Decomposition techniques for modelling the levels of particulate matter PM10 air pollutant in the city of silistra, Bulgaria

Cite as: AIP Conference Proceedings **2302**, 060019 (2020); https://doi.org/10.1063/5.0033631
Published Online: 03 December 2020

E. Veleva, I. Zheleva, and I. Georgiev

Decomposition Techniques for Modelling the Levels of Particulate Matter PM10 Air Pollutant in the City of Silistra, Bulgaria

E. Veleva[1,a)], I. Zheleva[b)] and I. Georgiev[c)]

Dept of Applied Mathematics and Statistics, Angel Kanchev University of Ruse, 7017 Ruse, 8 Studentska str., Bulgaria
www.uni-ruse.bg

[1] Corresponding author: [a)]eveleva@uni-ruse.bg
[b)]izheleva@uni-ruse.bg, [c)]irgeorgiev@uni-ruse.bg

Abstract. For the studied average monthly values of the levels of the air pollutant PM10 in Silistra in the period 01.2015 – 12.2019, two modern methods for decomposition were used - X-13ARIMA-SEATS and STL. The trend-cycle and seasonal component of the series were estimated in a total of 24 different ways – 8 models with the X-13ARIMA-SEATS approach with two seasonal adjustment options each – X11 and SEATS and 8 with the STL method. A comparative analysis was made between them, both in terms of estimating the components of the decomposition and in terms of the quality of approximation of the predicted values for the first six months of 2020 to the actually observed ones. In 23 out of 24 assessments of the trend-cycle component, a decreasing trend is observed, followed by a slightly increasing trend in the last year and several months of the period 2015–2019. The STL method yielded better forecast results for the first six months of 2020, using the default settings in the corresponding functions of R programming language. The estimated trend-cycle component by STL method is significantly smoother than that by method X-13ARIMA-SEATS.

INTRODUCTION

It is well known that the ambient air contamination by fine particulate matter (PM) is very serious problem ([1]-[10]). Nowadays this pollutant cause more and more respiratory problems, asthma, lung cancer and premature death. According to the European Commission surveys every year poor air quality causes the premature death of more people than the road accidents. All this make the scientific research for examination of the air contaminations in Bulgaria very important recently. That is why nowadays many studies in this area appeared in the scientific literature ([1]-[17]).

This paper is one continuation of our investigations of PM10 pollution for Bulgarian Danube region ([1]-[5], [8]-[13]) especially for the city of this region – Silistra ([5]).

Silistra Municipality is located in the north-eastern part the Republic of Bulgaria. To the north, the Danube River marks the border with Romania. The climate of the Silistra region is characterized by a moderate continental character and falls within the Danube climatic subregion. Characteristic of this area is the hot summer, the early onset of spring and the severe cold in winter. The absolute minimum temperature reached is -32^{0}C, and the maximum is 40.4^{0}C. The average annual air temperature is 11.6^{0}C. The stable retention of air temperature above 10^{0}C begins in the first ten days of April and lasts until the end of October - about 200 days per year. Recently the contamination of air with particulate matter in Silistra is going up, as in the whole territory of transborder Danube region Bulgaria - Romania ([1], [8], [9]).

PM10 levels for Silistra mark an increase during the autumn-winter period compared to the levels during the spring-summer period. Also the specific meteorological conditions during the winter seasons reduce the possibility of dissipation of atmospheric pollutants ([5], [10]). Thus the biggest peak of PM10 levels for the autumn-winter period is usually observed in January months. It is usually in January that the number of days with exceedance of the limit values of the PM10 levels is maximum observed compared with the other months in the year ([5], [10]).

There are several research papers for PM10 measurements modelling and forecasting for some regions in Bulgaria ([1]-[3], [6], [7], [11], [14]-[16]). In these publications different statistical methods for PM10 data modelling and studying are used such as ARIMA, SARIMA, GARCH and etc. In this paper we examine the application of two modern approaches to the decomposition of time series as a product of the trend-cycle and seasonal component - X-13ARIMA-SEATS and STL method. The different options of the two approaches are compared, and the total number of considered models is 24. The quality of approximation of the forecast values for the first six months of 2020, obtained with each of the models to the actually observed average monthly levels of the PM10 air pollutant in Silistra in 2020, was studied.

Application of Mathematics in Technical and Natural Sciences
AIP Conf. Proc. 2302, 060019-1–060019-9; https://doi.org/10.1063/5.0033631
Published by AIP Publishing. 978-0-7354-4036-4/$30.00

To examine the PM10 pollution in the city of Silistra, Bulgaria we use official data from the monitoring stations in the city [18].

1. DATA DESCRIPTION

We consider 60 average monthly values of PM10, measured by Bulgarian Ministry of Environment and Water in the five years period 2015–2019 (the training data set), shown on Figures 1 and 2. The average monthly levels of PM10 in the first six months of 2020 are used as a test data.

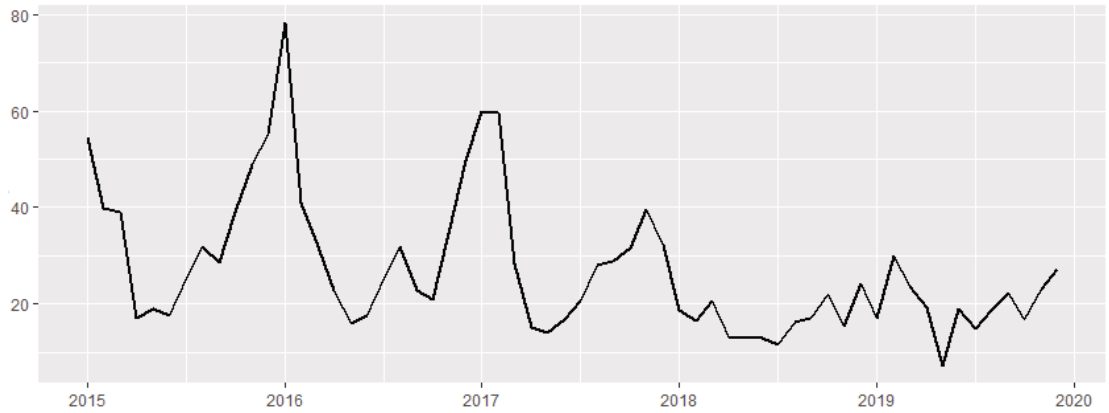

FIGURE 1. Average monthly values of PM10 levels for the period 01.2015–12.2019

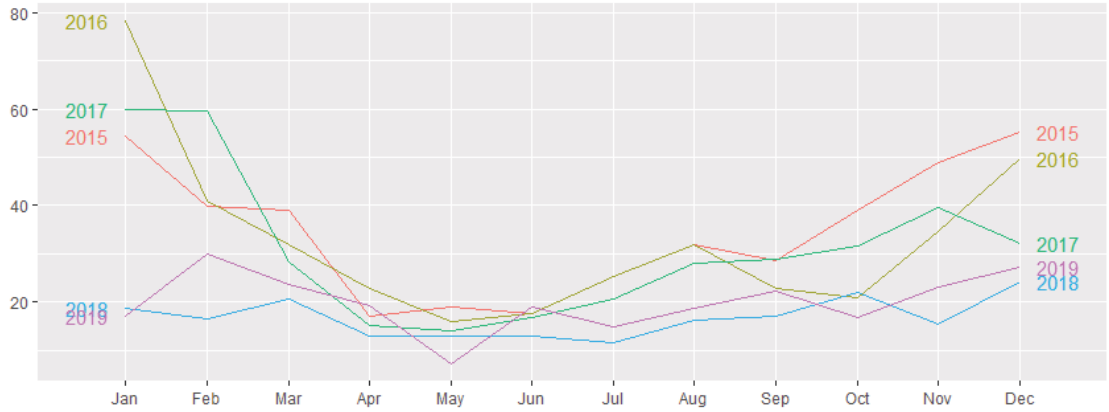

FIGURE 2. Seasonal plot of the data

The data obviously have a trend and a seasonal pattern. From Figure 1 it can be seen that the series has a decreasing tendency and together with it a decrease of the seasonal wave is noticed. Figure 2 shows the seasonal plot of the data. Here, too, it can be seen that in the last two years there has been a significant decrease in the amplitude of the seasonal wave.

A commonly used approach (see [19,20]) to study and explore the historical changes over time of data containing seasonality is to perform decomposition, allowing to separate the so-called trend-cycle and seasonal components. For data where the amplitude of the seasonal wave is proportional to the levels of the measured values, a multiplicative decomposition is recommended of the form

$$Y_t = T_t S_t R_t , \tag{1}$$

where T_t, S_t and R_t are the trend-cycle, the seasonal and the remainder components, respectively. The decomposition (1) can be written in the equivalent form

$$\log Y_t = \log T_t + \log S_t + \log R_t ,$$

which is an additive decomposition of the series $\log Y_t$. The series

$$Y_t^{s.adj} = Y_t / S_t = T_t R_t , \tag{2}$$

the so-called seasonally adjusted data, contain only the trend-cycle and the remainder components. They are often used to model and forecast the trend-cycle component of the original data Y_t.

2. MULTIPLICATIVE DECOMPOSITION OF THE DATA

In the classical decomposition method, dating back to the 1920s, the trend-cycle component is first estimated by smoothing using moving averages, and then the seasonal component is estimated by averaging for each season of the detrended values for that season ([19]). For example, with monthly data, the seasonal index for January is the average of all the detrended January values in the data. Thus, the method assumes that the seasonal component repeats from year to year.

In recent years, new, improved decomposition procedures have been developed and applied ([19], [20]).

The *seasonal* package in R programming language ([21]) provides an easy-to-use and full-featured R-interface to X-13ARIMA-SEATS, the newest seasonal adjustment software developed by the United States Census Bureau. The procedure handles monthly, quarterly or bi-annual time series; the additive or multiplicative decomposition; trading day variation, holiday effects and the effects of known predictors. The algorithm firstly models the data with so-called regARIMA models – regression models with ARIMA (autoregressive integrated moving average) errors. If no regressors are used, the regARIMA model reduces to an ARIMA model. There are built-in regressors for estimating various flow and stock trading day effects, holiday effects, certain kinds of disruptions in the series or sudden changes in level, whose effects need to be temporarily removed from the data before extracting the seasonal and trend-cycle components. It is also possible to incorporate user-defined regression variables into the model fitted. The choice of the ARIMA model for the errors can be made automatically by the algorithm or set by the user. The use of a regARIMA model with or without regressors aims to extend the series with forecasts (and backcasts) in order to improve the seasonal adjustments of the most recent (and the earlies) data. The user can choose between two seasonal adjustment options: X11 and SEATS. X11 uses an iterative approach based on smoothing by moving averages to estimate the components of a time series. SEATS (Signal Extraction in ARIMA Time Series) estimates and forecasts the components of a time series using signal extraction techniques applied to ARIMA models. A demo website ([22]) supports interactive modelling of custom data.

The STL (Seasonal and Trend decomposition using Loess) method, developed in 1990 (see [23]), is another, often used, modern method for decomposition. Key to the STL approach is Loess–LOcal regrESSion smoothing. It is a versatile and robust method, which can be used for any type of seasonality (see [19]). The trend-cycle and seasonal components are allowed to change over time and the rate of change can be controlled by the user. The *stl* function in R has two main parameters to be chosen – the trend-cycle window (*t.window*) and the seasonal window (*s.window*). Both should be odd numbers and refer to the number of consecutive years to be used in smoothing. The *mstl* function, *forecast* package, provides an automated STL decomposition choosing *t.window* automatically. The default value for *s.window* is 13. The *stlf* function uses *mstl* to decompose the series, then it calculates the seasonally adjusted values (2), for which it can automatically choose an ARIMA model to compute point and interval forecasts and finally "reseasonalises" the predictions, using the estimated by *mstl* seasonal component for the last year (sub period).

2.1. Decomposition By X-13ARIMA-SEATS Method

The function *seas* is the core function of the *seasonal* package. The function *view* of the *seasonalview* package provides a graphical user interface to the *seasonal* package, identical to that of the site [22].

First, we run the *seas* command using the default options - automatic regARIMA model selection, automatic data transformation selection (Logarithmic, Square root or No transformation), outlier detection, testing both seasonal adjustment methods – X11 and SEATS separately. As a result we get the model ARIMA$(1,0,0)(1,0,0)_{12}$, with one additive outlier at 05.2019 (the observed value is only 7.13 μg/m^3) and Logarithmic transformation of the original data. The function *summary* gives the values of the AICc (Corrected Akaike's Information Criterion) and BIC (Schwarz's Bayesian Information Criterion),

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2(k+2)(k+3)}{N-k-3}\ , \quad \mathrm{AIC} = N\log\left(\frac{SSE}{N}\right) + 2(k+2)\ , \quad \mathrm{BIC} = N\log\left(\frac{SSE}{N}\right) + (k+2)\log(N)$$

where $N = 60$ is the number of observations used for estimation, k is the number of predictors in the model and *SSE* is the sum of squared errors of the one-step-ahead predictions of the chosen ARIMA model on the training data set. The results are given in the first row of Table 1. The forecasted values $\hat{Y}_t$ for the first six months of 2020 are shown in the first row of Table 2. Given the actual observed values in 2020 (the test data set), shown in the last row of Table 2 we can calculate one of the most commonly used criteria for the quality of the predictions - RMSE, MAE and MAPE,

$$\text{RMSE} = \sqrt{mean(e_t^2)}\ ,\ \text{MAE} = mean\left(|e_t|\right),\ \ \text{MAPE} = mean\left(|p_t|\right),\ \ p_t = 100\,e_t / Y_t,\ \ e_t = Y_t - \hat{Y}_t. \qquad (3)$$

The obtained Test RMSE, Test MAE and Test MAPE are given in the last three columns in row 1 of Table 1. The two possible adjustment methods – X11 and SEATS produce different decompositions of the form (1). If in (3) we substitute $\hat{Y}_t = T_t S_t$, the predicted by the decomposition values for Y_t on the training data set, we will obtain the Training RMSE, MAE and MAPE for each of the two methods separately, also shown in row 1 of Table 1. Row 2 of Table 1 contains the corresponding results for the same model but choosing the option "no outlier detection."

TABLE 1. The results with different ARIMA models and options

No	Model	Adjustment method	AICc	BIC	Train. RMSE	Train. MAE	Train. MAPE	Test RMSE	Test MAE	Test MAPE
1	$(1,0,0)(1,0,0)_{12}$	X11	410.2	419.5	5.5295	3.3439	14.0836	8.0530	6.7299	44.9944
		SEATS			2.6554	1.5565	7.6948			
2	$(1,0,0)(1,0,0)_{12}$ no outlier detection	X11	423.7	431.3	5.2509	3.1721	12.2148	6.1898	5.1043	28.9704
		SEATS			2.2549	1.5157	6.1364			
3	(1,0,0)	X11	426.4	434.1	5.1555	3.2455	14.0334	9.2483	7.7221	55.1096
		SEATS			3.6109	2.4313	10.9428			
4	$(1,0,0)(0,0,1)_{12}$	X11	424.8	434.2	5.0935	3.2408	14.0808	9.2435	8.0962	55.2230
		SEATS			3.3996	2.2935	10.4848			
5	$(2,0,0)(1,0,0)_{12}$	X11	437.9	445.5	5.2108	3.1609	12.2322	6.2388	5.4687	31.3832
		SEATS			2.5132	1.6577	6.6258			
6	$(1,0,1)(1,0,0)_{12}$	X11	438.1	445.6	5.2403	3.1671	12.2147	6.1927	5.2578	29.8216
		SEATS			2.6708	1.7483	6.9950			
7	$(0,0,2)(0,1,1)_{12}$	X11	324.5	332.5	4.6143	3.0601	11.9626	5.0568	3.1003	11.9623
		SEATS			2.8069	1.8302	6.9860			
8	$(0,0,2)(0,1,1)_{12}$ no outlier detection	X11	339.6	346.2	5.1024	3.2514	12.4985	5.4046	4.0646	16.8854
		SEATS			2.2142	1.3464	4.9735			

TABLE 2. The actual and the forecasted values

Model No	Jan.20	Feb.20	Mar.20	Apr.20	May.20	Jun.20
1	23.00	28.18	25.31	23.05	22.85	22.72
2	22.52	27.76	24.76	22.45	14.75	22.16
3	25.80	28.04	26.30	25.17	26.81	24.55
4	24.81	31.02	26.41	24.86	26.20	23.63
5	23.26	28.58	25.57	23.10	15.33	22.23
6	22.73	28.16	25.18	22.74	14.95	21.97
7	23.61	26.42	21.18	14.38	8.89	14.27
8	21.97	28.29	23.59	17.19	8.94	17.15
Actual	**33.53**	**25.64**	**21.26**	**21.74**	**9.35**	**14.27**

The trend-cycle component for this automatically chosen by the procedure model ARIMA$(1,0,0)(1,0,0)_{12}$, with or without outlier detection, obtained by each of the two possible methods X11 and SEATS is shown on Figure 3. Each of the four lines of Figure 3 has a declining trend until the second half of 2018, followed by an upward trend in recent years and several months. The lines corresponding to the SEATS option are smoother than those obtained by the X11 method. Figure 4 shows the corresponding 4 lines depicting the seasonal component in the decomposition (1). Here, each of the four lines shows a fading seasonal component. With the X11 method, we have approximately the same form of seasonal wave over the years, as opposed to choosing the SEATS option. Rows 3 to 6 inclusive in Table 1 show the corresponding results when manually entering in the *seas* command each of the 4 other ARIMA models suggested by the procedure as the next 4 best models chosen by the BIC criterion. These models are shown as options in the graphical user interface via command *view* or

when using the function *fivebestmdl*. The values of the criteria corresponding to these four models are analogous to those of the first two rows of Table 1.

Command *auto.arima* in *forecast* package returns best ARIMA model according to either AIC, AICc or BIC value. Using this function on the original data Y_t with the option *lambda*=0 (indicating the logarithmic transformation of the data), we obtain the model ARIMA(0,0,2)(0,1,1)$_{12}$. Then in the function *seas* we manually input this model with the option *arima.model*="(0,0,2)(0,1,1)". Rows 7 and 8 of Table 1 contain the corresponding results by X11 and SEATS adjustment methods, with or without outlier detection.

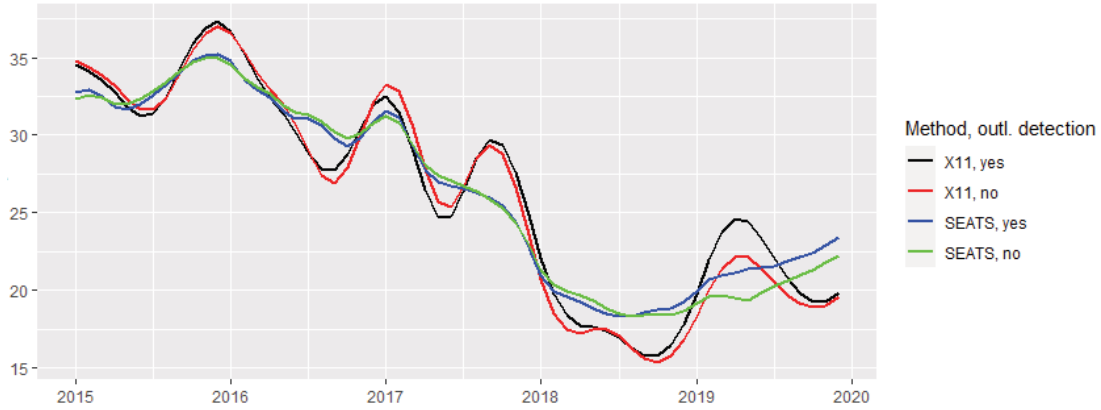

FIGURE 3. The trend-cycle component for the model ARIMA(1,0,0)(1,0,0)$_{12}$

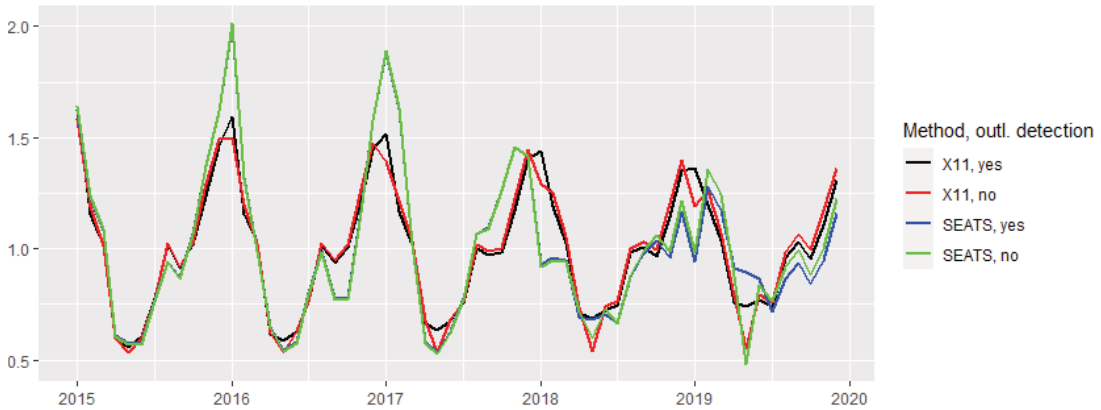

FIGURE 4. The seasonal component for the model ARIMA(1,0,0)(1,0,0)$_{12}$

Row 7 contains the minimum values in Table 1 of AICc, BIC, Test RMSE, Test MAE and Training RMSE, Training MAE, Training MAPE for the X11 method. Row 8 contains the minimum values of Training RMSE, Training MAE and Training MAPE for the SEATS method.

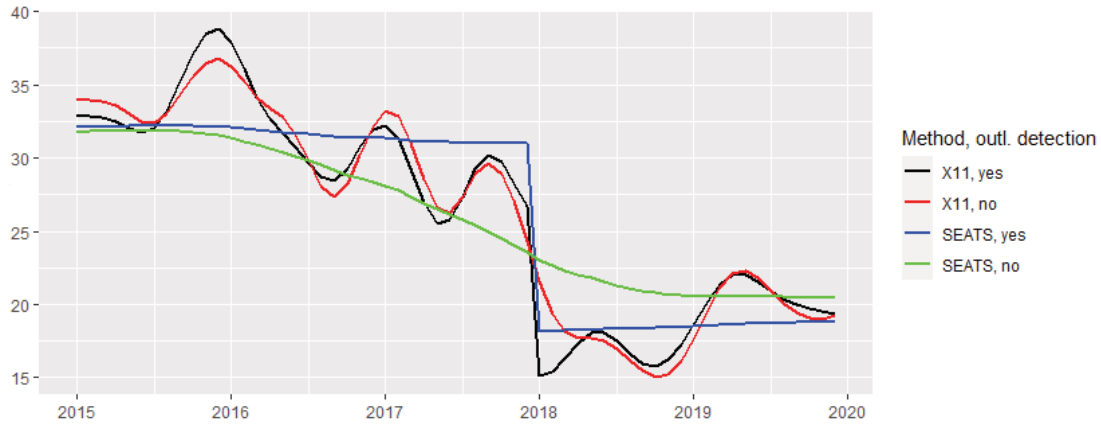

FIGURE 5. The trend-cycle component for the model ARIMA(0,0,2)(0,1,1)$_{12}$

The actual test data and the predictions for the first six months of 2020 by all 8 models are given in Table 2. The predictions in the Table that are closest to the respective actual values are colored green. Model 7 has 3 green values, followed by Model 2, Model 3 and Model 8 with 1 green value each.

Figure 5 shows the trend-cycle component by Models 7 and 8, with X11 and SEATS options each. The blue and black lines, corresponding to Model 7 show decreasing tendency until the nearly end of 2017 then we see a sharp decline and rising trend since the beginning of 2018. The red and green lines corresponding to Model 8 (the same model but without outlier detection), X11 and SEATS options, show, like Figure 3, a declining trend until the second half of 2018, followed by an increasing trend over the remaining year and several months. The four corresponding lines in Figure 6 representing the seasonal component are almost identical. Here, too, we see a decrease in the amplitude of the seasonal wave, retaining almost the same shape over the years. In both Figures 5 and 6, the green line corresponding to SEATS method, no outlier detection is that which differs more sharply from the other three lines. But exactly for Model 8, SEATS adjustment method, the product $\hat{Y}_t = T_t S_t$ is closest to the observed values Y_t, with minimum values of Training RMSE, Training MAE and Training MAPE in Table 1.

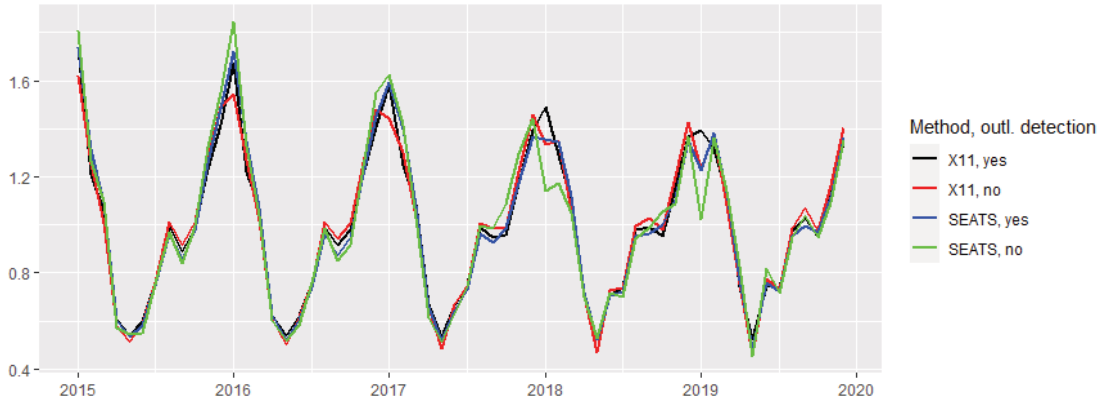

FIGURE 6. The seasonal component for the model ARIMA(0,0,2)(0,1,1)$_{12}$

2.2. Decomposition by STL Method

The results in Table 3 are obtained by the functions *mstl* and *stlf* in R at different preset values for the parameter *s.window*. Smaller values allow greater flexibility of the seasonal component, which is averaged over a smaller number of consecutive years. At higher values for *s.window*, greater uniformity of the seasonal wave in the different sub periods is achieved. This can be seen in Figure 7, which shows the seasonal component evaluated at four different values of *s.window* – 3, 5, 9 and 13 (the default value).

At *s.window = periodic* the estimated seasonal wave is the same for all sub periods. Figure 8 shows the trend-cycle components corresponding to the seasonal components of Figure 7. We see that when the seasonal component is more flexible, more closely following the data, the corresponding trend component has a smoother line and vice versa. Such a pattern can also be seen in Figures 3 and 4.

TABLE 3. The results with different values for the parameter *s.window*

Model No	s.window	ARIMA	AICc	BIC	Train. RMSE	Train. MAE	Train. MAPE	Test RMSE	Test MAE	Test MAPE
9	3	(1,1,0)	-108.84	-104.90	2.3471	1.5720	6.5332	6.1946	4.2795	18.2701
10	5	(0,1,1)	-12.79	-8.85	5.5239	3.7412	14.0859	5.1585	3.3770	13.7939
11	7	(1,1,1)	-3.47	2.32	6.0218	3.9885	15.1064	4.5114	2.9827	12.3816
12	9	(1,1,1)	2.06	7.86	6.4205	4.2770	16.0824	4.2574	2.8128	12.0020
13	11	(1,1,1)	5.23	11.02	6.5788	4.4201	16.5588	4.1307	2.7783	12.2172
14	13	(0,1,1)	7.03	10.97	6.6879	4.5103	16.8789	4.1016	2.7100	12.0955
15	15	(0,1,1)	8.07	12.01	6.7541	4.5643	17.0723	4.0735	2.6966	12.1617
16	periodic	(0,1,1)	10.53	14.47	6.8204	4.6541	17.4595	3.9824	2.6617	12.3418

After calculating the seasonally adjusted values (2), command *stlf* with the option *method="arima"* automatically finds a non-seasonal ARIMA model with a minimum value for the AICc criterion. The choice of

model for each row in Table 3, the values of AICc and BIC criteria are given in columns 3 to 5. The forecasted values, obtained with the selected model and multiplied by the respective estimated seasonal component for the last year are given in Table 4. Based on them and the actual observed values in the first six months of 2020, by formulas (3) are calculated Test RMSE, Test MAE and Test MAPE, also shown in Table 3. It can be seen that the worst predictions were obtained for *s.window* = 3, and the lowest values for Test RMSE and Test MAE correspond to *s.window = periodic*.

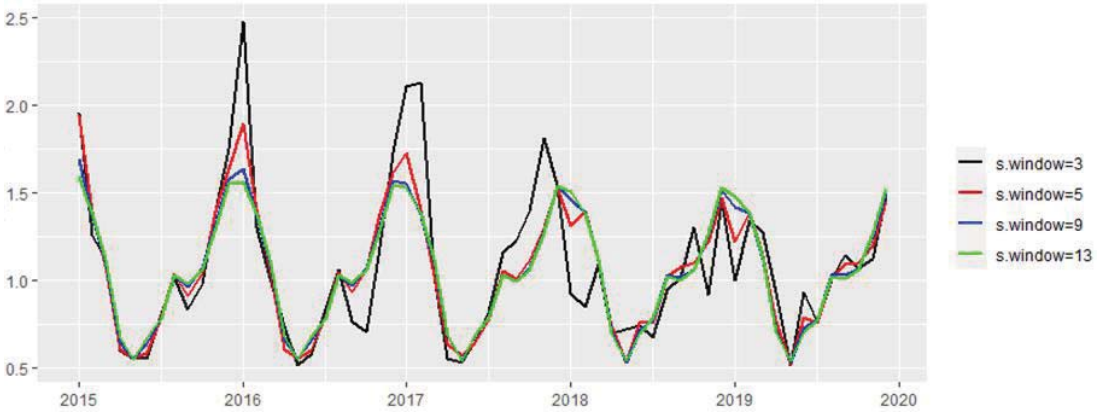

FIGURE 7. The seasonal component by STL decomposition method

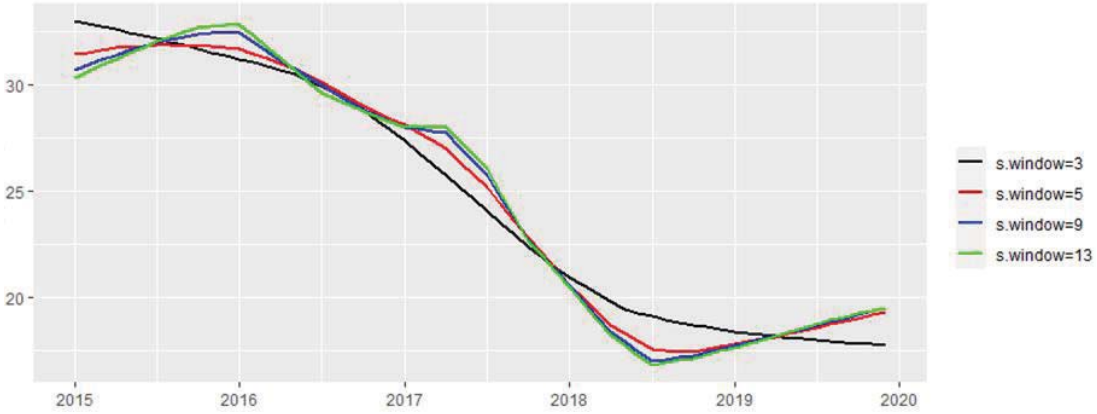

FIGURE 8. The trend-cycle component by STL decomposition method

For *s.window* = 9, 11, 13, 15 and *periodic*, close to each other predicted values, values for Test RMSE, Test MAE and Test MAPE and values for AICc and BIC criteria were obtained. Therefore, in order to obtain good prognostic results for future PM10 levels in Silistra, the default value *s.window* = 13 would be appropriately chosen.

TABLE 4. The actual and the forecasted values

Model No	Jan.20	Feb.20	Mar.20	Apr.20	May.20	Jun.20
9	19.81	25.53	24.83	17.96	10.05	18.08
10	23.15	26.28	21.76	14.69	10.08	15.23
11	25.48	26.61	21.59	14.28	10.17	14.54
12	26.76	26.53	21.55	13.91	10.23	14.05
13	27.52	26.51	21.60	13.72	10.29	13.79
14	27.95	26.27	21.44	13.49	10.25	13.55
15	28.20	26.25	21.45	13.42	10.27	13.46
16	28.92	26.27	21.54	13.29	10.34	13.26
Actual	**33.53**	**25.64**	**21.26**	**21.74**	**9.35**	**14.27**

A comparison between Tables 1 and 3 with respect to Test RMSE, Test MAE and Test MAPE shows that the STL method combined with a non-seasonal ARIMA prediction model gives more accurate forecasts. Only for

model ARIMA(0,0,2)(0,1,1)$_{12}$ (Models 7 and 8) in Table 1 the values of Test RMSE, Test MAE and Test MAPE are close to those in Table 3, unlike the other five automatically generated models of the procedure X-13ARIMA-SEATS.

The values in columns Training RMSE, Training MAE and Training MAPE of Table 3 are filled in analogously to those in Table 1 - putting in formulas (3) $\hat{Y}_t = T_t S_t$. Figures 7 and 8 show some similarity of the decomposition at higher *s.window* values with that using the X11 method, no outlier detection in Figures 3 - 6. There is also an analogy between the lines of Figures 7 and 8, corresponding to *s.window*=3 with the decomposition by SEATS, no outlier detection of Figures 3 to 6. This observation is also confirmed when comparing the values of Training RMSE, Training MAE and Training MAPE in Tables 1 and 3. The values at *s.window* = 5 in Table 3 are similar to those for the X11 method in Table 1. The others, at *s.window* = 7, 9, 11, 13, 15 and *periodic* are higher, but by the STL decomposition the line of the trend-cycle component is significantly smoother than that by the X11 adjustment method of the procedure X-13ARIMA-SEATS, which is more strongly influenced by the changes in the observed values of PM10 air pollutant.

CONCLUSIONS

For the studied average monthly values of the levels of the air pollutant PM10 in Silistra in the period 01.2015–12.2019, two modern methods for decomposition of type (1) were used - X-13ARIMA-SEATS and STL. The trend-cycle and seasonal component of the series were estimated in a total of 24 different ways – 8 models with the X-13ARIMA-SEATS approach with two seasonal adjustment options each – X11 and SEATS and 8 with the STL method. A comparative analysis was made between them, both in terms of estimating the components of the decomposition and in terms of the quality of approximation of the predicted values for the first six months of 2020 to the actually observed ones. In 23 out of 24 assessments of the trend-cycle component, a decreasing trend is observed, followed by a slightly increasing trend in the last year and several months of the period 2015–2019. The STL method yielded better forecast results for the first six months of 2020, using the default settings in the corresponding functions of R programming language. The estimated trend-cycle component by STL method is significantly smoother than that by method X-13ARIMA-SEATS.

ACKNOWLEDGMENTS

This paper contains results of the work on project No 2020 - FNSE – 04, financed by "Scientific Research Fund of Ruse University, Bulgaria."

REFERENCES

1. I. Zheleva, E. Veleva, and M. Filipova, "Analysis and modeling of daily air pollutants in the city of Ruse, Bulgaria," in *AMiTaNS'17*, AIP CP1895, edited by M.Todorov (American Institute of Physics, Melville, NY, 2017), paper 030007, https://doi.org/10.1063/1.5007366.
2. E. Veleva and I. Zheleva, "GARCH models for particulate matter PM10 air pollutant in the city of Ruse, Bulgaria," in *AMiTaNS'18*, AIP CP2025, edited by M.Todorov (American Institute of Physics, Melville, NY, 2018), paper 040016, doi:10.1063/1.5064900.
3. E. Veleva and I. Zheleva, "Statistical modeling of particle mater air pollutants in the city of Ruse, Bulgaria," in *MATEC Web of Conferences* **145**, 01010 (2018).
4. I. Tsvetanova, I. Zheleva, and M. Filipova, "Statistical study of the influence of some atmospheric characteristics upon the particulate matter (PM10) air pollutant in the city of Ruse, Bulgaria," in *AMiTaNS'18*, AIP CP2025, edited by M.Todorov (American Institute of Physics, Melville, NY, 2018), paper 110006, doi: 10.1063/1.5064949.
5. I. Tsvetanova, I. Zheleva, and M. Filipova, "Statistical study of the influence of the atmospheric characteristics upon the particulate matter (PM10) air pollutant in the city of Silistra, Bulgaria," in *AMiTaNS'19* AIP CP2164, edited by M.Todorov (American Institute of Physics, Melville, NY, 2019), paper 120014.
6. S. Gocheva-Ilieva, A. Ivanov, and I. Iliev, "Exploring key air pollutants and forecasting particulate matter PM10 by a two-step SARIMA approach," in *AIP CP2106*, (American Institute of Physics, Melville, NY, 2019), paper 020004.
7. S. Gocheva-Ilieva and A. Ivanov (2019) Assaying stochastic SARIMA and generalized regularized regression for particulate matter PM10 modeling and forecasting, *International Journal of Environment and Pollution* **66**, 41-62.
8. M. Filipova, I. Zheleva, and P. Roussev (2013 Characteristics of PM air pollution along Bulgaria - Romania Danube region, *Ecologica* **71**, 215-217.
9. M. Filipova, I. Zheleva, I. Tsvetanova, D. Stefanova, and P. Rusev, "Analysis of the state of ambient air in the border region Bulgaria-Romania," in *International Symposium "The Environmental and The Industry" SIMI 2016, Bucharest*, Symposium Proceedings, pp.440-450, http://www.simiecoind.ro/doi10-21698simi-2016-0062-analysis-of-the-state-of-ambient-air-in-the-border-region-bulgaria-romania/.

10. I. Tsvetanova, I. Zheleva, and M. Filipova, "Statistical study of the influence of the atmospheric characteristics upon the particulate matter (PM10) air pollutant in the city of Silistra," in *AMiTaNS'19*, AIP CP2164, edited by M.Todorov (American Institute of Physics, Melville, NY, 2019), paper 120014.
11. E. Veleva and I. Zheleva, "Statistical modeling of particle mater air pollutants in the city of Ruse, Bulgaria," in *MATEC Web of Conferences* **145**, 01010, (2018), *NCTAM 2017*, doi: https://doi.org/10.1051/matecconf/ 201814501010.
12. I. Zheleva and M. Filipova, "Atmospheric characteristics statistic study of Ruse region, Bulgaria," in *AMiTaNS'16*, AIP CP1773, edited by M.Todorov (American Institute of Physics, Melville, NY, 2016), paper 110019, doi: http://dx.doi.org/10.1063/1.4965023.
13. A. Stefanova, I. Zheleva, M. Filipova, and I. Tsvetanova, "Examination of the possibility of transborder pollution in the days with registered exceedances of pollutant dust in the city of Ruse, Bulgaria," in *MATEC Web of Conferences* **145**, 01004 (2018), *NCTAM 2017*, https://doi.org/10.1051/matecconf/201814501004.
14. A. Ivanov, D. Voynikova, S. Gocheva-Ilieva, and D. Boyadzhiev, "Parametric time series analysis of daily air pollutants of city of Shumen, Bulgaria," in *AMiTaNS'12* AIP CP1487, edited by M.Todorov (American Institute of Physics, Melville, NY, 2012), pp. 386-396, http://dx.doi.org/10.1063/1.4758982.
15. D. Voynikova, S. Gocheva-Ilieva, A. Ivanov, and I. Iliev, "Studying the effect of meteorological factors on the SO2 and PM10 pollution levels with refined versions of the SARIMA model," in *AMiTaNS'15*, AIP CP1684, edited by M.Todorov (American Institute of Physics, Melville, NY, 2015), paper 100005.
16. M. Stoimenova, D. Voynikova, A. Ivanov, S. Gocheva-Ilieva, and I. Iliev, "Regression trees modeling and forecasting of PM10 air pollution in urban areas," in *AMiTaNS'17* AIP CP1895, edited by M.Todorov (American Institute of Physics, Melville, NY, 2017), paper 030005.
17. A. Stefanova, "Analysis of the inventory methods for pollutants in ambient air in the EIA reports (environmental impact assessment) for investment proposals implemented on the territory of RIEW Ruse for the 2013-2016 period," National Research and Development Institute for Industrial Ecology, Romania, INCD-ECOIND, http://www.dspace.incdecoind.ro/handle/123456789/372.
18. Data from the measured PM10 concentrations by automated measuring station, Ruse, Bulgaria, http://www.riosv-ruse.org/mesechna-spravka-za-nivata-na-fpch10.html.
19. R.J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd edn (OTexts, Melbourne, Australia, 2018), OTexts.com/fpp2.
20. C. Sax and D. Eddelbuettel (2018) Seasonal Adjustment by X-13ARIMA-SEATS in R, *Journal of Statistical Software*, **87**(11), 1-17, doi: 10.18637/jss.v087.i11.
21. R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
22. http://www.seasonal.website/.
23. R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning (1990) STL: A Seasonal-trend decomposition procedure based on loess, *Journal of Official Statistics* **6**, 3-73.