Evolutionary Statistical System for applying in Forest Fire Spread Prediction^{*}

Germán Bianchini, Miguel Mendez-Garabetti and Paola Caymes-Scutari

Laboratorio de Investigación en Cómputo Paralelo/Distribuido (LICPaD) Departamento de Ingeniería en Sistemas de Información, Facultad Regional Mendoza - Universidad Tecnológica Nacional. (M5502AJE) Mendoza, Argentina

Abstract. Several propagation models have been developed to predict forest fire behaviour. They can be grouped into empirical, semi-empirical, and physical models. These models can be used to develop simulators and tools for preventing and fighting forest fires. Nevertheless, in many cases the models present a series of limitations related to the need for a large number of input parameters. Furthermore, such parameters often have some degree of uncertainty due to the impossibility of measuring all of them in real time. Therefore, they have to be estimated from indirect measurements, which negatively impacts on the output of the model. In this paper we present a method which combines Statistical Analysis with Parallel Evolutionary Algorithms (taking advantage of the computational power provided by High Performance Computing) to improve the quality of model's output.

1 Introduction

Different propagation models have been developed to predict fire behaviour. They can be classified into empirical, semi-empirical, and physical models [8]. The probable fire behaviour is predicted in empirical models from average conditions and accumulated knowledge obtained from laboratory and outdoor experimental fire or from historical fires. Semi-empirical (semi-physical or laboratory models) are those models based on a global energy balance and on the assumption that the energy transferred to the unburned fuel is proportional to the energy released by the combustion of the fuel; one of the most important among these models is the pioneering work of Rothermel (1972 and 1983) [21, 22]. Finally, physical (theoretical or analytical) models are based on physical principles and have the potential to accurately predict the parameters of interest over a broader range of input variables than empirically based models do. These models can be used to develop simulators and tools for preventing and fighting forest fires. Some old and current examples are Behave-Plus [1], FARSITE [9], FIREMAP [2], FireStation [15], WRF-Fire [16], XFire [14], etc.

According to Fons [10] the relevant factors that affect the rate of spread and shape of a forest fire front are the fuel type (type of vegetation), humidity, wind

^{*} This work has been supported by Conicet under project PIP 11220090100709, by UTN under project UTN1194 and by ANPCyT under project PICT PRH-00242.

speed and direction, forest topography (slope and natural barriers), and fuel continuity (vegetation thickness). Therefore, models require a set of input parameters, including vegetation type, moisture contents, wind conditions, and so on, and they provide the evolution of the fire line in the successive simulation steps. However, the result obtained after the direct application of a simulator (known as Classical Prediction and explained in Section 2) usually differs from reality because of the difficulty of providing accurate input values to the model. Given this uncertainty, we propose an alternative method, that tries to determine the possible fire behaviour based on Statistical Analysis [19] and Parallel Evolutionary Algorithms (PEAs) [18] as optimization method. This method corresponds to an improvement of a previous methodology based on Statistical Analysis and High Performance Computing, which has been modified by the combination with Evolutionary Algorithms to improve the prediction level and reduce the execution time.

Clearly, the simulation of the spread of forest fires is a challenge from the computational point of view, given the complexity of the models involved, the need for efficient numerical methods and resource management for results. In this context, the method presented in this paper is an important tool for the prevention and prediction of forest fires, as it provides more complete information about the potential fire behaviour. This is a general method which could be applied on different propagation models (e.g. floods, snow avalanches, landslides, etc.), but here we only present its application to forest fire prediction.

In the remaining sections of this paper we describe the direct use of a simulator in section 2 (known as Classical Prediction); section 3 shows the predecessor of the current method (Statistical System for Forest Fire Management - S^2F^2M [4,5]) and section 4 describes the new methodology, implemented in a system called Evolutionary Statistical System (ESS) [3]. In section 5 we compare both methods using a set of real cases of forest fires and also we comment on the obtained results related to the execution time and the speed-up obtained when we work on a cluster computer. Finally, we present the main conclusions.

2 Classical Prediction

Classical Prediction approach is depicted in Fig. 1. In this scheme, FS corresponds to the underlying fire simulator, which will be seen as a black box. RFL0 is the real fire line at time t_0 (initial fire front), whereas RFL1 corresponds to the real fire line at t_1 . If the prediction process works, after executing FS (which should be fed with the corresponding input parameters and RFL0) the predicted fire line at time t_1 (PFL) should coincide with the real fire line (RFL1).

As we mentioned previously, models require static parameters (information about topography), parameters that can change very slowly (type of vegetation), parameters that can change frequently (moisture content), and parameters that are completely dynamic (like wind conditions). The simulator will not work properly without this set of parameters. The precision of these parameters is a very important point in prediction of the behaviour, and in many cases it is



Fig. 1. Diagram of Classical Prediction of forest fire propagation (FS: Fire Simulator; PFL: Predicted Fire Line; RFLX: Real Fire Line on time X)

impossible to carry out some types of measurements, particularly in a real fire situation.

Generally, the obtained prediction using this approach does not match the reality. One reason for the discrepancy between real and simulated propagation stems from the difficulty of feeding the model with accurate input values. Uncertainties in the input variables can have a substantial impact on the result errors and should be considered.

In this context, the prediction of the fire line behaviour cannot be considered to be reliable for two reasons: on the one hand, the difficulties in making an accurate estimate of the parameters and, on the other hand, the resulting prediction is based on a single simulation, which does not constitute a reasonable basis for making a decision given the uncertainty of the parameters.

3 Statistical Method for Uncertainty Reduction

The statistical method for uncertainty reduction [4–6] has been the result of the combination of various research projects. This method has as its bases the concepts of statistical analysis and distributed computing. Basically, the method finds a pattern of behaviour of the model without performing a specific analysis of each scenario (where a particular setting of the input parameter values defines an individual scenario). All the possible scenarios are discretely generated considering a certain domain by a factorial experiment [19] and the model is evaluated with each set of values. The results are combined to determine the trend in the behaviour of the model, adjusting to the current observation of it. The pattern found is then taken to predict the next step.

This method requires a large number of operations, and therefore is very time demanding. For this reason, we applied a parallel computing scheme for its implementation. Because of this, we used multiple computational resources working in parallel to reduce the time. Keeping in mind the nature of the problem, we applied a Master-Worker paradigm [12, 17], because the problem we face can be divided into multiple partitions and the same calculations can be applied over each data subset. Therefore, we face a problem that can be solved using domain decomposition: a main processor can calculate each combination of parameters and send them to a set of Workers. These Workers carry out the simulation in parallel, taking into account several combinations of parameters, and return the partial results to the Master, which aggregates all these individual results at each iteration. Also, the Master process is responsible for the statistical stage and it is in charge of the remaining prediction technique.

A scheme of a whole prediction system is presented in Fig. 2. As we can see, the process of prediction needs a calibration stage at the beginning (time period that goes from t_0 to t_1) to firstly obtain a K_{ign} value (Key Ignition value) to start up the prediction chain. For every *i* from 1 to *n*, both the prediction operation for time t_i and the calibration stage to obtain the K_{ign} to be used in time t_{i+1} will overlap at time t_i . This situation is the one depicted in Fig. 2. As can be observed, the output generated by the **SS** box (Statistical Stage) is used for a double purpose. On the one hand, the probability maps are used as an input of the **SK** box (Search K_{ign}) to search for the current K_{ign} , which will be used at the next prediction time. In this stage, a Fitness Function (**FF**) is used to evaluate the probability map. On the other hand, the output of **SS** box enters the Fire Prediction box (**FP**), which will be in charge of generating the prediction map taking into account the K_{ign} evaluated at previous time. This process will be repeated during the execution as the system is fed with new information about the fire situation.



Fig. 2. Detailed diagram of $S^2 F^2 M$ (FS: Fire Simulator; CS: Calibration Stage; SS: Statistical Stage; SK: Search Kign Stage; FF: Fitness Function; FP: Fire Prediction; PFL: Predicted Fire Line; RFLX: Real Fire Line at time X)

Although the statistical method can be used to solve various Grand Challenge Problems, as a case of study, the method has been applied on a behavioural model of forest fire propagation. As a result, we developed a system called Statistical System for Forest Fire Management (S^2F^2M) , which is the product of the combination of applying the proposed method with the simulator fireSim (fireSim is the implementation of a fire behaviour simulator based on the Rothermel model [21] and implemented with the library fireLib [7]). For a detailed description of the method, we suggest the reader to consult [4, 5].

4 Evolutionary Statistical System

The improvement and modification of the statistical method discussed in the previous section has resulted in a new method that combines the strength of three components: uncertainty reduction, evolutionary algorithms and parallelism, that is why the new method has been called Evolutionary Statistical System [3]. The improvement of the method is related to the introduction of features of PEAs in the calibration step of the statistical method. As we seen in the previous section, the statistical phase of the methodology includes all the results of a series of cases that arise as some combination of the possible resulting values (within valid ranges) of the parameters that exhibit uncertainty. Clearly, there is a certain percentage of cases that do not contribute significant values to the global result, whether they are now redundant, or because they are too far from reality (and thus, could be considered as negative cases that ultimately degrade the result provided by the method). To avoid this problem is that we have decided to apply the Evolutionary Algorithms (EAs), whose basis is explained in more details in the next section.

4.1 Evolutionary Algorithms

Evolutionary algorithms (EAs) mimic the concept of natural biological evolution: they operate on a population of potential solutions applying the principle of survival of the fittest [11]. In each iteration EAs create a new set of approaches through a process of selecting individuals according to the level of fitness for the problem domain (through the fitness function that quantifies this feature) and perform a recombination of them using operators that mimic natural genetics. This process leads to the evolution in the population of individuals that have best adapted to the environment just as happens in natural adaptation.

The EAs model natural processes such as selection, crossover, mutation, migration, locality and the notion of neighbourhood, working on populations of individuals rather than on unique solutions. Thus, the search can be performed in parallel, thus providing a number of potential solutions instead of one. This scheme is known as Parallel Evolutionary Algorithms (PEAs). According to the amount of populations involved in the algorithm, the treatment and the operators PEAs can be classified in three broad groups: Unique Population and Parallel Evaluation, Unique Population and Overlapped Neighbourhoods, and Multiple Populations and Migration. In this work, we consider the first group.

In each generation the fitness of each individual in the population is evaluated in parallel. Multiple individuals are stochastically selected from the current population (depending on fitness), and modified (by recombination or by random mutation) to form a new population. The fitness is defined in terms of the genetic representation and measures of quality of the solution represented. The execution of the PEA may finalize by various criteria. One method is to finish after a predetermined number of iterations. Another way is to check whether the measure of population quality has improved or not after a certain number of generations. Another is to finish when all individuals are identical, which can only happen when not using the mutation.

Evolutionary algorithms are a powerful tool for solving different kinds of problems [20]. However, sometimes this type of methodology iterates for a long time and does not converge or converges to a local optimum. This is one of the reasons why it is interesting combine the use of evolutionary methods with parallel computing. However, given the use of evolutionary algorithms in optimization problems, where they have found very good results, we propose the application of this methodology in combination with statistical methods, as discussed in the following section.

4.2 Methodology of the Evolutionary Statistical System

The Evolutionary Statistical System (ESS), classified as Data-Driven methods with Multiple Overlapping Solution, is an improvement of the S^2F^2M method previously commented. It combines the original uncertainty reduction method implemented in S^2F^2M with the advantages that offer the Parallel Evolutionary Algorithms (PEAs), dealing with a population of scenarios relevant to the study. ESS, like its predecessor, is based on statistics, mainly on the concept of factorial experiment [19], where the combination of several factors (input parameters) defines a scenario. In this case, each scenario is represented by an individual in a population of possible solutions.



Fig. 3. Diagram of ESS (FS: Fire Simulator; PEA: Parallel Evolutionary Algorithm; OS: Optimization stage; SS: Statistical System; SK: Search K_{ign} ; FF: Fitness Function; CS: Calibration stage; FP: Fire Prediction; PFL: Predicted Fire Line, RFLX: Real Fire Line on time X)

A scheme of ESS is presented in Fig. 3. As can be observed, the system is divided in two general stages: an Optimization Stage (OS) that implements the

parallel evolutionary algorithm (**PEA** box), and the Calibration Stage (**CS**) that is in charge of the statistical method. **OS** iterates until the population reaches a certain level of quality. For each individual **FS** and the fitness are calculated in parallel. Then, every individual will be included in the Statistical System (**SS** box). Similarly to S^2F^2M , the output of **SS** (a probability map) has a double purpose. On the one hand, the probability maps are used as the input of the **SK** box (Search K_{ign}) to search for the current K_{ign} (a key number used to make a prediction), which will be used at the next prediction time. In this stage, a Fitness Function (**FF**) is used to evaluate the probability map. On the other hand, the output of **SS** box enters the Fire Prediction box (**FP**). **FP** will be in charge of generating the prediction map taking into account the K_{ign} evaluated at previous time. All this process will be repeated during the execution as the system is fed with new information about the fire situation.

The architecture of the ESS is based on the Master-Worker paradigm [12, 17]: In each iteration the Master distributes an individual per Worker; the simulation of the model and the evaluation of fitness function are applied over each individual (tasks carried out by the Workers), returning the results to the Master. This process is repeated until every individual in the population is treated. Finally the Master evolves the population, aggregates the partial results and makes the prediction for each time step.

5 Experimental Results

This section compares the results obtained after applying the original statistical method (S^2F^2M) and the Evolutionary Statistical System (ESS) described in this paper. To that end we have used four cases of controlled burns. They were made in the field (Fig. 4), particularly in a hill of Serra de Lousã (Gestosa, Portugal). The burns were part of the SPREAD project [23]. These experiments were very useful to collect experimental data, to support the development of new concepts and models, and to validate existing methods or models in various fields of fire management. We have not included the results of the classical method because the values obtained are low and do not contribute information to this work. In addition, previous studies have shown that the values obtained by applying the statistical method surpasses the quality of the prediction achieved by the classic approach [5].

Along the progress of burning, discrete steps were defined to represent the progress of the fire front. Therefore, we consider various time instants $t_0, t_1, t_2...$ etc. In Table 1 can be appreciated the characteristics (size and slope) of the land used for each experiment. In order to gather as much information as possible about the fire-spread behaviour, a camera recorded the complete evolution of the fires. The videos obtained were analyzed and several images were extracted every certain period of time. From the images, the corresponding fire contours were obtained and converted into a suitable format so they could be interpreted by the methods.



Fig. 4. Real fire during the burns in the Gestosa area.

Table 1. Dimensions and slopes of the plots used in experiments.

Experiment	Width (m)	Length (m)	Slope ($^{\circ}$)
1	58	50	21
2	89	91	21
3	95	123	21
4	20	30	6

In experiments 1 and 2 the cell size was 1 m^2 , and in experiments 3 and 4 the cell size was 0.333 m^2 . The remaining parameters such as wind conditions and moisture content were variable.

5.1 The fitness function

It is necessary to define a criterion to compare the prediction resulting from each method with the real situation. To evaluate the system response we have defined a fitness function. Since the simulator uses an approximation based on cells, the fitness function is defined as a quotient. The following equation shows the expression:

$$Fitness = \frac{(\#cells \cap -\#IgnitionCells)}{(\#cells \cup -\#IgnitionCells)}$$

where $\#cells \cap$ represents the number of cells in the intersection between the simulation results and the real map, $\#cells \bigcup$ is the number of cells in the union of the simulation results and the real situation, and #IgnitionCells represents the number of burned cells before starting the simulation.

A fitness value equal to one corresponds to the perfect prediction because it means that the predicted area is equal to the real burned area. On the other hand, a fitness equal to zero indicates the maximum error because, in this case, our experiment did not coincide with reality at all.

5.2 Comparison

According to the information already known about the experiments and the models of Rothermel [21] for some of the parameters, certain ranges have been specified (in particular those parameters that exhibit uncertainty). A part of this information has been measured during the experiment, and the remainder has been taken from standard values used by BehavePlus [1].

The experiments 1, 3 and 4 belong to cases of fires started at the base of the field through pyrotechnic devices in a linear way. Meanwhile, in experiment 2, the fire originated in a single point. After execution of the methods, the fitness values found are shown in Figure 5. We can see that in all four cases, ESS performs better compared to the original version of the method. However, at certain times, the values found may be similar or even slightly lower than the results of the original method (this happens in Experiment 2 at minute 7.5 and in Experiment 3 at minute 12).



Fig. 5. Fitness comparison between the $S^2 F^2 M$ and ESS for four experiments.

It is important to emphasize that for calibration and prediction purposes these methods need one real fire line more than the classical prediction, so we cannot provide suggestions at the first time t_1 , i.e., along the first step of these methods it is only possible to apply the calibration stage whose result will be used in t_2 . Thus, from t_2 to t_n , every step t_i of the methods executes both **CS** and **FP**, basing their **FP** in the K_{ign} provided by the previous step t_{i-1} (see Fig. 2 and 3). This is the reason why the figure shows the results from t_2 considered in each experiment.

Another important point to highlight is to mention that the shown values for ESS are the average of ten executions. In the case of S^2F^2M , this is not necessary because it gives a deterministic output.

5.3 Parallelism and Speedup

The results were obtained by executing both systems on a LINUX cluster (12 processors AMD64 2G RAM and Gigabit Ethernet 1000 Mbps) under an MPI environment [13]. The performance gain has been analyzed using the measure known as Speedup [12], which is defined as the ratio of the time taken to solve a problem on a single processing element to the time required to solve the same problem on a parallel computer with p identical processing elements.

The numbers of processors used were 1, 2, 4, 6, 8, 10 and 12 (although the graph of speedup is usually designed with p equal to successive powers of 2). Figure 6 shows the values obtained as an average of all experiments.



Fig. 6. Speedup for the methods.

The continuous line represents the linear case (or ideal Speedup). As we can see, both methods have a speedup relatively good $(S^2F^2M$ a bit better than ESS). For the purposes of a fair comparison, in both cases were performed the same number of simulations. Thus, in addition to the graph, the execution times are also similar (ESS takes on average 10% less execution time). However, in actual executions, ESS usually takes even less time because in principle, the number of iterations depends on when it finds individuals who meet the expected fitness, and this usually happens before in ESS that in S^2F^2M (remember that S^2F^2M is deterministic and exhaustive method, while ESS is a not deterministic one). For instance, for Experiment 4, ESS can take around 35 *min* to find individuals with fitness equal to 0.85, or it can spend 140 *min* looking for individuals with fitness equal to 0.95. In conclusion, there is a trade-off between time and quality, and depends on the user to configure certain parameters to emphasize either the time restriction or the expected quality.

6 Conclusions

In this work, a method is described, which represents a major enhancement compared to previous methodologies. As we have seen, the techniques that combine high performance computing with statistical methods have excellent ability to solve or reduce the problem of uncertainty in input parameters. For this reason, it is of great interest the ongoing research on this subject, so as to optimize and evolve on the approaches and methods already developed to maximize the results achieved. Then, from $S^2 F^2 M$ we have arrived at the concept of Evolutionary Statistical System (ESS). To do this, we combined the power of the statistical calculation with capabilities provided by parallel evolutionary algorithms, achieving results that actually improve the original methodology $S^2 F^2 M$ based solely on statistical calculation and high performance computing. Both methods have been described throughout the present work. They correspond to methods to reduce uncertainty in the input parameters, in this case applied to the prediction of forest fires spread. Given the costs, risks and obvious difficulties for design multiple fires in real plots to obtain reliable data for experimentation and validation of the methods, the experiments were conducted on four real fires considering different instants of time in each case. In addition to significantly improve the accuracy of the prediction quality of the classical method, one of the most important features of both methods is that they are general enough to be used on different models (floods, avalanches, etc.). Thus, the combination of evolutionary computation, parallelism and uncertainty reduction is a promising option for tackling various Grand Challenge Problems, as in this case it is the prediction of forest fire behaviour.

In this first approach of ESS, we decide apply parallelism only in the evaluation of the individuals, with the goal of gradually increase the degree of parallelism to compare the results offered by each alternative of PEAs. Further study should focus on the analysis and tuning of the method to obtain the best possible results and compare it with other methods.

References

 Andrews, P.L., Bevins, C.D., Seli, R.C.: BehavePlus fire modeling system, version 2.0: User's Guide. Gen. Tech. Rep. RMRS-GTR-106WWW. Ogden, UT: Dept. of Agriculture, Forest Service, Rocky Mountain Research Station (2003) pp. 132

- Ball, G.L., Guertin, D.P.: FIREMAP, in Nodvin, S. C. and Waldrop, T. A., Fire and the Environment: Ecological and Cultural Perspectives: Proceedings of an International Symposium. Knoxville, TN. USDA Forest Service, Southeastern Forest Experiment Station, Asheville, NC. General Technical Report SE-69 (1991) 215–218
- Bianchini, G., Caymes Scutari, P.: Uncertainty Reduction Method Based on Statistics and Parallel Evolutionary Algorithms. Proceedings of High-Performance Computing Symposium - 40 JAHO (HPC 2011, ISSN: 1851-9326) (2011) 1–4
- Bianchini, G., Denham, M., Cortés, A., Margalef, T., Luque, E. (2010): Wildland Fire Growth Prediction Method Based on Multiple Overlapping Solution, Journal of Computational Science, 1 Issue 4 (2010) 229–237
- Bianchini, G., Cortés, A., Margalef, T., Luque, E.: Improved prediction methods for Wildfires using High Performance Computing: A comparison. LNCS 3991, Part I (2006) 539–546
- Bianchini, G., Denham, M., Cortés, A., Margalef, T., Luque E.: Improving forestfire prediction by applying a statistical approach. Forest Ecology and Management. (234, supplement 1) (2006) S210
- Bevins, C. D.: FireLib User Manual & Technical Reference. (2004) http://www.fire.org (Accessed on May 2012)
- Bodrožić, L., Stipanicev, C., Šeric, M.: Forest fires spread modeling using cellular automata approach. Modern trends in control. (eds V. Hladky, J. Paralic, J. Vašcak). (2006) 23–33 (Košice, Slovakia: equilibria)
- Finney M.A.: FARSITE: Fire Area Simulator-model development and evaluation. Res. Pap. RMRS-RP-4, Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. (1998) pp. 47
- 10. Fons W.: Analysis of fire spread in light forest fuels, J. Agric. Res. 72 (1946) 93–121
- Goldberg, D.E.: Genetic and evolutionary algorithms. Come of age. Communications of the ACM, 37 (3) (1994) 113–119
- 12. Grama A., Gupta A., Karypis G., Kumar V.: Introduction to Parallel Computing. Second Edition. Pearson (2003)
- 13. Gropp W., Lusk E., Skjellum A.: Using MPI Portable Parallel Programming with the Message-Passing Interface. Second edition. The MIT Press (1999)
- 14. Jorba, J., Margalef, T., Luque, E., Campos da Silva, J., Viegas, D.X.: Parallel Approach to the Simulation of Forest Fire Propagation. Proc. 13 Inter. Symposium "Informatik fur den Umweltshutz" der Gesellshaft Fur Informatik (GI) (1999) 68–81
- Lopes, A.M.G., Cruz, M.G., Viegas, D.X.: FireStation An integrated software system for the numerical simulation of wind field and fire spread on complex topography. Environmental Modelling & Software, 17 (3) (2002) 269–285
- Mandel, J., Beezley, J.D., Kochanski, A.K.: Coupled atmosphere-wildland fire modeling with WRF 3.3 and SFIRE 2011, Geoscientific Model Development (GMD) 4 (2011) 591–610
- 17. Mattson, T., Sanders, B., Massingill, B.: Patterns for Parallel Programming. Addison-Wesley (2005)
- Michalewics, Z.: Genetic Algorithms + Data Structures = Evolution Programs. Third, Revised and Extended Edition. Springer (1999)
- 19. Montgomery, D.C., Runger G.C.: Probabilidad y Estadística aplicada a la Ingeniería. Limusa Wiley (2002)
- Nelson, K.M.: Applications of evolutionary algorithms in mechanical engineering. (1997) http://digitool.fcla.edu/dtl_publish/34/12514.html (Accessed on May 2012)
- Rothermel, R. C.: A mathematical model for predicting fire spread in wildland fuels, Res. Pap. INT-115, US Dept. of Agric., Forest Service, Intermountain Forest and Range Experiment Station. (Ogden, UT.) (1972)

- 22. Rothermel, R. C.: How to predict the spread and intensity of forest fire and range fires. Gen. Tech. Rep. INT-143, US Dept. of Agric., Forest Service, Intermountain Forest and Range Experiment Station. (Ogden, UT.) (1983)
- 23. Viegas, D.X. (Coordinator) et al., Project Spread Forest Fire Spread Prevention and Mitigation (2004) http://www.algosystems.gr/spread/ (Accessed on May 2012)