P-ISSN 2586-9000
E-ISSN 2586-9027
Homepage : https://tci-thaijo.org/index.php/SciTechAsia
Science & Technology Asia
Vol.28 No.4 October - December 2023
Page: [108-114]
Original research article

Explosive Behavior Detection of PM2.5 During Wildfire Period Based on BSADF Test

Kanisorn Sawangsawai[1], Pat Vatiwutipong[2,*]

[1]*Kamnoetvidya Science Academy, Rayong 21210, Thailand*
[2]*Independent Researcher, Bangkok 10600, Thailand*

Received 28 June 2023; Received in revised form 17 October 2023
Accepted 3 November 2023; Available online 27 December 2023

ABSTRACT

Time series is a type of data that is popular in statistical analysis. Explosive behavior, which is an immediately skyrocketing time series, is one of the important behaviors of time series that is often used in many tasks. In addition, a particular tool that has been used frequently in the last few years is the Backward Supremum Augmented Dickey-Fuller Test (BSADF Test). BSADF is developed for mainly use in the stock market, but when it is used with PM 2.5 data in which wildfires occurred, it is observed that BSADF cannot detect the explosive behavior in a short time series of the data. This problem leads to the development of a method based on BSADF to detect explosive behavior in a short period of a time series, so this new method can face various types of data. From investigating the BSADF test by using different sizes of windows in synthetic data that was generated by the ARMA process, it has been noticed that decreasing of windows will affect the BSADF test by increasing the BSADF value a little in the explosive behavior period, whereas other periods have been increased numerously. So, by using the difference in the amount of gap of the BSADF value in different sizes of windows, it led to a new test statistic. The new test statistic is outperformed compared to the BSADF test both in synthetic and real data, it could detect explosive behavior in a short period of time series when wildfire occurred and not over-detect explosive behavior in other periods.

Keywords: BSADF test; Explosive behavior; PM2.5 data

1. Introduction

Time series is a type of data that is popular in statistical analysis nowadays, such as the daily stock market index at the close of trading each day, company revenue each year, or the amount of rainfall each hour. There are many different types of patterns and behaviors in time series. Time series analysis was developed primarily for financial sciences such as stock market trend

***Corresponding author:** pvatiwutipong@gmail.com

analysis or predicting the price of assets. However, nowadays, other disciplines have increasingly applied time series analysis to analyze behavior and predict data trends [1]. Explosive behavior, a region of data which skyrockets in intensity compared to others, is one important feature which can occur in time series data. Detecting the explosive behavior of time series can be very useful in analyzing time series, such as stock prices in the stock market data.

A tool used for analysis of explosive behavior in the last few years is the Backward Supremum Augmented Dickey-Fuller Test (BSADF Test) [2]. BSADF was developed for use in the financial 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. The observations found that the PM 2.5 data fluctuated faster and more strongly than the asset price data in the financial market. In some data segments, the behavior of time series bursts dramatically over a short period. When using a too-large window size, the length of the smallest time series interval used in regression analysis, BSADF will be unable to detect. However, when narrowing the data window, BSADF was found to be extremely sensitive. This causes it to detect more data with explosive behavior in the time series than actually exists. 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.

2. Research Methodology

2.1 Data preparation

The 2018 daily PM2.5 concentration in Butte County, California, United States collected by The United States Environmental Protection Agency from 1 January 2018 to 31 December 2018 for 346 days (Fig. 1), shows an example of exploding behavior which corresponds to a wildfire event, as shown by the red circle.

2.1 Statistical methods

2.2.1 ADF Test

The Augmented Dickey-Fuller test (ADF test) [3] is used to test whether time series data is stationary or non-stationary by testing the null hypothesis that whether a unit root is presented in the data as expressed in the following equation:

$$\Delta y_t = \mu + \beta y_{t-1} + \sum_{i=1}^{p} \phi_i y_{t-1} + \varepsilon_t ,$$

for some value of lag parameter p.

2.2.2 BSADF Test

The Backward Supremum Augmented Dickey-Fuller Test (BSADF Test) [2] detects the explosive behavior of the time series data by performing a supremum ADF test [4] on a backward expanding sequence of data as expressed in the following equation:

$$BSADF_{r_2}\left(r_0\right) = \sup_{r_1 \in [0, r_2 - r_0]} \left\{ ADF_{r_1}^{r_2} \right\},$$

where the window size is the length of the smallest time series interval used in regression analysis.

2.2.3 Proposed test statistics

A new test statistic was created following from the results observed in section 3.1 Analysis of PM2.5 data using BSADF Test. This is represented by the expression:

$$\frac{BSADF_{standard}}{\left(BSADF_{x(standard)} - BSADF_{standard}\right) + y},$$

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.

2.2.4 ARMA model

Autoregressive–Moving Average model (ARMA model) [5] at order (p,q) is presented by the expression:

$$X_t - \phi_0 - \phi_1 X_1 - \cdots - \phi_p X_p = W_t + \theta_1 W_{t-1} + \cdots + \theta_q W_{t-q},$$

where W_t is white noise. $\text{ARMA}(p,q)$ will stay in the stationary state when all roots of equation $1 - \phi_0 x - \cdots - \phi_p x^p = 0$ are more than 1.

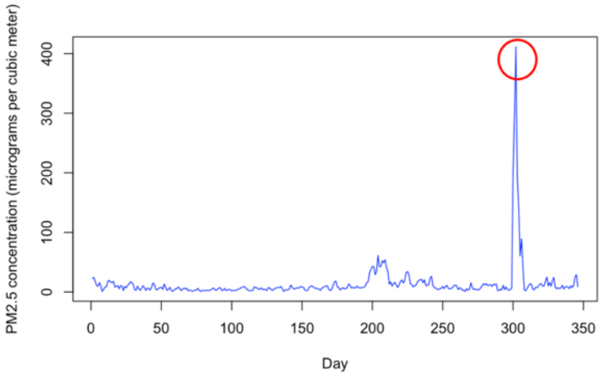


Fig. 1. Graph of PM2.5 concentration versus index in 2018 (Butte County, California, USA).

3. Experiment and Discussion

3.1 Analysis of PM2.5 data using BSADF Test

PM2.5 data were analyzed in RStudio with package reticulate, exuber, TSA, and vars in R language, using BSADF with different window sizes, proposed in [2], these being either standard, half standard, or one-third of standard window size. The real data shown in Fig. 1 was separated into 2 parts, a wildfire period (red circle) and other periods. The results shown in Table 1 were evaluated by confidence values in intervals of 90%, 95%, and 99%.

As indicated in Fig. 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 Fig. 1 which show explosive behavior (red circle) are observed as being below the critical value in Fig. 2. In addition, when the window size is decreased, other data periods show over-detection, i.e., explosive behavior, as indicated in Fig. 3.

Since few examples of real data such as in Fig. 1 exist, synthetic data sets are needed to evaluate the effect of different window sizes. The inexplosive part (Part 1) was modeled by the ARMA process, while the real data was utilized for the explosive part (Part 2) as shown in Fig. 4. The Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF) and Extended Autocorrelation Function (EACF), [4-6], were used to indicate the order of ARMA process. The results were shown in Fig. 5, where "x" denotes that the sample autocorrelation is significantly different from zero and "o" denotes that the sample autocorrelation is not significantly different from zero, which show that the model that best fits the data is ARMA(1, 2).

Then, three types of synthetic data were created, each with different parameter of ARMA process that obtained from different ranges of data, as shown in Table 2. 1000 sets of each data type were created using library math, numpy, pandas, and statsmodels in Python language. The synthetic data show that p-value of type 1 data is very small since the coefficient of AR1 is very close to 1 representing that the data is almost non-stationary. Finally, each set of data was analyzed by using BSADF test. Fig. 6 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, the effect is quite pronounced. The magnitude of these intensity differences (gaps) between traces for different window sizes thus creates a set of new test statistics.

Table 1. Performance of BSADF on the PM2.5 Data (Butte County, California, USA).

Window size	Number of the detected explosive behavior					
	90%		95%		99%	
	Wildfire	**Others**	**Wildfire**	**Others**	**Wildfire**	**Others**
standard	0	2	0	2	0	2
half standard	0	3	0	2	0	2
one-third of standard	0	8	0	8	0	5

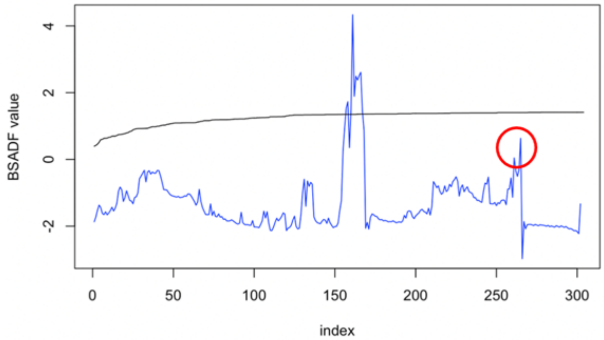


Fig. 2. Graph of BSADF Test versus index for a standard window size, 95% confidence interval.

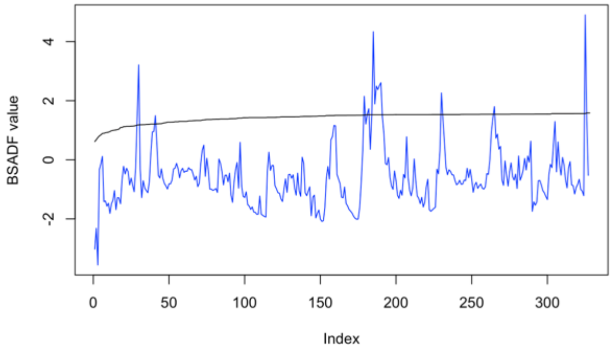


Fig. 3. Graph of BSADF Test versus index for a one-third standard window size, 95% confidence interval.

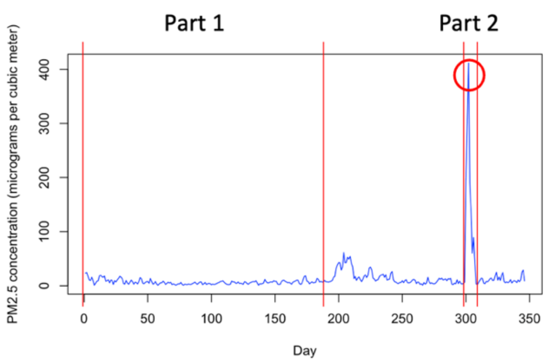
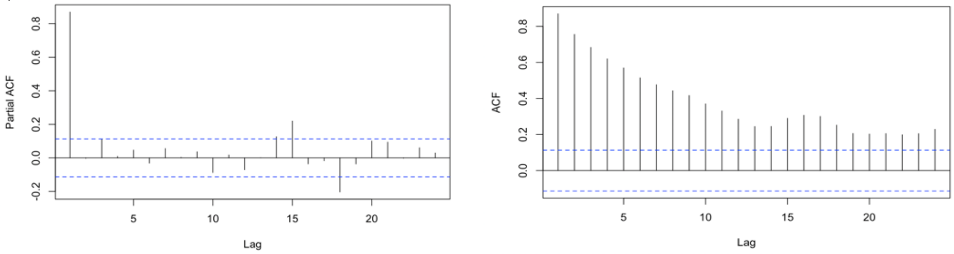


Fig. 4. Graph of PM2.5 concentration separated into an inexplosive part (Part 1) and an explosive part (Part 2).




```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x x x x x  x  x  x
1 o x o o o o o o o o o  o  x  x
2 o x o o o o o o o o o  o  o  x
3 o x o o o o o o o o o  o  o  x
4 o x o o o o o o o o o  o  o  o
5 x x o o o o o o o o o  o  o  x
6 x x o o o o o o o o o  o  o  x
7 x x o o o o o o o o o  o  o  x
```

Fig. 5. Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF) and Extended Autocorrelation Function (EACF) of the Part 1 data, where "x" denotes that the sample autocorrelation is significantly different from zero and "o" denotes that the sample autocorrelation is not significantly different from zero.

3.2 Analysis of PM2.5 data using proposed statistic

A new test statistic was created following on from the previous results that show the magnitude of BSADF intensity differences (gaps) between traces for different window is quite pronounced in the inexplosive region but absent in the explosive region. Therefore, the new test statistic is represented by the expression:

$$\frac{BSADF_{standard}}{\left(BSADF_{x(standard)} - BSADF_{standard}\right) + y},$$

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. After running Monte Carlo analysis on 1000 sets of synthetic data while varying x at 0.3, 0.4, 0.5, 0.6, and 0.7, and varying y at 0.001, 0.01, 0.1, 1, and 10, and running 10000 times to derive the critical value. The x and y constants best fit this set of data, which were selected from the value that the new test statistics can detect explosive behavior and perform the least false detection in the inexplosive range, were 0.5 and 1, respectively. The critical values obtained from the Monte Carlo method are shown in Table 3.

Table 2. Parameters of ARMA(1, 2) process obtained from different ranges of data.

Type	Range of data	Coefficient of AR1	Coefficient of MA1	Coefficient of MA2	Mean	Variance	t-test statistics	*p*-value
1	1 January 2018 to 25 July 2018	0.989	-0.290	-0.539	7.200	8.683	-9.6519	<0.00001
2	21 January 2018 to 25 July 2018	0.511	0.189	-0.121	6.184	7.106	-0.0028	0.4992

3	22 January 2018 to 25 July 2018	0.232	0.485	0.078	6.109	6.951	0.0014	0.4996

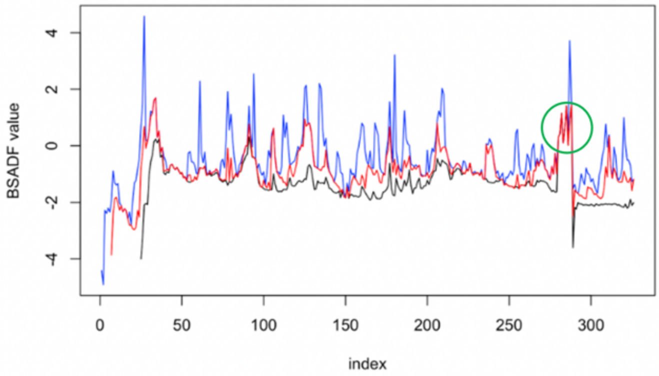


Fig. 6. Graph of BSADF values versus index for data in which the black line represents the standard window, the red line represents half of the standard window, and the blue line represents one-third of the standard window.

Table 3. Critical value of three types of synthetic data.

Type 1		**Type 2**		**Type 3**	
90%	**95%**	**90%**	**95%**	**90%**	**95%**
0.270	0.109	0.102	0.015	0.102	0.014

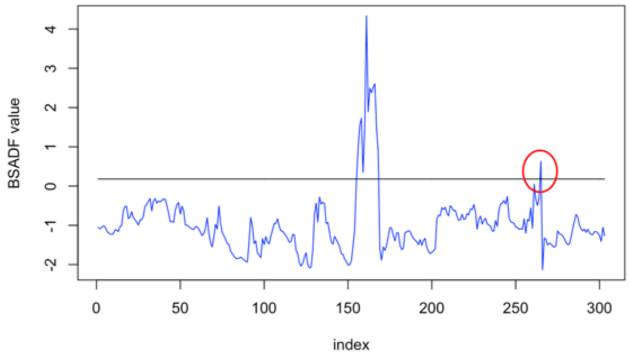


Fig. 7. Graph of new test statistics.

The real PM2.5 data in Fig. 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. As outlined by the red circle in Fig. 7, the new test statistics, as given by

$$\frac{BSADF_{standard}}{\left(BSADF_{x(standard)} - BSADF_{standard}\right)+1},$$

can detect explosive behavior over a short period of a time series. Moreover, over-detection of explosive behavior does not occur.

4. Conclusion

The detection of explosive behavior of PM2.5 data during wildfire events is crucial for protecting public health and mitigating the negative impact on the environment. In this paper, we propose a new test statistic based on BSADF to detect explosive behavior in PM2.5 during wildfire events. The proposed test statistic has the advantage of being more efficient than the existing BSADF test for both synthetic and real data.

This provides a better tool for environmental monitoring during wildfire events. The proposed test statistic has the added benefit of not over-detecting explosive behavior during periods when no wildfires occur. Thus, the proposed test statistic provides a more accurate and efficient way of detecting PM2.5 releases over short periods of time, such as in wildfire events, while minimizing false detection during other periods.

In future research, more data sets should be used to conduct the research to find universal critical values, x and y constants for each category of data set.

References

[1] Wei WWS. (2013). Time Series Analysis. Oxford Handbooks Online.

[2] Phillips PCB, Shi S, Yu J. Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. International Economic Review, 2015;56(4):1043-78.

[3] Said SE, Dickey DA. Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika, 1984;71:599-608.

[4] Phillips PCB, Wu Y, Yu J. Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values? International Economic Review, 2011;52(1):201-26.

[5] Tsay RS, Tiao GC. Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. Journal of the American Statistical Association, 1984;79(385):84-96.

[6] Dikey DA, Fuller WA. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association, 1979;74(366a):427-31.