Atmospheric Research 93 (2009) 446–456

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmos

Assessing sounding-derived parameters as storm predictors in different latitudes

José Luis Sánchez [a], José Luis Marcos [a], Jean Dessens [b], Laura López [a,*], Carlos Bustos [c], Eduardo García-Ortega [a]

[a] *Laboratory of Atmospheric Physics, IMA, University of León, Spain*
[b] *ANELFA, Toulouse, France*
[c] *Government of Mendoza, Argentina*

ARTICLE INFO

Article history:
Received 8 January 2008
Received in revised form 24 November 2008
Accepted 26 November 2008

Keywords:
Forecast
Hailstorms
Logistic regression model
Spain
France
South America

ABSTRACT

Many thermodynamic parameters and indices are currently being used as thunderstorm predictors because of their high correlations with the beginning and development of convection. Many of these indices have been developed for one specific area and their forecasting accuracy has generally been assessed in that zone and not in others. It is a highly intriguing question whether there are parameters or indices that may function adequately as thunderstorm predictors, as far as the Probability of Detection is concerned, irrespective of the latitude of the study zone.

In order to approach this issue the present study focuses on data from 1692 sounding days in León (Spain), Zaragoza (Spain), Bordeaux (France) and Mendoza (Argentina). Specific discriminant models have already been developed for these areas.

When comparing the results found by the different models constructed for each of the four study zones it may be noticed that there are no indices that function extremely well in all of the zones. Rather, a common ingredient pattern is observed for the beginning of convection): atmospheric instability and moist layers in the low atmosphere. It may also be concluded that sounding data alone are not enough to detect accurately the triggering mechanism, which is the third ingredient necessary for convection.

The aim of this paper is to build a logistic equation integrating the four study zones. The stepwise method was employed with this purpose because it allows for the gradual inclusion of variables in the final equation according to their discriminating power. The results obtained suggest that Showalter Index and 850 hPa Dew Point Temperature are the variables that best characterize preconvective conditions irrespective of the geographic area considered. The values for POD (Probability of Detection) and FAR (False Alarm Ratio) are acceptable, but they are clearly lower than the ones obtained by each of the models in the study zone for which they were developed.

1. Introduction

It is becoming increasingly evident that weather forecasts are the most important service that the societies demand from their National Meteorological Departments (Schmeits et al., 2005; Weiher, 2007). More accurate weather forecasts enable people to make better decisions in various situations: in activities where lives or properties may be at risk (Nicholls, 2001; Ebi et al., 2004), in planning daily activities or leisure activities (Lazo and Chestnut, 2002), or in any economic activity that depends to some extent on the weather (Murphy, 1994; Nicholls, 1996; Katz and Murphy, 1997; Mjelde et al., 1998; Brooks and Douglas, 1998; Lazo, 2005; Leston et al., 2007). Therefore, economic assessment and research in

* Corresponding author. Group for Atmospheric Physics, Institute for the Environment, University of León, 24071 León, Spain.
E-mail address: llopc@unileon.es (L. López).

0169-8095/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.atmosres.2008.11.006

communication and social perception should become an integral part of weather forecast systems (Doswell, 2003; Llasat-Botija et al., 2007).

Forecasting the weather is a complex task because decisions must be made in situations with a certain degree of uncertainty. Several authors (Tsonis and Elsner, 1989; Elsner and Tsonis, 1992; Wilks, 1991) have discussed the uncertainty associated to atmospheric phenomena, arguing that it represents one of the main challenges in short-term forecasts of severe convective events due to their intrinsic stochastic nature (Schmeits et al., 2005). To solve this problem we make use of statistical models, which are very valuable tools providing one clearly defined response on the basis of a number of meteorological data or variables measured in a particular place. In other words, the aim is to establish automatic procedures to make the task of the forecaster easier.

Many studies have employed statistical regression equations to forecast severe weather events, but the coefficients of these equations vary according to the geographic location of the study zone and according to the season of the year. In fact, the climatic conditions characteristic of a particular place must be reflected somehow in the equations (Glahn and Lowry, 1972; Charba, 1979; Dubrovsky, 1994; Hughes, 2004).

Logistic regression models offer their results in terms of Risk/No Risk (binary forecast models) and many authors have used them to forecast storms (Angus et al., 1988; Dubrovsky, 1994; Billet et al., 1997; Sánchez et al., 1998, 2001b, 2007a; Schmeits et al., 2005; López et al., 2007).

In a binary forecast a relationship is established between a predictand, i.e. the occurrence of a thunderstorm or hailfall, and a group of predictors which comprises a number of meteorological variables and stability indices. Obviously, the forecasting power of the equations will depend on the appropriate selection of the variables, and consequently, these variables must account for the physical mechanisms involved in convection. According to Johns and Doswell (1992), three ingredients are necessary:

i) A sufficiently moist and deep layer in the low or mid atmosphere.
ii) Conditional instability.
iii) A triggering mechanism to start convection.

The three ingredients are required, and even though they influence convective activity in different ways, their interaction determines the likelihood of a particular phenomenon occurring or not occurring. In general, the data necessary to construct the different models is provided by sounding balloons, on the one hand, and by meteorological variables measured at the surface. Numerous studies claim that the ideal situation is the use of parameters calculated on the basis of sounding data to characterize preconvective conditions and the actual beginning of convection. The values for moisture and instability required for convection may thus be determined (Schultz, 1989; Johns and Doswell, 1992; McNulty, 1995; Tudurí and Ramis, 1997; Llasat et al., 1997; Sánchez et al., 1998, 2001a, 2007a; López et al., 2007). It is much more difficult though to establish the triggering mechanisms, which are best determined using mesoscale models (García-Ortega et al., 2007).

One important issue with respect to the thermodynamic indices is that most of them have been developed and validated as diagnostic variables and not as forecasting variables. Doswell and Schultz (2006) define a diagnostic variable being used as a forecast parameter as one that allows a forecaster to make an accurate weather forecast based on the current values of that variable. That is, it is expected that for a proper forecast parameter there should be a time-lagged correlation between the parameter and the weather being forecast.

Over the years a large number of indices provided by radio soundings have been defined and tested to characterize situations that could favor the formation and development of thunderstorms: Showalter (Showalter, 1953), Lifted Index (Galway, 1956), K (George, 1960), Total Totals (Miller, 1975), CAPE (Moncrieff and Miller, 1976) and others. The values reached by these indices generally express ranges of the probability for a particular meteorological phenomenon to occur or not, making it possible to quantify the uncertainty implicit in any forecasts (Murphy, 1977).

It is essential to verify the quality of the prediction, so it is necessary to compare the values of the indices with the occurrence or non-occurrence of the phenomenon. For this purpose, a number of skill scores are employed combining in different ways the hits and misses of the various indices. The most widely used skill scores are Probability of Detection (POD), False Alarm Ratio (FAR), Frequency of Misses (FOM), and True Skill Score (TSS), among others.

The various stability indices have been used separately, and have later been compared with respect to their forecasting power in the case of storms and other related phenomena in many parts of the world: Anthes (1976); Schultz (1989) in the U.S.A.; Jacovides and Yonetani (1990), Collier and Lilley (1994); Huntrieser et al. (1997); Haklander and Van Delden (2003); Marinaki et al. (2006) in Europe, or Takashi and Hiroshi (2005) in Japan.

No index may characterize the state of the atmosphere on its own (Blanchard, 1998) and contradictory results may be obtained when considering the indices separately (Schultz, 1989). It is preferable to select combinations of variables grouped together in statistic regression models. This procedure has proved to be adequate, as a considerable improvement in the forecasting accuracy has been noticed (Bowman et al., 1969).

In the case of regression models there are clearly defined statistical techniques to select the variables that will better fit the experimental data to the mathematical functions. Because the climatology in the study zone is important for determining the frequency of storms and their characteristics, the question rises of whether there are variables that are systematically included in the equations due to their relative importance in characterizing preconvective situations. According to Schultz (1989), considering stability indices separately, it is not likely that any of them will work better everywhere, and similarly, it is not likely that one particular stability index will be a better storm forecaster than any of the other indices.

The present paper will use a logistic regression model to analyze common patterns in all of them, in particular with respect to the systematic use of the same variables when studying the characteristics of storms or hailstorms in various parts of the world: the province of León and the Central Ebro Valley in Spain, several parts of France, and the region of Mendoza in Argentina.

The results suggest that there are indeed common variables in different study zones, and the possible causes and implications of this fact will be analyzed in detail.