Supplementary for

A 23-Year Nationwide Study Revealing Aerosol-Driven Light Rain Shifts in China's Emission Control Era

Rou Zhang[1], Xiaoxiao Huang[1], Pu Wang[1], Guiquan Liu[1], Mengyu Liu[1], Songjian Zou[1], Lu Chen[2], Fang Zhang[1*]

[1]*School of Ecology and Environment, College of Artificial Intelligence, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China*

[2]*School of Urban and Planning, Yancheng Teachers University, Yancheng, 224051, China*

Correspondence to F. Zhang (zhangfang2021@hit.edu.cn)

This PDF file includes:

Method descriptions

Figs. S1 to S17

References

Machine learning methods

The XGBoost (eXtreme Gradient Boosting) model is an advanced machine learning algorithm that has gained significant popularity and achieved state-of-the-art results in various predictive modeling tasks (Chen and Guestrin, 2016). It belongs to the family of gradient boosting algorithms and is known for its efficiency, flexibility, and high performance. XGBoost is designed to handle both classification and regression problems. It works by sequentially adding weak prediction models, typically decision trees, to an ensemble in a process known as boosting. Each subsequent model is built to correct the mistakes made by the previous models, gradually improving the overall predictive accuracy. What sets XGBoost apart is its focus on optimization and regularization techniques. It incorporates a regularized objective function that combines a loss function and a penalty term to control model complexity and prevent overfitting (Gui et al., 2020; Si and Du, 2020; Wong et al., 2021).

Figures

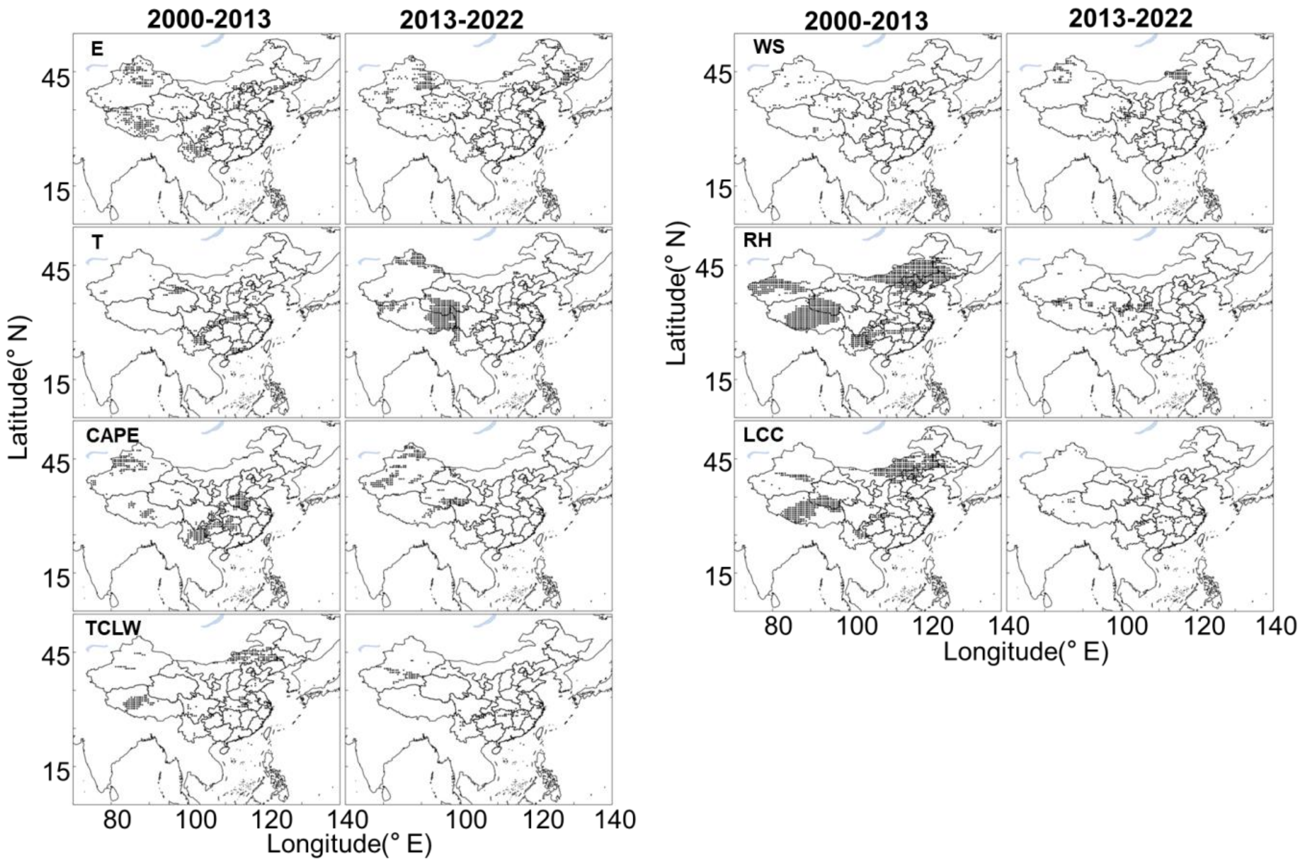

Fig. S1. The significance of the fitted variation trends of other influencing factors in China during the two periods of 2000 - 2013 and 2013 - 2022 (black dots indicate passing the 95% significance test).

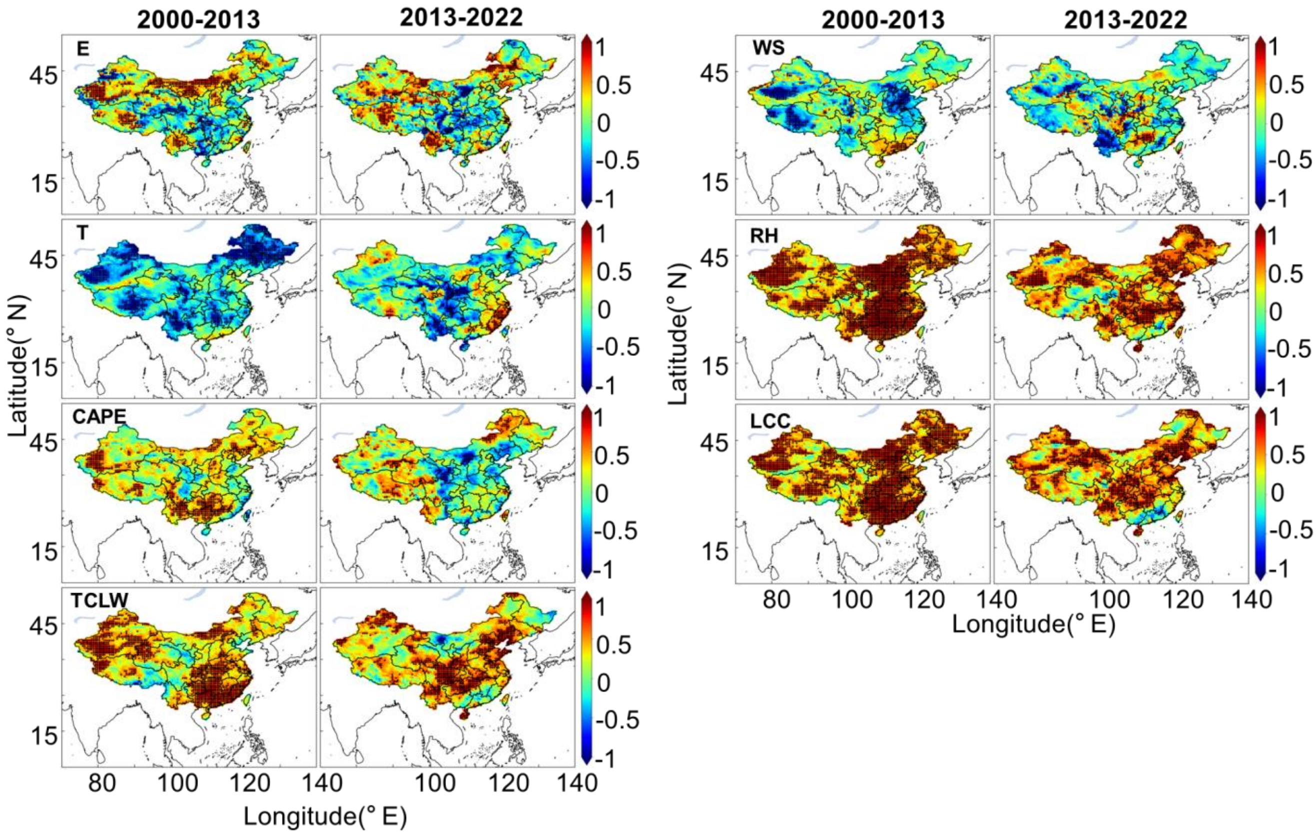

Fig. S2. correlation of between Meteorological factors and light rain frequency in 2000 - 2013 and 2013 - 2022.

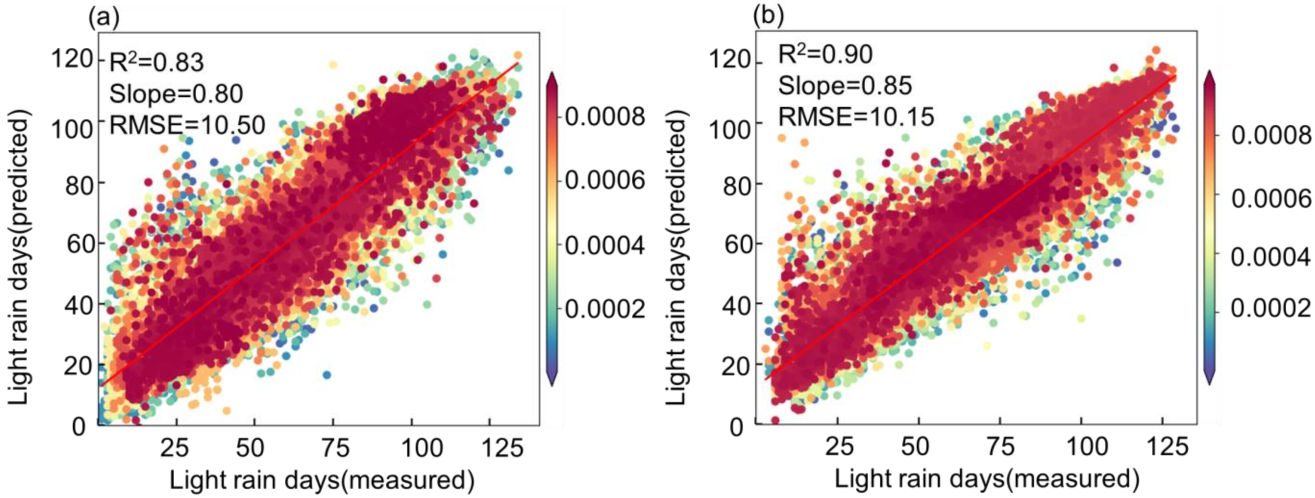

Fig. S3. The predicted values of light rain days during **(a)** 2000 - 2013 and **(b)** 2013 - 2022 by the XGBoost method.

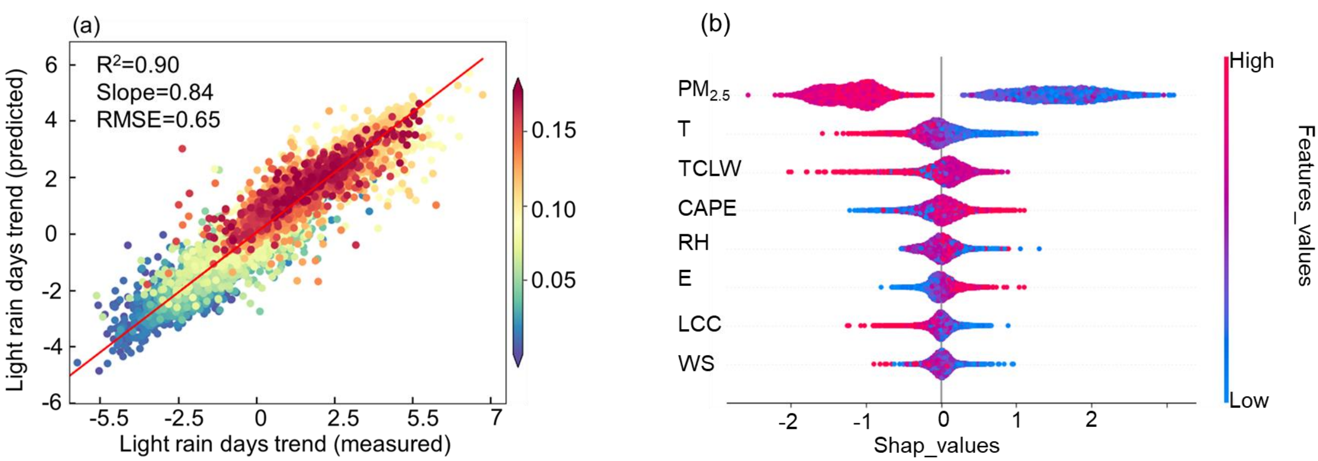

Fig. S4. The predicted values of light rain days trend by the XGBoost method (a) and SHAP summary plot (b).

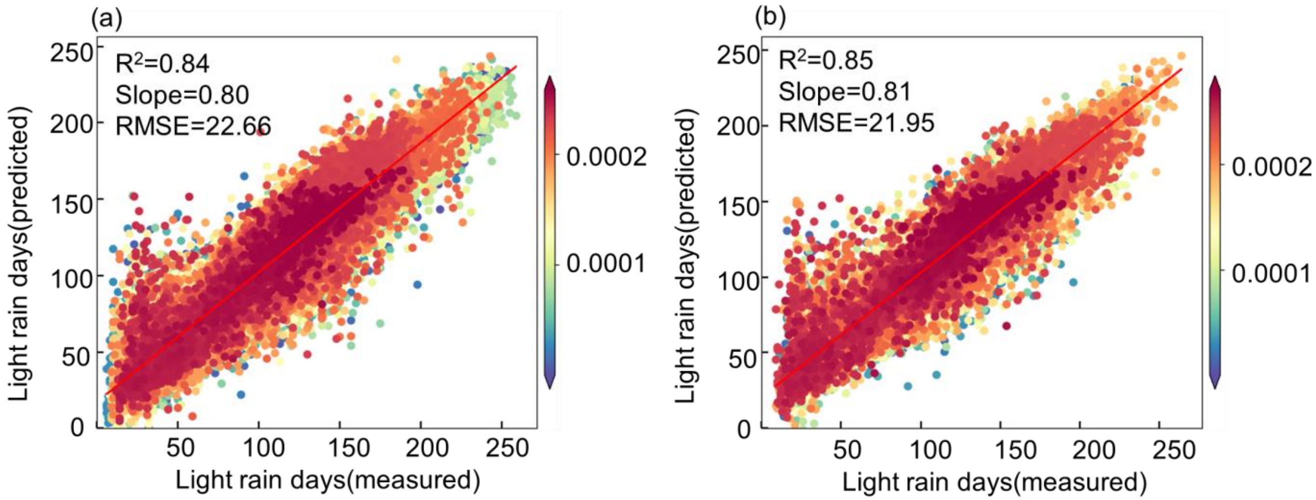

Fig. S5. The predicted values of light rain days in the warm season (Jun.-Oct.) during **(a)** 2000 - 2013 and **(b)** 2013 - 2022 by the XGBoost method.

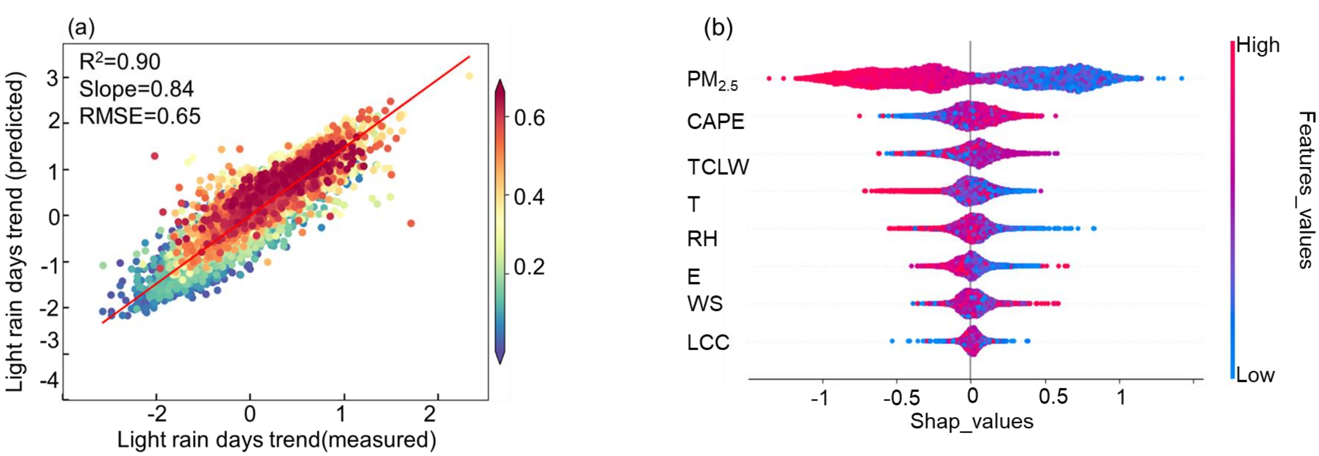

Fig. S6. The predicted values of light rain days trend in the warm season (Jun.-Oct.) by the XGBoost method **(a)** and SHAP summary plot **(b)**.

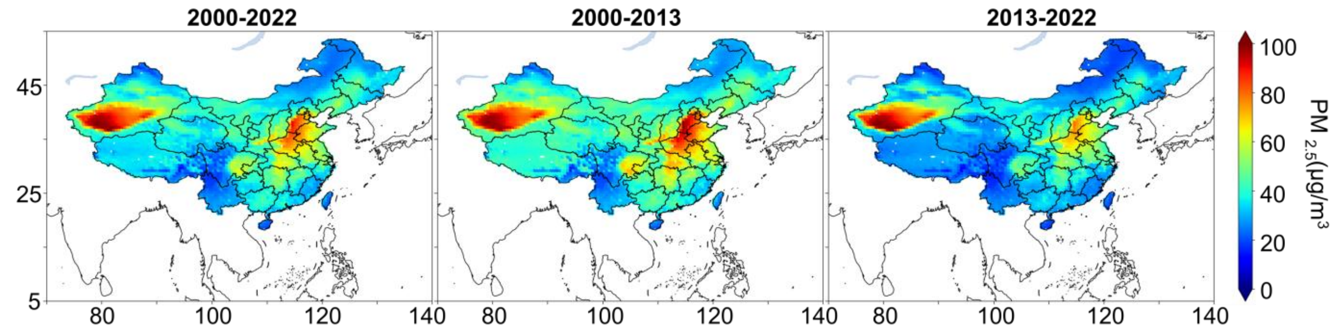

Fig. S7. Spatial distribution of average $PM_{2.5}$ in 2000 - 2022,2000 - 2013 and 2013 - 2022 respectively.

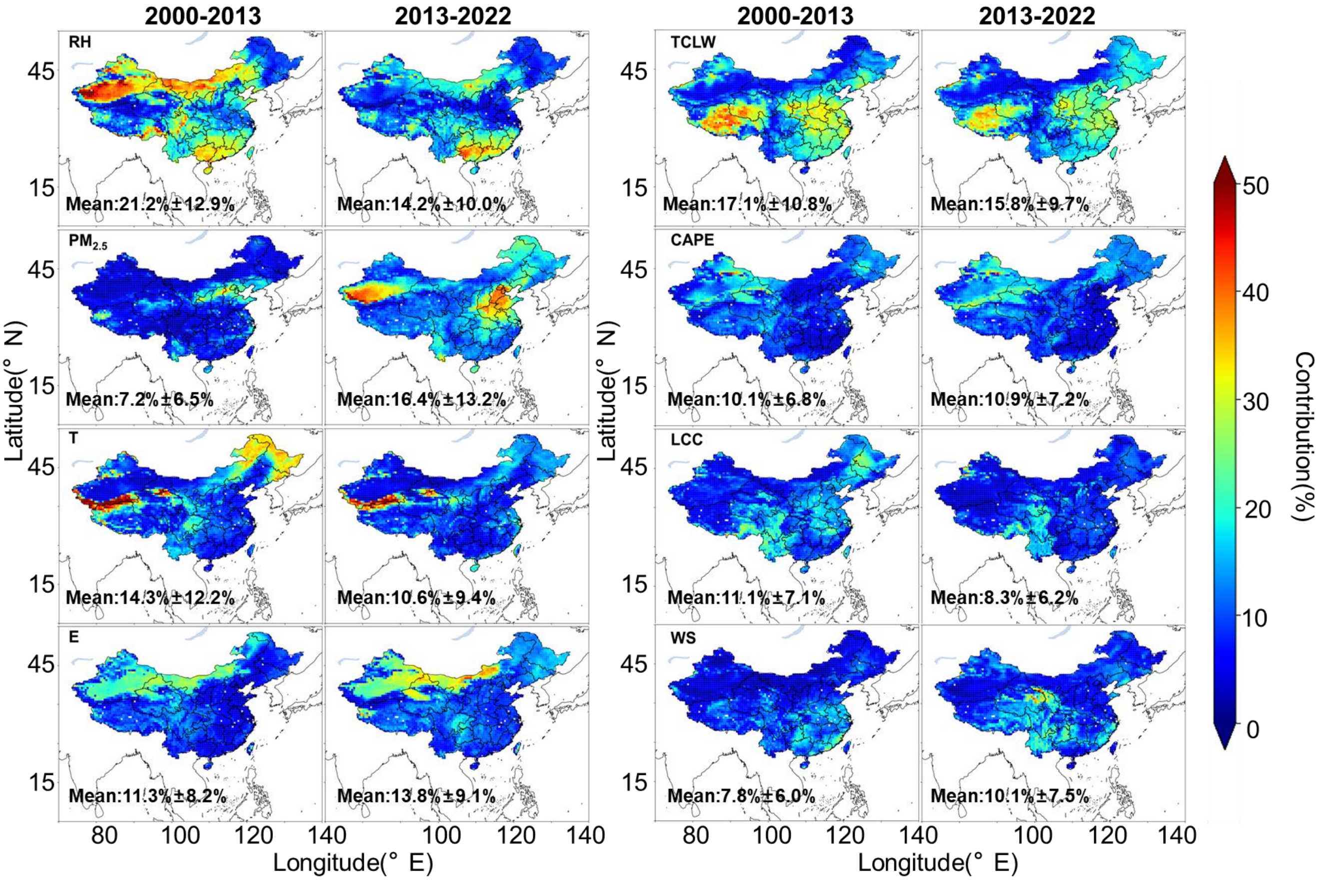

Fig. S8. The relative contribution of relative humidity (RH), $PM_{2.5}$ mass concentration, temperature (T), evaporation (E), total column liquid water (TCLW), CAPE, low cloud cover (LCC) and wind speed (WS) to light rain days in the warm season during 2000 - 2013 and 2013 - 2022.

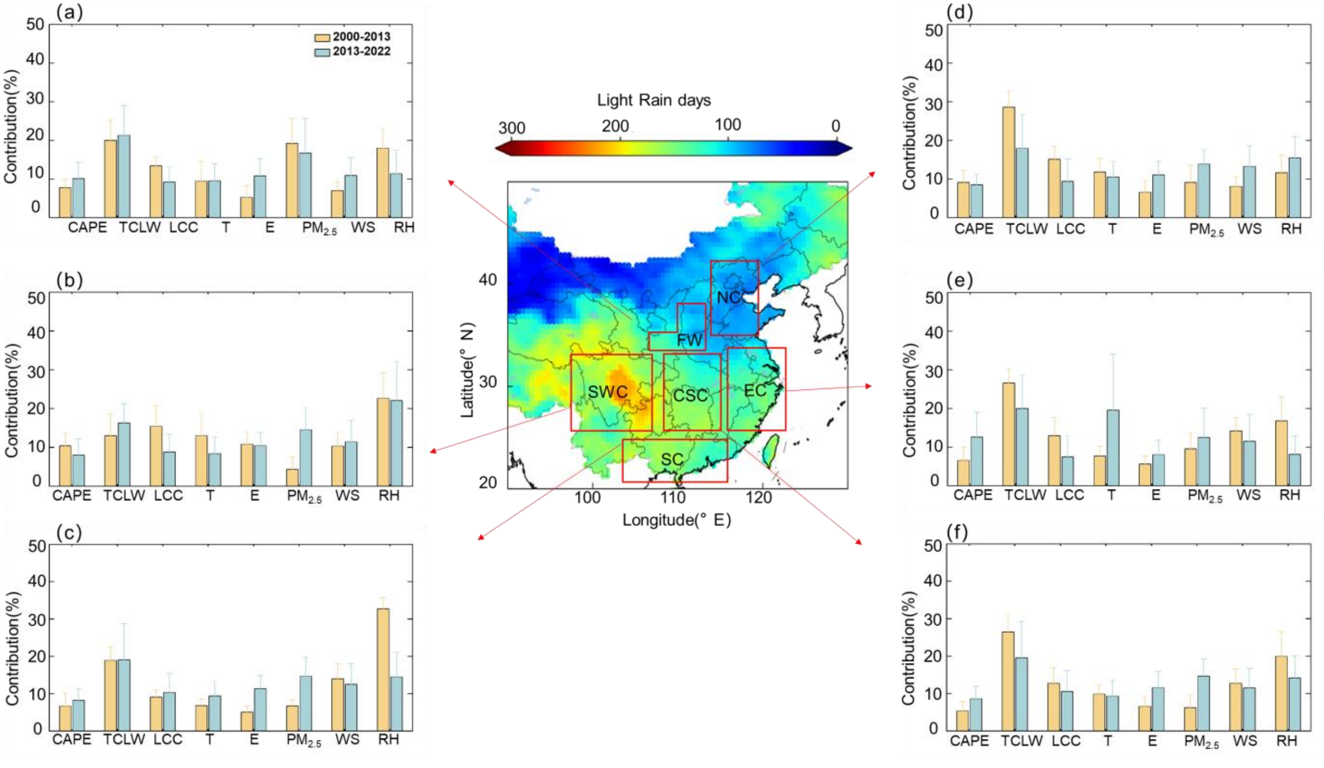

Fig. S9. Comparison of the contribution of individual factors to light rain days in the warm season over 2000 - 2013 and 2013 - 2022 in the selected six regions of China.

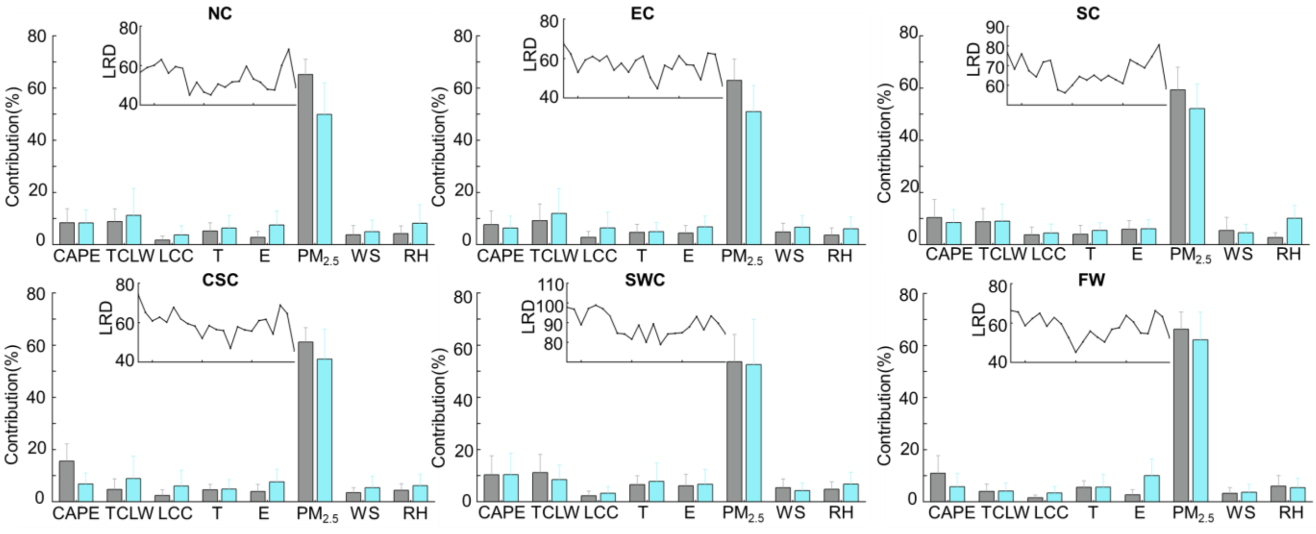

Fig. S10. Contribution of variations of each individual factor to the long-term trends of light rain days of warm season in the selected six regions of China. The small graphs embedded in the middle represent the average trends of light rain days of warm season over 2000 - 2022 in the six regions.

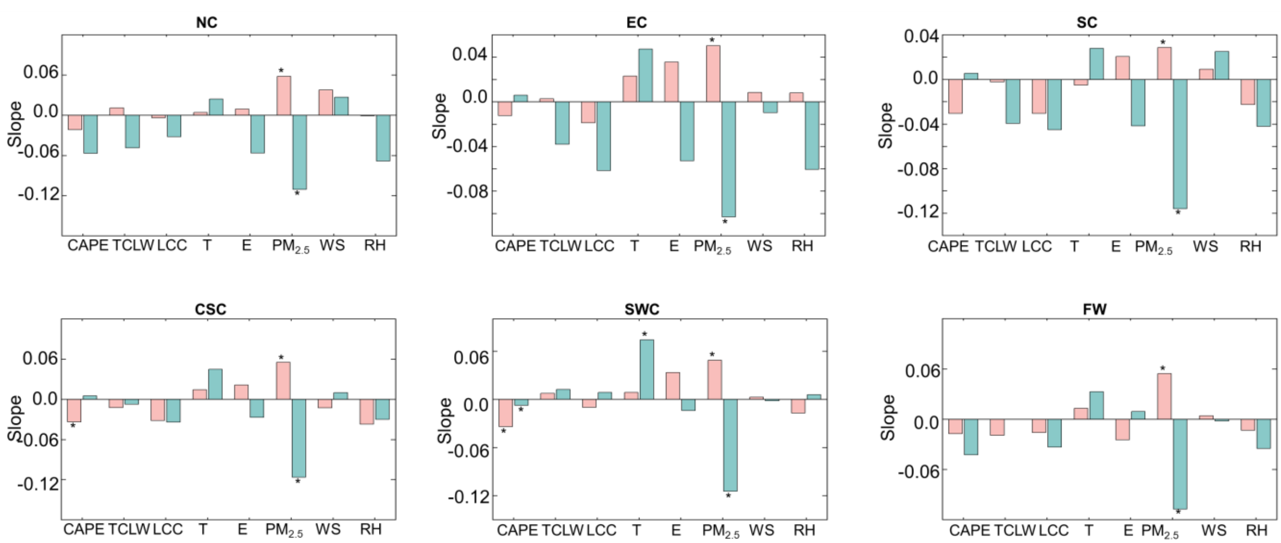

Fig. S11. The long-term trends of each individual factor of warm season in the selected six regions of China. The black asterisk (*) represent the trend pass the 95% significance test.

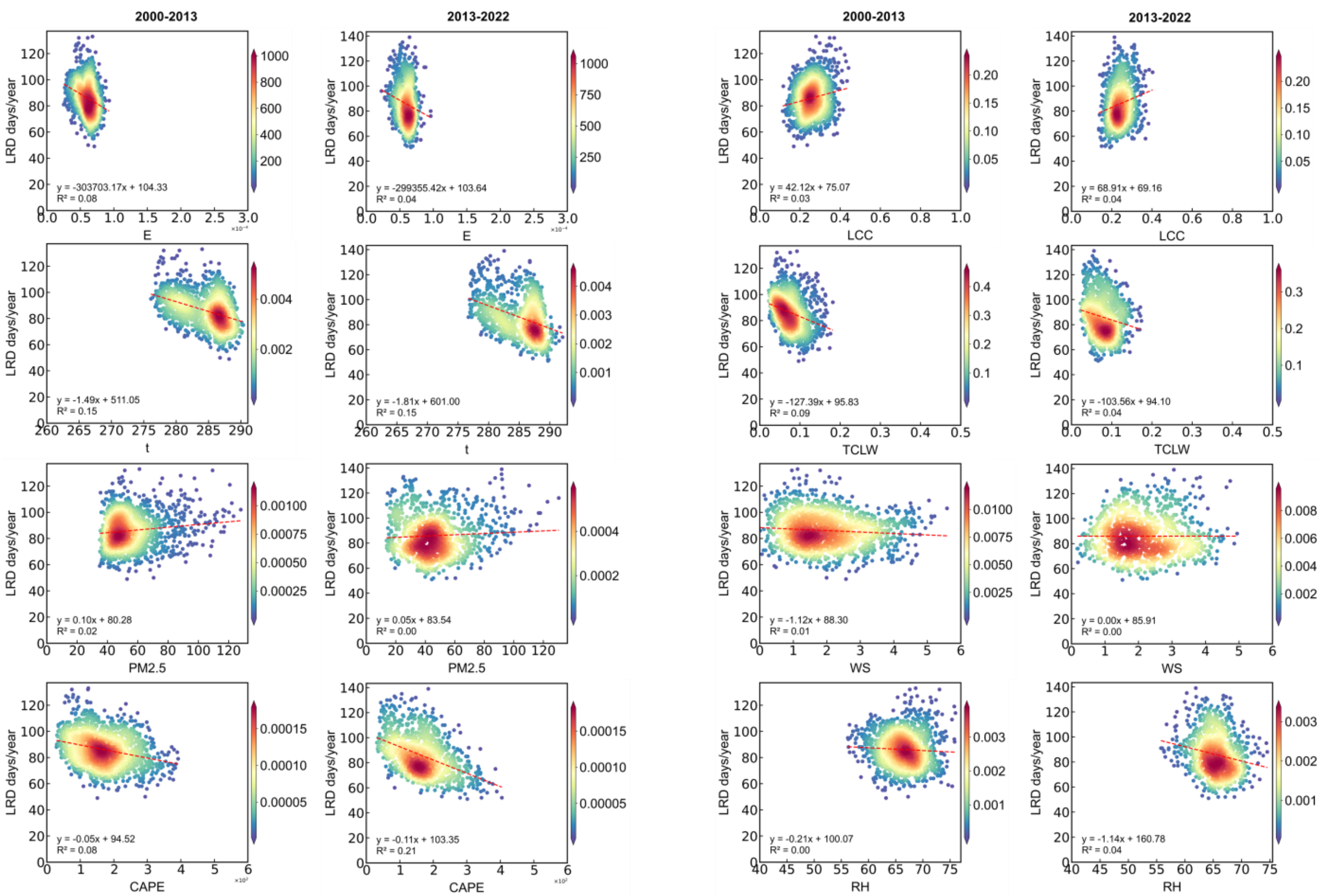

Fig. S12. Scatter plots showing the relationships between various factors and the number of light rain days in NC.

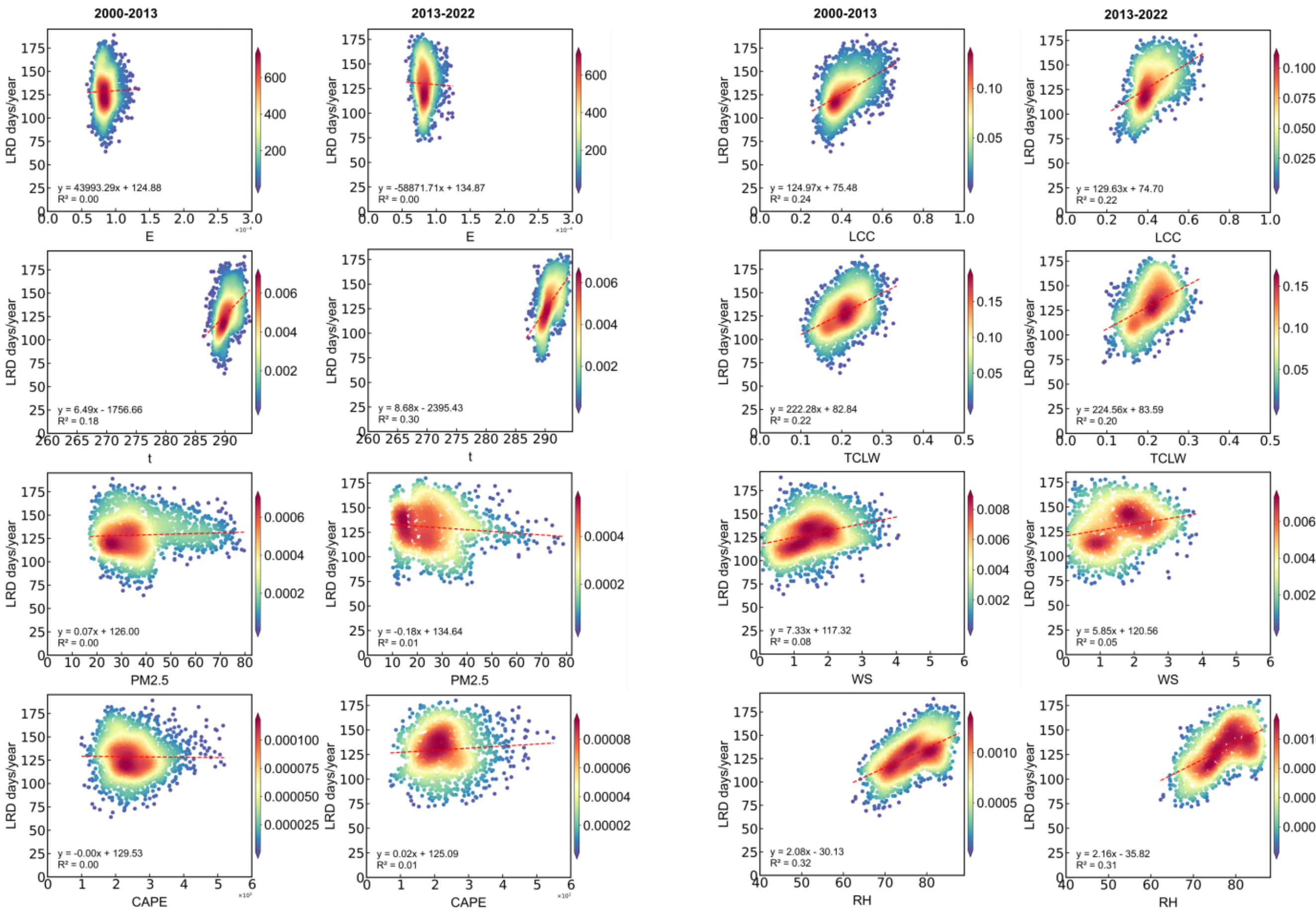

Fig. S13. Scatter plots showing the relationships between various factors and the number of light rain days in EC.

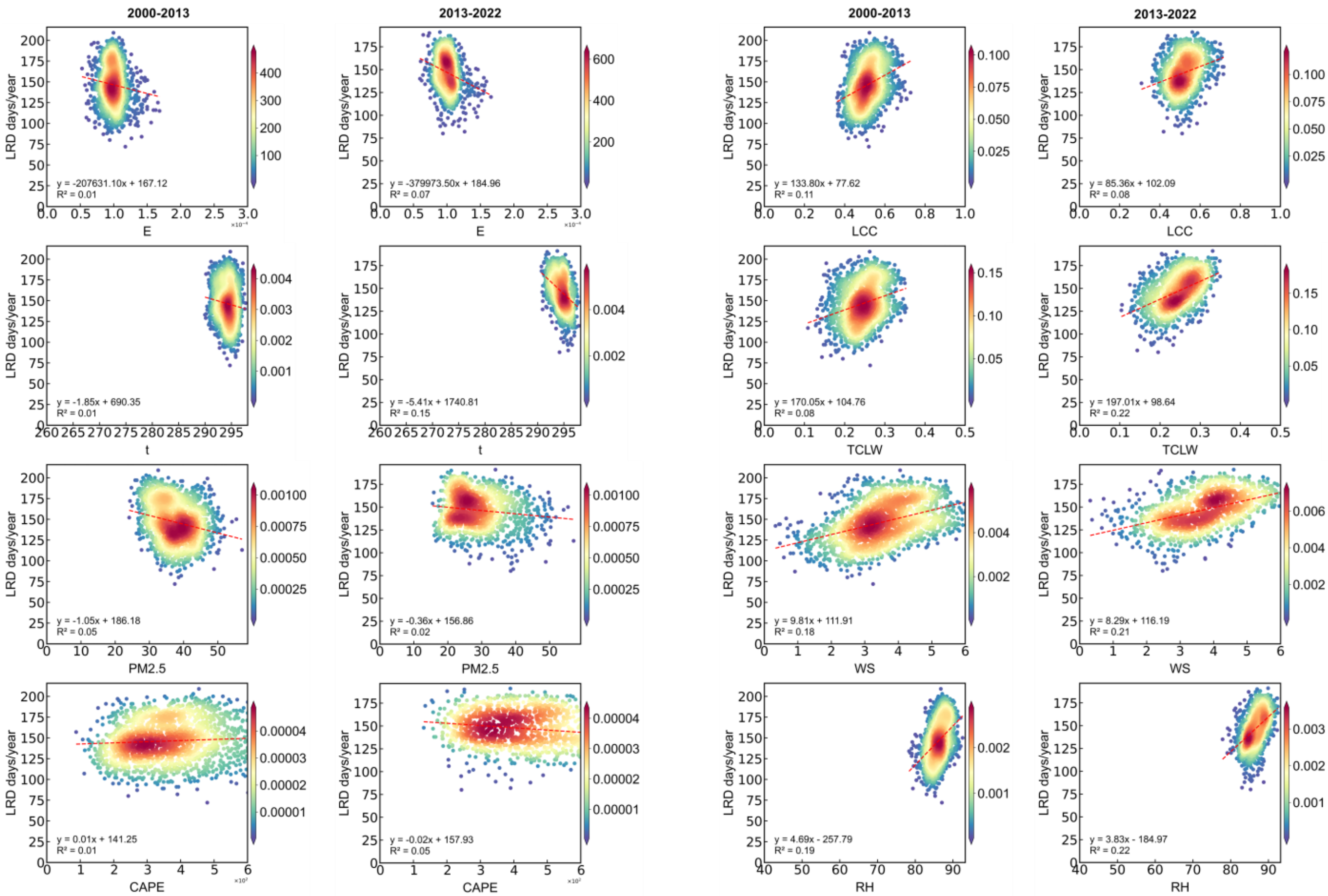

Fig. S14. Scatter plots showing the relationships between various factors and the number of light rain days in SC.

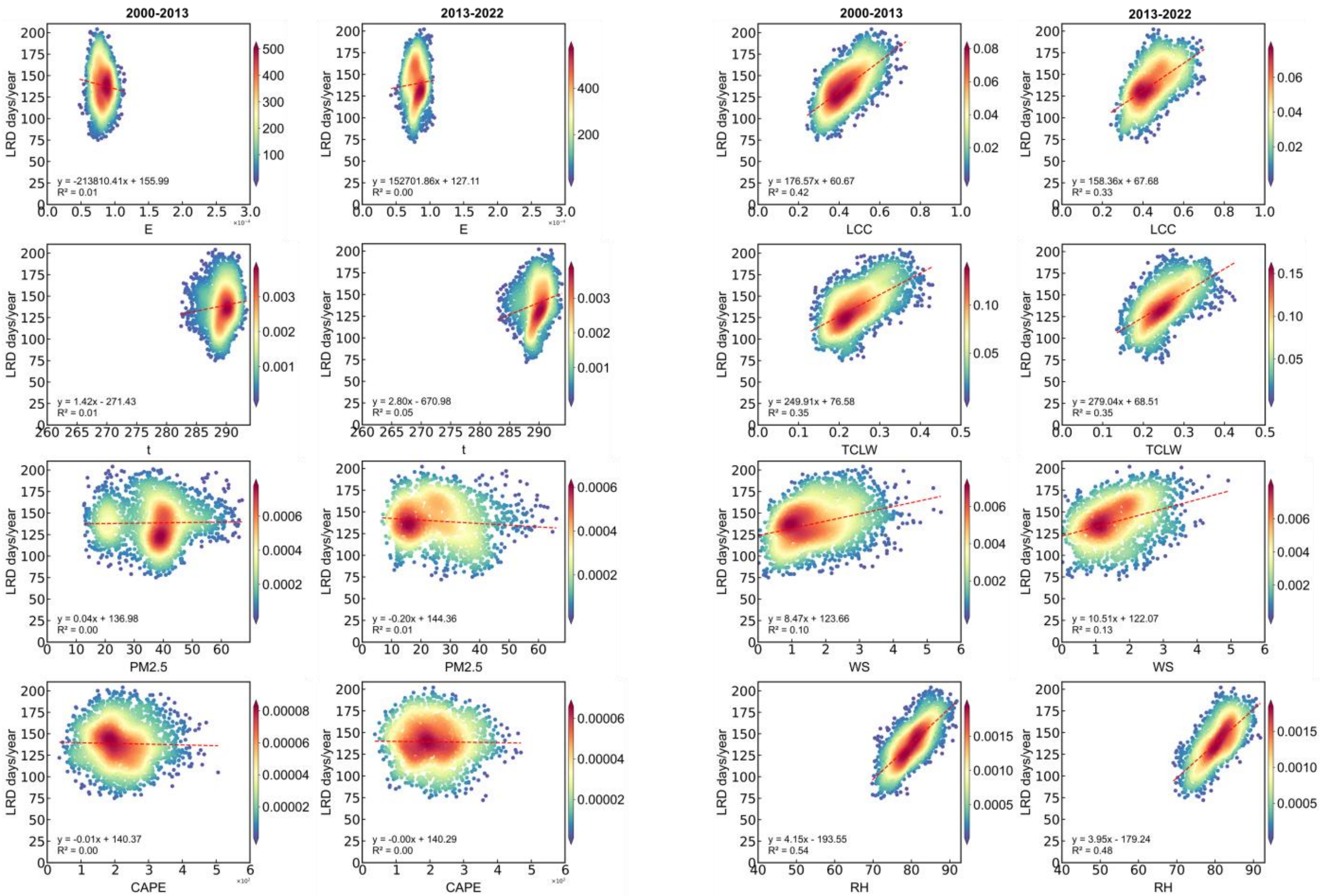

Fig. S15. Scatter plots showing the relationships between various factors and the number of light rain days in CSC.

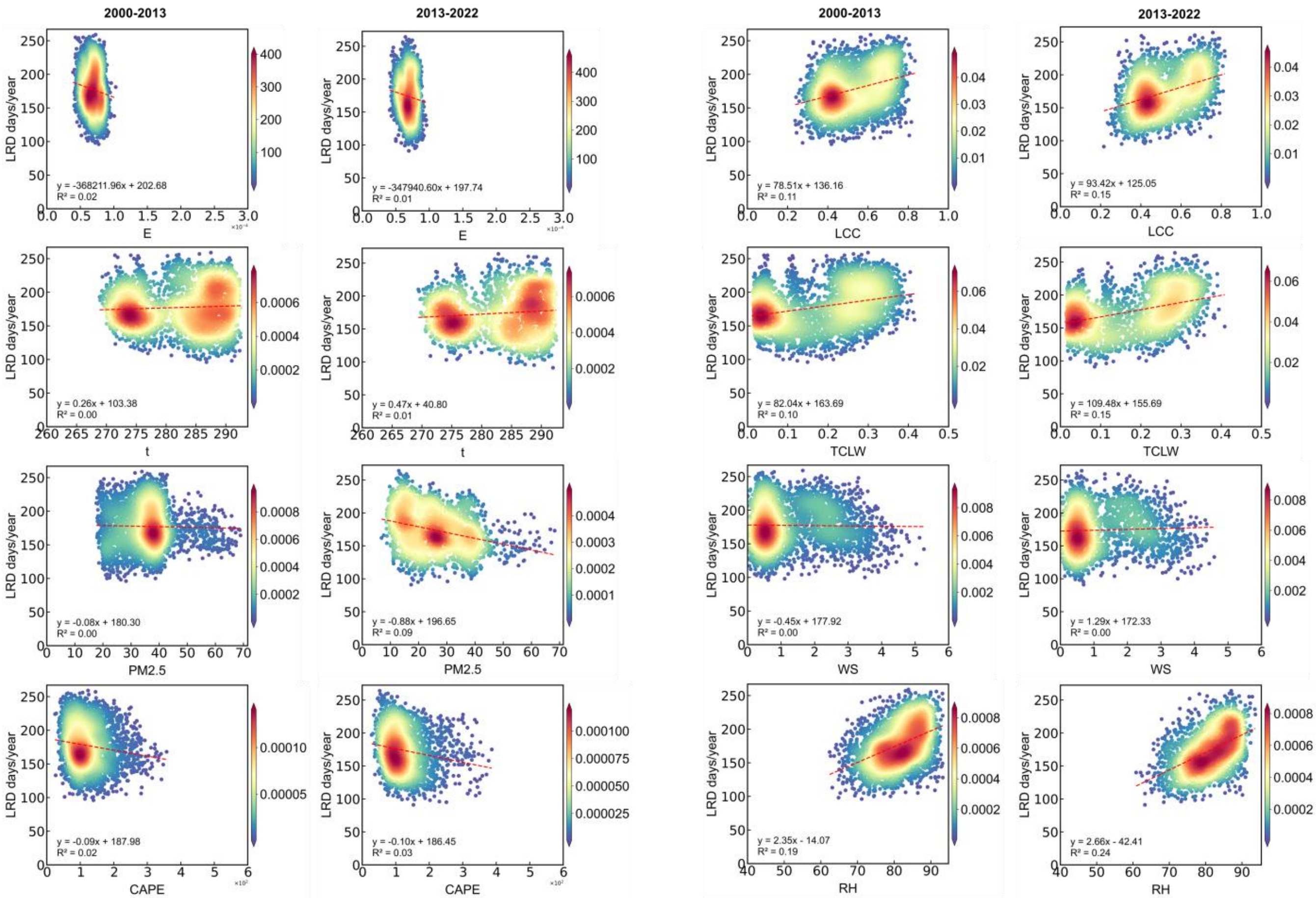

Fig. S16. Scatter plots showing the relationships between various factors and the number of light rain days in SWC.

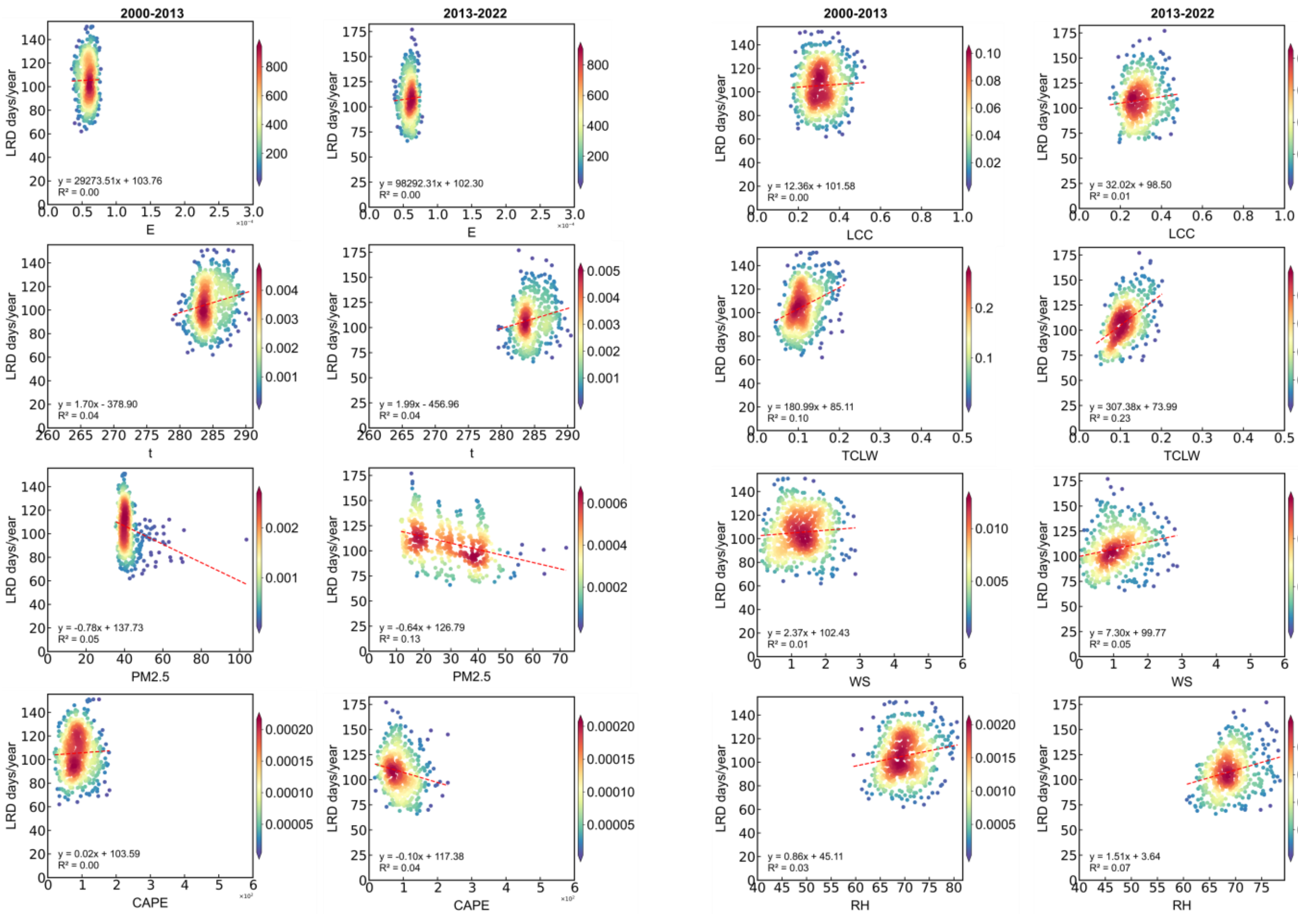

Fig. S17. Scatter plots showing the relationships between various factors and the number of light rain days in FW.

References

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785

Gui, K., Che, H., Zeng, Z., Wang, Y., Zhai, S., Wang, Z., Luo, M., Zhang, L., Liao, T., Zhao, H., Li, L., Zheng, Y., Zhang, X., 2020. Construction of a virtual PM2.5 observation network in China based on high-density surface meteorological observations using the Extreme Gradient Boosting model. Environment International 141, 105801. https://doi.org/10.1016/j.envint.2020.105801

Si, M., Du, K., 2020. Development of a predictive emissions model using a gradient boosting machine learning method.

Environmental Technology & Innovation 20, 101028. https://doi.org/10.1016/j.eti.2020.101028

Wong, P., Lee, H.., Chen, Y., Zeng, Y., Chern, Y., Chen, N., Candice Lung, S.C., Su, H., Wu, C., 2021. Using a land use regression model with machine learning to estimate ground level PM2.5. Environmental Pollution 277, 116846. https://doi.org/10.1016/j.envpol.2021.116846