Procedia Computer Science
Volume 80, 2016, Pages 418–427
ICCS 2016. The International Conference on Computational Science

# Error Function Impact in Dynamic Data-Driven Framework Applied to Forest Fire Spread Prediction

Carlos Carrillo, Tomàs Artés, Ana Cortés, and Tomàs Margalef

Computer Architecture and Operating Systems Department
Universitat Autònoma de Barcelona, Spain.
{carles.carrillo, tomas.artes}@e-campus.uab.cat
{ana.cortes, tomas.margalef}@uab.cat

**Abstract**

In order to use environmental models effectively for management and decision-making, it is vital to establish an appropriate measure of confidence in their performance. There are different ways and different methodologies to establish and measure the confidence of the models. In this paper, we focus on the forest fire spread prediction. Simulators implementing forest fire spread models require diverse input parameters to deliver predictions about fire propagation. However, the data describing the actual scenario where the fire is taking place are usually subject to high levels of uncertainty. In order to minimize the impact of the input-data uncertainty a Two-Stage methodology was developed to calibrate the input parameters in (1) an adjustment stage so that the calibrated parameters are used, and (2) the prediction stage to improve the quality of the predictions. Is in the adjustment stage where the error formula plays a crucial role, because different formulas implies different adjustments and, in consequence, different wild fire spread predictions. In this paper, different error functions are compared to show the impact in terms of prediction quality in DDDAS for forest fire spread prediction. These formulas have been tested using a real forest fire that took place in Arkadia (Greece) in 2011.

*Keywords:* error function, forest fire, simulation, FARSITE

## 1 Introduction

Natural hazards represent a challenge for society since each year they cause important losses worldwide. In the last decades, physical and mathematical models have been developed and implemented in simulators to predict the behaviour of certain natural hazards [14] [12] [13] [16]. Such models require inputs parameters to describe the environment where the disaster takes place but, unfortunately, it is impossible to obtain the data that populates these models without error. This data uncertainty is due to the difficulty in gathering precise values at the right places where the catastrophe is taking place, or because the hazard itself distorts the measurements. So, in many cases the unique alternative consists of working with interpolated, outdated, or

 Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2016


doi:10.1016/j.procs.2016.05.342

even absolutely unknown values. Obviously, this fact results in a lack of accuracy and quality on the provided predictions. In this work, we focus on forest fire spread prediction as a natural hazard study case. Forest fire propagation simulators require both static and dynamic input data. One defines static data as those parameters that keep quite stable during the propagation simulation process. Topography and vegetation maps are examples of this data typology. On the other hand, dynamic input data are those parameters that are affected by the meteorology as, for example, humidity and wind. Obviously, both sets of parameters have a direct impact in the forecast quality, however, the dynamic ones are the most sensitive parameters [1] [15]. In order to minimize the uncertainty in all this input data and improve the accuracy of the delivered predictions, a Two-Stage Dynamic Data Driven Forest Fire Prediction System was developed [7]. The dynamic data driven approaches for forest fire spread prediction seek to drive the model forecast using control variables. That is, to enhance basic forest fire spread simulations with the knowledge obtained from a calibration/adjustment stage [4] [5]. In this Calibration stage, the obtained simulations are evaluated in order to weight them according to some fitness/error function, which determines the similitude of a given simulation to the observed real fire propagation. Once the best scored simulation is selected, the configuration of the control variables associated to that winner, is applied for prediction purposes in the near future [2]. The problem arises when trying to establish an appropriate level of confidence in the way that the error, in the Calibration stage, is computed. Currently, there exist a large set of techniques to determine the quality of the simulations depending on the underlying problem [3] so, this work aims to analyze how the fitness/error function used in the Calibration stage affects the prediction of the burned area.

This paper is organized as follows. In section 2 the Dynamic Data Driven Two-Stage methodology is briefly explained. Section 3 describes the different functions proposed to compute the simulation error. Section 4 presents the experimental results and, finally, section 5 summarizes the main conclusions and the future work.

## 2 DDDAS for Forest Fire Spread Prediction

As we have mentioned, to overcome the input uncertainty problem in the case of forest fire spread prediction, a DDDAS Two-stage methodology has been used [8]. This strategy is based on extracting relevant information about real observed fire spread to dynamically adjust certain parameters of the simulation process. Then, those calibrated parameters will be introduced in the simulation system to drive the forecast of the evolution of the forest fire (see figure 1). As a search technique, in the Calibration stage, a Genetic Algorithm (GA) has been applied, where a random initial population of individuals (input parameters setting) is generated. Each individual is simulated and the resulting forest fire spread is compared to the real observed fire evolution to compute the fitness of each individual (called error in this case). Then, according to the quality of the simulations evaluated in the Calibration stage, the individuals are ranked and the genetic operators (elitism, selection, crossover, mutation) are then applied to generate a new population. The process is repeated a certain number of iterations and the best individual at the end of the process is selected to drive the prediction of the near future forest fire spread. This dynamic data-driven forest fire spread prediction system has been designed to be simulator independent, so any forest fire simulator could easily be included in a plug&play fashion. In particular, the forest fire spread simulator used is this work has been the so called FARSITE ([10]).

Evaluating the simulations quality is a key point in this scheme, because different error formulas imply different rankings of the individuals. Since the obtained ranking is used to