

Uncertainty characterization in a combined IR/Microwave scheme for remote sensing of precipitation

CARLO DE MARCHI¹, ARIS GEORGAKAKOS² &
CHRISTA PETERS-LIDARD³

¹ School of Natural Resources and Environment, University of Michigan, 2205 Commonwealth Blvd, Ann Arbor, Michigan 48105, USA

² School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, Georgia 30332-0355, 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 data from the precipitation radar aboard the TRMM satellite with infrared/visible (IR/VIS) images by geostationary satellites. The approach estimates half-hour precipitation based on IR/VIS data, storm stage, and terrain, and quantifies the uncertainty of the precipitation estimates by computing their full probability distribution. The probabilistic characterization is composed of a binomial distribution for the probability of rain and a lognormal distribution for the conditional rain intensity. Temporal and spatial autocorrelations are fully accounted for by using spatially optimal estimator methods (kriging). The procedure is tested in the Lake Victoria basin over the period 1996–1998 against data from more than one hundred rain gauges, showing lower bias and better correlation with ground data than commonly used methods and reproducing precipitation variability over a range of temporal and spatial scales.

Key words sequential simulation; kriging; Lake Victoria; Nile River

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). Despite these advances, however, 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's mean square error at some spatial and temporal scale. However, this information is of little help to users who need to aggregate precipitation over larger areas and/or periods, or who need to assess the uncertainty in the hydrological response of a basin. This methodology (ProbRain)

addresses these needs by generating reliable estimates of precipitation and quantifying 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 precipitation estimate uncertainty.

AVAILABLE DATA

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

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 (respectively 60 gauges during 1966–

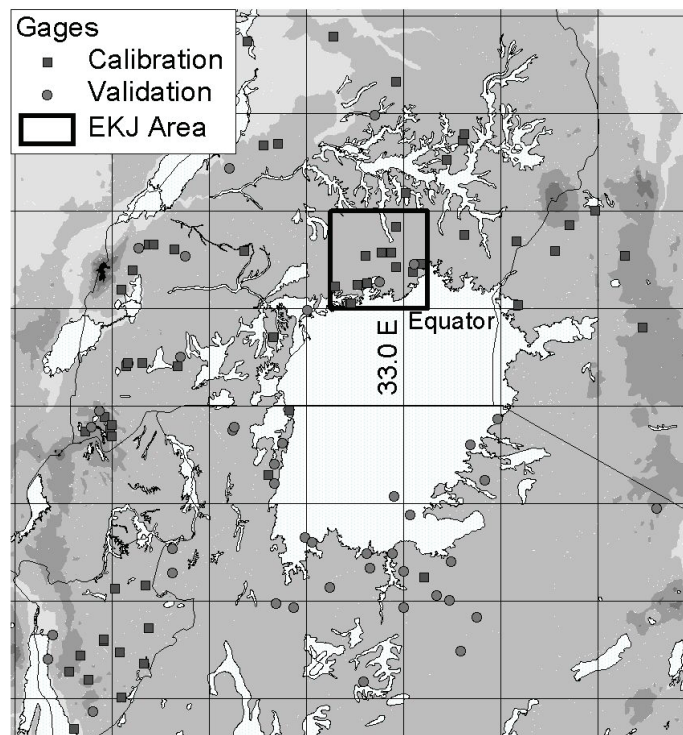


Fig. 1 Distribution of raingauges in the Lake Victoria basin used for calibration and validation.

1998, and 42 gauges during 1996–1997). The satellite data consisted of 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 $0.05^\circ \times 0.05^\circ$ ($\sim 5.5 \times 5.5$ 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. For this purpose, ProbRain uses an artificial neural network (ANN) to identify the onset of deep convective storms at the pixel level. Further, it uses the number of half-hour time steps elapsed since the storm onset as an indication of the storm stage (DeMarchi, 2006).

Estimation of instantaneous precipitation as 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 $IR-IR_{5 \times 5}$, where $IR_{5 \times 5}$ 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 are 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. Here, such relations are developed considering contemporaneous data from a multiyear data set, thus increasing their resolution in the IR/VIS/storm-stage space and their ability to represent the variability of precipitation patterns. Similarly, these relations are developed by aggregating data from the entire Lake Victoria area instead of the more typical $1^\circ \times 1^\circ$ or $2^\circ \times 2^\circ$ resolutions. On the other hand, rain rates are partitioned according to terrain into four classes (<1000 m; $1000 < h < 2000$ m; >2000 m; lake pixels) to better account for the steep orographic variations that are typical of this region. The calibration of the procedure involved the selection of the lookup table resolution and the minimum number of samples used for partitioning it. Further, precipitation rates for IR above 258°K , or for VIS below 40% albedo, were set to zero (DeMarchi, 2006).

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

An ensemble of several hundreds of precipitation realizations is generated for each pixel and each half-hour, while the precipitation ensemble for longer periods is obtained by aggregating the half-hour precipitations 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 a half-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 $t - 1$ using equation (2):

$$\begin{cases} P_0(x, t | z(x, t - 1)) = P_0(x, t) + \frac{C_I(x, t, t - 1)}{C_I(x, t - 1, t - 1)} (i_0(x, t - 1) - P_0(x, t - 1)) \\ \mu_{LNR}(x, t | z(x, t - 1)) = \mu_{LNR}(x, t) + \frac{C_{LNR}(x, t, t - 1)}{C_{LNR}(x, t - 1, t - 1)} [z(x, t - 1) - \mu_{LNR}(x, t - 1)] \\ \sigma_{LNR}^2(x, t | z(x, t - 1)) = \sigma_{LNR}^2(x, t) (1 - \rho_{LNR}^2(x, t, t - 1)) \end{cases} \quad (2)$$

where $P_0(x, t | z(x, t - 1))$ is the conditional probability of no-rain at time t given the precipitation at $t - 1$; $P_0(x, k)$ is the prior probability of no-rain at time k from satellite observations; $i_0(x, t - 1)$ is the rain indicator at $t - 1$ (1 for no rain, 0 otherwise); $C_I(x, t, t - 1)$ is the covariance between indicator residuals at time t and time $t - 1$; $C_I(x, t - 1, t - 1)$ is the indicator residuals variance at time $t - 1$; $\mu_{LNR}(x, t | z(x, t - 1))$ is the conditional mean rain intensity at time t given the precipitation at $t - 1$; $\mu_{LNR}(x, k)$ is the prior conditional mean rain intensity at time k from satellite observations; $C_{LNR}(x, t, t - 1)$ is the covariance between precipitation residuals at time t and time $t - 1$; $C_{LNR}(x, t - 1, t - 1)$ is the precipitation residuals variance at time $t - 1$; $\sigma_{LNR}^2(x, t | z(x, t - 1))$ is the conditional rain intensity variance at time t given the precipitation at $t - 1$; $\sigma_{LNR}^2(x, t)$ is the prior conditional rain intensity variance at time t from satellite observations; and $\rho_{LNR}^2(x, t, t - 1)$ is the correlation coefficient between precipitation residuals at time t and time $t - 1$.

The one-step covariances $C_I(x, t, t - 1)$ and $C_{LNR}(x, t, t - 1)$ are assumed independent of x and t , but are allowed to vary with storm stage to account for the weaker relation

during the convective stages and the stronger relation during the stratiform phases. The values of the one-step covariances were estimated from the spatial TRMM PR images considering the precipitation correlation in adjacent pixels at storm stage t and $t - 1$ as determined by the ANN (DeMarchi, 2006).

Most hydrological applications where remotely sensed precipitation would be useful 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 a Bayesian sequential Gaussian simulation procedure estimates the precipitation intensity (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 stationary (DeMarchi, 2006).

Precipitation probability and precipitation intensity variograms used in the two sequential simulations were obtained from the TRMM images in the entire 1998–1999 data set. Comparisons with single TRMM images showed that the prior distributions of no-rain probability and precipitation intensity were too uneven at the $0.05^\circ \times 0.05^\circ$ resolution to guarantee an adequate reconstruction of the precipitation spatial correlation using the kriging approach. However, smoothing the prior rain field by decreasing ProbRain spatial resolution to $0.15^\circ \times 0.15^\circ$ was not possible for the Lake Victoria study because of the difficulty of comparing precipitation estimates at this scale with single gauge data, especially at the daily resolution. Further, this resolution would not have allowed us to derive the precipitation temporal correlation from the TRMM images with the method illustrated earlier. Thus, ProbRain uses the original resolution, but emulates the $0.15^\circ \times 0.15^\circ$ resolution by applying the modifications in precipitation probability and average conditional precipitation rate determined by the kriging techniques for a pixel to the eight adjacent pixels. This operation considerably improves the simulation of the precipitation fields (DeMarchi, 2006).

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 the heterogeneity of the application region (DeMarchi, 2006). In spite of this, ProbRain features higher correlation and lower bias and MAE than GPI for the

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

	Calibration (1996–1998, 60 gauges)			Comparison with 3B42 (1998, 48 gauges)			
	GPI	ProbRain <i>t</i> -Uncorr.	ProbRain <i>t</i> -Correl.	GPI	3B42	ProbRain <i>t</i> -Uncorr.	ProbRain <i>t</i> -Correl.
Bias/Gauge	0.75	−0.01	0.01	0.67	0.48	−0.01	0.01
Cor _{Day}	0.38	0.44	0.44	0.42	0.41	0.45	0.45
MAE _{Day}	1.57	1.08	1.09	1.51	1.41	1.09	1.10
C95 _{Day}	-	0.91	0.94	-	-	0.91	0.94
Cor ₁₀	0.60	0.64	0.64	0.65	0.63	0.64	0.64
MAE ₁₀	0.96	0.54	0.55	0.89	0.79	0.56	0.56
C95 ₁₀	-	0.79	0.92	-	-	0.79	0.92
Cor _M	0.71	0.74	0.73	0.75	0.72	0.73	0.73
MAE _M	0.84	0.38	0.39	0.77	0.64	0.40	0.41
C95 _M	-	0.74	0.88	-	-	0.74	0.87

Cor_{Day} (Cor₁₀, Cor_M): correlation between the average of the estimation ensemble and gauge data for daily (10-day, monthly) precipitation; MAE_{Day} (MAE₁₀, MAE_M): mean absolute error between the average of the estimation ensemble and gauge data for daily (10-day, monthly) precipitation; C95_{Day} (C95₁₀, C95_M): frequency with which gauge data fall within the estimate 95%-confidence interval for daily (10-day, monthly) precipitation;

calibration data sets. Similar results are obtained for the validation period (DeMarchi, 2006). ProbRain performs also better than Adjusted GPI (TRMM product 3B42; Adler *et al.*, 2000), even when the latter is adapted to consider only the pixels containing rain gauges.

When the precipitation temporal correlation is not explicitly modelled (Table 1, ProbRain *t*-Uncorr.), the methodology does not properly account for the precipitation uncertainty, as shown by the fact that much less than 95% of gauge data fall within the estimation 95% confidence interval, especially at the monthly resolution. However, the procedure adopted here for representing the precipitation temporal correlation, yields good results at the daily and 10-day level (95% compliance rate ~ 0.9 – 0.95), while somewhat still decaying at the monthly level. This is likely due to the fact that the temporal correlation functions derived from the TRMM data set are just approximations of the true same-pixel one-step temporal correlation. Further, the precipitation at time *t* may depend upon previous precipitation values over areas larger than just the same pixel. Unfortunately, TRMM images do not provide enough information for the development of such extended relations. Such high temporal and spatial relations require radar or sufficient raingauge data.

Multi-pixel estimation

The ability of ProbRain to estimate the mean areal precipitation (MAP) and its uncertainty was tested over the $1^\circ \times 1^\circ$ 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 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-

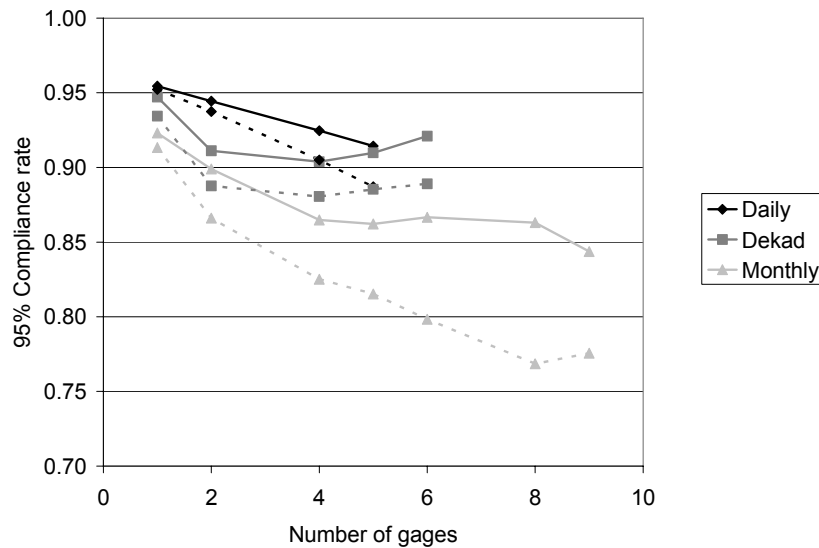


Fig. 2 ProbRain multi-pixel 95% compliance rates for the EKJ area during 1996–1997 including spatial correlation (solid lines) and neglecting it (dotted lines). Estimation ensemble of 500 realizations.

measured and satellite-estimated precipitation over all possible combinations of gauges in the square. For the combined gauges, the daily/decadal/monthly gauge-derived precipitation is computed by aggregating the daily/decadal/monthly precipitation measured at each single gauge in the combination. In contrast, the satellite-derived precipitation is composed of 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. The satellite–gauge correlation (MAE) significantly increases (decreases) from 0.49 to 0.65 (106% to 77%) at the daily level and from 0.75 to 0.89 (35% to 21%) even when averaging precipitation over less than ten pixels (DeMarchi, 2006). On the other hand, Fig. 2 shows that, if the precipitation spatial correlation is not considered, the ability to represent precipitation uncertainty severely declines when increasing the number of pixels over which precipitation is aggregated. The inclusion of the precipitation spatial correlation substantially improves the situation, although not completely. 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, but to the fact that more numerous combinations are normally spread over larger areas (DeMarchi, 2006), and it is caused by the rapid decrease with distance of the average half-hour precipitation autocorrelation. Considering different correlation functions for large storms (which feature large areas of uniform stratiform precipitation) and for smaller storms will likely improve the representation of the precipitation uncertainty.

CONCLUSIONS

ProbRain uses multi-year data sets of contemporaneous TRMM PR and geostationary IR/VIS data as the basis for estimating precipitation and quantifies the estimation

uncertainty by producing an ensemble of equally probable estimates. The ensemble average performs better than GPI and AGPI in estimating precipitation, while the ensemble distribution successfully accounts for the estimate uncertainty over a range of temporal and spatial scales. The correct assessment of precipitation uncertainty for periods longer than a half-hour and areas larger than a single pixel could not be based only on the implicit spatial and temporal correlation provided by remote sensing data, but required explicit modelling of the spatial and temporal precipitation auto-correlation. 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 improving the detail of the radiation/precipitation relation. Further, decreasing ProbRain’s spatial resolution to $0.15^\circ \times 0.15^\circ$ should improve the representation of the spatial and temporal correlation precipitation field.

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. Met.* **39**, 2007–2023.
- Adler, R. F., Kidd, C., Petty, G., Morrissey, M. & Goodman, M. H., (2001) Intercomparison of global precipitation products: the third Precipitation Intercomparison Project (PIP-3). *Bull. Am. Met. 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: Geostatistical Software Library and User’s Guide*. Oxford University Press, 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 System 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. Hydromet.* **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. Met.* **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.