What Meteorological Characteristics Do Affect Weather Forecasting? *

Hyeon-Ju Jeon **[0000-0002-2400-8360], Jeon-Ho Kang, In-Hyuk Kwon

Data Assimilation Group, Korea Institute of Atmospheric Prediction Systems (KIAPS), 35, Boramae-ro 5-gil, Dongjak-gu, Seoul 07059, Korea
{hjjeon,jhkang,ihkwon}@kiaps.org

Abstract. Earth system consists of complex physics-based models to cover the various physical properties and stochastical patterns. In order to overcome the high complexity of existing methods, several studies attempt to predict the weather based on a large amount of observation data. However, they only focus on improving the prediction performance using deep learning models, and few studies analyze various meteorological characteristics. In this paper, we analyze what meteorological characteristics should be reflected in deep learning techniques to understand atmospheric phenomena. To predict weather conditions, we consider the following characteristics: (i) spatial features, (ii) temporal features, and (iii) the diversity of scale and shape of meteorological phenomena. The convolutional neural network is extended for extracting spatial patterns in the three-dimensional atmospheric fields, and a convolutional attention module is adopted to reflect the distinguished form of each atmospheric phenomenon. Then, dynamic patterns are extracted by the recurrent neural network-based time series model. We evaluated the performance of the proposed model with existing models by predicting the forecast fields. The experimental results showed that the synergy of various meteorological properties has a positive effect on weather prediction.

Keywords: Weather Forecast · Meteorological Feature Analysis · Spatio-temporal Prediction · Convolutional Attention module

1 Introduction

In modern society, weather forecasting is inseparable from the planning and management strategies of various resources in the retail markets, agriculture,

 In: N. D. Vo, O.-J. Lee, H. G. Lim, H.-J. Jeon, P.-M. Nguyen, B. Q. Tuyen, J. J. Jung, T. T. Le (eds.): Proceedings of the 3rd International Conference on Human-centered Artificial Intelligence (Computing4Human 2022), Hanoi, Viet Nam, 16 to 17-December-2022, published at http://ceur-ws.org

** Corresponding author.

aviation, and other industrial fields [19]. Weather forecasting is also the basis for dealing with issues related to smart cities and intelligent transportation [17] such as traffic flow prediction, medical treatment, intelligent buildings, and so on. The primary method for weather prediction is numerical-based systems (i.e., Numerical weather prediction), which solve large systems consisting of nonlinear mathematical equations based on specific physical models [7]. As enhancing technical expertise, the range and quality of observation data (e.g., weather satellite, aircraft data, ground-based global navigation satellite systems (GNSS), smart Seoul data of things (S-DoT), and so on) available to use are increasing, which contributes to improved forecasting performance [12].

Therefore, there are various studies not only for improving prediction accuracy by assimilating new observation data into the numerical-based systems [15] but also for applying deep learning methods [1, 8] that effectively predict using large amounts of historical data are conducted. The conventional prediction system still have difficulty with error caused by information loss in the process of generating initial conditions, and an incomplete understanding of atmospheric physical processes [2]. The novel data-driven approaches overcome the limitations of the conventional prediction systems which require highly complex physics-based models since various physical properties and the stochastic patterns of the Earth system need to be reflected [13].

This study aims to analyze what meteorological characteristics should be reflected in deep learning techniques to understand atmospheric phenomena. We propose a prediction model considering (i) spatial features, (ii) temporal features, and (iii) the diversity of scale and shape of meteorological phenomena. In recent years, a convolutional neural network (CNN) [10] captures complex non-linear relationships of neighboring regions based on a large amount of atmospheric spatial datasets. However, the existing two-dimensional CNN model cannot consider the spatial distribution in the vertical direction. Although Kamangir et al. [11] introduces a three-dimensional CNN, an additional dimension is only used to consider temporal dependency with neighboring regions. Therefore, in this study, three-dimensional CNN is applied to reflect both vertical and horizontal spatial distribution, and the correlation between meteorological variables in neighboring regions is also reflected by extending the channel. Then, we extract dynamic patterns of the spatial distribution with multiple variables by the time series model. In addition, although various deep learning backbones such as VGG network [9] and U-net [20] have been applied to the prediction model and showed high accuracy, various experiments that could analyze meteorological characteristics were not performed [18]. Therefore, we expand the model by applying the attention module [23] to reflect multi-scale meteorological phenomena with distinct forms, and conduct experiments that can analyze which meteorological characteristics are important for weather forecasting. This result can be used in various tasks such as observation evaluation or preprocessing in the future.

2 Methodology

In this study, we propose a novel weather forecasting model that considers (1) spatial features, (2) temporal features, and (3) the variety of scales and shapes of atmospheric phenomena. First, we prepare the data set from the NWP data in three-dimensional form. Then, we modify and extend the conventional CNN and GRU models to analyze the various meteorological characteristics.

2.1 Data

The numerical-based prediction model defines a coordinate system that divides the earth into three-dimensional grids [22]. Meteorological variables are measured within each grid, and atmospheric characteristics are predicted based on interactions with adjacent grids. Here, the data assimilation technique [16] is used to quantify the uncertainty of observations. In other words, the technique allows us to explain the changes in atmospheric characteristics without large amounts of historical data, and prediction models can obtain higher accuracy using assimilated information, called analysis field, than using observed data directly. Also, various types of meteorological variables collected from different observed species are assimilated into four variables (i.e., u-component of wind (U), v-component of wind (V), temperature (T), and relative humidity (R)) through the data assimilation process. Thus, data assimilation systems are dependent on prediction models. In this study, we analyzed which meteorological features are important to evaluate the effect of the analysis field of the KIM variational data assimilation system (KVAR) [14] on the Korean Integrated Model (KIM) [7].

2.2 Problem Statement

The prediction model needs to consider extracting the spatial and temporal meteorological features at the same time. Thus, the assimilated data in a specific geographic region over time is used as the input sequence for the prediction task. The proposed model can be defined as:

$$Y_{t+1}^{(i)} = f(X_t^{(i)}, \cdots, X_{t-n}^{(i)}), X_t^{(i)} \in \mathbf{R}^{V \times D \times W \times H} \tag{1}$$

where $Y_{t+1}^{(i)}$ indicates the target forecasts in time $t+1$ at i^{th} geographic region, and $f(\cdot)$ refers to the proposed model which extracts the meteorological features from historical data. $X_t^{(i)}$ indicates the assimilated data composed of meteorological variables $V = \{v_1, \cdots, v_j\}$ at i^{th} geographic region in time t. Each assimilated data $X_{t,v_j}^{(i)} \in \mathbf{R}^{L \times W \times H}$ for $j = 1, 2, \cdots, J$ (i.e. the number of meteorological variables) represented by a tensor, which consists of a $D \times W \times H$ grid map, where D, W, and H indicate the level, latitude, and longitude of the analysis fields, respectively. The sequence length of the assimilated data is represented by n.

2.3 Architecture of the Prediction Model

Considering Multi-scale Spatial Features The 3D-CNN part of our prediction model extracts spatial features by reflecting the interaction between meteorological variables V in the adjacent region. The 3D-CNN extends the convolution kernel to three dimensions to perform convolution operations on the earth grid map. The operation of 3D-CNN can be expressed as follows:

$$C_l^{k,x,y,z} = \partial\left(\sum_{v=0}^{V_l-1}\sum_{d=0}^{D_l-1}\sum_{w=0}^{W_l-1}\sum_{h=0}^{H_l-1} W_{l,k}^{vdwh} C_{l-1}^{v(x+d)(y+w)(z+h)} + b_l^k\right) \tag{2}$$

where $W_{l,j}^{vdwh}$ indicates the value of the cell at position (d, w, h) of v^{th} variables in the k^{th} kernel connected to the l^{th} layer. $C_l^{k,x,y,z}$ represents the value of the cell in the v^{th} variables at position (d, w, h) in the $l-1^{th}$ layer. D_l,W_l, and H_l denote the level, latitude, and longitude of the three-dimensional kernel. The activation function is ∂, and b is a bias parameter.

One three-dimensional kernel only captures a local feature, and multiple layers with multiple three-dimensional kernels can be used to extract global features in the grid map. We used a three-dimension convolutional network with $3\times3\times3$ kernel which shows the best results considering the consumption of parameters and time according to the existing study [4].

The meteorological data has three-dimensional patterns with different spatial resolutions which can be captured by using different sizes of kernels and receptive fields. Dilated convolutional layer using an expanded receptive field extracts multiscale contextual information without loss of resolution or coverage [9]. Therefore, we propose a multiscale 3D dilated convolutional module to reflect more complicated spatial features of meteorological data such as various scales of atmospheric phenomena.

Considering Various Shapes of the Atmospheric Phenomena Each atmospheric phenomenon has its own distinguishing shape. For example, the rain front has an elongated shape on both sides, and the typhoon has a round shape with a huge volume. To consider the various shapes of the phenomena, we apply the attention module that extracts detailed and wide spatial features even for regions that exceed the size of the convolutional filter. In this study, we propose an attentive 3D-convolutional neural network that extracts meaningful regions with important meteorological variables. The proposed attentive 3D-CNN consists of two parts; variable attention module and spatial attention module.

The variable attention module generates the feature vectors which are reflected in the inter-variable relationship in meteorological characteristics. The module can effectively learn meaningful variables in each grid since the feature vector is the result of training with more heavy weights on the important variables. The variable attention is computed as:

$$A_v(C) = \sigma(g(AvgPool(C)) + g(MaxPool(C))), \tag{3}$$

where $C \in \mathbb{R}^{k,x,y,z}$ is the intermediate feature vector as input from the previous CNN modules, σ denotes the ReLU activation function, and $g(\cdot)$ represents three dense layers.

And then, we produce a spatial attention vector by reflecting the interactions with each grid map. Here, we use both the average and max pooling layers simultaneously to extract the distinguishable spatial information of the meteorological features. Different from the variable attention, the spatial attention module concentrates on the informative region which is complementary to the variable attention. To focus on the informative regions we apply the pooling layer along the variable axis. And then, the convolutional layer is used to produce a spatial feature vector that compresses where to highlight. Thus, spatial attention is formulated as:

$$A_s(C) = \sigma(f([AvgPool(C)); MaxPool(C)]), \tag{4}$$

where σ represents the ReLU activation function and $f(\cdot)$ indicates a convolutional layer.

According to input variables in a grid map, two attention modules aggregate complementary attention, focusing on the meaningful variables and regions, respectively. In the previous study [23], a sequential arrangement of the two modules achieved a better performance than a parallel arrangement. In addition, the variable attention module before the spatial attention module has shown slightly better performance than the reverse order. We arranged the sequence of our proposed model following the previous study.

Considering Temporal Features The recurrent neural network (RNN) is widely used to extract temporal features in speech recognition and action recognition tasks. Most of the weather prediction tasks (e.g., precipitation, wind power, solar irradiance) which can be defined as time series problems have adopted various RNN models (i.e., LSTM or GRU) to avoid the long-term dependency problem [3,21]. Although a structure such as FC-LSTM, ConvLSTM, and CRNN in which the output of a fully connected layer or a convolutional layer is fed into an RNN has been developed to handle spatiotemporal data [5], it has difficulty in considering the spatial data in the vertical direction. In addition, the LSTM model has more parameters than the GRU model [6]. In this paper, an attentive three-dimensional convolutional GRU that is improved for time series learning of three-dimensional space is introduced to capture temporal information in our model. The GRU layers consist of the reset and update gates. The reset gate r_t at time t is applied to control the amount of temporal information of the previous time periods that will be forgotten or remembered. Likewise, the update gates u_t at time t are used to control how much past information should be reserved. The proposed three-dimensional convolutional GRU can be formulated as:

$$r_t = \sigma(\theta_r[C_t^{(N)}, h_{t-1}]), \tag{5}$$

$$u_t = \sigma(\theta_u[C_t^{(N)}, h_{t-1}]), \tag{6}$$

$$c_t = \phi(\theta_c[C_t^{(N)}, (r_t * h_{t-1})]), \tag{7}$$

$$h_t = u_t * h_{t-1} + (1 - u_t) * c_t, \tag{8}$$

where $\sigma(\cdot)$ refers to a sigmoid function. $\phi(\cdots)$ indicates the hyperbolic tangent function. $*$ denotes the Hadamard product. $C_t^{(N)}$ is a spatial feature matrix captured by N_t using the last convolutional layer. θ_r, θ_u, and θ_c are learnable weight matrices. After the GRU layer, one fully connected layer is used to predict the future atmospheric states from h_t with linear activation.

Attentive 3D Convolutional GRU Model The attentive 3DCNN extracts the spatial features of the meteorological variables related to the assimilated data. Then, the GRU captures the progressive relationship of the e atmospheric motion. The two modules are combined to form a 3DCGRU model for weather forecasting, where the features extracted by the 3DCNN, the position of the atmospheric phenomena, and the interaction of the multiple variables are concatenated as the input of the GRU model. As the result, the 3DCGRU model can understand not only the horizontal advection but also the vertical convection of the atmospheric motion.

In general, the proposed method learns spatiotemporal features from historical data to achieve more accurate predicted weather in time t while maintaining the performance of each model. The 3DCGRU model computes and back-propagates the loss to update the weights and biases in the training process. Therefore, the loss function is formulated as:

$$L = \sum_{i,t} \left(Y_{t+1}^{(i)} - \widehat{Y_{t+1}^{(i)}} \right)^2, \tag{9}$$

where, $Y_{t+1}^{(i)}$ represents the real forecast states at i^{th} geographical region in time $t+1$, and $\widehat{Y_{t+1}^{(i)}}$ refers to the predicted forecast states at i^{th} geographical region in time $t+1$.

3 Evaluation

3.1 Experimental setup

In this study, we used four accuracy metrics to evaluate the performance of the proposed models: root mean square error (MSE), mean absolute error (MAE), R-squared (R^2), and variance (σ). These metrics can be measured as follows:

$$MSE = \left[\frac{1}{T+I} \sum_{\forall t,i} \left(Y_{t+1}^{(i)} - c \right)^2 \right]^{\frac{1}{2}}, \ MAE = \frac{1}{T+I} \sum_{\forall t,i} \left| Y_{t+1}^{(i)} - \widehat{Y_{t+1}^{(i)}} \right|,$$
$$R^2 = 1 - \frac{\sum_{\forall t,i} \left(Y_{t+1}^{(i)} - \widehat{Y_{t+1}^{(i)}} \right)^2}{\sum_{\forall t,i} \left(Y_{t+1}^{(i)} - \bar{Y} \right)^2}, \ \sigma = 1 - \frac{\sigma \left(Y_{t+1}^{(i)} - \widehat{Y_{t+1}^{(i)}} \right)}{\sigma \left(Y_{t+1}^{(i)} \right)} \tag{10}$$

where $Y_{t+1}^{(i)}$ and $\widehat{Y_{t+1}^{(i)}}$ indicate the true forecast states and predicted forecast states at i^{th} region in time t, respectively. T and I denote the total number of time points and grid point, respectively. $\bar{Y}$ denotes the average $Y_{t+1}^{(i)}$ for $\forall t \in [0, T]$, $\forall i \in [0, I]$, and $\sigma(\cdot)$ indicates variance.

3.2 Experimental results

We evaluate the proposed model by comparing its accuracy with the baseline methods. The models were trained to predict the forecast field at time t by analyzing the previous analysis field from $t-1$ to $t-8$. We evaluated the performance of the proposed and existing deep learning models. Table 1 shows the experimental results.

Table 1: A performance comparison of the proposed model with the baseline models. The bold means models with the best performance on each metric.

	MAE	MSE	R^2	σ
GRU	0.03	0.02	0.94	0.95
2DCGRU	0.02	0.02	0.96	0.96
3DCGRU	0.02	0.01	0.97	0.97
3DCGRU (w/att.)	**0.01**	**0.01**	**0.99**	**0.99**

In case of GRU which consider only the time series of meteorological variables excluding spatial features showed the lowest accuracy. If we provide only the time sequence of variables, the model cannot obtain enough information for understanding weather conditions. Comparing the performance of 3DCGRU and 2DCGUR, 3DCGRU which the model learns the vertical structure of the atmosphere showed better results. Therefore, we can assume that vertical convection can be an informative feature for the proposed model. The proposed prediction model outperformed the other baseline models in every evaluation metric. Thus, the high performance of the proposed model supports that the model could understand the meteorological properties by overcoming the variety scale and shapes the atmospheric phenomena have.

4 Conclusion and future work

In this paper, we have proposed a weather prediction model with multiple meteorological variables. We have defined the atmospheric field by the three-dimensional grid. By extending the convolutional filter to three-dimension, the spatial distribution was captured by considering not only horizontal advection but also vertical convection. The variable and spatial attention module are adopted to extract the relationships with multiple variables in the neighboring grids. The experimental results show that the proposed model can reflect various features of atmospheric phenomena and synergy of various meteorological properties has a positive effect. However, it is difficult to determine the importance of various features. Future research will focus on the systematic designing of our experiments with other conventional prediction methods. Also, we extend our study to estimate the how sensitive forecast field changes as the analysis field changes.

Acknowledgements This work was supported by the R&D project "Development of a Next-Generation Data Assimilation System by the Korea Institute of Atmospheric Prediction System (KIAPS)," funded by the Korea Meteorological Administration (KMA2020-02211)

References

1. An, S., Na, I., Oh, T.J., Kim, J.: Neural network emulator optimizer: A preliminary study on korean microphysicis parameterization model. In: Vo, N.D., Lee, O.J., Bui, K.H.N., Lim, H.G., Jeon, H.J., Nguyen, P.M., Kim, J.T., Tuyen, B.Q., Jung, J.J., Vo, T.A. (eds.) In proceedings of the 2nd International Conference on Human-centered Artificial Intelligence (Computing4Human 2021). CEUR Workshop Proceedings, vol. 3026, pp. 110–119. Da Nang, Viet Nam (Oct 2021)
2. Buehner, M., Du, P., Bédard, J.: A new approach for estimating the observation impact in ensemble–variational data assimilation. Monthly Weather Review **146**(2), 447–465 (jan 2018). https://doi.org/10.1175/mwr-d-17-0252.1
3. Castro, R., Souto, Y.M., Ogasawara, E., Porto, F., Bezerra, E.: STConvS2s: Spatiotemporal convolutional sequence to sequence network for weather forecasting. Neurocomputing **426**, 285–298 (feb 2021). https://doi.org/10.1016/j.neucom.2020.09.060
4. Cheng, W., Wang, Y., Peng, Z., Ren, X., Shuai, Y., Zang, S., Liu, H., Cheng, H., Wu, J.: High-efficiency chaotic time series prediction based on time convolution neural network. Chaos, Solitonsv & Fractals **152**, 111304 (nov 2021). https://doi.org/10.1016/j.chaos.2021.111304
5. Fu, K., Li, H., Deng, P.: Chaotic time series prediction using DTIGNet based on improved temporal-inception and GRU. Chaos, Solitons & Fractals **159**, 112183 (jun 2022). https://doi.org/10.1016/j.chaos.2022.112183
6. Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: A search space odyssey. IEEE Transactions on Neural Networks and Learning Systems **28**(10), 2222–2232 (oct 2017). https://doi.org/10.1109/tnnls.2016.2582924

7. Hong, S.Y., Kwon, Y.C., Kim, T.H., Kim, J.E.E., Choi, S.J., Kwon, I.H., Kim, J., Lee, E.H., Park, R.S., Kim, D.I.: The korean integrated model (KIM) system for global weather forecasting. Asia-Pacific Journal of Atmospheric Sciences **54**(S1), 267–292 (jun 2018). https://doi.org/10.1007/s13143-018-0028-9
8. Jeon, H.J., Choi, M.W., Lee, O.J.: Day-ahead hourly solar irradiance forecasting based on multi-attributed spatio-temporal graph convolutional network. Sensors **22**(19), 7179 (sep 2022). https://doi.org/10.3390/s22197179
9. Jeon, H.J., Lim, H.G., Shung, K.K., Lee, O.J., Kim, M.G.: Automated cell-type classification combining dilated convolutional neural networks with label-free acoustic sensing. Scientific Reports **12**(1), 19873 (nov 2022). https://doi.org/10.1038/s41598-022-22075-6
10. Ju, Y., Sun, G., Chen, Q., Zhang, M., Zhu, H., Rehman, M.U.: A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. IEEE Access **7**, 28309–28318 (Mar 2019). https://doi.org/10.1109/access.2019.2901920
11. Kamangir, H., Collins, W., Tissot, P., King, S.A., Dinh, H.T.H., Durham, N., Rizzo, J.: FogNet: A multiscale 3d CNN with double-branch dense block and attention mechanism for fog prediction. Machine Learning with Applications **5**, 100038 (sep 2021). https://doi.org/10.1016/j.mlwa.2021.100038
12. Kang, J.H., Chun, H.W., Lee, S., Ha, J.H., Song, H.J., Kwon, I.H., Han, H.J., Jeong, H., Kwon, H.N., Kim, T.H.: Development of an observation processing package for data assimilation in KIAPS. Asia-Pacific Journal of Atmospheric Sciences **54**(S1), 303–318 (jun 2018). https://doi.org/10.1007/s13143-018-0030-2
13. Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V.: Machine learning for the geosciences: Challenges and opportunities. IEEE Transactions on Knowledge and Data Engineering **31**(8), 1544–1554 (aug 2019). https://doi.org/10.1109/tkde.2018.2861006
14. Kwon, I.H., Song, H.J., Ha, J.H., Chun, H.W., Kang, J.H., Lee, S., Lim, S., Jo, Y., Han, H.J., Jeong, H., Kwon, H.N., Shin, S., Kim, T.H.: Development of an operational hybrid data assimilation system at KIAPS. Asia-Pacific Journal of Atmospheric Sciences **54**(S1), 319–335 (jun 2018). https://doi.org/10.1007/s13143-018-0029-8
15. Lee, S., Song, H.J., Kwon, I.H., Kang, J.H.: Impacts of aeolus horizontal line-of-sight (HLOS) wind assimilation on the korean integrated model (KIM) forecast system. Atmospheric Science Letters (oct 2022). https://doi.org/10.1002/asl.1138
16. Mack, J., Arcucci, R., Molina-Solana, M., Guo, Y.K.: Attention-based convolutional autoencoders for 3d-variational data assimilation. Computer Methods in Applied Mechanics and Engineering **372**, 113291 (dec 2020). https://doi.org/10.1016/j.cma.2020.113291
17. Neirotti, P., Marco, A.D., Cagliano, A.C., Mangano, G., Scorrano, F.: Current trends in smart city initiatives: Some stylised facts. Cities **38**, 25–36 (jun 2014). https://doi.org/10.1016/j.cities.2013.12.010
18. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat: Deep learning and process understanding for data-driven earth system science. Nature **566**(7743), 195–204 (feb 2019). https://doi.org/10.1038/s41586-019-0912-1
19. Štulec, I., Petljak, K., Naletina, D.: Weather impact on retail sales: How can weather derivatives help with adverse weather deviations? Journal of Retailing and Consumer Services **49**, 1–10 (jul 2019). https://doi.org/10.1016/j.jretconser.2019.02.025

20. Trebing, K., Stanczyk, T., Mehrkanoon, S.: SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. Pattern Recognition Letters **145**, 178–186 (May 2021). https://doi.org/10.1016/j.patrec.2021.01.036
21. Wang, P., Wang, P., Wang, C., Xue, B., Wang, D.: Using a 3d convolutional neural network and gated recurrent unit for tropical cyclone track forecasting. Atmospheric Research **269**, 106053 (may 2022). https://doi.org/10.1016/j.atmosres.2022.106053
22. Weyn, J.A., Durran, D.R., Caruana, R.: Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. Journal of Advances in Modeling Earth Systems **12**(9) (sep 2020). https://doi.org/10.1029/2020ms002109
23. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: Convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) Proceedings of the 15th European Conference Computer Vision (ECCV 2018). Lecture Notes in Computer Science, vol. 11211, pp. 3–19. Springer International Publishing, Munich, Germany (Sep 2018)