RECONSTRUCTION of GAPS in FLOW SERIES USING SINGULAR SPECTRUM ANALYSIS (SSA)

Kübra AĞAÇ*, Özlem BAYDAROĞLU and Kasım KOÇAK

Istanbul Technical University, Graduate School of Science Engineering and Technology, Atmospheric Sciences, Istanbul, Turkey (*agac@itu.edu.tr)

❖ INTRODUCTION

Meteorological data obtained from observations include gaps due to deficiency and/or errors of measurements. This condition causes several problems such as prediction inaccuracy, model inconsistency and data mining.

Here, we use Singular Spectrum Analysis (SSA) to fill the gaps in flow records for a univariate record and it leads to make temporal correlations. Then for a multivariate record, Multi-channel SSA (M-SSA) is used to fill the missing data for to make spatial and temporal correlation. These SSA and M-SSA are a-state-of-art models. Results and performance criteria are encouraging.

❖ METHOD

SSA: Singular Spectrum Analysis (SSA) (Broomhead and King, 1986a; Fraedrich, 1986; Ghill et al., 2002) is a state-of-the-art spectral, data-adaptive and nonparametric method. SSA decomposes an original time series to trend (if exists), oscillatory and noise components by way of a singular value decomposition. The term singular spectrum comes from the spectral (eigenvalue) decomposition of a matrix into its spectrum of eigenvalues (Elsner and Tsonis, 1996). The basic SSA algorithm has stages of decomposition and reconstruction. The decomposition stage requires embedding and singular value decomposition. In this stage, there are two basic steps. The former is embedding the sampled time series in a vector space of dimension M, the latter is computing the MxM lag-covariance matrix of the data (Vautard and Ghill, 1989). The reconstruction stage demands the grouping to make subgroups of the decomposed trajectory matrices and diagonal averaging to reconstruct the new time series from the subgroups (Myung, 2009).

❖ DATA

Monthly river flows of two different dams data (one of is between 1938 and 2014 and the other one is between 1960 and 2014) have been used and 10%, 30%, 50%, 70% of data were removed then SSA was applied. For this calculation SSA-MTM Toolkit was used (http://web.atmos.ucla.edu/tcd//ssa/). Correlation coefficients, RMSE (Root Mean Square Error), RE (Relative Error), NRMSE (Normalized Root Mean Square Error) were found for different amount of gaps.

❖ RESULTS

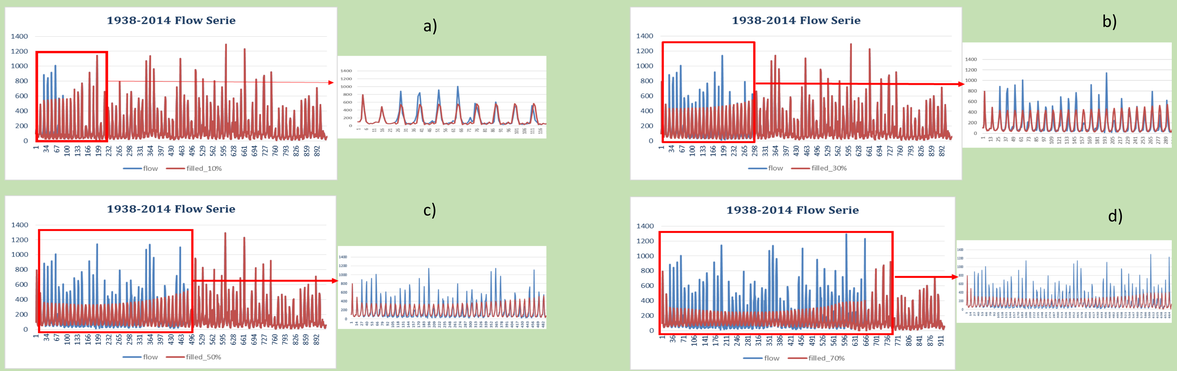

Figure.1: Original and reconstructed flow series between 1938-2014 with a) 10% b) 30% c) 50% d)70% filled gaps

In Fig.1 original and reconstructed flow series were plotted and correlation coefficient between original and filled part were calculated (Table1). Also, **Root Mean Square Error (RMSE), Relative Error (RE)** and **Normalized Roort Mean Square Error (NRMSE)** were obtained (Table2).

❖ RESULTS

Table.1: Correlation coeff. between original and filled part of flow series.

	Correlation Coefficient (r)
Flow-10% filled	0.98
Flow-30% filled	0.94
Flow-50% filled	0.88
Flow-70% filled	0.78

Table.2: RMSE, RE, NRMSE

	10% filled	30% filled	50% filled	70% filled
RMSE	38.83	67.95	98.93	136.16
Relative E.	0.19	0.34	0.49	0.68
NRMSE	0.03	0.05	0.08	0.11

❖ RESULTS

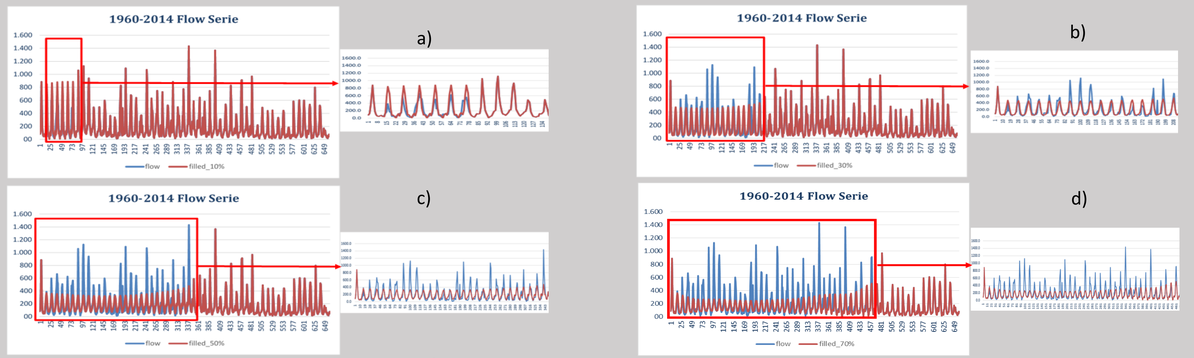

Figure.2: : Original and reconstructed flow series between 1960-2014 with a) 10% b) 30% c) 50% d)70% filled gaps

In Fig.2 original and reconstructed flow series were plotted and correlation coefficient between original and filled part were calculated (Table3). Also, **Root Mean Square Error (RMSE), Relative Error (RE)** and **Normalized Roort Mean Square Error (NRMSE)** were obtained (Table4).

Table.1: Correlation coeff. between original and filled part of flow series.

	Correlation Coefficient (r)
Flow-10% filled	0.96
Flow-30% filled	0.93
Flow-50% filled	0.84
Flow-70% filled	0.73

Table.2: RMSE, RE, NRMSE

	10% filled	30% filled	50% filled	70% filled
RMSE	59.97	78.16	119.46	154.25
Relative E.	0.28	0.37	0.56	0.72
NRMSE	0.04	0.06	0.08	0.11

❖ CONCLUSION

Two different flow series were reconstructed after different lenght of gaps removing process. Correlation coefficients and error terms were calculated. According to NRMSE, for both series 10 and 30 % of data were reconstructed successfully but when the gap length is increased the gap filling starts to be weaker as expected. But for this process, SSA MTM-Toolkit can predict the oscillation in the series accomplishedly. Later with another studies beside SSA also M-SSA (Multi-Channel SSA) will used for gap filling.

❖ REFERENCES

Broomhead, D.S., and G.P. King, 1986a. Extracting qualitative dynamics from experimental data, Physica D, 20, 217–236.
Elsner and Tsonis, 1996. Singular Spectrum Analysis: A New Tool in Time Series Analysis, Springer Science & Business Media
Ghil, M., R. M. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, et al. (2002) "Advanced spectral methods for climatic time series", Rev. Geophys. 40(1), 3.1–3.41.
No Kang Myung,2009. Singular Spectrum Analysis. Msc Thesis, University of California.
Montgomery, D.C.; Johnson, L.A. Forecasting and Time Series Analysis; McGraw-Hill: New York, 1976.