ALFA BK UNIVERSITY
FACULTY OF INFORMATION TECHNOLOGIES
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
International Scientific Conference

UDK:519.246.8
711.45:004.78
DOI:10.46793/ALFATECHproc25.122T
Original scientific paper

Performing EDA Techniques for Time Series Forecasting in Smart Cities

Ninoslava TIHI[1], Miloš TODOROV[2], Marko PAVLOVIĆ[3], Filip KOKALJ[4]

Abstract: In the age of smart cities, where interconnected technologies produce substantial data, time series forecasting has become a crucial instrument for efficient urban management. Exploratory Data Analysis (EDA) is an essential initial phase in comprehending and organising data for precise and insightful predictions. This study examines the utilisation of exploratory data analysis approaches for time series data in the framework of smart cities. It emphasises techniques for detecting trends, seasonality, autocorrelations, and correlations in data streams from sources like environmental monitoring systems. The research investigates ways to identify patterns, test hypotheses, and confirm assumptions using visual and quantitative methods. Employing diverse visualisation tools, statistical summaries, and decomposition techniques enhances the comprehension and preprocessing of time series datasets. The objective of the research is to perform and evaluate the efficacy of standard tools such as time plots, autocorrelation and correlation analysis, seasonality decomposition, and identifying trend patterns in conjunction with sophisticated techniques such as seasonal decomposition of time series (STL), correlation heatmaps, and trend detection techniques. The findings emphasise EDA's significance as a fundamental step in empowering researchers and practitioners to make informed decisions, ensuring strong analytical results, and developing resilient time-series forecasting models for contemporary urban issues.

Keywords: EDA techniques; environmental data; preprocessing time series; smart cities; time series forecasting.

1 INTRODUCTION

The concept of smart cities, which use data-driven solutions to improve urban living, has gained popularity due to the rapid growth of urbanisation and improvements in technology. Time series forecasting is important for smart cities because it allows for predictions about infrastructure, resource allocation, and public services. Forecasting accurately can help improve energy management, traffic flow, pollution control, drought management and catastrophe preparedness, which will ultimately lead to a better quality of life for those living in cities.

Exploratory Data Analysis (EDA) is a crucial first step in accomplishing these goals. It helps to analyse time series data, discover trends, and find patterns in data that can affect predictions. EDA approaches assist in transforming raw data into useful insights by visualising trends, seasonality, and correlations within the data. This technique is particularly important for time series data since temporal dependencies and anomalies can have a major effect on the predictive models.

Anestis Kousis [1] conducts a bibliometric study of research on Data Mining (DM) technology used in smart city applications. It seeks to identify the primary data mining approaches and the evolution of the study area throughout time. The research encompasses 197 publications published from 2013 to 2021, using the Scopus database and the Biliometrix library in R. The results indicate a diverse array of DM technologies used across all levels of a smart city initiative, with several ML algorithms utilised at the instrumentation, middleware, and application layers. DM for smart cities is an expanding scientific domain garnering global attention in study and cooperation.

Chen [2] presents a hybrid method (STL-LSTM) to enhance short-term metro ridership forecasting. It integrates Seasonal-Trend decomposition using Loess (STL) with the LSTM neural network to alleviate irregular oscillations. The original series is divided into three sub-series, and the LSTM neural network forecasts each one. The whole production is consolidated. The STL-LSTM model attains superior accuracy compared to the single LSTM, support vector regression, and the EMD-LSTM model.

The Fourth Industrial Revolution's data science is changing cities' technology and operations. Automating and intelligent city systems need actionable city data insights and data-driven models. Sarker [3] studies "Smart City Data Science," which uses sensor, internet-connected device, and other data to improve decision-making and citizen services. Artificial intelligence, especially machine learning analytical modelling, may improve municipal data understanding and computer operations. The report provides academics, industry experts, and policymakers with 10 open research problems for data-driven smart cities.

Martin Strohbach [4] investigates the possibilities of a cohesive Big Data analytical framework for Internet of Things (IoT) and Smart City applications. It offers a summary of Big Data and Internet of Things technologies, showcases a case study in the smart grid sector, and introduces a preliminary version of the framework that tackles the challenges of volume and velocity. The results are derived from comprehensive outcomes of the EU-funded projects BIG and the German-funded project PEC. The objective is to decrease development expenses and facilitate new services for residents and urban policymakers.

Green IoT seeks to diminish energy usage by lowering the energy demands of IoT devices. Mohammad Shorfuzzaman [5] seeks to develop prediction models using sensor data to analyse energy use in an IoT-enabled smart home setting. Two methodologies are employed: a comprehensive energy consumption model using linear and non-linear regression approaches, and a multi-step time-series model applying the autoregressive integrated moving average (ARIMA) methodology. The LSTM regression model surpasses conventional linear and ensemble regression models because to its significant variability. The suggested predictive models will assist customers in reducing energy use and enable energy suppliers to more effectively plan and anticipate future energy demand, hence promoting sustainable urban development.

The rising levels of particulate matter (PM2.5) in air pollution inside smart cities globally present a substantial risk to both the populace and the ecosystem. In Malaysia, traffic congestion significantly contributes to air pollution in urban areas such as Kuala Lumpur and Johor Bahru. Machine learning methods have been extensively researched worldwide for air pollution prediction, although their use in Malaysia remains limited. This work intends to use machine learning methods to forecast PM2.5 levels in smart cities using the Malaysia Air Pollution dataset. The Rishanti Murugan [6] study evaluates Multi-Layer Perceptron (MLP) and Random Forest algorithms, concluding that Random Forest exhibits superior accuracy in forecasting PM2.5 air pollution indices in smart cities.

This article explores the use of EDA approaches for forecasting time series in smart cities. It aims to conduct and evaluate the effectiveness of techniques such as time plots, seasonality decomposition of time series using the LOESS method, ACF and PACF functions, and trend pattern identification by using various trend detecting tests. This approach helps build smarter, more efficient, and more sustainable cities by connecting data analysis with predictive modelling.

2 TIME SERIES DATA IN SMART CITIES

Understanding time series data in the context of smart cities is essential for optimising urban management, improving resource efficiency, and elevating inhabitants' quality of life. Smart cities produce enormous volumes of time series data through a variety of IoT devices, sensors, and interconnected systems, which is an essential element of urban operations. Primary sources of such data include transportation and traffic systems, environmental monitoring, smart buildings and infrastructure, healthcare systems, public safety and security, and energy and utilities.

Time series data is a set of observations or measurements gathered at regular, consecutive periods over time. It is utilised to monitor alterations, trends, patterns, or behaviours of a variable across time. Time series generally consists of three (trend, seasonality, and residuals) or four components (trend, seasonality, cyclical, and residuals) depending on the characteristics of the data and the phenomena being analyzed. If seasonality is clear, residual noise is sufficient, and there is no cyclical behaviour, a time series can be decomposed into three components. However, the presence of cyclical patterns in addition to trend, seasonality, and residuals shows that the time series consists of four components.

The key components of a time series are:

- **Trend:** It represents a long-term direction or changes (linear or otherwise) of the data over an extended period. It can indicate overall increase, decrease, or stability in the data. For example, global warming manifests in temperature data as a long-term increase in the annual average across many regions in the world.
- **Seasonality:** It refers to the regular and repeating patterns in the data that occur at regular periods within a time series. In time series analysis, seasonality appears as peripdic fluctuations that repeat over fixed time intervals like days, weeks, months, or years. For example, a high air temperature or higher electricity demand due to air conditioning during the summer months.
- **Cycles:** A cycle represents non-fixed fluctuations usually caused by economic or social factors. Unlike seasonality, cycles may vary in length and do not follow a certain fixed interval. For example, recessions affecting housing markets.
- **Irregularities:** The irregular component, sometimes referred to as the residual or noise, is the remainder after the seasonal and trend components have been estimated and removed. It results from short-term variations in the series which are neither predictable nor systematic.

3 EXPLORATORY DATA ANALYSIS TECHNIQUES

EDA techniques are used for examining the relationships between features and the goal variables. There are many EDA techniques which can be categorized for univariate, bivariate and multivariate data. Techniques related to univariate analysis are descriptive statistics which include mean, median, mode, variance, standard deviation, skewness, kurtosis, visualization techniques which include histogram, boxplots, density plots etc. Techniques related to bivariate analysis include statistical measures such as correlation coefficient (Pearson, Spearman, Kendall) and covariance, and visualization techniques like scatter plots, pair plots, box plots, heatmaps, correlation matrices etc. Multivariate data analysis include techniques such as Principal Component Analysis (PCA), clustering and factor analysis, and visualization techniques such as pair plots, 3D scatter plots, paralell coordinates plots etc. However, there are certain time series EDA techniques specifically used for time series data like pattern detection techniques which include trend analysis, seasonality analysis, seasonal decomposition, autocorrelation (ACF) and partial autocorrelation (PACF), stationarity testing (Unit root tests), and visualization techniques like line plots, lag plots, seasonal decomposition plots etc.

3.1 Decomposition of time series

Time series can contain a series of patterns combined together, and it is possible to decompose the time series into these components.

Decomposition represents the process of disaggregating a time series into its fundamental components, facilitating the detection of patterns and comprehension of behaviour, hence enhancing the selection of predictive models [7].

It is sometimes essential to execute some adjustments or transformations by removing the source of variance in order to be able to simplify the observed patterns before performing the actual decomposition. There are four categories of adjustments: calendar adjustment, population adjustment, inflation adjustment, and mathematical transformations. Adjustments, such as calendar modifications, are applied to data exhibiting seasonality and variations due to basic calendar effects [8].

After the transformations are performed, the process of decomposition starts with choosing one of the two decomposition models, specifically the additive model or the multiplicative model [9].

The classical or additive decomposition can be written in the form of the sum of the three components, namely the seasonal, trend, and irregular components:

$$y_t = S_t + T_t + E_t \quad (1)$$

where:

y_t = data in period t,

S_t = seasonal component in period t,

T_t = trend component in period t,

E_t = irregular component uin period t.

This model is mainly used when the variation of the seasonal component is relatively constant over time. However, if the variations increase over time, it is better to use a multiplicative decomposition model that can be represented as a product of all three components:

$$y_t = S_t * T_t * E_t \quad (2)$$

A procedure used to decompose time series into its three components is called the STL method. This method was first presented by Robert Cleveland, William Cleveland, Jean McRae, and Irma Terpenning in 1990. STL stands for "Seasonal and Trend decomposition using Loess," where Loess is the method used to estimate non-linear relationships. The decomposed time series components can be presented using plots to visually determine the existence of a trend or seasonality [10].

The STL technique possesses the subsequent characteristics [10]:

• features a simplistic design

• enables rapid computation of extensive time series

• enables the specification of the seasonal and trend component periods

• allows for the decomposition of time series with missing values

• supports only additive decomposition, excluding multiplicative decomposition

3.2 Trend analysis

In addition to visualisation techniques using time series decomposition, the trend can also be determined by statistical tests for trend detection. The primary objective in identifying a trend within a time series is to determine the presence of a trend type, that is, to determine whether the trend is linear, monotonic, or any trend, assuming that the data are uncorrelated.

The statistical tests used for this objective are:

• Classic t-test for assessing the presence of a linear trend

• Mann-Kendall test for evaluating the existence of a monotonic trend

• WAVK test for assessing the presence of any non-monotonic trend via local regression

The classic t-test tests the presence of a linear trend in the data under the assumption that the time series may be autocorrelated, whiich is generally the most common case.

The hypotheses used for the tests are as follows [11]:

H0 (null hypothesis): No trend exists in the data.

H1 (alternative hypothesis): A linear trend exists in the data.

For determining a monotonic trend, the Mann-Kendall Trend test is used. Since the linear trend is also a monotonic trend, similar results can be expected as with the classic t-test [12].

The hypotheses used for this test are as follows:

H0 (null hypothesis): No trend exists in the data.

H1 (alternative hypothesis): A monotonic trend exists in the data.

The WAVK test using local regression is used to determine any non-monotonic trend.

The hypotheses used for this test are as follows [11]:

H0 (null hypothesis): No trend exists in the data.

H1 (alternative hypothesis): There is a non-monotonic trend in the data.

3.3 Seasonality analysis

Seasonality appears as recurring patterns in a time series linked to particular time periods. These patterns may result frrom multiple variables, including meteorological conditions, holidays, or consumer behaviour. Recognising seasonality enables data scientists to reveal fundamental trends and enhance forecast accuracy.

Some of the common methods for identifying seasonality include visual methods which include autocorrelation analysis and decomposition of time series. Although these methods may be effective, there are statistical tests such as the Kruskal-Wallis, and Canova-Hansen test which provide a robust statistical approach to confirm the presence of seasonality.

The Kruskal-Wallis test is a non-parametric statistical method used to determine whether statistically significant differences exist among three or more groups within a dataset. It is similar to Wilcoxon's Rank Sum test in that we are comparing the sum of ranks applied to the data. The test statistic is calculated as [13]:

$$K = \frac{12}{N*(N+1)} \sum_{i=1}^{k} \frac{{R_i}^2}{n_i} - 3*(N+1) \quad (3)$$

where:

K = calculated value of K statistics,

R_i = the sum of ranks for the ith group,

N = data set

Canova and Hansen proposed a test statistic that the seasonal pattern is stable. It introduces Lagrange multiplier tests of the null hypothesis of no unit roots at seasonal frequencies against the alternative of a unit root at either a single seasonal frequency or a set of seasonal frequencies. The tests complement those of Dickey, Hasza, and Fuller and Hylleberg, Engle, Granger, and Yoo that examine the null of seasonal unit roots [15].

Autocorrelation or ACF function is crucial in the identification of time series processes. It represents the correlation between the values of a time series and its lags over a period of time. The auto-correlation function at lag n, indicated by ρ_n is denoted as:

$$\rho_n = \frac{\gamma_n}{\gamma_0} \quad (4)$$

where:

ρ_n = autocorrelation at lag n,

γ_n = covariance of lag n,

γ_0 = standard deviation of lag n.

4 RESEARCH METHODOLOGY

This research utilised the R programming language and RStudio IDE (version 2021.09.0 Build 351) as the software tools. R is a statistical programming language extensively utilised in academia for various statistical computations, data analysis, and data visualisation. CRAN, a comprehensive R archive network, works as a central repository for R software packages, which consist of functions, data, and documentation that enhance the capabilities of the R programming language. CRAN is available through its website (http://cran.r-project.org), where users can search for packages, download R, and obtain documentation regarding R and its packages [16].

The raw data were gathered by multiple sensors from the sensor station situated in Kać, in the municipality of Novi Sad, from January 2014 to December 2020. Upon data collection, they are transferred to the online portal. Subsequently, data collected from various sensors can be stored in Excel format. The sensors are part of a network category known as "Proactive Networks." The sensors collect data hourly, resulting in a total of 288 measurements over a 24-hour period. The sensors can be classified into seven groups. The initial category has six sensors exclusively designed to measure soil moisture (SM1, SM2, SM3, SM4, SM5, SM6). Additional categories include a humidity sensor (AH1), wind speed (WS1), wind direction (WD1), air temperature (AT1), precipitation (PP1), and battery power status (BT1). Only the air temperature time series will be used for applyinh EDA techniques and data analysis [17].The dataset consists of four attributes: time, device, ID value, and value. Table 1 [18] provides descriptions of all the features in the dataset.

Table 1 –Structure of the data set

Attribute name	Description
Time	date and time of the measured parameter
Device	the device identification number of the measured parameter
ID value	sensor identification number of the measured parameter
Value	the measured parameter value

Data are retrieved from the online portal in Excel format and then saved in a data frame. A data frame is a two-dimensional, tabular data structure used in R and other computer languages for the management and analysis of data in a structured fashion for statistical and data science applications. The preparation technique initially involves filtering out the parameters that wont be used and then aggregating the values of air temperature using the "mean" function to obtain the average values. Ultimately, the data are saved as a data table and organised in ascending order by months and years. Upon preparation of the data, EDA techniques may be applied [18].

5 RESULTS AND DISCUSSION

The first technique that will be used is the decomposition of time series by using the STL method. The STL method is implemented in the R language using the "stl()" function found in the R package *stats*. This function is used to decompose the air temperature time series into its components using the additive decomposition model, and then, if necessary, those components can be removed. This function contains two main arguments, s.window and t.window, which control how fast the trend and seasonal component can change. The first argument represents the number of consecutive years used to evaluate the value in the seasonal component. It has no default value and must be defined, and the value "periodic" is usually set. The second argument is the number of consecutive observations of the time series that are used to estimate the values in the trend component. If this value is not explicitly defined, the default value is taken [8].

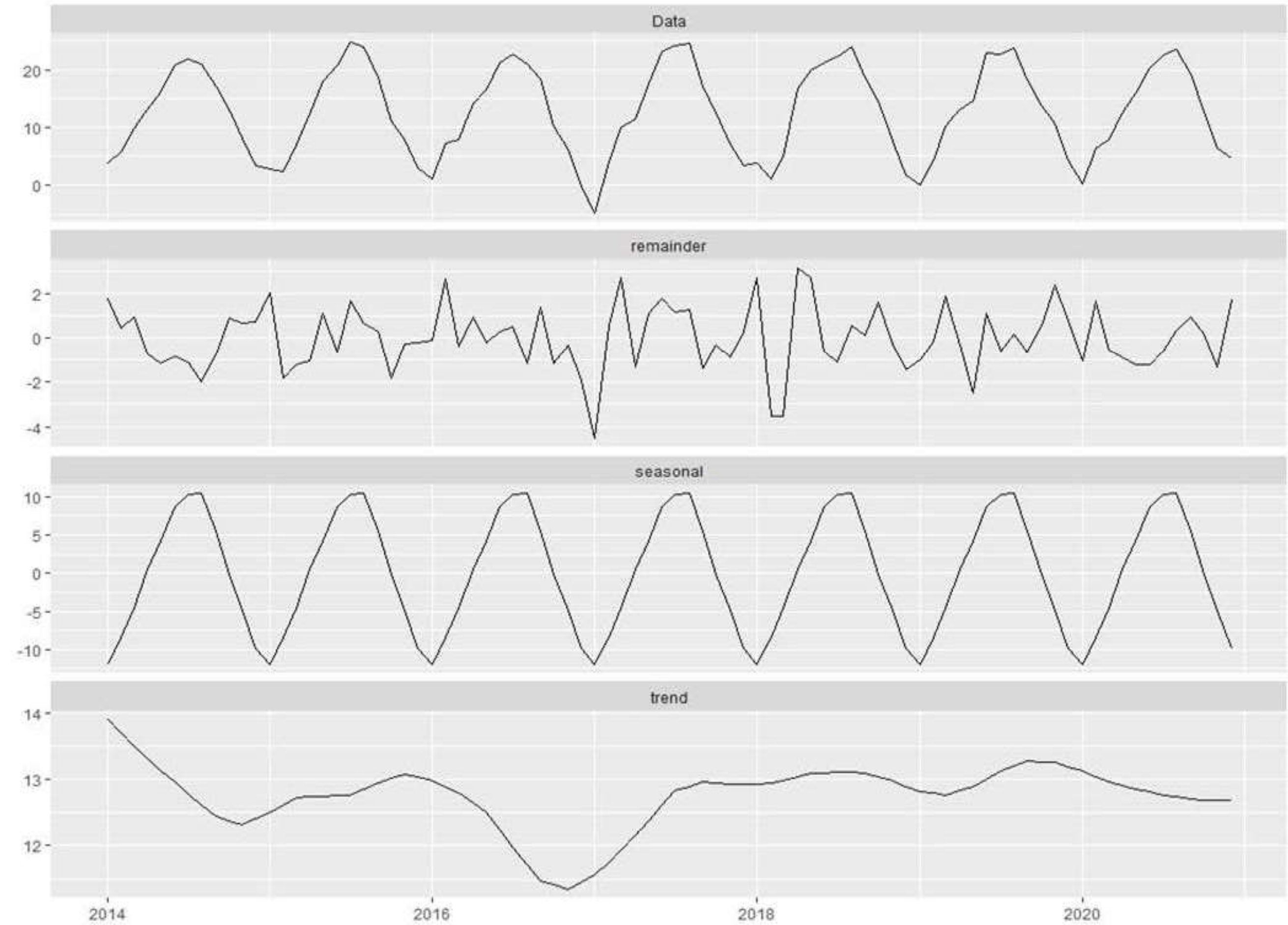


Figure 1. Decomposed air temperature time series with all its components

Figure 1 shows four panels. The bottom three panels show the three components (trend, seasonality, and error) separately. If these three components were added together, the raw data shown in the first panel above could be reconstructed. As can be seen in the figure, the seasonal component shown in the third panel is very obvious, and it changes slowly so that the pattern is quite similar for each year. The remainder component shown in the second panel symbolises the residual left over after subtracting the two remaining components from the data. The last panel shows the trend of the data series, which seems non-linear.

The classic t-test tests the presence of a linear trend in the data under the assumption that the time series may be autocorrelated, which is generally the most common case. A classical t-test is performed using the "notrend_test()" function from the *funtimes* package which will test the data set of the temperature time series for a significance level of α = 0.05. If the p-value of the test is less than the significance level, then the null hypothesis can be rejected, and it can be said that there is statistically significant evidence that a linear trend is present in the data.

Figure 2 shows the results of the executed classical t-test, where it can be seen that the p-value of 0.914 is significantly higher than the significance level of 0.05, which indicates that the null hypothesis cannot be rejected. Therefore, the conclusion is that there is no linear trend in the time series.

```
        Sieve-bootstrap Student's t-test for a linear trend

data:  temperatura.TS
Student's t value = 0.35744, p-value = 0.914
alternative hypothesis: linear trend.
sample estimates:
$AR_order
[1] 10

$AR_coefficients
      phi_1       phi_2       phi_3       phi_4       phi_5       phi_6       phi_7       phi_8 
 0.77476739 -0.10086897  0.11274414 -0.09014890 -0.02198085 -0.04585700 -0.10057103  0.12518475 
      phi_9      phi_10 
-0.12970698  0.45326079 
```

Figure 2. The results of the classical t-test for air temperature time series

To determine a monotonic trend, the Mann-Kendall trend test is used. Since the linear trend is also a monotonic trend, similar results can be expected as with the classic t-test. The "notrend_test()" function performs the Mann-Kendall Trend test, using the value MK as the test parameter.

```
data:  temperatura.TS
Mann--Kendall's tau = 0.033276, p-value = 0.888
alternative hypothesis: monotonic trend.
sample estimates:
$AR_order
[1] 10

$AR_coefficients
      phi_1       phi_2       phi_3       phi_4       phi_5       phi_6       phi_7       phi_8       phi_9      phi_10 
 0.77476739 -0.10086897  0.11274414 -0.09014890 -0.02198085 -0.04585700 -0.10057103  0.12518475 -0.12970698  0.45326079 
```

Figure 3. The results of Mann-Kendall Trend test for air temperature time series

Figure 3 shows the execution results of the Mann-Kendall trend test. The test statistic is 0.033276, and the p-value is 0.888. Since this p-value is much higher than 0.05, the null hypothesis of the test cannot be rejected, which leads to the conclusion that even a monotonic trend is not present in the air temperature time series.

A WAVK test using local regression is used to determine any non-monotonic trend. The test is performed using the "notrend_test()" function, where the value "WAVK" is taken as the test parameter.

```
data:  temperatura.TS
WAVK test statistic = 7.4696, moving window = 8, p-value = 0.911
alternative hypothesis: (non-)monotonic trend.
sample estimates:
$AR_order
[1] 10

$AR_coefficients
      phi_1       phi_2       phi_3       phi_4       phi_5       phi_6       phi_7       phi_8 
 0.77476739 -0.10086897  0.11274414 -0.09014890 -0.02198085 -0.04585700 -0.10057103  0.12518475 
      phi_9      phi_10 
-0.12970698  0.45326079 
```

Figure 4. The results of WAVK test for air temperature time series

In Figure 4, you can see the results of the WAVK test, where the p-value of 0.911 is much higher than the significance level of 0.05 and the null hypothesis cannot be rejected. The test suggests that there is no non-monotonic trend in the time series, which can be visualized on the graph by adding a dotted red line showing the trend (Figure 5).

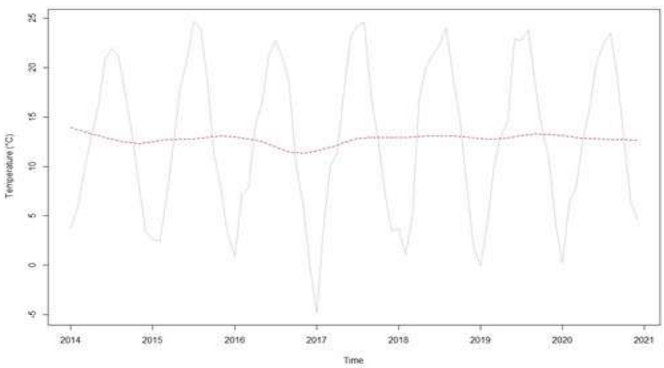


Figure 5. Dsiplay of air temperature time series with trend visualisation

Autocorrelation diagrams, or ACF diagrams, show the correlation between a certain time series and its lags. If there is a correlation between the series and its lags it implies that there are some trends or a seasonal component

in it, and therefore its statistical properties are not constant over time. The "acf()" and "pacf()" functions in R can display the autocorrelation plots separately from each other. If we want to display the combination of the time series together with its autocorrelation and partial autocorrelation on one graph, we use the "PlotACF()" function from the *DescTools* package. The lag.max parameter indicates the number of lags that will be displayed on the diagrams, and in our case it is 36.

Figure 6 displays a graph of the air temperature time series along with ACF and PACF diagrams. This shows that the time series has seasonality and that the ACF function is slowly decreasing, which means that the data is not stationary. There are important positive and negative autocorrelations with a lot of initial lags that are well outside the 5% significance limits shown by the blue dashed line in the ACF plot. These lags are 1, 5, 6, 7, 11, 12, and 13. The PACF diagram only shows jumps at delays 1, 2, 3, 4, 5, and 6. The other delays are within the acceptable range so this could be due to the carry-over correlation from the initial and earlier delays.

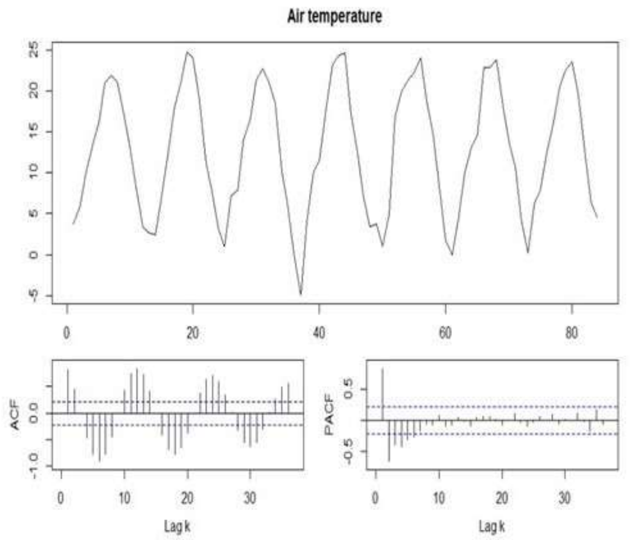
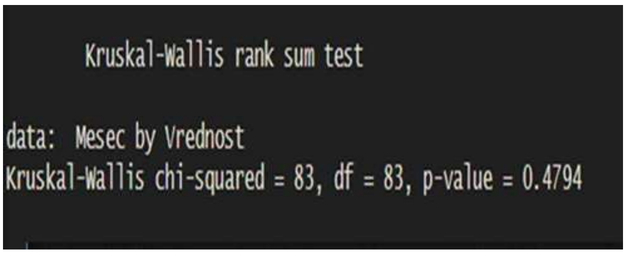


Figure 6. Display of air temperature time series with its ACF and PACF diagrams

Figure 7 shows the results of the Kruskal-Wallis rank sum test for the air temperature time series with two groups: Mesec (engl. Month) and Vrednost (engl. Value). The values are compared across months only. This test suggests that there are no significant differences between the two groups, which means that the average values of air temperature time series do not differ significantly across months because the p-value of 0.4794 is higher than the significance level of 0.05.

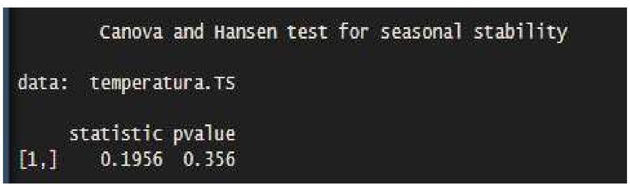


```
        Kruskal-Wallis rank sum test

data:  Mesec by Vrednost
Kruskal-Wallis chi-squared = 83, df = 83, p-value = 0.4794
```

Figure 7. Results of Kruskal-Wallis rank sum test for the air temperature time series

The Canova-Hansen test for seasonal stability is part of the *uroot* package. The test is performed using the "ch.test()" function, where the frequency is 12 since the data are monthly averages. As you can see in Figure 8, the Canova Hansen test showed that the p-value was 0.356, which is higher than the significance level of 0.05This means that there is a 95% chance that the null hypothesis is not true. Therefore, this test suggests that there is no unit root in the seasonal component nor strong evidence of a seasonality pattern in the time series.



```
        Canova and Hansen test for seasonal stability

data:  temperatura.TS

     statistic pvalue
[1,]    0.1956  0.356
```

Figure 8. Results of Canova Hansen test for the air temperature time series

5 CONCLUSION

Exploratory data analysis (EDA) is a powerful technique that not only improves forecasting precision but also guarantees the application of suitable modelling techniques—such as ARIMA, LSTMs, or Prophet—according to the inherent patterns in the data. Furthermore, finding time series features like stationarity and detecting trend and seasonality patterns during exploratory data analysis guarantees data integrity, which is crucial for making accurate and reliable predictions.

EDA techniques were performed on air temperature time series for detecting trend and seasonality patterns. Since no seasonality was found by the Canova-Hansen test, alternative seasonality tests and visual methods were also taken into account when searching for seasonality.Visual inspections using STL decomposition and ACF and PACF functions indicated strong seasonality repeating patterns included in the air temperature time series. According to the seasonality and trend tests performed, the conclusion was that the data is non-stationary, and there were no signficant differences between tested Month and Value groups. Also, a strong seasonality pattern was detected in the time series which both the STL decomposition and ACF function confirmed. On the other side, trend tests indicated that there was no trend in the time series - no linear, monotonic or any trend.

6 REFERENCES

[1] Kousis, A. &Tjortjis, C. (2021). *Data Mining Algorithms for Smart Cities:A Bibliometric Analysis*. Algorithms, 14(8), pp.242, doi: 10.3390/a14080242

[2] Chen, D., Zhang, J. & Jiang, S. (2020). *Forecasting the Short-Term Metro Ridership With Seasonal and Trend Decomposition Using Loess and LSTM Neural Networks*, IEEE Access, vol. 8, pp. 91181–91187, 2020, doi: 10.1109/ACCESS.2020.2995044

[3] Sarker, I. H. (2022). *Smart City Data Science: Towards data-driven smart cities with open research issues*, Internet of Things, vol. 19, pp. 100528, doi: 10.1016/j.iot.2022.100528

[4] Strohbach, M., Ziekow, H., Gazis, V. & Akiva, N. (2014). *Towards a Big Data Analytics Framework for IoT and Smart City Applications'*,book:Modeling and Processing for Next-Generation Big-Data Technologies, Editors: Xhafa, F., Barolli, L., Barolli A. & Papajorgji, P. Springer International Publishing, pp. 257–282. doi: 10.1007/978-3-319-09177-8_11.

[5] Shorfuzzaman, M. & Hossain, M. S. (2022). *Predictive Analytics of Energy Usage by IoT-Based Smart Home Appliances for Green Urban Development*, ACM Trans. Internet Technol., 22(2), pp. 1–26, doi: 10.1145/3426970.

[6]Murugan, R. & Palanichamy, N. (2021). *Smart City Air Quality Prediction using Machine Learning*, 5th International Conference on Intelligent Computing and Control Systems (ICICCS),India:IEEE, pp.1048–1054, doi: 10.1109/ICICCS51141.2021.9432074

[7] Auffarth, B. (2021). *Machine Learning for Time-Series with Python*, Packt Publishing, Birgmingham, the UK, ISBN 978-1-80181-962-6.

[8] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*, 3nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on 28.01.2025.

[9] Gakhov, A. V. (2018).*An Introduction to Time Series Forecasting with Python*, PyCon UA 2018, Kharkov, Ukraine. DOI: 10.13140/RG.2.2.18053.86249.

[10]Cleveland, R. B., Cleveland, W. S., McRae, J. E., &Terpenning, I. J. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess.*Journal of Official Statistics; Stockholm*, Statistics Sweden (SCB), 6(1), pp.3-73, ISSN 0282423X.

[11]https://cran.rproject.org/web/packages/funtimes/vignettes/trendtests.html, accessed 28.01.2025.

[12] Alhaji, U. U., Yusuf, A. S., Edet, C. O., Oche, C. O., &Agbo, E. P. (2018). Trend Analysis of Temperature in Gombe State Using Mann Kendall Trend Test.*Journal of Scientific Research and Reports*, 20(3), pp.1-9, ISSN: 2320-0227, DOI: 10.9734/JSRR/2018/42029

[13] Gauthier, T. D., Hawley, M. E. (2015). Introduction to Environmental Forensics (Third Edition); Chapter 5 - Statistical Methods. Academic Press; Editor(s): Brian L. Murphy, Robert D. Morrison, pp. 99-148, ISBN 9780124046962, https://doi.org/10.1016/B978-0-12-404696-2.00005-9.

[14] Mohr, D. L., Wilson, W. J., Freund, R.J. (2022). Statistical Methods (Fourth Edition), Chapter 14 - Nonparametric Methods. Academic Press; Editor(s): Donna L. Mohr, William J. Wilson, Rudolf J. Freund, pp. 651-683, ISBN 9780128230435, https://doi.org/10.1016/B978-0-12-823043-5.00014-X.

[15] Canova, F. & Hansen, B. E. (1995). *Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability*. Journal of Business & Economic Statistics, 13(3), pp.237–252. https://doi.org/10.1080/07350015.1995.10524598

[16] Tihi, N., Popov, S., Bondžić, J. & Dujović, M. (2021). *Visualization of Big Data as Urban Drought MOonitoring Support in Smart Cities*, Fresenius Environmental Bulletin (FEB), 30(01), pp. 719–723.

[17] Tihi, N.&Popov, S. (2023). *Selection of the Best ARIMA models for Urban Drought Prediction*, Fresenius Environmental Bulletin (FEB), 32(06), pp.2564–2572.

[18] Tihi, N. & Popov, S. (2024). A Comparison of Arima and Random Forest Time Series Models for Urban Drought Prediction, in Proceedings of the International Scientific Conference - Sinteza 2024, Beograd, Serbia: Singidunum University, pp. 51–56. doi: 10.15308/Sinteza-2024-51-56.

Contact information:

Ninoslava TIHI, MSc, lecturer
(Corresponding author)
1979
The Higher Education Technical School of Professional Studies in Novi Sad,
21000, Novi Sad, Serbia
J-Mail:tihi@vtsns.edu.rs
https://orcid.org/0009-0004-8009-8120

Miloš TODOROV, PhD, Assistant professor
1982
Faculty of Mathematics and Computer Science, Alfa BK University,
11000, Belgrade, Serbia
E-Mail:milos.todorov@alfa.edu.rs
https://orcid.org/0000-0002-5614-3057

Marko PAVLOVIĆ, Undergraduate student
Faculty of Business and Law, MB University11000, Belgrade, Serbia
E-Mail: markuspavlovic98@gmail.com
https://orcid.org/0009-0009-5526-1244

Filip KOKALJ, PhD, Associate professor
1973
Faculty of Mechanical Engineering, University of Maribor
2000, Maribor, Slovenia
E-Mail:filip.kokalj@um.si
https://orcid.org/0000-0001-9234-7834