

EGU2020-15217

<https://doi.org/10.5194/egusphere-egu2020-15217>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



Merging of satellite rainfall estimates from diverse sources with K nearest neighbour in sparsely gauged basins

Biswa Bhattacharya¹ and Junaid Ahmad²

¹IHE Delft Institute for Water Education, Integrated Water Systems and Governance, Delft, Netherlands (b.bhattacharya@un-
ihe.org)

²Punjab Irrigation Department, Lahore, Pakistan (enrjunaid07@gmail.com)

Satellite based rainfall estimates (SBRE) are used as an alternative to gauge rainfall in hydrological studies particularly for basins with data issues. However, these data products exhibit errors which cannot be always corrected by bias correction methods such as Ratio Bias Correction (RBC). Data fusion or data merging can be a potentially good approach in merging various SBREs to obtain a fused dataset, which can benefit from all the data sources and may minimise the error in rainfall estimates. Data merging methods which are commonly applied in meteorology and hydrology are: Arithmetic merging method (AMM), Inverse error squared weighting (IESW) and Error variance (EV). Among these methods EV is popular, which merges bias corrected SBREs using the minimisation of variance principle.

In this research we propose using K Nearest Neighbour (KNN) as a data merging method. KNN has a particular advantage as it does not depend upon any specific statistical model to merge data and presents a great flexibility as the value of K (the number of neighbours to be chosen) can be varied to suit the purpose (for example, choosing different K values for different seasons). In this research it is proposed to compute the distances of bias corrected SBREs of the training data from the gauge data and to assign the SBRE with the minimum distance as the class C where $C = 1, 2, 3, \dots$, number of SBREs. In validation each data point consisting of a value of each SBRE may be compared with the data points from the training set and the class of the data point(s) closest to this data point is assigned as the class of the validation data point.

The KNN approach as a data merging method was applied to the Indus basin in Pakistan. Three satellite rainfall products CMORPH, PERSIANN CDR and TRMM 3B42 with $0.25^\circ \times 0.25^\circ$ spatial and daily temporal resolution were used. Based on the climatic and physiographic features the Indus basin was divided into four zones. Rainfall products were compared at daily, weekly, fortnightly, monthly and seasonally whereas spatial scales were gauge location, zonal scales and basin scale. The RBC method was used to correct the bias. The KNN method with $K=1, 3$ and 5 was used and compared with other merging methods namely AMM, IESW and EV. The results were compared in two seasons i.e. non-wet and wet season. AMM and EV methods performed similarly whereas IESW performed poorly at zonal scales. KNN merging method outperformed all other merging methods and gave lowest error across the basin. The daily normalised root mean square error at

the Indus basin scale was reduced to 0.3, 0.45 and 0.45 respectively with KNN, AMM and EV whereas this error was 0.8, 0.65 and 0.53 respectively in CMORPH, PERSIANN CDR and TRMM datasets. The KNN merged product gave lowest error at daily scale in calibration and validation period which justifies that merging with KNN improves rainfall estimates in sparsely gauged basins.

Key words: Merging, data fusion, K nearest neighbour, KNN, error variance, Indus.