Geophysical Research Abstracts,
Vol. 11, EGU2009-4959-1, 2009
EGU General Assembly 2009

Precipitation Data Merging using a Machine Learning Based Fusion

A.C. Turlapaty (1,2), V.G Anantharaj (1), N.H Younan (1,2), and F.J. Turk (3)

(1) Geosystems Research Institute,Mississippi State University, Starkville, United States (anish@gri.msstate.edu), (2) Electrical and Computer Engineering, Mississippi State University, Starkville, United States, (3) Naval Research Laboratory, Marine Meteorology Division, Monterey, CA

Precipitation is one of the most important processes studied by climatologists and hydrologists. It plays a vital role in global water cycle and it is of critical importance in understanding the influence of climate on society. There is a need of precipitation data with high spatial and temporal resolutions on a global scale to better understand its role in global water cycle as well as hydrological applications. Though, several satellite based precipitation products are available for scientists, the spatial and temporal resolutions do not meet the requirements for many hydrological applications. The Global Precipitation Measurement (GPM) mission will estimate precipitation on a global scale with improved spatial and temporal resolution, thus, generating a precipitation product useful for both short term and long term climate and hydrometeorological studies.

In order to evaluate the value added by the GPM-era precipitation products, a Rapid Prototyping Capability (RPC) project has been sponsored by NASA. One of the objectives of this project is to develop and test a data fusion methodology to merge the satellite precipitation products available from different rainfall estimation algorithms. Data fusion is generally used in fusing information from a set of sensors with a common final goal, for example, target identification. In the past, data fusion has been used to merge satellite and ground based rainfall data. In this study, a fusion method is developed to merge precipitation data available from four different products. The final objective is to develop a product which is better than any individual product at any given spatial or temporal location. The precipitation data from Arkansas Red Basin River Forecast Center (ABRFC) region is used as reference data. The fusion method is based on binary classification of the input data. A combination of vector transfer function and a two-layer neural network is used as classifier. Initially, the input rainfall data are arranged into vectors with each four precipitation values. Then, these vectors are transformed and scaled into a new vector space using a scaled exponential transfer function. These new vectors are used as inputs to the neural network. The rainfall information from a small portion of reference data from the summer of 2007 is used as target vector in the training process. The trained neural network is used to classify the input vector data and the resulting binary classification data is multiplied with the average dataset of all individual products to produce a final merged product. In order to validate the merged product, it is compared with the reference data using statistical skill scores like Heidke skill score (HSS), critical success score, and bias score. At a given spatial location, if the rainfall time series has a better HSS compared to all other products, then it is counted as a success. Based on this criterion, the merged product has a maximum success rate of 90% during the summer, a minimum rate of 60% during the winter, and 85% during the fall and the spring seasons.