# Reducing the uncertainty associated with water resources planning in a developing country basin with limited runoff data through AI rainfall–runoff modelling

**ADEBAYO J. ADELOYE**
*School of the Built-Environment, Heriot-Watt University, Riccaton, Edinburgh EH14 4AS, UK*
a.j.adeloye@hw.ac.uk

**Abstract** A major bane of water resources assessment in developing countries is the insufficiency or total lack of hydrometeorological data, resulting in huge uncertainties and ineffectual performance of water schemes. This study reports on the application of the Kohonen Self-organizing Map (KSOM) unsupervised artificial neural networks in harnessing the multivariate correlations between the rainfall and runoff for an inadequately gauged basin in southwest Nigeria, for the sole purpose of extending the runoff records, and through them, reducing water resources planning uncertainty associated with the use of short data records. The extended runoff records were then analysed to determine possible abstractions from the main river source at different exceedence probabilities. The study demonstrates the successful use of emerging tools in reducing the uncertainty associated with lack or insufficiency of data for water resources planning assessment.

**Key words** water resources assessment; hydrological data; Kohonen Self Organising Map (KSOM); reliability; water abstractions; Nigeria

## INTRODUCTION

Planning and management of water resources systems require that adequate and reliable hydro-meteorological data are available (Adeloye, 1996). However, this is often not the situation in most developing countries where the runoff data are either unavailable or when available, they are too short and riddled with numerous gaps and outliers. One commonly adopted solution in such situations is to reconstruct and extend the available runoff data records using the typically much longer rainfall data records from rainfall–runoff modelling such as regression analysis. However, traditional rainfall–runoff modelling, including regression, can only analyse for one dependent variable and is unfeasible for prediction if the predictor is missing (Adeloye, 2009).

To overcome this problem, artificial intelligence (AI), multivariate techniques unhindered by missing values and noise in the available data 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 reduce a multi-dimensional data array into a two dimensional array of features or best matching units (BMUs), which then form the basis for multivariate prediction. This approach was used to reconstruct and extend monthly rainfall and runoff data for stations within the lower Osun basin in southwest Nigeria for the purpose of water resources assessment.

In the next Section, further details about the KSOM modelling are given. This is then followed by the methodology applied for the Osun catchment study. Finally, the results are presented and discussed.

## KSOM MODELLING

The KSOM (also called feature map or Kohonen map) is one of the most widely used artificial neural network algorithms (Kohonen *et al.*, 1996) and its principal goal is to transform an incoming signal pattern of arbitrary dimension into a two-dimensional (2-D) discrete map. It involves clustering the input patterns in such a way that similar patterns are represented by the same output neurons, or by one of its neighbours.

The KSOM consists of two layers: the multi-dimensional input layer and the competitive or output layer; both of these layers are fully interconnected as illustrated in Fig. 1. The output layer

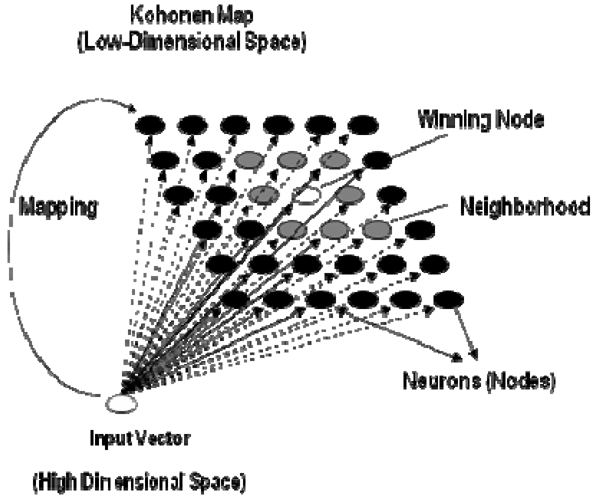


**Fig. 1** Illustration of the winning node and its neighbourhood in the Kohonen Self-organizing Map.

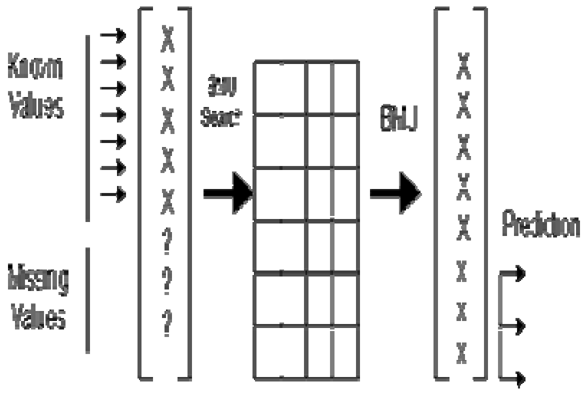


**Fig. 2** Prediction of missing components of the input vector using the Kohonen Self-organizing Map (BMU = best matching unit).

consists of $M$ neurons arranged in a 2-D grid of nodes. Each node or neuron $i$ ($i$ = 1, 2, …, $M$) is represented by an $n$-dimensional weight or reference vector $Wi = [w_{i1}, \ldots, w_{in}]$, where $n$ is the dimension of each input vector, i.e. the maximum number of variables in the input vector. In training, an input vector is randomly selected and its most similar reference vector, i.e. BMU, is identified by determining the closest to it in terms of the Euclidian distance, $D_i$:

$$D_i = \sqrt{\sum_{j=1}^{n} m_j \left(x_j - w_{ij}\right)^2}\,; i = 1,2,\ldots,M \tag{1}$$

where $x_j$ is the $j$th element of the current input vector; $w_{ij}$ is the $j$th element of the code vector $i$; and $m_j$ is the so called "mask" which is used to include in ($m_j = 1$), or exclude from ($m_j = 0$), the calculation of the Euclidian distance, the contribution of a given element $x_j$ of the input vector. Once identified, the elements of the BMU are updated and the process continues until some stated stopping criteria are reached. Rustum & Adeloye (2007) present additional comprehensive information about training the KSOM, including various indices for assessing its performance.

The application of the KSOM for prediction purposes is illustrated in Fig. 2 (see also Rustum & Adeloye, 2007). First, the model is trained using the available data set. Then the depleted vector, i.e. with the predictand either missing or deliberately removed, is presented to the KSOM to identify its BMU using the computed Di (equation (1)). The values for the missing variables are then obtained as their corresponding values in the BMU.

# METHODOLOGY

## Case study and data

The Osun basin in southwest Nigeria is highly developed with many water abstractions and impounding schemes, as shown in Fig. 3. The latest development being proposed on the river is an off-line, pumped storage scheme at Igbonla, just before the river enters the Lagos lagoon. To achieve this, probabilities associated with different pumped abstraction scenarios, as shown in Table 1, for filling the reservoir will need to be determined. Without runoff data at Igbonla, however, this proved an impossible task.

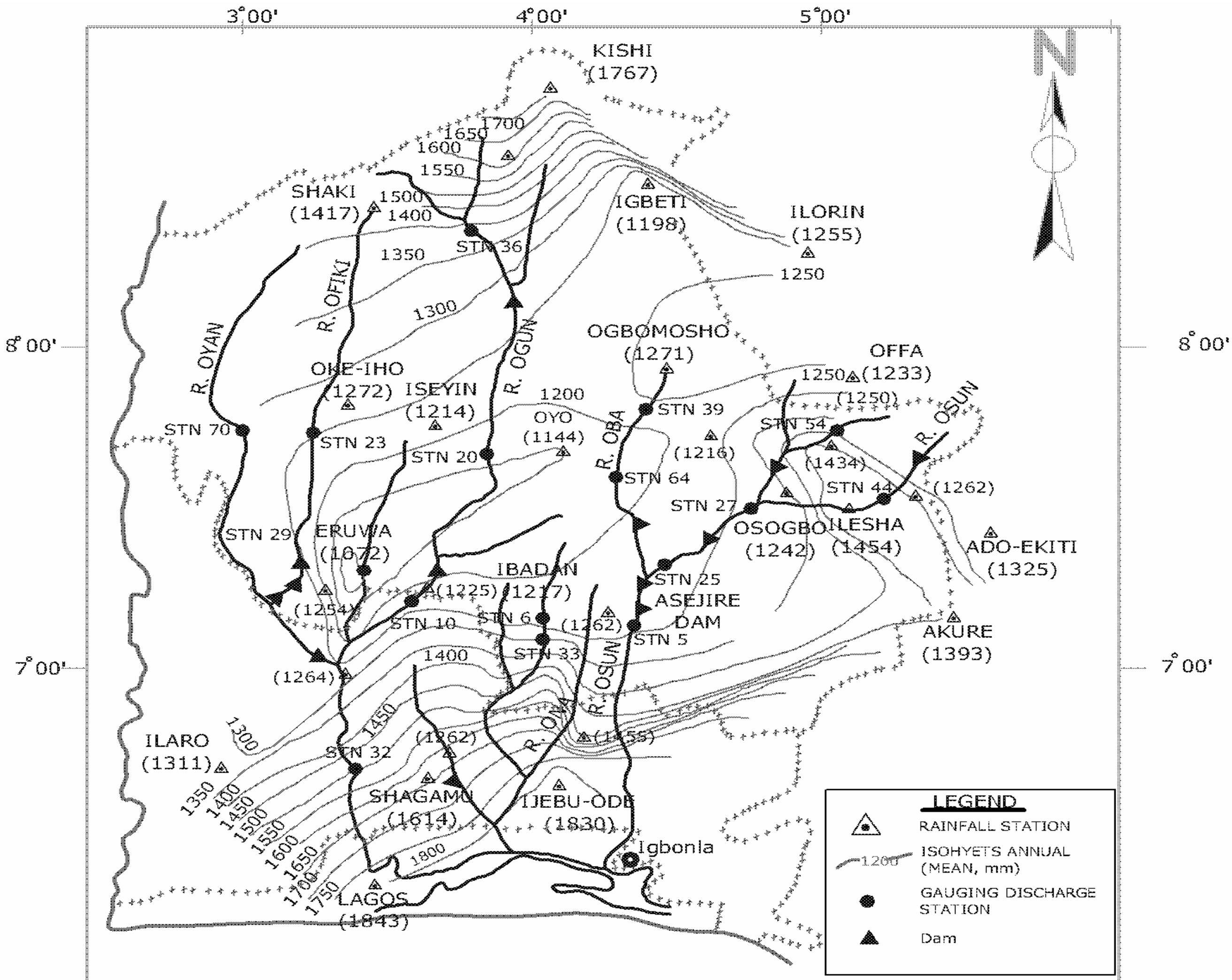


**Fig. 3** The Osun basin map showing isohyets and location of main towns, dams, runoff and rainfall stations.

**Table 1** Proposed abstraction scenarios from River Osun at Igbonla.

| Target year | Desired daily pumping rate (MCM) |
|---|---|
| 2011 | 0.63 |
| 2015 | 1.20 |
| 2020 | 1.82 |

Rainfall data are widely available within the basin, including: monthly data for Osogbo and Ibadan (1970–1983); Lagos Marina, Ijebu-Ode and Abeokuta (1987–2006, albeit with several missing values); and Lagos Island, Ikeja and Ibadan (1941–2006). Thus potentially, it could be assumed that rainfall data are available for 1941–2006, i.e. 66 years, which can be used as the basis for extending/reconstructing monthly runoff records at desired locations in the basin.

Stations with runoff records in the basin include STN 44 (monthly, 1970–1979) and STN 25 (monthly, 1973–1983), both upstream of the Igbonla site and with far too short record lengths to carry out any meaningful statistical analysis. One way of improving the situation is to harness the multivariate correlations between the rainfall and runoff, and thus extend the runoff records. All other gauging sites in Fig. 3 have no time series runoff records.

### KSOM modelling

The KSOM analysis used monthly rainfall at four stations, namely Ibadan (no. 65208), Ikeja (no. 65201), Lagos Island (no. 65203), and Osogbo (no. 65215) and the two runoff stations (25 and 44) for the multivariate, KSOM modelling. This led to the reconstruction of runoff at stations 25 and 44 for the period 1941–2006. The self-organizing map (SOM) toolbox for Matlab 5 was used for this case study. The toolbox was developed by the SOM team at the Helsinki University of Technology, Finland (http://www.cis.hut.fi).

### Runoff reconstruction and quantiles at Igbonla

Simple area-ratio scaling (Loucks *et al.*, 1981) was used for transposing runoff data to Igbonla. Since STN 25 is closer to Igbonla, its reconstructed (1941–2006) data formed the basis of the transposition using the respective catchment areas, i.e. 8174 km$^2$ at Igbonla and 4325 km$^2$ at STN 25. Allowance was also made for the Asejire dam (catchment area = 5646 km$^2$) and its downstream compensation releases, which was taken to be 10% of STN 25 mean runoff. Finally, runoff quantiles at Igbonla were determined by fitting the 3-p lognormal distribution as described by McMahon & Adeloye (2005) to the 1-month annual low-flow series.

## RESULTS AND DISCUSSIONS

### Rainfall–runoff analysis using the KSOM

The KSOM component planes which visually illustrate the correlations between the various variables are shown in Fig. 4. The runoff component planes at both stations 44 and 25 are broadly similar with coincidental occurrences of high and low runoff magnitudes. However, in relation to the rainfall stations, runoff at both stations 44 and 25 appears to correlate more with the Osogbo rainfall. This would suggest that Osogbo rainfall would be a much better predictor for the runoff at both stations 44 and 25. Information such as this if available *a priori* would have made Osogbo rainfall to be sole predictor variable for runoff at the sites if a univariate regression approach were to be adopted.

The time series plots of the monthly data are shown in Fig. 5 both for the observed and those estimated by the KSOM and confirm the good performance of the KSOM. In general, the Nash-Sutcliffe efficiency was above 94%, which is very good.

### Low flow quantiles

The probabilities corresponding to the planned abstractions rates (see Table 1) as estimated from the fitted 3-parameter lognormal distribution are summarised in Table 2. As seen in Table 2, except for the 2011 abstraction, where the probability of achieving the required abstraction or higher was 59%, all the other abstraction scenarios had unacceptably low probabilities. Table 3 shows the abstractions at Igbonla for a range of acceptable reliabilities, from which it is clear that achieving such reliabilities would entail large water shortages especially in years 2015 and 2020.

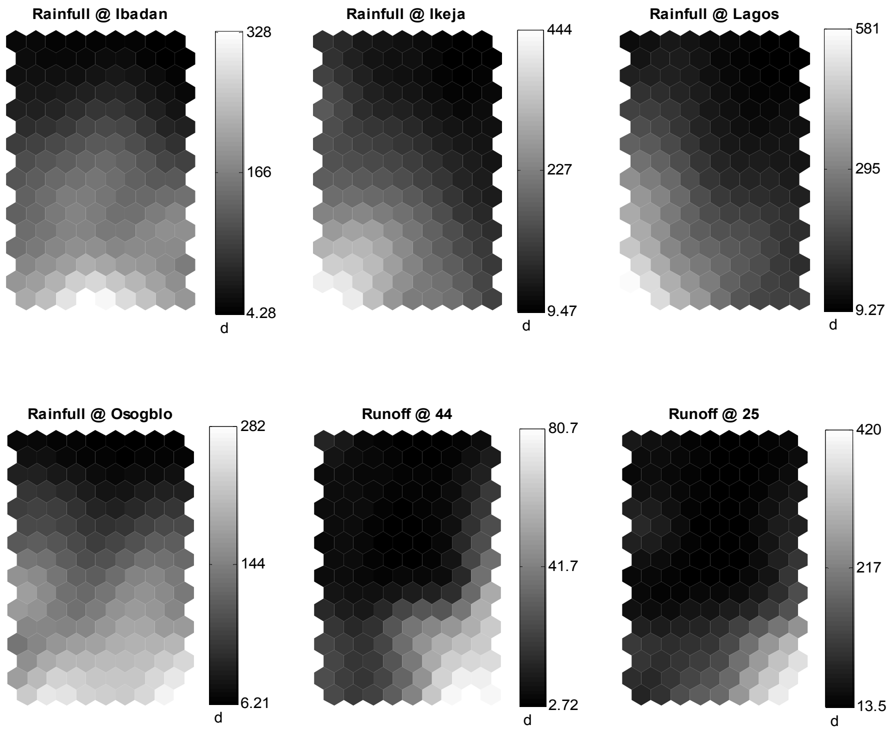


**Fig. 4.** KSOM component planes for rainfall and runoff data, Osun Basin, Nigeria.

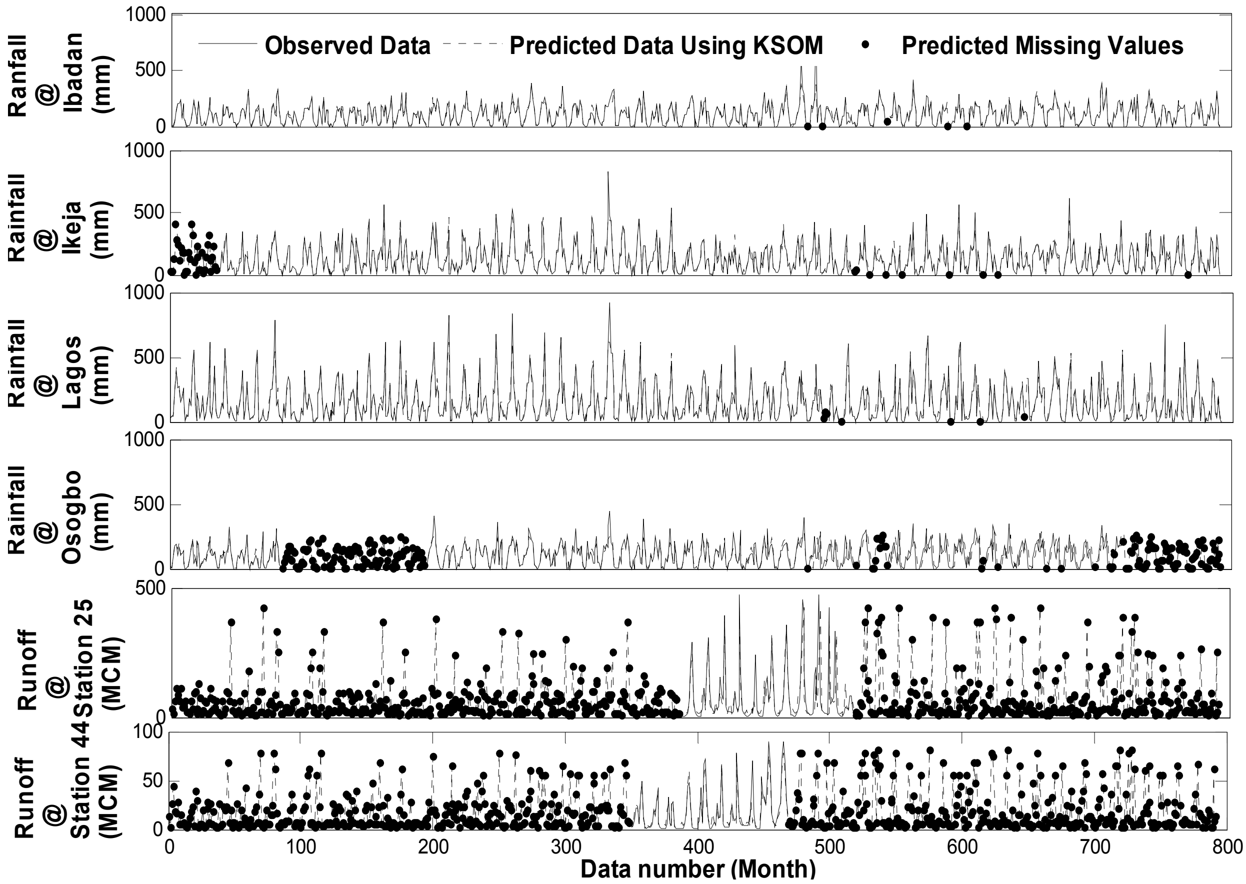


**Fig. 5** Performance of the KSOM in predicting monthly rainfall and runoff.

**Table 2** Assessed probabilities of projected abstractions at Igbonla.

| Year | Abstraction rate, q (MCM/month) | Cumulative probability | Exceedence probability (%) |
|---|---|---|---|
| 2011 | 19.16 | 0.41 | 58.67 |
| 2015 | 36.50 | 0.971 | 2.92 |
| 2020 | 55.36 | 0.998 | 0.1808 |

**Table 3** Assessed shortfall at Igbonla for acceptable exceedence probability levels.

| Exceedence probability (%) | Possible abstraction rate (MCM.month$^{-1}$) | Shortfall (%) | | |
|---|---|---|---|---|
| | | Year 2011 | Year 2015 | Year 2020 |
| 98 | 14.09099 | 26.5 | 61.4 | 74.5 |
| 95 | 14.8307 | 22.6 | 59.4 | 73.2 |
| 90 | 15.64421 | 18.3 | 57.1 | 71.7 |
| 80 | 16.86673 | 12.0 | 53.8 | 69.5 |

## CONCLUSIONS

This study has demonstrated the usefulness of the KSOM as a viable rainfall–runoff data extension tool in situations where the available data for water resources assessment are limited. Because of its powerful clustering ability, the implementation of the KSOM is unhindered by missing values; indeed, as seen in this work, one of the outcomes of the KSOM modelling is the provision of reliable estimates for the missing values. Also, as a consequence of this, data extension and reconstruction using the KSOM can be accomplished even when some of the predictor variables are unavailable, something that will be impossible with traditional rainfall–runoff modelling tools such as univariate regression. The methodology thus offers huge potential for combating the wicked problem of inadequate and insufficient data for meaningful water resources planning and management, and the associated uncertainty.

On the basis of the results obtained here, the runoff at Igbonla will only provide the projected abstraction rates with a low level of reliability. In the UK, it is customary for supply sources to have at least 98% reliability. Achieving such a high level of reliability at Igbonla would require significant reductions in the abstraction rates, as shown in Table 3, or supplementing the supply with water from other sources, e.g. groundwater which is abundant in the basin or desalination which is feasible but potentially expensive.

## REFERENCES


Adeloye, A. J. (1996) An opportunity loss model for estimating value of streamflow data for reservoir planning. *Water Resources Management* **10**(1), 45–79.

Adeloye, A. J. (2009) The relative utility of multiple regression and ANN models for rapidly predicting the capacity of water supply reservoirs. *Env. Modelling & Software* **24**(10), 1233–1240.

Kohonen, T., Simula, O. & Visa, A. (1996) Engineering applications of the self organising map. *IEEE* **84**(10), 1358–1384.

Loucks, D. P., Stedinger, J. R. & Haith, D. A. (1981) *Water Resources Systems Planning and Analysis*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, USA.

McMahon, T. A. & Adeloye, A. J. (2005) *Water Resources Yield*. Water Resources Publications LLC (www.wrpllc.com), Colorado, USA.

Rustum, R. & Adeloye, A. J. (2007) Replacing outliers and missing values from activated sludge data using Kohonen self organising map. *J. Environ. Engng, ASCE* **133**(9), 909–916.