

# MULTI-SCALE PRECIPITATION RETRIEVAL AND VALIDATION IN AFRICAN MONSOON SYSTEMS

F. Chopin<sup>1\*</sup>, J. C. Bergès<sup>2</sup>, M. Desbois<sup>1</sup>, I. Jobard<sup>1</sup>, T. Lebel<sup>3</sup>

<sup>1</sup> Laboratoire de Météorologie Dynamique IPSL/CNRS Ecole Polytechnique, Palaiseau, 91128 France

<sup>2</sup> PRODIG Université Paris 1, 75005 France

<sup>3</sup> IRD\*\*, LTHE, BP 53, 38041 Grenoble Cedex 9, France

## 1. INTRODUCTION

The international program AMMA (African Monsoon Multidisciplinary Analysis), which will be in intensive phase from 2005 to 2007 over West Africa, has a crucial need of precipitation measurements at scales ranging from the small basin to the regional scale, and from instantaneous values to monthly totals. This need includes estimations of the errors corresponding to each scale. In order to get relevant information in regions with very sporadic rainfall, the use of a high time sampling, which can only be provided by the geostationary satellites, is required. In a first step, as no sufficient ground validation data are available, the training of a neural network system is performed using four months of space time data between Meteosat and TRMM. Several Meteosat-derived parameters, including radiances and space-time characteristics, constitute the entries, while Precipitation Radar (PR) 2A25 rain estimates are used to train the outputs. These outputs are considered as rainfall probabilities. The second step of this research consists in multiplying these rainfall probabilities with potential rainfall intensities, calibrated by a GPCP reference dataset. Estimated rainfall intensity images are provided with geostationary satellite time and space resolution.

## 2. DATABASE

Our experimental database combines a Meteosat-7 dataset, TRMM precipitation radar observations and the 1°x 1° grid synthesis daily GPCP data (1dd). A raingauge dataset provided by the IRD\*\* (Ali et al., 2003) is used to validate our algorithm.

The considered geographical zone is West Africa during the 2000 rainfall period (June to September). The TRMM data used is the 3G68Land dataset.

## 3. NEURAL NETWORK AND RAINFALL PROBABILITIES

A neural network rainfall probability estimation has been chosen rather than direct rainfall intensity calculation because of its closer relationship with the information extracted from the geostationary satellite images. More information about rainfall estimation and rainfall classification difficulties using a feed forward neural network are detailed in Bergès and Chopin (2004)

### 3.1. Neural Network implementation

A three layer feed forward network has been selected. Assuming that rainfall probability is a continuous phenomenon it has been proved that it could be approximated by such a network (Hornik et al. 1989, Funahashi, 1989).

This neural network takes into account eight inputs which are deduced from Meteosat-7 : The Infrared (IR) temperature, its difference with the water vapour (WV) temperature, IR and WV local variance, IR and WV temporal difference, longitude, and latitude.

Output reference values are derived from TRMM 2A25 3G68Land grid data product. Rainfall presence is coded as a binary value and the outputs of this neural network are directly interpreted as rainfall probability.

It is essential to break the correlation between two neighbouring pixels due to the image spatial coherence. If not, following elements from learning dataset do not compensate their random dispersion and reinforce their common patterns : the direct consequence is to overweight the last image of the learning dataset. To correct this artefact, input dataset are scrambled, separating contiguous pixels in the learning phase.

Once weights have been estimated in order to get differences between estimated and reference values as low as possible, this network can be computed on any geostationary images dataset.

---

\* Corresponding author's address: Franck Chopin, Laboratoire de Météorologie Dynamique IPSL/CNRS, Ecole Polytechnique, 91128 Palaiseau Cedex France  
E-Mail : [chopin@lmd.polytechnique.fr](mailto:chopin@lmd.polytechnique.fr)

---

\*\* Institut de Recherche pour le Développement

### 3.2. Results and validation

When the training is accomplished, the rainfall probabilities have to be computed for every satellite slot on our interest period. Figure 1 is a resulting rainfall probability image.

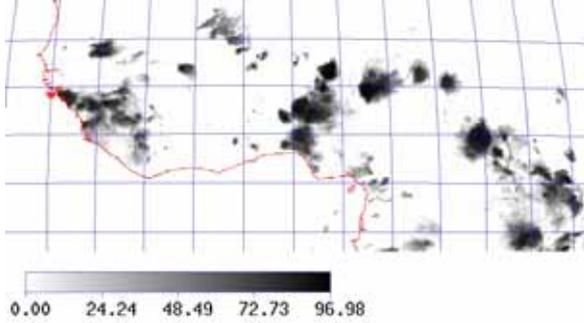


Figure 1 : Rainfall probability image ( $P_r$ ) - September 30<sup>th</sup> 2000 slot 39

To validate these results, the feed forward neural network has been compared with two other algorithms frequently used to evaluate rainfall probabilities.

The first one is an infrared temperature threshold. If a pixel infrared temperature is colder than a threshold “ $s$ ” it is considered as a rainy pixel otherwise not. The proportion of rainy cases according to TRMM 2A25 product has been estimated. To minimize the bias, the threshold value “ $s$ ” has been selected in order to fit with this proportion of cases.

The second algorithm is a probability matching method : The proportion of rainy cases has been calculated on the TRMM 2A25 dataset for each infrared temperature value. These proportions are considered as rainfall probabilities. So, the rainfall probability of a pixel “ $p$ ” depends on its infrared temperature.

The feed forward network results have been computed for this experiment considering 10 or 1000 times the learning dataset to show the bias differences related to this change.

Bias absolute values obtained with the TRMM 2A25 reference dataset (from June to September 2000) have been calculated for each day, slot and 1° in longitude and latitude. Then, means of bias absolute values on these four partition classes are evaluated. Results of this validation are presented on Figure 2.

As it can be expected, the simplest method, the temperature threshold, gives slightly poorer results than the probability matching method in any partition. Feed forward neural network results are appreciably better than those associated with the two classical algorithms. The accuracy of these neural network results is increasing with the learning iteration number. This is particularly important concerning the

latitude partition. The second neural network (1000) seems to consider the rainfall meridional gradient of the Sahelian region.

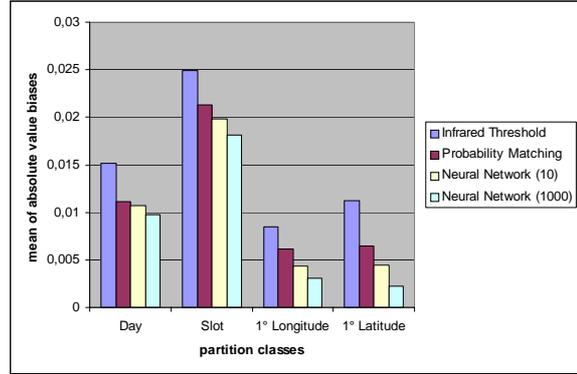


Figure 2 : Mean of absolute value biases on four partition classes for four procedures

### 4. SLIDING RESCALING ALGORITHM (SRA)

In order to produce estimated rainfall intensities, a Sliding Rescaling Algorithm (SRA) is applied. The starting point of this algorithm is a probability matching formula (1). The estimated rainfall intensity  $I_e$  is the product between the corresponding rainfall probability image  $P_r$ , assessed by the feed forward neural network described above, and a potential rainfall intensity  $I_p$ .

$$I_e = I_p \times P_r \quad (1)$$

#### 4.1. The potential rainfall intensity ( $I_p$ )

The Goes Precipitation Index (GPI) developed by Arkin and Mesner (1987) and The Adjusted GPI (AGPI) proposed by Adler et al. (1991) already use calibration coefficients to estimate rainfall intensities from infrared temperatures. While the GPI algorithm applies a constant rain rate of 3 mm/h, the AGPI is calibrated with satellite microwave data. This calibration coefficient, is based on a ratio between monthly GPI and microwave rainfall estimations and thus is space dependent. This adjustment suffers from the poor reliability of the microwave rainfall over land and from a very low repetitivity of microwave observations.

The purpose here, is to calibrate as precisely as possible our model and thus to minimize the information loss. Time dependent potential rainfall images are needed at the geostationary satellite space resolution. To obtain these images ( $I_p$ ), a calibration dataset made of cumulated rainfall estimations over the whole area of interest on short time periods is essential.

#### 4.2. The reference rainfall intensity ( $I_r$ )

To compute the potential rainfall intensity  $I_p$  images, a reference rainfall intensity  $I_r$  dataset is necessary. In this study, it has been decided to use the 1dd GPCP data. It has to be quoted that this estimator is more directly related to rainfall but does not allow a follow up of phenomena as fine as geostationary satellite images in time and space.

#### 4.3. Calculation of potential rainfall intensities

From the relation (1), we define a downscaling step in time and space equation (2). This step allows to evaluate the potential rainfall intensity in a circular window of area  $A$  for a given period  $T$ :

$$\iint_{AT} I_r(a,t) da dt = I_p(A,T) \cdot \iint_{AT} P_r(a,t) da dt \quad (2)$$

The calculation of a potential rainfall intensity  $I_p(a,T)$  is made possible for each pixel " $a$ " of the geostationary satellite image thanks to the utilisation of a sliding window of area  $A$ . It has been decided to create one  $I_p$  image for each day " $d$ " in the studied period defining the period  $T$  with the 15 days before and after " $d$ ". One  $I_p$  image corresponding to July 1<sup>st</sup> 2000 with a 25 pixels circular sliding window of area  $A$ , and a 31-day period  $T$  is represented on figure 3.

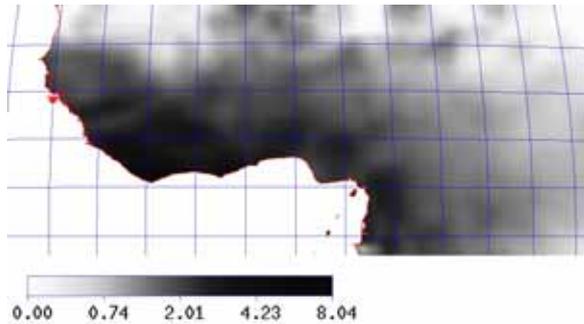


Figure 3 : Potential rainfall intensity image ( $I_p$ ) of the July 1<sup>st</sup> 2000 calculated on the period : June 16<sup>th</sup> – July 16<sup>th</sup> 2000

#### 4.4. Estimated rainfall intensity

Once the daily potential rainfall intensity  $I_p$  images are computed, the estimated rainfall intensity at time  $t$  during day  $d$  and position  $a$  can be calculated with the upscaling formula (3) deduced from the equation (1):

$$I_e(a,t) = I_p(a,d) \times P_r(a,t) \quad (3)$$

Let  $P_r(a,d)$  be the cumulative rainfall probability during day  $d$  and position  $a$ . The estimated rainfall accumulation during a period  $T$  can be easily computed with the equation (4) :

$$I_e(a,T) = \sum_{d \in T} I_p(a,d) \times P_r(a,d) \quad (4)$$

The estimated rainfall accumulation (Figure 4) is deduced from the relation (4). The advantage is that the geostationary satellite images resolution has been here preserved.

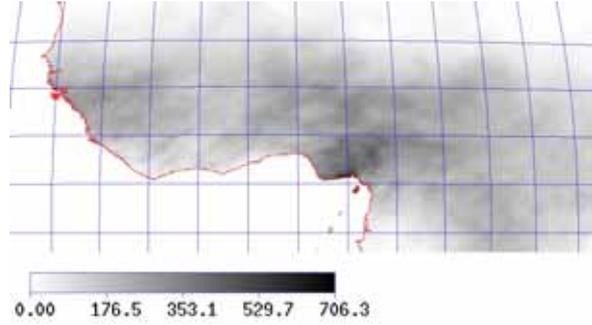


Figure 4 : Estimated rainfall intensity image ( $I_e$ ) on July 2000 (in mm)

#### 4.5. Validation

A  $1^\circ \times 1^\circ$  raingauge dataset provided by the IRD is used to compare and validate this sliding rescaling algorithm (SRA). Results of this procedure have been gridded with  $1^\circ \times 1^\circ$  cells in order to make the comparison.

The statistical parameters used to validate these results are the "bias", the "root mean square difference" (RMSD), the "normalised root mean square difference" (NRMSD), the "skill score index", and the "explained variance" ( $R^2$ ).

The monthly results of the validation are listed in Table 1.

Table 1 : Monthly validation results

Month	Method	Bias	Rmsd	nrmsd	$R^2$	Skill
July (280 cells)	GPCP	18,48	2,57	1,66	<b>0,8</b>	0,71
	SRA	<b>2,95</b>	<b>2,47</b>	<b>1,59</b>	0,75	<b>0,73</b>

Mean raingauge rainfall accumulation in July : 152,46 mm  
(Min : 24,96 mm Max : 449,01 mm)

August (280 cells)	GPCP	27,67	3,36	1,87	<b>0,76</b>	0,63
	SRA	<b>4,56</b>	<b>2,77</b>	<b>1,54</b>	<b>0,76</b>	<b>0,75</b>

Mean raingauge rainfall accumulation in August : 159,89 mm  
(Min : 15 mm Max : 406,15 mm)

July - August (1120cells)	GPCP	23,07	2,11	1,78	<b>0,77</b>	0,67
	SRA	<b>3,75</b>	<b>1,85</b>	<b>1,56</b>	0,75	<b>0,74</b>

Mean raingauge rainfall accumulation from July to August : 156,14 mm  
(Min : 15 mm Max : 449,01 mm)

The decades results of the validation are listed in Table 2.

Table 2 : Decades validation results

Decade	Method	Bias	Rmsd	Nrmsd	R <sup>2</sup>	Skill
<b>1<sup>st</sup> decade of July</b> (280 cells)	GPCP	20,52	1,93	3,29	0,7	-1,23
	SRA	<b>8,11</b>	<b>1,04</b>	<b>1,77</b>	<b>0,75</b>	<b>0,35</b>
Mean raingauge rainfall accumulation : 33,14 mm (Min : 0 mm Max : 107,56 mm)						
<b>2<sup>nd</sup> decade of July</b> (280 cells)	GPCP	4,11	1,17	1,43	0,65	0,44
	SRA	<b>1,48</b>	<b>0,87</b>	<b>1,06</b>	<b>0,75</b>	<b>0,69</b>
Mean raingauge rainfall accumulation : 50,48 mm (Min : 1,68 mm Max : 105,66 mm)						
<b>3<sup>rd</sup> decade of July</b> (280 cells)	GPCP	<b>-6,6</b>	1,6	1,59	0,5	0,39
	SRA	<b>-7,1</b>	<b>1,45</b>	<b>1,45</b>	<b>0,55</b>	<b>0,5</b>
Mean raingauge rainfall accumulation : 69,29 mm (Min : 14,51 mm Max : 203,35 mm)						
<b>1<sup>st</sup> decade of Aug</b> (280 cells)	GPCP	4,05	1,34	1,55	0,67	0,62
	SRA	<b>-2,35</b>	<b>1,24</b>	<b>1,44</b>	<b>0,68</b>	<b>0,68</b>
Mean raingauge rainfall accumulation : 55,95 mm (Min : 0 mm Max : 176,04 mm)						
<b>2<sup>nd</sup> decade of Aug</b> (280 cells)	GPCP	4,53	1,53	1,77	0,5	0,31
	SRA	<b>-0,32</b>	<b>1,3</b>	<b>1,51</b>	<b>0,57</b>	<b>0,49</b>
Mean raingauge rainfall accumulation : 57,36 mm (Min : 8,99 mm Max : 132,49 mm)						
<b>3<sup>rd</sup> decade of Aug</b> (280 cells)	GPCP	20,37	1,94	2,52	0,79	-0,31
	SRA	<b>8,51</b>	<b>1,08</b>	<b>1,41</b>	<b>0,8</b>	<b>0,59</b>
Mean raingauge rainfall accumulation : 45,29 mm (Min : 3,52 mm Max : 111,22 mm)						
<b>all decades</b> (2240 cells)	GPCP	7,81	0,53	1,94	0,6	0,39
	SRA	<b>2,75</b>	<b>0,41</b>	<b>1,51</b>	<b>0,67</b>	<b>0,63</b>
Mean raingauge rainfall accumulation : 45,25 mm (Min : 0 mm Max : 203,35 mm)						

As it can be seen, the SRA results are better than those deduced from the GPCP for four of the five statistical parameters concerning the monthly validation results and for all of these concerning the decades validation results. The explained variance (R<sup>2</sup>) is the only parameter giving superior results with the GPCP at monthly scales only. However, biases results between rainfall estimations of SRA and raingauge dataset are quite low.

The rescaling algorithm seems to be more efficient than the GPCP on this validation dataset.

## 5. CONCLUSION

A two steps algorithm based on the utilisation of rainfall probability images obtained thanks to TRMM precipitation radar and geographical potential rainfall intensity provides rainfall estimations with the geostationary satellite image resolution in time and space. This method has been validated for the monthly rainfall accumulation and seems to be better than GPCP on our dataset in particular in term of bias. Several time scales will be studied in the future from the decade to the instantaneous values of intensity. The proposed algorithm can be directly applied to Meteosat-8 : The 12 channels provided by this satellite will allow to improve the rainfall probability estimation and reinforce the reliability of our results.

## References

- Adler, R. F., Negri, J. M., Hakkarinen, I. M., 1991 : Rain estimation from combining geosynchronous infrared and low-orbit microwave data. *Paleogeogr. Paleoclimatol. Paleoecol.*, **90**, 87-92.
- Arkin, P. A., Meisner, B., 1987 : The relationship between large scale convective rainfall and cold cloud over the Western Hemisphere during 1982-1984. *Mon. Weather Rev.*, **115** (1), 51-74.
- Bergès, J.C., Chopin, F., 2004 : Satellite rainfall estimation using a feed forward neural network : improvements and limitations, *14<sup>th</sup> International Conference on Clouds and Precipitation*, **1**, 273-276.
- Funahashi, K. I., 1989 : On the approximate realisation of continuous mapping by neural networks, *Neural Networks*, **2**, 182-192.
- Hornik, K., Stinchcombe, M., White, H., 1989 : Multilayer feedforward networks are universal approximators, *Neural Networks*, **2**, 359-366.
- Ali, A., Lebel, T., Amani, A., 2003 : Invariance in the spatial structure of Sahelian Rain Fields at Climatologic scales, *Journal of Hydrometeorology* **4** (6), 996-1011.