Remote Sensing for Environmental Monitoring and Change Detection (Proceedings of Symposium HS3007 at IUGG2007, Perugia, July 2007). IAHS Publ. 316, 2007.

Probabilistic estimation of precipitation combining geostationary and TRMM satellite data

CARLO DE MARCHI[1], ARIS GEORGAKAKOS[2] & CHRISTA PETERS-LIDARD[3]

1 *School of Natural Resources and Environment, University of Michigan, 2205 Commonwealth Blvd., Ann Arbor, Michigan 48105, USA*
demarchi@umich.edu

2 *School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, Georgia 30332, USA*

3 *Hydrological Sciences Branch, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA*

Abstract This paper presents a methodology for estimating precipitation that combines precipitation rates observed by the TRMM satellite with infrared/-visible (IR/VIS) images by geostationary satellites. The method detects IR patterns associated with convective storms and characterizes their evolution phases. Precipitation rates are estimated for each phase using IR/VIS and terrain information. The approach is shown to improve the integration of TRMM precipitation rates and IR/VIS data by differentiating major storms from smaller events and noise, and by separating the precipitation regime characteristic of each storm phase. Further, the procedure explicitly quantifies the uncertainty of the precipitation estimates by computing their probability distribution. The methodology was tested in the Lake Victoria basin during the period 1996–1998 against data from >100 raingauges, showing lower bias and better correlation with ground data than commonly used methods, and reproducing the variability of precipitation over a range of temporal and spatial scales.

Key words convective storms; kriging; Lake Victoria; Nile River; precipitation uncertainty; remote sensing

INTRODUCTION

In many world regions, remote sensing from satellites provides the only economically or physically viable system for measuring precipitation. A major advance in this sense has been the deployment of an increasing number of satellites, such as the Tropical Rainfall Measurement Mission (TRMM), able to reliably measure rain rates, although at a low temporal frequency. Procedures combining this information with the more frequent data on cloud dynamics provided by geostationary satellites have notably improved the capabilities of remote sensing of precipitation (e.g. Hsu *et al*., 1999; Adler *et al*., 2000; Joyce *et al*., 2004).

However, despite these advances remote sensing of precipitation is still affected by considerable uncertainty, even at coarse temporal and spatial scales (Adler *et al*., 2001). Yet, most remote sensing estimates of precipitation provide no information on the estimation error. If information is provided, it is normally in the form of the estimation mean square error at some spatial and temporal scale. This information, however, is not very useful to users who need to aggregate precipitation over larger areas and/or periods, or need to assess the uncertainty of the basin hydrological

Copyright © 2007 IAHS Press

response. To address these needs, we developed a methodology (ProbRain) that produces reliable estimates of precipitation and quantifies the associated uncertainty over any temporal and spatial scale of interest. ProbRain combines the precise, but infrequent, precipitation data generated by the TRMM Precipitation Radar with the infrared (IR), visible (VIS), and water vapour (WV) images continuously produced by geostationary satellites to provide precipitation estimates at a variety of temporal and spatial scales. In contrast to most other merging techniques, the combination of the TRMM and geostationary data does not produce a single "optimal" value, but a full ensemble of equally probable values that can be used to assess the uncertainty in the precipitation estimate. Further, to reduce the precipitation uncertainty, ProbRain uses Artificial Neural Networks (ANN) to recognize IR patterns associated with convective storms and their evolution phases. Precipitation rates are then estimated for each phase based on IR, VIS, WV, and terrain information.

AVAILABLE DATA

ProbRain has been evaluated in an area surrounding Lake Victoria that extends from 28ºE to 37ºE and from 5ºS to 4ºN (Fig. 1). The lake covers roughly the central 10% of this region, with mountains up to 5000 m rising east and west of it. The climate of the region is equatorial, but elevation and lake influence contribute to moderate temperatures all year round. The lowlands in the southern side of the area are considerably warmer and drier than the rest of the basin.

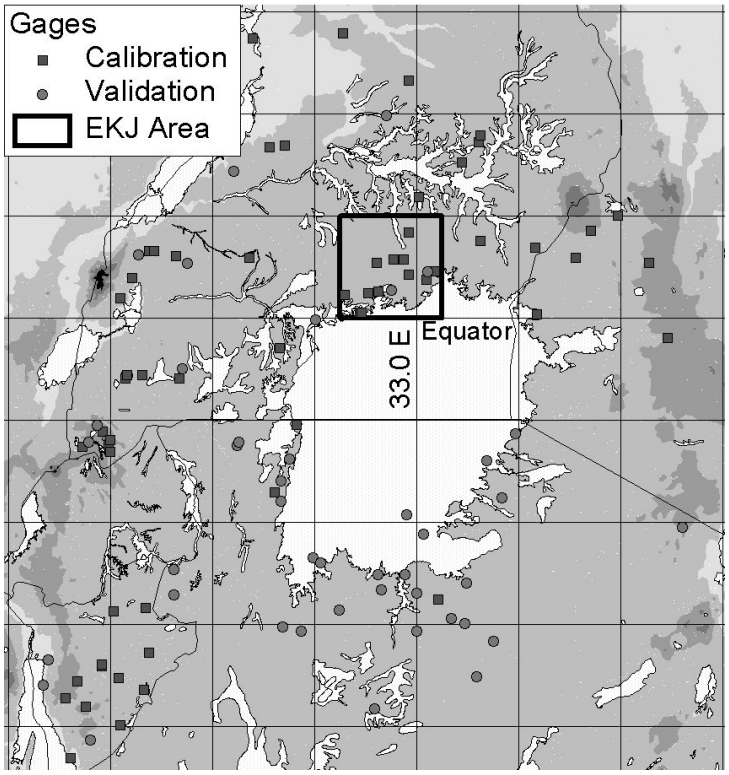

Fig. 1 Distribution of raingauges in the Lake Victoria basin used for calibration and validation.

Precipitation is driven mainly by the migration of the Inter Tropical Convergence Zone with the related northeast and southeast monsoons, but it is also heavily influenced by land–lake interactions and orography. Figure 1 shows the distribution of raingauges used for calibration and validation of the procedure (60 gauges during 1996–1998 and 42 gauges during 1996–1997, respectively).

Satellite data included Meteosat digital images in the IR, VIS, and WV bands, covering 1996–1999 at half-hour temporal resolution, and TRMM PR data covering 1998–1999. Satellite data were quality controlled and re-sampled to a common 0.05º× 0.05º (~5.5º× 5.5 km) regular grid.

METHODOLOGY

Identification of convective pixels

It is generally recognized that convective storms feature three distinct phases—developing, mature, and dissipating—each with a distinct rain regime. Using specific relations between cloud characteristics and precipitation for these different phases of the convective storm and for different types of storms should yield better precipitation estimates. ProbRain uses an ANN to identify the presence and temporal evolution of deep convective storms at the pixel level. This was achieved by training the ANN to recognize IR sequences from geostationary satellites associated with TRMM-detected convective pixels (Fig. 2(a)). The IR patterns associated with strongly convective precipitation were representative of the onset of convective storms, thereby providing a basis for differentiating the different stages of the storm (Fig. 2(b)). These are defined here as the number of time-steps elapsed from the moment the ANN recognizes the onset of the storm. Convective storm activity over a pixel terminates when the IR temperature returns above 253ºK. The best results (Validation Probability of Detection = 0.81, False Alarm Ratio = 0.04, and Area-Weighted Error Score = 0.23) were obtained by using 3-hour-long IR sequences (Fig. 2(b),(c)) and an ANN with two hidden layers, each composed of six nodes (Fig. 2(a)).

Estimation of instantaneous precipitation as a function of observed radiation

ProbRain utilizes a "lookup table" approach similar to those of King *et al*. (1995) and Kurino (1997) for relating IR/VIS/WV/Storm stage data from geostationary satellites to half-hourly precipitation rates. The look-up tables used in this research are indexed by orography, IR, VIS during daytime or IR-WV during night-time, storm stage, and month. The variable IR-$IR_{5\times5}$, where $IR_{5\times5}$ is the average IR over a 5 × 5 pixels sub-region, is used when neither VIS nor WV data are available. The look-up tables were partitioned into contiguous intervals containing at least 150 TRMM PR samples, a number deemed sufficient to generate meaningful statistics. In most techniques merging IR/VIS and microwave data, these relations are derived from data measured during the month preceding the estimate period. In contrast, we develop such relations considering contemporaneous data from a multiyear data set, thus increasing their

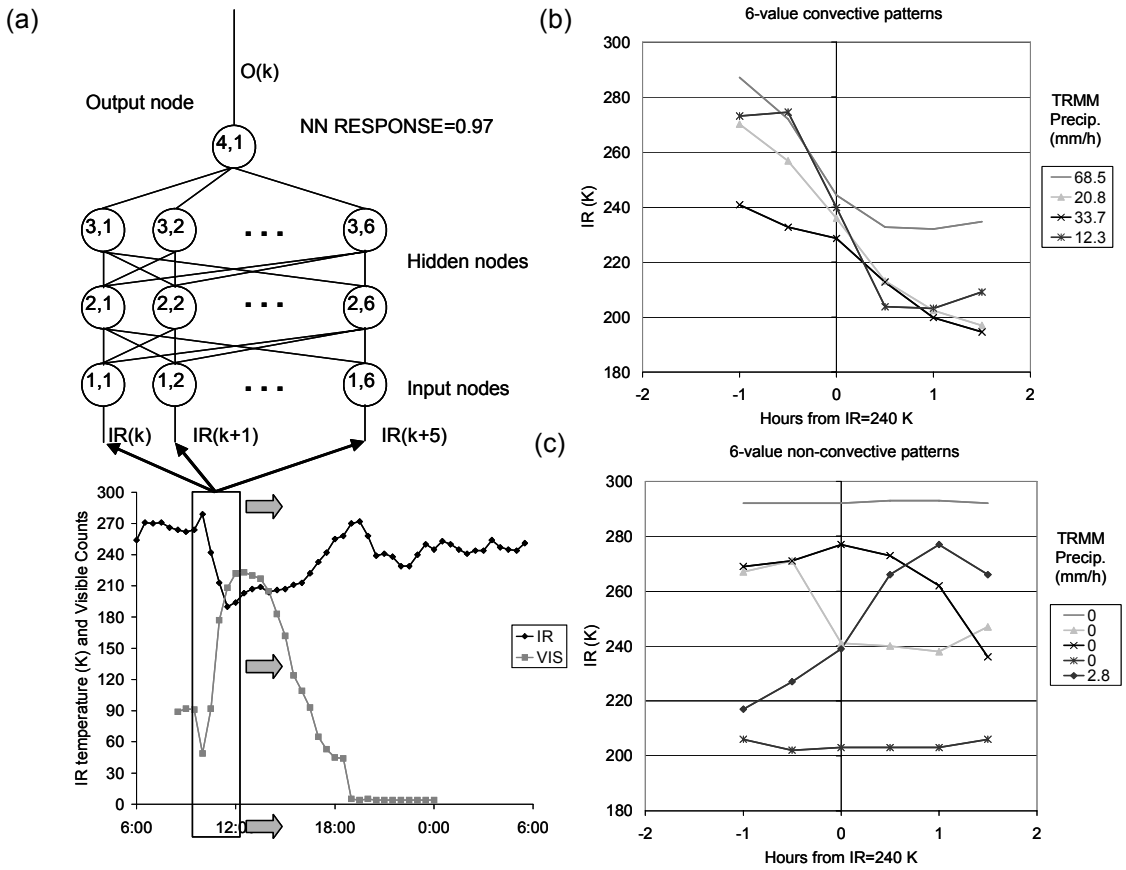

Fig. 2 (a) Schematic of Multi-layered Feed-Forward Neural Network application for convective storm identification; (b) convective patterns; (c) non-convective patterns.

resolution in the IR/VIS/Storm-stage space and their ability to represent the variability of precipitation patterns. Similarly, we developed these relations aggregating data from the entire Lake Victoria area instead of the more typical 1º × 1º or 2º × 2º resolutions. On the other hand, we partitioned rain rates according to terrain into four classes (<1000 m; 1000< h <2000 m; >2000 m; lake pixels) to better represent the steep orography differences that are typical of this region. Figure 3 shows that precipitation rates for similar IR, but different storm stage, can vary substantially, with more intense rain corresponding to the early stages (developing and mature phases). The calibration of the procedure involved the selection of the look-up table resolution and the minimum number of samples used for partitioning it. Further, precipitation rates for IR above 258ºK or for VIS below 40% albedo were set to zero.

Uncertainty characterization

As mentioned in the introduction, one of the main goals in developing ProbRain was to provide a complete characterization of the uncertainty inherent in precipitation estimates. To achieve this goal, the half-hour precipitation over a single pixel is treated not as a deterministic value, but as a random variable with distribution described by

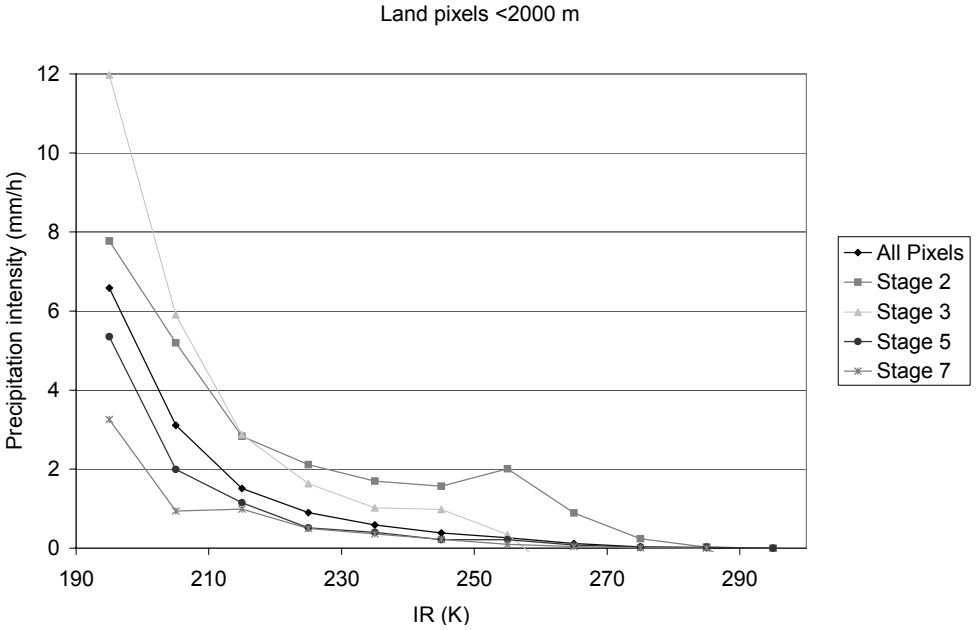

Fig. 3 Average precipitation intensity as a function of IR and storm stage.

equation (1), as in Bell (1987). Unlike Bell, in ProbRain parameters P_0, μ_{LNR}, and σ_{LNR} vary on a pixel-by-pixel basis according to the observed IR/VIS/WV and look-up tables illustrated in the previous section.

$$F_R(z) = \begin{cases} 0 & z < 0 \\ P_0 & z = 0 \\ P_0 + N(\ln(z), \mu_{LNR}, \sigma_{LNR}) * (1 - P_0) & z > 0 \end{cases} \quad (1)$$

where $F_R(z)$ is the cumulative density function of the rain rates; P_0 is the probability of no rain; and $N(ln(z), \sigma_{LNR}, \sigma_{LNR})$ is the lognormal distribution of positive rain rates.

An ensemble of several hundred precipitation realizations is generated for each pixel and each half-hour. The precipitation ensemble for longer periods is obtained by cumulating the half-hour precipitation rates pertaining to the same realization. Precipitation phenomena are characterized by strong temporal correlation that must be reproduced in order to properly account for the precipitation variability over periods longer than half an hour. Here, the precipitation distribution parameters at time t determined from the satellite-observed radiation are modified according to the same-pixel precipitation at time t–1 with the procedure described in DeMarchi (2006).

Most hydrological applications where remotely sensed precipitation would be useful to pertain to areas larger than a single 5 × 5 km pixel, requiring the aggregation of precipitation over several pixels. However, correct characterization of the statistical properties of precipitation for multiple pixels requires generating a spatially correlated random precipitation field. The challenge in integrating the remote sensing information in the generation of a spatially correlated random field is that it changes the local

unconditional precipitation distribution on a pixel-by-pixel basis. This means that, at the very least, the random field to be simulated does not have a stationary mean and possibly not even a stationary covariance.

In ProbRain the spatial correlation is modelled with a two-step procedure similar to that proposed by Barancourt *et al*. (1992) and by Pardo-Iguzquiza *et al*. (2006). In the first step, a one-threshold sequential indicator simulation with previous means discriminates rain/no-rain pixels, while in the second step the precipitation intensity is estimated using a Bayesian sequential Gaussian simulation procedure (Deutsch & Journel, 1998). The overall precipitation random field is obtained by multiplying the results of these two independent steps. As in Fiorucci *et al*. (2001), the precipitation intensity mean and variance is allowed to vary on a pixel-by-pixel basis, while the correlation coefficient is assumed to be stationary (DeMarchi, 2006). Precipitation probability and precipitation intensity variograms for the two sequential simulations were obtained using the entire 1998–1999 data set of TRMM images.

RESULTS AND DISCUSSION

Single pixel estimation

Table 1 reports the basin-wide average statistics of the comparison between satellite-based precipitation estimates and single-gauge data. Correlation is not very high for any of the examined techniques, partially because of the poor quality of many gauge data and heterogeneity of the application area (DeMarchi, 2006). In spite of this, ProbRain features higher correlation and lower bias and MAE than GPI for both the calibration and validation data sets. It also performs better than the Adjusted GPI (TRMM product 3B42, Adler *et al*., 2000), even when the latter is adapted to consider only the pixels containing raingauges. The representation of the precipitation

Table 1 ProbRain single-pixel calibration and validation results with an ensemble of 500 realizations.

	Calibration (1996–1998, 60 gauges)		Validation (1996–1997, 42 gauges)		Comparison with 3B42 (1998, 48 gauges)		
	GPI	ProbRain	GPI	ProbRain	GPI	3B42	ProbRain
Bias/Gauge	0.75	0.01	0.87	0.04	0.67	0.48	0.01
Cor_{Day}	0.38	0.44	0.41	0.47	0.42	0.41	0.45
MAE_{Day}	1.57	1.09	1.69	1.13	1.51	1.41	1.10
$C95_{Day}$	–	0.94	–	0.95	–	–	0.94
Cor_{10}	0.60	0.64	0.63	0.70	0.65	0.63	0.64
MAE_{10}	0.96	0.55	1.10	0.59	0.89	0.79	0.56
$C95_{10}$	–	0.92	–	0.89	–	–	0.92
Cor_{M}	0.71	0.73	0.75	0.78	0.75	0.72	0.73
MAE_{M}	0.84	0.39	0.98	0.43	0.77	0.64	0.41
$C95_{M}$	–	0.88	–	0.84	–	–	0.87

Cor_{Day} (Cor_{10}, Cor_{M}): correlation between the average of the estimation ensemble and gauge data for daily (10-day, monthly) precipitation; MAE_{Day} (MAE_{10}, MAE_{M}): mean absolute error between the average of the estimation ensemble and gauge data for daily (10-day, monthly) precipitation as fraction of the average gauge data; $C95_{Day}$ ($C95_{10}$, $C95_{M}$): frequency with which gauge data fall within the 95% estimate–confidence interval for daily (10-day, monthly) precipitation.

uncertainty, on the other hand, is very good at the daily and 10-day level (95% Compliance rate >0.9), while somewhat decaying at the monthly level. This is possibly due to the limited TRMM data set used to characterize the radiation–precipitation relation and temporal correlation.

Multipixel estimation

The ability of ProbRain to estimate the mean areal precipitation (MAP) and its uncertainty was tested over the 1° × 1° EKJ area on the northern shore of Lake Victoria, which features the highest density of gauges in the basin (Fig. 1). Even in this area, however, the number of gauges is insufficient for reliably assessing the MAP, especially at the daily resolution, for which the number of gauges with consistent data is just six and the precipitation spatial correlation is low (DeMarchi, 2006). Thus, we compare gauge-measured and satellite-estimated precipitation over all possible combinations of gauges in the square. For each combination of gauges, the daily/10-day/monthly gauge-derived precipitation is computed by aggregating the daily/10-day/monthly precipitation measured at each single gauge in the combination. On the other hand, the satellite-derived precipitation is composed by an ensemble of values, obtained by averaging satellite estimates belonging to the same realization of the precipitation random field for the pixels containing the gauges. Figure 4(a) shows that the satellite-gauge correlation (MAE) significantly increases (decreases) even when averaging precipitation over few pixels. The decline in the ability to represent precipitation uncertainty when aggregating precipitation over larger numbers of gauges is not actually due to the number of gauges in itself, but to the fact that more numerous gauge combinations are normally spread over larger areas (DeMarchi, 2006) and is caused by the rapid decrease with distance of the average half-hour precipitation autocorrelation (Fig. 4(b)). Considering different correlation functions for the large storms, which feature extensive areas of uniform stratiform precipitation, and for the smaller events would likely improve the representation of the precipitation uncertainty.

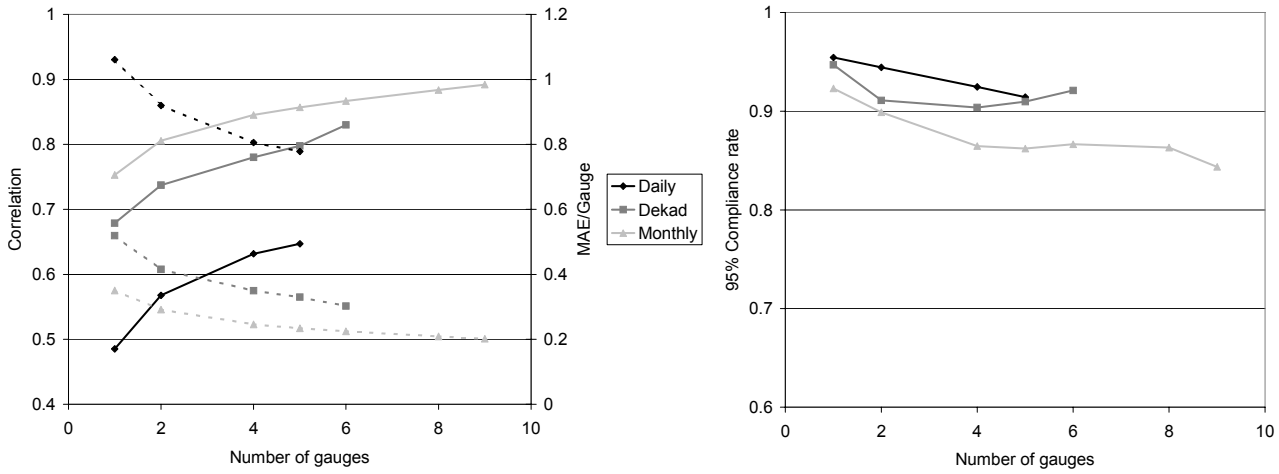

Fig. 4 ProbRain Multipixel performances for the EKJ area during 1996–1997 with 500 realizations. (a) Correlation (solid) and MAE/Gauge (dotted); (b) 95%-Compliance rate.

CONCLUSIONS

ProbRain uses multi-year data sets of contemporaneous TRMM PR and geostationary IR/VIS data for improving the detail of the radiation/precipitation relations. This approach showed better performances than GPI and AGPI in estimating precipitation and can be used for periods in which TRMM data are not available. Further, ProbRain consistently characterizes the estimation uncertainty over any area and period of interest. The extension of the multi-year data set from 1998–1999 to 1998–2006 and the use of the more numerous TRMM TMI data should further improve ProbRain performances by allowing increase of the detail of the radiation/precipitation relation. Precipitation distributions could be localized in space and time by comparing the multi-year unconditional and satellite-measured precipitation mean and variance for similar conditions and correcting for the differences using, for example, a Bayesian scheme.

Acknowledgements This research has been carried out at the Georgia Water Resources Institute, Georgia Institute of Technology, and was supported by several sponsored projects. These include the Nile River Basin Water Resources Management Program (UN-FAO; A. Georgakakos PI), Project NRA-02-ES-05 (NASA; A. Georgakakos and C. Peters-Lidard PIs), and the Water Resources Institute Program (USGS; A. Georgakakos PI). The authors are also grateful to EUMESTAT for providing a large part of Meteosat data and to Dr E. Anagnostou for kindly availing the TRMM data.

REFERENCES

Adler, R. F., Huffman, G. J., Bolvin, D. T., Curtis, S. & Nelkin, E. J (2000) Tropical rainfall distributions determined using TRMM combined with other satellite and rain gage information. *J. Appl. Meteor.* **39**, 2007–2023.

Adler, R. F., Kidd, C., Petty, G., Morrisey, M. & Goodman, M. H., (2001) Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3). *Bull. Amer. Meteor. Soc*. **82**, 1377–1396.

Barancourt, C., Creutin, J. D. & Rivoirard, J. (1992) A method for delineating and estimating rainfall fields. *Water Resour. Res.* **28**(4), 1133–1144.

Bell, T. L. (1987) A space-time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res.* **92**(D8), 9631–9643.

De Marchi, C. (2006) Probabilistic estimation of precipitation combining geostationary and TRMM satellite data. Doctoral Dissertation, Georgia Institute of Technology, Atlanta, Georgia, USA.

Deutsch, C. V. & Journel, A. G. (1998) *GSLIB: Geostatiscical Software Library and User's Guide*. Oxford University Press, New York, New York, USA.

Fiorucci, P., La Barbera, P., Lanza, L. G. & Minciardi, R. (2001) A geostatistical approach to multisensor rain field reconstruction and downscaling. *Hydrol. Earth Sys. Sci.* **5**(2), 201–213.

Hsu, K., Gupta, H. V., Gao, X. & Sorooshian, S. (1999) Estimation of physical variables from multi-channel remotely sensed imagery using a neural network: Application to rainfall estimation. *Water Resour. Res*. **35**(5), 1605–1618.

Joyce, R. J., Janowiak, J. E., Arkin, P. A. & Xie, P. (2004) CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* **5**, 487–503.

King, P. W. S., Hogg, W. D. & Arkin, P. A. (1995) The role of visible data in improving satellite rain-rate estimates. *J. Appl. Meteor.* **34**, 1608–1621.

Kurino, T. (1997) A satellite infrared technique for estimating "Deep/Shallow" precipitation. *Adv. Space Res*. **19**(3), 511–514.

Pardo-Iguzquiza, E., Grimes, D. I. F. & Teo, C. (2006) Assessing the uncertainty associated with intermittent rain fields. *Water Resour. Res*. **42**(1), 1412.