Probabilistic Forecast Method For Wildfires In Catalonia (1990-2005)

Bota-Moliner, L.
Altava-Ortiz, V.
Llasat, M. C.
Group of Analysis of Adverse Meteorological Situations (GAMA)
Department of Astronomy and Meteorology, University of Barcelona
Av. Diagonal 647, 08028 Barcelona, Spain
lbota@am.ub.es
valtava@am.ub.es
carmell@am.ub.es

Castro, F.X.
Montserrat, D
Forest Fire Prevention Service, Environmental Department, Catalonia Government
Finca Torreferrussa, Ctra. de Sabadell a Sta. Perpètua de Mogoda, Km. 4,5
08130 Sta. Perpètua de Mogola, Spain
francesc.castro@gencat.net
d.montserrat@gencat.net

Abstract: In Catalonia (NE Spain) the forest and wildland surface per inhabitant has increased in the last decades and the fight against wildfires focuses efforts from population and authorities. That makes necessary the development of a tool that allowed the prediction of meteorological situations enhancing fire risk. The probabilistic method proposed, the Analogous Method, is based on the sorting of meteorological situations similar to the current one over a selection scope delimited by two domains, so situations in the past that best reproduce situations in the present are obtained . In order to carry out this task, various meteorological parameters are taken into account, such as temperature, relative humidity and geopotencial height at 1000, 850 and 500 hPa, obtained from the NCEP/NCAR meteorological reanalysis for the period 1958-2005. Regarding to wildfire information, we have got two data series, which combined provide forest fire data from 1990 to 2005. At the present work, various episodes that took place in 2003 (when 700 incidents episodes burnt a total area of almost 10000 forest hectares) have been studied from that perspective. The selected summer days are 10 and 29 August, located in the context of the heat wave that affected great part of Europe, and, in addition, a winter day (31 January) linked to a cold-dry spell advection has been also analyzed. A validation of the Analogous Method outputs is carried out from different points of view: statistical parameters, relation with wildfire situations in the past and ROC curves. Results show that the method developed allows to identify, for forest fire events, more than 50% of analogous situations in the past with wildfire occurrence and improves up to 80% to the prediction of surface temperature thresholds overcoming.

Keywords: forest fire, analogous, probabilistic method, Catalonia

1. Introduction

In Catalonia (NE Spain), the forest and wildland surface per inhabitant has increased in the last decades, until reaching one of the highest values in Europe. Otherwise, wildfires have been and are still cause of worry due to the ecologic and economic damages that entail. As shown in figure 1, among news regarding to natural phenomena risks that appear in press, those related to forest fires and other agrometeorologic events hold the second position, after the floods category.

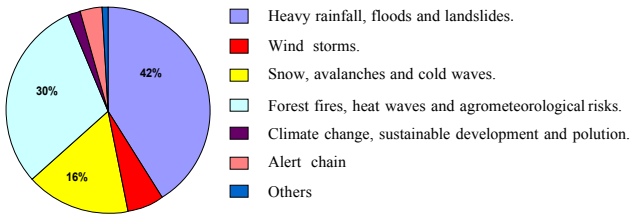

Figure 1. *Proportion of news appearing in press between 1985 and 2005 in Catalonia according to 7 kind of natural risks.*

No wonder why the fight against wildfires concentrates everyday more and more efforts from population and authorities. From the three aspects considered in this fight (that is, prediction, prevention and extinction), in the present work we will take advantage of the close relation between forest fires outbreak and meteorological conditions to improve the forecast of suitable situations for wildfire occurrence. This task will be carried out by means of the application of the Analogous Method. The main idea of this method was suggested by Lorentz (1969) as "Two atmospheric states are deemed "analogous" when there is certain resemblance between them" and was firstly applied by Electricité de France (Duband, 1980) for quantitative rainfall forecasting. From then on, the Analogous Method has been developed and spread to different countries and several application fields, such as temperature extremes in Netherlands (Kruizinga *et al.,* 1983) and cyclone tracks in Australia (Fraedrich *et al.*, 2003). The work we submit in this paper gives continuity to the task initiated the last decade in Spain (Llasat *et al.*, 1997; Gibergans-Báguena, 2001) and more recently developed by Altava-Ortiz *et al.*, 2006 and Bota, 2006. Next to this introduction section, in the present paper we will describe how the most appropriate variables have been selected and its database has been built (section 2). Hereafter, section 3 contains an accurate explanation of the Analogous Method. The results obtained will be shown in section 4, where the analysis of three forest fire events that took place in 2003 can be found, and in section 5, dedicated to validation by ROC curves. To finalize, conclusions and future work will be shown in section 6.

2. Data and variable selection

Forest fires have to do with different aspects of meteorology and climate, as shown in various studies carried out all around. For instance, Takle *et al.* (1994) came up with the connection between wildfire and geopotencial height in the NE US. In Australia, Speer *et al.* (1996) analyzed the meteorological conditions for wildfires of January 1994. Viegas *et al.* (1994) established in Portugal a relationship between rainfall and burned area. Now centred in our region study, Catalonia (NE Iberian Peninsula), we count on the accurate

contributions by Llasat (1997) and Montserrat (1998 and 1999), with special attention on the weather conditions at the initial stage of wildfires.

Based on prior studies, together with the experience on the daily prediction of fire risk by the Forest Fire Prevention Services, we set the variables and domains we will work with from two basic premises: geopotencial height fields at different levels in a wide space domain is useful to compare synoptic situations and the combination of geopotencial height, relative humidity and temperature at a smaller size domain contributes to identify thermodynamic phenomena at a regional level as well as to do a dynamic downscaling over Catalonia.

The first space domain, namely First Selection Window or FSW, covers a wide spatial range over Europe, in a synoptic scale, and the variables of work chosen in that domain are geopotencial height at 1000, 850 and 500 hPa. Contained in the first domain, another one, namely Second Selection Window or SSW, has been established in order to pay special attention to our study region and mesoscale phenomena that can take place. Geopotencial height, temperature and relative humidity at 850 hPa are the variables used in the SSW. Figure 2 summarizes the information above. A database has been built from the NCEP/NCAR meteorological reanalysis for the period 1958-2005. Regarding to wildfire information, we have got two data series, which combined provide forest fire data from 1990 to 2005 in Catalonia.

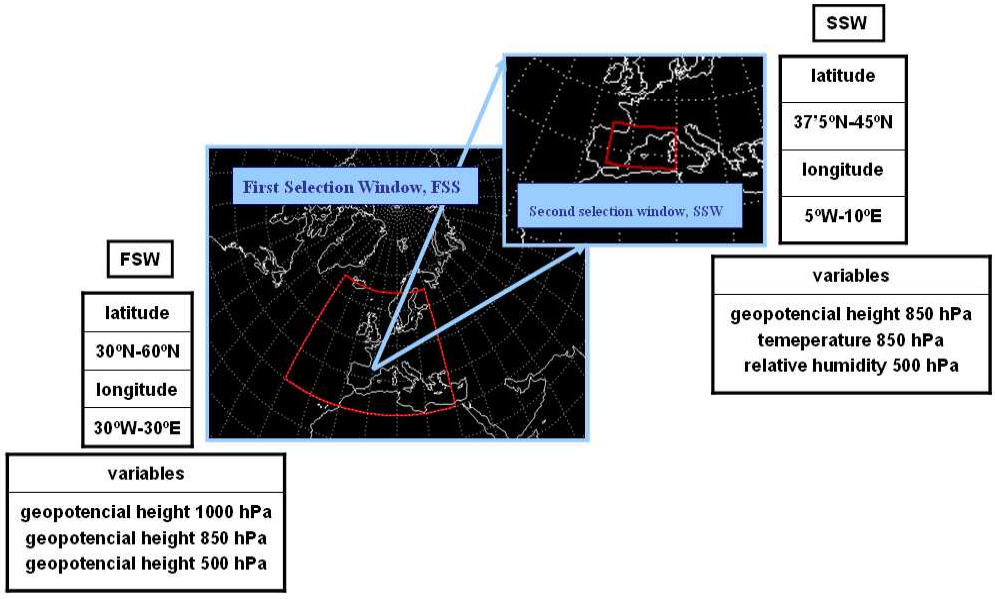

Figure 2. *Spatial domains, FSS and SSW, their geographic limits and meteorological variables corresponding to each one.*

As said above, the FSW is an extended area and represents a considerate number of points where data is available. This fact is an advantage in atmosphere description, but it holds a trouble due to the great number of freedom degrees. So great number of freedom degrees hinders the process, making the analogous criteria application not feasible. To decrease the number of freedom degrees and solve the problem, but trying to maintain a

good system description, a *principal component analysis* is applied. The initial six Principal Components (PC) keep more than 84% of system variance, which is accepted as a reasonable level. On the other hand, prior to the intervention of the SSW on the analogous selection, a *nearest neighbours interpolation* is performed on the vertical of Catalonia, so that, special importance is given to the study region, without losing surrounding information.

3. Methodology

The analogous method takes into account the weather observed in the past associated to certain meteorological contexts. When a meteorological situation is forecasted, several past situations, the analogous, are selected by applying some similarity criteria. After that, the model outputs are calculated using the meteorological observations corresponding to selected situations. Different criteria to consider two states as analogous can be applied in addition to different variables (as seen in section above) to do the selection. At the present work, two selection criteria take part in the selection of "analogous situations". The first one is a criterion of proximity, Euclidean distance, in a space of n dimensions (n is related with the number of freedom degrees). The second one is a criterion of correlation, Pearson's correlation coefficient, between variables that characterize an atmospheric state. Only the states that have parameters higher than a prefixed threshold are selected as analogous. The criterion of proximity, applied at the FSW, takes into account geopotencial height values at 1000, 850 and 500 hPa levels, which describe the meteorological state of the atmosphere from the synoptic point of view. The considered analogous situations are those that satisfy condition 1:

$$\sum_{j=1}^{6}\left(Z_{ij}^{1000} - Z_{tj}^{1000}\right)^2 < d^2 \qquad (1)$$

where d is the radius of a six-ball in a 6-dimensional space (figure XX illustrates that condition in a 2 dimension space), Zij is the value of the variable Zj on the day i and Ztj is the variable value on the forecasted day (test day).

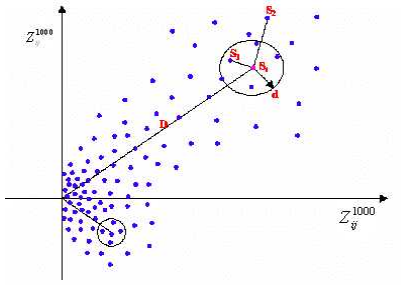

Figure 3. *Criterion of proximity. S_1 would be considered as analogous situation because lies inside the n-ball. S_2, instead, would be refused for being too far accordong to the established criterion.(Gibergans-Báguena., 2001)*

To calculate the distance "d" the procedure is as follows. The radius is taken as a function of the distance from the origin to the test day, we call it D, and is obtained as:

$$D^2 = \sum_{i=1}^{6}(Z_i^{1000})^2 \qquad (2)$$

Since $Z_1, Z_2 \ldots Z_n$ are considered independent and identically distributed, the variable D^2 follows a χ^2 distribution law with N degrees of freedom. For a law of χ^2 with *n* degrees of freedom, the function of density is given by the expression:

$$f(x) = \frac{x^{(n/2)-1}\, e^{-x/2}}{2^{n/2}\, \Gamma(\frac{n}{2})}, \quad with \quad \Gamma(n) = (n-1)! \tag{3}$$

Since just the first six PC have been retained, the density function takes the next form:

$$f(x) = \frac{x^2\, e^{-x/2}}{16} \tag{4}$$

that corresponds to a Gamma function with parameters (λ=3 and ρ=2). Any quartile of this function of density has a typical deviation σ_D proportional to the size of the sample, m, that is, to the number of analogous situations for each test day. If 30 analogous days are considered then:

$$\sigma_D = \frac{1}{f(x)}\sqrt{\frac{p(1-p)}{m}} = \frac{1}{f(x)}\sqrt{\frac{p(1-p)}{30}} \tag{5}$$

in which, p is the probability calculated by interpolation based on a table of the Gamma distribution with six degrees of freedom.

For a number of variables between six and eight, Duband (1980) proposes the following expression:

$$d^2 = \omega^2 \sigma_D^2 \quad \Rightarrow \quad d = \omega \sqrt{\frac{1}{f(x)}\sqrt{\frac{p(1-p)}{30}}} \tag{6}$$

in which ω is a parameter that should be increased if the number of analogous situations is lower than 30.

Once calculated the Euclidean distance, The "criterion of correlation" is applied to select those days that have high correlation coefficient with the test day. The correlation coefficient used is the linear Pearson's one and the values to calculate the correlation coefficient are 18 PC, 6 at every pressure level mentioned previously. The expression for the Pearson coefficient and the vectors containing the values are:

$$r_i = \frac{\sum_{j=1}^{18}(Z_{ij} - \overline{Z_i})(Z_{tj} - \overline{Z_t})}{\sqrt{\sum_{j=1}^{18}(Z_{ij} - \overline{Z_i})^2}\sqrt{\sum_{j=1}^{18}(Z_{tj} - \overline{Z_t})^2}} \tag{7}$$

$$\left(Z_{i1}^{1000}, Z_{i2}^{1000}, \ldots, Z_{i6}^{1000}, Z_{i1}^{850}, Z_{i2}^{850}, \ldots, Z_{i6}^{850}\, Z_{i1}^{500}, Z_{i2}^{500}, \ldots, Z_{i6}^{500} \right) \quad \text{for each day } i \tag{8}$$

$$\left(Z_{t1}^{1000}, Z_{t2}^{1000}, \ldots, Z_{t6}^{1000}, Z_{t1}^{850}, Z_{t2}^{850}, \ldots, Z_{t6}^{850}\, Z_{t1}^{500}, Z_{t2}^{500}, \ldots, Z_{t6}^{500} \right) \quad \text{for the test day.} \tag{9}$$

Candidate days that satisfied the next conditions (Duband, 1981) are selected as analogous:

$$u^2 = \frac{d^2}{r^2} < 6 \quad i \quad r^2 > 0.25 \tag{10}$$

If conditions as shown in (10) are too restrictive and the minimum analogous number is under 30, values of *d* will be increased in 0,5 units and values of *r* will be decreased in 0,03 units as many times as necessary until reaching the minimum desired, 30 analogous for each test day, in order to avoid the possibility of ignoring a similar day in the FSW but with low correlation and high d.

Next step in the analogous selection comes with the intervention of the SSW, where a second criterion of correlation is applied over the first step outputs. At that moment, the Pearson coefficient is calculated separately for each variable, geopotencial height, temperature a nd relative humidity at 850 hPa. Finally, best correlated days taking into account r_{gh} , r_t and r_{rh} are considered the most similar to the test day.

4. Study cases

The main forest fires recorded in Catalonia in 2003 are located in the context of three *critic forest fire episodes* (synoptic episodes during which meteorology conditions become in favour of forest fire activity). First event analyzed (31 January) fits in a *North situation*, the second one (10 August) took place in a *South situation* and the third one (29 August) corresponds to a West Situation, known as Ponent in Catalonia. Figure 4 shows the location of each event and the municipalities affected.

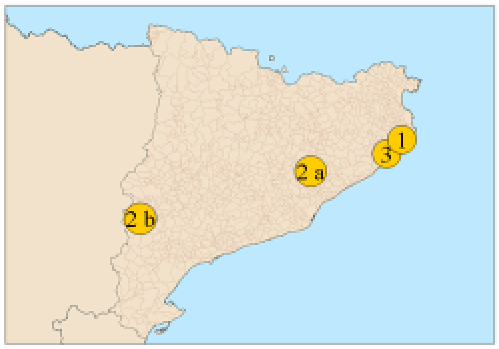

Figure 4. *Map of Catalonia (NE Spain).Circled numbers locate the area affected in the corresponding study case.*
1: Santa Cristina d'Aro and St. Feliu de Guíxols (Baix Empordà)
2 a: St. Llorenç Savall (Vallès Occidental)
2 b: La Granja d'Escarp i Seròs (Segrià)
3: Castell-Platja d'Aro, Calonge and Santa Cristina d'Aro (Baix Empordà)

31 January 2003 event was recorded in the middle of an intense North situation; actually it was one of the strongest North situations that have taken place in the last decade. The complete event lasted from 23 January to 5 February. Topography at 850 hPa has revealed that at UTC North wind speed exceeded 20 m/s over Catalonia. Besides, the evolution of meteorology variables in its entirety shows that the wind reinforces after the pass of a frontal system. This episode was exceptional as much its persistence as its intensity. In consequence, almost 400 ha were burned, road closures were necessary and power cuts occurred.

Regarding to the analogous obtained in the FSW, 57% are connected with other forest fire events in the past. Without any restriction in the analogous search, most of the analogous dates obtained belong to wintertime, that shows the capability of the Analogous Method to discriminate between situations. Now from the SSW point of view, as it can be

observed in table 1, geopotencial height was the meteorology variable that in this case pointed to wildfire occurrence in the past by the analogous selection.

analogous date	geopotencial height Pearson coefficient	number of forest fires	burned area (Ha)
11-4-1994	0,9937	6	108
29-1-2001	0,9936	1	0
12-3-1998	0,9884	4	8
14-2-1996	0,9880	0	0
9-12-1992	0,9852	0	0

Table 1. *Best 5 analogous dates obtained for 31 January 2003 by geopotencial height at 850 hPa in the SSW and forest fire activity related to them.*

The second case study analyzed, 10 August 2003, actually went on from 3 until 15 August and, as it can be seen in figure 4, different regions were affected in Catalonia. IT was not a remarkable episode concerning temperatures, but relative humidity made the difference, as it will be shown in table 2. From South situations like this one, paradoxically, either floods or forest fires can be expected. As a result of the wildfires that finally took place, more than 4000 ha were burned and 5 members of a family were killed.

As to the 30 best analogous in the FSW, 93% had recorded forest fire activity in the past. That result really makes us aware of what may happen if weather conditions persist, as it happened indeed. Among the best five analogous selected by the SSW taking into account relative humidity (the key meteorology variable in this event) it is remarkable 22 July 1991, when a great forest fire took place.

analogous date	relative humidity Pearson coefficient	number of forest fires	burned area (Ha)
24-6-1994	0,9670	18	54
1-6-2002	0,9410	1	0
1-8-1993	0,9272	5	6
22-7-1991	0,9075	656	7
17-7-1999	0,8973	5	15

Table 2. *Best 5 analogous dates obtained for 10 August 2003 by relative humidity at 850 hPa in the SSW and forest fire activity related to them.*

The third case study we show in this paper, 29 August 2003, took place in a West situation context (Ponents), widely known and analyzed in Catalonia due to the problems it usually generates in specific regions. These situations are especially dangerous in Catalonia in late summer, because winds coming from west, after crossing great part of the Iberian Peninsula, bring warm and dry air masses all over the country. This was a brief episode and a quick response was necessary, despite the changing weather conditions made difficult its prediction. Finally, because of that event, 332 ha were burned, which is not a great extension, but the evacuation of 12000 people that involved had an important social impact.

Given the importance of regional conditions, only 53% of best analogous in the FSW were linked to fire occurrence in the past. Nevertheless, results obtained in the SSW

prove the usefulness of the analogous search and selection by combining both spatial domains. Below these lines, tables 3 and 4, provide information about the final selection of analogous, which pointed to an important episode in the past.

analogous date	relative humidity Pearson coefficient	number of forest fires	burned area (Ha)
14-9-1993	0,9670	8	10
3-5-1996	0,9488	0	0
14-9-1994	0,9464	7	5211
11-8-1994	0,9436	16	772
4-7-2000	0,9264	1	5

Table 3. *Best 5 analogous dates obtained for 29 August 2003 by geopotencial height at 850 hPa in the SSW and forest fire activity related to them.*

analogous date	relative humidity Pearson coefficient	number of forest fires	burned area (Ha)
17-5-1994	0,9964	0	0
6-10-1993	0,9957	1	0
9-8-1999	0,9952	14	31
14-5-1994	0,9915	1	2
10-8-1994	0,9902	22	9160

Table 4. *Best 5 analogous dates obtained for 29 August 2003 by temperature at 850 hPa in the SSW and forest fire activity related to them.*

There was an exceptional event in August 1994 that burned a total area bigger than 10.000 ha. This is the episode in the past that appears among the best 5 analogous days to our case study, from geopotencial and temperature in the SSW point of view. In 2003 a similar situation could have happen again, but there were other factors, such as state of the vegetation, that fortunately didn't facilitate fire spreading.

5. Validation

The analogous search process is subjected to evaluation by ROC curves (Relative Operation Characteristic), which are diagrams of detection and false alarm probabilities. That kind of validation has been applied at as diverse fields as medicine, Swets *et al.* (2000), and education, Bacallao (1997).

Before starting the validation it is needed to see what the most applying variable to this task is. In our case, always keeping in mind that, beside meteorology aspects, there are other factors that play an important role in forest fire genesis, we have chosen surface temperature.

The obtaining of ROC curves requires firstly the construction of a data file with the surface temperature values of the best analogous days selected from the reference period (1/january/1990- 31/december/2005) for each day in the validation period (1/april/2003-30/september/2003). From now on, the analogous values will be considered as predictions for the test day values, which are supposed to be observations. Next to it, a Gaussian

distribution function is adjusted and percentages are calculated. Afterwards comes the filling of a contingence table, from the comparison between a threshold value, predictions and observations. In the present work, the threshold for surface temperature is fixed at 30ºC, since overcoming or not of that value can be very influent on forest fire outbreak in the study region (personal communications). Finally, detection and false alarm probability is obtained from the contingence tables for each grid point and plotted to give way to ROC curves, as shown in figures 5 and 6.

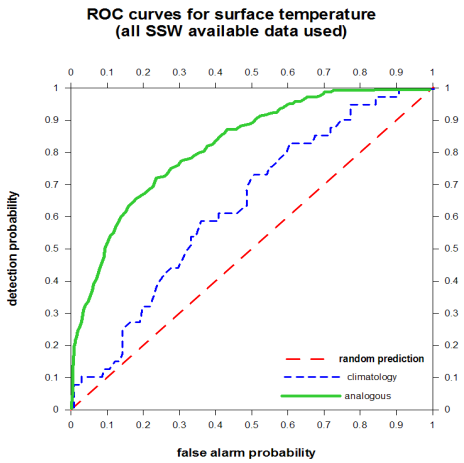

Figure 5 . *ROC curves obtained for surface temperature taking into account all grid points in the SSW.*

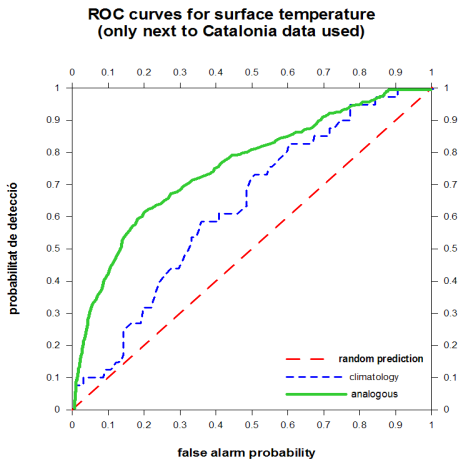

Figure 6 . *ROC curves obtained for surface temperature considering just the closest area surrounding Catalonia.*

The fact that the curve corresponding to the analogous (straight green) is, in most of the range from 0 to 1, above the climatology curve (dashed blue), means that the prediction of overcoming a threshold temperature or not is better when an analogous selection is carried out before the estimation of the temperature. The area under the curves is an objective value to assess that improvement. In the case where all available data in the SSW is used (figure5), the area under climatology is 0'61 and the area under analogous is 0'76, so an improvement of 15% is obtained. If we focus in the area surrounding Catalonia (figure 6), the area under the curves goes from 0'67 to 0'80, which means an improvement of 13%.

6. Conclusions and future work

The conclusions we have drawn from the analysis of results obtained are shown next.

The selection of analogous situations, besides linked to wildfire occurrence in the past, improves the forecasting of meteorological situations likely to give rise to forest fire in the present or in short-term. As shown by the ROC curves, concerning to surface temperature, the improvement with regard to climatology can increase from 67% to 80%.

The space domains FSW and SSW complement each other in the analogous search and selection. The First Selection Window acts as a primary filter and emphasizes in the analogous in which synoptic conditions is the most remarkable fact. Next to the first selection, the intervention of the Second Selection Window provides an accurate selection and highlights thermodynamic aspects at a more regional scale.

Relative humidity is, in general, the variable with the lowest correlation coefficients, due to its intrinsic spatial inhomogeneity. Nevertheless, is proved to be useful to discriminate between synoptic conditions likely to produce either forest fires or floods in our study region.

Concerning the case studies shown in this paper and others not included, it can be stated that among the best 5 analogous finally selected for each variable there is always at least one that reproduces accurately the situation of the test day in Catalonia. Regarding to the connection between analogous dates and forest fires events in the past: among the 30 best analogous selected by the FSW, from 53% up to 93 % coincide with wildfire occurrence and this percentage goes from 60% to 93% as for the 15 analogous selected by the SSW.

Automatic search of analogous and cartography generation helps at daily operative prediction, as well as at research and diagnose task.

The satisfactory results achieved and the chances that reveals the Analogous Method reinforce a line of future work that might start by the introduction of new predictor fields, comparison between predictions from D to D+5 obtained by the analogous method and mesoscale numerical model outputs and obtaining variables validation based on observational data from meteorological stations.

Acknowledgements. The authors acknowledge the European INTERREG III-B AMPHORE project GRIMFOMET+MEDIFIRE and the Catalonian Environmental Department for making possible the development of the present work. We also want to thank Antoni Tudela from the Wildfires Prevision Centre, for providing the data base which is used at the present work, as well as the staff of the Spanish environmental ministry for supplying information regarding forest fire occurrence. Thanks to our colleagues in the research team, GAMA (Grup d'Anàlisi de situacions Meteorològiques Adverses), for their support in several aspects during the development of that work.

7. References

Altava-Ortiz, A., Barrera, A., Llasat, M.C., Prat, M.A., Gibergans-Bàguena, J., Barnolas, M., 2006. Application of the MM5 and the analogous method to heavy rainfall event, the case of 16-18 October 2003 in Catalonia (NE Spain). Advances in Geosciences, 7, 313–319.

Bacallao, J., 1997. Las curvas ROC y las medidas de detectabilidad para la validación de predictors del rendimiento docente. Rev. Cubana Educ. Med. Sup., 10 (1).

Bota Moliner, L., 2006. Millora de la predicció de situacions meteorològiques favorables als grans incendis forestals a Catalunya mitjançant el mètode dels anàlegs. Treball per a l'obtenció del Diploma d'Estudis Avançats. Departament d'Astronomia i Meteorologia, Universitat de Barcelona. Publicació interna.

Duband, D., 1980. Dynamic selection of analogue flow patterns to produce quantitative precipitation forecasts. WMO symposium on Probabilistic and statistical methods in weather forecasting, Nice-France, 487- 492.

Fraedrich, K., Raible, C.C., Sielman, F., 2003. Analog ensemble forecasts of tropical cyclone tracks in the Australian region. Weather and Forecasting, 18, 3-11.

Gibergans-Báguena, J., 2001. Métodos hidrometeorológicos aplicados a la predicción de precipitación diaria en Cataluña. Universidad de Barcelona, Tesis Doctoral, Publicación Interna.

Kruizinga, S., Murphy, A.H., 1983. Use of an analogue procedure to formulate objectively probabilistic temperature forecasts in the Netherlands. Monthly Weather Review, 111, 2244-2254.

Llasat Botija, M.C., 1997. Meteorología agrícola i forestal a Catalunya. Secció d'estadística i informació agrària. Gabinet tècnic, DARP.

Llasat, M.C., Gibergans Báguena, J., Guilgaud, S., Rodríguez, J.Y., 1997. The use of the vertical thermodynamic data in the daily rain classification. Application to the improvement of the Analog Technique. Actas de INM/WMO International Symposium on Cyclones and Hazardous Weather in the Mediterranean Regions, pp 395-402.

Lorentz, E.N., 1969. Atmospheric predictability revealed by naturally occurring analogues. J. of Atmospheric Sciences, 26, 636-646.

Montserrat, D.,1999. La predicció dels grans incendis: Situacions sinòptiques relacionades amb l'inici de grans incendis forestals a Catalunya (1980-1998). Dins: Peix, J. Foc Verd II. Barcelona, Direcció General del Medi Natural, DARP, 166-180.

Speer, M.S., Leslie, L.M., Colquhoun, J.R., Mitchell, E., 1996. The Sydney Australia Wildfires of January 1994- Meteorological conditions and high resolution numerical modelling experiments. International Journal of Wildland Fire 6(3) 145 – 154.

Takle, E., Bramer, D.J., Heilman, W.E., Thompson, M.R., 1994. A synoptic climatology for forest-fires in the NE US and future implications from GCM simulations. International Journal of Wildland Fire 4(4) 217 – 224.

Swets, J. A., Dawes, R. M., Monahan, J., 2000. Better decisions through science. Scientific American, 82-87.

Viegas, D.X., Viegas, M.T., 1994. A relationship between rainfall and burned area for Portugal. Int. J. Wildland Fire (4): 11-16.