A MODEL FOR AVALANCHE FORECASTING ON THE BONAIGUA PASS, SPAIN USING CLASSIFICATION TREES

Pablo Gorospe[1], Ivan Moner[2]*, Montse Bacardit[2] and Jordí Gavaldà[2]

[1]Universitat Barcelona, Barcelona, Spain
[2]Centre de Lauegi, Conselh Generau d'Aran, Vielha, Spain

ABSTRACT: The highway C-28 is located in the Central Pyrenees and links the Aran valley with Catalonia along 20 km over the Bonaigua Pass. It constitutes a key access route for winter visitors. Most of the slopes affecting the road face to the south, with heights varying between 1600 and 2300 meters. We started from 12 years of meteorological and avalanche data collected by the local avalanche warning service of Aran Valley. Weather data were obtained from two automatic weather stations and a flowcapt, whereas avalanche activity was manually recorded in a GIS. We selected several weather parameters including snow drift, elapsed time, trend and categorical parameters. Using a classification tree method, we have developed a model to determine periods of significant avalanche activity in terms of the predefined avalanche day concept. The model is performed for the entire road in a combined analysis and also for three individual sub-areas within the Pass. Results showed that conventional factors describing snow depth were more significant than temperature and precipitation factors. Derived snow drift parameters from snow depth and water precipitacion showed more importance than drift data from the flowcapt. Radiation and wind direction variables had low importance in all the tests. The detailed analysis by sub-areas has not achieved the objective due to the reduction of the database. However, it has allowed to confirm the differences between one of the sub-areas and the dynamics of the rest of the highway.

KEYWORDS: Pyrenees, avalanche forecasting, modeling, classification trees.

1. INTRODUCTION

The C-28 highway is one of the two main roads that connects the Aran Valley with Catalonia. It is also a key communication route with France (Fig. 1). The Aran Valley has an oceanic climate with strong influence from the humid Atlantic Ocean flows that leave abundant precipitation in the valley. The average liquid precipitation is 1000 mm per year. Sometimes they can reach 1500 mm in heights above 2000 m. These are characterized by a homogeneous distribution throughout the year. During the winter season precipitation as snow above 2200 m reach between 600 and 700 cm per year, on average. The route of the C-28 through the Bonaigua Pass extends along 20 km and passes through 58 avalanche paths. Of the total, only 25 paths have a common activity. Most of these starting zones are only able to generate D2 or D3 avalanches, but in many places the road is carved in the hill side, which means that every single avalanche will certainly hit the road. The highest point of the route (2075 m) divides two climatically distinct sectors. On the one hand, the northern branch of the road has a marked influence of the Atlantic flows affecting the northern side of the Pyrenees. On the other hand, the southern branch is also dominated by the Atlantic rainfall but these are much less abundant. The topography of Ruda Valley, near Bonaigua Pass, channels flows from the north and northwest. This causes an intense transport of snow on the northern slope and generates wind slabs south of the pass (Gavaldà and Benet, 2014). Because of this the road suffers frequent cuts. Damage to vehicles and their occupants may also occur. These incidents have a direct impact on winter tourism, the main economic activity of the valley. On both faces of the route C-28 there are few defense works. Snow and avalanche wind fences have been set up in one of the existing avalanche prone areas on the southern leg, but at the remaining avalanche prone areas in this leg there are no protective structures. There is also a barrier system in order to close it down whenever avalanche risk exists. Until the 2003-2004 season, the management of the road was based on the regional avalanche risk bulletin that makes the Institut Cartografic i Geo-

* *Corresponding author address:*
Ivan Moner, Conselh Generau d'Aran
Passeg dera libertat 16. 25530 Vielha, Spain
tel: +34 973641801
email: i.moner@aran.org.

logic de Catalunya (ICGC) for the Catalan Pyrenees. In the 2003-2004 season the local avalanche forecasting was launched. The highway management has seen a significant improvement since it is based on the local forecasting (Gavaldà and Benet, 2014). The local prediction is based on data obtained from four automatic stations, daily snowpack observations and sporadic field outings. Since the 2008-2009 winter, preventive avalanche triggering has started, using the device known as DaisyBell® (Gavaldà and Moner, 2009). The prediction center has also made a detailed mapping of the starting zones affecting the road and the recorded avalanche events. They have differentiated seven sectors with different dynamics of avalanches. Each sector entails several starting zones with similar behavior (Fig. 2).

The main objective of this work is to use this database to determine the most significant meteorological variables and potentially, develop a model to predict avalanche days on this stretch of road C -28 using classification trees based on the algorithm of Breiman (1984). Knowledge of the dynamics of avalanches in different sectors of the route has allowed to distinguish three sub-areas (Fig. 2) with a more homogeneous behavior. In addition to the analysis for the whole route, an individualized study of the sub-areas has been carried out to determine differences with the general model and other relevant variables.

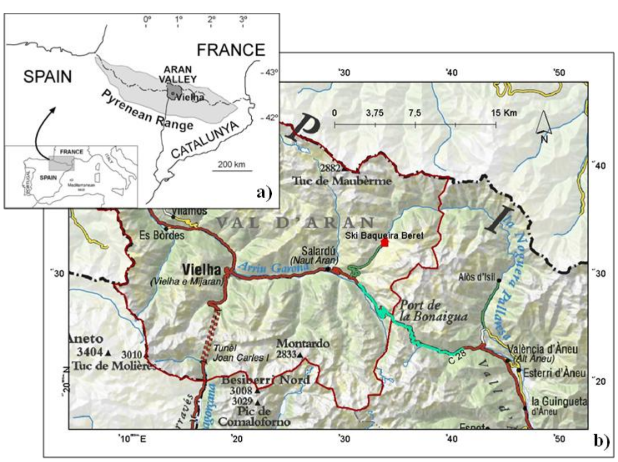

Fig. 1: a) Aran Valley situation. From: (Gavaldà and Benet, 2014) b) Situation of the C -28. Study zone passing through the Bonaigua pass. © Cartographic base property of Institut Cartogràfic i Geològic de Catalunya, disponible en www.icgc.cat

2. METHODOLOGY

The statistical tool chosen to develop the model is a classification tree developed in the work "Classification and regression trees" (CART) de Breiman (1984). This tool stands out over other methods of analysis for its ease of use and graphical representation of results. This simplifies processing and provides a more intuitive interpretation than other methods (Trujillo et al., 2008). Comparative studies between classification trees and other methods such as logistic regression are frequent. Misclassification rates in CART are usually larger than in logistic regression. Logistic regression gives better results in the case of samples with low correlation variables and small populations. However, when working with larger populations and variables with high correlation, misclassification rates decrease in CART models to approach the results obtained by logistic regression (Serna, 2009). Other comparative studies between statistical methods, as Díaz (2012), recommend using CART against the logistic regression in case of ignoring the functional form of the model for better results. There are many articles published in recent decades aiming to develop a statistical model to facilitate avalanche risk management. The tools of statistical analysis that the authors have chosen are varied: Multivariate discriminant function analysis (Bois et al., 1975) or analysis of variance and canonical discriminant analysis (Floyer and McClung, 2003) are just a few examples. Also classification and regression trees have been used in many occasions (Davis et al., 1992), (Davis and Elder, 1994), (Davis et al., 1999), (Rosenthal et al., 2002). More recently, Hendrikx et al. (2005) and Murphy and Hendrikx (2014) have complemented this method using cross-validation for test sample.

To implement a model of daily prediction, it is necessary to generate a daily database with meteorological variables and avalanche events. We performed a data processing of the 12 seasons, both meteorological and avalanche events. Weather data for analysis are taken from weather stations because they are the only source with a complete daily record since 2003. Of the four stations available (Fig. 2) only data from two stations has been used, Bonaigua and Flowcapt. Bonaigua station is in a small closed valley NW-SE direction; this means that the measured wind direction is greatly influenced by the topography. Because of this the parameters of speed and wind direction as well as transport of snow are taken from Flowcapt.

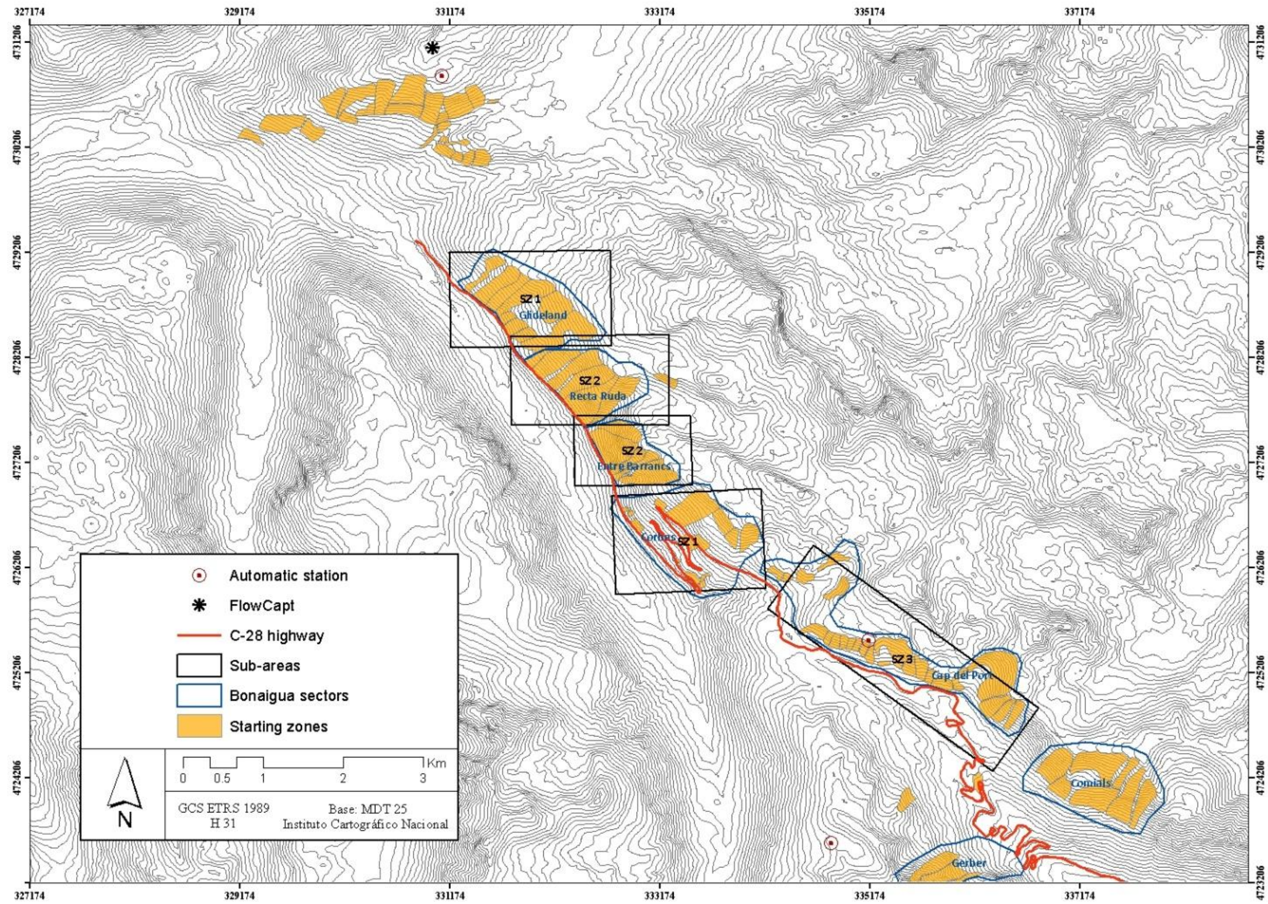

Fig. 2: Location map showing the C-28 Highway (Red) in Bonaigua Pass. Starting zones are show as orange polygons.

Once the primary variables between 2003 and 2015 were obtained from the automatic stations, the derived variables were calculated. These variables include both continuous and categorical data. Table 1 shows all the variables that have been used in the analysis and their characteristics.

On the other hand, avalanches that have occurred in the Bonaigua Pass since the winter of 2003 to present are collected in a GIS map. Each event is assigned an identification code, area code where avalanche has occurred, drop date, notes about the event and photographs.

The events that cut the road and the deposits located within 20 m of the road have been selected from the GIS map. This distance is set because it is the best to select events with enough entity to be included in the sample and it's in line with the criteria of the experts of the center. For this analysis an avalanche day has been defined as any day when at least one avalanche with the above criteria occurred. After this process there were a total of 177 avalanche days. Following the work Floyer and McClung (2003), the sectors where the pass was divided have been classified into three subareas (Fig. 2). Table 2 specifies what sectors are included and their characteristics. To develop the model we will use only the days between the 2004-2005 season and the 2014-2015 season. Each season only covers the months from November to April, inclusive. Of this time series some days should be excluded according to different criteria. First, non avalanches days included in periods when the road was closed were eliminated. Without access these days cannot be classified correctly. Also, those days lacking data or with mistakes were excluded. Finally, the avalanche days in which the DaisyBell was used were also excluded. According Floyer and McClung (2003) this type of days should be removed from the series since control operations are not regular and are only made in good weather episodes.After all filtering, total avalanche days were 137.

PARAMETER	VARIABLE NAME	TIME PERIOD
Average high temp.	AvHigh_XX	24 48,72
Average low temp.	AvLow_XX	24,48,72
High temp trend	High_temp_trend	
Low temp. trend	Low_temp_trend	
Total snow depth	Total_Snow	
Average new snow	AvSnow_XX	48,72
Sum new snow	SUMSnow_XX	24,48,72
Snowfall day	Day_snow	
Days since last snowfall	Day_since_Snow	
Average water	AvWater_XX	48,72
Sum of water	SumWater_XX	24,48,72
Wind direction for the prior 24h	Dir_24	
Average max wind speed	AvMaxSpeed_XX	24,48,72
Drift (water equivalent*ave wind speed)	Drift_XX	24,48,72
Drift (new snow depth*ave wind speed)	SD_Drift_XX	24,48,72
Sum drift Flowcat	SUMFlow_XX	24,48,72
Sum solar radiation	SumRad_XX	48,72
Relative humidity trend	HR_Trend	
Sum relative humidity	SumHRM_XX	48,72
Avalanches day before	AvaObs_1prev	

Table 1: Continuous and categorical meteorological parameters used in the analysis that were obtained from the Bonaigua automatic station and the flowcapt.

Sub-areas	Sectors	Avalanche dynamics	population
1	Glideland + corbes	Glides	165 days
2	Entre barrancs + Recta Ruda	Wet snow avalanches and recent snow avalanches	200 days
3	Cap del Port	Wind slabs	78 days

Table 2: Sub-areas of the pass with their characteristics and the total number of days that had learned the model in each analysis.

The last step is to ramdomly select a number of non-avalanche days equal to the previous. With an equal number of days of both types, the two classes have the same statistical weight. In this selection we must make sure that the days have been chosen uniformly throughout the time series. The statistical software, CART of Saldford Systems®, has been chosen to develop the classification trees. This is based on the algorithm developed in publishing Breiman (1984). CART is a non - parametric binary segmentation method which aims to generate homogeneous populations. The decision tree is developed from historical data and it's used to classify new data. To perform the verification test model a method of cross-validation and the Gini index as partition criterion have been used.

3. RESULTS

For the whole of the road two trees were obtained. The first was performed by introducing in the software all variables calculated. This first step allows us to observe which variables are relevant. In this way you can refine the model by adding only the most efficient variables to make the tree more accurate. In this first phase, a large tree with a large number of nodes was obtained. In Fig. 3 you can see the first branches of this tree. This tree uses the variables total_snow and sumsnow_72. After a 72-hour period if the thickness of new snow exceeds 18 cm it will be classified as avalanche day. Those days that do not exceed these values would be included in the node on the left. This node is again divided with the total_snow variable. The partition value 155.35 cm classifies those days with a greater thickness as avalanches days and those with lower value as non-avalanches days. The high statistical significance of the summation of snow for 72 hours corresponds with the dynamics of avalanches in the Pass. A snowfall period when a critical value of new snow is reached is related with more avalanches activity. This tree classified as avalanches day any day with more than 155 cm of total snow and with little amount of recent snow. This model cannot be used to manage the highway. This total thickness is reached annually in the automatic station Bonaigua at an advanced point in the season. This only tells us at what point of the season we find ourselves in. Avalanche activity will be more intense the more advanced the season is. The same conclusion was reached by the authors Floyer and McClung (2003).

To develop the next tree those non-significant variables were excluded as well as total_snow parameter. The result is the tree of Fig. 4. The percentage of successes it has achieved in the cross-validation test is 75.74 % for overall data, 80.00% for non-avalanche days and 71.53 % for avalanche days. This model uses variables sumsnow_72, drift_24, avaobs_1prev and Hr_trend. In the node 2 the days with a value less than 25.05 are the days when the snowfall has been sparse. By contrast those with greater value than 25.05 will present little snow but more than in node 3. From here the tree has two branches, each of which represents a common avalanche situation in Bonaigua Pass. The left branch that belongs to node 3 (Fig. 4) is characteristic of avalanches days of wet snow and glides very common in the hillsides facing south. The terminal node 2 has been classified as an avalanches day. In these days, wet snow avalanche could happen because the starting zones are covered of snow. Often in episodes of atmospheric stability glides left starting zones clean.

The right branch belonging to the node 4 represents dry snow avalanches. The Hr_trend variable represents the trend of the relative humidity in a 24 hour period. It is related to the arrival or the pass of a storm. The terminal node 3 represents the arrival of new rainfall. The terminal node 4 includes those days of atmospheric stability.

The three branches of the tree allow follow the evolution front rainfall. Each branch represents different situations, more rainfall or anticyclonic weather.

Sub-areas analysis obtained limited results due to the reduction of the population of the study in all three models. Subareas 1 and 2 include slopes with an avalanche activity very similar to the whole road. Therefore the model obtained in both cases has the same structure as the general tree (Fig. 5). Sub-area 3 is the most different. It is characterized by avalanches of new snow and wind slabs. The results of this tree are inconclusive because of the small population. However, the variables used (AvLow_48 and High_temp_trend) indicate that the behavior of this area is very different from the rest of the pass (Fig. 6). The structure of this tree has nothing to do with those set above.

4. DISCUSSION

The southern orientation of the slopes, the latitude and altitude at which the study area is, causes a higher incidence of wet snow avalanches. Based on this, the temperature was expected to be more relevant. However the importance of it has been limited. The explanation that has been found is that in days of high temperatures very often the starting areas have no snow and avalanches therefore not occur. In many cases the slopes are cleaned after snowfall with good weather and the snow height remains low until the next snowfall.

As mentioned earlier, Subarea 3 is the one with more differences with the rest of the route. However the small population has not allowed us to develop a consistent model for this area. All the "drift" variables were intended for use in this subarea. Of the different calculation methods, the most significant for the general model have been the Drift_XX. These are calculated from liquid precipitation. Those calculated from the thickness of new snow have given worse results. Neither transport values provided by the flowcapt presented statistical significance. Despite this, in the interpretation of the models it has been found that these "drift" parameters are an indicator of precipitation and not of transport. Cut values as low point in that direction.

The set of variables that have been shown most significant are those that express snow depth. In the first model for the whole route the total_snow variable presented a good level of classification, although it is not optimal for goals. In all trees sumsnow_72 is responsible for the first partition. Also the variables "drift" that indicate the amount of precipitation on the output zones. For all this, it is clear the need for a variable that expresses the amount of snow on the starting zones. Of the categorical variables that have been used only avaobs_1prev it has been significant both in the general model and sub-areas. The rest has not proved decisive. This variable is important because it confirms that the conditions for avalanches occur. This data is very important in predicting glides.

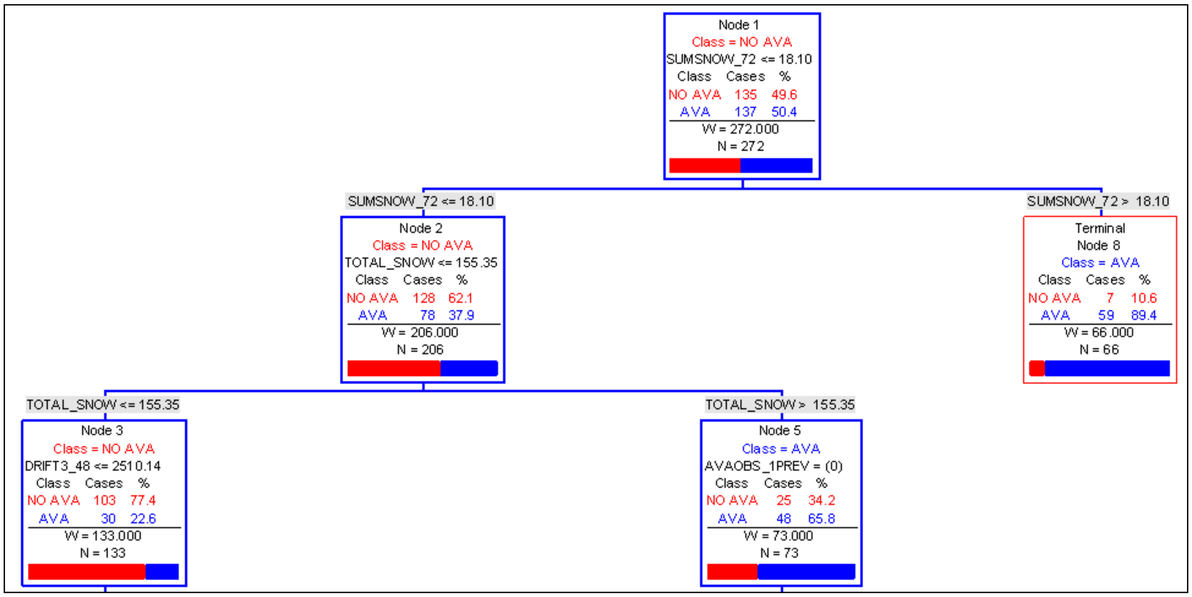

Fig. 3: First classification tree for the whole road.

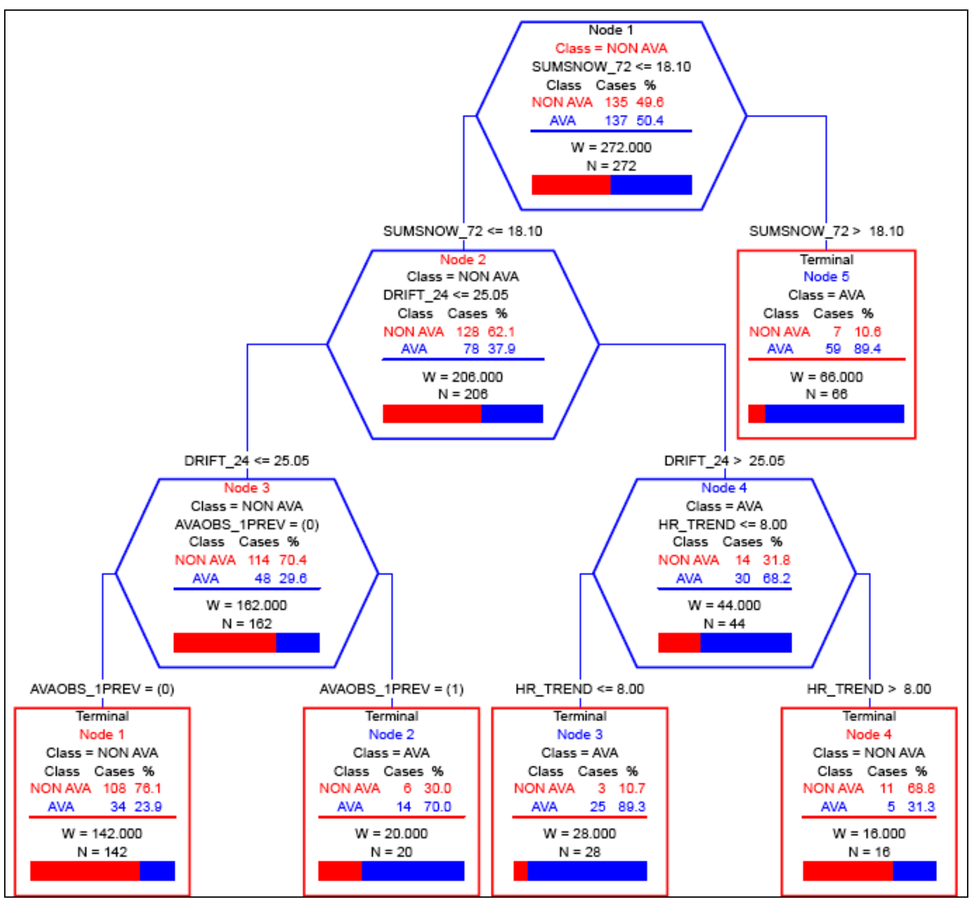

Fig. 4: Second classification tree for the whole road.

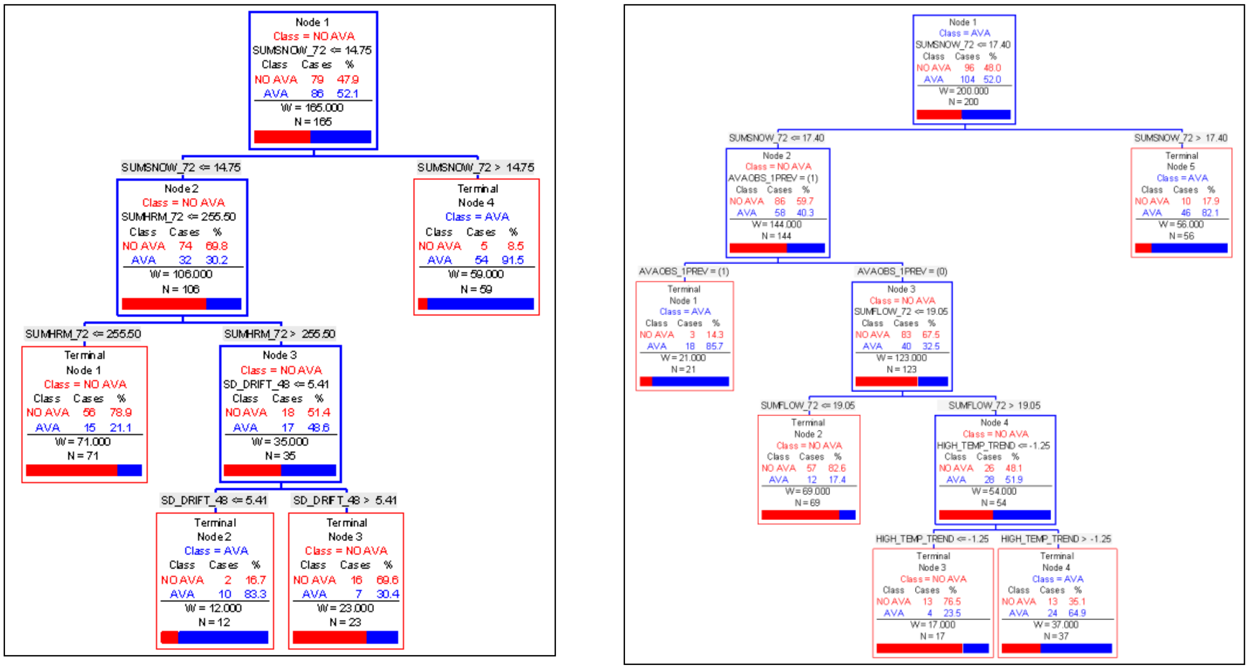

Fig. 5: Classification trees for sub-area 1 and 2.

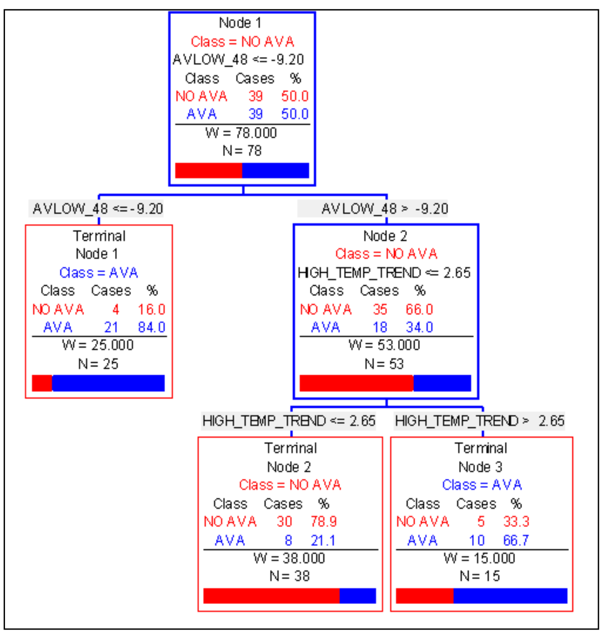

Fig. 6: Classification tree for sub-area 3.

The analysis by sub - areas has failed to generate different models to the general tree. This similarity is not surprising considering that the three datasets represent similar dynamics associated with wet snow avalanches. For sub-area 3, the only one that presents a notable different dynamic, the small study population has prevented to get its own model.

5. CONCLUSIONS

The tree implemented for the entire road has obtained a 75.77 % overall hit. A fairly high result considering that the model only works with data from automatic stations. Compared to other jobs with longer series of manual observations, the obtained model has more limited results. This makes visible the need to include manual observations in the future. To do this we must create a more continuous and significant snow data series.

In the model discussed, the various branches of its structure allow us to follow the different phases of a period of atmospheric instability, from its beginning to its end. Each of these branches represent the different situations of avalanches that occur in the pass, dry snow or wet snow avalanches after rainfall.

Sub-areas analysis has not achieved its goal. For sub-areas 1 and 2 the similarity of both models with the general model is justified because their avalanche dynamics are identical.

In the case of sub-area 3 the sparse population has prevented to obtain good results. The evidence represents a very different dynamic to the rest of the port which raises the need for further analysis with longer time series.

Drift_XX variables have proven to be the most significant within the set of different "drift" variables. However, its role in the models is to provide information of precipitation and it does not work as snow transport parameters. Transport variables provided by the flowcapt show no statistical significance for the overall model.

The exclusion of the temperature parameters of all models has been one of the main revelations of the analysis. However it is justified by the lack of snow that often starting zones have during episodes of rising temperatures.

Snowpack thick data have proved especially effective in the statistical analysis. The snow depth data provided by the Bonaigua automatic station differs greatly with the reality of the starting zones. Clearly, a complete time series of snow thickness in starting zones would significantly increase the effectiveness of the model.

ACKNOWLEDGEMENTS

We would like to thank the Institut Cartogràfic i Geològic de Catalunya and the Servei Meteorològic de Catalunya for the meteorological data. We also thank Dr Joan Manuel Vilaplana for his continuing support of this project.

REFERENCES

Bois, P., Obled, C., and Good, W., 1975, Multivariate data analysis as a tool for day-to-day avalanche forecast, *in* Proceedings of the International Symposium on Snow Mechanics, I.A.H.S., Grindelwald, Switzerland, p. 391–403.

Breiman, L. (Ed.), 1984, Classification and regression trees: Wadsworth International Group, Belmont, Calif.

Davis, R.E., and Elder, K., 1994, Application of classification and regression trees: Selection of avalanche activity indices at Mammoth Mountain, *in* Proceedings of International Snow Science Workshop, Snowbird, Utah, p. 285–294.

Davis, R.E., Elder, K., and Eddy Bouzaglou, 1992, Applications of classification tree methodology to the avalanche data management and forecasting, *in* Proceedings of International Snow Science Workshop, Breckenridge, Colorado, p. 126–133.

Davis, R.E., Elder, K., Howlett, D., and Bouzaglou, E., 1999, Relating storm and weather factors to dry slab avalanche activity at Alta, Utah, and Mammoth Mountain, California, using classification and regression trees: Cold Regions Science and Technology, v. 30, no. 1–3, p. 79–89, doi: 10.1016/S0165-232X(99)00032-4.

Díaz, J.F., 2012, Comparación entre árboles de regresión CART y regresión lineal [Master thesis]: Universidad Nacional de Colombia, 79 p.

Floyer, J.A., and McClung, D.M., 2003, Numerical avalanche prediction: Bear Pass, British Columbia, Canada: Cold Regions Science and Technology, v. 37, no. 3, p. 333–342, doi: 10.1016/S0165-232X(03)00074-0.

Gavaldà, J., and Benet, J., 2014, Ten years of avalanche forecasting on Bonaigua abd Beret roads, Aran Valley,Spain, *in* 17th International Road Weather Conference, Andorra.

Gavaldà, J., and Moner, I., 2009, Handling of avalanche risk for the “Bonaigua” and “Beret” roads, Aran Valley, Spain, *in* Proceedings of International Snow Science Workshop, Davos, Switzerland, p. 577–582.

Hendrikx, J., Owens, I., Carran, W., and Carran, A., 2005, Avalanche activity in an extreme maritime climate: The application of classification trees for forecasting: Cold Regions Science and Technology, v. 43, no. 1–2, p. 104–116, doi: 10.1016/j.coldregions.2005.05.006.

Murphy, M., and Hendrikx, J., 2014, Operational use of a statistical forecasting tool on the Seward Highway, Alaska, *in* Proceedings of International Snow Science Workshop, Banff, Alberta, Canada.

Rosenthal, W., Kelly Elder, and Robert E. Davis, 2002, Operational decision tree avalanche forecasting, *in* Proceedings of International Snow Science Workshop, Penticton, British Columbia, p. 152–158.

Serna, S.C., 2009, Comparación entre Árboles de regresión y clasificación y regresión logística [Master thesis]: Universidad Nacional de Colombia, 60 p.

Trujillo, J., Antonio Sarria-Santamera, Aureli Esquerda, Mariona Badia, Matilde Palma, and Jaume March, 2008, Aproximación a la metodología basada en árboles de decisión (CART). Mortalidad hospitalaria del infarto agudo de miocardio: Gaceta Sanitaria, v. Gac Sanit vol.22 no.1.