ICIC Express Letters
Part B: Applications
Volume **12**, Number **4**, April **2021**

ICIC International ©2021 ISSN 2185-2766
pp. 395–401

CROP PEST PREDICTION USING CLIMATE ANOMALY MODEL BASED ON DEEP-LSTM METHOD

Teguh Wahyono[1,2], Yaya Heryadi[1], Haryono Soeparno[3]
and Bahtiar Saleh Abbas[4]

[1]Doctor of Computer Science, Binus Graduate Program
[3]Bioinformatics and Data Science Research Center
[4]Department of Industrial Engineering
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
{ teguh.wahyono; yayaheryadi; haryono; bahtiar }@binus.edu

[2]Faculty of Information Technology
Satya Wacana Christian University
Jl. Diponegoro 52-60, Salatiga, Central Java 50711, Indonesia
teguh.wahyono@uksw.edu

Received August 2020; accepted November 2020

Abstract. *Most research on pest prediction used climate data as time-series historical data that could directly be used to predict the presence or occurrence of pest. However, the climate variable had a strong seasonality component in nature, which had similar characteristics in terms of its long range data as well as complex structures that require a more proper approach in its utilization. This research proposes a different approach. It uses climate data to predict pest attack by considering the possibility of anomaly in the data first. The initial step is detecting the climate anomaly, and then the anomaly data are used to predict the pest attack. The method used in this research is Deep-Long Short-Term Memory (LSTM), which has a strength in analyzing the time-series data and storing long-term memory. The results of the research show that the use of climate anomaly data for predicting the pest attack produced higher accuracy values. The testing accuracy values produced from the proposed model were 0.0698 for Mean Square Error (MSE) and 0.2606 for the Root Mean Square Error (RMSE). These results are the best performance compared to other models and settings in this study.*
Keywords: Crop pest prediction, Climate anomaly detection, LSTM, Deep learning

1. **Introduction.** Pest attack is one of major factors that causes crop failures in Indonesia, besides drought and natural disasters [1]. During the harvesting period of 2017/2018 for instance, approximately 63,000 ha of rice fields or almost 0.5% of the total farm lands in Indonesia were attacked by pests and experienced drought. From 63,075 ha of the fields attacked by pests, 20,152 ha experienced crop failures [2]. According to Food and Agriculture Organization (FAO), the pest attack causes the reduction of farming productivity until 40% every year [3].

Paddy is one of national staple plants that is vulnerable to pest attack. Although the size of the affected areas is decreasing from year to year, it is still relatively wide. Rice pest attacks always occur from year to year, where the total reach approximately 400,000, ha per year [4]. Due to the key role of paddy as a commodity, pest control for this plant is one of the most important priorities for the Indonesian Government, particularly the Agriculture Department, in order to attain food self-sufficiency. One of paddy main pests is the white rice stemborer (scirpophaga innotata). Table 1 below shows the attacked area of paddy field. It can be seen that based on the attacked area in 2019, white rice

DOI: 10.24507/icicelb.12.04.395

TABLE 1. The area coverage of major pests of paddy plants from 2017-2019

Pest	2017	2018	2019
White rice stemborer	87,197	47,747	46,233
Brown plant hopper	36,696	22,658	16,795
Rats	74,274	41,484	43,797
Blast	80,172	17,949	16,456
Bacterial leaf blight	48,412	18,798	17,660

stemborer was on the first rank, which attacked 46,233 ha of paddy fields. In the next positions were rats, bacterial leaf blight, brown plant hopper, and blasts [4].

Paddy pest attacks are highly influenced by climate factors. Temperature, rainfall, humidity, and photoperiodicity have direct impacts on insects' life cycles, span of life, and diapause ability. Rain that falls in the dry season, unsynchronized planting time, and unstable weather conditions can also stimulate pest attacks. Some occurrences show that *El-Nino* and *La-Nina* can stimulate the growth of pests such as rice stemborer species and brown plant hopper (Nilaparvata lugens Stal.) in West Java and Central Java, grasshoppers in Lampung, and tungro in Central Java, NTB, and South Sulawesi. Wet-dry season phenomenon also caused brown plant hopper (Nilaparvata lugens Stal.) attack to 50,000 ha of rice fields in Indonesia in 2017, in which 80% of the affected fields were located in Java Island [5].

Based on the aforementioned reasons, this research will develop a prediction model of paddy plants pest attacks by using climate data as the predicting factor. Some research has been done previously. Some methods have been developed for crop pest attacks prediction using climate factor, e.g., simple linear regression and multiple linear regression [6], Bayesian network [7], neural network [8], recurrent neural network [9], and long short term memory [10,11]. Most of the research used climate data as time-series historical data that could directly be used as a variable to predict. Meanwhile, the climate variable has a strong seasonality component, which had similar characteristics for longer data and complex structures so that a proper approach is needed [12,13].

Therefore, this research proposes a different approach by using the climate data to predict pest attacks by considering anomaly occurrences in the data. The anomaly occurrences should be taken into account since they have strong influence to other events around them [13]. To give anomaly label in the climate data and predict the pest attacks, a Long Short-Term Memory (LSTM) deep learning method is used. This method is chosen because it has the ability to store long-term memory, analyze the time-series data, and has been proven to work effectively in some prediction research [10,11,13].

The results of this study indicate that the use of climate anomaly data as one of the additional features to predict pest attacks can improve the accuracy of the model. The results of this study contributes to the development of feature selection techniques, which are an important factor in improving the performance of prediction models. The rest of the paper is organized as follows. The first section is the introduction and background of the research. The second section is the research method. The third section is the results and discussion, and the last section is the conclusion and future work.

2. Methodology.

2.1. Proposed method. This research proposes a model for predicting rice pest attacks by considering the possibility of anomaly in the data. The proposed model can be seen in Figure 1. The dataset used in this study included climate data and pest attacks. The climate data included temperature, humidity, and rainfall within 15 year period, from 2002 to 2017. The climate data were obtained from the Weather Observation Station of

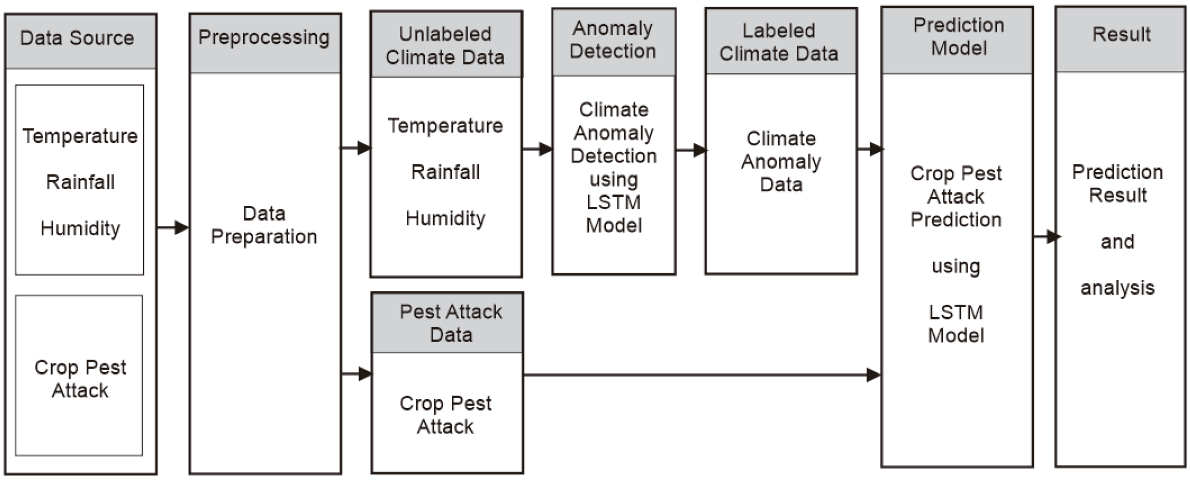

Figure 1. Proposed model: Pest prediction model using climate anomaly data

Indonesian Air Force in Adi Sumarmo Solo. Meanwhile, the pest attack data used in this study were data of rice stemborer species' attacks taken from pest population movement monitoring in Boyolali Regency.

The research activity started with data preparation. Data preparation activities were handling missing values and normalizing data. The missing data were replaced by using a median method, which was finding the median value and replacing all the missing values with the median value. The next step was detecting the climate anomaly, and then the anomaly data were used to predict the pest attack using LSTM model. The anomaly detection would give labels on the climate data, whether or not they have anomaly. After that, the labeled climate data would be used to predict pest attacks.

To measure the performance of the prediction model, Mean Square Error (MSE) and Root Mean Square Error (RMSE) were used. MSE was used to measure the consistency of the model with square differences between actual data and predicted data (Equation (1)).

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(X_{obs,i} - X_{pred,i})^2, \tag{1}$$

where n is number of data, X_{obs} is observed values and X_{pred} is the predicted values.

RMSE was used to measure the accuracy level of the model's prediction results. The value of the errors generated by the prediction model could be seen from the RMSE value (Equation (2)).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{pred,i})^2}{n}}, \tag{2}$$

where n is number of data, X_{obs} is observed values and X_{pred} is the predicted values. The smaller the MSE and RMSE values are, the smaller the distance between predictions and reality is. This means that prediction performance is also getting better.

2.2. **LSTM model.** The architecture of LSTM model is portrayed in Figure 2, where an LSTM network has three gates, called input gate, forget gate, and output gate [11,14]. Forget gate will determine the information that should be thrown away from a cell state, to diminish dependency on long-term memory.

The first equation is the formula for the forget gate (f).

$$\vec{f_t} = \sigma\left(\vec{x}_t W_{xf} + \vec{h}_{t-1} W_{hf} + \vec{c}_{t-1} W_{cf} + \vec{b}_f\right), \tag{3}$$

where x is the input units, t is discrete time ($t = 1, 2, \ldots, n$), σ is the logistic sigmoid activation function, W is the weights, c is memory cell units, h is the hidden units and b

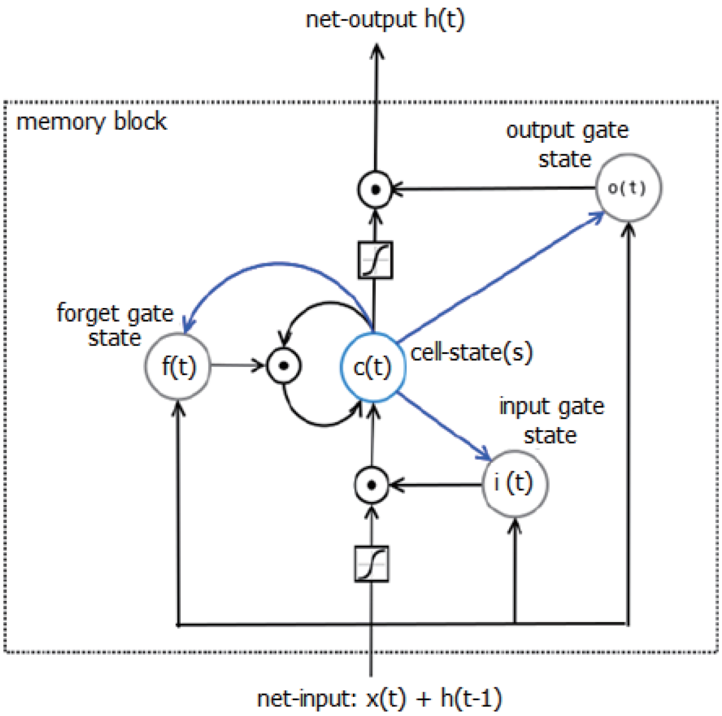

FIGURE 2. LSTM architecture [14]

is the bias. The new information that will be stored in the cell state will be determined by the input gate (i) with the formula as shown in the equation below.

$$\vec{i}_t = \sigma\left(\vec{x}_t W_{xi} + \vec{h}_{t-1} W_{hi} + \vec{c}_{t-1} W_{ci} + \vec{b}_i\right). \tag{4}$$

Further, cell units (c) will be upgraded to become new cell states with the formula shown in equation below.

$$\vec{c}_t = \vec{f}_t \circ \vec{c}_{t-1} + \vec{i}_t \circ \tanh\left(\vec{x}_t W_{xc} + \vec{h}_{t-1} W_{hc} + \vec{b}_c\right), \tag{5}$$

where ($\circ$) denotes an element-wise multiplication.

The parts of the produced cell states will be decided by the output gate. Below is the equation of the output gate (o).

$$\vec{o}_t = \sigma\left(\vec{x}_t W_{xo} + \vec{h}_{t-1} W_{ho} + \vec{c}_t W_{co} + \vec{b}_o\right). \tag{6}$$

The output value of the cell, which is the hidden activation is derived from the calculation of two terms, shown as follows:

$$\vec{h}_t = \vec{o}_t \circ \tanh\left(\vec{c}_t\right). \tag{7}$$

Therefore, LSTM has similar working principles, where the size output of LSTM is the length of sequential data times the number of declared hidden neuron.

3. Result and Discussion.

3.1. **Climate anomaly detection.** Climate anomaly was detected by first developing a Deep-LSTM method, a stack of several LSTM layers as a deep learning model, that has been proven to improve the accuracy of the model in previous research [15-17]. Deep LSTM was then used to extract hierarchical patterns in time series data through a training process. The extracted pattern was used to detect anomaly. The deviant data, which were different from the general pattern, would then be marked or labeled as anomaly data. The trained LSTM model was used to find anomaly data from each climate attribute. In the initial phase, the anomaly detection was used in the temperature attribute.

Anomaly in the time series data was detected by measuring the Root Mean Square Deviation (RMSD) value, which compared prediction data and actual observation data. Data anomaly happened when the deviation value was high. In other words, there was a high deviation between the prediction data and the observation data. Many works show that RMSD was often used in adaptive thresholding model to detect anomaly in time series data [18]. In this research, the data were considered as anomaly if the RMSD value was more than 2.

Figure 3 shows the values that are considered as anomaly data. The anomaly dots or points happened if there were high deviations between the data generalization of the model and the actual observation data or the RMSE value was more than 2 as adaptive threshold. After the above model was gained, the next step was labeling every dot whether or not it was anomaly.

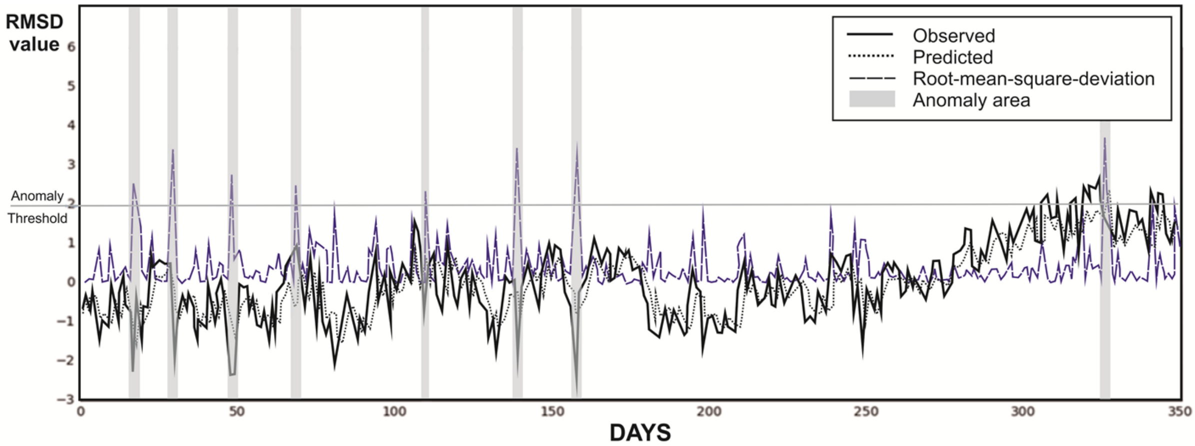

FIGURE 3. RMSD graph for the model

3.2. **Pest prediction model using climate anomaly data.** Pest prediction model proposed in this study was designed by considering data anomaly occurrences in the climate data to be used as a pest attack prediction variable. Thus, the data used as an input in this model were the number of weather anomaly occurrences during the 2-week periods, which was obtained from anomaly detection process in the previous phase. In this experiment, the anomaly data used were the number of anomalies in the temperature data. The temperature data were chosen because the data had a stable movement without high fluctuation, compared to other weather data.

The next phase was to define and fit the data to the LSTM model. The LSTM model used in this study had 50 neurons in the first hidden layer and 1 neuron in the output layer to predict future pest attacks. After several trials, the best model could be attained that had the condition of 50 training periods (epoch) with the batch value of 72. In this LSTM model, densely connected layer was added to reduce a network complexity and increase the accuracy. Mean Absolute Error (MAE) loss function was also applied. Adam version optimizer was used to reduce stochastic gradient. The training process produced a loss graphic as shown in Figure 4.

To measure the performance of the proposed model, a comparison was made with other models. Thus, there are two models compared. The first model was pest attack prediction by using climate data which have anomaly labels, as was done in previous experiments in this study. The second model was a direct prediction of pest attack by using climate data without anomaly labeling. To improve the prediction performance, the epoch value of each model was adjusted. Several studies have shown that choosing the right epoch value can improve the performance of the model.

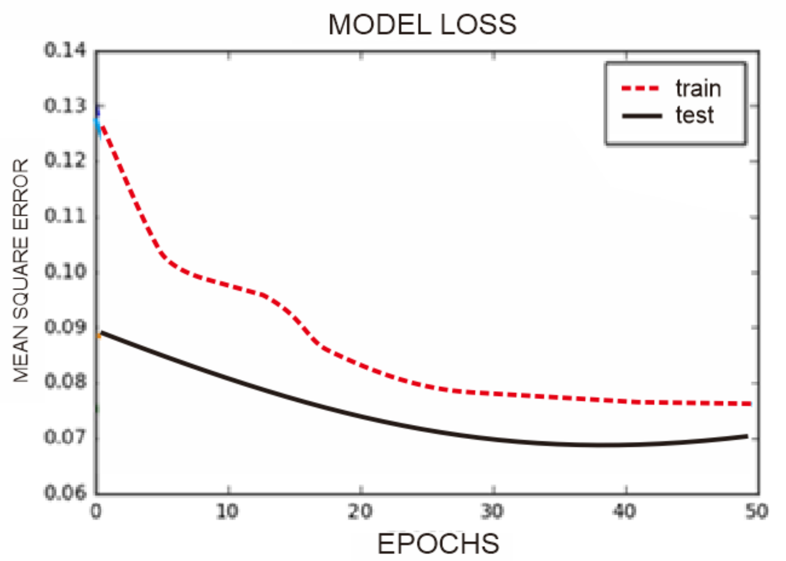

FIGURE 4. Model loss of the proposed model

The summary of this experiment is presented in Table 2. From the table, it can be seen that from several experiments conducted, model 1 has a smaller MSE and RMSE values compared to model 2. This means that pest prediction models using climate anomaly data have better performance compared to pest prediction models using climate data without anomaly labeling. It can be argued that by using the climate anomaly data, we can increase the accuracy of the prediction of pest attacks. The conclusion aligns with some previous studies, that we should pay more attention to climate anomaly since it has a strong influence toward various other occurrences [12,13,19].

TABLE 2. Experiment result resume

Model	**Epoch**	**Training**		**Testing**		**Result**
		MSE	**RMSE**	**MSE**	**RMSE**	
1st model	10	0.0971	0.3125	0.0740	0.2722	
	50	0.0926	0.3044	0.0698	0.2643	
	100	0.0909	0.3015	0.0698	0.2606	Best performance
2nd model	10	0.0977	0.3116	0.0773	0.2780	
	50	0.0934	0.3056	0.0742	0.2724	
	100	0.0924	0.3041	0.0717	0.2679	

After several experiments, it turns out that the best performance was obtained by the proposed model with 100 epochs. The accuracy values produced from the training process of the model were 0.0909 for MSE and 0.3015 for RMSE. The accuracy values produced from the testing process were 0.0698 for MSE and 0.2606 for the RMSE. These results are the best performance compared to other models and settings.

4. **Conclusions.** As an agricultural country, Indonesia should pay attention to climate anomaly occurrences and pest attacks since they have big influences toward its agricultural productivity. This study recommends a prediction model of pest attacks by considering climate anomaly as the input variable. Focusing on anomaly occurrences is important as the climate variable has a strong seasonality component in nature, which has similar characteristics for longer data and also a complex structure. The anomaly occurrences should also be considered since they have a strong influence toward some other phenomena. The

anomaly detection was done by using Deep-LSTM method and the pest attack prediction model was built by considering weather anomaly data that have been detected. The results show that pest prediction models using climate anomaly data have better performance compared to pest prediction models using climate data without anomaly labeling. Further studies can be done by implementing various methods to observe the effectivity of climate anomaly data in developing the pest attacks prediction model. Moreover, additional data set is needed to improve the model performance.

Acknowledgment. The authors would like to thank The Center for Food Crop and Horticulture Protection of Central Java Province that has given the permission to access the data of pest attack in Boyolali Regency of Central Java Province, Indonesia.

REFERENCES

[1] S. Y. J. Prasetyo, S. Subanar, E. Winarko and B. S. Daryono, ESSA: Exponential smoothing and spatial autocorrelation, methods for prediction of outbreaks pest in Indonesia, *International Review Computer and Software*, vol.10, no.4, pp.362-375, 2014.
[2] F. F. Rachman, *63000 Hectares of Rice Fields Were Attacked by Planthoppers, Detik Finance*, https://finance.detik.com/berita-ekonomi-bisnis/d-3627569, 2018.
[3] M. S. Rizaludin, *Prevention First, before a Pest Outbreak: Time to Adopt Agroecological Farming*, The Jakarta Post, https://www.thejakartapost.com/academia/opinion, 2020.
[4] Ministry of Agriculture, *Crop Pest Attack Report until October 2019*, Jakarta, Republic of Indonesia, 2019.
[5] KOMPAS, *Brown Plant Hopper Attacks*, Kompas Newspaper, https://kompas.id/baca/ekonomi/2017/07/21/serangan-wereng-cukup-parah/, 2017.
[6] R. Mouly, T. N. Shivananda and A. Verghese, Weather based forecasting models for prediction of leafhopper population *Idioscopus nitidulus* Walker; (Hemiptera: Cicadellidae) in mango orchard, *Journal of Entomology and Zoology Studies*, vol.5, no.1, pp.163-168, 2017.
[7] A. K. Tripathy et al., Data mining and wireless sensor network for agriculture pest/disease predictions, *World Congress on Information and Communication Technologies*, pp.1229-1234, 2011.
[8] C. Ma, Y. Liang and X. Lyu, Weather analysis to predict rice pest using neural network and D-S evidential theory, *Proc. of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pp.277-283, 2019.
[9] S. Jain and D. Ramesh, Machine learning convergence for weather based crop selection, *Proc. of IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp.1-6, 2020.
[10] Q. Xiao, W. Li, Y. Kai et al., Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network, *BMC Bioinformatics*, 2019.
[11] T. Wahyono, Y. Heryadi, H. Soeparno and B. S. Abbas, Enhanced LSTM multivariate time series forecasting for crop pest attack prediction, *ICIC Express Letters*, vol.14, no.10, pp.943-949, 2020.
[12] J. Kawale et al., *Anomaly Construction in Climate Data: Issues and Challenges*, Technical Report, Department of Computer Science, University of Minnesota, 2011.
[13] A. Singh, *Anomaly Detection for Temporal Data Using Long Short-Term Memory (LSTM)*, Master Thesis, KTH Information and Communication Technology, Stockholm, Sweden, 2017.
[14] C. Herta, *LSTM (Long Short Term Memory), Lecture Notes and Teaching Material*, http://www.christianherta.de/lehre/dataScience/machineLearning/neuralNetworks/LSTM.php, 2019.
[15] Y. Omae, M. Mori, T. Akiduki and H. Takahashi, A novel deep learning optimization algorithm for human motions anomaly detection, *International Journal of Innovative Computing, Information and Control*, vol.15, no.1, pp.199-208, 2019.
[16] J. Brownlee, Long short-term memory networks with python: Develop sequence prediction models with deep learning, *Machine Learning Mastery*, 2017.
[17] Hendry, R.-C. Chen, L.-H. Li and Q. Zhao, Using deep learning to learn user rating from user comments, *International Journal of Innovative Computing, Information and Control*, vol.14, no.3, pp.1141-1149, 2018.
[18] C. Chio and D. Freeman, *Machine Learning and Security*, O'Reilly Media, Inc., Gravenstein Highway North, 2017.
[19] M. P. Tingley, A Bayesian ANOVA scheme for calculating climate anomalies with applications to the instrumental temperature record, *J. Climate*, vol.25, pp.777-791, 2012.