# KSOM clustering as a possible cure for the wicked water problem of inadequate data for water resources planning

**ADEBAYO J. ADELOYE**[1] **& RABEE RUSTUM**[2]
1 *School of the Built Environment, Heriot-Watt University, Edinburgh EH14 4AS, UK*
a.j.adeloye@hw.ac.uk
2 *College of Engineering, Dammam University, Dammam, Saudi Arabia*

## INTRODUCTION: THE KEY WICKED WATER PROBLEM

A common wicked problem during the planning and management of water resources systems is the lack of the necessary runoff data, resulting in huge uncertainties and gross inadequacy of schemes to deliver (Adeloye, 1996). A possible solution is to reconstruct/extend the runoff data using correlation and regression with the often much longer rainfall records. However, regression can only analyse for one dependent variable and is infeasible for prediction if the predictor is missing. The latter situation is prevalent in most developing countries where the available data records are often littered with gaps and missing values.

To overcome this problem, artificial intelligence techniques can be used. The Kohonen self organising map, KSOM (Kohonen *et al*., 1996) is a class of unsupervised artificial neural networks whose powerful clustering capability can lead to multi-variate prediction without being encumbered by missing values.

## KSOM TOOL FOR SOLVING THE WICKED DATA INADEQUACY PROBLEM

Application of the KSOM for estimating missing water quality data was recently reported by Rustum & Adeloye (2007). The procedure involves clustering the patterns in an input data set in such a way that similar patterns are represented by the same output vector or best matching unit (BMU), which is that for which the Euclidian distance is a minimum. The missing values in the input vector are then assumed to take on the corresponding values in the BMU (Fig. 1).

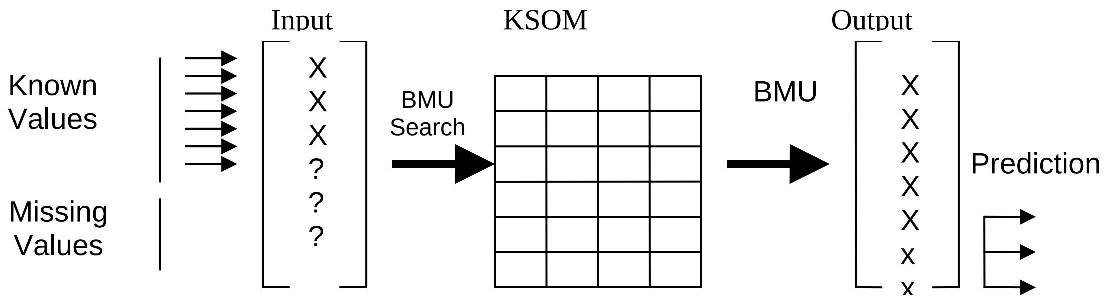


**Fig. 1** Schematic illustration of prediction of missing components of an input data vector using the KSOM. BMU is the best matching unit.

## CASE STUDY

The Osun basin in Nigeria is highly developed and the latest proposed development is a pumped-storage scheme at its lower reaches just before the river enters the Lagos lagoon. However, there were no measured runoff data at the proposed abstraction site with which to estimate flow probabilities, although there were short (<10 years) runoff data at two upstream stations, in addition to four much longer rainfall stations. The KSOM clustering of the runoff and rainfall records was thus carried out, which led to extended runoff records at the two upstream stations, from which the runoff at the proposed abstraction site was derived.

## RESULTS

Figure 2 shows the time series of the observed and KSOM-predicted data, from which it is quite clear that the performance of the KSOM is very good. The prediction efficiency of the KSOM was generally above 0.94 for all the predictands (see Adeloye & Rustum, 2009). Figure 2 also contains the in-filled missing values, which generally accord with the observed trends.

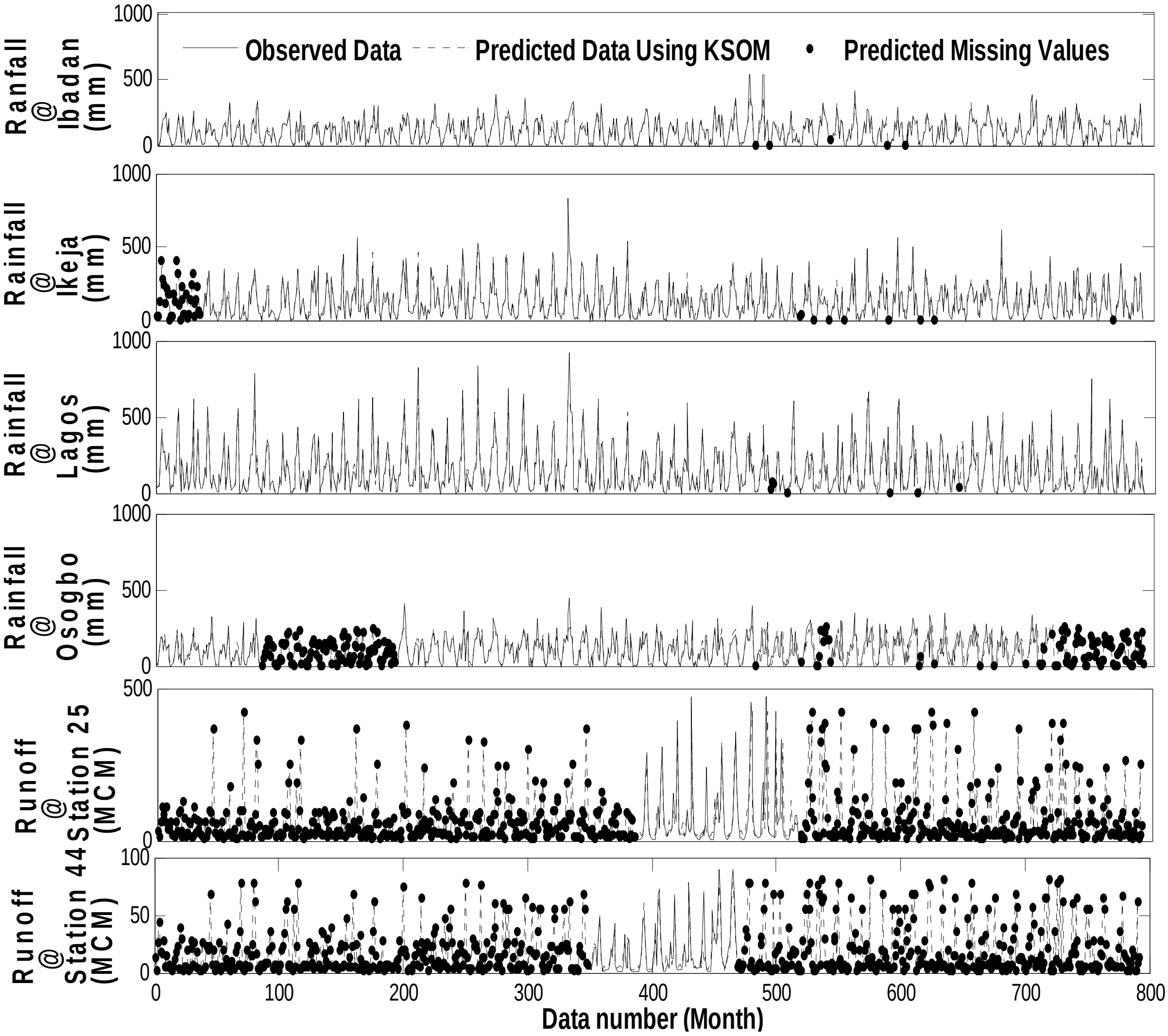


**Fig. 2** Time series plots of observed and KSOM-predicted data at four rainfall sites and two runoff sites in Osun basin, southwest Nigeria.

## KEY LESSONS

The KSOM has been demonstrated as a very useful tool for solving the problem of lack or inadequacy of data. KSOM is a multivariate approach and is unaffected by missing values; indeed one of its important outcomes is the prediction of such missing values.

## REFERENCES

Adeloye, A. J. (1996) *Water Resour. Manage.* **10**(1), 45–79.

Adeloye, A. J. & Rustum, R (2009) Lagos (Nigeria) Igbonla water supply project.

Kohonen, T., Simula, O. & Visa, A. (1996) IEEE **84**(10), 1358–1384.

Rustum, R. & Adeloye, A. J. (2007) *J. Environ. Engng ASCE* **133**(9), 909–916.