Journal of Hydrology 550 (2017) 103–117

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Assessing the applicability of TMPA-3B42V7 precipitation dataset in wavelet-support vector machine approach for suspended sediment load prediction

Sushil Kumar Himanshu [a,*], Ashish Pandey [a], Basant Yadav [b]

[a] Department of Water Resources Development and Management, Indian Institute of Technology Roorkee, India
[b] Department of Civil Engineering, Indian Institute of Technology Delhi, India

ARTICLE INFO

Article history:
Received 11 January 2017
Received in revised form 30 March 2017
Accepted 25 April 2017
Available online 27 April 2017

Keywords:
TRMM
IMD
SVM
WASVM
Suspended sediment load
Prediction

ABSTRACT

In the present study, the latest Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) research product 3B42V7 has been evaluated over gauge-based India Meteorological Department (IMD) gridded dataset employing statistical and contingency table methods for two South Indian watersheds. A comparative analysis of TMPA-3B42V7 with IMD gauge-based gridded dataset was carried out on daily, monthly, seasonal and yearly basis for 16 years (1998–2013). The study revealed that TMPA estimates performed reasonably well with the gauge-based gridded dataset, however, some significant biases were also observed. It has been observed that TMPA overestimates at very light rain, but underestimates at light, moderate, heavy and very heavy rainfall intensities. Further, the TMPA estimates was evaluated for prediction of daily suspended sediment load (SL) employing Support Vector Machine (SVM) with wavelet analysis (WASVM). Initially, 1-day ahead SL prediction was performed using best WASVM model. The results showed that 1-day predictions were very precise and shows a better agreement with the observed SL data. Later, the developed WASVM model was used for the prediction of SL for the higher leads period. The statistical analysis shows that the developed WASVM model could predict the target value successfully up to 6-days lead and found to be not suitable for higher lead specifically in the selected watersheds with similar hydro-climatic conditions like the ones selected in this study. Predictions results of WASVM model is superior to conventional SVM model and could be used as an effective forecasting tool for hydrological applications. The study suggest that the use of TMPA precipitation estimates can be a compensating approach after suitable bias correction and have potential for SL prediction in data-sparse regions.

1. Introduction

Prediction of the suspended sediment loads in rivers or streams is very crucial for sustainable water resources and environmental systems (Kisi, 2009; Yang et al., 2009). Suspended sediments are a governing factor for the design and operation of hydraulic structures, like canals, diversions and dams. The phenomenon of sediment transport and erosion in rivers and watersheds are complex environmental and hydrological problems (Mirbagheri et al., 1988). In the water resources management study, sediments in river water has a significant impact on the exploitation of surface water resources globally. Therefore, its accurate prediction could be an important index to design the future water resources management policies. Prediction of SL is very complicated and nonlinear in nature, it depends on a number of complex factors such as flow, precipitation, topography and soil characteristics of the river basin or watershed (Pandey et al., 2016). This area has been widely explored by the researchers and numerous techniques has been proposed to simulate and predict the SL. However, the complex and non-linear transportation process of SL in rivers makes these techniques over parameterized, complicated and time consuming (Nourani, 2009; Rajaee, 2011; Himanshu et al., 2017).

Moreover, accurate spatio-temporal distribution of rainfall is required for reliable hydrological predictions and is crucial for water resources applications (Kumar et al., 2014). Spatio-temporal distribution of rainfall is measured by rain-gauge networks which should be optimally located. In most of the developing countries, rain-gauge networks are usually unevenly and

* Corresponding author.
E-mail addresses: sushilkumarhimanshu@gmail.com (S.K. Himanshu), ashisfwt@gmail.com (A. Pandey), basant1488@gmail.com (B. Yadav).

http://dx.doi.org/10.1016/j.jhydrol.2017.04.051
0022-1694/

sparsely distributed, due to the complex geography, climatic variability and other limited conditions (Meng et al., 2014). The above shortcomings can be overcome using satellite-based precipitation estimates i.e., Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Sorooshian et al., 2000), Climate Prediction Centre (CPC) Morphing algorithm (CMORPH; Joyce et al., 2004), TRMM Multi-satellite Precipitation Analysis (TMPA; Huffman et al., 2007; Huffman and Bolvin, 2011), Global Precipitation Measurement (GPM; Kidd and Huffman, 2011) etc. All these products have different procedures for estimation of precipitation, although they all combine the information from infrared and microwave sensors on geostationary satellites and low earth orbiting to enhance the consistency, coverage, timeliness and accuracy of high resolution rainfall data (Behrangi et al., 2011). The TMPA precipitation estimates is one of the most widely used satellite precipitation dataset for hydro-meteorological research and applications.

The TMPA precipitation estimates developed by the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC), provide precipitation estimates at spatial resolution of 0.25-degree by 0.25-degree grid in three distinct timescales (3-hourly, daily and monthly) in two versions (near-real-time and research quality estimates). The purpose of TMPA algorithm is to produce the estimates of quasi-global precipitation which benefit from high quality microwave/infrared sensors aboard several earth observation satellites, root-mean-square precipitation-error estimates and ground-based gauge observations. The near-real-time estimates are suitable for monitoring activities, they provide very quick, less accurate dataset; whereas research quality estimates provide more precise dataset, more appropriate for research (Liu, 2015a). There are two successive versions of research quality estimate: version-6 (3B42V6) and the latest version-7 (3B42V7). The TMPA precipitation estimates have been used widely in different hydro-meteorological research and applications such as climatological studies (Kumar et al., 2014), flood prediction (Li et al., 2009), hydrological modeling (Artan et al., 2007; Stisen and Sandholt, 2010; Kumar et al., 2015), rainfall erosivity estimation (Vrieling et al., 2010), soil erosion estimation (Sanchez-Moreno et al., 2014), soil moisture prediction (Gupta et al., 2014) etc. Different studies reveals that the latest 3B42V7 algorithm has much improved accuracy upon 3B42V6 (Chen et al., 2013; Qiao et al., 2014; Yong et al., 2014; Liu, 2015b). When compared with gauge-based ground rainfall, 3B42V7 estimates revealed reduced bias across the Indian monsoon region (Prakash et al., 2015) and performed superior than other coexisting multi-satellite precipitation products (Prakash et al., 2014). Although TMPA precipitation estimates has been evaluated widely on large watersheds/river basin for stream flow simulations, limited studies has been reported for sediment load estimation. Overall, TMPA precipitation estimates can be a compensating approach for hydrological modeling in data-sparse regions.

Physical based modeling has always been the primary choice of the researchers while dealing with complex problem like sediment load prediction. However, recent trends shows that the machine learning models had been employed to model nonlinear processes which are complex in nature (Sivapragasam et al., 2008). Machine learning models perform well while simulating the complex nonlinear processes and produces accurate and efficient results. Further, these techniques also provide an additional advantage whereby the associated difficulties like slow computing, source code and coupling problems can be avoided (Yadav et al., 2016). Artificial Neural Network (ANN) technique has been applied widely to study the problem of SL prediction (Jain, 2001; Licznar and Nearing, 2003; Cobaner et al., 2009; Rajaee et al., 2009; Lafdani et al., 2013). Similarly, other techniques like genetic programming (Koza, 1992; Muttil and Lee, 2005), fuzzy logic (Klir and Yuan, 1995; Kecman, 2001) were also employed in various SL problems. Some recent techniques like Support Vector Machine (SVM; Vapnik and Lerner, 1963; Vapnik and Chervonenkis, 1964; Cristianini and Shawe-Taylor, 2000) and Extreme Learning Machine (ELM; Huang et al., 2004, 2006), have also been used effectively in hydrological modeling. However, the most widely used is the SVM, which is based on the principle of structural risk minimization of upper bound to the generalization. Previously SVM has been used extensively for different hydrological applications such as, flood stage forecasting (Liong and Sivapragasam, 2002; Yu et al., 2006), water levels prediction in lakes/reservoir (Khan and Coulibaly, 2006; Hipni et al., 2013), rating curve extension (Sivapragasam and Muttil, 2005), long term discharges prediction (Lin et al., 2006; Wang et al., 2009; Kisi and Cimen, 2011), canal's settling basins removal efficiency estimation (Singh et al., 2008), statistical downscaling (Tripathi et al., 2006; Chen et al., 2010), runoff and sediment yield simulation (Misra et al., 2009), probabilistic reservoir operation (Karamouz et al., 2009); optimal in situ bioremediation (Sudheer et al., 2013; Yadav et al., 2016), ground water levels prediction (Yoon et al., 2011) etc. As compared to other methods SVM is found to perform better (Adamowski and Chan, 2011; Sudheer et al., 2013).

Further, despite the flexibility and usefulness of machine learning methods in modeling hydrological processes, they have some drawbacks with highly non-stationary responses or seasonality (Nourani et al., 2014). On the other hand, there may not be enough data to train machine learning models alone. In this case combinations of models of different types (i.e., hybrid models) could be a solution. Research in hybrid modeling is aimed at developing algorithms to ensure optimal combinations of various machine learning models, and testing the resulting models in various situations. To handle such a situation, a technique called wavelet analysis has been used in various studies. Wavelets are mathematical articulation which decomposes the original time series into different components. By providing useful information at several levels, the wavelet components help to improve the forecasting capability of a model. Compared to traditional Fourier transforms, performance of wavelet transforms has been proved better (Adamowski and Chan, 2011). Wavelet has been used successfully to solve various hydrological problems in coupled form with ANN (Partal and Cigizoglu, 2009; Rajaee et al., 2011; Liu et al., 2013; Nourani et al., 2014). Similarly, Wavelets transform with SVM are also used extensively for different applications such as river flow forecast (Kisi and Cimen, 2011; Kalteh, 2013; Komasi and Sharghi, 2016), sediment load estimation (Haji et al., 2013; Himanshu et al., 2017), precipitation forecasting (Kisi and Cimen, 2012; Wei, 2012; Feng et al., 2015), groundwater level prediction (Adamowski and Chan, 2011; Suryanarayana et al., 2014) etc. All these studies suggest that hybrid machine learning models perform well while simulating the complex nonlinear processes and produces accurate and efficient results.

Looking to the aforementioned, the present study was carried out to evaluate the latest TMPA-3B42V7 precipitation estimates with gauge-based IMD gridded dataset. The time series of SL was decomposed into various components using wavelets analysis and the decomposed components were used as inputs to SVM model. This study aims to evaluate the 3B42V7 for prediction of daily SL employing SVM with wavelet analysis (WASVM) as they may provide an alternative source of rainfall data for ungauged or data-sparse regions. The present study was the first of its kind for application of WASVM for prediction of SL employing satellite-based precipitation estimates.