

A new method to quickly identify outliers in air quality monitoring data

October 30 2018



The PM2.5 monitoring instruments at State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry (LAPC), Institute of Atmospheric Physics, Chinese Academy of Sciences. Credit: TANG Xiao

Ambient air quality monitoring data comprise the most important source for public awareness of air quality, and are widely used in many research fields, such as improving air quality forecasting and the analysis of haze episodes. However, there are outliers among such monitoring data, due to instrument malfunctions, the influence of harsh environments, and the limitation of measuring methods.

In practice, manual inspection is often applied to identify these outliers. However, as the amount of data grows rapidly, this [method](#) becomes increasingly cumbersome.

To deal with the problem, Dr. Wu Huangjian and Associate Professor Tang Xiao from the Institute of Atmospheric Physics, Chinese Academy of Sciences, propose a fully automatic outlier detection method based on the [probability](#) of residuals. The method adopts multiple regression methods, and the regression residuals are used to discriminate outliers. Based on the standard deviations of the residuals, probabilities of the residuals can be calculated, and the observations with small probabilities are tagged as outliers and removed by a computer program. Their findings are published in *Advances in Atmospheric Sciences*.

"By introducing the probabilities of residuals, multiple rules can be used for identifying outliers on the same framework," says Dr. Wu. "For example, by assuming that the residuals of spatial regression and temporal regression obey a bivariate normal distribution, spatial and temporal consistencies can be simultaneously evaluated for better

identification of outliers".

The method can flag potentially erroneous data in the hourly observations from 1436 stations of the China National Environmental Monitoring Center (CNEMC) within a minute. Indeed, it has been used in CNEMC's [air quality](#) forecasting system, and is going to be integrated into the data management system. The hope is that outliers in the system's real-time air [quality](#) data will be removed in the near future.

The method is published in *Advances in Atmospheric Sciences*.

More information: Huangjian Wu et al, Probabilistic Automatic Outlier Detection for Surface Air Quality Measurements from the China National Environmental Monitoring Network, *Advances in Atmospheric Sciences* (2018). [DOI: 10.1007/s00376-018-8067-9](https://doi.org/10.1007/s00376-018-8067-9)

Provided by Chinese Academy of Sciences

Citation: A new method to quickly identify outliers in air quality monitoring data (2018, October 30) retrieved 11 May 2024 from

<https://phys.org/news/2018-10-method-quickly-outliers-air-quality.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--