

Analysis of the Relationship between Banded Orographic Convection and Atmospheric Properties Using Factorial Discriminant Analysis and Neural Networks

A. GODART

*Laboratoire d'étude des Transferts en Hydrologie et Environnement, Université de Grenoble
(CNRS, UJF, IRD, INPG), Grenoble, France*

E. LEBLOIS

Cemagref, UR Hydrologie-Hydraulique, Lyon, France

S. ANQUETIN

*Laboratoire d'étude des Transferts en Hydrologie et Environnement, Université de Grenoble
(CNRS, UJF, IRD, INPG), Grenoble, France*

N. FREYCHET

*Laboratoire des Ecoulements Géophysiques et Industriels, Université de Grenoble
(UJF, CNRS, INPG), Grenoble, France*

(Manuscript received 10 March 2009, in final form 25 September 2009)

ABSTRACT

The relationship between banded orographic convection and atmospheric properties is investigated for a region in the south of France where the associated rainfall events are thought to represent a significant portion of the hydrologic input. The purpose is to develop a method capable of producing an extensive database of banded orographic convection rainfall events from atmospheric sounding data for this region where insufficient rain gauge data and little or no suitable radar or satellite data are available. Two statistical methods—discriminant factorial analysis (DFA) and neural networks (NNs)—are used to determine 16 so-called elaborated nonlinear variables that best identify rainfall events related to banded orographic convection from atmospheric soundings. The approach takes rainfall information into account indirectly because it “learns” from the results of a previous study that explored meteorological and available rainfall databases, even if incomplete. The new variables include wind shear, low-level moisture fluxes, and gradients of the potential temperature in the lower layers of the atmosphere, and they were used to create an extensive database of banded orographic convection events from the archive of atmospheric soundings. Results of numerical simulations using the nonhydrostatic mesoscale (Méso-NH) meteorological model validate this approach and offer interesting perspectives for the understanding of the physical processes associated with banded orographic convection. DFA proves to be useful to determine the most discriminant factors with a physical meaning. Neural networks provide better results, but they do not allow for physical interpretation. The best solution is therefore to use the two methods together.

1. Introduction

The Mediterranean Sea basin is considered to be one of the regions of the world that is most vulnerable to

climate change. Many water-related environmental issues will be affected by possible changes in the year-to-year variability of precipitation and/or changes in small-scale rainfall properties. Before studying the effects of climate change, it is necessary to better understand the processes leading to rainfall variability, particularly in mountainous areas. Orography is a major factor in water resource distribution at global and regional scales (Banta 1990;

Corresponding author address: Sandrine Anquetin, LTHE, BP 53 X, 38041 Grenoble CEDEX 09, France.
E-mail: sandrine.anquetin@hmg.inpg.fr

Barros and Lettenmaier 1994). Relief deviates atmospheric flows, resulting in a wide range of mesoscale processes (Smith 1979; Houze et al. 1993).

In a climatological study of precipitation in the Alps, Frei and Schär (1998) showed that climatic rainfall anomalies often occur in the French Mediterranean region in autumn, producing very heavy rainfalls. Moreover, Bois et al. (1997) mapped extreme rainfalls in the French Mediterranean region over different accumulation periods (hourly to daily). The extreme rainfalls were estimated for a 100-yr return period and were obtained by means of a long-term hourly rain gauge database. Results showed very different rainfall patterns at different locations. On the daily time scale, the accumulated rainfall depended significantly on relief, whereas no significant relationship was observed for the hourly time step. On the daily time scale, intermittency was lower in the mountainous area than on the hourly time scale. These results suggest that convection properties explain the variability of rainfall in space and time.

The present study focuses on rainfall events related to banded orographic convection, thought to represent a significant portion of the hydrologic input to the French Mediterranean region. Based on Anatol S-band radar observations (Pointin et al. 1988) collected in the mountainous area of the Cévennes region for the 1986–1988 field experiment, Miniscloux et al. (2001) demonstrated the prevailing location of shallow convection organized in bands over the Massif Central. The term “band” is used according to the definition in the American Meteorological Society’s *Glossary of Meteorology* [quoted by Fuhrer and Schär (2007)] and refers to rainfall patterns made up of a number of narrow and elongated areas (length-to-width ratio equal to at least 4), in which high accumulated rainfall areas alternate with dry areas. In the region studied here, the rainfall intensity is moderate, less than 10 mm h^{-1} , but accumulated rainfall amounts can be large because the rainbands can last a long time when associated with stationary flow. Moreover, synoptic-scale steadiness fixes the location of the rainbands and yields high cumulative precipitation. The bands are oriented parallel to the flow (Gysi 1998; Anquetin et al. 2003) and are associated with shallow convection. The vertical extension of the clouds does not exceed 6 km, as deduced from the vertical profiles of reflectivity estimated with an inverse method by Andrieu and Creutin (1995). Miniscloux et al. (2001) also pointed out the correlation between the structure of the observed rain pattern and the valley width. Many theoretical studies, essentially based on simulations, have been carried out to determine the spatial and temporal characteristics of such rainbands and to identify the causes (Kirshbaum and Durran 2004, 2005a). Numerical

studies have also shown the role of small-scale topographic obstacles in triggering and anchoring the rainbands [Cosma et al. (2002) for the southeastern part of France; Colle (2004) and Kirshbaum and Durran (2005b) for Oregon].

From a climatological point of view, the distribution of rain accumulation over long periods is clearly linked to this type of banded orographic convection. In a recent paper, Godart et al. (2009) produced an initial database of these events based on the joint use of atmospheric and rainfall criteria to identify the events. Atmospheric criteria were applied to soundings from Nîmes, whereas rainfall criteria were applied to the hourly rainfall database. However, few events were identified because of the incomplete hourly rainfall database. It was therefore necessary to develop a different identification strategy to extend the database of banded orographic convection events. This is the aim of the present study that makes use of the results of Godart et al. (2009) and thus indirectly makes use of the information from available rain gauge data. Rainfall fields are considered as a random function conditioned by atmospheric flow characteristics. They are partially explained by information contained in soundings. Analysis of the initial database from Godart et al. (2009) made it possible to identify some of the main properties of soundings associated with banded orographic convection. The present study proposes to refine these properties using linear and nonlinear statistical concepts. Statistical descriptors of banded orographic convective rainfall are thus extracted from the soundings. The main objectives of this paper are 1) to determine the capability of statistical methods to identify rainfall events associated with banded orographic convection on the basis of atmospheric soundings and 2) to statistically select the atmospheric variables that best characterize the banded orographic convection events. The results will help improve our understanding of the atmospheric processes involved and to extend the initial database of banded orographic convection rainfall events produced by Godart et al. (2009). In addition, the results will be used to help design an observation system for the Hydrological Cycle in Mediterranean Experiment (HyMeX) field experiment (available online at www.hymex.org) that will enhance our understanding of the processes associated with banded orographic convection.

To deal with the nonlinearity of the relationship between precipitation and atmospheric fields, linear and nonlinear statistical tools are considered using simple or nonlinearly elaborated variables. The linear statistic tool is discriminant factorial analysis (DFA), whereas the nonlinear statistic tool is a multilayer perceptron neural network (NN). Discriminant factorial analysis was used by Ghosh et al. (2004) for the classification of

thunderstorm and nonthunderstorm days in Calcutta and recently by Mercer et al. (2008) for the statistical modeling of downslope windstorms in Boulder, Colorado. Neural networks have been widely used (Lau and Widrow 1990; Elsner and Tsonis 1992; Masters 1993; Ripley 1996; Bishop 1995). Gardner and Dorling (1998), Marzban (2000, notes from Short Course on Artificial Intelligence Methods in Atmospheric and Oceanic Sciences: Neural Networks, Fuzzy Logic, and Genetic Algorithms, AMS Annual Meeting, Seattle, Washington, 10–11 January 2004, available online at <http://www.nhn.ou.edu/~marzban>), and Marzban and Witt (2001) have reviewed their applications in the atmospheric sciences. Manzato (2005) used a NN and DFA with soundings to provide inputs to forecast short-term thunderstorms. Comparison between the two methods suggests that a NN usually provides better results. Classification and pattern recognition are other fields of atmospheric sciences in which the application of neural networks has provided positive results. Examples include cloud classification by Bankert (1994), Peak and Tag (1994), and Pankiewicz (1995), Atlantic blocking detection (Verdecchia et al. 1996), and thunderstorm classification (Manzato 2005). These applications show that neural networks are a valuable approach for extracting patterns from noisy data. Advantages and disadvantages in comparison with other statistical techniques are discussed in Marzban and Stumpf (1996). Despite the widespread use of neural networks, few comparisons with standard linear techniques have been published. Gardner and Dorling (1999), McGinnis (1994), Navone and Ceccatto (1994), and Weichert and Bürger (1998) have shown that neural networks usually give better results. Thus, one of the objectives of the present study is also to propose a comparison between the performance of discriminant factorial analysis and neural networks as classifiers of banded convection events. It will be discussed why the two methods should be used jointly.

The datasets and methods are detailed in section 2. Section 3 deals with the search for the most discriminating variables and comparison of the methods. Section 4 presents studies concerning the sensitivity of discriminant analysis and neural networks to input variables and sampling. Section 5 indicates the relationship between banded orographic convection and atmospheric properties and presents the final classification of rainfall events. Section 6 presents the conclusions and perspectives.

2. Datasets and methods used

a. Region of interest and databases used

The Cévennes region (Fig. 1) is located southeast of the Massif Central mountains in France. The relief is a

southeasterly facing slope delimited by the Mediterranean shore and the Rhône valley. The elevation varies from sea level up to 1500 m over roughly 30 km. This hilly mass is dissected by deep and narrow valleys (500 m deep, 10 km wide) oriented northwest–southeast (Fig. 1). According to elevation, three sectors can be identified: 1) a lower terrace (elevation below 200 m) called the “plain”; 2) a hilly sector (elevation between 200 and 500 m) called the “piedmont”; and 3) a mountainous sector (elevation above 500 m) called the “mountains” (Fig. 1).

Atmospheric properties come from the database of soundings from Nîmes (Fig. 1) at 0000 and 1200 UTC between 1976 and 2005 and were retrieved from the University of Wyoming Web site (available at <http://weather.uwyo.edu/>). These data were then used as inputs to the DFA and NN. Godart et al. (2009) proposed a methodology to identify banded orographic convection events using the sounding database and an hourly rain gauge dataset (see Fig. 1 for the locations of the rain gauges). The identification process was based on what were referred to as 1) dynamic criteria (from soundings) and 2) rainfall criteria. The dynamic criteria are 1) a southerly wind at the sounding station of Nîmes; 2) a mean wind speed between ground and 800 hPa that exceeds 7 m s^{-1} ; 3) directional shear below 800 hPa corresponding to a rotation of the wind vector with elevation of less than 15° km^{-1} . The rainfall criteria are based only on hourly rainfall data and are as follows: 1) the mean rainfall intensity within the mountainous area must be higher than in the plains region and 2) the intermittency (percentage of zeros) must decrease with elevation. Among the 21 944 available soundings, only 880 satisfy the three dynamic criteria. Hourly rainfall data are only available for 350 of them. Using a daily rainfall archive, 59 of the remaining 530 ($880 - 350$) soundings can be identified as dry days and can thus be eliminated. No hourly rainfall information is available for the 471 remaining days. The aim of this study is to identify soundings corresponding to banded orographic convection among all 880 soundings without the use of the hourly rainfall database.

To implement the statistical classifications, a sample of known soundings is required. The results of Godart et al. (2009) show that, among the 350 soundings satisfying the dynamic criteria and for which hourly rainfall data were available, only 121 soundings, subsequently referred as group 1, satisfied the rainfall criteria for “banded orographic convection.” The remaining 229 soundings, subsequently referred as group 2, constitute the “other rainy events.”

The first step is therefore dedicated to the calibration of the statistical tools used to discriminate between these two known groups. Improved discriminant atmospheric

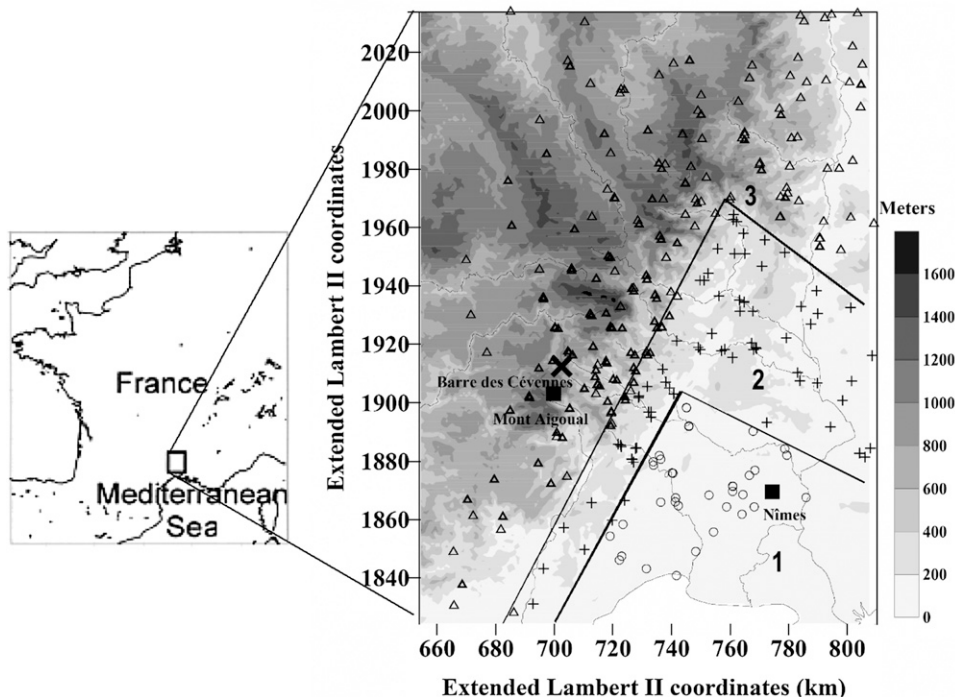


FIG. 1. (left) General location of the Cévennes–Vivarais region in France. (right) Relief of the studied area. The digital elevation model (DEM) resolution is 75 m. The two weather stations are Mont Aigoual and Nîmes (filled squares). The three topographic sectors are numbered 1 (plain), 2 (piedmont), and 3 (mountains). The rain gauges for the three sectors are labeled with circles (plain), plus signs (piedmont), and triangles (mountains). The radar location is indicated by the X at Barre des Cévennes.

variables need then to be identified. Next, the statistical methods are used to classify the whole sounding archive. Only the soundings that fit the dynamic criteria [880, as pointed out in Godart et al. (2009)] will be classified.

b. Statistical approaches used

1) DFA

Discriminant factorial analysis is a classical method for 1) detecting variables that discriminate between k groups and 2) classifying cases into groups with accuracy better than a random classification. In the first case, the goal is to identify new variables that are more effective at separating the k groups. These new variables are linear combinations of the original variables and are called either “discriminant axes” or “discriminant functions.” They explain the largest differences between the group means. The principle consists in finding the $(k - 1)$ borders that best separate the gravity centers of each group by optimizing the grouping around the gravity center. This is done by maximizing the interclass inertia while minimizing the intraclass variability. Any unknown individual is assigned to the group where the Mahalanobis distance between the gravity center and the individual coordinates is minimized.

The correlation coefficient between the variables and the discriminant axis is the basis for the classification. The variables presenting the highest correlation coefficient (positive or negative) are interpreted as the most discriminant variables for the groups.

2) THE MULTILAYER PERCEPTRON NEURAL NETWORK

The multilayer perceptron, a kind of NN, has the same objectives as the DFA. The difference is in the use of nonlinear combinations of input variables to classify individuals into several groups. A multilayer perceptron is a very standard tool that is documented in any textbook on neuromimetic algorithms (Bennani 2006). The perceptron is a set of three layers of neurons—input, hidden, and output—as presented in Fig. 2. Values of the input neurons are values of the atmospheric variables. Values of the hidden neurons correspond then to the sum of weighted connections from the input neurons followed by a nonlinear yet derivable transformation of the inputs received with a transfer function σ [Fig. 2; in this paper, the term “logistic function” (Manzato 2005) is used]. The same procedure is used to produce inputs for the output neurons, which perform a linear transformation of their inputs v_k , and the result will be

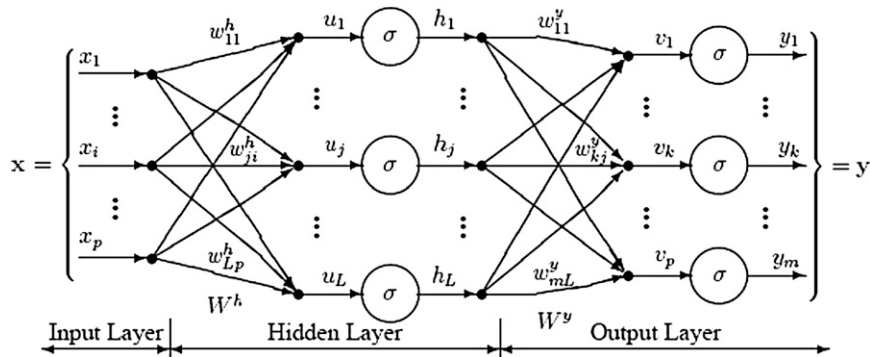


FIG. 2. Example of a multilayer perceptron structure composed of the input, hidden, and output layers. Each neuron in each layer is connected to neurons in the next layer by weighed connections (w_{ij}).

considered as the final output for a given situation noted y_k in Fig. 2. The number of output neurons depends on the number of groups in the case of classification studies (in our case, two output neurons describe the two groups). The backpropagation algorithm (Rumelhart et al. 1986a) is used to find the best weighted connections, which make it possible to fit the initial classification during the training phase (individuals whose groups are known are specified). In this study, the backpropagation algorithm is stopped after 2 000 000 iterations. The backpropagation correction step in the neural network appears to be very similar to the adjunct model, in the sense of data assimilation activities in atmospheric sciences.

The architecture (i.e., the number of neurons and the nonlinear function used) of the NN is highly problem dependent. Usually, optimality is defined as the smallest network that adequately captures the relationship in the training data (Stanley 1988; David Shepard Associates 1990; Caudill 1991; Fletcher and Goss 1993; Cheng and Titterington 1994; Ripley 1994; Wierenga and Kluytmans 1994; Venugopal and Baets 1994). Maier and Dandy (2000) and Manzato (2005) reviewed modeling issues and steps that should be followed in the development of the NN. In most studies and in particular the present one, the number of hidden layers is chosen to be one. A small number of neurons in the hidden layer limits the success of the training phase (Stanley 1988), whereas numerous neurons lead to overfitting (Fig. 3). Any additional neuron can improve data fitting but possibly without any statistical sense. With overfitting, the NN accurately reproduces the noise present in the data used for training, but it cannot be used for the validation phase (Fig. 3).

In this study, three architectures are tested using three different numbers of neurons in the hidden layer. This number ranges from the arithmetic to geometric means of the number of input and output neurons. Since the

number of output neurons is fixed at two, for each set of inputs, the retained architecture will be the one that gives the best performances. The number of input neurons corresponds to the number of atmospheric variables used to discriminate between the two groups. The validation phase assesses the ability of the NN to be generalized to untrained data (Weigend et al. 1990; Masters 1993; Cheng and Titterington 1994; Rogers and Dowla 1994; Amari et al. 1997). This ability will be affected by the variability of data within the training and validation samples (Flood and Kartam 1994; Minns and Hall 1996) and by the criterion used to stop the process. It is therefore required to run the NN on different combinations of training and validation samples.

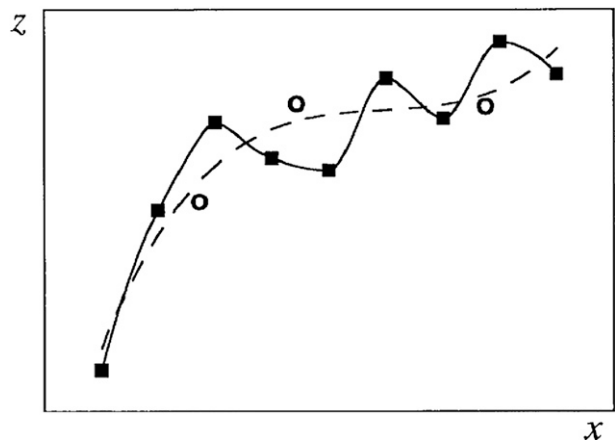


FIG. 3. Illustration of overfitting. Solid squares represent data used for training, and empty circles represent data used for validation. The dashed line shows good model training that filters out the noise and allows for effective validation. The thick line shows a typical case of overfitting: the model has perfectly learned the data during the training phase but is unable to approach data in the validation phase.

c. Sampling of data

To train the DFA and the NN, two sampling approaches are considered and compared.

In the first sampling approach, 80% of the 121 group 1 soundings and 80% of the 229 group 2 soundings are randomly selected for training. Validation is then carried out on the remaining 20% of soundings in the two groups.

In the second sampling approach, the same number of soundings is considered in each group. Indeed in DFA, the sampling plays a major role in the determination of the mean of the different groups and the dispersion matrix. Generally, it is better to have the same number of data in each group (Cacoullos 1973). Moreover, Bennani (2006) explained that the NN training depends on the number of data in each group. Since 121 soundings belong to group 1, 121 soundings among the 229 of group 2 are randomly selected. These two groups of 121 soundings constitute the initial sample. Then, as was done previously, 80% of the soundings are used for the training phase and the remaining 20% for the validation phase.

Note that DFA and NN performances are strongly dependent on the sample. For this reason, 500 samplings are considered. Average performance and standard deviations are then used in the analysis.

3. Determination of the discriminant atmospheric variables and comparison of the two statistical methods

A preliminary analysis of the 350 soundings, based on a Student t test of the mean and standard deviation, showed (Godart 2009) significant differences between the two groups for a few variables at a given level chosen arbitrarily or integrated over different layers of the atmosphere. These included low-level moisture fluxes, wind speed and direction, relative humidity, precipitable water, and wind shear. This result is encouraging but does not demonstrate a correlation between variables. This first analysis suggests that there are differences between the two groups and confirms the possibility of finding discriminant variables built on a combination of these first variables.

This section presents the method used to select a few variables to discriminate banded orographic convection events with the two statistical methods (DFA and NN) using the first sampling approach. A comparison of these two methods is also proposed.

a. Presentation of the results

The evaluations of the training and validation phases are based on the hit rate score (H , Eq. 1), the probability of detection [POD, Eq. (2)], and the specificity [SPE,

TABLE 1. Contingency table.

		Obs	
		Group 1	Group 2
Classification	Group 1	a	b
	Group 2	c	d

Eq. (3)]. The hit rate score gives the percentage of soundings that are assigned to their expected groups. The POD (SPE) defines the percentage of soundings belonging to group 1 (group 2) and assigned to group 1 (group 2) by the statistical method. Equations are given below and are computed using the contingency table (Table 1; Doswell III et al. 1990),

$$H = \frac{a + d}{n} \times 100 \quad \text{with } n = a + b + c + d, \quad (1)$$

$$\text{POD} = \frac{a}{(a + c)} \times 100, \quad \text{and} \quad (2)$$

$$\text{SPE} = \frac{d}{(b + d)} \times 100. \quad (3)$$

Then, the results of the statistical tools are checked to see if they provide an improvement over the classification given by random assignment, which is completed using the Heidke skill score [originally Doolittle (1888)'s association ratio]. Since the DFA and NN are run on 500 different samplings for each test, only the mean values and their associated standard deviations are discussed.

b. Use of the DFA and the NN with basic thermodynamic variables as inputs

In this section, we consider the extent to which basic variables derived from soundings are sufficient to explain the deterministic part of the relationship between the soundings and the rainfall field. Geopotential height, temperature, relative humidity, and wind speed, subsequently referred to as ZTUVp, are the all projected in the south–north direction every 50 hPa between 950 and 400 hPa. There are thus 48 input variables for the DFA and the NN.

The result for the DFA during the training phase is presented in Fig. 4 for one sampling (among the 500 samplings). Because two groups are considered, there is only one discriminating axis. The lines above the symbols are another representation of the results where a probability density function (PDF) for the location of sounding is assumed using a kernel-based density estimation. It is built by summing each elementary Gaussian function of each individual weighted by $1/N$. The standard deviation is common to all elementary Gaussian functions and tuned to make the PDF easy to read. This

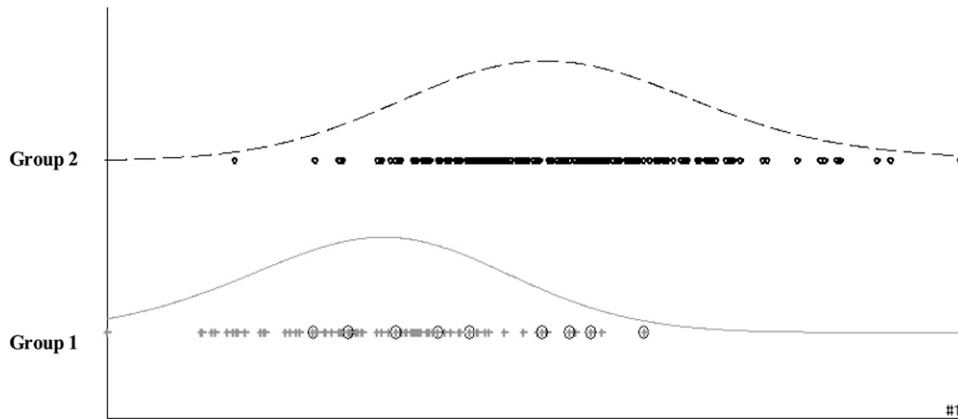


FIG. 4. Results of DFA with ZTUVp variables as inputs. Crosses and the solid line correspond to the “banded orographic events”; circles and the dashed line correspond to the “other events.” The symbols are coordinates of the soundings on the discriminant axis and the lines are the PDF. Circled symbols correspond to soundings not well classified. The vertical axis has no physical or statistical significance; it is only used to individually represent the two groups.

is more a graphical representation than a real statistical model. The scores and their associated standard deviations are given in Table 2 for the training and validation phases. Note that the DFA performance falls during the validation phase, making this approach inadequate to classify the soundings in Group 1. The value of Heidke’s score indicates that the DFA method performs like a random classification.

The NN is used with 17 neurons in the hidden layer. The mean scores are given in Table 2. The hit rate scores are lower and the POD is higher during the validation phase than for the DFA. Overall, the NN gives better scores than the DFA. It is able to consider variables elaborated from the basic thermodynamic inputs, unlike the DFA. Rainfall processes are highly nonlinear, and the combination of the nonlinear transfer function and the hidden layer enables the NN to approximate the relationships between predictors and predictands more precisely than a linear method. Linear methods can be adapted to find nonlinear relationships by using nonlinear combinations of real variables (Kuligowski and Barros 1998). However, these adaptations require extensive trial-and-error selection on the part of the user, whereas the neural network requires much less subjective intervention to find these relationships.

Results obtained with the NN are encouraging in the search for variables explaining the deterministic part of the relationship between soundings and rainfall fields. However, interpretation remains very difficult. No conclusion is possible as far as the most discriminative variables are concerned, because we do not know how NN intrinsically elaborates its own nonlinear variables. The DFA results are also difficult to interpret in our case because of the large number of input variables. The projection of

the individuals in the factorial space shows no significant correlation between the variables and the discriminative axis. A search on elaborated variables is thus necessary to make the results interpretable.

c. Search for the best discriminative elaborated variables

Because of the numerous possible combinations of input variables and the long computational time required to run the NN, the analysis is first completed with the DFA. This analysis relies on the assumption that the NN will give at least similar results since elaborated variables are used as inputs (Weichert and Bürger 1998). Once the most discriminative set of variables is found, the NN is used with these inputs and both results are compared. On the basis of previous studies dealing with banded orographic rainfall (Lin et al. 2001; Kirshbaum et al. 2007a,b; Godart et al. 2009), we first select some known discriminative thermodynamic variables.

Banded orographic convection is associated with large values of relative humidity (>80%) in the lowest layers

TABLE 2. Mean scores (%) obtained during the training and validation phases with the DFA and the NN using the ZTUVp variables as inputs. The values in parentheses are the standard deviations.

	Hit rate	POD	SPE	Heidke’s score
DFA				
Training	95.1 (0.9)	86.2 (2.6)	99.8 (0.3)	
Validation	65.6 (2.4)	5.5 (4.7)	96.6 (2.8)	0
NN				
Training	70.8 (18.6)	66.7 (36.3)	73 (36.5)	
Validation	60.4 (12.9)	52 (34.8)	64.8 (34.4)	0.158

of the atmosphere (<5000 m). Its vertical profile presents a rapid decrease in the upper atmosphere. Several variables representing humidity may be derived from the soundings: precipitable water, potential precipitable water, (i.e., precipitable water multiplied by the relative humidity), moisture flux, and potential moisture flux (i.e., moisture flux multiplied by the relative humidity). The potential moisture flux has been found significant to explain rainfall (Bontron and Obled 2005) since it characterizes not only the quantity of water in atmospheric layers but the proximity to saturation as well. Each of these variables is averaged over the whole sounding profile or within different atmospheric layers. Table 3 summarizes the performance of the DFA using these different variables as inputs, during the training phase. The term “RH profile” refers to the relative humidity over 200-m intervals. The best result is achieved with the potential moisture flux in 0–1000- and 1000–3000-m layers where the hit rates reach 70%. A principal component analysis (PCA) carried out by Godart (2009) on the whole set of variables associated with the humidity shows that all variables are strongly correlated with the first axis and between themselves. The first axis of the PCA explains 66% of total variance. Consequently, it is reasonable to consider only two variables for the characterization of the moisture as far as soundings of the two groups are concerned.

The same study is performed for the wind field. The orographic rainfall has a clearer banded organization when the unidirectional wind shear is strong, which limits the extension of the convection (Kirshbaum and Durran 2004, 2005a; Yates 2006). On the other hand, the directional wind shear favors convection but inhibits the banded organization by suppressing the vertical coherence of bands (Kirshbaum and Durran 2004, 2005a; Yates 2006). A high wind speed (>10 m s $^{-1}$) is also important (Kirshbaum and Durran 2005a; Godart et al. 2009). The mean wind speed, the directional wind shear, and the unidirectional wind shear are considered either over the whole sounding or for different atmospheric layers. The best performance for the classification in two groups (Godart 2009) is achieved when the directional wind shear, mean wind speed, and unidirectional wind shear between 1500 and 3000 m are used together. The hit rate reaches 72.1%. A principal component analysis on these three variables highlights their decorrelation (Godart 2009). The three axes explain 43%, 30%, and 27% of the total variance. The three axes are thus necessary. The information cannot be reduced.

The stability of lower layers is an important factor for the banded organization (Kirshbaum and Durran 2004, 2005a). Instability confined in the first few kilometers explains the limited development of convection associated

with banded orographic events. The gradient of the equivalent potential temperature and the wet Brunt–Väisälä frequency N_m^2 (Emanuel 1994) are used to characterize the state of stability–instability of the atmosphere. According to the different computed performance characteristics (not shown here), the discriminating variables retained are the five first gradients of equivalent potential temperature between layers of 1000-m depth. Other variables, such as the CAPE or Froude number, would have been used but they are redundant with previous variables (at least for the neural network).

In the end, 10 variables are selected: mean wind speed, directional wind shear, unidirectional wind shear between 1500 and 3000 m, potential moisture flux between 0 and 1000 m, potential moisture flux between 1000 and 3000 m, and gradient of equivalent potential temperature within the ranges of 0–1000, 1000–2000, 2000–3000, 3000–4000, and 4000–5000 m. The mean scores of the DFA and their associated standard deviations are given in Table 4 for the training and validation phases. The projection of the individuals in the factorial space presents a strong positive correlation coefficient between both the potential moisture flux within the range of 0–1000 m and the mean wind speed, and the discriminative axis. These two variables are thus stronger during banded orographic convection events than during “other rainy events.” The hit rate and POD mean scores are better during the validation phase than when the ZTUVp variables are used (Table 2). DFA is an effective method for the classification if inputs are physically based variables. The NN is used with five neurons in the hidden layer. Table 4 shows that the obtained hit rate and the SPE mean scores are better than when the ZTUVp variables are used (Table 2). As opposed to DFA results, the POD score drops. For the elaborated variables, the hit rate score with the NN is slightly lower than the hit rate score for DFA. The standard deviation indicates that the two methods now have the same accuracy.

To gain some performance, the input of the synoptic circulation was analyzed using the geopotential height (Z) given at 50-hPa intervals between 950 and 400 hPa. Compared to Table 4, scores for both statistical approaches (not shown here) were better during the training phase and similar during the validation phase. We then tried removing some levels starting at 400 and then with 450, 550 hPa, and so on to check the change in performance. Until 700 hPa, the performance did not change significantly. Figures 5 and 6 summarize all the performance characteristics for the different sets of input variables, during the training (Fig. 5) and validation (Fig. 6) phases. The best results are obtained when the 16 variables are used as inputs to the two statistical approaches (Fig. 5).

TABLE 3. Mean scores (%) obtained during the training phase with the DFA, using different input variables. The best results are indicated in boldface.

Variables (layers in meters)	Hit rate	POD	SPE
RH profile	71.7	47.4	96.1
Precipitable water			
0–1000	59.7	58.8	60.7
1000–2000	49.3	41.2	57.3
2000–3000	49.1	1.0	97.2
1000–3000	54.1	33.0	75.3
0–1500	58.8	54.6	62.9
1500–3000	52.8	43.3	62.4
0–3000	57.9	55.7	60.1
0–6000	56.8	54.6	59.0
0–1500 and 1500–3000	59.5	50.5	68.5
0–1000 and 1000–3000	65.7	76.3	55.1
0–1000 and 1000–2000	63.4	63.9	62.9
0–3000 and 3000–6000	59.1	53.6	64.6
0–1000, 1000–2000, 2000–3000	61.4	47.4	75.3
Mean	57.5	48.8	66.2
Potential precipitable water			
0–1000	64.1	67.0	61.2
1000–2000	52.6	49.5	55.6
2000–3000	49.1	13.4	84.8
1000–3000	51.8	46.4	57.3
0–1500	60.0	58.8	61.2
1500–3000	52.4	52.6	52.2
0–3000	61.0	61.9	60.1
0–6000	58.8	59.8	57.9
0–1500 and 1500–3000	61.2	57.7	64.6
0–1000 and 1000–3000	57.8	56.7	59.0
0–1000 and 1000–2000	63.0	67.0	59.0
0–3000 and 3000–6000	60.6	58.8	62.4
0–1000, 1000–2000, 2000–3000	62.5	59.8	65.2
Mean	58.1	54.6	61.6
Moisture flux			
0–1000	68.8	62.9	74.7
1000–2000	59.3	55.7	62.9
2000–3000	63.6	59.8	67.4
1000–3000	61.8	56.7	66.9
0–1500	67.1	63.9	70.2
1500–3000	63.7	63.9	63.5
0–3000	66.6	62.9	70.2
0–6000	64.4	59.8	69.1
0–1500 and 1500–3000	68.2	71.1	65.2
0–1000 and 1000–3000	67.5	72.2	62.9
0–1000 and 1000–2000	69.4	74.2	64.6
0–3000 and 3000–6000	68.5	73.3	63.8
0–1000, 1000–2000, 2000–3000	69.7	74.2	65.2
Mean	66.0	65.4	66.7
Potential moisture flux			
0–1000	68.5	62.9	74.2
1000–3000	58.0	52.6	63.5
1000–2000	58.1	51.5	64.6
2000–3000	55.5	47.4	63.5
0–1500	66.7	64.9	68.5
1500–3000	59.0	55.7	62.4
0–3000	65.0	59.8	70.2
0–6000	62.8	58.8	66.9
0–1500 and 1500–3000	66.7	72.2	61.2
0–1000 and 1000–3000	70.0	74.2	65.7

TABLE 3. (Continued)

Variables (layers in meters)	Hit rate	POD	SPE
0–1000 and 1000–2000	68.1	72.2	64.0
0–3000 and 3000–6000	64.6	61.5	67.8
0–1000, 1000–2000, 2000–3000	67.9	76.3	59.6
Mean	63.9	62.3	65.5

Similar results are found for validation of the two methods (Fig. 6). Therefore, the signature of the synoptic circulation is significant in the classification. In Fig. 7, Heidke's skill score is used to compare the different classifications to a random classification. Except for the ZTUVp classification, classifications based on elaborated variables give at least a 20% more accurate assignment to expected groups than a random classification. This is confirmed by the χ^2 test.

d. Discussion

In this section, it is shown that the nonlinear method is found to give significantly better results when it uses the ZTUVp inputs, which indicates its capability to intrinsically elaborate nonlinear variables. However, when elaborated variables are used in the two statistical approaches, no significant difference is observed. Physical reasons explain the misperformance of the two approaches and lead to the obtained discrepancy between the new classifications and the classification used as reference. To produce the reference classification in Godart et al. (2009), the hourly rainfalls are used for 24 hours around the sounding time. However, what is the temporal representativeness of soundings? Banded orographic convection events may be triggered under particular large-scale atmospheric conditions and small-scale atmospheric disturbances associated with the relief scale. Moreover, other environmental factors not represented in a sounding play a major role, such as the initial wetness of the ground, synoptic state, and the type and number of aerosols (Leroy 2007). These limits probably explain the results.

TABLE 4. Mean scores (%) obtained during the training and validation phases with the DFA and the NN using the 10 elaborated variables as inputs. The values in brackets are the standard deviations.

	Hit rate	POD	SPE	Heidke's score
DFA				
Training	78.6 (1.7)	69.5 (4.9)	83.4 (3.2)	
Validation	68.9 (4.6)	51.4 (10.9)	78.1 (6.9)	0.288
NN				
Training	78 (2.4)	62 (10.7)	86.5 (4.5)	
Validation	67.6 (4.6)	47 (13.5)	78.4 (7.8)	0.248

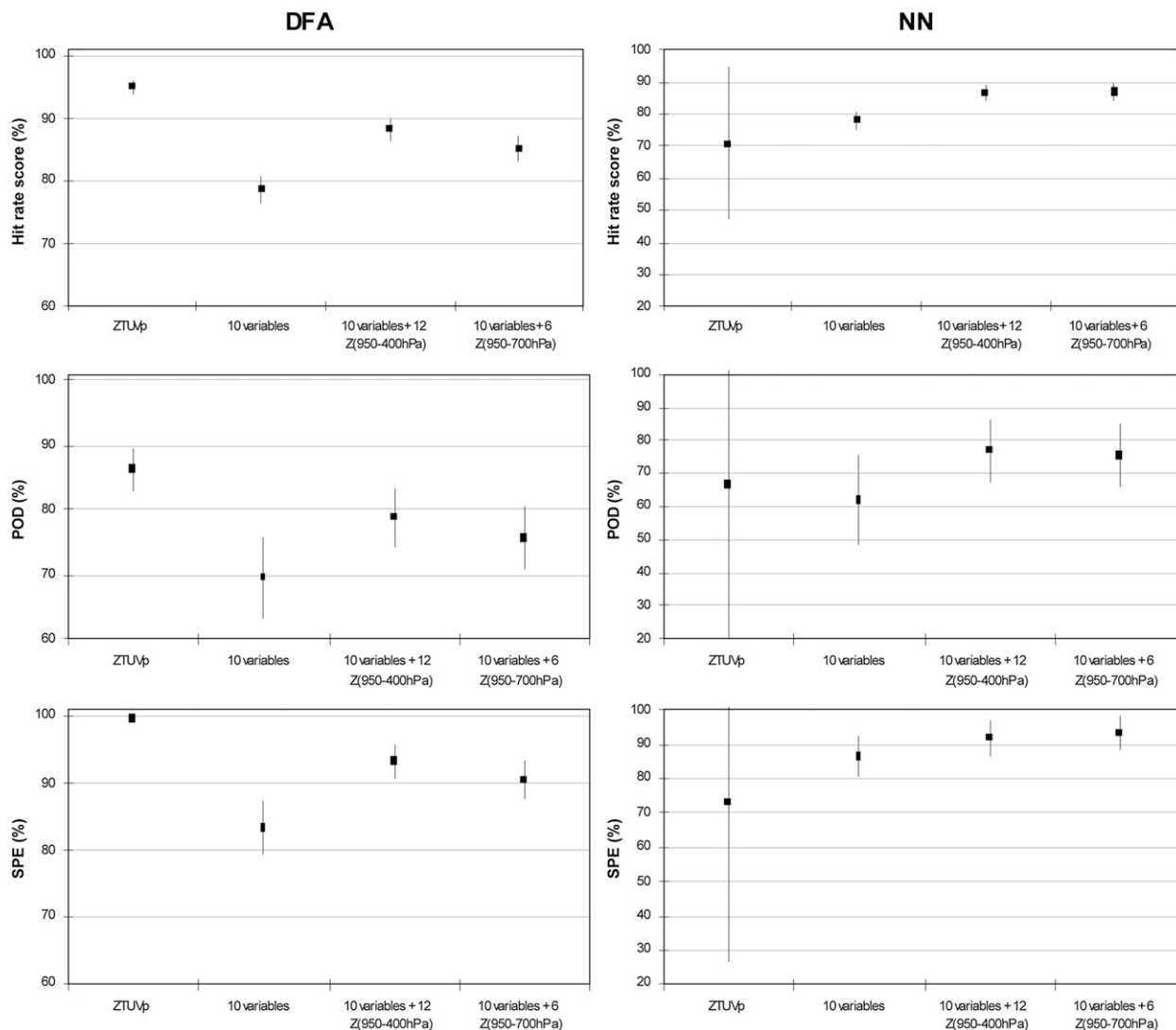


FIG. 5. Performance of the (left) DFA and (right) NN, during the training phase, using different input variables. Squares correspond to the mean score for the 500 samplings. Vertical lines indicate the interval ($\pm 1.28 \times$ standard deviation). Note that the vertical scales are not the same for the DFA and NN.

4. Sensitivity studies

Since the final goal of this study is to classify the whole sounding archive, sensitivity to the number of data in the sample and to the number of input variables was studied to test the robustness of the previous conclusions.

a. Sensitivity to the number of soundings in each group

In this section, the second type of sampling is tested. The sample is thus composed of 121 soundings in each group, with 80% used for training, and the remaining 20% for validation. In line with the conclusions of the previous section, the 16 variables are used as inputs to the statistical tools. The results of the DFA and NN after

500 samplings are summarized in Fig. 8. Whatever the type of sampling, the hit rate score for the training phase remains similar for the DFA, whereas the second sampling type improves the results of the NN. However, for the validation phase, the hit scores are slightly lower for both statistical methods when the second sampling type is used. This confirms the sensitivity of the statistical approaches to the range of the input data (Flood and Kartam 1994; Minns and Hall 1996). However, in both the training and validation phases, the POD score is found to be better for the second sampling type. The difference between the POD and SPE scores is smaller; therefore, the two groups are equally retrieved by the methods. This was not the case in the previous section since the number of individuals was larger in group 2 than in group 1.

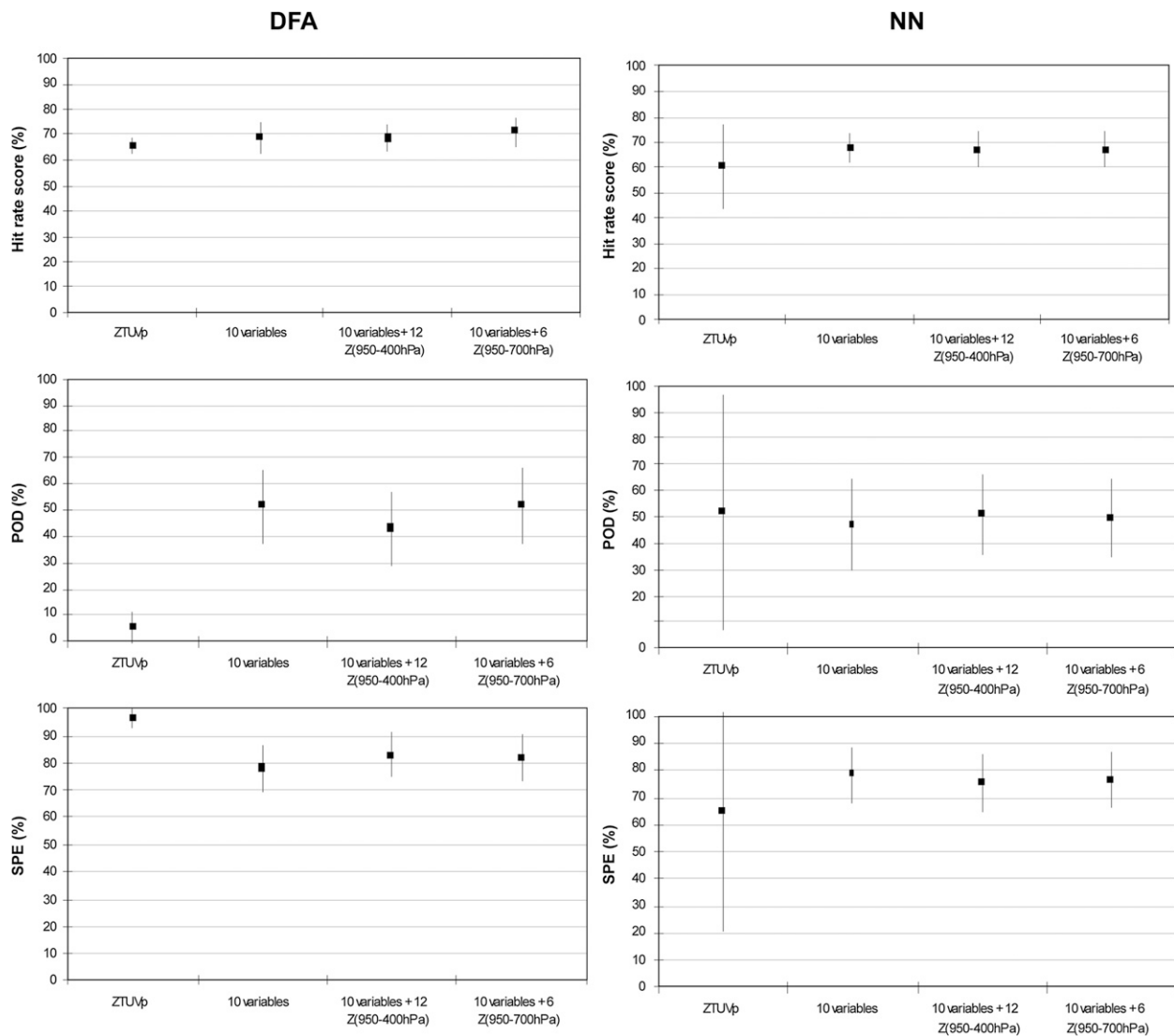


FIG. 6. As in Fig. 5, but for the validation phase.

Therefore, the probability of detection of group 2 was much larger. An equal number of individuals in each group is therefore required to provide a classification for which the probability to retrieve each group is the same.

b. Sensitivity to the number of input variables

Ghosh et al. (2004) suggested that for both the DFA and NN approaches, the accuracy of the results could be improved by reducing the number of the input variables. The dimensionality may be reduced either by reducing the number of input variables or by combining these variables by means of PCA. In this section, we will check to see if the 16 elaborated variables are all pertinent and whether or not the number of variables can be reduced with PCA. The first type of sampling (80% of 121 soundings for group 1 and 80% of 229 soundings for group 2) is used.

A pertinent variable is defined as a variable that leads to a deterioration of model performance when eliminated during the training phase. A number of tests were carried out with the elimination of one of the 16 variables. For both methods, the hit rate score was always lower than when the 16 variables were used. The gradient of equivalent potential temperature, low-level moisture flux, and geopotential height between 850 and 700 hPa were found to be the variables leading to the largest loss of performance.

Next, PCA was used to reduce the number of input variables. Only factors explaining at least 80% of the total variance were retained. Figure 9 summarizes the performance of DFA during the training and validation phases for the variables and their corresponding factors with the PCA. Globally, whatever the set of input variables,

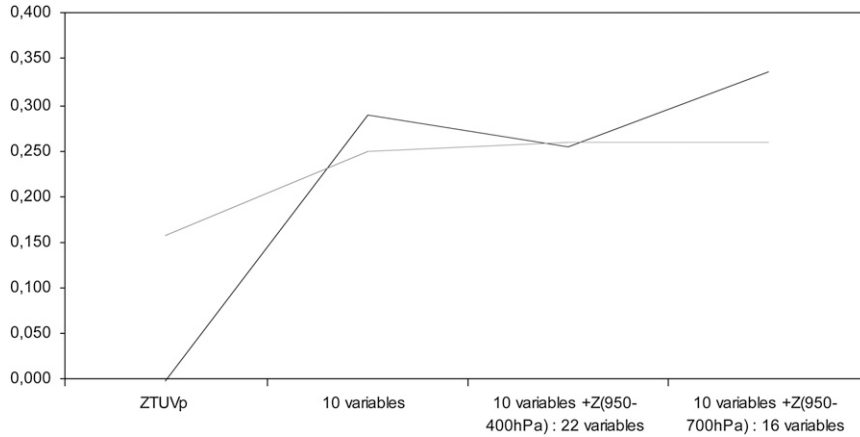


FIG. 7. Heidke's skill score for the DFA (black line) and the NN (gray line) for the different input variables used.

performance drops when PCA is used. This was also true for the NN except when the ZTUVp variables were used as input.

In conclusion, the number of input variables cannot be reduced when using elaborated variables since they are all discriminant for our classification problem.

c. Sensitivity to suspicious soundings

For the 500 samplings, a given sounding S may be sampled n times in the training sample and $(500 - n)$ times in the validation sample. A suspicious sounding is defined as a sounding that is not well classified in the validation phase for both methods. It means that for more than $(500 - n)/2$ times, it appears in the wrong group

whatever the statistical method. This section presents the effect of the iterative elimination of these suspicious soundings on performance. The data are composed of the 16 elaborated variables, and the two groups have the same number of individuals (i.e., s sampling type with 121 soundings in both groups 1 and 2).

The procedure starts from the previous results (Fig. 9), without any samples eliminated. At this stage, 48 suspicious soundings (20 in group 1 and 28 in group 2) were detected among the 242. At the first iteration, these 48 soundings were removed from the sample. The results of the validation phase, for both statistical approaches, were improved by more than 10%, but 16 new suspicious soundings were detected. This procedure was repeated

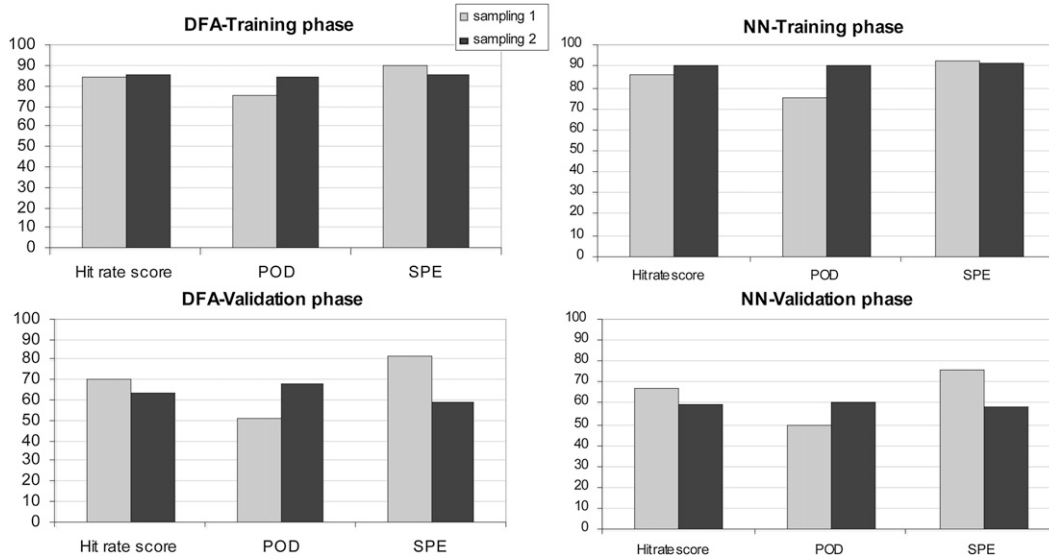


FIG. 8. Comparison of the mean performance of the two sampling approaches for the (top) training phase and (bottom) validation phase. The 16 elaborated variables are used as inputs to the statistical method: (left) DFA and (right) NN.

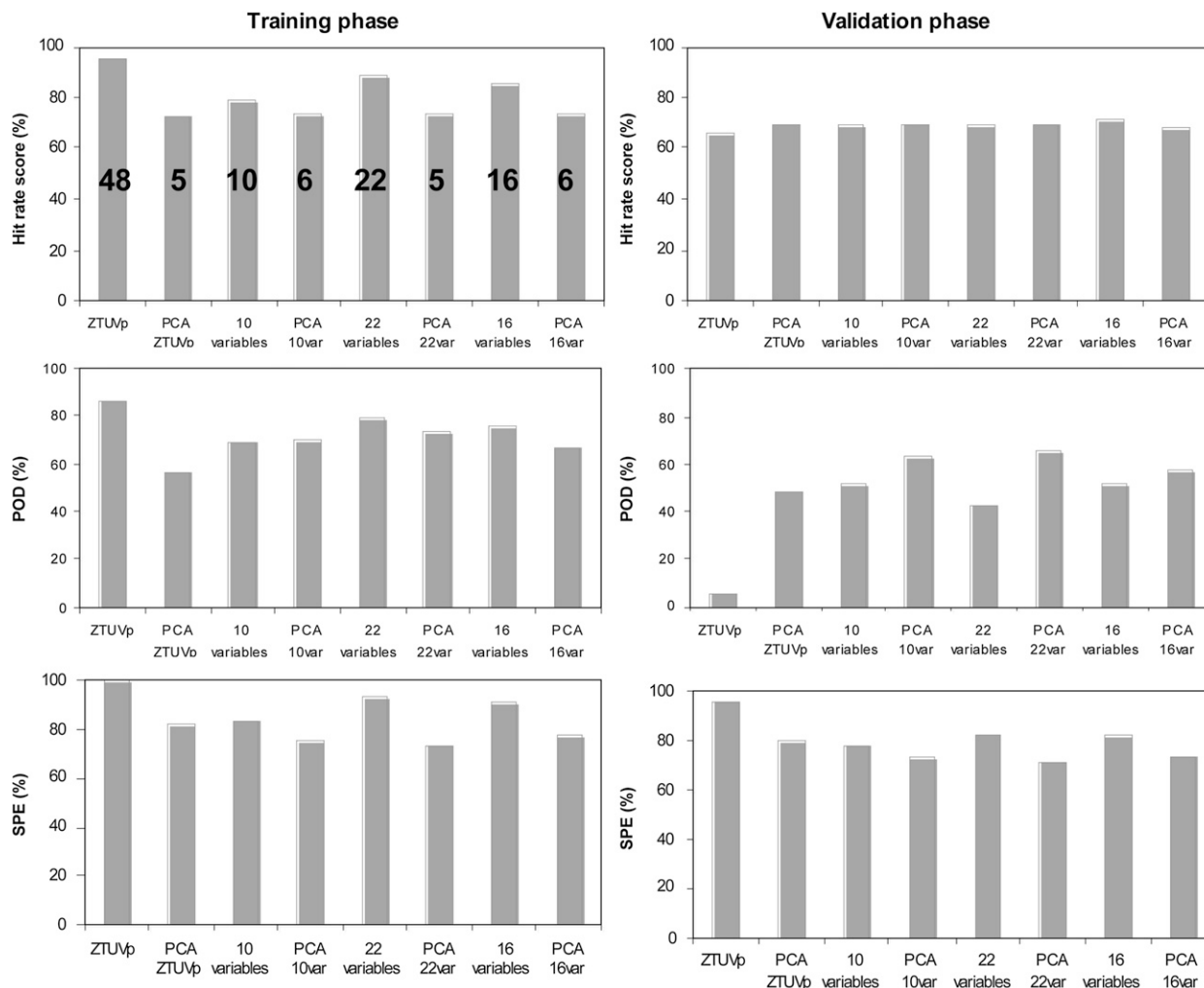


FIG. 9. Mean performance of the DFA during the training and validation phases for the different variables used and their corresponding factors in a PCA. The numbers in the upper-left panel give either the number of variables or the number of considered factors. The 22 variables are the 10 variables together with the 12 geopotential heights between 950 and 400 hPa. The 16 variables are the 10 variables together with the six geopotential heights between 950 and 700 hPa. The PCA x variables correspond to factors of the PCA performed on the x variables.

3 times. At the third iteration, no more suspicious soundings were detected and performance improved by more than 20% for the DFA and more than 30% for the NN. In the end, 68 suspicious soundings were removed among the 242. This sample of 174 soundings constitutes the final calibrated input that is now used to train the two statistical methods used to classify the whole sounding archive.

In the end, the DFA and the NN agree for 141 soundings among the 174 soundings available.

5. Relationship between banded orographic convection and atmospheric properties

This section presents a preliminary interpretation of the DFA to highlight the main atmospheric factors associated

with banded orographic convection. Next, the two statistical methods are used to classify the soundings. The obtained classification is then validated as proposed in Godart et al. (2009).

The sample of 174 soundings was used to calibrate the DFA and NN statistical methods. The hit rate score was 95.9% for the DFA, whereas it reach 100% (perfect calibration) for the NN. The representation of the DFA and the correlation coefficient of each variable on the axis is given in Fig. 10. This last representation gives an interpretation of the relationship between variables derived from soundings and banded orographic convection. Banded orographic convection events have the smallest coordinates on the axis (Fig. 10). The high potential moisture flux (PMF) in the lowest layers of the

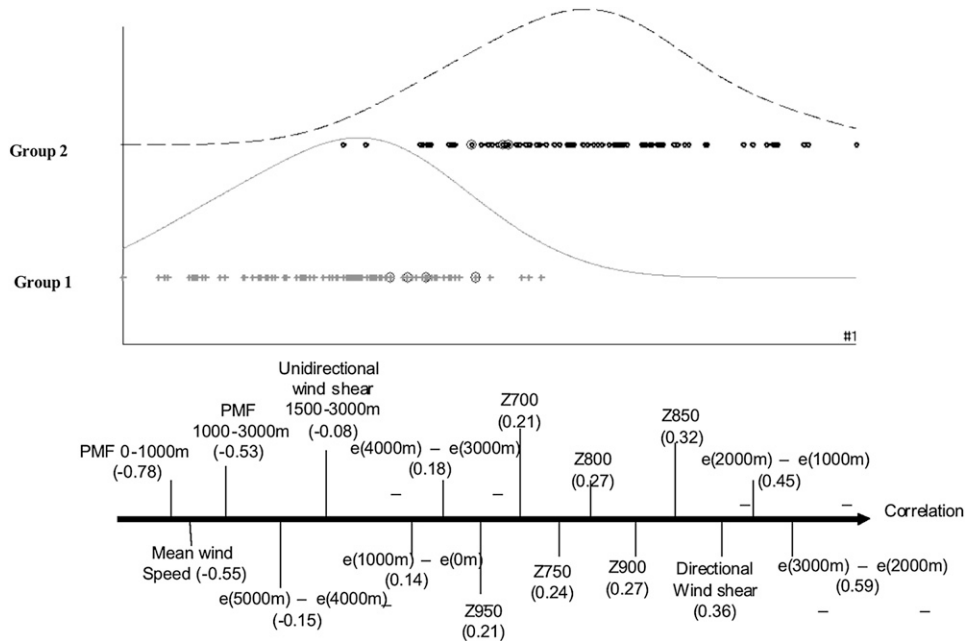


FIG. 10. (top) Results for the DFA with the 16 elaborated variables, which is the same number of soundings in each group for the training phase and without suspicious soundings. See also the legend of Fig. 4. (bottom) Representation of variables on the first axis according to their weight (correlation coefficient with the discriminant axis). PMF stands for potential moisture flux, e is the equivalent potential temperature, and Z is the geopotential height.

atmosphere (0–1000 and 1000–3000 m), the high mean wind speed, the gradient of equivalent potential temperature between 4000 and 5000 m, and the unidirectional wind shear between 1500 and 3000 m are negatively correlated with this axis. This means that the higher these variables are, the lower the coordinates on the axis will be. Consequently, the higher these variables are, the higher the probability of a banded convective orographic event (since they have the lowest coordinates on the axis). On the other hand, the directional wind shear and the gradient of equivalent potential temperature between 1000–2000 m and 2000–3000 m are positively correlated with the axis. The higher these variables are, the higher the coordinates on the axis will be and the higher the probability of an “other rainy event” (characterized by the higher coordinates in the axis).

It is more difficult to interpret results obtained with the NN.

The final step is the classification of the complete sounding archive that satisfies the dynamic criteria. Among the 880 soundings, 350 fit the rainfall criteria (Godart et al. 2009) and 59 soundings are affected to dry days according to the daily rainfall database. This leaves 466 soundings for which no rainfall information is available. The DFA and the NN methods are calibrated with the final sample of 174 soundings. Then, the statistical methods are applied to the remaining soundings, that is to

say 642 soundings [466 + (350 – 174)]. According to the DFA, 207 soundings (32.2%) correspond to banded orographic events, whereas the NN classifies 187 soundings (29.1%) as banded orographic events. The two methods agree on 514 soundings (80%). The DFA and the NN agree on banded orographic events for 133 soundings.

As was done by Godart et al. (2009), 26 numerical simulations performed with the nonhydrostatic meteorological (Meso-NH) model (Lafore et al. 1998) were used to check the classification. For clarity, only three results are illustrated, but the conclusions were the same for all the simulations. The MASDEV4.7.3 Meso-NH version was used with a grid-nesting configuration in its two-way formulation (Stein et al. 2000). The larger domain had a 4-km resolution, whereas the inner domain has a 1-km resolution. The model was forced with the different soundings as input. Lateral boundaries were opened and the top boundary was a rigid horizontal lid with an absorbing layer to avoid reflection of gravity waves. The parameterizations used in the simulations are presented in Table 5. Only results in the inner domain are presented in Fig. 11. Each simulation was performed over 24 hours and the generic sounding (Godart et al. 2009) was used as a reference because it was demonstrated to be characteristic of banded orographic convection. The 24-h accumulated simulated rainfalls are presented for 21 November 1989 (Fig. 11b),

TABLE 5. Physical parameterization for the simulations. ISBA is the Interactions between Soil, Biosphere, and Atmosphere model. TKEL, BL89, and ICE3 refer to different options of the Méso-NH package (available online at <http://mesonh.aero.obs-mip.fr/mesonh/>).

Physical mechanism	4 km	1 km
Soil-atmosphere exchange model	ISBA model (Noilhan and Planton 1989)	
Turbulent scheme	TKEL and BL89 (Bougeault and Lacarrère 1989)	
	1D	3D
Deep and shallow convection scheme	Kain-Fritsch-Bechtold scheme	
Microphysical scheme	ICE3 (Caniaux et al. 1994)	
No. of grid points: n_x, n_y, n_z	$80 \times 90 \times 45$	$120 \times 120 \times 45$
Vertical resolution (m)	60 (ground) \rightarrow 1000 (top)	
Horizontal grid size (km)	4	1
Time step (s)	12	3

classified by DFA and NN as a banded orographic convective event, and 13 October 1983 (Fig. 11c), classified as an “other rainy event” by both statistical methods. Results of the simulations confirm the obtained classification. The 21 November 1989 rainfall pattern (Fig. 11b) exhibits locations with more accumulation alternating with drier locations, which is characteristic of the banded organization. These prevailing locations are confirmed when studying instantaneous rainfall (not shown here). The 13 October 1983 rainfall pattern is not banded, rainfall intensities are weaker, and only localized over major mountains (Fig. 11c).

6. Conclusions and perspectives

Studies on banded orographic events are necessary to improve our knowledge of how they occur and to quantify their contribution to the total hydrological input in mountainous areas. Nevertheless, these studies remain difficult since continuous observations are rare. For example, in the Cévennes–Vivarais region, in the southeastern part of France, operational weather radar networks are not designed to observe low-level convection within complex terrain, and the density of rain gauges is insufficient to address this issue.

In this paper, two statistical methods are used to build an extensive database of banded orographic convective rainfall events in this part of France. In a previous study, some of these events were identified by the combined use of meteorological and rainfall criteria. However, the lack of hourly rainfall data corresponding to all the meteorological data in the archive led to an incomplete database that is impossible to use for a climatic study. This work therefore attempted to define atmospheric factors derived from soundings that can be used as inputs to identify these events without the use of the rainfall archive.

The problem of the nonlinearity in the relationship between the atmospheric factors and rainfall fields was solved in two steps. Using basic variables derived from

the soundings as input to the statistical tools, the first step consisted of evaluating the effectiveness of the neural network (NN), a nonlinear statistical model, to capture the nonlinearity in comparison to the discriminant factorial analysis (DFA). The next step consisted of elaborating on nonlinear variables based on a priori knowledge. These variables were then used as input to the DFA and the NN. The advantages of each method were evaluated to provide the best classification of the banded orographic convective events.

Since the final goal of this study was to classify all the soundings between 1976 and 2005, the work was organized into four steps: 1) analysis of the effectiveness of the NN to render the nonlinearity between the atmospheric environment and the rainfall; 2) combination of basic variables to form elaborated variables that best identify banded orographic convective rainfall events; 3) application of the statistical tools on a calibrated sample; and 4) classification of the entire set of soundings. The three first steps used a sample that had been previously evaluated and validated for this aim. It consisted of 350 soundings at Nîmes, divided into 121 soundings validated as banded orographic convection events and 229 soundings of “other rainy events.” When the basic thermodynamic variables were used as input, the neural network gave much better results than the DFA. However, when using elaborated variables, both methods gave similar results.

On the basis of several DFA performance analyses, 16 elaborated nonlinear variables were found to best discriminate the banded orographic convection events. These 16 variables were established using a priori knowledge of the processes associated with this rainfall pattern organization. They include wind shear, moisture fluxes, and gradients of the equivalent potential temperature in the lower layers of the atmosphere. The synoptic circulation expressed by the geopotential heights is also required to identify these events. Using these 16 elaborated variables as input, the two statistical models perform equally. The nonlinearity of the NN is therefore

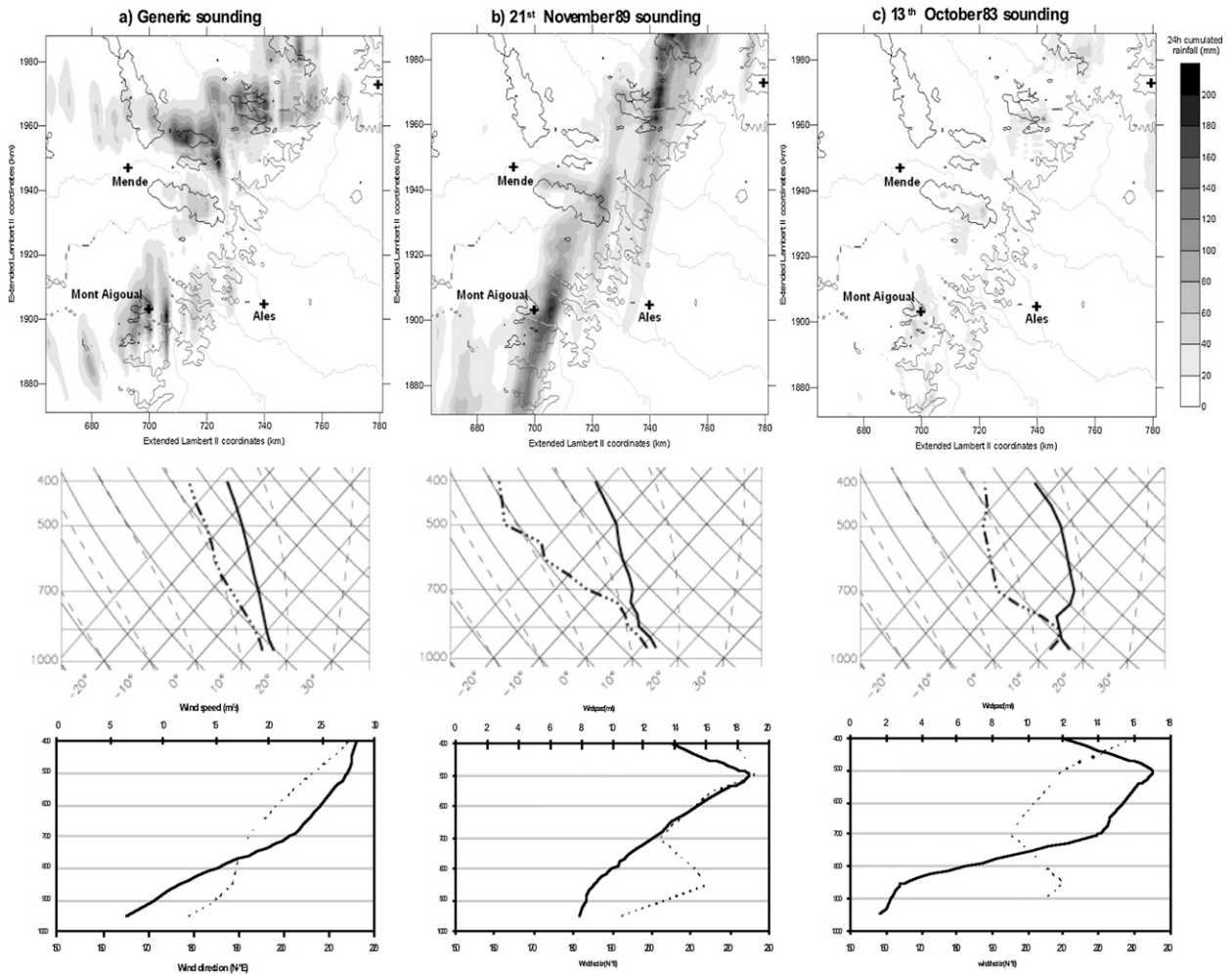


FIG. 11. (top) The 24-h accumulated simulated rain (mm) for the (a) generic sounding (Godart et al. 2009), (b) 21 Nov 1989 sounding, and (c) 13 Oct 1983 sounding. Solid lines indicate the relief contours (dark gray line for 500-m contour, black line for 1300-m contour). The main river network (light gray) and major locations (crosses) are also represented. (middle) Skew plots for the corresponding sounding (solid line: temperature, dashed line: dewpoint temperature). (bottom) Wind profile (solid line: wind direction, dashed line: wind speed).

no longer necessary since the necessary information for the classification is contained in these variables. At this stage, linear and nonlinear approaches give equivalent results. However, the NN outputs are much more difficult to interpret physically than the DFA outputs.

The sensitivity studies concerning the type of sampling revealed the need to carry out classification with an equal number of individuals in each group. The probability of retrieving each group is therefore the same.

Next, the final set of individuals used to calibrate the two statistical models was obtained from an iterative procedure that eliminated soundings that were classified in the wrong group by both statistical methods. Less than 30% of the soundings were removed from the initial sample. The results of the classification of this calibration

set show the relationship between the different atmospheric variables and banded orographic convection. High mean wind speed, a strong potential moisture flux between 0 and 1000 and between 1000 and 3000 m, a stable layer between 4000 and 5000 m, and high unidirectional wind shear between 1500 and 3000 m are the most relevant variables associated with these events.

In the end, 3 simulations, among 26 (Godart 2009), using three different soundings as input to the Méso-NH atmospheric model were used to illustrate the results of the final classification of the entire set of soundings. The simulated rain fields validate the classification. These simulated results are currently part of an article in preparation that is aimed mainly at understanding the atmospheric processes involved. These simulations have been

extended by academic sensitivity studies highlighting the contributing role of the different processes (mean wind speed, moisture flux, stability, etc.) on the rainfall pattern.

Moreover, it is planned to implement an extensive observation system within the Hydrological Cycle in Mediterranean Experiment (HyMeX; available online at <http://www.cnrm.meteo.fr/hymex/>; see Anquetin et al. 2008) to gather detailed observations that will be used to validate our results. The observation system will include an X-band radar network, which is mainly located in the mountains, radar pointers, and low-level soundings in the valleys to observe the lowest layer of the atmosphere. Moreover, a high-density network of disdrometers will be deployed to sample rainfall variability below the bands. This unique dataset will help improve our understanding of the processes associated with shallow convection and the transition between shallow and deep convection. This observation system will be first implemented in the southeast of France, then moved to Crete, where such conditions are often observed. In terms of climatological perspectives, the final classification may be used to evaluate the contribution of such events to the total hydrologic input to the region (Godart 2009).

Acknowledgments. The current study was supported by the French national LEFE/IDAO program from INSU. Data were provided by Météo-France, the University of Wyoming, and the OHMCV. The numerical simulations were carried out on the IDRIS numerical platform.

REFERENCES

- Amari, S., N. Murata, K.-R. Muller, M. Finke, and H. H. Yang, 1997: Asymptotic statistical theory of overtraining and cross-validation. *IEEE Trans. Neural Networks*, **8**, 985–996.
- Andrieu, H., and J.-D. Creutin, 1995: Identification of vertical profiles of radar reflectivity for hydrological applications using an inverse method. Part 1: Formulation. *J. Appl. Meteor.*, **34**, 225–239.
- Anquetin, S., F. Minsicloux, J.-D. Creutin, and S. Cosma, 2003: Numerical simulation of orographic rainbands. *J. Geophys. Res.*, **108**, 8386, doi:10.1029/2002JD001593.
- , and Coauthors, cited 2008: HyMeX: Shallow orographic convection contribution to the water resources in Mediterranean: Proposition of an observation device within the framework of HyMeX. [Available online at <http://lthel21.hmg.inpg.fr/PagePerso/anquetin/Hymex.html>.]
- Bankert, R., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. *J. Appl. Meteor.*, **33**, 909–918.
- Banta, R. M., 1990: The role of mountain flows in making clouds. *Atmospheric Processes over Complex Terrain, Meteor. Monogr.*, No. 45, Amer. Meteor. Soc., 173–228.
- Barros, A. P., and D. P. Lettenmaier, 1994: Dynamic modeling of orographically induced precipitation. *Rev. Geophys.*, **32**, 265–284.
- Bennani, Y., 2006: *Apprentissage Connexionniste*. Hermès, 361 pp.
- Bishop, C., 1995: *Neural Networks for Pattern Recognition*. Clarendon Press, 482 pp.
- Bois, P., C. Obled, M.-F. Saintignon, and H. Mailloux, 1997: *Atlas Expérimental des Risques de Pluies Intenses: Cévennes-Vivarais*. CNRS, 24 pp.
- Bontron, G., and C. Obled, 2005: A probabilistic adaptation of meteorological model outputs to hydrological forecasting. *Houille Blanche*, **1**, 23–28.
- Bougeault, P., and P. Lacarrère, 1989: Parameterization of orography-induced turbulence in a mesobeta-scale model. *Mon. Wea. Rev.*, **117**, 1872–1890.
- Cacoullos, T., 1973: *Discriminant Analysis and Applications*. Academic Press, 434 pp.
- Caniaux, G., J.-L. Redelsperger, and J.-P. Lafore, 1994: A numerical study of the stratiform region of a fast-moving squall line. Part I. General description and water and heat budgets. *J. Atmos. Sci.*, **51**, 2046–2074.
- Caudill, M., 1991: Neural network training tips and techniques. *AI Expert*, **6**, 56–61.
- Cheng, B., and D. Titterton, 1994: Neural networks: A review from a statistical perspective. *Stat. Sci.*, **9**, 2–54.
- Colle, B. A., 2004: Sensitivity of orographic precipitation to changing ambient conditions and terrain geometries: An idealized modeling perspective. *J. Atmos. Sci.*, **61**, 588–606.
- Cosma, S., E. Richard, and F. Miniscloux, 2002: The role of small-scale orographic features in the spatial distribution of precipitation. *Quart. J. Roy. Meteor. Soc.*, **128**, 75–92.
- David Shepard Associates, 1990: *The New Direct Marketing: How to Implement a Profit-Driven Database Marketing Strategy*. Dow Jones-Irwin, 535 pp.
- Doolittle, M., 1888: Association ratios. *Bull. Philos. Soc. Wash.*, **7**, 122–127.
- Doswell, C. A., III, R. Davies-Jones, and D. L. Keller, 1990: On summary measures of skill in rare event forecasting based on contingency tables. *Wea. Forecasting*, **5**, 576–585.
- Elsner, J., and A. Tsonis, 1992: Nonlinear prediction, chaos, and noise. *Bull. Amer. Meteor. Soc.*, **73**, 49–60.
- Emanuel, K. A., 1994: *Atmospheric Convection*. Oxford University Press, 580 pp.
- Fletcher, D., and E. Goss, 1993: Forecasting with neural networks: An application using bankruptcy data. *Inf. Manage.*, **24**, 159–167.
- Flood, I., and N. Kartam, 1994: Neural networks in civil engineering. I: Principles and understanding. *J. Comput. Civ. Eng.*, **8**, 131–148.
- Frei, C., and C. Schär, 1998: A precipitation climatology of the Alps from high-resolution raingauge observations. *Int. J. Climatol.*, **18**, 873–900.
- Fuhrer, O., and C. Schär, 2007: Dynamics of orographically triggered banded convection in sheared moist orographic flows. *J. Atmos. Sci.*, **64**, 3542–3561.
- Gardner, M., and S. Dorling, 1998: Artificial neural networks: A review of applications in the atmospheric sciences. *Atmos. Environ.*, **32**, 2627–2636.
- , and —, 1999: Neural network modelling and prediction of hourly NO_x and NO₂ concentrations in urban air in London. *Atmos. Environ.*, **33**, 709–719.
- Ghosh, S., P. Sen, and U. De, 2004: Classification of thunderstorm and non-thunderstorm days in Calcutta (India) on the basis of linear discriminant analysis. *Atmosfera*, **17**, 1–12.
- Godart, A., 2009: Les précipitations orographiques organisées en bandes dans la région Cévennes-Vivarais. Caractérisation et

- contribution au régime pluviométrique. Ph.D. thesis, University Joseph Fourier–Grenoble, 336 pp.
- , S. Anquetin, and E. Leblois, 2009: Rainfall regimes associated with banded convection in the Cévennes-Vivarais area. *Meteor. Atmos. Phys.*, **103**, 25–34, doi:10.1007/s00703-008-0326-3.
- Gysi, H., 1998: Orographic influence on the distribution of accumulated rainfall with different wind directions. *Atmos. Res.*, **47–48**, 615–633.
- Houze, R., W. Schmid, R. Fovell, and H. H. Schiesser, 1993: Hailstorms in Switzerland: Left movers, right movers, and false hooks. *Mon. Wea. Rev.*, **121**, 3345–3370.
- Kirshbaum, D. J., and D. R. Durran, 2004: Factors governing cellular convection in orographic precipitation. *J. Atmos. Sci.*, **61**, 682–698.
- , and —, 2005a: Atmospheric factors governing banded orographic convection. *J. Atmos. Sci.*, **62**, 3758–3774.
- , and —, 2005b: Observations and modelling of banded orographic convection. *J. Atmos. Sci.*, **62**, 1463–1479.
- , G. H. Bryan, R. Rotunno, and D. R. Durran, 2007a: The spacing of orographic rainbands triggered by small-scale topography. *J. Atmos. Sci.*, **64**, 4222–4245.
- , —, —, and —, 2007b: The triggering of orographic rainbands by small-scale topography. *J. Atmos. Sci.*, **64**, 1530–1549.
- Kuligowski, R. J., and A. P. Barros, 1998: Experiments in short-term precipitation forecasting using artificial neural networks. *Mon. Wea. Rev.*, **126**, 470–482.
- Lafore, J. P., and Coauthors, 1998: The Meso-NH Atmospheric Simulation System. Part I: Adiabatic formulation and control simulation. *Ann. Geophys.*, **16**, 90–109.
- Lau, C., and B. Widrow, 1990: Special issue on neural networks. *Proc. IEEE*, **78**, 1411–1414.
- Leroy, D., 2007: Développement d'un modèle de nuage tridimensionnel à microphysique détaillée - Application à la simulation de cas de convection profonde. Ph.D. thesis, University Blaise Pascal, 214 pp.
- Lin, Y. L., S. Chiao, T. A. Wang, M. L. Kaplan, and R. P. Weglarz, 2001: Some common ingredients for heavy orographic rainfall. *Wea. Forecasting*, **16**, 633–660.
- Maier, H., and G. Dandy, 2000: Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environ. Modell. Software*, **15**, 101–124.
- Manzato, A., 2005: The use of sounding-derived indices for a neural network short-term thunderstorm forecast. *Wea. Forecasting*, **20**, 896–917.
- Marzban, C., 2000: A neural network for tornado diagnosis. *Neural Comput. Appl.*, **9**, 133–141.
- , and G. Stumpf, 1996: A neural network for tornado prediction based on Doppler radar-derived attributes. *J. Appl. Meteor.*, **35**, 617–626.
- , and A. Witt, 2001: A Bayesian neural network for severe-hail size prediction. *Wea. Forecasting*, **16**, 600–610.
- Masters, T., 1993: *Practical Neural Network Recipes in C++*. Academic Press, 493 pp.
- McGinnis, D. L., 1994: Predicting snowfall from synoptic circulation: A comparison of linear regression and neural network in methodologies. *Neural Nets: Applications in Geography*, B. Hewitson and R. G. Crane, Eds., Kluwer Academic Publishers, 79–99.
- Mercer, A., M. Richman, H. Bluestein, and J. Brown, 2008: Statistical modeling of downslope windstorms in Boulder, Colorado. *Wea. Forecasting*, **23**, 1176–1194.
- Miniscola, F., J. D. Creutin, and S. Anquetin, 2001: Geostatical analysis of orographic rainbands. *J. Appl. Meteor.*, **40**, 1835–1854.
- Minns, A., and M. Hall, 1996: Artificial neural networks as rainfall-runoff models. *Hydrol. Sci. J.*, **41**, 399–417.
- Navone, H. D., and H. A. Ceccatto, 1994: Predicting Indian monsoon rainfall: A neural network approach. *Climate Dyn.*, **10**, 305–312.
- Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536–549.
- Pankiewicz, G. S., 1995: Pattern recognition techniques for the identification of cloud and cloud systems. *Meteor. Appl.*, **2**, 257–271.
- Peak, J., and P. Tag, 1994: Segmentation of satellite imagery using hierarchical thresholding and neural networks. *J. Appl. Meteor.*, **33**, 605–616.
- Pointin, Y., D. Ramond, and J. Fournet-Fayard, 1988: Radar differential reflectivity Z_{DR} : A real-case evaluation of errors induced by antenna characteristics. *J. Atmos. Oceanic Technol.*, **5**, 416–423.
- Ripley, B., 1994: Neural networks and related methods of classification. *J. Roy. Stat. Soc.*, **56**, 409–456.
- , 1996: *Pattern Recognition and Neural Networks*. Cambridge University Press, 403 pp.
- Rogers, L., and F. Dowla, 1994: Optimization of groundwater remediation using artificial neural networks with parallel solute transport modelling. *Water Resour. Res.*, **30**, 457–481.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986a: Learning internal representations by error propagation. *Foundations*, D. E. Rumelhart and J. L. McClelland, Eds., Vol. 1, *Parallel Distributed Processing*, MIT Press, 318–362.
- Smith, R. B., 1979: The influence of mountains on the atmosphere. *Advances in Geophysics*, Vol. 21, Academic Press, 87–230.
- Stanley, J., 1988: *Introduction of Neural Networks: Computer Simulations of Biological Intelligence*. California Scientific Software, 222 pp.
- Stein, J., E. Richard, J. P. Lafore, J. P. Pinty, N. Asencio, and S. Cosma, 2000: High-resolution non-hydrostatic simulations of flash-flood episodes with grid-nesting and ice-phase parametrization. *Meteor. Atmos. Phys.*, **72**, 203–221.
- Venugopal, V., and W. Baets, 1994: Neural networks and statistical techniques in marketing research: A conceptual comparison. *Mark. Intell. Plann.*, **12**, 30–38.
- Verdecchia, M., G. Visconti, F. D'Andrea, and S. Tibaldi, 1996: A neural network approach for blocking recognition. *Geophys. Res. Lett.*, **23**, 2081–2084.
- Weichert, A., and G. Bürger, 1998: Linear versus nonlinear techniques in downscaling. *Climate Res.*, **10**, 83–93.
- Weigend, A., D. Rumelhart, and B. Huberman, 1990: Predicting the future: A connectionist approach. *Int. J. Neural Syst.*, **1**, 193–209.
- Wierenga, B., and J. Kluytmans, 1994: Neural nets versus marketing models in time series analysis: A simulation study. *Marketing: Its Dynamics and Challenges*, J. Bloemer, J. Lemmink, and H. Kasper, Eds., EMAC, 1139–1153.
- Yates, E., 2006: Convection en région Cévennes-Vivarais: Etude de données pluviométriques, simulations numériques et validation multi-échelles. Ph.D. thesis, Grenoble Polytechnic Institute, 235 pp.