Pattern Recognition Letters 31 (2010) 1184–1200

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Precipitation data fusion using vector space transformation and artificial neural networks

Anish C. Turlapaty [a,b,*], Valentine G. Anantharaj [b], Nicolas H. Younan [a,b], F. Joseph Turk [c]

[a] Department of Electrical and Computer Engineering, Mississippi State University, MS 39762, United States
[b] Geosystems Research Institute, Mississippi State University, MS 39762, United States
[c] Naval Research Laboratory, Marine Meteorology Division, Monterey, CA 93943, United States

ARTICLE INFO

Article history:
Available online 4 January 2010

Keywords:
Pattern recognition
Optimization
Data merging
Convergence
Artificial neural networks

ABSTRACT

We have developed a new methodology to fuse several precipitation datasets, available from different estimation techniques. The method is based on artificial neural networks and vector space transformation function. The final merged product is statistically superior to any of the individual datasets over a seasonal period. The results have been tested against ground-based measurements of rainfall over a study area. This method is shown to have average success rates of 85% in the summer, 68% in the fall, 77% in the spring, and 55% in the winter.

1. Introduction

Precipitation is an important component of the global energy and water cycle; it is one of the main variables predicted in weather forecast models. Moreover, it is a key process in short-term meteorological and long-term climatological studies. Precipitation events are a driving force behind the hydrological phenomenon, such as floods and storms (Addison et al., 2002; Pidwirny, 2006). The amount of rainfall in a given location can be measured by rain gages and estimated over a given area by remote sensing techniques, both from ground-based and space-borne platforms. Rainfall estimates from land-based radars are usually limited to continental land areas and the coastal zone. Even then, there is no uniform radar coverage across the global land areas. For example, the continental United States and Western Europe have a much better coverage than the African continent. Similarly, the *in situ* measurements are also not uniformly distributed. For practical reasons, most of the *in situ* measurements are reported only on a cumulative daily basis whereas radar estimates are routinely available every hour and estimates of rain intensities even more frequently as necessary. Space-borne remote sensing platforms collectively provide nearly global coverage. Most of the low Earth orbiting (LEO) satellites have better coverage near the poles than the tropical regions, and any given geostationary (GEO) satellite observes continuously the region in view and has the capability to make measurements more frequently. Hence, the LEO and GEO satellites have different spatio-temporal sampling patterns. Besides, the LEO and GEO platforms have different kinds of instruments on-board for precipitation estimates. The rainfall estimates from geostationary satellites are typically based on infra-red (IR) measurements of cloud top temperatures whereas most of the LEO satellites make passive microwave (PMW) measurements that can be used to make more accurate estimates of rainfall. The Tropical Rainfall Measurement Mission (TRMM) has active precipitation radar (PR) on-board, capable of making high quality sensing of precipitation (Levizzani, 2008; Levizzani et al., 2007).

A number of high resolution precipitation products (HRPP) from satellite observations are routinely produced by various research and operational agencies across the world. In addition, short-term precipitation forecasts are also available from global numerical weather prediction (NWP) models. However, every one of these HRPP products has inherent advantages and limitations, and their performance varies across the seasons. Ebert et al. (2007) verified twelve sets of rainfall products (including products from numerical weather prediction models) against ground data in the United States, Australia, and Western Europe. The comparisons performed in the Australian region showed that satellite-based rainfall estimation algorithms had greater skill during the summer in tropical regions. However, models were effective during the winter in mid-latitude regions. Both types of products were not very skillful with heavy rainfall. In the continental United States (CONUS), the agreement between the combined IR-PMW rainfall and ground data varied with geographic locations; with a closest agreement in the central states. Also, the products based on PMW only data had similar characteristics as the PMW + IR combined products. However,

* Corresponding author. Address: Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762, United States. Tel.: +1 662 617 1568; fax: +1 662 325 7692.
E-mail address: anish@gri.msstate.edu (A.C. Turlapaty).

0167-8655/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2009.12.033

the rain rates derived from IR only data were severely underestimated in many locations. In the Western Europe studies, the most important finding was that the Climate Prediction Center morphing (CMORPH) algorithm's dataset outperformed all other satellite-based rainfall estimation algorithms, thus, suggesting that CMORH, which is a combination of several PMW estimates and finally merged with IR data, is an effective technique (Ebert et al., 2007).

The Global Precipitation Measurement (GPM) is a flagship mission, involving a group of international partners including the National Aeronautics and Space Administration (NASA) and the Japanese Aerospace Exploration Agency (JAXA). It consists of a constellation of satellites along with a "GPM Core" satellite (to be launched in 2013). The "GPM Core" satellite will have dual-frequency precipitation radar (PR) and a PMW radiometer called the GPM Microwave Imager (GMI). The GPM PR will be used to calibrate the measurements from the rest of the GPM constellation of satellites planned to be launched by the GPM mission partners. The goal is to obtain reliable observations of rainfall data on a global scale with high spatial and temporal resolutions (NASA GPM, 2009). These measurements will benefit researchers studying both short-term and long-term meteorological phenomena, specifically over oceans and in areas with little ground based measurements. The GPM is also conceived to be a "science mission with broad societal applications." Besides supporting weather and climate research, the high resolution precipitation products based on GPM are envisioned to benefit a number of applications including human health, disaster management, and agriculture. For instance, the GPM measurements will be useful for some of the Southeast Asian countries, where floods and storms are major issues; as they do not have required ground based measurement infrastructure to observe and respond to these contingencies in a timely manner. Since the GPM-era precipitation products will be based on measurements from a constellation of satellites, it is necessary to develop novel data fusion techniques to merge observations from satellite instruments, with different technical characteristics, capable of monitoring different physical characteristics of the precipitation process.

Satellite-based estimation of precipitation is usually not a direct measurement of rainfall, but it is based on the observation of a closely related physical entity. For instance, in the case of microwave-based observations, the scattering properties of water drops in the atmosphere are measured; which are related to the amount of rainfall if there are multiple sets of data available. Thus, the accuracy, coverage, resolution, and consistency of any single precipitation product may not be the best compared to the corresponding properties of any other product at every point in space and time. The idea of multi-sensor data fusion is to combine the information from all the available measurements and synthesize a new product, which is comparatively better than any given instance of any of the individual dataset over a period of time.

Multi-sensor data fusion has recently emerged as an established engineering discipline mainly due to research done and funded by the Department of Defense (DoD). Data fusion has been successfully used in military applications, such as target tracking, identification, and battle assessment. Moreover, fusion has been applied in non-military applications, such as remote sensing, medical diagnosis, overseeing machine building, and robotics. In simple words, a fusion process for target identification consists of the following: (a) a set of homogenous or heterogeneous sensors that observes a phenomenon generating a set of observations (Hall and Llinas, 1998); (b) a set of attributes or features that is extracted from each series of observations to develop a feature set; and (c) using a suitable classification method to extract a feature set, which, in turns, helps in identifying the target. One of the emerging fields of data fusion is in the area of remote sensing for applications such as image fusion and rainfall data merging.

Recently, fusion of precipitation data is successfully employed for improving several precipitation and related products (Chiang et al., 2007; Joyce et al., 2004; Turk et al., 2008). Nirala (2003) proposed a merging method to use rainfall datasets from multiple satellite-based sensors to improve the quality of precipitation estimation. Two of these satellite-based products are from the Advanced Microwave Sounding Unit-B (AMSU-B) and TRMM Microwave Imager (TMI). The importance of high resolution satellite-based rainfall data is recognized as most of the physical models do not perform well in some regions of the world, e.g. Indonesia (Joyce et al., 2004). Rainfall estimates from IR data have reasonable spatio-temporal coverage but have limited accuracy. The IR measurements, representing the cloud top temperatures, are fitted to a model relating it to rain rates. Hence, the actual rain rates may be different from that estimated ones by taking into consideration the relationship between the rain rate and cloud top temperatures. Generally, more accurate rainfall estimates are derived from microwave observations, which are available from PMW sensors from the LEO satellite but with limited spatial and temporal sampling. Many of the HRPP algorithms have adopted innovative techniques to merge PMW data from different satellites and, in some cases, IR data from geostationary as well. The traditional HRPP blending methods are generally based on either (a) adjustment based techniques, where instantaneous PMW data from Special Sensor Microwave Imagers (SSMI) are merged with IR data from a geostationary satellite or (b) motion based techniques, where IR data at higher spatial and temporal resolutions are used to estimate propagation vectors for microwave data. This approach does not require the IR temperature-rainfall relationship assumption (Turk et al., 2008). An example of a motion-based method is CMORPH, a morphing method developed to merge these two sets of observations using propagating vector matrices. The resulting merged rainfall product was better than products derived solely from PMW or IR data (Joyce et al., 2004). Recently, a study was done in the Wu-Tu region of Taiwan to merge satellite rainfall data with gauge observations and flash flood forecasting as the final goal (Chiang et al., 2007). The method consisted of a linear model for precipitation merging and a hydrological model for flood forecasting. The hydrological model was implemented using recurrent neural networks (RNN). Calibration of the model was done using a set of historical stream flow events. For flood forecasting applications, Chiang et al. (2007) found that the satellite data contributed only around four to five percent toward the merged precipitation. Nevertheless, high resolution precipitation products derived from satellite remote sensing have the potential for improving several other hydro-meteorological and water management applications, such as monitoring water availability.

The objective of this study is to develop an intelligent methodology for merging different observation sets. The proposed approach is different from the traditional HRPP merging methods as it does not require any assumptions in terms of cloud-actual rainfall relations. The fusion tool tries to learn the underlying patterns in the rainfall information from different HRPPs, relating them to actual rainfall over a relatively short period of time, for example, a single season in the year, and use that knowledge to merge precipitation over a longer period, for instance, the entire year. The precipitation observations available from several satellites and ground-based sensors are used to develop downscaled spatio-temporal data on a uniform grid with a spatial resolution of 0.1° by 0.1° and a time resolution of 1 h or less. The goal is to merge these different HRPP datasets into an improved product that is better than any individual dataset at any time or location. Our approach is also fundamentally different from the methodologies adopted by other HRPP algorithms; our technique uses a set of estimated HRPP rather than directly using the data obtained from the instruments onboard the satellites. In the subsequent sections, this