A Data Mining Approach to Predict Forest Fires using Meteorological Data

Paulo Cortez[1] and Aníbal Morais[1]

Department of Information Systems/R&D Algoritmi Centre, University of Minho,
4800-058 Guimarães, Portugal,
pcortez@dsi.uminho.pt
WWW home page: http://www.dsi.uminho.pt/~pcortez

Abstract. Forest fires are a major environmental issue, creating economical and ecological damage while endangering human lives. Fast detection is a key element for controlling such phenomenon. To achieve this, one alternative is to use automatic tools based on local sensors, such as provided by meteorological stations. In effect, meteorological conditions (e.g. temperature, wind) are known to influence forest fires and several fire indexes, such as the forest Fire Weather Index (FWI), use such data. In this work, we explore a Data Mining (DM) approach to predict the burned area of forest fires. Five different DM techniques, e.g. Support Vector Machines (SVM) and Random Forests, and four distinct feature selection setups (using spatial, temporal, FWI components and weather attributes), were tested on recent real-world data collected from the northeast region of Portugal. The best configuration uses a SVM and four meteorological inputs (i.e. temperature, relative humidity, rain and wind) and it is capable of predicting the burned area of small fires, which are more frequent. Such knowledge is particularly useful for improving firefighting resource management (e.g. prioritizing targets for air tankers and ground crews).
Keywords: Data Mining Application, Fire Science, Regression, Support Vector Machines.

1 Introduction

One major environmental concern is the occurrence of forest fires (also called wildfires), which affect forest preservation, create economical and ecological damage and cause human suffering. Such phenomenon is due to multiple causes (e.g. human negligence and lightnings) and despite an increasing of state expenses to control this disaster, each year millions of forest hectares (ha) are destroyed all around the world. In particular, Portugal is highly affected by forest fires [7]. From 1980 to 2005, over 2.7 million ha of forest area (equivalent to the Albania land area) have been destroyed. The 2003 and 2005 fire seasons were especially dramatic, affecting 4.6% and 3.1% of the territory, with 21 and 18 human deaths.

Fast detection is a key element for a successful firefighting. Since traditional human surveillance is expensive and affected by subjective factors, there has been an emphasis to develop automatic solutions. These can be grouped into three major categories [1]: satellite-based, infrared/smoke scanners and local sensors (e.g. meteorological). Satellites have acquisition costs, localization delays and the resolution is not adequate for

all cases. Moreover, scanners have a high equipment and maintenance costs. Weather conditions, such as temperature and air humidity, are known to affect fire occurrence [15]. Since automatic meteorological stations are often available (e.g. Portugal has 162 official stations), such data can be collected in real-time, with low costs.

In the past, meteorological data has been incorporated into numerical indices, which are used for prevention (e.g. warning the public of a fire danger) and to support fire management decisions (e.g. level of readiness, prioritizing targets or evaluating guidelines for safe firefighting). In particular, the Canadian forest Fire Weather Index (FWI) [24] system was designed in the 1970s when computers were scarce, thus it required only simple calculations using look-up tables with readings from four meteorological observations (i.e. temperature, relative humidity, rain and wind) that could be manually collected in weather stations. Nevertheless, nowadays this index highly used not only in Canada but also in several countries around the world (e.g. Argentina or New Zealand). Even though Mediterranean climate differs from those in Canada, the FWI system was correlated with fire activity in southern Europe countries, including Portugal [26].

On the other hand, the interest in Data Mining (DM), also known as Knowledge Discovery in Databases (KDD), arose due to the advances of Information Technology, leading to an exponential growth of business, scientific and engineering databases [8]. All this data holds valuable information, such as trends and patterns, which can be used to improve decision making. Yet, human experts are limited and may overlook important details. Moreover, classical statistical analysis breaks down when such vast and/or complex data is present. Hence, the alternative is to use automated DM tools to analyze the raw data and extract high-level information for the decision-maker [10].

Indeed, several DM techniques have been applied to the fire detection domain. For example, Vega-Garcia et al. [25] adopted Neural Networks (NN) to predict human-caused wildfire occurrence. Infrared scanners and NN were combined in [1] to reduce forest fire false alarms with a 90% success. A spatial clustering (FASTCiD) was adopted by Hsu et al. [14] to detect forest fire spots in satellite images. In 2005 [19], satellite images from North America forest fires were fed into a Support Vector Machine (SVM), which obtained a 75% accuracy at finding smoke at the 1.1-km pixel level. Stojanova et al. [23] have applied Logistic Regression, Random Forest (RF) and Decision Trees (DT) to detect fire occurrence in the Slovenian forests, using both satellite-based and meteorological data. The best model was obtained by a bagging DT, with an overall 80% accuracy.

In contrast with these previous works, we present a novel DM forest fire approach, where the emphasis is the use of real-time and non-costly meteorological data. We will use recent real-world data, collected from the northeast region of Portugal, with the aim of predicting the burned area (or size) of forest fires. Several experiments were carried out by considering five DM techniques (i.e. multiple regression, DT, RF, NN and SVM) and four feature selection setups (i.e. using spatial, temporal, the FWI system and meteorological data). The proposed solution includes only four weather variables (i.e. rain, wind, temperature and humidity) in conjunction with a SVM and it is capable of predicting the burned area of small fires, which constitute the majority of the fire occurrences. Such knowledge is particularly useful for fire management decision support (e.g. resource planning).

The paper is organized as follows. First, we describe the forest fire data in Section 2. The adopted DM methods are presented in Section 3, while the results are shown and discussed in the Section 4. Finally, closing conclusions are drawn (Section 5).

2 Forest Fire Data

The forest Fire Weather Index (FWI) is the Canadian system for rating fire danger and it includes six components (Figure 1) [24]: Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI), Buildup Index (BUI) and FWI. The first three are related to fuel codes: the FFMC denotes the moisture content surface litter and influences ignition and fire spread, while the DMC and DC represent the moisture content of shallow and deep organic layers, which affect fire intensity. The ISI is a score that correlates with fire velocity spread, while BUI represents the amount of available fuel. The FWI index is an indicator of fire intensity and it combines the two previous components. Although different scales are used for each of the FWI elements, high values suggest more severe burning conditions. Also, the fuel moisture codes require a memory (time lag) of past weather conditions: 16 hours for FFMC, 12 days for DMC and 52 days for DC.

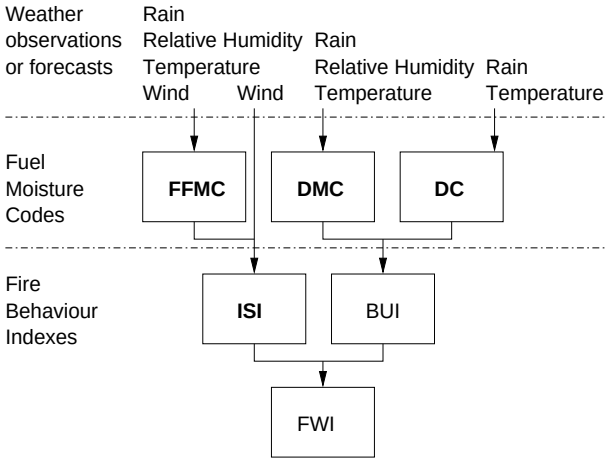

Fig. 1. The Fire Weather Index structure (adapted from [24])

This study will consider forest fire data from the Montesinho natural park, from the Trás-os-Montes northeast region of Portugal (Figure 2). This park contains a high flora and fauna diversity. Inserted within a supra-Mediterranean climate, the average annual temperature is within the range 8 to 12°C. The data used in the experiments was collected from January 2000 to December 2003 and it was built using two sources. The first database was collected by the inspector that was responsible for the Montesinho fire occurrences. At a daily basis, every time a forest fire occurred, several features

were registered, such as the time, date, spatial location within a 9×9 grid (x and y axis of Figure 2), the type of vegetation involved, the six components of the FWI system and the total burned area. The second database was collected by the Bragança Polytechnic Institute, containing several weather observations (e.g. wind speed) that were recorded with a 30 minute period by a meteorological station located in the center of the Montesinho park. The two databases were stored in tens of individual spreadsheets, under distinct formats, and a substantial manual effort was performed to integrate them into a single dataset with a total of 517 entries. This data is available at: `http://www.dsi.uminho.pt/~pcortez/forestfires/`.

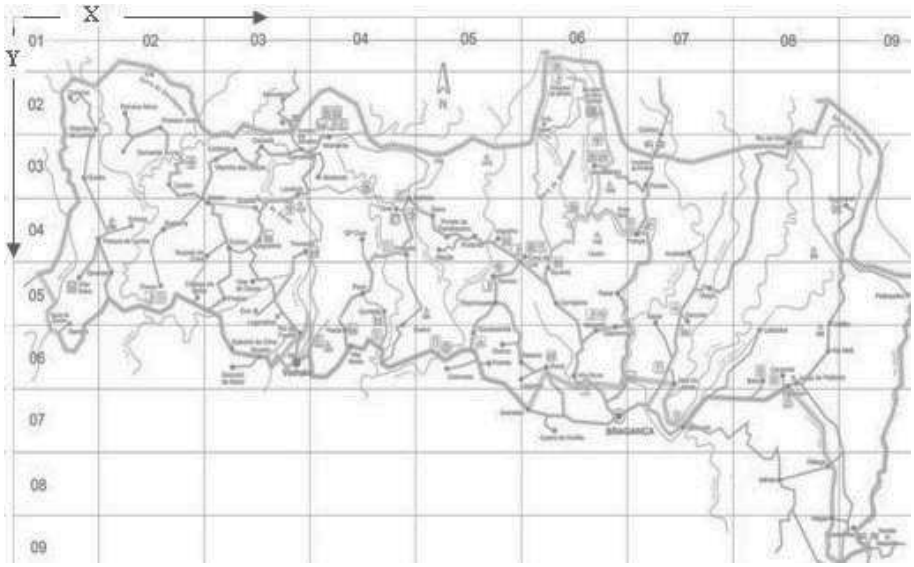

Fig. 2. The map of the Montesinho natural park

Table 1 shows a description of the selected data features. The first four rows denote the spatial and temporal attributes. Only two geographic features were included, the **X** and **Y** axis values where the fire occurred, since the type of vegetation presented a low quality (i.e. more than 80% of the values were missing). After consulting the Montesinho fire inspector, we selected the **month** and **day** of the week temporal variables. Average monthly weather conditions are quite distinct, while the day of the week could also influence forest fires (e.g. work days vs weekend) since most fires have a human cause. Next come the four FWI components that are affected directly by the weather conditions (Figure 1, in bold). The BUI and FWI were discarded since they are dependent of the previous values. From the meteorological station database, we selected the four weather attributes used by the FWI system. In contrast with the time lags used by FWI, in this case the values denote instant records, as given by the station sensors when the fire was detected. The exception is the **rain** variable, which denotes the accumulated precipitation within the previous 30 minutes.

The burned **area** is shown in Figure 3, denoting a positive skew, with the majority of the fires presenting a small size. It should be noted that this skewed trait is also present in other countries, such as Canada [18]. Regarding the present dataset, there are 247 samples with a zero value. As previously stated, all entries denote fire occurrences and zero value means that an area lower than $1ha$/100 = $100m^2$ was burned. To reduce skewness and improve symmetry, the logarithm function $y = ln(x + 1)$, which is a common transformation that tends to improve regression results for right-skewed targets [20], was applied to the **area** attribute (Figure 3). The final transformed variable will be the output target of this work.

Table 1. The preprocessed dataset attributes

Attribute	Description
X	x-axis coordinate (from 1 to 9)
Y	y-axis coordinate (from 1 to 9)
month	Month of the year (January to December)
day	Day of the week (Monday to Sunday)
FFMC	FFMC code
DMC	DMC code
DC	DC code
ISI	ISI index
temp	Outside temperature (in $^{\circ}$C)
RH	Outside relative humidity (in %)
wind	Outside wind speed (in km/h)
rain	Outside rain (in mm/m^2)
area	Total burned area (in ha)

3 Data Mining Models

A regression dataset D is made up of $k \in \{1, ..., N\}$ examples, each mapping an input vector $(x_1^k, \ldots, x_A^k)$ to a given target y_k. The error is given by: $e_k = y_k - \widehat{y}_k$, where $\widehat{y}_k$ represents the predicted value for the k input pattern. The overall performance is computed by a global metric, namely the *Mean Absolute Deviation (MAD)* and *Root Mean Squared (RMSE)*, which can be computed as [27]:

$$\begin{aligned} MAD &= 1/N \times \textstyle\sum_{i=1}^{N} |y_i - \widehat{y}_i| \\ RMSE &= \sqrt{\textstyle\sum_{i=1}^{N} (y_i - \widehat{y}_i)^2 / N} \end{aligned} \tag{1}$$

In both metrics, lower values result in better predictive models. However, the $RMSE$ is more sensitive to high errors. Another possibility to compare regression models is the Regression Error Characteristic (REC) curve [2], which plots the error tolerance (x-axis), given in terms of the absolute deviation, versus the percentage of points predicted

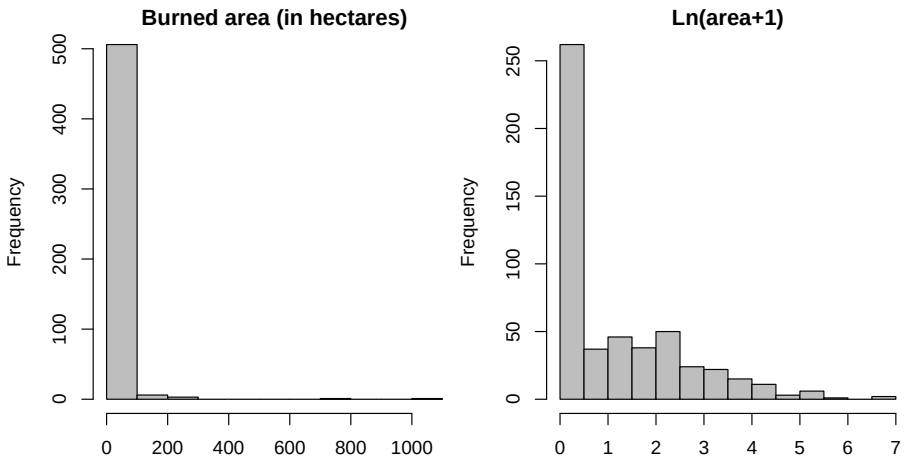

Fig. 3. The histogram for the burned area (left) and respective logarithm transform (right)

within the tolerance (y-axis). The ideal regressor should present a REC area close to 1.0.

Several DM algorithms, each one with its own purposes and capabilities, have been proposed for regression tasks. This work will consider five DM models. The Multiple Regression (MR) model is easy to interpret and this classical approach has been the widely used [11]. Yet, it can only learn linear mappings. To solve this drawback, one alternative is to use methods based on tree structures, such as Decision trees (DT) and Random Forests (RF), or nonlinear functions, such as Neural Networks (NN) and Support Vector Machines (SVM).

The DT is a branching structure that represents a set of rules, distinguishing values in a hierarchical form [4]. This representation can translated into a set of IF-THEN rules, which are easy to understand by humans. The RF [3] is an ensemble of T unpruned DT, using random feature selection from bootstrap training samples. The RF predictor is built by averaging the outputs of the T trees. In general, RF exhibits a substantial improvement over a single DT.

NN are connectionist models inspired by the behavior of the human brain. In particular, the multilayer perceptron is the most popular NN architecture. It consists of a feedforward network where processing neurons are grouped into layers and connected by weighted links [12]. This study will consider multilayer perceptrons with one hidden layer of H hidden nodes and logistic activation functions and one output node with a linear function [11]. Since the NN cost function is nonconvex (with multiple minima), NR runs will be applied to each neural configuration, being selected the NN with the lowest penalized error. Under this setting, the NN performance will depend on the value of H.

SVM present theoretical advantages over NN, such as the absence of local minima in the model optimization phase. In SVM regression, the input $x \in \Re^A$ is transformed into a high m-dimensional feature space, by using a nonlinear mapping. Then, the *SVM*

finds the best linear separating hyperplane in the feature space:

$$\widehat{y} = w_0 + \sum_{i=1}^{m} w_i \phi_i(x) \tag{2}$$

where $\phi_i(x)$ represents a nonlinear transformation, according to the kernel function $K(x, x') = \sum_{i=1}^{m} \phi_i(x)\phi_i(x')$. To estimate the best SVM, the ϵ-insensitive loss function (Figure 4) is often used [22]. The popular Radial Basis Function kernel, which presents less hyperparameters and numerical difficulties than other kernels (e.g. polynomial or sigmoid), will also be adopted [13]:

$$K(x, x') = exp(-\gamma||x - x'||^2),\ \gamma > 0 \tag{3}$$

The SVM performance is affected by three parameters: C – a trade-off between the model complexity and the amount up to which deviations larger than ϵ are tolerated; ϵ – the width of the ϵ-insensitive zone; and γ – the parameter of the kernel. Since the search space for the three parameters is high, the C and ϵ values will be set using the heuristics proposed in [5]: $C = 3$ (for standardized inputs) and $\epsilon = 3\widehat{\sigma}\sqrt{\frac{ln(N)}{N}}$, where and $\widehat{\sigma}$ is the standard deviation as predicted by a 3-nearest neighbor algorithm.

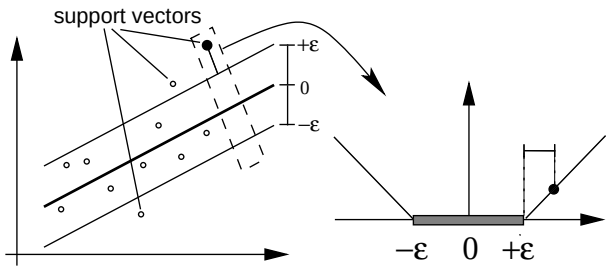

Fig. 4. Example of a linear SVM regression and the ϵ-insensitive loss function (adapted from [22])

Due to their performance in terms of predictive knowledge, RF, NN and SVM are gaining an attention within the DM field [27]. However, these methods require more computation and use representations that are more difficult to interpret when compared with the more simple MR and DT models. Nevertheless, it is still possible to provide explanatory knowledge for RF, NN and SVM in terms of input relevance [3][16].

4 Experimental Results

All experiments reported in this study were conducted using the **RMiner** [6], an open source library for the **R** statistical environment [21] that facilitates the use of DM techniques in classification and regression tasks. In particular, the **RMiner** uses the **randomForest** (RF algorithm by L. Breiman and A. Cutler), **nnet** (for the NN) and and **kernlab** (LIBSVM tool [13]) packages.

Before fitting the models, some preprocessing was required by the MR, NN and SVM models. The nominal variables (i.e. discrete with more than two non-ordered values), such as the **month** and **day**, were transformed into a *1-of-C* encoding, as advised in [13]. Also, for the NN and SVM methods, all attributes were standardized to a zero mean and one standard deviation [11]. Next, the regression models were fitted. The MR parameters were optimized using a least squares algorithm, while the DT node split was adjusted for the reduction of the sum of squares. Regarding the remaining methods, the default parameters were adopted for the RF (e.g. $T = 500$), the NN were adjusted using $NR = 3$ trainings and $E = 100$ epochs of the BFGS algorithm and the Sequential Minimal Optimization algorithm was used to fit the SVM. After fitting the DM models, the outputs were postprocessed using the inverse of the logarithm transform. In few cases, this transformation may lead to negative numbers and such negative outputs were set to zero.

To infer about the impact of the input variables, four distinct feature selection setups were tested for each DM algorithm: **STFWI** – using spatial, temporal and the four FWI components; **STM** – with the spatial, temporal and four weather variables; **FWI** – using only the four FWI components; and **M** – with the four weather conditions. To access the predictive performances, thirty runs of a 10-fold [17] (in a total of 300 simulations) were applied to each tested configuration. Regarding the NN and SVM hyperparameters, a internal 10-fold grid search (i.e. using only training data) was used to find the best $H \in \{2, 4, 6, 8, 10\}$ and $\gamma \in \{2^{-9}, 2^{-7}, 2^{-5}, 2^{-3}, 2^{-1}\}$. After selecting the H/γ value, the NN/SVM model was retrained with all training data. Table 2 shows the median values of the selected H and γ parameters.

Table 2. The best hyperparameters for NN and SVM (median values)

	Feature Selection Setup			
DM Model	STFWI	STM	FWI	M
NN	4	6	4	4
SVM	2^{-5}	2^{-3}	2^{-3}	2^{-3}

The results are shown in Table 3 in terms of the mean and respective t-student 95% confidence intervals [9]. For benchmarking purposes, the naive average predictor (first row) was also added to the table. Under the MAD criterion, all DM methods outperform the naive benchmark. Within a given feature selection, the SVM tends to produce the best predictions (except for the STM setup). Another interesting result is the non relevance of the spatial and temporal variables, since when removed the SVM performance improves. In effect, the best configuration is given by the **M** setup and SVM model and paired t-tests against all other models confirmed the statistical significance of this result. For the SVM, it is better to use weather conditions rather than FWI variables. This is interesting outcome, since the meteorological variables can be acquired directly from the weather sensors, with no need for accumulated calculations. However, from the $RMSE$ point of view, the best option is the naive average predictor. This ap-

parent contradiction is justified by the nature of each error criteria, i.e. the $RMSE$ is more sensitive to outliers than the MAD metric.

A more detailed analysis to the quality of the predictive errors is given by using REC curves (Figure 5). To simplify the visualization, only three models are plotted: M–SVM, the best MAD configuration; M–RF, the second best meteorological based method (in terms of the MAD value); and Naive, the best $RMSE$ model. From the REC analysis, the M–SVM is clearly the best solution, with the highest area. Although there is only a 0.22 difference in terms of the average MAD values, the M–SVM and M–RF curves are distinct, with the former model presenting the best predictions for an admissible absolute error up to 2.85. For example, 46% of the examples are accurately predicted if an error of $1ha$ is accepted and this value increases to 61% when the admissible error is $2ha$. Regarding the naive predictor, it is the worst method, surpassing the other alternatives only after an absolute error of 13.7.

Table 3. The predictive results in terms of the MAD errors ($RMSE$ values in parentheses; underline – best model; **bold** – best within the feature selection)

DM	**Feature Selection Setup**			
Model	STFWI	STM	FWI	M
Naive	18.61±0.01 (**63.7**±0.0)	18.61±0.01 (**63.7**±0.0)	18.61±0.01 (**63.7**±0.0)	18.61±0.01 (**63.7**±0.0)
MR	**13.07**±0.01 (64.5±0.0)	**13.04**±0.01 (64.4±0.0)	13.00±0.00 (64.5±0.0)	13.01±0.00 (64.5±0.0)
DT	13.46±0.04 (64.4±0.1)	13.43±0.06 (64.6±0.0)	13.24±0.03 (64.4±0.0)	13.18±0.05 (64.5±0.0)
RF	13.31±0.02 (64.3±0.0)	**13.04**±0.01 (64.5±0.0)	13.38±0.05 (64.0±0.1)	12.93±0.01 (64.4±0.0)
NN	13.09±0.04 (64.5±0.0)	13.92±0.60 (68.9±8.5)	13.08±0.05 (64.6±0.1)	13.71±0.69 (66.9±3.4)
SVM	**13.07**±0.04 (64.7±0.0)	13.13±0.02 (64.7±0.0)	**12.86**±0.00 (64.7±0.0)	**12.71**±0.01 (64.7±0.0)

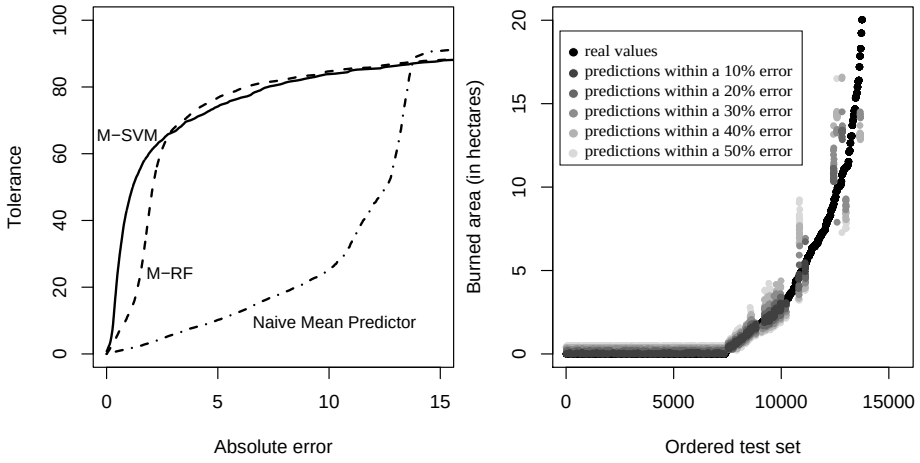

Fig. 5. The REC curves for the **M**-SVM, **M**-RF and Naive models (left); and the real values (black dots) and M-SVM predictions (gray dots) along the $y-$axis output range (right)

To complement the REC analysis, another plot is presented for the M–SVM configuration (Figure 5). The intention is to observe how the errors are distributed along the output range. The real values (black dots) of the test set were ordered (x-axis) according their burned area (y-axis). It should be noted that x-axis ranges from 1 to 517×30 runs = 15510. To clarify the analysis, the y-axis was set within the range $[0, 20ha]$. The M–SVM predictions are also shown in the figure, using a gray scale that is dependent on the accuracy. In general, the gray dots denote predictions within a relative error that ranges from 10% (darker grey) to 50% (lighter grey). The exception is when when the real values are below $1ha$. In this case, the gray scale corresponds to absolute differences (from $0.1ha$ to $0.5ha$). The plot shows that the M–SVM performance is better when predicting small fires (e.g. within the $[0, 3.2ha]$ range).

Regarding the input relevance procedure, the whole 517 records were used to fit the M–SVM model. Then, a sensitivity analysis [16] procedure was performed by measuring the variance (V_a) produced by the output when a given input attribute x_a varies through its entire range with L levels (here set to $L = 5$). Let $\overline{y}_{aL_i}$ be the average output when the attribute $x_a = L_i$ and all other inputs are set to their original values (from the dataset). Then $V_a = \sum_{i=1}^{L} (\overline{y}_{aL_i} - \overline{\overline{y}_{aL_i}})^2/(L-1)$. These variances can be relativized, by using the expression: $R_a = V_a / \sum_{j=1}^{A} V_j$ (Table 4). This procedure indicates that all weather conditions affect the model, with the outside temperature being the most important feature, followed by the accumulated precipitation (rain).

Table 4. The sensitivity analysis values for the weather inputs of the M–SVM model

	temp	**RH**	**wind**	**rain**
V_a	9.95	0.56	0.64	2.45
R_a	73.2%	4.1%	4.7%	18.0%

5 Conclusions

Forest fires cause a significant environmental damage while threatening human lives. In the last two decades, a substantial effort was made to build automatic detection tools that could assist Fire Management Systems (FMS). The three major trends are the use of satellite data, infrared/smoke scanners and local sensors (e.g. meteorological). In this work, we propose a Data Mining (DM) approach that uses meteorological data, as detected by local sensors in weather stations, and that is known to influence forest fires. The advantage is that such data can be collected in real-time and with very low costs, when compared with the satellite and scanner approaches. Recent real-world data, from the northeast region of Portugal, was used in the experiments. The database included spatial, temporal, components from the Canadian Fire Weather Index (FWI) and four weather conditions. This problem was modeled as a regression task, where the aim was the prediction of the burned area. Five different DM algorithms, including Support

Vector Machines (SVM), and four feature selections (using distinct combinations of spatial, temporal, FWI elements and meteorological variables) were tested.

The proposed solution, which is based in a SVM and requires only four direct weather inputs (i.e. temperature, rain, relative humidity and wind speed) is capable of predicting small fires, which constitute the majority of the fire occurrences. The drawback is the lower predictive accuracy for large fires. To our knowledge, this is the first time the burn area is predicted using only meteorological based data and further exploratory research is required. As argued in [18], predicting the size of forest fires is a challenging task. To improve it, we believe that additional information (not available in this study) is required, such as the type of vegetation and firefighting intervention (e.g. time elapsed and firefighting strategy). Nevertheless, the proposed model is still useful to improve firefighting resource management. For instance, when small fires are predicted then air tankers could be spared and small ground crews could be sent. Such management would be particularly advantageous in dramatic fire seasons, when simultaneous fires occur at distinct locations.

This study was based on an off-line learning, since the DM techniques were applied after the data was collected. However, this work opens room for the development of automatic tools for fire management support. Indeed, in the future we intend to test the proposed approach by using an on-line learning environment as part of a FMS. This will allow us to obtain after some time a valuable feedback from the firefighting managers, in terms of trust and acceptance of this alternative solution. Another interesting possibility would be the use of weather forecasts, in order to build proactive responses. Since the FWI system is widely used around the world, further research is need to confirm if direct weather conditions are preferable than accumulated values, as suggested by this study. Finally, since large fires are rare events, outlier detection techniques [28] will also be addressed.

6 Acknowledgments

We wish to thank Manuel Rainha for providing the spatial, temporal and FWI data. We also thank the Bragança Polytechnic Institute for the meteorological station database.

References

1. B. Arrue, A. Ollero, and J. Matinez de Dios. An Intelligent System for False Alarm Reduction in Infrared Forest-Fire Detection. *IEEE Intelligent Systems*, 15(3):64–73, 2000.
2. J. Bi and K. Bennett. Regression Error Characteristic curves. In *Proceedings of 20th International Conference on Machine Learning (ICML)*, pages 43–50, Washington DC, USA, 2003.
3. L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, 2001.
4. L. Breiman, J. Friedman, R. Ohlsen, and C. Stone. *Classification and Regression Trees*. Wadsworth, Monterey, CA, 1984.
5. V. Cherkassy and Y. Ma. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. *Neural Networks*, 17(1):113–126, 2004.
6. P. Cortez. RMiner: Data Mining with Neural Networks and Support Vector Machines using R. In R. Rajesh (Ed.), *Introduction to Advanced Scientific Softwares and Toolboxes*, In Press.

7. European-Commission. Forest Fires in Europe. Technical report, Report N-4/6, 2003/2005.
8. U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. *Advances in Knowledge Discovery and Data Mining*. MIT Press, 1996.
9. A. Flexer. Statistical evaluation of neural networks experiments: Minimum requirements and current practice. In *Proceedings of the 13th European Meeting on Cybernetics and Systems Research*, volume 2, pages 1005–1008, Vienna, Austria, 1996.
10. D. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining*. MIT Press, Cambridge, MA, 2001.
11. T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, NY, USA, 2001.
12. S. Haykin. *Neural Networks - A Compreensive Foundation*. Prentice-Hall, New Jersey, 2nd edition, 1999.
13. C. Hsu, C. Chang, and C. Lin. A Practical Guide to Support Vector Classification. http://www.csie.ntu.edu.tw/˜cjlin/papers/guide/ guide.pdf, July, Dep. of Comp. Science and Information Eng., National Taiwan University, 2003.
14. W. Hsu, M. Lee, and J. Zhang. Image Mining: Trends and Developments. *Journal of Intelligent Information Systems*, 19(1):7–23, 2002.
15. J. Terradas J. Pinol and F. Lloret. Climate warming, wildfire hazard, and wildfire occurrence in coastal eastern Spain. *Climatic Change*, 38:345–357, 1998.
16. R. Kewley, M. Embrechts, and C. Breneman. Data Strip Mining for the Virtual Design of Pharmaceuticals with Neural Networks. *IEEE Transactions on Neural Networks*, 11(3):668–679, May 2000.
17. R. Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Montreal, Quebec, Canada, August 1995.
18. K. Malarz, S. Kaczanowska, and K. Kulakowski. Are forest fires predictable? *International Journal of Modern Physics*, 13(8):1017–1031, 2002.
19. D. Mazzoni, L. Tong, D. Diner, Q. Li, and J. Logan. Using MISR and MODIS Data For Detection and Analysis of Smoke Plume Injection Heights Over North America During Summer 2004. *AGU Fall Meeting Abstracts*, pages B853+, December 2005.
20. S. Menard. *Applied Logistic Regression Analysis*. SAGE, 2nd edition, 2001.
21. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. URL: http://www.R-project.org, ISBN 3-900051-00-3.
22. A. Smola and B. Scholkopf. A tutorial on support vector regression. Technical Report NC2-TR-1998-030, University of London, UK, 1998.
23. D. Stojanova, P. Panov, A. Kobler, S. Dzeroski, and K. Taskova. Learning to Predict Forest Fires with Different Data Mining Techniques. In D. Mladenic and M. Grobelnik, editors, *9th International multiconference Information Society (IS 2006)*, Ljubljana, Slovenia, 2006.
24. S. Taylor and M. Alexander. Science, technology, and human factors in fire danger rating: the Canadian experience. *International Journal of Wildland Fire*, 15:121–135, 2006.
25. C. Vega-Garcia, B. Lee, P. Woodard, and S. Titus. Applying neural network technology to human-caused wildfire occurence prediction. *AI Applications*, 10(3):9–18, 1996.
26. D. Viegas, G. Biovio, A. Ferreira, A. Nosenzo, and B. Sol. Comparative Study of various methods of fire danger evalutation in southern Europe. *International Journal of Wildland Fire*, 9:235–246, 1999.
27. I.H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco, CA, 2005.
28. J. Zhao, C. Lu, and Y. Kou. Detecting Region Outliers in Meteorological Data. In *Proceedings of the 11th ACM International Symposium on Advances in Geographic Information Systems*, pages 49–55, New Orleans, Louisiana, USA, 2003.