LONG-LEAD FORECASTS OF WINTERTIME AIR STAGNATION INDEX IN SOUTHERN CHINA USING OCEANIC MEMORY EFFECTS

Chenhong Zhou
Department of Computer Science
Hong Kong Baptist University
`20482795@life.hkbu.edu.hk`

Xiaorui Zhang & Meng Gao
Department of Geography
Hong Kong Baptist University

Shanshan Liu
University of Science and
Technology of China

Yike Guo
The Hong Kong University
of Science and Technology

Jie Chen [*]
Department of Computer Science
Hong Kong Baptist University
`chenjie@comp.hkbu.edu.hk`

ABSTRACT

Stagnant weather condition is one of the major contributors to air pollution as it is favorable for the formation and accumulation of pollutants. To measure the atmosphere's ability to dilute air pollutants, Air Stagnation Index (ASI) has been introduced as an important meteorological index. Therefore, making long-lead ASI forecasts is vital to make plans in advance for air quality management. In this study, we found that autumn Niño indices derived from sea surface temperature (SST) anomalies show a negative correlation with wintertime ASI in southern China, offering prospects for a prewinter forecast. We developed an LSTM-based model to predict the future wintertime ASI. Results demonstrated that multivariate inputs (past ASI and Niño indices) achieve better forecast performance than univariate input (only past ASI). The model achieves a correlation coefficient of 0.778 between the actual and predicted ASI, exhibiting a high degree of consistency.

1 INTRODUCTION

Air pollution has become a focus problem which various countries are concerned about. The growth of atmospheric contaminants damages vegetation and crops and even is highly related to some serious human health diseases (Kampa & Castanas, 2008; Al-Saadi et al., 2005; Bai et al., 2018; Masood & Ahmad, 2021). Apart from direct emissions of air pollutants, meteorological conditions are also important to the accumulation and dispersion of pollutants. Air stagnation occurs usually with descending air, low wind speeds, low precipitation, and the compressed boundary layer, which is favorable for the formation and accumulation of atmospheric contaminants (Wu et al., 2017; Gao et al., 2019). Hence, the Air Stagnation Index (ASI) has been proposed to assess the atmosphere's ability to dilute air pollutants (Horton et al., 2012; 2014; Huang et al., 2018). Studies have shown that air stagnation highly correlates with the concentrations of air pollutants (Liao et al., 2006; Huang et al., 2018). Therefore, accurate forecasts of ASI are important and valuable for managing air quality and enabling advance planning.

However, most existing works usually forecast the next several hours or days of air pollution levels in advance by using common meteorological variables (e.g., wind speed, wind direction, humidity, temperature, and rainfall), ground-level observations, and satellite data, etc (Al-Saadi et al., 2005; Bai et al., 2018; Li et al., 2011; Harishkumar et al., 2020; Ham et al., 2019; Chae et al., 2021; Zhang et al., 2020; Xiao et al., 2020; Chen et al., 2021; Kurt & Oktay, 2010; Ham et al., 2021; Chang et al., 2020; Wu & Lin, 2019). At present, long-lead seasonal or multi-year air pollution forecasts are still under exploration, because the ability of long-lead forecasts is highly dependent on finding strongly correlated climate factors and appropriate forecast algorithms. Previous studies have linked the
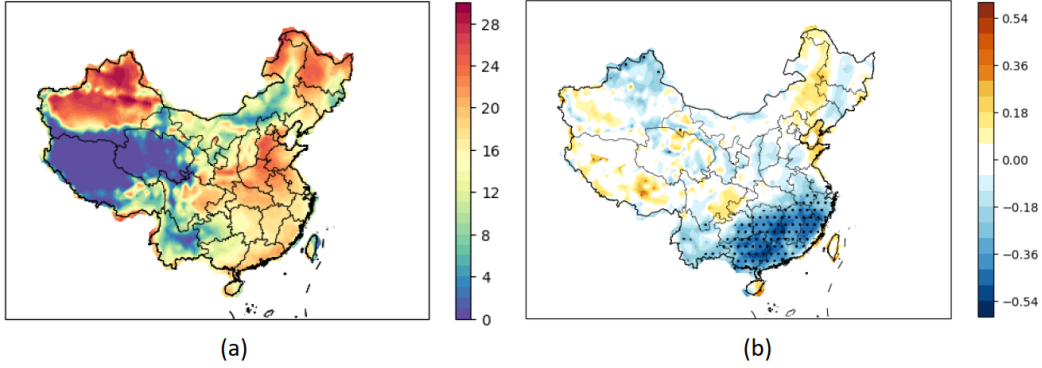
---

[*]Corresponding author.

Figure 1: (a) Multi-year (1950–2020) wintertime monthly mean (From December of one year to February of the following year) ASI distribution across China. (b) Correlation coefficients between autumn Niño 3.4 and wintertime ASI over China, where black dots denote significance (p <0.05).

long-lead El Niño/Southern Oscillation (ENSO) (Ham et al., 2019; 2021), Indian wintertime aerosol pollution (Gao et al., 2019) and wintertime $PM_{2.5}$ concentrations over East Asia (Jeong et al., 2021) with ocean memory effects. They are based on the fact that sea surface temperatures (SST) vary slowly and the presence of decadal oceanic variations as well as their coupling to the atmosphere would modulate the interannual variability of climate change and air pollution. Given these linkages, we aim to investigate the relationship between ocean memory effects and ASI forecasts.

Contributions   We first use the meteorological variables from the ERA5 reanalysis dataset (Hersbach et al., 2020) to generate a long-term ASI dataset over China during a long period from 1950 to 2020. To investigate the feasibility of long-lead ASI forecasts, we analyze the correlations between ASI and ENSO-related indices (Niño 1+2, Niño 3, Niño 3.4, Niño 4) calculated from SST anomalies. The correlation map shows that Niño indices are negatively associated with seasonal variations of ASI in southern China. Furthermore, we develop the Long Short-Term Memory (LSTM) using Niño indices as predictors to obtain skillful long-lead forecasts of wintertime ASI.

Pathways to Climate Impact   We show the statistical predictability of wintertime ASI in southern China using machine learning with ocean memory effects as predictors. Our encouraging results suggest that SST patterns play an important role in long-lead forecasts of wintertime ASI, which is useful for analyzing the impact of climate patterns on air pollution.

2   METHODOLOGY

Data   There is no available ready-made ASI dataset across China, so we first generate a long-term ASI dataset during the period of 1950 to 2020 using the meteorological variables from ERA5 reanalysis dataset (Hersbach et al., 2020). Following the computation process of ASI in (Garrido-Perez et al., 2021), we downloaded these meteorological variables from the ERA5 reanalysis dataset, including convective available potential energy, boundary layer height, convective inhibition, daily wind speed at different heights, and daily accumulated precipitation, etc, to compute the ASI across China. According to the definitions in (Garrido-Perez et al., 2021), we determine whether air stagnation occurs by judging whether the meteorological condition meets some predefined thresholds of daily meteorological fields. Hence, ASI on one day is a binary value, i.e. stagnant/non-stagnant, and the monthly average ASI can be easily obtained by accumulating the values of ASI in one month. A long-term ASI dataset is finally generated and we display the multi-year wintertime monthly mean ASI distribution across China in Figure 1 (a).

Four ENSO-related indices: Niño 1+2, Niño 3, Niño 3.4, and Niño 4, derived from SST anomalies are downloaded from `https://www.cpc.ncep.noaa.gov/data/indices/`. To prove the feasibility of predicting wintertime ASI using Niño indices, we need to investigate the rela-
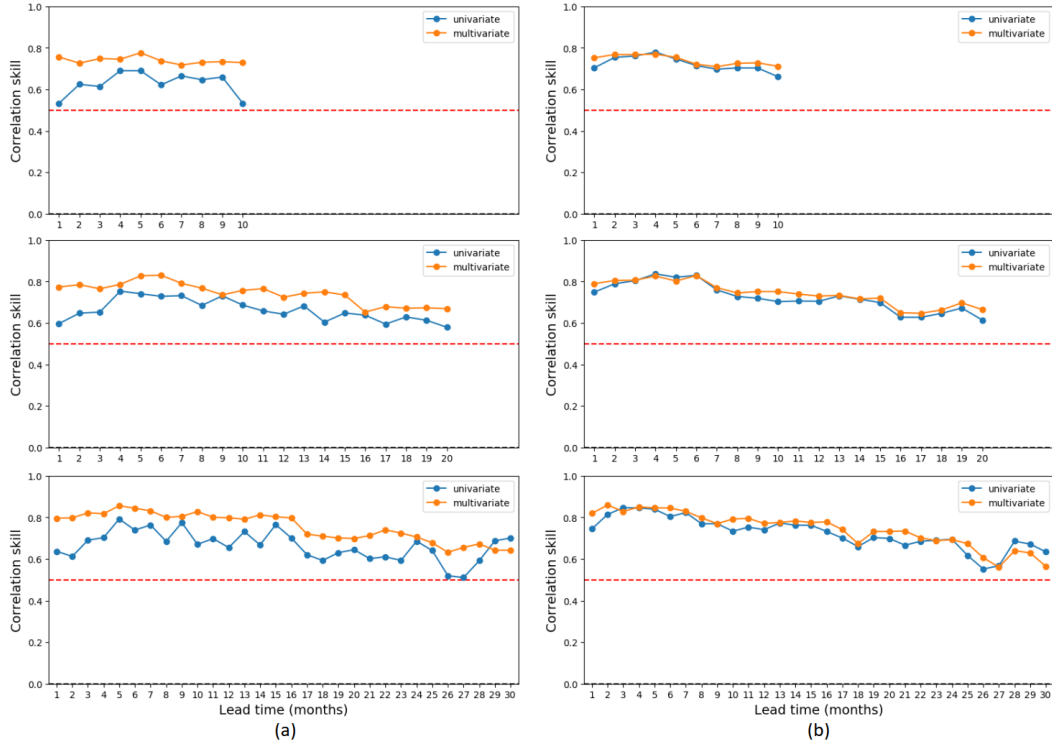
Figure 2: Correlation skills of using the univariate (past ASI) and multivariate (past ASI plus Niño indices) predictors as inputs of the LSTM-based model, (a) the length of input sequences is 3; (b) the length of input sequences is 6.

tionships between them before constructing a forecast model. More specifically, we calculate the Pearson correlation coefficients between wintertime ASI across China and autumn Niño indices. Figure 1 (b) shows the correlation coefficient map where black dots mean statistically significant relationships. It is observed that autumn Niño 3.4 index shows strong negative correlations with wintertime ASI in southern China. The negative correlations between Niño 3.4 index and air pollution in southern China have also been discussed in (Zhao et al., 2018; Cheng et al., 2019). Therefore, we select the statistically significant region, i.e., seven provinces in southern China, including Guizhou, Hunan, Jiangxi, Fujian, Zhejiang, Guangxi, and Guangdong. Then we calculate the average monthly ASI of this region and obtain the spatially average wintertime ASI time-series data. The Pearson correlation coefficients between this ASI time-series data and autumn Niño 1+2, Niño 3, Niño 3.4, and Niño 4 are -0.42, -0.50, -0.46, and -0.53, respectively. These correlations suggest that making seasonal or even interannual predictions of wintertime ASI is possible.

Model and Experimental Setup   We develop a forecast model by exploiting the LSTM cell which is good at capturing the long-term dependency of time-series data (Hochreiter & Schmidhuber, 1997). This LSTM-based model is composed of one input layer, two-layer LSTMs (each LSTM layer with 32 hidden states), and one output layer. To explore the effects of September-October-November (SON) Niño indices on foreseeing future December–January–February (DJF) ASI over southern China, we make comparisons between univariate input and multivariate inputs. The univariate input refers to past DJF ASI, while multivariate inputs refer to past DJF ASI plus SON Niño indices. The input sequence length is denoted as k, which means the input sequence from time $\tau - k + 1$ months to time τ (in months). The output layer is a fully connected (FC) layer whose variable is the future ASI at time t ($t >= \tau + 1$). The training period is from 1950 to 1999 used to train the LSTM model and the period for validating the forecast skill is from 2000 to 2020. Adam optimizer (Kingma & Ba, 2014) with a learning rate of 0.001 and 0.01 is used for univariate and multivariate inputs respectively, and mean-squared error (MSE) is chosen as our loss function.
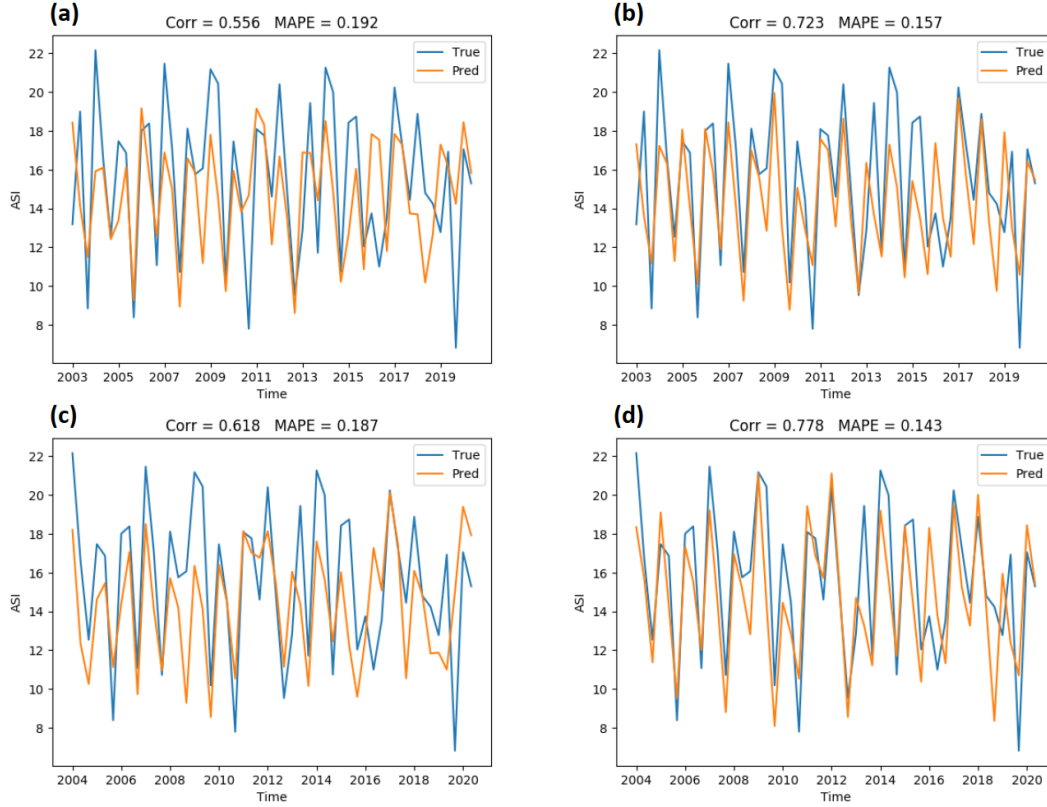
3

Figure 3: Time series of the true and the predicted DJF ASI in southern China: using univariate input with a sequence length of 6 (a) or 9 (c); using multivariate inputs with a sequence length of 6 (b) or 9 (d).

3 EXPERIMENTS

Evaluation To evaluate the forecast skill of the LSTM-based model, we adopt the temporal anomaly correlation coefficient C in (Ham et al., 2019). C is a function of the forecast lead months l and measures the linear correlation between the actual and the predicted values. In addition, we also use Pearson correlation coefficients (Corr) and Mean absolute percentage error (MAPE) to compare the true and the predicted time series during the validating period.

Results Figure 2 shows the correlation skill of the LSTM-based model at lead times of 10, 20, and 30 months (the first, second, and third row) using the univariate (blue line) and multivariate inputs (orange line) with the input sequence length of 3 (a) and 6 (b) respectively. Both curves of univariate and multivariate inputs in Figure 2 show a downward trend along with increasing lead time, which is reasonable. Figure 2 (a) indicates that multivariate inputs achieve significantly better forecast skills than univariate input at lead times of 10, 20, and 30 months (both lines almost above 0.5), while Figure 2 (b) shows the univariate input can achieve an approximate forecast performance to multivariate inputs when the input sequence length is 6. Niño indices have been proven to be beneficial to improve the forecast ability of wintertime ASI when the input sequence length is short.

Furthermore, Figure 3 shows the true and predicted DJF ASI during the validation period using univariate and multivariate inputs with different input sequence lengths. We can observe that multivariate inputs achieve higher correlation coefficients and lower MAPE than univariate input under the same sequence length by comparing Figure 3 (a) and (b), (c) and (d). Finally, multivariate inputs with a sequence length of 9 achieve the best predictive performance with a high correlation coefficient of 0.778 and a low MAPE of 0.143.

4 CONCLUSION AND FUTURE WORK

In this work, we have explored leveraging ocean memory effects to achieve long-lead forecasts of wintertime ASI. We first find negative correlations between autumn Niño indices and wintertime ASI in southern China, indicating the prospects for a prewinter forecast. Based on these correlations, an LSTM-based forecast model has been developed. Experimental results show that Niño indices are beneficial to help improve the forecast performance of wintertime ASI, especially when the input sequence is short. In future work, we will conduct more investigations: (1) try more powerful machine learning models, such as Transformer and ConvLSTMs, to further enhance the prediction accuracy; (2) directly use SST anomalies as predictors which may provide global and more useful information for long-lead forecasts.

REFERENCES

Jassim Al-Saadi, James Szykman, R Bradley Pierce, Chieko Kittaka, Doreen Neil, D Allen Chu, Lorraine Remer, Liam Gumley, Elaine Prins, Lewis Weinstock, et al. Improving national air quality forecasts with satellite aerosol observations. *Bulletin of the American Meteorological Society*, 86(9):1249–1262, 2005.

Lu Bai, Jianzhou Wang, Xuejiao Ma, and Haiyan Lu. Air pollution forecasts: An overview. *International journal of environmental research and public health*, 15(4):780, 2018.

Sangwon Chae, Joonhyeok Shin, Sungjun Kwon, Sangmok Lee, Sungwon Kang, and Donghyun Lee. PM_{10} and $PM_{2.5}$ real-time prediction models using an interpolated convolutional neural network. *Scientific Reports*, 11(1):1–9, 2021.

Yue-Shan Chang, Hsin-Ta Chiao, Satheesh Abimannan, Yo-Ping Huang, Yi-Ting Tsai, and Kuan-Ming Lin. An LSTM-based aggregated model for air pollution forecasting. *Atmospheric Pollution Research*, 11(8):1451–1463, 2020.

Yarong Chen, Shuhang Cui, Panyi Chen, Qiangqiang Yuan, Ping Kang, and Liye Zhu. An LSTM-based neural network method of particulate pollution forecast in China. *Environmental Research Letters*, 16(4):044006, 2021.

Xugeng Cheng, Richard Boiyo, Tianliang Zhao, Xiangde Xu, Sunling Gong, Xiaoning Xie, and Ke Shang. Climate modulation of niño3.4 SST-anomalies on air quality change in Southern China: Application to seasonal forecast of haze pollution. *Atmospheric Research*, 225:157–164, 2019.

Meng Gao, Peter Sherman, Shaojie Song, Yueyue Yu, Zhiwei Wu, and Michael B McElroy. Seasonal prediction of indian wintertime aerosol pollution using the ocean memory effect. *Science advances*, 5(7):eaav4157, 2019.

Jose M Garrido-Perez, Ricardo García-Herrera, and Carlos Ordóñez. Assessing the value of air stagnation indices to reproduce pm10 variability in europe. *Atmospheric Research*, 248:105258, 2021.

Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo. Deep learning for multi-year ENSO forecasts. *Nature*, 573(7775):568–572, 2019.

Yoo-Geun Ham, Jeong-Hwan Kim, Eun-Sol Kim, and Kyoung-Woon On. Unified deep learning model for el niño/southern oscillation forecasts by incorporating seasonality in climate data. *Science Bulletin*, 66(13):1358–1366, 2021.

KS Harishkumar, KM Yogesh, Ibrahim Gad, et al. Forecasting air pollution particulate matter ($PM_{2.5}$) using machine learning regression models. *Procedia Computer Science*, 171:2057–2066, 2020.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, and et al Horányi. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Daniel E Horton, Noah S Diffenbaugh, et al. Response of air stagnation frequency to anthropogenically enhanced radiative forcing. *Environmental Research Letters*, 7(4):044034, 2012.

Daniel E Horton, Christopher B Skinner, Deepti Singh, and Noah S Diffenbaugh. Occurrence and persistence of future atmospheric stagnation events. *Nature climate change*, 4(8):698–703, 2014.

Q. Huang, X. Cai, J. Wang, Y. Song, and T. Zhu. Climatological study of the boundary-layer air stagnation index for china and its relationship with air pollution. *Atmospheric Chemistry and Physics*, 18(10):7573–7593, 2018.

Jaein I Jeong, Rokjin J Park, Sang-Wook Yeh, and Joon-Woo Roh. Statistical predictability of wintertime $PM_{2.5}$ concentrations over East Asia using simple linear regression. *Science of The Total Environment*, 776:146059, 2021.

Marilena Kampa and Elias Castanas. Human health effects of air pollution. *Environmental pollution*, 151(2):362–367, 2008.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Atakan Kurt and Ayşe Betül Oktay. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Systems with Applications*, 37(12):7986–7992, 2010.

Can Li, N Christina Hsu, and Si-Chee Tsay. A study on the potential applications of satellite data in air quality monitoring and forecasting. *Atmospheric Environment*, 45(22):3663–3675, 2011.

Hong Liao, Wei-Ting Chen, and John H Seinfeld. Role of climate change in global predictions of future tropospheric ozone and aerosols. *Journal of Geophysical Research: Atmospheres*, 111 (D12), 2006.

Adil Masood and Kafeel Ahmad. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production*, 322:129072, 2021.

Ping Wu, Yihui Ding, and Yanju Liu. Atmospheric circulation and dynamic mechanism for persistent haze events in the beijing–tianjin–hebei region. *Advances in Atmospheric Sciences*, 34(4): 429–440, 2017.

Qunli Wu and Huaxing Lin. Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network. *Sustainable Cities and Society*, 50: 101657, 2019.

Fei Xiao, Mei Yang, Hong Fan, Guanghui Fan, and Mohammed AA Al-Qaness. An improved deep learning model for predicting daily $PM_{2.5}$ concentration. *Scientific reports*, 10(1):1–11, 2020.

Qi Zhang, Jacqueline CK Lam, Victor OK Li, and Yang Han. Deep-air: A hybrid CNN-LSTM framework for fine-grained air pollution forecast. *arXiv preprint arXiv:2001.11957*, 2020.

S. Zhao, H. Zhang, and B. Xie. The effects of El Niño–Southern Oscillation on the winter haze pollution of China. *Atmospheric Chemistry and Physics*, 18(3):1863–1877, 2018.