Latent Three-dimensional Variational Data Assimilation with Convolutional Autoencoder and LSTM for Flood Forecasting

Kun Wang[⋆,1,2], Gabriele Bertoli[⋆,3], Kai Schröter[4], Enrica Caporali[3], Matthew D. Piggott[2], Yanghua Wang[1,2], and Rossella Arcucci[2,5]

[1] Resource Geophysics Academy, Imperial College London, South Kensington, London SW7 2AZ, UK
[2] Department of Earth Science and Engineering, Imperial College London, South Kensington, London SW7 2AZ, UK
[3] Department of Civil and environmental Engineering, University of Florence, via di Santa Marta, 3. Firenze, Italy
[4] Leichtweiß-Institute for Hydraulic Engineering and Water Resources, Division Hydrology and River Basin Management, Technische Universität Braunschweig, Beethovenstr. 51a, 38106 Braunschweig, Germany
[5] Data Science Institute, Imperial College London, London, UK

Abstract. Fast and accurate flood forecasting models are fundamental for managing flood risk and mitigating the negative impacts that floods can have on the society and the environment. For the pan-European area, currently, the European Flood Awareness System (EFAS) is the official flood forecasting and early warning system. It's forecasts derive from a process-based rainfall-runoff model (LISFLOOD), which requires large amounts of high quality hydro-meteorological data, that usually are not uniformly available, affecting the quality of the outputs. Running process-based models at high spatial resolution for large spatial scales requires high-performance computer clusters to deliver timely forecasts. Recently, the use of machine learning-based forecasting models like long short-term memory (LSTM) as surrogate models for traditional process-based models has gained popularity. Compared to traditional process-based models, machine learning-based forecasting models offer the advantage of lower computational resource requirements and greater tolerance for variations in input data quality. Additionally, machine learning-based compression models, such as convolutional autoencoder (CAE), can further reduce computational costs by compressing the data. Moreover, large-scale models, frequently exhibit lower accuracy in forecasting river discharge in smaller watersheds, in which data may be less available and rivers are smaller, but still significant, and cannot be overlooked. These watersheds exhibit a fast response to rainfall events, thus requiring fast river discharge forecasts to guarantee enough lead time to take action in case of an imminent flood. To enhance forecasting accuracy, data assimilation (DA)—a technique that integrates data from multiple sources to optimize forecasting outcomes—can be effectively employed to improve

[⋆] co-first author

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

the precision of river discharge forecasting. In this work, we propose a latent three-dimensional variational data assimilation (3D-Var) method combined with machine learning models to deliver fast and accurate river discharge forecasting. We tested our method on the real-world datasets (EFAS and Lamah-CE) and achieved an average 53.6% improvement in forecasting accuracy measured by Mean Squared Error (MSE) compared to LSTM forecasting, while delivering one-day lead-time river discharge forecasting in approximately one minute for an area of around 30,000 km^2.

Keywords: flood forecasting · three-dimensional variational data assimilation · convolutional autoencoder · long short-term memory.

1 Introduction

River flow forecasting is key in flood risk management, being fundamental for early warning systems, thus improving emergency preparedness and supporting decision-making for mitigating potential impacts to people, infrastructures and environment. Traditional hydrological and hydraulic models, such as widely used conceptual models based on physical data, are currently the most established techniques for forecasting river floods. However, these models often rely on extensive parameter calibration and suffer from computational inefficiencies, particularly when applied at large spatial scales [9, 26]. The complexity of river systems and precipitation events, the varying watershed characteristics and the scarcity of high quality data, have revealed significant limitations in traditional methods, also for efficient real-time data assimilation [8, 37, 42, 18]. Recent advances in machine learning (ML) offer a promising alternative by enabling faster and more adaptive flood forecasting, potentially learning complex hydrological patterns without requiring physical modeling, reducing the computational resources necessary for data preparation for model set-up [24] and not requiring the subsequent calibration phase. Recent studies have demonstrated that ML-based models can outperform traditional methods in terms of forecasting speed and accuracy [39, 7]. Achieving accurate and timely flood forecasting on multiple spatial scales is a significant step forward in flood risk management.

Three-dimensional variational data assimilation (3D-Var) is widely used in flood forecasting as a method of fusing observations to correct the background fields forecasted by models [23, 31, 35, 34, 15]. However, when dealing with high-dimensional data, 3D-Var is typically computationally intensive and thus requires an implementation in a reduced space [4, 10], often called control space [3] or latent space [5]. In recent years, many researchers have utilized compression models such as Autoencoder (AE) [1, 41] and convolutional autoencoder (CAE) [32, 2, 33] in combination with the 3D-Var method, a technique known as latent data assimilation (DA) [17, 16, 13, 14, 30]. By reducing the dimensionality of the field, this approach significantly accelerates the computation of 3D-Var while achieving the desired accuracy. In addition, traditional hydrological models, such as LISFLOOD [43], require significant computational effort and long simulation times to model river discharge. As a result, a common approach today in the research is to use machine learning-based models as surrogate models to enhance computational efficiency [20, 46, 29, 11, 44, 12, 47, 45]. As a widely used surrogate model in flood forecasting, long short-term memory (LSTM) [22] features a specialized gating mechanism that effectively handles long-term dependencies, making it well-suited for capturing complex temporal patterns in hydrological data. In this paper, to achieve fast and accurate flood forecasting, we propose a flood forecasting method that integrates the 3D-Var technique with a machine learning based surrogate model using a LSTM to achieve faster and more accurate flood forecasts. We validate our method using real datasets, including the European Flood Awareness System (EFAS) [36] and the LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe (Lamah-CE) [28].

The following sections of this article are organized as follows: Section 2 describes the dataset and preprocessing, Section 3 introduces the proposed method, Section 4 presents the experimental results, and Section 5 summarizes the paper outcomes while outlining potential directions for future research.

2 Datasets and preprocessing

This section provides an overview of the datasets utilized in this research to validate our proposed method. The first step involved identifying hydrological data with the highest available spatial and temporal resolution, suitable for subsequent experiment and comparison. River discharge simulations from the EFAS have been selected to be studied alongside river discharge observations, provided by the LamaH-CE dataset. Data sets like LamaH-CE and the EFAS are suitable for such investigations, as their combination of data quality, extensive coverage, consistency across large domains, diversity, and public availability provides a solid foundation for improving the knowledge of hydrological processes and developing new modeling techniques.

2.1 Datasets description

EFAS is a pan-European project aimed at providing early flood forecasting and warnings to support disaster risk management and mitigation. EFAS covers the entire European continent, releasing flood forecasts at a spatial resolution of 1 arcmin. Its forecasts are based on the LISFLOOD [43] hydrological model, an integrated and distributed rainfall-runoff and river routing model specifically designed for large-scale applications. LISFLOOD simulates hydrological processes such evapotranspiration, soil moisture dynamics, snow melt, obtaining the runoff for each cell of the grid. Runoff values are then routed into the river network, allowing for representation of river flow, which is one of the LISFLOOD outputs, expressed in cubic meters per second. Historical runs of EFAS river discharge are based on reanalysis of meteorological data, and over the LamaH-CE area, they are available from 1991 to nowadays. The LamaH-CE dataset is a comprehensive

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

resource for large-scale hydrological research and modeling, focused on, but not limited to, river discharge observations. Covering a temporal range from 1981 to 2017, it provides high-resolution hydro-meteorological data for 859 gauged catchments in Central Europe, representing diverse climatic and geographical conditions. River discharge observations, expressed in cubic meters per second, are retrieved from river gauges, and stored at daily and hourly resolutions, enabling detailed analyses of river flow and hydrological processes. The dataset also includes advanced basin delineations that allow inter-catchment analyses and the study of hydrological connectivity. The main differences between these two datasets are that the LamaH-CE structure follows the hydrological watersheds hierarchy and provides river discharge observations at each river gauge location (punctual), with hourly time resolution, while EFAS has a gridded structure with 1 arcmin spatial resolution, and provides simulated river discharge values at 6-hour time resolution. No direct link between the EFAS grid and the LamaH-CE river discharge data is defined. Preprocessing of the two datasets is necessary for comparing and jointly processing both data sets.

2.2 Data preprocessing

LamaH-CE observations, provided in single CSV files for each river gauge, have been firstly aggregated into a unique matrix, in which each row contains all the available river discharge records for a specific river gauge, and each column contains all the values recorded at the corresponding time. EFAS data are instead delivered in NetCDF4 file format, a multidimensional raster file, containing 1 year of simulations for each file. Before proceeding with data extraction and time alignment of the EFAS data, a direct relation between river gauge observations of the LamaH-CE and the cells of the EFAS simulations must be determined.

Spatial alignment Finding relationships between the two datasets implies establishing a link between cells of the EFAS raster and the locations of the river gauges of the LamaH-CE. Only specific cells of the raster grid (i.e. the most representatives of the observed flow) can be directly compared with the observations, thus, all the other cells that store river flow values far from the gauges, will not be considered in this work. The resulting data structure has been consequently designed to adopt only the LamaH-CE gauges (i.e. observation locations) as spatial reference. Aligning the raster data from EFAS with the punctual values of the LamaH-CE dataset is not straightforward. The resolution of the EFAS grid (1 arcmin), designed for large-scale modeling, does not allow to accurately reproduce the actual river network, introducing several uncertainties when finding the cell that corresponds to the gauge location at a river. Furthermore, the actual river width in this region is less than 1 arcmin, and especially in plains it is common to have more than one river covered by a single EFAS cell, complicating the cell selection procedure. Moreover, due to the grid geometry, the river cells do not exactly follow the river bends. Consequently, in many cases, the EFAS cells do not overlap with the closest river gauge, even if they represent the flow

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

that is measured by that sensor. This misalignment is also partially explainable by the Digital Elevation Models (DEM) used to determine the river network in the two datasets having different resolutions, thus resulting in slightly different paths of the rivers. To overcome this problem, an automated hydrological approach has been adopted to determine the most suitable cells of the EFAS grid to be associated with the corresponding river gauges. The adopted approach derives from the hydrological definition of watersheds: they may have multiple inlets but they have only a single outlet, and river gauges are installed in proximity of the outlets to measure the streamflow from the watershed. Therefore, the EFAS cell reporting the maximum streamflow within a given basin, is assumed to be the most representative for the river gauge, and it is then selected to be associated with the river gauge of that watershed. For each watershed in the LamaH-CE dataset, an automatic procedure selected such cells associating them with the respective river gauges. To avoid errors for small watersheds, LamaH-CE watersheds have been aggregated to at least 100 km^2 catchment area, taking hydrological connectivity into account. After the described processing, only a few misalignment cases related to peculiar and uncommon catchment shapes, have been manually fixed or removed.

The spatial alignment procedure result is a selection of EFAS cells for the next extraction of river discharge values from the native NetCDF4 files, for proceeding with the temporal alignment of the two datasets.

Temporal alignment The final data structure has been designed to adopt the EFAS time resolution (6 hours) and format. EFAS time is in Unix format, also referred to as Epoch time. The Unix time system, implemented as an integer counter, is a standard format for timekeeping in databases. It ensures consistency and comparability across datasets, making it appropriate for setting a unique and comparable timestamp format between EFAS and LamaH-CE data. Consequently, LamaH-CE timestamps were converted (from human-readable format) to EFAS time format. EFAS streamflow data are only available as 6 hour average values, and we opted to keep this time resolution instead of the hourly resolution of the LamaH-CE dataset, because disaggregating EFAS values to hourly resolution would introduce additional uncertainty to our analyses. In the present work, hourly LamaH-CE observations have been averaged over the past 6 hours, keeping the same daily time steps of the EFAS dataset (at 00:00, 06:00, 12:00, 18:00). Regarding the year coverage of the two datasets, a common time window has been defined, and both datasets have been aligned to cover a 25 years interval.

After the described procedures, data from the two original datasets are linked, allowing both an observed and a simulated river discharge to be associated with each time step and location, therefore enabling the next steps of the study.

Data selection and variables A subset region of the LamaH-CE dataset of 58 river gauges, located over approximately 30000 km^2 in eastern Austria, has been selected as test case for this study. In the following sections, the simulated

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

river discharge values from EFAS are identified as the state field $\mathbf{X}_i$ (raw EFAS data over the study area - grid of 128 x 128 cells), and the vectors $\mathbf{x}_i$ contain the EFAS cell values associated to the LamaH-CE gauges (after spatial alignment: 58 cells) for each time step i. River flow observations for each gauge from the LamaH-CE dataset are identified by the vectors $\mathbf{y}_i$, (58 gauges) for each time step i.

3 Method

This section provides a detailed description of the latent 3D-Var method we proposed, which is combined with machine learning models.

Firstly, a compression model is introduced to compress the state field $\mathbf{X}_i$ into the latent space. The structure employed for this purpose is a CAE, shown in Eq. 1,

$$\begin{aligned} \widehat{\mathbf{X}}_i &= \mathcal{D}(\mathcal{E}(\mathbf{X}_i)) \\ \tilde{\mathbf{x}}_i &= \mathcal{E}(\mathbf{X}_i), \end{aligned} \tag{1}$$

where $\mathbf{X}_i \in \mathbb{R}^{128\times128}$ represents the state field at the i_{th} time step, $\mathcal{E} : \mathbb{R}^{128\times128} \to \mathbb{R}^{128\times1}$ represents the encoder of the CAE, $\mathcal{D} : \mathbb{R}^{128\times1} \to \mathbb{R}^{128\times128}$ represents the decoder of the CAE, $\widehat{\mathbf{X}}_i \in \mathbb{R}^{128\times128}$ represents the reconstructed state field at the i_{th} time step and $\tilde{\mathbf{x}}_i \in \mathbb{R}^{128\times1}$ represents the state vector in the latent space at the i_{th} time step. The dimension of $\mathbf{X}_i$ is 128×128, while the dimension of $\tilde{\mathbf{x}}_i$ is 128×1, indicating that $\tilde{\mathbf{x}}_i$ has a significantly smaller size compared to $\mathbf{X}_i$. The process of training the CAE model is to minimize the distance between the state field and the reconstructed state field, ensuring that the compression process retains as much relevant information as possible, shown in Eq. 2.

$$\mathcal{L} = ||\mathbf{X}_i - \widehat{\mathbf{X}}_i||^2 \tag{2}$$

where $\mathcal{L}$ represents the loss function. The advantage of the CAE structure lies in its convolutional layers, which have the capability to extract spatial features and capture the nonlinear correlations within the state field. Additionally, the CAE effectively removes noisy information from the state field [27, 40], enhancing the quality of the compressed representation. After the model is trained, the state field is passed through the encoder to obtain the state vector in the latent space, which is then used in the subsequent processes for forecasting and assimilation.

Secondly, to improve forecasting efficiency, a machine learning surrogate model is designed to accelerate the forecasting process. The surrogate model is based on the LSTM network, a variant of recurrent neural network. The unique gating mechanism of LSTM allows it to effectively handle long-term dependencies in time series data, making it well-suited for river discharge forecasting. In this experiment, the 4-to-4 LSTM forecasting is performed within the latent space, shown in Eq. 3,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

$$\widehat{\tilde{\mathbf{x}}}_{i+4:i+7} = \mathcal{M}_{LSTM}(\tilde{\mathbf{x}}_{i:i+3}), \tag{3}$$

where $\widehat{\tilde{\mathbf{x}}}_i$ represents the forecasted vector in the latent space at the i_{th} time step, $\mathcal{M}_{LSTM}$ represents the LSTM model. The process of training the LSTM model is to minimize the distance between the forecasting and the simulation, shown in Eq. 4.

$$\mathcal{L} = ||\widehat{\tilde{\mathbf{x}}}_{i+4:i+7} - \tilde{\mathbf{x}}_{i+4:i+7}||^2 \tag{4}$$

The trained LSTM serves as the offline model in the subsequent 3D-Var process, allowing for a reduction in computational costs by eliminating the need to update parameters during the real-time forecasting.

Thirdly, to enhance the accuracy of LSTM forecasting and improve the efficiency of the 3D-Var process, a latent 3D-Var framework is developed to refine the LSTM outputs. The cost function of latent 3D-Var is shown in Eq. 5,

$$\begin{aligned} J(\tilde{\mathbf{x}}_{i:i+3}) = &\frac{1}{2}(\tilde{\mathbf{x}}_{i:i+3} - \tilde{\mathbf{x}}^b_{i:i+3})^T \tilde{\mathbf{B}}^{-1}(\tilde{\mathbf{x}}_{i:i+3} - \tilde{\mathbf{x}}^b_{i:i+3}) \\ &+ \frac{1}{2}(\mathcal{H}(\tilde{\mathbf{x}}_{i:i+3}) - \mathbf{y}_{i:i+3})^T \mathbf{R}^{-1}(\mathcal{H}(\tilde{\mathbf{x}}_{i:i+3}) - \mathbf{y}_{i:i+3}), \end{aligned} \tag{5}$$

where $\tilde{\mathbf{x}}^b_i \in \mathbb{R}^{128\times1}$ represents the background vector in the latent space at the i_{th} time step, $\tilde{\mathbf{B}} \in \mathbb{R}^{128\times128}$ represents the background error covariance matrix in the latent space [6], $\mathbf{y}_i \in \mathbb{R}^{58\times1}$ represents the observation at the i_{th} time step, $\mathcal{H} : \mathbb{R}^{128\times1} \rightarrow \mathbb{R}^{58\times1}$ represents the observation operator which is composed of a decoder and a location selection operator and $\mathbf{R} \in \mathbb{R}^{58\times1}$ represents the observation error covariance matrix [25]. We estimate these two covariance matrices on the training and validation datasets. The background error covariance matrix in the latent space is derived from the forecasting error of the LSTM model. The observation error covariance matrix is estimated as the covariance matrix of the observations, and we approximate this matrix by retaining only the diagonal elements, resulting in a diagonal covariance matrix. In the implementation of latent 3D-Var, the four time steps forecasted by the LSTM (referred to as background vectors in DA) are used as inputs to the latent 3D-Var. Additionally, four corresponding observations at the same time steps are incorporated for assimilation. Then the assimilation process obtains the analysis vector by minimizing the cost function J, shown in Eq. 6.

$$\tilde{\mathbf{x}}^a_{i:i+3} = argmin(J(\tilde{\mathbf{x}}_{i:i+3}))) \tag{6}$$

where $\tilde{\mathbf{x}}^a_i \in \mathbb{R}^{128\times1}$ represents the analysis vector in the latent space at the i_{th} time step. Finally, the analysis vector is passed through the decoder to generate the analysis field at the i_{th} time step $\mathbf{x}^a_i \in \mathbb{R}^{128\times128}$.

The training and assimilation processes are conducted on an Intel Core i9-14900HX CPU and a NVIDIA GeForce RTX 4070 GPU. During the training process, the first 70% of the data from the datasets is used as the training set,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

the next 10% is allocated for validation, and the remaining 20% is reserved as the test set to evaluate the performance of the model and the latent 3D-Var framework. The loss function used for both the CAE and LSTM models is Mean Squared Error (MSE), and the optimization is performed using the Adam optimizer.

4 Results

This section evaluates the performance of the proposed latent 3D-Var compared wth the LSTM model.

4.1 Numerical result

The first evaluation metric used is MSE, shown in Eq. 7,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{y}_i)^2, \tag{7}$$

where $\mathbf{x}_i \in \mathbb{R}^{58\times 1}$ represents the forecasting or assimilation result at the i_{th} time step. A lower MSE indicates that the forecasted values are closer to the observations, which is crucial for providing accurate flood forecasting. The observed data is derived from the Lamah-CE dataset, meaning the calculations are restricted to the values of elements in the state field that correspond to observation points. As shown in the Fig. 1, we have selected three weeks of results here, with 28 result points for each week (four result points per day).

In Fig. 1, the blue line represents the MSE between latent 3D-Var and the observations, and the red line represents the MSE between LSTM and the observations. The dot lines indicate the time steps at which the 3D-Var assimilation is performed. It is evident that after performing 3D-Var, the MSE of latent 3D-Var are significantly lower than those of LSTM. Numerically, the MSE of latent 3D-Var is 53.6% lower than that of LSTM on average, with the maximum reduction over 90%. Thus, the results demonstrate that latent 3D-Var can significantly correct LSTM forecasting, thereby enhancing the accuracy of flood forecasting.

The second evaluation metric used is the Standard Deviation (STD), which reflects the distribution of the residuals, shown in Eq. 8.

$$STD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - \bar{e})^2}, \tag{8}$$

where e_i represents the error between the forecasting or assimilation result at the i_{th} time step and the observations and $\bar{e}$ represents the mean value of the error. In the context of flood forecasting, a lower STD signifies that the results are more reliable and stable, which is crucial for informed decision-making in flood risk management. The result of STD is shown in Fig. 2.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

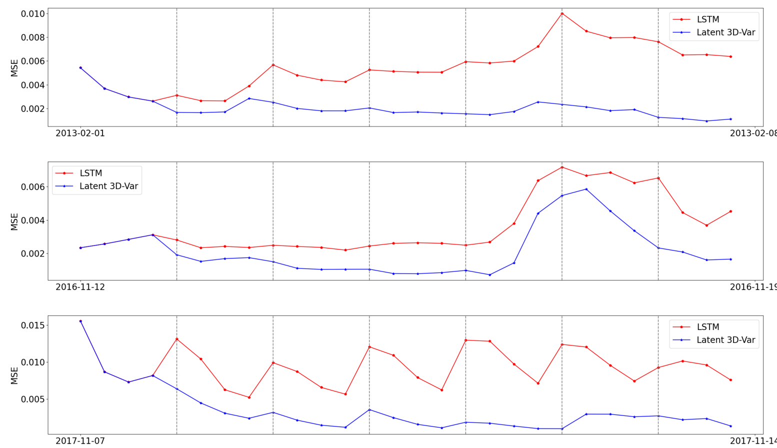

Fig. 1. The MSE of latent 3D-Var and LSTM for randomly selected weeks (The blue line represents the MSE between latent 3D-Var and the observations, and the red line represents the MSE between LSTM and the observations. The dot lines indicate the time steps at which the 3D-Var assimilation is performed.)

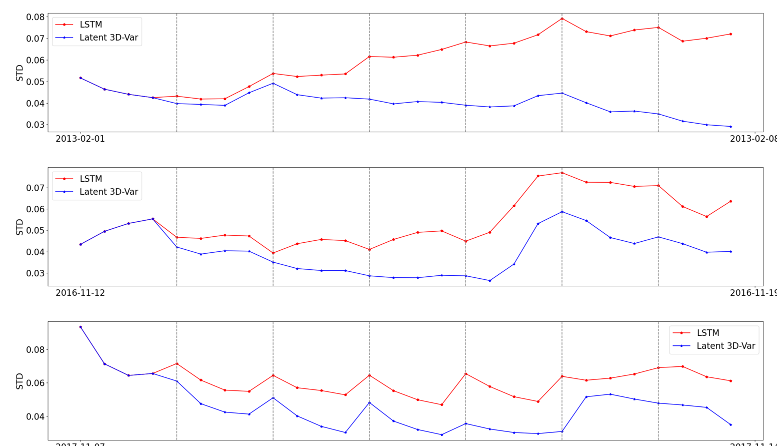

Fig. 2. The STD of latent 3D-Var and LSTM for randomly selected weeks (The blue line represents the STD between latent 3D-Var and the observations, and the red line represents the STD between LSTM and the observations. The dot lines indicate the time steps at which the 3D-Var assimilation is performed.)

In Fig. 2, the STD of latent 3D-Var is significantly reduced, with values consistently lower than those of LSTM after the initial four time steps. On average, the reduction is 27.2%, with a maximum reduction of 59.6%. These results indicate that the STD of the latent 3D-Var corrected outputs is considerably lower

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

than that of the LSTM results. This suggests that the latent 3D-Var method produces more stable and reliable forecasting, making it a more suitable basis for decision-making in the flood forecasting and warning.

4.2 Visualization result

To visualize the results of latent 3D-Var and LSTM, the absolute error at four different moments of a single day (randomly chosen) is spatially plotted in Fig. 3, with lighter colored circles (river gauges locations) signifying better forecasting. LSTM results, reported in the upper row, in general show larger errors than latent 3D-Var, with error spikes on individual gauges, and approximately twice the variance of the latent 3D-Var forecasting. Therefore, not only the latent 3D-Var is performing better than LSTM in forecasting river flow, but also its performances show less variability and are thus more homogeneous across different areas and watersheds.

4.3 Execution time

Additionally, the execution time required for latent 3D-Var to correct the four time steps (one day), using a NVIDIA GeForce RTX 4070 GPU, is approximately 60 seconds to cover 58 river gauges. This time is significantly faster compared to conventional 3D-Var methods, greatly enhancing computational speed. Furthermore, this processing time is well-suited for the needs of flood forecasting.

5 Conclusion

In this paper, we propose a latent 3D-Var method combined with machine learning to address the challenges of achieving fast and accurate flood forecasting. The designed CAE model efficiently compresses the state field and reduces its dimensionality, thereby decreasing the computational load of both the forecasting model and the 3D-Var process. Additionally, the CAE acts as a denoising mechanism, minimizing error accumulation during the forward process. Simultaneously, the methods employs LSTM as the surrogate model, significantly improving computational efficiency compared to traditional hydrological models. The observations are then assimilated with the LSTM forecasting to yield more accurate results. The results demonstrate that this method greatly speeds up the process and reduces errors. More importantly, for the task of flood forecasting, the results of latent 3D-Var are much closer to the real values, while also being more reliable and stable. At the same time, latent 3D-Var has a very short running time of approximately one minute, making it well-suited to meet the needs of flood forecasting. In hydrological terms, the dataset subset (58 sensors – 26 years) used in this study is relatively small but deemed sufficient for developing a new machine learning-integrated 3D-Var model for river discharge forecasting. However, comprehensive validation is required following this initial

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

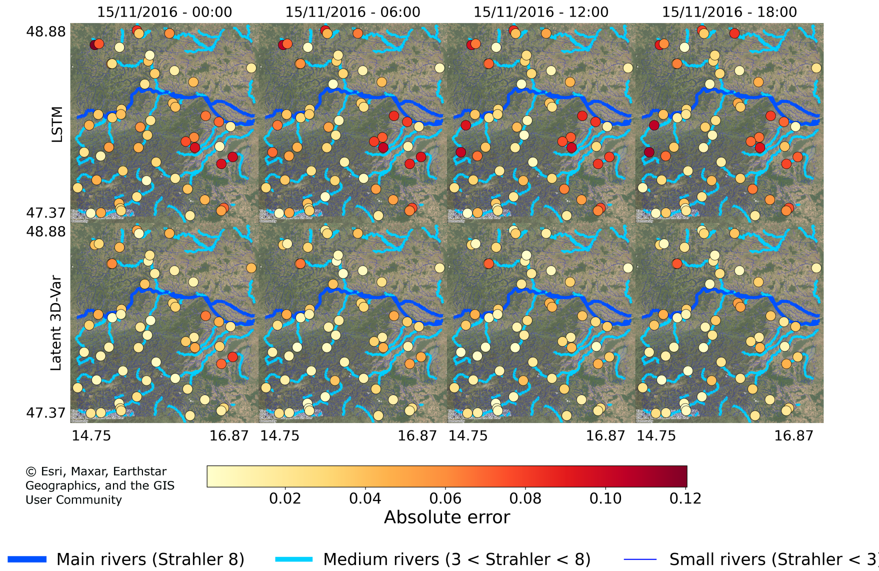

Fig. 3. In this figure a spatial evaluation of absolute error in river discharge forecasting across a test area is shown. River gauges in the study area are represented by the colored circles. On x-axes and y-axes are reported the geographic coordinates of the extent of the represented area. The upper row of images reports LSTM forecasting errors, while the lower row shows 3D-Var forecasting errors, which demonstrates significantly improved accuracy compared to LSTM. Each column corresponds to different time steps (00:00, 06:00, 12:00, and 18:00) on November 15, 2016. The upper color scale at the bottom indicates the magnitude of absolute error, with lower values (lighter colors) signifying better forecasting. 3D-Var consistently exhibits lower error, confirming its superior performance in river discharge forecasting. Satellite images from "© Esri, Maxar, Earthstar Geographics, and the GIS User Community" [19] are displayed in the background. The river network is classified according to the Strahler stream order, which assigns higher values to rivers with more tributaries. The dark blue river represents a section of the Danube.

step. Future studies could incorporate longer time series and larger areas to expand the model's learning base [37, 9, 26], enhance hydrological understanding, and improve the usability of the results. Also, to deepen the understanding of the forecasting capabilities, metrics like the Nash-Sutcliffe Efficiency (NSE) [38] and the Kling-Gupta Efficiency (KGE) [21], which are commonly used metrics to assess the performance of hydrological models, can be adopted to better align the results to the hydrological science. Additionally, extreme events, whether driven by climate change or unique to specific sub-regions climate, present challenges. The 6-hour time step adopted in our work is considered as standard for large spatial scale simulations, and it's particularly suitable for large river basins with long rainfall-runoff response time. Unfortunately, this time resolution is less

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

suitable for intense but short flooding events, which in the smaller watersheds can cause many damages and severely endanger the population. Conversely, it can be used for detailed analyses of extreme drought events in long-range forecasts, which evolve over extended time scales. For example, it can be leveraged to enhance the forecasting of reservoirs storage level, contributing to an improved and more reliable water resource management. Consequently, this broadens the applicability of the obtained results to the analysis of multiple natural hazards.

Acknowledgments. The authors are grateful to the sponsors at the Resource Geophysics Academy, Imperial College London, for supporting this research.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Amendola, M., Arcucci, R., Mottet, L., Casas, C.Q., Fan, S., Pain, C., Linden, P., Guo, Y.K.: Data assimilation in the latent space of a neural network. arXiv preprint arXiv:2012.12056 (2020)
2. Amendola, M., Arcucci, R., Mottet, L., Casas, C.Q., Fan, S., Pain, C., Linden, P., Guo, Y.K.: Data assimilation in the latent space of a convolutional autoencoder. In: International Conference on Computational Science. pp. 373–386. Springer (2021)
3. Arcucci, R., Mottet, L., Pain, C., Guo, Y.K.: Optimal reduced space for variational data assimilation. Journal of Computational Physics **379**, 51–69 (2019)
4. Arcucci, R., Xiao, D., Fang, F., Navon, I.M., Wu, P., Pain, C.C., Guo, Y.K.: A reduced order with data assimilation model: Theory and practice. Computers & Fluids **257**, 105862 (2023)
5. Arcucci, R., Zhu, J., Hu, S., Guo, Y.K.: Deep data assimilation: integrating deep learning with data assimilation. Applied Sciences **11**(3), 1114 (2021)
6. Bannister, R.N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography **134**(637), 1951–1970 (2008)
7. Bentivoglio, R., Isufi, E., Jonkman, S.N., Taormina, R.: Deep learning methods for flood mapping: a review of existing applications and future research directions. Hydrology and Earth System Sciences **26**(16), 4345–4378 (2022). https://doi.org/10.5194/hess-26-4345-2022, https://hess.copernicus.org/articles/26/4345/2022/
8. Blöschl, G., [...], Zhang, Y.: Twenty-three unsolved problems in hydrology (uph) – a community perspective. Hydrological Sciences Journal **64**(10), 1141–1158 (2019). https://doi.org/10.1080/02626667.2019.1620507
9. Brunner, M.I., Slater, L., Tallaksen, L.M., Clark, M.: Challenges in modeling and predicting floods and droughts: A review. WIREs Water **8**(3), e1520 (2021). https://doi.org/https://doi.org/10.1002/wat2.1520
10. Casas, C.Q., Arcucci, R., Wu, P., Pain, C., Guo, Y.K.: A reduced order deep data assimilation model. Physica D: Nonlinear Phenomena **412**, 132615 (2020)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

11. Chen, J., Li, Y., Zhang, C., Tian, Y., Guo, Z.: Urban flooding prediction method based on the combination of lstm neural network and numerical model. International Journal of Environmental Research and Public Health **20**(2), 1043 (2023)
12. Cheng, S., Bocquet, M., Ding, W., Finn, T.S., Fu, R., Fu, J., Guo, Y., Johnson, E., Li, S., Liu, C., et al.: Machine learning for modelling unstructured grid data in computational physics: a review. arXiv preprint arXiv:2502.09346 (2025)
13. Cheng, S., Chen, J., Anastasiou, C., Angeli, P., Matar, O.K., Guo, Y.K., Pain, C.C., Arcucci, R.: Generalised latent assimilation in heterogeneous reduced spaces with machine learning surrogate models. Journal of Scientific Computing **94**(1), 11 (2023)
14. Cheng, S., Liu, C., Guo, Y., Arcucci, R.: Efficient deep data assimilation with sparse observations and time-varying sensors. Journal of Computational Physics **496**, 112581 (2024)
15. Cheng, S., Lucor, D., Argaud, J.P.: Observation data compression for variational assimilation of dynamical systems. Journal of computational science **53**, 101405 (2021)
16. Cheng, S., Prentice, I.C., Huang, Y., Jin, Y., Guo, Y.K., Arcucci, R.: Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. Journal of Computational Physics **464**, 111302 (2022)
17. Cheng, S., Zhuang, Y., Kahouadji, L., Liu, C., Chen, J., Matar, O.K., Arcucci, R.: Multi-domain encoder–decoder neural networks for latent data assimilation in dynamical systems. Computer Methods in Applied Mechanics and Engineering **430**, 117201 (2024)
18. Cloke, H., Pappenberger, F.: Ensemble flood forecasting: A review. Journal of Hydrology **375**(3), 613–626 (2009). https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.06.005, https://www.sciencedirect.com/science/article/pii/S0022169409003291
19. Esri: Esri world imagery basemap, https://www.arcgis.com, accessed: 2025-02-17
20. Fang, Z., Wang, Y., Peng, L., Hong, H.: Predicting flood susceptibility using lstm neural networks. Journal of Hydrology **594**, 125734 (2021)
21. Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F.: Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. Journal of Hydrology **377**(1-2), 80–91 (2009). https://doi.org/10.1016/j.jhydrol.2009.08.003
22. Hochreiter, S.: Long short-term memory. Neural Computation MIT-Press (1997)
23. Hou, T., Kong, F., Chen, X., Lei, H.: Impact of 3dvar data assimilation on the prediction of heavy rainfall over southern china. Advances in Meteorology **2013**(1), 129642 (2013)
24. Huang, H., Lei, X., Liao, W., Liu, D., Wang, H.: A hydrodynamic-machine learning coupled (hmc) model of real-time urban flood in a seasonal river basin using mechanism-assisted temporal cross-correlation (mtc) for space decoupling. Journal of Hydrology **624**, 129826 (2023). https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.129826, https://www.sciencedirect.com/science/article/pii/S0022169423007680
25. 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 the representation error in data assimilation. Quarterly Journal of the Royal Meteorological Society **144**(713), 1257–1278 (2018)
26. Jew Das, Velpuri Manikanta, K.N.T., Umamahesh, N.V.: Two decades of ensemble flood forecasting: a state-of-the-art on past developments, present applications

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

and future opportunities. Hydrological Sciences Journal **67**(3), 477–493 (2022). https://doi.org/10.1080/02626667.2021.2023157
27. Jung, Y., Kim, T., Han, M.R., Kim, S., Kim, G., Lee, S., Choi, Y.J.: Ovarian tumor diagnosis using deep convolutional neural networks and a denoising convolutional autoencoder. Scientific Reports **12**(1), 17024 (2022)
28. Klingler, C., Schulz, K., Herrnegger, M.: LamaH-CE: LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe. Earth System Science Data **13**(9), 4529–4565 (2021). https://doi.org/10.5194/essd-13-4529-2021, https://doi.org/10.5194/essd-13-4529-2021
29. Le, X.H., Ho, H.V., Lee, G., Jung, S.: Application of long short-term memory (lstm) neural network for flood forecasting. Water **11**(7), 1387 (2019)
30. Liu, C., Fu, R., Xiao, D., Stefanescu, R., Sharma, P., Zhu, C., Sun, S., Wang, C.: Enkf data-driven reduced order assimilation system. Engineering Analysis with Boundary Elements **139**, 46–55 (2022)
31. Liu, Y., Liu, J., Li, C., Liu, L., Wang, Y.: A wrf/wrf-hydro coupled forecasting system with real-time precipitation–runoff updating based on 3dvar data assimilation and deep learning. Water **15**(9), 1716 (2023)
32. Mack, J., Arcucci, R., Molina-Solana, M., Guo, Y.K.: Attention-based convolutional autoencoders for 3d-variational data assimilation. Computer Methods in Applied Mechanics and Engineering **372**, 113291 (2020)
33. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J.: Stacked convolutional autoencoders for hierarchical feature extraction. In: Artificial Neural Networks and Machine Learning–ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part I 21. pp. 52–59. Springer (2011)
34. Mazzarella, V., Maiello, I., Ferretti, R., Capozzi, V., Picciotti, E., Alberoni, P., Marzano, F., Budillon, G.: Reflectivity and velocity radar data assimilation for two flash flood events in central italy: A comparison between 3d and 4d variational methods. Quarterly Journal of the Royal Meteorological Society **146**(726), 348–366 (2020)
35. Mazzarella, V., Maiello, I., Capozzi, V., Budillon, G., Ferretti, R.: Comparison between 3d-var and 4d-var data assimilation methods for the simulation of a heavy rainfall case in central italy. Advances in Science and Research **14**, 271–278 (2017)
36. Mazzetti, C., Carton de Wiart, C., Gomes, G., Russo, C., Decremer, D., Ramos, A., Grimaldi, S., Disperati, J., Ziese, M., Schweim, C., Sanchez Garcia, R., Jacobson, T., Salamon, P., Prudhomme, C.: River discharge and related historical data from the european flood awareness system, v5.0 (2023). https://doi.org/10.24381/cds.e3458969, accessed on 15-11-2024
37. Ming, X., Liang, Q., Xia, X., Li, D., Fowler, H.J.: Real-time flood forecasting based on a high-performance 2-d hydrodynamic model and numerical weather predictions. Water Resources Research **56**(7) (2020). https://doi.org/https://doi.org/10.1029/2019WR025583
38. Nash, J.E., Sutcliffe, J.V.: River flow forecasting through conceptual models: Part 1—a discussion of principles. Journal of Hydrology **10**(3), 282–290 (1970). https://doi.org/10.1016/0022-1694(70)90255-6
39. Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T.Y., Weitzner, D., Matias, Y.: Global prediction of extreme floods in ungauged watersheds. Nature **627**(8004), 559–563 (2024). https://doi.org/10.1038/s41586-024-07145-1, https://doi.org/10.1038/s41586-024-07145-1

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4

40. Nishio, M., Nagashima, C., Hirabayashi, S., Ohnishi, A., Sasaki, K., Sagawa, T., Hamada, M., Yamashita, T.: Convolutional auto-encoder for image denoising of ultra-low-dose ct. Heliyon **3**(8) (2017)
41. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. nature **323**(6088), 533–536 (1986)
42. Sood, A., Smakhtin, V.: Global hydrological models: a review. Hydrological Sciences Journal **60**(4), 549–565 (2015). https://doi.org/10.1080/02626667.2014.950580
43. Van Der Knijff, J., Younis, J., 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**(2), 189–212 (2010)
44. Wang, H., Zhou, H., Cheng, S.: Dynamical system prediction from sparse observations using deep neural networks with voronoi tessellation and physics constraint. Computer Methods in Applied Mechanics and Engineering **432**, 117339 (2024)
45. Wang, Z., Xiao, D., Fang, F., Govindan, R., Pain, C.C., Guo, Y.: Model identification of reduced order fluid dynamics systems using deep learning. International Journal for Numerical Methods in Fluids **86**(4), 255–268 (2018)
46. Zhang, L., Qin, H., Mao, J., Cao, X., Fu, G.: High temporal resolution urban flood prediction using attention-based lstm models. Journal of Hydrology **620**, 129499 (2023)
47. Zhu, J., Hu, S., Arcucci, R., Xu, C., Zhu, J., Guo, Y.k.: Model error correction in data assimilation by integrating neural networks. Big Data Mining and Analytics **2**(2), 83–91 (2019)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_4