2021 International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 25-27, 2021

Forest Fire Prediction Using Machine Learning Techniques

Preeti T, Dr.Suvarna Kanakaraddi, Aishwarya Beelagi, Sumalata Malagi, Aishwarya Sudi
KLE Technological University
Hubli, Karnataka, India
preeti@kletech.ac.in; suvarnagk@kletech:ac:in; aishwaryabeelagi30@gmail:com;
uma.malagi@gmail.com; 1997aishwarya13s@gmail.com

2021 International Conference on Intelligent Technologies (CONIT) | 978-1-7281-8583-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/CONIT51480.2021.9498448

***Abstract*—Forest Fire Prediction is a key component of forest fire control. This is a major environmental problem that creates ecological destruction in the form of a threatened landscape of natural resources that disrupts the stability of the ecosystem, increases the risk for other natural hazards, and decreases resources such as water that causes global warming and water pollution. Fire Detection is a key element for controlling such incidents. Prediction of forest fire id expected to reduce the impact of forest fire in the future. Many fire detection algorithms are available with different approach towards the detection of fire. In the existing work processes the fire affected region is predicted based on the satellite images. To predict the occurrences of a forest fire the proposed system processes using the meteorological parameters such as temperature, rain, wind and humidity were used. Random forest regression and Hyperparameter tuning using RandomizedSearchCV algorithm we used a various sub-samples of dataset on which it fits several decision trees and uses averaging to improve the predictive accuracy and control over-fitting. Based on the analysis of the models with all the selected meteorological parameters can represent the forest fire events.This paper discusses about a comparative study of different models for predicting forest fire such as Decision Tree, Random Forest, Support Vector Machine, Artificial Neural networks (ANN) algorithms. The study of calculation of RandomizedSearchCV coefficient using Hyperparameter tuning gives best results of Mean absolute error(MAE) 0.03, Mean squared error(MSE) 0.004, Root mean squared error(RMSR) 0.07**

***Index Terms*—Decision Tree, Random Forest, SVM, ANN, Forest Fire Prediction.**

I. Introduction

Forest fires are a matter of concern as a result the cause in depth injury to surroundings, property and human life. Hence, it's crucial to notice the fire at Associate in nursing earlier stage. One of the most reasons of the incidence of forest fires is heating as a result, the increasing in average temperature of the world. The opposite reasons are because of lightning, throughout thunderstorms, and human negligence. Annually a mean of one.2 million acres of the forest within the U.S. get destroyed because of the wildfires. In The Asian nation forest fires have exaggerated by hundred and twenty-fifth between the years 2016 and 2018.Nowadays, there are numerous technologies for fireplace models to predict the unfold of 5 fires, like physical models and mathematical models. These models rely upon knowledge assortment throughout forest fires simulations, and sciences laboratory experiments to specify and predict fireplace growth in several areas. Recently, simulation tools are wont to predict forest fires, however simulation tools round-faced some issues like the accuracy of the computer file and the simulation tool execution time. the machine learning could be a sub-branch of a computing (AI) to be told computers side.Machine learning may be divided into 2 classes: supervised, unattended and reinforcement. Supervised machine learning algorithms are as regression, Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Decision trees. In the unattended learning, the information attributes don 't seem to be tagged. This leads that the formula should outline the labels. The structure of the information set and also the relationship between the options is going to be learned by the formula.[1]

• The main motivation for forest fire prediction is to provide proper resource allocation and to help in best possible way to firefighters of Fire Management team.
• The main factors of fire are Meteorological conditions. the climatic information is gotten from nearby sensors which are fused in the closest meteorological stations.
• Land with a possible high fire risk has many indicators that can be used to measure the forecast by closely evaluating the indications.
• Every year, fire destroys millions of hectares of land. These fires have burned vast areas and generate more carbon monoxide than total vehicle traffic.
• Monitoring potential danger areas and early warning of fire can greatly reduce response time, as well as the potential for damage and firefighting costs.

II. Literature survey

A. Forest Fire Prediction using Artificial Intelligence

George E. Sakr et al. (2010), An approach to the study of forest fire prediction methods based on artificial intelligence has been suggested. Forest fire risk forecast algorithm is built on help vector machines. Lebanon data were used for the application of the algorithm and has proven the ability to correctly estimate the risk of fire.

B. Forest Fire Prediction using Image Mining Technique

Divya T L et al. (2015) in their paper have presented the by analysing a series of pixel values, an image mining technique

can be used to predict the spread of a forest fire. The proposed model uses the satellite images for forest fire prediction.

C. Forest Fire Prediction using Artificial Neural Network

Nizar HAMADEH LARIS EA et al. (2015), in this paper authors have considered an area called Lebanon to predict the occurrence of forest fire. Temperature, relative humidity, and wind speed are among the parameters. These parameters force Artificial Neural Networks to evolve in order to anticipate forest fires.

D. Forest Fire Prediction using Linear Regression

Mukhammad Wildan Alauddin et al. (2018) For forest fire prediction, multiple linear regression has been proposed. Temperature, humidity, wind, and rain are among the factors involved. Different techniques such as gauss-jordan, gauss-seidel, and least-squares are used to calculate various linear regression coefficients. Comparative analysis of the methods is done and the results are discussed[2]. This Section discusses about the Literature survey.

–

III. Proposed System

This Section discusses about the Block Diagram of the Proposed System.

A. Block Diagram of Proposed method

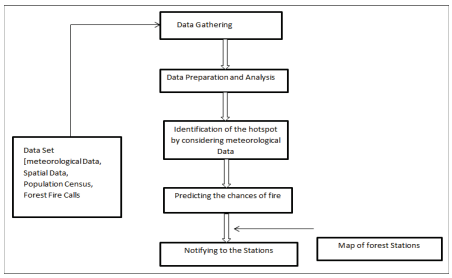

Fig. 1. Block diagram of Forest Fire Prediction.

The Proposed System Block Diagram discusses where we had gathered the data set from Kaggle which consists of meteorological data then we have performed the exploratory analysis that pre-processing where we will try to remove the noisy data and converting that categorical data to numerical data so it will be easy to understand that dataset. After the preprocessing technique and then hotspot location is identified based on the meteorological data available in the data set then apply the models to predict the chances of occurrence of fire and send the notification to the nearest station.

IV. Methodology

A. SYSTEM DESIGN

System architecture or system architecture is a computational paradigm that describes the structure, behavior and views of the system. A system architecture may be composed of system modules and subsystems that will collaborate to execute the overall system.

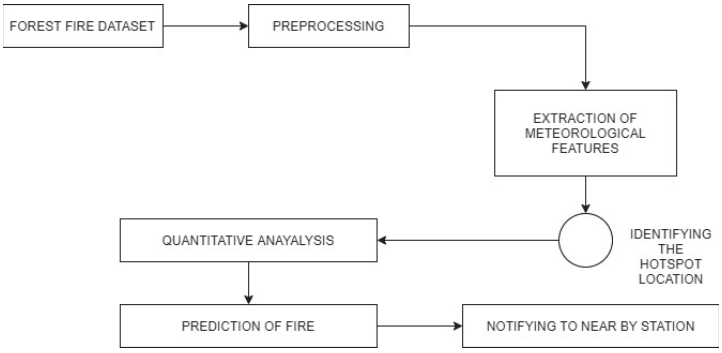

Fig. 2. High Level Diagram of Forest Fire Prediction.

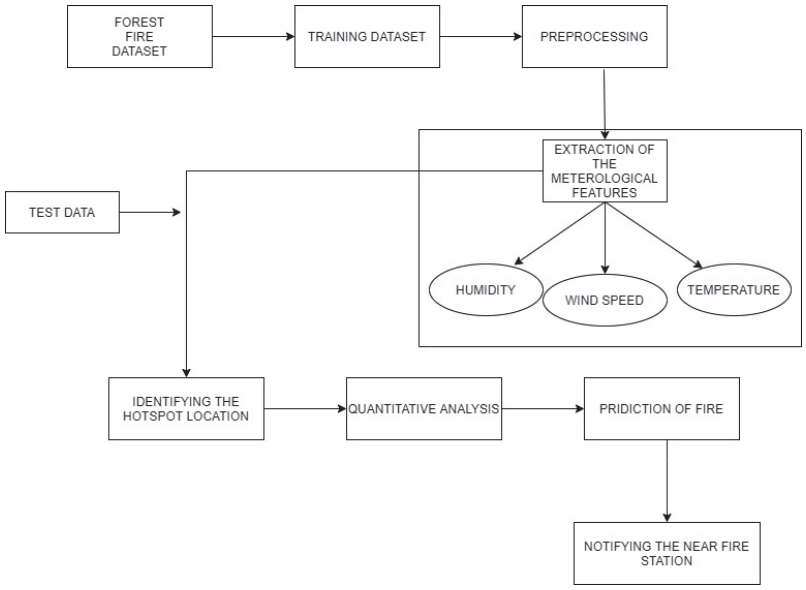

Fig. 3. Low Level Diagram of Forest Fire Prediction.

– A few key points regarding detailed design are given below:-
  - The data utilized in this paper collected from kaggle. The dataset contains 517 observations and 13 variables from the natural park of Montesano in the European republic. For each incident weekday, month coordinates, and hence the burnt house square measure registered, in addition to many earth science data such as rain, temperature, humidity, and wind. Progress reads input and develops a regression model assisted by abstraction, time and weather variables.
  - After data collection data pre-processing takes place in which dataset to be formed in standard format.
  - After data preparation suitable model to be selected based on the dataset.

- ∗ In this project, we are using regression techniques used for prediction are Random forest (RF), Decision Tree (DT) and Support Vector Regression (SVR) and Naive Bayes.
- ∗ After model implementation model evaluate.
- ∗ Predicting the data and accuracy of each model[3].

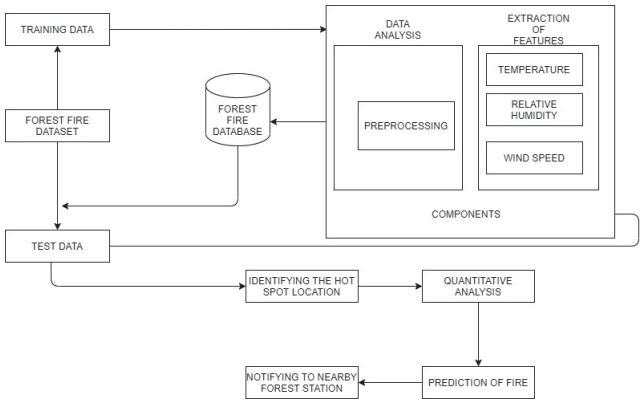

Fig. 4. Architecture Level Diagram of Forest Fire Prediction.

– A few key points regarding detailed design are given below:-
  - ∗ An architectural diagram is a diagram of the system used to explain the overall outline of the software system and the interactions, limits and boundaries between the elements.
  - ∗ After splitting of data into training data and test data, for training data we do preprocessing and extraction of features and the data will be stored in database.
  - ∗ from the database we take data for testing data and test it and quantitative analysis will be carried out and the result will be displayed prediction of fire.

This Section discusses about the Development.

1) Exploratory Analysis: One of the best methods used in data science these days is Exploratory Data Analysis. There is a difference between Data Analysis and Exploratory Data Analysis which people fail to understand in the initial days of their career. Exploratory Data Analysis is a tribute to inferential statistics which uses random data to protect fairly rigid with rules and formulas. We will explore a Data set and perform the exploratory data analysis. To begin this exploratory analysis, we first import libraries and then define data plotting functions using matplotlib. And we'll load the Forest Fire dataset and carry the Data Exploration process were are trying to display the following things:

• Head of the dataset: The head of the dataset function will help us to display only the top 5 records.

• column of the dataset: The column of dataset attribute tells us several observations and variables we have made in the data set. It is used to check the dimension of data. The forest fire data set has 518 observations and 13 variables in the data set.

• Null values: It displays the null values in the dataset

• Information of the dataset: It gives us complete information about the dataset to understand and to analyze it.

Now we're able to read the information, and we can plot a graph to visualize data.

• Distribution graph(histogram/bar graph) was plotted for column data that display the unique values between 1 and 50.

• Correlation matrix was plotted which shows columns where there are more than 1 unique values and also that we can evaluate the dependence between two variables and also calculate how the two variables move together.

This Section discusses about the Exploratory Analysis[4].

2) PreProcessing Analysis: When we go through and talk about data, Normally we think about any big databases with a massive number of rows and columns. Although this is likely to be the case, it is not necessarily the case that the data may be in too many different forms: Structured tables, images, audio files, videos, etc. Data: Data Preprocessing is the stage in which the data is converted, or encoded, to get it to such a state that now the computer can quickly parse it.

Preprocessing is one of the features that offers a number of functions and transform classes to translate raw data vectors into representation and also to convert raw data into a clean data set i.e. (when data is gathered from different sources, in turn, it is collected from the raw format which is not feasible for the analysis).

• correlation matrix was plotted for the preprocessed data were tried to display the correlation between the meteorological features like relative humidity, wind speed, temperature, rain by this we understand the dependence between two variable and also measures how two variables move together.

• By this result, we can see that relative humidity and temperature are less correlated.

• If the sum is greater than 0, there is a positive correlation. If the value is less than zero it is a negative correlation, so that is why we can assume that relative humidity and temperature are less correlated since it has a negative value of correlation.

• We tried converting the categorical data to numerical data which is called one-hot encoding.

One hot encoding: Machine learning models require both input and output variables to be in numeric form. This ensures that if the data includes categorical data, you must encode it into numbers before you can adjust and evaluate the model.

• We tried converting categorical data of month and day to numerical data as 1 and 0.

• After encoding we tried to split our dataset in the ratio of 75:25 as training dataset and testing dataset. Then we

tried in Fitting Random Forest Regression to the dataset and predict the result.[5]

3) Decision Tree: The decision tree constructs a regression model inside the tree structure category. It splits down the data set into smaller associated degrees with smaller subsets The decision tree is rendered top-down from the root node and entails partitioning the knowledge into subsets of instances of identical values (homogenous).

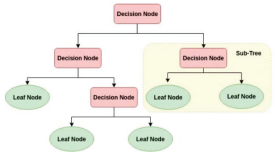

Fig. 5. Decision Tree.

4) Support Vector Machine: The support Vector Machine also can be used as a regression technique, maintaining all the most options that characterize the algorithmic rule (maximal margin). The Support Vector Regression (SVR) uses constant principles because the SVM for classification, with solely many minor variations.

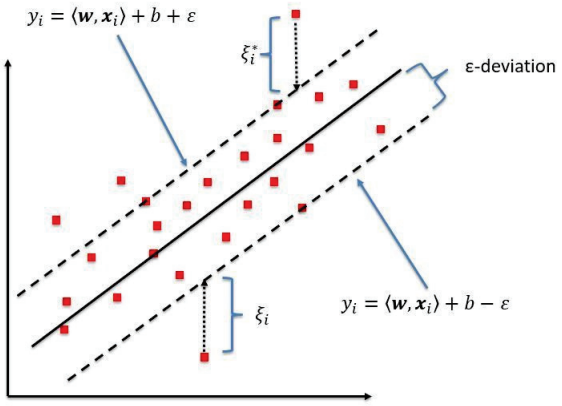

Fig. 6. Support Vector Repressor.

5) Random Forest Regression: Random forest may be a supervised learning rule that uses an ensemble learning methodology for a classification and regression. The trees in random forest the area unit run in a parallel. A random forest may be a meta-estimator (i.e. incorporates the outcomes of several predictions) that aggregates several call trees, with some helpful modifications: The number of choices that will be divided on each node is limited to a certain proportion of the total. This Section discusses about the Preprocessing Analysis.

V. Result Analysis

The collected data are used to train the system and tend to do the prediction. We'll be analysing a temperature, humidity,rain,wind speed and associated attributes and can

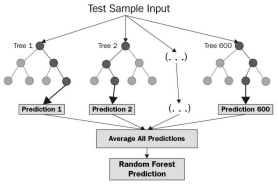

Fig. 7. Random Forest tructure.

predict the fire in forest. The regression techniques used for prediction are random forest (RF), Decision tree(DT) and support vector regression (SVR).The planned models were enforced in the python platform. Random forest, the decision tree and Support Vector Regressor model and is used for implementation, and a comparative study has been created, results obtained from the model coaching and testing. Here calculated accuracy and MAE for random forest, svr, Decision tree models[6]

The mean square error tells you how close a regression curve is to a collection of points. It will this by taking the distances from the points to the regression curve (these distances are the "errors") and squaring them. The squaring is important to get rid of any negative signs. It additionally offers a lot of weights to larger variations. It's known as the mean square error as finding the common of a collection of errors.

Variance is a measure of how much the observed values vary from the average of the expected values, i.e. their deviation from the mean of the predicted values. The target is to have a low valuation.

Before applying the normalization the graphs of Variance of models.

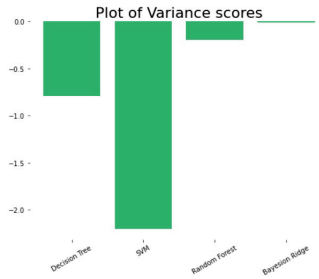

Fig. 8. Result of before applying normalization of Variance graph.

Normalization converts the data to a range between 0 and 1. Standardization transforms the data in such a way that the resulting distribution has a mean of 0 and a standard deviation of 1. After applying the normalisation, the MSE Model Variance graphs are applied.

A. Random Forest Regression

Experiment with Mean Square Error and the result are:

- MEAN ABSOLUTE ERROR : 0.04
- MEAN SQUARED ERROR : 0.01

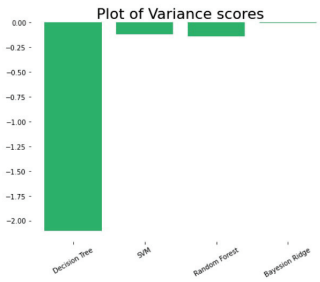

Fig. 9. Result of After applying normalization of Variance graph.

• ROOT MEAN SQUARED ERROR : 0.11

Experiment Hyper parameter tuning using RandomizedSearchCV and the result are:
• MEAN ABSOLUTE ERROR : 0.03
• MEAN SQUARED ERROR : 0.004
• ROOT MEAN SQUARED ERROR : 0.07

Rxperiment with Artificial neural network model and the result are:
• MEAN ABSOLUTE ERROR : 0.71
• MEAN SQUARED ERROR : 3.96
• ROOT MEAN SQUARED ERROR : 1.99

This Section discusses about the result Mean Square Error.

B. Principle component analysis

An important machine learning method for dimensionality reduction is called Principal Component Analysis. It is a technique that uses basic matrix operations from linear algebra and statistics to measure the projection of the original data to the same or lesser number.

Using principal component analysis we tried to plot a graph of the affected area of fire and non-affected area of the fire.

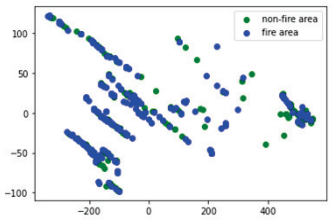

Fig. 10. Principle component analysis.

Exploratory Analysis Results:
1. Distribution graphs of month column verses count We've tried to show the column graph values that are between 1 and 50 unique values.

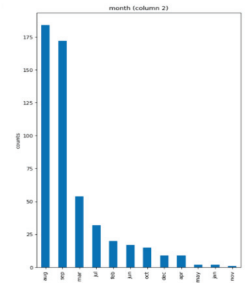

Fig. 11. Distribution graph

2. Distribution graphs of day column verses count We Tried to display the graph values of columns that have between 1 and 50 unique values.

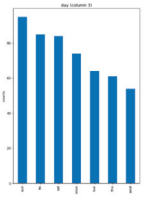

Fig. 12. Distribution graph

3. Distribution graphs of wind column verses count We Tried to display the graph values of columns that have between 1 and 50 unique values.

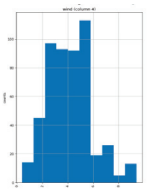

Fig. 13. Distribution graph

4. Distribution graphs of rain column verses count We've tried to show the column graph values that are between 1 and 50 unique values.

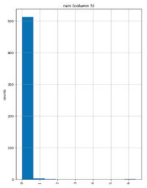

Fig. 14. Distribution graph

5. Graph of Scatter and density plots We tried to show columns with more than 1 unique value.

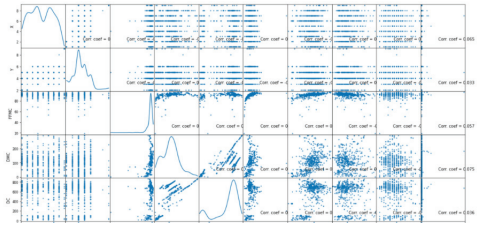

Fig. 15. Scatter plot graph

6. Graph of Correlation We tried to show columns with more than 1 unique value.

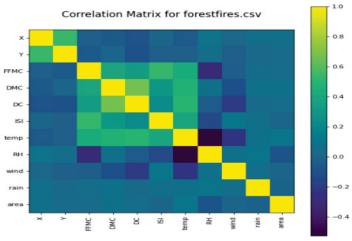

Fig. 16. Corrrealtion graph

7. Graph of correlation of features We tried to display the dependency between two variables.

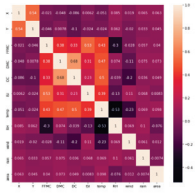

Fig. 17. Correlation graph

8. Graph of Comparison of different Models.

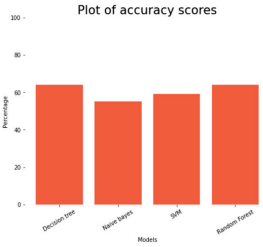

Fig. 18. Distribution graph

VI. Conclusion and Future Scope

Experiments are concluded in order to have a different number of training instances set and evaluation instances set for forest fire prediction. The factors causing the frequency of fire are investigated in this project. Meteorological factors (Temperature, Relative Humidity and Wind Speed) are taken into account. Extreme temperatures, moderate humidity, high wind speeds, significantly raise the chance of burning. It is also found that the number of fires in forests is higher than in other surface areas. As the risk of forest fire increases significantly in the forest, data mining techniques are to be used for fire prediction purposes. This project can be further expanded to do better so that the models are better equipped and the effects are better. We may also have a UI built for the application to provide some real-time performance. The workflow of the UI model could be, the user may enter the local and the zip code. Using the zip code, we'll get latitude and longitude using any APIs and consume the coordinates as parameters, get the weather conditions like peak temperature, min temperature, humidity, wind speed, etc. for a given day.

.

References

[1] A. Alonso-Betanzos, O. Fontenla-Romero, B. Guijarro-Berdi nas, E. Hern´andez-Pereira, M. Inmaculada Paz Andrade, E. Jim´enez, J. LuisLegido Soto, and T. Carballas, "An intelligent system for forest fire risk prediction and fire fighting management in Galicia,"Expert systems with applications, vol. 25, no. 4, pp. 545–554, 2003.

[2] N. Aronszajn, Introduction to the theory of Hilbert spaces. Stillwater, Oklahoma:Reasearch [sic] Foundation, 1950.

[3] T. Cheng and J. Wang, "Applications of spatio-temporal data mining and knowledge for forest fire," in ISPRS Technical Commission VII Mid Term Symposium, Enschede, 2006, pp. 148–153.

[4] "Integrated Spatio-temporal Data Mining for Forest Fire Prediction," Transactions in GIS, vol. 12, no. 5, pp. 591–611, 2008

[5] K. Clarke, J. Brass, and P. Riggan, "A cellular automaton model of wildfire propagation and extinction," Photogrammetric Engineering and Remote Sensing, vol. 60, no. 11, pp. 1355–1367, 1994.

[6] Z. Li, Y. Kaufman, C. Ithoku, R. Fraser, A. Trishchenko, L. Giglio, J. Jin, and X. Yu, "A review of AVHRR-based active fire detection algorithms: Principles, limitations, and recommendations," Global and Regional Vegetation Fire Monitoring from Space.

[7] J. Han, K. Ryu, K. Chi, and Y. Yeon, "Statistics Based Predictive Geo-spatial Data Mining: Forest Fire Hazardous Area Mapping Application," Lecture notes in computer science, pp. 370–381, 2003.

[8] R. Jaiswal, M. Saumitra, D. Kumaran, and S. Rajesh, "Forest fire risk zone mapping from satellite imagery and gis," International Journal of Applied Earth Observation and Geo-information, vol. 4, pp. 1–10,2002.

[9] G. Mitri and I. Gitas, "A semi-automated object-oriented model for burned area mapping in the Mediterranean region using Landsat-TM imagery," International Journal of Wildland Fire, vol. 13, no. 3, pp. 367–376, 2004.

[10] A. Muzy, T. Marcelli, A. Aiello, P. Santoni, J. Santucci, and J. Balbi, "An object oriented environment applied to a semi-physical model of fire spread across a fuel bed," in Actesde la conf´erence ESS 2001 conference, 2001, pp. 641–643.