Atmospheric Research 164–165 (2015) 194–209

Contents lists available at ScienceDirect

Atmospheric Research

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

Spatial distribution of thermodynamic conditions of severe storms in southwestern Europe

E. Gascón [a,*], A. Merino [a], J.L. Sánchez [a], S. Fernández-González [b], E. García-Ortega [a], L. López [a], L. Hermida [a]

[a] Atmospheric Physics Group, IMA, University of León, 24071 León, Spain
[b] Department of Astrophysics and Atmospheric Sciences, Faculty of Physical Sciences, Universidad Complutense de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 28 April 2015
Received in revised form 19 May 2015
Accepted 21 May 2015
Available online 28 May 2015

Keywords:
Thermodynamic
Hailstorm
Stability index
Binary logistic regression
Cluster analysis

ABSTRACT

The Mid-Ebro Valley (MEV) is an area of the northeast Iberian Peninsula with a large number of hailstorms throughout the year. Therefore, new forecasting tools are required to improve the spatiotemporal detection of such storms. Using a database of 100 days with severe storms (SSs) over a 13-year period between 2001 and 2013, we obtained vertical profiles predicted by the WRF mesoscale model. A total of 31 indexes describing conditions of humidity, stability, helicity or precipitable water were obtained from the profiles and input to a binary logistic regression model. The regression model was applied using the forward stepwise method, which indicated that the stability indexes were the most accurate for SS in the study area. Of the 31 indexes, 5 were selected: Showalter index (SI), wind speed at 500 hPa (SPD500), dew point temperature at 850 hPa (Td850), relative helicity between 0 and 3 km (SREH3km), and wet bulb zero height (WBZ). Combination of these indexes in a logistic equation gives the probability of SS risk/no risk in the study area. Results of the logistic equation show a Probability of Detection (POD) of 0.94 and a False Alarm Ratio (FAR) of 0.22. The second part of the article describes regionalization of the study area by SS spatial distribution according to the logistic equation. Thus, using multivariate techniques, we used principal component analysis (PCA) in T-mode and posterior cluster analysis, getting four clusters according to the spatial distribution of SS thermodynamic behavior and the distribution of storms observed via radar data. Clusters 1 and 2 showed probabilities of hail occurrence that were lower than Clusters 3 and 4, mainly affecting the MEV and the eastern end of the study area. Likewise, the predicted hail area was more extensive in those last two clusters. These results provide a new tool that complements those previously developed for this study area, toward improving SS prediction and pinpointing these storms in space and in time.

1. Introduction

One of the major problems in predicting severe hailstorms is accurately detecting their location at small spatial and temporal scales. The development of these storms usually occurs at small spatial (1–10 km) and temporal (1–12 h) scales (Kunz, 2007), such that identification of the start of convection and of storm intensity remains an issue and a challenge to weather forecasters. Hailstorms form from localized severe convection and produce intense precipitation that can result in natural disasters and significant negative socioeconomic effects, especially in mountainous terrain that is favorable for triggering convection (Barthlott et al., 2006; de la Torre et al., 2011, 2015; García-Ortega et al., 2012).

Some authors have defined three general requirements for such convective development (Johns and Doswell, 1992; Houze, 1993). These are a sufficiently deep layer with instability, sufficient moisture at the lowest levels of the atmosphere, and a trigger mechanism that activates the convective process. To analyze these requirements, we must take into account the physical mechanisms involved in the development of convection, which operate at different scales and interact with each other. Convective processes depend mainly on the synoptic configuration or large-scale processes, whereas mesoscale processes mainly determine the initiation of convection (Doswell, 1987). In an attempt to identify this convective initiation and associated thermodynamic characteristics, various thermodynamic indexes have been defined over recent decades, which indicate potential storm development according to air mass properties such as humidity, temperature and helicity (Kunz, 2007). These thermodynamic or stability indexes are usually applied to soundings to study a particular convective process. However, soundings are available only 2–4 times per day, generally at 0000 and 1200 UTC, and their launch sites may be separated by hundreds of kilometers. For this reason, tools are required that enable more accurate spatiotemporal determination of pre-convective conditions for severe storm episodes. A key tool used in recent years by many researchers to avoid

* Corresponding author.
E-mail address: egass@unileon.es (E. Gascón).

http://dx.doi.org/10.1016/j.atmosres.2015.05.012

the drawbacks of radiosondes (Steinheimer and Haiden, 2007; Litta et al., 2012; Merino et al., 2013) is the numerical model, because it can forecast vertical profiles at each grid point in a given area with the desired spatiotemporal resolution.

Moreover, in recent years, various algorithms based on stability indexes have been developed, which are used as new tools in the prediction of severe storms. Various studies have used statistical regression equations for this purpose, which by combining a number of indexes increase predictability to more than that by use of each index individually. It is important to note that coefficients of these indexes in an equation are dependent on the geographic location and the season (Charba, 1979; Hughes, 2004; Sánchez et al., 2009). The same equation cannot be applied to a region without first validating it using real data. Specifically, binary logistic regressions can distinguish the risk probability of an episode by dividing the results of occurrence/no occurrence, and are widely used to predict severe storms with or without hail (Dubrovsky, 1994; Sánchez et al., 1998a, 1998b, 2009; Schmeits et al., 2005; López et al., 2007; Hanlon et al., 2014; Onderlinde and Fuelberg, 2014). However, the main drawback of this method is that it provides no information about the spatial distribution of storms.

Mathematical techniques such as cluster analysis are used to classify the spatial distribution of severe storm episodes. This type of technique permits classification of events defined by a set of variables. These clusters are characterized according to whether their components are similar to each other or different from components in other groups. This approach is widely used in meteorological studies to establish synoptic (Straus, 2010; García-Ortega et al., 2012), mesoscale (Merino et al., 2013, 2014) or thermodynamic (Tudurí and Ramis, 1997) patterns, to define a particular spatial distribution and explain certain behaviors of a weather event.

The main objective of the present study is to create a new tool for the hailstorms forecast in the Mid-Ebro Valley (MEV) of Spain. This region is characterized by a high frequency of such storms, more than 60 days annually between May and September (Merino et al., 2013). Since 1997, the Atmospheric Physics Group (GFA) of the University of León has been researching tools to assist in the prediction of severe storms, especially hailstorms. Authors such as Sánchez et al. (1998a, 1998b, 2000), López et al. (2007) and Merino et al. (2015) have used binary logistic regression to create these tools, using sounding or satellite data. García-Ortega et al. (2011) and Merino et al. (2013) used cluster analysis to classify synoptic and mesoscale patterns, respectively, that determine the distribution of severe hailstorms in the MEV, using data from numerical models.

Here, we use binary logistic regression to obtain the optimum stability indexes that identify pre-convective hailstorms events, from thermodynamic analysis using numerical models, in this case the Weather Research and Forecasting (WRF) mesoscale model. The results obtained with the regression equation are used in cluster analysis to classify the various spatial distributions given by the probability equation, which represent the thermodynamic behavior in the study area. By combining these techniques, we can create a prediction tool that complements the synoptic and mesoscale cluster analyses of García-Ortega et al. (2011) and Merino et al. (2013) toward predicting both the probability and distribution of hailstorms in the MEV.

The organization of this article is as follows. Section 2 provides a detailed description of the study area and database used for the prediction model and observation data. Section 3 describes in detail the methods, separated into four steps. The results of these steps are presented in Section 4. In Section 5, we discuss the results and briefly discuss earlier studies. Finally, Section 6 presents our conclusions.

2. Database

The database used consists of 100 days with severe storms (SSs) between 2001 and 2013 in the Aragón region (Spain), observed within a 140-km radius around Zaragoza Airport (41°39′58″ N, 1°02′30″ W). These SS days were chosen from a database of a C-band radar belonging to the GFA, which is very close to the airport. The criteria for selecting the studied SS days were the presence of convective precipitation (verified by radar) with a convection trigger occurring between 1200 and 1500 UTC during May–September and hail occurrence as verified by hail pads, radar data or ground observers. In addition, 100 days with no storms (NS) in the same area, randomly chosen between May and September over a period of 13 years (2001–2013), were used to create a binary logistic regression equation.

2.1. Study area

The study area corresponds to the so-called MEV and includes two major mountainous regions, the Pyrenees to the north and Iberian System to the south (Fig. 1). The northwest part of the area borders the Bay of Biscay and the southeast extends to the Mediterranean Sea, key features that predispose the area to the effects of air masses with different characteristics (García-Ortega et al., 2007). The study area was between 40.30°N and 43°N and between 2°W and 1°E (Fig. 1). The area has altitudes from sea level through 2300 m in the Iberian System and 3400 m in the Pyrenees. The climatic characteristics of the area are strongly linked to its complex topography (López et al., 2007).

2.2. Numerical model

We used the WRF model (Skamarock et al., 2005) in this study. Two nested grid domains were defined, with 27 and 9 km horizontal resolutions. Vertical resolution was 30 sigma levels. Initial and boundary conditions were obtained from National Centers for Environmental Prediction (NCEP) reanalysis data with 1° resolution. Temporal resolution of WRF model outputs was 3 h for the first domain and 1 h for the second. The duration of each simulation was 24 h, beginning at 0000 UTC each day. García-Ortega et al. (2011) and Merino et al. (2013) conducted researches with WRF for severe storms in the same study area. They concurred that for this type of event, the best parameterizations are the Goddard moisture (Tao et al., 1989; Tao and Simpson, 1993) and Kain–Fritsch cumulus (Kain, 2004) schemes, and so, these were used in our study. For the remaining parameterizations, we used the Dudhia shortwave scheme (Dudhia, 1989) and Noah land surface model (Chen and Dudhia, 2001).

From the WRF simulations, we obtained vertical profiles of temperature (T), dew point temperature (Td) and relative humidity (RH) at 1200 UTC each day at Zaragoza airport, which was the most representative moment of the atmospheric conditions before convective initiation (Merino et al., 2013) and any precipitation.

From the WRF vertical profiles (100 days SS and 100 days NS), 50 days from each group were used to construct the regression equation. The other 50 days were used to validate this regression equation.

A total of 31 stability indexes were calculated for each WRF vertical profile at Zaragoza airport for the 100 NS and 100 SS days, to be used as predictors in the binary logistic regression. The choice of stability index was based on studies that characterized these variables as good predictors of pre-convective conditions in the study area (Castro et al., 1992; Sánchez et al., 1998a, 1998b; López et al., 2007; Merino et al., 2013) and other regions of severe convection. Table 1 lists the 31 indexes, which characterize the vertical profiles for temperature, humidity, instability, precipitable water and helicity.

2.3. Observations and ground measurements

Profiles of T, Td and RH from soundings launched at 1200 UTC from Zaragoza airport were used to validate the model vertical profiles. Since