Multivariate Multi-step Agrometeorological Forecast Model for Rapid Spray

Guobin Shi¹², Chun Wang¹³*  
¹ College of Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China  
² College of Computer Science, Guangdong University of Petrochemical Technology, Maoming 525000, China  
³ South Subtropical Crops Research Institute, Chinese Academy of Tropical Agricultural Sciences, National Soil Quality Zhanjiang Observation and Experimental Station, Zhanjiang, 524003, China  

Corresponding author: Chun Wang (e-mail: wangchun1963@126.com).  
This work was supported in part by the Nature Science Foundation of Hainan Province under Grant 519MS097, in part by the Key Assignment of Rural Science and Technology Commissioners in Guangdong Province in 2020 under Grant Yuekehannongzi [2020]409, and in part by the Fundamental Research Funds of Chinese Academy of Tropical Agricultural Sciences under Grant 19CXTD-31.

ABSTRACT The timing of spray application plays an essential role in daily pesticides management. Proper wind speed, air temperature, and relative humidity are the main external factors to improve the efficacy of pesticides, reduce the amount of spray drift and environmental pollution. Very few previous studies have focused on the need for short-term weather prediction in rapid spraying decisions. In this paper, a Convolutional-LSTM encoder-decoder (ConvLSTM-AE) hybrid model for multivariate output and multi-step prediction with short time intervals is proposed to predict these three agrometeorological variables in advance. This model can predict daily weather conditions at 15-minute intervals and track the changes of time-varying systems effectively. This method was also compared with other methods such as CNN, multi-head CNN, LSTM encoder-decoder, and CNN-LSTM encoder-decoder models. The results show that the proposed model outperforms other models and is suitable for daily weather forecasts in a short time interval. The obtained rapid and accurate prediction results provide a reliable basis for precise spray timing in actual farming.

INDEX TERMS ConvLSTM-EA, multi-step, multivariate output, walk-forward validation.

I. INTRODUCTION  
The development of energy-saving, environmental protection, and environmentally friendly agriculture is highly recommended in contemporary society. In agricultural spraying operations, spray drift has a negative impact on the environment; even 85% loss was reported in different vineyard structures and sprayers [1]. Thus, reducing spray drift in orchards is a major concern for researchers. Some researchers focused on the form of spray vehicles such as airblast sprayers or air-assisted sprayers and developed spraying apparatus to improve spray quality and reduce drift loss [2, 3]. Carpenter et al. [4] proposed a new spray system called permanent pesticide application system, representing a fixed spray system model. As a further development, solid set canopy delivery system (SSCDS) was widely studied as the fix spray system that offered comparable performance as the airblast sprayer, improving the efficiency of the spray operation and reducing drift [5-9]. The influence of temperature, humidity, and wind speed on the drift was also analyzed and studied to some degree [10-12]. Some researchers have applied intelligent spray systems to target spray to control drift [13, 14]. Rakesh et al. [15] developed an automatic SSCDS that took advantage of a relatively short time window to automatically control spray in minutes, allowing rapid spray operations under suitable ambient temperature, humidity, and wind speed. Developing new spraying automation technology for precision agriculture makes the spraying process more rapid, significantly reducing the spraying operation time. The efficient spraying quality is more dependent on external weather conditions. Previous studies have shown that it is beneficial to accurately control spraying quality under appropriate external temperature, humidity, and wind speed conditions. Still, the proper timing of spraying needs to be judged according to the weather conditions, which requires weather prediction in advance and corresponding preparation before spraying.

II. LITERATURE REVIEW
As a time series problem, time series data of agrometeorological forecast contains air temperature, dew point, relative humidity, wind speed, solar radiation, etc., which is affected by the complex interplay of some factors. Because of the importance of meteorological conditions for spraying operation, it is necessary to develop a reasonable forecast of meteorological conditions in a short time interval to provide reliable guidance before automatic spraying operation. Scholars have developed lots of researches about temperature, humidity, and wind speed prediction in recent years. Chen et al. [16] set an unscented Kalman filter integrated with a support vector regression model to update the short-term estimation of wind speed sequence. Cao et al. [17] compared a recurrent neural network with ARIMA and found multivariate models perform better than univariate models, and the recurrent neural network models outperform the ARIMA models. Chen et al. [18] proposed an EnsemLSTM model that achieved better performance in short-term wind speed forecasting. Park et al. [19] fabricated an LSTM-based hybrid model that can predict the temperature with better performance than DNN-based or LSTM-based models. Kreuzer et al. [20] showed ConvLSTM had good performance in short-term temperature forecasts with multivariate inputs. However, these prediction models only achieve univariate output predictions. Liu et al. [21] built a large VAR model to forecast three weather variables: hourly solar radiation, temperature, and wind speed using a very short-term forecasting style. Nevertheless, these predictions should be based on the assumption that these variables are stable, and the model can learn directly from lag observations, limiting the application scope of this method. Lu et al. applied the NNARX hybrid model [22], and Shi et al. [23] used an improved BP neural network to forecast indoor air temperature and relative humidity. Still, the fluctuation of external environmental factors in the natural environment is more significant than that of the indoor environment, so the model’s adaptability needs further study. Hewage et al. [24] took a stacked LSTM layer to forecast weather parameters over a given time. Zhang et al. [25] developed a new method for temperature and solar radiation prediction, which used a linear combination of the observation and the forecast data. These methods use hourly data and cannot meet more granular forecasting needs. Hence, it is necessary to predict in more fine-grained units of time, build appropriate methods, optimize and compare the performance of prediction models. In the past, few literature studies focused on the effects of short-term agrometeorological temperature, humidity, wind speed and made a combined prediction between multiple factors to guide the decision of rapid spray in the orchard. Therefore, it is necessary to study this problem to meet the needs of the spraying decision. Table 1 highlights the main characteristics of the existing schemes developed for weather forecasting.

The neural network can automatically incorporate nonlinear relationships between predictor variables and forecast distribution parameters in a data-driven way without presupposed link functions. The trained neural network can gain insight into the relationship of meteorological variables, which make predictions well. Deep learning algorithms, such as Convolutional Neural Networks (CNN) or Recurrent Neural Networks (RNN), are potential solutions for time series prediction. CNN was widely used in spatio-temporal features extraction, such as image and video, as well as time series prediction. In terms of time series prediction, CNN benefits from multi-layer perceptron; it supports multivariate input, multivariate output and can automatically learn complex functional relationships. By learning and extracting features from many input sequences most relevant to the prediction, the model achieves satisfactory prediction results [26]. As a variant of CNN, the multi-head convolution is a CNN in which each time series is processed on an utterly independent convolution. Each head was designed using one-dimensional CNN to extract features from each input channel and combine them together, called the convolution head. Multi-head CNN combined multi-dimensional input data well to represent the relationship between multivariate, which can be used for multivariate prediction [27]. LSTM model was first proposed by Sepp et al. [28, 29]. As an extension of RNN, the unique structure of LSTM can capture the dependency features of sequence data and solve the vanishing gradient problem in neural network training. LSTM model, LSTM variant, and other improved methods were widely used in water quality forecasting [30], air quality prediction [31], traffic prediction [32, 33], and so on. Some hybrid models methodology, such as LSTM-AE, CNN-LSTM, and ConvLSTM, showed better performance than that of the single model [34-36]. These deep learning algorithms provide an ideal choice for solving time series problems.

The weather data at a long interval, such as several hours interval, can not reflect the actual condition on time, so it is necessary to develop several minutes interval continuous prediction to meet the needs of fast spray. Whereas short-time intervals will increase the number of daily prediction steps and the cumulative prediction error, suitable models should be built to deal with the peculiar time-series data to obtain accurate predictions. These peculiar time series data usually include the following characteristics:
1) Non-linear relationships among variables;
2) Periodicity and outliers exist in data;
3) Prediction needs for multi-input, multi-output and multi-step.

Therefore, the model should be built to deal with the periodic, short time interval, nonlinear relationships, multi-step and multivariable problems of meteorological data. Based on the analysis above, this study established a multi-step, short time interval multi-variable combined prediction model to meet the need of forecast weather conditions in
advance, compared the performance of the prediction results with these neural network-based algorithms: CNN, multi-head CNN, LSTM Encoder-Decoder (LSTM-AE), Hybrid CNN with an LSTM-AE (CNN-LSTM-AE) structure model, and a hybrid convolutional LSTM encoder-decoder (ConvLSTM-AE) model. These models are investigated, evaluated, and compared using different statistical indicators to assure that the final model meets the goal that is built for.

The statistical indicators include mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean squared error (MSE), the root mean square error (RMSE), and symmetric mean absolute percentage error (SMAPE), respectively. The final results of this study clarified the feasibility of the models, verified the excellence of the method, and provided an ideal weather prediction model to guide spraying operation in actual farming.

| Aim                  | Technique                  | Merits/outcomes                                                                 | Demerits                                                                 | dataset                                      |
|----------------------|---------------------------|--------------------------------------------------------------------------------|---------------------------------------------------------------------------|----------------------------------------------|
| Wind speed prediction| Unscented Kalman filter based state-space support vector regression approach. | The proposed method has better performance in both one-step-ahead and multistep-ahead predictions than ANNs, SVR, autoregressive, and autoregressive integrated with Kalman filter models. | (i) Needs to develop the predictive model-based control and optimization strategies for wind farm operation. (ii) Only univariate output prediction. | Center for Energy Efficiency and Renewable Energy at University of Massachusetts including 700 samples. |
| Wind speed forecasting| Univariate and multivariate ARIMA models, recurrent neural network. | Multivariate models perform better than univariate models and that the recurrent neural network models outperform the ARIMA models. | Needs to examine the improvement in wind speed forecasting by incorporating other covariates such as neighboring turbines, wind direction and centralize pressure. | Wind Engineering Research Field Laboratory (WERFL) at Texas Tech University including 26,999 observations within a 15-minute interval. |
| Short-term wind speed forecasting | LSTMs (Long Short Term Memory neural networks), SVRM (support vector regression machine) and EO (extremal optimization algorithm). | The proposed EnsemLSTM can achieve a better forecasting performance than ARIMA, SVR, ANN, KNN, GBRT. | Needs to consider more interrelated features like weather conditions, human factors, and power system status. | A wind farm in Inner Mongolia, China. Wind speed data sampled per ten minutes from November 23, 2012 to November 28, 2012. |
| Temperature prediction | Long short-term memory (LSTM) neural network. | The proposed model is higher than that of the Unified Model (UM) Local Data Assimilation and Prediction System (LDAPS) for 24 h temperature predictions. | Needs to improve prediction accuracy during the season-changing period. | Three different locations in South Korea. Hourly data from November 1981 to December 2016. |
| Short-term temperature forecasts | 2D-convolutional LSTM (ConvLSTM) network. | ConvLSTM performance better than SARIMA, seasonal naive approach, LSTM, and LSTM (multi). | The time granularity and relevance of the various input channels still need further research. | Five different weather stations in Germany with hourly data of different time series length. |
| Temperature, solar and wind prediction | Vector autoregression (VAR) model. | VAR approach is appropriate for very short-term forecasting of hourly solar radiation, temperature, and wind speed. | Needs to customize the model and its autoregressive structure to better exploit the correlations in the data. | National Solar Radiation Database (NSRDB). Data from 1990 to 2008. |

Table 1: Main characteristics of the existing weather forecasting schemes
| Aim | Technique | Merits/outcomes | Demerits | dataset |
|-----|-----------|----------------|----------|---------|
| Indoor Temperature and Relative Humidity | Improved backpropagation (BP) neural networks. | The BP model was more effective in temperature prediction than that proposed by Mba. | The fluctuation of external environmental factors in the natural environment is more complex than indoor environment. | Chongqing, China. Data from January 2017 to January 2018. |
| Ten weather parameters | Long Short-Term Memory (LSTM) layered model. | Proposed model produces better results compared to WRF model. | Needs to compare and increase the accuracy with more MIMO approach. | Global Forecast System (GFS) GRIB data from 1st of January 2018 to 31st of May 2018. |
| Temperature and solar radiation | Linear combination of various weather forecast sources. | A large number of online weather forecast and local observation data combination make accurate predictions on temperature and solar radiation in short-term (24 hours) prediction. | Needs to develop a dynamic adaptation of the model parameters in order to account for changes in climate and/or forecast models. | Leicestershire, England. Data from four websites of bi-hourly intervals. 2004 to 2006. |

### III. METHODOLOGY

#### A. CNN

CNN with multivariate input was called multi-channels deep convolutional neural networks model. In this multi-channel input model, every individual one-dimensional time series is one of the channel inputs; the model learns features from the time series input variables by using different kernel mappings. This setting can solve the problem that the output sequence expresses some characteristics of the observations and expresses the trend of the future values.

The inputs are fed into a feature extractor to learn features through convolution layers, activation and pooling layers combination. After the pooling layer, the distilled feature maps are then flattened into 1-D tensor form with a multivariable combined output. Final data was reshaped into prediction results with 96 steps ahead and 3-D vectors. 32 and 16 filters were selected for two convolution layers separately; all layers are separated by Rectifier Linear Units (ReLU), kernel size is set to 24, the batch size is set to 96, and model is stopped training at 100 epochs.

#### B. MULTI-HEAD CNN

A specially designed structure is responsible for extracting features from 8 input sequences. Each input variable has a separate sub-model to extract features in the sequence separately. Then the outputs from each model with features are merged into one long vector. Finally, data are interpreted by fully connected layers and reshaped into multi-step and multi-output variables. This processing method is the same as that of CNN data processing. Parameters of multi-head CNN model are set as follows: the first two convolution layer uses 32 filters, kernel size is set to 16, the next convolution layer uses 128 filters with a kernel size of 5, all layers use ReLU activation, the training epoch is set to 100.

#### C. LSTM ENCODER-DECODER

LSTM algorithm performs well for time series data prediction. An LSTM cell contains an input gate, a forget gate, and an output gate. This gated structure model enables LSTM to solve the learning problems related to sequential data. It can be calculated using the following equations:

**Input Gate:**

\[
i_t = \delta_p(W_i x_t + U_i h_{t-1} + b_i) \quad (1)
\]

**Forget Gate:**

\[
f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)
\]

**Cell:**

\[
c_t = i_t \tilde{c}_t + f_t c_{t-1} \quad (3)
\]

**Output Gate:**

\[
o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)
\]

\[
h_t = o_t \tan h (C_t) \quad (6)
\]

Where \(i_t, f_t, o_t\) are input gate, \(W_i, W_f, W_o, U_i, U_f, U_o\) are weight matrices respectively; \(b_i, b_f, b_o\) are bias vectors respectively; \(h_t, h_{t-1}\) are layer output at time step \(t\) and \(t-1\) respectively; \(C_o, \tilde{C}_t\) and \(C_{t-1}\) are the alternative state of cell at time step \(t\) and \(t-1\) respectively; \(\delta_p\) is the gate activation function.
function, which normally is the sigmoid function, and the tanh is the hyperbolic tangent function.

LSTM-AE adds an autoencoder model (AE) to the traditional multi-layer LSTM [35]. AE consists of an encoder and a decoder, which provides a solution for multi-step and multi-output problems. Features from the input sequence are learned through LSTM and encoded to feature vectors for the encoder function. For the decoder, it reconstructs the input to accumulate the internal state while outputting the prediction sequence. The final optimized model for prediction uses one LSTM layer and one encoder-decoder LSTM hybrid architecture. The first LSTM layer obtains time series data and extracts features. The encoder-decoder LSTM architecture reconstructs the previous outputs with ninety-six time steps and one feature. Then it outputs a sequence with ninety-six time steps and three features; a repeat vector layer and time distributed wrapped dense layer can achieve this. An LSTM layer inside encoder-decoder architecture generates the output values sequentially. Model parameters are set as follows: the number of units in LSTM is set to 200, sigmoid is used as activation function, the batch size is set to 96, and the number of epochs is set to 100.

D. CNN-LSTM ENCODER-DECODER
CNN-LSTM-AE utilizes a network structure similar to LSTM-AE. It obtains the trend of time series through LSTM-AE and realizes multi-step and multivariate output. By inserting the CNN structure before LSTM-AE, the critical features of input variables are further extracted to obtain the complex features among multivariate variables [37]. CNN output results are flattened into a 1-D tensor and sent to the repeat vector layer; this layer repeats the incoming inputs with a specific number of time steps to output multi-step results. The subsequent traditional LSTM layer is replaced by a CuDNNLSTM layer optimized with 100 units which GPU can accelerate to speed up the training process. Finally, one time distributed wrapped dense layer is added for multi-output values, and the epoch is set to 50.

E. CONVLSTM ENCODER-DECODER
Compared with the CNN-LSTM model, the proposed ConvLSTM and LSTM-AE hybrid structure is shown in Fig. 1. ConvLSTM is a variant of LSTM that includes convolution operations inside the LSTM cell; This unique RNN model can obtain spatial relations between multiple inputs and long-term dependence on time series data; ConvLSTM replaces matrix multiplication with convolution operations at each gate in the LSTM cell [36]. This process is done with a ConvLSTM2D layer; ConvLSTM outputs were flattened into a 1D tensor and passed through an encoder-decoder structure. By doing so, it captures underlying multidimensional relationships by performing convolution operations in multidimensional data. In this paper, data was reshaped to 5D tensor and fed to a single ConvLSTM2D layer with eight filters and kernel size as 1×5, the sigmoid function is used as the activation function, the batch size is set to 1024, the number of units in CuDNNLSTM is set to 10, and training epoch uses 50.

![FIGURE 1. An overview of optimized ConvLSTM-AE hybrid network structure](image)

F. WALK FORWARD VALIDATION
Multi-step prediction can be achieved by iteration of the single-step forecast, but the error accumulation of one-step ahead will significantly degrade the prediction performance. A specialized walk-forward validation method was designed to solve the multi-step time series forecasting problem before data was fed into the model. Records with a minimum available interval of 15 minutes were divided according to the daily basis; models were provided with data at intervals of days to predict multiple future steps accurately. The implementation method is described as follows: the in-sample data are set to 192, and the out-of-sample data are set to 96. This method uses two days (192 steps) of in-sample data and walks forward one day (96 steps) at a time.

G. EVALUATION METHOD
In order to evaluate the experimental results of the model mentioned above, MSE, RMSE, MAE, MAPE (%), and
SMAPE (%) were chosen for statistical metrics. Definition of as follows:

\[
\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (Y_{pred(i)} - Y_{true(i)})^2
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (Y_{pred(i)} - Y_{true(i)})}
\]

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |(Y_{pred(i)} - Y_{true(i)})|
\]

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{|Y_{pred(i)} - Y_{true(i)}|}{Y_{true(i)}}\right] \times 100
\]

\[
\text{SMAPE} = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{2 \times (Y_{pred(i)} - Y_{true(i)})}{|Y_{pred(i)}| + |Y_{true(i)}|}\right] \times 100
\]

Where \(Y_{pred(i)}\) and \(Y_{true(i)}\) are the predicted value and actual value, respectively, and \(N\) is the number of all predicted values.

IV. RESULTS AND DISCUSSION

A. INPUT VARIABILITY ANALYSIS

All experiments were conducted on the Google Colab platform, which provided free access to a 2x2.2 GHz CPU and a Tesla T4 GPU environment. Meteorological datasets were collected from the Washington State Agricultural Weather Network (AgWeatherNet. http://weather.wsu.edu/), data were recorded every 15 minutes. Data preprocessing deleted redundant data and replaced a small amount of outlier or missing values such as NAN or values beyond a given range with linear interpolation between normal adjacent values and screened out parts with large amounts of abnormal data. Finally, continuous data were selected, normalized, and fed into the model. The final data included Data Time, Air Temperature (AT), Dew Point (DP), Relative Humidity (RH), Leaf Wetness (LW), Wind Direction (WD), Wind Speed (WS), Total Precipitation (TP), and Solar Radiation (SR) of continuous subseries, each meteorological variables with every 15 minutes intervals from 01/01/2018 to 2/28/2019 generated more than 40k records. The data was divided into two parts, data from 01/01/2018 to 12/31/2018 was used for model training and validation, and the rest was used for testing. The wind direction values were encoded (integer encoded) for model training. The weather variables values were normalized between [0, 1] by Min-Max scaling except Data Time; the formula is shown in (12). The parameters selected are shown in Table 2.

\[
\chi' = \frac{x - \min(x)}{\max(x) - \min(x)}
\]

Many methods can be used to analyze time series characteristics. Time series decomposition is one of the methods for analyzing these data. The decomposition model reduces a time series into three components: trend, seasonal effects, and noise. One primary multiplicative model decomposition is shown in Formula (13). Where the trend is the increasing or decreasing value in the series; seasonality is the repeating short-term cycle in the series; noise is the random variation in the series. This multiplicative model can explain the increasing trend of the long-term AT observed values with 20k data in Fig. 2(c). AT series in Fig. 2(a), 2(b), 2(c) represented decomposition with different periodicity when the frequency was set to daily, weekly and monthly, respectively. These decomposed data clearly show multi-periodicity, nonlinear trend, and multiple seasonality in AT subseries, the same as RH subseries. Therefore, it is suitable for extracting features from the daily data to build a prediction model.

\[
y(t) = \text{Trend} \times \text{Seasonality} \times \text{Noise}
\]
The Pearson’s correlation coefficient method was also used to analyze the correlation between AT, RH, WS, and other factors in the environment. This method was measured after data was normalized, and the correlation results were shown in Table 3. The analysis results showed that AT had a strong correlation with RH (-0.78) and DP (0.72), a moderate positive correlation with SR (0.52), a weak correlation with LW (-0.38), a very weak correlation with WS (0.19). RH had a moderate correlation with LW (0.58) and SR (-0.56), a weak correlation with WS (-0.33) and DP (-0.21). WS has a weak correlation with SR (0.22).

The correlation analysis of multiple input variables showed that there was an apparent nonlinear relationship among the variables. As a result, the traditional technique for time series predictions such as ARIMA is inappropriate for solving nonlinear problems. In contrast, artificial neural networks (ANN) such as CNN, LSTM, and hybrid models are better choices.

The interquartile ranges (IQR) were computed to compare the variability of input variables. The results shown in Table 4 indicated the IQR of the normalized input variables was in the range of 0 to 0.95. RH had the largest IQR (0.57), which indicated the variability was relatively high and may have a greater impact than other variables predicted by the prediction model. The IQR of AT (0.28) and SR (0.26) was lower than that of RH. WS (0.08) had the lowest IQR.

For example, Fig. 3 shows the AT curves from April 1, 2018, to April 7, 2018. In general, the AT curves for seven consecutive days reached the lowest value at about four to seven o’clock and then increased gradually, reaching the maximum value of a day around 12 to 14 o’clock, then these curves showed a downward trend to the lowest value, the reason depends mainly on direct solar radiation. However, the variation range and time of daily AT curves are also affected.

**TABLE 3. CALCULATION RESULTS OF CORRELATION**

|       | AT   | DP   | RH   | LW   | WS   | SR   |
|-------|------|------|------|------|------|------|
| Obs   | 0.724001 | -0.78457 | -0.38049 | 0.185617 | 0.517334 |
| RH    | -0.78457 | -0.206041 | 0.584606 | -0.33357 | -0.55929 |
| WS    | 0.185617 | -0.04213 | -0.33357 | -0.26171 | 0.218328 |

**TABLE 4. NORMALIZED TARGET AND INPUT VARIABLES.**

|       | AT   | DP   | RH   | WS   | SR   |
|-------|------|------|------|------|------|
| Q25%  | 0.29398 | 0.44522 | 0.37443 | 0.09783 | 0  |
| Q50%  | 0.42565 | 0.54494 | 0.64027 | 0.13768 | 0.00529 |
| Q75%  | 0.57899 | 0.63764 | 0.9457 | 0.18116 | 0.25661 |
| IQR   | 0.28491 | 0.19242 | 0.57127 | 0.08333 | 0.25661 |
by other factors like RH, wind currents, cloud cover, etc. Some involve factors can be measured by measuring instruments, whereas others may find unsuitable to be measured in a small region. These factors lead to the differences in daily AT variation, and the multivariable prediction model can reflect the correlation between variables. Fig. 3 shows that the maximum daily AT difference values was 22.8 °F (12.7°C), and the maximum daily average AT difference was 12.2 °F (6.8°C). The significant differences require the prediction model have a strong ability to follow the trend and predict effectively.

![Graph showing AT curves for 7 consecutive days](image)

**FIGURE 3.** The AT curves for 7 consecutive days

### B. PARAMETER SETTING

Five network models were trained using the preprocessed dataset to compare the differences in performance. Hyperparameters of models were tuned with the Grid Search method in the model training process. Possible parameter values were selected for combination, and the performance metric was measured by validation on the validation dataset. The ideal parameter combination was finally determined to learn features and reduce overfitting, and the relevant parameters were tested more than 20 times to achieve the best model results. Above mentioned models use the MSE for performance evaluation and use "Adam" optimizer for minimization algorithm. The final parameters selection of these multivariate output models are based on minimizing RMSE of AT, RH, and WS, respectively. Table 5 shows a model structure and a comparison of the parameters of these models. ConvLSTM-AE model had the minimum number of the parameters, thus reducing the model training time. In contrast, the multi-head CNN model has the maximum number of parameters due to the stack of a large number of convolutional layers.

| Table 5. Five model structures and their parameters. |
|---------------------------------------------------|
| No  | Type of network | Type of layer | Output sequence | Number of parameters | Total number of parameters |
|-----|-----------------|---------------|-----------------|---------------------|--------------------------|
| 1   | CNN             | Conv1D        | (None, 169, 32) | 6176                | 112,276                  |
|     |                 | Maxpooling1D  | (None, 84, 32)  | 0                   |                          |
|     |                 | Conv1D        | (None, 61, 16)  | 12304               |                          |
|     |                 | Maxpooling1D  | (None, 30, 16)  | 0                   |                          |
|     |                 | Flatten       | (None, 480)     | 0                   |                          |
|     |                 | Dense         | (None, 100)     | 48100               |                          |
|     |                 | Reshape       | (None, 288)     | 29088               |                          |
| 2   | Multi-head CNN  | InputLayer    | (None, 192, 1)  | 0                   |                          |
|     |                 | Conv1D        | (None, 177, 32) | 544×8               |                          |
|     |                 | Conv1D        | (None, 162, 32) | 16416×8             |                          |
|     |                 | MaxPooling1D  | (None, 81, 32)  | 0                   |                          |
|     |                 | Conv1D        | (None, 77, 128) | 20608×8             |                          |
|     |                 | MaxPooling1D  | (None, 38, 128) | 0                   |                          |
|     |                 | Flatten       | (None, 4864)    | 0                   |                          |
|     |                 | Concatenate   | (None, 38912)   | 0                   |                          |
|     |                 | Dense         | (None, 200)     | 7782600             |                          |
|     |                 | Dense         | (None, 100)     | 20100               |                          |
|     |                 | Dense         | (None, 288)     | 29088               |                          |
|     |                 | Reshape       | (None, 96, 3)   | 0                   |                          |
| 3   | LSTM-AE         | LSTM          | (None, 200)     | 167200              | 488,603                  |
|     |                 | RepeatVector  | (None, 96, 200) | 0                   |                          |
|     |                 | LSTM          | (None, 96, 200) | 320800              |                          |
|     |                 | TimeDistributed | (None, 96, 3) | 603                 |                          |
| 4   | CNN-LSTM-AE    | Conv2D        | (None, 2, 94, 4)| 100                 | 116,403                  |
|     |                 | MaxPooling2D  | (None, 1, 47, 4)| 0                   |                          |
|     |                 | Flatten       | (None, 188)     | 0                   |                          |
|     |                 | RepeatVector  | (None, 96, 188)| 0                   |                          |
|     |                 | CuDNNLSTM     | (None, 96, 100) | 116000              |                          |
|     |                 | TimeDistributed | (None, 96, 3) | 303                 |                          |
C. MODELING PREDICTION ASSESSMENT

Based on the analysis above, multi-step ahead prediction models were performed. Multi-step predicted and observed values of AT of different models are shown in Fig. 4, Fig. 4(a) shows the predicted values versus observed values of the five models. Fig. 4(b-f) represent scatter plots of AT prediction values of different models. Compared with the actual values, the distribution of the CNN scatter in Fig. 4(b) is the most dispersed, which leads to a relatively large prediction error. The multi-head CNN model ($R^2=0.81$) in Fig. 4(c) performs better than the CNN model ($R^2=0.37$) with a better correlation because the multi-head structure with multiple convolutional networks in each channel can learn more features, which helps the model perform better. For LSTM-AE, the scatter in Fig. 4(d) is closer to the standard line (slope of 1.04), but predicted values deviate too much from the observed values. The CNN-LSTM-AE ($R^2=0.86$) in Fig. 4(e) has a better correlation than the multi-head CNN model, and the overall predicted value was higher than the observed value. Fig. 4(f) showed that the predicted values of ConvLSTM-AE agreed with the observed values well, with a slope of 0.72 and $R^2$ of 0.92. ConvLSTM-AE in Fig. 4(f) shows a slimmer distribution form compared with CNN-LSTM-AE, and apparent data aggregation in ConvLSTM-AE indicates it could follow the range of predicted values more effectively, and the scatter is closer to the standard line, the cumulative error of multi-step prediction does not increase significantly. Therefore, ConvLSTM-AE for AT prediction is more reasonable.

![Diagram showing multi-step AT forecasts of different models](image)

**FIGURE 4.** Multi-step AT forecasts of different models. (a) is the predicted values versus observed values for different models, (b-f) are the scatter plots of AT prediction values for different models.
Fig. 5 represents the RH forecasts of these models. Fig. 5(a) shows the predicted values versus observed values; Fig. 5(b-f) are scatter plots of RH prediction values of different models. The scatter plots of the observed and predicted RH show that the CNN-LSTM-AE (R²=0.83) in Fig. 5(c) and ConvLSTM-AE (R²=0.85) in Fig. 5(f) performed better than CNN (R²=0.44) in Fig. 5(b) and multi-head CNN model (R²=0.73) in Fig. 5(c). LSTM-AE, CNN-LSTM-AE, and ConvLSTM-AE models show a relatively small dispersion of observed values compared to other methods. ConvLSTM-AE performs best for RH prediction.

![Graph showing RH forecasts of different models](image)

**FIGURE 5.** Multi-step RH forecasts of different models. (a) is the predicted values versus observed values for different models, (b-f) are the scatter plots of RH prediction values for different models.

Fig. 6 shows WS forecasts of these models. Fig. 6(a) indicates that WS’s stochastic uncertainty and fluctuations changed frequently, and the instability affects the corresponding prediction. The observed values of CNN (R²=-0.01) in Fig. 6(b) and multi-head CNN (R²=0.11) in Fig. 6(c) are more dispersive than other models, which generate a larger error than LSTM-AE in Fig. 6(d) and hybrid models. Scatter plots in Fig. 6(b-f) indicate that LSTM-AE and hybrid models can better learn the time series trend and reduce the influence of abnormal fluctuations. Although the performance of ConvLSTM-AE in terms of R2 (0.17) in Fig. 6(f) is lower than that of CNNLSTM (R2=0.26) in Fig. 6(e), it achieves higher data density than CNN-LSTM-AE, and predict values are normally distributed.
The optimal model performance metrics of these five predictive models are listed in Table 6. For multi-input, multi-output (MIMO) prediction, the multivariate output evaluation demonstrates a comparative performance among these models. It is found that ConvLSTM-AE has the lowest RMSE of AT (1.88) and WS (0.95) among these models, which outperforms other models. SMAPE can achieve the same conclusion. As for RMSE of RH, CNN-LSTM-AE and LSTM-AE perform better than the ConvLSTM-AE model in this single test. This result indicates that the ConvLSTM model balances the multivariable output results of each indicator so that the error of continuous prediction outputs of each variable value does not change too much. Compared MIMO with the single-input and single-output (SISO) ConvLSTM-AE model, the MIMO prediction of ConvLSTM-AE still showed better performance than SISO prediction with the help of the correlation between variables. MIMO indicators outperform SISO indicators. Results demonstrate multivariate models perform better than univariate models.

Multi-step ahead prediction is also used to measure the stability of these models for seven consecutive days' tests. The average RMSE of AT, RH, and WS in Table 7 indicates that ConvLSTM-AE has the lowest RMSE, superior to other models. The results also indicate ConvLSTM-AE is stable; this model still maintains a good predictive ability and can be used for multi-step ahead prediction on consecutive days.
TABLE 7. RMSE FOR MULTI-STEP AHEAD PREDICTION OF 7 CONSECUTIVE DAYS

| Model               | Mean AT | Mean RH | Mean WS |
|---------------------|---------|---------|---------|
| CNN                 | 6.695481| 17.96799| 3.275301|
| Multi-head CNN      | 14.34884| 24.15716| 5.108632|
| LSTM-AE             | 6.321618| 16.08127| 2.629418|
| CNN-LSTM-AE         | 7.056461| 15.05143| 2.494531|
| ConvLSTM-AE         | 6.149656| 13.58034| 2.477793|

Different in-sample data with walk-forward validation were also tested, as shown in Table 8. In-sample data of one day, two days, four days, and six days were used for one day-ahead prediction with the ConvLSTM-AE model. In general, the RMSE of the proposed model indicates that corresponding continuous training data increase with the number of in-sample data increasing from one day to two days, which is beneficial for the model to learn data features.

However, when the in-sample data increased from two days to four days and six days, more and more variation values of in-sample data were involved, leading to a messy variation trend, more significant prediction error, and worse performance. Therefore, relevant parameters need to be adjusted according to specific prediction needs, and two days of in-sample data are best for the prediction purpose.

TABLE 8. RMSE OF IN-SAMPLE DAYS FOR ONE DAY-AHEAD PREDICTION

| ConvLSTM-AE model | In-sample     | RMSE of AT | RMSE of RH | RMSE of WS |
|-------------------|---------------|------------|------------|------------|
|                   | 1 day         | 2.31       | 4.51       | 1.18       |
|                   | 2 days        | 1.88       | 3.85       | 0.95       |
|                   | 4 days        | 1.89       | 5.65       | 1.04       |
|                   | 6 days        | 2.56       | 5.52       | 0.99       |

V. CONCLUSION

In this work, five models with different parameters were designed, tested, and optimized through the Colab platform with GPU acceleration, and the optimized models average training time was significantly reduced; values are as follows: CNN 136 s, multi-head CNN 875 s, LSTM-AE 1096 s, CNN-LSTM-AE 100 s, ConvLSTM-AE 126 s. Although the ConvLSTM-AE model is relatively complex, with GPU acceleration and parameters optimization, the model has an obvious advantage in training time, which is 1 to 9 times faster than other models. The test results also indicate that the ConvLSTM-AE model with larger filters and kernel sizes will learn more abnormal weather values, leading to the over-fitting of the forecast results, increasing the error of the forecast results. Therefore, the ConvLSTM-AE model achieved desirable prediction results by using smaller odd-sized filter sizes and kernel sizes. The walk-forward validation method, multi-step prediction, multi-input, and multi-output prediction is also realized with five models; these models’ performance was also obtained. The prediction results show that the proposed model can better obtain the internal relationship of spatio-temporal series data and solve the trend prediction of time-varying signals. The proposed ConvLSTM-AE model has generated more precise and reliable forecasting results than other reference models. It could satisfy the practical requirements for spray timing selection. For future work, we will investigate further improving the model’s performance by using the hybrid algorithm to conduct data preprocessing more effectively and filter the outliers for the final forecasting series.

REFERENCES

[1] O. Viret, W. Siegfried, E. Holliger, and U. Raisigl, “Comparison of spray deposits and efficacy against powdery mildew of aerial and ground-based spraying equipment in viticulture,” Crop Prot., vol. 22, no. 8, pp. 1023-1032, Sep. 2003.
[2] D. L. Peterson, and H. W. Hogmire, “Tunnel sprayer for dwarf fruit trees,” Transactions of the ASAE, vol. 37, no. 3, pp. 709-715, Jun. 1994.
[3] H. Zhu, R. D. Brawee, R. Derksen, and R. D. Fox, “A Specially Designed Air-Assisted Sprayer to Improve Spray Penetration and Air Jet Velocity Distribution Inside Dense Nursery Crops,” Trans. ASABE, vol. 49, no. 5, pp. 1285-1294, Sep. 2006.
[4] T. G. Carpenter, D. L. Reichard, and S. M. Wilson, “Design and Feasibility of a Permanent Pesticide Application System for Orchards,” Trans. ASABE, vol. 28, pp. 350-355, Oct. 1984.
[5] A. Sharda, M. Karkkee, and Z. Qin, “Fluid dynamics of a solid set canopy spray delivery system for orchard applications,” in the 2013 ASABE Annual International Meeting, Kansas City, Missouri, 2013.
[6] Grieshopp, M. Gut, L. Wise, and J. Smith, “development of solid set delivery system for high density apples,” New York Fruit Quarterly, vol. 32, pp. 16-20, Nov. 2015.
[7] A. Sharda, M. Karkkee, Q. Zhang, and I. Ewanlow, “Effect of emitter type and mounting configuration on spray coverage for solid set canopy delivery system,” Comput Electron Agric, vol. 112, pp. 184-192, Mar. 2015.
[8] S. M. Niemann, M. D. Whiting, and I. Hanrahan, “Efficacy of a solid set canopy delivery system (SSCDS) in sweet cherry and apple,” Acta Hortic., no. 1130, pp. 145-150, Nov. 2016.
[9] R. Sinha, “Drift potential from a solid set canopy delivery system and an axial–fan air-assisted sprayer during applications in grapevines,” Biosyst. Eng., vol. 188, no. 2, pp. 207-216, Nov. 2019.
[10] P. Owen-Smith, J. Wise, and M. Grieshop, “Season Long Pest Management Efficacy and Spray Characteristics of a Solid Set Canopy Delivery System in High Density Apples,” Insects, vol. 10, no. 7, pp. 193, Jun. 2019.
[11] J.-F. Praat, J. Abber, and D. Markelov, “The effect of canopy development and sprayer position on spray drift from a pipfruit orchard,” N. Z. Plant Prot., vol. 53, Aug. 2000.
[12] P. Owen-Smith, R. Perry, J. Wise, and R. Z. Raja Jamil, “Spray coverage and pest management efficacy of a solid set canopy delivery system in high density apples,” Pest Manag. Sci., vol. 75, no. 11, pp. 3050-3059, Mar. 2019.
[13] D. Stajnko, P. Berk, M. Lešnik, and V. Ježič, “Programmable ultrasonic sensing system for targeted spraying in orchards,” Sensors, vol. 12, no. 11, pp. 15500-15519, Nov. 2012.
[14] Y. Chen, H. Zhu, and E. Ozkan, “Development of a Variable-Rate Sprayer with Laser Scanning Sensor to Synchronize Spray Outputs to Tree Structures,” Trans. ASABE, vol. 55, no. 3, pp. 773-781, May. 2012.
[15] R. Sinha, “Automated solid set canopy delivery system (SSCDS) for large scale spray application in tree-fruit crops,” Trans. ASABE, vol. 62, no. 3, pp. 585-592, Apr. 2019.
[16] K. Chen, and J. Yu, “Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression
approach,” Appl. Energ., vol. 113, pp. 690-705, Jan. 2014.

[17] Q. Cao, B. T. Ewing, and M. A. Thompson, “Forecasting wind speed with recurrent neural networks,” Eur. J. Oper. Res., vol. 221, no. 1, pp. 148-154, Aug. 2012.

[18] J. Chen, G.-Q. Zeng, W. Zhou, and W. Du, “Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization,” Energ. Convers. Manage., vol. 165, pp. 681-695, Jun. 2018.

[19] Park, H. S. Kim, Lee, and Song, “Temperature Prediction Using the Missing Data Refinement Model Based on a Long Short-Term Memory Neural Network,” Atmosphere, vol. 10, no. 11, pp. 718, Nov. 2019.

[20] D. Kreuzer, M. Munz, and S. Schlüter, “Short-term temperature forecasts using a convolutional neural network — An application to different weather stations in Germany,” Machine Learning with Applications, vol. 2, pp. 100007, Dec. 2020.

[21] Y. Liu, M. C. Roberts, and R. Sioshansi, “A vector autoregression weather model for electricity supply and demand modeling,” J. Mod. Power Syst. Clean Energy, vol. 6, no. 4, pp. 763-776, July. 2018.

[22] T. Lu, and M. Viljanen, “Prediction of indoor temperature and relative humidity using neural network models: Model comparison,” Neural. Comput. Appl., vol. 18, pp. 345-357, May. 2009.

[23] X. Shi, W. Lu, Y. Zhao, and P. Qin, “Prediction of Indoor Temperature and Relative Humidity Based on Cloud Database by Using an Improved BP Neural Network in Chongqing,” IEEE Access, vol. 6, pp. 30559-30566, Jun. 2018.

[24] P. Hewage, A. Behera, M. Trovati, and E. Pereira, “Long-Short Term Memory for an Effective Short-Term Weather Forecasting Model Using Surface Weather Data,” in AIAG, Hersonissos, Crete, Greece, 2019, pp. 382-390.

[25] Y. Zhang, and V. Hanby, “Short-term prediction of weather parameters using online weather forecasts,” in IBPSA, Beijing, China, 2007, pp. 1411-1416.

[26] J.-B. Yang, N. N hut, P. San, and X. li, “Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition,” in ICAI, Buenos Aires, Argentina, 2015, pp. 3995-4001.

[27] S. Ö. Ark, H. Jun, and G. Diamos, “Fast Spectrogram Inversion Using Multi-Head Convolutional Neural Networks,” IEEE Signal Process. Lett., vol. 26, no. 1, pp. 94-98, Nov. 2018.

[28] S. Hochreiter, and J. Schmidhuber, “Long Short-Term Memory,” Neural Comput., vol. 9, no. 8, pp. 1735-80, Dec. 1997.

[29] S. Hochreiter, and J. Schmidhuber, “LSTM can solve hard long time lag problems,” in Advances in Neural Information Processing Systems 9, NIPS, Denver, CO, USA, 1996, pp. 473-479.

[30] H. Zhuhua, Y. Zhang, Y. Zhao, and M. Xie, “A Water Quality Prediction Method Based on the Deep LSTM Network Considering Correlation in Smart Mariculture,” Sensors, vol. 19, no. 6, pp. 1420, Mar. 2019.

[31] K. Rao, G. Devi, and N. Ramesh, “Air Quality Prediction in Visakhapatnam with LSTM based Recurrent Neural Networks,” Int. J. Intell. Syst., vol. 11, no. 2, pp. 18-24, Feb. 2019.

[32] W. Xiangxue, X. Lunhui, and C. Kaixun, “Data-Driven Short-Term Forecasting for Urban Road Network Traffic Based on Data Processing and LSTM-RNN,” Arab. J. Sci. Eng., vol. 44, pp. 3043-3060, Jun. 2018.

[33] Z. Cui, R. Ke, and Y. Wang, “Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction,” ArXiv, Jan. 2018.

[34] M. J. Bappy, C. Simons, L. Nataraj, and B. Manjunath, “Hybrid LSTM and Encoder-Decoder Architecture for Detection of Image Forgeries,” IEEE Trans. Image Process., vol. 28, no. 7, pp. 3286 - 3300, Jan. 2019.

[35] Z. Khan, T. Hussain, A. Ullah, and S. Rho, “Towards Efficient Electricity Forecasting in Residential and Commercial Buildings: A Novel Hybrid CNN with a LSTM-AE based Framework,” Sensors, vol. 20, no. 5, pp. 1399, Mar. 2020.

[36] W. Luo, W. Liu, and S. Gao, “Remembering history with convolutional LSTM for anomaly detection,” in ICME, Hong Kong, China, 2017, pp. 439-444.

[37] S. Vosoughi, P. Vijayaraghavan, and D. Roy, “Tweet2Vec: Learning Tweet Embeddings Using Character-level CNN-LSTM Encoder-Decoder,” in SIGIR, Pisa, Italy, 2016, pp. 1041-1044.