https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Hydrology and Earth System Sciences Discussions

Quantifying the uncertainty of precipitation forecasting using probabilistic deep learning

Lei Xu[1], Nengcheng Chen[1,2], Chao Yang[1]

[1]National Engineering Research Center for Geographic Information System, School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China

[2]State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China

Correspondence to: Chao Yang (yangchao@cug.edu.cn)

Abstract. Precipitation forecasting is an important mission in weather science. In recent years, data-driven precipitation forecasting techniques could complement numerical prediction, such as precipitation nowcasting, monthly precipitation projection and extreme precipitation event identification. In data-driven precipitation forecasting, the predictive uncertainty arises mainly from data and model uncertainties. Current deep learning forecasting methods could model the parametric uncertainty by random sampling from the parameters. However, the data uncertainty is usually ignored in the forecasting process and the derivation of predictive uncertainty is incomplete. In this study, the input data uncertainty, target data uncertainty and model uncertainty are jointly modeled in a deep learning precipitation forecasting framework to estimate the predictive uncertainty. Specifically, the data uncertainty is estimated a priori and the input uncertainty is propagated forward through model weights according to the law of error propagation. The model uncertainty is considered by sampling from the parameters and is coupled with input and target data uncertainties in the objective function during the training process. Finally, the predictive uncertainty is produced by propagating the input uncertainty and sampling the weights in the testing process. The experimental results indicate that the proposed joint uncertainty modeling and precipitation forecasting framework exhibits comparable forecasting accuracy with existing methods, while could reduce the predictive uncertainty to a large extent relative to two existing joint uncertainty modeling approaches. The developed joint uncertainty modeling method is a general uncertainty estimation approach for data-driven forecasting applications.

1 Introduction

Precipitation is a key hydrometeorological variable in earth system science, and is the main driving

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

factor of floods and droughts (Xu et al., 2019). In the year of 2019, the flood disaster driven by extreme precipitation caused a direct economic loss of 29.6 billion dollars in China, and the drought disaster led to a crop production loss of 23.6 billion kilograms (http://www.mwr.gov.cn/sj/#tjgb). Accurate precipitation forecasting is vital for the early warning of flood and drought, smart city management and agricultural water resources allocation (Van Den Hurk et al., 2012; Pozzi et al., 2013). However, the precipitation forecasting problem suffers from uncertainties from data, algorithms and random factors (Reeves et al., 2014; Kobold and Sušelj, 2005; Xu et al., 2020b). The predictive uncertainty is a measurement of the spread of precipitation forecasting and could indicate how much the forecasted precipitation values fluctuate around the mean (Papacharalampous et al., 2020). Therefore, the uncertainty range should be given when generating precipitation forecasting results.

The precipitation forecasting methods can be divided into two categories: numerical weather forecasting and statistical machine learning. Numerical models consider the physical process of earth system and could simulate the interactions between atmospheres, oceans and lands (Sikder and Hossain, 2016; Molinari and Dudek, 1992). Numerical models have strong physical meaning and are the dominant ways of operational precipitation forecasting. However, the forecasting ability of numerical models is limited due to the uncertainty in initial and boundary conditions, the imperfection of parameterization schemes and the uncertainty in parameters (Reeves et al., 2014). With the development of computer technology and machine learning algorithms, using random data-driven techniques for precipitation forecasting is becoming popular in recent years (Shi et al., 2015; Trebing et al., 2021; Sønderby et al., 2020). The accuracy of data-driven methods is comparable to currently advanced numerical models in short-term (e.g. from hours to weeks) precipitation forecasting. For example, the convolutional long-short term memory network is shown to outperform the physical optical flow method in precipitation nowcasting based on radar images (Shi et al., 2015). Another deep learning model called MetNet showed advantages over traditional numerical models in terms of the forecasting accuracy and running time for hourly precipitation prediction (Sønderby et al., 2020). The data-driven methods also exhibit appealing results in subseasonal to seasonal precipitation forecasting relative to numerical models (Boukabara et al., 2019; Chantry et al., 2021; Hwang et al., 2019). A key drawback of data-driven precipitation forecasting method is the lack of physical meaning, also known as black-box model. Despite this feature, data-driven statistical machine learning methods have been widely used for parameter calibration, data

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

processing, submodel replacement and process understanding among physical simulations (Ardabili et al., 2019; Sahoo et al., 2017; Reichstein et al., 2019). The data-driven learning techniques are strong complements to numerical models for the improvement of precipitation forecasting accuracy.

The predictive uncertainty in precipitation forecasting arises mainly from data and models (Gal, 2016). The data uncertainty comes from external observation conditions, instruments and processing algorithms. The data uncertainty is usually examined by perturbing initial conditions in numerical models and producing a perturbed multi-model ensemble, which is widely seen in hydrometeorological ensemble forecasting (Xu et al., 2019; Gneiting and Raftery, 2005; Duan et al., 2019; Vitart et al., 2017). The data uncertainty is rarely investigated in data-driven precipitation forecasting and is often assumed to be accurate without error. The model uncertainty is often represented by an ensemble of perturbed model physics and parameters in numerical weather forecasting (Vitart et al., 2017; Kirtman et al., 2014; Taylor et al., 2012). In data-driven models, the model uncertainty is generally modeled by random regularization of parameters (Gal, 2016; Kendall and Gal, 2017). For linear regression, the parametric uncertainty is indicated by the standard deviation of trained parameters. In deep learning, the network layers could be randomly abandoned to prevent overfitting and generate a forecasted ensemble by Monte Carlo sampling (Kendall and Gal, 2017; Srivastava et al., 2014; Loquercio et al., 2020; Ghahramani, 2015).

The data and model uncertainties should be considered jointly in an integrated modeling framework to get the predictive uncertainty, as the data and model uncertainties could both inflate the predictive spread considerably (Gal, 2016; Kendall and Gal, 2017). It is expected that, the forecasting result would be more or less different if the used data and parameters are randomly sampled from the population. Data uncertainty is usually assumed as a constant or Gaussian distribution and could be propagated into final forecasting through error forward propagation (Loquercio et al., 2020; Xu et al., 2020a). If the data uncertainty is unknown, it can be learned from the training process by considering the data uncertainty as a trainable parameter (Kendall and Gal, 2017). However, the joint learning of data errors and model weights will increase the number of training parameters and may mix the error flow from data and parameters. A prior estimation of data uncertainty could help unravel the data error and facilitate the training process. On the other hand, previous forecasting studies usually model the input data uncertainty and ignore the uncertainty in the target (predictand) data (Kendall and Gal, 2017; Loquercio et al., 2020). The uncertainty in the target dataset also plays an important role in the parameter training process and

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Hydrology and Earth System Sciences
Open Access
Discussions

could influence the forecasting accuracy.

There are two ways to estimate the data uncertainty. One is to use in-situ ground stations to calculate the systematic and random errors within the data and the other is to use multisource datasets to compute random error by intercomparison (Xu et al., 2021; Gruber et al., 2016; Sun et al., 2018). The in-situ validation method is limited to the number and density of ground stations and are suitable for small areas with enough station coverage. The second method is independent of the in-situ stations and requires multisource datasets with independent error distribution (Gruber et al., 2016). There are numerous precipitation datasets from various sensors and models and could be used to calculate precipitation data error at a large spatial scale (Xu et al., 2020b; Sun et al., 2018). Three-cornered hat (TCH) and triple collocation (TC) are two commonly used methods to evaluate the random error among multisource datasets, which do not require ground measurements as references (Premoli and Tavella, 1993; Mccoll et al., 2014; Stoffelen, 1998). The basic assumption of TCH and TC methods is the stationarity of both the raw dataset and its error, which may not be always satisfied for real-world data. Most of the existing studies assume that the used multisource datasets obey the stationarity condition when using TCH or TC methods (Xu et al., 2020b; Gruber et al., 2016; Gruber et al., 2017), which is useful for the determination of relative prior random error.

In this study, we aim to quantify the predictive uncertainty of data-driven precipitation forecasting by fully considering the uncertainty from data and models. The data uncertainty is estimated by the TCH method a priori and is assumed as Gaussian distribution. The data uncertainty is propagated within model training by the law of error propagation. The parametric uncertainty is modeled by randomly abandoning some network layers during the training process. The data and model uncertainties are jointly considered in the objective function within a deep learning encoder-decoder framework. The forecasting experiments are conducted to see whether the accuracy of precipitation forecasting can be improved by joint data-model uncertainty modeling relative to several uncertainty processing strategies from the existing studies.

2 Study area and data

The study area is located at southern and northern China, East Asia (Figure 1). The annual rainfall decreases from the southeast to the northwest, with an approximately average rainfall of 1500 mm in the

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

southeast regions and 300 mm in the northwest areas. Most of the southern areas feature a subtropical monsoon climate and the rainfall is relatively larger in summer and smaller in winter. From June to July in 2020, extreme precipitation hit the southern China (Wei et al., 2020) and caused a direct economic loss of 13.2 billion dollars. The precipitation forecasting in southern area of China is very challenging and meaningful. Previous studies use numerical models for precipitation forecasting in this area and show some values (Yuan et al., 2012; Luo et al., 2017). The northern area of China features the temperate moon and continental climates, with an annual rainfall of 400 to 800 mm and the main rainy season of July and August. Here we would like to explore the possibility of weekly precipitation forecasting by a data-driven deep learning method.

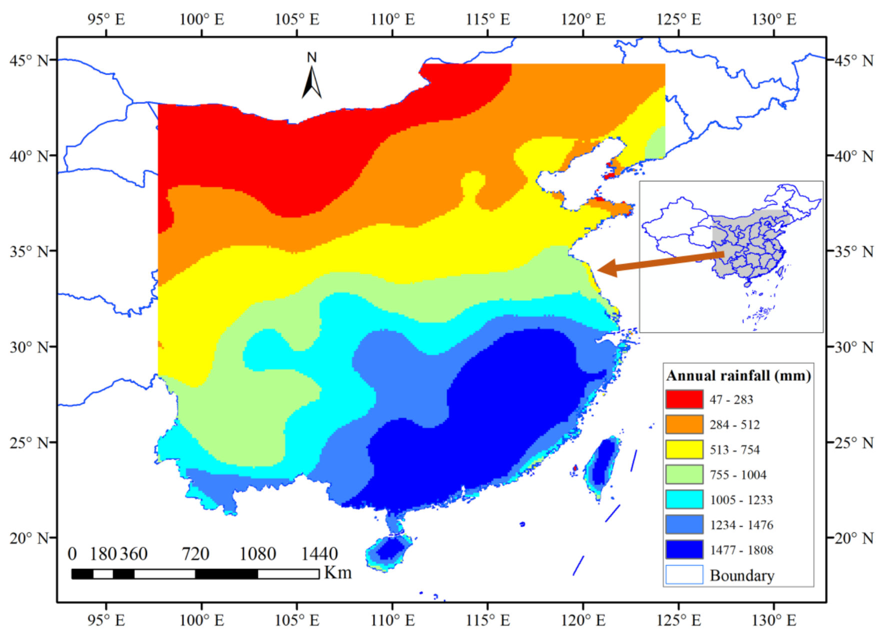


Figure 1: The study area for short-term precipitation forecasting.

Multisource precipitation datasets are used here to obtain the data error and to measure the precipitation forecasting ability of different uncertainty processing strategies, including the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al., 2017), the National Centers for Environmental Prediction Reanalysis version 2 (NCEP R2) (Saha et al., 2014) and the European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA-5) (Hersbach et al., 2020) datasets from 1980 to 2020. The surface 2-meter temperature and the geopotential height at 500 hPa datasets are collected from the three datasets accordingly as predictors. All these daily datasets

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

are converted to weekly data and are bilinearly interpolated into 0.25° resolution. In the forecasting process, the temperature and geopotential height predictors in the historical three consecutive weeks are used to forecast the precipitation in the target week.

3 Methods

3.1. Estimation of data uncertainty

The TCH method (Xu et al., 2020b; Premoli and Tavella, 1993) is used to estimate the uncertainty in temperature, geopotential height and precipitation datasets. The collected three datasets are all reanalysis data, which is generated from different physical models and data assimilation algorithms. The different reanalysis datasets and their errors are generally not closely correlated and are regarded as collocated datasets for the uncertainty estimation, similar with existing studies (Xu et al., 2021; Mccoll et al., 2014; Gruber et al., 2017). In the TCH algorithm, one arbitrary dataset is chosen as the reference among the three datasets, and then the differencing operation is conducted between the reference and the other two datasets to get the differencing series. The covariance of the differencing series is connected to the variance-covariance matrix of precipitation datasets through matrix transformation. The parameters of the variance-covariance matrix are iteratively resolved by minimizing the global correlation of the covariance of the differencing series. A detailed introduction of TCH method could refer to Premoli and Tavella (1993) and Xu et al. (2020b).

The uncertainties of the predictors and predictands are estimated seasonally by the TCH method. The weekly datasets are grouped according to the weekly climatology and then used to estimate the uncertainty. For example, all the precipitation datasets which belongs to the first week of each year are concatenated to apply the TCH method in order to get the uncertainty of the datasets on the first week of each year. Similarly, the data uncertainty on the second week, third week and until the fifty-two week is evaluated sequentially. This strategy enables a time-variant uncertainty estimation, which is more reasonable as the precipitation climatology is different for different seasons. The NECP R2 and ERA-5 data are used to assist the uncertainty estimation of MERRA-2 data by the TCH method, and the precipitation forecasting experiments are conducted based on MERRA-2 data to evaluate the proposed forecasting framework.

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

3.2. Variational Bayesian inference

Here we introduce the variational inference theory (Hoffman et al., 2013), which is a standard Bayesian modeling technique for the estimation of model uncertainty. Given the input data $X=\{x_1,\ldots,x_N\}$ and the output data $Y=\{y_1,\ldots,y_N\}$, the Bayesian regression is to find suitable parameters within the function $y=f^w(x)$ which could generate the output Y according to the input X. The parameters w is assumed to obey a prior distribution $p(w)$ before the observations are known. When the observed data is obtained, it is possible to determine which parameters are more suitable for the function according to the data. A likelihood distribution $p(y|x,w)$ is defined to describe the probability of y generated by x and w. For example, a Gaussian likelihood function is defined as

$$p(y|x,w) = \mathcal{N}(y; f^w(x), \tau^{-1}I) \tag{1}$$

where τ^{-1} is the observation noise.

Given the input data X and the output data Y, the Bayesian theorem is to find the posterior distribution of parameters in the parameter space.

$$p(w|X,Y) = \frac{p(Y|x,w)p(w)}{p(Y|X)} \tag{2}$$

where the numerator $p(Y|X)$ is the normalization factor, also named as model evidence.

$$p(Y|X) = \int p(Y|X,w)p(w)dw$$

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Hydrology and Earth System Sciences
Open Access
Discussions

(3)

The solution of Equation (3) needs to marginalize the likelihood over w, which is tractable analytically for some simple models such as Bayesian regression, while is intractable for complex models such as deep learning methods (Gal, 2016).

Given the new input data x', the forecasted value is generated by the integral of probability over the parameter space, which is called the inference process.

$$p(y'|x',X,Y) = \int p(y'|x',w)p(w|X,Y)dw$$

(4)

Since the posterior distribution of parameters $p(w|X,Y)$ cannot be obtained analytically, an approximate analytical distribution $q_\theta(w)$ could be defined, with θ as the parameter to be estimated, to be as close as the posterior distribution. The Kullback-Leibler (K-L) divergence (Kullback and Leibler, 1951) is an indicator to measure the similarity of two distributions, also known as relative entropy. The objective function is to minimize the K-L divergence between the two distributions.

$$KL(q_\theta(w)||p(w|X,Y)) = \int q_\theta(w) log \frac{q_\theta(w)}{p(w|X,Y)} dw$$

(5)

The optimal variational distribution $q'_\theta(w)$ is obtained when the K-L divergence is minimized. The estiamted variational distribution could be regarded as the posterior distribution of parameters and then the predictive distribution could be generated.

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

$$p(y'|x',X,Y) \approx \int p(y'|x',w)q'_\theta(w)dw =: q'_\theta(y'|x') \tag{6}$$

The above inference process is the variational Bayesian inference. Variational inference replaces the integral of the likelihood with optimization, which simplifies the estimation of posterior distribution.

3.3. Monte Carlo sampling

Monte Carlo method is a kind of stochastic simulation technology, proposed by Stanislaw Ulam and John von Neumann during the second world war (Von Neumann and Ulam, 1951). Monte Carlo methods are used to estimate unknown parameters by random sampling and are widely applied in mathematics, physics, game theory and finance (Brooks, 1998; Jacoboni and Lugli, 2012; Metropolis and Ulam, 1949; Rubinstein and Kroese, 2016).

In Equation (4), the posterior distribution $p(w|X,Y)$ cannot be solved analytically. Assume U_i as the weight matrix $K_i \times K_{i-1}$ from i-1 layer to i layer, i.e. $w=\{U_i\}_{i=1,\ldots,L}$, a variational weight distribution $q(w)$ is defined to randomly replace the columns with zero (dropout process).

$$U_i = H_i \cdot diag\left([z_{i,j}]_{j=1}^{K_i}\right) \tag{7}$$

$$[z_{i,j}] \sim Bernoulli(p_i), i = 1, \ldots, L, j = 1, \ldots, K_{i-1} \tag{8}$$

where p_i and H_i are variational parameters; $z_{i,j}$ is a binary variable, with a value of zero representing the abandoning of jth unit in i-1 layer and a value of one the keeping, based on the *Bernoulli* distribution at the probability p_i.

The predictive distribution is estimated after minimizing the K-L divergence.

$$q(y'|x') = \int p(y'|x',w)q(w)dw$$

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

(9)

The predictive mean and variance can be obtained after repeating the dropout process multiple times.

$$\mathbb{E}_{q(y'|x')}(y') = \int y' q(y'|x') dy' =$$
$$\int y' \mathcal{N}(y'; \hat{y}'(x', U_1, \ldots, U_L), \tau^{-1} I) Bern(U_1) \cdots Bern(U_L) dU_1 \cdots dU_L dy' =$$
$$\int \hat{y}'(x', U_1, \ldots, U_L) Bern(U_1) \cdots Bern(U_L) dU_1 \cdots dU_L \approx \frac{1}{T}\sum_{t=1}^{T} \hat{y}'(x', \widehat{U}_{1,t}, \ldots, \widehat{U}_{L,t})$$

(10)

$$Var_{q(y'|x')}(y') \approx \tau^{-1} I + \frac{1}{T}\sum_{t=1}^{T} \hat{y}'(x', \widehat{U}_{1,t}, \ldots, \widehat{U}_{L,t})^T \hat{y}'(x', \widehat{U}_{1,t}, \ldots, \widehat{U}_{L,t}) -$$
$$\mathbb{E}_{q(y'|x')}(y')^T \mathbb{E}_{q(y'|x')}(y')$$

(11)

where $u_{t,i}$ is the forecasted value for *i*th pixel and *t*th ensemble. The calculation of predictive variance is based on the standard deviation of the ensemble, which represents the spread of the forecasted values.

The above Monte Carlo sampling and dropout process is the Monte Carlo dropout technique, which is used to obtain the model uncertainty here.

3.4. Joint data and model uncertainties modeling

Dropout is a Bayesian method to model the model uncertainty in forecasting (Srivastava et al., 2014). However, the data uncertainty also needs to be considered. Kendall and Gal (2017) regarded the data uncertainty as a trainable parameter and jointly considered data and model uncertainties. However, the predictand data uncertainty is ignored and the learning of data uncertainty increases the number of training parameters. Here we propose an integrated modeling framework to fully incorporate the data and model uncertainties during the training process (Figure 2). First, the data uncertainties of predictors and predictands are estimated by the TCH method and are assumed as Gaussian distribution.

$$\sigma = TCH(D_i), i = 1,2,3$$

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Hydrology and Earth System Sciences
Open Access
Discussions

(12)

$$\sigma_x \sim \mathcal{N}(0, \sigma)$$

(13)

$$\sigma_y \sim \mathcal{N}(0, \sigma)$$

(14)

The data uncertainty is randomly sampled T times to generate an ensemble of predictors and predictands. In the meantime, the parameters are randomly dropped out for T times to construct a parametric ensemble. The perturbed data and parameter values are jointly used to calculate the training loss. The objective function is expressed as follows (Kendall and Gal, 2017), which is obtained from the likelihood of a Gaussian process (Srivastava et al., 2014).

$$\mathcal{L}(\theta, p) = -\frac{1}{N}\sum_{i=1}^{N} \log p\left(y_{i,\sigma} \middle| f^{\widehat{U}_i}(x_{i,\sigma})\right) + \frac{1-p}{2N}\|\theta\|^2$$

(15)

$$\sigma = \sqrt{(\sigma_x^{(l)})^2 + \sigma_y^2}$$

(16)

where N is the sample size; p is the dropout probability; $\widehat{U}_i \sim q'_\theta(U)$; θ is the parameter to be estimated; σ_x and σ_y are the data uncertainty for predictor and predictand, respectively. The negative log-likelihood function can be deduced according to the objective function.

$$-\log p\left(y_{i,\sigma} \middle| f^{\widehat{U}_i}(x_{i,\sigma})\right) \propto \frac{1}{2\sigma^2}\left\|y_i - f^{\widehat{U}_i}(x_i)\right\|^2 + \frac{1}{2}\log(\sigma^2)$$

(17)

where σ is the regression noise, with the mean of zero in a Gaussian distribution.

The objective function consists of a mean square error (MSE) term adjusted by data uncertainty and a regularization term, which is the negative logarithm of the Gaussian likelihood function. The objective function includes an uncertainty parameter σ^2, which is determined by the sum of propagated uncertainty and target data uncertainty. The minimization of the negative log-likelihood function could be reached by differentiating the optimization function and setting to zero.

$$\frac{\partial}{\partial \sigma^2}\left[\frac{1}{2\sigma^2}\left\|y_i - f^{\theta_i}(x_i)\right\|^2 + \frac{1}{2}\log\left(\sigma^2\right)\right] = 0 \tag{18}$$

$$\Rightarrow -\frac{1}{2\sigma^4}\left\|y_i - f^{\theta_i}(x_i)\right\|^2 + \frac{1}{2\sigma^2} = 0 \tag{19}$$

$$\Rightarrow \sigma^2 = \left\|y_i - f^{\theta_i}(x_i)\right\|^2 \tag{20}$$

where the minimum value of the negative log-likelihood function could be reached when the data variance equals to the square of the difference between the forecasted value and the observation.

Once the network weights are determined according to the objective function, the new input data uncertainty is propagated and the weights are randomly sampled to produce the forecasted ensemble. The predictive mean and variance are calculated from the predictive ensembles.

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T} y_{t,i} \tag{21}$$

$$Var_i \approx \frac{1}{T}\sum_{t=1}^{T} y_{t,i}^2 - \left(\frac{1}{T}\sum_{t=1}^{T} y_{t,i}\right)^2 + \frac{1}{T}\sum_{t=1}^{T} \sigma_{t,i}^2$$

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

(22)

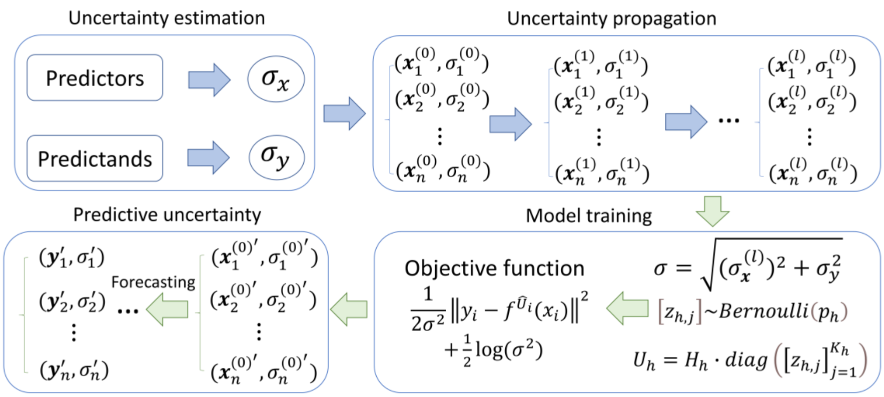


Figure 2: The proposed integrated data-model uncertainty modeling framework in precipitation forecasting.

3.5. The deep learning forecasting framework

In deep learning, encoder-decoder is a commonly used forecasting model (Badrinarayanan et al., 2017; Cho et al., 2014). In the encoder process, an input signal is converted into a one-dimension vector with fixed length. In the decoder process, the one-dimension vector is transformed into the target data with variable length. The available networks used for encoder and decoder processes are arbitrary and depend on the specific problem, such as convolutional neural network (CNN), recurrent neural network (RNN) and long-short term memory (LSTM) network (Hochreiter and Schmidhuber, 1997; Goodfellow et al., 2016). Here we designed a deep learning encoder-decoder model for weekly precipitation forecasting (Figure 3). The temperature and geopotential height data for previous three weeks are regarded as inputs, with an image size of 64×64×6. In the encoder process, the input image is down-sampled by a series of convolution, pooling and dropout operations, resulting in a one-dimension vector (1×1×2048). In the decoder process, the one-dimension vector is up-sampled by deconvolution, dropout and convolution operations, resulting in a forecasted precipitation image (64×64×1). The down-sampling and up-sampling procedures are used to learn the nonlinear mapping relationships between predictors and predictands.

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

In the training process, the optimization algorithm is set to Adam (Kingma and Ba, 2014), which is a stochastic learning algorithm based on adaptive moment. The network learning rate is set to 0.001 and the stopping rule of iteration is that the validating error does not decrease for at least 100 times. The data uncertainty is propagated forward according to the law of uncertainty propagation and the dropout process is repeated 10 times with a dropout rate of 0.5. The random seed is set to 1 to enable the reproducibility of the experiment. The experimental data spans from 1980 to 2020 (2139 weeks), of which 60%, 20% and 20% of the data are used for training, validating and testing, respectively. The optimal model parameters are determined based on the minimal validating loss.

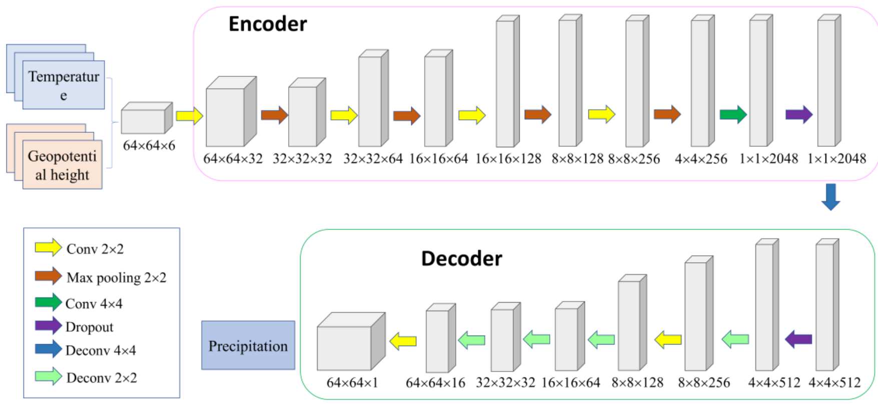


Figure 3: The developed deep learning model for precipitation forecasting.

The proposed deep learning framework for precipitation forecasting is demonstrated in Figure 4. The input data and its uncertainty are prepared and are considered as inputs for the forecasting model. The model weights are initialized and the input uncertainty is propagated forward according to the weights. The loss function value is calculated according to the forecasted value, the propagated uncertainty, target data and its uncertainty. The forecasting model is trained according to the optimization algorithm and then the trained model is obtained. Next the test data is used to produce the forecasted value and variance based on the model weights and uncertainty propagation. Finally, the forecasted value and variance are evaluated and compared with several precipitation forecasting methods.

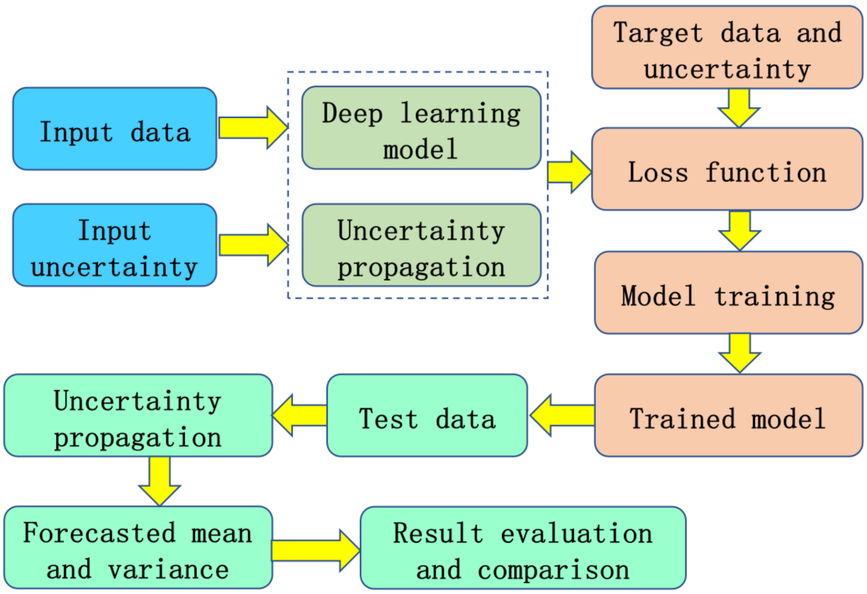


Figure 4: The proposed deep learning framework for precipitation forecasting.

We designed a series of comparison experiments to investigate the effect of different uncertainty processing strategies on the forecasting performance. The precipitation forecasting experiment without considering uncertainty is used as the baseline (Experiment 1). The mean square error is used as the loss function and the data and model uncertainties are not considered in Experiment 1. The uncertainty sources are incorporated differently into the experiments, including predictor uncertainty (Experiment 2), predictor and predictand uncertainties (Experiment 3), model uncertainty based on Srivastava et al. (2014)'s method (Experiment 4), data and model uncertainties based on Kendall and Gal (2017)'s method (Experiment 5), data and model uncertainties based on Loquercio et al. (2020)'s method (Experiment 6), and data and model uncertainties (Experiment 7) based on the proposed framework here. The data uncertainty only includes the propagated uncertainty from the input data in Equation (16) in Experiment 2, while the propagated uncertainty from the input data and the target data uncertainty are both included in Experiment 3. In Experiment 4, the data uncertainty is ignored and the model parameters are randomly sampled for 10 times to get the model spread. In Experiment 5, the input uncertainty is regarded as the trainable parameter and the trained uncertainty value is used as the data uncertainty in Equation (16), and the model uncertainty is considered the same way as Experiment 4. In Experiment 6, the input data uncertainty is propagated and the model uncertainty is modeled by sampling the parameters. In

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Experiment 7, the input uncertainty is propagated and the target data uncertainty is included in Equation (16) and the model uncertainty is represented by multiple sampling process.

The root mean square error (RMSE) statistic is used to measure the difference between forecasted value and true value.

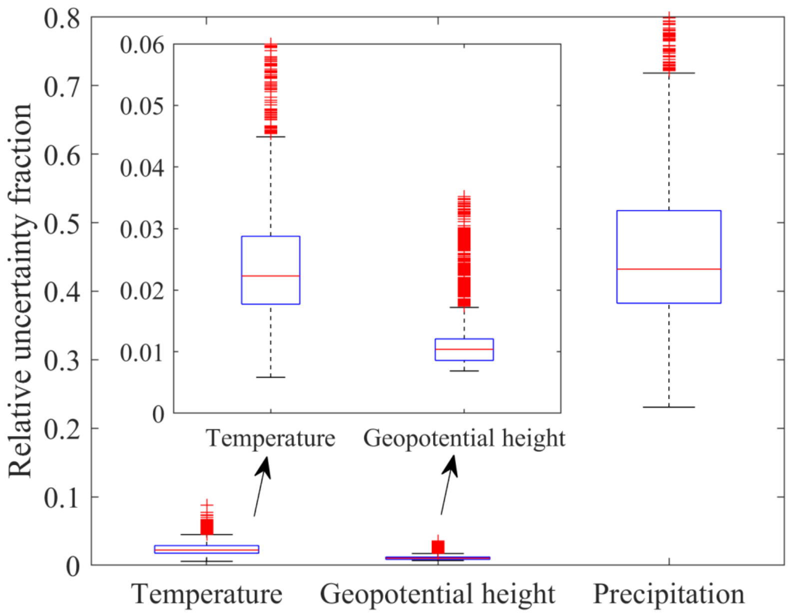
$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} \qquad (23)$$

where y_i is the true value or observation and $\hat{y}_i$ is the forecasted value; n is the sample size.

4 Results and discussion

4.1. The uncertainty of input and output datasets

The data uncertainty of predictors and predictand is calculated based on the TCH method and is shown in Figure 5. The precipitation data uncertainty is much higher than



https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Figure 5: The data uncertainty calculated by the TCH method. The uncertainty distribution is plotted according to the uncertainty over all the pixels of the study area. The red cross indicates the outliers.

the temperature and geopotential height variables, with a median of ~43% relative uncertainty fraction for precipitation, ~2% for temperature and ~1% for geopotential height. Therefore, the precipitation data suffer from greater uncertainty relative to the input data and the predictand uncertainty should not be ignored in the training process. The combination of the propagated input uncertainty and the predictand uncertainty is used as the adjusted parameter to regularize the loss function, which is relatively reasonable as the data with larger uncertainty should contribute less to the total training loss. It should be noted that the predictor and predictand data are normalized to [0,1] before uncertainty estimation to ensure a fair comparison of uncertainty value. The high uncertainty for precipitation data is related to strong spatiotemporal heterogeneity of precipitation and the high inconsistency among the reanalysis data (Xu et al., 2020b), while temperature and geopotential height data are much more homogeneous in space and time.

4.2. Overall precipitation forecasting performance

As for the predictive uncertainty, the forecasting method that only considers model uncertainty (Srivastava et al., 2014) obtains the minimum predictive uncertainty (Table 1). However, the data uncertainty is not considered when only sampling from the parameters and thus the impact of data error on forecasting is not evaluated. In Kendall and Gal (2017)'s method, the data uncertainty is regarded as a trainable parameter and the model uncertainty is modeled by random parameter sampling. Whether the learned parameter value for data uncertainty parameter could represent the real data error needs further investigation. Loquercio et al. (2020) used the law of uncertainty propagation to propagate the data uncertainty and sampled the parameters randomly during training. In our proposed method, the input data uncertainty, target data uncertainty and model uncertainty are jointly coupled by uncertainty propagation and random parameter sampling. The average predictive uncertainty (3.460) based on the proposed method is smaller than the Loquercio et al. (2020)'s and Srivastava et al. (2014)'s methods. In this regard, the proposed method could reduce the predictive uncertainty of precipitation forecasting to some extent, when jointly modeling data and model uncertainties. The proposed method could slightly improve the precipitation forecasting performance and could improve the reliability of precipitation forecasting by reducing the uncertainty.

When only the input uncertainty is modeled in the forecasting model, the predictive uncertainty is 12.950. If the target data uncertainty is coupled with input uncertainty, the predictive uncertainty is substantially reduced (3.022). In Equation (20), when the predictive error on the right side of the equation reaches local minimum and remains unchanged basically, the left side of the equation includes the input

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

uncertainty propagation and the target data uncertainty. When new data is used to make prediction, the predictive uncertainty is generated by the input uncertainty and the law of error propagation. Thus, when only the input uncertainty is modeled in Equation (20), the left side of this Equation equals to the propagated uncertainty from the input data. If the left side of the Equation (20) is replaced from the propagated input uncertainty with the combination of propagated input uncertainty and target uncertainty, the propagated input uncertainty after replacement will be smaller than that of no replacement, i.e. $(\sigma_x^{(l)})^2 = \sigma^2 - \sigma_y^2 < \sigma^2 = \left\|y_i - f^{\hat{U}_i}(x_i)\right\|^2$.

Table 1. The accuracy of precipitation forecasting based on different uncertainty processing strategies. The best RMSE is shown in bold for each column.

Uncertainty processing	RMSE	Uncertainty
No uncertainty	25.357	-
Predictor uncertainty	25.500	12.950
Predictor and predictand uncertainties	25.932	3.022
Model uncertainty (Srivastava et al., 2014)	25.368	1.778
Data and model uncertainties (Kendall and Gal, 2017)	25.225	4.914
Data and model uncertainties (Loquercio et al., 2020)	25.324	13.169
Data and model uncertainties (This study)	**25.199**	3.460

4.3. Spatial patterns of precipitation forecasting

In Figure 6, the spatial patterns of the RMSE for precipitation forecasting demonstrate some similarities and differences between different uncertainty processing strategies. Overall, the spatial distribution of RMSE is similar with each other and is smaller in the northwest region but larger in the southeast region. In the places where the annual rainfall is abundant, the water cycle process is accelerated and the precipitation observations may suffer from large uncertainty. The difficulty of forecasting extreme high precipitation volume also increases the average RMSE in the southeast region relative to the northwest region (Yuan et al., 2012; Huang et al., 2013). There are some differences of the forecasting error among different forecasting methods in local areas. For example, the forecasting performance based on our proposed method could outperform the methods in Experiments 1, 2 and 4 and is comparable with the methods in Experiments 3, 5 and 6 for the local areas covered by black circles in Figure 6.

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Hydrology and Earth System Sciences
Discussions

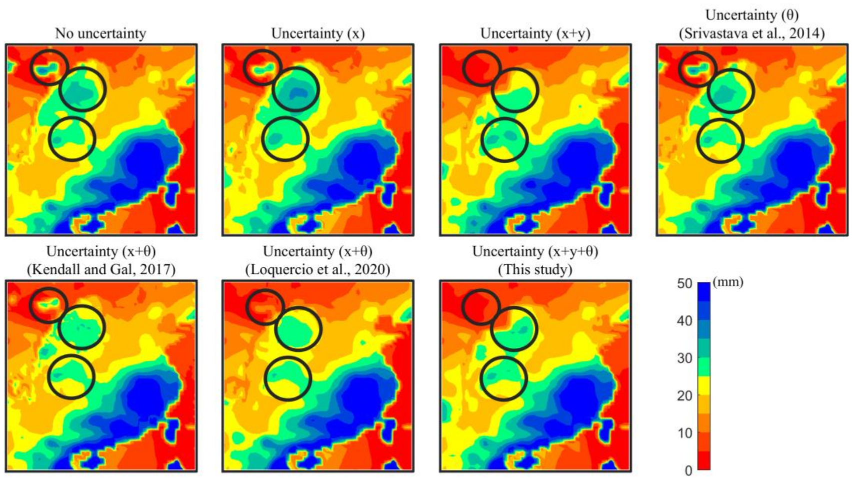


Figure 6: The spatial patterns of RMSE for precipitation forecasting. In this figure, x means the modeling of input uncertainty; x+y represents the modeling of input and output uncertainty; x+y+θ indicates the modeling of input uncertainty, output uncertainty and model uncertainty. The black circles represent the highlighted areas.

Figure 7 demonstrates the impact of different uncertainty processing methods on the predictive uncertainty of precipitation. If only the input uncertainty is considered in the forecasting model, the predictive uncertainty is large in the central and southwest regions. The predictive uncertainty could be substantially reduced when incorporating the target data uncertainty besides the input uncertainty. The modeling of model uncertainty only could produce the minimum predictive uncertainty spatially in the experiment. The predictive uncertainty is relatively small based on Kendall and Gal (2017)'s method. In Loquercio et al. (2020)'s method, the predictive uncertainty is close to the result of input uncertainty modeling in space, suggesting that the coupled modeling of input and model uncertainties fails to help reduce the forecasting spread. Our proposed method could include the input, target and model uncertainties jointly and could help reduce the predictive uncertainty to a large extent, relative to the methods in Experiment 2 and 6.

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

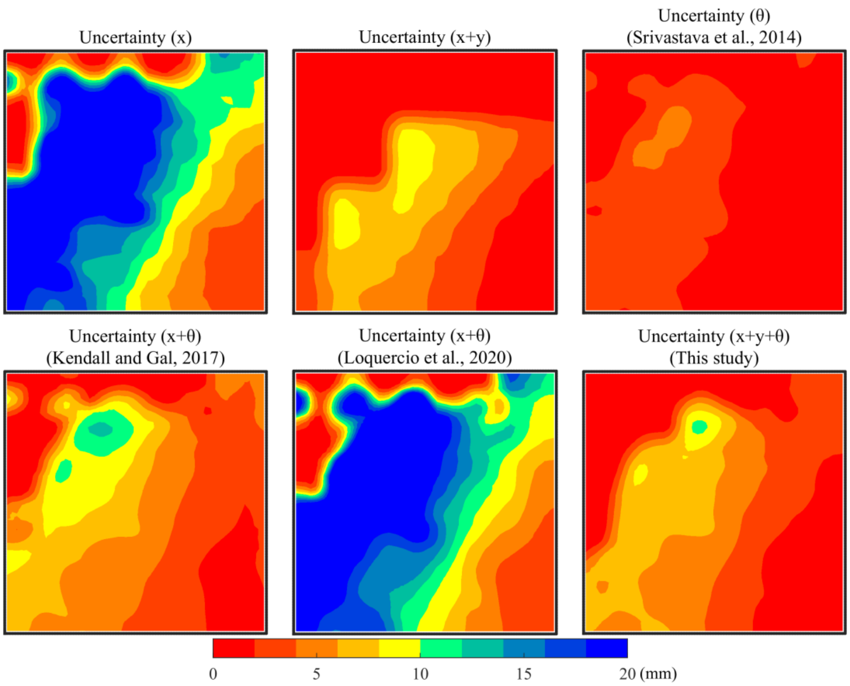


Figure 7: The spatial patterns of uncertainty for precipitation forecasting.

4.4. Uncertainty analysis and discussion

In precipitation forecasting, data and model uncertainties both bring uncertainty to the forecasting result. The higher the data and model uncertainties, the more divergent the forecasting, suggesting the forecasting less reliable. Therefore, the data and model uncertainties should be jointly considered in the forecasting process (Gal, 2016; Kendall and Gal, 2017; Loquercio et al., 2020; Parrish et al., 2012). Although the predictive error is close to each other among different forecasting methods in Figure 6 and Table 1, the predictive uncertainty has some discrepancies. The modeling of input uncertainty only in the forecasting model would bring high predictive uncertainty and ignore the target data uncertainty. The joint modeling of input and target uncertainties could reduce the predictive uncertainty substantially, which is related to the change of the variance in Equations (16) and (20) corresponding to the minimum value of the forecasting error term. The propagation of input uncertainty is constrained by refining the uncertainty representation in Equation (16) after incorporating the target uncertainty term and thus changing the weight training process.

The proposed method in this study could model the input uncertainty, target uncertainty and model uncertainty jointly and could reduce the predictive uncertainty relative to Kendall and Gal (2017)'s and Loquercio et al. (2020)'s methods. The developed method does not increase the training parameter and is a general forecasting uncertainty method for geophysical applications such as temperature forecasting, runoff forecasting and wind speed forecasting, especially for data-driven forecasting models (Ham et al., 2019; Zheng et al., 2020; Hossain et al., 2015).

In numerical precipitation forecasting systems, ensemble forecasting is commonly used to quantify the predictive uncertainty (Duan et al., 2019). In ensemble forecasting, the model parameters and data are perturbed to produce a forecasted ensemble and thus the data and model uncertainties are both considered. However, it would be time-consuming and cost-expensive to conduct large-sample sampling for complex physical models. In our developed method, the law of error propagation is used to propagate the data uncertainty. The uncertainty propagation of convolution, max-pooling and deconvolution in the deep learning forecasting model is tractable in an analytical form. However, the uncertainty propagation process is generally intractable analytically for complex statistical or physical models. Therefore, the theory and implementation technology for uncertainty modeling require further development, such as surrogate modeling, Monte Carlo methods, polynomial chaos expansions and Bayesian approaches (Linde et al., 2017; Sudret et al., 2017; Zhu and Zabaras, 2018; Schiavazzi et al., 2017; Nitzler et al., 2020).

5 Conclusion

In this study, we proposed a data-model uncertainty coupling framework to estimate the predictive uncertainty of precipitation forecasting. In this framework, the predictor and predictand uncertainties are estimated a prior by the TCH method and are assumed as Gaussian distribution. The predictor uncertainty is propagated forward during training and testing processes by the law of error propagation. The model uncertainty is represented by randomly abandoning model weights from deep learning layers. The data and model uncertainties are jointly modeled in the objective function during training and are also used during the testing process. The loss function is constructed by the MSE statistic adjusted by data uncertainty and a regularization term based on logarithmic data uncertainty. In the loss function, the adjusting parameter is determined by the combination of the square of predictor and predictand uncertainties. The forecasted ensembles are used to calculate the predictive mean and variance to estimate

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

the predictive uncertainty of precipitation.

The weekly precipitation forecasting in southern and northern China is used as an example to examine the effectiveness of the proposed joint uncertainty modeling framework. Temperature and geopotential height data in previous three weeks are used to forecast the precipitation in the target week. The forecasting model is developed based on an encoder-decoder deep neural network, with multivariate spatiotemporal predictor data as inputs and spatiotemporal precipitation data as output. The results exhibit comparable precipitation forecasting accuracy for the proposed method with several existing uncertainty processing strategies, while the predictive uncertainty is reduced relative to two data-model uncertainty modeling methods. The reduction of predictive uncertainty is significative for quantitative precipitation forecasting from a data-driven view.

The data-driven precipitation forecasting method has limitations in the interpretation part relative to numerical weather prediction. The precipitation forecasting accuracy for numerical models could still be improved by improving the parameterization schemes and resolving the uncertainties in observations, parameters and models. The proposed uncertainty modeling framework may also provide some insights for the uncertainty quantification in numerical prediction models. For example, the uncertainty propagation for the input data and the coupling with target data uncertainty could be used in a data assimilation scheme to estimate the propagated uncertainty in weather forecasting.

Data-driven precipitation forecasting could be used as a tool to assist regional prediction and warning of extreme weather events together with numerical models. The proposed joint data-model uncertainty modeling framework could help estimate the forecasting spread and is a general approach to derive predictive uncertainty for geophysical forecasting applications. Further research should focus on the non-Gaussian uncertainty modeling for complex integrated statistical-physical models.

Data availability

The meteorological data are publicly available and can be obtained via the website https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/ for MERRA-2, https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html for NCEP R2 and https://climate.copernicus.eu/climate-reanalysis for ERA-5.

Author contributions

LX conceptualized and wrote the paper. CY provided supervision of this study. NC gave support in developing the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgment

This research was supported by the National Key R&D Program (2018YFB2100500) and the National Natural Science Foundation of China program (41890822).

References

Ardabili, S., Mosavi, A., Dehghani, M., and Várkonyi-Kóczy, A. R.: Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review, International Conference on Global Research and Education, 52-62,

Badrinarayanan, V., Kendall, A., and Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39, 2481-2495, 10.1109/TPAMI.2016.2644615, 2017.

Boukabara, S.-A., Krasnopolsky, V., Stewart, J. Q., Maddy, E. S., Shahroudi, N., and Hoffman, R. N.: Leveraging modern artificial intelligence for remote sensing and NWP: Benefits and challenges, B. Am. Meteorol. Soc., 100, ES473-ES491, 2019.

Brooks, S.: Markov chain Monte Carlo method and its application, Journal of the royal statistical society: series D (the Statistician), 47, 69-100, 1998.

Chantry, M., Christensen, H., Dueben, P., and Palmer, T.: Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI, 2021.

Cho, K., Van Merriënboer, B., Bahdanau, D., and Bengio, Y.: On the properties of neural machine translation: Encoder-decoder approaches, arXiv preprint arXiv:1409.1259, 2014.

Duan, Q., Pappenberger, F., Wood, A., Cloke, H. L., and Schaake, J.: Handbook of hydrometeorological ensemble forecasting, Springer2019.

Gal, Y.: Uncertainty in deep learning, University of Cambridge, 1, 4, 2016.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., and Reichle, R.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), J. Clim., 30, 5419-5454, 2017.

Ghahramani, Z.: Probabilistic machine learning and artificial intelligence, Nature, 521, 452-459, 2015.

Gneiting, T. and Raftery, A. E.: Weather forecasting with ensemble methods, Science, 310, 248-249, 2005.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y.: Deep learning, 2, MIT press Cambridge2016.

Gruber, A., Dorigo, W. A., Crow, W., and Wagner, W.: Triple collocation-based merging of satellite soil moisture retrievals, IEEE Transactions on Geoscience and Remote Sensing, 55, 6780-6792, 2017.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., and Wagner, W.: Recent advances in (soil moisture) triple collocation analysis, International Journal of Applied Earth Observation and Geoinformation, 45, 200-211, 2016.

Ham, Y.-G., Kim, J.-H., and Luo, J.-J.: Deep learning for multi-year ENSO forecasts, Nature, 573, 568-572, 2019.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., and Schepers, D.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999-2049, 2020.

Hochreiter, S. and Schmidhuber, J.: Long short-term memory, Neural Comput., 9, 1735-1780, 1997.

Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J.: Stochastic variational inference, Journal of Machine Learning Research, 14, 2013.

Hossain, M., Rekabdar, B., Louis, S. J., and Dascalu, S.: Forecasting the weather of Nevada: A deep learning approach, 2015 international joint conference on neural networks (IJCNN), 1-6,

Huang, Y., Xue, J., Wan, Q., Chen, Z., Ding, W., and Zhang, C.: Improvement of the surface pressure operator in GRAPES and its application in precipitation forecasting in South China, Advances in Atmospheric Sciences, 30, 354-366, 2013.

Hwang, J., Orenstein, P., Cohen, J., Pfeiffer, K., and Mackey, L.: Improving subseasonal forecasting in the western US with machine learning, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2325-2335,

Jacoboni, C. and Lugli, P.: The Monte Carlo method for semiconductor device simulation, Springer Science & Business Media2012.

Kendall, A. and Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision?, arXiv preprint arXiv:1703.04977, 2017.

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

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., Van Den Dool, H., Saha, S., Mendez, M. P., and Becker, E.: The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction, B. Am. Meteorol. Soc., 95, 585-601, 2014.

Kobold, M. and Sušelj, K.: Precipitation forecasts and their uncertainty as input into hydrological models, Hydrol. Earth. Syst. Sc, 9, 322-332, 2005.

Kullback, S. and Leibler, R. A.: On information and sufficiency, The annals of mathematical statistics, 22, 79-86, 1951.

Linde, N., Ginsbourger, D., Irving, J., Nobile, F., and Doucet, A.: On uncertainty quantification in hydrogeology and hydrogeophysics, Adv. Water Resour., 110, 166-181, https://doi.org/10.1016/j.advwatres.2017.10.014, 2017.

Loquercio, A., Segu, M., and Scaramuzza, D.: A general framework for uncertainty estimation in deep learning, IEEE Robotics and Automation Letters, 5, 3153-3160, 2020.

Luo, Y., Zhang, R., Wan, Q., Wang, B., Wong, W. K., Hu, Z., Jou, B. J.-D., Lin, Y., Johnson, R. H., and

Chang, C.-P.: The southern China monsoon rainfall experiment (SCMREX), B. Am. Meteorol. Soc., 98, 999-1013, 2017.
McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., and Stoffelen, A.: Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, Geophys. Res. Lett., 41, 6229-6236, 2014.
Metropolis, N. and Ulam, S.: The monte carlo method, J. Am. Stat. Assoc., 44, 335-341, 1949.
Molinari, J. and Dudek, M.: Parameterization of convective precipitation in mesoscale numerical models: A critical review, Mon. Weather. Rev., 120, 326-344, 1992.
Nitzler, J., Biehler, J., Fehn, N., Koutsourelakis, P.-S., and Wall, W. A.: A generalized probabilistic learning approach for multi-fidelity uncertainty propagation in complex physical simulations, arXiv preprint arXiv:2001.02892, 2020.
Papacharalampous, G., Tyralis, H., Koutsoyiannis, D., and Montanari, A.: Quantification of predictive uncertainty in hydrological modelling by harnessing the wisdom of the crowd: A large-sample experiment at monthly timescale, Adv. Water Resour., 136, 103470, https://doi.org/10.1016/j.advwatres.2019.103470, 2020.
Parrish, M. A., Moradkhani, H., and DeChant, C. M.: Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation, Water Resour. Res., 48, 2012.
Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J. V., Heim, R. R., Brewer, M. J., Svoboda, M., and Westerhoff, R.: Toward global drought early warning capability: Expanding international cooperation for the development of a framework for monitoring and forecasting, B. Am. Meteorol. Soc., 94, 776-785, 2013.
Premoli, A. and Tavella, P.: A revisited three-cornered hat method for estimating frequency standard instability, IEEE Transactions on instrumentation and measurement, 42, 7-13, 1993.
Reeves, H. D., Elmore, K. L., Ryzhkov, A., Schuur, T., and Krause, J.: Sources of uncertainty in precipitation-type forecasting, Weather and forecasting, 29, 936-953, 2014.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 2019.
Rubinstein, R. Y. and Kroese, D. P.: Simulation and the Monte Carlo method, John Wiley & Sons2016.
Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., and Iredell, M.: The NCEP climate forecast system version 2, J. Clim., 27, 2185-2208, 2014.
Sahoo, S., Russo, T., Elliott, J., and Foster, I.: Machine learning algorithms for modeling groundwater level changes in agricultural regions of the US, Water Resour. Res., 53, 3878-3895, 2017.
Schiavazzi, D. E., Doostan, A., Iaccarino, G., and Marsden, A. L.: A generalized multi-resolution expansion for uncertainty propagation with application to cardiovascular modeling, Computer Methods in Applied Mechanics and Engineering, 314, 196-221, https://doi.org/10.1016/j.cma.2016.09.024, 2017.
Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c.: Convolutional LSTM network: A machine learning approach for precipitation nowcasting, arXiv preprint arXiv:1506.04214, 2015.
Sikder, S. and Hossain, F.: Assessment of the weather research and forecasting model generalized parameterization schemes for advancement of precipitation forecasting in monsoon-driven river basins, Journal of Advances in Modeling Earth Systems, 8, 1210-1228, 2016.
Sønderby, C. K., Espeholt, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., Agrawal, S., Hickey, J.,

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

and Kalchbrenner, N.: Metnet: A neural weather model for precipitation forecasting, arXiv preprint arXiv:2003.12140, 2020.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, The journal of machine learning research, 15, 1929-1958, 2014.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, Journal of geophysical research: oceans, 103, 7755-7766, 1998.

Sudret, B., Marelli, S., and Wiart, J.: Surrogate models for uncertainty quantification: An overview, 2017 11th European Conference on Antennas and Propagation (EUCAP), 19-24 March 2017, 793-797, 10.23919/EuCAP.2017.7928679,

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K. L.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons, Rev. Geophys., 56, 79-107, 2018.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485-498, 2012.

Trebing, K., Stan̉czyk, T., and Mehrkanoon, S.: Smaat-unet: Precipitation nowcasting using a small attention-unet architecture, Pattern Recog. Lett., 145, 178-186, 2021.

van den Hurk, B., Doblas-Reyes, F., Balsamo, G., Koster, R. D., Seneviratne, S. I., and Camargo, H.: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe, Clim. Dynam., 38, 349-362, 2012.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., and Fuentes, M.: The subseasonal to seasonal (S2S) prediction project database, B. Am. Meteorol. Soc., 98, 163-173, 2017.

Von Neumann, J. and Ulam, S.: Monte carlo method, National Bureau of Standards Applied Mathematics Series, 12, 36, 1951.

Wei, K., Ouyang, C., Duan, H., Li, Y., Chen, M., Ma, J., An, H., and Zhou, S.: Reflections on the catastrophic 2020 Yangtze River Basin flooding in southern China, The Innovation, 1, 100038, 2020.

Xu, L., Chen, N., and Zhang, X.: Global drought trends under 1.5 and 2 C warming, Int. J. Climatol., 39, 2375-2385, 2019.

Xu, L., Abbaszadeh, P., Moradkhani, H., Chen, N., and Zhang, X.: Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index, Remote Sens. Environ., 250, 112028, 2020a.

Xu, L., Chen, N., Moradkhani, H., Zhang, X., and Hu, C.: Improving Global Monthly and Daily Precipitation Estimation by Fusing Gauge Observations, Remote Sensing, and Reanalysis Data Sets, Water Resour. Res., 56, e2019WR026444, 2020b.

Xu, L., Chen, N., Zhang, X., Moradkhani, H., Zhang, C., and Hu, C.: In-situ and triple-collocation based evaluations of eight global root zone soil moisture products, Remote Sens. Environ., 254, 112248, 2021.

Yuan, X., Liang, X.-Z., and Wood, E. F.: WRF ensemble downscaling seasonal forecasts of China winter precipitation during 1982–2008, Clim. Dynam., 39, 2041-2058, 2012.

Zheng, G., Li, X., Zhang, R.-H., and Liu, B.: Purely satellite data–driven deep learning forecast of complicated tropical instability waves, Science advances, 6, eaba1482, 2020.

Zhu, Y. and Zabaras, N.: Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification, Journal of Computational Physics, 366, 415-447,

https://doi.org/10.5194/hess-2021-432
Preprint. Discussion started: 25 August 2021

Hydrology and Earth System Sciences
Discussions

https://doi.org/10.1016/j.jcp.2018.04.018, 2018.