Neural Network as Transformation Function in Data Assimilation*

Kun Wang[1,2], Matthew D. Piggott[2], Yanghua Wang[1,2], and Rossella Arcucci[2,3]

[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] Data Science Institute, Imperial College London, London, UK
{k.wang23,r.arcucci}@imperial.ac.uk

Abstract. Variational Data Assimilation (DA) is a technique aimed at mitigating the error in simulated states by integrating observations. Variational DA is widely employed in weather forecasting and hydrological modeling as an optimization technique for refining dynamic simulation states. However, when constructing the cost function in variational DA, it is necessary to establish a transformation function from simulated states to observations. When observations come from ground sensors or from remote sensing, representing such a transformation function with explicit expressions can sometimes be challenging or even impossible. Therefore, considering the strong mapping capabilities of Neural Network (NN)s in representing the relationship from simulated states to observations, this paper proposes a method utilizing a NN as the transformation function. We evaluate our method on a real dataset of river discharge in the UK and achieved a 39% enhancement in prediction accuracy, measured by Mean Square Error (MSE), compared to the results obtained without DA.

Keywords: Variational data assimilation · Neural network · Mapping

1 Introduction

In the simulation of dynamic systems, prediction errors arise from simulation processes. To mitigate these errors, incorporating observations is necessary. However, it is important to acknowledge that observation errors also exist within the observations. DA, as an optimization method, effectively addresses this issue. By leveraging observations, DA seeks to improve simulated states, thereby bringing simulation states closer to real values [1]. Variational DA, as a subtype of DA, updates simulated states by minimizing the cost function between simulated states and observations [2]. A pivotal component in the cost function of variational DA is the transformation function. This function represents the transition from simulated states to observations and is typically encoded using

* Supported by Resource Geophysics Academy, Imperial College London

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_23

a selection matrix (to represent location information) or a physical formula (to represent the relationship between physical quantities). Variational DA, with its capability to address DA issues arising from nonlinear relationships between simulated states and observations, is progressively gaining popularity in practical applications [3]. This trend is fueled by the common occurrence of nonlinear relationships between simulated states and observations in real-world scenarios. Variational DA finds widespread application in hydrology [4–8]. These applications showcase the effectiveness of variational DA in hydrology for improving the accuracy of simulation results.

However, in practice, establishing the relationship between observations obtained and simulated states (whether based on location or physical relations) using explicit expressions is challenging, rendering it difficult to obtain the transformation function. For instance, there might be no explicit physical expression to represent the relationship from simulated states to observations, or the resolution of simulated states may differ from that of observations. Hence, this presents a challenging issue in the practical implementation of DA. To address this challenge, Cheng et al. [9] proposed a method to compress simulated states and observations into a shared latent space. However, the execution of this approach in an online update necessitates retraining the model whenever updates occur in the inputs to the parametric model, thereby demanding additional computational resources. NNs, as a machine learning method, can be trained to capture the relationship between two correlated variables [10]. In this paper, leveraging the inherent capabilities of NNs, we propose an offline method that employs a NN as the transformation function to address the aforementioned challenge. We evaluated our method using a real-world dataset comprising two sources: European Flood Awareness System (EFAS) [11] and National River Flow Archive (NRFA) [12]. EFAS is utilized for simulation, while NRFA serves as observations. We opt for MSE as our evaluation metric. Our data assimilation method exhibits an average improvement of 15% in accuracy compared to the raw simulation results.

The organization of the remainder of this article is as follows. Section 2 describes the test case. Section 3 introduced the method we proposed. Section 4 presents the results of our experiments. Section 5 summarizes the entire paper and proposes future work.

2 Test case

To validate the feasibility of our method, real-world data is selected as our test case. River discharge from the UK is selected for DA in this experiment. The data chosen for the simulated states consists of simulated river discharge data provided by EFAS. This data provides daily river discharge for the whole of the UK at a spatial resolution of 5 km $\times$ 5 km. The NRFA provides data on daily river flows from 1600 river stations within the UK, which serve as our observations in this experiment.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_23

The data in the EFAS dataset consists of 2D images, where each pixel represents the river discharge for the corresponding area. The size of the data is 200×200 pixels, with only 14120 pixels containing valid values. The number of stations suitable for experimentation in the NRFA dataset is 924. The details of these two datasets are shown in Table 1. In this experiment, DA is conducted every three days for simulated states.

Table 1. the detail of two datasets

name	EFAS	NRFA
spatial resolution	5 km $\times$ 5 km	
temporal resolution	one-day	one-day
size of sample	14120×1	924×1

3 Proposed Method

3.1 Idea

The cost function in variational DA is shown in given by

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - \mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x})), \qquad (1)$$

where $\mathbf{x}_b$ represent simulated states, $\mathbf{y}$ represents observations, $\mathbf{B}$ and $\mathbf{R}$ represent the covariance matrix of simulated states and observations, and $\mathcal{H}$ represents the transformation function.

From Eq. 1, it is evident that the transformation function $\mathcal{H}$ can be regarded as a mapping relation from simulated states to observations. Given the nonlinear mapping capabilities, NNs represent a rational choice for serving as the transformation function. Therefore, in practical applications where it is challenging to represent the relationship from simulated states to observations with explicit expressions, utilizing NNs for mapping purposes presents a viable solution.

3.2 Implementation

In the experiment, the dataset of simulated states can be represented by $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ and the dataset of observations can be represented by $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n\}$, where n represents the number of samples, which corresponds to individual time step. The input of the NN is the simulated state $\mathbf{x}_i$ and the output of the NN is the observation $\mathbf{y}_i$. The structure of the transformation function obtained via the NN approach is depicted in Fig. 1.

In designing the structure of the NN model, the input $\mathbf{x}_i$ of the network is a vector of size 14120×1, and the output $\mathbf{y}_i$ is a vector of size 924×1. Given the

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_23

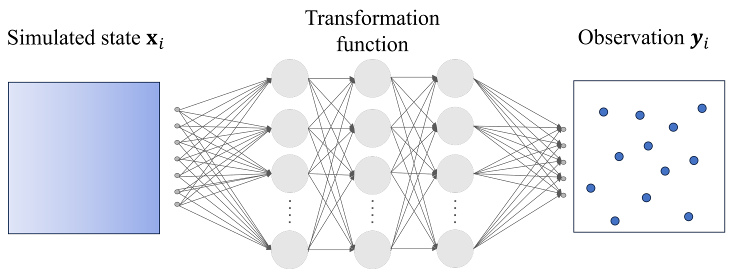

Fig. 1. Structure of transformation function

shape of the input and output vectors, Multi-Layer Perceptron (MLP) is chosen as the network architecture for this experiment. In the MLP structure, which comprises 4 fully connected layers, the activation function used for each layer is the LeakyReLU.

The training of the transformation function is shown in Algorithm. 1, where $k_{\max}$ represents the number of iterations, ϵ represents the threshold of training loss and $\mathcal{L}$ represents the loss function.

Algorithm 1 training transformation function

Inputs:$\mathbf{X}, \mathbf{Y}, \mathcal{H}$
Parameters: $\epsilon, k_{\max}, \mathbf{x}_i \in \mathbf{X}, \mathbf{y}_i \in \mathbf{Y}$
$k = 0$
while $k < k_{\max}$ and $\mathcal{L} > \epsilon$ **do**
  $\mathbf{y}_i = \mathcal{H}(\mathbf{x}_i)$
  $\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} ||\mathbf{y}_i - \mathcal{H}(\mathbf{x}_i)||^2$
  $k = k + 1$
end while
output: $\mathcal{H}$

In the training process, a training strategy that combines mini-batch and Stochastic Gradient Descent (SGD) is employed, which aids in achieving results with smaller error [13]. This approach enhances the NN's ability to more accurately represent the mapping relationship from simulated states to observations. The NN is trained on these two datasets for 300 epochs with a batch size of 4. SGD with a momentum of 0.9 and the weight decay of 1e-4 is chosen as the optimizer. In addition, the cosine annealing strategy is utilized to update the learning rate, starting with an initial learning rate of 1e-2 and a cycle length of 150 epochs [14].

The loss function utilized in our method is MSE, selected to minimize the discrepancy between observations and the mapping result of simulated states to observations via the NN.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_23

4 Result

This section assesses the performance of our proposed method. In this experiment, MSE is selected as the evaluation metric by measuring the error between the ground truth and simulated results. A temporal span of 30 days is selected for the presentation of our findings. The predictive model utilized is Long Short-Term Memory (LSTM) [15]. The result of MSE is shown in Fig. 2.

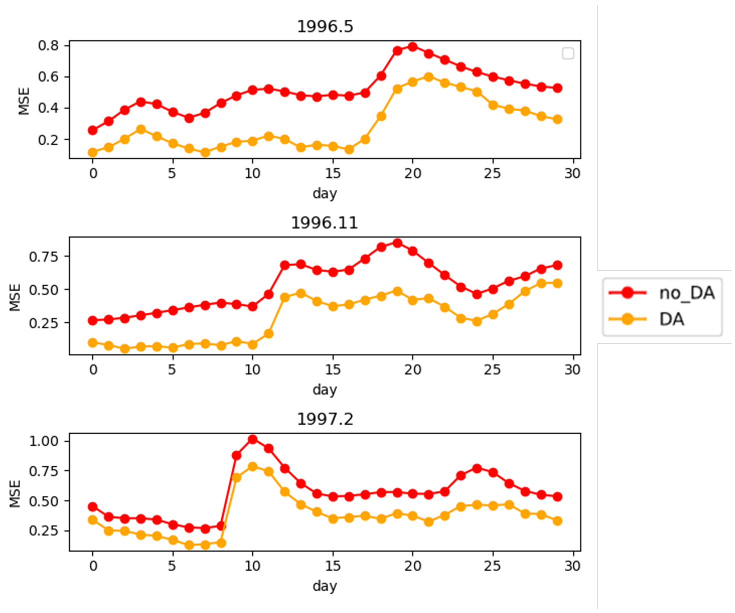

Fig. 2. MSE of our proposed method and the simulated result without DA

In Fig. 2, the red dots represent the MSE between the simulated results without DA and the ground truth, while the orange dots represent the MSE between our proposed method and the ground truth. From Fig. 2, it is evident that the results of our proposed method outperform the simulation results without DA in terms of the MSE metric, with an average improvement of 39%. Additionally, the highest improvement observed exceeds 50%. This metric demonstrates that our method effectively enhances the accuracy of the assimilation results.

Additionally, the Structural Similarity Index (SSIM) is used as an auxiliary evaluation metric. SSIM is an evaluation metric that assesses the structural similarity between two images. The formula for SSIM is given by

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + c_1)(2\sigma_{\mathbf{xy}} + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + c_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + c_2)}, \tag{2}$$

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_23

where $\mu_{\mathbf{x}}$ represents the pixel sample mean of simulated states, $\mu_{\mathbf{y}}$ represents the pixel sample mean of observations, $\sigma_{\mathbf{x}}$ represents the variance of the simulated states, $\sigma_{\mathbf{y}}$ represents the variance of the observations, $\sigma_{\mathbf{xy}}$ represents the covariance of simulated states and observations, c_1 and c_2 represents two variables to stable the division. SSIM is employed here to assess the similarity between the states obtained after DA and the ground truth. The result is shown in Fig. 3.

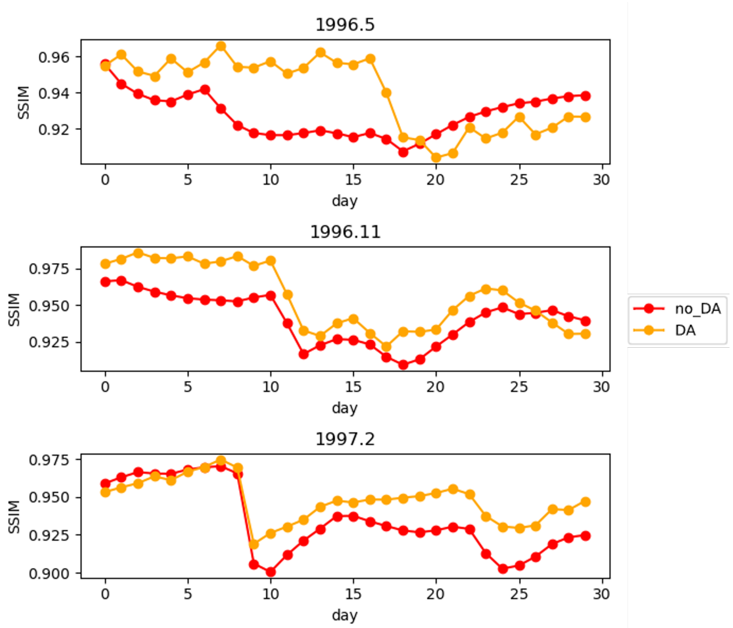

Fig. 3. SSIM of our proposed method and the simulated result without DA

The SSIM of our proposed method is higher than the simulated results without DA. This result indicates that the states obtained by our proposed method are structurally closer to the ground truth, meaning that the variance and mean of the river discharge in the results are closer to the true values.

5 Conclusion

The inability to represent the transformation function with an explicit expression poses a challenge for DA in practical applications. In this paper, we propose to address this issue by employing a NN as the transformation function. This approach leverages the excellent nonlinear mapping properties inherent in NNs. We conduct a test of the proposed method on a real dataset, and the experimental results demonstrate that our approach improves the evaluation metric of MSE by 39% compared to the results obtained without DA. These results validate the

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_23

feasibility of using NN as the transformation function. The utilization of a NN as the transformation function expands the range of observation types viable for the DA. In future work, addressing the issue of sparse observation distribution in data assimilation, which leads to localized correction of the simulated state, will be imperative.

Acronyms

EFAS European Flood Awareness System
DA Data Assimilation
LSTM Long Short-Term Memory
MLP Multi-Layer Perceptron
MSE Mean Square Error
NN Neural Network
NRFA National River Flow Archive
SGD Stochastic Gradient Descent
SSIM Structural Similarity Index

References

1. Lahoz, William A., and Philipp Schneider. "Data assimilation: making sense of Earth Observation." Frontiers in Environmental Science 2 (2014): 16.
2. Cheng, Sibo, et al. "Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review." IEEE/CAA Journal of Automatica Sinica 10.6 (2023): 1361-1387.
3. Bannister, Ross N. "A review of operational methods of variational and ensemble-variational data assimilation." Quarterly Journal of the Royal Meteorological Society 143.703 (2017): 607-633.
4. Reichle, Rolf H., Dennis B. McLaughlin, and Dara Entekhabi. "Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications." IEEE Transactions on Geoscience and Remote Sensing 39.8 (2001): 1708-1718.
5. Cheng, Sibo, et al. "Error covariance tuning in variational data assimilation: application to an operating hydrological model." Stochastic Environmental Research and Risk Assessment 35.5 (2021): 1019-1038.
6. Alvarado-Montero, R., et al. "Multi-parametric variational data assimilation for hydrological forecasting." Advances in Water Resources 110 (2017): 182-192.
7. Cheng, Sibo, et al. "A graph clustering approach to localization for adaptive covariance tuning in data assimilation based on state-observation mapping." Mathematical Geosciences 53.8 (2021): 1751-1780.
8. Cheng, Sibo, and Mingming Qiu. "Observation error covariance specification in dynamical systems for data assimilation using recurrent neural networks." Neural Computing and Applications 34.16 (2022): 13149-13167.
9. Cheng, Sibo, et al. "Generalised latent assimilation in heterogeneous reduced spaces with machine learning surrogate models." Journal of Scientific Computing 94.1 (2023): 11.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_23

10. May, Robert, Graeme Dandy, and Holger Maier. "Review of input variable selection methods for artificial neural networks." Artificial neural networks-methodological advances and biomedical applications 10.1 (2011): 19-45.
11. Matthews, Gwyneth, et al. "Evaluating the impact of post-processing medium-range ensemble streamflow forecasts from the European Flood Awareness System." Hydrology and Earth System Sciences 26.11 (2022): 2939-2968.
12. National River Flow Archive Homepage, https://nrfa.ceh.ac.uk/. Last accessed 7 Feb 2024
13. Li, Mu, et al. "Efficient mini-batch training for stochastic optimization." Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014.
14. Johnson, Olanrewaju V., et al. "ps-CALR: Periodic-Shift Cosine Annealing Learning Rate for Deep Neural Networks." IEEE Access (2023).
15. Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63775-9_23