

Machine Learning–based Weather Support for the 2022 Winter Olympics

Jiangjiang XIA^{1,2}, Haochen LI^{3,4}, Yanyan KANG⁵, Chen YU³, Lei JI⁶, Lve WU⁷, Xiao LOU¹, Guangxiang ZHU⁸, Zaiwen Wang⁶, Zhongwei YAN^{1,2}, Lizhi WANG¹, Jiang ZHU¹, Pingwen ZHANG³, Min CHEN⁶, Yingxin ZHANG⁵, Lihao GAO⁹, and Jiarui HAN¹⁰

¹*Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

²*University of Chinese Academy of Sciences, Beijing 100049, China*

³*School of Mathematical Sciences, Peking University, Beijing 100871, China*

⁴*School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China*

⁵*Beijing Weather Forecast Center, Beijing 100089, China*

⁶*Institute of Urban Meteorology, China Meteorological Administration, Beijing 100089, China*

⁷*School of Information and Communication Engineering, Beijing Information Science and Technology University, Beijing 100101, China*

⁸*Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China*

⁹*Chinese Academy of Meteorological Sciences, Beijing 100081, China*

¹⁰*Training Centre, China Meteorological Administration, Beijing 100081, China*

(Received 21 February 2020; revised 7 May 2020; accepted 28 May 2020)

Citation: Xia, J. J., and Coauthors, 2020: Machine learning–based weather support for the 2022 Winter Olympics. *Adv. Atmos. Sci.*, 37(9), 927–932, <https://doi.org/10.1007/s00376-020-0043-5>.

1. A key support for the 2022 Winter Olympics

The XXIV Olympic Winter Games are scheduled to take place from 4 to 22 February 2022, followed by the Paralympic Games from 4 to 13 March, in Beijing and towns in the neighboring Hebei Province, China. Weather plays an extremely important role in the outcome of the games (Chen et al., 2018). It can not only cause a difference between a medal or not, but affect the safety of athletes. Success of the Winter Olympics will greatly depend on weather conditions at the outdoor competition venues, dealing with many weather elements including the snow surface temperature, apparent temperature, gust wind speed, snow, visibility, etc. To ensure that the scheduled games go smoothly, it is imperative to have hourly or even every 10-minutely forecasts as well as updated weather-related risk assessments at the venues for the next 240 hours. So far, the Beijing/Hebei Meteorological Observatory has already started intelligent weather forecasting at 3-km resolution based on the results of numerical weather prediction (NWP) models. However, these experiments have suggested that the current forecasting techniques are incapable of capturing the complex mountain weather variations around some venues. The forecasting capability of NWP is constrained partly by limited knowledge of the local weather mechanisms.

In recent years, artificial intelligence (AI) has accelerated the development of meteorology. The integration of AI and conventional physical simulation can help to improve the objective forecasting capability of NWP models. As an application of AI, machine learning is also expected to extend the forecasting system's ability to automatically learn and improve from past experience and novel understanding of the physical mechanism. It could be used as a post-processing method for NWP model outputs, by matching NWP forecasts against observations through a regression function. It takes advantages of both physical simulation (NWP models) and data-driven (machine learning) approaches, e.g., using physical simulation to predict a wide range of environmental atmospheric trends and using data-driven algorithms to capture highly localized details, thus improving the forecasting capability. Many studies have shown the effectiveness of machine learning as a post-processing method for NWP. It fosters potential solutions to the improvement of weather forecasting at venues in a complex mountain environment for the Winter Olympics.

* Corresponding author: Jiangjiang XIA
Email: xiajj@tea.ac.cn

2. Progress and problems

Aimed at providing high-quality mountain weather forecasts for the Winter Olympics, a data science team was organized in August 2018, involving the Beijing Weather Forecast Center (BWFC), the Institute of Atmospheric Physics (IAP) of the Chinese Academy of Sciences, the School of Mathematical Sciences (SMS) of Peking University, and the Beijing Institute of Urban Meteorology (IUM). This team is composed of members experienced in fields of artificial intelligence, weather forecasting, and meteorological big data. Five groups were set up with specific targets as follows:

- (1) Understanding the needs of weather forecasting for the 2022 Winter Olympics, led by BWFC and jointly conducted by all the partners;
- (2) Data Acquisition and Understanding, led by BWFC;
- (3) Data Preprocessing and Feature Engineering, led by IAP;
- (4) Machine Learning Model construction, led by SMS;
- (5) Application Test, mainly led by IUM.

This group work is partly supported by the National Key Research and Development Program of China, via a project led by BWFC. Figure 1 shows the cyclic linkages among the work of the different groups.

A few studies have been conducted since the organization of the team. Li et al. (2019) proposed an NWP model (ECMWF model) output machine learning (MOML) method. Several experiments showed significant improvement of the forecast accuracy of weather elements for the Olympic venues based on machine-learning methods (e.g., Sun et al., 2019; Yu et al., 2020). The MOML method has been integrated into the existing meteorological operational system. MOML is a post-processing method, which matches NWP forecasts against a long series of verifying observations through a regression function that uses a machine learning algorithm (Fig. 2). The MOML method uses machine learning algorithms including multiple linear regression, support vector regression, random forest, gradient boosting decision tree, XGBoost, deep learning, etc. It takes the NWP model output (wind speed, wind direction, temperature, humidity, pressure, precipitation, etc.) and the geographic information (such as latitude and longitude, altitude, slope, aspect, distance from the coastline, underlying surface type, etc.) of each weather station (or each grid) as the input (“features”), to predict the meteorological elements (near-surface air temperature, wind speed, precipitation, etc., which are also termed “labels”). MOML constructs datasets by feature engineering based on spatial and temporal data, which makes full use of the spatial and temporal structure of a point on the grid (Li et al., 2019). The machine learning-based MOML method could learn the general physical laws from a large number of model output data and geographic information data. Experimental results showed that the MOML method increased the forecasting accuracy by more than 10%, compared with conventional model output statistics (MOS) and ECMWF model results (Fig. 3).

Of note is that the linear regression model can perform better than other models for short-term forecasting; only for mid-to-long-term forecasting can machine learning and deep learning take full advantage (Li et al., 2019). Two reasons are considered. On the one hand, the NWP is not very biased for short-term forecasts, such that some linear relationship can be applied well between its output and observation. On the other hand, machine learning, especially deep learning, requires a large amount of data. Although our training dataset is quite large, it remains insufficient. For example, there are about 20 stations around the venue on Haituo Mountain, but only four stations have relatively long weather records back to 2014.

Though some progress has been made, the achievements so far have not satisfied the required forecast accuracy. One of the major difficulties is the wind forecast. The Winter Olympics requires much stricter constraints on wind speed and direction than on other weather elements. The most common snow sports all have specific requirements for wind speed and direction. Alpine skiing requires a gust wind speed below 14 m s^{-1} , with an alert threshold of 11 m s^{-1} for average wind speed; a wind speed above 13 m s^{-1} will greatly affect sports involving sleighs; ski jumping has a desirable wind speed of no more than 4 m s^{-1} and requires uphill/downhill wind speed change to be within 1 m s^{-1} . Moreover, all kinds of wind direction will impact the result of outdoor games and even the safety of athletes. Wind forecasting is even more important in the mountain area, since the uphill and downhill winds differ significantly. Unfortunately, since the weather observatories at the moun-

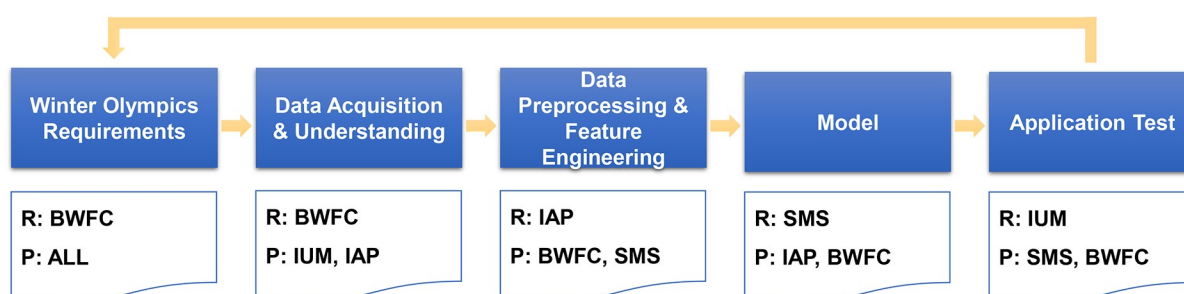


Fig. 1. Machine learning applications lifecycle for the 2022 Winter Olympics.

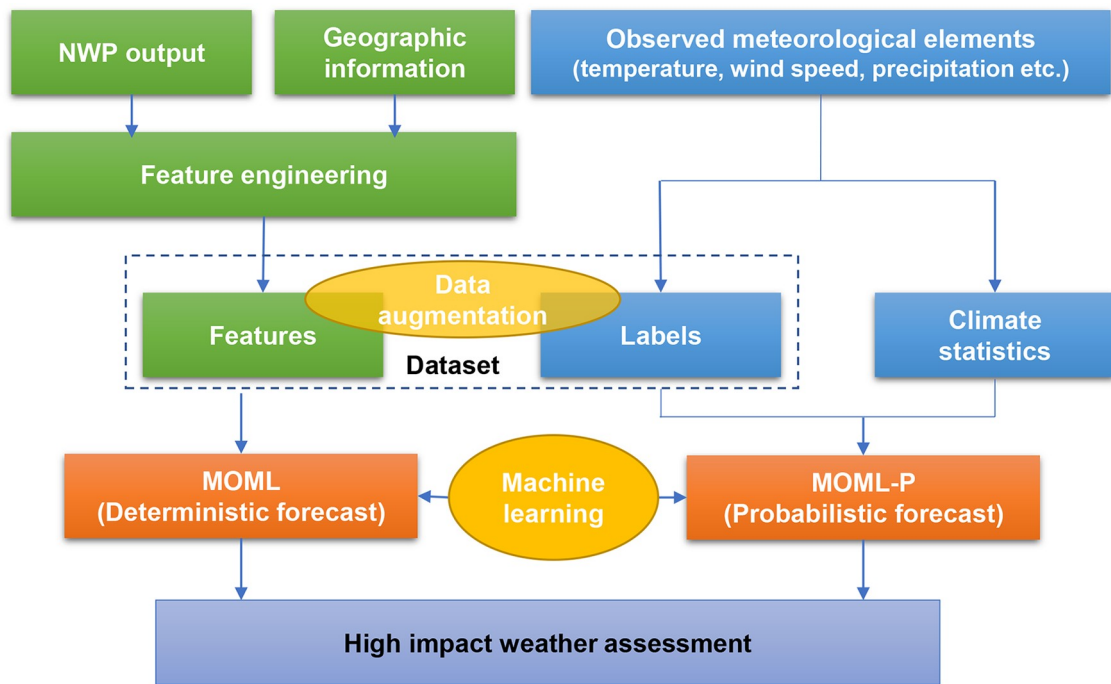


Fig. 2. Framework for applying machine learning to NWP post-processing.

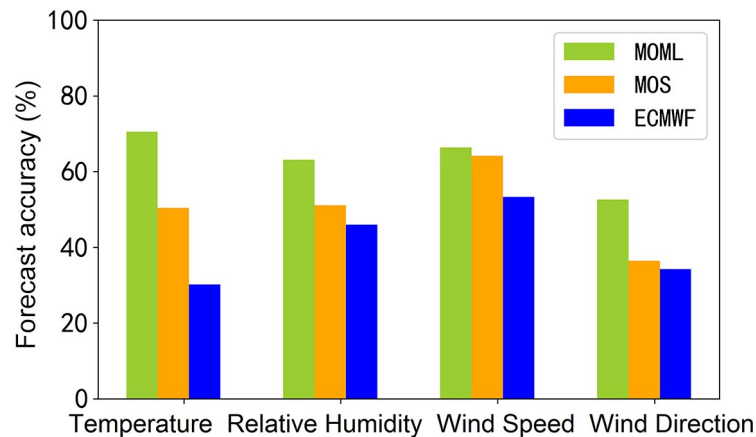


Fig. 3. Comparison of forecast accuracy for four meteorological elements averaged over four stations (located on Haituo Mountain) at a 7-day forecasting lead time among MOML, MOS and ECMWF model output.

tainous Olympic venues have not been established for long, there are no sufficient data on winter winds for training a machine-learning model, leading to inaccurate forecasting and large uncertainties. Therefore, it is a serious challenge for the interdisciplinary data science team to improve the forecasting of local wind speed and direction for the mountainous venues.

3. Focus of further work

Additional machine-learning model tuning and relevant feature engineering are critical and still under way. Moving forward, we seek to improve the machine learning-based forecast accuracy (especially for the wind forecast) by carrying out the following three tasks.

First, apply MOML with refined numerical model prediction results. The Rapid-refresh Multi-scale Analysis and Prediction System-Short Term (or RMAPS-ST), which aims at forecasting the weather condition in northern China, will be considered as an alternative weather forecasting system. The system provides hourly forecasting results for 0-3 days at a horizontal grid resolution of 3 × 3 km, the spatiotemporal resolution of which is much higher than that of the ECMWF model. It has played a key role in weather forecasting operations in Beijing. A professional meteorological research team from IUM has been working on developing and maintaining RMAPS-ST for many years (Fan et al., 2013; Xie et al., 2019).

Second, integrate semi-intelligent forecasts with forecasters' experience and data augmentation techniques. Forecasters have a wealth of weather knowledge and long-term accumulated forecasting experience, especially for some severe weather processes. Forecasters can use the results of advanced artificial intelligence models (e.g., MOML) to combine with prior knowledge to make more accurate forecasts. As extreme and turning weather records are rare and not sufficient for machine-learning training, data augmentation techniques (Perez and Wang, 2017; Shorten and Khoshgoftaar, 2019) could be used to increase samples of such rare and abnormal weather events.

Finally, produce probability predictions by integrating the MOML with climate statistics. There is a limit to predictability for any NWP model or MOML, and no perfect forecast can be achieved on any scale. Ensemble forecasting has advantages over deterministic forecasting (Bougeault et al., 2010), by giving the possibility of certain weather conditions and describe the uncertainty of weather variations. In this area of study, we will calculate the climate statistics for each meteorological element (near-surface air temperature, wind speed, precipitation, etc.) based on high-temporal-resolution (5-min) historical records (Fig. 3). For example, we will calculate the probability of occurrence of an event with wind speeds beyond a certain criterion for each station. Then, using machine learning to match the MOML output against the probability, build an MOML-P model (MOML probabilistic forecast model). Studies have already shown that probabilistic forecasting can be well applied to users' decision-making, bringing considerable social and economic benefits to users. The integrated MOML-P model is a hopeful solution.

4. Views beyond the Olympics

As a technique for implementing AI, machine learning, especially deep learning, which has greatly accelerated the development of AI, is quite powerful. This tool has been needed for developing objective meteorological prediction. Taking its data-driven advantage, a machine-learning framework will make optimal use of all available predictors and enable the model to routinely incorporate new predictors as they become available (Veillette et al., 2013). In particular, machine learning is feasible to address the challenges of nonlinear and uncertainty problems of complicated weather and climate systems (Xu, 2018). In fact, by using machine learning, meteorologists have made some breakthroughs.

For example, nowcasting has become one of the most influential streams of predictive studies employing machine learning (especially deep learning) models. Shi et al. (2015) proposed the Convolutional Long-Short Term Memory (ConvLSTM) model for precipitation nowcasting based on radar data. ConvLSTM merges the Convolutional Neural Network for extracting abstract features from images and LSTM for processing sequential data to solve the spatiotemporal sequence forecasting problem. Thus, it made a major breakthrough. Many subsequent studies have been conducted and demonstrated that deep learning models can significantly outperform traditional statistical algorithms (Shi et al., 2017; Wang et al., 2017, 2018, 2019; Guo et al., 2019; Jing et al., 2019). However, this kind of deep learning model is mainly an extrapolation of radar reflectivity without considering other meteorological fields, e.g. the wind, temperature, and humidity, thus having inherent limitation for forecasting. Some works have employed multi-meteorological fields for nowcasting, e.g. the convective initiation, by using machine learning (Veillette et al., 2013; Han et al., 2017; Liu et al., 2019; Zhang et al., 2019; Zhou et al., 2019). However, most of these forecast studies focused on a time point rather than series. Relating information from multiple sources and integrating multimodal knowledge is one of the core problems in general AI. Multimodal machine learning (Ngiam et al., 2011; Srivastava and Salakhutdinov, 2012; Baltrušaitis et al., 2019) has been proposed to address this issue by learning shared features over multiple-modalities data and making predictions for multiple tasks in a uniform framework. It could be employed to address weather forecast issues too.

To improve the accuracy of forecasting, combining the numerical model and AI is an efficient approach (Reichstein et al., 2019; Toms et al., 2019). AI techniques can be used as model-post processing methods, as done in the present project. Alternatively, they can be integrated into the numerical model. For example, machine learning was used to optimize parameterization schemes in a numerical model based on observations and targeted high-resolution simulations (Schneider et al., 2017). It was also interesting to build a fast AI model, which closely reproduces a far more expensive simulation, referred to as model emulation (Prudden et al., 2017). Model emulation is of potential value for a wide range of applications in meteorological and climate studies.

Machine learning has been used for years as a black-box solution. In recent years, it has become a tool to help to explore underlying mechanisms for weather prediction and climate diagnosis (e.g., Gao et al., 2019; Ham et al., 2019; Tong et al., 2019; Wei et al., 2020). In addition, the data-driven advantages of machine learning also facilitate characterizing the impact of weather and climate change on human society, economies, etc., thus constructing impact assessment and auxiliary decision-making models for various users (McGovern et al., 2017). In particular, it is of great potential in developing a machine learning-based replacement for conventional weather consultation.

Application of machine learning in meteorological and climate sciences is still in its infancy. It requires not only abundant resources (e.g., for big data analysis) but also multidisciplinary collaborations. The 2022 Winter Olympics offers a good opportunity for developing and training a multidisciplinary group of young researchers in meteorology, computer science,

mathematics and operational forecasting to promote the development of machine learning–based weather forecasting.

Acknowledgements. This work was supported by the National Key Research and Development Program of China (Grant No. 2018YFF0300104), Beijing Academy of Artificial Intelligence, and the Open Research Fund of the Shenzhen Research Institute of Big Data (Grant No. 2019ORF01001).

REFERENCES

- Baltrušaitis, T., C. Ahuja, and L. P. Morency, 2019: Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**(2), 423–443, <https://doi.org/10.1109/TPAMI.2018.2798607>.
- Bougeault, P., and Coauthors, 2010: The THORPEX interactive grand global ensemble. *Bull. Amer. Meteorol. Soc.*, **91**, 1059–1072, <https://doi.org/10.1175/2010BAMS2853.1>.
- Chen, M. X., and Coauthors, 2018: Enhanced weather research and forecasting in support of the Beijing 2022 Winter Olympic and Paralympic Games. *WMO Bulletin*, **67**(2), 58–61.
- Fan, S. Y., H. L. Wang, M. Chen, and H. Gao, 2013: Study of the data assimilation of radar reflectivity with the WRF 3D-Var. *Acta Meteorologica Sinica*, **71**, 527–537, <https://doi.org/10.11676/qxxb2013.032>. (in Chinese)
- Gao, L. H., F. Y. Wei, Z. W. Yan, J. Ma, and J. J. Xia, 2019: A study of objective prediction for summer precipitation patterns over eastern China based on a multinomial logistic regression model. *Atmosphere*, **10**(4), 213, <https://doi.org/10.3390/atmos10040213>.
- Guo, H. Y., M. X. Chen, L. Han, W. Zhang, R. Qin, and L. Y. Song, 2019: High resolution nowcasting experiment of severe convections based on deep learning. *Acta Meteorologica Sinica*, **77**(4), 715–727, <https://doi.org/10.11676/qxxb2019.036>. (in Chinese)
- Ham, Y. G., J. H. Kim, and J. J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, <https://doi.org/10.1038/s41586-019-1559-7>.
- Han, L., J. Z. Sun, W. Zhang, Y. Y. Xiu, H. L. Feng, and Y. J. Lin, 2017: A machine learning nowcasting method based on real-time reanalysis data. *J. Geophys. Res. Atmos.*, **122**, 4038–4051, <https://doi.org/10.1002/2016JD025783>.
- Jing, J. R., Q. Li, and X. Peng, 2019: MLC-LSTM: Exploiting the spatiotemporal correlation between multi-level weather radar echoes for echo sequence extrapolation. *Sensors*, **19**(18), 3988, <https://doi.org/10.3390/s19183988>.
- Li, H. C., C. Yu, J. J. Xia, Y. C. Wang, J. Zhu, and P. W. Zhang, 2019: A model output machine learning method for grid temperature forecasts in the Beijing area. *Adv. Atmos. Sci.*, **36**(10), 1156–1170, <https://doi.org/10.1007/s00376-019-9023-z>.
- Liu, Z. J., and Coauthors, 2019: Local severe storm tracking and warning in pre-convection stage from the new generation geostationary weather satellite measurements. *Remote Sensing*, **11**(4), 383, <https://doi.org/10.3390/rs11040383>.
- McGovern, A., K. L. Elmore, D. J. Gagne II, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteorol. Soc.*, **98**, 2073–2090, <https://doi.org/10.1175/BAMS-D-16-0123.1>.
- Ngiam, J., A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, 2011: Multimodal deep learning. *Proceedings of the 29th International Conference on Machine Learning (ICML)*, Bellevue, ICML.
- Perez, L., and J. Wang, 2017: The effectiveness of data augmentation in image classification using deep learning. arXiv preprint, arXiv:1712.04621.
- Prudden, R., N. Robinson, A. Arribas, and C. Ewen, 2017: Machine learning in weather forecasting. Expert Articles in ODBMS.org. [Available from <http://www.odbms.org/2017/07/machine-learning-in-weather-forecasting/>]
- Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195–204, <https://doi.org/10.1038/s41586-019-0912-1>.
- Schneider, T., S. W. Lan, A. Stuart, and J. Teixeira, 2017: Earth System Modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys. Res. Lett.*, **44**, 12 396–12 417, <https://doi.org/10.1002/2017GL076101>.
- Shi, X. J., Z. R. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Proceedings of the 28th International Conference on Neural Information Processing Systems*, Cambridge, MIT Press, 802–810.
- Shi, X. J., Z. H. Gao, L. Lausen, H. Wang, and D. Y. Yeung, 2017: Deep learning for precipitation nowcasting: A benchmark and a new model. *Proceedings of the 31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, Curran Associates Inc., 5617–5627.
- Shorten, C., and T. M. Khoshgoftaar, 2019: A survey on image data augmentation for deep learning. *Journal of Big Data*, **6**(1), 60, <https://doi.org/10.1186/s40537-019-0197-0>.
- Srivastava, N., and R. Salakhutdinov, 2012: Multimodal learning with deep Boltzmann machines. *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Red Hook, Curran Associates Inc., 2222–2230.
- Sun, Q. D., R. L. Jiao, J. J. Xia, Z. W. Yan, H. C. Li, J. H. Sun, L. Z. Wang, and Z. M. Liang, 2019: Adjusting wind speed prediction of numerical weather forecast model based on machine learning methods. *Meteor. Mon.*, **45**(3), 426–436, <https://doi.org/10.7519/j.issn.1000-0526.2019.03.012>. (in Chinese with English abstract)
- Toms, B. A., E. A. Barnes, and I. Ebert-Uphoff, 2019: Physically interpretable neural networks for the geosciences: Applications to earth system variability. arXiv preprint arXiv:1912.01752.
- Tong, X., Z. W. Yan, J. J. Xia, and X. Lou, 2019: Decisive atmospheric circulation indices for July–August precipitation in North China based on tree models. *Journal of Hydrometeorology*, **20**, 1707–1720, <https://doi.org/10.1175/JHM-D-19-0045.1>.
- Veillette, M. S., H. Iskenderian, P. M. Lamey, and L. J. Bickmeier, 2013: Convective initiation forecasts through the use of machine learn-

- ing methods. 11th Conf. on Artificial and Computational Intelligence and its Applications to the Environmental Sciences. [Available from <https://pdfs.semanticscholar.org/8020/b88ea36667d45d3d5f7fee99aa1b33c1f10d.pdf>]
- Wang, Y. B., M. S. Long, J. M. Wang, Z. F. Gao, and P. S. Yu, 2017: PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. *Proceedings of the 31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, Curran Associates Inc., 879–888.
- Wang, Y. B., Z. F. Gao, M. S. Long, J. M. Wang, and P. S. Yu, 2018: PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning. *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, ICML, 5123–5132.
- Wang, Y. B., J. J. Zhang, H. Y. Zhu, M. S. Long, J. M. Wang, and P. S. Yu, 2019: Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. arXiv preprint arXiv: 1811.07490.
- Wei, W. G., Z. W. Yan, and P. D. Jones, 2020: A decision-tree approach to seasonal prediction of extreme precipitation in eastern China. *International Journal of Climatology*, **40**(1), 255–272, <https://doi.org/10.1002/joc.6207>.
- Xie, Y. H., S. Y. Fan, M. Chen, J. C. Shi, J. Q. Zhong, and X. Y. Zhang, 2019: An assessment of satellite radiance data assimilation in RMAPS. *Remote Sensing*, **11**, 54, <https://doi.org/10.3390/rs11010054>.
- Xu, X. F., 2018: From physical model to intelligent analysis: A new exploration to reduce the uncertainty of weather forecast. *Meteor. Mon.*, **44**(3), 341–350, <https://doi.org/10.7519/j.issn.1000-0526.2018.03.001>. (in Chinese)
- Yu, C., H. C. Li, J. J. Xia, H. Q. Z. Wen, and P. W. Zhang, 2020: A data-driven random subfeature ensemble learning algorithm for weather forecasting. *Commun. Comput. Phys.* (in press)
- Zhang, W., L. Han, J. Z. Sun, H. Y. Guo, and J. Dai, 2019: Application of multi-channel 3D-cube successive convolution network for convective storm nowcasting. *Proceedings of 2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, IEEE.
- Zhou, K. H., Y. G. Zheng, B. Li, W. S. Dong, and X. L. Zhang, 2019: Forecasting different types of convective weather: A deep learning approach. *J. Meteor. Res.*, **33**(5), 797–809, <https://doi.org/10.1007/s13351-019-8162-6>.