Deep Spatio-Temporal Attention Model for Grain Storage Temperature Forecasting

†Shanshan Duan, †Weidong Yang, ‡Xuyu Wang, §Shiwen Mao, and †Yuan Zhang
†*Henan Key Laboratory of Grain Photoelectric Detection & Control, Key Laboratory of Grain Information Processing & Control, College of Information Science & Engineering, Henan University of Technology, Zhengzhou, Henan, China*
‡*Department of Computer Science, California State University, Sacramento, CA 95819-6021 USA*
§*Department of Electrical & Computer Engineering, Auburn University, Auburn, AL 36849-5201 USA*
18903838857@163.com, Yangweidong@haut.edu.cn, xuyu.wang@csus.edu, smao@ieee.org, zhangyuan@haut.edu.cn

2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS) | 978-1-7281-9074-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICPADS51040.2020.00082

***Abstract*—Temperature is one of the major ecological factors that affect the safe storage of grain. In this paper, we propose a deep spatio-temporal attention mode to predict stored grain temperature, which exploits the historical temperature data of stored grain and the meteorological data of the region. In this proposed model, we use the Sobel operator to extract the local spatial factors, and leverage the attention mechanism to obtain the global spatial factors of grain temperature data and temporal information. In addition, a convolutional neural network (CNN) is used to learn features of external meteorological factors. Finally, the spatial factors of grain pile and external meteorological factors are combined to predict future grain temperature using long short-term memory (LSTM) based encoder and decoder models. Experiment results show that the proposed model achieves higher predication accuracy compared with the traditional methods.**

***Keywords*-Grain Temperature; Spatio-temporal Model; Meteorological Metrics; Attention Mechanism; Deep Learning.**

I. Introduction

With the world population growth (predicted to grow from the present 7.4 billion to 9.1 billion by 2050), the demand for food is expected to increase by 70% by 2050 due to the increases in both population and per capita consumption. Annual cereal production will be required to increase to over 3 billion tons by 2050 to meet this demand [1]. Based on the forecast of Food and Agriculture Organization (FAO), world cereal stocks have been raised to 929 million tons by the close of seasons in 2021 [2]. In addition, the COVID-19 pandemic greatly impacts the economic growth, energy markets, food production and demand. Therefore, safe storage of grain is of unprecedented importance.

The temperature and moisture level of grain are the two important parameters that affect the quality of grain storage. Effective techniques to measure or predict these factors would be highly desirable. For example, Yang *et al.* proposed economic, contactless, and fast wheat moisture and mildew detection systems utilizing commodity WiFi and machine learning [3]–[5]. In addition, many traditional methods have been applied for predicting the temperature of grain pile, such as linear regression, autoregressive integrated moving average (ARIMA), exponential smoothing, and nonparametric regression. A 3D transient heat, mass, momentum, and species transfer model was proposed for the stored grain ecosystem with a finite element method [6]. Furthermore, machine learning methods have been used to improve the temperature predication performance. We have exploited the meteorological factors to predict temperature changes on the surface of grain pile using different machine learning methods (e.g., Support vector machine (SVM) and adaptive boosting (AdaBoost)) [7], [8]. However, these methods do not consider the spatial-temporal features in stored grain temperature data, which are collected from temperature sensors deployed at different locations (spatial) at different time instances (temporal). Thus there is considerable room for improving the predication performance.

Recently, deep learning and spatio-temporal models have been widely used in prediction problems, especially in computer vision, natural language processing (NLP), and big data in smart city [9]. For example, spatio-temporal model was applied to visual question answering [10] and utilized for solving the challenging problem of automatic video captioning [11]. In intelligent transportation systems, spatio-temporal model was used to address the diversity and complexity of urban traffic prediction [12]. The attention mechanism has been effectively used in intelligent traffic prediction [13], [14] and geo-sensor data prediction [15] for its salient advantages of having fewer parameters, fast execution, and good performance. In addition, a deep hybrid spatio-temporal dynamic neural network (DHSTNet) was used to predict both inflows and outflows in every region of a city [16]. A spatio-temporal position prediction Model (SPPM) based on long short-term memory (LSTM) was developed to predict the next spatio-temporal trajectory point of mobile user [17]. In cellular networks, a spatial modeling depth model based on auto-encoder was proposed for spatiotemporal modeling and prediction [18].

Motivated by the above interesting works, in this paper, we propose a novel *deep spatio-temporal attention model* to predict the temperature of stored grain. The proposed model is used to predict the trend of change of grain temperature. In fact, the grain temperature is an important ecological factor of grain storage, which is affected by external meteorological factors. The grain temperature in a granary will change along with the external meteorological factors, but with a certain

2690-5965/20/$31.00 ©2020 IEEE
DOI 10.1109/ICPADS51040.2020.00082

time delay. When the temperature of grain in the granary exceeds a certain threshold, it will cause grain heating, mildew, insect damage, etc, thus resulting in the loss of grain in the storage process. The proposed model is used to accurately predict the grain temperature in the granary. Once the trend of grain temperature changes significantly, ventilation, cooling, and other manual operations in the granary should be conducted to prevent the loss of grain.

Compared with the traditional spatio-temporal model, the attention mechanism in the proposed model can effectively reduce the number of parameters in the deep spatio-temporal model, and adpatively implement feature selection at different time instances for different temperature sensors, which help to improve the temperature forecasting performance. In addition, the computation of the attention mechanism in each step does not depend on the results in the previous step, which is amenable for parallel processing and achieving scalability in GPU-accelerated deep models.

The proposed deep spatio-temporal attention model includes three main parts. (i) A spatio encoder based on LSTM and convolutional neural network (CNN): specifically, we first extract the local and global spatial factors that affect the target temperature time series through the Sobel operator and attention mechanism, respectively. We then obtain the time dependence through LSTM. (ii) Time attention mechanism: used to capture the temporal correlation between spatial factors, which is able to adaptively select the hidden states. (iii) In a general fusion module, the spatial factors of grain pile and external meteorological factors are combined to predict grain temperature using LSTM based encoder and decoder models. We use real grain temperature data and meteorological data of the granary location to evaluate the proposed prediction method. The results show that the proposed model outperforms the ARIMA and LSTM baseline schemes with significant improvements.

The remainder of this paper is organized as follows. Section II presents the forecasting problem and preliminary analysis. Section III introduce the proposed deep spatio-temporal attention network model. Section IV validates the performance of the proposed method using real world data. Section V summarizes this paper.

II. Grain Temperature Forecasting Problem and Preliminary Analysis

A. Problem Statement

In our experiments, 200 temperature sensors are deployed in the grain pile, each of which can generate a time series of temperature measurements. We consider such time series as the predicted target sequence, which is characterized by the spatial factors that affect the temperature of the target location and the external meteorological factors that affect the temperature of the entire grain pile. Given a time window of duration T, let $\mathbf{Y} = \left(\mathbf{y}^1, \mathbf{y}^2, ..., \mathbf{y}^{200}\right) \in \mathbb{R}^{200\times T}$ denote the data of all temperature series during the past T days, where $\mathbf{y}^i \in \mathbb{R}^T$. The predication problem is that given the historical temperature data from each temperature sensor and the meteorological data of the corresponding region, develop a model to predict the temperature of sensor i for the next τ days, denoted as $\hat{\mathbf{y}}^i = \left(\hat{y}^i_{T+1}, \hat{y}^i_{T+2}, ..., \hat{y}^i_{T+\tau}\right)^\top \in \mathbb{R}^\tau$.

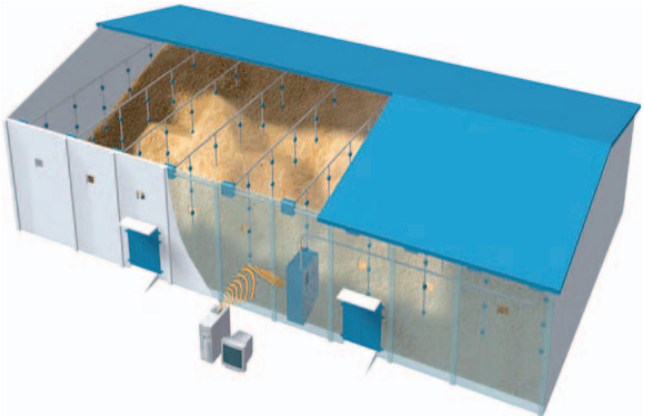

Figure 1. A tall flat granary where a temperature sensor network is deployed.

B. Data Collection and Preprocessing

In this paper, the data used follows the national standard and the distribution principle of temperature measurement cables in a large flat warehouse. The dataset consists of wheat temperature data from a large warehouse in Zhumadian, Henan Province, China and the corresponding meteorological data from the China Meteorological Data Network. The dataset includes daily grain temperature measured by the temperature sensors from July 3, 2015 to December 17, 2018. The frequency of measuring temperature data with a temperature sensor is three times a week. Before modeling and forecasting, we interpolate the grain temperature data to obtain one temperature value per day, which is synchronized with the daily meteorological data for the same area. Fig. 1 shows the tall flat granary with a temperature sensor network. 200 temperature sensors are deployed, where 10 rows of sensors are arranged on the X-axis direction (46 m long) and 5 rows of sensors on the Y-axis direction (21 m wide). The height of the grain pile in the granary is 6 m, and 4 layers of sensors are deployed along the Z-axis.

There are some missing values and abnormal values in collected data because of sensor failure and operational problems. In addition, the collected meteorological data is inconsistent in unit order of magnitude. Accordingly, we need to address such missing and abnormal values in grain temperature data, and normalize all the data. The 3σ criterion is used to detect and remove outliers in this paper. The datasets should be normalized to ensure that sensor temperature data and meteorological data have the same unit. The StandardScaler is used to normalize the datasets.

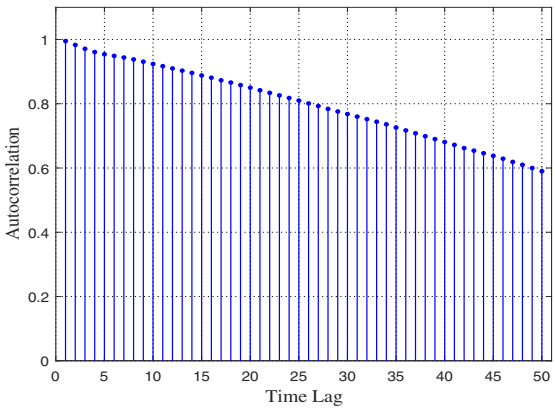

Figure 2. Temporal autocorrelation coefficients at different time lags.

C. Data Analysis

Consider the sensor at spatial coordinates (x, y, z), denoted by $s(x, y, z)$. The temperature data series D obtained from the sensor is regarded as a time series. We test the correlation of data from the time and space domains, respectively, and summarize our main findings in the following.

- *The temperature data series D of the stored grain pile presents a non-zero autocorrelation in the time domain.* The autocorrelation coefficient [19] of the temperature samples in the grain pile is defined as:

$$R_p = \frac{\sum_{t=1}^{d} (y_t - \bar{y})(y_{t+\xi} - \bar{y})}{\sum_{i=1}^{d} (y_t - \bar{y})^2}, \quad (1)$$

where y_t and $y_{t+\xi}$ are the grain temperature measurements at adjacent times t and $t + \xi$, respectively; $\bar{y}$ is the average grain temperature; and d is the number of samples. The unit of temperature in this paper is °C. Fig. 2 shows the autocorrelation at time lag $\xi = 0, 1, ..., 50$ for the grain temperature data of any point sensor. We can see that there is strong correlation between the values of grain temperature data at any time. Therefore, the time series of grain temperature has non-zero time autocorrelation in the time domain.

- *The temperature data series D has a non-zero spatial correlation among different sensor locations.* According to the Pearson correlation coefficient defintion [20], the correlation between the data series from two temperature sensors in a grain pile is defined as:

$$R_{pp} = \frac{\sum_{t=1}^{d} (y_{it} - \bar{y}_i)(y_{jt} - \bar{y}_j)}{\sqrt{\sum_{t=1}^{d} (y_{it} - \bar{y}_i)^2} \sqrt{\sum_{t=1}^{d} (y_{jt} - \bar{y}_j)^2}}, \quad (2)$$

where y_{it} and y_{jt} are the grain temperature values at two temperature sensors i and j at time t, respectively; $\bar{y}_i$ and $\bar{y}_j$ are the average temperatures at the two temperature sensors, respectively. We examine the correlation among different sensors and present the results among nine neighboring temperature sensors in Table I, where (Row, Column, Layer) represents the row, column, and layer coordinates of the temperature sensor location in the granary. We can see that the correlation coefficients between different sensors vary greatly in a certain range. For instance, the correlation value of temperature data of sensors (1, 2, 2) and (1, 2, 3) is 0.523, while that of sensors (1, 2, 3) and (1, 2, 4) is 0.872, even though sensors (1, 2, 2) and (1, 2, 3) have the same spatial relationship as sensors (1, 2, 3) and (1, 2, 4). These results show that the grain temperature is highly correlated spatially.

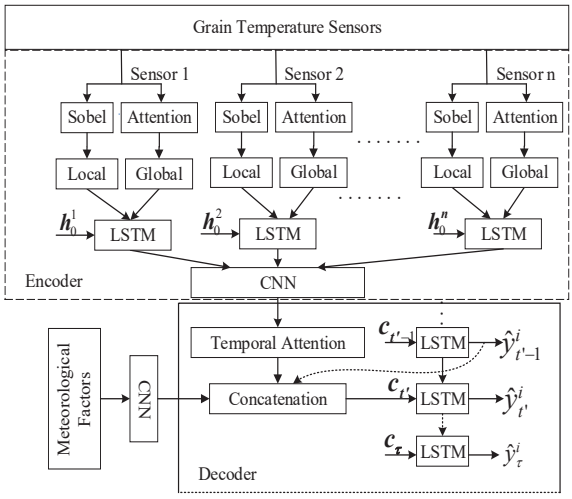

Figure 3. The proposed deep spatio-temporal attention model. (Local: local factors; Global: global factors; $\hat{y}_t^i$: the predicted value at time t; c_t: the context vectors at time t; h_0^i: the initial state of each LSTM model.)

III. Deep Spatio-Temporal Attention Model

We design a deep spatio-temporal attention model to solve the temperature forecasting problem, which leverages historical data of grain temperature collected by the sensors and the historical data of meteorological factors. Fig. 3 presents the framework of our approach, which includes feature representation (i.e., local spatial factors, global spatial factors, and external meteorological factors), temporal attention, and encoder and decoder. We will discuss these modules in detail in the following.

A. Feature Representation

1) Local Spatial Factors: In order to facilitate the study of local temperature change in the grain pile, temperature field and temperature gradient are introduced here. In general, temperature field is a function of space coordinates and time, which is denoted by $F = f(x, y, z, t)$, where (x, y, z) are the space coordinates and t is time. The isotherm and

Table I
CORRELATION MATRIX OF THE DATA SERIES FROM NINE SENSORS

(Row, Column, Layer)	(1, 1, 2)	(1, 1, 3)	(1, 1, 4)	(1, 2, 2)	(1, 2, 3)	(1, 2,4)	(2, 1, 2)	(2, 1, 3)	(2, 1, 4)
(1, 1, 2)	1.000	0.559	0.481	0.995	0.510	0.190	0.997	0.636	0.296
(1, 1, 3)	0.559	1.000	0.964	0.573	0.995	0.838	0.553	0.989	0.906
(1, 1, 4)	0.481	0.964	1.000	0.491	0.969	0.929	0.475	0.937	0.961
(1, 2, 2)	0.995	0.573	0.491	1.000	0.523	0.196	0.996	0.649	0.304
(1, 2, 3)	0.510	0.995	0.969	0.524	1.000	0.872	0.504	0.980	0.932
(1, 2, 4)	0.190	0.838	0.929	0.196	0.872	1.000	0.184	0.781	0.981
(2, 1, 2)	0.997	0.553	0.475	0.996	0.504	0.184	1.000	0.632	0.290
(2, 1, 3)	0.636	0.989	0.937	0.649	0.980	0.781	0.632	1.000	0.867
(2, 1, 4)	0.296	0.906	0.961	0.304	0.932	0.981	0.290	0.867	1.000

isothermal surface are caused by the temperature difference inside the grain pile, i.e., the change of temperature. There is a temperature change rate along the isotherm and the normal direction of the isothermal surface in mathematical sense, which has the maximum value among all changes, called the *temperature gradient*. The components of the three coordinate axes represent the magnitude and direction of the temperature gradient, namely:

$$\nabla F = \frac{\partial F}{\partial x} \cdot \mathbf{i} + \frac{\partial F}{\partial y} \cdot \mathbf{j} + \frac{\partial F}{\partial z} \cdot \mathbf{k}. \tag{3}$$

In a limited area, because the distance between the isotherms in temperature points cannot be calculated accurately, the temperature gradient values cannot be obtained directly. Therefore, the Sobel operator [21] is introduced to calculate the approximate gradient at each point in the temperature field. Using this operator at all the points will produce the corresponding gradient vector. Sobel operator consists of two sets of 3×3 matrices, which are horizontal and vertical, respectively [22]. Let $\mathbf{A}$ be an image (i.e., a plane matrix) from the temperature field F. To obtain the approximate gradient vector (G_x, G_y) of image $\mathbf{A}$, we use two 3×3 kernels to convolute with image $\mathbf{A}$ as follows.

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \circledast \mathbf{A} \tag{4}$$

$$G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \circledast \mathbf{A}. \tag{5}$$

Then, the gradient amplitude G and direction Θ of $\mathbf{A}$ are:

$$G = \sqrt{G_x^2 + G_y^2} \quad \text{and} \quad \Theta = \arctan\left(\frac{G_y}{G_x}\right). \tag{6}$$

When calculating the gradient of the edge points of the grain pile, the temperature points are interpolated first. $\mathbf{A}$ is a 3×3 temperature plane matrix centered on the predicted target point. The granary is regarded as a cube with a matrix $\mathbf{A}$ in $x-y$ plane, $x-z$ plane and $y-z$ plane, respectively. Then, the gradient directions are calculated in the $x-y$ plane, $x-z$ plane, and $y-z$ plane, respectively, where 10 temperature values are used in the gradient direction for each plane. Totally, 30 points on the three planes are employed as the local influence factors of the target points. The output vector of the local space factor is:

$$\mathbf{l}_t = \left(y_t^{grad,1}, y_t^{grad,2}, ..., y_t^{grad,L}\right)^\top, \tag{7}$$

where $y_t^{grad,\ell}$ is the computed ℓth temperature value in the direction of the temperature gradient at the target point (e.g., L is 30 for the temperature sensor network).

2) Global Spatial Factors: By analyzing the correlation between the temperature points (see Section II-C), we conclude that there are some temperature time series with weaker correlation. Because the temperature series change over time, the weight of the impact of other temperature points on a target point also changes dynamically. Thus, directly using each temperature time series as inputs to the encoder to capture the relationship between different temperature sensors, will cause high computational overhead and reduce the learning ability of the model. Moreover, the weights are affected by local conditions at other temperature points. To address this issue, we leverage the attention mechanism to explore the dynamic relationship between different temperature sensors.

The proposed model consists of an encoder and a decoder. Let $\mathbf{h}_{t-1} \in \mathbb{R}^m$ and $\mathbf{s}_{t-1} \in \mathbb{R}^m$ denote the hidden state and cell state of the encoder at time $t-1$, respectively. The attention score between a temperature sensor p and the target sensor i can be calculated as follows [11], [15]:

$$a_t^p = \mathbf{v}_g^\top \tanh\left(\mathbf{W}_g\left[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}\right] + \mathbf{U}_g \mathbf{y}^p + \mathbf{V}_g \mathbf{y}^i + \mathbf{b}_g\right), \tag{8}$$

$$\delta_t^p = \frac{\exp\left(a_t^p\right)}{\sum_{j=1}^{200} \exp\left(a_t^j\right)}, \tag{9}$$

where $[\cdot;\cdot]$ represents the concentration operation, $\mathbf{v}_g, \mathbf{b}_g \in \mathbb{R}^T, \mathbf{W}_g \in \mathbb{R}^{T \times 2m}, \mathbf{U}_g \in \mathbb{R}^{T \times T}$, $\mathbf{V}_g \in \mathbb{R}^{T \times T}$ are model parameters. By considering the local influence factors of the target temperature point and other temperature sensors, the attention mechanism can make a prediction by selecting the temperature point of the related temperature sensor adaptively. Moreover, the hidden state $\mathbf{h}_{t-1}$ and cell state $\mathbf{s}_{t-1}$ in the encoder at time $t-1$ allow the historical temperature dynamically propagate over time. After the attention score

is obtained, the output vector of the global space factor at time step t is calculated as:

$$\mathbf{g}_t = \left(\delta_t^1 y_t^{i,1}, \delta_t^2 y_t^{i,2}, ..., \delta_t^{200} y_t^{i,200}\right)^\top. \tag{10}$$

3) External Meteorological Factors: In this paper, we consider the meteorological features as external factors, including daily average air temperature (TEM), air pressure (PRS), 0 cm ground surface temperature (GST), relative humidity (RHU), daily average evaporation (EVP), sunshine duration (SSD), wind speed (WIN), and precipitation (PRE). The correlation coefficient between the average surface temperature of grain temperature and meteorological factors is shown in Table II. We can see that there is a strong correlation between the average surface temperature of grain pile and the meteorological factors.

As shown in Fig. 2, the meteorological factors will be learned by a CNN. In this paper, there are eight meteorological elements at each time t so that the input layer of CNN has eight neurons. For better feature fusion, a unique hidden layer is added and its number of neurons is set to 32. In fact, the original meteorological elements are nonlinear dimensionally increased, the number of output neurons is 64, and the output of LSTM is concatenated to complete the hidden feature.

B. Temporal Attention

Generally, the encoder-decoder performance will decline with the increase in the length of time series. In order to capture the temporal correlation between spatial factors, we also exploit the time attention mechanism to determine which hidden states are of interest [23], [24]. It updates new hidden states according to the current input and related hidden states, and establish a dynamic time correlation model between different time intervals in the target sequences. In order to obtain the attention vector, we calculate the context vector $\mathbf{c}_{t'}$ of each output time period t', which depends on the annotation vector of the encoder $(h_1, h_2, ..., h_T)$. The encoder maps the input spatial factors to these annotation vectors, and each annotation contain information about the entire input time sequence. The context vector $\mathbf{c}_{t'}$ is calculated as a weighted sum of the annotation vectors $\mathbf{h}_o$, as:

$$\mathbf{c}_{t'} = \sum_{o=1}^{T} \gamma_{t'}^o \mathbf{h}_o, \tag{11}$$

where the weight $\gamma_{t'}^o$ of each annotation vector $\mathbf{h}_o$ is calculated as

$$\gamma_{t'}^o = \frac{\exp\left(u_{t'}^o\right)}{\sum_{j=1}^{T} \exp\left(u_{t'}^j\right)}, \text{ and} \tag{12}$$

$$u_{t'}^o = \mathbf{v}_d^\top \tanh\left(\mathbf{W}_d' \left[\mathbf{d}_{t'-1}; \mathbf{s}_{t'-1}'\right] + \mathbf{W}_d \mathbf{h}_o + \mathbf{b}_d\right), \tag{13}$$

where $\mathbf{W}_d \in \mathbb{R}^{m\times m}$, $\mathbf{W}_d' \in \mathbb{R}^{m\times 2n}$, $\mathbf{v}_d$, $\mathbf{b}_d \in \mathbb{R}^m$ are model parameters, $\mathbf{d}_{t'-1} \in \mathbb{R}^n$ and $\mathbf{s}_{t'-1} \in \mathbb{R}^n$ are the hidden state and cell state of the decoder at time $t'-1$, respectively. $\gamma_{t'}^o$ or its associated quantity $u_{t'}^o$ indicates the importance of the annotation vector $\mathbf{h}_o$ in determining the next hidden state compared to the previous hidden state.

C. Encoder and Decoder

We use LSTM to encode local feature $\{\mathbf{l}_t\}_{t=1}^T$ and global feature $\{\mathbf{g}_t\}_{t=1}^T$, and concatenate them to get the spatial feature $\mathbf{m}_t = [\mathbf{l}_t; \mathbf{g}_t]$. It is used as input to the LSTM model, which can update the hidden state $\mathbf{h}_t$ at time t as follows.

$$\mathbf{h}_t = \mathrm{LSTM}\left(\mathbf{m}_t, \mathbf{h}_{t-1}\right). \tag{14}$$

In order to further reduce the impact of noise on the decoder, the output of each LSTM is passed through the CNN module to reduce its dimension. The output of CNN is the final feature vector of the encoder. After obtaining the external meteorological feature $\mathbf{ex}_t$, the context vector $\mathbf{c}_t$ and the previous temperature prediction value $\hat{y}_{t'-1}^i$, we feed them to the decoder LSTM, and update the hidden state of the decoder with $\mathbf{d}_{t'} = f_d\left(\mathbf{d}_{t'-1}, \left[\hat{y}_{t'-1}^i; \mathbf{ex}_{t'}; \mathbf{c}_{t'}\right]\right)$, where $f_d(\cdot)$ is the LSTM unit used in the decoder and t' is a future time step. Then, we feed the context vector $\mathbf{c}_{t'}$ and the hidden state $\mathbf{d}_{t'}$ to the LSTM unit. Finally, the decoder is trained to predict the next grain temperature, given by:

$$\hat{y}_{t'}^i = \mathbf{v}_y^\top \left(\mathbf{W}_m \left[\mathbf{c}_{t'}, ; \mathbf{d}_{t'}\right] + \mathbf{b}_m\right) + b_y, \tag{15}$$

where $\mathbf{W}_m \in \mathbb{R}^{n\times(m+n)}$ and $\mathbf{b}_m \in \mathbb{R}^n$ map the concatenated $[\mathbf{c}_{t'}; \mathbf{d}_{t'}] \in \mathbb{R}^{m+n}$ to the size of the decoder hidden state; $\mathbf{v}_y \in \mathbb{R}^n$ and $b_y \in \mathbb{R}$ are linear transformation parameters, which can predict the final output. The Adam optimizer is used to train the model to minimize the root mean square error (RMSE) between the predicted temperature value $\hat{\mathbf{y}}^i$ and the real temperature value $\mathbf{y}^i$.

IV. Performance Evaluation

A. Implementation

In this paper, model training is implemented using the tensorflow library. The local feature dimension of spatial factor is 30, and the global feature dimension is 199. We use the Sobel operator and attention mechanism to extract the local and global feature of spatial factors, respectively. The feature dimension of the external weather factor is 8 and a CNN with one layer is used to extract the external features. In the encoding stage, 12 temperature time series are selected for data expansion to help supervise the learning and training model. This method is suitable for global features and external input data. In the decoding phase, 6 time series are used for data expansion and data enhancement. The encoding of spatial factors simultaneously exploits the relationship between different sensors. 30 local influence factors of each temperature point are used as its local input. In the model training, the initial learning rate is 0.001, and the coefficient of avoiding gradient explosion is 2.5. Data

Table II
CORRELATION MATRIX OF GRAIN TEMPERATURE AND METEOROLOGICAL FACTORS

	TEM	GST	PRS	RHU	SSD	WIN	EVP	GRAIN TEMPERATURE
TEM	1.000	0.983	-0.910	0.139	0.280	-0.153	0.469	0.923
GST	0.983	1.000	-0.888	0.119	0.351	-0.180	0.507	0.913
PRS	-0.910	-0.888	1.000	-0.222	-0.195	0.126	-0.342	-0.813
RHU	0.139	0.119	-0.222	1.000	-0.536	-0.265	-0.566	0.294
SSD	0.280	0.351	-0.195	-0.536	1.000	0.03	0.646	0.128
WIN	-0.153	-0.180	0.126	-0.265	0.03	1.000	0.178	-0.188
EVP	0.469	0.507	-0.342	-0.566	0.646	0.178	1.000	0.365
GRAIN TEMPERATURE	0.923	0.913	-0.813	0.294	0.128	-0.188	0.365	1.000

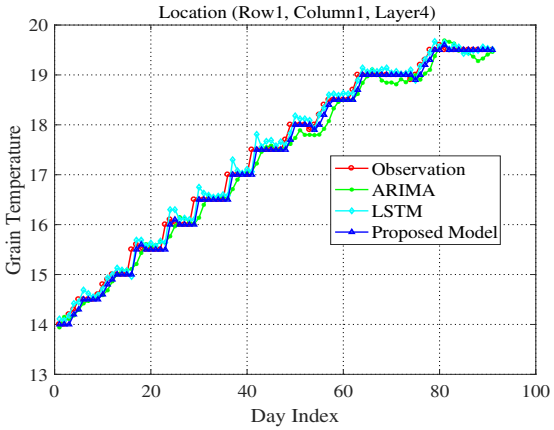

Figure 4. Grain temperature prediction at location (row 1, column 1, layer 4) of stored grain pile.

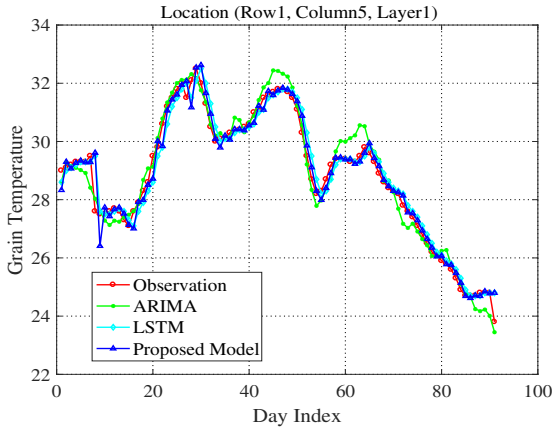

Figure 5. Grain temperature prediction at location (row 1, column 5, layer 1) of stored grain pile.

expansion and dropout [25], [26] are used to avoid overfitting, and the ratio of dropout is set to 0.3.

B. Model Results and Comparison

To verify the performance of the proposed model, we predict the grain temperature at different locations, including the boundary and near the center of the grain pile. Two baseline methods, i.e., ARIMA and LSTM, are used for comparison purpose. The results are shown in Figs. 4 to 8. Fig. 4 and Fig. 5 show the results at two locations close to the eastern boundary of the granary, where Fig. 4 is for the sensor close to the northern side of the granary and Fig. 5 is for the sensor close to the southern side of the granary. Also, Fig. 6 presents the results for the sensor in the middle of the grain pile. Figs. 7 and 8 present the results for the sensors close to the west boundary of the granary, where Fig. 7 is for the sensor on the upper layer of the grain pile and Fig. 8 is for the sensor on the lower layer of the grain pile. The results show that the proposed model can accurately predict temperature for all the positions in the grain pile. Moreover, the proposed model outperforms the two baseline schemes (i.e., ARIMA and LSTM) with a superior predication performance.

We use RMSE as performance metric for all the models. The RMSE results of the three prediction models are

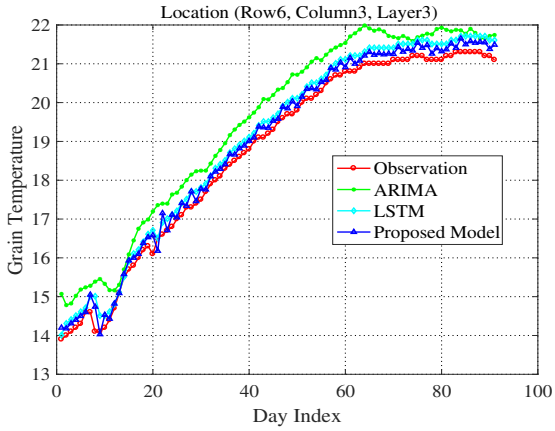

Figure 6. Grain temperature prediction at location (row 6, column 3, layer 3) of stored grain pile.

presented in Table III. For the sensor (1, 1, 4) located at the lower layer of the southeast boundary, the RMSE of the proposed model is 0.1464, which achieves a reduction of 26.87% and 5.73% over ARIMA and LSTM, respectively. For the sensor (1, 5, 1) located at the upper layer of the northeast boundary, the RMSE of the proposed model is 0.3756, which achieves a reduction of 7.49% and 7.42% over ARIMA and LSTM, respectively. For the sensor (6, 3,

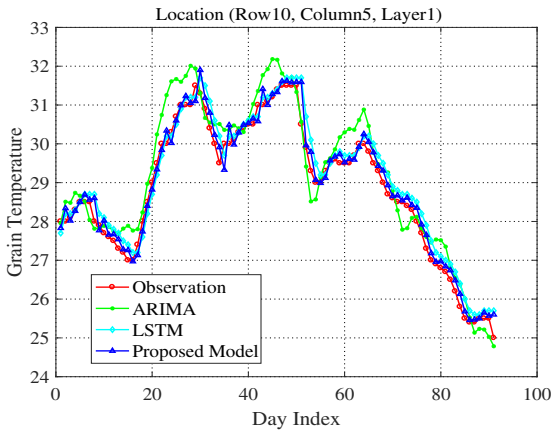

Figure 7. Grain temperature prediction at location (row 10, column 5, layer 1) of stored grain pile.

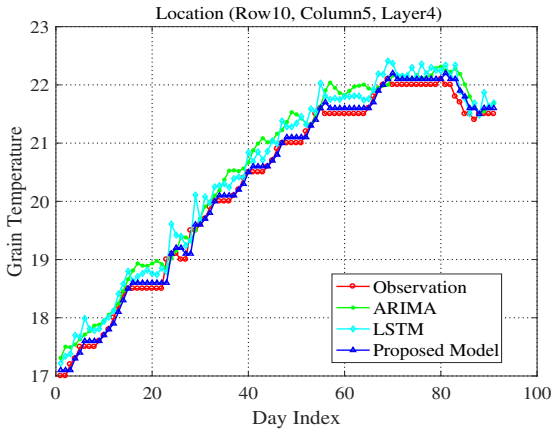

Figure 8. Grain temperature prediction at location (row 10, column 5, layer 4) of storage grain pile.

Table III
RMSE OF FORECAST RESULTS AT DIFFERENT LOCATIONS

Location	(1, 1, 4)	(1, 5, 1)	(6, 3, 3)	(10, 5, 1)	(10, 5, 4)
ARIMA	0.2002	0.4060	0.7713	0.5330	0.3087
LSTM	0.1553	0.4057	0.3495	0.3720	0.2785
Proposed	0.1464	0.3756	0.2502	0.2676	0.1204

3) located at the center of the grain pile, the RMSE of the proposed model is 0.2502, which achieves a reduction of 67.56% and 28.41% over ARIMA and LSTM, respectively. For the sensor (10, 5, 1) located at the upper layer of the northwest boundary, the RMSE of the proposed model is 0.2676, which achieves a reduction of 49.79% and 28.06% over ARIMA and LSTM, respectively. For the sensor (10, 5, 4) located at the lower layer of the northwest boundary, the RMSE of the proposed model is 0.1204, which achieves a reduction of 61% and 56.77% over ARIMA and LSTM, respectively.

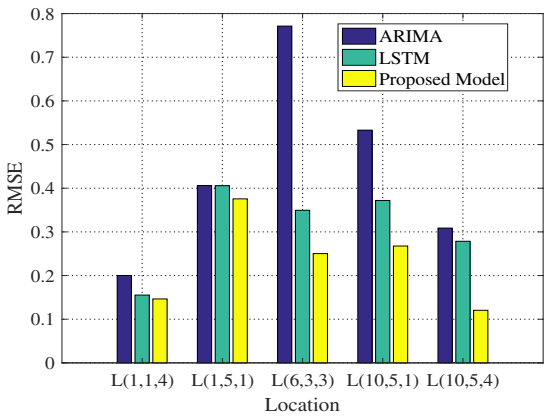

Figure 9. RMSEs of different forecasting methods at different locations.

It can be seen that the prediction performance of the LSTM method with meteorological factors is better than that only using historical data, and the method proposed in this paper has the best prediction performance for all the locations. Fig. 9 presents the RMSE results obtained with the three methods. The results verify the effectiveness of the proposed deep spatial-temporal attention model.

V. CONCLUSIONS

In this paper, we proposed a deep spatial-temporal attention model for forecasting the temperature of stored grain, which consisted of an attention mechanism, a CNN model, and an LSTM-based encoder and decoder model. In the proposed model, the Sobel operator was used to extract the local spatial factors, and the attention mechanism was applied to obtain the global spatial factors of the grain temperature data and temporal information. The CNN was used to learn features of external meteorological factors. Finally, the spatial factors of grain pile and external meteorological factors were combined to predict future grain temperature. The experiment results demonstrated that the proposed model achieved high prediction accuracy than the traditional methods (i.e., ARIMA only using historical data and LSTM using meteorological factors).

ACKNOWLEDGMENT

This work is supported in part by the National Key Research and Development Program of China (2017YFD0401001), the Natural science project of Henan Education Department (19HASTIT027), the Central Plains Scholar Project (172101510005), Open fund of Key Laboratory of Grain Information Processing and Control (KFJJ-2017-201), Open Fund of Key Laboratory of Grain Photoelectric Detection and Control (KFJJ-2016-302), the Natural Science Foundation of Henan Province (182300410036), and by the NSF under Grants ECCS-1923163 and ECCS-1923717, the Wireless Engineering Research and Education Center (WEREC) at Auburn University, Auburn, AL, USA.

REFERENCES

[1] C. Le Mouël and A. Forslund, "How can we feed the world in 2050? A review of the responses from global scenario studies," *European Review of Agricultural Economics*, vol. 44, no. 4, pp. 541–591, Aug. 2017.

[2] A. Ifpri, P. Pinstrup, R. Pandya-Lorch, and M. W. Rosegrant, "The world food situation: Recent developments," *Emerging Issues, and Long Term Prospects*, 2020.

[3] W. Yang, X. Wang, A. Song, and S. Mao, "Wi-Wheat: Contact-free wheat moisture detection using commodity WiFi," in *Proc. IEEE ICC 2018*, Kansas City, MO, May 2018, pp. 1–6.

[4] W. Yang, X. Wang, S. Cao, H. Wang, and S. Mao, "Multi-class wheat moisture detection with 5GHz Wi-Fi: A deep LSTM approach," in *Proc. ICCCN 2018*, Hangzhou, China, July/Aug. 2018, pp. 1–9.

[5] P. Hu, W. Yang, X. Wang, and S. Mao, "MiFi: Device-free wheat mildew detection using off-the-shelf WiFi devices," in *Proc. IEEE GLOBECOM 2019*, Waikoloa, HI, Dec. 2019.

[6] J. Lawrence, D. E. Maier, and R. L. Stroshine, "Three-dimensional transient heat, mass, momentum, and species transfer in the stored grain ecosystem: Part I. Model development and evaluation," *Transactions of the ASABE*, vol. 56, no. 1, pp. 179–188, Jan. 2013.

[7] S. Duan, W. Yang, X. Wang, S. Mao, and Y. Zhang, "Grain pile temperature forecasting from weather factors: A support vector regression approach," in *Proc. IEEE/CIC ICCC 2019*, Changchun, China, Aug. 2019, pp. 255–260.

[8] S. Duan, W. Yang, X. Wang, S. Mao, and Y. Zhang, "Forecasting of grain pile temperature from meteorological factors using machine learning," *IEEE Access*, vol. 7, pp. 130 721–130 733, Dec. 2019.

[9] X. Wang, X. Wang, and S. Mao, "RF sensing for Internet of Things: A general deep learning framework," *IEEE Communications*, vol. 56, no. 9, pp. 62–69, Sept. 2018.

[10] Y. Jang, Y. Song, Y. Yu, Y. Kim, and G. Kim, "TGIF-QA: Toward spatio-temporal reasoning in visual question answering," in *Proc. IEEE CVPR 2017*, Honolulu, HI, July 2017, pp. 2758–2766.

[11] M. Zanfir, E. Marinoiu, and C. Sminchisescu, "Spatio-temporal attention models for grounded video captioning," in *Proc. 2016 Asian Conference on Computer Vision*, Taipei, Taiwan, Nov. 2016, pp. 104–119.

[12] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, Aug. 2019, pp. 1720–1730.

[13] X. Shi, H. Qi, Y. Shen, G. Wu, and B. Yin, "A spatial-temporal attention approach for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, to appear.

[14] C. Park, *et al.*, "STGRAT: A spatio-temporal graph attention network for traffic forecasting," *arXiv preprint arXiv:1911.13181*, Nov. 2019.

[15] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "GeoMAN: Multi-level attention networks for geo-sensory time series prediction," in *Proc. IJCAI'18*, Stockholm, Sweden, July 2018, pp. 3428–3434.

[16] A. Ali, Y. Zhu, Q. Chen, J. Yu, and H. Cai, "Leveraging spatio-temporal patterns for predicting citywide traffic crowd flows using deep hybrid neural networks," in *Proc. IEEE ICPADS'19*, Tianjin, China, Dec. 2019, pp. 125–132.

[17] S. Tian, X. Zhang, Y. Zhang, Z. Cao, and W. Cao, "Spatio-temporal position prediction model for mobile users based on LSTM," in *Proc. IEEE ICPADS'19*, Tianjin, China, Dec. 2019, pp. 967–970.

[18] J. Wang, J. Tang, Z. Xu, Y. Wang, G. Xue, X. Zhang, and D. Yang, "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in *Proc. IEEE INFOCOM'17*, Atlanta, GA, May 2017, pp. 1–9.

[19] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. New York, NY: Springer, 2016.

[20] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*, J. Benesty, J. Chen, Y. Huang, and I. Cohen, Eds. New York, NY: Springer, 2009, pp. 1–4.

[21] R. C. Gonzales and R. E. Woods, *Digital image processing*. Upper Saddle River, NJ: Prentice Hall, 2002.

[22] Y. Zhang, X. Han, H. Zhang, and L. Zhao, "Edge detection algorithm of image fusion based on improved Sobel operator," in *Proc. IEEE 2017 Information Technology and Mechatronics Engineering Conference*, Chongqing, China, Oct. 2017, pp. 457–461.

[23] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, Sept. 2014.

[24] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. ICML'15*, Lille, France, July 2015, pp. 2048–2057.

[25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, June 2014.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS 2012*, Lake Tahoe, NV, Dec. 2012, pp. 1097–1105.