Analysis of Missing Data for River Flow Using Statistical Tool Group Method of Data Handling (GMDH)

Conference Paper · December 2016

4 authors:

Shanker Lal Meghwar
Mehran University of Engineering and Technology
4 PUBLICATIONS   **0** CITATIONS

SEE PROFILE

Waqas Ahmed
Mehran University of Engineering and Technology
10 PUBLICATIONS   **9** CITATIONS

SEE PROFILE

Rakhshinda Bano
Mehran University of Engineering and Technology
8 PUBLICATIONS   **2** CITATIONS

SEE PROFILE

Muneer Ahmed Memon
Mehran University of Engineering and Technology
2 PUBLICATIONS   **0** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Water Resources View project

Remote Sensing in Civil Engineering View project

Analysis of Missing Data for River Flow Using Statistical Tool Group Method of Data Handling (GMDH)

Shanker Lal Meghwar, Waqas Ahmed, Rakhshinda Bano, Muneer Ahmed Memon
US-Pakistan Center for Advanced Study in Water(USPCASW), Mehran UET, Jamshoro, Sindh, Pakistan
shanker.uspcasw@admin.muet.edu.pk, wapathan.uspcasw@faculty.muet.edu.pk,
rbano.uspcasw@faculty.muet.edu.pk, sscientist.uspcasw@admin.muet.edu.pk

Abstract

Water resources management requires handling of large datasets. Transforming the large water resources datasets into the information requires preprocessing to deduce it into the valuable information. While transforming the data, researchers face following problems: 1) the datasets are not sufficient; 2) hard to locate; 3) and are inconsistent or non-compatible. Recent development in the data driven modeling has given researchers to overcome these problems. This research paper presents the quantitative analysis of the missing flow data of **Guddu Barrage** using group method of data handling (GMDH). This study is aimed at finding the missing data using the most appropriate approach of statistics in GMDH and to compare the missing data with actual data for checking and validation of model in GMDH. Flow data having upstream (US) and downstream (DS) flow of sixteen years (i.e. 1998 to 2013) is trained in GMDH using Curve Fitting (CF) and Load Forecast(LF) Analysis. After data mining, accuracy of mined data is assessed using available statistical indices in GHDH i.e. mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2). It is observed from analyzed data that: for US flow data, R^2 is 0.942 and 0.089, MAE is 9793.85 and 63488, RMSE is 25216.8 and 100371 in LF and CF respectively. On the other hand, for DS flow data, R^2 is 0.959 and 0.0974, MAE is 6013.8 and 53967, RMSE is 19581 and 92853 in LF and CF respectively.

Keywords
Missing data; Water Resource Management; Statistical Parameters; Data driven modelling