

EGU2020-18471

<https://doi.org/10.5194/egusphere-egu2020-18471>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Anomaly Detection by STL Decomposition and Extended Isolation Forest on Environmental Univariate Time Series

İsmail Sezen¹, Alper Unal¹, and Ali Deniz²

¹Istanbul Technical University, Eurasia Institute of Earth Sciences, Istanbul, Turkey

²Istanbul Technical University, Institute of Science and Technology, Istanbul, Turkey

Atmospheric pollution is one of the primary problems and high concentration levels are critical for human health and environment. This requires to study causes of unusual high concentration levels which do not conform to the expected behavior of the pollutant but it is not always easy to decide which levels are unusual, especially, when data is big and has complex structure. A visual inspection is subjective in most cases and a proper anomaly detection method should be used. Anomaly detection has been widely used in diverse research areas, but most of them have been developed for certain application domains. It also might not be always a good idea to identify anomalies by using data from near measurement sites because of spatio-temporal complexity of the pollutant. That's why, it's required to use a method which estimates anomalies from univariate time series data.

This work suggests a framework based on STL decomposition and extended isolation forest (EIF), which is a machine learning algorithm, to identify anomalies for univariate time series which has trend, multi-seasonality and seasonal variation. Main advantage of EIF method is that it defines anomalies by a score value.

In this study, a multi-seasonal STL decomposition has been applied on a univariate PM10 time series to remove trend and seasonal parts but STL is not resourceful to remove seasonal variation from the data. The remainder part still has 24 hours and yearly variation. To remove the variation, hourly and annual inter-quartile ranges (IQR) are calculated and data is standardized by dividing each value to corresponding IQR value. This process ensures removing seasonality in variation and the resulting data is processed by EIF to decide which values are anomaly by an objective criterion.