Using naive Bayes classifier for classification of convective rainfall intensities based on spectral characteristics retrieved from SEVIRI

SLIMANE HAMEG, MOURAD LAZRI* and SOLTANE AMEUR

Laboratoire Lampa, University Mouloud Mammeri, Tizi Ouzou, Algeria.

**Corresponding author. e-mail: m_lazri@yahoo.fr*

This paper presents a new algorithm to classify convective clouds and determine their intensity, based on cloud physical properties retrieved from the Spinning Enhanced Visible and Infrared Imager (SEVIRI). The convective rainfall events at 15 min, 4×5 km spatial resolution from 2006 to 2012 are analysed over northern Algeria. The convective rain classification methodology makes use of the relationship between cloud spectral characteristics and cloud physical properties such as cloud water path (CWP), cloud phase (CP) and cloud top height (CTH). For this classification, a statistical method based on ‘naive Bayes classifier’ is applied. This is a simple probabilistic classifier based on applying ‘Bayes’ theorem with strong (naive) independent assumptions. For a 9-month period, the ability of SEVIRI to classify the rainfall intensity in the convective clouds is evaluated using weather radar over the northern Algeria. The results indicate an encouraging performance of the new algorithm for intensity differentiation of convective clouds using SEVIRI data.

1. Introduction

In general, intense rainfall events, both in the tropical and extratropical regions, are dominated by the presence of convective rain cells. The estimation of convective precipitation and its contribution to total precipitation is an important issue in meteorology.

In the Mediterranean region, significant rainfall amounts are recorded during warm periods from April to October of the year. The main source of this rainfall during these periods is the convective systems (Lazri *et al.* 2014b). Mediterranean storms are often characterised by heavy precipitation, intense wind shear and deep atmospheric lows, and heavily affect coastal and continental regions. In fact, daily rainfall exceeding 200 mm has been recorded in these regions (Lazri *et al.* 2015). These convective events take place mostly

during autumn and summer. The causes of such severe events are generally attributed to the geographic complexity of the Mediterranean basin especially at the local scale, where the orography plays an important role in the destabilisation of the westerly oceanic systems reaching the area.

Over the past decades, several rain rate retrieval methods based on passive imager observations have been developed for convective rainfall. The methods based on geostationary satellites use thermal infrared observations and relate cloud top height (CTH) (e.g., Adler and Negri 1988; Huffman *et al.* 2001). These methods perform reasonably well over areas where rainfall is governed by deep convection, but are less favourable when the convective cloud systems also constitute a part of low clouds convection which can turn into stratiform cloud. Indeed, this part of precipitation is generally underestimated because their top temperatures

Keywords. Convective cloud; SEVIRI; weather radar; naive Bayes classifier.

cannot be discriminated from no-raining cloud top temperature.

These composed cloudy convective systems can be identified and classified according to their intensities using their physical properties. In this case, only IR thermal data are not sufficient, more information is necessary for satellite precipitation classification. Therefore, besides CTH information, cloud water path (CWP) and cloud phase (PC) should be used for more reliable classification of convective clouds (Lazri *et al.* 2013a, b; Kühnlein *et al.* 2014). CWP represents the amount of water vertically integrated into the cloud and depends on the diameter of raindrops and the cloud thickness. CWP is directly related to the rainfall intensity. Also, ice particles in the upper part of the clouds (Cloud Phase) are needed to support convective rain generation with the so-called Bergeron–Findeisen process (Houze 1993).

Passive imagers have been used extensively to retrieve cloud water path, cloud thermodynamic chase and cloud top height from satellite radiances (Roebeling *et al.* 2006; Thies *et al.* 2010; Lazri *et al.* 2012). Validation studies confirmed that CWP, CP and CTH can be retrieved with high accuracy from Spinning Enhanced Visible and Infrared Imager (SEVIRI) onboard METEOSAT-8 (Roebeling *et al.* 2008).

However, few studies have investigated the potential of using cloud physical property retrievals from the SEVIRI onboard the geostationary METEOSAT to classify precipitation intensity although this classification can be of relevance for the development of techniques for convective precipitation estimation.

This paper presents a method for convective rainfall intensity separation based on cloud physical properties retrieved from Meteosat Second Generation (MSG) satellite. Meteorological satellites are the only instruments capable of providing large amount of rainfall measurements in remote areas where data are difficult or impossible to collect from the ground. The method proposed here is based on a naive Bayes classifier. This classifier is selected due to its classification efficacy and implementation simplicity (Hall and Holmes 2003). As a result, it is common to provide classification accuracy achieved by a naive Bayes classifier as a comparison baseline (Hall and Holmes 2003; Polat and Gunes 2006). The training and validity of this scheme are carried out by comparisons with weather radar observations. In order to realise an inter-comparison, estimates of precipitation from convective systems classification are carried out.

The structure of this paper is as follows: the underlying datasets and methods are presented in section 2. Section 3 gives a procedure of the

application of the model and results. The paper is closed with conclusions in section 4.

2. Data and methods

2.1 Satellite and radar data

The dataset used in this study is MSG/SEVIRI data and corresponding ground-based radar data. Aiming to achieve comparability of these types of data, spatial and temporal collocation between satellite and radar observations have been carried out.

2.1.1 Satellite observations

MSG is a new series of European geostationary satellites that is operated by EUMETSAT. The first MSG *Meteosat-8* is a spinning stabilised satellite that carries the 12-channel SEVIRI instrument with three channels at visible and near-infrared wavelengths between 0.6 and 1.6 μm , eight channels at infrared wavelengths between 3.8 and 14 μm , and one high-resolution visible channel. The nadir spatial resolution of SEVIRI is $1 \times 1 \text{ km}^2$ for the broadband high-resolution channel and $3 \times 3 \text{ km}^2$ for other channels. Over northern Algeria, the satellite viewing zenith angle of SEVIRI is about 45° , and consequently the spatial resolution is reduced to about $4 \times 5 \text{ km}^2$.

2.1.2 Meteorological radar observations

The radar-based precipitation data of the Setif Weather Service is used for the development and validation of the new convective rainfall classification technique. The radar of Setif, installed near the town of Setif, at $36^\circ 11' \text{N}$, $5^\circ 25' \text{E}$ with an altitude of 1700 m, is one of the seven Algerian meteorological radars. This is a Radar AWSR 81C in C-band; its operational frequency is 5.6 GHz. The displacement in azimuth is between 0 and 360° in continuous and the movement in inclination is -1° to 90° . Its polarisation is linear and horizontal.

Radar data are collected at a temporal resolution of 15 min and a spatial resolution of 1 km in a format of 512×512 pixels. The physical parameter of the radar is the reflectivity factor, denoted by Z (dBZ). Each pixel is coded on four bits. Thus, it consists of 16 classes representing different reflectivity intensities [04, 12, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70 (dBZ)]. The relationship used to convert the reflectivity factor Z (mm^6/m^3) into precipitation intensity R (mm/h) is given by equation (1):

$$Z = 382 R^{1.85}. \quad (1)$$

2.1.3 Spatial and temporal collocation of satellite and radar observations

In order to compare MSG data with ground radar data, the radar data is re-mapped onto the METEOSAT grid due to spatial shift between satellite and radar images that can be of up to 20 km. The high-resolution radar field ($1 \times 1 \text{ km}^2$) was degraded and adapted to the satellite resolution ($4 \times 5 \text{ km}^2$). Every satellite pixel, not filled by radar observations, was rejected from the analysis. In the study region, the resolution $4 \times 5 \text{ km}^2$ is assumed constant due to small overlapped area observed by both sensors.

The time lag between the radar and the satellite is about 3 min. This small time difference does not require synchronisation between the two data types.

The data were divided into a training and a validation dataset. The training dataset was used for the development of the technique and consists of only convective precipitation events from 2006 to 2012. The validation dataset was considered for the appraisal of the proposed technique and consisted of convective scenes from April 2013 to October 2013.

2.2 Methodology

The convective rain classification methodology makes use of the relationship between cloud spectral characteristics and cloud physical properties such as cloud water path (related to the particle size and the cloud thickness), cloud phase and cloud top height. For this classification, a statistical method based on naive Bayes classifier (NBC) was applied. This is a simple probabilistic classifier based on applying Bayes theorem with strong (naive) independent assumptions. The classifier was trained by utilising the cloud radiance database prepared from the MSG satellite.

To achieve this classification, the method works in two phases:

- Extraction of cloud spectral characteristics
- Application of classifier NBC

2.2.1 Extraction of cloud spectral characteristics

The spectral characteristics from the SEVIRI used to gain information about cloud physical properties (CTH, CWP and CP) are brightness temperature in IR10.8 channel (infrared at $10.8 \text{ }\mu\text{m}$), reflectance in VIS0.6 channel (visible at $0.6 \text{ }\mu\text{m}$) and NIR1.6 channel (near-infrared at $1.6 \text{ }\mu\text{m}$) and the brightness temperature differences (see table 1) between the thermal bands of MSG SEVIRI ($\text{BTD}_{3.9-10.8}$, $\text{BTD}_{3.9-7.3}$, $\text{BTD}_{8.7-10.8}$, $\text{BTD}_{10.8-12.0}$, $\text{BTD}_{6.2-10.8}$).

These spectral parameters derived from SEVIRI used for classification are summarised in table 1.

- Brightness temperature $\text{BT}_{10.8}$ is an indication of the vertical extent of the cloud because, in general, brightness temperature of the convective system depends on the cloud top height (CTH).
- The CWP is used to identify raining clouds. Indeed, it provides information on the effective droplet radius (r_e) and the optical thickness τ of clouds. The effective particle radius (r_e) defined by the ratio between the third and the second power of the droplet spectrum is taken in place of the actual droplet spectrum. The cloud optical thickness (τ) defined by the integration of the extinction coefficient integrated over the cloud geometrical thickness is considered representatively for the cloud geometrical thickness. The CWP represents the amount of water vertically integrated in the cloud and depends on the diameter of raindrops and the thickness of the cloud formed by these drops. The relationship is given by the following equation:

$$\text{CWP} = \frac{2}{3} \rho \cdot \tau \cdot r_e, \quad (2)$$

where ρ (g/m^3) is the density of water in the clouds (Thies *et al.* 2008a, b; Nauss and Kokhanovsky 2006).

The CWP can therefore be used as a delimiter between the non-raining and raining clouds (e.g., Nauss and Kokhanovsky 2006).

Table 1. Spectral parameters used for intensity differentiation of convective clouds.

Spectral parameters	
$\text{BTD}_{3.9-10.8}$	Brightness temperature difference between the IR3.9 and the IR10.8 μm channels
$\text{BTD}_{3.9-7.3}$	Brightness temperature difference between the IR3.7 and the WV7.3 μm channels
$\text{BT}_{10.8}$	Brightness temperature at the IR10.8 μm channel
$\text{BTD}_{10.8-12.0}$	Brightness temperature difference between the IR10.8 and the IR12.0 μm channels
$\text{BTD}_{6.2-10.8}$	Brightness temperature difference between the WV6.2 and the IR10.8 μm channels
$\text{BTD}_{8.7-10.8}$	Brightness temperature difference between the IR8.7 and the IR10.8 μm channels
$\text{R}_{0.6}$	Reflectance in VIS0.6
$\text{R}_{1.6}$	Reflectance in NIR1.6

During day-time, the CWP (i.e., values of r_e and τ) considered for a rainfall intensity differentiation can be retrieved on a pixel basis using a combination of two solar channels, namely the VIS0.6 and NIR1.6 channels of MSG (Lazri *et al.* 2013a, b; Thies *et al.* 2008a). High values of reflectance $R_{0.6}$ correspond to high optical depth of cloud and low values of reflectance $R_{1.6}$ indicate large particles in the cloud. This means that a large CWP is obtained when high values of $R_{0.6}$ coincide with low values of $R_{1.6}$. It should be noted that the retrievals are limited to satellite and solar viewing zenith angles smaller than 72° .

To compute implicitly the CWP, it is decided to use the original reflectance of VIS0.6 and NIR1.6 channels, instead of the computed values of effective droplet radius and optical thickness of clouds. Here, CWP classes are constructed from respective combinations of $R_{0.6}$ *vs.* $R_{1.6}$ derived from the comparison of SEVIRI scenes between April 2006 and October 2006 with corresponding meteorological radar reflectivity intensities (figure 1).

During night-time, combinations of brightness temperature differences $BTD_{3.9-10.8}$ and $BTD_{3.9-7.3}$ are used to infer implicit information about CWP (Lazri *et al.* 2013a, b; Thies *et al.* 2008b). Indeed, for thin clouds with small or large particles (small or medium CWP), respectively brightness temperature differences reach the highest values. Thick clouds with small particles (medium CWP) lead to small values of brightness temperature differences. In contrast, large particles together with a high optical

thickness (high CWP) result in medium values of brightness temperature differences. A raining cloud indicates mean values of brightness temperature differences.

Therefore, suitable combinations of brightness temperature differences ($BTD_{3.9-7.3}$ and $BTD_{3.9-10.8}$) of MSG SEVIRI to infer qualitative information on CWP are used. Here, CWP classes are constructed from respective combinations of $\Delta T_{3.9-10.8}$ *vs.* $\Delta T_{3.9-7.3}$ derived from the comparison of SEVIRI scenes between April 2006 and October 2006 with corresponding meteorological radar reflectivity intensities (figure 1).

The CWP takes values from 1 to 12 according to the coincidences of $R_{0.6}$ and $R_{1.6}$ during the daytime or $BTD_{3.9-10.8}$ and $BTD_{3.9-7.3}$ during night-time.

- (c) The brightness temperature difference $BTD_{10.8-12.0}$ (hereafter called $BTD1$) provides information on cloud optical thickness. It is used to discriminate optically thick clouds from optically thin clouds (Inoue 1987). Indeed, due to their black-body characteristics, optically thick cloud shows smaller $BTD1$, while optically thin cloud shows larger $BTD1$. In general, optically thick clouds produce precipitation.
- (d) The brightness temperature difference $BTD_{8.7-10.8}$ (hereafter called $BTD2$) can be utilized to gain information about the CP (Strabala *et al.* 1994). Radiative transfer simulations show that for ice clouds, $BTD2$ tends to be positive, whereas for low-level water clouds, $BTD2$ tends to be small negative (Baum and Platnick 2006). Mean $BTD2$ for the convective system

Figure 1. Cloud water path as a function of (a) $R_{VIS0.6}$ *vs.* $R_{NIR1.6}$ and (b) $\Delta T_{IR3.9-IR10.8}$ *vs.* $\Delta T_{IR3.9-WV7.3}$ (Lazri *et al.* 2013b).

are positive, increasing almost steadily with time. This implies an increasing ice formation in the convective system throughout its entire lifecycle.

- (e) Brightness temperature differences $\text{BTD}_{6.2-10.8}$ (hereafter called BTD3) is largely used in distinguishing between high-level and low-level/mid-level clouds (Lazri *et al.* 2014a, b). It usually takes very small negative values for upper level thick clouds. Positive differences may occur when water vapour is present in the stratosphere above the cloud top, which is a sign of convective cloud tops (Fritz and Laszlo 1993; Schmetz *et al.* 1997). Mean BTD3 at the time of maximum extent of the deep convective clouds is very close to zero, with a time evolution in fully agreement with the $\text{BT}_{10.8}$ pattern. This implies that intense convective cloud tops are penetrating the tropopause.

2.2.2 Naïve Bayes algorithm

The Naïve Bayes algorithm (Hanson *et al.* 1991) is a classification algorithm based on Bayes rule, that assumes the attributes $X = \{X_1, \dots, X_n\}$ are all conditionally independent of each other, given Y . The value of this assumption is that it dramatically simplifies the representation of $P(X/Y)$, and the problem of estimating it from the training data. Considering X contains n attributes which are conditionally independent of one another given Y , the representation is given by equation (3):

$$P\left(\frac{X_1, \dots, X_n}{Y}\right) = \prod_{i=1}^n P\left(\frac{X_i}{Y}\right). \quad (3)$$

Assuming in general that Y is any discrete-valued variable and the attributes $X = \{X_1, \dots, X_n\}$ are any discrete or real-valued attributes. The goal is to train a classifier that will output the probability distribution over possible values of Y , for each new instance X that needs to be classified. The expression for the probability that Y will take on its k th possible value, according to Bayes rule, is

$$P(Y = y_k/X_1, \dots, X_n) = \frac{P(Y = y_k) P(X_1, \dots, X_n/Y = y_k)}{\sum_j P(Y = y_j) P(X_1, \dots, X_n/Y = y_j)}, \quad (4)$$

where the sum is taken over all possible values y_j of Y . Now, assuming X_i are conditionally independent given Y , one can use the equation to rewrite as:

$$P(Y = y_k/X_1, \dots, X_n) = \frac{P(Y = y_k) \prod_i P(X_i/Y = y_k)}{\sum_j P(Y = y_j) \prod_i P(X_i/Y = y_j)}, \quad (5)$$

which is the fundamental equation for the Naïve Bayes classifier. The $P(Y = y_k/X_1, \dots, X_n)$ is the posterior probability of class membership, i.e., the probability that X belongs to $Y = y_k$.

Given a new instance $X_{\text{new}} = (X_1, \dots, X_n)$, this equation shows how to calculate the probability that Y will take on any given value, given the observed attribute values of X_{new} and given the distributions $P(Y)$ (prior probability) and $P(X_i/Y)$ (likelihood) and estimated from the training data. If most probable value of Y is to be found, then the Naïve Bayes classification rule is given by

$$y \leftarrow \arg \max_{y_k} \frac{P(Y = y_k) \prod_i P(X_i/Y = y_k)}{\sum_j P(Y = y_j) \prod_i P(X_i/Y = y_j)} \quad (6)$$

which simplifies to the following expression (because the denominator does not depend on y_k).

$$y \leftarrow \arg \max_{y_k} P(Y = y_k) \prod_i P(X_i/Y = y_k). \quad (7)$$

2.2.3 Calibration procedure

The calibration procedure is that the SEVIRI retrievals of precipitation are compared against corresponding weather radar observations (reference data). Indeed, the training of classifier used here (NBC) is carried out by comparing the four classes y_i namely non-raining, slightly convective, moderately convective and strongly convective derived from radar data with the corresponding combination of attributes (X_i) $\{X_1 = \text{BT}_{10.8}$, $X_2 = \text{CWP}$, $X_3 = \text{BTD1}$, $X_4 = \text{BTD2}$ and $X_5 = \text{BTD3}\}$ derived from MSG (table 2). The comparison is done for an independent calibration dataset comprising observations between 2006 and 2012. It should be noted that the time of year chosen for the calibration is between April and October. During this time of year, the main source of rainfall is convective systems. The coincidences between spectral parameters and classes of radar data are given in table 2.

Thus, from calibration table (table 2), different probabilities useful for classification were calculated using Bayes theory. These probabilities are the following:

Prior probability represented by $P(Y_i)$, likelihood represented by $P(X_j(k)_{Y_i}/Y_n)$ and posterior probability given by $P(Y_n/X_j(k)_{Y_i})$ where $i = 1, 2, 3$ or 4 , $n = 1, 2, 3$ or 4 , $j = 1, 2, 3, 4$ or 5 , and $k = 1$ to a, b, c or d . a, b, c and d are the number of pixels belong to class $Y1$, $Y2$, $Y3$ and $Y4$, respectively.

These probabilities will be used for classification of validation data collected between April 2013 and October 2013. From calibration procedure, it is thus possible to draw up a classification of

Table 2. *Classes Y_i and corresponding attributes X_i .*

Classes (Y_i) derived from radar data		Corresponding attributes (X_i) derived from MSG data				
		X_1	X_2	X_3	X_4	X_5
Non-raining	$Y1 < 0.1 \text{ mm/h}$	$X_1(1)_{Y1}$	$X_2(1)_{Y1}$	$X_3(1)_{Y1}$	$X_4(1)_{Y1}$	$X_5(1)_{Y1}$
		.	.	.	.	.
		.	.	.	.	.
		$X_1(a)_{Y1}$	$X_2(a)_{Y1}$	$X_3(a)_{Y1}$	$X_4(a)_{Y1}$	$X_5(a)_{Y1}$
		.	.	.	.	.
Slightly convective	$0.1 \text{ mm/h} \leq Y2 < 1.7 \text{ mm/h}$	$X_1(1)_{Y2}$	$X_2(1)_{Y2}$	$X_3(1)_{Y2}$	$X_4(1)_{Y2}$	$X_5(1)_{Y2}$
		.	.	.	.	.
		.	.	.	.	.
		$X_1(b)_{Y2}$	$X_2(b)_{Y2}$	$X_3(b)_{Y2}$	$X_4(b)_{Y2}$	$X_5(b)_{Y2}$
		.	.	.	.	.
Moderately convective	$1.7 \text{ mm/h} \leq Y3 < 7.1 \text{ mm/h}$	$X_1(1)_{Y3}$	$X_2(1)_{Y3}$	$X_3(1)_{Y3}$	$X_4(1)_{Y3}$	$X_5(1)_{Y3}$
		.	.	.	.	.
		.	.	.	.	.
		$X_1(c)_{Y3}$	$X_2(c)_{Y3}$	$X_3(c)_{Y3}$	$X_4(c)_{Y3}$	$X_5(c)_{Y3}$
		.	.	.	.	.
Strongly convective	$Y4 \geq 7.1 \text{ mm/h}$	$X_1(1)_{Y4}$	$X_2(1)_{Y4}$	$X_3(1)_{Y4}$	$X_4(1)_{Y4}$	$X_5(1)_{Y4}$
		.	.	.	.	.
		.	.	.	.	.
		$X_1(d)_{Y4}$	$X_2(d)_{Y4}$	$X_3(d)_{Y4}$	$X_4(d)_{Y4}$	$X_5(d)_{Y4}$
		.	.	.	.	.

the rainfall events in the light of their greater or lesser convective character. However, the assumption considered here holds valid for the present scope of work or at least for climates with the same characteristics as the climate of the study area.

3. Application of NBC to convective rainfall intensity separation and results

For an evaluation of the new proposed method for rainfall intensity differentiation, the new scheme is applied to a validation dataset consisting of convective precipitation events from April to October 2013. The precipitation events chosen for the evaluation study are independent from the above-mentioned precipitation events used for calibration procedure. It should be noted that Radar data are used here only for validating the developed scheme.

To assign each pixel representing a combination of values (X_1, X_2, X_3, X_4, X_5) to a class, posterior probabilities for each class are calculated using the relationship (4). Thus, four posterior probabilities are obtained; posterior Y_1, Y_2, Y_3 and Y_4 . The pixel is then assigned to a class Y_1, Y_2, Y_3 or Y_4 whose posterior probability is higher (section 2.2.2).

3.1 Case of a single rainfall situation

The NBC is applied to an instantaneous situation of convective events of April 5, 2013 (11:45 UTC) to

classify the rainfall according to its intensities. To gain a first visual impression of the performance of the proposed scheme, the results of the classification of convective clouds are presented in figure 2. Figure 2(a) shows the brightness temperature in the channel IR10.8. Figure 2(b) shows areas classified together by RADAR and developed scheme.

For this instantaneous situation, by comparison with ground radar data, the correct classification rate obtained here is evaluated at about 89%.

3.1.1 Case of situations observed between April and October 2013

Validation scenes of rainfall during the period from April to October 2013 were classified using the developed scheme. Altogether 516 scenes were observed by radar, which counted 2,185,463 raining pixels (see table 3). In general, strongly convective area covers about 25% of the whole rain area. While the moderately convective and slightly convective area cover about 40% and 35% of the whole rain area, respectively.

For evaluation, the SEVIRI retrievals are compared against weather radar observations. Indeed, we used the Probability of Detection (POD), Probability of False Detection (POFD) False Alarm Ratio (FAR) the bias, the Critical Success Index (CSI) and the Percentage of Corrects (PC). These parameters are calculated from table 4, in which

Figure 2. Rain area differentiation for the scene of April 5, 2013 (11:45 UTC). (a) BT_{10.8} image and (b) rain area differentiation by RADAR and developed scheme.

Table 3. Number of raining pixels during the period from April to October 2013.

	Slightly convective	Moderately convective	Strongly convective
Number of raining pixels	724,410	839,797	621,256

a , b , c and d were values from a contingency table (table 4). The optimum values of these statistical parameters are given in table 5.

The statistical parameters are calculated by using equations (8–13).

- The probability of detection (POD) measures the fraction of observed events that were correctly identified and is given by

$$\text{POD} = \frac{a}{a + c}. \quad (8)$$

- The probability of false detection (POFD) indicates the fraction of pixels incorrectly identified by the satellite method and is given by:

$$\text{POFD} = \frac{b}{b + d}. \quad (9)$$

- The false alarm ratio (FAR) measures the fraction of estimated events that were actually not events and is calculated by:

$$\text{FAR} = \frac{b}{a + b}. \quad (10)$$

- The frequency BIAS index (Bias), which is calculated as:

$$\text{Bias} = \frac{a + b}{a + c}. \quad (11)$$

- The critical success index (CSI) measures the fraction of observed and/or estimated events that were correctly diagnosed. It is given by:

$$\text{CSI} = \frac{a}{a + b + c}. \quad (12)$$

- The percentage of corrects (PC) is the percentage of correct estimations and is calculated as:

$$\text{PC} = \frac{a + d}{n}. \quad (13)$$

The evaluation of method performances is accomplished by analysing the calculated verification scores previously described. Table 5 shows an overview of the values of verification score for the classes, slightly convective, moderately convective and strongly convective.

3.2 Discussion of results

The pixels of the slightly convective fields detected by the proposed scheme are in good agreement with those identified in the radar data. But the detected pixels are somewhat underestimated by the developed scheme (bias: 0.6). The POD indicates that on average about 65% of the pixels classified as slightly convective precipitation by the radar are consistently identified by the satellite scheme together with a very low percentage of wrongly classified pixels indicated by the POFD (19%). The FAR (32%) average value is relatively moderate. The CSI (63%) indicates a good degree of correctly classified pixels. The positive impression concerning the classified area of slightly convective precipitation is further supported by the

Table 4. Overview of the combinations in a contingency table.

		Observed by radar		Total
		Class (i)	No class (i)	
Identified by satellite method	Class (i)	<i>a</i>	<i>b</i>	<i>a + b</i>
	No class (i)	<i>c</i>	<i>d</i>	<i>c + d</i>
	Total	<i>a + c</i>	<i>b + d</i>	<i>a + b + c + d = n</i>

Table 5. Results of evaluation parameters.

	POD (%)	POFD (%)	FAR (%)	Bias	CSI (%)	PC (%)
Strongly convective	83	09	22	1.1	73	97
Moderately convective	74	12	29	1.5	66	93
Slightly convective	65	19	32	0.6	63	91
Total	75	19	29	1.2	66	92
Optimal values	100	00	00	1	100	100

good values of PC (91%). Therefore, from these statistics, it is concluded that the scheme has skill for detection of slightly convective precipitation area.

Concerning the moderately convective precipitation area, the percentage detected by the satellite-based technique is also in good agreement with those identified in the radar data (table 5). Overall, the detected area is slightly overestimated by the developed scheme (bias: 1.5). The POD value may be considered satisfactory due to the small fraction of this moderately convective over the whole area of rain. On average, 74% of the pixels classified as moderately convective precipitation by the radar are consistently identified by the satellite scheme. The very low percentage of wrongly classified pixels indicated by the POFD underlines this good impression. At the same time, the FAR average value is relatively moderate (29%). As for slightly convective, the CSI (66%) indicates a satisfying degree of correctly classified pixels and the good values of the PC (93%) further support the positive impression concerning the classified area of moderately convective precipitation. According to the statistical results, the detection of moderately convective precipitation indicates a slightly better performance compared to the class slightly convective.

The area percentage of the strongly convective with high rainfall intensities classified by the proposed scheme is almost consistent with those detected by the radar (table 5). The good POD (83%) value is accompanied by a very low POFD (9%) value and the low FAR (22%) is even lower than for the moderately convective precipitation area. This classification of strongly convective area reveals a positive performance compared to other classes. Indeed, the classification of strongly convective precipitation area is almost unbiased. The

CSI (77%) indicates a good degree of correctly classified pixels and the overall positive impression for the classified strongly convective precipitation area is underlined by the good values of the PC (97%). The calculated parameters point to a slightly better performance scheme regarding the classified convective cores.

Overall, these results show good performance for convective rainfall intensity differentiation. The joint combination of information about the cloud physical properties retrieved from the various channels of SEVIRI has better classified convective rainfall intensities.

3.2.1 Rainfall estimation and intercomparison

To gain an impression of the performance of the developed scheme, estimations of precipitation from convective rainfall intensities classified using NBC are carried out. Several parameters of evaluation, such as; the correlation coefficient (*r*), the root mean square difference (RMSD), the bias, mean absolute difference (MAD) and the percent difference (PD) between the radar measurements and presented technique are calculated. They are given by:

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N (E_i - V_i) \quad (14)$$

$$\text{MAD} = \frac{1}{N} \sum_{i=1}^N |E_i - V_i| \quad (15)$$

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N (E_i - V_i)^2} \quad (16)$$

$$\text{PD} = \frac{1}{N} \sum_{i=1}^N \frac{|E_i - V_i|}{V_i} \quad (17)$$

where E_i is the estimated value and V_i is the validation values. The diagrams of dispersion for the two methods are given in figure 3 with associated statistics in table 6.

Figure 3. Comparison results between satellite estimates and radar measurements.

Table 6. *Bias, RMSD, PD, MAD and correlation coefficient (r) between satellite estimates and radar measurements.*

	Mean(mm)		RMSD (mm)	MAD (mm)	Bias (mm)	PD	r
	Satellite	Radar					
Rainfall April to October 2013	80.56	84.57	20.12	13.22	−4.01	3.13	0.83

The results are very interesting. Indeed, compared to radar data, the RMSD, MAD and PD indicate values 20.12 mm, 13.22 mm and 3.13, respectively. A slight underestimation is shown by the Bias (−4.01 mm). A good correlation is showed by the correlation coefficient of 0.83.

The performance of the developed scheme is also assessed with respect to EPSAT-SG (Estimation of Precipitation by SATellites Second Generation), GPI (GOES Precipitation Index) presented by Jobard *et al.* (2011) and Cold Cloud Duration (CCD) technique used by the TAMSAT (Tropical Applications in Meteorology using satellite data) Group at the University of Reading (Milford and Dugdale 1989). The presented technique is also compared to CCPD (Cold Cloud Phase Duration) developed by Lazri *et al.* (2014a, b). It should be noted that these methods are designed and applied to convective precipitation.

The EPSAT-SG technique consists of combining a rainfall probability and a rainfall potential intensity to derive the final precipitation estimate. The rainfall probabilities are provided by merging

the SEVIRI multispectral channels and Tropical Rainfall Measuring Mission (TRMM) precipitation radar (PR) information, using a feed forward neural network. The algorithm is described in Berges *et al.* (2010).

The GPI method has been described in Arkin and Meisner (1987). The GPI technique estimates tropical rainfall using cloud top temperature as the sole predictor. The GPI rainfall estimates R (mm) are obtained by multiplying the fractional coverage (FC) of cold clouds by period P and rainfall rate RR . The fractional coverage of cold clouds is determined as the percentage of satellite pixels in a reasonably large grid box, with temperature below the 235 K brightness temperature threshold; the FCs are averaged during the observation period P evaluated in hours, and the constant precipitation rate RR is 3 mm/h.

The CCD technique attempts to link the life duration of convective cloud to rainfall recorded on the ground to establish a relationship. It assumes that there is a linear relationship between the length of time that convective clouds are present

Table 7. *Statistical intercomparison.*

	BIAS (mm)	RMSD (mm)	Mean (mm)	Correlation	Spatial resolution
GPI	45–47	25–32	84.57	0.65–0.70	1.0°
CCD	–7.09	35.96	84.57	0.66	Pixel
EPSAT-SG	8.5–11	29–31	84.57	0.82–0.85	0.1°
CCPD	–3.49	30.27	84.57	0.83	Pixel
Presented technique	–4.01	20.12	84.57	0.83	Pixel

and the amount of rain that falls. The method uses the brightness temperature to select a temperature threshold to identify regions of convection.

For CCPD method, the life cycle phase of a given convective cloud (growth–decay) is evaluated through the different internal dynamics of cloud. The CCPD analyses the evolution of three parameters; namely the average of cloud top temperature, the vertical extent of cloud and cloud water path to identify different phases of life cycle of convective clouds. Two precipitation rates are affected by the growth phase and decay phase respectively.

Table 7 shows the statistical intercomparison results. These statistics have been recomputed for a mean rain of 84.57 mm to compare with the presented technique which recorded a mean rain of 84.57 mm.

Indicated by the value of correlation coefficient (0.83), the developed scheme shows a good correlation which is similar to EPSAT-SG and CCPD. In contrast, the developed technique performs better than CCD and GPI, which presents value of correlation coefficient between 0.65 and 0.70.

In terms of RMSD, the developed scheme seems to perform better than the GPI EPSAT-SG, CCD and CCPD. Indeed, the RMSD shows a value of 20.12 mm against values in the range 25–31 mm for others methods. But the presented technique shows a slight underestimation than CCPD. This is indicated by values of Bias of –3.49 for CCPD method and –4.01 for the developed scheme. Unlike EPSAT-SG method that indicates a slight overestimate the GPI method shows a significant overestimation.

In the whole, compared to developed scheme, methods consistently showed larger errors. In terms of bias, RMSD and correlation coefficient, the presented technique performs better.

4. Conclusion

A new consistent day and night technique based on NCB is proposed using cloud physical properties. The aim of the present study is to investigate the potential of MSG SEVIRI for improved convective rainfall intensity differentiation.

The present study demonstrated enormous potential of increased spectral resolution of MSG/SEVIRI. Indeed, the integration of multispectral information on the physical properties of cloud improved the differentiation of the intensity of convective precipitation. The detection of the strongly convective precipitation area relies on information about the CTH gained from BT_{10.8} and BT_{D3}. It is expected that cloud areas with a higher BT_{D3} and a lower top temperature are characterised by higher rainfall intensities. The identification of slightly and moderately convective precipitation area is based on information about the CWP and the CP. It is assumed that raining areas characterised by higher CWP values and a higher amount of ice particles in the upper parts of the cloud produce higher rainfall intensities. Information about the CP is gained by considering the temperature differences BT_{D1} and BT_{D2}. Information about the CWP is incorporated by considering the VIS0.6 and NIR1.6 channels during daytime and the brightness temperature differences BT_{D3.9–10.8}, BT_{D3.9–7.3}, BT_{D2} and BT_{D3} during night-time.

The evaluation of the technique's performance is carried out by comparing with radar data. Indeed, the results obtained are interesting.

The performance of the developed scheme is also assessed with respect to EPSAT-SG, GPI, CCD and CCPD methods. The proposed technique is more efficient than these techniques, especially, the techniques based only on information about the cloud top temperature. The algorithm shows encouraging performance for convective rainfall intensity differentiation and estimation.

There are important methodological implications from this work. The convective rainfall estimation technique could be improved by exploiting this study. In addition, the developed scheme could be generalised for other regions that lack rain gauge measurements.

References

- Adler R F and Negri A J 1988 A satellite infrared technique to estimate tropical convective and stratiform rainfall; *J. Appl. Meteor.* **27** 30–51.
- Arkin P A and Meisner B N 1987 The relationship between large-scale convective rainfall and cold cloud over the

- western hemisphere during 1982–1984; *Mon. Wea. Rev.* **115** 51–74.
- Baum B A and Platnick S 2006 Introduction to MODIS cloud products; In: *Earth science satellite remote sensing: Science and instruments* (eds) Qu J J, Gao W, Kafatos M, Murphy R E and Salomonson V V (Springer, New York), 78p.
- Berges J C, Chopin F, Jobard I and Roca R 2010 EPSAT-SG: A satellite method for precipitation estimation: Its concept and implementation for AMMA experiment; *Ann. Geophys.* **28** 289–308.
- Fritz S and Laszlo I 1993 Detection of water vapour in the stratosphere over very high clouds in the tropics; *J. Geophys. Res.* **98**(D12) 22,959–22,967.
- Hall M A and Holmes G 2003 Benchmarking attribute selection techniques for discrete class data mining; *IEEE Transactions on Knowledge and Data Engineering* **15** 1437–1447.
- Hanson R, Stutz J and Cheeseman P 1991 Bayesian Classification Theory; NASA, Technical Report FIA-90-12-7-01.
- Houze 1993 *Cloud Dynamics*; Academic Press, 573p.
- Huffman G J, Adler R F, Morrissey M M, Bolvin D T, Curtis S, Joyce R, McGavock B and Susskind J 2001 Global precipitation at one-degree daily resolution from multisatellite observations; *J. Hydrometeor.* **2** 36–50.
- Inoue T 1987 A cloud type classification with NOAA-7 split window measurements; *J. Geophys. Res.* **92** 3991–4000.
- Jobard I, Chopin F, Berges J C and Roca R 2011 An inter-comparison of 10-day satellite precipitation products during West African monsoon; *Int. J. Remote Sens.* **32**(9) 2353–2376.
- Kühnlein M, Appelhans T, Thies B and Nauss T 2014 Improving the accuracy of rainfall rates from optical satellite sensors with machine learning – A random forest-based approach applied to MSG SEVIRI; *Remote Sens. Environ.* **141** 129–143.
- Lazri M, Ouallouche F, Ameur S, Brucker J M and Mohia Y 2012 Identifying convective and stratiform rain by confronting SEVIRI sensor multispectral infrared to radar sensor data using neural network; *Sensors & Transducers Journal* **145**(10) 19.
- Lazri M, Ameur Z, Ameur S, Mohia Y, Brucker J M and Testud J 2013a Rainfall estimation over a Mediterranean region using a method based on various spectral parameters of SEVIRI-MSG; *Adv. Space Res.* **52** 1450–1466.
- Lazri M, Ameur S, Brucker J M, Testud J, Hamadache B, Hameg S, Ouallouche F and Mohia Y 2013b Identification of raining clouds using a method based on optical and microphysical cloud properties from Meteosat second generation daytime and nighttime data; *Appl. Water. Sci.* **3** 1–11.
- Lazri M, Ameur S and Mohia Y 2014a Instantaneous rainfall estimation using neural network from multispectral observations of SEVIRI radiometer and its application in estimation of daily and monthly rainfall; *Adv. Space Res.* **53**(1) 138–155.
- Lazri M, Ameur S, Brucker J M and Ouallouche F 2014b Convective rainfall estimation from MSG/SEVIRI data based on different development phase duration of convective systems (growth phase and decay phase); *Atmos. Res.* **147–148** 38–50.
- Lazri M, Ameur S, Brucker J M, Lahdir M and Sehad M 2015 Analysis of drought areas in northern Algeria using Markov chains; *J. Earth Syst. Sci.* **124** 61–70.
- Milford J R and Dugdale G 1989 Estimation of rainfall using geostationary satellite data; In: *Application of remote sensing in agriculture, Proceedings of 4Sth Easter School in Agricultural Science*, University of Nottingham, July 1989, Butterworth, London, 16p.
- Nauss T and Kokhanovsky A A 2006 Discriminating raining from non-raining clouds at mid-latitudes using multispectral satellite data; *Atmos. Chem. Phys.* **6** 5031–5036.
- Polat K and Gunes S 2006 Automated identification of diseases related to lymph system from lymphography data using artificial immune recognition system with fuzzy resource allocation mechanism (fuzzy-AIRS); *Biomedical Signal Processing and Control* **1** 253–260.
- Roebeling R A, Feijt A J and Stammes P 2006 Cloud property retrievals for climate monitoring: Implications of differences between SEVIRI on METEOSAT-8 and AVHRR on NOAA-17; *J. Geophys. Res.* **11** D20210.
- Roebeling R A, Deneke H M and Feijt A J 2008 Validation of cloud liquid water path retrievals from SEVIRI using one year of CloudNET observations; *J. Appl. Meteor. Climatol.* **47** 206–222.
- Schmetz J, Tjemkes S A, Gube M and van de Berg L 1997 Monitoring deep convection and convective overshooting with Meteosat; *Adv. Space Res.* **19** 433–441.
- Strabala K I, Ackerman S A and Menzel W P 1994 Cloud properties inferred from 8–12- μ m data; *J. Appl. Meteorol.* **33** 212–229.
- Thies B, Nauss T and Bendix J 2008a Discriminating raining from non-raining clouds at mid-latitudes using Meteosat second generation daytime data; *Atmos. Chem. Phys.* **8** 2341–2349.
- Thies B, Nauss T and Bendix J 2008b Delineation of raining from non-raining clouds during nighttime using Meteosat-8 data; *Meteorol. Appl.* **15** 219–230.
- Thies B, Turek A, Nauss T and Bendix B 2010 Weather type dependent quality assessment of a satellite-based rainfall detection scheme for the mid-latitudes; *Meteorol. Atmos. Phys.* **107** 81–89.

MS received 16 June 2015; revised 27 December 2015; accepted 4 April 2016

Corresponding editor: PRADEEP KUMAR THAPLIYAL