Journal of Hydrology 624 (2023) 129944

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

A surrogate modeling method for distributed land surface hydrological models based on deep learning

Ruochen Sun [a,b,c], Baoxiang Pan [d], Qingyun Duan [a,b,c,*]

[a] The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing, China
[b] College of Hydrology and Water Resources, Hohai University, Nanjing, China
[c] China Meteorological Administration Hydro-Meteorology Key Laboratory, Hohai University, Nanjing, China
[d] Institute of Atmospheric Physics, Chinese Academy of Science, Beijing, China

ARTICLE INFO

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Qiuhua Liang, Associate Editor

Keywords:
Surrogate model
Convolutional neural network
LSTM
VIC model
Soil moisture

ABSTRACT

Surrogate modelling methods have been used to estima[illegible]tially distributed parameters of land surface hydrological models due to their computational efficie[illegible]ver, traditional surrogate techniques have trouble in modelling high-dimensional input–output maps. [illegible]dition, their applications to a distributed model over large spatial and temporal domains are very unlikely to produce consistent spatial pattern and dynamic relationship of model input and output, leading to unbalanced model performance across different spatiotemporal domains. To address these issues, we employ a deep autoregressive neural network to construct an accurate and reliable surrogate system of distributed dynamic model outputs. The network uses a deep convolutional encoder-decoder architecture to take full advantage of the deep convolutional layers in high-dimensional image processing and spatial feature extraction. Moreover, the autoregressive strategy makes the network good at handling dynamic relationship between the time-varying model input and output. We apply this deep learning (DL) network to building a surrogate model of the Variable Infiltration Capacity (VIC) model simulated soil moisture over the Huaihe river basin of China. The results show that the adopted deep autoregressive neural network can provide an accurate surrogate model of the high-dimensional dynamic relationship between 15 input fields and 1 output field, each covering 408 grids, over multiple years. This surrogate model significantly outperforms the popular long short-term memory (LSTM) model, achieving an average mean squared error (MSE) of 0.26 and an average R^2 of 0.76 using the test datasets. The surrogate's MSE and R^2 performance represents an average improvement of 43% and 33%, respectively, compared to the LSTM model. In addition, it shows better spatial performance than the LSTM model, demonstrating its superior performance in approximating spatial patterns of VIC simulated soil moisture. The accuracy of this surrogate method and its ability to handle high-dimensional data is promising for advancing parameter uncertainty quantification of distributed land surface hydrological models, which usually have high dimensionality due to the large number of variables involved, including distributed parameters, input forcing variables, and output variables.

1. Introduction

Land surface hydrological models have progressed to a stage where they can represent land surface and hydrological processes in a spatially detailed manner. Such models are important tools to provide insights into the water and energy balances as well as the spatiotemporal heterogeneity of multiple energy fluxes and hydrological states. Spatially explicit simulations and predictions offered by the distributed models have been widely applied in flood forecasting, drought monitoring, policy making, climate impact studies on spatial scales from regional to even global (Cuo et al., 2013; Gou et al., 2021; Hao et al., 2017; Kauffeldt et al., 2016).

The land surface hydrological models inherently involve some degree of uncertainty. These models usually involve a large number of spatially-varying parameters which can significantly affect the simulation and prediction accuracy (Lohmann et al., 2004). In practice, direct measurement of the parameters is infeasible and their default values based on land surface characteristics are not always appropriate (Huang

* Corresponding author at: The National Key Laboratory of Water Disaster Prevention, College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China.
E-mail address: qyduan@hhu.edu.cn (Q. Duan).

https://doi.org/10.1016/j.jhydrol.2023.129944
Received 9 February 2023; Received in revised form 16 May 2023; Accepted 8 July 2023
Available online 19 July 2023

et al., 2016; Rosero et al., 2010). Therefore, improving the accuracy and reliability of the models usually relies on parameter uncertainty quantification (UQ), which can help to identify which parameters have the most impact on model predictions, and can provide estimates of these parameters based on observed data. Using mathematical algorithms for sensitivity analysis and parameter optimization has a long history in hydrological community. However, traditional methods, especially automatic optimization algorithms often require up to tens of thousands of model runs, which makes themselves computationally impractical for modern land surface hydrological models (Gong et al., 2015). As an alternative to alleviate the computational burden, much less computationally expensive surrogate models can be employed to approximate and replace the original model. Therefore, surrogate methods have become increasingly popular in the fields of hydrology and water resources over the last decade (Asher et al., 2015; Garzón et al., 2022; Razavi et al., 2012). Our group has developed a series of surrogate model based optimization methods, including the ASMO (Adaptive Surrogate Modeling based Optimization) method for single objective optimization (Wang et al., 2014), the Multi-objective ASMO (MO-ASMO) method (Gong et al., 2016), the MO-ASMO based on grid sampling (MO-ASMOGS) which is designed for calibration of distributed models (Sun et al., 2021a), the MO-ASMO for constrained hybrid problems (MO-ASMOCH) suited for optimization problems with various kinds of decision variables and different constraints (Sun et al., 2022), the ASMO-parameter optimization and distribution estimation (ASMO-PODE) which is designed for estimation of the posterior parameter distribution (Gong and Duan, 2017). In addition, many other studies have also applied surrogate methods to model calibration, sensitivity and uncertainty analysis of land surface hydrological models (Gan et al., 2018; Gou et al., 2020; Huang et al., 2016; Lu et al., 2018; Ricciuto et al., 2018).

When it comes to parameter UQ of distributed land surface hydrological models, the estimation of distributed parameters remains a significant obstacle, particularly in large-scale hydrologic modelling (Archfield et al., 2015; Clark et al., 2016), owing to the high dimensionality that stems from the heterogeneous nature of the parameters involved (Pokhrel and Gupta, 2010). Generally, estimating spatially distributed parameters requires some form of spatial regularization or parameter regionalization, which can reduce the dimensionality. Spatial regularization usually takes two types based on either the use of spatially constant parameter multipliers applied to a priori parameter fields or the application of spatially constant coefficients to transfer functions, which relate geophysical attributes to model parameters (Mizukami et al., 2017). The global multipliers or coefficients to transfer functions can then be used to build surrogate models (Gou et al., 2021; Sun et al., 2021a). However, relying on spatially constant values in surrogate modelling and calibration may lead to disparate, discontinuous or even contradictory effects in model predictions across various spatial domains. In other words, one of the biggest challenges often encountered in distributed model calibration is that improvements in one simulation region often led to degradations of performance for others (Jiang et al., 2020; Sun et al., 2021a; Yang et al., 2019). Therefore, incorporating distributed parameter fields directly into surrogate modelling and further calibration remains a problem that needs to be addressed.

The surrogate models used in the aforementioned parameter UQ studies are usually response surface models, which use data-driven functions to relate model outputs or objective functions of interest to the model parameters. The widely used data-driven models include multivariate adaptive regression splines, Gaussian process, radial basis functions, polynomial chaos expansion and artificial neural networks. However, almost all of the above-mentioned surrogate methods have trouble in effectively modelling high-dimensional input–output maps (Lu and Ricciuto, 2019; Razavi et al., 2012; Zhu et al., 2019). In such cases, the primary problem is that the number of training samples required to build such approximation models can be excessively large. Training a surrogate model with too many samples can be very time consuming and obtaining these samples means equal number of original model evaluations. Therefore, these surrogate modelling methods become less attractive or even infeasible when dealing with high-dimensional problems (Razavi et al., 2012). One way of reducing the computational cost of training a surrogate model is using adaptive sampling strategies, where the training samples for surrogate model construction are adaptively selected (Gong et al., 2016; Ma et al., 2022; Müller et al., 2015; Wang et al., 2014). Such adaptive approaches can somewhat reduce the number of original model simulations, but the improvement is relatively marginal in problems of high-dimensional input–output maps (Sun et al., 2022). Another approach to partially alleviate the high-dimensional issue is applying dimensionality reduction techniques to model parameters (Razavi et al., 2021) or model outputs (Lu and Ricciuto, 2019).

Recently, deep learning (DL) has gained comprehensive attention across scientific disciplines, including the geophysical community (Pan et al., 2021; Reichstein et al., 2019; Yu and Ma, 2021). A review of DL in hydrology and water resources can be found in Shen (2018). Due to the multi-layered architectures, deep networks have the capacity to represent extremely complex functions and capture spatiotemporal features hidden in the big data. Furthermore, modern training techniques and the fast development of parallel processing power of graphical processing units (GPUs) make the training of the deep networks feasible (Shen et al., 2018). For surrogate modelling, the DL models have great potential to provide accurate approximations for highly nonlinear functions through a hierarchy of hidden layers of flexible structure. Therefore, DL models can handle the curse of dimensionality well based on a series of nonlinear transformations of the input into low-dimensional feature maps.

There have been some studies on using DL models for surrogate modelling in the hydrology community. For example, using DL models for surrogate modelling in uncertainty quantification of groundwater models are increasing (Mo et al., 2019a; Mo et al., 2019b; Tripathy and Bilionis, 2018; Zhu and Zabaras, 2018). In contrast, the studies of using DL models as surrogates for distributed land surface hydrological models are relatively few. Two successful applications of the long short-term memory (LSTM) model to the VIC (Variable Infiltration Capacity) model (Liang et al., 1994) were reported recently (Gu et al., 2020; Tsai et al., 2021). Tsai et al. (2021) proposed a novel differentiable parameter learning framework to learn the spatially-varying parameters of VIC, in which the differentiable feature of the LSTM surrogate played a key role. Although LSTM has been very popular in the hydrological community for time series prediction, it may not perform well in spatiotemporal sequence prediction problems (Shi et al., 2015). In other words, the grid-level training mode of LSTM may not guarantee the accurate representation of spatial patterns, which is a unique and relevant source of information for distributed land surface hydrological models (Dembélé et al., 2020). If the surrogate model fails to accurately approximate the spatial response of the original model, it may result in an imbalanced performance during the calibration process.

In summary, traditional surrogate methods face challenges in addressing distributed calibration of land surface hydrological models due to the difficulties in handling high dimensionality, spatial pattern, and dynamic change of model input–output relationships. To address these issues, we use an image-to-image regression strategy to produce a reliable and accurate surrogate model. We try to mimic the preparation and simulation processes of the original model, in which the distributed meteorological forcings of each time step as well as the parameter fields are provided to simulate the water and energy fluxes of this time step. In addition, the state variables of last time step are also used to simulate themselves of the current time step. Therefore, both the spatial pattern and dynamic feature of the model response could be well approximated by the surrogate model. To implement the idea, we adopt a dense convolutional network (DenseNet, Huang et al., 2017) based deep autoregressive neural network (ARnet hereafter). The network takes full

advantage of the dense convolutional layers and the encoder-decoder architecture in high-dimensional image processing and spatial feature extraction. In addition, the autoregressive strategy makes it good at handling dynamic relationship between the time-varying model input and output. We use this DL network to build a surrogate system of VIC simulated soil moisture over the Huaihe river basin of China. The results will show that the proposed DL network can effectively serve as an accurate surrogate, even with input dimensions as high as 6120. In addition, it significantly outperforms the popular LSTM surrogate in approximating spatial patterns of model outputs.

To the best of our knowledge, ARnet has not been previously applied in the surrogate modelling of distributed land surface hydrological models. By introducing this DL network, the surrogate model can directly incorporate high-dimensional parameter fields, without requiring the sampling of grids from the domain as is necessary with the widely used LSTM. Additionally, the ARnet model is highly suited for reproducing spatial patterns of the model outputs, which are often ignored in previous surrogate modelling methods. Therefore, the application of the ARnet model in this context is novel and it has great potential for advancing distributed calibration of land surface hydrological models.

The following section mainly describes the DL networks used in this study (section 2). Section 3 gives a brief introduction of the study area, data and evaluation methods. Section 4 presents detailed results of the performance of two different DL networks, followed by discussion in Section 5. Section 6 gives the conclusions of this study.

2. Methodology

2.1. Land surface hydrological model

This study aims to build an accurate and efficient surrogate system of a distributed land surface hydrological model outputs. The wildly used VIC model is selected to test the proposed DL-assisted surrogate modelling method. VIC (Liang et al., 1994) is a distributed, grid-based, macroscale model that solves full water and energy balances. The grid cells of a study region are simulated independently of each other, there is no communication between grid cells. From the state-space point of view (DeChant and Moradkhani, 2014), the reduced structural equation of the VIC model and other distributed hydrological models can be written, in a generic sense, as:

$$s_{i,t} = f\left(x_{i,t}, s_{i,t-1}, \theta_i\right) \tag{1}$$

where $f()$ represents the model structure, $x_{i,t}$ is the forcing vector of the i th grid at time t, $s_{i,t}$ represents the model state vector of the i th grid at time t, θ_i is the parameter vector of the i th grid. Therefore, the land–atmosphere fluxes, and the water and energy balances of each grid cell are simulated at a prescribed time step with the meteorological forcings, parameters and initial conditions. In this study, the latest VIC 5.1 version (Hamman et al., 2018) is used.

2.2. LSTM

The LSTM network is a special kind of recurrent neural network (RNN) capable of learning long-term dependencies. It was first introduced by Hochreiter and Schmidhuber (1997), and were refined and popularized by many people in following work. In contrast to other types of RNNs, LSTM does not have a problem with exploding or vanishing gradients. The key to the LSTM model is a memory cell known as a cell state that maintains its state over time. Information can be added to or removed from the cell state in LSTM and is regulated by gates (input gate, forget gate and output gate). These gates optionally let the information flow in and out of the cell. Therefore, the cell states are analogous to state vectors in a traditional dynamic model, which makes LSTM potentially suitable for modeling hydrological processes like streamflow, snowmelt, soil moisture which have relatively long timescales (Fang et al., 2019; Kratzert et al., 2018; Kratzert et al., 2019).

Fig. 1 gives a schematic diagram of LSTM. LSTM is capable of processing the entire sequence of data. Given an input sequence $x = (x[1], x[2], \cdots x[T])$ with T time steps, where each element $x[t]$ is a vector containing input features at time t, then the cell state $C[t]$ and the hidden state $H[t]$ at time step t are calculated as follows:

$$C[t] = C[t-1] \odot f[t] + g[t] \odot i[t] \tag{2}$$

$$H[t] = o[t] \odot \tanh(C[t]) \tag{3}$$

where $i[t]$, $f[t]$ and $o[t]$ are the input gate, forget gate, and output gate, respectively, $g[t]$ represents the input node, $\tanh()$ is the hyperbolic tangent function, $\odot$ denotes the elementwise product operator. $i[t]$, $f[t]$, $o[t]$ and $g[t]$ are calculated as follows:

$$i[t] = \sigma(x[t]W_{xi} + H[t-1]W_{hi} + b_i) \tag{4}$$

$$f[t] = \sigma(x[t]W_{xf} + H[t-1]W_{hf} + b_f) \tag{5}$$

$$o[t] = \sigma(x[t]W_{xo} + H[t-1]W_{ho} + b_o) \tag{6}$$

$$g[t] = \tanh(x[t]W_{xc} + H[t-1]W_{hc} + b_c) \tag{7}$$

where $\sigma()$ is the sigmoid function, W_{xi}, W_{hi}, W_{xf}, W_{hf}, W_{xo}, W_{ho}, W_{xc}, W_{hc} are weight parameters, b_i, b_f, b_o, b_c are bias parameters.

Generally speaking, the data feeding into the LSTM gates are the input at the current time step and the hidden state of the previous time step. As a result of the sigmoid activation, all values of the three gates are in the range of $(0,1)$. Intuitively, the input gate determines how much of the input node's value should be added to the current memory cell internal state. The forget gate determines whether to keep the current value of the memory or flush it. And the output gate determines whether the memory cell should influence the output at the current time step.

2.3. Deep autoregressive neural network

In this study, we employ a DenseNet based deep autoregressive neural network ARnet (Fig. 2) to mimic the modelling processes of the distributed land surface hydrological models such as VIC. The adopted network architecture is inspired by Mo et al. (2019a), which has shown a great potential in efficiently obtaining accurate approximations of a synthetic contaminant transport model with time-varying inputs and time-dependent outputs. In comparison, the land surface hydrological models are much more complex, thus the nonlinear mapping between high-dimensional inputs and outputs are more difficult to model. Therefore, it is worthy to study the feasibility and applicability of the deep autoregressive neural network for surrogate modelling of surface hydrological system. We will also compare it with the widely used LSTM to see if this deep autoregressive network could produce better spatial predictability, which is of great importance to modelers when applying a distributed model.

The core idea of using the deep autoregressive neural network is to convert the surrogate modelling of a distributed model to an image-to-image regression task to make full use of the advantages of convolutional neural networks (CNNs) in image processing. Assume that the simplified state-space equation of a distributed hydrological model (VIC in this study) defined in section 2.1 is solved over a spatial domain of size $n_x \times n_y$, where n_x and n_y denote the number of grid cells along the horizontal direction (longitude) and the vertical direction (latitude) of the domain. The input of the deep autoregressive network is M images corresponding to meteorological forcings, parameter maps and model states, while the output is N images corresponding to model outputs. Therefore, the input–output mapping can be described by the network as:

$$f' : R^{M \times n_x \times n_y} \rightarrow R^{N \times n_x \times n_y} \tag{8}$$

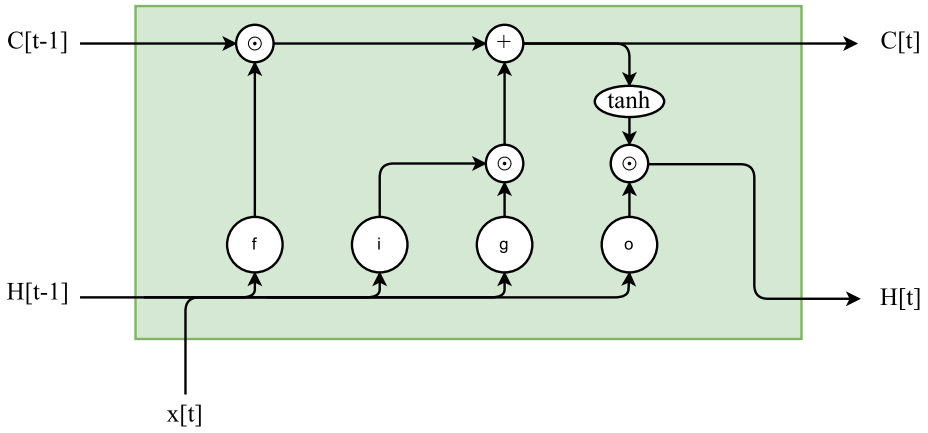

Fig. 1. Schematic diagram of a memory cell of LSTM.

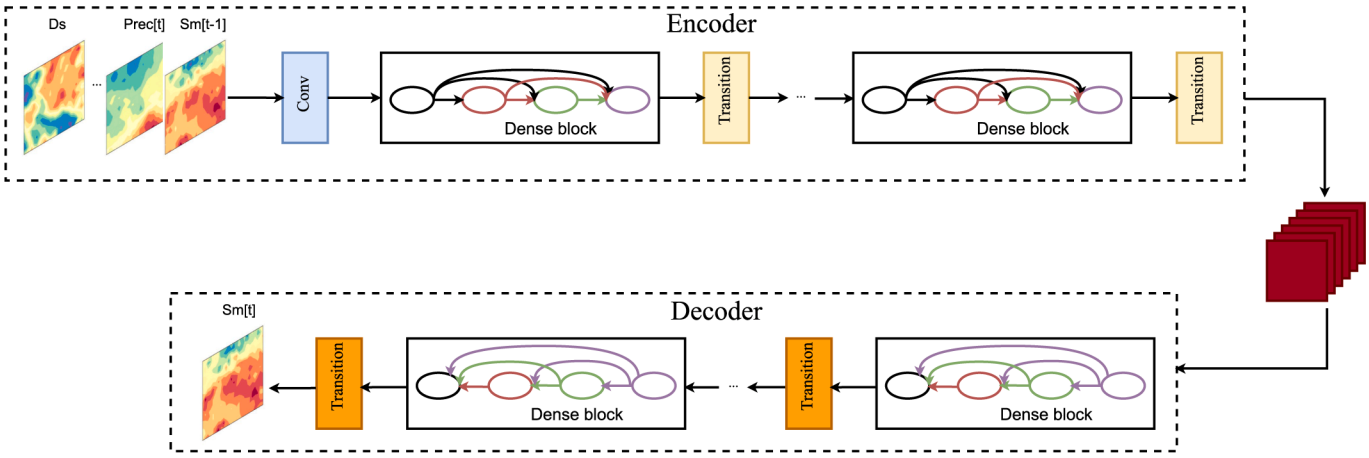

Fig. 2. Overview of the network architecture of the deep autoregressive neural network.

It is easily found that the original model and the surrogate model have similar forms of processing input data to output data. In another word, modelers need to provide distributed input data (spatial images) to a land surface hydrological model when doing regional modelling and they can get distributed model output, the same applies to the deep autoregressive neural network based surrogate model.

The above-mentioned image-to-image regression problem can be efficiently solved by using a dense convolutional encoder-decoder network (Zhu and Zabaras, 2018). In this network architecture, the high-dimensional image regression is performed through a coarse-refine process. In other words, high-level coarse features from the input images are firstly extracted using an encoder and then refined to output images through a decoder. In addition, to avoid using massive training data, a dense convolutional network called DenseNet (Huang et al., 2017) is employed without any fully connected layers. Compared to traditional convolutional networks, DenseNet is distinguished by its connectivity pattern, where each layer is connected to all the preceding layers, as well as its use of concatenation operations to retain and reuse features from previous layers, thereby reducing the number of parameters that need to be learned. This helps the network to generalize better to new data and reduces the likelihood of overfitting, which can occur when there is not enough training data to accurately learn the underlying patterns. The main components of a DenseNet are dense blocks and transition layers. In the dense block, the outputs of any layer are directly connected to all subsequent layers. As a result, the l th layer receives the feature maps of all preceding layers in the dense block as input, that is, $x_l = H_l([x_0, x_1, \cdots x_{l-1}])$, where $[x_0, x_1, \cdots x_{l-1}]$ denotes the concatenation of the output features produced in layers $0, \cdots, l-1$, H_l is a composite non-linear transformation function of operations. In this study, H_l consists of three consecutive operations: batch normalization (BN), rectified linear units (ReLU), and convolution (Conv). There are two parameters determining the structure of a dense block, namely, the number of layers within the block (L) and the growth rate k, which is the number of feature maps produced by each layer. Thus the l th layer within a dense block has $k_0 + k \times (l-1)$ input features. The transition layer is referred to as the layer between two dense blocks, which controls the number of feature maps so that it is not too large, since the expansion of x_l can be quite high-dimensional. In addition, it also helps to change the size of feature maps through down-sampling in the encoder and up-sampling in the decoder. The transition layers used in our study consist of BN, ReLU, Conv or transposed Conv. Interested readers can refer to the original paper of DenseNet (Huang et al., 2017) for more details.

Fig. 2 shows the DenseNet based dense convolutional encoder-decoder network architecture. Basically, it consists of two deep DenseNet networks which are called encoder and decoder respectively. The encoder-decoder architecture is a widely used structure in deep learning. Encoder-decoder architectures can handle inputs and outputs which both consist of variable-length sequences and thus are suitable for sequence-to-sequence problems. In our network, the encoder is used to extract high-level feature maps from the model input data, which are then passed through a decoder to produce output images.

For developing surrogate models of dynamic systems, it is important to allow prediction at any time steps. The metrological forcings and output of the land surface hydrological models are time-varying, hence it is important for the surrogate model to capture the dynamic response between the input and output. Based on the state-space framework in Eq.

(1), we think that representing this dynamic system in an autoregressive manner could be appropriate. In another word, the current model states depend on the meteorological forcings of the current time step, the time-independent parameters and the model states of the previous time step, which contains the information about the previous inputs and states. Therefore, one forward model run with T time steps can produce T training samples.

3. Experimental design

3.1. Study region and data

The study area is similar to that of our previous study (Sun et al., 2016), which is the upper basin of Bengbu hydrologic station over the Huaihe River basin. The difference is that the study area in this study is expanded to a rectangle region (30.75°-35°N, 111.75°-117.75°E) which exactly covers the upper Huaihe River basin (Fig. 3). The average annual precipitation is approximately 888 mm, of which 50–75% falls during the warm season (June–September) influenced by the East Asian summer monsoon. In this study, the VIC model is operated with a spatial resolution of 0.25°× 0.25°, thus the study region has dimensions of 24 × 17.

The data used in this study are mainly for driving the VIC model. For the meteorological forcings, the precipitation data is the CMORPH (Climate Precipitation Center Morphing) gauge-satellite merged precipitation analysis over China (Shen et al., 2014). This dataset has a spatial–temporal resolution of 0.1°, hourly. Other required meteorological variables as inputs come from the China Meteorological Forcing Dataset (CMFD) dataset with a temporal resolution of three hours and a spatial resolution of 0.1° (He et al., 2020). All the meteorological variables are processed to have a temporal resolution of three hours and a spatial resolution of 0.25°. In order to obtain distributed parameter maps, this study uses the MPR (Multiscale Parameter Regionalization) method (Samaniego et al., 2010), which links model parameters with related geophysical features at the fine scale through transfer functions and then upscales the parameters to the desired model resolution. Instead of determining the transfer function coefficients by calibration, we randomly sample different coefficients within their ranges to get many parameter maps for surrogate model training. The geophysical features are from the China dataset of soil hydraulic parameters (Dai et al., 2013) and the Soil Database of China for Land Surface Modeling (Shangguan et al., 2013). The transfer functions and upscaling operators used in VIC model are listed in Gou et al. (2021).

3.2. DL networks

In this study, we use the VIC model to simulate daily soil moisture in the study region from 2008 to 2015, and we try to build surrogate models for the daily input–output relationship. The common input of the two DL based surrogate models consists of 7 meteorological forcing variables of VIC (precipitation, air temperature, atmospheric pressure, incoming shortwave radiation, incoming longwave radiation, vapor pressure, wind speed) plus 7 model parameters including variable infiltration curve parameter (b), maximum velocity of baseflow (Dsmax), fraction of Dsmax where non-linear baseflow begins (Ds), fraction of maximum soil moisture where non-linear baseflow occurs (Ws), the second soil layer thickness (d2), the third soil layer thickness (d3), the second soil layer drainage parameter (E2). For the ARnet model, the soil moisture of last day is also used as input. Table 1 gives a summary of the input and output of the two DL based surrogate models.

We randomly generate 1500 distributed fields of the 7 parameters using the above-mentioned transfer functions with randomly sampled function coefficients, and then run the VIC model 1500 times. The model data are split into training data pool (1200 input–output samples), validation data pool (200 input–output samples) and testing data pool

Table 1
The input and output of the ARnet and LSTM based surrogate models.

Surrogate model	Input	Output
ARnet	$\{x_{D,t}\}_{i=1}^{7}, s_{D,t-1}, \{\theta_D\}_{j=1}^{7}$	$s_{D,t}$
LSTM	$\{x_{g,t-lag+1}\}_{i=1}^{7}, \cdots, \{x_{g,t}\}_{i=1}^{7} \{\theta_g\}_{j=1}^{7}$	$s_{g,t-lag+1}, \cdots, s_{g,t}$

$\{x\}_{i=1}^{7}$ denotes the i th meteorological variable, $\{\theta\}_{j=1}^{7}$ denotes the j th parameter, s denotes soil moisture. The subscript D represents the study domain and the subscript g represents one grid cell. The subscripts t and lag denote the t th time step and the sequence length respectively.

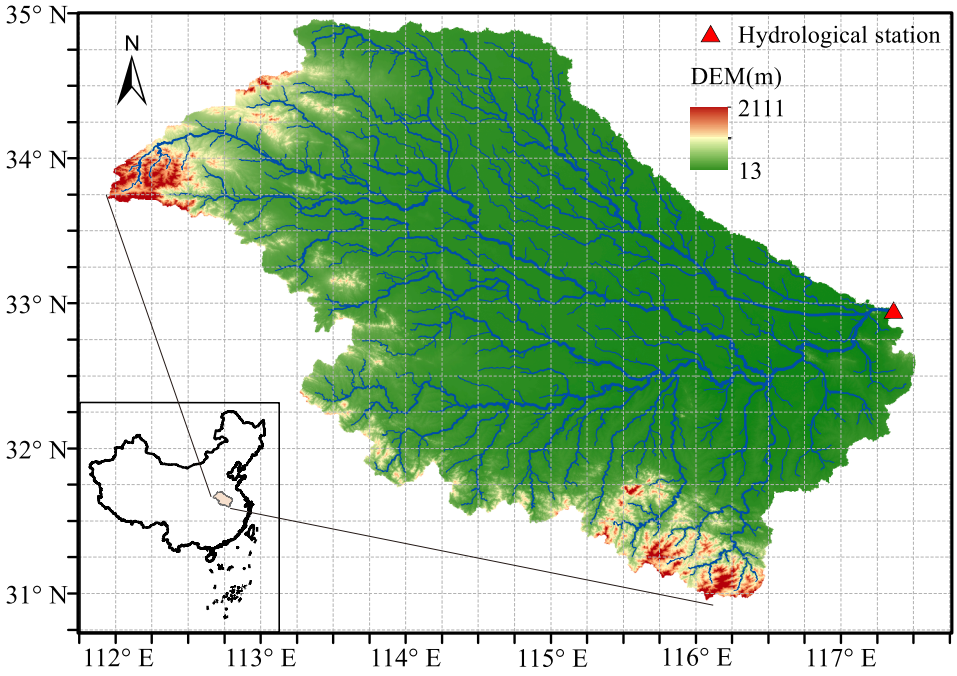

Fig. 3. The rectangular study area covering the upper Huaihe river basin. Each small box formed by the intersected dashed lines represents a 0.25°× 0.25° grid cell.

(100 input–output samples). In addition, the time periods of 2009–2010, 2013–2014 and 2015 are set for training, validation and testing of the deep neural networks, respectively. The validation dataset is used to perform early stopping, which monitors the performance of the deep learning models on a held-out validation set during the training and terminate the training conditional on the validation performance to avoid overfitting. The test dataset is used to evaluate the performance of the trained models independently.

For the ARnet model, the input should be images, we first assess how the size of training samples impacts the performance of the ARnet model. We generate four training sample sets with sizes of 100, 400, 800, 1200 from the training data pool, and 200 model evaluations from the validation data pool to train the model. The results which will be described in section 4 show that the ARnet model achieves relatively good performance with 400 training samples. Therefore, the training sample size is set to 400 for further comparison and evaluation. In our ARnet model, the encoder has two dense blocks with 5 and 10 internal layers, and the decoder has one dense block with 5 layers. The growth rate k is set to 40. The convolution kernels are the same for all convolutional layers within the same dense block. The first convolutional layer shown in Fig. 2 has 48 filters, resulting in 48 feature maps. Other detailed information about the ARnet is shown in Table 2. We use batch size of 200 and mean squared error (MSE) loss.

The ARnet model considers the soil moisture of past day as input during training. Once the model is trained and used to simulate soil moisture for a desired time period (the test period in this study), sequential prediction of the model output can be performed from t_1 to t_n with the forcings of each time step, parameter fields and an initial soil moisture field. This means that the soil moisture simulated by ARnet at time t_{n-1} will be used as an input to simulate soil moisture at time t_n. Considering that the primary role of the surrogate model is to reflect the relationship between model outputs and different parameters, which facilitates subsequent parameter UQ studies, we can use VIC simulation results obtained using default parameters as inputs to the ARnet model. This means that the soil moisture simulated by VIC with default parameters at time t_{n-1} will be used as an input to simulate soil moisture at time t_n. In this study, we will assess the two ARnet simulation modes, which are denoted as ARnet_WV (with VIC) and ARnet_NV (no VIC), using the test datasets.

For the LSTM model, the input–output data pair of each grid cell can be used for training. The LSTM model was trained using the data of 400 model evaluations from the training data pool and 200 model evaluations from the validation data pool, consistent with the data used in ARnet. Therefore, we actually have 400 × 408 (grid cells) = 163200 grid cells with 2 years of data for training. To determine the hyperparameters, which are the number of memory cells and the length of the input sequence in our study, we performed a grid search over a range of parameter values, following a similar approach to Kratzert et al. (2019). Finally, the LSTM has 256 memory cells and is run in sequence-to-sequence mode with sequence length (look-back window size) of 30 days. We use batch size of 400 and MSE loss.

For both the LSTM and ARnet training, we set the maximum training epoch to 200 to ensure that early stopping occurs. The widely used Adam algorithm (Kingma and Ba, 2014) is used for parameter learning of the neural networks. The models are trained on a NVIDIA GeForce RTX 3090 GPU with 24 GB video memory.

Table 2
The network architecture of ARnet.

Layers	Description	Output shape ($O \times H \times W$)[a]
Input	Input images	15 × 17 × 24
Convolution	48 filters, kernel 7 × 8 (s2p3)[b]	48 × 9 × 12
Dense block	L5k40, kernel 3 × 3 (s1p1)	248 × 9 × 12
Transition layer	124 filters, kernel 1 × 1 (s1p0) 124 filters, kernel 3 × 4 (s2p1)	124 × 5 × 6
Dense block	L10k40, kernel 3 × 3 (s1p1)	524 × 5 × 6
Transition layer	262 filters, kernel 1 × 1 (s1p0) 262 filters, kernel 3 × 4 (s2p1), transposed conv	262 × 9 × 12
Dense block	L5k40, kernel 3 × 3 (s1p1)	462 × 9 × 12
Transition layer	231 filters, kernel 1 × 1 (s1p0) 1 filter, kernel 3 × 4 (s2p1) transposed conv	17 × 24

[a] O denotes the number of output feature maps, $H \times W$ is the dimension size of the feature map.
[b] s and p represent stride and padding of the convolutional layer.

3.3. Evaluation methods

Land surface hydrological models tend to be simulated on a regional or even global scale with many grid cells for several years. Therefore, we usually need to build a surrogate system over large spatial and temporal domains. However, traditional evaluation of distributed modelling remains focused on pixel-wise or spatially aggregated metrics. The neglect of a specific focus on spatial patterns in distributed model evaluation is a paradox considering an increasing acknowledgement of the role of spatial patterns in the hydrological systems (Mendiguren et al., 2017). Therefore, it is highly necessary to use a variety of performance metrics to comprehensively evaluate the model performance.

The evaluation methods used in our study not only include classical verification metrics like the coefficient of determination (R^2), correlation coefficients (CC), Kling-Gupta efficiency (KGE; Gupta et al., 2009), but also consists of two bias-insensitive spatial performance metrics, namely empirical orthogonal function (EOF) analysis and the fractions skill score (FSS). Here, we provide a concise overview of two spatial performance metrics.

The EOF method is a popular technique used in fields such as atmosphere, climate, ocean, and hydrology science to identify potential spatial patterns of variability and how they change over time. Koch et al. (2015) introduced a novel approach of performing a joint EOF analysis on a combined data matrix containing both reference and simulated data. This approach allows for the identification of spatiotemporal patterns of variability for both datasets, and also enables the calculation of a similarity score between them. The difference between the loadings at each time step can serve as an indicator of spatial similarity, with the loading deviation weighted according to the corresponding EOF's variance contribution to ensure a reliable pattern similarity score. The EOF-based similarity score between a reference map and a modelled map at time t can be formulated as follows:

$$S^t_{EOF} = \sum_{i=1}^{n} w_i |(load^s_i - load^o_i)| \quad (9)$$

where w_i is the variance contribution of the i th EOF, n is the total number of orthogonal modes (EOFs), $load^s_i$ and $load^o_i$ are the corresponding loadings of simulated and observed maps, respectively. Smaller score indicates a higher degree of similarity between the reference map and the modelled map.

The FSS is a scale-selective performance metric developed by Roberts and Lean (2008) for assessing the accuracy of precipitation forecasts at different spatial scales, for a given threshold. It uses the nearest neighbour approach to select the relevant scales for analysis. FSS calculates the fractional coverage of binary events that exceed the defined threshold, within a given spatial window. Typically, percentile thresholds are used to convert continuous data into binary fields to remove the impact of bias and focus on spatial accuracy. The main steps for obtaining the FSS are as follows: 1) convert the reference and modelled spatial patterns into binary fields for a specific threshold; 2) for each grid in each binary field, compute the fraction of grids with a value of 1 within a given square window of length n; 3) calculate the MSE between the referenced and modelled fraction fields; 4) obtain the final FSS by normalizing the MSE from step 3 with the largest possible MSE that can

be obtained from the modelled and referenced fractions. The FSS at a spatial scale of n, for a given threshold, can be expressed as:

$$FSS_{(n)} = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(O_{(n)i} - S_{(n)i})^2}{\frac{1}{N}(\sum_{i=1}^{N}O_{(n)i}^2 + \sum_{i=1}^{N}S_{(n)i}^2)} \quad (10)$$

where $O_{(n)}$ and $S_{(n)}$ are the resultant fields of referenced fractions and modelled fractions, respectively. N is the total number of valid grids in the domain. The FSS ranges from 0 (complete mismatch) to 1 (perfect match).

Interested readers can refer to our previous study (Sun et al., 2021b) for more details about the two spatial evaluation methods. The evaluations of the LSTM and ARnet based surrogate models are all performed using the 100 model evaluations from the test dataset.

4. Results

We first evaluate how the size of training samples impacts the performance of the ARnet model. Fig. 4 displays the MSE values for ARnet on the validation data set, with varying sizes of the training data set. The figure illustrates that even with only 400 training samples, the model achieves a relatively low MSE value of 0.047, despite the problem having 6120 (15 × 408) input dimensions. As the training sample size increases to 1200, the model's MSE value further improves to 0.043. To ensure that both the ARnet and LSTM models are trained on the same data, we ultimately set the training sample size to 400. The performance metrics of the two DL models during training have little difference (not shown). Both models show good performance in giving overall accurate prediction of soil moisture. When using the surrogate model for data–model integration studies, it is desired to analyse the surrogate accuracy independently. Therefore, the results shown later are all based on the test dataset.

Fig. 5 shows the average soil moisture values for 2015, as evaluated by VIC and two surrogate models over 100 test data across all grids. Generally, both ARnet_WV and ARnet_NV perform better than LSTM. The points of ARnet_WV basically lie along the diagonal of the figure, except that it has dispersed data points of large values. ARnet_NV has a more dispersed scatter plot with slight underestimation. In comparison, the points of LSTM are much more dispersed with a lot of overestimation and underestimation. Fig. 6 further shows the time series comparison at a randomly chosen grid cell. Although there are biases, it can be seen that LSTM modelled soil moisture has consistent trend with the VIC modelled soil moisture, meaning that LSTM can capture the dynamic characteristics of input–output relationship. What surprises us is that ARnet models also perform very well in reproducing the dynamic characteristics, since the input of this surrogate model does not explicitly contain information of time series. Despite not including any information from the VIC model, ARnet_NV does not significantly degrade the temporal performance. This indicates that the autoregressive strategy used in the ARnet plays a key role.

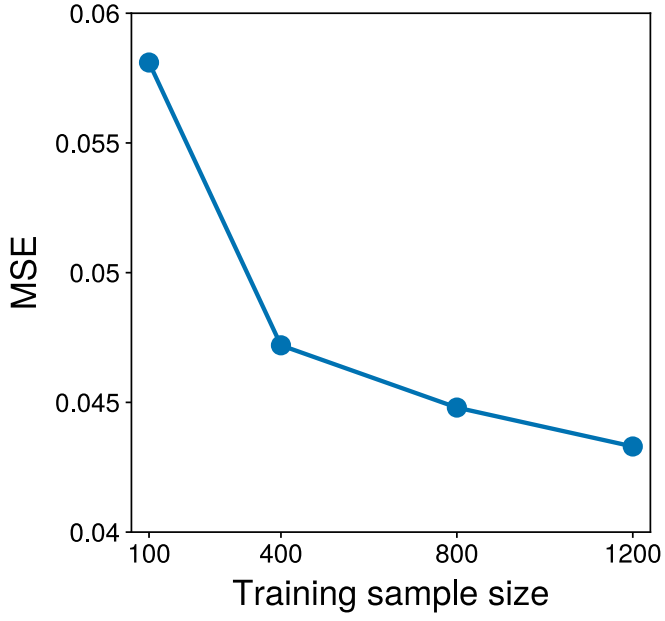

Fig. 4. Comparison of the MSE values of ARnet evaluated on validation samples with different number of training samples.

Fig. 7 shows the average KGE calculated over the entire domain in evaluating 100 test data. ARnet still performs better than LSTM with higher KGE during the testing period. Basically, the KGE values of ARnet_WV are larger than 0.5 with relatively small fluctuation, displaying a uniformly good approximation in the testing year. It also can be observed that although ARnet_NV has a poor performance during September, its KGE values increase again after this period, indicating that the sequential prediction mode of ARnet could also be reliable. Since the daily input–output relationship is modelled in this study, the variation of atmospheric drivers is large and clearly it has an impact on surrogate accuracy. However, the impact is less significant on the ARnet model, which indicates that we can expect a good approximation when using the surrogate model at any specific time for data–model analysis.

Fig. 8 depicts the distribution of the average soil moisture over 2015 of a random test sample based on VIC and two surrogate models. It can be seen that two ARnet modes have smaller approximation errors compared to those of LSTM. The LSTM cannot capture the overall pattern of VIC modelled soil moisture. ARnet modelled soil moisture is very similar to that of VIC, indicating that the ARnet model is good at learning spatial information. We further show the spatial maps of CC and KGE between the DL models predicted soil moisture and the VIC simulated soil moisture of a random test sample during the testing period (Fig. 9). There are about 95% of grid cells with CC larger than 0.7 and about 80% of grid cells with KGE larger than 0.65 for the LSTM model. This shows that the widely used LSTM has the ability to give reasonable approximation of the VIC model in most of the grid cells, however, its performance is even inferior to ARnet_NV. For ARnet_NV prediction in Fig. 9, about 97% of grid cells have CC values larger than 0.7 and about 88% of grid cells have KGE values larger than 0.65.

Since the two DL models are evaluated based on 100 test samples from the test dataset, their mean MSE and R^2 values are calculated, along with the variances. Table 3 shows that ARnet_WV has the lowest mean MSE of 0.1 with a small variance of 0.067, indicating its stable and superior performance in terms of prediction accuracy. ARnet_NV has a mean MSE of 0.26 with a variance of 0.082, which suggested a relatively stable prediction accuracy. LSTM, however, has the highest mean MSE of 0.46 with a larger variance of 0.11, implying that its prediction capability is less stable than the other two models. In terms of R^2, ARnet_WV also demonstrats the highest mean value of 0.91 with a small variance of 0.039. ARnet_NV has a mean R^2 of 0.76 with an even smaller variance of 0.037, while LSTM has the lowest mean R^2 of 0.57 with a larger variance of 0.088. Despite not using any auxiliary input from the VIC model, ARnet still achieves a significant improvement over the LSTM model in terms of both MSE (by 43%) and R^2 (by 33%).

In addition to the classical evaluations, this study also performs true spatial evaluations based on two bias-insensitive spatial performance methods, namely EOF analysis and FSS. It is necessary to perform spatial evaluations since one important objective of distributed land surface hydrologic modelling is to produce a spatially consistent hydrologic assessment. Fig. 10 presents the EOF-based similarity scores for the two surrogate models. Low values of the EOF-based metric are preferred, indicating the spatial performances of the DL model predicted soil moisture and the VIC modelled soil moisture are very similar. The time series of EOF-based scores show certain degree of fluctuation, which is less obvious in winter and much more obvious in summer. The small

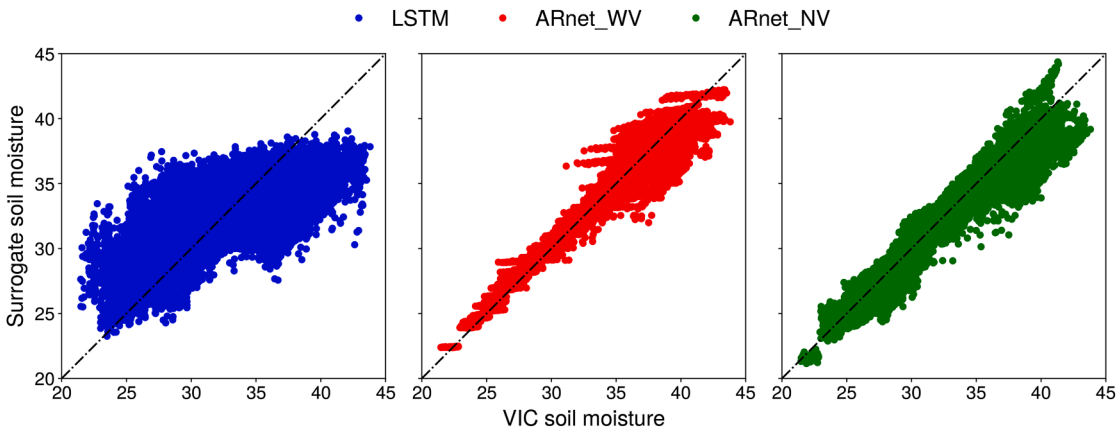

Fig. 5. Averaged daily soil moisture over 2015 simulated by VIC and two surrogate models in evaluating 100 test data for the entire domain.

spatial variability of soil moisture in winter may be the cause. It is also clear that ARnet predicted soil moisture has better spatial performance than LSTM predicted soil moisture in terms of the EOF-based metric. FSS is a scale-dependent verification method which has been widely used to evaluate precipitation forecasts. In our study, if a threshold percentile is below 50%, the method will use the grids that fall below the value (low phase). On the contrary, it will focus on grids exceeding the threshold (high phase) when it is above 50%. Fig. 11 shows the FSS curves for the average soil moisture patterns of the testing period using different percentile thresholds. Higher FSS values are preferred. At all thresholds the FSS values increase gradually with the scale used to compute the fractions getting larger, which is an essential feature of the method (Sun et al., 2021b). However, the increasing rate differs a lot among different thresholds. The differences in the spatial performance of the two surrogate models are clear. ARnet and LSTM have similar skills in capturing the patterns of low soil moisture with larger scales. LSTM performs much worse than ARnet for the high thresholds, indicating that LSTM has poor skill in reproducing the localized features of high soil moisture. In comparison, ARnet shows an excellent spatial performance. ARnet_NV shows comparable FSS values to ARnet_WV, particularly for high soil moisture thresholds. Its skill of capturing localized features is very stable, showing the great advantage of the network architecture in processing images.

5. Discussion

This study proposes a deep autoregressive neural network based surrogate method for distributed land surface hydrological modelling. This method converts the distributed surrogate modelling to an image-to-image regression task, which also mimics the modelling processes of the distributed land surface hydrological models. We compare this ARnet with the widely used LSTM model, which uses input–output data of single grid cell as a training sample.

Due to the large number of hyperparameters associated with ARnet, we have not conducted extensive experiments to optimize the options, unlike LSTM. We set the hyperparameter values mainly based on the suggestions of Zhu and Zabaras (2018). However, we assess the impact of some key hyperparameter options on model performance in this study. Table 4 presents the validation MSE values of ARnet for various hyperparameters related to the network architecture. The hyperparameters include dense block layers, growth rate, and features after the first convolutional layer. By default (hyperparameters in Table 2), the model achieves a validation MSE of 0.0433. However, decreasing the number of dense block layers to (4, 8, 4) or (3, 6, 3) resulted in slightly higher validation MSE values of 0.0475 and 0.0547, respectively. When decreasing the growth rate to 32 and 16, the validation MSE values were 0.0479 and 0.0665, respectively. Furthermore, when using 32 or 16 features after the first convolutional layer, the validation MSE values were 0.0467 and 0.0567, respectively. Further increasing the hyperparameter values would significantly increase the number of network parameters needed to be trained. The results indicate that the default hyperparameter setup of ARnet has produced a relatively good performance. The results also suggest that the choice of hyperparameters can have an impact on the performance of ARnet, and extensive experiments can be carried out to optimize the hyperparameters to achieve better results in future work.

This study sets the spatial resolution of VIC to 0.25°, thus producing a regularly discretized box with size of 24 × 17 to cover the simulation domain. Since the model resolution directly affects the input dimensions, we conduct another experiment to assess how discretization of the domain would affect the ARnet performance. In this experiment, we set the spatial resolution of VIC to 0.125°, thus increasing the box size to 48 × 34. We use the same ARnet configuration for training and testing. Fig. 12 shows the distribution of the average soil moisture over all test samples based on VIC and two ARnet modes. Similarly, ARnet is capable of reproducing the spatial distribution of VIC modelled soil moisture effectively, even with an increase in VIC resolution resulting in more spatial details. Other evaluations (not presented) have also shown similar results to those obtained at the coarse resolution. Therefore, the discretization of the study domain does not seem to have a significant impact on the accuracy of the ARnet model. However, further increasing the input dimensions of ARnet results in a nearly 4.5-fold increase in training time.

The results show that the ARnet has a very distinct advantage over LSTM of spatial performance. This is attributed to the core idea of image-to-image regression used in the ARnet. The spatial features and information hidden in the input images can be extracted and handled well through a series of dense blocks and transition layers due to their inherent convolutional structure. In comparison, the LSTM model does not take spatial correlation into consideration. People may argue that LSTM can maintain the spatial pattern as the input–output data of different grid cells used for training have already contained spatial information. However, the spatial continuity and consistency cannot be guaranteed without considering spatial information explicitly. The spatial evaluation methods have clearly shown the deficiency of LSTM in spatial performance. Shi et al. (2015) also showed that LSTM performed not well in spatiotemporal sequence forecasting problems and the improved ConvLSTM (convolutional LSTM) network captures spatiotemporal correlations better.

The use of the autoregressive strategy in the ARnet enables the network to well capture the dynamic relationship between the time-varying inputs and outputs. The model states of the previous time step used as input contain information about the previous inputs and states, which play a similar role as the memory cell in LSTM. In this study, we

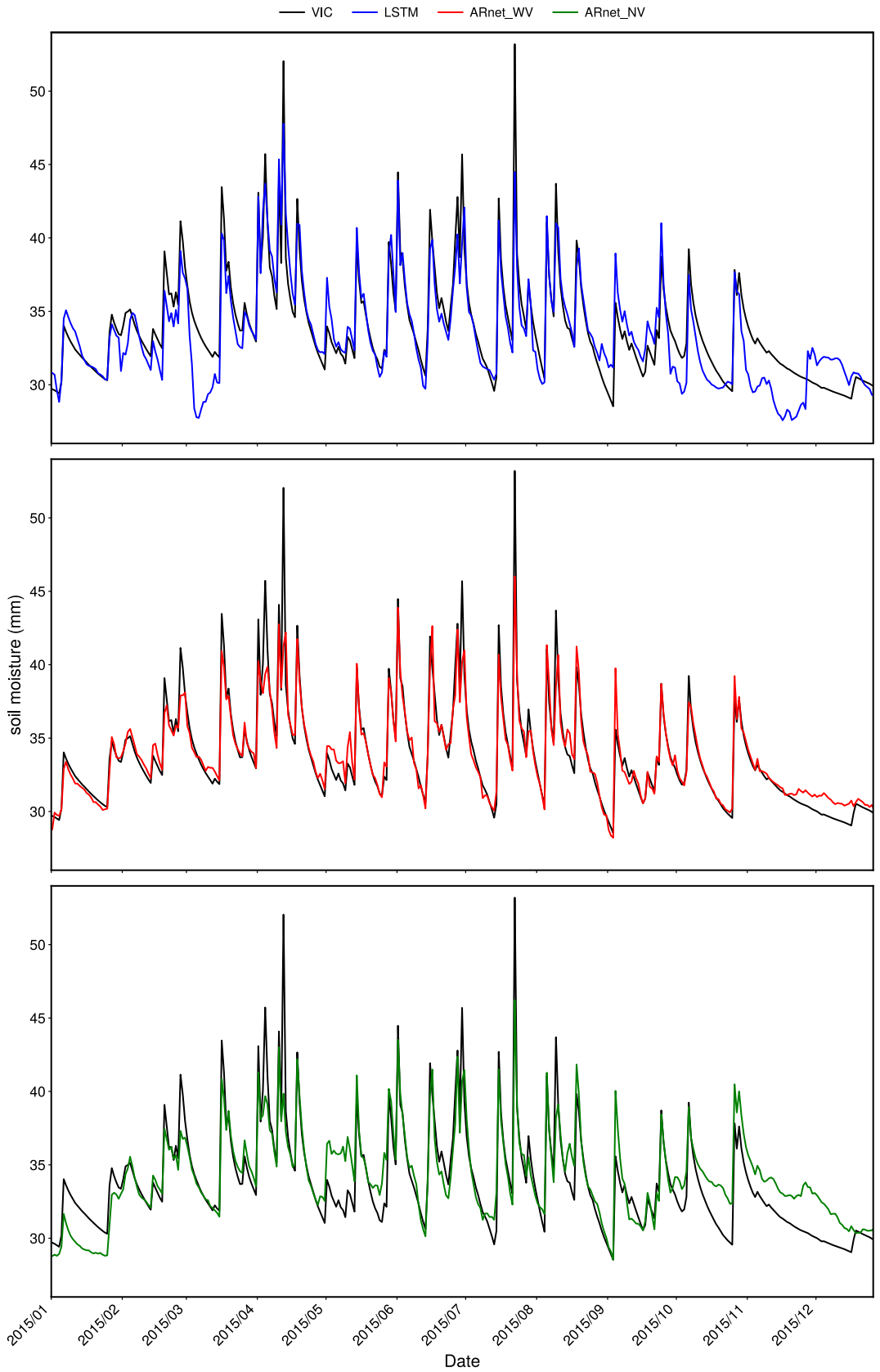

Fig. 6. Time series comparison of daily soil moisture simulated by VIC and two surrogate models of a random test sample in 2015 at a randomly chosen grid cell.

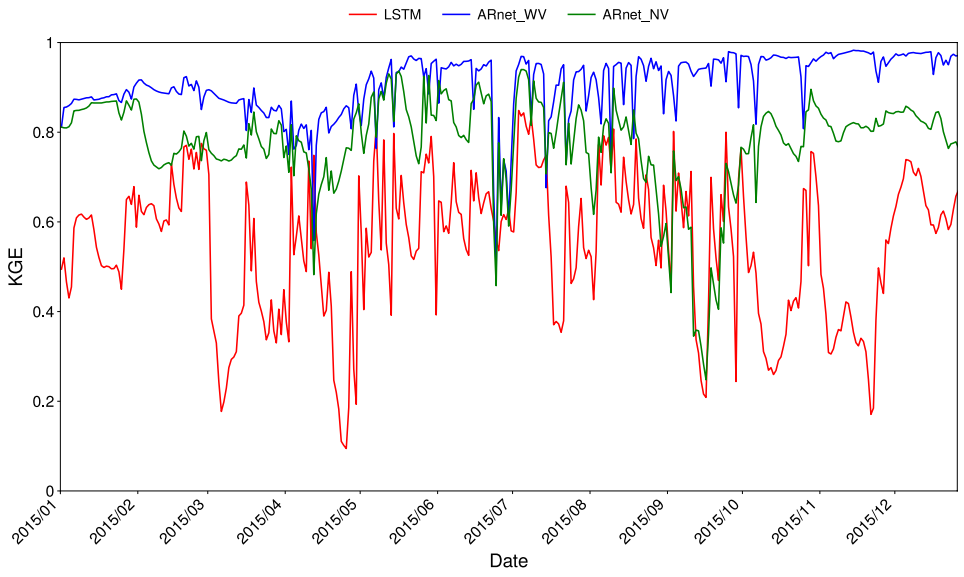

Fig. 7. Time series of the averaged KGE calculated over the entire domain for the two surrogate models in evaluating 100 test data.

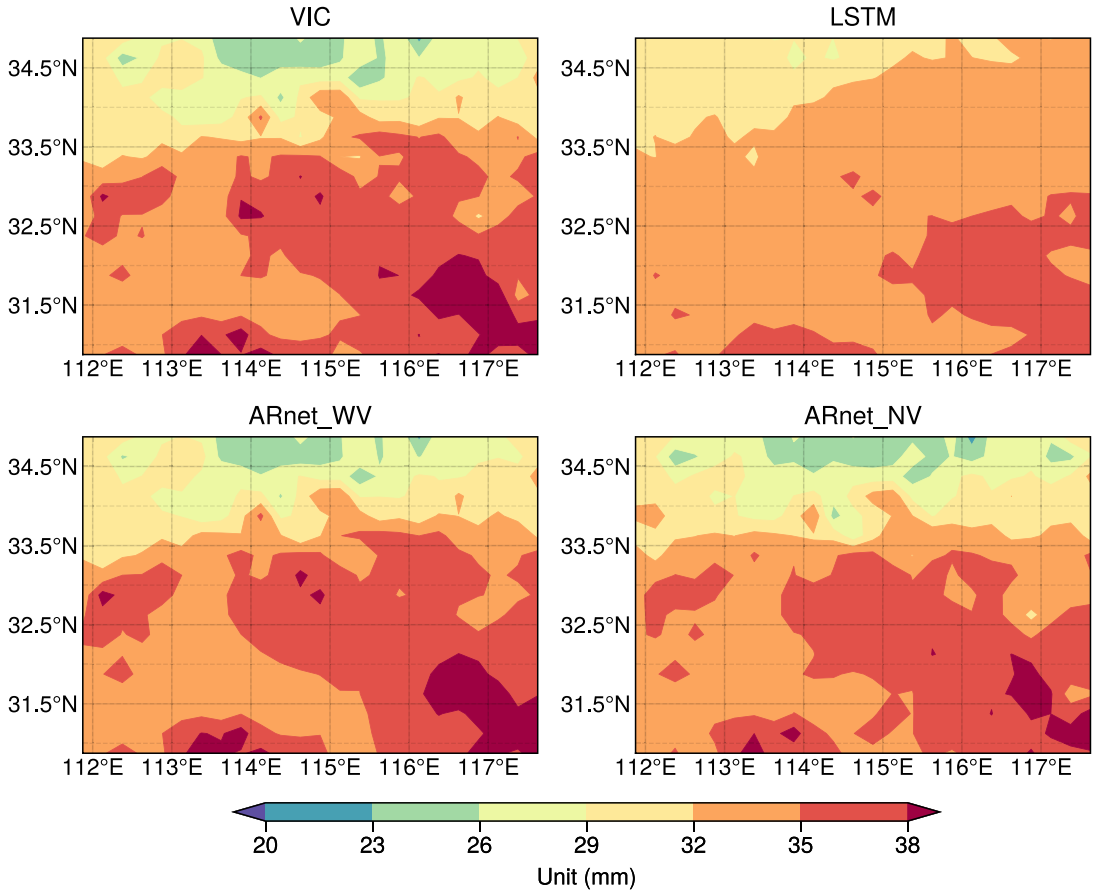

Fig. 8. Spatial distributions of the averaged soil moisture over 2015 of a random test sample based on VIC and two surrogate models.

only choose soil moisture as the model output, but the ARnet can easily handle multiple output variables without modifications of the network architecture. We think the ARnet is very suitable for approximating state variables like soil moisture, soil temperature, snow water equivalent. For non-state variables, the network architecture may still work, and we will test its applicability in our future study.

Another advantage of ARnet for approximating distributed land surface hydrological models is the computational efficiency. In our study, the training time of ARnet is about 140 s for 1 epoch and that of LSTM is about 90 min for 1 epoch. Since the input–output data of each grid cell can be used as a training sample for LSTM, the sample size could be extremely large when the study region is large. A common way of increasing the training efficiency of LSTM is to only sample a subset of grid cells out of the entire domain for training (Gu et al., 2020; Tsai

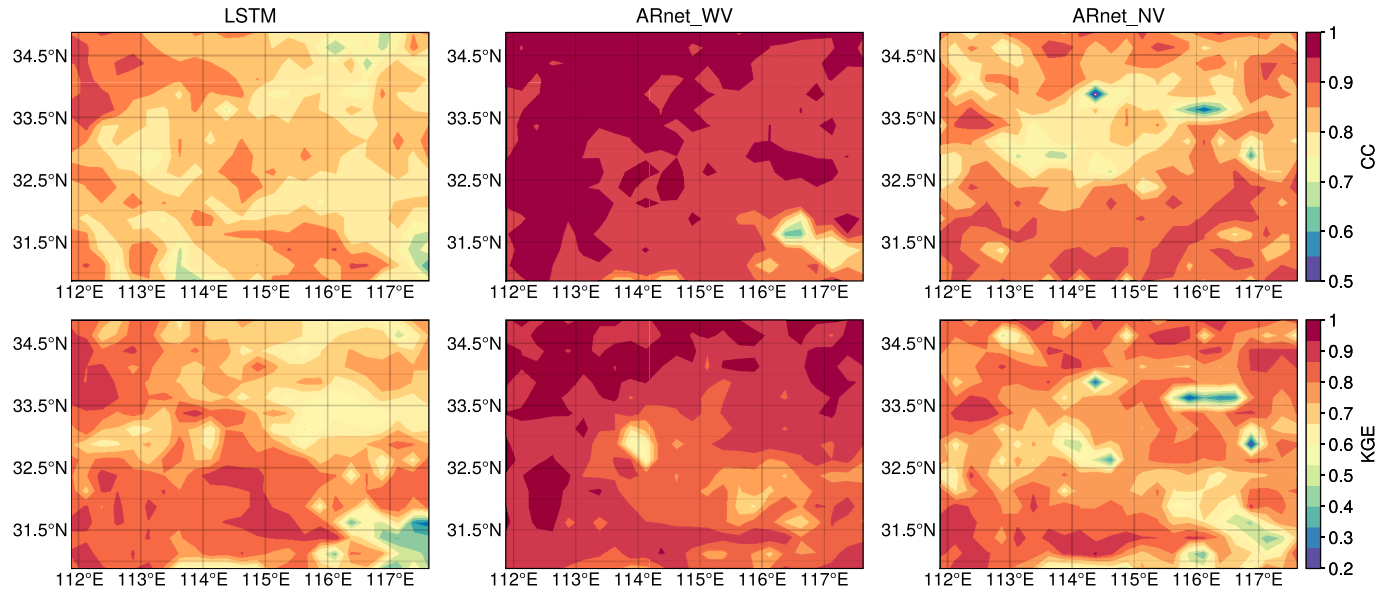

Fig. 9. Spatial distributions of CC (first row) and KGE (second row) between the surrogate models predicted soil moisture and the VIC simulated soil moisture of a random test sample during 2015.

Table 3
Mean and variance of MSE and R^2 for 100 test samples of LSTM and ARnet models.

		MSE	R^2
ARnet_WV	mean	0.10	0.91
	variance	0.067	0.039
ARnet_NV	mean	0.26	0.76
	variance	0.082	0.037
LSTM	mean	0.46	0.57
	variance	0.11	0.088

et al., 2021). In comparison, the ARnet directly processes the input images, and the dense block used in the network can substantially enhance information propagation without requiring a huge amount of data to train the network.

As described in the introduction, surrogate modelling methods are necessary in solving parameter UQ problems associated with complex and computationally demanding land surface hydrological models. To apply surrogate models in the downstream tasks of UQ, it is common practice to match the model simulation with the observation. The ARnet model has demonstrated good performance in simulating soil moisture during a completely independent test period. After being trained, the ARnet model can replace the original model in simulating soil moisture across the entire domain. Whenever we require outputs at a specific location, for different times, we can easily extract the necessary information from the surrogate system. Therefore, the soil moisture simulated by ARnet can be utilized to match the observations during the calibration process. DL based surrogate models can be used in combination with traditional parameter UQ methods (Mo et al., 2019a) or more advanced differentiable parameter learning framework (Tsai et al., 2021). In our future work, we will focus on developing a distributed calibration method based on ARnet to resolve the bottleneck of distributed modelling.

6. Conclusions

In this study, we employ a deep autoregressive neural network

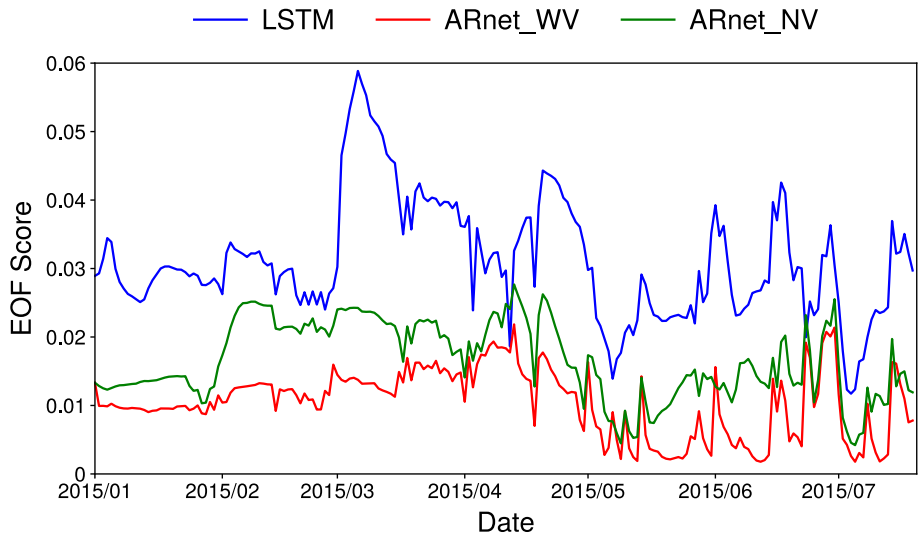

Fig. 10. Comparison of the EOF-based similarity scores for the LSTM and ARnet predicted soil moisture over 2015.1–2015.8.

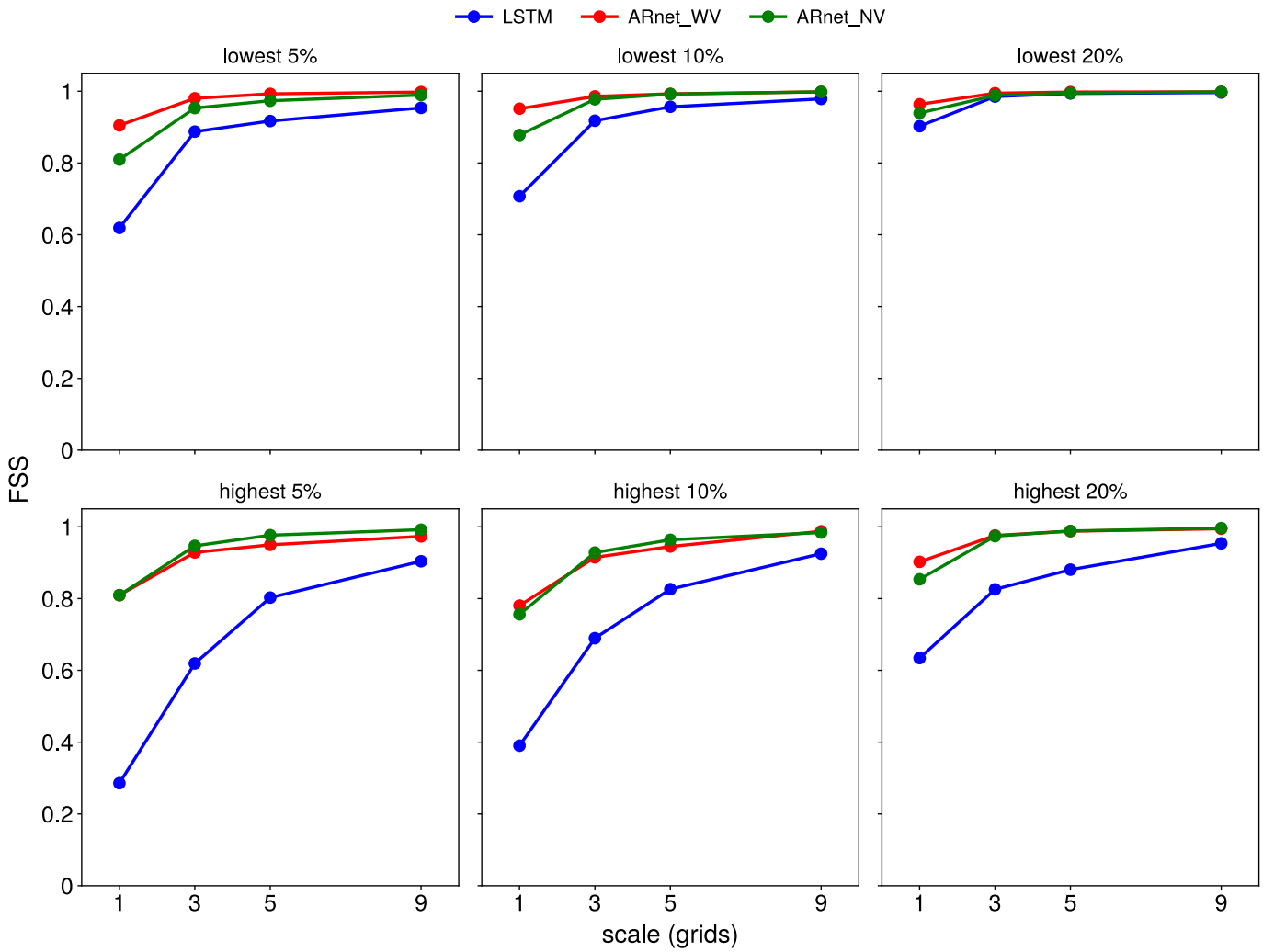

Fig. 11. Plots of FSS against neighborhood length for the averaged soil moisture patterns of 2015 of the two surrogate models using different percentile thresholds.

Table 4
Validation MSE values of ARnet for different network architecture hyperparameters.

Hyperparameters	Options	MSE
Dense block layers	default	0.0433
	(4, 8, 4)	0.0475
	(3, 6, 3)	0.0547
Growth rate	default	0.0433
	32	0.0479
	16	0.0665
Features after first Conv	default	0.0433
	32	0.0467
	16	0.0567

(ARnet) to construct an accurate and reliable surrogate model of the distributed land surface hydrological model VIC simulated soil moisture. In this network, high-dimensional input and output fields are handled as images. The ARnet uses a DenseNet based encoder-decoder network architecture to take full advantage of the convolution in image processing. It can reduce the input dimensionality implicitly through a cascade of nonlinear transformations while having a reliable capability to reconstruct the output images. In addition, the autoregressive strategy makes the ARnet good at handling complex, dynamic relationship between the time-varying input and output. We apply the ARnet to build a surrogate system of the VIC modelled daily soil moisture in the upper Huaihe river basin. We also compare the performance of ARnet with that of the widely used LSTM model. The major conclusions are as follows:

(1) The ARnet can provide an accurate approximation for the mapping between high-dimensional inputs (meteorological fields and parameter fields) and outputs (soil moisture fields) of the VIC model. In addition, ARnet also performs very well in reproducing the dynamic characteristics of the soil moisture due to the autoregressive strategy.
(2) The ARnet surrogate significantly outperforms the LSTM surrogate in approximating spatial patterns of VIC modelled soil moisture based on the spatial evaluations. Using the dense convolutional encoder-decoder could effectively capture the spatial information and thus increase the spatial predictive skill.

It is worth noting that the ability of the DL networks to approximate the dynamic behaviour of complex distributed models is very useful for parameter UQ problems such as sensitivity analysis and model calibration. More research efforts are needed to develop DL-based UQ methods for reducing uncertainties, handling equifinality and improving predictive skill of distributed land surface hydrological models in a changing environment.

CRediT authorship contribution statement

Ruochen Sun: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Investigation, Writing –

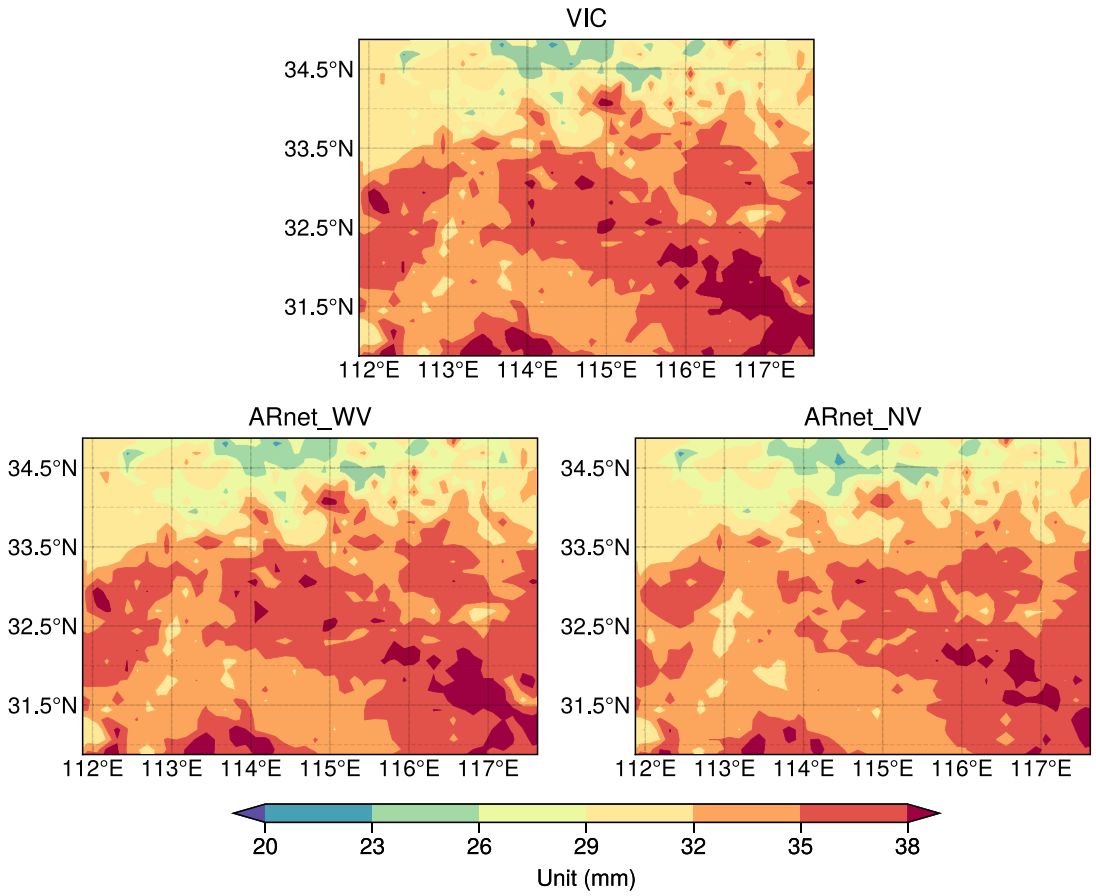

Fig. 12. Spatial distributions of the average soil moisture over all test samples based on VIC and two ARnet modes with finer resolution.

original draft, Writing – review & editing. **Baoxiang Pan:** Conceptualization, Software. **Qingyun Duan:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study is jointly supported by the National Key R&D Program of China (2021YFC3201102), the National Natural Science Foundation of China (42101046, 51979004), the Key Scientific and Technological Project of the Ministry of Water Resources. P.R.C (SKS-2022001), and the Scientific Research Start-up Foundation for High-level Talents of Hohai University (522020012).

References

Archfield, S.A., Clark, M., Arheimer, B., Hay, L.E., McMillan, H., Kiang, J.E., Seibert, J., Hakala, K., Bock, A., Wagener, T., Farmer, W.H., Andréassian, V., Attinger, S., Viglione, A., Knight, R., Markstrom, S., Over, T., 2015. Accelerating advances in continental domain hydrologic modeling. Water Resour. Res. 51 (12), 10078–10091. https://doi.org/10.1002/2015wr017498.

Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resour. Res. 51 (8), 5957–5973. https://doi.org/10.1002/2015WR016967.

Clark, M.P., Schaefli, B., Schymanski, S.J., Samaniego, L., Luce, C.H., Jackson, B.M., Freer, J.E., Arnold, J.R., Moore, R.D., Istanbulluoglu, E., Ceola, S., 2016. Improving the theoretical underpinnings of process-based hydrologic models. Water Resour. Res. 52 (3), 2350–2365. https://doi.org/10.1002/2015WR017910.

Cuo, L., Zhang, Y., Gao, Y., Hao, Z., Cairang, L., 2013. The impacts of climate change and land cover/use transition on the hydrology in the upper Yellow River Basin, China. J. Hydrol. 502, 37–52. https://doi.org/10.1016/j.jhydrol.2013.08.003.

Dai, Y., Shangguan, W., Duan, Q., Liu, B., Fu, S., Niu, G., 2013. Development of a China Dataset of Soil Hydraulic Parameters Using Pedotransfer Functions for Land Surface Modeling. J. Hydrometeorol. 14 (3), 869–887. https://doi.org/10.1175/jhm-d-12-0149.1.

DeChant, C.M., Moradkhani, H., 2014. Hydrologic Prediction and Uncertainty Quantification. In: Eslamian, S. (Ed.), Handbook of Engineering Hydrology: Modeling, Climate Change, and Variability. CRC Press, pp. 387–414.

Dembélé, M., Hrachowitz, M., Savenije, H.H.G., Mariéthoz, G., Schaefli, B., 2020. Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets. Water Resour. Res. 56 (1) https://doi.org/10.1029/2019wr026085.

Fang, K., Pan, M., Shen, C., 2019. The Value of SMAP for Long-Term Soil Moisture Estimation With the Help of Deep Learning. IEEE Trans. Geosci. Remote Sens. 57 (4), 2221–2233. https://doi.org/10.1109/tgrs.2018.2872131.

Gan, Y., Liang, X.-Z., Duan, Q., Ye, A., Di, Z., Hong, Y., Li, J., 2018. A systematic assessment and reduction of parametric uncertainties for a distributed hydrological model. J. Hydrol. 564, 697–711. https://doi.org/10.1016/j.jhydrol.2018.07.055.

Garzón, A., Kapelan, Z., Langeveld, J., Taormina, R., 2022. Machine Learning-Based Surrogate Modeling for Urban Water Networks: Review and Future Research Directions. Water Resour. Res., 58(5), e2021WR031808. https://doi.org/10.1029/2021WR031808.

Gong, W., Duan, Q., 2017. An adaptive surrogate modeling-based sampling strategy for parameter optimization and distribution estimation (ASMO-PODE). Environ. Model. Softw. 95, 61–75. https://doi.org/10.1016/j.envsoft.2017.05.005.

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Dai, Y., Ye, A., Miao, C., 2015. Multi-objective parameter optimization of common land model using adaptive surrogate modeling. Hydrol. Earth Syst. Sci. 19 (5), 2409–2425. https://doi.org/10.5194/hess-19-2409-2015.

Gong, W., Duan, Q., Li, J., Wang, C., Di, Z., Ye, A., Miao, C., Dai, Y., 2016. Multiobjective adaptive surrogate modeling-based optimization for parameter estimation of large, complex geophysical models. Water Resour. Res. 52 (3), 1984–2008. https://doi.org/10.1002/2015wr018230.

Gou, J., Miao, C., Duan, Q., Tang, Q., Di, Z., Liao, W., Wu, J., Zhou, R., 2020. Sensitivity Analysis-Based Automatic Parameter Calibration of the VIC Model for Streamflow Simulations Over China. Water Resour. Res. 56 (1) https://doi.org/10.1029/2019wr025968.

Gou, J., Miao, C., Samaniego, L., Xiao, M., Wu, J., Guo, X., 2021. CNRD v1.0: A High-Quality Natural Runoff Dataset for Hydrological and Climate Studies in China. Bull. Am. Meteorol. Soc. 102 (5), E929–E947. https://doi.org/10.1175/bams-d-20-0094.1.

Gu, H., Xu, Y.-P., Ma, D.i., Xie, J., Liu, L.i., Bai, Z., 2020. A surrogate model for the Variable Infiltration Capacity model using deep learning artificial neural network. J. Hydrol. 588, 125019.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377 (1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hamman, J.J., Nijssen, B., Bohn, T.J., Gergel, D.R., Mao, Y., 2018. The Variable Infiltration Capacity model version 5 (VIC-5): infrastructure improvements for new applications and reproducibility. Geosci. Model Dev. 11 (8), 3481–3496. https://doi.org/10.5194/gmd-11-3481-2018.

Hao, Z., Yuan, X., Xia, Y., Hao, F., Singh, V.P., 2017. An Overview of Drought Monitoring and Prediction Systems at Regional and Global Scales. Bull. Am. Meteorol. Soc. 98 (9), 1879–1896. https://doi.org/10.1175/bams-d-15-00149.1.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X., 2020. The first high-resolution meteorological forcing dataset for land process studies over China. Sci Data 7 (1), 25. https://doi.org/10.1038/s41597-020-0369-y.

Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Huang, M., Ray, J., Hou, Z., Ren, H., Liu, Y., Swiler, L., 2016. On the applicability of surrogate-based Markov chain Monte Carlo-Bayesian inversion to the Community Land Model: Case studies at flux tower sites. J. Geophys. Res. Atmos. 121 (13), 7548–7563. https://doi.org/10.1002/2015jd024339.

Jiang, L., Wu, H., Tao, J., Kimball, J.S., Alfieri, L., Chen, X., 2020. Satellite-Based Evapotranspiration in Hydrological Model Calibration. Remote Sens. 12 (3), 428. https://doi.org/10.3390/rs12030428.

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., Thielen, J., 2016. Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. Environ. Model. Softw. 75, 68–76. https://doi.org/10.1016/j.envsoft.2015.09.009.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Koch, J., Jensen, K.H., Stisen, S., 2015. Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. Water Resour. Res. 51, 1225–1246. https://doi.org/10.1002/2014WR016607.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., Herrnegger, M., 2018. Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. Hydrol. Earth Syst. Sci. 22 (11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., Nearing, G., 2019. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrol. Earth Syst. Sci. 23 (12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019.

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J. Geophys. Res. Atmos. 99 (D7), 14415–14428. https://doi.org/10.1029/94JD00483.

Lohmann, D., Mitchell, K.E., Houser, P.R., Wood, E.F., Schaake, J.C., Robock, A., Cosgrove, B.A., Sheffield, J., Duan, Q., Luo, L., Higgins, R.W., Pinker, R.T., Tarpley, J.D., 2004. Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. J. Geophys. Res. Atmos. 109 (D7) https://doi.org/10.1029/2003JD003517.

Lu, D., Ricciuto, D., 2019. Efficient surrogate modeling methods for large-scale Earth system models based on machine-learning techniques. Geosci. Model Dev. 12 (5), 1791–1807. https://doi.org/10.5194/gmd-12-1791-2019.

Lu, D., Ricciuto, D., Stoyanov, M., Gu, L., 2018. Calibration of the E3SM Land Model Using Surrogate-Based Global Optimization. J. Adv. Model. Earth Syst. 10 (6), 1337–1356. https://doi.org/10.1002/2017ms001134.

Ma, R., Xiao, J., Liang, S., Ma, H., He, T., Guo, D., Liu, X., Lu, H., 2022. Pixel-level parameter optimization of a terrestrial biosphere model for improving estimation of carbon fluxes with an efficient model–data fusion method and satellite-derived LAI and GPP data. Geosci. Model Dev. 15 (17), 6637–6657. https://doi.org/10.5194/gmd-15-6637-2022.

Mendiguren, G., Koch, J., Stisen, S., 2017. Spatial pattern evaluation of a calibrated national hydrological model – a remote-sensing-based diagnostic approach. Hydrol. Earth Syst. Sci. 21 (12), 5987–6005. https://doi.org/10.5194/hess-21-5987-2017.

Mizukami, N., Clark, M.P., Newman, A.J., Wood, A.W., Gutmann, E.D., Nijssen, B., Rakovec, O., Samaniego, L., 2017. Towards seamless large-domain parameter estimation for hydrologic models. Water Resour. Res. 53 (9), 8020–8040. https://doi.org/10.1002/2017wr020401.

Mo, S., Zabaras, N., Shi, X., Wu, J., 2019a. Deep Autoregressive Neural Networks for High-Dimensional Inverse Problems in Groundwater Contaminant Source Identification. Water Resour. Res. 55 (5), 3856–3881. https://doi.org/10.1029/2018WR024638.

Mo, S., Zhu, Y., Zabaras, N., Shi, X., Wu, J., 2019b. Deep Convolutional Encoder-Decoder Networks for Uncertainty Quantification of Dynamic Multiphase Flow in Heterogeneous Media. Water Resour. Res. 55 (1), 703–728. https://doi.org/10.1029/2018WR023528.

Müller, J., Paudel, R., Shoemaker, C.A., Woodbury, J., Wang, Y., Mahowald, N., 2015. CH_4 parameter estimation in CLM4.5bgc using surrogate global optimization. Geosci. Model Dev. 8 (10), 3285–3310. https://doi.org/10.5194/gmd-8-3285-2015.

Pan, B., Anderson, G.J., Goncalves, A., Lucas, D.D., Bonfils, C.J.W., Lee, J., Tian, Y., Ma, H.Y., 2021. Learning to Correct Climate Projection Biases. J. Adv. Model. Earth Syst. 13 (10) https://doi.org/10.1029/2021ms002509.

Pokhrel, P., Gupta, H.V., 2010. On the use of spatial regularization strategies to improve calibration of distributed watershed models. Water Resour. Res. 46 (1) https://doi.org/10.1029/2009wr008066.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 (7) https://doi.org/10.1029/2011wr011527.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J.H.A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., Maier, H. R., 2021. The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. Environ. Model. Softw. 137, 104954.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566 (7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Ricciuto, D., Sargsyan, K., Thornton, P., 2018. The Impact of Parametric Uncertainties on Biogeochemistry in the E3SM Land Model. J. Adv. Model. Earth Syst. 10 (2), 297–319. https://doi.org/10.1002/2017MS000962.

Roberts, N.M., Lean, H.W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon. Wea. Rev. 136, 78–97. https://doi.org/10.1175/2007MWR2123.1.

Rosero, E., Yang, Z.-L., Wagener, T., Gulden, L.E., Yatheendradas, S., Niu, G.-Y., 2010. Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season. J. Geophys. Res. Atmos. 115 (D3) https://doi.org/10.1029/2009JD012035.

Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res. 46 (5) https://doi.org/10.1029/2008wr007327.

Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., Liu, H., Liu, S., Miao, C., Zhang, Y., 2013. A China data set of soil properties for land surface modeling. J. Adv. Model. Earth Syst. 5 (2), 212–224. https://doi.org/10.1002/jame.20026.

Shen, C., 2018. A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. Water Resour. Res. 54 (11), 8558–8593. https://doi.org/10.1029/2018wr022643.

Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-L., Kifer, D., Fang, Z., Fang, K., Li, D., Li, X., Tsai, W.-P., 2018. HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community. Hydrol. Earth Syst. Sci. 22 (11), 5639–5656. https://doi.org/10.5194/hess-22-5639-2018.

Shen, Y., Zhao, P., Pan, Y., Yu, J., 2014. A high spatiotemporal gauge-satellite merged precipitation analysis over China. J. Geophys. Res. Atmos. 119 (6), 3063–3075. https://doi.org/10.1002/2013jd020686.

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., Woo, W. C., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, 28.

Sun, R., Duan, Q., Huo, X., 2021a. Multi-Objective Adaptive Surrogate Modeling-Based Optimization for Distributed Environmental Models Based on Grid Sampling. Water Resour. Res., 57(11), e2020WR028740. https://doi.org/10.1029/2020WR028740.

Sun, R., Yuan, H., Liu, X., Jiang, X., 2016. Evaluation of the latest satellite–gauge precipitation products and their hydrologic applications over the Huaihe River basin. J. Hydrol. 536, 302–319. https://doi.org/10.1016/j.jhydrol.2016.02.054.

Sun, R., Duan, Q., Wang, J., 2021b. Understanding the spatial patterns of evapotranspiration estimates from land surface models over China. J. Hydrol. 595, 126021 https://doi.org/10.1016/j.jhydrol.2021.126021.

Sun, R., Duan, Q., Mao, X., 2022. A multi-objective adaptive surrogate modelling-based optimization algorithm for constrained hybrid problems. Environ. Model. Softw. 148, 105272 https://doi.org/10.1016/j.envsoft.2021.105272.

Tripathy, R.K., Bilionis, I., 2018. Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification. J. Comput. Phys. 375, 565–588. https://doi.org/10.1016/j.jcp.2018.08.036.

Tsai, W.P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., Shen, C., 2021. From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. Nat. Commun. 12 (1), 5988. https://doi.org/10.1038/s41467-021-26107-z.

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., Miao, C., 2014. An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. Environ. Model. Softw. 60, 167–179. https://doi.org/10.1016/j.envsoft.2014.05.026.

Yang, Y., Pan, M., Beck, H.E., Fisher, C.K., Beighley, R.E., Kao, S.C., Hong, Y., Wood, E.F., 2019. In Quest of Calibration Density and Consistency in Hydrologic Modeling: Distributed Parameter Calibration against Streamflow Characteristics. Water Resour. Res. 55 (9), 7784–7803. https://doi.org/10.1029/2018wr024178.

Yu, S., Ma, J., 2021. Deep Learning for Geophysics: Current and Future Trends. Rev. Geophys. 59 (3) https://doi.org/10.1029/2021rg000742.
Zhu, Y., Zabaras, N., 2018. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. J. Comput. Phys. 366, 415–447. https://doi.org/10.1016/j.jcp.2018.04.018.
Zhu, Y., Zabaras, N., Koutsourelakis, P.-S., Perdikaris, P., 2019. Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. J. Comput. Phys. 394, 56–81. https://doi.org/10.1016/j.jcp.2019.05.024.

A surrogate modeling method for distributed land surface hydrological models based on deep learning

Sun, Ruochen; Pan, Baoxiang; Duan, Qingyun

01 mingxi zhang Page 1

30/8/2024 7:53

02 mingxi zhang Page 1

30/8/2024 7:54