

How **BLUE** is the Sky? Estimating the Air Quality Data in Beijing During the Blue Sky Day Period (2008-2012) by the Bayesian LSTM Approach

EPRG Working Paper 1912

Cambridge Working Paper in Economics 1929

Yang Han, Victor OK Li, Jacqueline CK Lam, and Michael Pollitt

Abstract Over the last three decades, air pollution has become a major environmental challenge in many of the fast growing cities in China, including Beijing. Given that any long-term exposure to high-levels of air pollution has devastating health consequences, accurately monitoring and reporting air pollution information to the public is critical for ensuring public health and safety and facilitating rigorous air pollution and health-related scientific research. Recent statistical research examining China's air quality data has posed questions regarding data accuracy, especially data reported during the Blue Sky Day (BSD) period (2000 – 2012), though the accuracy of publicly available air quality data in China has improved gradually over the recent years (2013 – 2017). To the best of our understanding, no attempt has been made to re-estimate the air quality data during the BSD period. In this paper, we put forward a machine-learning model to re-estimate the official air quality data during the BSD period of 2008 – 2012, based on the PM_{2.5} data of the Beijing US Embassy, and the proxy data covering Aerosol Optical Depth (AOD) and meteorology. Results have shown that the average re-estimated daily air quality values are respectively 64% and 61% higher than the official values, for air quality index (AQI) and AQI equivalent PM_{2.5}, during the BSD period of 2008 to 2012. Moreover, the re-estimated BSD air quality data exhibit reduced statistical discontinuity and irregularity, based on our validation tests. The results suggest that the proposed data re-estimation methodology has the potential to provide more justifiable historical air quality data for evidence-based environmental decision-making in China.

Keywords Blue Sky Day (BSD), Air Quality, Beijing, Data Irregularity, Bayesian LSTM, Data Estimation

JEL Classification C53, C63, Q53

Contact yhan@eee.hku.hk ; vli@eee.hku.hk ; jcklam@eee.hku.hk

Publication March 2019

Financial Support This research is supported in part by the Theme-based Research Scheme of the Research Grants Council of Hong Kong, under Grant No. T41-709/17-N, and by the Seed Fund for Basic Research from the University of Hong Kong, under Grant No. 201611159182.