IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10282636

Physical Analysis of the Impact of Polarization Parameters on Deep-learning Networks for Nowcasting

Kexin Zhu[1], *Haonan Chen*[1], and *Lei Han*[2]

[1]*Colorado State University, Fort Collins, CO*
[2] *College of Information Science and Engineering, Ocean University of China, Qingdao, China*

***Abstract*—The task of nowcasting by deep learning using multivariate, rather than just reflectivity, is limited by poor interpretability. The previous experiment designed MCT (Multivariate Channel Transformer), a deep learning model capable of nowcasting with dual-polarization radar data. Four analytical methods are designed to further explore the contribution of polarization parameters: (i) Case studies of different meteorological processes. (ii) A permutation test ranking the significance of each variable. (iii) Visualization of the feature maps obtained by forward propagation of the input data. (iv) Data downscaling of polarimetric radar data. The results show that the polarization parameters serve as a guide to predict the location and shape of strong reflectivity, as well as the energy retention of strong echoes at 40-50 dBZ. The contributions of Z_{dr} and K_{dp} are more evident in the prediction results after 30 min, and the importance of K_{dp} exceeds that of Z_{dr} in case of strong convective weather.**

***Index Terms*—polarization parameter, nowcasting, permutation test, feature map, data downscaling.**

I. Introduction

The use of deep learning as a method for nowcasting is gaining traction [1] [2] [3], yet the end-to-end opacity of deep learning itself limits its reliability. Dual polarimetric radar has been widely used in meteorological tasks such as quantitative precipitation estimation and water condensate phase identification [4] [5] [6], however, its application in nowcasting tasks is not yet common. Therefore, the analysis of the physical impact of the polarization parameters in the task of nowcasting through dual-polarization radar data and the deep learning approach to nowcasting not only verify the contribution of the dual polarization parameters to nowcasting but also enhance the interpretability of the deep learning to accomplish the meteorological task.

The purpose of this paper is to further explore the contribution of the input polarization variables toward the model performance, unlocking the black box and empowering models with interpretability. The remainder of this article is organized as follows.

Section II examines two meteorological processes, and analyzes the impact of polarization parameters under deep learning. Section III ranks the contributions of the polarization parameters at different prediction moments by conducting the Breiman permutation tests for 20 cases. Section IV visualizes the feature maps during forward propagation to exhibit the feature extraction process of different parts of the network with or without the introduction of dual-polarization parameters. In section V, Z, Z_{dr}, K_{dp} are downscaled to two-dimensional space by the t-SNE method using 4096 cases to present the relationship between the variables. Conclusion is presented in Section VI.

II. Cases Study

In this section, the polarization parameter channels in the original MCT-3 are removed and retrained to obtain the MCT-R model. With different characteristics of meteorological processes (movement and increase of strong convection) [7], representative cases are analyzed and the laws are obtained analytically.

A. Strong echos movement

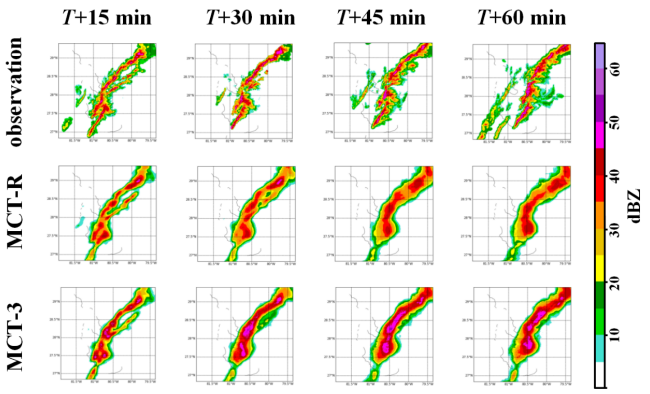

Fig. 1. Prediction and observation of case A in KMLB area where T is 0000 UTC on 16 Feb 2021.

First, the squall line is used as an example to demonstrate the characteristics of the movement of strong convection. Such echoes are characterized by a predominantly moving energy concentration of 40-50 dBZ (see Fig. 1), with no significant growth or dissipation trend occurring. Observing the forecasting, the introduction of the polarization parameter enables MCT-3 to predict high reflectivity more accurately, as seen in the retention for the stronger energy fraction (more than 40 dBZ). The challenge of maintaining strong echoes as the prediction time increases is often encountered in deep

learning, which is mitigated by the polarization parameter in this case.

B. Strong echos growth

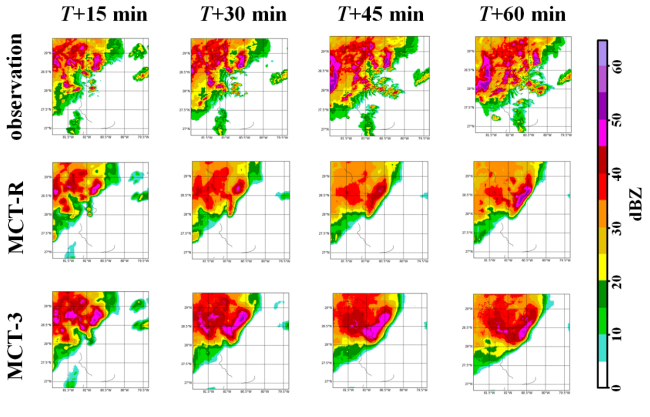

Fig. 2. Prediction and observation of case B in KMLB area where T is 1517 UTC on 11 April 2021.

Next, a case of strong convective growth is analyzed. In the case shown in Fig. 2, the intensity peaks up to 55 dBZ and tends to increase. MCT-R predicts almost nothing in the range of true values greater than 35 dBZ, but MCT-3 covers it more accurately and the shape of MCT-3 predictions for strong echoes (35-45 dBZ) is generally much closer. It is worth noting that neither model predicts new strong cloud masses, which is a difficulty in deep learning graph-to-graph prediction, but MCT-3 is closer to the true value for 35-45 dBZ.

Combined with the analysis of the two precipitation types in Section II, the following conclusions can be drawn:

The polarization parameter has a positive contribution in predicting the intensity of the reflectivity factor and maintaining high energy (40-50 dBZ) and forecasting the location of the echoes.

III. PERMUTATION TEST

The basic idea of the Permutation Test (PT) is that the input variable x_i is reordered to obtain the control result $Y^{'}$, which is compared with the predicted result Y before the permutation to obtain the difference ΔY. A larger ΔY reflects a more dominant x_i. This scheme allows more flexibility to examine the contribution of individual variables without retraining (see Section II). In this work, two permutation tests are conducted to assess the impact of each variable using CSI (threshold=35 dBZ) as a measure, with 20 cases selected.

Breiman's approach is to disrupt the different variable channels separately to different samples and rank them according to the degradation of the model performance to validate the factor pair forecasting results [8]. Scrutinizing the performance ranking after permutation in Fig. 3, it is par of the course that taking away the reflectivity causes the sharpest decrease in model prediction performance since Z is the output itself. The second-ranking of performance loss at all four prediction moments is K_{dp}, indicating that it is of greater importance

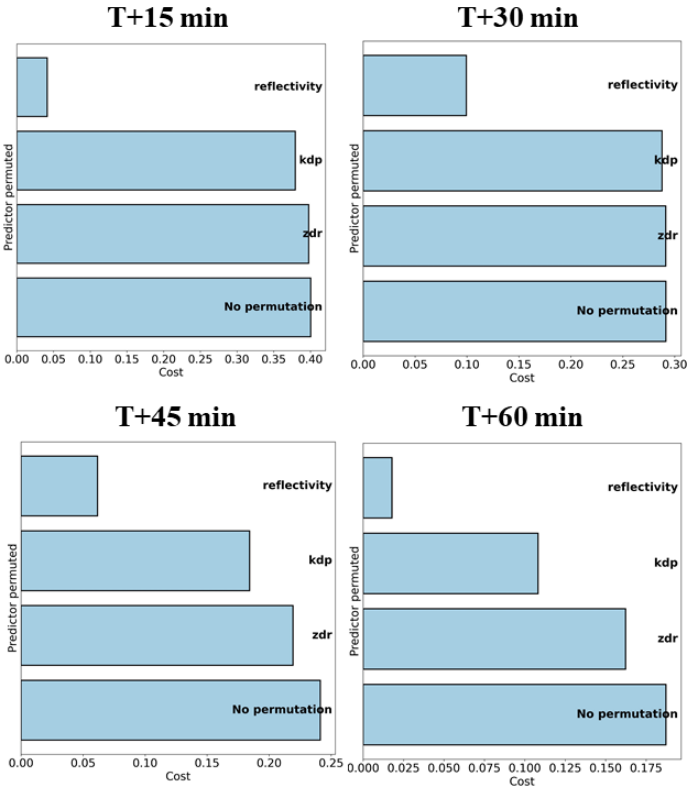

Fig. 3. Breiman's approach in four display moments. The length of the bar represents the forecast performance of the model after removing each variable separately, and the variables are ranked according to their performance loss.

than Z_{dr}, which is in third place. Additionally, it can be seen that the contribution of K_{dp}/Z_{dr} manifests itself more in the results of the 30 min-60 min prediction duration because the metrics of the first 30 min show that the loss of CSI with K_{dp} and Z_{dr} removed is minor and almost close to the results without permutation.

To summarize, the understanding of the contribution of the polarization parameters through the permutation test in this paper is as follows:

(1). The contribution of K_{dp} is higher than that of Z_{dr}.

(2). The effect of polarization covariance is more significant at the forecast duration of the latter 30 min.

IV. FEATURE MAP VISUALIZATION

The feature map presents how much the neural unit is activated during forward propagation [9]. MCT belongs to multi-scale coding and decoding neural network, and different scales of network extract information with different focus. Comparing the feature maps of multiple depths, the learning process of the neural network could be intuited.

Using the MCT model as the object of study, a strong convective forecast process on February 13, 2021 is analyzed, and Fig. 4 visualizes the characteristic maps at different scales, and the conclusions that can be drawn from the observations:

(1). Comparing the shallow and deep semantics (down1/4, up 2/3/4), the focus of each layer is different, but the more

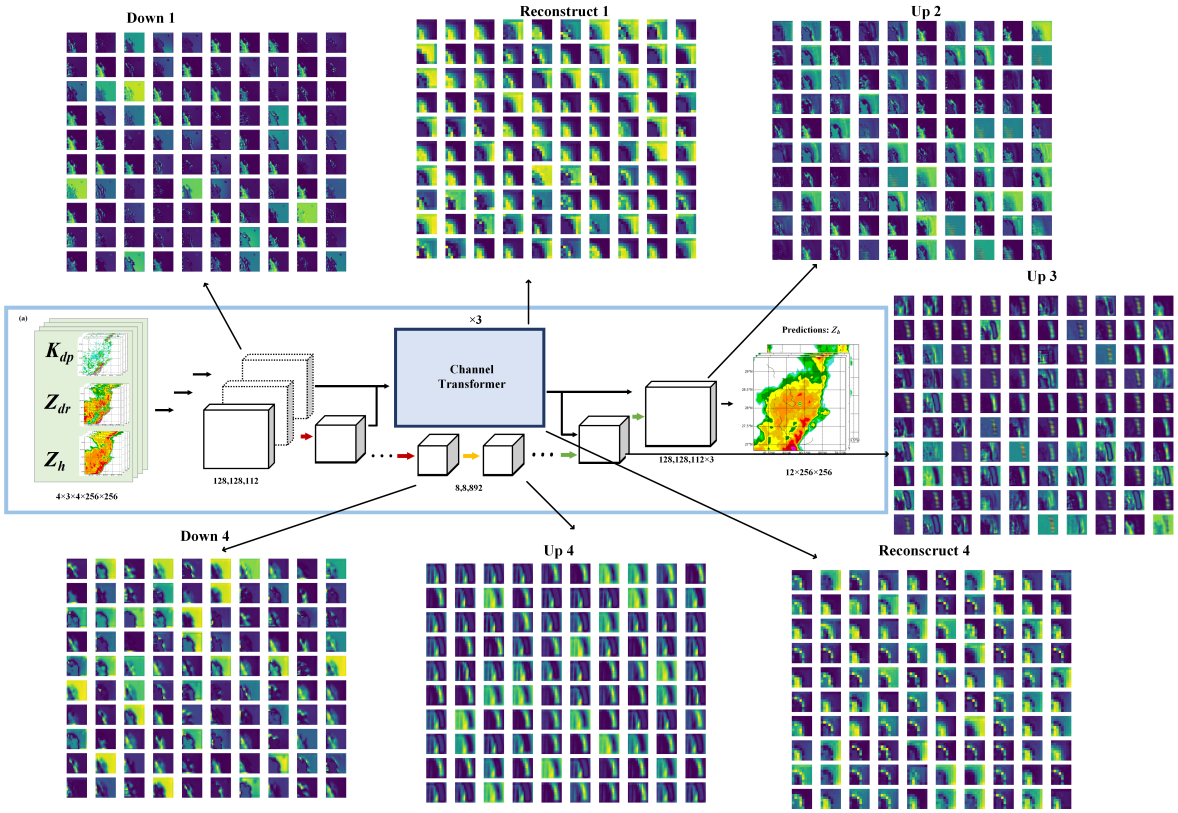

Fig. 4. Visualization results of feature maps at different semantic scales of the Feb. 13, 2021 case. It includes the encoder parts (Down), decoder parts (Up) and Channel Transformer parts (Reconstruct).

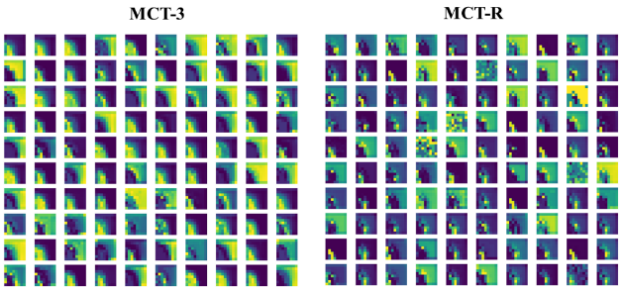

Fig. 5. Comparison of feature maps of MCT-3/R in Channel Transformer reconstruction layer 1 of the Feb. 13, 2021 case.

shallow the information, the more details it is (down1, up2), while the deeper the information the more abstract and important it is. (2). Comparing the encoder-decoder information (down/up), the encoder-decoder focus is not exactly the same. The encoder extracts features that are still closer to the storm shape, while the decoder extracts to more abstract features. (3). Focusing on the semantic information generated by the Transformer, the focus generated at different semantic layers is different, but all are more abstract and important (reconstruct 1/4). The semantic information generated by the attention mechanism in the same layer (reconstruct 4, up/down 4) is different from the output of the encoder, indicating that the attention mechanism in the Transformer plays a better role in providing more information.

Fig. 5 shows a comparison of the feature maps of the reconstructed layer of the attention mechanism with and without the polarization parameter to observe their contribution. From this, the semantic information extracted by each layer is not equivalent and the MCT-R sometimes suffers from invalid features such as random distribution, which also demonstrates that the presence of polarization parameters may enrich the data information and enhance the ability of the model to extract abstract information.

V. DATA DOWNSCALING VISUALIZATION

When the three physical variables are characterized in terms of space, the strength of correlation of physical locations can be well represented; however, such a representation is more redundant if the characterization between data is desired. Therefore, this paper considers mapping the samples characterized by complex high-dimensional features (spatial dimensionality, H × W) to a low-dimensional, more easily understood space. In this way the distance and aggregation between samples provides a visual representation of the similarity between samples.

Therefore, this paper attempts to implement data dimensionality reduction using t-Distributed Stochastic Neighbor

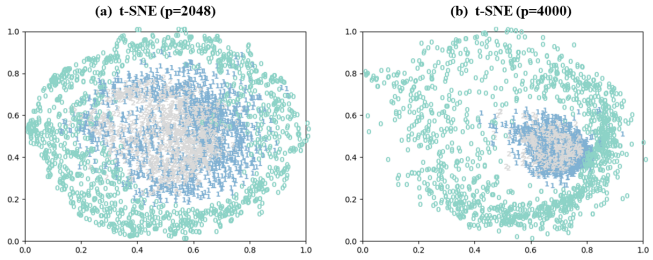

Fig. 6. t-SNE visualization results. (a) t-SNE with 2048 perplexity. (b) t-SNE with 4000 perplexity. The more p is, the global perspective the method presents. 0 (green) represents the horizontal reflectivity Z_h, 1 (blue) represents the differential reflectivity Z_{dr}, and 2 (gray) represents the differential phase shift K_{dp}.

Embedding (t-SNE) [10]. The t-SNE can keep an eye on the overall characteristics as well as the local details, distributing similar samples at closer distances in a lower dimensional, easily human-observable space, while pulling apart clusters with t-distribution. This algorithm contains the parameter perplexity p. p characterizes the consideration of the number of surrounding samples when calculating the similarity of two samples in the original space, which needs to be smaller than the number of samples. the larger p is, the more macroscopic the perspective considered by the algorithm. In this subsection, 4092 samples in the test set are taken for visualization, the dimension of former space is 256*256 and the dimension of latter space is 2. In Fig. 6, 0 (green) represents the horizontal reflectivity Z_h, 1 (blue) represents the differential reflectivity Z_{dr}, and 2 (gray) represents the differential phase shift K_{dp}. It can be seen that the distribution of (a) and (b) are similar, both forming three concentric circles, with the sample of Z_h in the outermost layer, then Z_{dr}, and finally K_{dp}. As p increases, the samples considered by t-SNE are more comprehensive, and the distance between Z_h and the polarization parameter Z_{dr}/K_{dp} is farther, indicating that the difference in features is greater. Also, the distance between K_{dp} and Z_h is greater than the distance between Z_{dr} and Z_h, indicating that the "K_{dp}-Z_h" mutual information is smaller, which is conducive to the model to get more angles and more comprehensive information, and also verifies the previous conclusion that the contribution of K_{dp} is higher.

VI. Conclusion

In this paper, four methods are designed to analyze dual polarization based on a deep learning nowcasting model MCT that incorporates attention mechanisms and dual polarization radar data, using selected precipitation data from the KMLB WSR88D. The focus of the study is to explore the physical significance of the dual polarization and its contribution to nowcasting through deep learning approaches. The main findings are summarized as follows:

(i). The polarization parameter contributes to the prediction of strong convection, the retention of energy in high intensity echoes, and the position determination.

(ii). The dual polarization data does provide additional information to assist the model and the contribution of K_{dp} is higher than that of Z_{dr} in the deep learning method like MCT.

(iii). In the experiment with 60 min forecast duration, the contribution of polarization data is more prominent at longer forecast durations.

Acknowledgment

The research was supported by the U.S. National Science Foundation (NSF) CAREER program. The work of Lei Han was supported by the National Natural Science Foundation of China (grant number 42275003). The authors gratefully acknowledge the support of NVIDIA Corporation for the donation of GPU used for this research.

References

[1] L. Espeholt, S. Agrawal, C. Sønderby, M. Kumar, J. Heek, C. Bromberg, C. Gazen, J. Hickey, A. Bell, and N. Kalchbrenner, "Skillful twelve hour precipitation forecasts using large context neural networks," *arXiv preprint arXiv:2111.07470*, 2021.

[2] L. Han, J. Sun, and W. Zhang, "Convolutional neural network for convective storm nowcasting using 3-d doppler weather radar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1487–1495, 2019.

[3] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," *Advances in neural information processing systems*, vol. 30, 2017.

[4] H. Chen and V. Chandrasekar, "Estimation of light rainfall using ku-band dual-polarization radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 5197–5208, 2015.

[5] H. Chen, V. Chandrasekar, and R. Bechini, "An improved dual-polarization radar rainfall algorithm (drops2. 0): Application in nasa ifloods field campaign," *Journal of Hydrometeorology*, vol. 18, no. 4, pp. 917–937, 2017.

[6] Y. Gou and H. Chen, "Combining radar attenuation and partial beam blockage corrections for improved quantitative application," *Journal of Hydrometeorology*, vol. 22, no. 1, pp. 139–153, 2021.

[7] N. Velásquez, "Assessment of deep convective systems in the colombian andean region," *Hydrology*, vol. 9, no. 7, p. 119, 2022.

[8] L. Breiman, *Random forests*. Springer, 2001, vol. 45.

[9] A. McGovern, R. Lagerquist, D. J. Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, "Making the black box more transparent: Understanding the physical implications of machine learning," *Bulletin of the American Meteorological Society*, vol. 100, no. 11, pp. 2175–2199, 2019.

[10] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.