Nat. Hazards Earth Syst. Sci. Discuss., 2, 3219–3249, 2014
www.nat-hazards-earth-syst-sci-discuss.net/2/3219/2014/
doi:10.5194/nhessd-2-3219-2014

Natural Hazards and Earth System Sciences
Open Access
Discussions

Evaluation of forest fire models on a large observation database

J.-B. Filippi[1], **V. Mallet**[2,3], **and B. Nader**[1]

[1]CNRS SPE, Università di Corsica, BP 52, 20250 Corte, France
[2]INRIA, BP 105, 78 153 Le Chesnay cedex, France
[3]CEREA (joint laboratory École des Ponts ParisTech & EDF R&D, université Paris Est), Marne-la-Vallée, France

Received: 16 April 2014 – Accepted: 17 April 2014 – Published: 8 May 2014

Correspondence to: J.-B. Filippi (filippi@univ-corse.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This paper presents the evaluation of several fire propagation models using a large set of observed fires. The observation base is composed of 80 Mediterranean fire cases of different sizes, which come with the limited information available in an operational context (burned surface and approximative ignition point). Simulations for all cases are carried out with 4 different front velocity models. The results are compared with several error scoring methods applied to each of the 320 simulations. All tasks are performed in a fully automated manner, with simulations ran as first guesses with no tuning for any of the models or cases. This approach leads a wide range of simulation performance, including some of the bad simulation results to be expected in an operational context. Regardless the quality of the input data, it is found that the models can be ranked based on their performance and that the most complex models outperform the more empirical ones. Data and source code used for this paper are freely available to the community.

1 Introduction

Model evaluation requires comparing predicted to observed values and is critical to establish model's potential errors and credibility. The first step to evaluate model performance is to be able to evaluate single simulation results against observations, such scoring methods have been the subject of several studies, initiated by Fujioka (2002) and recently compiled and extended in Filippi et al. (2013). Basically it is clear from all these studies that a single value cannot be representative of a model performance, as it only gives limited insights on all aspects of performance, while the analysis of a human eye provides a better understanding of what was good and what went wrong.

Nevertheless, the problem of evaluating model performance must be tackled as it is important to know if a parametrisation or a new formulation is superior, and to continue the process of enhancing models, codes and data. From an operational point of

NHESSD
2, 3219–3249, 2014

view, it also appears that models are used with a clear lack of systematic evaluation as noted recently by Alexander and Cruz (2013). A major step in such model evaluation is to compare observed rate of spread (ROS) with simulated ROS, as much data exists in the literature. Cruz and Alexander (2013) carried out such comparison with the clear and reassuring conclusion that well empirical and semi-empirical built models may provide a good ROS approximation. Our study focuses on the use of these models to simulate the overall two-dimensional fire spread and its corresponding burned area. While evaluating the absolute model performance is yet out of the question, it is proposed here to evaluate specific model performance. This specific evaluation will be linked to a typical model usage and to a territory. The typical usage proposed corresponds to the plausible "first guess" case where just an ignition location is known with no direct observation of the wind or fuel moisture near the fire. The selected area is the Mediterranean island of Corsica where numerous wildfires occur every year in a variety of configurations. The test consists in running simulations with four different models and a large number of observations, and compiling the results in the form of comparison scores between simulated and observed fires. These simulations must be run in a fully automated manner, without observation biases introduced by manual adjustment of fuel, wind or ignition location in order to enhance results. The overall results are the distributions of scores, using different scoring methods, which fulfills the goal of ranking the models according to their specific use.

The four different models are presented in the first section along with the simulation method used to compute the front propagation. The second section details the evaluation method, data preprocessing and numerical set-up. The results are presented and discussed in the last section along with focuses on specific cases.

2 Models description

Fire propagation modeling can refer to a vast family of codes, formulations, systems or even datasets (Sullivan, 2009a, b). As this study focuses on the evaluation of large

NHESSD
2, 3219–3249, 2014

scale fire simulation, our selected definition of "fire model" is the formulation of the fire-front velocity. A velocity is obviously not enough to obtain fire progression and burned areas. A fire-front solver code and input data are needed. These two are the same for all models and described in the next section. Note that the proposed model selection is unfortunately not exhaustive of all existing formulations, but rather focuses on representing some kind of evolution in the model types.

Depending on their complexity, the models can take into account the terrain slope, the atmospheric properties (wind velocity $\boldsymbol{v}$, air density ρ_a and temperature T_a), a spatial characterization of the fuels (mass loading σ, density of live/dead $\rho_{l,d}$, height e, surface to volume ratio S_v, emissivity ϵ_v and moisture content m defined as the fraction of water over total weight) and the fuel combustion properties (ignition temperature T_i, calorific capacity $c_{p,v}$, combustion enthalpy Δh, stoichiometry s and mass exchange rate due to pyrolysis $\dot{\sigma}$). Each model prognoses the fire front velocity V in the normal direction to the front $\boldsymbol{n}$, pointing toward the unburned fuel.

2.1 Three-percent model

The first and most simple model makes the strong assumption that the fire is propagating at three percents of the wind velocity, as long as there is fuel available, regardless of the vegetation changes or the terrain slope. In practice, in order to compute the velocity everywhere on the fire front, the wind normal to the front $W_s = \boldsymbol{v} \cdot \boldsymbol{n}$ is taken here as wind velocity, and $V = 0.03 W_s$. This "rule of thumb model" is sometimes used by firefighters, with caution because of its lack of reliability. It will serve here the purpose of being, hopefully, the lowest reference in terms of performance.

2.2 Rothermel model

The quasi-empirical Rothermel model (Rothermel, 1972) forms the basis of the United States National Fire Danger Rating System and fire behavior prediction tool BEHAVE (Andrews, 1986). It builds on earlier works from Byram and Fons (1952) and is based

NHESSD
2, 3219–3249, 2014

on a heat balance developed by Frandsen (1971). It highly relies on a set of parameters obtained from wind tunnels experiments in artificial fuel beds (Rothermel and Anderson, 1966) and Australian experiments (McArthur, 1966). This model is widely used also in Mediterranean countries for various purposes (fire risk and behavior), but usually requires some adjustments by experts in order to be fully efficient on all fuel types. Since this study is created as a blind test for each model, these adjustments were not performed here; hence the default values given in Rothermel (1972) such as moisture of extinction ($M_\chi = 0.3$) and mineral damping ($\eta_s = 1$) were used.

This quasi-physical model uses also a number of fitted parameters (in US Customary units) that read:

$$V = \frac{I_r \xi (1 + \phi_V + \phi_P)}{(\rho_d 250 + 1.116 m \exp(-138/S_v))} , \quad (1)$$

with reaction intensity

$$I_r = R' \sigma \Delta H \eta_s (1 - 2.59(m/M_\chi) + 5.11(m/M_\chi)^2 - 2.59(m/M_\chi)^3) . \quad (2)$$

A propagating flux ratio given by

$$\xi = (192 + 0.2595 S_v)^{-1} \exp(0.792 + 0.681\sqrt{S_v}(\beta + 0.1)) . \quad (3)$$

The wind factor is

$$\phi_V = 7.47 \exp(-0.133 S_v^{0.55}) W_s^{0.02526 \times S_v^{0.54}} . \quad (4)$$

The slope factor is

$$\phi_P = 5.275 \beta^{-0.3} \alpha^2 . \quad (5)$$

An optimal packing ratio with

$$\beta_{op} = 3.338 S_v^{-0.8189} . \quad (6)$$

NHESSD
2, 3219–3249, 2014

The packing ratio is given by

$$\beta = \rho_d / \rho_l \, . \tag{7}$$

The maximum reaction rate is

$$R'_{Max} = S_v^{1.5} \times (1/(495 + 0.0594 S_v^{1.5})); \, . \tag{8}$$

The reaction rate is given as

$$R' = (R'_{Max}(\beta/\beta_{op}))^A \exp(A(1 - \beta/\beta_{op})) \, , \tag{9}$$

where

$$A = 1/(4.774 S_v^{0.1} - 7.27) \, . \tag{10}$$

2.3 Balbi model

The Balbi model (Balbi et al., 2009) like Rothermel can be classified as a quasi-physical model. Its formulation is based on the assumption that the front propagates as a radiating panel in the direction normal to the front. The model verifies that for a specific wind, terrain and fuel configuration the absorbed energy equals the combustion energy directed toward the unburned fuel. This energy is the sum of "radiant part" from the flame and a "conductive" part within the fuel layer. Assumption is also made that only a given portion χ_0 of the combustion energy is released as radiation because the flame is viewed as a tilted radiant panel with an angle γ toward the unburned fuel. The equation governing the propagation velocity of the front reads

$$V = V_0(\epsilon_v, T_i, e, \sigma, m, T_a) + \chi_0 \Delta h \dot{\sigma} f(\lambda, \gamma) \, , \tag{11}$$

where $V_0 = \epsilon_v B T_i^4 e / 2\sigma[c_{p,v}(T_i - T_a) + m\Delta h_w]$ is the contribution of the vegetation undergoing pyrolysis (B is the Boltzmann constant and Δh_w the water evaporation enthalpy).

NHESSD
2, 3219–3249, 2014

The second term accounts for the propagation by radiation and reads

$$f(V,\gamma) = \frac{R}{2+\mu\tau\cos\gamma}(1+\sin\gamma-\cos\gamma)\mathcal{H}_{\mathbb{R}^+}(\gamma)\,, \tag{12}$$

where $\mathcal{H}_{\mathbb{R}^+}$ is the Heavyside function for positive reals and μ is an evolution coefficient of the ratio between radiated energy and released combustion energy. The volume-to-surface ratio is noted S_v and τ is the burning duration given by the Andersen model (Anderson, 1969). The flame tilt angle γ depends on the slope angle α and wind $\boldsymbol{v}$ projected onto the normal to the front:

$$\tan\gamma = \tan\alpha + \rho_a(\boldsymbol{v}\cdot\boldsymbol{n})/2(1+s)\dot{\sigma} \text{ with } \boldsymbol{n}\,. \tag{13}$$

A major assumption of Balbi and Rothermel models is that the fire is always traveling at a stationary speed that verifies $V = \kappa/\tau$, and that all energy is absorbed within the fuel bed for the computation of V_0. Because of these assumptions, front velocity is not dependent on the local fire state (previous intensity, front curvature, depth). It cannot accelerate or go to extinction, and it is only dependent on the local fuel, wind and terrain properties. These assumptions are required to compute a priori potential rate of spread without knowing explicitly the local front depth λ or its curvature κ, such as in the BEHAVE tool. Later versions of Rothermel added sub-models for acceleration or extinction. A more fundamental approach was developed for the Balbi model with a non-stationary formulation.

2.4 Balbi non-stationary model

By using the front tracking solver, local front depth λ and curvature κ are always available as numerical diagnostics of the front. The introduction of these variables in the model was rather simple as it removed strong assumptions. The updated formulation reads:

$$V = V_0(\epsilon_v, T_i, e, \lambda, \sigma, m, T_a) + \chi_0\Delta h\dot{\sigma}f(\lambda,\gamma)\,, \tag{14}$$

NHESSD
2, 3219–3249, 2014

with

$$V_0 = \left(1 - e^{\frac{\lambda\beta_d}{4}}\right) \epsilon_v B T_i^4 e / 2\sigma[c_{p,v}(T_i - T_a) + m\Delta h_w] \,, \tag{15}$$

and β_d a radiation dumping ratio that depends on fuel packing. The second term removes the requirement for the stationary speed with

$$f(\kappa, \lambda, \gamma) = (1 - \cos\kappa)\frac{\lambda}{2 + \mu\lambda\cos\gamma}(1 + \sin\gamma - \cos\gamma)\mathcal{H}_{\mathbb{R}^+}(\gamma) \,. \tag{16}$$

The main disadvantage of the model is that it is now tight to a solver able to locally and constantly diagnose λ and κ with a reasonable numerical cost, introducing in the process some additional numerical errors. The solver used for the study is the front tracking code ForeFire.

2.5 Fire propagation method

The fire propagation solver ForeFire (Filippi et al., 2009) uses a front tracking method that relies on a discretization with Lagrangian markers. The outward normal $\boldsymbol{n}_i$ of marker i defines the direction of propagation, and v_i is the local front speed. The maximum distance d_m allowed between two consecutive markers is called the perimeter resolution. If two markers are further away than this distance, a re-mapping of the front is carried in order to keep the resolution constant. A filtering distance $d_f = d_m/2$ is also needed to avoid over-crossing of two markers and potential inversion of the normal. The advection scheme is an first-order Euler scheme in space:

$$\begin{aligned} \boldsymbol{x}_i^{(n+1)} &= \boldsymbol{x}_i^{(n)} + \delta l \cdot \boldsymbol{n}_i^{(n)} \,, \\ t_i^{(n+1)} &= t_i^{(n)} + \frac{\delta l}{v_i^{(n)}} \,, \end{aligned} \tag{17}$$

where the $\cdot^{(n+1)}$ refers to the value of a variable at the next step, and $\cdot_i$ the marker index. Spatial increment δl determines the resolution of the front propagation and should

NHESSD
2, 3219–3249, 2014

be smaller than the smallest space scale influencing the fire propagation, which are usually fire breaks such as roads, i.e., in typical simulations $\delta l \approx 1$ m.

3 Evaluation method

A selection of 80 fire cases have been compiled into an observation database for this study. For each fire, the required initial data is preprocessed to generate the initial conditions and the data required by the selected propagation model. The simulations are then run by distributing the computation of the different cases. Once a simulation result is available for an observation/simulation pair, the comparison is computed with the scores introduced in Sect. 3.3.

3.1 Observation database

The observation database is composed of 80 fires that all occurred in the Mediterranean island of Corsica between 2003 and 2006. These cases were extracted from the Prométhée database (http://www.promethee.com/) a repository of wildfire observations managed by the "Institut Géographique National" (http://www.ign.fr/). Data within this database offers precise burned surface for many wildfires, along with an information on the ignition date and sometimes location. Nevertheless, this information cannot always be trusted, so each case was reviewed with a field expert in order to validate ignition points and date, which reduced the amount of relevant observations. The selection of the Corsica island was made because field expertise was available, as well as adequate data and homogeneity in fire dynamics given the relatively limited dataset (mostly shrubs and Mediterranean maquis). Fire sizes in the selection ranged from one hectare to several hundreds hectares, in order to be representative of all potential model uses.

NHESSD
2, 3219–3249, 2014

3.2 Data preprocessing

The first step in order to launch a simulation is to compile and format data in a way that it can be processed by the fire propagation solver. For the fire simulation code ForeFire, the input data is composed of a configuration file, an elevation field, one or several fields of wind direction and speed, a land use field and a fuel field.

The design of the study implied that the data was not tailored for any model or test case. All of the preprocessing is thus automatically done for the ignition location and date. We are aware that this automatic generation will generate input data that might seem totally unrealistic compared to the observation. In particular, wind direction and fuel state can be significantly different from the real values. For the considered fires, or for any fire that may ignite at anytime, anywhere, the exact inputs to the simulation models are not observed or stored. This study tries to rely on the best available data provided in an operational context.

The simulation configuration defines the size of the domain, the date and the numerical set-up. The rest of the input data is extracted from data files with the same domain extent and date using the GDAL library and its tool `ogr2ogr` (GDAL Development Team, 201x) using a conformal projection.

The elevation field originates from the IGN BDAlti at 25 m resolution. It is originally available in DEM ASCII format.

Wind information originates from the nearest of 12 automatic 10 m-high weather stations of the Météo France network, relatively evenly distributed across the island. The wind speed and direction is taken at nearest time with data available from this nearest station.

Wind fields is extrapolated with the WindNinja mass consistent code (Forthofer, 2007). WindNinja inputs consist of the elevation field and the wind station data. The wind field is outputted at the same resolution as the elevation field and is used for the whole duration of the simulation (under the strong assumption that the wind does not change direction during the fire).

NHESSD
2, 3219–3249, 2014

Fuel distribution data is taken in vector format (shape file) from the IGN IFEN ("Inventaire Forestier National"), with locally 10 fuel classes corresponding to the main burning vegetation in Corsica. Fuel parameters are derived from Anderson et al. (1981) for the grass and pine forest types, while the Proterina parametrisation (Santoni et al., 2011) is used for the shrubs and maquis. As little is known about the fuel state, the fuel moisture is set as moderate to high water stress for every fuel model, which corresponds to high fire danger in summer day.

3.3 Comparison scores

Our evaluation relies on scoring methods that were analyzed, and for two of them proposed, in Filippi et al. (2013). We denote $\mathcal{S}(t)$ the burned surfaces at time t in the simulation. The final simulation time is t_f. At the end of the observed fire t_f^o ($^\mathrm{o}$ stands for observation), the observed burned surface is $\mathcal{S}(t_\mathrm{f}^\mathrm{o})$. Ω is the simulation domain. The arrival time at some point X is denoted $\mathcal{T}(X)$ for the simulation and $\mathcal{T}^\mathrm{o}(X)$ for the observation. Finally, the area of a surface $\mathcal{S}$ is denoted $|\mathcal{S}|$. We assume that the ignition time is 0.

We relied on the following scores, with $t_\mathrm{u} = \max(t_\mathrm{f}, t_\mathrm{f}^\mathrm{o})$:

$$\text{Sørensen Similarity Index} = \frac{2\left|\mathcal{S}^\mathrm{o}(t_\mathrm{u}) \cap \mathcal{S}(t_\mathrm{u})\right|}{|\mathcal{S}^\mathrm{o}(t_\mathrm{u})| + |\mathcal{S}(t_\mathrm{u})|}, \tag{18}$$

$$\text{Jaccard Similarity Coefficient} = \frac{\left|\mathcal{S}^\mathrm{o}(t_\mathrm{u}) \cap \mathcal{S}(t_\mathrm{u})\right|}{|\mathcal{S}^\mathrm{o}(t_\mathrm{u}) \cup \mathcal{S}(t_\mathrm{u})|}, \tag{19}$$

$$\text{Kappa Statistics} = \frac{P_\mathrm{a} - P_\mathrm{e}}{1 - P_\mathrm{e}}, \tag{20}$$

where

$$P_\mathrm{a} = \frac{\left|\mathcal{S}^\mathrm{o}(t_\mathrm{u}) \cap \mathcal{S}(t_\mathrm{u})\right|}{|\Omega|} + \frac{\left|\Omega \backslash (\mathcal{S}^\mathrm{o}(t_\mathrm{u}) \cup \mathcal{S}(t_\mathrm{u}))\right|}{|\Omega|}, \tag{21}$$

NHESSD
2, 3219–3249, 2014

and

$$P_e = \frac{\left|\mathcal{S}^o(t_u)\right| |\mathcal{S}(t_u)|}{|\Omega|^2} + \frac{\left|\Omega\backslash\mathcal{S}^o(t_u)\right| |\Omega\backslash\mathcal{S}(t_u)|}{|\Omega|^2}, \tag{22}$$

$$\mathrm{ATA} = 1 - \frac{1}{\left|\mathcal{S}(t_f)\cup\mathcal{S}^o(t_f^o)\right| \max(t_f, t_f^o)} \left[\int_{\mathcal{S}(t_f)\cap\mathcal{S}^o(t_f^o)} \max(\mathcal{T}(X) - \mathcal{T}^o(X), 0)\mathrm{d}X \right. \tag{23}$$

$$\left. + \int_{\mathcal{S}(t_f)\backslash\mathcal{S}^o(t_f^o)} \max(t_f^o - \mathcal{T}(X), 0)\mathrm{d}X + \int_{\mathcal{S}^o(t_f^o)\backslash\mathcal{S}(t_f)} (t_f - \mathcal{T}^o(X))\mathrm{d}X \right],$$

$$\mathrm{SA} = 1 - \frac{1}{t_f} \left[\int_{]0,t_f^o]} \frac{\left|\mathcal{S}(t)\backslash\mathcal{S}^o(t_f^o)\right|}{|\mathcal{S}(t)|}\mathrm{d}t + \int_{[t_f^o,t_f[} \frac{\left|\mathcal{S}^o(t_f^o)\backslash\mathcal{S}(t)\right|}{\left|\mathcal{S}^o(t_f^o)\right|}\mathrm{d}t \right]. \tag{24}$$

Sørensen similarity index and Jaccard similarity coefficient compare the final simulated and observed areas. Kappa statistics computes the frequency with which simulation agrees with observation (P_a), with an adjustment (P_e) that takes into account agreement by chance. The arrival time agreement and the shape agreement both take into account the dynamics of the simulation. Even though there is only one observation at the end of the fire (i.e., the final burned surface), these scores were designed to partially evaluate the time evolution of the simulation dynamics. Further details on these scores may be found in Filippi et al. (2013).

3.4 Simulation set-up

The 4 models were run using the ForeFire code, with 80 cases it resulted in a total of 320 simulations. Each simulation is set-up using the ignition point that defines where to carry out the simulation. The simulation domain is centered on the ignition point. In

NHESSD
2, 3219–3249, 2014

north-south and east-west directions, the domain size is about four times the extension of the fire. The spatial increment δl depends on the fire size. If the final observed burned area is A, then $\delta l = \max(1, \log_{10} A - 4)$ m, if A is in m^2. The filtering distance d_f is set to $20\delta l$.

All simulations were carried out at most until the burned area equals the observed final burned area. In practice, on small fires, the area burned in the simulation may be larger because the stopping criterion is checked every 7 min (so the over-development cannot be more than 7 min of fire propagation). The simulation can also stop earlier if the front velocity is zero everywhere (stopped fire).

4 Results

In this section, we investigate the performance of the models on all fire simulations. All simulations are concisely described in appendix, with plotted contours (simulations and observation). The performance of each model for individual case is hardly analyzed since it can vary a lot, depending on the quality of the data. We essentially draw conclusions that are supported by the overall performance.

4.1 Distribution of the scores

An important aspect of the comparison is to select and understand the way scores are presented. At first, let consider the distribution of the 80 scores for each model and scoring method. The 80 scores are here sorted in ascending order and plotted. It is important to note that the sorting is carried out independently for each model; so the number in the abscissae corresponds to a rank in the list of sorted simulations (per model), but not to the same fire case in all 4 distinct evaluated models. The Fig. 1 shows the distributions for all scoring methods. The distributions for the classical scores, Sørensen similarity index, Jaccard similarity coefficient and kappa coefficient, are very similar with worst score distributions for the 3-percent model. The

NHESSD
2, 3219–3249, 2014

Rothermel models gives significantly better results. The Balbi model arguably brings additional improvements over Rothermel simulations. The non-stationary Balbi model clearly provides the best overall results for this data-set. The same ranking is found with the distributions of shape agreements (Fig. 1e). It is harder to discriminate the models with the distributions of the arrival time agreements (Fig. 1d). The 3-percent model still seems to be the worst model. The non-stationary Balbi model is still among the most reliable models, but its stationary version performs better at the high end of the distribution. We point out that all observed fires are assumed to last at most 24 h, which is a very rough approximation to the actual duration of the fire and, in these conditions, arrival time agreement is a less relevant scoring method than the other scores as it strongly penalizes over-prediction.

In Fig. 1, the scores are sorted in ascending order, independently for each model. It is therefore impossible to compare the performance for each individual fire and to make sure that a model consistently shows better performance. The score distributions are shown in Fig. 2 with an ascending sorting based on the mean score across the models. The sorting is therefore the same for all models and their scores remain paired in the plot. There is clearly a large variability in the performance of the models at each individual fire. There is however an overall ascending tendency and, more importantly, for most fires, the non-stationary Balbi model shows the best performance and the 3-percent model the worst. It means there is a consistent improvement for all types of fires and conditions. The conclusion is not so clear concerning the arrival time agreement, since it may lack relevance in this context. The large variance in the models performance suggests that the generation of large ensembles of simulations may be needed in the simulation process of a single fire. Indeed, with a single simulation for a single fire and weak data quality, a poor simulation is likely to occur.

Based on all the distributions, a conclusion is that the non-stationary Balbi model consistently gives the best results. The second model is then the stationary Balbi model. It is worth noting that the switch to a non-stationary model, where the fire depth is taken into account, leads to very significant improvements, at least as high as the

NHESSD
2, 3219–3249, 2014

changes due to the fire spread rate. This suggests that the representation of the fire in the numerical model is a key aspect of a simulation and making it more "physical" by removing assumptions to the model can enhance its versatility and performance. The Rothermel model is ranked third in this study but we should mitigate this rank because we used the original formulation of Rothermel, and the Balbi formulation has been developed in the context of Corsican fires. The study is more about the method to rank models. In order to conclude on the Rothermel rank, a renewed implementation of the Rothermel model would be required, as well as an extended database of fires. The 3-percent model is clearly the worst of the four. This suggests that more physics is needed to attain state-of-the-art models performance, even in a context where the input data may be of poor quality. Overall, it is interesting to note that, despite the poor data quality (which we might expect in operational forecasts), the models are clearly ranked, hence even a poor database can help objective model development.

4.2 A look at the individual simulations

For many simulations, the input data is really poor, hence a dramatically low performance. When the data is probably reliable, the models may perform reasonably well, which is detected by the scoring methods. For each scoring method and model, we found the simulation with the highest performance across the 80 fires. These selected cases are shown in Table 1.

The Sørensen, Jaccard and kappa scores identify the same "best" cases. It is consistent with the strong similarities found between the scores distributions for these scoring methods (Sect. 4.1). The dynamic-aware scores (arrival time agreement and shape agreement) select a variety of fire cases. It is more difficult to understand the reason for these selections from the final contours, since these scores take into account the full dynamics. Let us consider more closely two fires which appear several times in Table 1: the Oletta case (21 August 2004, N3 in appendix) and the Santo Pietro di Tenda (1 July 2003, T0 in appendix). In Fig. 3, we plotted the former case. In this case, it is clear that the Balbi model shows a much better performance than the other models.

NHESSD
2, 3219–3249, 2014

The overall shape of the final simulated contour by Balbi model has similar aspect ratio than the observed one, while the Rothermel model and the 3-percent model show large over-burning at the head of the fronts. So, even when the wind direction is correct, there can still be large differences between the models simulations. When we only consider the Balbi simulations for this fire case, we might conclude that a model can reach good prediction skill once proper input data is provided. Nevertheless the performance of the other models questions the reliability of the input data. It is likely that the input data might be erroneous and compensating for deficiencies in the Balbi model. It is fairly possible that the Balbi model would under-perform with the exact data. This suggests that exposing the actual prediction skill of fire simulation requires the use of several models and the exploration of the full probability distribution of the uncertain data.

In Fig. 4, we plotted the case for which the Balbi model attained its best shape agreement. The visual agreement with the observed contour does not seem as good as in the previous fire case. Hence the dynamics of the fire must have significantly contributed to the shape agreement. This suggests that taking into account the front dynamics in the evaluation can cast different light on a model.

4.3 Forecast reliability

In an operational context, the ability of a model to predict that a location will burn can be of high importance. One way to assess the reliability of a model in this context is to evaluate its burn probability. When the model predicts that a fire will burn a given location, we compute the frequency with which this actually happens. Conversely, we are interested in the frequency with which a model correctly predicts that a location will burn. We refer to this indicator as the detection skill. The detection skill may be perfect if the model burns very large regions: any location that is actually burned in reality will be burned in the simulation as well. On the contrary, the first indicator (burn probability) will be perfect if the fire stops right after it started: at the ignition point, the fire was observed, so the location was burned, just like the model "predicted". Consequently, we computed both the burn probability and the detection skill. See Table 2.

NHESSD
2, 3219–3249, 2014

The burn probability and detection skill confirm previous results. The non-stationary Balbi is still the best model, followed by the (stationary) Balbi model and the Rothermel model. When the non-stationary Balbi model predicts that a location will be burned, the real fire burned it in 29 % of all cases. When the real fire reaches a given location, the non-stationary Balbi model would have predicted it in 42 % of all cases. It is noteworthy that the performance varies a lot from one model to another. It is obvious that the 3-percent model with a detection skill of 11 % is of very low value compared to the 42 % of the non-stationary Balbi model.

Overall, the performance may not be good enough for a model to be reliable in operational applications. The performance spread among the models suggests that further development on the model has the potential to strongly improve the results. Nonetheless, it is clear from the simulation results that the input data plays a key role. In particular, an erroneous wind direction spoiled so many simulations that good local meteorological forecasts alone would be likely to dramatically improve performance.

5 Conclusions

The objective of this paper was to evaluate firespread models and their relevance in a realistic operational context with limited information. We considered a database of 80 fires whose final burned surfaces were observed. We simulated these fires in a purely automated manner, using only poor data-set that may be available in this operational context. The meteorological values were taken at the closest observation station, even though the actual wind direction at the exact fire location may be significantly different. The vegetation cover and the associated fuel load were not tuned by any means for any cases or models.

Despite the crude application setting, we were able to rank these firespread models. The non-stationary version of the Balbi model gave overall the best results. Its higher performance (compared to the other models) was consistently observed across most of the 80 cases. The stationary version of Balbi model showed a significantly lower

NHESSD
2, 3219–3249, 2014

performance. This suggests that the numerical treatment of the fire front is a key aspect, just like the rate of spread. The Rothermel model, (in its first original form and without tuning) was ranked third. Finally, we observed that the most empirical model, taking 3 percents of the wind velocity, is clearly not a good option. Overall, it is guessed that current firespread models may benefit from a better physical description, even with poor data quality.

We evaluated the skill of the models to forecast that a certain location will be burned. We also evaluated whether the locations burned in a simulation is likely to be burned in reality. In both cases, there is a wide spread in the models skills. In addition, the skill of the best model may not be high enough for a reliable use in operational context. Therefore we conclude that further work on the physical models is still needed and improved input data, especially for the wind direction, are obviously needed. For instance, largest gains in performance may be obtained from local micro-meteorological forecasts that properly forecast wind direction.

Considering the high uncertainties in both models and the input data, the simulation context is barely deterministic. There is a need for probabilistic approaches such as developed by Finney et al. (2011). With models results being so variable, a promising direction may be the use of ensembles of models, together with perturbed input data. Such developments could be used for uncertainty quantification, risk assessment and ensemble-based forecasting.

Appendix A

Simulations for all fires

This section shows the simulation results of the 4 models, for all considered fires. Information about the fires is collected in Table 3. The results are displayed in Figs. 5–7.

Acknowledgements. This research is supported by the Agence Nationale de la Recherche, project ANR-09-COSI-006, IDEA http://anridea.univ-corse.fr/.

NHESSD
2, 3219–3249, 2014

References

Alexander, M. E. and Cruz, M. G.: Are the applications of wildland fire behaviour models getting ahead of their evaluation again?, Environ. Modell. Softw., 41, 65–71, doi:10.1016/j.envsoft.2012.11.001, 2013. 3221

Anderson, H.: Heat transfer and fire spread, USDA Forest Service research paper INT, Intermountain Forest and Range Experiment Station, Forest Service, US Dept. of Agriculture, 1969. 3225

Anderson, H., Forest, I., and Range Experiment Station (Ogden, U.): Aids to determining fuel models for estimating fire behavior, General technical report INT, US Dept. of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, 1981. 3229

Andrews, P. L.: BEHAVE: fire behavior prediction and fuel modeling system – BURN subsystem, Tech. rep., USDA Forest Service, Intermountain Research Station, 1986. 3222

Balbi, J., Morandini, F., Silvani, X., Filippi, J., and Rinieri, F.: A physical model for wildland fires, Combust. Flame, 156, 2217–2230, doi:10.1016/j.combustflame.2009.07.010, 2009. 3224

Byram, G. and Fons, W.: Thermal conductivity of some common forest fuels, Tech. rep., USDA Forest Service, Division of Fire Research, 1952. 3222

Cruz, M. G. and Alexander, M. E.: Uncertainty associated with model predictions of surface and crown fire rates of spread, Environ. Modell. Softw., 47, 16–28, doi:10.1016/j.envsoft.2013.04.004, 2013. 3221

Filippi, J.-B., Morandini, F., Balbi, J. H., and Hill, D. R.: Discrete event front-tracking simulation of a physical fire-spread model, Simulation, 86, 629–646, doi:10.1177/0037549709343117, 2009. 3226

Filippi, J.-B., Mallet, V., and Nader, B.: Representation and evaluation of wildfire propagation simulations, Int. J. Wildland Fire, doi:10.1071/WF12202, 2013. 3220, 3229, 3230

Finney, M., Grenfell, I., McHugh, C., Seli, R., Trethewey, D., Stratton, R., and Brittain, S.: A method for ensemble wildland fire simulation, Environ. Model. Assess., 16, 153–167, doi:10.1007/s10666-010-9241-3, 2011. 3236

Forthofer, J.: Modeling Wind in Complex Terrain for Use in Fire Spread Prediction, Thesis, Colorado State University, 2007. 3228

Frandsen, W. H.: Fire spread through porous fuels from the conservation of energy, Combust. Flame, 16, 9–16, doi:10.1016/S0010-2180(71)80005-6, 1971. 3223

NHESSD
2, 3219–3249, 2014

Fujioka, F. M.: A new method for the analysis of fire spread modeling errors, Int. J. Wildland Fire, 11, 193–203, 2002. 3220
GDAL Development Team: GDAL – Geospatial Data Abstraction Library, Version 1.10.0, Open Source Geospatial Foundation, available at: www.gdal.org, 2013. 3228
McArthur, A.: Weather and Grassland Fire Behaviour, Leaflet, Forest Research Institute, Forestry and Timber Bureau, 1966. 3223
Rothermel, R.: A mahematical model for predicting fire spread in wildland fuels, Tech. rep. USDA Forest Service, 1972. 3222, 3223
Rothermel, R. and Anderson, H.: Fire spread characteristics determined in the laboratory, US Forest Service research paper INT, Intermountain Forest & Range Experiment Station, Forest Service, US Dept. of Agriculture, 1966. 3223
Santoni, P., Filippi, J. B., Balbi, J.-H., and Bosseur, F.: Wildland Fire Behaviour Case Studies and Fuel Models for Landscape-Scale Fire Modeling, J. Combust., 613424, doi:10.1155/2011/613424, 2011. 3229
Sullivan, A. L.: Wildland surface fire spread modelling, 1990–2007, 1: Physical and quasi-physical models, Int. J. Wildland Fire, 18, 349, doi:10.1071/WF06143, 2009a. 3221
Sullivan, A. L.: Wildland surface fire spread modelling, 1990–2007, 2: Empirical and quasi-empirical models, Int. J. Wildland Fire, 18, 369, doi:10.1071/WF06142, 2009b. 3221

NHESSD
2, 3219–3249, 2014

Table 1. The best simulations, according to the scoring method and the fire model. "ATA" stands for "arrival time agreement". "Shape" refers to "shape agreement". The final contours may be found in Figs. 5–7, and further information about the fires is available in Table 3.

Score	Non-stationary Balbi	Stationary Balbi
Sørensen	N3 – Oletta (21 Aug 2004)	N3 – Oletta (21 Aug 2004)
Jaccard	N3 – Oletta (21 Aug 2004)	N3 – Oletta (21 Aug 2004)
Kappa	N3 – Oletta (21 Aug 2004)	N3 – Oletta (21 Aug 2004)
ATA	N3 – Oletta (21 Aug 2004)	H0 – Corbara (1 Jul 2003)
Shape	B0 – Santo Pietro di Tenda (1 Jul 2003)	B0 – Santo Pietro di Tenda (1 Jul 2003)
Score	Original Rothermel	3-percent
Sørensen	E0 – Sisco (24 Jul 2003)	B1 – Calenzana (30 Jun 2005)
Jaccard	E0 – Sisco (24 Jul 2003)	B1 – Calenzana (30 Jun 2005)
Kappa	E0 – Sisco (24 Jul 2003)	B1 – Calenzana (30 Jun 2005)
ATA	E0 – Sisco (24 Jul 2003)	S1 – Ersa (17 Oct 2003)
Shape	J3 – Olmeta di Tuda (21 Aug 2004)	L0 – Ajaccio (22 Jul 2006)

NHESSD
2, 3219–3249, 2014

Table 2. Burn probability and detection skill for the 4 models, all fires included. The burn probability is the frequency with which the model correctly predicts that a location will be burned in reality. The detection skill is the frequency with which a burned location (in observations) is also burned in the simulation.

Indicator	Non-stationary Balbi	Stationary Balbi	Original Rothermel	3-percent
Burn probability	0.29	0.23	0.21	0.12
Detection skill	0.42	0.35	0.26	0.11

NHESSD
2, 3219–3249, 2014

Table 3. Information about all the 80 fires shown in Figs. 5–7. The size is the final burned area in hectares. WS stands for "wind speed", in $m\,s^{-1}$. WD is the wind direction in degrees, defined clockwise, with 0 corresponding a westerly wind (90 southerly). The resolution in meters is in column "Res". We finally put the best scores (among the 4 models) found for Jaccard similarity coefficient (BJ), arrival time agreement (BA) and shape agreement (BS).

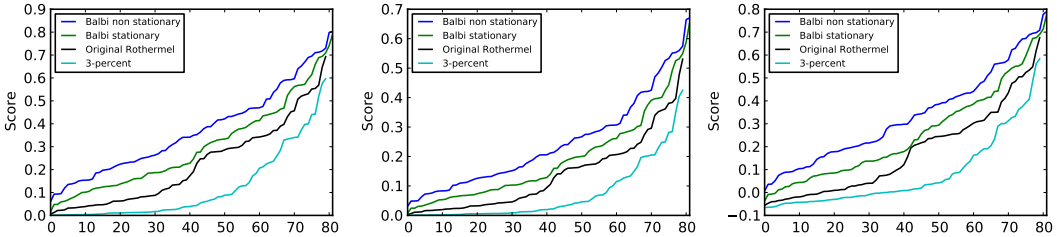
#	Name	Date	Size	WS	WD	Res.	BJ	BA	BS
A0	Aullene	7 Aug 2003	990 ha	9	60	133 m	0.15	0.7	0.69
A1	Oletta	9 Aug 2003	53 ha	3	275	27 m	0.29	0.86	0.65
A2	Olmi Cappella	29 Aug 2003	192 ha	15	210	60 m	0.12	0.63	0.46
A3	Pietracorbara	13 Aug 2004	1082 ha	19	260	139 m	0.47	0.8	0.63
B0	Santo Pietro di Tenda	1 Jul 2003	1310 ha	3	150	136 m	0.53	0.88	0.95
B1	Calenzana	30 Jun 2005	1418 ha	6	320	214 m	0.42	0.85	0.77
B2	Calenzana	19 Jul 2005	17 ha	5	220	13 m	0.26	0.75	0.65
B3	Solaro	13 Jan 2004	28 ha	2	129	13 m	0.29	0.71	0.6
C0	Volpajola	21 Jun 2003	49 ha	5	130	23 m	0.15	0.58	0.3
C1	Murzo	9 Aug 2003	180 ha	5	210	52 m	0.031	0.76	0.28
C2	Ghisonaccia	2 Aug 2003	95 ha	7	40	45 m	0.34	0.71	0.82
C3	Vivario	9 Jul 2007	56 ha	4	140	30 m	0.13	0.62	0.41
D0	Corscia	1 Jan 2009	57 ha	1	0	23 m	0.42	0.78	0.7
D1	Biguglia	2 Aug 2003	782 ha	7	20	103 m	0.42	0.71	0.89
D2	Casaglione	26 May 2008	2 ha	5	110	5 m	0.084	0.61	0.22
D3	Aleria	5 Sep 2004	62 ha	5	30	24 m	0.29	0.71	0.57
E0	Sisco	24 Jul 2003	441 ha	1	276	79 m	0.53	0.99	0.79
E1	Linguizzetta	9 Aug 2003	33 ha	4	140	16 m	0.38	0.74	0.75
E2	Aleria	20 Jun 2004	80 ha	4	90	28 m	0.17	0.79	0.4
E3	Prunelli di Fiumorbo	4 Jul 2003	26 ha	5	270	18 m	0.53	0.8	0.84
F0	Olmeta di Tuda	21 Aug 2004	80 ha	7	240	26 m	0.4	0.77	0.84
F1	Calenzana	19 Jul 2005	75 ha	8	250	38 m	0.28	0.66	0.82
F2	Nocario	11 Apr 2006	81 ha	4	240	26 m	0.15	0.69	0.38
F3	Ventiseri	21 Aug 2004	28 ha	7	260	19 m	0.19	0.65	0.48
G0	Porto Vecchio	21 Aug 2006	3 ha	3	210	5 m	0.077	0.54	0.39
G1	Afa	22 Jul 2003	242 ha	5	220	61 m	0.16	0.7	0.66
G2	Loreto di Casinca	20 Apr 2005	41 ha	7	260	22 m	0.21	0.67	0.61
G3	Calenzana	17 Oct 2008	59 ha	7	210	25 m	0.1	0.55	0.23
H0	Corbara	1 Jul 2003	10 ha	8	210	12 m	0.28	0.81	0.88
H1	Vero	15 Jul 2003	535 ha	3	210	96 m	0.11	0.6	0.64
H2	Canale di Verde	4 Jan 2003	34 ha	3	240	14 m	0.25	0.73	0.77
H3	Altiani	12 Aug 2003	23 ha	4	130	16 m	0.21	0.66	0.56
I0	Patrimonio	11 Feb 2005	12 ha	5	120	13 m	0.21	0.74	0.41
I1	Pruno	20 Apr 2005	116 ha	7	260	37 m	0.25	0.67	0.65
I2	Olcani	1 Jan 2006	69 ha	12	260	37 m	0.13	0.64	0.29
I3	Sartene	14 Jul 2008	0.25 ha	10	270	1 m	0.13	0.62	0.35
J0	Calenzana	8 Aug 2003	16 ha	2	275	12 m	0.17	0.6	0.37
J1	Propriano	24 Aug 2008	2.8 ha	4	210	5 m	0.5	0.76	0.71
J2	Canari	1 Jul 2003	94 ha	7	239	57 m	0.31	0.66	0.79
J3	Olmeta di Tuda	21 Aug 2004	18 ha	6	320	12 m	0.66	0.83	0.89
K0	Calenzana	31 Jul 2005	91 ha	8	240	50 m	0.13	0.78	0.61

NHESSD
2, 3219–3249, 2014

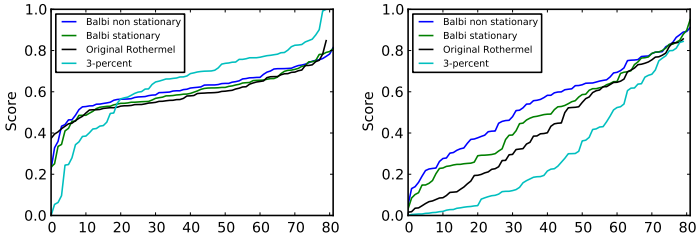
Table 3. Continued.

#	Name	Date	Size	WS	WD	Res.	BJ	BA	BS
K1	Calenzana	15 Jul 2004	102 ha	4	260	32 m	0.049	0.78	0.33
K2	Calenzana	30 Aug 2006	143 ha	7	230	26 m	0.17	0.65	0.49
K3	Piana	22 Aug 2008	2.2 ha	3	280	6 m	0.086	0.59	0.052
L0	Ajaccio	22 Jul 2006	13 ha	4	190	15 m	0.55	0.9	0.84
L1	Bastelicaccia	13 Jul 2008	0.3 ha	4	300	2 m	0.27	0.69	0.46
L2	Propriano	19 Jul 2006	0.5 ha	4	210	1 m	0.21	0.56	0.39
L3	Oletta	2 Jul 2003	1126 ha	7	280	168 m	0.41	0.77	0.8
M0	Soveria	29 Aug 2003	56 ha	5	350	26 m	0.17	0.56	0.41
M1	Sisco	14 Aug 2003	382 ha	1	258	42 m	0.24	0.62	0.31
M2	Coti Chiavari	6 May 2003	185 ha	7	110	46 m	0.31	0.83	0.65
M3	Lumio	12 Aug 2003	201 ha	6	240	51 m	0.34	0.74	0.54
N0	Santa Maria Poggio	3 Aug 2003	58 ha	1	321	29 m	0.55	0.77	0.88
N1	Barbaggio	1 Jul 2003	517 ha	7	239	80 m	0.56	0.78	0.84
N2	Rutali	14 Jan 2004	59 ha	2	72	31 m	0.23	0.65	0.83
N3	Oletta	21 Aug 2004	186 ha	13	260	44 m	0.67	0.81	0.9
O0	Propriano	27 Aug 2008	0.25 ha	3	220	2 m	0.4	0.83	0.54
O1	Calenzana	25 Jul 2004	465 ha	6	200	72 m	0.48	0.77	0.75
O2	Calvi	8 Jul 2009	78 ha	8	210	36 m	0.44	0.67	0.81
O3	San Giovanni di Moriani	4 Aug 2003	10 ha	5	340	10 m	0.57	0.78	0.84
P0	Luri	13 Jan 2004	50 ha	2	270	26 m	0.36	0.71	0.66
P1	Saint Florent	5 Aug 2003	25 ha	3	284	17 m	0.21	0.84	0.43
P2	Peri	23 Jul 2009	749 ha	2	280	152 m	0.066	0.76	0.59
P3	Poggio d'Oletta	11 Sep 2004	53 ha	4	140	26 m	0.38	0.82	0.51
Q0	Borgo	20 Jul 2003	40 ha	1	20	17 m	0.24	0.67	0.56
Q1	Sisco	14 Aug 2003	357 ha	4	238	43 m	0.074	0.79	0.23
Q2	Calenzana	6 Sep 2004	27 ha	10	20	19 m	0.05	0.59	0.16
Q3	Bonifacio	28 Jul 2003	477 ha	8	280	104 m	0.11	0.6	0.58
R0	Propriano	19 Jul 2006	3.5 ha	4	210	7 m	0.14	0.56	0.28
R1	Tolla	9 Aug 2003	942 ha	2	325	146 m	0.4	0.73	0.73
R2	Santa Lucia di Mercurio	6 Sep 2004	19 ha	7	0	20 m	0.31	0.68	0.67
R3	Alata	21 Jul 2008	0.12 ha	6	230	1 m	0.1	0.7	0.13
S0	Omessa	31 Aug 2003	115 ha	4	140	33 m	0.42	0.57	0.63
S1	Ersa	17 Oct 2003	156 ha	3	90	69 m	0.24	1	0.79
S2	Manso	8 Jul 2004	11 ha	7	230	8 m	0.083	0.62	0.32
S3	Olmeta di Tuda	4 Aug 2003	13 ha	1	260	9 m	0.16	0.69	0.52
T0	Santo Pietro di Tenda	1 Jul 2003	110 ha	7	239	37 m	0.55	0.77	0.83
T1	Calenzana	6 Sep 2004	28 ha	10	40	20 m	0.14	0.57	0.37
T2	Bisinchi	2 Aug 2005	13 ha	5	40	12 m	0.28	0.7	0.56
T3	Montegrosso	1 Jul 2003	43 ha	5	265	21 m	0.27	0.76	0.7

NHESSD
2, 3219–3249, 2014



(a) Sørensen similarity index (b) Jaccard similarity (c) Kappa



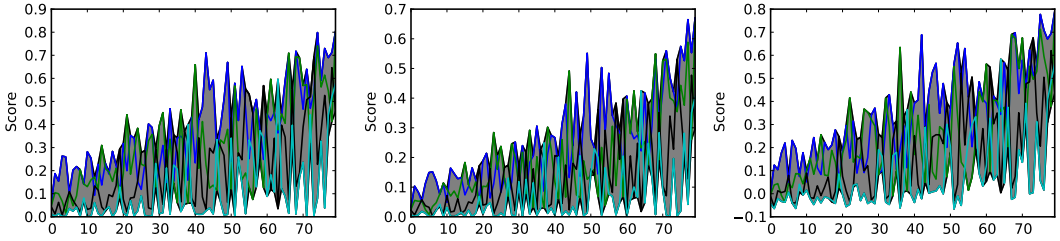
(d) Arrival time agreement (e) Shape agreement

Fig. 1. The distribution of the scores for the 80 cases and for the 4 models. The sorting was carried out independently for each model, hence one cannot compare individuals scores.

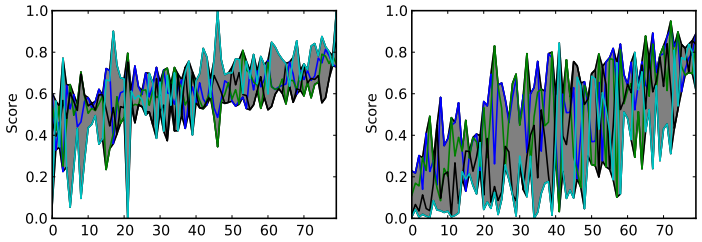
NHESSD
2, 3219–3249, 2014

Evaluation of forest fire models on a large observation database

J.-B. Filippi et al.



(a) Sørensen similarity index (b) Jaccard similarity (c) Kappa



(d) Arrival time agreement (e) Shape agreement

Fig. 2. The distribution of the scores for the 80 cases and for the 4 models. The sorting was carried out with respect to the mean score across the models, hence at a given abscissa, one can read the four scores for a given fire case. The gray area is delimited by the upper and lower bounds of the lines.

NHESSD
2, 3219–3249, 2014

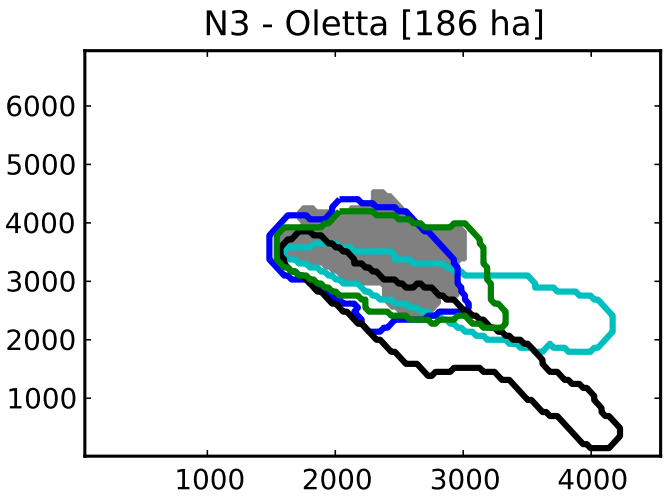


Fig. 3. Case of Oletta fire (31 August 2004, N3 in appendix) as simulated by the non-stationary Balbi model (blue), the stationary Balbi model (green), the Rothermel model (black) and the 3-percent model (cyan).

NHESSD
2, 3219–3249, 2014

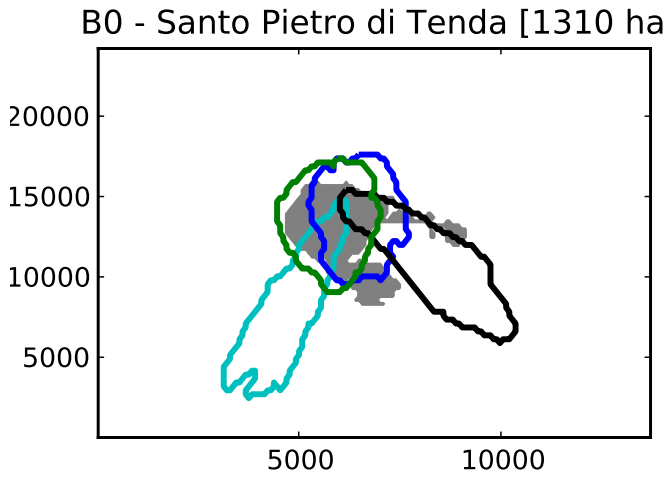


Fig. 4. Case of Santo Pietro di Tenda fire (1 July 2003, T0 in appendix) as simulated by the non-stationary Balbi model (blue), the stationary Balbi model (green), the Rothermel model (black) and the 3-percent model (cyan).

NHESSD
2, 3219–3249, 2014

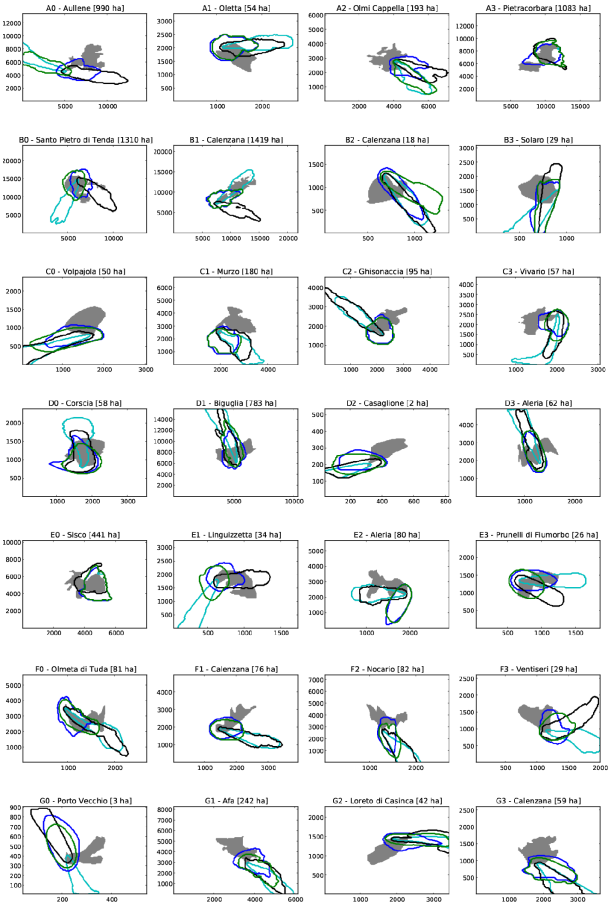


Fig. 5. Simulation for all fires (1/3). The non-stationary Balbi model is in blue; the stationary Balbi model is in green; the Rothermel model is in black; the 3-percent model is in cyan. The gray area is the observed final burned area. The coordinates are meters.

NHESSD
2, 3219–3249, 2014

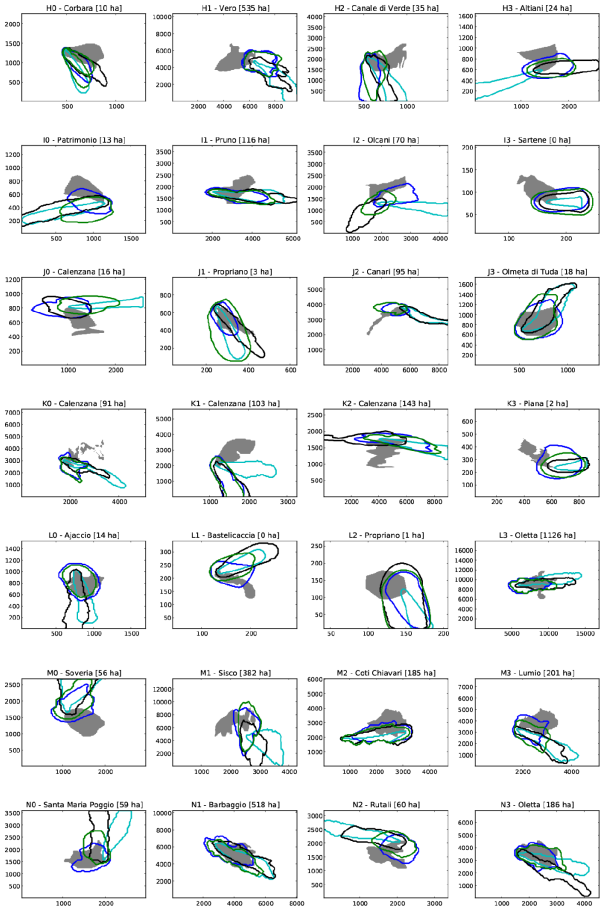


Fig. 6. Simulation for all fires (2/3). The non-stationary Balbi model is in blue; the stationary Balbi model is in green; the Rothermel model is in black; the 3-percent model is in cyan. The gray area is the observed final burned area. The coordinates are meters.

NHESSD
2, 3219–3249, 2014

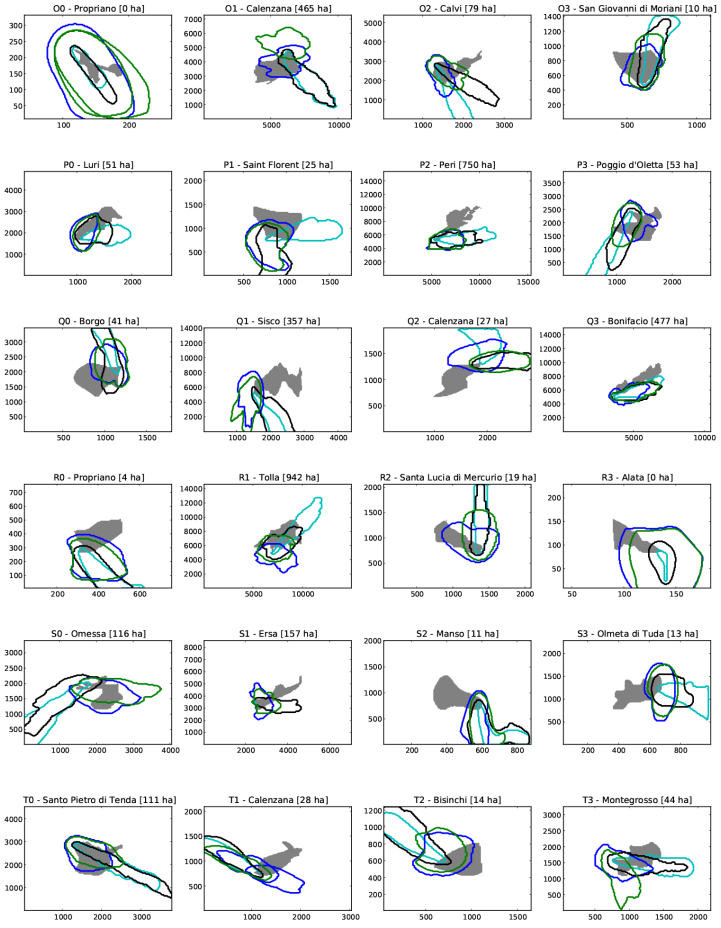


Fig. 7. Simulation for all fires (3/3). The non-stationary Balbi model is in blue; the stationary Balbi model is in green; the Rothermel model is in black; the 3-percent model is in cyan. The gray area is the observed final burned area. The coordinates are meters.

NHESSD
2, 3219–3249, 2014