DISCRIMINANT ANALYSIS APPLIED ONTO HAIL DETECTION USING RADAR

López L., Sánchez J.L., García-Ortega E. and Marcos J. L.,

Laboratory of Atmospheric Physics, University of León, Spain, llopc@unileon.es

(Dated: April 20, 2007)

I. INTRODUCTION

There are a number of tools available – and more are currently being developed — to discriminate, within a particular storm, areas with hail precipitation using data provided by conventional meteorological radar systems (Waldvogel and Federer, 1979; Greene and Clark, 1972; Rasmussen and Wilhelmson, 1983; Kizmiller and Breindenbach, 1993; Witt et al., 1998). However, at a supraregional level most of the methods that have traditionally been employed to identify hailstorms often present ambiguous results. (Edward and Thompson, 1998). Caution is required when it comes to extrapolate the various identification models to areas other than the ones where they were developed.

On the other hand, many currently available systems for data extraction and treatment make it relatively easy to obtain a large number of variables derived from radar parameters for each storm analyzed and at different stages in its development. The questions are now: is it possible to select and/or classify these variables according to their ability to discriminate hailstorms from non-hail storms? And if so, would the combination of several of these variables enable us to develop new and improved discriminating tools? These issues have taken us to the *stepwise* method used to develop radar-based hail–detection-products implemented in the northeast of the Iberian Peninsula.

II. STUDY ZONE

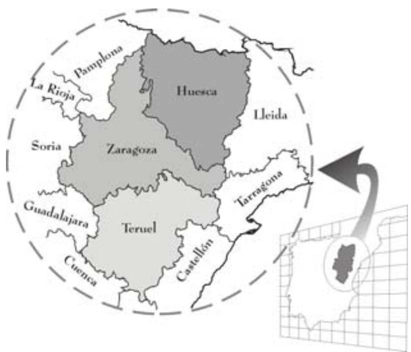

FIG. 1: Study zone.

The study zone (Fig. 1) lies in the northeast of the Iberian Peninsula, more precisely in the Valley of the River Ebro. It comprises nearly the whole of the Region of Aragón and part of the province of Lérida (about 40 kilometers of the area known as *Ponent de Lleida*).

This zone lies between 39º 51´ and 42º 55´ N, and 2º 06´ W - 0º 44´ E.

The region has an important convective activity, and the radar imagery shows that there are approximately 60 storm days every summer. The damages caused by hailstorms in the study zone amount annually to about 100 M€, which indicates the relatively high frequency of this phenomenon and their important economic impact.

The *Laboratory of Atmospheric Physics* at the University of León, Spain, owns a portable C-band radar which was installed in the study zone to gather data during the experimental campaign. The radar was set up 10 km SW of the city of Zaragoza, and its range was estimated at a radius of 140 km from that point. TITAN Software *(Thunderstorm, Identification, Tracking, Analysis and Nowcasting)* was used, providing data for a pixel size of 1 km, both horizontally and vertically.

With respect to the methodology employed, whenever a particular storm cell was seen to reach a maximum reflectivity of 35 dBZ or more, the local observer was phoned up to determine the type of precipitation registered on the ground, thus distinguishing between rain and hail. For this study a total of 729 towns and villages were included all over the study zone, and there was at least one meteorological informer in each.

The methodology described above provided a database of the *ground truth.* All in all, a radar image and the corresponding ground truth of 702 instances were gathered, 308 of which corresponded to hail precipitation and 394 to rain, according to the information provided by the *in situ* observers.

III. METHODOLOGY AND RESULTS

In discriminant analyses it is convenient that the independent variables fulfill a number of preliminary conditions: normality, linearity and multicollinearity. In addition, the model must comply with the assumption of homoscedasticity, and the population means of the two groups must differ significantly (Hair, 1999). Once these preliminary conditions had been noticed, the independent variables were selected that would lead to the most powerful and stable estimation possible.

The aim was to select the most appropriate variables to discriminate hail tracks from non-hail tracks, and the *stepwise* method was employed (Carrasco and Hernán, 1993). With this method the independent variables are introduced one by one in a discriminant function according to their discriminating power.

The whole sample of tracks was arbitrarily divided into two parts in order to set up the model. One part of the sample – formed by 276 non-hail track and 202 hail tracks - was used to set up the discriminant equation, and the remaining third of the total sample was later used to validate the results.

To control the stepwise introduction of the variables, the Squared Mahalanobis Distance was used. The minimum threshold value was established at 0.01, and the value 0.1 was established for eliminating variables (Hair, 1999). In each one of the iterations in the analysis the Wilks *lambda* decreased from 0.423 to 0.369 in the last step. The Fisher-Snedecor *F* statistic shows that these changes were significant in all the steps.

In the end, six variables fulfilled the minimum criteria to be considered significant discriminators on the basis of their Wilks *lambda* and the minimum values of Mahalanobis D^2.

The six variables selected (Altitude of maximum reflectivity, D dBZ max/dt, Top, Square root of VIL, Square root of the maximum reflectivity and Square root of the inclination) were included in the final discriminant function. The total variance explained by the discriminant function is 0.632 (Hair, 1999), and the function obtained shows a canonical correlation value of 0.795.

It must be taken into account that the chances of belonging to the hail group or the non-hail group need not be the same. Because of this, the Fisher's linear discriminant functions calculated for each group considered the various *a priori* possibilities for each of the groups studied.

In order to assess the forecasting power of the discriminant function, contingency tables were calculated for both the main sample and the validation sample. The validation sample consisted of 224 tracks. Finally, different precision indices were calculated to account for the goodness of the forecast.

The probability of detection is 0.923 in the main sample, and 0.868 in the validation sample. Even though the score is somewhat lower in the latter case, as expected, it can still be considered a very satisfactory result. The False Alarm Ratio was very low (0.08 and 0.123, respectively). The frequency of unforeseen events is 0.054, which guarantees that very few events will go undetected.

The results for HSS and TSS are 0.7685 and 0.7679 in the validation sample. The value of these indices lies between –1 and 1, which would correspond to a perfect forecast. It can be seen that the values found point towards a very good performance of our model.

IV. AKNOWLEDGMENTS

The present study has been supported by the CICYT Project SEVERUS (CGL2006-13372-C02-01).

V. REFERENCES

Carrasco, J. L. y M. A. Hernán; 1993: Estadística multivariante en las ciencias de la vida. Editorial Ciencia, Madrid, 347 pp.

Edwards, R. E., y Thompson, R. L., 1998: Nationwide Comparisons of Hail Size with WSR-88D Vertically Integrated Liquid Water and Derived Thermodynamic Sounding Data, *Wea. And Forecasting*, **13**, 277-285.

Greene, D. R., y R.A. Clark, 1972: Vertically integrated liquid- A new analysis tool. *Mon. Wea. Rev*., **100**, 548-552.

Hair J. F., R. E. Anderson, R. L. Tatham, W. C. Black, 1999: Análisis Multivariante. Prentice Hall Iberia. Madrid. 1999

Kitmiller, D. H., y J. P. Breidenbach, 1993: Probabilistic nowcasts of large hail based on volumetric reflectivity and storm environmental characteristics. Preprints, 26th Conf. on Radar Meteorology, Norman, OK, Amer. Meteor. Soc., 157-159

Rasmussen, E.N. y R.B. Wilhelmson, 1983: Relationship between storm characteristics and 1200 GMT hodographs low-level shear and stability. Preprints, 13th Conf. on Severe Local Storms. Tulsa, OK, Amer. Meteor. Soc., J5-J8.

Walvogel, A., Federer, B., and Grimm, P., 1979: Criteria for the Detection os Hail Cells, *J. Appl. Meteor.*, **18**, 1521-1525.