International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6148
***Article Received**: 22 November 2022 **Revised**: 26 December 2022 **Accepted**: 28 January 2023*

Comparison of Imputation Methods for Univariate Time Series

Geeta Chhabra*
Dy. Director, Data Informatics & Innovation Division, Ministry of Statistics & PI, Govt. of India, India, geeta_chhabra@rediffmail.com

Abstract
In order to predict and forecast with greater accuracy, handling “missing values” in “time series” information is crucial. Complete and accurate historical data are essential. There are many research studies on multivariate time series imputation, however due to the lack of associated factors, imputation in univariate time series data is rarely taken into consideration. It is natural that “missing values” could arise because almost all scientific disciplines that collect, store, and monitor data use "time series" observations. Therefore, time series characteristics must be considered in order to develop an effective and acceptable method for dealing with missing data. This work uses the statistical package R to assess and measure the effectiveness of imputation methods in the context of "univariate time series" data. The “imputation algorithms” explored are evaluated using “root mean square error”, “mean absolute error” and “mean absolute percent error”. Four types of “time series” are taken into consideration. According to experimental findings, “seasonal decomposition” performs better on the time series having seasonality characteristic, followed by “linear interpolation”, and “kalman smoothing” provides values that are more similar to the original time series data set and have lower error rates than other imputation techniques.

Keywords: Univariate, time series, kalman, interpolation, missing values.

I. Introduction

Numerous disciplines, including economics (Yang, 2012), energy research (Mohamad et al., 2021), environmental studies (Hadeed et al., 2020), signal processing (Stankovic et al., 2014), traffic engineering (Ran et al., 2015) and ecology (Hossie et al., 2021), among others, can benefit from “time series” data analysis. In order to facilitate the implementation of policies or the deployment of control mechanisms, time series data can be analysed using a number of methodologies that explain emergent data patterns and forecast future behaviour. The accuracy of the data and the comprehensiveness of the supplemental information are requirements for information extraction from time series data.

Missing observations can frequently happen while measuring, collecting, or creating data as a consequence of various factors, such as communication failures, data-generating device failures and power failures. In analytical research, “missing data” may result in flawed and unwanted results, such as incorrect projections or poor policy judgements (Phan, 2020). Therefore, techniques to substitute “missing data” are required.

In "time series", "missing data" can be replaced by either "imputation-based" or "model-based" techniques. The “model-based” techniques are different from “direct imputation” in that they solve likelihood equations applied to “missing data”. The “imputation-based” techniques on the other hand, estimate missing values by either completely removing them or replacing with appropriate values via general approach. In contrast to “model-based methods” for “multivariate time series”, most of techniques replacing the “missing observations” in “univariate time series” are “imputation-based”. The “univariate time series” imputation methods are classified as: “univariate algorithms”, “univariate time series algorithms” and “multivariate algorithms” for “lagged data” (Moritz et al., 2017).

Other research, that does not explicitly take into account the statistical features of “time series”, highlighted many drawbacks of "univariate time series" imputation approaches, which were also noted by Moritz et al. (2015). One of the most popular approaches for univariate time series is "last observation carried forward", which combines "interpolation" and "arithmetic mean". In general, more reliable imputation techniques are needed for univariate time series, especially ones that can make better use of the statistical properties of the observations. In terms of imputation and prediction accuracy, we compared “mean”, “last observation carried forward”, “kalman smoothing”, “seasonal decomposition using interpolation”, “seasonal imputation using mean”, “moving average” and “moving average with exponential weighting”, “linear interpolation”, “spline interpolation”, and “stine interpolation” for single variate data, namely tsAirgap, tsNH4, and tsHeating, are all available in the R-package imputeTS. The same techniques was used on real-time consumer price index data downloaded from the M/o Statistics website. The precision is expressed as

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6148
***Article Received**: 22 November 2022 **Revised**: 26 December 2022 **Accepted**: 28 January 2023*

"root mean square error", "mean absolute error and "mean absolute percentage error".

II. Related work

Fewer studies have been conducted on the imputation of "missing data" for "univariate time series".

In their article "A Method for Improving Imputation and Prediction Accuracy of Highly Seasonal Univariate Data with Large Periods of Missingness," Chaudhry et al. (2019) used LTE spectrum data, which is highly seasonal and univariate. They used Kalman filtering, which is defined as Kalman smoothing on an ARIMA model's state space representation, and MICE. They converted the univariate data to multivariate for MICE imputation. They evaluated their proposed method using mean absolute percentage error (MAPE) metrics.

In his article titled "Imputation Methods in Time Series with a Trend and a Consecutive Missing Value Pattern", Wongoutong et al. (2021), compared ten real datasets to assess how well imputation methods performed under three different scenarios involving artificial missing data in "time series" with different ratios of missing values. The evaluation of six methods to impute "missing values"—"interpolation", "kalman", "moving average", "last observation carried forward", "mean" and "linear trend at point"— were explored in terms of "root-mean-square error" and "mean absolute percentage error". The "interpolation", "kalman" and "linear trend at point" imputation methods outperformed the other three by an average of 80% when compared to the "mean" imputation method and 30-60% when compared to the "last observation carried forward" and "moving average" methods. They came to the conclusion that for "time-series" with trend, "interpolation", "kalman" and "linear trend at point" performed better for imputing successive "missing values".

Han et al. (2022) proposed a "univariate imputation" approach for integrating decomposition method with imputation algorithms in their article titled "Univariate imputation method for recovering missing data in wastewater treatment process." To cope with the nonstationary properties of wastewater treatment process data, the "time series" is first divided into "seasonal", "trend", and "remainder" using "seasonal-trend decomposition". Second, estimates of its missing values are provided by using "support vector regression" to roughly estimate "nonlinearity" of "trend" and "remainder", respectively. Based on its periodic pattern, a "self-similarity decomposition" is used to fill the "seasonal component". Third, the imputation result is created by combining all of the imputed results. The imputation performance is then assessed using six time series of the wastewater treatment process and based on two indicators, compared to seven other methods. The experimental findings show that, the suggested "univariate imputation" is better for "time series" of wastewater treatment processes with various missing ratios.

In their article "On imputation approaches in univariate time series," Rantou et al. (2017) used the statistical programme R to assess the effectiveness of "imputation algorithms" in case of "univariate time series data". The "imputation methods" are evaluated by three fundamental types of "time series" and error metrics namely; "mrse" and "mape".

In their work "Local Average of Nearest Neighbors: Univariate Time Series Imputation," Flores et al. (2019) introduced two imputation techniques for the "missing data" in "univariate time series". These algorithms have used two "algorithms" based on "means" of "nearest neighbours". The first is the neighbourhood average. Neighbors determines the "missing value" by averaging the values of the neighbour before it and the neighbour after it. The second one is "Local Average of Neighbors Neighbors+(LANN+)" that uses the distance between neighbours.

In their article "Efficiency of Imputation Techniques in Univariate Time Series," Twumasi-Ankrah et al. (2019) used "imputation method" for "univariate time series" missing values, depending on specific error metric and characteristics.

III. Missing Data Mechanism

The distribution of the gaps will depend on what produces missing data. In two ways, comprehending this distribution might be beneficial, can be used as information for choosing suitable "imputation algorithm" and by using a realistic simulator that will eliminates "missing data" from the test dataset. A simulator of this kind will assist in producing data for which, the real value is known, so that effectiveness of "imputation algorithm's" can be evaluated.

"Missing data" mechanisms specify the relationship between variables that are observed and those that are missing. There are three basic groupings: "missing entirely at random (MCAR)", "missing at random (MAR)", and "missing not at random (MNAR)". A variable is missing if it is neither dependent on observed variables and nor on itself. For instance, a house's number of fireplaces is independent of itself. It is MNAR if the missingness of a variable is related to itself. For instance, the proximity to the market may be a key consideration when renting or purchasing a home because it is quite convenient to get there on walking. It is MAR if the absence of a variable depends on another variable. The use of correlations with other variables by imputation algorithms is made possible by MAR, which leads to better results than MCAR and MNAR. For instance, if a home lacks a garage, the garage's capacity or quality will always be lacking (Baddoo et al. 2021).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6148
***Article Received**: 22 November 2022 **Revised**: 26 December 2022 **Accepted**: 28 January 2023*

The picture of “missing data” mechanisms for “univariate”, “temporal series” appears little different as the data only appears to have one variable, while time is implicitly assumed that is considered to be a variable when developing a dataset's mechanism. Another distinction is that “time series” imputation methods can use “time series” properties also besides variables to estimate missing values. As a result, it is much simpler to estimate “missing values” for MCAR data. MAR and MCAR are essentially equivalent for “univariate time series” imputation.

IV. Time series imputation for univariate data

A "univariate time series", is a “time series” that has only one observation that is progressively recorded at equal time intervals. Imputation is the process of substituting estimations for “missing data”. The following studies employs “imputation techniques” for “univariate time series”;

4.1 Mean Imputation

Mean of the “observed values” of the “non-missing observations” is calculated that replaces “missing values” with mean. This method comes from R's imputeTS packages. It uses the function "na_mean" to substitute the “missing values” in “time series”. The following formula is used to estimate the value;

$$\overline{x_i} = \sum_{i=0}^{k} y_i \,/ n_k \qquad (1)$$

where n_k is the number of observations and y_i, is the observed values.

4.2 Last Observation Carried Forward (LOCF)

On orderly sorted dataset, algorithm locates the first “missing value” and then it is imputed with non-missing value immediately preceding the missing data. The method in R employs imputeTS package, and the function is "na_locf" (Phan et al., 2020).

4.3 Kalman smoothing

The estimates of unknown variables are produced by a series of measurements that are observed over time, including inaccuracies and noise. These estimates, which use joint probability distribution over the variables for each timeframe, are more precise than those based on a single measurement alone. The algorithm follows two steps. For the prediction phase, Kalman smoothing generates estimates of the current state variables along with their uncertainty. When the subsequent measurement is taken, these estimates are updated using a weighted average, with a higher weight given to more precise estimates. It is iterative algorithm. Without requiring any prior knowledge, it may operate in real time using only the most recent input measurements, the previously calculated state, and its uncertainty matrix. It assumes that the errors have “normal distribution” (Jeong, 2021).The “kalman smoothing” to operate on state-space models is of the form;

$$y_t = Z\alpha_t + \epsilon_t \epsilon_t \sim N(0, \sigma^2) \qquad (2)$$

$$\alpha_{t1} = T\alpha_t + n_t n_t \sim N(0, \sigma_1^2) \qquad (3)$$

$$\text{where} \quad \alpha_t \sim N(a_t, \sigma_t^2)$$

where y_t = observed data and α_t = unobserved.

The measurement equation, y_t means the “observed data” is related to the “unobserved” states, α_{t1}, transition equation, implies the “unobserved” states evolve over time in a particular way.“Kalman smoothing” uses “algorithm” to find best estimates of α_t. The “kalman smoothing” has been applied to the entire time period to get the estimates of the states a_t, σ_t^2 at t=1,2,. ,T. It employs imputeTS package and the function “na_kalman” to replace “missing values” in R.

4.4 Seasonally Decomposed Missing Value Imputation

The algorithm starts with a “Loess Seasonal Decomposition” of “time series”. "Time series" is split into "seasonal", "trend" and "irregular". The original series' "seasonal" component is then removed. In a subsequent step, the deseasonalized series is subjected to the selected imputation algorithm, such as na_locf, na_ma, etc. As a result, the algorithm is unaffected by "seasonal" patterns. The "seasonal" component is reintroduced into the de-seasonalized series after the NA gaps are replaced. It uses impute TSpackage in R with function "na_seadec".

4.5 Moving Average (MA) Imputation

The average is calculated using the same number of data points on either side of the central value in this algorithm. This means that if a "missing value" occurs at position i of a "time series" data set, the average is computed using observations i-1, i-2, and i+1, i+2.The package imputeTS and method "na_ma" are used in R.

The weighting factors decreases exponentially in an “exponential weighted moving average”. The observations immediately adjacent to a central value i have a weight of 1^2, the observations further away (i-2, i+2) have a weight of ½^2, the observations further away (i-3,i+3) have a weight of ½^3, and so on.

4.6 Imputation by Interpolation

The na_nterpolation function with the parameters "linear," "spline," or "stine" interpolation is used by the imputeTS package. The "least square principle" minimises the sum of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6148
***Article Received**: 22 November 2022 **Revised**: 26 December 2022 **Accepted**: 28 January 2023*

squares of errors for a polynomial of a given degree. The algorithm for "linear interpolation" is as follows:

If k is the number of "missing data" points in a given "time series" dataset, and $a_1, a_2, \ldots a_{k-1}$ are constants, and y_t represents the "missing observation" at time t, consider fitting a polynomial of fixed degree k.

$$y_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_{k-1} t^k \quad (4)$$

This is calculated using the time series data's observed values. The matrix approach is used to obtain the values of a_0, a_1, $a_2 \ldots a_{k-1}$. Missing values are calculated at each iteration, t=1,2,3,.., T.

V. Error Metric

We evaluate the effects of three error metrics on imputation strategies.

5.1 Root Mean Square Error (RMSE)

It measures the spread of predicted errors over actual data points means, that it indicates how far or close an estimated model's predicted values are to the actual data points(Bokde et al., 2018). The formula is as follows:

$$\text{RMSE}=\sqrt{\sum_{i=1}^{N}\left(\frac{Y-\overline{Y}}{N}\right)^2} \quad (5)$$

where N is the sample size, Y is the actual data and $\overline{Y}$ is the predicted data.

5.2 Mean Absolute Error (MAE)

It is the average of absolute errors, which is the, magnitude of difference between the actual and predicted values. It tells how large an error is from the predicted, is expected on average.

$$\text{MAE}=\frac{1}{N}\sum\left|Y-\overline{Y}\right| \quad (6)$$

5.3 Mean Absolute Percentage Error (MAPE)

It measures the, percentage forecast error. It is used to calculate the forecast accuracy using:

$$\text{MAPE}=\left(\frac{1}{N}\sum\frac{\left|Y-\overline{Y}\right|}{\left|\overline{Y}\right|}\right)X\ 100\% \quad (7)$$

VI. Experimental Analysis

The performance of 10 imputation algorithms is examined in this study using three reference time series datasets from the imputeTS package (Moritz et al, 2015) and one real time series data set downloaded from M/o Statistics & PI, Govt. of India website, http://www.mospi.gov.in. These datasets are widely used in the literature having well-known properties by all time series data.

6.1 Datasets

Following are the four datasets that were used in this study:

1.Air passengers: The dataset comes from "Time series analysis: forecasting and control" (Box et al., 2015) and includes 144 monthly total passengers of international airline from 1949 to 1960. The dataset shows a strong trend as well as seasonality. There are two time series provided for comparing imputation algorithm results with this series.One series with no missing values that can be used as the basis for further analysis and the "imputation algorithm" can be applied on another NA-based series.

2. Wastewater system: The "time series" was created using data from the 2014 GECCO Industrial Challenge (Martina et al., 2014). It has 4552 rows, measured in 10-minute increments from 30.11.2010 to 01.01.2011. There are two time series provided for comparing imputation algorithm results with this series. One series with no missing values that can be used as the basis for further analysis, and another is NA-based series that is used for "imputation algorithms". The dataset shows significant seasonality but no trend.

1. **Heating systems supply temperature:** The "time series" was created using data from the GECCO Industrial Challenge 2015. (Moritz et al., 2015), and it was measured in 1 minute steps from 18.11.2013 - 05:12:00 to 13.01.2015 - 15:08:00. There are 606837 rows in the Time Series. "Recovering missing information in heating system operating data" was the topic of this Challenge. The goal is to effectively substitute "missing values" in sensor data from a heating system. There are two time series provided for comparing imputation algorithm results with this series. One series with no missing values that can be used as the basis for further analysis and another NA-based series that is used for "imputation algorithms".There is no trend or seasonality in the dataset.

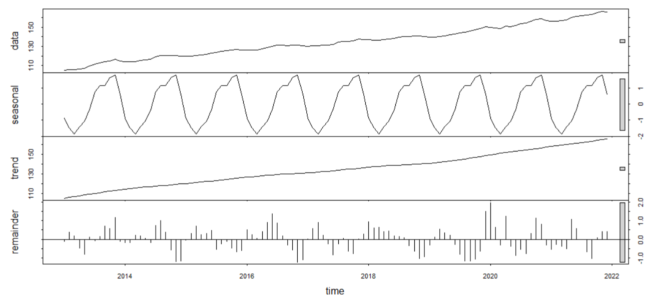

Figure 1: Loess seasonal decomposition of Consumer Price I ndex Dataset

4.Time series of consumer price index: The data has been downloaded from M/o Statistics & PI, Govt of India website,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6148
***Article Received**: 22 November 2022 **Revised**: 26 December 2022 **Accepted**: 28 January 2023*

http://www.mospi.gov.in. It is a time series data yearly measured from 2013 to 2021. It has 108 rows. Missing values around 10 % has been artificially simulated. These missing values are then imputed and compared with the actual data. The dataset exhibits strong trend and seasonality (Figure 1).

6.2 Line plot to visualize the missing values distribution

The "ggplot_na_distribution" function from imputeTS pack age depicts the distribution of "missing values" within "ti me series". As a result, "time series" is plotted and whenever there is NA, the back color appears differently. The plot fo r time series data can be seen below (Figure 2).

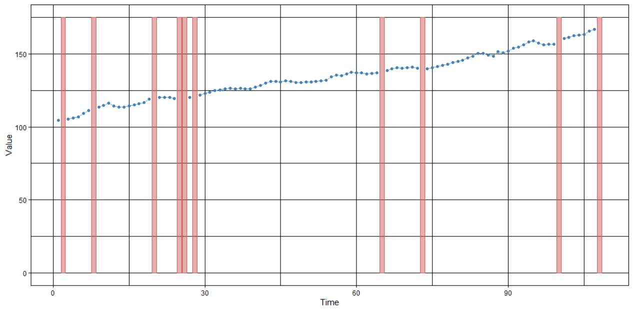

Figure 2: Distribution of Missing Values for Consumer Price Index Dataset
(Time Series with Highlighted Missing Region)

6.3 Missing Values Statistics

Summary statistics for the "missing values" distribution in "univariate time series" are printed by the "statsNA" function in imputeTS package and it is summarized for each dataset in Table1.

Table 1:Summary of Datasets

Dataset	Length of Time Series	Missing Values	Percentage of Missing Values	No of Gaps	Avg Ga p Size	Longest NA gap (seri es of consecutive NA s)	Most frequent gap size
tsAirgap	144	13	9.03	11	1.1818	3 in a row	1 NA in a row (occurri ng 10 times)
tsNH4	4552	883	19.4	155	5.6968	157 in a row	1 NA in a row (occurri ng 68 times)
tsHeating	60683 7	57391	9.46	2087	27.4993	258 in a row	2 NA in a row (occurri ng 104 times)
tscpi	108	10	9.26	9	1.1111	2 in a row	1 NA in a row (occurri ng 8 times)

VII. Conclusions

Given that most statistical techniques assume that the data is complete and free of "missing values". Missing data constitute the first challenge when developing prediction models. It might not be viable or even optimal to handle missing data in "univariate time series" using typical "imputation algorithms". Since they differ from multivariate, non-time series datasets in several ways, "univariate time series" need specific consideration. To carry out an effective imputation, time dependencies must be used in place of covariates.

Table 2: Comparison of accuracy

Method	Air Passenger			Waste Water Management			Heating System Supply Temperature			Consumer Price Index		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Mean	34.63	8.90	0.0318	3.79	1.44	0.1108	5.69	1.43	0.0234	5.34	1.30	0.0010
LOCF	10.66	2.81	0.0104	2.01	0.45	0.0619	3.44	0.63	0.0116	0.38	0.10	0.0007

Kalman Smoothing	**3.21**	**0.75**	**0.0028**	1.29	0.37	0.2055				0.18	0.04	0.0003
Seasonal decomposition	**1.92**	**0.47**	**0.0018**	**0.83**	**0.22**	**0.1688**	2.66	0.45	0.0080	**0.11**	**0.03**	**0.0002**
seasonal imputation using mean	34.77	9.33	0.0345	3.29	1.25	0.1024	4.61	1.09	0.0184	5.26	1.27	0.0010
Simple moving average	9.67	2.37	0.0085	1.36	0.34	0.0413	2.82	0.46	0.0081	0.32	0.090	0.0006
Exponential weighted moving average	8.90	1.92	0.0070	1.20	0.29	0.0368	2.80	0.45	0.0080	0.21	0.05	0.0004
Linear interpolation	6.09	1.57	0.0057	**1.06**	**0.26**	**0.0326**	**2.39**	**0.38**	**0.0065**	**0.13**	**0.04**	**0.0003**
Spline interpolation	5.50	1.40	0.0052	2.06	0.49	0.1187	5.64	0.68	0.0265	0.15	0.37	0.0003
Stine interpolation	6.04	1.51	0.0056	1.16	0.29	0.0351	**2.45**	**0.36**	**0.0064**	**0.11**	**0.03**	**0.0002**

The methods namely "mean", "last observation carried forward", "kalman smoothing", "seasonal decomposition using interpolation", "seasonal imputation using mean", "moving average" and "moving average with exponential weighting", "linear interpolation", "spline interpolation" and "stine interpolation" has been applied on tsAirgap, tsNH4, and tsHeating data, available in the R-package imputeTS. The precision is expressed as "rmse", "mae" and "mape" (Table 2). "Seasonal decomposition" performs better than other techniques on the time series having seasonality characteristic namely tsAirgap and tsNH4 having seasonality followed by "linear interpolation", and "kalman kmoothing". The same techniques were also applied on real data i.e. consumer price index data that exhibhits strong seasonality. In this case, also "seasonal decomposition using interpolation" followed by "linear interpolation" performs better. Also, plots at figure 3 and figure 4 clearly shows that imputed values are quite close to the truth values in case of consumer price index data.

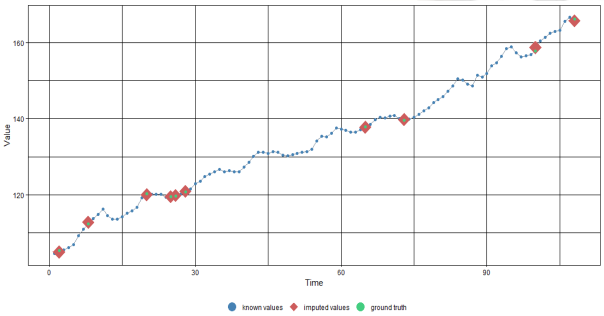

Figure 3: "Imputed Values for Consumer Price Index Data ", (Visualization of missing value replacements in case of seas onal decomposition)

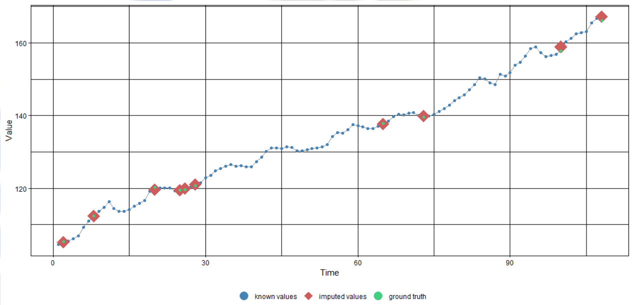

Figure 4: "Imputed Values for Consumer Price Index Data ", (Visualization of missing value replacements in case of Kal man Filtering)

The primary goal of this paper was to compare and quantify the performance of "imputation techniques" while dealing with "univariate time series". When handling "missing data" in "univariate time series", the results of our experiment indicates that seasonal decomposition perfoms well with the data having seasonality characteristics followed by "linear interpolation" and "kalman structural models using smoothing" as the most effective algorithms.

References

[1]. Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, *9*(1), 207. https://doi.org/10.32614/rj-2017-009

[2]. Moritz, S., Sardá, A., Bartz-Beielstein,T., Zaefferer, M. and Stork, J. (2015.). Comparison of different Methods for Univariate Time Series Imputation in R. https://doi.org/10.48550/arXiv.1510.03924.

[3]. Chaudhry, A., Li, W., Basri, A., & Patenaude, F. (2019, January 3). A Method for Improving Imputation and

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v11i2s.6148
Article Received: 22 November 2022 Revised: 26 December 2022 Accepted: 28 January 2023

Prediction Accuracy of Highly Seasonal Univariate Data with Large Periods of Missingness. *Wireless Communications and Mobile Computing*, 1–13. https://doi.org/10.1155/2019/4039758.

[4]. Wongoutong, C., (2021, October). Imputation Methods in Time Series with a Trend and a Consecutive Missing Val ue Pattern. *Thailand Statistician*, 19(4), 866-879. http://st atassoc.or.th.

[5]. Han, H., Sun, M., Han, H., Wu, X., & Qiao, J. (2022, April). Univariate imputation method for recovering missing data in wastewater treatment process. *Chinese Journal of Chemical Engineering*. https://doi.org/10.1016/j.cjche.2022.01.033

[6]. Rantou, K.,-E., Karagrigoriou, A., Vonta I. (2017).On imputation methods in univariate time series. Mathematics in Engineering, Science and Aerospace (MESA), 8(2).

[7]. Flores, A., Tito, H., & Silva, C. (2019). Local Average of Nearest Neighbors: Univariate Time Series Imputation. *International Journal of Advanced Computer Science and Applications*, *10*(8). https://doi.org/10.14569/ijacsa.2019.0100807.

[8]. Baddoo, T. D., Li, Z., Odai, S. N., Boni, K. R. C., Nooni, I. K., & Andam-Akorful, S. A. (2021, August 7). Compar ison of Missing Data Infilling Mechanisms for Recoverin g a Real-World Single Station Streamflow Observation. *I nternational Journal of Environmental Research and Pub lic Health*, *18*(16), 8375. https://doi.org/10.3390/ijerph18 168375

[9]. Phan, T. T. H., Poisson Caillault, M., Lefebvre, A., & Bigand, A. (2020, November). Dynamic time warping-based imputation for univariate time series data. *Pattern Recognition Letters, 139*, 139–147. https://doi.org/10.1016/j.patrec.2017.08.019.

[10]. Twumasi-Ankrah, S.,Odoi, B., Pels, A., W. & Gyamfi, H.,E. (2019). Efficiency of Imputation Techniques in Univariate Time Series. *International Journal of Science, Environment and Technology*, *8*(3), 430–453.

[11]. Hadeed, J.,S., O'Rourke, M., K., Burgess, L.J., Harris, R., B., Canales, A., R. (2020).Imputation methods for addressing missing data in short-term monitoring of airpollutants. *Science of Total Environ*ment, 730, https://doi.org/10.1016/j.scitotenv.2020. 139140.

[12]. Mahmoudvand, R., & Rodrigues, P. C. (2016, March). Missing value imputation in time series using Singular Spectrum Analysis. *International Journal of Energy and Statistics*, *04*(01), 1650005. https://doi.org/10.1142/s2335680416500058.

[13]. Bokde, N., Ãlvarez, M., F., Beck, W. M., Kulat, K. (2018).A novel imputation methodology for time series based on pattern sequence forecasting, *Pattern Recognition Letters*, 116, 88-96, https://doi.org/10.1016/j.patrec.2018.09.020.

[14]. Jeong,H.,-Y.,Hong,S.,-H.,Jeon,J.,-S.,Lim,S.,-C.,Kim J.,-C., Park C.,-Y.(2021). A Research for Imputation Method of Photovoltaic Power Missing Data to Apply Time Series Models. *Journal of Korea Multimedia Society*, 24(9), 1251-1260, https://doi.org/10.9717/kmms.2021.24.9. 1251.

[15]. Phan,T.,-T.,-H. (2020). Machine Learning for Univariate Time Series Imputation, *2020 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, 1-6, DOI: 10.1109/MAPR49794.2020.9237768.

[16]. Ran, B., Tan, H., Feng, J., Liu, Y., & Wang, W. (2015). Traffic Speed Data Imputation Method Based on Tensor Completion. *Computational Intelligence and Neuroscience*, *2015*, 1–9, https://doi.org/10.1155/2015/364089.

[17]. Yang, Y.(2012). Modelling Nonlinear Vector Economic Time Series, Ph.D. dissertation, Aarhus University.

[18]. Hossie, T. J., Gobin, J., & Murray, D. L. (2021, August 19). Confronting Missing Ecological Data in the Age of Pandemic Lockdown. *Frontiers in Ecology and Evolution*, *9*. https://doi.org/10.3389/fevo.2021.669477.

[19]. Mohamad, N. B., Lim, B. H., & Lai, A. C. (2021, April 1). Imputation of Missing Values for Solar Irradiance Data under Different Weathers using Univariate Methods. *IOP Conference Series: Earth and Environmental Science*, *721*(1), 012004. https://doi.org/10.1088/1755-1315/721/1/012004

[20]. Stankovic, L., Stankovic, S., & Amin, M. (2014, January). Missing Samples Analysis in Signals for Applications to L-Estimation and Compressive Sensing. *Signal Processing*, *94*(1), 401–408. https://doi.org/10.1016/j.sigpro.2013.07.002

[21]. Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015). Time series analysis: forecasting and control. *Fifth Edition. John Wiley and Sons*.

[22]. Martina, F., Andreas, F., Oliver, F., Olaf, M., Thomas, B.-B., and Klaus, W. (2014, July 16). GECCO Industrial Challenge 2014 Dataset: A water quality dataset for the 'Active protection against pollution of the surface water' competition at the *Genetic and Evolutionary Computation Conference 2015*, *Vancouver*, *Canada*. http://www.spotseven.de/geccochallenge/gecco-challenge-2014.

[23]. Moritz, S., Martina, F., Andreas, F., Christopher, S., and Thomas, B.-B., (2015, May 1). GECCO Industrial Challe nge 2015 Dataset: A heating system dataset for the 'Reco vering missing information in heating system operating d ata' competition at the *Geneticand Evolutionary Computa tion Conference 2015, Madrid, Spain*. http://doi.org/10.52 81/zenodo.3884899.