Statistical parameter estimation for a cellular automata wildfire model based on satellite observations

E. Couce & W. Knorr
Department of Earth Sciences, University of Bristol, UK

Abstract

The importance of understanding the impact of wildfires on natural ecosystems has given rise to the development of realistic computer models for the simulation of wildfires. Stochastic models based on simplified equations and local interactions, such as Cellular Automata (CA) models, are particularly popular as an alternative to more computationally demanding deterministic models. However, the challenges associated with observing wildfires under natural conditions, and the highly non-linear nature of fire spread makes it extremely difficult to parameterize them. In this work we present a method for adjusting the behaviour of one such CA model from the statistical analysis of satellite data of more than 750,000 African wildfires detected in 2003. Statistical metrics are developed to characterize agreement between model and satellite observations. The average probability of fire transmission amongst cells and the spatial scale of the model are adjusted so that maximum agreement is found between model output and the observed extension and statistical distribution of the real fires. While the results obtained are only valid for the particular CA model used and within the geographical limits of the region studied, we believe the process could be adapted to fine-tune and validate other CA models in regions where enough fire observations are available.

Keywords: fire spread model, cellular automata, parameter estimation, African savanna wildfires, satellite observations.

1 Introduction

The importance of wildfires for natural ecosystems, together with the socio-economic danger they represent, have lead to a great deal of effort invested in the

simulation and modelling of fire behaviour. There is strong demand for accurate fire simulations that would provide an important tool for fire fighters and other people involved in fire management. However the modelling and understanding of wildfires is a highly complex problem, governed by non-linear equations and depending on more factors than can possibly be known at any time and which affect the fire behaviour in non-trivial ways. Because of this, deterministic models based on theoretical partial differential equations have only found limited success in the description of wildfires, and this only at the cost of large amounts of computer resources and processing time [1, 2]. Thus, stochastic models based on simplified empirical or semi-empirical equations have remained a popular alternative.

One of the main obstacles to the development of a realistic model for the simulation of the spread of wildfires lays in its validation with data from real fires. This is mainly due to the fact that the knowledge of the factors required by the models (such as humidity, wind information and vegetation type and state) is limited for the wildfires on record. Many published fire spread models are not compared directly to real data. Instead, a model's validity is often discussed by comparing its output to that of some better-known model, usually for only one or two particular cases. If real data are used in the comparison, they are typically from a single wildfire and of high spatial and temporal resolution. While this is no doubt a very relevant initial test, it can hardly be considered enough for a process as complex as wildfire spread, particularly if the model is to be applied to a wider range of conditions.

In this work we compare the output of a new stochastic CA model for the spread of wildfires with statistics of the areas of real fires detected by satellite within an extensive region of Africa during the 2003 fire season (nearly 800,000 wildfires). The model's parameters are initialized with observations of vegetation type, wind, temperature, precipitation, and FAPAR (fraction of plant-absorbed photosynthetically active radiation) from the study region. They are then adjusted to better reproduce the histogram of the observed fire areas, in order to establish the model's optimal spatial resolution and average probability for fire spread. While this technique does not replace the need to compare a model's output with data of high spatial and temporal resolution, we believe it represents a significant step towards comprehensive validation of fire spread models.

2 Methods

2.1 Fire data

The fire data used in our study was obtained by the Moderate Resolution Imaging Spectroradiometer (MODIS) carried by the Terra (EOS AM) and Aqua (EOS PM) satellites, with a spatial resolution of 500x500 m^2. The study region (Figure 1) corresponds to the MODIS tiles h19v10 and h20v10, an area of approximately 1200x2400 km^2, with latitude spanning 10.00ºS to 20.00ºS, and

longitude between 10.15ºE and 31.92ºE. It includes portions of Angola, Zambia, Namibia, Botswana, Zimbabwe, Congo, Zaire and Mozambique. Data were available for the fire seasons (April to November) of 2000 to 2004.

Figure 1:   Study region, indicated by rectangle.

Fires were identified with a generic algorithm being developed by Rebelo et al. [3] and Rebelo [4]. This algorithm detects areas exhibiting sudden changes based on discrepancies between expected and observed bi-direction reflectance (BRDF) observations. In the study region, one of the most common causes of these sudden changes are wildfires, although several additional tests are used to separate burning from phenological changes within the pixel. The algorithm is similar to that of the MODIS fire product (see Roy *et al.* [5, 6]).

The algorithm detects the day a sudden change in BRDR suggests the onset of a wildfire in the region. Therefore the only information available is the probable day a fire starts for a pixel, and nothing is known about the fire's duration. Nevertheless considering the type of vegetation present in the study region, predominantly savanna and shrubland, these wildfires would tend to propagate fast and have short duration.

2.2 Additional geographical information

The Type 2 MODIS Land Cover Product (MOD12Q1) was used to identify types of vegetation present within the study region [7, 8]. Five different types of vegetation were considered in the study: savanna, woody savanna, grassland, open shrubland and deciduous broadleaf forest. Together they encompass over 80% of the total area analyzed, with savanna and woody savanna representing 45% and 20% respectively.

In order to estimate the potential amount of burnable fuel available in each grid cell, we employed the monthly, gridded 0.5° by 0.5° FAPAR product of Gobron *et al.* [9] for 2002-2003, generated from an analysis of the data recorded by the Sea-viewing Wide Field-of-view Sensor (SeaWiFS).

Some wind information was available for the period from February 1999 to October 2001 from measurements carried out by a station of the FLUXNET network located near Maun, Botswana [10, 11].

For the climate input we used daily precipitation and daily minimum and maximum temperatures. The values of those variables and of solar incoming radiation for the period 2000 to 2004 were generated on a global 2° latitude by 2° longitude grid using the method of Nijssen *et al.* [12], based on daily station data from the Summary of the Day Observations (Global CEAS), National Climatic Data Center, and monthly gridded data. Monthly gridded temperature was obtained from the data set of Jones *et al.* [13, 14], with gaps filled from data of Hansen *et al.* [15, 16]. Monthly gridded precipitation data came from a 1.0° version of Chen *et al.* [17].

2.3 The model

The model used in the study [18] is laid out on a rectangular 2-dimensional lattice. It takes the cells initially on fire as input, and reproduces the possible evolution of the fire over successive time steps. The fire spread relies on the computation of semi-empirical probabilities of fire transmission from cells on fire and is stochastic in nature. The probabilities are computed based on climate factors, vegetation type, wind intensity, topography, fire intensities, and fuel content of each cell. Although the main form of propagation occurs among neighbouring cells, propagation from other cells is also possible by the emission of sparks, influenced both by wind and topographic conditions. The duration of the fire on each cell is computed from the amount of fuel existing on the cell and the fire's intensity. The model does not present spurious symmetry, and the results obtained appear realistic and successfully reproduce features of real wildfires, such as spotting.

For this analysis, the probability of transmission was expressed as the product of independent factors reflecting the effects of vegetation type, climate, wind strength, and average fuel load respectively. The effect of the vegetation type was implemented by re-scaling the percentages of the area burned within each vegetation type against data from a previous exhaustive study of the region during 2000 to 2004 (Roy *et al.* [5]). Climatic factors affecting fire spread make use of the widely adapted Nesterov Index [19], which takes into account the maximum daily temperatures of any series of consecutive days without significant rainfall. A preliminary analysis of the fire behavior in the area for the period for which wind data were available suggested a relationship between the probability of transmission and the maximum daily wind strength. Finally, the average litter load in the region of interest was approximated as the integral of the losses in leaf mass, estimated from time-integrated decreases in satellite-derived FAPAR.

2.4 Methods

Since the spatial resolution of the fire data (250,000 m^2 per pixel) is much coarser than the model's lattice cells, the comparison with the model output was

carried out statistically considering the total burned area of each pixel. Burning pixels were considered to belong to the same fire if they were both contiguous in space (neighbouring cells) and time (the changes must have been detected either the same day, or with one day difference). A histogram with the fire areas of the nearly 800,000 fires detected in 2003 was compared with histograms obtained from a large number of model simulations.

For each model run, a fire starting point within the study area and a start data between April to November of 2003, the period for which all required data were available, were chosen at random. The random election of a date is justified, since the amount of detected fires barely changed between the different months of the fire season analyzed (not so the number of burning pixels, which clearly peaks in July and August, implying more extensive fires during those months).

The initial distribution of fuel load on the model lattice, i.e. at a much higher spatial resolution than the available satellite observations, was again generated through a random process for each run. The average fuel values were thereby kept below 30%, giving rise to fast, low intensity wildfires, similar to the ones observed for typical savanna conditions.

The contributions of type of vegetation, climate, and fuel load to the average probability of transmission were computed as a function of vegetation type, litter load, and Nesterov index. For the computation of the effect of the wind, a random value for the maximum daily wind strength was generated, based on the monthly averages and standard deviations values measured at the Botswana station.

The final probability of fire transmission between cells was obtained from the product of the contributions of vegetation type, climate factor, and fuel load multiplied by a global constant k, which was optimised by maximizing agreement between modelled and observed fire extension histograms. k was allowed to take on several values within the interval $(0, 1)$. The model was run multiple times for each value of k. Over 350,000 simulations were run in total, all employing a 100x100 lattice.

The equivalence between the number of cells that are counted as burned in a simulation and a burned pixel from the satellite fire data depends both on the spatial resolution of the model (which was allowed to change) and the percentage of the area of a pixel that needs to burn in order for the fire to be detected. For the satellite fire detection algorithm employed, the latter falls somewhere between 10 and 20%, and in this study it was set at 15% for comparison with the model output.

3   Results

The comparison between the histogram derived from observations and model output was carried out using the Kullback-Leibler divergence [20], which has previously been applied to the testing of other ecological models [21, 22].

The number of bins for the histogram is limited by the number of satellite data pixels that correspond to the area represented by the model's 100x100

lattice. We found an optimal representation at seven bins, considering number of events per bin and number of data points of the two histograms.

Figure 2 shows the variation of the Kullback-Leibler divergence as a function of the model's resolution (A, B) and the value of k (A, C). We can see that the KL divergence value does not vary much with resolution as long as it stays above ca. 22–23 meters, although the fit does get progressively worse as the spatial resolution gets coarser. The dependence on k is much stronger, with a very well defined minimum at 0.46.

The optimal model histogram (corresponding to k=0.46 and a spatial resolution of 26x26 m^2) is shown in Figure 3 for comparison with that of the satellite observations from the study area (in black). We find good agreement between the histograms, with a final value for the Kullback-Leibler divergence of 0.015.
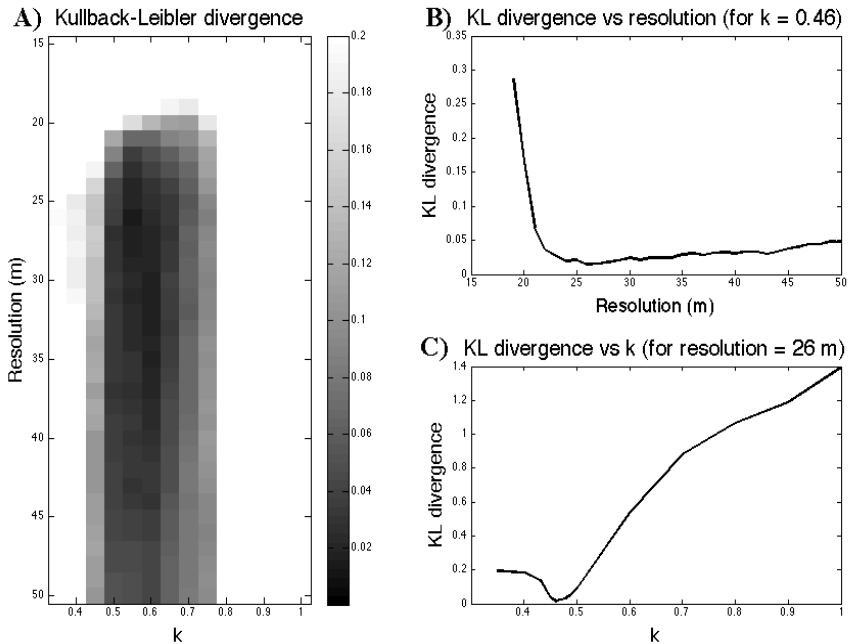

Figure 2:   A) Kullback-Leibler divergence as a function of the probability factor k and the model's spatial resolution (indicated by the length corresponding to the side of a cell). The minimum is found for k=0.46 and a resolution of 26x26 m^2. B) KL divergence as a function of the spatial resolution, for k=0.46. C) KL divergence as a function of k, for a resolution of 26x26 m^2.
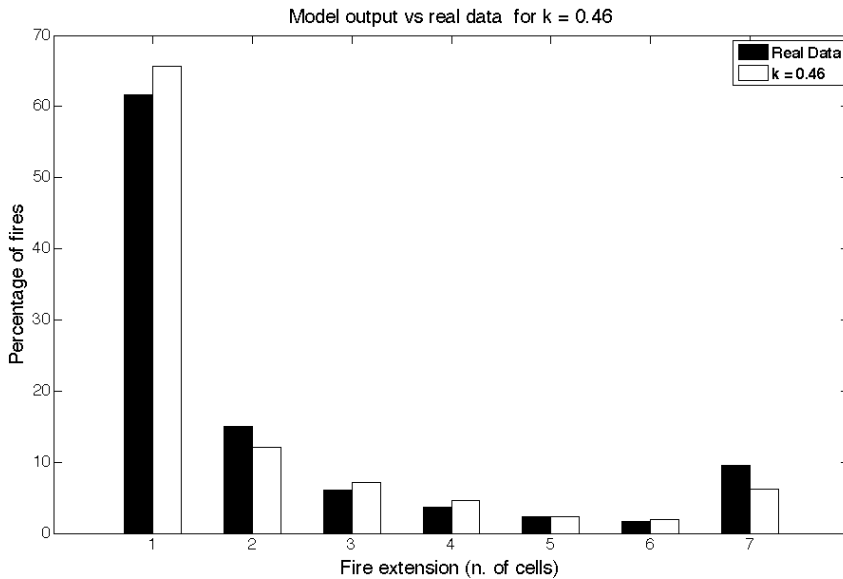
Figure 3: Comparison of the histograms with the percentage of fires detected as a function of their extension, for real data (black) and the optimal model output (white), for k=0.46 and a spatial resolution of 26x26 m^2. Fire extension is measured in pixels, each corresponding to 250,000 m^2. For these two distributions the Kullback-Leibler divergence reaches the minimum value of 0.015.

4 Conclusions

In this work we have presented a new stochastic model for the prediction of the spread of wildfires based on cellular automata on a square grid, and its application to African savanna fires. After running multiple simulations with random conditions from the extensive study region, the fire area distribution of the model was compared to the satellite-derived fire data for the fire season of 2003 for a large region of southern Africa, with over 750,000 detected wildfires. The method has allowed the selection of the optimal average probability of fire spread and spatial resolution of the model. A good agreement has been found, with a value for the Kullback-Leibler divergence of 0.015 for a 7-bin histogram of the frequency each fire area was detected/simulated.

We believe this result lend valuable credibility to the model in an extensive set of conditions, in particular given that fire spread models are rarely compared to such an extensive set of real wildfire data. However it does not eliminate the need of further tests. The direct comparison of model output with the evolution of real fires with data available at high spatial and temporal resolution would greatly benefit the adjustment of the model's behaviour, particularly regarding the effects of topography and wind on the probability of transmission.

Acknowledgements

References

[1] Bradley, J.H. & Clymer, A.B., Difficulties in the simulation of wildfires. *1993 International Emergency Management and Engineering Conference*: pp. 161–171, 1984.

[2] Karafyllidis, I. & Thanailakis, A., A model for predicting forest fire spreading using cellular automata. *Ecological Modelling*, **99(1)**, pp. 87–97, 1997.

[3] Rebelo, L., Lewis, P., et al., A temporal-BRDF model-based approach to change detection. *Geoscience and Remote Sensing Symposium, IGARSS '04*, 2004.

[4] Rebelo, L.M., The development of a generic change detection scheme: An application to the identification and delineation of fire affected areas. PhD thesis, Department of Geography, Remote Sensing Unit, London, University College London, 2005.

[5] Roy, D.P., Lewis, P.E., et al., Burned area mapping using multi-temporal moderate spatial resolution data–a bi-directional reflectance model-based expectation approach. *Remote Sensing of Environment*, **83(1–2)**, pp. 263–286, 2002.

[6] Roy, D.P., Jin, Y., et al., Prototyping a global algorithm for systematic fire-affected area mapping using MODIS time series data. *Remote Sensing of Environment*, **97(2)**, pp. 137–162, 2005.

[7] Strahler, A., Muchoney, D., et al., MODIS Land cover product: Algorithm Theoretical Basis Document. 1999.

[8] Hansen, M.C., Defries, R.S., et al., Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, **21(6–7)**, pp. 1331–1364, 2000.

[9] Gobron, N., Pinty, B., et al., Evaluation of fraction of absorbed photosynthetically active radiation products for different canopy radiation transfer regimes: Methodology and results using Joint Research Center products derived from SeaWiFS against ground-based estimations. *Journal of Geophysical Research-Atmospheres*, **111(D13)**, 2006.

[10] Veenendaal, E.M., Kolle, O., et al., Seasonal variation in energy fluxes and carbon dioxide exchange for a broad-leaved semi-arid savanna (Mopane woodland) in Southern Africa. *Global Change Biology*, **10(3)**, pp. 318–328, 2004.

[11] Arneth, A., Veenendaal, E.M., et al., Water use strategies and ecosystem-atmosphere exchange of CO2 in two highly seasonal environments. *Biogeosciences*, **3(4)**, pp. 421–437, 2006.

[12] Nijssen, B., Schnur, R., et al., Global retrospective estimation of soil moisture using the VIC land surface model, 1980–1993. *Journal of Climate*, **14**, pp. 1790–1808, 2001.

[13] Jones, P.D., New, M., et al., Surface air temperature and its variations over the last 150 years. *Reviews of Geophysics*, **37**, pp. 173–199, 1999.

[14] Jones, P.D., Osborn, T.J., et al., Adjusting for sampling density in grid box land and ocean surface temperature time series. *Journal of Geophysical Research-Atmospheres*, **106(D4)**, pp. 3371–3380, 2001.

[15] Hansen, J., Ruedy, R., et al., GISS analysis of surface temperature change. *Journal of Geophysical Research-Atmospheres* **104(D24)**, pp. 30997–31022, 1999.

[16] Hansen, J., Ruedy, R., et al., A closer look at United States and global surface temperature change. *Journal of Geophysical Research-Atmospheres,* **106(D20)**, pp. 23947–23963, 2001.

[17] Chen, M.Y., Xie, P.P., et al., Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, **3(3)**, pp. 249–266, 2002.

[18] Couce, E., A stochastic cellular automata model for the spread of wildfires: casestudy of the African savanna fires. MSc thesis, Dept. of Earth Sciences, University of Bristol, Bristol, 2008.

[19] Nesterov, V.G., Fire Frequency Index and Methods of Its Determination, Goslesbumaga, Moscow, 1949.

[20] Kullback, S. & Leibler, R.A., On information and sufficiency. *Annals of Mathematical Statistics*, **22(1)**, pp. 79–86, 1951.

[21] Burnham, K.P. & Anderson, D.R., Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research,* **28(2)**, pp. 111–119, 2001.

[22] Burnham, K.P. & Anderson, D.R., *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second Edition,* Springer Science, New York, 2002.