B}}^{-1}}$
    $+\|\tilde{\mathbf{y}}_{t-3:t} - \mathcal{H}_o(\tilde{\mathbf{x}}_{t-3:t})\|_{\mathbf{R}^{-1}}$
    $+\|\tilde{\mathbf{y}}_{t+1:t+4} - \mathcal{H}_o(\mathcal{M}(\tilde{\mathbf{x}}_{t-3:t}))\|_{\mathbf{R}^{-1}}$
    $\tilde{\mathbf{x}}^a_{t-3:t} = \text{argmin}(J(\tilde{\mathbf{x}}_{t-3:t}))$
    $k = k + 1$
  end while
end if
$\tilde{\mathbf{x}^b}_{t+1:t+4} = \mathcal{M}(\tilde{\mathbf{x}^a}_{t-3:t})$
$\mathbf{x}^a_{t-3:t} = \mathcal{D}(\tilde{\mathbf{x}^a}_{t-3:t})$
$t = t + 4$
Output: Analysis field of the last time period: $\mathbf{x}^a_{t-3:t}$ and background field of the current time period: $\tilde{\mathbf{x}^b}_{t+1:t+4}$.

After applying 4D-Var with eight time steps of observations and four time steps of the background field, the analysis field for the current four time steps is computed.

D. Baseline Methods

To evaluate the performance of AI-empowered latent 4D-Var, two baseline models are developed for comparison, allowing for a comprehensive assessment of the proposed method's effectiveness.

1) Voronoi-Based LSTM: Voronoi tessellation, which utilizes NNs to reconstruct fields from sparse observations, is commonly employed in scenarios with limited data [59]. The first baseline model, known as Voronoi-based LSTM, employs

Voronoi tessellation to reconstruct the state field from observations at the time step of DA. This reconstructed state field is then compressed using an encoder, and the compressed representation is utilized as the input for the LSTM to make predictions. In this experiment, U-Net is chosen as the backbone architecture for the Voronoi reconstruction network design [60]. The U-Net structure excels in multiscale feature extraction, retains critical information, and learns representations efficiently, which enhances its effectiveness compared to traditional CNN architectures proposed by Fukami et al.[59]. Consequently, this allows for improved learning of the state field features in this experiment. In training the Voronoi reconstruction network, the state field is set to zero at locations without observations and then processed with Voronoi interpolation to construct an interpolated field, which serves as the input to the network. The network's output is the reconstructed state field. This model leverages the spatial interpolation from observed points to generate a more complete state field. The model parameters are trained by minimizing the distance between the interpolated field (after processing) and the original state field. This objective ensures that the model learns to accurately reconstruct the state field by capturing spatial relationships based on available observations. Therefore, this model can use the observation directly as the input, allowing it to reconstruct the Voronoi-interpolated field through the NN. In implementing the Voronoi-based LSTM, the reconstructed field from the observation is used as the LSTM input, and the predicted field is compared with the ground truth to evaluate this baseline's performance.

2) AI-Empowered Latent Three-dimensional Variational Data Assimilation (3D-Var): The second baseline model is the AI-empowered latent 3D-Var. Unlike the AI-empowered latent 4D-Var, which incorporates the temporal distribution of observations, the AI-empowered latent 3D-Var focuses solely on spatial information. However, this model offers a lower computational cost compared to the AI-empowered latent 4D-Var.

In the AI-empowered latent 3D-Var, the four time steps of observations are used to assimilate the background field at the same assimilation time. The cost function of the AI-empowered latent 3D-Var is expressed in the formula as shown in the following:

$$\mathcal{J}(\tilde{\mathbf{x}}_{t:t+3}) = \left\|\tilde{\mathbf{x}}_{t:t+3} - \tilde{\mathbf{x}^b}_{t:t+3}\right\|^2_{\tilde{\mathbf{B}}^{-1}} + \|\tilde{\mathbf{y}}_{t:t+3} - \mathcal{H}_o(\tilde{\mathbf{x}}_{t:t+3})\|^2_{\mathbf{R}^{-1}}. \quad (5)$$

In AI-empowered latent 3D-Var, the configurations for $\tilde{\mathbf{B}}$, $\mathbf{R}$, and the observation operator $\mathcal{H}_o$ remain consistent with those used in the AI-empowered latent 4D-Var.

E. Implementation Details

In this experiment, the NN models are trained using mean square error (MSE) as the loss function. The Adam optimizer is utilized with default settings, employing a learning rate of 1e-3, and a batch size of 32 is implemented throughout the training process. The training process is conducted on an A100 40 GB GPU server operating under Ubuntu 22.04.5.

IV. Numerical Results and Discussions

In this section, five evaluation metrics are employed to evaluate the performance of the proposed AI-empowered latent 4D-Var, comparing it with the EFAS historical simulation and two baseline models. In addition, the flood event that occurred in the study area in June 2013 is selected as a case study for further analysis.

A. Numerical Results

In this experiment, data from 2013 to 2017 is selected as the test set to assess the performance of the proposed AI-empowered latent 4D-Var and the DA methods are implemented using the TorchDA package [61]. Five evaluation metrics—Mean Square Error (MSE), relative L_2 norm, R^2, standard deviation (STD) and Kling–Gupta efficiency (KGE)—are employed to compare performance against the EFAS historical simulation and two baseline models, highlighting the advantages of the proposed method. Ground truth data are sourced from observations in the Lamah-CE dataset. The predicted state field results are compared with the ground truth values at the corresponding observation locations and time steps, using normalized river discharge as the physical quantity. The first four evaluation metrics are averaged monthly, with each month consisting of 120 data points and the Kling-Gupta Efficiency (KGE) is calculated by averaging the results from each observation location across the entire test set to assess the performance of each method. The aggregated results of the first four evaluation metrics are presented in Fig. 4. In this figure, different colors represent the model results in comparison to the ground truth: black for EFAS historical simulation, red for LSTM, orange for Voronoi-based LSTM, blue for AI-empowered latent 3D-Var, and green for AI-empowered latent 4D-Var.

Evaluating the four metrics collectively, AI-empowered latent 4D-Var demonstrates the strongest performance. In contrast, the results from the other methods, including EFAS historical simulation, AI-empowered latent 3D-Var, and Voronoi-based LSTM, fall short of the performance exhibited by the proposed method. Compared to EFAS historical simulation, AI-empowered latent 4D-Var incorporates observations to further enhance the accuracy of predictions. In addition, unlike pure machine learning methods, such as LSTM and Voronoi-based LSTM, DA methods integrate physical constraints, which enhances forecasting accuracy and improves the stability of trend predictions. Notably, AI-empowered latent 4D-Var surpasses AI-empowered latent 3D-Var, largely due to its ability to account for the temporal distribution of observations and the forward model, both of which are crucial for improving time series forecasting accuracy.

The first metric analyzed is the MSE, which measures the difference between the estimator and the estimated quantity. It calculates the distance between the model's predicted values and the ground truth, with a larger distance indicating greater forecasting error. As shown in Fig. 4(a), the MSE values for AI-empowered latent 4D-Var are relatively stable and consistently below 0.002, reflecting a low error rate. This performance is the best among all of the methods tested. AI-empowered

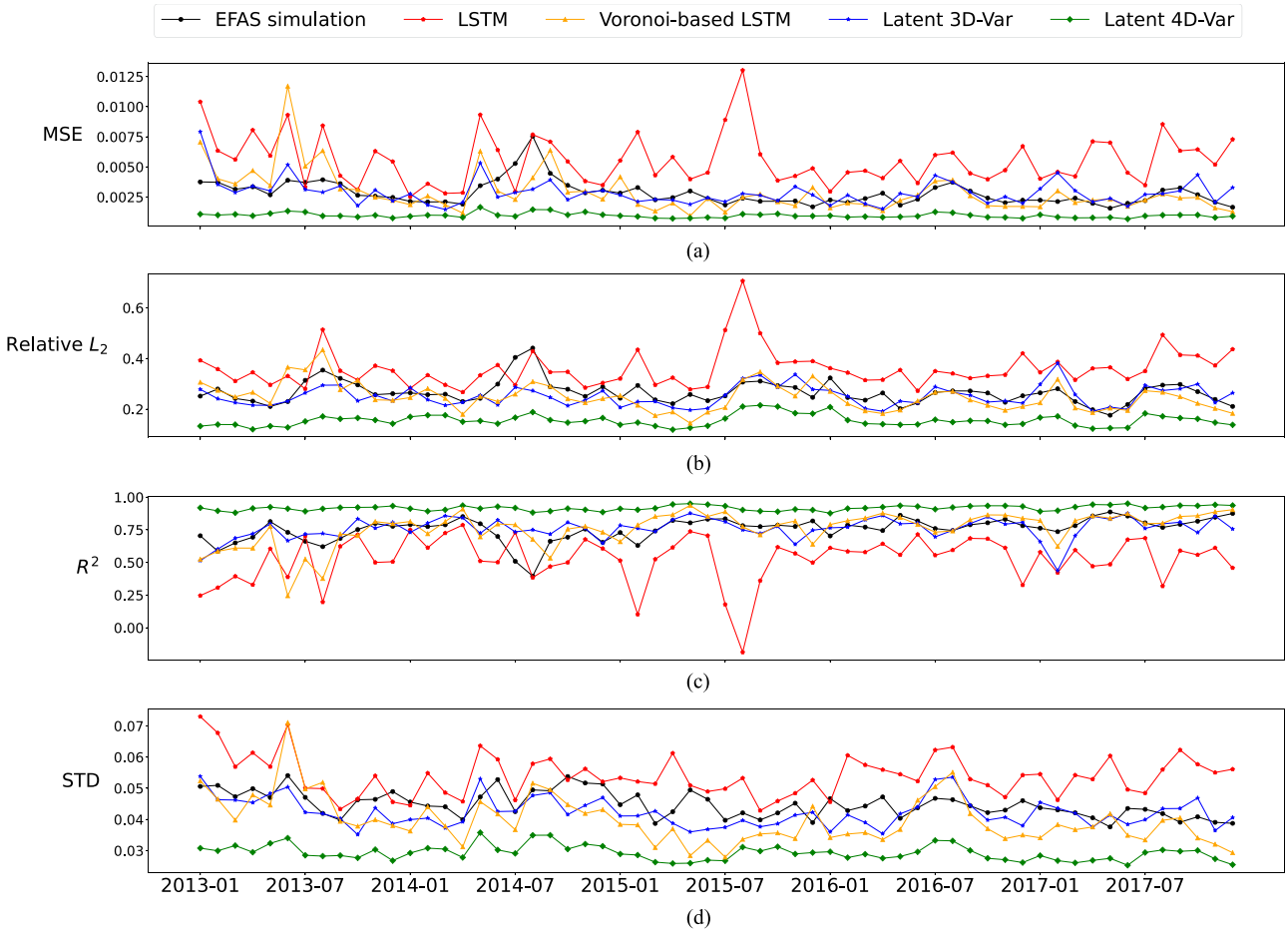


Fig. 4. Results on the four evaluation metrics (MSE, Relative L_2, R^2 and STD) of five methods for the test set (2013–2017), with observation serving as the ground truth. The colors represent each method: black for EFAS historical simulation, red for LSTM, orange for Voronoi-based LSTM, blue for AI-empowered latent 3D-Var, and green for AI-empowered latent 4D-Var.

latent 3D-Var exhibits performance similar to that of the EFAS historical simulation and Voronoi-based LSTM. However, due to previously mentioned limitations in these methods, their MSE performance is lower than that of AI-empowered latent 4D-Var. Consequently, by integrating physical constraints and considering the temporal distribution of observations along with the forward model, the AI-empowered latent 4D-Var method both corrects LSTM forecasting errors significantly and enhances overall accuracy and stability.

The relative L_2 norm serves as an evaluation metric similar to MSE but accounts for error relative to the ground truth. This enables it to measure the magnitude of forecasting errors in proportion to the true values, making it especially useful for comparing datasets of different scales. Consequently, the relative L_2 norm effectively reflects the model's performance under significant data variability, making it particularly valuable for assessing our method's accuracy during seasonal variations in river discharge magnitudes. As seen in Fig. 4(b), the AI-empowered latent 4D-Var results are all less than 0.2, so there is a good stability in AI-empowered latent 4D-Var on all seasons with good performance. In contrast, other methods can exhibit considerable fluctuations due to seasonal variations in river discharge magnitudes. Therefore, AI-empowered latent 4D-Var can generate more stable predictions across these variations by incorporating the time dimension.

R^2 is an evaluation metric that indicates how well the model fits the data. A value close to 1 signifies a better fit, meaning that the model explains a significant portion of the variability in the data. In this experiment, R^2 is used to assess the correlation between the model's predictions and the ground truth. As shown in Fig. 4(c), the R^2 value for AI-empowered latent 4D-Var consistently remains close to 1, with minimal fluctuations compared to AI-empowered latent 3D-Var. This indicates a strong correlation between the results of AI-empowered latent 4D-Var and the ground truth, demonstrating that the predictions generated by our proposed method align closely with actual river discharge observations. The fluctuations observed in the values of pure machine learning methods, compared to DA methods, arise from the absence of actual physical meaning in these models. Without physical constraints, the predicted values often fail to align with meaningful real-world values. This underscores the advantage of AI-empowered latent 4D-Var, which effectively incorporates physical constraints, bringing predictions more in line with real-world laws.

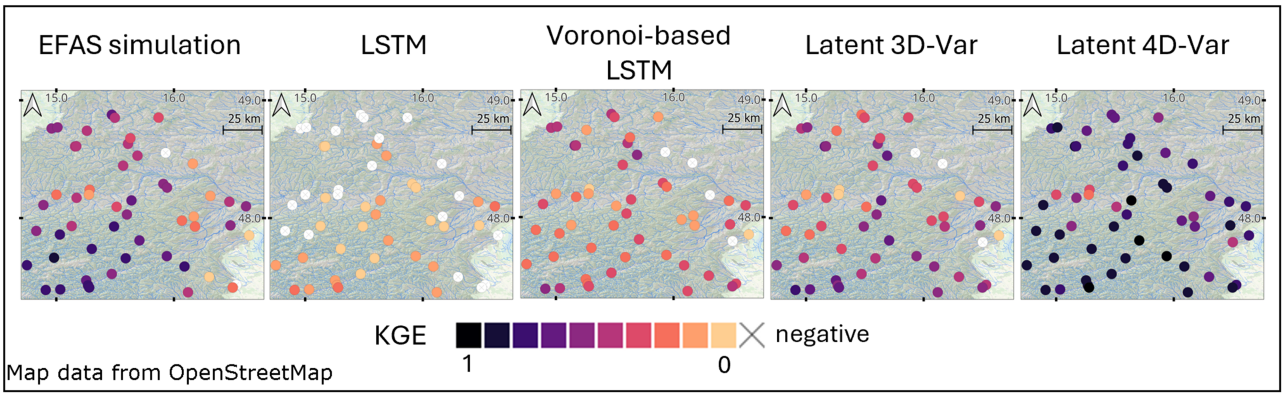


Fig. 5. Maps showing the KGE metric for each model, computed for the whole testing period over the study area. From left to right, it reported the EFAS historical simulation, LSTM, Voronoi-based LSTM, AI-empowered latent 3D-Var, and AI-empowered latent 4D-Var. Darker colors represent better performance. Negative values are marked in white. The background map is from OpenStreetMap.

Standard Deviation (STD) is a statistical measure that quantifies the dispersion of data points in a dataset relative to the mean. A smaller STD indicates that the predicted fluctuations are lower than those of the true values, suggesting that the model more effectively captures the characteristics of the actual data. As shown in Fig. 4(d), the AI-empowered latent 4D-Var demonstrates greater stability and consistency in forecasting compared to other methods. This resilience to seasonal variations can be attributed to AI-empowered latent 4D-Var's consideration of the temporal distribution of observations and the effects of forward models.

The KGE [62] is a hydrological metric used to assess the performance of a model in simulating observed flow, capable of capturing multiple performance dimensions, and it is defined as

$$\mathrm{KGE} = 1 - \sqrt{\left(\frac{\sigma_s}{\sigma_o} - 1\right)^2 + \left(\frac{\mu_s}{\mu_o} - 1\right)^2 + (\rho_{s,o} - 1)^2} \quad (6)$$

where σ_s and σ_o represent the STD of the simulated and observed data, respectively, μ_s and μ_o denote the means of the simulated and observed data, respectively, and $\rho_{s,o}$ is the correlation coefficient between the simulated and observed data.

KGE values closer to 1 indicate better fit, with optimal correlation, bias, and variability representation by the model with respect to the observations. Values between 0,6 and 0,9 are generally considered representative of good modeling performances for hydrological applications. Values lower than 0,5 might be still acceptable, depending on the specific cases, but usually are indices of discrepancies among the model and the observations. In this experiment, KGE is used to assess the agreement between the considered model's outcomes and the LamaH-CE observations, and the results are shown in Fig. 5. The higher KGE in AI-empowered latent 4D-Var underscores the advantage of integrating DA, which strengthens both correlation and physical relevance by aligning model outputs more closely with observed hydrologic patterns. Negative values of KGE indicate that the model is less descriptive than the mean value of the dataset. There are few negative KGE values, mostly coming from the LSTM forecasting. Voronoi-based LSTM and AI-empowered latent 3D-Var show only four and three negative values, respectively, but the overall performance is largely better than LSTM. No negative values are resulting from AI-empowered latent 4D-Var.

TABLE IV
COMPARISON OF THE RESULTS OF THE FIVE EVALUATION INDICATORS ACROSS THE TEST DATASET

	EFAS simulation	LSTM	Voronoi-based LSTM	Latent 3D-Var	Latent 4D-Var
MSE ↓	0.00279	0.00556	0.00289	0.00285	0.00096
Relative L_2↓	0.26585	0.35939	0.24780	0.25120	0.15495
R2 ↑	0.75563	0.52654	0.76050	0.75936	0.91683
STD ↓	0.04479	0.05397	0.03981	0.04228	0.02929
KGE ↑	0.81432	0.67678	0.79239	0.81370	0.93816

To comprehensively evaluate the AI-empowered latent 4D-Var, we calculate the average results of these five metrics over the entire test set, as presented in Table IV. In the table, the numbers marked in red highlight the best-performing methods for each indicator. As demonstrated, AI-empowered latent 4D-Var outperforms all other methods across all indicators, indicating its superiority in terms of accuracy, stability, and correlation. On the contrary, the worst performer is LSTM, primarily because it is a pure machine learning model that does not consider physical constraints, leading to significant bias when compared to real observations. In addition, LSTM fails to incorporate supplementary observations that could provide valuable information, further contributing to its poor performance.

The improvement of Voronoi-based LSTM over LSTM can be attributed to Voronoi-based LSTM's use of observations as additional information. However, as a pure machine learning method, Voronoi-based LSTM does not incorporate physical constraints, resulting in only a slight advantage over LSTM and still falling short of AI-empowered latent 4D-Var.

Moreover, the results of EFAS historical simulation are generated through hydrological modeling without incorporating observations from gauge stations into the predictions. As a result, the performance is not as robust as that of AI-empowered latent 4D-Var.

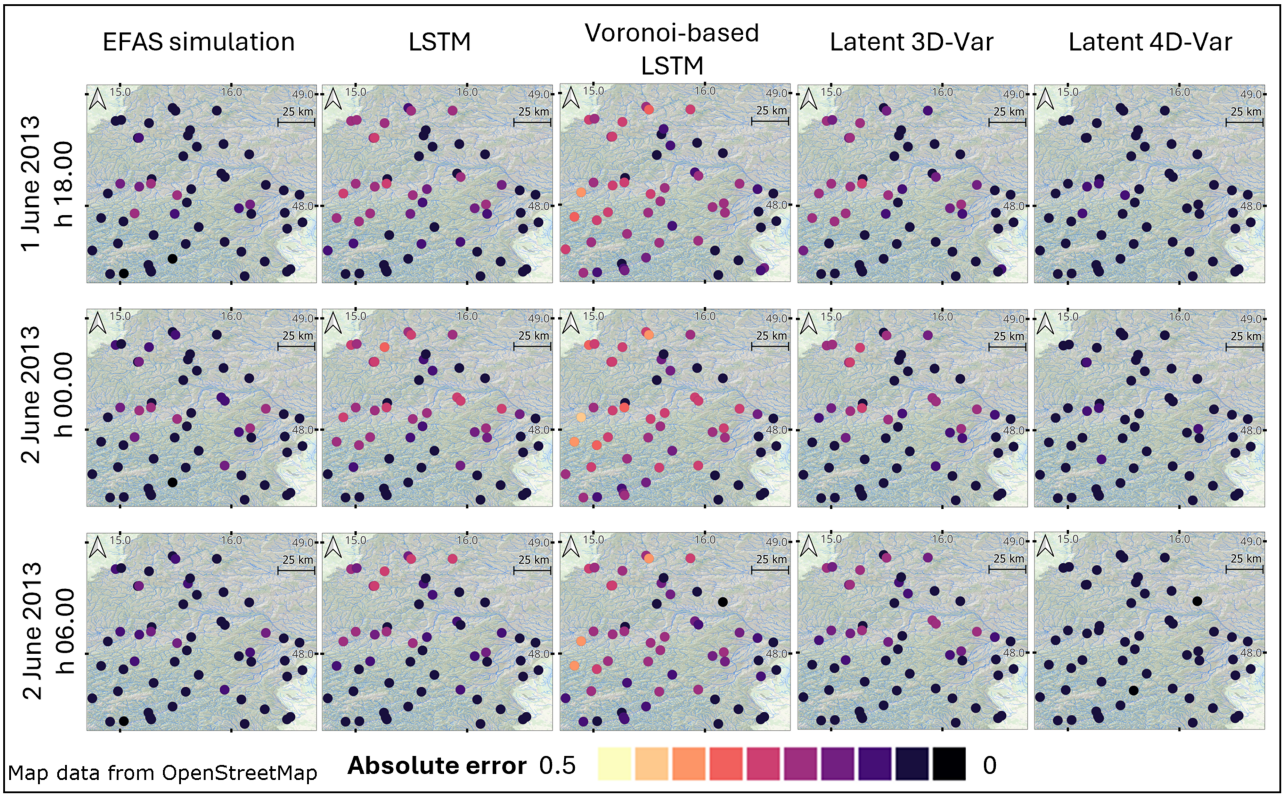


Fig. 6. Map of absolute error relative to EFAS historical simulation, LSTM, Voronoi-based LSTM, AI-empowered latent 3D-Var, and AI-empowered latent 4D-Var across the entire testing area. The map includes three time steps of the flood event: June 1, 2013, at 18:00; June 2, 2013, at 00:00; June 2, 2013, at 06:00. The maps are arranged in a matrix format, with the five models represented as columns and the three time steps as rows. Absolute error is displayed at each gauge station, with darker colors indicating lower error values. The background map is from OpenStreetMap.

On the other hand, AI-empowered latent 3D-Var, as a DA method, yields better results than pure machine learning methods and EFAS historical simulation. However, it does not account for the temporal distribution of observations or the forward model, which explains why it does not perform as well as AI-empowered latent 4D-Var.

In numerical comparisons, the AI-empowered latent 4D-Var significantly outperforms other methods, demonstrating improvements of 65.59% to 82.73% in MSE, 37.47% to 56.89% in the relative L_2 norm, and 26.42% to 45.73% in STD. It also achieves better performance in R^2 by 20.56% to 74.12% and in KGE by 15.20% to 38.62% . These results indicate superior predictive accuracy and variability explanation compared to other methods.

In terms of execution time, the AI-empowered latent 4D-Var completes one-day forecasting in about 100 s on an A100 40 GB GPU.

B. Test Case

We analyze the model outcomes during a flood event that occurred within the time frame used for the testing (2013–2017). We focus on the June 2013 flood, which was caused by an atmospheric condition typical of flood events in the upper Danube [63], characterized by persistent rainfall and a stalling low atmospheric pressure. Moreover, the soil moisture levels registered at the end of May 2013 exceeded all the historical records [64], and ground waters levels were particularly high. All these factors led to high runoff values and complex patterns of high river flows in the area of the upper Danube basin [63].

In Fig. 6, we show the map of absolute error relative to EFAS historical simulation, LSTM, Voronoi-based LSTM, AI-empowered latent 3D-Var, and AI-empowered latent 4D-Var, for the all testing areas. We reported the three initial time steps of the flood event: the 1 June 2013 at 18.00, 2 June 2013 at 00.00 (flood-wave peak), and the 2 June 2013, at 06.00. Maps are presented in the shape of a matrix, which has the five tested models as columns and the three time steps as rows. Absolute error is reported for each gauge station, with darker colours for lower error values.

Overall, the western and some central gauges seem to be characterized by slightly lower performances with respect to the eastern ones, across all methods. Peculiar characteristics of those watersheds, and the complex runoff pattern associated to the analyzed flood event, can partially explain this; moreover, some of those values are directly influenced by hydrological connections not included in the study area. In all the models, the first and the second time steps show slightly higher error values with respect to the first, because the distance between the predictions and the observations generally decreased after the flood-wave peak, which occurred at the second time step. The AI-empowered latent 4D-Var, reported in the last column,

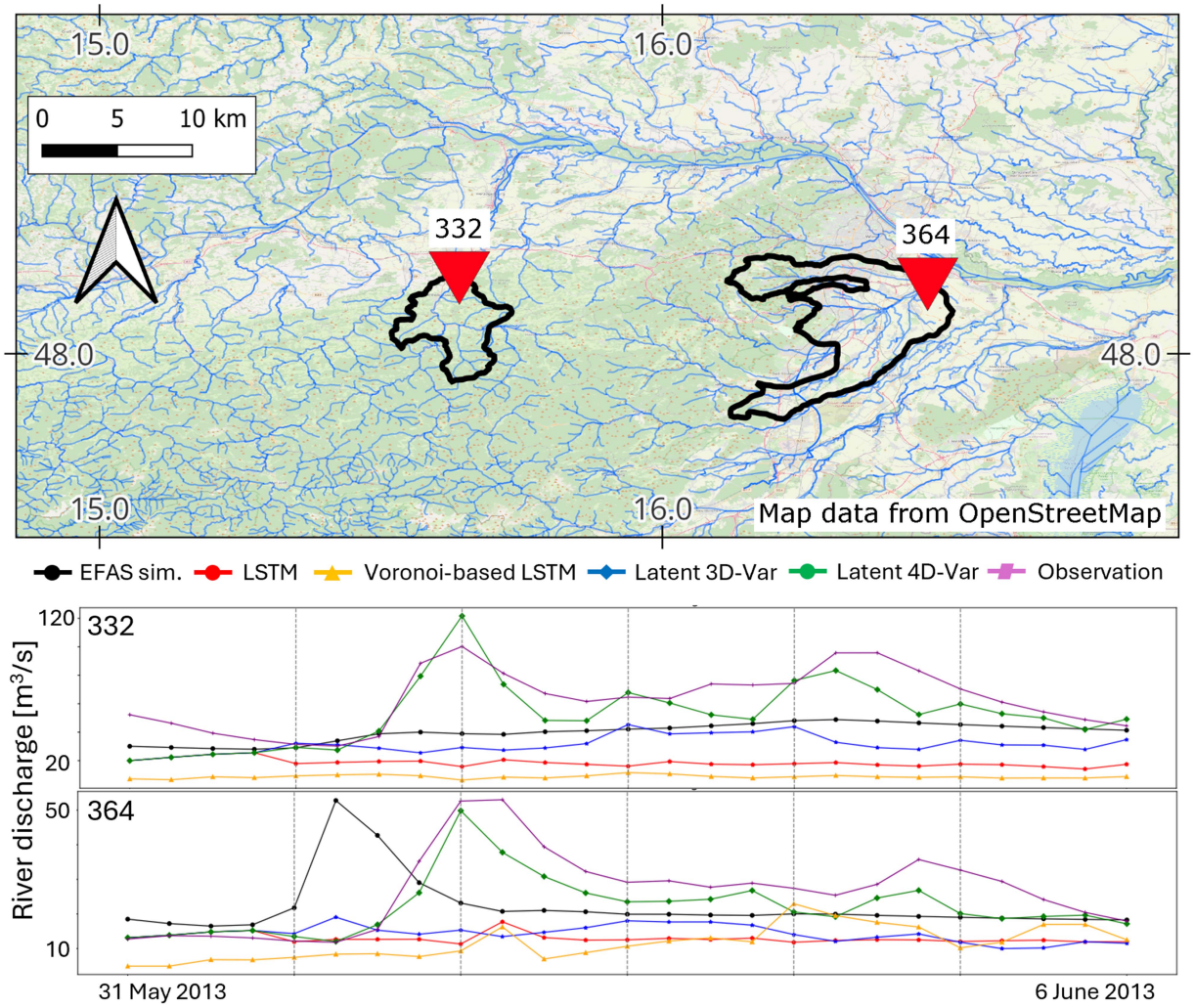


Fig. 7. Location of gauge stations, and the forecasted river discharge for each model (EFAS historical simulation, LSTM, Voronoi-based LSTM, AI-empowered latent 3D-Var, and AI-empowered latent 4D-Var) compared to observed values. Watersheds and river network shown in the map are from LamaH-CE. The background map is from OpenStreetMap.

outperforms all the other models in almost all the gauges and for each time step, overall providing better predictions of river discharge for the analyzed area. The Voronoi-based LSTM have lower forecasting performances with respect to all the other methods: Voronoi correlation may incorrectly catch the complex hydrological connections among the watersheds, thus negatively affecting the predictions. LSTM and AI-empowered latent 3D-Var show similar forecasting capabilities, although AI-empowered latent 3D-Var improved the LSTM results for almost all the gauges.

For a more detailed analysis, we focus on gauge 332 (from the "ID" field of the LamaH-CE classification) on the Traisen River, and Gauge 364 on the Schwechat River. Both rivers are tributaries of the Danube in Lower Austria. Even if they were not particularly hit by this flood event, they are exemplary for comparing models' performance. These rivers can be considered midsize in the LamaH-CE dataset, and have similar characteristics in terms of watershed area and length: the Traisen drains a 915 km^2 basin with a length of 82.6 km, while the Schwechat drains an area of 1182 km^2 and has a length of 62 km. In Fig. 7, the locations of the gauges are shown, and below it, the predicted river discharge is reported for all the models against the observed one.

For both gauges, the AI-empowered latent 4D-Var shows a significant improvement with respect to the other methodologies, although a residual difference remains between the observed river discharge and the predictions. Also, improvements are visible in the shape of the flood wave, which is better captured by the AI-empowered latent 4D-Var, despite some differences in the slopes, which mostly regard the falling limb, and the successive steps of the flood. The other methods seem to not sense the flood wave at all. For the Traisen River, forecasting from AI-empowered latent 4D-Var would have been the only ones to likely raise a flood alert, making a critical difference for public safety, civil protection efforts, and the implementation of flood-risk mitigation measures. EFAS predictions for the Schwechat river sensed the flood-wave but with incorrect timing, and the predicted wave limb fall immediately before the observed flood-wave rising limb, thus not sensing the actual flood event and delivering predictions opposite to the observed trend,

which is extremely dangerous in terms of flood risk mitigation. AI-empowered latent 4D-Var is again the best model in sensing the flood.

V. Conclusion and Future Work

This article introduces AI-empowered latent 4D-Var, a method that significantly improves the speed and accuracy of river discharge forecasting. The method first applies an NN-based compression model to compress the river discharge state field into a latent space, followed by an NN-based surrogate model that replaces the traditional hydrological model, substantially accelerating forward model predictions. In addition, the 4D-Var process is computed in the latent space, further enhancing computational efficiency. Unlike traditional 4D-Var, which requires tangent linear and adjoint models to compute the cost function, this method leverages NNs to directly solve the problem without approximation, greatly reducing computational demands and increasing processing speed. The NN models are trained on data from 1992 to 2012 and tested on data from 2013 to 2017. The numerical results demonstrate a significant improvement in accuracy, stability, and correlation over both the EFAS historical simulation dataset and our two baseline models—Voronoi-based LSTM and AI-empowered latent 3D-Var—indicating a substantial enhancement in predictive performance. Moreover, our method completes next day forecasting (four time steps) in only 500 iterations, or roughly 100 s. This rapid processing time is highly efficient and well-suited for timely flood forecasting. We use the June 2013 flood event in the study area as a test case to evaluate the proposed method. By comparing predictions to those from the EFAS historical simulation and the two baseline models, our method not only forecasts the occurrence of the flood on time, but also provides more accurate results. Therefore, the results indicate that the method is highly effective for forecasting flooding events in practical applications. In addition, compared to existing methods, our approach requires lower computational resources while achieving higher accuracy, making it more suitable for real-world applications. Future work will focus on incorporating additional variables relevant to river discharge, such as rainfall and soil moisture, into the forward model to enhance forecasting accuracy.

Moreover, gauge stations in our study are strategically located at watershed outlets delineated in the LamaH-CE dataset [53] and effectively represent the hydrology of these areas. We leveraged reliable datasets from data-rich regions to build and test our model but acknowledge the importance of extending it to data-scarce areas. But working with shorter time series and sparser gauge stations, potentially with record gaps, may lower the model's performance and increase uncertainties. These issues can be mitigated using transfer learning, where models pretrained in data-rich contexts are fine-tuned for data-scarce regions. This task is a part of our ongoing work, and results will be presented in a forthcoming publication. Furthermore, during the DA process, the concepts presented in [51] could be utilized to introduce a Voronoi-related cost function, with a newly designed cost function aimed at achieving faster convergence. This improvement could significantly enhance the effectiveness of our forecasting.

References

[1] *CRED*, "2023: Disasters in numbers," Brussels, 2024. Accessed: Nov. 12, 2024. [Online]. Available: https://files.emdat.be/reports/2023_EMDAT_report.pdf

[2] L. Alfieri et al., "Global projections of river flood risk in a warmer world," *Earth's Future*, vol. 5, no. 2, pp. 171–182, 2017.

[3] *UNDRR*, "GAR 2022: Our world at risk," 2022. [Online]. Available: https://www.undrr.org/gar/gar2022-our-world-risk-gar

[4] *U. A. C. of Engineers*, "HEC-HMS," (n.d.). [Online]. Available: https://www.hec.usace.army.mil/software/hec-hms/

[5] *multiple*, "Soil & water assessment tool," (n.d.). [Online]. Available: https://swat.tamu.edu/

[6] J. Van Der Knijff, J. Younis, and A. De Roo, "LISFLOOD: A GIS-based distributed model for river basin scale water balance and flood simulation," *Int. J. Geographical Inf. Sci.*, vol. 24, no. 2, pp. 189–212, 2010.

[7] P. Smith et al., "On the operational implementation of the European flood awareness system (EFAS)," in *Flood Forecasting*. Amsterdam, The Netherlands: Elsevier, 2016, pp. 313–348.

[8] L. Alfieri et al., "GloFAS–Global ensemble streamflow forecasting and flood early warning," *Hydrol. Earth Syst. Sci.*, vol. 17, no. 3, pp. 1161–1175, 2013.

[9] H. G. Damavandi et al., "Accurate prediction of streamflow using long short-term memory network: A case study in the Brazos river basin in Texas," *Int. J. Environ. Sci. Develop.*, vol. 10, no. 10, pp. 294–300, 2019.

[10] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using long short-term memory (LSTM) networks," *Hydrol. Earth Syst. Sci.*, vol. 22, no. 11, pp. 6005–6022, 2018.

[11] F. Kratzert, D. Klotz, M. Herrnegger, A. K. Sampson, S. Hochreiter, and G. S. Nearing, "Toward improved predictions in ungauged basins: Exploiting the power of machine learning," *Water Resour. Res.*, vol. 55, no. 12, pp. 11344–11354, 2019.

[12] W. Xu et al., "Coupling deep learning and physically based hydrological models for monthly streamflow predictions," *Water Resour. Res.*, vol. 60, no. 2, 2024, Art. no. e2023WR035618.

[13] K. Bevan, "Facets of uncertainty: Epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication," *Hydrological Sci. J.*, vol. 61, no. 9, pp. 1652–1665, 2016, doi: 10.1080/02626667.2015.1031761G.B.

[14] G. Blöschl, "Twenty-three unsolved problems in hydrology (UPH)–A community perspective," *Hydrological Sci. J.*, vol. 64, no. 10, pp. 1141–1158, 2019, doi: 10.1080/02626667.2019.1620507.

[15] *UNDRR*, "GAR special report 2024 forensic insights for future resilience learning from past disasters," 2024. [Online]. Available: https://www.undrr.org/gar

[16] F. Pappenberger, H. L. Cloke, D. J. Parker, F. Wetterhall, D. S. Richardson, and J. Thielen, "The monetary benefit of early flood warnings in Europe," *Environ. Sci. Policy*, vol. 51, pp. 278–291, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1462901115000891

[17] K. Jafarzadegan, P. Abbaszadeh, and H. Moradkhani, "Sequential data assimilation for real-time probabilistic flood inundation mapping," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 9, pp. 4995–5011, 2021.

[18] T. H. Nguyen et al., "Improvement of flood extent representation with remote sensing data and data assimilation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4206022.

[19] J. Oh and M. Bartos, "Model predictive control of stormwater basins coupled with real-time data assimilation enhances flood and pollution control under uncertainty," *Water Res.*, vol. 235, 2023, Art. no. 119825.

[20] S. Cheng et al., "Machine learning for modelling unstructured grid data in computational physics: A review," *Inf. Fusion*, vol. 123, 2025, Art. no. 103255.

[21] E. Belanger and A. Vincent, "Data assimilation (4D-VAR) to forecast flood in shallow-waters with sediment erosion," *J. Hydrol.*, vol. 300, no. 1/4, pp. 114–125, 2005.

[22] V. Mazzarella, I. Maiello, V. Capozzi, G. Budillon, and R. Ferretti, "Comparison between 3D-VAR and 4D-VAR data assimilation methods for the simulation of a heavy rainfall case in central Italy," *Adv. Sci. Res.*, vol. 14, pp. 271–278, 2017.

[23] X. Lai, Q. Liang, H. Yesou, and S. Daillet, "Variational assimilation of remotely sensed flood extents using a 2-D flood model," *Hydrol. Earth Syst. Sci.*, vol. 18, no. 11, pp. 4325–4339, 2014.

[24] F. Rabier and Z. Liu, "Variational data assimilation: Theory and overview," in *Proc. ECMWF Seminar Recent Develop. Data Assimilation Atmos. Ocean*, Reading, U.K., 2003, pp. 29–43.

[25] S. Hatfield, M. Chantry, P. Dueben, P. Lopez, A. Geer, and T. Palmer, "Building tangent-linear and adjoint models for data assimilation with neural networks," *J. Adv. Model. Earth Syst.*, vol. 13, no. 9, 2021, Art. no. e2021MS002521.
[26] R. Dong, H. Leng, J. Zhao, J. Song, and S. Liang, "A framework for four-dimensional variational data assimilation based on machine learning," *Entropy*, vol. 24, no. 2, p. 264, 2022, Art. no. 264.
[27] J. Lever et al., "Facing & mitigating common challenges when working with real-world data: The data learning paradigm," *J. Comput. Sci.*, vol. 85, 2025, Art. no. 102523.
[28] M. Beauchamp, Q. Febvre, H. Georgenthum, and R. Fablet, "4DVarNet-SSH: End-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry," *Geoscientific Model Develop.*, vol. 16, no. 8, pp. 2119–2147, 2023.
[29] R. Fablet, Q. Febvre, and B. Chapron, "Multimodal 4DVarNets for the reconstruction of sea surface dynamics from SST-SSH synergies," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4204214.
[30] C. Wu, K. W. Chau, and Y. S. Li, "River stage prediction based on a distributed support vector regression," *J. Hydrol.*, vol. 358, no. 1/2, pp. 96–111, 2008.
[31] J. C. Refshaard and B. Storm, "Mike she," 1995.
[32] H. Gu, Y.-P. Xu, D. Ma, J. Xie, L. Liu, and Z. Bai, "A surrogate model for the variable infiltration capacity model using deep learning artificial neural network," *J. Hydrol.*, vol. 588, 2020, Art. no. 125019.
[33] R. Sun, B. Pan, and Q. Duan, "A surrogate modeling method for distributed land surface hydrological models based on deep learning," *J. Hydrol.*, vol. 624, 2023, Art. no. 129944.
[34] N. Fraehr, Q. J. Wang, W. Wu, and R. Nathan, "Assessment of surrogate models for flood inundation: The physics-guided LSG model vs. state-of-the-art machine learning models," *Water Res.*, vol. 252, 2024, Art. no. 121202.
[35] M. A. A. Mehedi, M. Khosravi, M. M. S. Yazdan, and H. Shabanian, "Exploring temporal dynamics of river discharge using univariate long short-term memory (LSTM) recurrent neural network at east branch of delaware river," *Hydrology*, vol. 9, no. 11, p. 264, 2022, Art. no. 202.
[36] Y. Sudriani, I. Ridwansyah, and H. A Rustini, "Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia," *IOP Conf. Series: Earth Environ. Sci.*, vol. 299, 2019, Art. no. 012037.
[37] Y. Zhang, Z. Gu, J. V. G. Thé, S. X. Yang, and B. Gharabaghi, "The discharge forecasting of multiple monitoring station for Humber river by hybrid LSTM models," *Water*, vol. 14, no. 11, p. 264, 2022, art. no. 1794.
[38] Z. M. Yaseen, S. O. Sulaiman, R. C. Deo, and K.-W. Chau, "An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction," *J. Hydrol.*, vol. 569, pp. 387–408, 2019.
[39] L. Qian, J. Li, C. Liu, J. Tao, and F. Chen, "River flow sequence feature extraction and prediction using an enhanced sparse autoencoder," *J. Hydroinformatics*, vol. 22, no. 5, pp. 1391–1409, 2020.
[40] Y. Zhao et al., "Physics-enhanced machine learning models for stream-flow discharge forecasting," *J. Hydroinformatics*, vol. 26, 2024, Art. no. jh2024061.
[41] R. Arcucci, J. Zhu, S. Hu, and Y.-K. Guo, "Deep data assimilation: Integrating deep learning with data assimilation," *Appl. Sci.*, vol. 11, no. 3, p. 264, 2021, Art. no. 1114.
[42] C. Buizza et al., "Data learning: Integrating data assimilation and machine learning," *J. Comput. Sci.*, vol. 58, 2022, Art. no. 101525.
[43] S. Cheng et al., "Machine learning with data assimilation and uncertainty quantification for dynamical systems: A review," *IEEE/CAA J. Automatica Sinica*, vol. 10, no. 6, pp. 1361–1387, Jun. 2023.
[44] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[45] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, 2010.
[46] S. Cheng, I. C. Prentice, Y. Huang, Y. Jin, Y.-K. Guo, and R. Arcucci, "Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting," *J. Comput. Phys.*, vol. 464, 2022, Art. no. 111302.
[47] M. Amendola et al., "Data assimilation in the latent space of a convolutional autoencoder," in *Proc. Int. Conf. Comput. Sci.* 2021, pp. 373–386.
[48] R. Maulik et al., "Efficient high-dimensional variational data assimilation with machine-learned reduced-order models," *Geoscientific Model Develop.*, vol. 15, no. 8, pp. 3433–3445, 2022.
[49] R. Arcucci, L. Mottet, C. Pain, and Y.-K. Guo, "Optimal reduced space for variational data assimilation," *J. Comput. Phys.*, vol. 379, pp. 51–69, 2019.
[50] H. Wang, H. Zhou, and S. Cheng, "Dynamical system prediction from sparse observations using deep neural networks with Voronoi tessellation and physics constraint," *Comput. Methods Appl. Mechan. Eng.*, vol. 432, 2024, Art. no. 117339.
[51] S. Cheng, C. Liu, Y. Guo, and R. Arcucci, "Efficient deep data assimilation with sparse observations and time-varying sensors," *J. Comput. Phys.*, vol. 496, 2024, Art. no. 112581.
[52] C. Mazzetti et al., "River discharge and related historical data from the European flood awareness system, v5.0," 2023.
[53] C. Klingler, K. Schulz, and M. Herrnegger, "LamaH| Large-sample data for hydrology and environmental sciences for central Europe," *Earth Syst. Sci. Data Discuss.*, vol. 2021, pp. 1–46, 2021.
[54] K. Wang et al., "Latent three-dimensional variational data assimilation with convolutional autoencoder and LSTM for flood forecasting," in *Proc. Int. Conf. Comput. Sci.*, Cham, Switzerland, 2025, pp. 43–56.
[55] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[56] S. Hochreiter, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[57] M. Amendola et al., "Data assimilation in the latent space of a neural network," 2020, *arXiv:2012.12056*.
[58] I. Loshchilov, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
[59] K. Fukami, R. Maulik, N. Ramachandra, K. Fukagata, and K. Taira, "Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning," *Nature Mach. Intell.*, vol. 3, no. 11, pp. 945–951, 2021.
[60] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, Munich, Germany, 2015, pp. 234–241.
[61] S. Cheng, J. Min, C. Liu, and R. Arcucci, "TorchDA: A python package for performing data assimilation with deep learning forward and transformation functions," *Comput. Phys. Commun.*, vol. 306, 2025, Art. no. 109359.
[62] H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling," *J. Hydrol.*, vol. 377, no. 1/2, pp. 80–91, 2009.
[63] G. Blöschl, T. Nester, J. Komma, J. Parajka, and R. A. Perdigão, "The Jun. 2013 flood in the upper Danube basin, and comparisons with the 2002, 1954 and 1899 floods," *Hydrol. Earth Syst. Sci.*, vol. 17, no. 12, pp. 5197–5212, 2013.
[64] K. Schröter, M. Kunz, F. Elmer, B. Mühr, and B. Merz, "What made the Jun. 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation," *Hydrol. Earth Syst. Sci.*, vol. 19, pp. 309–327, 2015, doi: 10.5194/hess-19-309-2015.



Kun Wang received the B.S. degree in electronic information engineering and the M.S. degree in electrical and electronic engineering from The University of Hong Kong, Hong Kong, in 2021 and 2022, respectively. He is currently working toward the Ph.D. degree in flood event modeling and forecasting using machine learning methods and data assimilation techniques with the Resource Geophysics Academy, Department of Earth Science and Engineering, Imperial College London, London, U.K.

His research interests include machine learning and data assimilation for hydrologic applications, with a focus on flood prediction and early-warning systems.



Gabriele Bertoli received the bachelor's degree in civil engineering from Politecnico di Milano, Milan, Italy, in 2019, and the master's degree in geoengineering from the University of Florence, Florence, Italy (110/110 cum Laude), in 2022, where he is currently working toward the Ph.D. degree in river flood forecasting combining hydrology and data science to improve the accuracy and lead time of flood forecasting.

He worked in national and international projects, including CASTLE and AG-WAMED. During his Ph.D., he has collaborated internationally with the Data Science Institute at Imperial College London and the Leichtweiß-Institute for Hydraulic Engineering and Water Resources Management (TU Braunschweig). His research interests include flood risk management and assessment, flood vulnerability, and water resources management.



Sibo Cheng received the Ph.D. degree in advanced data assimilation/error covariance specification in data assimilation from Université Paris-Saclay, Gif-sur-Yvette, France, in 2020.

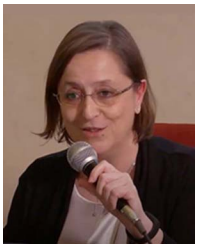
He is currently a Junior Professor (Chaire de Professeur Junior) with CEREA, ENPC, Institut Polytechnique de Paris, Palaiseau, France, and an Honorary Research Fellow with Imperial College London, London, U.K. His research interests include machine learning for dynamical systems, reduced-order modeling, and inverse problems in environmental science and physics, with applications spanning geosciences (including wildfire modeling and air pollution) and fluid dynamics.

Dr. Cheng is a Principal investigator for several national research projects funded by the Agence Nationale de la Recherche, focusing on AI for Earth sciences. He is also the Theme Leader for Data Assimilation within the SAMA Working Group (comprising over 20 permanent researchers) with the Institut Pierre-Simon Laplace, Guyancourt, France.



Kai Schröter received the diploma and the doctoral degree (Dr.-Ing.) in civil engineering from Technische Universität Darmstadt, Darmstadt, Germany, in 2002 and 2008, respectively, and the habilitation degree (Dr. rer. nat. habil) from the University of Potsdam, Potsdam, Germany, in 2020.

He is currently a Full Professor of hydrology and river basin management with the Leichtweiß-Institute for Hydraulic Engineering and Water Resources, Technische Universität Braunschweig, Braunschweig, Germany.



Enrica Caporali received the Laurea cum laude degree in civil engineering, Hydraulic section, in 1988 and the Ph.D. degree in methods and technology for environmental monitoring at University of Florence, in 1994. She is currently a Full Professor of hydraulic and maritime structures and hydrology - CEAR 01/B at UNIFI, Department of Civil and Environmental Engineering (DICEA). She is currently a Member of the Academic Senate of UNIFI, a Vice-President of the School of Engineering, a Coordinator of the international and interdisciplinary master's degree in geoengineering, developed in the framework of the UNESCO Chair on "Prevention and Sustainable Management of Geo-Hydrological Hazards", of which she is a Program Coordinator for Education and Chair Associate, and a Scientific Coordinator with the Spatial Data Laboratory of DICEA. Her research interests include water resources management and flood risk mitigation, also with reference to cultural heritage protection, water-energy-food nexus and water related ecosystem services, environment monitoring and protection, hydrological extreme events, and hydro-sedimentological modeling at river basin scale.

Prof. Caporali was the recipient of Postdoctoral fellowship at UNIFI and at the Massachusetts Institute of Technology, Cambridge, MA, USA.



Matthew D. Piggott received the master's and Ph.D. degrees in mathematics from the University of Bath, Bath, U.K., in 1998 and 2002, respectively.

He was with Imperial College London, London, U.K., in 2001 to work on adaptive mesh techniques and finite element methods for computational fluid dynamics with an emphasis on geophysical flows. He was a Professor of computational geoscience and engineering in 2016. His research interests include machine learning and data science, optimization and inversion, digital twins, remote sensing, renewable energy systems, natural hazards, and numerical analysis, specifically in numerical methods development for the solution of partial differential equations. He was the Head of the Computational Geoscience and Engineering Research Section and Director of the M.Sc.in environmental data science and machine learning. He is currently the Director of postgraduate studies for masters level courses with the Department of Earth Science and Engineering.



Yanghua Wang received the Ph.D. degree in geophysics from Imperial College London, London, U.K., in 1997.

Since 2004, he has been a geophysics Professor with Imperial College London.. Since 2019, he has been also the Director with Resource Geophysics Academy, Imperial College London.

Dr. Wang is a Fellow with the Royal Academy of Engineering, London, U.K., and an International Member of Chinese Academy of Engineering.



Rossella Arcucci received the Ph.D. degree in computational and computer science from University of Naples Federico II and EuroMediterranean Centre on Climate Change (CMCC), in 2012.

She is currently an Associate Professor in data learning and AI for Good with Imperial College London, London, U.K., a Director of research with the Imperial Data Science Institute, and a Director with the Ada Lovelace Academy at Imperial. She is an Elected Member of the World Meteorological Organization and Visiting Professor with the European Space Agency PhiLab, Frascati, Italy. Her research focuses on developing AI models for climate, health, and environmental impact. She also collaborates with the Leonardo Centre on Business for Society, Imperial College Business School. She is also a Co-Investigator of several grants and projects.

Dr. Arcucci was the recipient of Marie Sklodowska-Curie fellow from European Commission Research Executive Agency in 2017.