Statistics and Its Interface Volume 16 (2023) 279–286

Modeling water table depth using singular spectrum analysis

Rahim Mahmoudvand*, Mehrdad Barati, Asghar Seif, Sahar Ranjbaran, and Paulo Canas Rodrigues

The majority of countries are facing or will face a serious water crisis. As a consequence, we observe a deterioration in the water quality such as the drop in the water table and a salinity increase. Therefore, it is highly recommended to conduct a regular monitoring program on groundwater levels in order to sustain this source. Water table depth (WTD) is an index of water availability that influences many soil characteristics. Consequently, there are concerns with WTD in both time and space. This paper shows how to build a model for water table depth using Singular Spectrum Analysis (SSA). The study area is located in the Ghahavand plain in the Hamedan province, western Iran. We used water table depth records that were collected by Hamedan regional water authority as part of their monitoring system program. The data were obtained monthly by measuring WTD of about 200 wells within the study area in the period between 1988 and 2016. There were many errors, inconsistencies and missing cells in the data file. So, we started with improving data quality and filling the missing cells. The other problem with the data was related to the well samples that have changed during the study horizon. Classically, we take a simple average on the observations at each time point to build a univariate time series. However, a descriptive analysis revealed that the heterogeneity in the value of the WTD in the study area has increased over time. So, we used box plot components to build model for WTD. We used both univariate SSA and multivariate SSA to capture the information within the box plot components. The performance of the proposal was accessed by using both in sample fitting errors and out of sample forecasting errors. The results suggest that the new approach provides an attractive alternative to the classical approach.

AMS 2000 subject classifications: Primary 62xx, 62M10; secondary 86xx, 86A05.
Keywords and phrases: Hamedan, Water table depth, Singular spectrum analysis.

1. INTRODUCTION

The water resource management, in every country, depends on quantitative understanding of hydrology and hydrogeology [28]. In spite of increasing awareness about the water resource problems in recent years, improvement in the area of decision-making is still limited [19, 20]. New approaches in decision-making help active stakeholders to consider environmental problems using computer modelling methods ([29]). Two important tools that can facilitate strategic decision making in complex environmental/natural resource management systems are: (i) participatory modelling (PM); and (ii) stakeholder engagement [30, 27, 13, 29, 31]. Water table depth (WTD) is simply the depth of phreatic water below the surface [12]. WTD affects the water availability and many other soil characteristics, see e.g., [18]. According to Iran's Tasnim news agency there are 90 sinkholes in the whole country, including 25 in the Hamadan province. A giant sinkhole about 60 metres deep has made a surprise appearance in western Iran (see Figure 1).

The sinkhole opened up on August 19, 2018 near the small village of Kerdabad, in the Kabudrahang county of the western province of Hamadan. This is not the first time this phenomenon has occurred in the region; it is actually the 13th giant sinkhole that appeared in Kabudrahang county in two decades. So far, none of them have opened up inside villages or cities, but some are nearby. It has terrified both locals and experts, who say this phenomenon shows that groundwater tables have been badly damaged.

Consequently, there are concerns with WTD in both time and space. Time series associated with climate inputs, such as rainfall, are usually nonlinear. Thus, using simple time series techniques for analysing climate inputs may be misleading in selecting the appropriate models. On the other hand, the majority of techniques are essentially confirmatory procedures, which is unrealistic in practice because the true/real model is unknown.

In this paper, we use a model free time series analysis technique so called Singular Spectrum Analysis (SSA). This technique extracts and builds the model in the process of time series analysis. SSA is a very flexible technique which does not depends on restrictive assumptions such as linearity, normality and stationary. It is based on the idea of classical time series analysis and mainly uses matrix algebra ([6]). SSA were successfully employed for analysing time series in various domains such as meteorological ([2, 7, 9]),

*Corresponding author.

Figure 1. A giant sinkhole about 60 metres deep that opened up on August 19, 2018 near the Kabudrahang county of the western province of Hamadan.

bio-medical ([26]), hydrological ([1, 11, 32]), physical sciences ([3]), economics and finance ([10, 24, 25]), engineering ([5]), among others.

Besides of the ability of the technique, it is important to adapt its use to the problem in hand. The data of this study is not a simple univariate time series, it includes multiple observations at each time step (measurements of the same phenomenon, but from different locations). Specifying an appropriate proxy for the value of time series at each time step is a problem that we address in this study. In particular, we consider the multivariate SSA applied to the monthly time series of the five box-plot components: lower bound, first quartile, median, third quartile, and upper bound.

This paper is structured as follows. A brief description of SSA is presented in Section 2. The detailed modelling framework of the proposed approach is provided in Section 3. Finally, we derive some concluding remarks in Section 4.

2. MATERIALS AND METHODS

2.1 A brief introduction to SSA

Here we explain briefly how to apply the SSA method; for more details see for example [6]. Consider a real-valued vector/time series $Y_T = (y_1, \ldots, y_T)'$ of length T. Let L $(2 \leq L \leq T/2)$ be an integer called the window length and set $K = T - L + 1$. In order to denoise (smooth) the time series Y_T, the following four steps must be performed.

Step 1: Embedding. Define the trajectory matrix $\mathbf{X}$ as:

$$\mathbf{X} = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ y_3 & y_4 & y_5 & \cdots & y_{K+2} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \cdots & y_T \end{bmatrix} \quad (1)$$

In the trajectory matrix, all the elements along the off diagonal are equal. Such matrix is called a Hankel matrix. This property shows that there is a one-to-one relationship between original time series and trajectory matrix.

Step 2: Singular Value Decomposition. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L \geq 0$ denote eigenvalues of $\mathbf{XX}'$ and U_i the normalized eigenvector corresponding to the eigenvalue λ_i $(i = 1, \ldots, L)$. Then, the singular value decomposition (SVD) of the trajectory matrix $\mathbf{X}$ can be written as:

$$\mathbf{X} = \mathbf{E}_1 + \ldots + \mathbf{E}_L; \quad (2)$$

where, $\mathbf{E}_i = \sqrt{\lambda_i}\, U_i V_i'$ with $V_i = \mathbf{X}' U_i / \sqrt{\lambda_i}$ $(i = 1, \ldots, L)$.

Step 3: Grouping. Considering the partition $\{1, \ldots, L\} = \bigcup_{j=1}^{m} \{I_j\}$ where $I_p \cap I_q = \varnothing, p \neq q$, we can find the following decomposition:

$$\mathbf{X} = \mathbf{X}_{I_1} + \ldots + \mathbf{X}_{I_m}; \quad (3)$$

where, $\mathbf{X}_{I_j} = \sum_{\ell \in I_j} \mathbf{E}_\ell$, $j = 1, ..., m$. In this paper, we consider two partitions, one associated to the signal and another to the noise. This means that there is an integer r between 1 and L such that:

$$\begin{aligned} \mathbf{X} &= \sum_{i=1}^{r} \mathbf{E}_i + \sum_{i=r+1}^{L} \mathbf{E}_i \\ &= \textbf{signal} + \textbf{noise} \end{aligned}$$

Step 4: Reconstruction. The last step in the basic SSA, transforms each matrix of the grouped decomposition (Step 3) into a new time series of length T. Indeed, we defined a Hankel operator in the first step which transforms a vector of dimension $T \times 1$ into a Hankel matrix of dimension $L \times K$. Here, we apply the inverse of that operator that can be performed with diagonal averaging. Using this operator, we obtain reconstructed series $\widetilde{Y}_T = (\widetilde{y}_1, \ldots, \widetilde{y}_T)'$ that is associated **signal** matrix.

It must be noted that the general purpose of the SSA is the decomposition of the original series into the sum of interpretable components such as trend and periodic components. Sometimes, however, it can also be used in specific tasks, such as: (i) extraction of signal from noise; (ii) extraction of oscillatory components; (iii) smoothing; and (iv) forecasting. To implement model fit (e.g. the first three tasks (i)-(iii) mentioned above), the above SSA algorithm can be used. However, in order to produce forecast by SSA, we require further computations, as described in the next subsection.

2.1.1 Forecasting by SSA

The basic SSA recurrent forecasting algorithm discussed in [6] should be regarded as the main forecasting algorithm. Although, there have been proposed several natural modifications to this algorithm which can provide better forecasts in specific situations (e.g [6, 14]), here we consider standard recurrent forecasting algorithm and denote it by SSA-R. Let $\widehat{y}_{h|T}$ denote the SSA-R forecast at the time point T, for the lead time h (or for h steps ahead). According to the SSA-R, the following recursive formula can be used to obtain forecasts:

$$\widehat{y}_{h|T} = \begin{cases} \widetilde{y}_h & h = 1, \dots, T \\ \sum_{t=1}^{L-1} a_t \widehat{y}_{(h-t)|T} & h = T+1, \dots \end{cases} \tag{4}$$

where, $(\widetilde{y}_1, \dots, \widetilde{y}_T)'$ denote the reconstructed time series,

$$R = (a_{L-1}, \dots, a_1)' = \frac{1}{1-\nu^2} \sum_{i=1}^{r} \pi_i U_i^{\nabla}$$

such that $\nu^2 = \sum_{i=1}^{r} \pi_i^2$, and π_i is the last component of the vector U_i, and $U_i^{\nabla}$is the vector containing the first $L-1$ components of U_i, for $i = 1, \dots, L$.

2.1.2 SSA choices

There are two basic, but very important, parameters that must be selected when performing SSA for forecasting. The first parameter is window length, L, and the second is the number of eigentriples/components to conduct the reconstruction, r. The window length is applied for the embedding and the other parameter is used for grouping and constructing the forecast engine. To select the window length, a value between 2 and $N/2$ should be considered, being proportional to the number of observations per period (e.g. proportional to 12 for monthly time series). There are several methods for choosing the number of components used for reconstruction, being two of the most popular methods based on the scree plot of the singular values and on the weighted correlation between components. We used the first method to determine an appropriate values for pair (L, r). For more details on the choice of these two parameters can be found, for example in [6, 15, 16, 5, 14, 22].

2.2 Multivariate SSA

Multivariate SSA, or MSSA, is a natural extension of SSA for analysing multivariate time series. Let $Y_t = \left[y_t^{(1)}, \dots, y_t^{(M)}\right]$, $t = 1, \dots, T$, denote a sample of a M-variate time series with length T. Note that it is possible to consider different number of observations for the individual time series in the multivariate framework ([8, 23]), but here we assume equal number of observations. Let us assume that Y_t can be written in terms of a signal plus noise model as:

$$\begin{aligned} \mathbf{Y}_T &= \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{bmatrix} = \begin{bmatrix} y_1^{(1)} & \dots & y_1^{(M)} \\ y_2^{(1)} & \dots & y_2^{(M)} \\ \vdots & \dots & \vdots \\ y_T^{(1)} & \dots & y_T^{(M)} \end{bmatrix} \\ &= \begin{bmatrix} s_1^{(1)} & \dots & s_1^{(M)} \\ s_2^{(1)} & \dots & s_2^{(M)} \\ \vdots & \dots & \vdots \\ s_T^{(1)} & \dots & s_T^{(M)} \end{bmatrix} + \begin{bmatrix} n_1^{(1)} & \dots & n_1^{(M)} \\ n_2^{(1)} & \dots & n_2^{(M)} \\ \vdots & \dots & \vdots \\ n_T^{(1)} & \dots & n_T^{(M)} \end{bmatrix}. \end{aligned}$$

The MSSA algorithm denoises (smooths) the multivariate time series $\mathbf{Y}_T = [Y_1, \dots, Y_T]'$ using the same steps as the univariate SSA, i.e. embedding, SVD, grouping and reconstruction. The only difference is related to the definition of trajectory matrix. Although there are several forms to define the trajectory matrix in MSSA ([8]), here we use the horizontal form defined as:

$$\mathbf{X} = \left[\mathbf{X}^{(1)} \dots \mathbf{X}^{(M)}\right] =$$

$$\begin{bmatrix} y_1^{(1)} & y_2^{(1)} & \dots & y_K^{(1)} & \dots & y_1^{(M)} & y_2^{(M)} & \dots & y_K^{(M)} \\ y_2^{(1)} & y_3^{(1)} & \dots & y_{K+1}^{(1)} & \dots & y_2^{(M)} & y_3^{(M)} & \dots & y_{K+1}^{(M)} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ y_L^{(1)} & y_{L+1}^{(1)} & \dots & y_T^{(1)} & \dots & y_L^{(M)} & y_{L+1}^{(M)} & \dots & y_T^{(M)} \end{bmatrix}.$$

where L and K are chosen similarly as before and $\mathbf{X}^{(j)}$ is an Hankel matrix for the column j of the $\mathbf{Y}_T$. This means that the trajectory matrix for the MSSA algorithm is a block Hankel matrix and this property is considered for the reconstruction step.

Denote by $\widetilde{Y}_T^{(j)} = (\widetilde{y}_1^{(j)}, \dots, \widetilde{y}_T^{(j)})'$ the reconstructed values of the time series j. The h-steps ahead forecasts obtained by the MSSA algorithm can be obtained with the following recursive formula:

$$\widehat{y}_{h|T}^{(j)} = \begin{cases} \widetilde{y}_h^{(j)} & h = 1, \dots, T \\ \sum_{t=1}^{L-1} a_t \widehat{y}_{(h-t)|T}^{(j)} & h = T+1, \dots \end{cases} \tag{5}$$

where $R = (a_{L-1}, \dots, a_1)'$ will be obtained from trajectory matrix $\mathbf{X}$ similarly as in the univariate SSA-R. The methodology of the multivariate SSA shows that it also needs two choices for its application in practice: the window length L, and the cutting point r. We can determine these values using the same approaches mentioned for the univariate SSA. It is worth mentioning that the complexity of the multivariate SSA model is smaller than the univariate SSA, when we apply both models for analysing multivariate time series. This happens because multivariate SSA needs two choices (L, r) whereas the univariate SSA needs $2M$ choices $\{(L_1, r_1), \dots, (L_M, r_M)\}$.

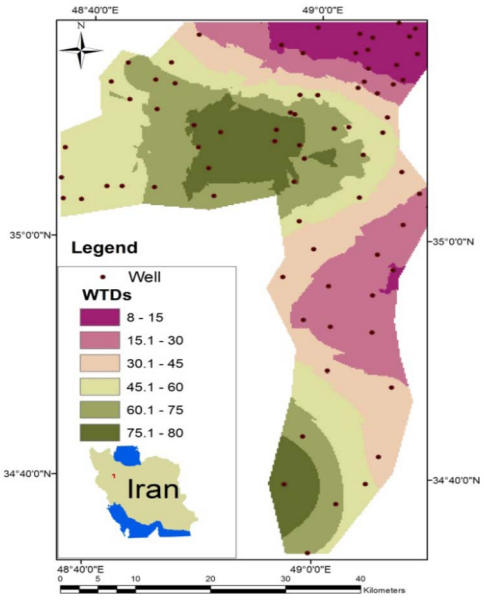

Figure 2. Location of wells in the study area in terms of WTD.

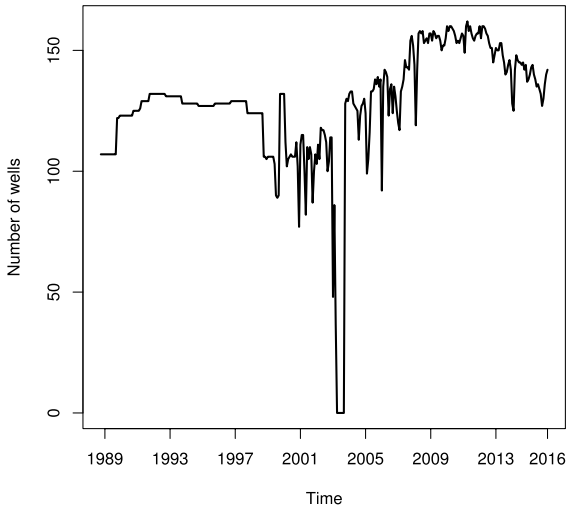

Figure 3. Number of sampled wells in each month between October 1988 and March 2016.

3. EMPIRICAL RESULTS

The study area is located in the Ghahavend plain, which is located in the East Hamedan province, Iran (34°24′19.87″ – 35°17′27.39″N, 48°33′46.79″ – 49°28′31.67″E). Its extension is about 50 km to east-west direction and about 100 km in the north-south direction, with represents an area of approximately 2500 km^2. The regional structure is of a broad asymmetrical syncline known as the "Ghahavand-Kabudrahang" basin. The regional structure is well represented by the Mesozoic to Cenozoic Groups. Static groundwater levels were measured at about 200 hand-dug and drilled wells within the study area. Figure 2 shows the digital elevation model and locations of the wells which were collected from Hamedan regional water authority, as part of their monitoring system program in the area, in March 2016. WTDs in the area increase from 8 to 80 meters from the north to the centre of the study area.

3.1 Description of data

The data for this study consists on monthly WTD records obtained from about 200 wells located in the study area. The data covers the period between October 1988 and March 2016, with a total number of 49232 records. The number of wells at each month (sample sizes) is shown in Figure 3. This figure shows that the majority of sample sizes range from 100 to 150 wells. These sample sizes are quite enough for our study, as guidelines suggest that a sample between 20 and 30 wells are enough to provide statistical confidence ([4]).

Suppose that a sample of size w_t wells is chosen at the time point t. Denote by $y_i(t)$ the WTD for well i at time t. Then, we calculated monthly time series by using the following formula:

$$y_t = \frac{1}{w_t} \sum_{i=1}^{w_t} y_i(t). \tag{6}$$

Using this formula, we computed observations $(y_1, \ldots, y_{330})$ corresponds to period between October 1988 and March 2016. The data between April 2003 and September 2003 (observations y_{175} to y_{180}) were missing. In order to impute these missing observations, the algorithm proposed by [21] and [17] was used. According to their approach, the time series data before and after the missing block of observations is used to produce two estimates for the missing observations, which are then combined, considering a given weighting scheme, into a single estimate for a greater precision. Here, we used $y_1, \ldots, y_{174}$ to obtain forecasts for observations y_{175} to y_{180}. In addition, we find another estimates for y_{175} to y_{180} by back casting observations $y_{330}, \ldots, y_{181}$. Finally, we combine two estimates by the procedure in [17] to produce the ultimate estimates. Figure 4 shows the monthly time series of WTD, including estimates for the missing observations.

The plot in Figure 4 shows several important results. A clear nonlinear trend emerges, indicating that the water resources suffered a decline over the years (approximately three times more depth in 2015 than in 1988). At the same time, the monthly figures will follow an almost identical pattern each year (e.g., more depth during summer than during any other time of the year). This data also illustrates the amplitude of the seasonal changes fluctuating along the years. It decreases with the year up to 1999, and suddenly increased thereafter up to about 2008, and decreases up to 2013, with a slightly increases thereafter. Taking another

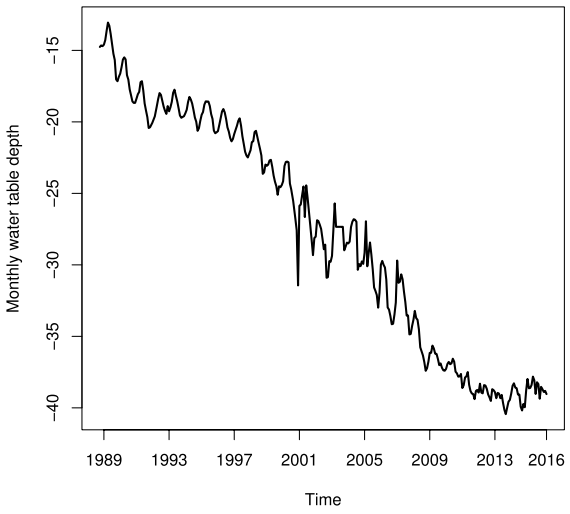

Figure 4. Time series of monthly water table depth between October 1988 and March 2016.

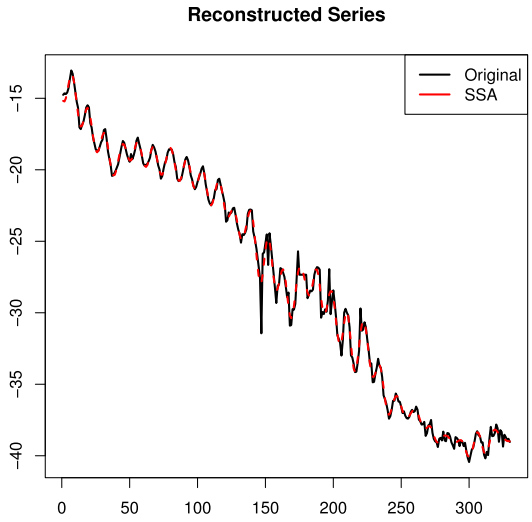

Figure 5. SSA model fit for the WTD data.

Table 1. Summary statistics (first quartile, mean, standard deviation and third quartile) for time series of WTD between October 1988 and March 2016.

Quantity	Q_1	Mean	SD	Q_3
value	-36.87	-27.32	20.08	-8.21

careful look at Figure 4, it reveals that the WTD series has four cycles:

(i) 1988 to 1992, in which the WTD have fallen sharply;
(ii) 1992 to 1999, in which the severity of decline in the WTD decreases in comparison with those of the period 1988 to 1992;
(iii) 1999 to 2013, in which we observe again a sharp decline in the WTD; and
(iv) 2013 to 2015, in which the direction of time series has turned upward.

In addition, some descriptive statistics for the time series of WTD are provided in Table 1. Notations Q_1 and Q_3 in this table, represent the first and third quartiles, respectively.

3.2 Model fitting

In order to fit a model to the monthly WTD data, using the SSA algorithm, we need to set up an appropriate pair of parameters (L, r). Considering the behaviour of the singular values of the trajectory matrix and weighted correlation between pairwise components, for $L = 12$ it can be seen that the first three components are enough to reconstruct the original noise-free time series. The first component captures the trend and the second and third components are associated to the periodic components of the series. These components taken together explain about 95% of the variability in the original time series. Therefore, we fit SSA model using $L = 12$ and $r = 3$. This model will be obtained by diagonal averaging of the matrix:

$$\widetilde{\mathbf{X}} = \mathbf{E_1} + \mathbf{E_2} + \mathbf{E_3} \tag{7}$$

where the matrices $\mathbf{E}_i$, $i = 1, 2, 3$, are defined as before. Figure 5 shows the original time series and the fitted values with that model.

3.3 Further strategies for WTD modelling

It is well known in hydrological studies that, statistical measures such as mean, minimum and maximum are used to construct time series. In the previous section, we used a simple average to construct the WTD time series. In the approach proposed in the section, we use the information from all sampled wells at each time point by considering the distribution of the WTD. Due to the problems in data collection and data quality, the proposed approach can be resistant to outliers. Accordingly, we propose to use box-plots of the WTD observed in the sampled wells at each time point. On one hand it provides more useful information than a statistical measure (e.g. mean values) and, on the other hand, it is less likely to be affected by outliers. Figures 6 and 7 show the box-plot time series for our data set. Note that there is a high number of observations for each of these box-plots. For instance, each box-plots in Figure 6 was constructed by using about 150 samples and, for Figure 7, this number increases to more than 1500 observations. This provides more confidence in the results and conclusions.

Figures 6 and 7 provide similar conclusions:

- There is a considerable decreasing trend,
- There are outliers in all time points (months and years),
- The variation of the distribution of the WTD increased by time,

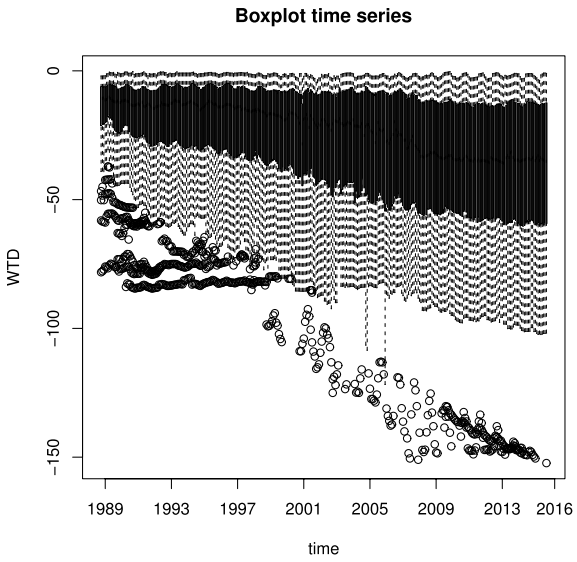

Figure 6. Box-plot of the WTD time series per month, between October 1988 and March 2016.

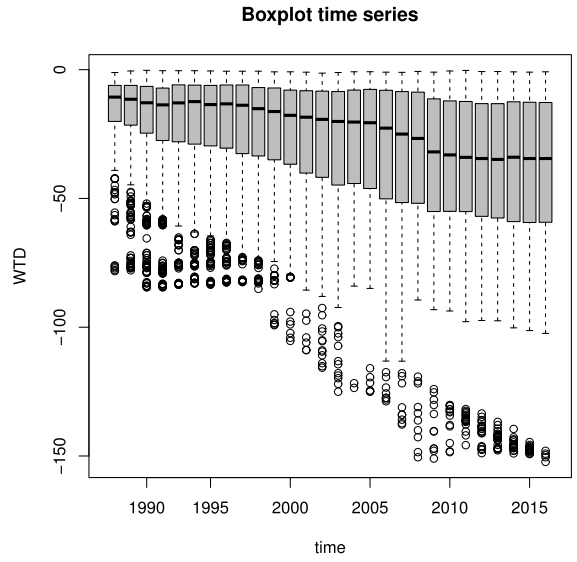

Figure 7. Box-plot of the WTD time series per year, between 1988 and 2016.

- The box-plots are asymmetric in the beginning of the time points and then they approach symmetric in more recent years and months.

This new description of the data may help us to produce a better model using other approaches such as multivariate time series analysis and functional data analysis. Indeed, we know that a box-plot can be constructed using five values: upper bound, third quartile (Q_3), median, first quartile (Q_1) and lower bound. It is also popular to use the following formulas for finding lower and upper bounds:

$$\begin{aligned} \text{upper bound} &= \min(\max(y), Q_3 + 1.5 * IQR) \\ \text{lower bound} &= \max(\min(y), Q_1 - 1.5 * IQR) \end{aligned}$$

where the inter-quartile range can be obtained as $IQR = Q_3 - Q_1$. So, we may consider this information in the framework of the multivariate SSA and produce a new model. This new model uses more information than a simple average as presented before. Therefore, we expect to obtain better results with this proposal.

Firstly, we built an univariate SSA model for each component of the monthly box-plot using the SSA choices given in Table 2.

Table 2. SSA choices for modelling the components of the boxplots.

Component	Lower	Q_1	Med	Q_3	Upper
L	12	12	12	12	12
r	1	3	3	3	5

Then we consider $L = 12$ and $r = 3$ to obtain the results for the multivariate SSA, using the time series of the five box-plot components. Fitting errors (in sample) and forecasting errors (out-of-sample) were evaluated using the root mean squared error (RMSE). In order to find the RMSE for out of sample observations, we examined 30 forecasts and compute RMSE using the following formula:

$$\text{RMSE} = \sqrt{\frac{1}{30}\sum_{i=1}^{30}\left(y_{T-i+1} - \widehat{y}_{h|T-h-i+1}\right)^2}, \tag{8}$$

where T is the length of time series and $\widehat{y}_{h|T-h-i+1}$ denotes h-steps ahead forecast with starting time point $T - h - i + 1$. The results for $h = 6, 12, 24$ steps-ahead are given in Table 3.

Table 3. Root mean square errors for the univariate and multivariate SSA algorithms, for the five box-plot components, considering model fit and out-of-sample model forecast for 6, 12 and 24 steps-ahead.

	Method	Lower	Q_1	Med	Q_3	Upper
Fit	SSA	3.404	0.686	0.776	0.329	0.205
	MSSA	2.613	0.688	0.775	0.330	0.288
6 steps	SSA	1.203	1.750	2.221	1.218	0.660
	MSSA	1.110	1.274	1.543	1.059	0.690
12 steps	SSA	1.597	2.234	2.660	1.330	0.716
	MSSA	1.433	1.618	1.847	1.150	0.777
24 steps	SSA	3.544	3.811	4.999	2.087	0.724
	MSSA	2.903	2.632	2.986	1.637	0.770

Table 3 shows that the MSSA algorithm produces more accurate results than SSA in both in-sample model fit and

out-of-sample model forecasting. The only exception is the time series of the box-plot upper bound, where SSA provided more accurate results than the MSSA for all cases. To better understand this fact, Table 4 shows the linear correlation between the box-plot components, where it is visible that the correlation between "Upper" and the other variables are lower than the other correlations. This is the main reason of a weaker performance of MSSA when compared with the SSA.

Table 4. Linear correlation between the five box-plot components, lower bound, first quartile, median, third quartile, and upper bound.

	Lower	Q_1	Med	Q_3	Upper
Lower	1.000	0.948	0.871	0.806	0.461
Q_1	0.948	1.000	0.958	0.891	0.489
Med	0.871	0.958	1.000	0.958	0.438
Q_3	0.806	0.891	0.958	1.000	0.499
Upper	0.461	0.489	0.438	0.499	1.000

4. CONCLUSION

In this paper, we analysed monthly WTD records that were collected by the regional water authority of Hamedan, in Iran, as part of their monitoring system program. We used two approaches for data preparation. The first approach employed a simple average on the observation for each time point/month. Whereas in the second approach we extracted the monthly time series for the five box-plot components, lower bound, first quartile, median, third quartile, and upper bound. We used SSA to model the fluctuation in both approaches, and the multivariate SSA for the second approach. We observed that the SSA algorithm fits the data very well. Generally speaking, the descriptive analysis of the data and the SSA fitted model showed that there is a decreasing trend in the WTD, which might be the result of not enough intervention by the related institutions to prevent the severity of this crucial problem in the area. It should be mentioned that the data sets for this study suffered from several problems including (1) many mistakes in registering the right name of the well; (2) missing information; and (3) lack of access to recent data. The first problem was solved by using the other information such as longitude and latitude of the wells, which was time consuming. The second problem was solved by using statistical methods for imputation of missing observations. The third problem required funds and specific official permission that could not be solve for this analysis.

The results showed that the propose strategy of considering the time series of the box-plot components in the MSSA algorithm prided better results that using the univariate SSA or the SSA for each of the five box-plot components. This strategy can be applied to WTD in other locations/countries, other applications in the field of hydrology, and also to other areas of application.

Received 10 June 2021

REFERENCES

[1] Awe, O. O., Mahmoudvand, R., and Rodrigues, P. C. (2020). Non-Negative Time Series Reconstruction via Singular Spectrum Analysis: A Case Study of Precipitation Dynamics in Nigeria. *Fluctuation and Noise Letters,* 2050045.

[2] Feliks, Y., Small, J., and Ghil, M. (2021). Global oscillatory modes in high-end climate modeling and reanalyses. *Climate Dynamics*, **57(11)**, 3385–3411.

[3] Ghanati, R; Kazemhafizi, M; Mahmoudvand, R and Fallahsafari, M. (2016). Filtering and parameter estimation of surface-NMR data using singular spectrum analysis. *Journal of Applied Geophysics,* **130** 118–130.

[4] Gilliom, R.J., Alley, W. M., and Gurtz, M.E. (1995). Design of the National Water-Quality Assessment Program—Occurrence and Distribution of Water-Quality Conditions: U.S. Geological Survey Circular 1112, 31 p.

[5] Golyandina N, and Zhigljavsky A. (2013). *Singular Spectrum Analysis for Time Series.* Springer. MR3024734

[6] Golyandina N, Nekrutkin V, and Zhigljavsky A. (2001) *Analysis of Time Series Structure: SSA and related techniques.* Chapman & Hall/CRC. MR1823012

[7] Groth A, Feliks Y, Kondrashov D, Ghil M (2017) Interannual variability in the North Atlantic ocean's temperature field and its association with the wind stress forcing. *J Clim* **30(7)**: 2655–2678

[8] Hassani, H., and Mahmoudvand, R. (2013) Multivariate singular spectrum analysis: A general view and new vector forecasting approach. *International Journal of Energy and Statistics,* **1(01)** 55–83.

[9] Hassani, H., Huang, X., Gupta, R., and Ghodsi, M. (2016). Does sunspot numbers cause global temperatures? A reconsideration using non-parametric causality tests. *Physica A: Statistical Mechanics and its Applications*, **460**, 54–65.

[10] Hassani, H., Coreman, J., Heravi, S., and Easaw, J. (2018). Forecasting Inflation Rate: Professional Against Academic, Which One is More Accurate. *Journal of Quantitative Economics*, **16(3)**, 631–646.

[11] Jahromi, F. K., Sabziparvar, A. A., and Mahmoudvand, R. (2021). Spectral analysis of soil temperature and their coincidence with air temperature in Iran. *Environmental Monitoring and Assessment,* **193(2)** 1–14.

[12] Knotters, M. (2001) *Regionalized Time Series Models for WTDs.* Thesis (Ph.D. in Hydrology). Wageningen University (WUR), Wageningen.

[13] Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G.,Geller, G., Quinn, N., and Blind, M. (2013). Integrated environmental modeling: avision and roadmap for the future. *Environ. Model. Softw.* **39** 3–23.

[14] Mahmoudvand, R. and Rodrigues, P.C. (2018). A New Parsimonious Recurrent Forecasting Model in Singular Spectrum Analysis, *Journal of Forecasting.* **37** 191–200. MR3771548

[15] Mahmoudvand, R.; and Zokaei, M. (2012). On the Singular Values of the Hankel Matrix with Application in Singular Spectrum Analysis. *Chilean Journal of Statistics.* **3(1)** 43–56. MR2922267

[16] Mahmoudvand, R.; Najari, N.; and Zokaei, M. (2013). On the Parameters for Reconstruction and Forecasting in the Singular Spectrum Analysis. *Communication in Statistics: Simulations and Computations.* **42** 860–870. MR3039618

[17] Mahmoudvand, R.; and Rodrigues, P. (2016). Missing value imputation in time series using Singular Spectrum Analysis. *International Journal of Energy and Statistics* doi: 10.1142/S2335680416500058.

[18] Omran, E. S. E. (2016). A stochastic simulation model to early predict susceptible areas to water table level fluctuations in North

Sinai, Egypt. *The Egyptian Journal of Remote Sensing and Space Science.* **19(2)** 235–257.
[19] Pahl-Wostl, C., Arthington, A., Bogardi, J., Bunn, S.E., Hoff, H., Lebel, L., Nikitina, E.,Palmer, M., Poff, L.N., and Richards, K. (2013). Environmental flows and water governance: managing sustainable water uses. *Curr. Opin. Env. Sust.* **5** 341–351.
[20] Pahl-Wostl, C., Jeffrey, P., Isendahl, N., and Brugnach, M. (2011). Maturing the new water management paradigm: progressing from aspiration to practice. *Water Resource Manage.* **25** 837–856.
[21] Rodrigues, P.C. and de Carvalho, M. (2013). Spectral modeling of time series with missing data. *Applied Mathematical Modelling.* **37** 4676–4684. MR3020603
[22] Rodrigues, P.C. and Mahmoudvand, R. (2020). A new approach for the vector forecast algorithm in singular spectrum analysis. *Communications in Statistics: Simulation and Computation.* DOI: 10.1080/03610918.2019.1664578 MR4068476
[23] Rodrigues, P.C. and Mahmoudvand, R. (2020). The benefits of multivariate singular spectrum analysis over the univariate version. *Journal of the Franklin Institute.* **355** 544–564. DOI: 10.1016/j.jfranklin.2017.09.008 MR3739602
[24] Rodrigues, P.C., Awe, O.O., Pimentel, J. and Mahmoudvand, R. (2020). Modelling the behaviour of currency exchange rates with singular spectrum analysis and artificial neural networks. *Stats.* **3** 137–157. DOI: 10.3390/stats3020012.
[25] Rodrigues, P.C., Pimentel, J., Messala, P., and Kazemi, M. (2020). Decomposition and forecasting of mutual investment funds using singular spectral analysis. *Entropy.* **22** 83. DOI: 10.3390/e22010083. MR4072097
[26] Sanei, S., and Hassani, H. (2015). *Singular spectrum analysis of biomedical signals.* CRC press.
[27] Stave, K. (2010). Participatory system dynamics modeling for sustainable environmental management: observations from four cases. *Sustainability* **2** 2762–2784.
[28] Stuart, M., Maurice, L., Heaton, T., Sapiano, M., Micallef Sultana, M., Gooddy, D., and Chilton, P. (2010). Groundwater residence time and movement in the Maltese islands: a geochemical approach. *Appl. Geochem.* **25** 609–620.
[29] Videira, N., Lopes, R., Antunes, P., Santos, R., and Casanova, J.L. (2012). Mapping maritime sustainability issues with stakeholder groups. *Syst. Res. Behav. Sci.* **29** 596–619.
[30] Voinov, A., and Bousquet, F. (2010). Modelling with stakeholders. *Environ. Model. Softw.* **25** 1268–1281.
[31] Voinov, A., Kolagani, N., McCall, K.M., Glynn, D.P., Kragt, E.M., Ostermann, O.F., and Ramu, P. (2016). Modeling with stakeholders - next generation. *Environ. Model. Softw.* **77**196e77,220.
[32] Wang, F., Shen, Y., Chen, Q., and Wang, W. (2021). Bridging the gap between GRACE and GRACE Follow-on monthly gravity field solutions using improved multichannel singular spectrum analysis. *Journal of Hydrology*, **594**, 125972.

Rahim Mahmoudvand
Department of Statistics
Bu-Ali Sina University
Iran
E-mail address: r.mahmodvand@gmail.com

Mehrdad Barati
Department of Geology
Bu-Ali Sina University
Iran
E-mail address: m.barati@basu.ac.ir

Asghar Seif
Department of Statistics
Bu-Ali Sina University
Iran
E-mail address: a.seif@basu.ac.ir

Sahar Ranjbaran
Department of Geology
Bu-Ali Sina University
Iran
E-mail address: s.ranjbaran@basu.ac.ir

Paulo Canas Rodrigues
Department of Statistics
Federal University of Bahia
Brazil
E-mail address: paulocanas@gmail.com