Coupled Dynamic Data-Driven Framework for Forest Fire Spread Prediction

Carlos Brun, Ana Cortés$^{(\boxtimes)}$, and Tomàs Margalef

Computer Architecture and Operating Systems Department,
Universitat Autònoma de Barcelona,
Edifici Q, 08193 Cerdanyola V., Spain
carlos.brun@caos.uab.cat, {ana.cortes,tomas.margalef}@uab.cat
http://caos.uab.es

Abstract. Predicting the potential danger of a forest fire is an essential task of wildfire analysts. For that reason, many scientists have focused their efforts on developing propagation models that predict forest fire evolution to mitigate the consequences of such hazards. These propagation models require a precise knowledge of the whole environment where the fire is taking place. In the context of natural hazards simulation, it is well known that, part of the final forecast error comes from the uncertainty in the input data. In this work, we use a Dynamic Data-driven methodology to overcome such problem. The core of the methodology is a calibration stage previous to the forecast where complementary models, data injection and intelligent systems are working in a symbiotic way to reduce the forecast errors at real time. This approach has been tested using a forest fire that took place in Arkadia (Greece) in 2011.

Keywords: Forest fire · Simulation · Data uncertainty · Dynamic data-driven

1 Introduction

Forest fires involve serious consequences from the environmental, economic and social point of view. For that reason, the scientific community has developed simulation tools with the aim of providing useful information about forest fire spread evolution to the people in charge of managing extinction resources. Most of the existing forest fire spread simulators [1–7] implement the spread kernel equations based on the Rothermel's model [8]. However, it is well known that the forecast fire evolutions provided by existing fire spread simulators do not exactly reproduce the real behaviour of the fire. The reason for such a difference ranges from the input parameters uncertainty to the imprecision of the model itself. In previous works [9–11], it has been stated that a pre-processing of the simulator input parameters based on a steering loop driven by real data acquisition and fire behaviour observation, could lead to enhanced forecast fire evolutions. This prediction scheme is the so call Two-Stage prediction system.

The Two-Stage strategy performs a forest fire prediction by previously executing a Calibration stage, which involves the most sensitive parameters. In this stage, the actual evolution of the forest fire is observed and a Genetic Algorithm (GA) is carried out to determine the set of parameters that best reproduces the recent evolution of the fire. This set of values is then used as input parameters in the Prediction stage. Since forest fires are a dynamic phenomena, which is quite affected by changing data such as meteorological information, the mentioned pre-processing data phase has been designed as a feedback loop where gathered data guides the simulation and, the simulation results at a time, could eventually drive the data collection. This way of work feeds the so called Dynamic Data Driven Application System [12,13]. A key point in this process is the evaluation of the simulation's quality, because it has a direct impact in both the Calibration stage and in the Prediction stage. The Calibration stage consists of a Genetic Algorithm (GA), which tries to minimize a predetermined fitness function. In the context of forest fire propagation, this fitness function is the difference between the real burnt area and the burnt area obtained by simulation. Eventually, a perfect match should be reflected into the fitness function as a global minimum/maximum value according to its definition. Up to now, the Two-Stage prediction system relies on the symmetric difference between sets as a fitness function (also called error function). This error function has been proven to provide good results when applied to predict forest fire to a maximum of regional size. However, when moving to forest fire classified as "dangerous" at European level, this fitness function was detected to be not enough accurate and, for that reason, an alternative error has been defined. However, not all the information needed by the prediction system could be considered in the calibration process. In particular, those data considered as static information such as elevation maps and fuel data, is typically obtained from public repositories. There is not an unique source of this kind of data and, consequently, the fire behaviour prediction delivered by a given forest fire spread model could vary according to the selected static data sources. Furthermore, gathering dynamic data such as real fire perimeter evolution and meteorological data could be a bottleneck to the system if there is no a clear way to proceed. For that reason, at European level, EFFIS (European Forest Fire Information System) raises as the EU common platform to provide all input data required in a forest fire simulation system (FFSS). Therefore, relying on the EFFIS data, one can design FFSS at European level based on standard basis. However, the resolution of these information is not always at the desired precision, so, it becomes mandatory to include complementary models to the forest fire spread model in order to obtain high resolution data, which takes into account the environment where the hazard is occurring. For that reason, the basic Two-Stage forest fire simulation system was enhanced by coupling two different models, a wind field model (WindNinja [14]), which takes into account the wind speed and wind direction variation due to the underlying topography and, a meteorological model (WRF [15]) to evaluate the time evolution of the meteorological variables. The resulting coupled prediction framework has been tested using a study case retrieved from

the EFFIS database. In Sect. 2, the data uncertainty problem related to forest fire spread forecast is introduced. The coupled dynamic data-driven prediction framework (DDD-FFSS) is described in Sect. 3, as well as, the proposed error equation for events at paneuropean level. The described DDD-FFSS is then applied to a forest fire that took place in Arkadia (Greece) in 2011 in Sect. 4. Finally, Sect. 5 summarizes the main conclusions of this work.

2   Data Uncertainty in Forest Fire Simulation

As it has been mentioned, fire behavior models require accurate input data to provide fire spread forecast as reliable as possible. Although the model sensitivity to the input data clearly depends on the nature of each required parameter, the precision and quality of all of them are not dismissible. In general, the data needed to perform the predictions can be divided into two main groups: static and dynamic data. The static input data is the one that keeps constant during the whole prediction interval, and the dynamic data changes during the fire spread simulation. According to this feature (static/dynamic), the way of gathering and processing the information to obtain the corresponding input files is quite different. In the case of static data, the pre-processing and organization of the required layers in the proper format could be done previously to the hazard occurrence. There are certain constraints related to the terrain dimension that should be considered in an accurate way, but the process of homogenize the precision, projection and datum could be done off-line, and have this data characterized and ready to be used when a crisis occurs. If the static data of a region is available before a fire occurrence, the efforts must be focused on those parameters that vary dynamically during the simulation or depend on the fire scenario studied. In this case, it is necessary to collect them in real-time, thanks to the different data sources and services that can provide this information. Obviously, this case is the most critical since we depend on third-parties frequency of data arrival, and data format. Therefore, the conversions and the injection must be done in an on-line mode, while the simulation is being carried out. As we have previously mentioned, at European level, the reliable third-party is EFFIS (European Forest Fire Information System) and, therefore, this is the data source we have used to perform this study. Subsequently, we shall describe how static and dynamic data related to forest fire evolution is considered by EFFIS.

2.1   Topography

This data is obtained by processing Digital Elevation Maps (DEMs). A DEM defines the height of the terrain in every cell of the map. This is a discretization of a continuous surface, taking into account measures in certain points of the terrain. Those maps are obtained from the ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) imaging instrument onboard NASAs Terra satellite that takes high-resolution images of the earth. These images are processed and raster files are extracted with the information needed to perform the fire spread simulations (elevation, aspect and slope). The ASTER map resolution is 30 m.

2.2   Fuel Map

The vegetation map (or fuel map) is a raster file that describes the predominant vegetation in every cell. The fuel model used for fire simulation purposes, is the standard fuel model defined by [14]. This information has been obtained from the fuel type map of Europe developed at the JRC. The classification scheme adopted for the fuel map encompasses 42 fuel types representing the variety of fuel complexes found in European landscapes. A cross-walk to the original set of 13 fire behaviour fuel models tabulated by Anderson fire spread model is done at the JRC [14].

2.3   Meteorological Data

The meteorological data used is the ECMWF (European Centre for Medium-Range Weather Forecasts) operational high-resolution single global deterministic model (ec16), with a horizontal resolution of about 16 km [16]. The model is initiated on both the 00 and 12 UTC analysis reaching to a 24-hour forecast horizon with archived time-step of 3 hour. It is worth mentioning, that this is the configuration used in this work for experimental purposes, but it is not the unique source of meteorological data processed at the JRC.

2.4   Fire Perimeter

A key point in any forest fire simulation system is the capability to feed the system with either a real initial fire perimeters or a precise ignition point. The EFFIS Burnt Area Map module is in charge of this data. To obtain the fire perimeter information, the JRC relies on the MODIS (Moderate Resolution Imaging Spectroradiometer) sensors systems, which are on board both the NASAs Terra and Aqua satellites. Each satellite requires to complete 3 orbits (approximately 3 h) to cover the whole Europe area, so it could be possible to obtain fire perimeters twice a day, one from each satellite. The image resolution provided by the MODIS system is of 250 m.

3   Dynamic Data-Driven Forest Fire Simulation System

As it has been mentioned, the Two-Stage prediction scheme uses the Calibration stage to search for those input parameters setting that, fitted to the underlying simulator, better reproduces the recent observed forest fire spread. The obtained parameter configuration will then be used to fit the simulator in the Prediction stage to forecast the near future evolution of the fire. As a search technique in the Calibration stage, a Genetic Algorithm (GA) is applied, where a random initial population of individuals (input parameter setting) is generated. Each individual is simulated using FARSITE [6] (the underlying forest fire spread simulator) and the resulting forest fire spread is compared to the real observed fire evolution to compute the GA's fitness function (called error in this case).

Then, according to the quality of the prediction, the individuals are ranked and the genetic operators are applied to generate the new population. The process is repeated a certain number of iterations and the best individual at the end of the process is selected to run the prediction at the Prediction stage. Since evaluating the prediction quality is a key point in this scheme, in the following subsection, we shall describe the selected fitness function (error) and, subsequently, we will introduce how the DDD-FFSS has been improved by coupling complementary models.

3.1   Quality Prediction Evaluation

In order to compare the obtained predictions with the real fire behavior, it is necessary to define metrics in order to determine the quality of the simulations what enables the capacity of ranking them. There exist several metrics to compare real and predicted values [17] and each one weights the events involved differently. The event notation used to define those error functions is depicted in Fig. 1.
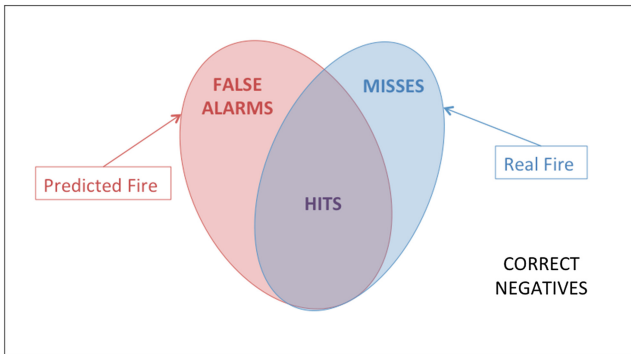

Fig. 1. Events involved in metrics related to forecast verification.

The cells around the map that have not been burnt by neither the real fire nor the predicted map are considered *Correct Negatives* (CN). Those cells that have been burnt in both maps are called *Hits*. The cells that are only present in the real fire and are not burnt by the predicted fire are called *Misses*. Finally, in the opposite case, the cells that the simulator predicts as burnt area but the real fire does not actually reach, are called *False Alarms* (FA). Besides these factors, some equations take into account the real map ($RealCell$), the simulated or predicted map ($SimCell$), the ignition map ($ICell$) or the total number of cells of the terrain ($TotalCell$). This notation may vary but the meaning remains the same.

Up until now, we have been using the *symmetric difference* between maps as error function. The values given by this error function are positive, but not in a closed interval, with the best value being 0 without an upper limit. We use the

concept of $UCell$ (all cells belonging to both real fire and simulated fire) and $InterCell$ (cells belonging to both real fire and simulated fire) as factors in the equation. To better understand and to simplify the equations, the initial fire is considered a point and, therefore, it can be removed from the original equations what simplifies the translation of Eq. 1 to the notation previously described.

$$Error = \frac{(UCell - ICell) - (InterCell - ICell)}{RealCell - ICell} =$$
$$= \frac{(Hits + Misses + FA) - (Hits)}{RealCell} =$$
$$= \frac{Misses + FA}{RealCell} \tag{1}$$

In fact, both real and simulated maps can also be transformed as a combination of *Hits*, *Misses*, and FA events as is shown in Eq. 2

$$RealCell = Hits + Misses$$
$$SimCell = Hits + FA \tag{2}$$

Equation 1 equally penalizes the *misses* and the *false alarms*. Another metric that was used to rank individuals was the *Critical Source Index* (CSI), which gives us the rate of *hits* achieved from 0 to 1, with 1 being the perfect match between maps. It also weight *misses* and *false alarms* in the same way.

$$Error = \frac{InterCell}{UCell} = \frac{Hits}{Hits + Misses + FA} \tag{3}$$

The fact of equally penalizing both factors is not suitable in our field because it is much more important to minimize the *misses* than to reduce the *false alarms*. The consequences of *misses* can cause severe damage, both to the environment and in human lives, while the *false positives* may represent an extra effort in fire-fighting resources.

The main problem of these metrics when applied to the Two-Stage methodology is concentrated into the Calibration stage. In this part of the methodology, we evaluate several scenarios, then we rank them using the error function, and finally, we select the best parameter set to perform the prediction. We have detected that when dealing with large forest fires at paneuroepan level, the individuals with less spread, tend to provide the best error values. Analyzing the shape of the other individuals, we observe that potential good predictions were discarded from the calibration process due to the high penalty generated by the *false alarms*. In order to solve this undesired effect, we changed Eq. 1 in order to minimize the effect of *false alarms*. The new error function is shown in Eq. 4.

$$Error = \frac{\frac{(UCell - ICell) - (InterCell - ICell)}{RealCell - ICell} + \frac{(UCell - ICell) - (InterCell - ICell)}{SimCell - ICell}}{2} =$$
$$= \frac{\frac{Misses + FA}{Real} + \frac{Misses + FA}{Sim}}{2} \tag{4}$$

The latest equation has shown better behavior in the Calibration stage than the other metrics. In those cases where the difference between the predicted burnt area and the real burnt are is the same in terms of magnitude but opposite according to the event related (*misses* or *false alarms*), the individuals that provide overestimated predictions have a better error than those individuals that underestimated the fire evolution. Therefore, the new error function was incorporated to the Dynamic Data-Driven Forest Fire Simulation System (DDD-FFSS) describe below.

3.2   Coupling Models to the DDD-FFSS

As it was previously introduced, we rely on DDD-FFSS to perform forest fire spread predictions. This system encapsulates the Two-Stage prediction scheme with the possibility of coupling/uncoupling complementary models to better consider available environmental conditions (wind, humidity,...). The basic scheme of the Two-Stage approach is depicted in Fig. 2 (2ST-BASIC). It is well known that one of the parameters that most affect fores fire propagation is the wind, for that reason, the efforts has initially been concentrated on this parameter. The complementary models included in DDD-FFSS are: a wind field model to consider the effect of the terrain on wind speed and wind direction and, a meteorological model whose output has been post-processed to deliver a meteorological wind speed and wind direction at the pinpointed centroid of the fire. Since the objective of this work is to test the DDD-FFSS approach at European level, it is unrealistic to consider that the data injected in the Calibration stage will only come from meteorological stations. In fact, the data fitted into the steering loop is provided by a metereological model but it also can be improved with real data obtained from meteorological stations and sensors. This information can be directly fitted into the Calibration stage of the forest fire spread prediction system. However, during the prediction stage such values are not available beforehand. So, it is necessary to introduce a meteorological model that can provide the expected values for the meteorological wind speed and wind direction used at the prediction stage (see Fig. 3, 2ST-MM). The last enhancement included in the system was to consider the influence of the topography on the wind components as it is shown in Fig. 4 (2ST-MM-WF). In this case, the information related to wind speed and wind direction used at the Calibration stage is introduced to the wind field model before running all forest fire spread simulations. In the Prediction stage, the meteorological data is provided by a certain meteorological model and then introduced to the wind field model to provide the corresponding wind field.

These prediction schemes have been tested at European level using fire cases from the EFFIS repository. In the following section, we show the results obtained in terms of quality improvement for a particular study case.

4   Experimental Study

The Mediterranean area is one of the European regions most affected for forest fires during high risk seasons. As we have previously mentioned, we rely on
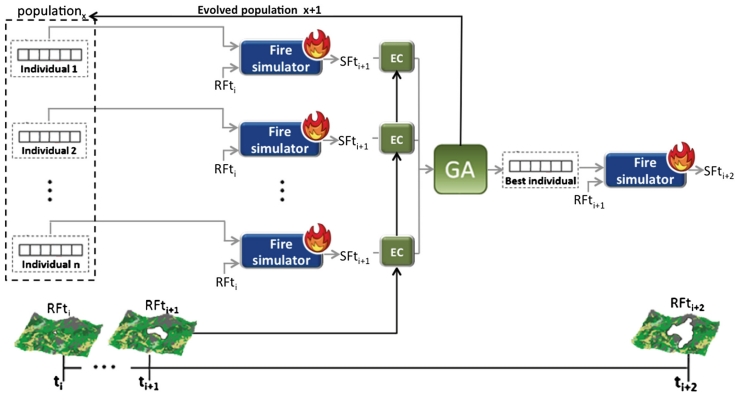
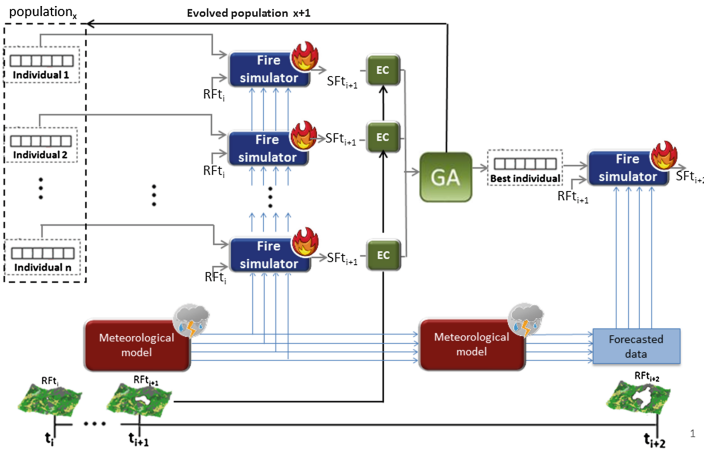**Fig. 2.** 2ST-BASIC prediction scheme

Fig. 3. 2ST-MM prediction scheme

EFFIS and JRC (Joint Research Centre) data sources to fit the dynamic data-driven prediction system described in the previous section. Therefore, we have selected as study case one event stored in the database of EFFIS. In particular, we have retrieved the information of a past fire that took place in Greece during the summer season of 2011 in the region of Arkadia, one of the seven prefectures of the Peloponnese peninsula in Greece. The forest fire began on the 26th of August and the total burnt area was 1,761 ha. The experimental results shown in this section were obtained using a computing platform, which consists of two PowerEdge C6145 nodes, each one including 4 AMD OpteronTM6376 with 16 cores each (128 cores).

In Fig. 5(a), the images provided by the MODIS system are shown for three different time instants:
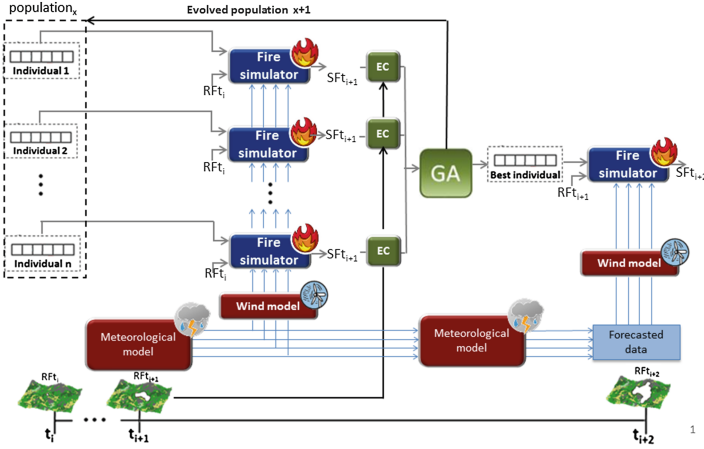
Fig. 4. 2ST-MM-WF prediction scheme

- t_0: August 26th at 09:43am obtained from the Terra satellite.
- t_1: August 26th at 11:27am obtained from the Aqua satellite.
- t_2: August 27th at 08:49am obtained from the Terra satellite.

The corresponding burnt areas (shapes) once the images have been processed, are shown in Fig. 5(b). These shapes are the information available at EFFIS. From these shapes, we obtain the real fire perimeters as are shown in Fig. 6.

In order to simplify the initial tests, the forecast meteorological data used for the simulations are the wind components (wind speed and wind direction), dew point and temperature of the pinpointing centroid of the observed fire. These meteorological data is provided with an frequency of 3 h, for that reason, the injection time step within the forest fire spread simulator has been set to 3 h. The prediction time horizons have been set according to the exact time the MODIS images have been obtained to be as fair as possible to the reality. In order to compare the prediction results in terms of quality when applying the DDD-FFSS coupling different complementary models, the system has been set to the three following configurations: 2ST-BASIC, 2ST-MM and 2ST-MM-WF. Keeping in mind that any configuration of the DDD-FFSS always implies the execution of the Calibration stage and then, the execution of the Prediction stage, it is necessary to describe how both stages use the available data perimeters. For calibration purposes, we used as initial perimeter *perimeter*1 from Fig. 6 and, as a reference perimeter, *perimeter*2 from Fig. 6. So, the simulations involved in the Calibration stage have been set to a time horizon around 2 h. In the case of the Prediction stage, the perimeter to be predicted is *perimeter*3 from Fig. 6 and, the initial perimeter is *perimeter*2. Therefore, the time horizon for the simulation at this stage has been set to 22 h. All these data inputs have been harmonized to fit a simulation grid map with a basic cell of 100 m × 100 m square.
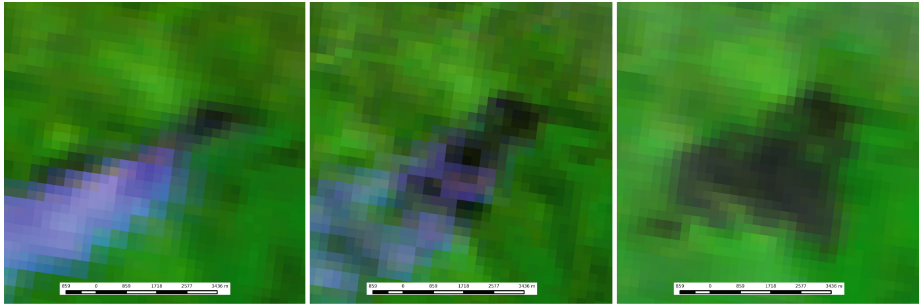
(a) MODIS images
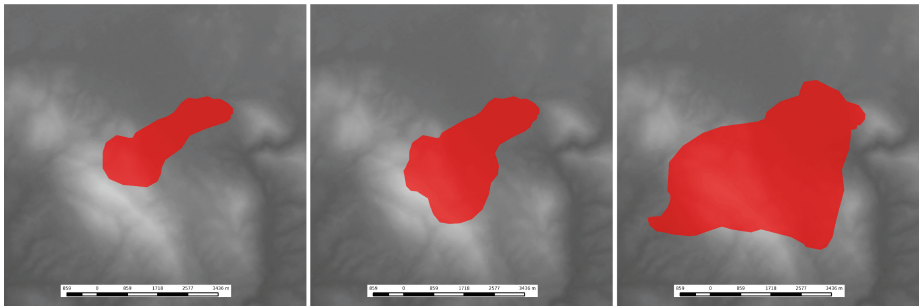
(b) Vectorial shapes as a result of MODIS images

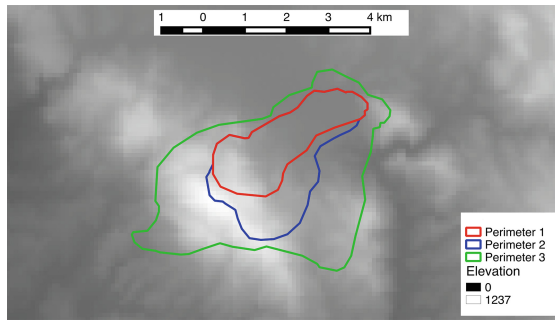**Fig. 5.** MODIS images and their corresponding extracting shapes

Fig. 6. Fire perimeters corresponding to Arkadia fire

The obtained results are shown in Fig. 7. As it can be observed, the 2ST-BASIC configuration is the one that provides the best error at the Calibration stage. Despite seeming this result contradictory to the claim of coupling models to obtain enhanced predictions, it is necessary to highlight that the prediction error in this case is the worst. To understand this, it is noteworthy that the interval between the first and second perimeters is around two hours, and there is only a single meteorological data sample in this interval. This lack of knowledge has a direct impact on the quality of the calibration. Figure 8 shows an example
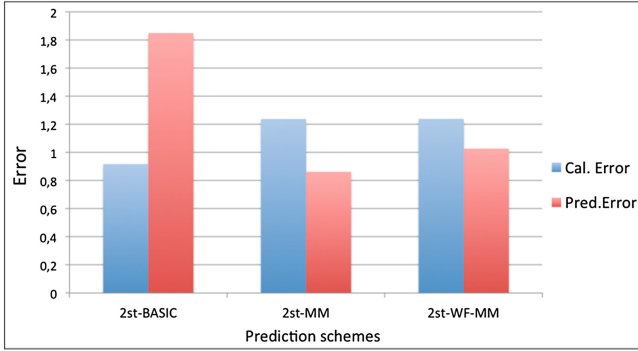
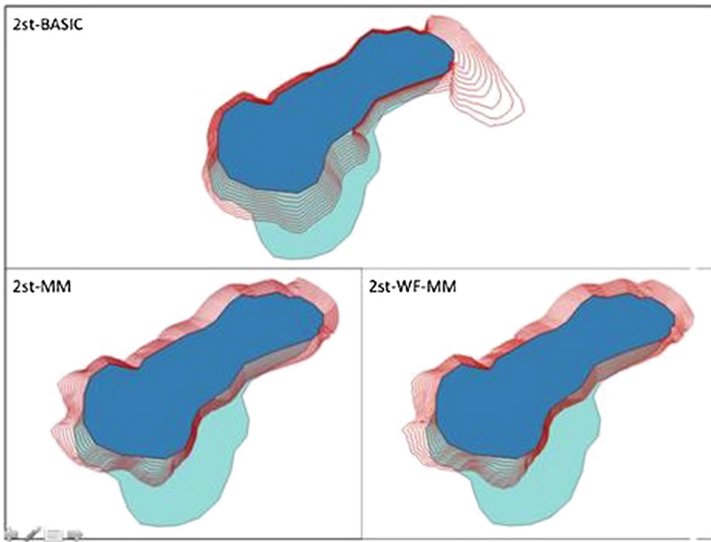**Fig. 7.** Calibration and prediction errors for every prediction scheme

Fig. 8. Calibration stage perimeters for each prediction scheme

of the best calibrated perimeter for each scheme. All three methods under-predict the fire behavior, and there are some possible causes for this fact. The measured wind could be less than the reality, and the schemes could not tune the other parameters to minimize this effect. This fact arise another potential problem related to the fuel models used. It is possible that the fuel model conversion from the European cover uses to the standard fire models resulted in low-propagation types. The main reason to support this idea was the behavior of the 2ST-BASIC scheme. Although not sensitive to sudden changes, this method usually finds calibrated winds that make the fire spreading quite close to the real fire, although the final shape can differ due to its uniform conditions.
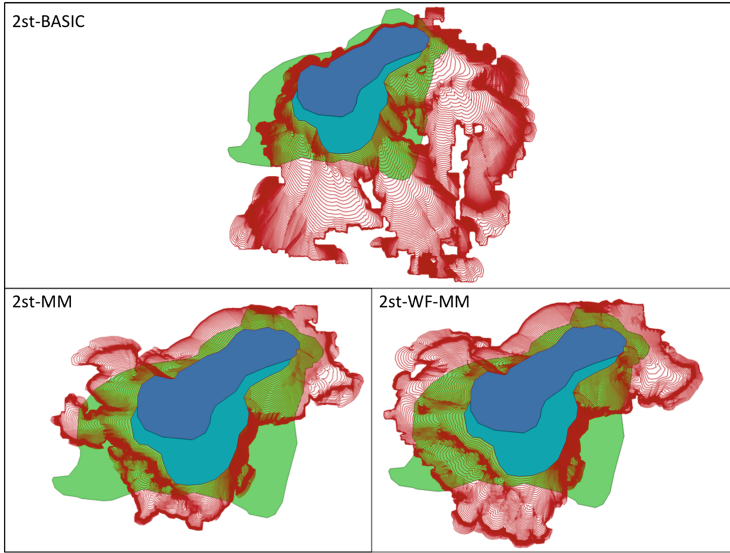
Fig. 9. Prediction stage perimeters for each prediction scheme

As it has been mentioned, this situation changes when we analyze the prediction stage that lasts around 22 h. In this case, the best prediction errors are the ones given by the 2ST-MM and 2ST-WF-MM schemes. The dynamic injection of meteorological data seems to be positive to the system and to provide good prediction shapes, as we can see in Fig. 9. Although in numerical terms the 2ST-MM is the best scheme, the 2ST-WF-MM gives back better perimeters and better covers the real burnt area. The 2ST-BASIC scheme uses the tuned weather values obtained in the calibration stage, which present a high wind speed value. This causes to excessively over-predict the real fire behavior.

5 Conclusions

Natural hazards, such as forest fire, are phenomena that require complex models to predict their evolution. In the particular case of forest fire, propagation models require input parameters that in some cases present a high degree of uncertainty. So, a Dynamic-Data Driven system was introduced to calibrate the input parameters based on the observation of the actual evolution of the fire. Moreover, some parameters are dynamic and present a temporal evolution that require the coupling of complementary models such as meteorological models to the Dynamic Data-Driven Forest Fire Framework. Finally it must be considered that all the input parameters must be introduced to the model at the same resolution what in some cases require the coupling of complementary models such as wind field models. The coupled Dynamic Data-Driven Forest Fire Framework has been used on the context of the European Forest Fire Information System

to analyse the potential improvement in forest fire spread prediction and it has demonstrate a quite significant improvement. In this context, a new error equation has been proposed to evaluate the prediction quality for large forest fires taking into account the factor of overestimated/underestimated predictions compared to the real forest fire spread.

References

1. Rothermel, R.C.: Behave and you can predict fire behavior. Fire Manag. Notes **44**(4), 11–15 (1983)
2. Finney, M.A., McHugh, C.W., Grenfell, I.C., Riley, K.L., Short, K.C.: A simulation of probabilistic wildfire risk components for the continental united states. Stoch. Env. Res. Risk Assess. **25**(7), 973–1000 (2011)
3. Andrews, P.L., Bevins, C.D., Seli, R.C., et al.: BehavePlus fire modeling system: version 2.0: user's guide. US Department of Agriculture (Forest Service), Rocky Mountain Research Station Fort Collins (2003)
4. Finney, M.A., et al.: An overview of flammap fire modeling capabilities. In: Proceedings of the Fuels Management-How to Measure Success, pp. 213–220 (2006)
5. Bevins, C.D.: Firelib user manual and technical reference (1996)
6. Finney, M.A., et al.: Farsite, fire area simulator-model development and evaluation (2004)
7. Lopes, A., Cruz, M., Viegas, D.: Firestationan integrated software system for the numerical simulation of fire spread on complex topography. Environ. Model. Softw. **17**(3), 269–285 (2002)
8. Rothermel, R.C.: A mathematical model for predicting fire spread in wildland fuels. Director (INT-115), 40 p. (1972)
9. Cencerrado, A., Rodríguez, R., Cortés, A., Margalef, T.: Urgency versus accuracy: dynamic driven application system natural hazard management. Int. J. Numer. Anal. Model. **9**, 432–448 (2012)
10. Abdalhaq, B., Cortés, A., Margalef, T., Luque, E.: Enhancing wildland fire prediction on cluster systems applying evolutionary optimization techniques. Future Gener. Comput. Syst. **21**(1), 61–67 (2005)
11. Bianchini, G., Denham, M., Cortés, A., Margalef, T., Luque, E.: Wildland fire growth prediction method based on multiple overlapping solution. J. Comput. Sci. **1**(4), 229–237 (2010)
12. Darema, F.: Dynamic data driven applications systems: a new paradigm for application simulations and measurements. In: Bubak, M., van Albada, G.D., Sloot, P.M.A., Dongarra, J. (eds.) ICCS 2004. LNCS, vol. 3038, pp. 662–669. Springer, Heidelberg (2004)
13. Darema, F.: Grid computing and beyond: the context of dynamic data driven applications systems. Proc. IEEE **93**(3), 692–697 (2005)
14. Potter, B., Butler, B.: Using wind models to more effectively manage wildfire. Fire Manage. **69**(2), 40–46 (2009)

15. Group, W.W.: Weather Research and Forecasting (WRF) model. Director (INT-115) (2008)
16. Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T.: The ECMWF ensemble prediction system: methodology and validation. Q. J. R. Meteorol. Soc. **122**(529), 73–119 (1996)
17. Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., et al.: Characterising performance of environmental models. Environ. Model. Softw. **40**, 1–20 (2013)