Model evaluation based on a large observation data set

Jean-Baptiste Filippi[1] Vivien Mallet[2,3] Bahaa Nader[1]

[1]SPE CNRS

[2]INRIA

[3]CEREA, joint laboratory ENPC - EDF R&D, Université Paris-Est

Numerical simulation of forest fires, Cargèse, May 2013

Topic

Evaluation of the performance of a propagation model

Questions

1. How to rank models? Can we identify the "best" model out of a pool of models?
2. How to evaluate the dynamics of the model when the observation is the final burned surface?
3. Can we evaluate a model regardless of the quality of its inputs?
4. Can we carry out probabilistic forecasts?

Notation

Observation

- Observation of final burned surface at time t_f^o
- Observed burned surface: $\mathcal{S}^o(t_f^o)$

Simulation

- Final simulation time: t_f^s
- Simulated burned surface at time t: $\mathcal{S}(t)$

Area

- $|\mathcal{S}|$ is the area of the surface $\mathcal{S}$
- Ω is the simulation domain

Scores

- Classical scores compare $\mathcal{S}^o(t_f^o)$ and $\mathcal{S}(t_f^s)$

Classical scores

Sørensen similarity index

$$\frac{2|\mathcal{S}^o(t_f^o) \cap \mathcal{S}(t_f^s)|}{|\mathcal{S}^o(t_f^o)| + |\mathcal{S}(t_f^s)|} \in [0, 1]$$

Jaccard similarity coefficient

$$\frac{|\mathcal{S}^o(t_f^o) \cap \mathcal{S}(t_f^s)|}{|\mathcal{S}^o(t_f^o) \cup \mathcal{S}(t_f^s)|} \in [0, 1]$$

Kappa coefficient

$$\frac{P_a - P_e}{1 - P_e} \leq 1$$

where

$$P_a = \frac{|\mathcal{S}^o(t_f^o) \cap \mathcal{S}(t_f^s)|}{|\Omega|} + \frac{|\Omega \backslash (\mathcal{S}^o(t_f^o) \cup \mathcal{S}(t_f^s))|}{|\Omega|}$$

$$P_e = \frac{|\mathcal{S}^o(t_f^o)||\mathcal{S}(t_f^s)|}{|\Omega|^2} + \frac{|\Omega \backslash \mathcal{S}^o(t_f^o)||\Omega \backslash \mathcal{S}(t_f^s)|}{|\Omega|^2}$$

Addition notation

- Arrival time: earliest time at which the front is known to have reached some point; $+\infty$ if the front never reaches the point
- Observed arrival time at X: $\mathcal{T}^o(X)$, say $\mathcal{T}^o(X) = t_f^o$ if $X \in \mathcal{S}^o(t_f^o)$
- Simulated arrival time at X: $\mathcal{T}(X)$, here, $\mathcal{T}(X) \leq t_f^s$ if $X \in \mathcal{S}(t_f^s)$

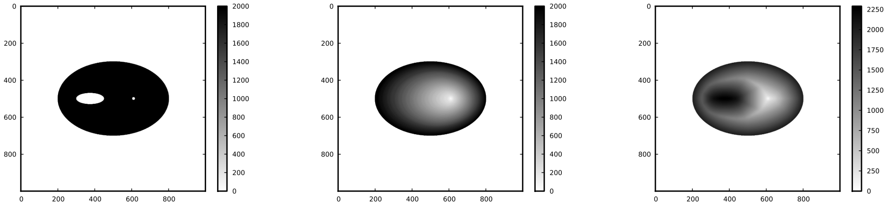
Arrival time agreement

$$1 - \frac{1}{|\mathcal{S}(t_f^s) \cup \mathcal{S}^o(t_f^o)| \max(t_f^s, t_f^o)} \left[\int_{\mathcal{S}(t_f^s) \cap \mathcal{S}^o(t_f^o)} \max(\mathcal{T}(X) - \mathcal{T}^o(X), 0) \mathrm{d}X \right.$$
$$+ \int_{\mathcal{S}(t_f^s) \backslash \mathcal{S}^o(t_f^o)} \max(t_f^o - \mathcal{T}(X), 0) \mathrm{d}X$$
$$\left. + \int_{\mathcal{S}^o(t_f^o) \backslash \mathcal{S}(t_f^s)} (t_f^s - \mathcal{T}^o(X)) \mathrm{d}X \right] \in [0, 1]$$

Shape agreement

$$1 - \frac{1}{t_f^s}\left[\int_{]0,t_f^o]} \frac{|\mathcal{S}(t)\setminus\mathcal{S}^o(t_f^o)|}{|\mathcal{S}(t)|}\mathrm{d}t + \int_{[t_f^o,t_f^s[} \frac{|\mathcal{S}^o(t_f^o)\setminus\mathcal{S}(t)|}{|\mathcal{S}^o(t_f^o)|}\mathrm{d}t\right] \in [0, 1]$$

Application of the scores to an idealized case



S = 0.981	S = 0.981
J = 0.962	J = 0.962
K = 0.977	K = 0.977
ATA = 0.985	ATA = 0.998
SA = 0.973	SA = 0.997

J.-B. Filippi, V. Mallet, and B. Nader (2013). “Representation and evaluation of wildfire propagation simulations”. In: *Under review for Int. J. Wild. Fire*

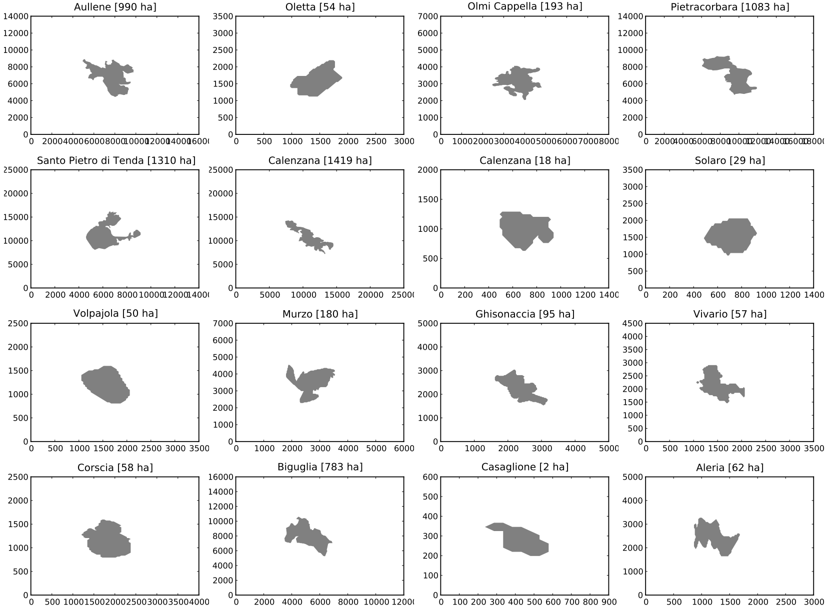
Scoring methods in Python at `http://sf.net/projects/pyfirescore/`

Simulation of 80 fire cases

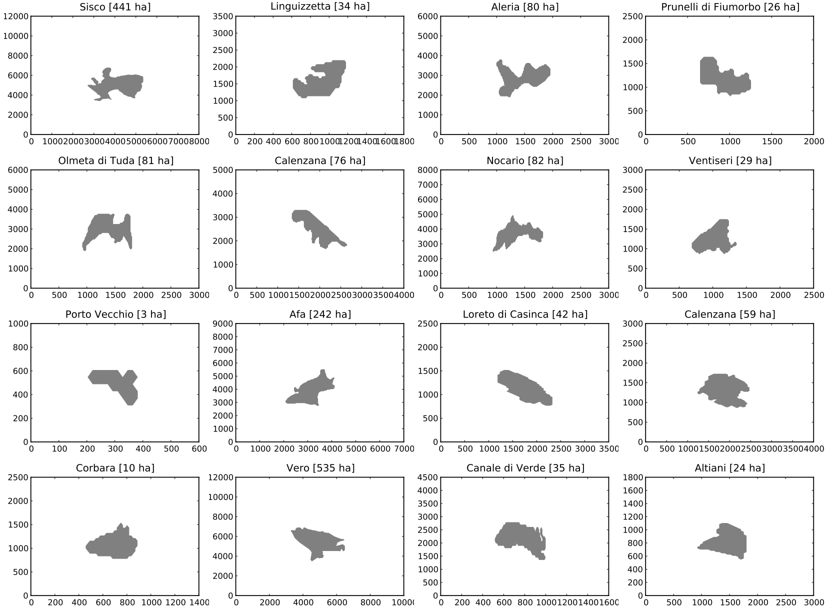
Fire cases

- Using observations from Prométhée, database of french fires
- Available data for a subset of the fires: date, ignition point, final contour
- Unfortunately, no data on firefights
- Considering 80 Corsican fires from 2003 to 2008

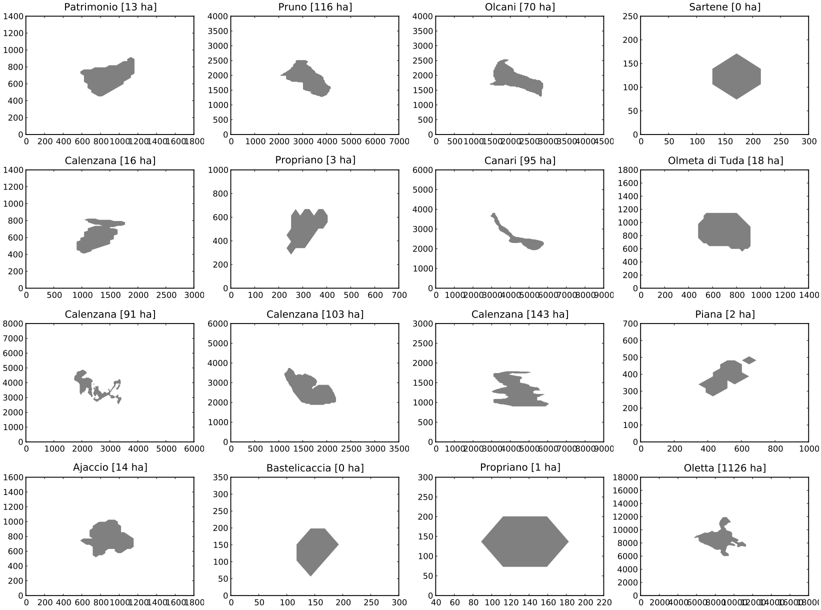
80 fire cases [1/5]



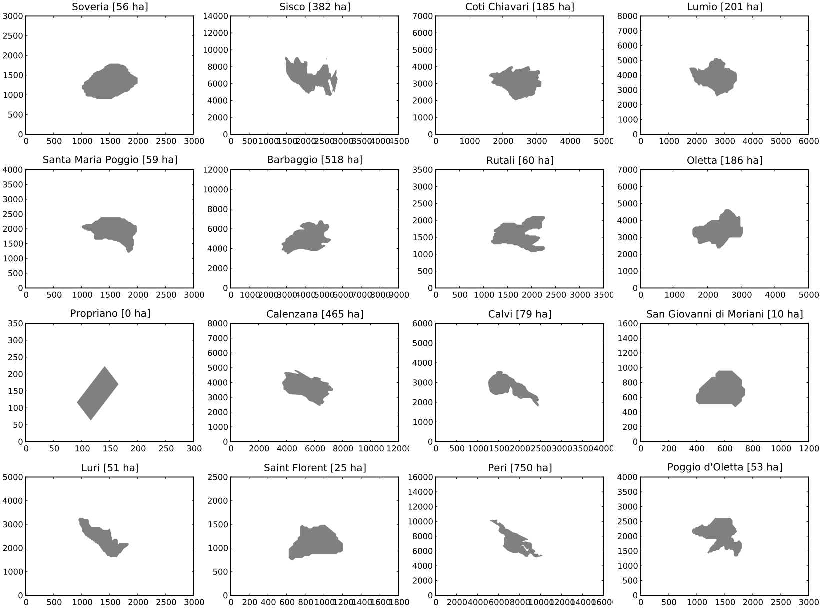
80 fire cases [2/5]



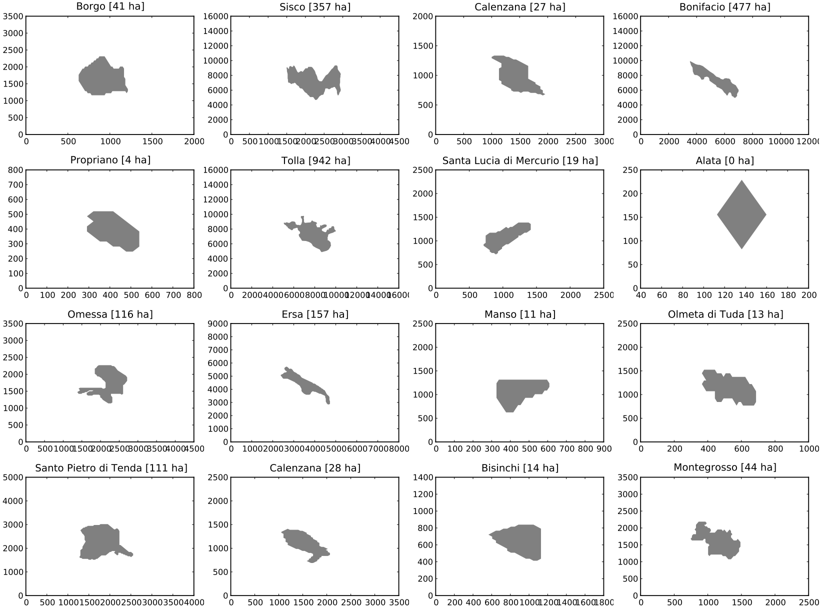
80 fire cases [3/5]



80 fire cases [4/5]



80 fire cases [5/5]

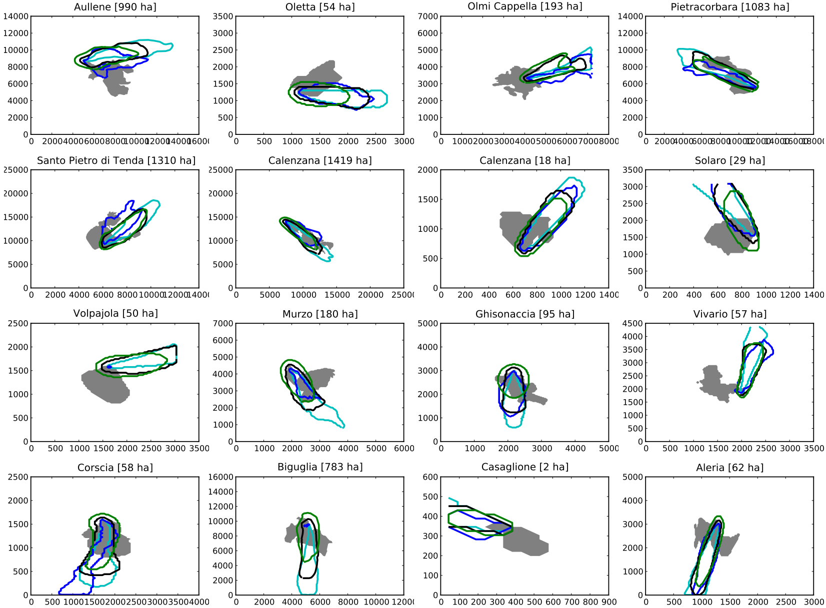


Simulation of 80 fire cases

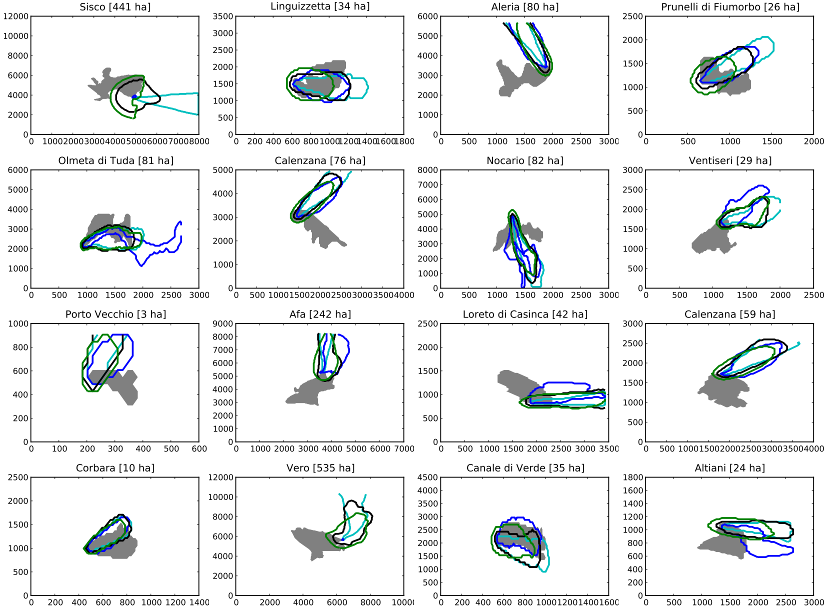
Simulations

- Land use cover: "inventaire forestier national" (IFN, from IGN) and global land cover (GLC)
- Elevation at 25 m resolution from IGN
- Wind velocity and direction
  - Taken at Ajaccio or Bastia meteorological station, or from ECMWF simulations
  - And then computed by Windninja (US Forest Service and Colorado State University)
- Running ForeFire (SPE), until the final burned area is attained
- Four models for the rate of spread
  - Balbi model (cf. his talk, tomorrow), "stationary"
  - Balbi model, "non-stationary", i.e., with front depth and dependence on the front geometry
  - Rothermel model
  - Simple rate of spread equal to 3% of wind velocity

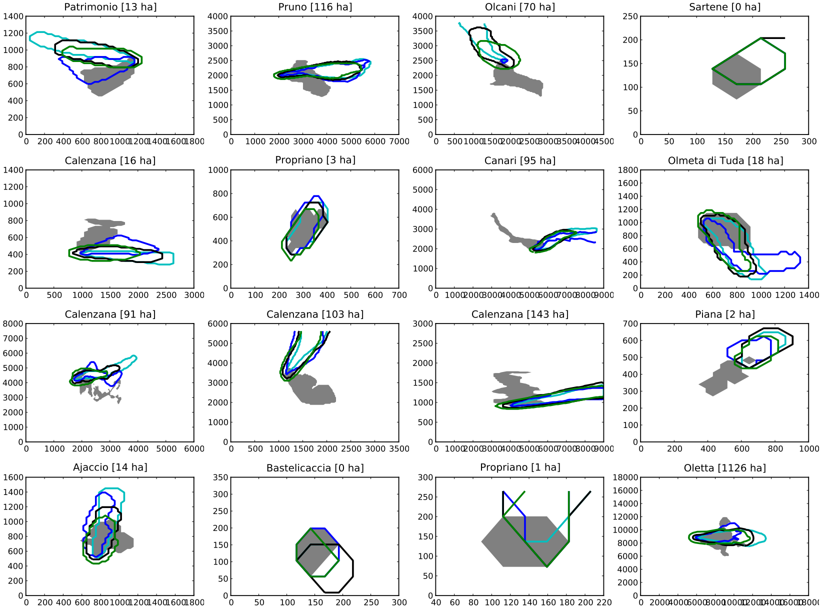
80 fire cases [1/5]



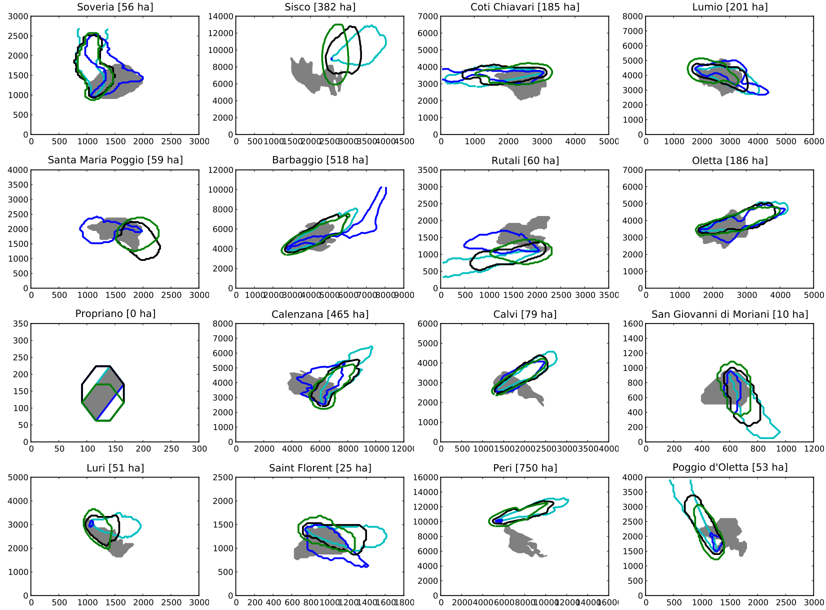
80 fire cases [2/5]



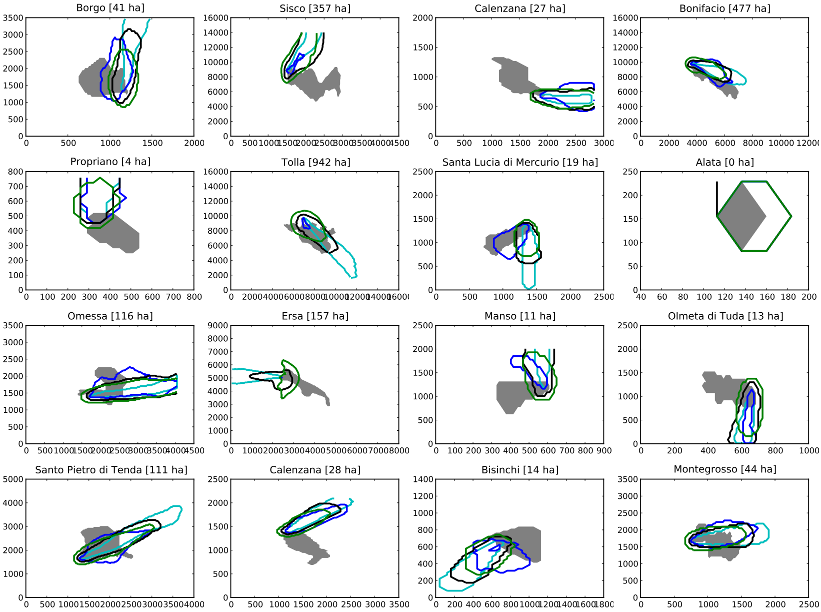
80 fire cases [3/5]



80 fire cases [4/5]

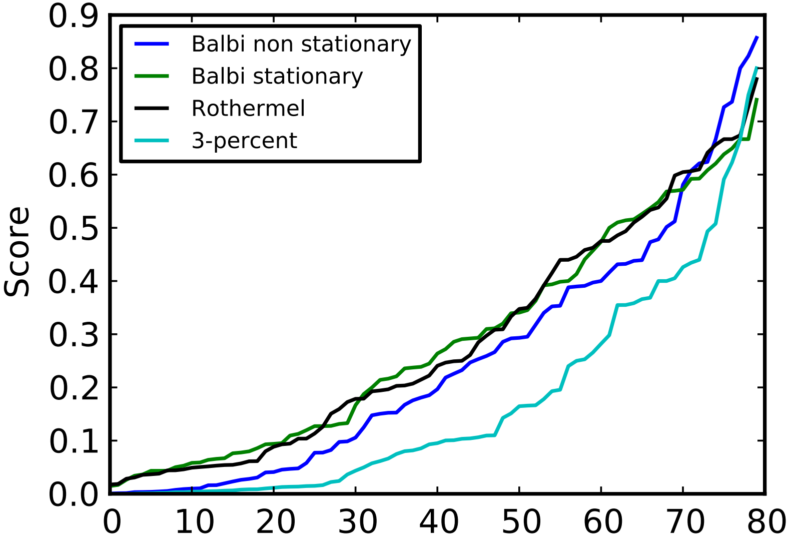


80 fire cases [5/5]



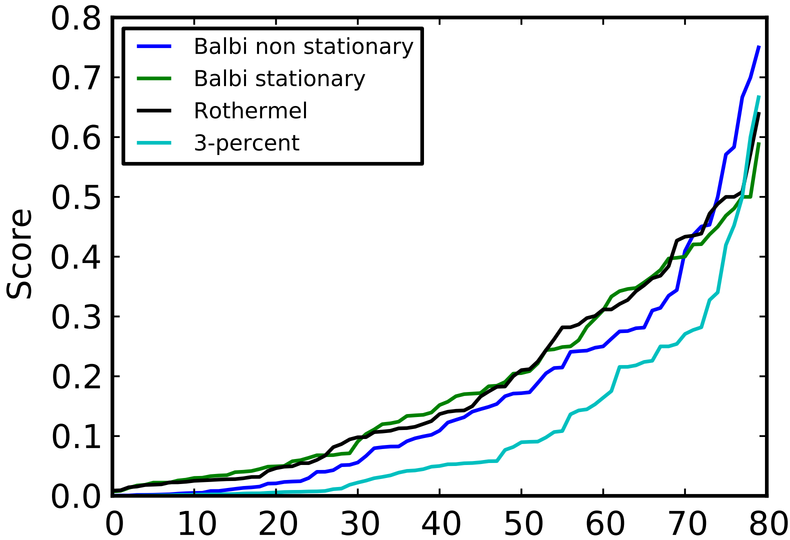
Distribution of the Sørensen similarity indexes

Scores sorted independently for each model



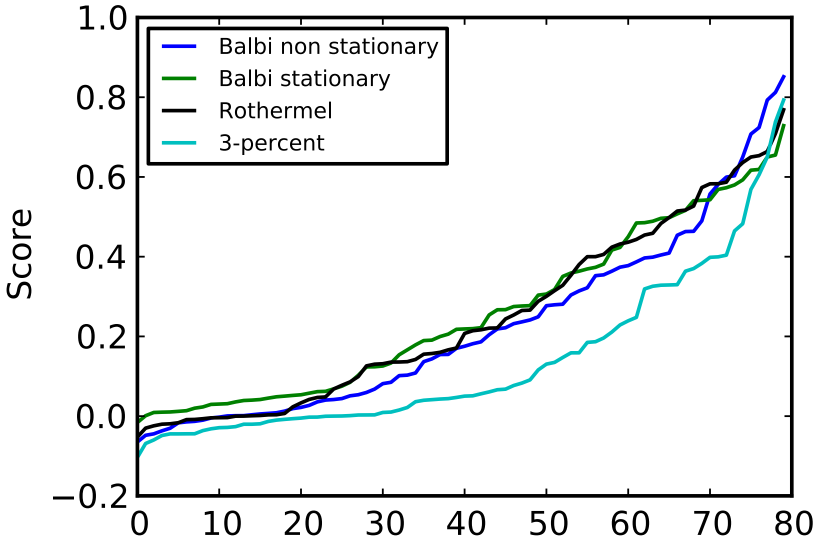
Distribution of the Jaccard similarity coefficients

Scores sorted independently for each model



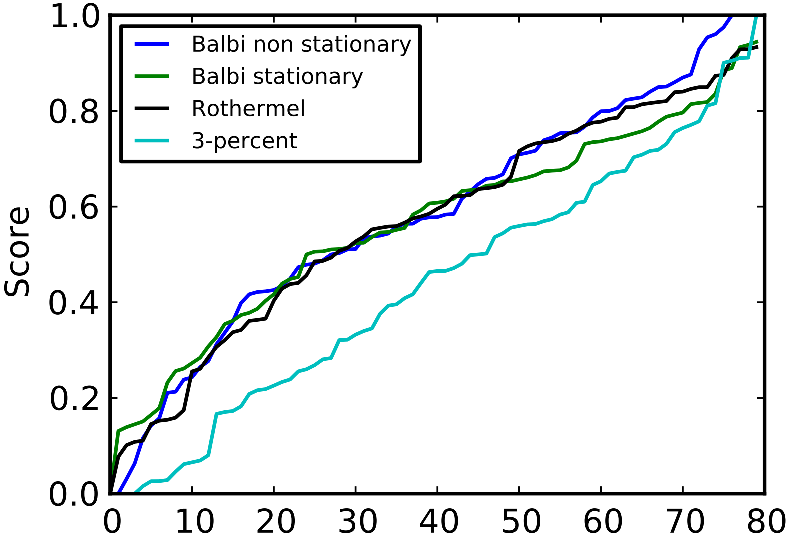
Distribution of the Kappa coefficients

Scores sorted independently for each model



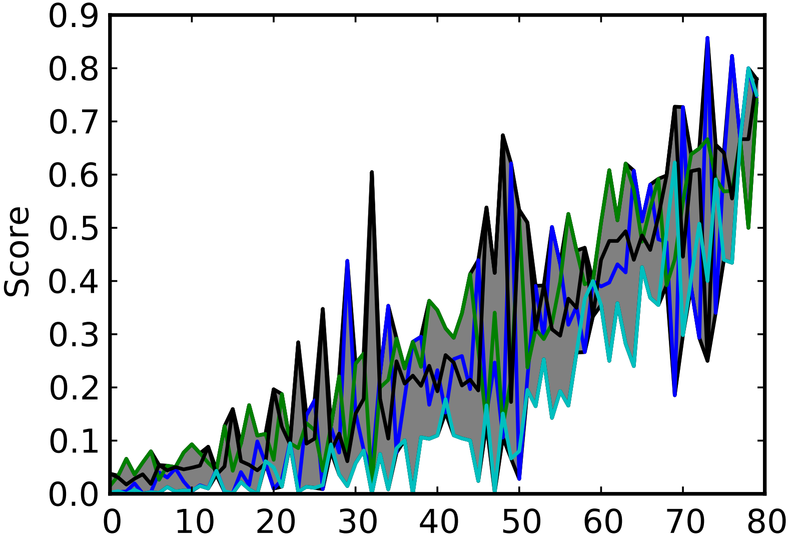
Distribution of the shape agreements

Scores sorted independently for each model



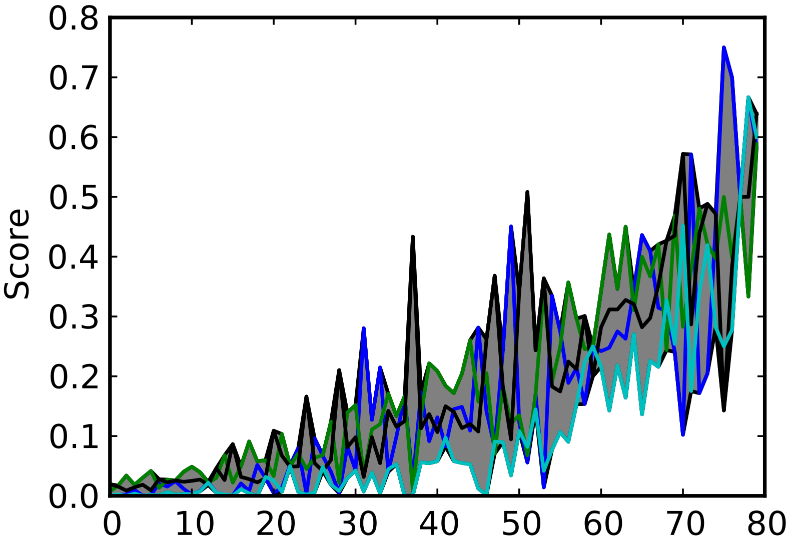
Distribution of the Sørensen similarity indexes

Scores sorted according to the average score



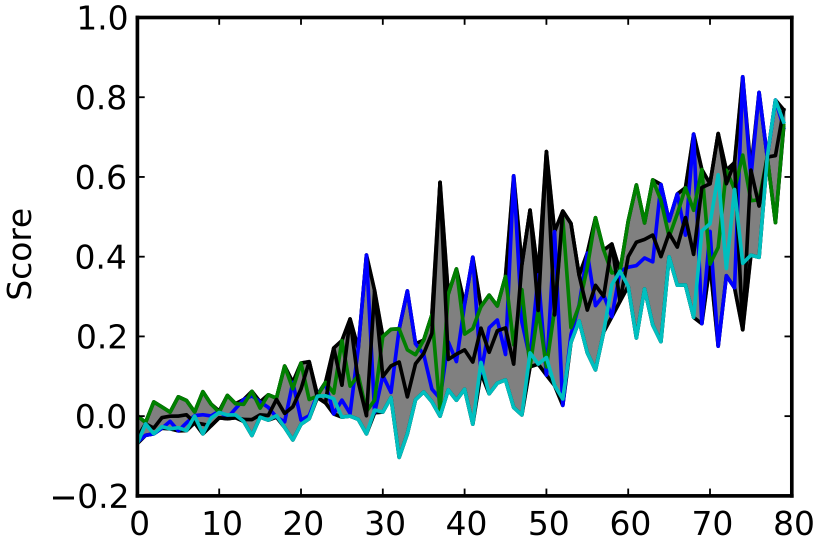
Distribution of the Jaccard similarity coefficients

Scores sorted according to the average score



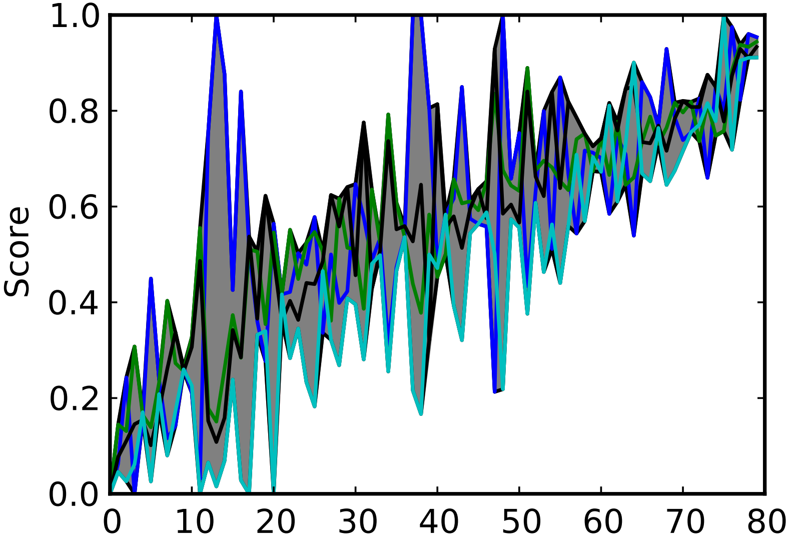
Distribution of the Kappa coefficients

Scores sorted according to the average score



Distribution of the shape agreements

Scores sorted according to the average score



Monte Carlo simulations

Pertubed parameters

- Wind direction: additive perturbation, clipped normal distribution, with zero mean and standard deviation of 40°
- Wind velocity: log-normal distribution, assuming $[\frac{1}{2}v, 2v]$ is a 0.95-confidence interval
- Final burned area: log-normal distribution, assuming $[\frac{2}{3}v, \frac{3}{2}v]$ is a 0.95-confidence interval
- Fuel load: log-normal distribution, assuming $[\frac{1}{1.1}v, 1.1v]$ is a 0.95-confidence interval
- Moisture content: log-normal distribution, assuming $[\frac{1}{1.1}v, 1.1v]$ is a 0.95-confidence interval

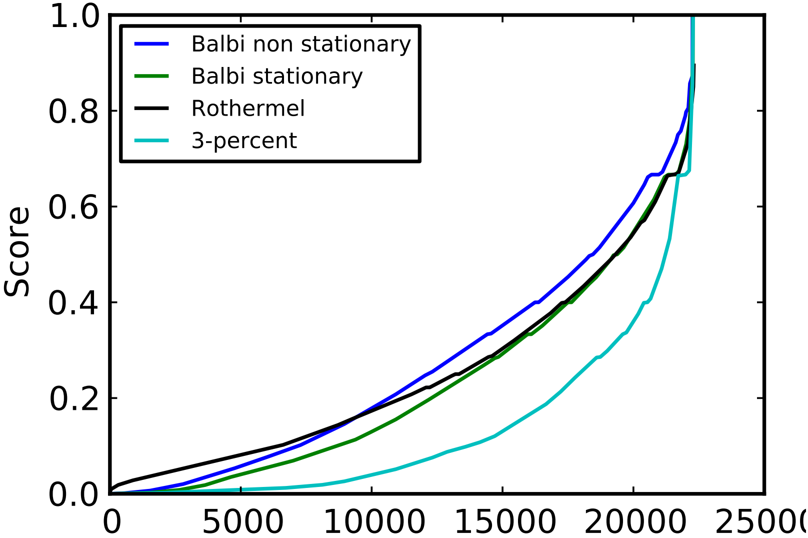
Monte Carlo simulations

Monte Carlo experiment

- Selecting the four models with equal probability
- Running on 75 cases (out of the 80 cases), with at least 1150 simulations per case
- About 89,000 simulations in total

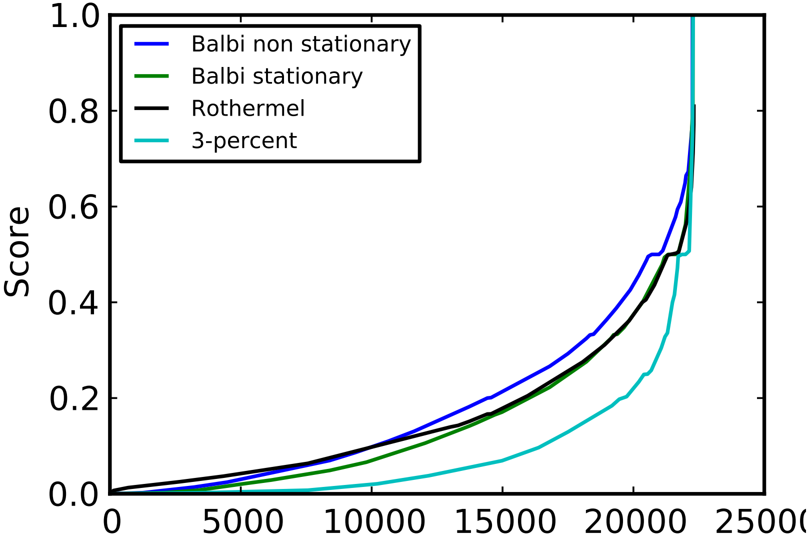
Distribution of the Sørensen similarity indexes

Scores sorted independently for each model



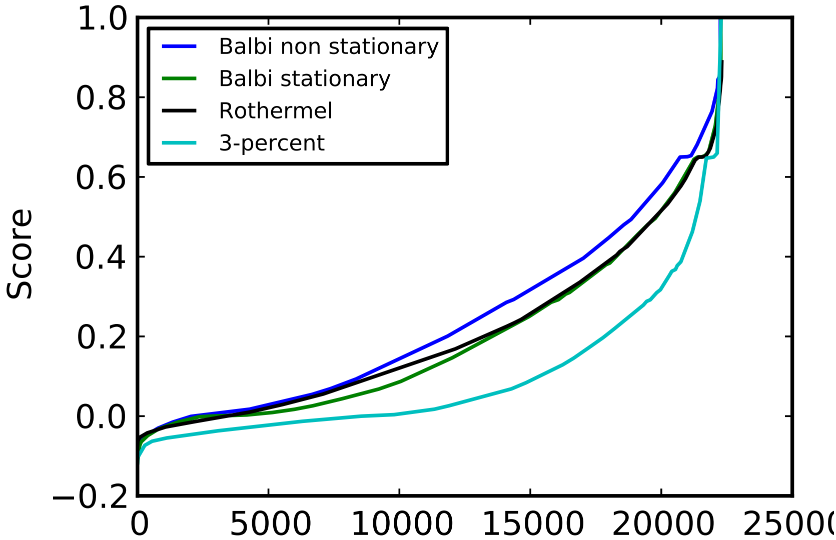
Distribution of the Jaccard similarity coefficients

Scores sorted independently for each model



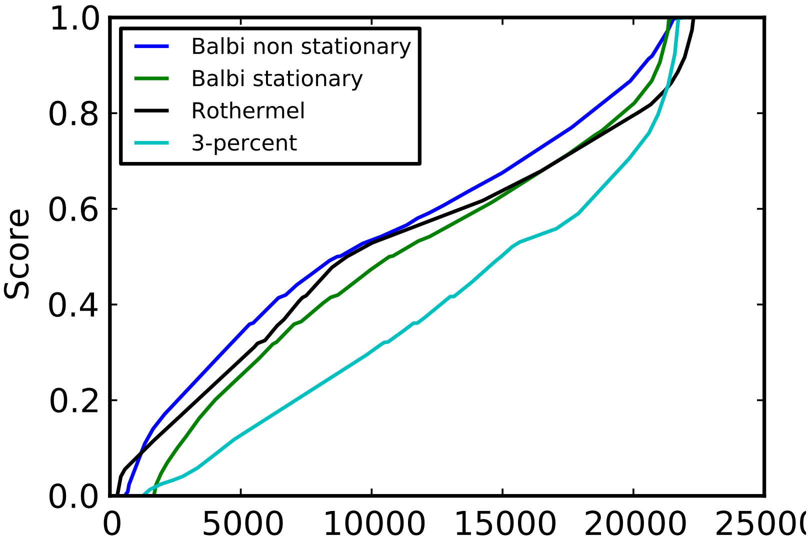
Distribution of the Kappa coefficients

Scores sorted independently for each model



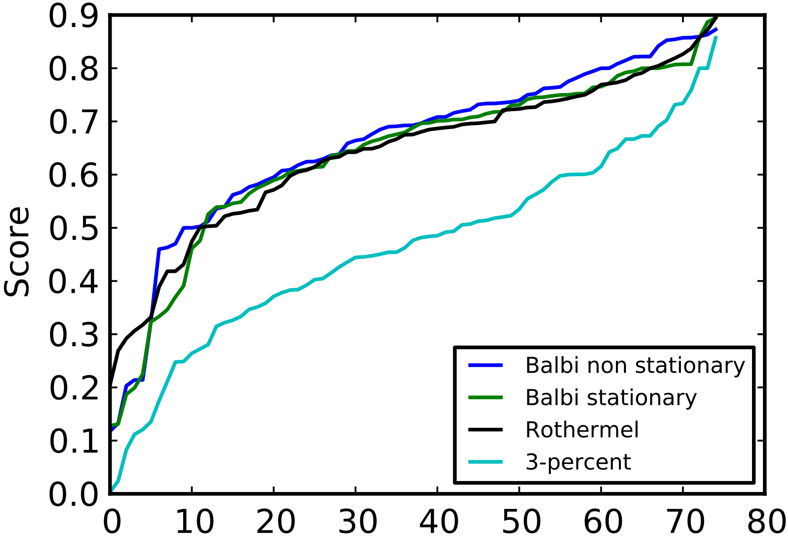
Distribution of the shape agreements

Scores sorted independently for each model



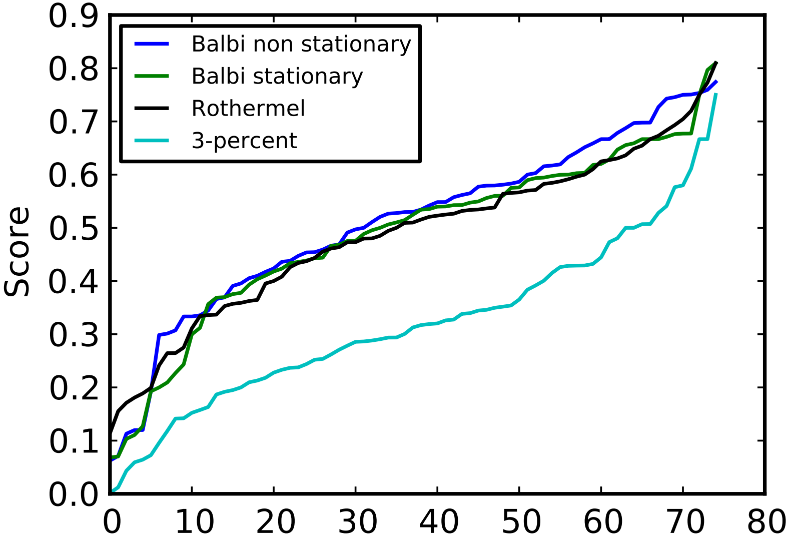
Distribution of the Sørensen similarity indexes

Scores of the best simulations, sorted independently for each model



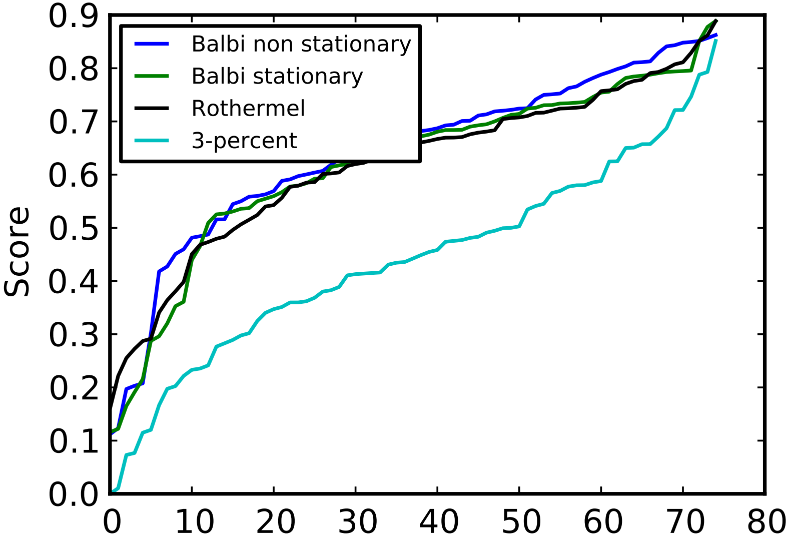
Distribution of the Jaccard similarity coefficients

Scores of the best simulations, sorted independently for each model



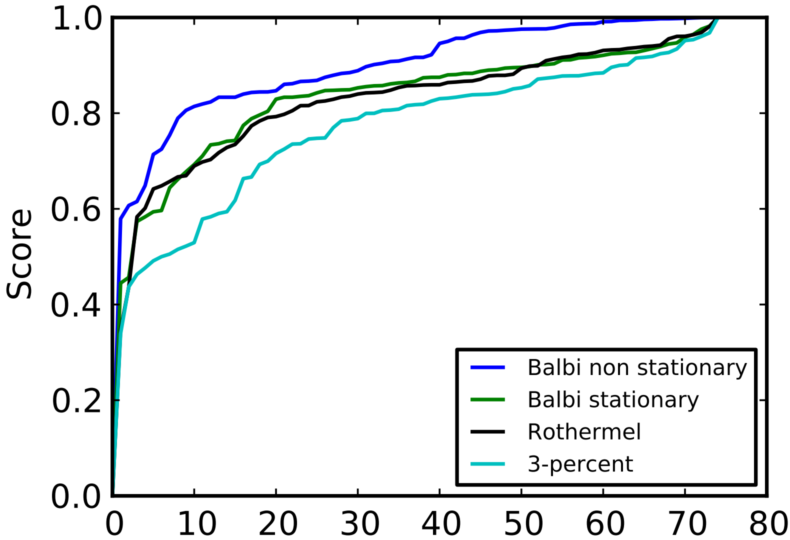
Distribution of the Kappa coefficients

Scores of the best simulations, sorted independently for each model



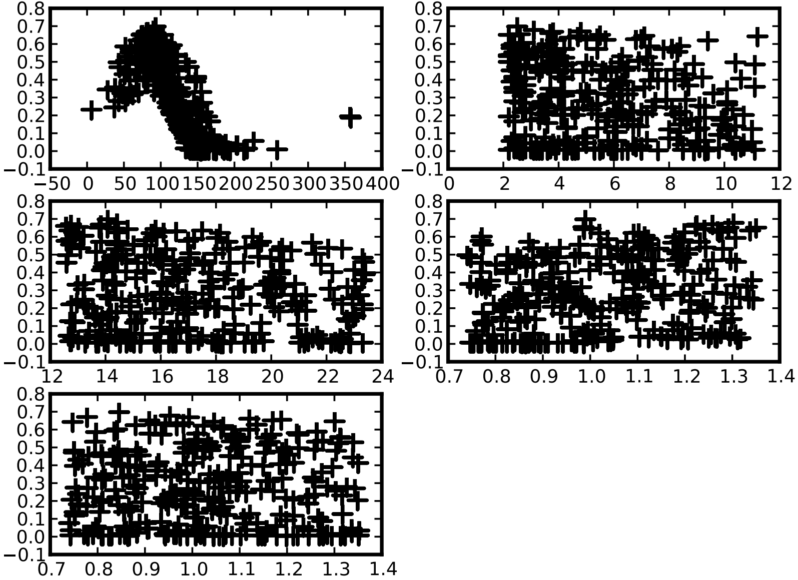
Distribution of the shape agreements

Scores of the best simulations, sorted independently for each model



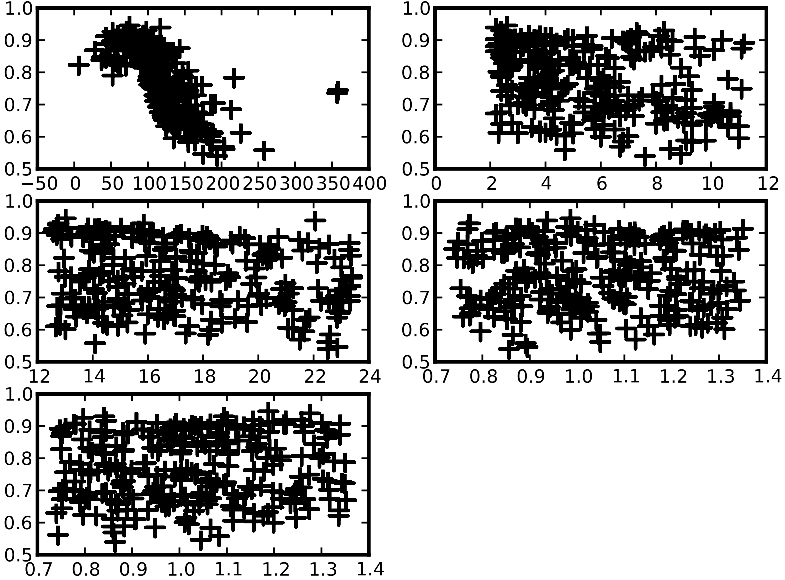
Distribution of the Jaccard similarity coefficients

Against input parameters, for case “Patrimonio (2005-02-11)”



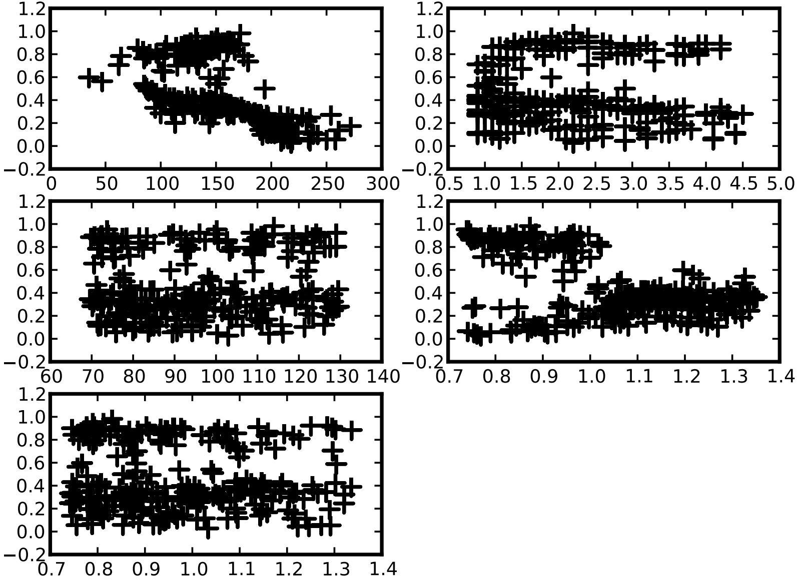
Distribution of the shape agreements

Against input parameters, for case "Patrimonio (2005-02-11)"



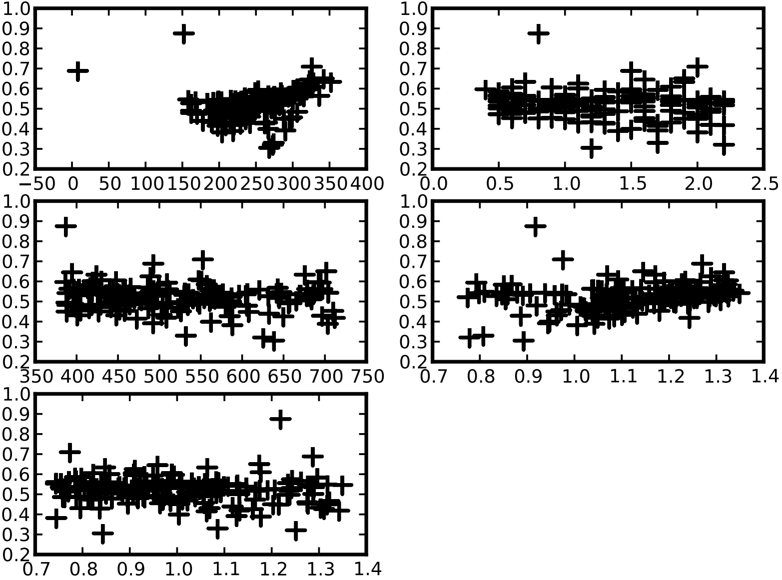
Distribution of the shape agreements

Against input parameters, for case "Olcani (2006-01-01)"

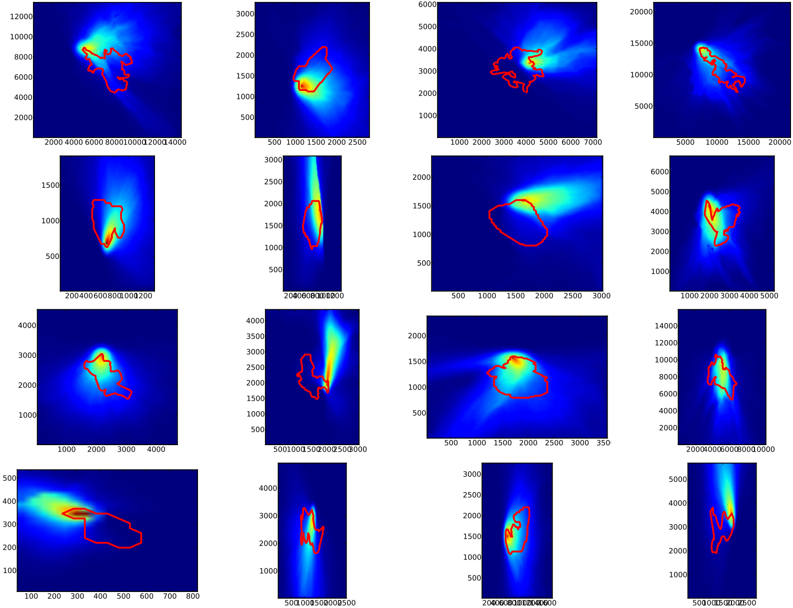


Distribution of the shape agreements

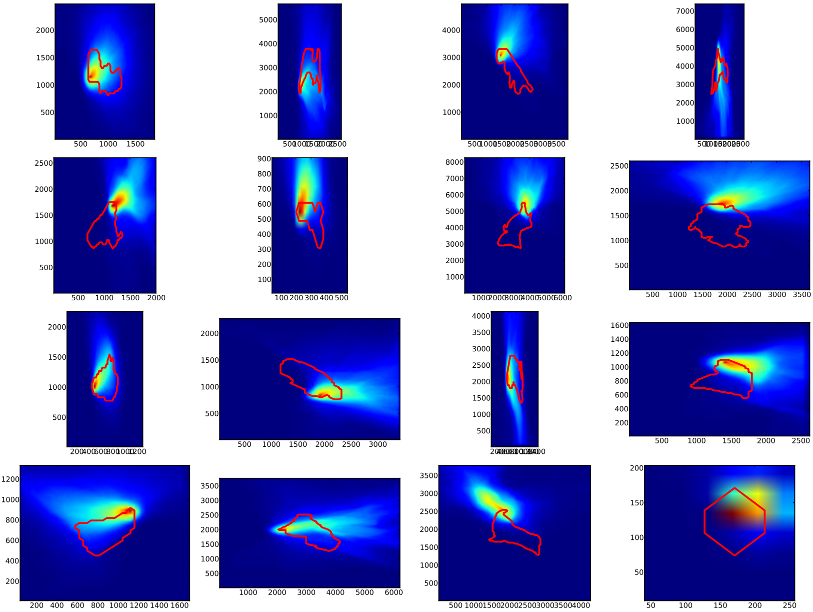
Against input parameters, for case “Sisco (2003-08-14)”



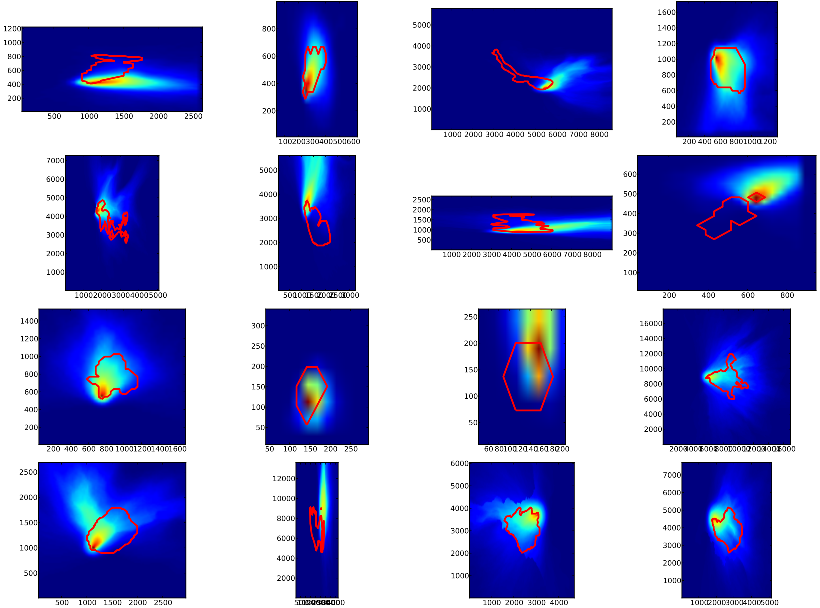
75 fire cases [1/5]



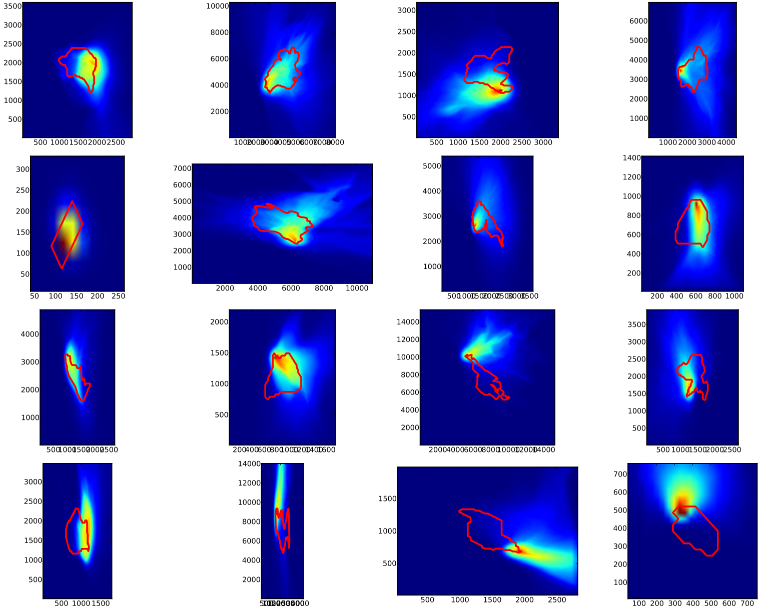
75 fire cases [2/5]



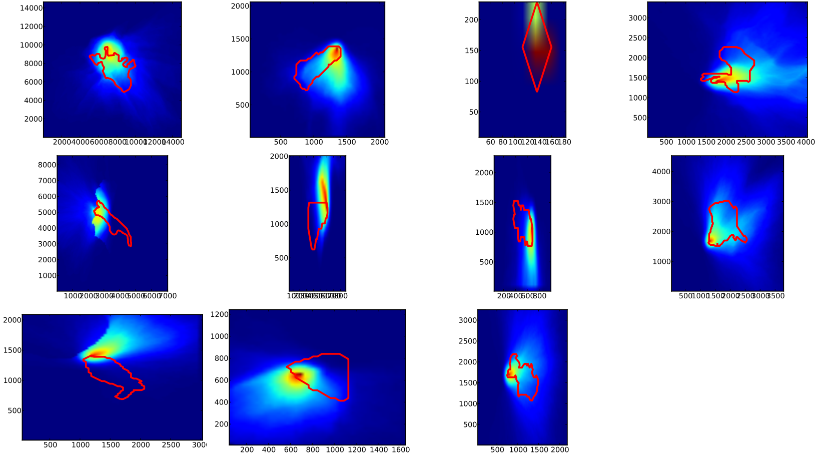
75 fire cases [3/5]



75 fire cases [4/5]

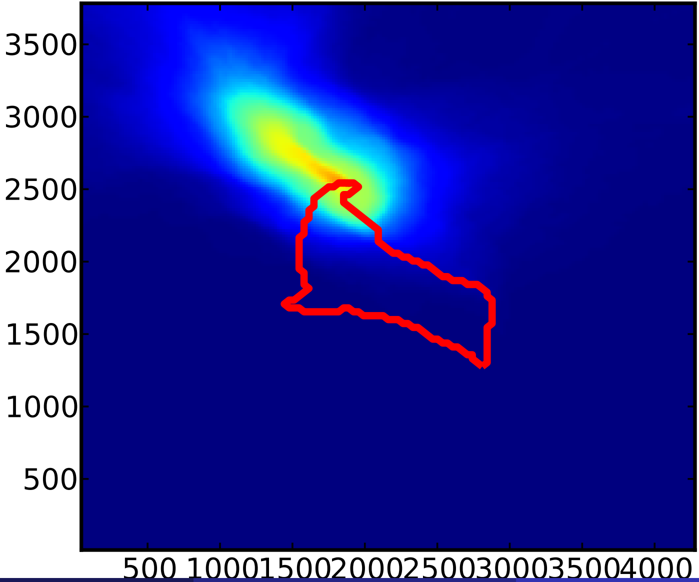


75 fire cases [5/5]



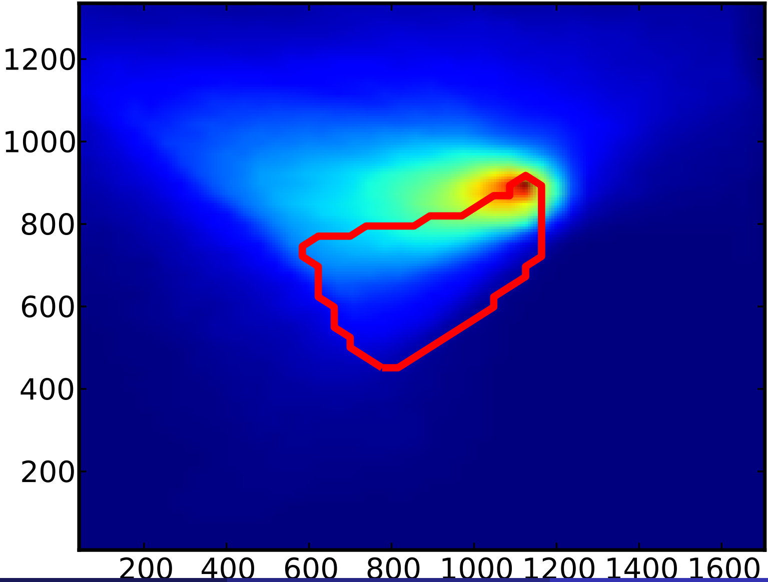
Probability density

For case "Olcani (2006-01-01)"

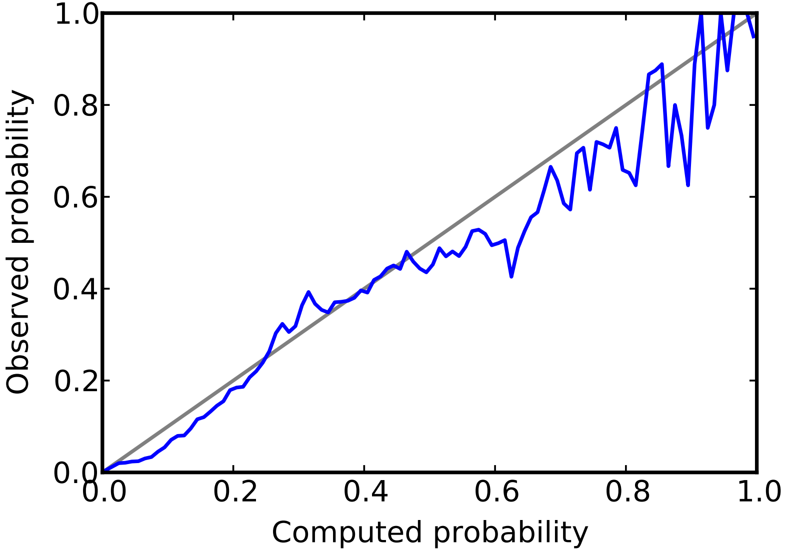


Probability density

For case “Patrimonio (2005-02-11)”



Reliability diagram



Conclusions

Scoring methods

- It is possible to take into account the model dynamics
  - Arrival time agreement and shape agreement

Evaluation on 80 fires

- It seems possible to rank models, without any (over)tuning
- 3-percent rate of spread significantly worse

Monte Carlo simulations

- May also be used to rank models
- Toward probabilistic forecasts and uncertainty estimation