

RESEARCH

Open Access



Time series data analysis under indeterminacy

Muhammad Aslam^{1*}

*Correspondence:
aslam_ravian@hotmail.com

¹ Department of Statistics,
Faculty of Science, King
Abdulaziz University,
21551 Jeddah, Saudi Arabia

Abstract

The existing semi-average method under classical statistics is applied to measure the trend in the time series data. The existing semi-average method cannot be applied when the time series data is in intervals or imprecise. In this paper, we will introduce a semi-average method under neutrosophic statistics to measure the trend in imprecise or interval data. The application of the proposed semi-average method will be given using the wind speed data. The efficiency of the proposed semi-average method under neutrosophic statistics will be given over the semi-average method under classical statistics in terms of information and adequacy.

Keywords: Semi average method, Trend, Imprecise data, Classical statistics, Forecasting

Introduction

The semi-average method is an important method in time series analysis that is used to analyze the trend in the time series data. This method is very simple and easy to apply in practice. In this method, the time series data in hand is divided into two parts and the corresponding average of each part is calculated. The semi-average method can be used in a variety of fields for estimation and forecasting purposes. The semi-average method is used to set the trend in the time series data and provide the forecasting using the data for future implications. Although this method is simple, easy, objective, and gives identical trend values it is a crude method. The application of the time series method in organizational research can be seen in [1]. The application of time series analysis can be seen in [2, 3]. The application of the method for geographical data can be seen in [4]. Kosiorowski et al. [5] proposed the Wilcoxon statistics for the time-series data.

The estimation and forecasting of wind speed cannot be done with the use of the appropriate statistical techniques. There are many studies on the use of statistical methods in the fields of energy and weather. The applications of the statistical distribution using the wind speed data can be seen in [6–12] also discussed the applications of statistical methods in wind speed forecasting.

It is important to note that classical statistics can be applied for forecasting and estimation purposes when the time series data have precise, certain and indeterminate observations. The use of such statistical techniques in an uncertain environment may

mislead the expert in estimating or forecasting wind speed. Therefore, statistical methods using fuzzy logic are applied to deal with this type of data. The applications of the fuzzy-based statistical methods in estimating or forecasting can be seen in [13–16].

Smarandache [17] discussed the advantages of neutrosophic logic over fuzzy logic. Based on this logic, the idea of neutrosophic statistics was introduced by [18]. More information on neutrosophic logic can be seen in [19, 20]. Smarandache [18, 21, 22] showed that neutrosophic statistics (NS) is quite effective to be applied in an uncertain environment than classical statistics. Smarandache [23] proved the efficiency of NS over interval statistics and classical statistics. The NS can be applied when imprecise; vague and uncertain observations are recorded in the time series data. Guan et al. [24] introduced a new perspective for time series forecasting, considering the quantification of inconsistency as a key characteristic. Abdel-Basset et al. [25] presented a novel neutrosophic forecasting approach using neutrosophic time series (NTS), transforming historical data into NTS data with truth, indeterminacy, and falsity functions. The proposed method includes determining neutrosophic logical relationship groups (NLRGs) and providing a deneutrosophication method for NTS. Singh et al. [26] applied a neutrosophic Set-Based Clustering Algorithm (NEBCA) to fMRI time series datasets, specifically focusing on working memory tasks and resting-state data. Aslam and Albassam [27] introduced the method of least squares under NS. More advantages and applications of NS can be seen in [28, 29].

Aslam and Albassam [27] made a valuable contribution by introducing the application of the least square method under NS for wind speed data forecasting. However, upon the thorough exploration of the literature on NS for time series analysis, we discovered a significant gap. Specifically, no previous research has focused on the utilization of the semi-average method under NS. This gap highlights the need for further investigation and presents an opportunity for us to introduce the semi-average method under NS in this paper. By addressing this gap, we aim to bring a novel approach to time series analysis within the framework of NS. Furthermore, we intend to demonstrate the practicality and effectiveness of our proposed method by applying it to wind speed data and comparing its performance against the semi-average method under classical statistics. Our work will make a significant contribution by introducing the application of the least square method under NS for wind speed data forecasting. Through an extensive review of NS literature, we identified a noteworthy research gap specifically, the absence of previous studies on the utilization of the semi-average method under NS. This comparative analysis will not only showcase the novelty of our approach but also establish its superiority over existing methods.

Preliminaries

Let $Y_{1N} = Y_{1L} + Y_{1U}I_{1N}; I_N \in [I_{1L}, I_{1U}]$, $Y_{2N} = Y_{2L} + Y_{2U}I_{2N}; I_N \in [I_{2L}, I_{2U}]$, ..., $Y_{nN} = Y_{nL} + Y_{nU}I_{nN}; I_N \in [I_{nL}, I_{nU}]$ present the time variables under neutrosophic statistics. Suppose that $n_N \in [n_L, n_U]$ be neutrosophic sample size and $I_N \in [I_{nL}, I_{nU}]$ be the measure of uncertainty/indeterminacy $I_N \in [I_{nL}, I_{nU}]$. Suppose that $Y_{1L}, Y_{2L}, \dots, Y_{nL}$ be the neutrosophic values of the time series. Let $X_{1N} = X_{1L} + X_{1U}I_{1N}; I_N \in [I_{1L}, I_{1U}]$ and $X_{2N} = X_{2L} + X_{2U}I_{2N}; I_N \in [I_{2L}, I_{2U}]$ present the coded values for the first and second

halves of the time data. For more details, see, [18, 21, 22]. These neutrosophic averages can be computed as follows.

Step-1: The neutrosophic average of the determinate part of the 1st half is calculated as

$$\bar{Y}_{1N} = \frac{1}{n_{1N}} \sum_{i=1}^{n_{1N}} Y_i \tag{1}$$

Step-2: The neutrosophic average of the indeterminate part of the 1st half is calculated as

$$\bar{Y}_{2N} = \frac{1}{n_{2N}} \sum_{i=1}^{n_{2N}} Y_i \tag{2}$$

Step-3: The neutrosophic form of both averages is given by.

$$\bar{Y}_{1N} = \bar{Y}_{1L} + \bar{Y}_{1UI}I_N; \bar{Y}_{2N} = \bar{Y}_{2L} + \bar{Y}_{2UI}I_N; I_N \in [I_L, I_U] \tag{3}$$

Semi-average method under indeterminacy

The neutrosophic time series data consists of the neutrosophic numbers having a lower and upper value of the variable of interest. By following the classical semi-average method, the neutrosophic time series data is divided into two halves. The neutrosophic average of the 1st half and the 2nd half are computed and placed at the center of each half. Let $\bar{Y}_{1N} \in [\bar{Y}_{1L}, \bar{Y}_{1U}]$ be the neutrosophic average of the 1st half and $\bar{Y}_{2N} \in [\bar{Y}_{2L}, \bar{Y}_{2U}]$ be the neutrosophic average of the 2nd half. Let X_{1N} and X_{2N} denote the coded values of each half, respectively. The proposed semi-average method under neutrosophic statistics is explained as

$$\begin{aligned} & \left(\hat{Y}_N - (\bar{Y}_{1L} + \bar{Y}_{1UI}I_N) \right) \\ &= \frac{((\bar{Y}_{2L} + \bar{Y}_{2UI}I_N) - (\bar{Y}_{1L} + \bar{Y}_{1UI}I_N))}{(X_{2N} - X_{1N})} \\ & (X_N - X_{1N}); \hat{Y}_N \in [\hat{Y}_L, \hat{Y}_U], I_N \in [I_L, I_U] \end{aligned} \tag{4}$$

The regression line under the proposed method is given by

$$\hat{Y}_N = a_N + b_N X_N; \hat{Y}_N \in [\hat{Y}_L, \hat{Y}_U] \tag{5}$$

where $a_N \in [a_L, a_U]$ is intercept under neutrosophic statistics and computed as

$$a_N = (\bar{Y}_{1L} + \bar{Y}_{1UI}I_N) - b_N X_{1N}; a_N \in [a_L, a_U], b_N \in [b_L, b_U], I_N \in [I_L, I_U] \tag{6}$$

where $b_N \in [b_L, b_U]$ is the slope of the regression line and can be computed as.

$$b_N = \frac{((\bar{Y}_{2L} + \bar{Y}_{2UI}I_N) - (\bar{Y}_{1L} + \bar{Y}_{1UI}I_N))}{(X_{2N} - X_{1N})}; b_N \in [b_L, b_U] \tag{7}$$

The necessary steps to show the derivation of the above equations are shown in the appendix.

Application using wind speed data

As mentioned earlier, the wind speed data is usually recorded as a minimum value and a maximum value. Therefore, we will use the wind speed data (mph) and apply the proposed semi-average method to it. The data is recorded from the Meteorology department in Punjab, Pakistan. The forecasting of the wind speed data recorded in the intervals can be done using the neutrosophic statistics adequately as compared to the semi-average method under classical statistics. The data of some selected months are taken from [30] are shown in Tables 1, 2 and 3.

Table 1 Trended value under neutrosophic statistics for January

January	Wind speed (mph)		Trend values		Neutrosophic form	Indeterminacy
	Y_L	Y_U	\hat{Y}_L	\hat{Y}_U		
1	0	6	0	5.85	$\hat{Y}_N = 0 + 5.85I_N$	$I_N \in [0, 1]$
2	0	7	0	6.28	$\hat{Y}_N = 0 + 6.28I_N$	$I_N \in [0, 1]$
3	0	6	0	6.71	$\hat{Y}_N = 0 + 6.71I_N$	$I_N \in [0, 1]$
4	0	3	0	7.14	$\hat{Y}_N = 0 + 7.14I_N$	$I_N \in [0, 1]$
5	0	7	0	7.57	$\hat{Y}_N = 0 + 7.57I_N$	$I_N \in [0, 1]$
6	0	12	0	8	$\hat{Y}_N = 0 + 8I_N$	$I_N \in [0, 1]$
7	0	7	0	8.43	$\hat{Y}_N = 0 + 8.43I_N$	$I_N \in [0, 1]$
8	0	7	0	8.86	$\hat{Y}_N = 0 + 8.86I_N$	$I_N \in [0, 1]$
9	0	8	0	9.29	$\hat{Y}_N = 0 + 9.29I_N$	$I_N \in [0, 1]$
10	0	12	0	9.72	$\hat{Y}_N = 0 + 9.72I_N$	$I_N \in [0, 1]$
11	0	9	0	10.15	$\hat{Y}_N = 0 + 10.15I_N$	$I_N \in [0, 1]$
12	0	14	0	10.58	$\hat{Y}_N = 0 + 10.58I_N$	$I_N \in [0, 1]$
13	0	21	0	11.01	$\hat{Y}_N = 0 + 11.01I_N$	$I_N \in [0, 1]$
1	0	8	0	11.44	$\hat{Y}_N = 0 + 11.44I_N$	$I_N \in [0, 1]$
1	0	6	0	11.87	$\hat{Y}_N = 0 + 11.87I_N$	$I_N \in [0, 1]$
16	0	9	0	12.3	$\hat{Y}_N = 0 + 12.3I_N$	$I_N \in [0, 1]$
17	0	5	0	12.73	$\hat{Y}_N = 0 + 12.73I_N$	$I_N \in [0, 1]$
18	0	6	0	13.16	$\hat{Y}_N = 0 + 13.16I_N$	$I_N \in [0, 1]$
19	0	10	0	13.59	$\hat{Y}_N = 0 + 13.59I_N$	$I_N \in [0, 1]$
20	0	9	0	14.02	$\hat{Y}_N = 0 + 14.02I_N$	$I_N \in [0, 1]$
21	0	10	0	14.45	$\hat{Y}_N = 0 + 14.45I_N$	$I_N \in [0, 1]$
22	0	100	0	14.88	$\hat{Y}_N = 0 + 14.88I_N$	$I_N \in [0, 1]$
23	0	15	0	15.31	$\hat{Y}_N = 0 + 15.31I_N$	$I_N \in [0, 1]$
24	0	12	0	15.74	$\hat{Y}_N = 0 + 15.74I_N$	$I_N \in [0, 1]$
25	0	9	0	16.17	$\hat{Y}_N = 0 + 16.17I_N$	$I_N \in [0, 1]$
26	0	12	0	16.6	$\hat{Y}_N = 0 + 16.6I_N$	$I_N \in [0, 1]$
27	0	12	0	17.03	$\hat{Y}_N = 0 + 17.03I_N$	$I_N \in [0, 1]$
28	0	9	0	17.46	$\hat{Y}_N = 0 + 17.46I_N$	$I_N \in [0, 1]$
29	0	9	0	17.89	$\hat{Y}_N = 0 + 17.89I_N$	$I_N \in [0, 1]$
30	0	9	0	18.32	$\hat{Y}_N = 0 + 18.32I_N$	$I_N \in [0, 1]$
31	0	8	0	18.75	$\hat{Y}_N = 0 + 18.75I_N$	$I_N \in [0, 1]$

Table 2 Trended value under neutrosophic statistics for February

February	Wind speed (mph)		Trend values		Neutrosophic form	Indeterminacy
	Y_L	Y_U	\hat{Y}_L	\hat{Y}_U		
1	0	5	0.0028	9.24	$\hat{Y}_N = 0.0028 + 9.24I_N$	$I_N \in [0, 1]$
2	0	7	0.0228	9.43	$\hat{Y}_N = 0.0228 + 9.43I_N$	$I_N \in [0, 1]$
3	0	6	0.0428	9.62	$\hat{Y}_N = 0.0428 + 9.62I_N$	$I_N \in [0, 1]$
4	0	14	0.0628	9.81	$\hat{Y}_N = 0.0628 + 9.81I_N$	$I_N \in [0, 0.99]$
5	0	9	0.0828	10	$\hat{Y}_N = 0.0828 + 10I_N$	$I_N \in [0, 0.99]$
6	0	8	0.1028	10.19	$\hat{Y}_N = 0.1028 + 10.19I_N$	$I_N \in [0, 0.99]$
7	0	6	0.1228	10.38	$\hat{Y}_N = 0.1228 + 10.38I_N$	$I_N \in [0, 0.99]$
8	0	12	0.1428	10.57	$\hat{Y}_N = 0.1428 + 10.57I_N$	$I_N \in [0, 0.99]$
9	0	12	0.1628	10.76	$\hat{Y}_N = 0.1628 + 10.76I_N$	$I_N \in [0, 0.98]$
10	0	12	0.1828	10.95	$\hat{Y}_N = 0.1828 + 10.95I_N$	$I_N \in [0, 0.98]$
11	0	7	0.2028	11.14	$\hat{Y}_N = 0.2028 + 11.14I_N$	$I_N \in [0, 0.98]$
12	0	14	0.2228	11.33	$\hat{Y}_N = 0.2228 + 11.33I_N$	$I_N \in [0, 0.98]$
13	0	15	0.2428	11.52	$\hat{Y}_N = 0.2428 + 11.52I_N$	$I_N \in [0, 0.98]$
14	2	21	0.2628	11.71	$\hat{Y}_N = 0.2628 + 11.71I_N$	$I_N \in [0, 0.98]$
15	0	15	0.2828	11.9	$\hat{Y}_N = 0.2828 + 11.9I_N$	$I_N \in [0, 0.98]$
16	0	12	0.3028	12.09	$\hat{Y}_N = 0.3028 + 12.09I_N$	$I_N \in [0, 0.97]$
17	0	12	0.3228	12.28	$\hat{Y}_N = 0.3228 + 12.28I_N$	$I_N \in [0, 0.97]$
18	0	9	0.3428	12.47	$\hat{Y}_N = 0.3428 + 12.47I_N$	$I_N \in [0, 0.97]$
19	3	23	0.3628	12.66	$\hat{Y}_N = 0.3628 + 12.66I_N$	$I_N \in [0, 0.97]$
20	0	21	0.3828	12.85	$\hat{Y}_N = 0.3828 + 12.85I_N$	$I_N \in [0, 0.97]$
21	0	15	0.4028	13.04	$\hat{Y}_N = 0.4028 + 13.04I_N$	$I_N \in [0, 0.97]$
22	0	7	0.4228	13.23	$\hat{Y}_N = 0.4228 + 13.23I_N$	$I_N \in [0, 0.97]$
23	0	7	0.4428	13.42	$\hat{Y}_N = 0.4428 + 13.42I_N$	$I_N \in [0, 0.97]$
24	0	12	0.4628	13.61	$\hat{Y}_N = 0.4628 + 13.61I_N$	$I_N \in [0, 0.97]$
25	0	12	0.4828	13.8	$\hat{Y}_N = 0.4828 + 13.8I_N$	$I_N \in [0, 0.97]$
26	0	9	0.5028	13.99	$\hat{Y}_N = 0.5028 + 13.99I_N$	$I_N \in [0, 0.96]$
27	0	9	0.5228	14.18	$\hat{Y}_N = 0.5228 + 14.18I_N$	$I_N \in [0, 0.96]$
28	0	16	0.5428	14.37	$\hat{Y}_N = 0.5428 + 14.37I_N$	$I_N \in [0, 0.96]$
29	3	21	0.5628	14.56	$\hat{Y}_N = 0.5628 + 14.56I_N$	$I_N \in [0, 0.96]$

Some necessary computations for the implication of the proposed method are shown as follows.

The regression line under neutrosophic statistics for January 2020 is expressed as

$$\hat{Y}_N = [0, 8.86] + [0, 0.43]X_N$$

The regression line under neutrosophic statistics for February 2020 is expressed as

$$\hat{Y}_N = [0.1428, 10.57] + [0.02, 0.19]X_N$$

The regression line under neutrosophic statistics for March 2020 is expressed as

Table 3 Trended value under neutrosophic statistics for March

March	Wind speed (mph)		Trend values		Neutrosophic form	Indeterminacy
	Y_L	Y_U	\hat{Y}_L	\hat{Y}_U		
1	0	12	0	17.36	$\hat{Y}_N = 0 + 17.36I_N$	$I_N \in [0, 1]$
2	0	6	0	17.07	$\hat{Y}_N = 0 + 17.07I_N$	$I_N \in [0, 1]$
3	0	6	0	16.78	$\hat{Y}_N = 0 + 16.78I_N$	$I_N \in [0, 1]$
4	0	12	0	16.49	$\hat{Y}_N = 0 + 16.49I_N$	$I_N \in [0, 1]$
5	3	25	0	16.2	$\hat{Y}_N = 0 + 16.2I_N$	$I_N \in [0, 1]$
6	2	25	0	15.91	$\hat{Y}_N = 0 + 15.91I_N$	$I_N \in [0, 1]$
7	0	13	0.53	15.62	$\hat{Y}_N = 0.53 + 15.62I_N$	$I_N \in [0, 0.97]$
8	0	10	1.4	15.33	$\hat{Y}_N = 1.4 + 15.33I_N$	$I_N \in [0, 0.91]$
9	5	15	2.27	15.04	$\hat{Y}_N = 2.27 + 15.04I_N$	$I_N \in [0, 0.85]$
10	0	12	3.14	14.75	$\hat{Y}_N = 3.14 + 14.75I_N$	$I_N \in [0, 0.79]$
11	0	13	4.01	14.46	$\hat{Y}_N = 4.01 + 14.46I_N$	$I_N \in [0, 0.72]$
12	3	37	4.88	14.17	$\hat{Y}_N = 4.88 + 14.17I_N$	$I_N \in [0, 0.66]$
13	5	23	5.75	13.88	$\hat{Y}_N = 5.75 + 13.88I_N$	$I_N \in [0, 0.59]$
14	3	14	6.62	13.59	$\hat{Y}_N = 6.62 + 13.59I_N$	$I_N \in [0, 0.51]$
15	0	7	7.49	13.3	$\hat{Y}_N = 7.49 + 13.3I_N$	$I_N \in [0, 0.44]$
16	0	12	8.36	13.01	$\hat{Y}_N = 8.36 + 13.01I_N$	$I_N \in [0, 0.36]$
17	0	7	9.23	12.72	$\hat{Y}_N = 9.23 + 12.72I_N$	$I_N \in [0, 0.27]$
18	0	9	10.1	12.43	$\hat{Y}_N = 10.1 + 12.43I_N$	$I_N \in [0, 0.19]$
19	0	8	10.97	12.14	$\hat{Y}_N = 10.97 + 12.14I_N$	$I_N \in [0, 0.10]$
20	0	7	11.84	11.85	$\hat{Y}_N = 11.84 + 11.85I_N$	$I_N \in [0, 0.001]$
21	0	16	12.71	11.56	$\hat{Y}_N = 12.71 - 11.56I_N$	$I_N \in [0, 0.09]$
22	0	8	13.58	11.27	$\hat{Y}_N = 13.58 - 11.27I_N$	$I_N \in [0, 0.17]$
23	0	6	14.45	10.98	$\hat{Y}_N = 14.45 - 10.98I_N$	$I_N \in [0, 0.24]$
24	0	7	15.32	10.69	$\hat{Y}_N = 15.32 - 10.69I_N$	$I_N \in [0, 0.30]$
25	0	21	16.19	10.4	$\hat{Y}_N = 16.19 - 10.4I_N$	$I_N \in [0, 0.36]$
26	0	12	17.06	10.11	$\hat{Y}_N = 17.06 - 10.11I_N$	$I_N \in [0, 0.41]$
27	0	17	17.93	9.82	$\hat{Y}_N = 17.93 - 9.82I_N$	$I_N \in [0, 0.45]$
28	3	10	18.8	9.53	$\hat{Y}_N = 18.8 - 9.53I_N$	$I_N \in [0, 0.49]$
29	3	7	19.67	9.24	$\hat{Y}_N = 19.67 - 9.24I_N$	$I_N \in [0, 0.53]$
30	0	12	20.54	8.95	$\hat{Y}_N = 20.54 - 8.95I_N$	$I_N \in [0, 0.56]$
31	0	14	21.41	8.66	$\hat{Y}_N = 21.41 - 8.66I_N$	$I_N \in [0, 0.60]$

$$\hat{Y}_N = [1.4, 15.33] + [0.87, -0.29]X_N$$

The regression line for January 2020 shows that the intercept of the regression line will be between 0 and 8.86. The slope of the regression for January 2020 is also in an indeterminate interval from 0 to 0.43. The regression line for February 2020 shows that the intercept of the regression line will be between 0.1428 and 10.57. The slope of the regression for February 2020 is also in the indeterminate interval from 0.02 to 0.19. The regression line for March 2020 shows that the intercept of the regression line will be between 1.4 and 15.33. The slope of the regression for March 2020 is also in indeterminate

interval from 0.87 to -0.29 . Using these regression lines, the trended values are calculated and placed in Tables 1, 2 and 3. From Tables 1, 2 and 3, it can be noted that the lower (minimum) value of wind speed is mostly zero. Therefore, a larger indeterminacy can be expected in forecasting using this indeterminate data. From Table 3, the fitted regression line for the 4th March 2020 is $\hat{Y}_N = 0.0628 + 9.81I_N; I_N \in [0, 0.99]$. From the trended regression equation, it can be forecast that the wind speed (mph) will be 0.0628 to 9.81 with the measure of indeterminacy 0.99. The fitted trend lines and the real wind speed (mph) for 3 months are shown in Fig. 1. From Fig. 1, it can be seen that for the month of January, the values of the first 11 days are close to the fitted regression lines while the values of the remaining days are away from the upper value of the trended line. In actual values of the months of February and March are away from the upper values of the trended lines. On the other hand, the lower values of actual data are very close to the lower values of trended lines for 3 months. Figure 1 clearly shows that indeterminate wind speed (mph) should be forecasted using the semi-average method under neutrosophic statistics. For this type of time series data, the use of the existing semi-average method under classical statistics may mislead decision-makers. Based on the wind speed data, it can be concluded that the proposed semi-average method under neutrosophic is adequate and suitable to use for forecasting purposes.

Competitive study based on wind speed data

The proposed semi-average method under neutrosophic is a generalization of the semi-average method under classical statistics. The equations of the proposed semi-average method reduce to equations under classical statistics when $I_L = 0$. The efficiency of the proposed semi-average method under neutrosophic statistics will be compared with the existing semi-average method under classical statistics in terms of information and adequacy. For the comparison purpose, we presented the trended values, their

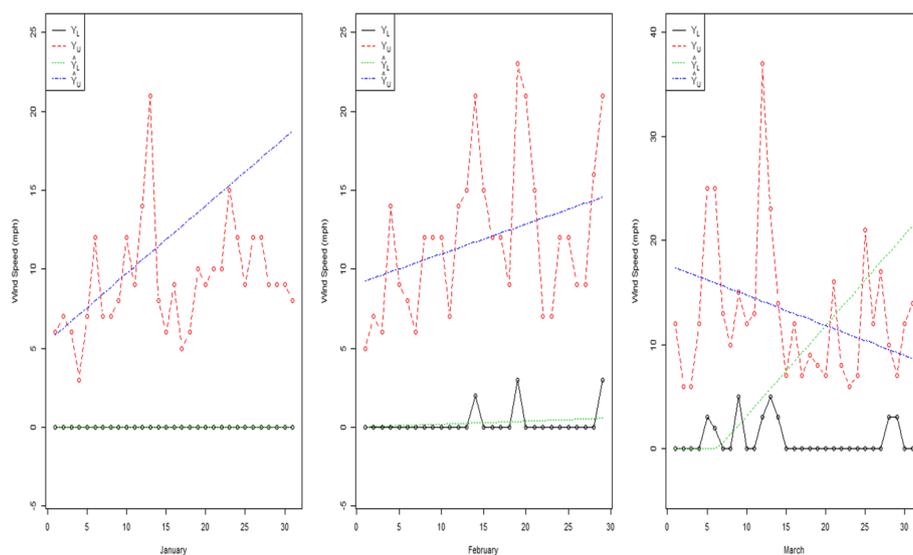


Fig. 1 The trend lines and actual observations for January, February and March 2020. Figure 1 is generated using R version 3.2.1 (<https://www.r-project.org/>)

neutrosophic forms, and the measure of indeterminacy in Tables 1, 2 and 3. Note here that in each neutrosophic form in Tables 1, 2 and 3, the first value presents the trend value using classical statistics. The second value of each neutrosophic form presents the indeterminate part. Each neutrosophic form reduces to the determinate part when $I_N = 0$. For example, the fitted regression line for 5th February 2020 is $\hat{Y}_N = 0.0828 + 10I_N$. In this neutrosophic form, the first value 0.0828 presents the forecasted value under classical statistics. The second value $10I_N$ denotes the indeterminate part and the measure of indeterminacy is $I_N \in [0, 0.99]$. From this analysis, it can be seen that this is the beauty of the proposed model in that it provides the forecasting results in indeterminate intervals rather the exact result in the presence of uncertainty. In addition, the proposed method gives additional information about the measure of indeterminacy that cannot be obtained from the analysis under classical statistics. Note here that there are higher measures of indeterminacy in the months of January and February. In March, after the 6th day, smaller values of measure of indeterminacy in trend values are found. From the analysis, it is clear that the existing semi-average method under neutrosophic statistics is more flexible than the classical method. In addition, in the presence of a high measure of indeterminacy, the forecasting may be affected and mislead the decision-makers. We also note that for interval data, it would be better to use neutrosophic statistics to forecast wind speed.

Conclusions

The paper introduced a modified version of the semi-average method within the context of neutrosophic statistics. This modified semi-average method serves as a generalization of the existing approach. Based on the methodology and application, it can be concluded that the proposed semi-average method is well-suited for situations where data is recorded in intervals. In contrast, utilizing the existing semi-average method for wind speed prediction with interval data may lead decision-makers astray. It is important to note that the proposed method has its limitations and can only be applied when dealing with imprecise observations within time series data. However, the neutrosophic semi-average method utilizes less information compared to other time series methods. The applications of the proposed method extend to various fields such as meteorology for weather forecasting, business, and medical science. As for further research, it is worth considering other time series methods within the framework of neutrosophic statistics. This would open up possibilities for additional advancements in the field. Other time series methods under neutrosophic statistics can be extended as future research.

Appendix

Let $Y_1 = a_1 + b_1I_{NY}$ and $Y_2 = a_2 + b_2I_{NY}$ for $I_{NY} \in [I_{LY}, I_{UY}]$ be two neutrosophic regression numbers obtained from regression analysis. The basic operation of these neutrosophic regression numbers can be performed as follows.

$$Y_1 + Y_2 = (a_1 + a_2) + (b_1 + b_2)I_{NY} = [a_1 + a_2 + b_1I_{LY} + b_2I_{LY}, a_1 + a_2 + b_1I_{UY} + b_2I_{UY}]$$

Or

$$Y_1 + Y_2 = [(a_1 + a_2) + (b_1 + b_2)I_{LY}, (a_1 + a_2) + (b_1 + b_2)I_{UY}]$$

Based on the above-mentioned information, let $Y_i = a_i + b_i I_{NY} (i = 1, 2, 3, \dots, n)$ be a group of neutrosophic numbers obtained from regression analysis having the lower and upper values for $I_{NY} \in [I_{LY}, I_{UY}]$, the neutrosophic average of these regression numbers can be calculated as follows

$$\bar{Y} = \bar{a} + \bar{b}I_{NY}, I_{NY} \in [I_{LY}, I_{UY}]$$

where $\bar{a} = \frac{1}{n} \sum_{i=1}^n a_i$ and $\bar{b} = \frac{1}{n} \sum_{i=1}^n b_i$

The development of Eq. (4), Eq. (5), Eq. (6) and Eq. (7) under neutrosophic statistics are given as follows

Suppose that $\bar{Y}_{1N} \in [\bar{Y}_{1L}, \bar{Y}_{1U}]$ and $\bar{Y}_{2N} \in [\bar{Y}_{2L}, \bar{Y}_{2U}]$ be neutrosophic average for the 1st half and the 2nd half, respectively. Let X_{1N} and X_{2N} be coded values for the 1st and 2nd half, respectively. Let (X_{1N}, \bar{Y}_{1N}) and (X_{2N}, \bar{Y}_{2N}) be two points from which the estimated line $\hat{Y}_N = a_N + b_N X_N$; $\hat{Y}_N \in [\hat{Y}_L, \hat{Y}_U]$ passes. Two constant a_N and b_N can be determined from the following expression

$$(\hat{Y}_N - \bar{Y}_{1N}) = \frac{(\bar{Y}_{2N} - \bar{Y}_{1N})}{(X_{2N} - X_{1N})} (X_N - X_{1N}); \hat{Y}_N \in [\hat{Y}_L, \hat{Y}_U], I_N \in [I_L, I_U]$$

By following the neutrosophic theory, the neutrosophic form of $\bar{Y}_{1N} \in [\bar{Y}_{1L}, \bar{Y}_{1U}]$ is expressed by

$$\bar{Y}_{1N} = \bar{Y}_{1L} + \bar{Y}_{1U}I_N; I_N \in [I_L, I_U]$$

The neutrosophic form of $\bar{Y}_{2N} \in [\bar{Y}_{2L}, \bar{Y}_{2U}]$ is expressed by.

$$\bar{Y}_{2N} = \bar{Y}_{2L} + \bar{Y}_{2U}I_N; I_N \in [I_L, I_U]$$

The above expression can be written as follows

$$(\hat{Y}_N - (\bar{Y}_{1L} + \bar{Y}_{1U}I_N)) = \frac{((\bar{Y}_{2L} + \bar{Y}_{2U}I_N) - (\bar{Y}_{1L} + \bar{Y}_{1U}I_N))}{(X_{2N} - X_{1N})} (X_N - X_{1N}); \hat{Y}_N \in [\hat{Y}_L, \hat{Y}_U], I_N \in [I_L, I_U]$$

$$(\hat{Y}_N - (\bar{Y}_{1L} + \bar{Y}_{1U}I_N)) = b_N(X_N - X_{1N})$$

$$\Rightarrow \hat{Y}_N = a_N + b_N X_N; \hat{Y}_N \in [\hat{Y}_L, \hat{Y}_U]$$

where

$$a_N = (\bar{Y}_{1L} + \bar{Y}_{1U}I_N) - b_N X_{1N}; a_N \in [a_L, a_U], b_N \in [b_L, b_U], I_N \in [I_L, I_U]$$

and

$$b_N = \frac{((\bar{Y}_{2L} + \bar{Y}_{2U}I_N) - (\bar{Y}_{1L} + \bar{Y}_{1U}I_N))}{(X_{2N} - X_{1N})}; b_N \in [b_L, b_U]$$

Acknowledgements

The author is deeply thankful to the editor and reviewers for their valuable suggestions to improve the quality and presentation of the paper.

Author contributions

MA wrote the paper.

Funding

None.

Availability of data and materials

The data is given in the paper.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

No conflict of interest regarding the paper.

Received: 2 March 2023 Accepted: 27 July 2023

Published online: 05 August 2023

References

- Jebb AT, Tay L. Introduction to time series analysis for organizational research: methods for longitudinal analyses. *Organ Res Methods*. 2017;20(1):61–94.
- Chatfield C, Xing H. *The analysis of time series: an introduction with R*. Boca Raton: CRC Press; 2019.
- McDowall D, McCleary R, Bartos BJ. *Interrupted time series analysis*. Oxford: Oxford University Press; 2019.
- Feyrer J. Trade and income—exploiting time series in geography. *Am Econ J Appl Econ*. 2019;11(4):1–35.
- Kosiorowski D, Rydlewski JP, Snarska M. Detecting a structural change in functional time series using local Wilcoxon statistic. *Stat Pap*. 2019;60(5):1677–98.
- Bidaoui H, et al. Wind speed data analysis using weibull and rayleigh distribution functions, case study: five cities Northern Morocco. *Procedia Manuf*. 2019;32:786–93.
- Alrashidi M, Rahman S, Pipattanasomporn M. Metaheuristic optimization algorithms to estimate statistical distribution parameters for characterizing wind speeds. *Renew Energy*. 2020;149:664–81.
- ul Haq MA, et al. Marshall-Olkin Power Lomax distribution for modeling of wind speed data. *Energy Rep*. 2020;6:1118–23.
- Mahmood FH, Resen AK, Khamees AB. Wind characteristic analysis based on Weibull distribution of Al-Salman site. *Iraq Energy Rep*. 2019;6:79–87.
- Akgül FG, Şenoğlu B. Comparison of wind speed distributions: a case study for Aegean coast of Turkey. *Energy Sources Part A Recovery Util Environ Eff*. 2019. <https://doi.org/10.1080/15567036.2019.1663309>.
- Zaman B, Lee MH, Riaz M. An improved process monitoring by mixed multivariate memory control charts: An application in wind turbine field. *Comput Ind Eng*. 2020;142: 106343.
- Yan J. A comparison wind power forecasting with different methods. *Integr Ferroelectr*. 2022;227(1):191–201.
- Grzegorzewski P. K-sample median test for vague data. *Int J Intell Syst*. 2009;24(5):529–39.
- Laib M, et al. Multifractal analysis of the time series of daily means of wind speed in complex regions. *Chaos Solitons Fractals*. 2018;109:118–27.
- Grzegorzewski P, Spiewak M. The sign test and the signed-rank test for interval-valued data. *Int J Intell Syst*. 2019;34(9):2122–50.
- Sezer OB, Gudelek MU, Ozbayoglu AM. Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. *Appl Soft Comput*. 2020;90: 106181.
- Smarandache F. Neutrosophy. Neutrosophic probability, set, and logic, proQuest information learning. *Ann Arbor Mich USA*. 1998;105:118–23.
- Smarandache F. Introduction to neutrosophic statistics. *Infinite Study*. 2014. <https://doi.org/10.13140/2.1.2780.1289>.
- Broumi S, et al. Interval-valued fermatean neutrosophic graphs. *Collect Papers Vol XIII Various Sci Topics*. 2022. <https://doi.org/10.31181/dmame0311072022b>.
- Broumi S, et al. Complex fermatean neutrosophic graph and application to decision making. *Decis Mak App Manag Eng*. 2023;6(1):474–501.
- Chen J, Ye J, Du S. Scale effect and anisotropy analyzed for neutrosophic numbers of rock joint roughness coefficient based on neutrosophic statistics. *Symmetry*. 2017;9(10):208.

22. Chen J, et al. Expressions of rock joint roughness coefficient using neutrosophic interval statistical numbers. *Symmetry*. 2017;9(7):123.
23. Smarandache F. Neutrosophic statistics is an extension of Interval Statistics, while plithogenic statistics is the most general form of statistics (second version). *Infinite Study*. 2022. <https://doi.org/10.54216/IJNS.190111>.
24. Guan H, et al. A neutrosophic forecasting model for time series based on first-order state and information entropy of high-order fluctuation. *Entropy*. 2019;21(5):455.
25. Abdel-Basset M, et al. A refined approach for forecasting based on neutrosophic time series. *Symmetry*. 2019;11(4):457.
26. Singh P, et al. Analysis of fMRI time series: neutrosophic-entropy based clustering algorithm. *J Adv Inf Technol*. 2022;13(3):224–9.
27. Aslam M, Albassam M. Forecasting of wind speed using interval-based least square method. *Front Energy Res*. 2022. <https://doi.org/10.3389/fenrg.2022.896217>.
28. Aslam M. Design of the Bartlett and Hartley tests for homogeneity of variances under indeterminacy environment. *J Taibah Univ Sci*. 2020;14(1):6–10.
29. Aslam M. On detecting outliers in complex data using Dixon's test under neutrosophic statistics. *J King Saud Univ Sci*. 2020;32(3):2005–8.
30. Aslam M. Retracted article: Forecasting of the wind speed under uncertainty. *Sci Rep*. 2022;10:20300.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
