



Bayesian Multi-modeling of Deep Neural Nets for Probabilistic Crop Yield Prediction

Peyman Abbaszadeh^{*}, Keyhan Gavahi, Atieh Alipour, Proloy Deb, Hamid Moradkhani^{*}

Center for Complex Hydrosystems Research, Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL, USA

ARTICLE INFO

Keywords:

Artificial Intelligence
Deep Learning
Model Uncertainty
Probabilistic Estimates

ABSTRACT

An imperative aspect of agricultural planning is accurate yield prediction. Artificial Intelligence (AI) techniques, such as Deep Learning (DL), have been recognized as effective means for achieving practical solutions to this problem. However, these approaches most often provide deterministic estimates and do not account for the uncertainties involved in model predictions. This study presents a framework that employs the Bayesian Model Averaging (BMA) and a set of Copula functions to integrate the outputs of multiple deep neural networks, including the 3DCNN (3D Convolutional Neural Network) and ConvLSTM (Convolutional Long Short-Term Memory), and provides a probabilistic estimate of soybean crop yield over a hundred counties across three states in the United States. The results of this study show that the proposed approach produces more accurate and reliable soybean crop yield predictions than the 3DCNN and ConvLSTM networks alone while accounting for the models' uncertainties.

1. Introduction

Soybean is one of the world's most significant protein and oil sources. Four countries including the U.S., Brazil, Argentina, and China, are the largest producers in the world with approximately 90% of total global soybean production (Schwalbert et al., 2020). The seasonal fluctuations of soybean produced by these countries have a significant impact on the global economy and financial market. According to the recent report by the United States Department of Agriculture (USDA, 2020), after Brazil, the U.S is currently the largest soybean producer in the world contributing to ~28.7% of global production. Soybean is the most produced crop in the US, followed by corn. Over 70% of soybeans grown in the country are turned into animal feed per year, while only 15% is produced for human consumption, and the rest is mainly exported.

Crop yield prediction plays a pivotal role in agricultural planning and management. It is also of great importance to food production and security at regional to global scales (Khaki and Wang, 2019; Pantazi et al., 2016). Reliable and timely crop yield prediction enables making timely import and export decisions to promote and reinforce national food security. This has become even more important as global warming

and the population continues to rise (Mueller et al., 2012; Rosenzweig et al., 2014; Karimiziarani et al., 2021). Empirical relationships have been widely used for crop yield prediction in agriculture (Doraiswamy et al., 2004; Funk and Budde, 2009; Johnson, 2014). These approaches mostly rely on the linearity assumption between the crop yield and other drivers such as canopy reflectance and weather data. Due to strong non-linearity and a high degree of autocorrelation among these variables, the yield forecasts are more prone to overfitting (Johnson et al., 2016; Schwalbert et al., 2020). To circumvent this problem, the Machine Learning (ML) models have been successfully utilized in many studies to predict crop yield (Chlingaryan et al., 2018).

The traditional ML techniques work based on feature extraction to predict crop yield (Ruß, 2009). However, it is often difficult to find optimal features when the ML algorithms are trained (Nevavuori et al., 2019; Abbaszadeh et al., 2018). As an alternative, Deep Learning (DL) techniques, such as Convolutional Neural Networks (CNNs) and Long Short Term Memory (LSTM) networks, have been developed in recent years and successfully used in some studies. For example, in CNNs, no features need to be precalculated as the feature extraction operation is already performed by the network's convolutional layers, and the optimal features are obtained in the training process (Nevavuori et al.,

^{*} Corresponding authors at: Center for Complex Hydrosystems Research, Department of Civil, Construction and Environmental Engineering, The University of Alabama, Tuscaloosa, AL, USA

E-mail addresses: pabbaszadeh@ua.edu (P. Abbaszadeh), kgavahi@crimson.ua.edu (K. Gavahi), aalipour@crimson.ua.edu (A. Alipour), pdeb1@ua.edu (P. Deb), hmoradkhani@ua.edu (H. Moradkhani).

<https://doi.org/10.1016/j.agrformet.2021.108773>

Received 18 June 2021; Received in revised form 6 October 2021; Accepted 8 December 2021

Available online 18 December 2021

0168-1923/© 2021 Elsevier B.V. All rights reserved.

2019). The advantage of DL techniques compared to traditional ML algorithms for crop yield prediction has been discussed in many studies (Chunjing et al., 2017; Khaki and Wang, 2019; Milioto et al., 2018; Nevavuori et al., 2019; Sa et al., 2018; You et al., 2017). Despite several applications of DL techniques in the field of crop yield prediction, most of the studies are based on deterministic prediction (Choudhury and Jones, 2014; Everingham et al., 2016; Jeong et al., 2016; Abbaszadeh et al., 2021), which do not provide any information on the uncertainty associated with the model predictions. The literature indicates that only a few studies have been conducted on probabilistic crop yield prediction, most of which are based on the traditional ML algorithms (Gyamerah et al., 2020; Salinas et al., 2020). However, these approaches may not be as effective and efficient as DL-based methods and mostly have suboptimal performance.

Although the probabilistic version of DL algorithms has been introduced in some studies in recent years (Krastanov and Jiang, 2017; Patel et al., 2016; Wang et al., 2017; Wu et al., 2016), they have not received much traction in the engineering literature mostly due to its high complexity and high resource requirements. To address this problem, in this study we aim to use DL techniques within a statistical framework to generate probabilistic crop yield predictions. The proposed framework is based on the Copula-Embedded Bayesian Model Averaging (COP-BMA) approach introduced and developed by Madadgar and Moradkhani (2014).

The multi-model ensemble is an effective approach to quantify prediction uncertainty due to the uncertainty in model formulation while producing more accurate and reliable predictions compared to any individual model simulation (Hagedorn et al., 2005). Model averaging has been recognized as an efficient strategy to combine an ensemble of models through a linear combination of different models. These include the Granger-Ramanathan averaging, Bates-Granger averaging (Granger and Ramanathan, 1984), AIC (Akaike Information Criterion), and BIC (Bayesian Information Criterion)-based model averaging (Buckland et al., 1997; Guthery et al., 2003; Hansen, 2008), which take the linear average of the deterministic model outputs and generate a combined single-value prediction. Despite the widespread use of these model averaging techniques, Hoeting et al. (1999) showed that the weights do not properly represent the actual contribution of single models, and instead proposed the Bayesian Model Averaging (BMA). In this method, the model weights are calculated based on Bayes' theorem through updating the prior information with the likelihood of model prediction given the observation (Raftery et al., 1997). Raftery et al. (2005a) used the Expectation-Maximization algorithm (EM) to estimate the model weights based on the performance of each model during a training period. Although the BMA approach has been rarely used in crop yield prediction studies (Huang et al., 2017), it has been widely employed in many other applications, such as hydrological modeling (Duan et al., 2007; Ajami et al., 2007; Madadgar and Moradkhani, 2013; Najafi and Moradkhani, 2016, 2015) climate projection (Miao et al., 2014), soil-plant simulation (Wöhling et al., 2015), and ecology (van Oijen et al., 2013).

In standard BMA (Raftery et al., 2005), the conditional Probability Distribution Function (PDF) of each model is assumed to follow a normal distribution. The literature showed that this assumption might not be valid for some forecast variables (Slougher et al., 2007), and the PDF should be carefully selected for better representing the posterior distribution of model outputs (Slougher et al., 2010). Moreover, data transformation is usually required to transform the model predictions to the space of posterior distribution. Madadgar and Moradkhani (2014) proposed an approach by integrating multivariate functions, called copula functions, into BMA to estimate the posterior distribution of model predictions without a need to assume the form of the posterior distribution and transform the model predictions. In this study, we use this approach within a framework to integrate the two state-of-the-art DL-based model predictions and generate probabilistic estimates of soybean yield for more than a hundred counties across three different

states in the US.

The rest of the paper is structured as follows. Section 2 describes the algorithms used in this study including the three-dimensional CNN (3DCNN), Convolutional LSTM (ConvLSTM), Bayesian Model Averaging (BMA), Copula-Embedded Bayesian Model Averaging (COP-BMA), and the proposed probabilistic crop yield prediction approach. Sections 3 and 4 summarize the datasets and the performance measures, respectively. Section 5 presents the results and discussions. Finally, the conclusion of this paper is provided in Section 6.

2. Material and Methods

This section first summarizes the two DL techniques, i.e., 3DCNN and ConvLSTM, and then introduces the proposed approach that works based on the COP-BMA algorithm. Note that here we briefly describe the 3DCNN and ConvLSTM approaches and for more information, the interested readers are referred to the original articles.

2.1. 3DCNN

CNNs were initially proposed to solve computer vision problems. The main idea of this approach dates back to the early 90s when LeCun et al. (1989) designed a CNN algorithm to recognize handwritten digits. CNN is a class of DL techniques that is known for its superb ability in classifying big data. Although 2DCNN mainly accounts for the spatial domain features, the 3DCNN has temporal direction providing spatiotemporal features of tensors (Zhong et al., 2018). 3DCNN can learn the correlation of temporal changes among the sequential images without a need to use any additional temporal learning method, which makes it a desirable approach for learning the spatiotemporal images (Ji et al., 2018; Pelletier et al., 2019).

A standard 3DCNN is composed of multiple layers including an input layer, a convolution layer, a pooling layer, a fully-connected layer, and an output layer. The convolutional layer is the main core of the CNN structure. Through the 3D filter, the convolutional operations are applied to the input data to produce the feature map through a set of activation functions (i.e., Rectified Linear Unit (ReLU) $f(x) = \max(0, x)$, Sigmoid function $f(x) = 1/(1 + e^{-x})$, and tanh function $f(x) = \tanh(x)$). The weights and biases of each filter are trained. The hyperparameters of convolutional operations include stride denoting the step size that filters move each time, padding (a process of adding layers of zeros to the input images to maintain the dimension of output as input), and filter size. Pooling layers are usually used between the convolutional layers to reduce the spatial size of data through down-sampling (Rao and Liu, 2020). The mathematical expression of the output value γ at position (x, y, z) on j th feature map in the i th 3D convolutional layer is as follows (Ji et al., 2013):

$$\gamma_{j,xyz}^{(i)} = \text{ReLU} \left(b_j^{(i)} + \sum_{m=1}^{M^{(i-1)}} \sum_{p=0}^{P^{(i)}-1} \sum_{q=0}^{Q^{(i)}-1} \sum_{r=0}^{R^{(i)}-1} w_{jm,pqr}^{(i)} \gamma_{m,(x+p)(y+q)(z+r)}^{(i-1)} \right) \quad (1)$$

where $\text{ReLU}(\cdot)$ indicates elementwise ReLU function, $b_j^{(i)}$ denotes the bias for the j th feature map, $w_{jm,pqr}^{(i)}$ is the (p, q, r) th value of the 3D filter for the j th feature map at the i th layer associated with the m th feature map in the $(i-1)$ th layer. $M^{(i-1)}$ represents the number of feature maps at $(i-1)$ th layer, and $P^{(i)}$, $Q^{(i)}$ and $R^{(i)}$ are the size of the 3D filter at i th layer.

The output of the convolution and pooling process is flattened into a single vector of values such that each represents a probability that a certain feature belongs to a label. Therefore, a Fully Connected (FC) layer, operated at the end of the 3DCNN layer, take the flattened tensor as the input and map it to the output vector. This process can be expressed through the following mathematical relationship:

$$\gamma^{(i)} = \sigma(W^{(i)}\gamma^{(i-1)} + b^{(i)}) \quad (2)$$

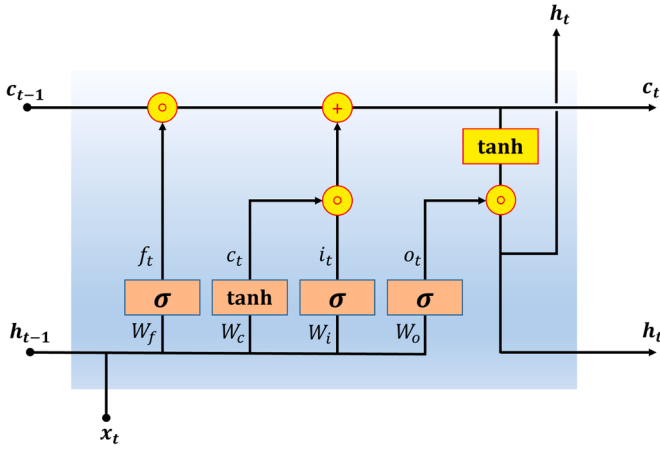


Figure 1. The inner structure of an LSTM cell. ‘ $\circ$ ’ denotes the Hadamard product. σ and $\tanh$ represent hyperbolic tangent and sigmoid activation functions.

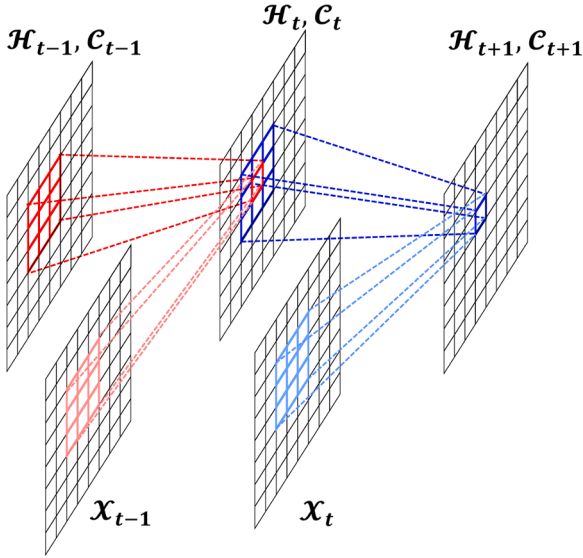


Figure 2. The inner structure of the ConvLSTM cell. [This Figure should be printed in color]

where $\gamma^{(i-1)}$ and $\gamma^{(i)}$ are the input and output of the adjacent i th layer. $\sigma(\cdot)$ is the activation function that operates elementwise. $W^{(i)}$ and $b^{(i)}$ are the trainable weight matrix and bias vector between the i th and $(i-1)$ th of the FC layer. For more information, we refer the readers to [Ji et al., \(2013\)](#).

2.2. ConvLSTM

The Long Short-Term Memory (LSTM) is a special configuration of the Recurrent Neural Network (RNN) structure that is known for its ability for modeling long-range dependencies ([Hochreiter and Schmidhuber, 1997](#)). The LSTM network encompasses three gates, including the forget gate, input gate, an output gate, that update and control the cell states. The gates use hyperbolic tangent and sigmoid activation functions. [Figure 1](#) illustrates the inner structure of an LSTM cell. Here, we only describe the key equations based on which the LSTM structure operates. For more information, we refer the readers to [Hochreiter and Schmidhuber \(1997\)](#).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (4)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (6)$$

$$h_t = o_t \circ \tanh(c_t) \quad (7)$$

Although the LSTM structure has been proven to be the most stable for handling temporal correlation, neither in input-to-state nor in state-to-state transitions the spatial information is encoded, which makes it inappropriate for handling spatiotemporal data. To cope with this problem, an extension of LSTM, which has convolutional layers in both input-to-state and state-to-state transitions, the ConvLSTM has been introduced (see [Figure 2](#)). This network is well-suited for the application of spatiotemporal sequence prediction. In ConvLSTM, all the inputs $\mathcal{X}_1, \dots, \mathcal{X}_t$, cell outputs $\mathcal{C}_1, \dots, \mathcal{C}_t$, hidden states $\mathcal{H}_1, \dots, \mathcal{H}_t$, and gates i_t, f_t and o_t are 3D tensors whose rows and columns represent the spatial dimension. Here, we only describe the ConvLSTM equations, and for more information, we refer the readers to [Shi et al., \(2015\)](#) who provided a comprehensive description of this approach. In these equations, ‘ $*$ ’ and ‘ $\circ$ ’ denote the convolution operator and the Hadamard product, respectively.

$$i_t = \sigma(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \circ \mathcal{C}_{t-1} + b_i) \quad (8)$$

$$f_t = \sigma(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \circ \mathcal{C}_{t-1} + b_f) \quad (9)$$

$$c_t = f_t \circ \mathcal{C}_{t-1} + i_t \circ \tanh(W_{xc} * \mathcal{X}_t + W_{hc} * \mathcal{H}_{t-1} + b_c) \quad (10)$$

$$o_t = \sigma(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \circ \mathcal{C}_t + b_o) \quad (11)$$

$$\mathcal{H}_t = o_t \circ \tanh(\mathcal{C}_t) \quad (12)$$

2.3. Bayesian Model Averaging

Bayesian Model Averaging (BMA) is a method that integrates the predicted forecast densities from multiple models to produce a new forecast Probability Density Function (PDF). Based on the law of total probability, the predictive distribution of a forecast variable y , given the observations Y during the training period and the independent predictions of k models, can be expressed by:

$$p(y|M_1, M_2, \dots, M_{k=k}, Y) = \sum_{i=1}^K p(M_i|Y)p(y|M_i, Y) \quad (13)$$

where $p(y|M_i, Y)$ is the posterior distribution of y given the model prediction M_i and training data Y . $p(M_i|Y)$ is the likelihood of model prediction given the observations Y during the training period, which also reflects the weight of each model M_i . Therefore, the outcome of the BMA approach is the weighted average of forecast PDF generated by each model. Since the model predictions are time-variant, [equation \(13\)](#) can be rewritten as:

$$p(y^t|M_1^t, M_1^t, \dots, M_k^t, Y) = \sum_{i=1}^K w_i p(y^t|M_i^t, Y) \quad (14)$$

Note that w represents the performance of the model during the training period. To solve this equation, it is usually assumed that the posterior distribution follows a Gaussian distribution with mean f_i^t and variance σ_i^2 , such that $p(y^t|f_i^t, Y) \sim g(y^t|f_i^t, \sigma_i^2)$. It should be noted that for non-Gaussian forecast variables, a power transformation (e.g., Box-Cox) is used to map them from their original space to a Gaussian space. Using the following log-likelihood function, the variance and weight of each forecast model can be estimated. [Raftery et al., \(2005\)](#) developed a procedure called the Expectation-Maximization (EM) algorithm to

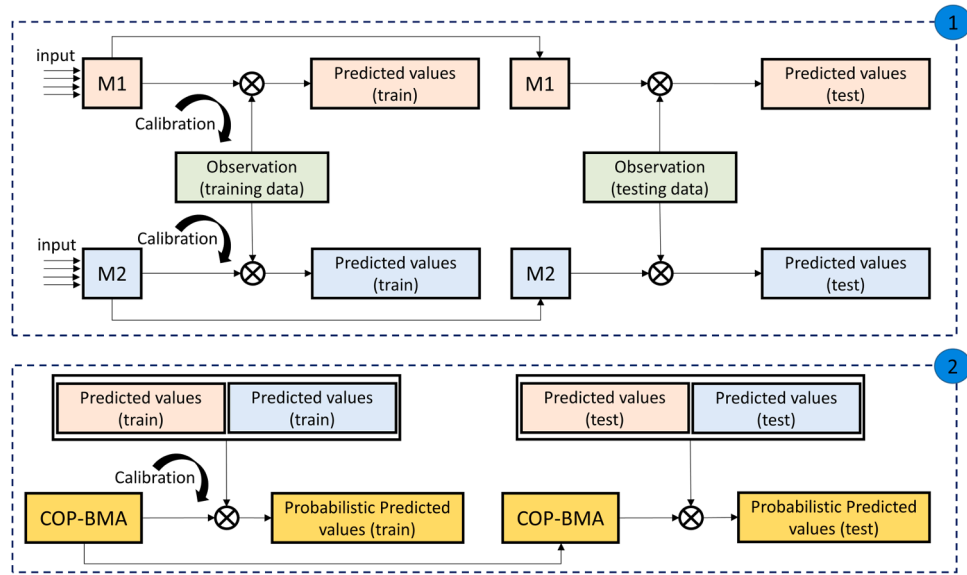


Figure 3. Schematic of the proposed framework. Step 1 represents two different Machine Learning (ML) models (i.e., M1 and M2) used to predict the testing data. The testing data is unseen during the training (model calibration) process. Step 2 shows the implementation of the COP-BMA model to integrate the predicted testing values obtained in Step 1 and generate probabilistic simulations.

maximize the [equation \(15\)](#), which is not analytically computable.

$$I(\theta) = \log \left(\sum_{i=1}^K w_i \cdot p(y|f_i, Y) \right) \quad (15)$$

EM algorithm iteratively updates the variance (σ_i^2) and weight (w_i) using the [equations \(18-21\)](#) until the termination criterion of $|I(\theta_{\text{iter}}) - I(\theta_{\text{iter}-1})|$ becomes less than ϵ . In the first iteration, the initial weight and variance for each model are set to:

$$w_{i,\text{iter}} = \frac{1}{K} \quad (16)$$

$$\sigma_{i,\text{iter}}^2 = \frac{1}{KT} \sum_{t=1}^T \sum_{k=1}^K (y^t - f_i^t)^2 \quad (17)$$

Then the algorithm proceeds based on the following equations.

$$w_{i,\text{iter}} = \frac{1}{T} \sum_{t=1}^T z_{i,\text{iter}}^t \quad (18)$$

$$\sigma_{i,\text{iter}}^2 = \frac{\sum_{t=1}^T z_{i,\text{iter}}^t \cdot (y^t - f_i^t)^2}{\sum_{t=1}^T z_{i,\text{iter}}^t} \quad (19)$$

$$z_{i,\text{iter}}^t = \frac{w_{i,\text{iter}-1} \cdot g(y^t | f_i^t, \sigma_{i,\text{iter}-1}^2)}{\sum_{i=1}^K w_{i,\text{iter}-1} \cdot g(y^t | f_i^t, \sigma_{i,\text{iter}-1}^2)} \quad (20)$$

$$I(\theta_{\text{iter}}) = \log \left(\sum_{i=1}^K w_{i,\text{iter}} \sum_{t=1}^T g(y^t | f_i^t, \sigma_{i,\text{iter}}^2) \right) \quad (21)$$

where T is the length of the training period and z is a latent variable.

2.4. Copula-Embedded Bayesian Model Averaging

As mentioned earlier, in the BMA approach the forecast PDFs are generally assumed to be a parametric distribution, such as Gaussian distribution. [Madadgar and Moradkhani \(2014\)](#) proposed an approach called the Copula-Embedded Bayesian Model Averaging (COP-BMA) that modifies the BMA predictive distribution through relaxing the assumption on parametric posterior distribution, which results in

increased multimodeling reliability. In this approach, $g(y|f_i, \sigma_i^2)$ is replaced with a group of multivariate copula functions. In the COP-BMA approach, the posterior distribution of forecast variables for each model $p(y^t | f_i^t, Y)$ is estimated using copula functions. Based on copula functions, a multivariate distribution $P(x_1 \dots x_n)$ can be expressed as follows:

$$P(x_1, \dots, x_i, \dots, x_n) = C[P(x_1), \dots, P(x_i), \dots, P(x_n)] = C(u_1, \dots, u_i, \dots, u_n) \quad (22)$$

where C is the Cumulative Distribution Function (CDF) of the copula and $P(x_i)$ is the marginal distribution of x_i being uniform on the interval $[0,1]$, which is denoted by u_i . The joint probability density function of $(x_1 \dots x_n)$ can be expressed by:

$$P(x_1, \dots, x_n) = c(u_1, \dots, u_n) \prod_{i=1}^n p(x_i) \quad (23)$$

where c represents the PDF of the copula. The conditional probability distribution of x_1 given x_2 is also defined as:

$$p(x_1 | x_2) = \frac{p(x_1, x_2)}{p(x_2)} \quad (24)$$

This equation can be revised by replacing the joint probability distribution of $p(x_1, x_2)$ in [equation \(24\)](#).

$$p(x_1 | x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \frac{c(u_1, u_2) \cdot p(x_1) \cdot p(x_2)}{p(x_2)} = c(u_1, u_2) \cdot p(x_1) \quad (25)$$

In this equation, x_1 and x_2 are considered as the forecasted variable y^t and the i th model prediction f_i^t , respectively. If the posterior distribution in [equation \(13\)](#) is replaced with the conditional probability distribution in (25), the predictive distribution of BMA becomes:

$$p(y^t | f_1^t, f_2^t, \dots, f_k^t, Y) = \sum_{i=1}^K w_i p(y^t | f_i^t, Y) = \sum_{i=1}^K w_i c(u_{y^t}, u_{f_i^t}) p(y^t) \quad (26)$$

Unlike the BMA approach where the posterior distribution $p(y|f_i, Y)$ is computed using the EM algorithm, in COP-BMA, it is directly achievable from [equation \(26\)](#). This equation also relaxes any assumption on the type of posterior distribution and also removes bias from model predictions. After defining the posterior distribution, we use the EM

algorithm with a few adjustments to estimate their weights.

$$w_{i,iter} = \frac{1}{T} \sum_{t=1}^T z_{i,iter}^t \quad (27)$$

$$z_{i,iter}^t = \frac{w_{i,iter-1} \cdot p(y^t | f_i^t)}{\sum_{i=1}^K w_{i,iter-1} \cdot p(y^t | f_i^t)} = \frac{w_{i,iter-1} \cdot c(u_{y^t}, u_{f_i^t}) p(y^t)}{\sum_{i=1}^K w_{i,iter-1} \cdot c(u_{y^t}, u_{f_i^t}) p(y^t)} \quad (28)$$

$$I(\theta_{iter}) = \log \left(\sum_{i=1}^K w_{i,iter} \sum_{t=1}^T c(u_{y^t}, u_{f_i^t}) p(y^t) \right) \quad (29)$$

2.5. The Proposed Approach

In this study for the first time utilizes the Bayesian multi-modeling to integrate different ML-based model outputs and generate the probabilistic soybean predictions while accounting for the uncertainties involved in the model predictions. Figure 3 illustrates the proposed framework. It uses the deterministic outputs from two different ML models (M1 and M2) to generate probabilistic predictions. Step 1 shows the training and testing of ML models, which are 3DCNN and ConvLSTM. Step 2 indicates how the COP-BMA approach is used for generating probabilistic simulations. It should be noted that the testing data is unseen during the ML model development process and therefore it can be properly used for validation of the predicted testing values in both steps 1 and 2. To implement the COP-BMA approach, those training datasets which had been used for ML model training/calibration in step 1 are used to train the COP-BMA model parameters. Then, we use the predicted deterministic values from ML models in step 1 as input to the trained COP-BMA model to generate probabilistic predicted testing values, which is hereafter referred to as COP-BMA results.

3. Datasets

In this study, the input variables of the ML models are MODIS (Moderate Resolution Imaging Spectroradiometer) surface reflectance, MODIS land cover, and MODIS land surface temperature. The MODIS/Terra surface reflectance product provides 7 bands of surface spectral reflectance at 500 m spatial resolution every 8 days. Each pixel contains the best possible surface reflectance observation value selected from all the acquisitions within the 8-day window. The product is publicly available at <https://lpdaac.usgs.gov/products/mod09a1v006/>. Here, we used all 7 bands of version 6 of this product as input to 3DCNN and ConvLSTM networks (Gavahi et al., 2021). The Terra and Aqua combined MODIS Land Cover type product provides yearly land cover types derived from six classification schemes at 500 m spatial resolution. The University of Maryland (UMD) classification (land cover type 2) was used in this study to mask the cropland areas. The dataset can be retrieved from <https://lpdaac.usgs.gov/products/mcd12q1v006/>. The MODIS Version 6 Land Surface Temperature (LST) provides an average 8-day per-pixel daytime and nighttime surface temperature at 1 km spatial resolution. The temperature is collected by using 7 thermal infrared bands using the LST algorithm. Both daytime and nighttime LSTs were used in the input layer of the deep NNs (i.e., 3DCNN and ConvLSTM) used in this study. This data is available at <https://lpdaac.usgs.gov/products/myd11a2v006/>. The models' output variable is the soybean yield data. This study aims at predicting the soybean yield data probabilistically across three states of Kansas, Louisiana, and Kentucky at the county scale. The dataset was collected from the USDA (United States Department of Agriculture) National Agricultural Statistical Services (NASS) repository available at https://www.nass.usda.gov/Quick_Stats/index.php. The soybean yield data time series were retrieved from 2003 to 2019. The first 15 years, from 2003 to 2017 (62 counties $\times$ 15 years = 930 data), were used for training the ML models and the COP-BMA model. The remaining two years of 2018 and 2019 (62

Table 1

Summary of performance measures used in this study

Performance Measure	Mathematical Representation
Bias	$\sum_{t=1}^T (y'_t - y_t)$
Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{T} \sum_{t=1}^T (y'_t - y_t)^2}$
Normalized Root Mean Square Difference (RMSD)	$\sqrt{\frac{1}{T} \sum_{t=1}^T ((y'_t - \bar{y}_t) - (y_t - \bar{y}_t))^2}$
Correlation Coefficient (R)	$\frac{1}{T-1} \sum_{t=1}^T \left(\frac{y'_t - \mu'}{\sigma'} \right) \left(\frac{y_t - \mu}{\sigma} \right)$
Mean Absolute Error (MAE)	$\frac{1}{T} \sum_{t=1}^T y'_t - y_t $
Coefficient of Variation (CV)	$\frac{\sigma'}{\mu'}$
Residual Standard Deviation (Std. Dev)	$\sqrt{\frac{\sum (y'_t - y_t)^2}{T-2}}$

counties $\times$ 2 years = 124 data), were used for testing the trained models and evaluating their predictive skills. In this study, 20% validation set was used for tuning deep NN models' hyperparameters. Cross-validation and early stopping were performed to avoid overfitting (more information can be found in our previous study, Gavahi et al., 2021).

4. Performance Measures

To assess the usefulness and effectiveness of the ML models used in this study and also the proposed COP-BMA approach, we use the performance measures listed in Table 1. In the following equations, y_t and y'_t are the observation and model prediction at time t , respectively. σ and σ' represent the standard deviation of the observation and model prediction, respectively. μ and μ' are the mean of observation and model prediction, respectively. T is the total time steps.

5. Results and Discussions

Figure 4 shows the performance of two deep learning algorithms and also their combined version, called the COP-BMA. Here we assess the usefulness of deep learning models in predicting the soybean crop yield and compare them with the COP-BMA approach through multiple deterministic performance measures (i.e., R, Bias, RMSE, and MAE). Except for MAE, all the measures indicate that the ConvLSTM performs better than the 3DCNN structure. Please note that this figure reports the predicted soybean values over 62 counties for the two years of 2018 and 2019. Depending on the nature of the predicted values (i.e., outliers, high values, and variance), the results of the MAE and RMSE can be different. The MAE is a linear score in which all the individual differences are weighted equally. While RMSE is a quadratic scoring rule which measures the average magnitude of the error and gives a relatively high weight to large errors. It should be noted that in this figure, all the model simulations for three states at the county level are collectively analyzed. The results clearly show that the COP-BMA approach outperforms both deep neural nets. Unlike the 3DCNN and ConvLSTM that provide a predicted value deterministically, the COP-BMA results in a probabilistic crop yield prediction. In this study, we have performed multiple experiments to find the optimum architecture for each 3DCNN and ConvLSTM model. The 3DCNN network consists of 5 blocks, each of which containing one Conv3D and one MaxPooling layer. The first block serves as a dimension reduction layer that includes two Conv3D layers with 9 and 3 filters. The extracted features from the last block are then connected to a flatten layer followed by two dense layers with 1024 and 1 neurons, respectively, and a dropout layer with 0.5 probability in between (Gavahi et al., 2021). For the ConvLSTM model, ConvLSTM2D layers with 10 and 5 filters with BatchNormalization and Maxpooling layers in between, were consecutively connected (LSTM block). Similar to the 3DCNN network, the spatiotemporal features extracted by the LSTM block were then flattened and then

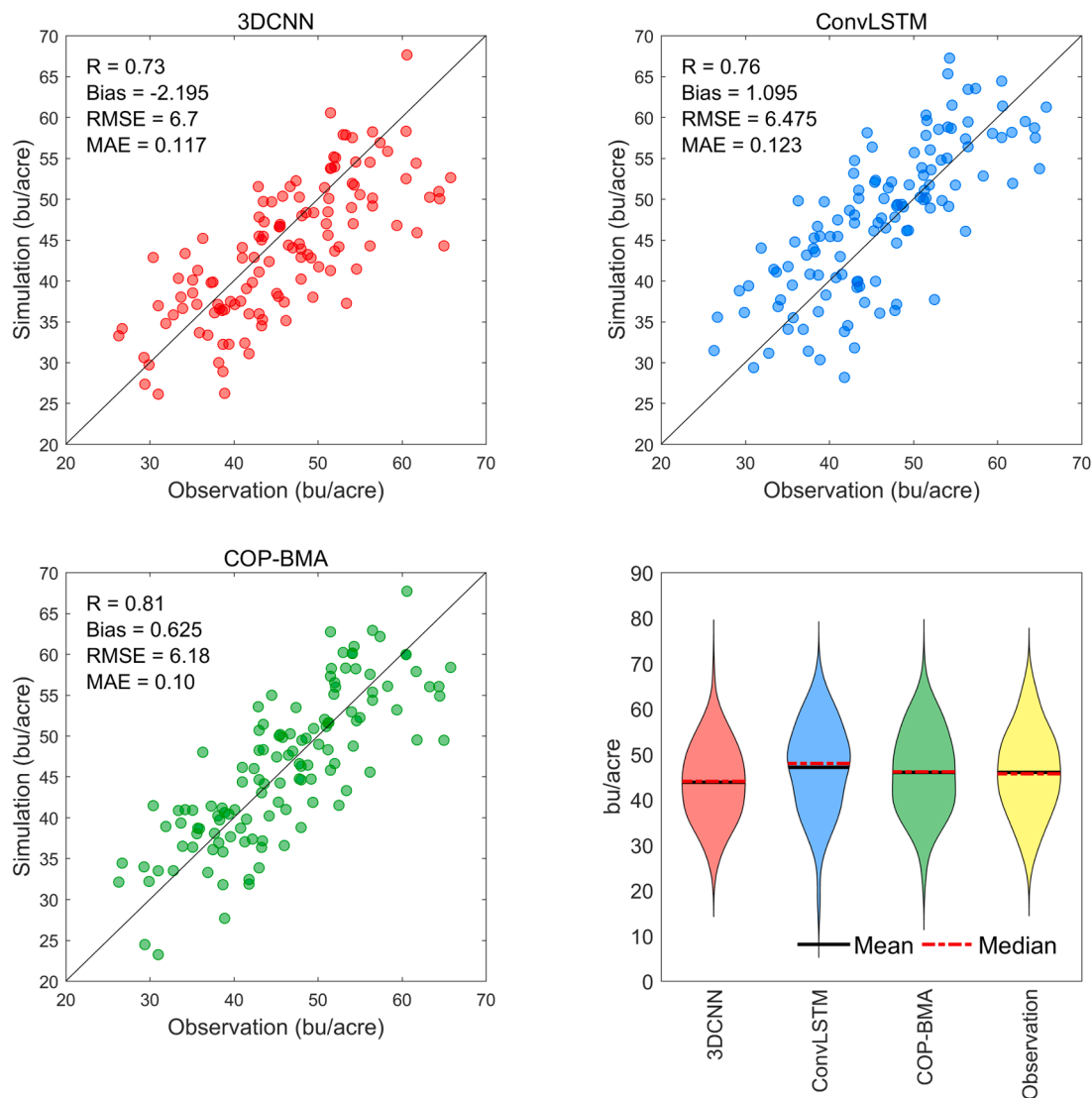


Figure 4. Performance of the three models used in this study, i.e., 3DCNN, ConvLSTM, and their combined version using the COP-BMA (3DCNN+ConvLSTM). The results are reported for the testing period. The boxplots also represent the model simulations' distributions compared to that of the observation. Bu/acre denotes “bushels per acre”.

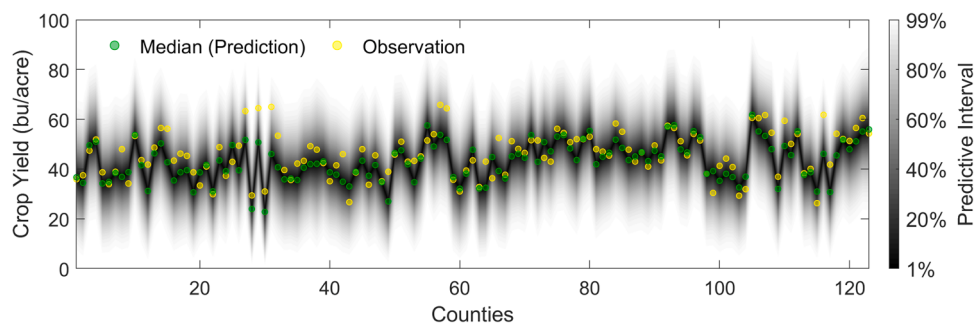


Figure 5. Probabilistic simulation of crop yield using the COP-BMA approach. The green points indicate the median of the predicted distribution, and the yellow points represent the observation. The shaded area around each point displays the predictive interval at different percentages ranging from 1% to 99%. [This Figure should be printed in color]

consecutively fed to a Dense (1024), Dropout (0.5), and Dense (1) layers. For a more detailed explanation of finding the optimum architecture of each network, we refer the interested readers to our previous work (Gavahi et al., 2021). At each time step, the COP-BMA integrates the deterministic predicted values from 3DCNN and ConvLSTM networks

within a Bayesian framework and provides a PDF, such that its median value and distribution represent the predicted value and its associated uncertainty, respectively. Also, Figure 4 reports the distribution of the predicted soybean crop yield values by three approaches and compares them with that of the observation data. The COP-BMA results in

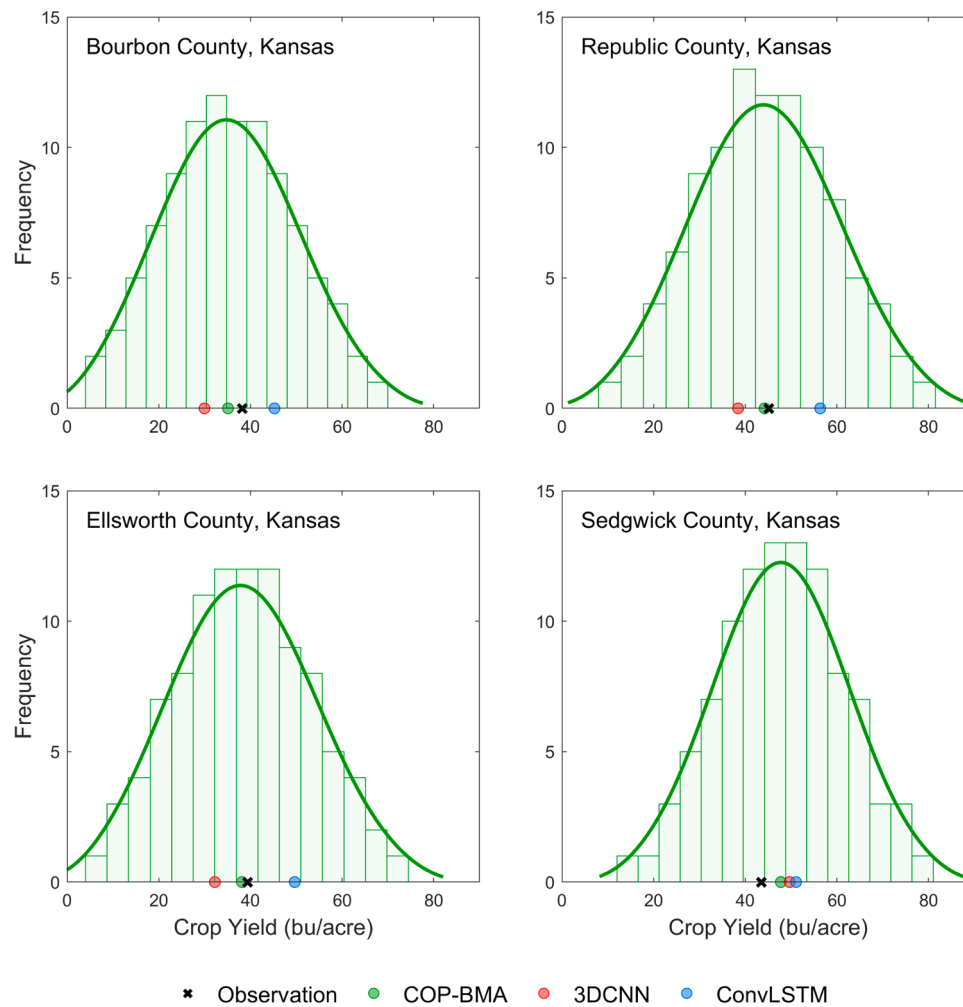


Figure 6. Accuracy of three models used in this study, i.e., 3DCNN, ConvLSTM, and their combined version using the COP-BMA (3DCNN+ConvLSTM) in predicting crop yield at four different counties in Kansas. The green curve shows the Kernel distribution fitted to the COP-BMA predicted simulations, and its corresponding median value is shown by a green point. [This Figure should be printed in color]

predicted values with a distribution very similar to that of the observation.

Figure 5 shows the probabilistic simulation of soybean crop yield provided by the proposed approach across more than 100 counties in three states (i.e., Kansas, Louisiana, and Kentucky). As this figure indicates, the median of the predicted distribution, shown as green points, closely follows the observation, and in most cases, they fall within the 20% predictive interval, which implies the higher accuracy of the predicted crop yield by the proposed methodology compared to the 3DCNN and ConvLSTM networks. The developed framework provides the deterministic estimates of the crop yield along with their associated uncertainty. Here, the uncertainty in crop yield prediction is associated with the structure of the deep neural networks including their parameterization, topology, and input variables. The median of the predicted soybean yield predictions along with the associated uncertainty interval helps the farmers and agriculture sectors to effectively decide and plan for the growing season while taking into account the risk of crop production failure. The proposed methodology is a general strategy that provides the possibility of taking advantage of multiple predictive crop models to generate more accurate crop yield values considering the uncertainty associated with the crop model choice. Therefore, regardless of the type of model used that is either physically-based data-driven models, or a combination of both, the proposed methodology can be used to integrate all the model outputs and provide more accurate and reliable crop yield predictions compared to the individual models. In

this study, although the uncertainty associated with the hydrometeorological and agrometeorological data are not directly taken into account, it is presumed that the uncertainty imposed by input data is already represented in the ML model outputs for which the Copula function is fitted.

In Figure 6, we display the results for four counties in Kansas State to clearly show how the developed approach results in improved prediction of soybean crop yield. It is important to note that in three cases (i.e., Bourbon County, Republic County, and Ellsworth County) the 3DCNN and ConvLSTM networks underestimated and overestimated the crop yield, respectively. However, the proposed COP-BMA approach by integrating the model outputs from two networks and providing probabilistic crop yield estimates resulted in a more accurate predicted value (the median of the fitted Kernel distribution). In Sedgwick County, both 3DCNN and ConvLSTM networks overestimated the crop yield, however, our probabilistic approach reduced the bias and increased the prediction accuracy. As per USDA (please see https://www.nass.usda.gov/Statistics_by_State/Kansas/Publications/County_Estimates/18KSsoy.pdf and https://www.nass.usda.gov/Statistics_by_State/Kansas/Publications/County_Estimates/19KSsoy.pdf for more information), for the years 2018 and 2019 soybean yields are observed to be 38.7 bu/ac and 35.9 bu/ac respectively for Bourbon County, Kansas. Similarly, for Republic and Sedgwick counties, the soybean yield is reported to be 51.4 bu/ac and 45.1 bu/ac, and 44.0 bu/ac and 36.6 bu/ac respectively for the corresponding 2018 and 2019 years. Although these three counties

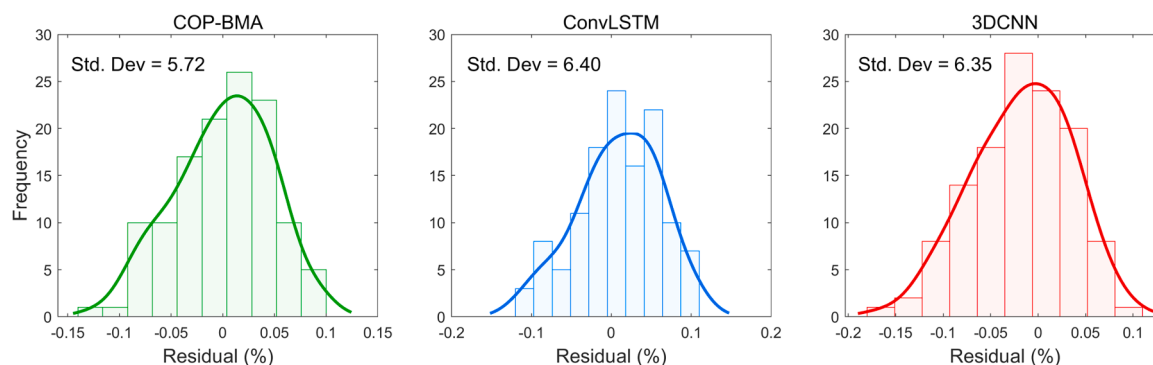


Figure 7. Residuals (%) histogram for three models used in this study, i.e., 3DCNN, ConvLSTM, and their combined version using COP-BMA (3DCNN+ConvLSTM). Std. Dev represents the standard deviation of residual values computed between the observation and model simulation.

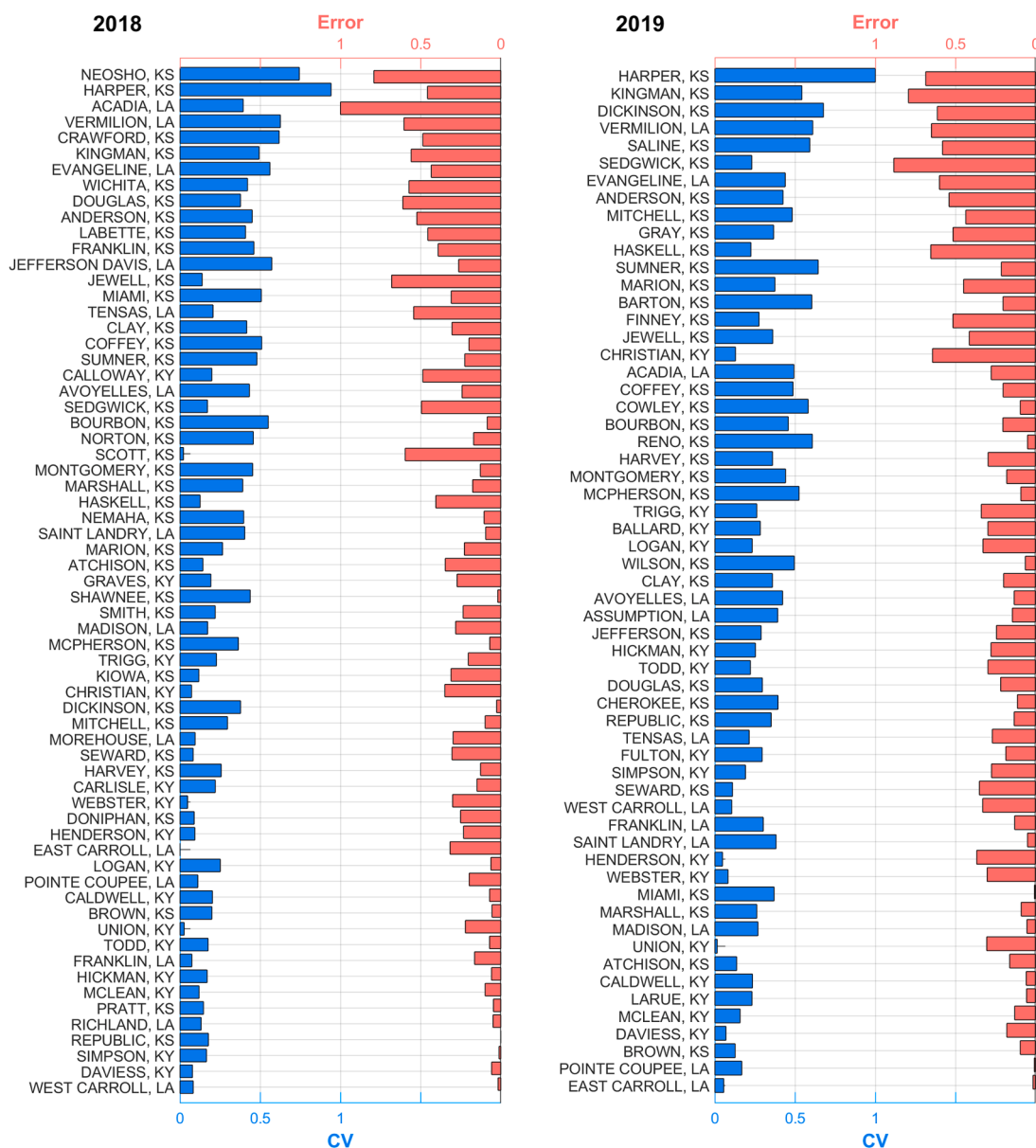


Figure 8. The probabilistic (CV) and deterministic (Error) performance measures were reported for the predicted crop yield across more than a hundred counties in the year 2018 and 2019. CV represents the coefficient of variation. [This Figure should be printed in color]

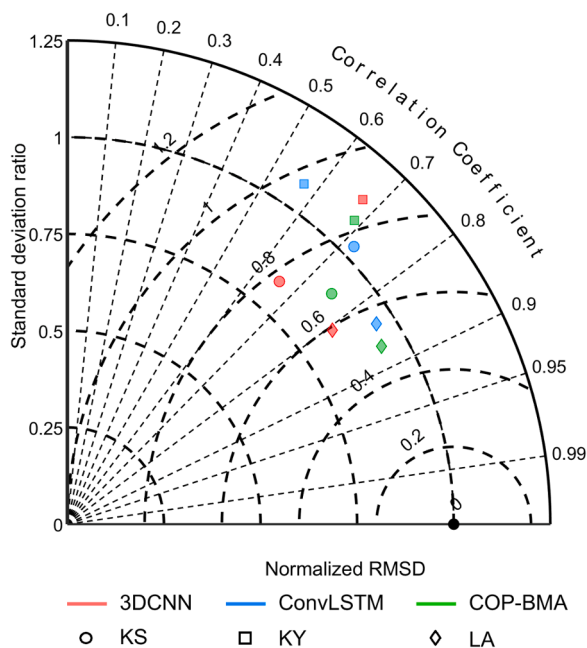


Figure 9. Taylor diagram showing three deterministic performance measures (i.e., RMSD, correlation coefficient, and standard deviation ratio) simultaneously. The black point displays the observation. Normalized standard deviation and correlation coefficient are on the radial axis and angular axis, respectively. The observation is shown with a black point on the horizontal axis. The red, blue, and green colors represent the three models used in this study. Each symbol indicates the state for which the crop yield was predicted. These results are reported collectively for the years 2018 and 2019. [This Figure should be printed in color]

are reported to experience mild to extreme drought during 2018 (which is likely to have contributed to the crop yield reduction compared to the long-term average) (<https://www.drought.gov/states/kansas/counties/bourbon>), all of these counties experienced extreme precipitation and flooding during the soybean growing season (starting from May to September) (Shorman, 2019). This is likely to have contributed to the higher soybean yield loss relative to the dry year (2018). Flooding in croplands affects crop yield directly by wilting the crops, whereas, extreme precipitation during the critical stages of soybean crops such as leaf development, and flowering indirectly affects the crop yield by restricting the plant growth (Kukul and Irmak, 2018). It is also worth mentioning that extreme events such as floods and droughts, which are known as frequent catastrophes in the US (Abbaszadeh et al., 2020; Alipour et al., 2020b, 2020a; Gavahi et al., 2020), can have a significant impact on the predictive capabilities of crop yield forecasting models.

Figure 7 represents the performance of the three models used in this study. It shows the residual standard deviation (in percent) computed between the observation and model simulation. This metric measures how much the data points spread around the regression line. The lower the residual standard deviation, the higher the accuracy of predicted soybean crop yield. The results indicate that the proposed approach has the lowest residual standard deviation compared to the 3DCNN and ConvLSTM networks. This implies the higher predictability of the proposed probabilistic crop yield simulation framework compared to its counterparts.

Figure 8 shows the probabilistic (coefficient of variation) and deterministic (absolute error) performance measures calculated for the predicted crop yield across more than a hundred counties in three states in the year 2018 and 2019. The coefficient of variation (CV) is a standardized measure of the dispersion of a probability distribution or frequency distribution. It is defined as the ratio of the standard deviation to the mean and represents the precision or relative variability of

estimates. For better interpretation and analysis of results, CV and Error are normalized between 0 and 1, such that the lower the CV and Error, the higher the performance of the proposed approach in predicting soybean crop yield. It is noted that Error is the absolute difference between the median of the predicted ensemble and the corresponding observed value. This figure indicates that our proposed approach, overall, has had the lowest and highest performance respectively in Neosho County (KS) and West Carroll County (LA) in 2018, and in 2019, similar results can be attributed to Harper County (KS) and East Carroll County (LA). The results also revealed that the reliability (CV) and accuracy (Error) of the proposed approach varies depending on the location of the counties where the hydroclimate and land surface conditions can be different, resulting in changes in the input variables of the deep neural networks and their ultimate performance, which accordingly affects the efficacy of the developed probabilistic approach.

The results also reveal that in general there is a relatively significant positive correlation (or trend) between the probabilistic and deterministic performance measures such that the counties whose error values are high are associated with larger predictive uncertainty intervals. More investigation showed that those counties with higher coverage of croplands (>90%) result in more reliable soybean yield prediction (with less predictive uncertainty interval) compared to the counties with less cropland coverage. For example, Brown county with 97% cropland coverage has predicted soybean yield with a CV of 0.21, while Bourbon County with 27% cropland coverage has the higher CV value of 0.55. The potential reason behind this is that the deep learning network's input image that has only nonzero values for cropland pixels contains more information for the network to learn from the inherent spatio-temporal patterns related to the crop yield.

Figure 9 compares the performance of the three models used in this study to simulate the soybean crop yield. The following Taylor diagram simultaneously reports three performance measures including standard deviation ratio, correlation coefficient, and normalized RMSD. The distance from the origin (shown as a black point) to each radial curve is the ratio of simulated to observed crop yield amplitude. Therefore, the model with a normalized standard deviation close to 1.0 produces a seasonal amplitude very close to the observation. The results indicate that in all three states the proposed approach outperforms the ConvLSTM and 3DCNN networks.

It is noted that in this paper the COP-BMA is a post-processing approach that we used to integrate the outputs of two deep NNs to provide more reliable and accurate prediction. The COP-BMA is a Bayesian approach that is used for multi-modeling. In this paper, we are not proposing a new ML technique, instead, we underscore the usefulness of the COP-BMA technique for multi-modeling of deep NNs and providing the probabilistic estimates of target values. This technique enables us to utilize multiple NN-based models and integrate their outputs to provide more accurate prediction while accounting for the uncertainty associated with the choice of the predictive model. The main idea behind multi-modeling is that the integrated outputs from multiple models is more accurate than those from individual models. Our analysis indicated that the COP-BMA is an efficient approach that performs well in integrating the outputs from multiple deep NNs. This Bayesian technique has been widely used for multi-modeling of physical hydrologic models, however, its applicability and usefulness for multi-modeling of data-driven models had not been explored before.

6. Conclusions

This study presents a new framework to integrate deterministic outputs from two deep neural network models to generate probabilistic simulation. The core of the proposed framework is the COP-BMA approach. COP-BMA integrates a group of multivariate Copula functions into Bayesian Model Averaging that relaxes any assumption on the shape of conditional PDF of each model and consequently provides more accurate and reliable predictive distributions. This study has been

conducted on probabilistic simulation of soybean crop yield across more than a hundred counties in three states in the United States. The results showed that the proposed approach provides more accurate predicted crop yield values compared to other deep neural networks such as ConvLSTM and 3DCNN while accounting for uncertainties involved in model predictions. Although the model outputs from these two networks were used in this study to generate probabilistic simulation, any other physically-based or data-driven based model outputs can also be used within the proposed framework. For future study, we plan to implement the developed approach to provide probabilistic yield predictions for different crops (such as maize and corn) across other states in the United States.

CRedit authorship contribution statement

Peyman Abbaszadeh: Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Keyhan Gavahi:** Software, Methodology, Data curation. **Atieh Alipour:** Software, Methodology, Visualization. **Proloy Deb:** Conceptualization, Writing – review & editing. **Hamid Moradkhani:** Supervision, Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors have no competing interests to declare.

Acknowledgments

Partial financial support for this research was provided by NSF-INFWEWS grant # 1856054.

References

- Abbaszadeh, P., Gavahi, K., Moradkhani, H., 2020. Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting. *Adv. Water Resour.* 145, 103721 <https://doi.org/10.1016/j.advwatres.2020.103721>.
- Abbaszadeh, P., Alipour, A., Asadi, S., 2018. Development of a coupled wavelet transform and evolutionary L evenberg-M arquardt neural networks for hydrological process modeling. *Comput. Intell.* 34 (1), 175–199. <https://doi.org/10.1111/coim.12124>.
- Abbaszadeh, P., Moradkhani, H., Gavahi, K., Kumar, S., Hain, C., Zhan, X., Duan, Q., Peters-Lidard, C., Karimzariani, M., 2021. High-Resolution SMAP Satellite Soil Moisture Product: Exploring the Opportunities. *Bulletin of the American Meteorological Society*. <https://doi.org/10.1175/BAMS-D-21-0016.1>.
- Alipour, A., Ahmadalipour, A., Abbaszadeh, P., Moradkhani, H., 2020a. Leveraging machine learning for predicting flash flood damage in the Southeast US. *Environ. Res. Lett.* 15, 024011 <https://doi.org/10.1088/1748-9326/ab6edd>.
- Alipour, A., Ahmadalipour, A., Moradkhani, H., 2020b. Assessing flash flood hazard and damages in the southeast United States. *J. Flood Risk Manag.* 13 <https://doi.org/10.1111/jfr3.12605>.
- Buckland, S.T., Burnham, K.P., Augustin, N.H., 1997. Model Selection: An Integral Part of Inference. *Biometrics* 53, 603. <https://doi.org/10.2307/2533961>.
- Chlingaryan, A., Sukkarieh, S., Whelan, B., 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* <https://doi.org/10.1016/j.compag.2018.05.012>.
- Choudhury, A., Jones, J., 2014. Crop yield prediction using time series models. *J. Econ. Econ. Educ. Res.* 15, 53–68.
- Chunjing, Y., Yueyao, Z., Yaxuan, Z., Liu, H., 2017. Application of convolutional neural network in classification of high resolution agricultural remote sensing images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives. International Society for Photogrammetry and Remote Sensing*, pp. 989–992. <https://doi.org/10.5194/isprs-archives-XLII-2-W7-989-2017>.
- Doraiswamy, P.C., Hatfield, J.L., Jackson, T.J., Akhmedov, B., Prueger, J., Stern, A., 2004. Crop condition and yield simulations using Landsat and MODIS, in: *Remote Sensing of Environment*. pp. 548–559. <https://doi.org/10.1016/j.rse.2004.05.017>.
- Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* 30 <https://doi.org/10.1016/j.advwatres.2006.11.014>.
- Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* 36 <https://doi.org/10.1007/s13593-016-0364-z>.
- Funk, C., Budde, M.E., 2009. Phenologically-tuned MODIS NDVI-based production anomaly estimates for Zimbabwe. *Remote Sens. Environ.* 113, 115–125. <https://doi.org/10.1016/j.rse.2008.08.015>.
- Gavahi, K., Abbaszadeh, P., Moradkhani, H., 2021. DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. *Expert Syst. Appl.* 184 <https://doi.org/10.1016/j.eswa.2021.115511>.
- Gavahi, K., Abbaszadeh, P., Moradkhani, H., Zhan, X., Hain, C., 2020. Multivariate assimilation of remotely sensed soil moisture and evapotranspiration for drought monitoring. *J. Hydrometeorol.* 21 <https://doi.org/10.1175/JHM-D-20-0057.1>.
- Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. *J. Forecast.* 3, 197–204. <https://doi.org/10.1002/for.3980030207>.
- Guthery, F.S., Burnham, K.P., Anderson, D.R., 2003. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. *J. Wildl. Manage.* 67 <https://doi.org/10.2307/3802723>.
- Gyamerah, S.A., Ngare, P., Ikpe, D., 2020. Probabilistic forecasting of crop yields via quantile random forest and Epanechnikov Kernel function. *Agric. For. Meteorol.* 280 <https://doi.org/10.1016/j.agrformet.2019.107808>.
- Hagedorn, R., Doblas-Reyes, F.J., Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.* <https://doi.org/10.1111/j.1600-0870.2005.00103.x>.
- Hansen, B.E., 2008. Least-squares forecast averaging. *J. Econom.* 146 <https://doi.org/10.1016/j.jeconom.2008.08.022>.
- Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. *Neural Comput.* 9. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: A tutorial. *Stat. Sci.* 14 <https://doi.org/10.1214/ss/1009212519>.
- Huang, X., Huang, G., Yu, C., Ni, S., Yu, L., 2017. A multiple crop model ensemble for improving large-scale yield prediction using Bayesian model averaging. *F. Crop. Res.* 211, 114–124. <https://doi.org/10.1016/j.fcr.2017.06.011>.
- Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.M., Gerber, J.S., Reddy, V.R., Kim, S.H., 2016. Random forests for global and regional crop yield predictions. *PLoS One* 11. <https://doi.org/10.1371/journal.pone.0156571>.
- Ji, S., Xu, W., Yang, M., Yu, K., 2013. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 221–231. <https://doi.org/10.1109/TPAMI.2012.59>.
- Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sens.* 10. <https://doi.org/10.3390/rs10010075>.
- Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* 141, 116–128. <https://doi.org/10.1016/j.rse.2013.10.027>.
- Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A., Bédard, F., 2016. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agric. For. Meteorol.* 218–219, 74–84. <https://doi.org/10.1016/j.agrformet.2015.11.003>.
- Karimzariani, M., Jafarzadeh, K., Abbaszadeh, P., Shao, W., Moradkhani, H., 2021. Hazard Risk Awareness and Disaster Management: Extracting the Information Content of Twitter Data. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2021.103577>.
- Khaki, S., Wang, L., 2019. Crop yield prediction using deep neural networks. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2019.00621>.
- Krastanov, S., Jiang, L., 2017. Deep Neural Network Probabilistic Decoder for Stabilizer Codes. *Sci. Rep.* 7 <https://doi.org/10.1038/s41598-017-11266-1>.
- Kukul, M.S., Irmak, S., 2018. Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. *Sci. Rep.* 8, 1–18. <https://doi.org/10.1038/s41598-018-21848-2>.
- LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L. D., 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* 1. <https://doi.org/10.1162/neco.1989.1.4.541>.
- Madadgar, S., Moradkhani, H., 2014. Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. *Water Resour. Res.* 50, 9586–9603. <https://doi.org/10.1002/2014WR015965>.
- Madadgar, S., Moradkhani, H., 2013. A Bayesian framework for probabilistic seasonal drought forecasting. *J. Hydrometeorol.* 14 <https://doi.org/10.1175/JHM-D-13-010.1>.
- Miao, C., Duan, Q., Sun, Q., Huang, Y., Kong, D., Yang, T., Ye, A., Di, Z., Gong, W., 2014. Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. *Environ. Res. Lett.* 9, 055007 <https://doi.org/10.1088/1748-9326/9/5/055007>.
- Milioto, A., Lottes, P., Stachniss, C., 2018. Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In: *Proceedings - IEEE International Conference on Robotics and Automation. Institute of Electrical and Electronics Engineers Inc.*, pp. 2229–2235. <https://doi.org/10.1109/ICRA.2018.8460962>.
- Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N., Foley, J.A., 2012. Closing yield gaps through nutrient and water management. *Nature* 490, 254–257. <https://doi.org/10.1038/nature11420>.
- Najafi, M.R., Moradkhani, H., 2016. Ensemble Combination of Seasonal Streamflow Forecasts. *J. Hydrol. Eng.* [https://doi.org/10.1061/\(asce\)he.1943-5584.0001250](https://doi.org/10.1061/(asce)he.1943-5584.0001250).
- Najafi, M.R., Moradkhani, H., 2015. Multi-model ensemble analysis of runoff extremes for climate change impact assessments. *J. Hydrol.* <https://doi.org/10.1016/j.jhydrol.2015.03.045>.
- Neuvauori, P., Narra, N., Lipping, T., 2019. Crop yield prediction with deep convolutional neural networks. *Comput. Electron. Agric.* 163 <https://doi.org/10.1016/j.compag.2019.104859>.
- Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L., Mouazen, A.M., 2016. Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* 121, 57–65. <https://doi.org/10.1016/j.compag.2015.11.018>.

- Patel, A.B., Nguyen, T., Baraniuk, R.G., 2016. A probabilistic framework for deep learning. *Advances in Neural Information Processing Systems*. Neural information processing systems foundation, pp. 2558–2566.
- Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens* 11. <https://doi.org/10.3390/rs11050523>.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005a. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133, 1155–1174. <https://doi.org/10.1175/MWR2906.1>.
- Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005b. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Mon. Weather Rev.* 133, 1155–1174. <https://doi.org/10.1175/MWR2906.1>.
- Raftery, A.E., Madigan, D., Hoeting, J.A., 1997. Bayesian Model Averaging for Linear Regression Models. *J. Am. Stat. Assoc.* 92, 179–191. <https://doi.org/10.1080/01621459.1997.10473615>.
- Rao, C., Liu, Y., 2020. Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization. *Comput. Mater. Sci.* 184 <https://doi.org/10.1016/j.commatsci.2020.109850>.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Müller, C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Khabarov, N., Neumann, K., Piontek, F., Pugh, T.A.M., Schmid, E., Stehfest, E., Yang, H., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc. Natl. Acad. Sci. U. S. A.* 111, 3268–3273. <https://doi.org/10.1073/pnas.1222463110>.
- Ruß, G., 2009. Data mining of agricultural yield data: A comparison of regression models. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 24–37. https://doi.org/10.1007/978-3-642-03067-3_3.
- Sa, I., Chen, Z., Popovic, M., Khanna, R., Liebisch, F., Nieto, J., Siegwart, R., 2018. WeedNet: Dense Semantic Weed Classification Using Multispectral Images and MAV for Smart Farming. *IEEE Robot. Autom. Lett.* 3, 588–595. <https://doi.org/10.1109/LRA.2017.2774979>.
- Salinas, D., Flunkert, V., Gasthaus, J., Januschowski, T., 2020. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *Int. J. Forecast.* 36, 1181–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>.
- Schwalbert, R.A., Amado, T., Corassa, G., Pott, L.P., Prasad, P.V.V., Ciampitti, I.A., 2020. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. For. Meteorol.* <https://doi.org/10.1016/j.agrformet.2019.107886>.
- 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, in: *Advances in Neural Information Processing Systems*.
- Shorman, J., 2019. Kansas floods in 2019 caused millions in damage, harmed dams | The Wichita Eagle. *The Wichita Eagle*.
- Slughter, J.M., Gneiting, T., Raftery, A.E., 2010. Probabilistic Wind Speed Forecasting Using Ensembles and Bayesian Model Averaging. *J. Am. Stat. Assoc.* 105, 25–35. <https://doi.org/10.1198/jasa.2009.ap08615>.
- Slughter, J.M.L., Raftery, A.E., Gneiting, T., Fraley, C., 2007. Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging. *Mon. Weather Rev.* 135, 3209–3220. <https://doi.org/10.1175/MWR3441.1>.
- van Oijen, M., Reyer, C., Bohn, F.J., Cameron, D.R., Deckmyn, G., Flechsig, M., Härkönen, S., Hartig, F., Huth, A., Kiviste, A., Lasch, P., Mäkelä, A., Mette, T., Minunno, F., Rammer, W., 2013. Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe. *For. Ecol. Manage.* 289, 255–268. <https://doi.org/10.1016/j.foreco.2012.09.043>.
- Wang, H., Yi, H., Peng, J., Wang, G., Liu, Y., Jiang, H., Liu, W., 2017. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Convers. Manag.* 153, 409–422. <https://doi.org/10.1016/j.enconman.2017.10.008>.
- Wöhling, T., Schöniger, A., Gayler, S., Nowak, W., 2015. Bayesian model averaging to explore the worth of data for soil-plant model selection and prediction. *Water Resour. Res.* 51, 2825–2846. <https://doi.org/10.1002/2014WR016292>.
- Wu, W., Chen, K., Qiao, Y., Lu, Z., 2016. Probabilistic short-term wind power forecasting based on deep neural networks. In: *2016 International Conference on Probabilistic Methods Applied to Power Systems, PMAPS 2016 - Proceedings*. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/PMAPS.2016.7764155>.
- You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian process for crop yield prediction based on remote sensing data. In: *31st AAAI Conference on Artificial Intelligence, AAAI 2017. AAAI press*, pp. 4559–4565.
- Zhong, Z., Li, J., Luo, Z., Chapman, M., 2018. Spectral-Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework. *IEEE Trans. Geosci. Remote Sens.* 56, 847–858. <https://doi.org/10.1109/TGRS.2017.2755542>.