

A new methodology for rain identification from passive microwave data in the Tropics using neural networks

Sahra Kacimi,^{a*}Nicolas Viltard^b and Pierre-Emmanuel Kirstetter^c

^a*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA*

^b*Université Versailles, Saint Quentin en Yvelines, CNRS/INSU, LATMOS-IPSL, Guyancourt, France*

^c*NOAA/National Severe Storms Laboratory, Norman, OK, USA*

*Correspondence to: S. Kacimi, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, MS 300-243 Pasadena, CA 91109, USA. E-mail: sahra.kacimi@jpl.nasa.gov

The detection of rainfall remains a challenge for the monitoring of precipitation from space. A methodology is presented to identify rain events from spaceborne passive microwave data using neural networks. We focus on BRAIN, the algorithm that provides instantaneous quantitative precipitation estimates at the surface, based on the MADRAS radiometer onboard the Megha-Tropiques satellite. A version of BRAIN using data from the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) has been used to compare several multilayer perceptrons (MLP) trained on different combinations of TMI brightness temperatures with the conventional GSCAT-2 algorithm approach used for rainfall detection. These classifiers were compared at a global scale to reference values from the TRMM Precipitation Radar (PR). They were also compared to ground measurements using two $1^\circ \times 1^\circ$ dense rain-gauge networks from different climatic zones in West Africa to assess the influence of rainfall types. At the global scale the MLPs provide better Probability of Detection than the GSCAT-2 decision tree but tend to have a higher False Alarm Rate. While no unique solution exists given the strong regional dependence of the classifiers' performances, the screen based on the 19, 21 and 85 GHz channels provides the best detection results at the instantaneous scales. As to accumulated rainfall, the screen that exhibits the lower bias relative to the PR makes use of the 37 and 85 GHz channels. The evaluation over West Africa using 10 years of TRMM overpasses shows that MLPs are in better agreement with both the PR and the gauges than GSCAT-2. The MLP trained on the 37 and 85 GHz channels increases the Probability of Detection by nearly 35% compared to the former screening over the two studied regions. Better results are obtained in the case of organized systems. Copyright © 2013 Royal Meteorological Society

Key Words: passive microwave; neural networks; Probability of Detection; bias; Bayesian inversion; rainfall estimation

Received 13 February 2012; Revised 20 December 2012; Accepted 3 January 2013; Published online in Wiley Online Library 21 March 2013

Citation: Kacimi S, Viltard N, Kirstetter P-E. 2013. A new methodology for rain identification from passive microwave data in the Tropics using neural networks. *Q. J. R. Meteorol. Soc.* **139**: 912–922. DOI:10.1002/qj.2114

1. Introduction

From the launch of the first Special Sensor Microwave/Imager (SSM/I) on the Defense Meteorological Satellite Program (DMSP F8), to the launch of the Tropical Rainfall Measuring Mission (TRMM) in 1997, the algorithms to retrieve rainfall from the measurements of passive microwave radiometers (PMR) on low Earth orbit have improved greatly (Stephens and Kummerow, 2007). The retrieval of a rain intensity from PMR data usually entails an explicit or implicit preliminary detection step, in which the algorithm identifies possible rain pixels.

Within the framework of the Megha-Tropiques (MT) mission that carries a conically scanning radiometer MADRAS (Microwave Analysis and Detection of Rain and Atmospheric Structures) with operating frequencies in the range 18.7–157 GHz, the retrieval of instantaneous rainfall will be performed by a Bayesian algorithm known as BRAIN (Bayesian Retrieval Algorithm Including Neural Networks) (Viltard *et al.*, 2006). BRAIN is a Bayes/Monte-Carlo based algorithm that retrieves the rain intensity at the surface from a vector of measured brightness temperatures (TB). To do so, the algorithm relies on a retrieval database, which contains near-surface precipitating rates with their associated TB vector. The construction of the retrieval database is described in Viltard *et al.* (2006). Once the retrieval database is built, BRAIN estimates the most probable rain rate (expected rain rate) corresponding to a measured TB vector, by computing a weighted average of the database elements (e.g. L'Ecuyer and Stephens, 2002). The weights are computed based on the distance between the input vector and each vector of the database in the TB space.

Rain retrieval in BRAIN does have an explicit preliminary step in which the no-rain pixels are identified and are therefore not processed. Identifying the presence of rain from PMR data is recognized as a crucial step for rain retrieval accuracy. Kirstetter *et al.* (2012) shows that about 30% of the total rain could be missed in BRAIN because of misclassified pixels. The screening is also meant to eliminate all ambiguous pixels from the retrieval with backgrounds similar to rain signatures but that are not actually rain falling at the surface (e.g. snow-covered ground). This filtering process will therefore be critical for instantaneous retrievals but even more so when developing accumulated rain products.

The difficulty in detecting rain over land surfaces is due to the high variability of the surface emissivity (ϵ) within the field of view (FOV) of the radiometers, that compromises the discrimination of the atmospheric contribution from the background in the radiometric signal. Since typical values of ϵ over land range from 0.6 to 0.95, the thermal emission by clouds leads to nearly the same brightness temperature as the surface, so that the signal from liquid rain is not easily identifiable. For this reason, most algorithms rely exclusively on the indirect relationship between surface precipitation and ice scattering in the upper part of clouds. For rain retrieval over land, BRAIN only uses the 37 GHz-vertical and both polarizations of the 85 GHz.

Over the last two decades, several screening methods have been proposed, among which the most widely known was proposed by Grody (1991) and is applicable over both land and ocean surfaces. Most of those methodologies were developed for SSM/I data and provide a discrimination of scattering surfaces from the scattering due to precipitation.

Usually, this step is followed by multiple tests on TB values to eliminate ambiguous situations, such as snow cover and deserts. However, none of these studies provides a quantitative evaluation of the impact of the screening method on rainfall estimation accuracy. The novelty of this article can be found in the quantitative analysis of the effects of the use of a new neural screen on the resulting average surface rain rates. This new rainfall identification method using artificial neural networks is implemented in the BRAIN algorithm, allowing us to test its impact on an already operational algorithm.

The first part of the article consists of a brief review of screening methodologies, including the one used in BRAIN described in section 3. The methodology used for the new detection is presented in section 4. In section 5, the results obtained with the new neural screen are evaluated on both a global and a regional scale. BRAIN performances are evaluated first against the TRMM Precipitation Radar (PR) and secondly against two sets of rain-gauge data. The comparison with the PR enables a quasi-global evaluation of the algorithm outputs, whereas the comparison with the rain-gauges provides a reliable evaluation over two different climatic regions in West Africa.

2. Background

As with rainfall retrieval algorithms, most screening methodologies can be grouped in two categories, either physical or statistical. The first category is called 'physical' in the sense that the possible rain is identified from physical characteristics of TB values (Grody, 1991; Hollinger, 1991; Adler *et al.*, 1994; Kniveton *et al.*, 1994; Petty, 1994; Ferraro *et al.*, 1998, 2000; Bennartz *et al.*, 2002; Greenwald and Christopher, 2002). Most of the physically based techniques were initially developed to avoid misinterpretation of cold surfaces as precipitation, an error which would lead to systematic bias in accumulated products. Ferraro *et al.* (1998) labelled these surfaces as 'scatterers' because their TB decreases with increasing frequency. As such, deserts, snow cover, semi-arid land or refrozen snow areas need to be properly identified before the retrieval process can take place. Surface types can also be classified from the difference between the Horizontal (H) and Vertical (V) polarization at a given frequency. Indeed, contrary to precipitation, most land surfaces exhibit a greater polarization difference at lower frequencies such as 19 GHz than at 85 GHz. Information on scattering and polarization is usually combined to discriminate precipitation over land surfaces leading to the computation of the so-called scattering index as demonstrated by Grody (1991). Clear-sky regions are identified when this index exceeds 5 to 10 K. This scattering index is usually equal to the difference between the estimated TB (85V) without the scattering effects of precipitation (estimated from a regression derived from global clear-sky observations) and the observed TB (85V). In the case of the TRMM 2A12 version 7 (Gopalan *et al.*, 2010), the scatterers are identified through a simple test: $(TB(22V)-TB(85V)) > 8$. These kinds of thresholds are also used in the Goddard Scattering (GSCAT-2) algorithm (Adler *et al.*, 1994) that was part of the former version of the GPROF algorithm version 4 (Kummerow *et al.*, 1996). It is also important to note that depending on the screening version, the horizontal polarization can be used instead of the vertical one. This comes from the use of SSM/I F08 data

which, starting in January 1989, have an unusable 85 GHz vertical polarization channel because of excess instrument noise.

The second category of methods uses a more statistically based approach. One way to implement this is through the use of artificial neural networks. In the remote sensing of precipitation, a number of satellite rainfall algorithms are based on neural tools (Zhang and Scofield, 1994; Hsu *et al.*, 1997; Mallet *et al.*, 2002; Tapiador *et al.*, 2004b). A review of neural networks applications in satellite rainfall estimation can be found in Tapiador *et al.* (2004a). In spite of the relative common use of neural networks for modelling inverse problems in rainfall estimation, their use for classification issues and especially the identification of possible rain from satellite data is less frequent. Grecu and Anagnostou (2001) and Moreau *et al.* (2002) proved the utility of neural networks for this particular purpose. As can be seen in the literature, the most common architecture of neural network used in the field of remote sensing is the multilayer perceptron. In this study, a multilayer perceptron is also used, but a radically different approach is presented. Indeed, we assume that the TBs without further modification or conversion are sufficient to enable rain identification. Furthermore, in the framework of instantaneous rainfall estimation, attention is given to the possible impact of the screen misclassification on the retrieved precipitation rate.

3. BRAIN screen

BRAIN is an instantaneous rainfall retrieval algorithm with currently running versions for various radiometers such as MADRAS, TRMM Microwave Imager (TMI), SSM/I, SSMI/S (Sounder) and the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E). Since rain retrieval from brightness temperatures is a very ill-posed problem, the algorithm is based on a simplified Bayesian scheme. The ensemble of possible solutions is stored in a retrieval database where each element consists of a vector of brightness temperatures with its corresponding near-surface rain rate. The retrieval is simply the computation of the weighted average of the elements found in the database. The weights are computed as a function of the distance between the measured vector of TB and each of the elements of the database. Each weight also takes into account the respective uncertainties of the database and of the measured vector itself (a thorough description can be found in Viltard *et al.* (2006), L'Ecuyer and Stephens (2002) and Kummerow *et al.* (2001)). In the retrieval process, each pixel of the radiometer is treated independently from the others. Before the retrieval itself, each pixel goes through a two-step process. First, its surface type is determined as ocean, land or coast. Second, its rainy or non-rainy quality is defined since rain occurrence is comparatively rare. All the pixels identified by the screen as possibly rainy are further processed to provide a surface rain rate estimate, while the others are set to 0 mm h^{-1} . The current screening methodology used in BRAIN is adapted from the one proposed by Adler *et al.* (1994) (GSCAT: Goddard SCATtering algorithm) that was improved in the GSCAT version 2. GSCAT was created using SSM/I data and is suitable for other radiometers, provided that the differences in frequency and geometry are not so large as to affect significantly the brightness temperature ranges. Studies further show that better performance is

achieved with the TMI measurements because of the reduced inhomogeneity encountered in the smaller pixels of the TRMM radiometer (Kummerow *et al.*, 2001). The core of GSCAT is implemented as a decision tree. An overview of the derived procedure used in BRAIN is given in Figure 1. The first threshold on TB 85H used here to filter no-rain pixels is set to 257 K, like in GSCAT. Unlike the screen developed by Grody (1991), here some pixels can be labelled as 'indeterminate'. Some weaknesses of this feature have been noted by Kummerow *et al.* (2001), showing that this procedure has the tendency to flag too many pixels as 'indeterminate' in the vicinity of rain areas.

The motivation to develop a new screen was raised by the study by Kirstetter *et al.* (2012) dealing with the validation and the error estimation of BRAIN estimates. A rainfall comparison over West Africa between rain-gauges, PR and BRAIN data was performed using 10 years of data. Results showed that BRAIN in its current version misses about 25% of rain in volume and 55% of rain occurrences. Although there is no direct evidence that the screen is entirely responsible for these discrepancies, it leaves no doubt that the BRAIN detection methodology can be improved, with a corresponding improvement in these scores.

4. Data characteristics and methodology

In this study, data from the TRMM mission were used to illustrate the potential of neural network techniques for identifying rain situations. Pixels can exist under three states: rainy, cloudy and clear. When no cloud is present within the FOV, the pixels are considered as clear. Cloudy pixels will have a certain amount of clouds, either warm or cold, but no rain on the ground, or at any level of the profile. Rain pixels are found when at least one level has precipitating particles. With this classification in mind, it is obvious that the so-called no-rain pixels can be either clear or cloudy. The reference for identifying rain and no-rain pixels here is the PR, which has a minimum theoretical detection threshold of approximately 17 dBZ or 0.7 mm h^{-1} . The PR profiles are processed so that each of them can be used in a radiative transfer model as in Viltard *et al.* (2000). The simulated TBs for each individual PR profile are convolved with the antenna pattern of each channel in order to provide TB at the proper frequency and spatial resolution.

An artificial neural network (NN) is trained to reproduce the behaviour of samples that make up the training database. The quality of those samples is therefore essential for the performance of the NN. Since BRAIN is meant to be used with the whole range of PMR that are flying presently or to be flown soon, the retrieval database is made of simulated brightness temperatures where the specifics of each instrument are accounted for. In order to develop such a screen for the different PMR, the learning phase is performed on the same dataset that was used to build the retrieval database, augmented by adding clear and cloudy cases. The simulation of TB is known to be affected by substantial errors due to incomplete or improper description of both the atmosphere and the surface. These errors can lead to systematic misclassification in a screening application. To ensure the best performance of the network, a calibration procedure is applied. This procedure is based on the approach proposed by Aires *et al.* (2010) and uses a multilayer perceptron (MLP), which in this case contains nine neurons in the input layer (i.e. the nine measured TBs

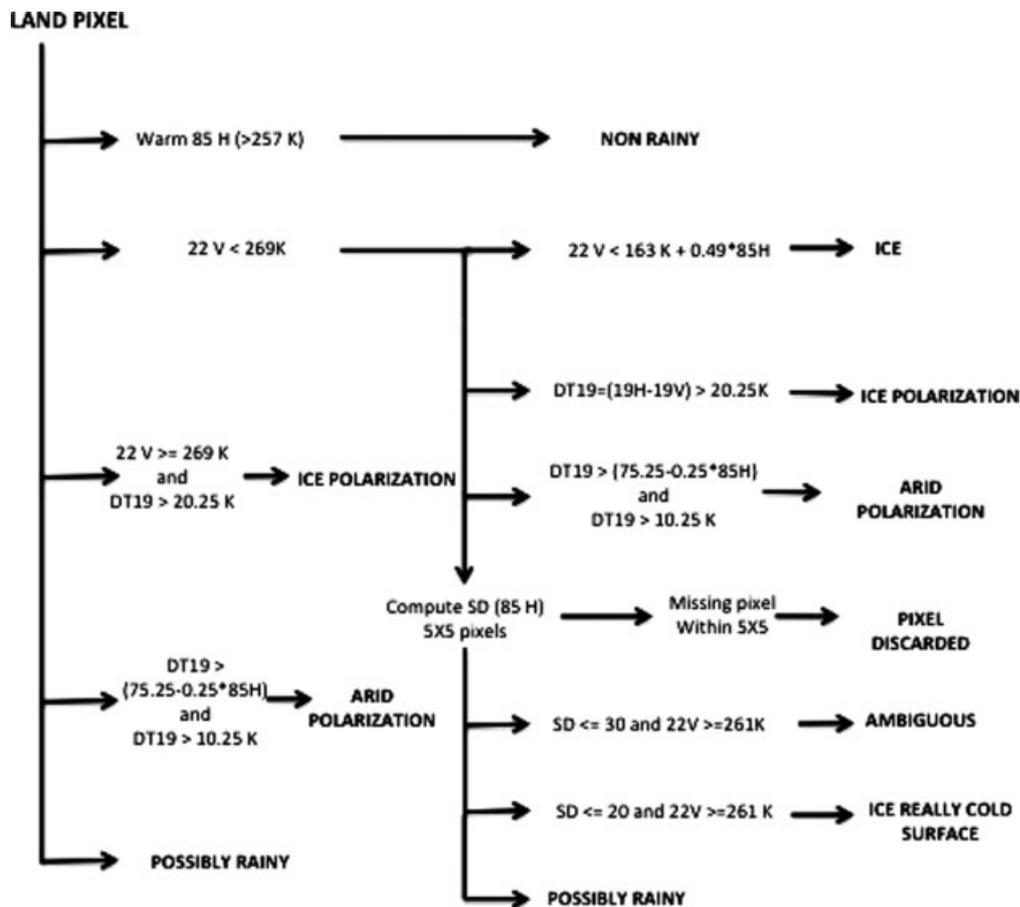


Figure 1. Current screening methodology over land surfaces for BRAIN algorithm.

from the TMI), ten neurons in the hidden layer and nine neurons in the output layer (i.e. the nine calibrated TBs). To maximize cross-platform consistency, the same calibration procedure will eventually be used for all radiometers.

The screening database is built from 10 years of randomly selected radiances from the nine TMI channels covering the whole tropical belt, with the rain/no-rain flag from the PR (Iguchi *et al.*, 2000). The PR pixels are averaged to the spatial resolution of the 37 GHz pixels of TMI (see Viltard *et al.* (2006) for more details). The pixels belong to the rain class in the sense of PR: there is signal at least at one level of the profile, which means at least 17 dBZ of backscattered power or about 0.7 mm h^{-1} of precipitation. This database contains about 60 000 points with 30 000 points for each class (rain/no-rain) to avoid favouring the naturally more-frequent no-rain cases within the screening procedure.

Figure 2 illustrates the TBs used in training the screen. Each channel is shown as a histogram of its corresponding brightness temperatures, either for rain or no-rain pixels of the database. As can be seen, none of the TB vectors from the dataset provides accurate separation between rain and no-rain pixels. Therefore, several associations of TB as network inputs were tested, resulting in six different combinations.

The screening MLP architecture consists of three successive layers. The number of units in the first layer depends on the variables retained for training. For classification, several configurations exist for the output layer. This layer contains only one neuron, and the class separation line is delimited by applying a threshold on the output. The optimal number of hidden neurons is obtained

using a heuristic approach. The selection of both the NN architecture and its parameters remains crucial. The risk of using a non-optimized network is that over-training problems and spurious parametrization can occur. To avoid those, the learning phase is followed by a test on a so-called training dataset and the robustness of the NN is further evaluated over a validation dataset. For all configurations of the input layer (i.e. number of TBs retained for learning), the optimal architecture is achieved when using 15 neurons in the hidden layer.

The network outputs are evaluated in different ways. First, the discrimination quality of a given MLP is captured through Receiver Operating Characteristics (ROC) curves and the calculation of the area under these curves (AUC). These tests are usually used to identify the best classifier (Bradley, 1997). Second, ROC curves represent an efficient way of visualizing the performance of a classifier in order to set a decision threshold. Indeed, the MLP provides a continuous output ranging from 0 to 1 that can be considered as a probability. The extreme values 0 and 1 correspond respectively to no-rain and rain labels. The ROC curve analysis enables the computation of a threshold that does not favour either of the two classes.

Once the optimal output threshold is determined, the percentage of agreement between the rain label from the NN (NNflag) and the PR (PRflag) is computed. Hence, pixel by pixel, the labels of the pixels provided by the NN and the PR are compared and gathered in four categories that are determined by the confusion matrix presented in Table 1. Those four categories are the cases of Hits, False Alarms (FA), Wrong Detections (WD) and Good Rejections

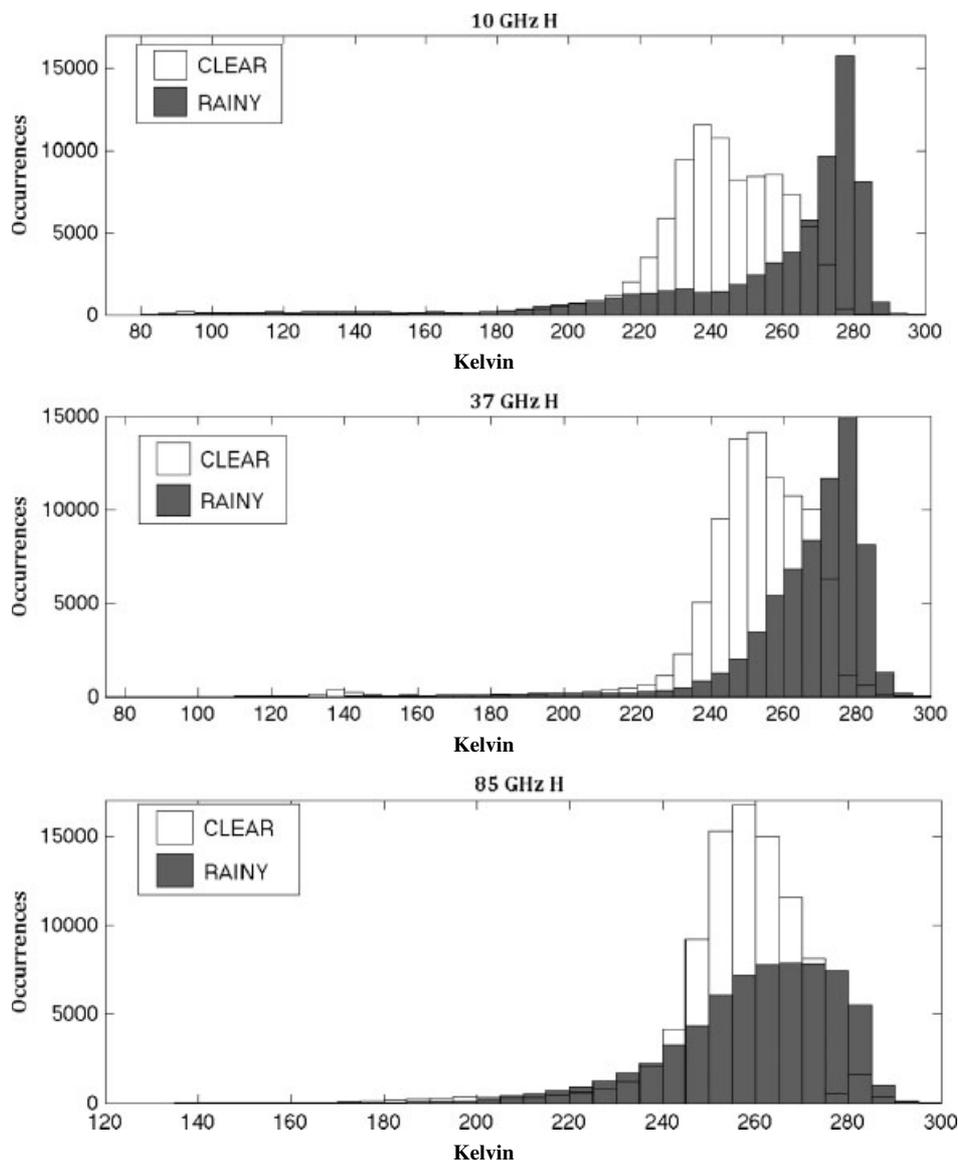


Figure 2. Histograms of TB (training database) for rainy/non-rainy discrimination.

Table 1. Confusion matrix and categories of classified pixel. Rain=R, No-rain=NR.

NNflag	PRflag	
	R	NR
R	Hits	False Alarms (FA)
NR	Wrong Detections (WD)	Good Rejections (GR)

(GR). A case of Hits occurs when both NN and PR identify a pixel as rainy, whereas a GR occurs when both of them identify a pixel as non-rainy. An FA corresponds to cases where the NN identifies a pixel as rainy contrary to the PR. A WD corresponds to the opposite situation. From the confusion matrix, several performance metrics can be calculated, particularly the sensitivity and specificity. In the literature, those indices are commonly mentioned as Hit rate and False Alarm rate (which will be explained in detail in the following section). Their computation is necessary for creating ROC curves. Indeed, each point of an ROC curve represents a sensitivity/specificity pair corresponding to a

particular decision threshold (i.e. applied to the output of the considered classifier, here NN).

In the specific issue of rainfall discrimination, Sensitivity (*Se*) and Specificity (*Sp*) respectively represent the ability of the network to detect rain pixels from no-rain ones and vice versa. *Se* and *Sp* are expressed as:

$$Se = \frac{Hits}{Hits + FA}$$

$$Sp = \frac{GR}{GR + WD}$$

Then sensitivity against 1-specificity is plotted on Figure 3. A classifier providing perfect discrimination (i.e. no overlap in the two classes) has an ROC plot that passes through the upper left corner where both sensitivity and specificity equal 1. The theoretical plot for a test with no discrimination is a 45° diagonal line from the lower left corner to the upper right corner. Qualitatively, the closer the plot is to the upper left corner, the higher the overall accuracy of the classifier. In Figure 3 the ROC curves obtained for six different classifiers are displayed. It can be noted that, with

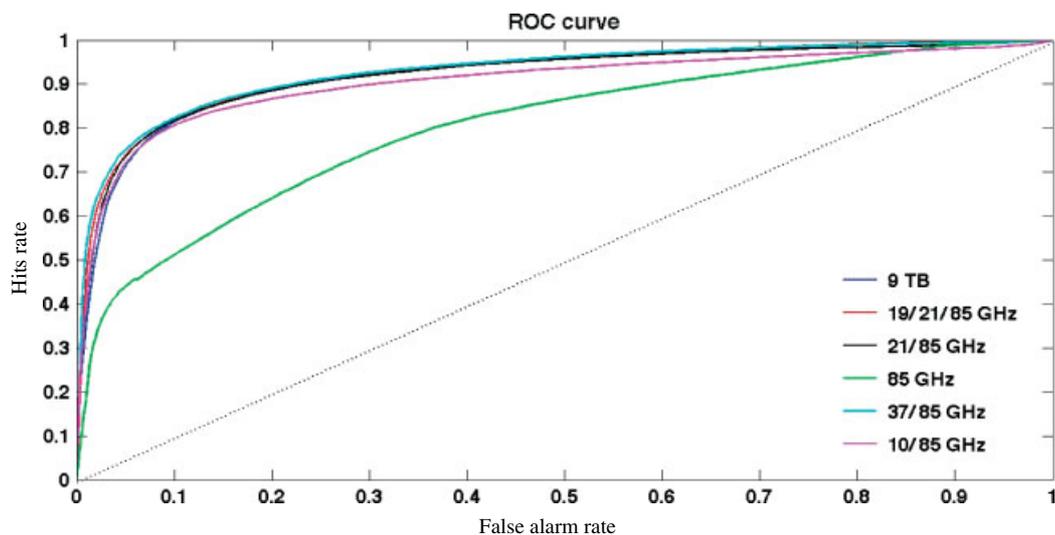


Figure 3. ROC curves obtained on calibrated TB.

Table 2. *AUC* for all MLP.

Variables (TB)	10.65–85.5 GHz	19.4–21–85 GHz	21.3–85 GHz	37–85.5 GHz	85.5 GHz	All TB
<i>AUC</i>	0.9047	0.9255	0.9209	0.9283	0.8020	0.9208

the exception of the MLP that was trained on the two 85 GHz and which exhibits an ROC curve below all the others, none of the classifiers stands out. Nevertheless, the ROC curve corresponding to the MLP that was trained on both 37 and 85 GHz frequencies seems to be lying slightly above and to the left of the other curves. To consolidate this qualitative analysis, the areas under all the ROC curves are computed. Results are shown in Table 2. As previously mentioned, the *AUC* is the most common global technique to quantify the accuracy of a classifier by a single number. Values of *AUC* usually range from 0.5 (meaning there is no apparent difference between the distributions of the two classes) to 1 (meaning that there is a perfect separation between the two classes). It can be seen in Table 2 that the model that is slightly ahead, as far as the *AUC* is concerned, is the one obtained with the 37 and 85 GHz TB (noted hereafter as MLP 37–85). It is worth noting that the channels which appear to be the best descriptors are the ones most often used in literature (see section 2). Several factors can impact the choice of those descriptors. At 37 and 85 GHz, scattering prevails and those channels are generally used to perform retrieval over land surfaces. In addition, the FOV for those channels is smaller than for the others, implying a better spatial resolution, which can help the detection of smaller convective cells and sharper gradients. Lower frequencies are often used in screening procedures. Indeed, the 21 GHz channel provides information on water vapour saturation in the neighbouring atmosphere and ice surfaces on the ground (Grody, 1991), while the 19 GHz channel provides information on polarization over desert areas.

To confirm the qualitative and quantitative analysis given by ROC plots and *AUC*, the total percentage of accurate classification is calculated at a fixed point and results are listed in Table 3. This point corresponds to the optimal threshold for each MLP that is evaluated through the corresponding ROC curve. In the present study, the optimal threshold is found to give the best compromise between specificity and sensitivity, which means that it spatially corresponds to the point closest to the top left corner on Figure 3. It is noticeable in Table 3 that the model providing the higher percentage of correct classification is the MLP 37–85 GHz.

The final step of the evaluation of our NN models is described in the following section. The MLP 37–85 was implemented in the BRAIN algorithm and the impact of this new screening methodology on the resulting estimates of rain rate was quantified and compared to the former decision tree at different spatial scales and over several TRMM orbits.

5. Results analysis

5.1. Global evaluation

BRAIN can be used either for a real-time application or to generate accumulated products. Therefore, depending upon the targeted application, the evaluation criteria might be different. Two levels of validation are considered for this study. First, a contingency table analysis is presented in Table 4. These results are especially relevant for instantaneous estimations where most applications require

Table 3. Total accuracy (%) for all MLP.

Variables (TB)	10.65–85.5 GHz	19.4–21–85 GHz	21.3–85 GHz	37–85.5 GHz	85.5 GHz	All TB
Total accuracy	83.9	91	90	91.5	82	89.2

Table 4. Scores (in %) for Decision tree (former screen) and MLP 37–85 above the sampled region.

Screen	Decision tree	MLP 2185	MLP 192185	MLP 3785
<i>Tropical belt</i>	<i>POD</i>	22.6%	63.8%	57.2%
	<i>FAR</i>	37.5%	48.8%	40.6%
	<i>WD_T</i>	68%	22.4%	31%

Table 5. Total mean bias for each screen.

	Decision tree	MLP 21–85	MLP 19–21–85	MLP 37–85
<i>Relative bias (%)</i>	121	65.36	82.6	56.8

a good description of the rainfall field structure and the spatial distribution of intensities. Second, mean values of the surface rainfall for each pixel category are computed. This gives a good indication of the possible bias that the screen would introduce in accumulated products. For instance, we assume that a high number of False Alarms corresponding to light rainfall intensities is less critical for the algorithm performance than a few cases of wrong detection of high-intensity rain.

Some classical scores to estimate the efficiency of a screen and its performance in terms of rain occurrence are calculated, namely:

- The Probability of Detection (*POD*)

$$POD = \frac{Hits}{Hits + WD}$$

- The False Alarm Rate (*FAR*)

$$FAR = \frac{FA}{Hits + FA}$$

- The total Wrong Detections percentage (*WD_T*)

$$WD_T = \frac{WD}{Hits + FA + WD}$$

In an ideal situation, the *POD* would have a maximum of 1 and the *FAR* and *WD_T* would equal 0. Those scores are generated for each MLP and compared to the former screening methodology.

At first, the various MLPs are compared to the existing decision tree on a global scale. As previously mentioned, some surfaces can lead to a misinterpretation by the screening procedure, such as snow cover. Thus, the impact of these surfaces on rainfall retrieval must be evaluated.

Fifty TRMM orbits were randomly selected and the BRAIN retrieval was performed over the entire tropical belt to produce the results shown in Table 5. Table 4 shows the performance of the screens in terms of occurrence. As expected, the decision tree exhibits the lowest *POD* and *FAR* but the greatest *WD_T*. All MLPs that were developed detect many more rain pixels, in agreement with the PR, than the former screen. The *FAR* values are quite different from one MLP to another with the lowest occurrences of False Alarms reached by the MLP 19–21–85. The highest *POD*

value is reached for the MLP 37–85 and is equal to 67.5%. This screen also provides the smallest number of wrong detections; however, this comes with a substantial number of False Alarms (73%).

Relative mean bias, obtained by considering pixels of all status (Hits, False Alarms and Wrong Detections) for each screen, is listed in Table 5. It appears from those values that the more restrictive the screen, the more significant the bias. This effect must be connected to the Bayesian estimation. Since this technique set up in BRAIN is a weighted average of its retrieval database elements, extremes are eroded, leading to overestimation of the light intensities and underestimation of the heavy ones. Hence, taking into account false alarms, the greater the number of detected low-rain pixels, the closer the mean value of the retrieved distribution will be to the lower intensities. Since BRAIN is globally overestimating the rain, the total bias with respect to PR will be reduced.

An important consideration in addressing rainfall detection in retrieval is the choice of the screen, which must be adapted to fit the envisioned application. As far as climatological applications are concerned, the first selection criterion is the relative bias, which must be minimized. As can be seen from results shown in Table 5, the MLP 37–85 appears to be the best screen. It provides a relative bias equal to 56.8%, which is lower than the one obtained when using the decision tree (121%). For instantaneous estimation, the mask that provides a relatively good *POD* while minimizing the *FAR* is the best. The results shown in Table 4 prove that only the MLP 19–21–85 exhibits a *FAR* which is close to the one provided by the decision tree. This MLP also exhibits a *POD* that is improved by nearly 35% from the *POD* of the decision tree.

Figure 4(b) presents the map obtained from rainfall mean values on a $1^\circ \times 1^\circ$ grid for August 2010 estimated by BRAIN using the MLP 37–85. The same map, but using the data from the PR, is presented in Figure 4(a). It can be seen that BRAIN and the PR are in good agreement in terms of rain areas. However, the map obtained from BRAIN data exhibits several regions of low surface rain rate corresponding to some false alarm cases. Clearly, some mountains with possible snow cover are not properly filtered out by the MLP. Those corresponding pixels, when kept in the retrieval process, can lead to high monthly-average rainfall intensities. Those cases are particularly visible above the Himalayas and the Andes. Given that it is extremely difficult to estimate rainfall over those regions, they are eliminated through a geographic screen that enables the deletion of surfaces exceeding 3500 m in altitude, although this solution is not entirely satisfactory. The map obtained when using this procedure is presented in Figure 4(c). Most of the ambiguous surfaces have been properly eliminated but a few pixels over the Andes persist. This preliminary solution could be improved, for example by applying a dynamic altitude threshold instead of a fixed one. The altitude of 3500 m seems to be an acceptable compromise between the deletion of false alarms and the preservation of hits over mountainous areas.

Some other regions with false alarms that apparently do not correspond to ambiguous surfaces are visible in Figure 4(b). It can be seen that South Africa and Australia present some localized regions of false alarms but with rather low intensities. The cause of this phenomenon is hard to determine at this point given the diversity of possible sources

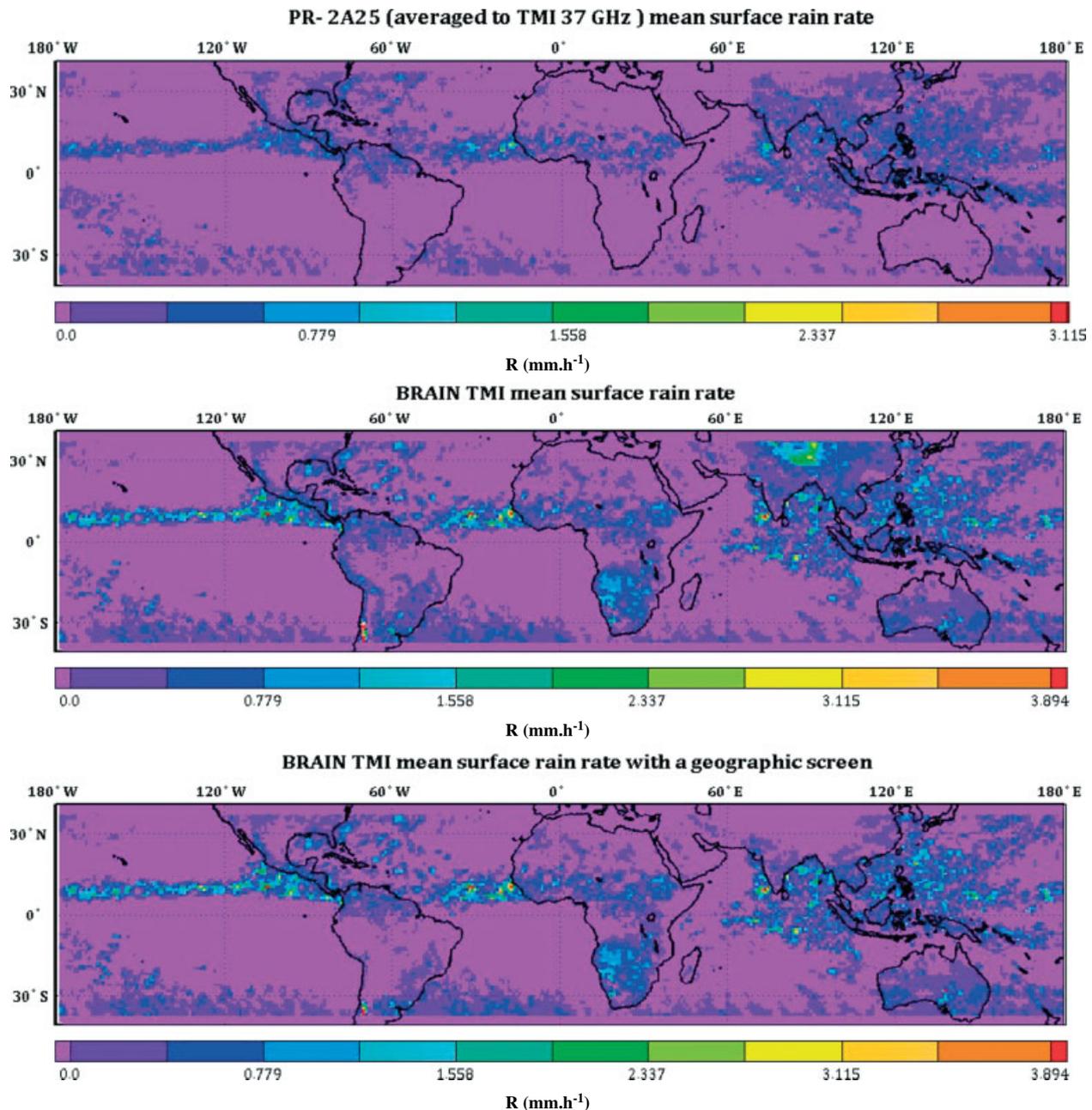


Figure 4. (a) Mean surface rain rates for August 2010, (b) from the PR, (c) from BRAIN-TMI using MLP 37–85, same as (b) but when using a geographic screen (above 3500 m of altitude filtered out).

for those false alarms. It is likely that some peculiar surface emissivity causes the MLP to fail in these regions.

5.2. Regional evaluation: Niger and Benin sites

The performances of BRAIN for rain detection were evaluated over West Africa (6°W/6°E/15°N/5°S) using both PR and ground data. The dataset from the PR used here consists of more than 1700 overpasses of TRMM between years 1998 and 2010.

The interest in this region has arisen from two major factors:

First, the recent study by Kirstetter *et al.* (2012) proposed an error model for BRAIN using data from rain-gauges that were available for this particular region and period. To allow the comparison of rainfall, rain-gauges were treated with kriging techniques in order to reach the same spatial

resolution as BRAIN-TMI (12.5 km). Similarly the PR pixels (2A25 v6: Iguchi *et al.*, 2000) within a 6 km radius of each TMI pixel were averaged together. Some uncertainty arises from this procedure because of the differences in scan geometry and the co-location errors between the various instruments. The practical minimum detectable rain rate by the PR is fixed by its sensitivity, which is about 0.7 mm h⁻¹ at its raw spatial resolution that is reduced to 0.3 mm h⁻¹ when averaged. Kirstetter *et al.* (2012) show that when compared to rain-gauges, PR is a good reference for rain presence (few false alarms). However, PR will eventually miss some very light rain (wrong detections). This light rain from rain-gauges is not considered as very robust (kriging errors are larger than the estimated rain rate).

Second, the focus on the Niger and Benin regions enables the comparison of two different climatic regimes and therefore different features of precipitation. Over Niger,

Table 6. Scores (in %) for Decision tree (former screen) and MLP 37–85 above the sampled region.

Screen		Decision tree	MLP 37–85
West Africa	POD	39.5	74
	FAR	3.7	24.3
	WD _T	59.6	21
NIGER	POD	35.1	82.5
	FAR	1.8	25.2
	WD _T	64.5	13.7
BENIN	POD	38.1	75.1
	FAR	4.1	24.4
	WD _T	60.8	20

convection is usually organized with strong updraughts creating many ice crystals with an easily identified scattering signal. Precipitation characteristics in Benin are more sensitive to local orography in the north, generating disorganized convection with cells, which might not be easy to capture within the FOV of the TMI channels. In addition, stratiform precipitation without a very defined scattering signature is more often encountered.

The values for *POD*, *FAR* and *WD_T* compared to the PR obtained with the MLP 37–85 (hereafter referred to as the MLP) and the former screen (GSCAT decision tree) are presented in Table 6. In general, the decision tree, which is by nature restrictive, exhibits low values of *POD* and *FAR* unlike the MLP, whose *POD* reaches 74% for the whole sample. It is not surprising due to the fact that the decision tree was created mainly to avoid false alarms related to high emissivity or polarized backgrounds. The *FAR* and *WD_T* are respectively increased and decreased when using the MLP since the NN screen detects many more rain pixels than the decision tree. The drawback of using a less stringent screen is that it naturally causes more false alarms. This can be explained not only by the nature of the discrimination between rain and no-rain from the radiometric signal in general, but also by the sensitivity of the PR. Given that the rain/no-rain limit is not clearly identified by any descriptor (cf. Figure 2), raising the occurrences of hits between PR and the NN models leads to an increase of the *FAR*.

Looking at the regional variability, the MLP appears to perform better in Niger than in Benin. This, as was previously mentioned, can be related to the more organized nature of convection in Niger, creating more intense TB depressions at 37 and 85 GHz. This regional sensitivity is not visible when using the decision tree. This aspect could be related to the use of 37 and 85 GHz channels that, as already mentioned, offer smaller pixels than the lower-frequency channels used within the tree.

This analysis of occurrences in Table 6 was further evaluated in terms of mean values of surface rainfall, and the results are presented in Table 7. Contrary to the decision tree, the values of mean surface rainfall for hits and total pixels are lower when using the MLP. These values highlight the fact that the MLP detects significantly more rainfall of light intensities than the decision tree. The *WD_T* are of the same order of magnitude for both screens, but it is noticeable that the values obtained with the MLP are slightly lower in terms of averaged rain intensity. Another unexpected result can be found in the mean values of false alarms that are higher for the decision tree than for the MLP. Owing to the fact that the MLP gives significantly more false alarms than

Table 7. Mean values (mm.h⁻¹) of surface categories for total region, Niger and Benin.

Screen		Decision tree	MLP 37–85
West Africa	Hits	5.2	3.9
	Wrong Detections(PR)	1.1	0.8
	False Alarms (BRAIN-TMI)	2.8	2.1
	Total	5.1	3.5
NIGER	Hits	7.8	4.8
	Wrong Detections(PR)	1.1	0.6
	False Alarms (BRAIN-TMI)	5.6	2.2
	Total	7.8	4.2
BENIN	Hits	4.3	3.6
	Wrong Detections(PR)	1.2	0.7
	False Alarms (BRAIN-TMI)	2.8	2.1
	Total	4.2	3.3

Table 8. Contingency table (%) for TMI over Niger and Benin areas.

Screen		Decision tree	MLP 37–85
Niger	Hits	21	42
	WD (GV)	30	9
	False Alarms (BRAIN-TMI)	0	8
	Correct rejection	49	41
Benin	Hits	18	28
	WD (GV)	28	18
	False Alarms (BRAIN-TMI)	1	6
	Correct rejection	53	49

Results are provided only for robust GV. Adapted from Kirstetter *et al.* (2012).

the decision tree, its *FA* mean value can be biased by the high amount of light-intensity rainfall that ends up in this category.

The design of the decision tree allows for different natures of false alarms. Indeed, the first threshold applied to the 85 GHz horizontal polarization (cf. Figure 1) is quite restrictive and strongly defines the no-rain pixel category. For that reason, this threshold limits the number of false alarms that are closely related to the ambiguous border between rain and no-rain situations, creating light-intensity false alarms. In the decision tree, the following tests identify the presence of rain in the pixels exhibiting a TB 85 GHz $H \leq 257$ K. Hence a false alarm occurs when a pixel passes through all the tests to eventually be identified as possible rain. By nature, the decision tree is therefore more susceptible to false alarms resulting from the misinterpretation of the background surface that can provide pixels of high-intensity rainfall. These pixels probably explain the higher mean rainfall value observed for false alarms when using the decision tree.

Finally, comparisons between BRAIN-TMI and rain-gauges (Ground Validation, GV) were made, using the same data over the same region. Results are listed in Table 8 and Table 9. It can be noted that in the contingency table (Table 8), occurrences of false alarms using the decision tree disappear. The number of hits doubles over Niger and increases by 10% over Benin when using the MLP. Values

Table 9. Hits and misses (Wrong Detections) (%) of BRAIN-TMI relative to GV rainfall with discarded rain volume by BRAIN-TMI due to misses relative to GV.

Screen		Decision tree	MLP 37–85
Niger	Hits	41	82
	WD	59	18
	Discarded rain volume	32	10
Benin	Hits	39	61
	WD	61	39
	Discarded rain volume	52	36

Adapted from Kirstetter *et al.* (2012).

of WD_T are considerably improved over Niger while they decrease by 18% over Benin.

In Table 9, only the cases of hits and wrong detections are taken into account. It is obvious that the sensitivity–specificity ratio is inverted for the two screens, the MLP being much more sensitive over both regions. As far as the discarded rain volume is concerned, the MLP enables the recovery of a substantial amount of rain, but the volume of missing rain remains more significant over Benin for both screens.

6. Conclusion

A series of masks based on artificial neural networks of the multilayer perceptron type were developed and tested to discriminate rain/no-rain regimes from passive microwave radiometers data over land. The training database was made of a series of profiles from the TRMM PR associated with their respective simulated brightness temperatures. Various combinations of input channels from TMI data were used, and provided a much better detection of the rain/no-rain boundaries than the original decision tree based on GSCAT-2. The use of ROC curves and their area helped to select the best combination of channels, which was then implemented in the BRAIN retrieval algorithm.

First, the performances of the new masks over the entire tropical region were evaluated on a number of randomly selected TRMM orbits, showing a strong regional dependence with local problems that need to be addressed separately. Generally, the Probability Of Detection is substantially improved but at the cost of a degraded False Alarm Rate while the wrong detections are kept under control. This shows the intrinsic difficulty in distinguishing the rain from the no-rain regions over a continental background due to the high variability of the surface emissivity.

Second, 10 years of BRAIN retrieval using the TRMM TMI data were compared to both PR data and rain-gauge data over two sites in Benin in Niger. This regional comparison showed that the performance of the mask is indeed affected by local conditions, although the MLP relying on the 37 and 85 GHz generally performed better than the older GSCAT-2 mask. The analysis also confirmed that the mask does indeed contribute substantially to the bias observed between the rain-gauges' data and the retrieval by controlling the number of light-rain pixels.

Depending on the rain retrieval algorithm application, features and performances of one or the other configuration of the multilayer perceptron must be considered. The use

of a combination of 37 and 85 GHz provides the lowest bias relative to the PR and therefore appears more adequate for climatological applications. On the contrary, the more complete combination of 19, 21 and 85 GHz exhibits a False Alarm Rate equivalent to the one obtained when using the former GSCAT-2 screen and appears better suited for real-time applications. The next step will be to develop a series of specialized networks to account for regional specifics. This raises the difficult task of avoiding discontinuities when switching from one mask to the other in adjacent regions.

References

- Adler RF, Huffman GJ, Keehn PR. 1994. Global tropical rain estimates from microwave-adjusted geosynchronous IR data. *Remote Sensing Rev.* **11**: 125–152.
- Aires F, Bernardo F, Brogniez H, Prigent C. 2010. An innovative calibration method for the inversion of satellite observations. *J. Appl. Meteorol. Clim.* **49**: 2458–2473.
- Bennartz R, Thoss A, Dybbroe A, Michelson DB. 2002. Precipitation analysis using the Advanced Microwave Sounding Unit in support of nowcasting applications. *Meteorol. Appl.* **9**: 177–189.
- Bradley AP. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recogn.* **30**: 1145–1159.
- Ferraro RR, Smith EA, Berg W, Huffman GJ. 1998. A screening methodology for passive microwave precipitation retrieval algorithms. *J. Atmos. Sci.* **55**: 1583–1600.
- Ferraro RR, Weng FZ, Grody NC, Zhao LM. 2000. Precipitation characteristics over land from the NOAA-15 AMSU sensor. *Geophys. Res. Lett.* **27**: 2669–2672.
- Gopalan K, Wang N-Y, Ferraro RR, Liu CT. 2010. Status of the TRMM 2A12 land precipitation algorithm. *J. Atmos. Oceanic Technol.* **27**: 1343–1354.
- Greco M, Anagnostou EN. 2001. Overland precipitation estimation from TRMM passive microwave observations. *J. Appl. Meteorol.* **40**: 1367–1380.
- Greenwald TJ, Christopher SA. 2002. Effect of cold clouds on satellite measurements near 183 GHz. *J. Geophys. Res.* **107**: 4170, DOI: 10.1029/2000JD000258.
- Grody NC. 1991. Classification of snow cover and precipitation using the Special Sensor Microwave Imager. *J. Geophys. Res.* **96**: 7423–7435.
- Hollinger JP. 1991. 'DMSP Special Sensor Microwave/Imager calibration/validation.' Naval Res. Lab., Washington, DC, NRL Final Rep., 314 pp.
- Hsu K-L, Gao XG, Sorooshian S, Gupta HV. 1997. Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteorol.* **36**: 1176–1190.
- Iguchi T, Kozu T, Meneghini R, Awaka J, Okamoto K. 2000. Rain-profiling algorithm for the TRMM Precipitation Radar. *J. Appl. Meteorol.* **39**: 2038–2052.
- Kirstetter P-E, Viltard N, Gosset M. 2012. An error model for instantaneous satellite rainfall estimates: Evaluation of BRAIN-TMI over West Africa. *Q. J. R. Meteorol. Soc.* Special issue on the Megha-Tropiques mission. DOI: 10.1002/qj.1964.
- Kniveton DR, Motta BC, Goodman HM, Smith M, LaFontaine FJ. 1994. The first WetNet precipitation intercomparison project: Generation of results. *Remote Sensing Rev.* **11**: 243–301.
- Kummerow CD, Olson WS, Giglio L. 1996. A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Trans. Geosci. Remote Sensing* **34**: 1213–1232.
- Kummerow CD, Hong Y, Olson WS, Yang S, Adler RF, McCollum J, Ferraro RR, Petty G, Shin D-B, Wilhelm TT. 2001. The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteorol.* **40**: 1801–1820.
- L'Ecuyer TS, Stephens GL. 2002. An uncertainty model for Bayesian Monte Carlo retrieval algorithms: Application to the TRMM observing system. *Q. J. R. Meteorol. Soc.* **128**: 1713–1737.
- Mallet C, Moreau E, Casagrande L, Klapisz C. 2002. Determination of integrated cloud liquid water path and total precipitable water from SSM/I data using a neural network algorithm. *Int. J. Remote Sensing* **23**: 661–674.
- Moreau E, Mallet C, Thiria S, Mabboux B, Badran F, Klapisz C. 2002. Atmospheric liquid water retrieval using a gated expert neural network. *J. Atmos. Oceanic Technol.* **19**: 457–467.
- Petty GW. 1994. Physical retrievals of over-ocean rain rate from multichannel microwave imagery. Part I: Theoretical characteristics

- of normalized polarization and scattering indices. *Meteorol. Atmos. Phys.* **54**: 79–99.
- Stephens GL, Kummerow CD. 2007. The remote sensing of clouds and precipitation from space: A review. *J. Atmos. Sci.* **64**: 3742–3765.
- Tapiador FJ, Kidd C, Hsu K-L, Marzano F. 2004a. Neural networks in satellite rainfall estimation. *Meteorol. Appl.* **11**: 83–91.
- Tapiador FJ, Kidd C, Levizzani V, Marzano FS. 2004b. A neural networks-based fusion technique to estimate half-hourly rainfall estimates at 0.1° resolution from satellite passive microwave and infrared data. *J. Appl. Meteorol.* **43**: 576–594.
- Viltard N, Kummerow CD, Olson WS, Hong Y. 2000. Combined use of the radar and radiometer of TRMM to estimate the influence of drop size distribution on rain retrievals. *J. Appl. Meteorol.* **39**: 2103–2114.
- Viltard N, Burlaud C, Kummerow CD. 2006. Rain retrieval from TMI brightness temperature measurements using a TRMM PR-based database. *J. Appl. Meteorol. Clim.* **45**: 455–466.
- Zhang M, Scofield RA. 1994. Artificial neural network techniques for estimating heavy convective rainfall and recognising cloud mergers from satellite data. *Int. J. Remote Sensing* **15**: 3241–3261.