Explosive Behavior Detection of PM 2.5 During Wildfire Period Based on BSADF Test

K. Sawangsawai, P. Vatiwutipong*

KVIS

APCYS 2022

Introduction

Time series is a type of data that is popular in statistical analysis nowadays. Explosive behavior, a region of data which skyrockets in intensity compared to others, is one important feature which can occur in time series data. A tool used for analysis of explosive behavior in the last few years is the Backward Supremum Augmented Dickey-Fuller Test (BSADF Test). BSADF was developed for use in the stock market, but it has the potential to be utilized for analyzing the burst release of PM2.5 particulates in wildfire situations. However, BSADF is not fully capable of analyzing the explosive behavior of PM2.5 over very short periods of time. This problem leads to the development of a method based on BSADF to detect explosive behavior of PM2.5 release in a short period of time series, with the aim that this new method can be applied to other various data types.

Methodology

1. Analysis of real PM2.5 data using BSADF

The 2018 daily PM2.5 concentration for Butte County, California, United States (image 1), shows an example of exploding behavior which corresponds to a wildfire event (as shown by the red circle). PM2.5 data were analyzed using BSADF with different window sizes, these being either standard, half standard, or one-third of standard window size. Which window size is the least length of data that used for regression analysis of BSADF test. The results were evaluated by confidence values in intervals of 90%, 95%, and 99%.

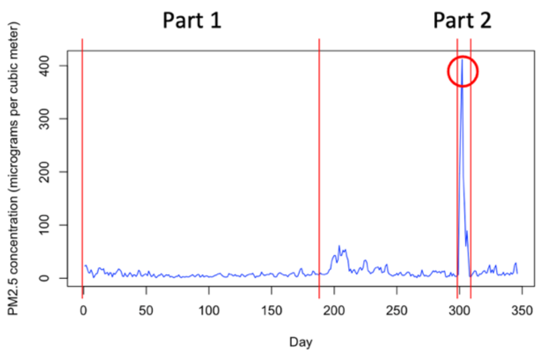

Image 1 - Graph of PM2.5 concentration versus index in 2018 (Butte County, California, USA)

2. Analysis of synthetic PM2.5 data using BSADF

Since few examples of real data such as in image 1 exist, synthetic data sets are needed to evaluate the effect of different window sizes. The real data shown in image 1 was separated into 2 parts, an inexplosive part (part 1) and the explosive region (part 2). The inexplosive part was mimicked by the ARMA process, while the real data was utilized for the explosive part. Three types of synthetic data were created, each with different coefficients of ARMA process that mimicked from different ranges of data, and 1000 sets of each data type were created. Finally, each set of data was analyzed, and the result was evaluated as in step 1 of the Methodology.

3. Analysis of synthetic PM2.5 data using the new test statistic

A new test statistic was created following on from steps 1 and 2 in the Results section. This is represented by the expression $\frac{BSADF_{standard}}{(BSADF_{x(standard)}-BSADF_{standard})+y}$ where $BSADF_{standard}$ represents the data set analyzed by BSADF at a standard window size, with x and y being constants which depend on the data set. Accordingly, x and y need to be varied as part of the Monte Carlo analysis. This was done 10000 times to derive the critical value of the new method and the constant that expressed the optimum model performance. These results were compared with those obtained using the BSADF test in step 2.

4. Analysis of real PM2.5 data using the new test statistic

The real PM2.5 data in image 1 was subjected to analysis using the new test statistic with suitable x and y constants from the data set (step 3). Results were compared with those from the BSADF test in step 1.

Conclusion

The new test statistic derived in this work is very efficient in comparison to the existing BSADF test for the detection of explosive behavior both in synthetic and real data. This may provide a more accurate way of detecting PM2.5 releases over short periods of time, such as in wildfire events, with the benefit that in other periods over-detection of explosive behavior does not occur.

Reference

[1] Wei, W. W. S. (2013). *Time Series Analysis. Oxford Handbooks Online.*

[2] Tsay, R. S. and Tiao, G. C. (1984). Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. Journal of the American Statistical Association, *79(385), 84-96.*

[3] Dickey, D. A. and Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association, 74(366a), 427–431.

[4] Said, S. E. and Dickey, D.A. (1984), Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika 71, 599-608.

[5] Phillips, P. C. B., Wu, Y., and Yu, J. (2011). Explosive Behavior In The 1990s Nasdaq: When Did Exuberance Escalate Asset Values? International Economic Review, 52(1):201–226.

[6] Phillips, P. C. B., Shi, S., and Yu, J. (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review,* 56(4):1043–1078.

Result

1. Analysis of real PM2.5 data by BSADF

As indicated in image 2, the BSADF test cannot be used to detect explosive behavior in a short period of time series of the data, for any window size. Data in image 1 which show explosive behavior (red circle) are observed as being below the critical value in image 2. In addition, when the window size is decreased, other data periods show over-detection, i.e., explosive behavior, as indicated in image 3.

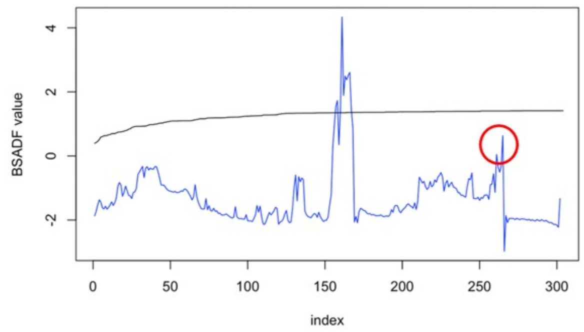

Image 2 - Graph of BSADF Test versus index for a standard window size, 95% confidence interval

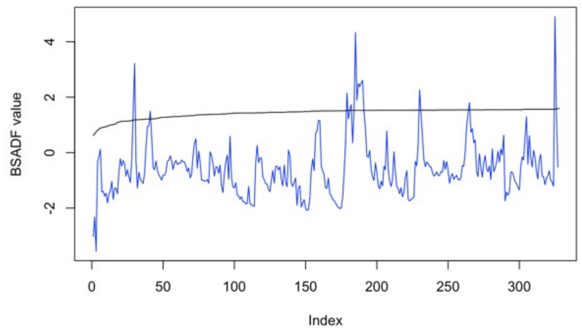

Image 3 - Graph of BSADF Test versus index for a one-third standard window size, 95% confidence interval

2. Analysis of synthetic PM2.5 data using BSADF

Image 4 highlights the impact of changing window size on the results obtained through data analysis using the BSADF test. In the explosive behavior regime (green circle) decreasing window size has little effect on BSADF value, although in other regions the effect as indicated by the difference in BSADF intensities between each trace, is quite pronounced. The magnitude of these intensity differences (gaps) between traces for different window sizes thus creates a set of new test statistics.

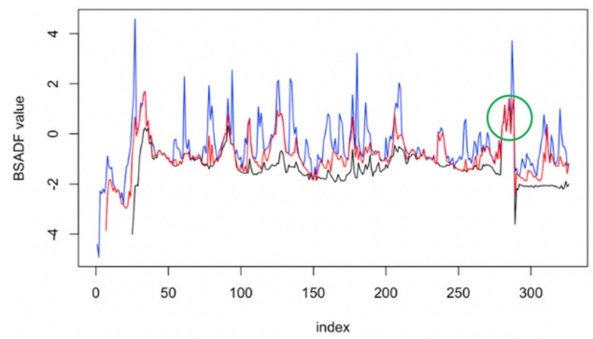

Image 4 - Graph of BSADF values versus index for data in which the black line represents the standard window, the red line represents half of the standard window, and the blue line represents one-third of the standard window.

3. Analysis of synthetic PM2.5 data using the new test statistic

From the new test statistics in section 2, the expression $\frac{BSADF_{standard}}{(BSADF_{x(standard)}-BSADF_{standard})+y}$ was derived after running Monte Carlo analysis (10000 times) on synthetic data. The x and y constants best fitting this set of data were 0.5 and 1, respectively.

4. Analysis of real PM2.5 data using the new test statistic

As outlined by the red circle in image 5, the new test statistics, as given by $\frac{BSADF_{standard}}{(BSADF_{0.5(standard)}-BSADF_{standard})+1}$, can detect explosive behavior over a short period of a time series. Moreover, over-detection of explosive behavior does not occur.

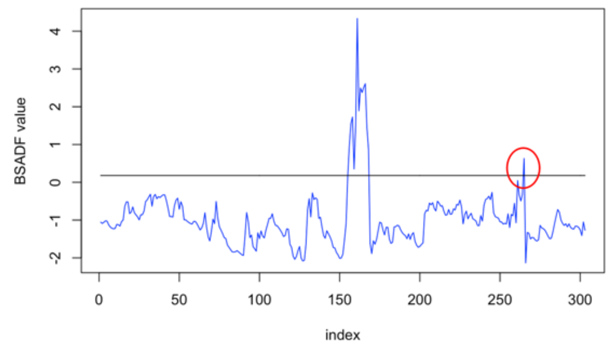

Image 5 - Graph of new test statistics