

Data-Driven Modeling to predict forest fire spread in the Patagonian region in Argentina

Lucas Becerra ^{a,b} ^{*}, Monica Malen Denham ^{c,d} , Alejandro B. Kolton ^{a,b,d} , Karina Laneri ^{a,b,d} 

^a Instituto Balseiro. Universidad Nacional de Cuyo, San Carlos de Bariloche, Río Negro, Argentina

^b Centro Atómico Bariloche. Comisión Nacional de Energía Atómica, San Carlos de Bariloche, Río Negro, Argentina

^c Centro Interdisciplinario de Telecomunicaciones, Electrónica, Computación y Ciencia Aplicada (CITECCA). Universidad Nacional de Río Negro, San Carlos de Bariloche, Río Negro, Argentina

^d Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

ARTICLE INFO

Keywords:

Forest fire modeling
Genetic algorithm
Reaction–diffusion–convection model
Wildfire simulation
Parameter optimization
Patagonian region

ABSTRACT

Wildfires are among the most severe disturbances affecting forest ecosystems, with over 50,000 hectares burned in Patagonia, Argentina, during 2025 alone. This study implements a Reaction–Diffusion–Convection (RDC) model to simulate wildfire spread in the Steffen and Martín Lakes area, a region severely impacted by fires. By integrating high-resolution maps of slope, wind velocity, and vegetation, we conducted three computational experiments of increasing complexity to simulate fire propagation across heterogeneous landscapes.

We employed a Genetic Algorithm (GA) to recover reference model parameters by maximizing the spatial overlap between simulated and reference burned areas. Subsequently, parameter estimates were refined using XGBoost to improve accuracy. Results demonstrate that the GA accurately recovers reference parameters across all scenarios, while the XGBoost fine-tuning significantly enhances accuracy in simpler cases. This integrated framework offers a systematic approach for estimating difficult-to-measure wildfire parameters, demonstrating the potential of hybrid computational methods for wildfire modeling and forest management.

1. Introduction

Forests represent one of the most important terrestrial ecosystems on the planet. They cover approximately a quarter of the Earth's land area (Ritchie, 2021) and provide essential functions, including productive, protective, recreative, and environmental functions (Führer, 2000). The loss of forests can lead to the loss of biodiversity, soil erosion by water and wind, contamination of ground and spring water, and even desertification. Despite their crucial roles, forests are increasingly threatened by wildfires, which represent one of the major natural and human-induced disturbances and are expected to intensify with climate change (Syphard and Keeley, 2015; Kitzberger et al., 2022).

Just in the Patagonian Region of Argentina, more than fifty thousand hectares of forest, plantations and residential areas have been lost due to wildfires in 2025, according to official reports (Gobierno Argentino, 2025).

The drivers behind wildfire ignition are diverse and region-dependent. In Northwestern Patagonia, there is a growing trend towards lightning-caused fires, as recently identified by Kitzberger and Bürgesser (2025). However, the majority of wildfires are still caused by humans (Costafreda-Aumedes et al., 2017). For instance, 88% of the

wildfires that occurred between 2016 and 2020 in the United States were human-caused (Hoover and Hanson, 2021).

Anticipating fire behavior is a valuable tool in wildfire management. Traditional approaches such as fire suppression and fuel manipulation are increasingly complemented with alternative strategies, given the growing extent and intensity of wildfires (Syphard and Keeley, 2015). In that sense, simulations provide a powerful tool to support decision-making. However, in the Patagonian region, wildfire propagation remains a challenge to simulate accurately, since widely used global models (such as Rothermel, 1972) are not adapted to local conditions.

In response to this gap, several efforts have been made to model wildfire propagation in Patagonia. For instance, Morales et al. (2015) proposed a cellular automata model optimized using fire scars from past events in the region. Later, Denham and Laneri (2018) extended this approach by implementing a parallel genetic algorithm (GA) to recover parameters from synthetic wildfires. A different perspective was adopted by Laneri et al. (2020), who developed a dynamical model for forest fire behavior in Patagonian landscapes. This model was implemented in a parallel and open-source simulator (Simulador

* Corresponding author.

E-mail address: lucasbecerra0210@gmail.com (L. Becerra).

de Incendios Patagonia, *SIP*) to improve the fire management in the region (Denham et al., 2022). Additionally, Barberá et al. (2025) contributed to the understanding of the physical drivers of fire through a series of ecological and biophysical studies. Together, these works have significantly expanded our understanding of wildfire dynamics in Patagonia.

Beyond regional studies, several works have explored hybrid approaches that combine physics-based wildfire models with machine learning techniques. For instance, Wadhvani et al. (2025) and Yu et al. (2025) implemented surrogate models based on deep neural network architectures to enhance the predictive performance of deterministic and stochastic wildfire models, respectively. Another example is (Li et al., 2024), who developed an interpretable hybrid ML framework that explicitly incorporates fuel availability, fuel flammability, and human suppression effects. These approaches aim to improve model predictions when the input parameters are assumed to be known. In contrast, Jellouli and Bernoussi (2022) employed a cellular automata model coupled with dynamic wind flow simulations, illustrating how rule-based approaches can also capture critical drivers of fire spread without relying on machine learning.

Nevertheless, parameter optimization remains a critical bottleneck for improving model accuracy, particularly in regions where field measurements are scarce. In this work, we address this limitation by optimizing a Reaction–Diffusion–Convection (RDC) model for the Patagonian region using a GA-based approach. To this end, we designed three experiments of increasing complexity to test the potential and the limitations of the GA in recovering reference parameters. However, the GA is a stochastic optimization method that requires a large number of computationally expensive simulations and does not guarantee convergence to a global optimum, particularly in high-dimensional parameter spaces. To address these limitations, we implemented a surrogate model (XGBoost) to enhance the exploration of the parameter space efficiently, enabling the evaluation of a much larger number of candidate solutions at a significantly lower computational cost. By combining an RDC model with a GA+XGBoost optimization scheme, this study provides a systematic framework to estimate parameters that are otherwise difficult to measure in the field, contributing to both theoretical understanding and practical fire management in real Patagonian landscapes.

In summary, the novel contributions of this work are: (a) the implementation of a reaction–diffusion–convection (RDC) wildfire model in a realistic Patagonian landscape using spatially explicit data (vegetation, wind, and slope); (b) the development of a controlled experimental framework based on synthetic wildfires to systematically evaluate parameter recovery; (c) a systematic assessment of the performance and limitations of genetic algorithms as the dimensionality and complexity of the parameter space increase; and (d) the integration of a surrogate model (XGBoost) to enhance the exploration of the parameter space.

2. Reaction–diffusion–convection model

The model used in this work is a *reaction–diffusion–convection* (RDC) model (Denham et al., 2022; Laneri et al., 2020), which describes the spread of fire in a two-dimensional medium in a phenomenological manner. The model does not attempt to capture the microscopic physics of combustion; rather, its purpose is to generate generic reaction–diffusion fronts representative of wildfire propagation at coarse-grained scales.

The system is described by the following set of equations:

$$\frac{\partial S}{\partial t} = -\beta SI, \quad (1)$$

$$\frac{\partial I}{\partial t} = \beta SI - \gamma I + D\nabla^2 I - \vec{v} \cdot \nabla I, \quad (2)$$

$$\frac{\partial R}{\partial t} = \gamma I. \quad (3)$$

The fields are defined on a two-dimensional spatial domain.

This spatiotemporal model is inspired by the SIR epidemiological framework (Murray, 2003; Kolton and Laneri, 2019; Zagarra et al., 2024). Here, S represents the local fraction of susceptible vegetation or available fuel, I represents the fire intensity, and R denotes the fraction of burnt vegetation. The diffusion term $\nabla^2 I$ accounts for the spatial propagation of burning activity, such as heat transfer, flame contact, or ember-induced ignition, within a coarse-grained description, rather than explicit thermal diffusion.

While the mathematical structure of the model is identical to that of spatial SIR epidemic models, its interpretation is fundamentally different. In the present context, S , I , and R represent local *state fractions* of the medium, rather than conserved populations. Consequently, there is no underlying local conservation law associated with mass or energy in the forest-fire interpretation. A key advantage of Eqs. (1)–(3) over, for instance, cellular automaton models is that they admit analytical results in the simplest homogeneous case, including closed-form expressions for the front velocity and width. Despite its simplicity, it has been shown (Zagarra et al., 2024) that fronts in this model belong to the same universality class as more realistic models of fire fronts in random media (Provatas et al., 1995), where thermal and chemical processes are taken into account explicitly.

Eq. (1) describes the dynamics of the susceptible subpopulation. This is governed by a nonlinear reaction term, where $\beta > 0$ is the ignition rate and depends on the type of vegetation. A larger β means that the vegetation ignites more easily.

Eq. (2) describes the dynamics of fire intensity, associated to burning vegetation. This subpopulation increases at the same rate β that susceptibles decrease and burns during a characteristic time γ^{-1} , transforming into the R fraction. After burning, those cells cannot burn again. The spread of fire is further modeled by a diffusion term with a diffusion constant D that for simplicity is assumed homogeneous across the grid. Additionally, an advective term was included into the model, where the advective velocity \vec{v} can be decomposed into the effects of wind \vec{w} and terrain slope $\vec{\nabla}h$:

$$\vec{v} = A\vec{w} + B\vec{\nabla}h \quad (4)$$

where A and B are homogeneous constants that weight the contributions of wind and slope of the terrain, respectively. As the fires tend to propagate in the wind direction and in the direction of increasing height it is expected that both, $A > 0$ and $B > 0$. Finally, Eq. (3) describes the dynamics of the burnt vegetation. The fields \vec{w} and $\vec{\nabla}h$ are input data for the simulations, while the parameters β , γ , D , A , and B are fitted using a genetic algorithm and a fitness function, as will be explained below.

For the numerical implementation, we discretize the set of Eqs. (1)–(3) on a two-dimensional grid using a finite differences scheme. Reaction and advection terms are integrated explicitly, while diffusion is handled with an alternating direction implicit (ADI) scheme (Press, 2007). These components are coupled through operator splitting: a half step of the reaction–advection update is followed by a half step of diffusion in one spatial direction, and the process is repeated in the orthogonal direction to complete the time step. Since the advection term is computed explicitly, the advection velocity \vec{v} must follow the Courant–Friedrichs–Lewy (CFL) condition (Press, 2007). The full discretization is provided in Appendix A.

3. Methods

3.1. Model fitting and synthetic data

In order to optimize the RDC model for the Patagonian region, we selected a representative landscape: the area of Steffen and Martin lakes, located at approximately 41°30′39.46″S, 71°39′26.34″W, in Río Negro Province, Argentina; a scenario that has proven to be susceptible to wildfires, with two significant events recorded in December

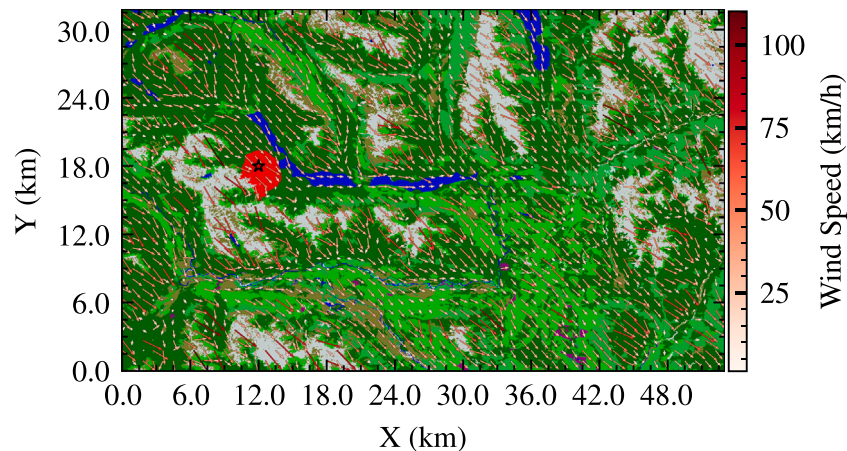


Fig. 1. Reference wildfire for the first experiment, simulated with $D = 10 \text{ m}^2 \text{ h}^{-1}$, $A = 1 \times 10^{-4}$, $B = 15 \text{ m h}^{-1}$, and ignition point at $(x, y) = (4.0, 6.0)$. Each green area corresponds to a different fuel type (Forest A, Forest B, exotic forest, pasture, and shrubland), with spatially heterogeneous β and γ values listed in Table 1. Red indicates the final burned area (final state after 500 steps, $dt = 0.5 \text{ h}$), yielding a burnt area of approximately 1015 ha; the star marks the ignition point; arrows indicate wind direction.

Table 1

Values for the ignition rate β and the extinction rate γ for each fuel type considered in the first and third experiment.

Fuel type	β value (h^{-1})	γ value (h^{-1})
Forest A	0.91	0.50
Forest B	0.72	0.38
Exotic forest	1.38	0.84
Pasture	1.94	0.45
Shrubland	0.75	0.14

2021 (Bari et al., 2022) and in December 2024 (Gobierno de Río Negro, 2025).

The simulation grid consists of 1748×1060 square cells, each cell covering an area of 900 m^2 (30 m resolution). The input data are raster ASCII maps including vegetation, wind, and slope information of the selected landscape. Slope rasters were derived from Digital Elevation Model (DEM) data. The wind direction for each cell was estimated from the average dominant wind direction during fire season and further adjusted for topographic effects using the WindNinja simulator (US Forest Service Research & Development, 2025). The vegetation map is a raster file with integer values that represent each fuel type. In our simulations, we considered five fuel types: two classes of native forest (Forest A, mainly conifers and broadleaf species, and Forest B, dominated by *Cupressus*), exotic forest, pasture, and shrubland.

We designed three different experiments of increasing complexity in the Steffen and Martin lakes scenario with the objective of testing the GA method to recover the RDC model parameters. In each experiment, using Eqs. (1), (2) and (3), we simulated a wildfire that will be called “reference wildfire”, which we generated with a set of arbitrary but plausible parameters, that we aimed to recover using the GA. For each one we performed 500 simulation steps with a stepsize $dt = 0.5 \text{ h}$, which is equivalent to 10.42 days of wildfire propagation. In the following, we describe the three experiments in detail.

In the first experiment, we optimized the parameters D , A , B , and the ignition point coordinates (x, y) , while keeping β and γ fixed but heterogeneous in space according to the vegetation map. The values of β and γ for each fuel type are listed in Table 1. The reference wildfire was simulated with the parameters $D = 10 \text{ m}^2 \text{ h}^{-1}$, $A = 1 \times 10^{-4}$, $B = 15 \text{ m h}^{-1}$, and ignition point at $(x, y) = (4.0, 6.0)$. The reference wildfire for this experiment is shown in Fig. 1. In total, we recovered five parameters (D, A, B, x, y) in this experiment.

In the second experiment, we optimized the same parameters (with the same values) as in the first one, but also including β and γ as

free parameters. For simplicity, the landscape was considered homogeneous, meaning that the same ignition and extinction rates were applied to all fuel types. The reference values used were $\beta = 1.5 \text{ h}^{-1}$ and $\gamma = 0.5 \text{ h}^{-1}$. The resulting reference wildfire is shown in Fig. 2. In total, we recovered seven parameters ($D, A, B, x, y, \beta, \gamma$) in this experiment.

In the third and more complex experiment, we added a different ignition and extinction rate for each fuel type as free parameters. The reference values are reutilized from Table 1. In order to optimize the five fuel types, we had to design a proper wildfire, meaning that the burnt area had to cover a representative area of each fuel type to be sensible to changes in the free parameters. To accomplish that challenge, we designed a wildfire with multiple ignition points so it could cover a representative area of each fuel type. To avoid an increase in the number of free parameters, we fixed the coordinates of the ignition points. For this case, we simulated a wildfire with three ignition points in the cells (1130, 290), (1300, 150), (620, 280). The other parameters are the same as those in the first and second experiments. The resulting reference wildfire is shown in Fig. 3. Its burnt area covered 29.15% of forest A, 44.09% of forest B, 3.62% of exotic forest, 13.18% of pasture and 9.96% of shrubland, ensuring a representative impact across all fuel types. In total we recovered thirteen parameters ($D, A, B, \beta_i, \gamma_i$) for $i = 1, 2, 3, 4, 5$ in this experiment.

The range of parameters to generate the initial population for the GA was set following different criteria. For the advection constants A and B , we took into account the maximum possible value given by the CFL condition. For a half stepsize of $dt/2 = 0.25 \text{ h}$ (the ADI scheme requires an evaluation at half time steps), a cell size of 30 m and taking the maximum value of wind and slope, the maximum values are $A_{\max} = 5.918 \times 10^{-4}$ and $B_{\max} = 24.853 \text{ m h}^{-1}$. For the diffusion constant, we set a maximum value of $D_{\max} = 100 \text{ m}^2 \text{ h}^{-1}$. For the β and γ rates, we set a maximum value of 2 h^{-1} and 1 h^{-1} , respectively. All these values were generated uniformly between zero and the maximum value. The ignition coordinates (in the first and second case) were limited to the burnt area of the reference wildfire.

As a summary, the details of the tree experiments are listed in Table 2.

3.2. Genetic algorithm

To estimate the optimal parameters of the RDC model, we implemented a GA. This population-based stochastic optimization method is well suited for high-dimensional, non-linear problems in which the objective function is computationally expensive to evaluate (Katoch et al., 2021).

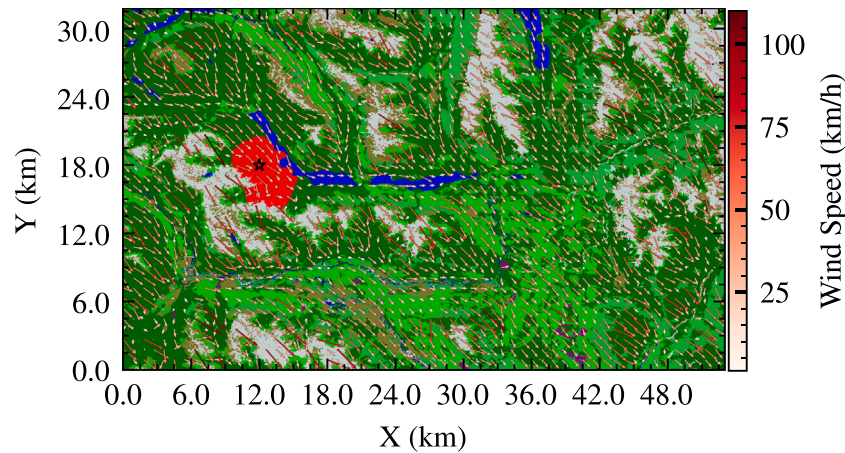


Fig. 2. Reference wildfire for the second experiment (Steffen-Martin landscape). The same reference values as the first experiment were used for ($D = 10 \text{ m}^2 \text{ h}^{-1}$, $A = 1 \times 10^{-4}$, $B = 15 \text{ m h}^{-1}$, and ignition point at $(x, y) = (400, 600)$), while homogeneous β and γ were introduced as additional tunable parameters with reference values ($\beta = 1.5 \text{ h}^{-1}$ and $\gamma = 0.5 \text{ h}^{-1}$), respectively. Red indicates burned area (final state after 500 steps, $dt = 0.5 \text{ h}$), yielding a burnt area of approximately 2342 ha.

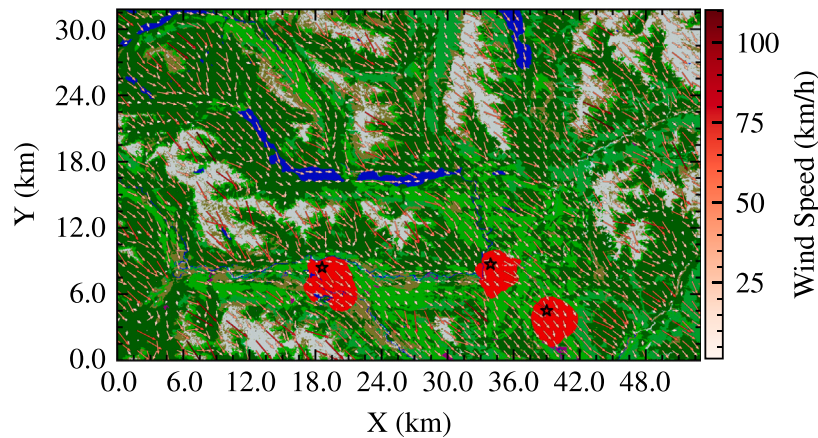


Fig. 3. Reference wildfire used in Experiment 3 (Steffen-Martin landscape). The simulation uses $D = 10 \text{ m}^2 \text{ h}^{-1}$, $A = 1 \times 10^{-4}$, and $B = 15 \text{ m h}^{-1}$, with three fixed ignition points at $(1130, 290)$, $(1300, 150)$, and $(620, 280)$. Each green area corresponds to a different fuel type (Forest A, Forest B, exotic forest, pasture, and shrubland), with heterogeneous and tunable β_i and γ_i values (Table 1). Red indicates the final burned area (final state after 500 steps, $dt = 0.5 \text{ h}$), yielding a burnt area of approximately 4241 ha; this reference fire was designed to burn a representative fraction of all fuel types.

Table 2

Summary of the three synthetic experiments designed to evaluate the proposed fitting methods. The parameters to be adjusted and their reference values are detailed, as well as the fixed parameters and their respective values.

Exp.	Parameter to adjust	Reference value	Fixed parameters	Reference value
1	D	$10 \text{ m}^2 \text{ h}^{-1}$	β	Heterogeneous according to Table 1
	A	1×10^{-4}	γ	Heterogeneous according to Table 1
	B	15 m h^{-1}	-	-
	(x, y)	$(400, 600)$	-	-
2	D	$10 \text{ m}^2 \text{ h}^{-1}$	-	-
	A	1×10^{-4}	-	-
	B	15 m h^{-1}	-	-
	(x, y)	$(400, 600)$	-	-
	β	1.5 h^{-1}	-	-
	γ	0.5 h^{-1}	-	-
3	D	$10 \text{ m}^2 \text{ h}^{-1}$	(x_1, y_1)	$(1130, 290)$
	A	1×10^{-4}	(x_2, y_2)	$(1300, 150)$
	B	15 m h^{-1}	(x_3, y_3)	$(620, 280)$
	β	Heterogeneous according to Table 1	-	-
	γ	Heterogeneous according to Table 1	-	-

The algorithm begins by generating an initial population of individuals, where each individual is defined by a set of parameters (genes). Every individual is then evaluated through a fitness function, which quantifies the similarity between the simulated fire spread and a reference map. In this work, we adopted the fitness function proposed in Denham (2009), which compares the number of burnt cells in common between the two maps:

$$\delta = \frac{A \cup A^* - A \cap A^*}{A}, \quad (5)$$

where A denotes the set of burnt cells in the reference map and A^* the set of burnt cells in the simulated map. Eq. (5) approaches zero as the simulated map becomes more similar to the reference. In the worst case in which the simulated map has a null intersection with the reference, $\delta = (A + A^*)/A = 1 + A^*/A > 1$ and the maximum value is given by the size of the map. We considered a burnt cell any cell with $R > 0.001$ at the end of the simulation.

Once the fitness of all individuals is computed, the GA applies three main operators: selection, where the best candidates are chosen according to their fitness; crossover, where pairs of individuals exchange some of their genes with a certain probability; and mutation, which introduces small random changes to the genes. These steps are repeated over successive generations, progressively improving the population. The population is constant in size throughout the generations.

Further implementation details of the GA are provided in Appendix B.

3.3. Parameters search refinement and error assessment

Since the GA requires a large number of computationally expensive wildfire simulations to explore the parameter space, and does not guarantee convergence to a global optimum, we implemented a refinement method to approximate the fitness function and find parameter combinations with lower fitness values. This approach takes advantage of the million simulations performed in the three experiments to train a machine learning model that can predict the fitness of new parameter sets without running the full simulation. Among available surrogate modeling approaches, XGBoost stands out for its ability to handle large datasets and complex relationships (Chen and Guestrin, 2016; Grinsztajn et al., 2022; Shwartz-Ziv and Armon, 2022).

We trained an XGBoost regressor (Chen and Guestrin, 2016) using the parameters as input features and the corresponding fitness values as targets. The dataset was split into training (80%) and testing (20%) sets. The model's performance was evaluated using the mean squared error (MSE) and the coefficient of determination (R^2). The trained model used 5000 estimators, a learning rate of 0.01, a maximum depth of 10, L1 regularization term of 0.1, and L2 regularization term of 1. This hyperparameter configuration was selected based on a grid search procedure, and applied consistently across the three experiments. Preliminary tuning efforts, particularly for the third experiment, did not lead to significant performance improvements.

With the trained surrogate model, we can rapidly evaluate the fitness of new parameter sets, enabling a more efficient exploration of the parameter space. We evaluated 100 million candidates using the surrogate model for each experiment. Then, we selected the top 50 candidates with the lowest predicted fitness and performed full wildfire simulations to compute their true fitness values. The accuracy of this approach depends on the surrogate model's ability to approximate the fitness landscape, which may degrade in highly complex or high-dimensional cases.

To quantify the uncertainty in the parameters estimated by the GA and the XGBoost model, we applied a bootstrap analysis to the final population of each experiment. This resampling technique allows us to approximate the sampling distribution of the minimum fitness values, and consequently the distribution of the best-fit parameters. For each bootstrap replicate, we sampled with replacement from the final population, extracted the best individual (minimum fitness), and

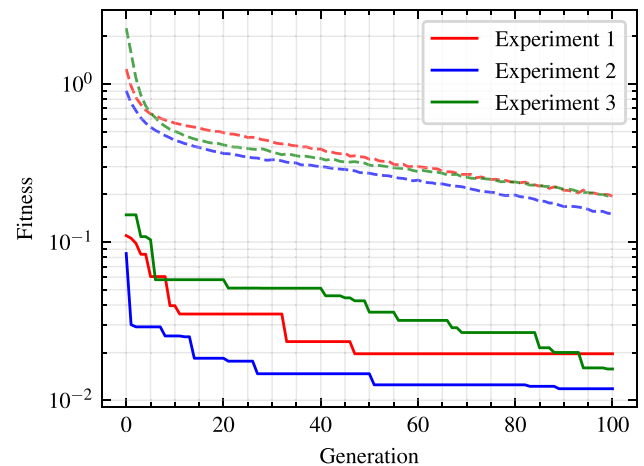


Fig. 4. Evolution of the fitness in the three experiments. The dashed line shows the mean fitness per generation, while the solid line indicates the minimum fitness (best individual). Elitism ensures that the best individual is preserved across generations.

repeated the process 1000 times. From these replicates we computed the mean and the 95% confidence interval, defined by the 2.5 and 97.5 percentiles.

The estimated uncertainty reflects the variability in the best-fit parameters arising from the stochastic nature of the genetic algorithm and the structure of the fitness landscape. In this context, the bootstrap confidence intervals (defined by the 2.5 and 97.5 percentiles of the estimated distribution) provide an empirical measure of the range of parameter values that yield similarly optimal solutions. A full propagation of parameter uncertainty into model outputs is beyond the scope of this work.

4. Results

We ran the genetic algorithm with 100 generations and a population size of 10000 individuals for each experiment, resulting in 1×10^6 wildfire simulations per experiment. Fig. 4 shows the evolution of the fitness along generations. The dashed line corresponds to the population mean, while the solid line represents the minimum fitness, i.e. the best individual of each generation. Due to elitism, the best individual is always preserved, which guarantees a non-increasing trend of the minimum fitness.

The bootstrap results for the three experiments are reported in Table 3. Each table lists the optimized parameters, their bootstrap mean, the corresponding 95% confidence intervals, the relative uncertainty, defined as the half-width of the confidence interval divided by the mean, and the relative error with respect to the true values.

The results reported in the Tables 3 show that the GA was able to recover some reference parameters with good accuracy across the three experiments.

After running the GA, we trained the XGBoost model for each experiment using the parameters as input features and the fitness values obtained from the GA million simulations as targets. The performance of the trained model for the three experiments is summarized in Table 4 and the results of the bootstrap analysis for the best individuals obtained after the refinement procedure are reported in Table 5.

In order to quantify the differences between the GA optimization and the XGBoost refinement, we compared the mean fitness values obtained from both methods in Table 6, using a bootstrap analysis.

Table 3

Results of the bootstrap analysis of the genetic algorithm for the three synthetic experiments. The uncertainty corresponds to the relative half-width of the confidence interval, calculated as $(\text{Upper Limit} - \text{Lower Limit}) / (2 \times \text{Mean}) \times 100$. The reference values for each experiment are found in Table 2.

Exp.	Parameter	Bootstrap Mean	Confidence Interval	Uncertainty (%)	Relative Error (%)
1	D ($\text{m}^2 \text{h}^{-1}$)	11.221	(10.436, 12.827)	10.65	12.21
	A	1.209×10^{-4}	$(0.934, 1.712) \times 10^{-4}$	32.18	20.90
	B (m h^{-1})	17.367	(14.345, 22.786)	24.30	15.78
	x	398.576	(396.000, 400.000)	0.50	0.36
	y	602.848	(600.000, 608.000)	0.66	0.47
2	D ($\text{m}^2 \text{h}^{-1}$)	34.056	(30.278, 34.670)	6.45	240.56
	A	1.107×10^{-4}	$(1.069, 1.121) \times 10^{-4}$	2.35	10.70
	B (m h^{-1})	17.955	(14.652, 18.511)	10.75	19.70
	x	402.784	(402.000, 404.000)	0.25	0.70
	y	600.423	(600.000, 604.000)	0.17	0.07
	β (h^{-1})	1.002	(0.990, 1.088)	4.89	33.20
	γ (h^{-1})	0.418	(0.411, 0.480)	8.25	16.4
3	D ($\text{m}^2 \text{h}^{-1}$)	11.321	(9.884, 11.899)	8.91	13.21
	A	1.008×10^{-4}	$(0.953, 1.036) \times 10^{-4}$	4.13	0.80
	B (m h^{-1})	14.527	(13.211, 15.075)	6.43	3.15
	β_1 (h^{-1})	0.527	(0.503, 0.621)	11.18	42.09
	β_2 (h^{-1})	0.587	(0.538, 0.609)	6.04	18.47
	β_3 (h^{-1})	0.898	(0.764, 1.068)	16.97	34.93
	β_4 (h^{-1})	1.969	(1.955, 2.000)	1.13	1.49
	β_5 (h^{-1})	0.658	(0.587, 0.771)	13.94	12.27
	γ_1 (h^{-1})	0.146	(0.129, 0.241)	38.33	70.80
	γ_2 (h^{-1})	0.259	(0.168, 0.286)	22.92	31.84
	γ_3 (h^{-1})	0.363	(0.209, 0.602)	54.61	56.79
	γ_4 (h^{-1})	0.530	(0.471, 0.559)	8.31	15.78
	γ_5 (h^{-1})	0.113	(0.100, 0.151)	22.66	84.93

Table 4

Performance metrics of the refined search for the three experiments. A cross-validation with 5 folds was performed to compute the mean and standard deviation of the metrics.

Experiment	MSE	R^2
First	$(1.14 \pm 0.02) \times 10^{-4}$	0.99746 ± 0.00004
Second	$(8.19 \pm 0.07) \times 10^{-4}$	0.98084 ± 0.00013
Third	$(236 \pm 4) \times 10^{-4}$	0.872 ± 0.003

The reported mean fitness corresponds to the expected value of the best fitness under the empirical sampling distribution obtained through bootstrap resampling, providing a robust estimate of the typical performance of each method. In contrast to reporting a single minimum fitness value, this approach reduces sensitivity to stochastic fluctuations and outliers in the optimization process.

The results in Table 6 show that the XGBoost refinement improves upon the GA results in the first experiment. In contrast, the GA achieves better performance in the second experiment, and significantly outperforms XGBoost in the third one. This behavior reflects the increasing complexity of the fitness landscape: while the surrogate model is effective in simpler scenarios, its performance deteriorates in more complex and high-dimensional cases, where accurately approximating the fitness function becomes more challenging.

5. Discussion

In this section, we analyze the performance of the genetic algorithm and the subsequent refinement method across the three experiments, focusing on which parameters were more reliably estimated and which ones showed larger uncertainties. By contrasting these outcomes, we highlight the strengths and limitations of the method in recovering the true model parameters.

In the first and second experiment, the ignition point coordinates were the best-estimated parameters. This outcome can be attributed to the fact that the search space for the coordinates was discrete and restricted to the burnt area of the reference wildfire. In the second experiment, however, the bootstrap mean for the diffusion constant exhibited a relative error larger than 90%, highlighting a difficulty

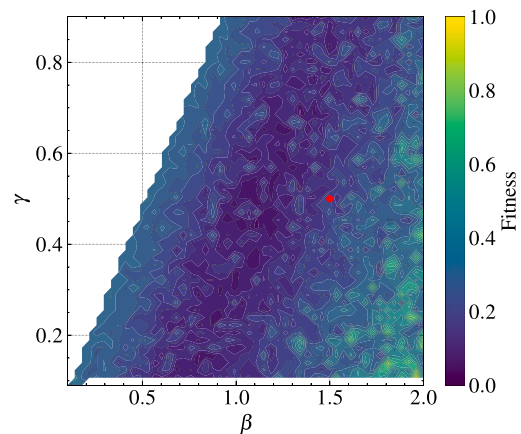


Fig. 5. Fitness landscape in relation to β and γ for the second case. It is shown that there is a valley of similar fitness values, which creates a strong correlation between both parameters and hinders their accurate estimation. The red point indicates the reference values $(\beta, \gamma) = (1.5, 0.5)$.

of optimizing this parameter. Moreover, the inclusion of β and γ as additional tunable parameters further increased the complexity of the fitness function, making convergence more challenging.

Fig. 5 illustrates the fitness landscape with respect to β and γ in the second experiment. The surface exhibits a highly irregular structure, with many points of similar fitness and no clear decreasing direction, which makes accurate estimation difficult. This behavior is consistent with the correlation matrix for the second experiment (Fig. 6), where β and γ show a strong correlation (0.93), indicating an almost linear relationship between them. This coupling can be attributed to the role of the basic reproduction rate in diffusive and non-diffusive epidemic models (Murray, 2003).

The correlation analysis supports the presence of practical non-identifiability of β and γ in the second experiment: many parameter

Table 5

Bootstrap analysis results for candidates found by the XGBoost model and validated with the RDC model for the three synthetic experiments. Reference values for each experiment are found in Table 2.

Exp.	Parameter	Bootstrap Mean	Confidence Interval	Uncertainty (%)	Relative Error (%)
1	D ($m^2 h^{-1}$)	10.206	(9.846, 10.361)	2.52	2.06
	A	1.008×10^{-4}	$(0.884, 1.352) \times 10^{-4}$	25.81	0.80
	B ($m h^{-1}$)	13.866	(12.622, 18.910)	22.78	7.56
	x	398.930	(396.185, 399.465)	0.41	0.27
	y	600.353	(599.637, 603.788)	0.34	0.06
2	D ($m^2 h^{-1}$)	22.772	(20.286, 25.631)	11.74	127.72
	A	1.522×10^{-4}	$(1.038, 1.615) \times 10^{-4}$	18.95	52.2
	B ($m h^{-1}$)	19.873	(15.646, 20.478)	12.16	32.49
	x	397.056	(394.909, 399.363)	0.56	0.74
	y	607.966	(597.516, 610.901)	1.10	1.33
	β (h^{-1})	1.040	(0.838, 1.128)	13.96	30.67
	γ (h^{-1})	0.337	(0.125, 0.424)	44.50	32.60
3	D ($m^2 h^{-1}$)	48.113	(37.493, 96.937)	61.78	381.13
	A	1.415×10^{-4}	$(1.178, 1.895) \times 10^{-4}$	25.33	41.50
	B ($m h^{-1}$)	18.833	(8.024, 23.183)	40.25	25.55
	β_1 (h^{-1})	0.823	(0.680, 0.918)	14.50	9.56
	β_2 (h^{-1})	0.935	(0.862, 0.966)	5.53	29.86
	β_3 (h^{-1})	0.737	(0.267, 0.904)	43.20	46.59
	β_4 (h^{-1})	1.310	(0.881, 1.467)	22.35	32.47
	β_5 (h^{-1})	0.206	(0.136, 0.476)	82.43	72.53
	γ_1 (h^{-1})	0.725	(0.541, 0.791)	17.27	45.00
	γ_2 (h^{-1})	0.731	(0.687, 0.826)	9.49	92.37
	γ_3 (h^{-1})	0.707	(0.334, 0.849)	36.33	15.83
	γ_4 (h^{-1})	0.470	(0.094, 0.834)	78.68	4.44
	γ_5 (h^{-1})	0.735	(0.683, 0.871)	12.78	425.00

Table 6

Comparison of mean fitness values obtained from the GA and XGBoost methods across the three synthetic experiments, based on bootstrap analysis. The XGBoost method outperforms the GA in the first experiment, while the GA achieves better performance in the second and third experiments.

Experiment	Genetic Algorithm	XGBoost
1	0.028	0.023
2	0.012	0.022
3	0.020	0.398

combinations yield similarly low fitness, forming an elongated valley in the objective landscape. This indicates that the available data are insufficient to uniquely constrain these parameters. Thus, the main limitation in our current setup is practical identifiability, although additional structural analysis would be required to formally rule out model intrinsic non-identifiability. By contrast, Fig. 7 shows the fitness landscape for the ignition coordinates in the first and second experiments. In both cases, in contrast to the previous figure, there is no ambiguity in the estimation of the coordinates. Despite the irregularity of the fitness landscape, the GA was able to converge to the correct ignition point.

Despite the fact that the diffusion constant was poorly estimated in the second experiment, the fitness of the best individual was very close to that of the reference wildfire, as can be seen in Fig. 4. This result suggests that different combinations of parameters can lead to similar fire spread patterns, at least in this regime of the parameter space. The difference between the reference and the best-estimated parameters for the second experiment is shown in Fig. 8.

In the third experiment, the diffusion and advection constants were accurately estimated. Not only that, but the associated relative errors were smaller than in the previous experiments. Nevertheless, the ignition and extinction rates for the different fuel types were poorly estimated in several cases. This result can be partly explained by the vegetation distribution in the landscape. Surprisingly, the coefficient for pasture was the best estimated, even though these were not the most prominent vegetation within the burnt area. Moreover, the outcome of a single GA run may lead to a misleading interpretation; however, this limitation was mitigated through the bootstrapping analysis.

The implementation of the refinement method using machine learning demonstrated promising results, particularly in the first experiment, where better estimations of the parameters were achieved compared to the GA without refinement. This outcome can be attributed to the high accuracy of the refined method in approximating the fitness function, as indicated by the R^2 value of 0.99746. In the second experiment, the results were comparable to those obtained with the GA, as the fitness landscape converged more clearly to a minimum. It is shown that the GA converged to a close minimum, and the XGBoost model was able to identify similar candidates. Then, the bias in the GA was transferred to the XGBoost model. In the third experiment, however, the refinement method produced worse results than the GA ($R^2 = 0.872$). The

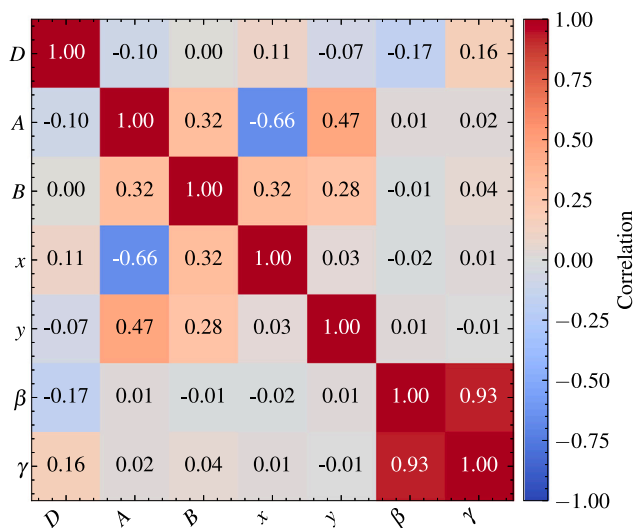


Fig. 6. Correlation matrix of the model parameters found in the second experiment. The correlation was calculated using the best 10000 individuals across all the generations. There is a strong correlation between β and γ , which leads to almost a linear relation between both parameters and hinders the optimization process.

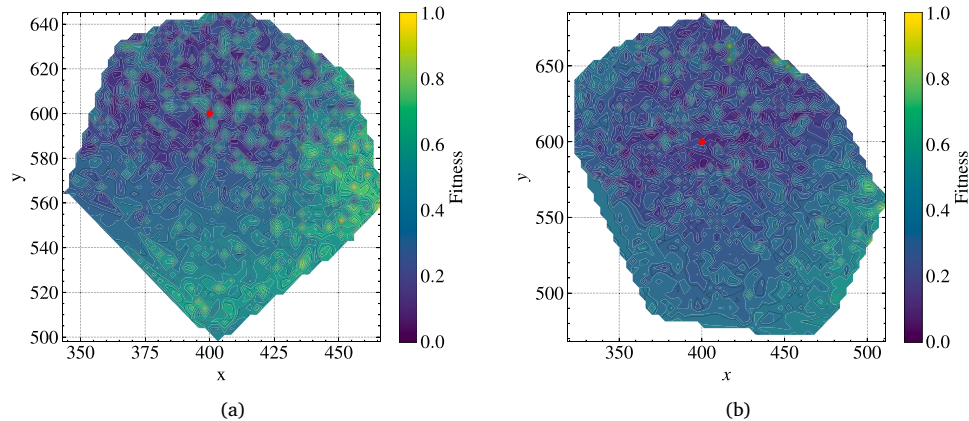


Fig. 7. Fitness landscape in relation to the first (left) and second (right) experiment. In contrast with Fig. 5, there is no ambiguity in the estimation of the ignition coordinates. The red point indicates the reference values $(x, y) = (400, 600)$.

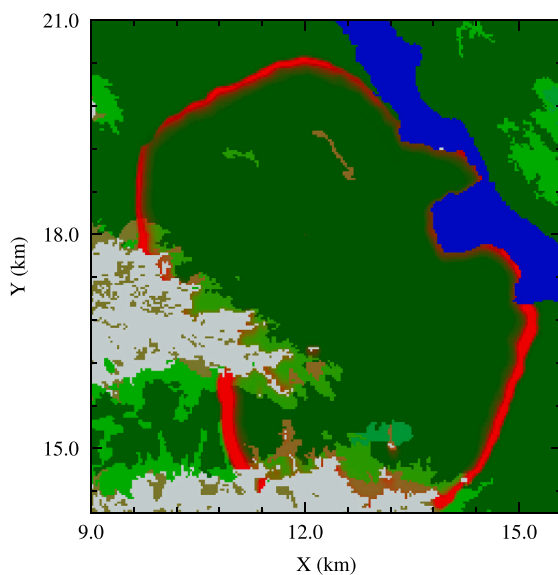


Fig. 8. Difference between the reference wildfire and the best-estimated wildfire in the second experiment. The difference is seen in the border of the burnt area, which indicates that both fire spread patterns are very similar.

increased complexity of the fitness landscape in the third experiment likely contributed to this reduced accuracy, as can be seen in Table 4. Further investigation is needed to enhance the refinement method for high-dimensional parameter spaces with complex fitness landscapes.

From an ecological perspective, the parameters of the RDC model have a direct interpretation in terms of wildfire behavior. For instance, the parameter β is associated with the effective spread rate of the fire, while γ represents the rate of fuel depletion or fire extinction. The diffusion term captures the local spread due to short-range processes such as radiation and spotting, whereas the convection term accounts for the influence of wind and slope in fire propagation.

In this work, we used a fitness function (Eq. (5)) based on the spatial overlap between the final burned areas of the simulated and reference fire spread patterns. In this case, the reference corresponds to a synthetic wildfire generated by the SIR model. Alternative approaches

in the literature incorporate temporal information into the fitness function (Kelso et al., 2015), which can capture aspects of the fire dynamics beyond the final burned extent. This is ecologically relevant because two wildfires may produce similar final burn scars while exhibiting different temporal evolutions, such as differences in front velocity or rate spread (Kolaitis et al., 2023).

In the model used in this work, however, time has a direct physical interpretation, as the system is based on a physical reaction–diffusion–convection formulation. Consequently, simulations are run for the same duration as the reference fire, which implicitly constrains the average propagation speed of the fire front. This partially reduces the space of admissible solutions, as similar burned areas obtained at the same simulation time cannot correspond to arbitrarily different average propagation velocities. However, this constraint is not sufficient to uniquely determine the underlying parameters, since different combinations can still produce similar spatial patterns and comparable effective propagation regimes. Differences in the detailed temporal dynamics of the fire spread may still exist. Exploring fitness functions that explicitly incorporate temporal information remains an interesting direction for future work.

A limitation of the present study is that the validation is based exclusively on synthetic wildfire scenarios. While this approach enables a controlled evaluation of parameter recovery under known conditions, it does not directly assess predictive performance in real wildfire events. In real settings, fire propagation is influenced by time-varying environmental factors, particularly wind, which can change significantly over the course of several days. In contrast, the current approach assumes a stationary wind field, representing a simplification that may limit its applicability to real-world scenarios. Addressing these limitations will require incorporating time-dependent environmental inputs and validating the model against observed wildfire data.

From a practical standpoint, these findings indicate that incorporating additional sources of information, such as temporal fire progression (e.g., isochrones) or time-varying environmental conditions, may be necessary to improve parameter estimation and model reliability. This is particularly relevant for decision-making processes in wildfire management, where accurate predictions of fire spread dynamics are essential for planning and risk assessment.

6. Conclusions

Wildfires are a major concern year after year in the Patagonian region. There is an urgent need to simulate forest fires propagation

in real landscapes, taking into account various layers of information — such as wind, slope, and vegetation — in order to test different scenarios for fire prevention. In this work, we calibrated a reaction–diffusion–convection model that simulates fire spread over real maps of the Patagonian region, particularly around Lake Steffen Martin, which is highly representative of the area. We progressively added complexity to the parameter search space in order to evaluate the performance of the fitting method when adjusting the simulations to the synthetically generated data. In the first experiment, we tuned five parameters, seven in the second, and thirteen in the third. To improve model fitting, we performed a bootstrap analysis over the results of the last generation, quantifying the uncertainties of the best individuals. Then we tuned the GA algorithm with machine learning techniques, to improve even more the fitting methodology. As a result, we were able to recover the original parameters with very good accuracy. The best performances were found in the ignition coordinates, followed by the diffusion and advection coefficients. The ignition and extinction rates remained the most challenging parameters. In the third experiment, where five fuel types were adjusted, only for shrubland and pasture these coefficients were accurately estimated.

Our systematic approach facilitates the estimation of field parameters in heterogeneous landscapes that are otherwise difficult to quantify. Though applied here to an RDC forest-fire model, the method is applicable in principle to a broad range of models based on cellular automata or partial differential equations. These results also highlight the importance of incorporating ecologically relevant information, such as temporal fire dynamics, to improve the interpretability and applicability of wildfire spread models in real-world scenarios.

The next step will involve fitting the model to real wildfire scars, particularly the Lake Martin fire, whose isochrones will enable the calibration of quantities such as the fire spread rate, diffusion coefficient, and the constants related to wind and slope. This extension will also require incorporating time-varying environmental conditions, especially wind, to better represent real fire dynamics. This work represents an important step towards predicting fire propagation in the Patagonian region. Combined with the already developed visual interface, it will serve as a valuable tool for planning and prevention efforts by firefighters and national fire management authorities.

CRediT authorship contribution statement

Lucas Becerra: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis. **Monica Malen Denham:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Alejandro B. Kolton:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis. **Karina Laneri:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

M. D., K. L., and A. K. are members of CONICET. M. D. and K. L. are part of the project grants PI UNRN 40-B-745 and Proyecto Federal de Innovación: Proyecto RN-2- PFI 2022. L.B. has a scholarship from Instituto Balseiro-UNCUYO-CNEA. We thank Marcelo Bari, Paula Presti and members of ICE for fruitful discussions. Computational resources were provided by the HPC cluster of the Physics Department at Centro Atómico Bariloche (CNEA) and UNC Supercómputo (CCAD), which is part of SNCAD, Argentina. We thank the reviewers for their valuable feedback.

Appendix A. Numerical implementation

The explicit scheme is implemented as follows:

$$S_{i,j}^{n+1/2} = S_{i,j}^n - \beta_{i,j} S_{i,j}^n I_{i,j}^n \frac{\Delta t}{2} \quad (\text{A.1})$$

$$I_{i,j}^{n+1/2} = I_{i,j}^n + \left(\beta_{i,j} S_{i,j}^n I_{i,j}^n - \gamma_{i,j} I_{i,j}^n \right) \frac{\Delta t}{2} - \vec{v} \cdot \nabla I_{i,j}^n \frac{\Delta t}{2} \quad (\text{A.2})$$

$$R_{i,j}^{n+1/2} = R_{i,j}^n + \gamma_{i,j} I_{i,j}^n \frac{\Delta t}{2} \quad (\text{A.3})$$

where the gradient operator $\nabla I_{i,j}^n = \left(\frac{\partial I_{i,j}^n}{\partial x}, \frac{\partial I_{i,j}^n}{\partial y} \right)$ is discretized using an upwind method:

$$\frac{\partial I_{i,j}^n}{\partial x} = \frac{1}{\Delta} \begin{cases} I_{i,j}^n - I_{i-1,j}^n, & \text{if } v_{x,i,j}^n > 0 \\ I_{i+1,j}^n - I_{i,j}^n, & \text{if } v_{x,i,j}^n \leq 0 \end{cases} \quad (\text{A.4})$$

$$\frac{\partial I_{i,j}^n}{\partial y} = \frac{1}{\Delta} \begin{cases} I_{i,j}^n - I_{i,j-1}^n, & \text{if } v_{y,i,j}^n > 0 \\ I_{i,j+1}^n - I_{i,j}^n, & \text{if } v_{y,i,j}^n \leq 0 \end{cases} \quad (\text{A.5})$$

with $\Delta = \Delta x = \Delta y$. The ADI scheme is implemented as follows:

$$I_{i,j}^{n+1/2} = I_{i,j}^n + \frac{1}{2} \alpha \left(\delta_x^2 I_{i,j}^{n+1/2} + \delta_y^2 I_{i,j}^n \right) \quad (\text{A.6})$$

$$I_{i,j}^{n+1} = I_{i,j}^{n+1/2} + \frac{1}{2} \alpha \left(\delta_x^2 I_{i,j}^{n+1/2} + \delta_y^2 I_{i,j}^{n+1} \right) \quad (\text{A.7})$$

where $\alpha \equiv \frac{D\Delta t}{\Delta^2}$ and δ_x and δ_y are the discretized diffusion operators in x and y directions, respectively.

$$\delta_x^2 I_{i,j}^n = I_{i+1,j}^n - 2I_{i,j}^n + I_{i-1,j}^n \quad (\text{A.8})$$

$$\delta_y^2 I_{i,j}^n = I_{i,j+1}^n - 2I_{i,j}^n + I_{i,j-1}^n \quad (\text{A.9})$$

The resulting matrix Eqs. (A.6) and (A.7) have the advantage to be tridiagonal, so they can be solved with a Thomas algorithm (Press, 2007) for example, that is highly parallelizable and fast in GPU, differing from the Crank–Nicolson method in which the matrix is pentadiagonal and more difficult to solve. The ADI scheme also has the advantages of being second order accurate in both time and space.

The RDC model was implemented on GPUs using the CuPy library (Nishino and Loomis, 2017). This implementation enables parallelization of the simulations and significantly reduces execution time. In the code, each grid cell is updated at every time step by a CUDA kernel running a large number of parallel threads. Moreover, to support the genetic algorithm, the program can update multiple maps simultaneously in batch mode.

Appendix B. Genetic algorithm implementation

The GA we implemented was a real-coded GA (Katoch et al., 2021). The algorithm operates on a population of candidate solutions, represented as real-valued vectors. Each individual in the population corresponds to a set of parameters for the fire propagation model. The fitness of each candidate is evaluated using Eq. (5).

After calculating the fitness for each individual, we applied three genetic operators: selection, crossover, and mutation.

The selection operator aims to choose the most promising candidates (parents) to generate offspring. We used tournament selection, where a group of N candidates (three in our case) is randomly sampled, and the best one (with minimum fitness) is selected as a parent.

The crossover operator combines the genetic information of two parents to generate new offspring. We implemented a one-point crossover: a random crossover point is chosen, and the genes are exchanged between the two parents to create two new individuals.

The mutation operator introduces small random changes to an individual's genetic material to maintain diversity in the population. We applied a Gaussian mutation, in which a random value drawn from a Gaussian distribution is added to a gene with a given probability.

Data availability

Data will be made available on request.

References

- Barberá, I., Cingolani, A.M., Tiribelli, F., Mermoz, M.A., Morales, J.M., Kitzberger, T., 2025. Biotic and physical drivers of fire in northwestern Patagonia. *Fire Ecol.* 21 (1), 1–21.
- Bari, M., Presti, P., Carp, A., Lipori, M., 2022. Técnicas para el abordaje de la evaluación y predicción del comportamiento del fuego: Experiencia del equipo conformado para el incendio Steffen–Martin, Parque Nacional Nahuel Huapi, diciembre 2021 – marzo 2022. Technical Report IF-2022-37711237-APN-PNNH#APNAC, Administración de Parques Nacionales, Argentina, San Carlos de Bariloche, Río Negro, Informe técnico del Parque Nacional Nahuel Huapi, Dirección Regional Patagonia Norte, Dirección Nacional de Conservación.
- Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.
- Costafreda-Aumedes, S., Comas, C., Vega-Garcia, C., 2017. Human-caused fire occurrence modelling in perspective: a review. *Int. J. Wildland Fire* 26 (12), 983–998.
- Denham, M.M., 2009. Predicción de la evolución de los incendios forestales guiada dinámicamente por los datos (Ph.D. thesis). Universitat Autònoma de Barcelona, Barcelona.
- Denham, M.M., Laneri, K., 2018. Using efficient parallelization in graphic processing units to parameterize stochastic fire propagation models. *J. Comput. Sci.* 25, 76–88.
- Denham, M.M., Waidelich, S., Laneri, K., 2022. Visualization and modeling of forest fire propagation in Patagonia. *Environ. Model. Softw.* 158, 105526.
- Führer, E., 2000. Forest functions, ecosystem stability and management. *Forest Ecol. Manag.* 132 (1), 29–38.
- Gobierno Argentino, 2025. Servicio nacional de manejo del fuego, reporte técnico de ocurrencia. URL: <https://www.argentina.gob.ar/seguridad/servicio-nacional-de-manejo-del-fuego/evaluacion-de-peligro-y-alerta-temprana/reporte>. (Accessed in April 2025).
- Gobierno de Río Negro, 2025. Incendio en los manzanos: un trabajo constante en condiciones adversas. URL: <https://rionegro.gov.ar/articulo/52573/incendio-en-los-manzanos-un-trabajo-constante-en-condiciones-adversas>. Fecha: 21 de enero de 2025.
- Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on typical tabular data? *Adv. Neural Inf. Process. Syst.* 35, 507–520.
- Hoover, K., Hanson, L.A., 2021. Wildfire Statistics. Technical Report, Congressional Research Service.
- Jellouli, O., Bernoussi, A.S., 2022. The impact of dynamic wind flow behavior on forest fire spread using cellular automata: application to the watershed BOUKHALEF (Morocco). *Ecol. Model.* 468, 109938.
- Katoch, S., Chauhan, S.S., Kumar, V., 2021. A review on genetic algorithm: past, present, and future. *Multimedia Tools Appl.* 80 (5), 8091–8126.
- Kelso, J.K., Mellor, D., Murphy, M.E., Milne, G.J., 2015. Techniques for evaluating wild-fire simulators via the simulation of historical fires using the australis simulator. *Int. J. Wildland Fire* 24 (6), 784–797.
- Kitzberger, T., Bürgesser, R.E., 2025. A novel fire regime driven by increased lightning activity and lightning ignition efficiency for northwestern Patagonia, Argentina. *Int. J. Wildland Fire* 34 (9).
- Kitzberger, T., Tiribelli, F., Barberá, I., Gowda, J.H., Morales, J.M., Zalazar, L., Paritsis, J., 2022. Projections of fire probability and ecosystem vulnerability under 21st century climate across a trans-Andean productivity gradient in Patagonia. *Sci. Total Environ.* 839, 156303.
- Kolaitis, D.I., Pallikarakis, C., Founti, M.A., 2023. Comparative assessment of wildland fire rate of spread models: effects of wind velocity. *Fire* 6 (5), 188.
- Kolton, A.B., Laneri, K., 2019. Rough infection fronts in a random medium. *Eur. Phys. J. B* 92 (6), <http://dx.doi.org/10.1140/epjb/e2019-90582-3>.
- Laneri, K.F., Waidelich, S., Zimmerman, V.B., Denham, M.M., 2020. First steps towards a dynamical model for forest fire behaviour in Argentinian landscapes. *J. Comput. Sci. Technol.*
- Li, F., Zhu, Q., Yuan, K., Ji, F., Paul, A., Lee, P., Radeloff, V.C., Chen, M., 2024. Projecting large fires in the western US with an interpretable and accurate hybrid machine learning method. *Earth's Futur.* 12 (10), e2024EF004588.
- Morales, J.M., Mermoz, M., Gowda, J.H., Kitzberger, T., 2015. A stochastic fire spread model for north patagonia based on fire occurrence maps. *Ecol. Model.* 300, 73–80.
- Murray, J.D., 2003. *Mathematical biology II: Spatial models and biomedical applications*, third ed. In: *Interdisciplinary Applied Mathematics*, vol. 18, Springer, New York, <http://dx.doi.org/10.1007/b98869>.
- Nishino, R., Loomis, S.H.C., 2017. Cupy: A numpy-compatible library for nvidia gpu calculations. In: 31st Conference on Neural Information Processing Systems, vol. 151, (7).
- Press, W.H., 2007. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press.
- Provatás, N., Ala-Nissila, T., Grant, M., Elder, K.R., Piché, L., 1995. Scaling, propagation, and kinetic roughening of flame fronts in random media. *J. Stat. Phys.* 81 (3), 737–759. <http://dx.doi.org/10.1007/BF02179255>.
- Ritchie, H., 2021. Forest area. Our World Data URL: <https://ourworldindata.org/forest-area>. Accessed in August 2025.
- Rothermel, R.C., 1972. *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*, vol. 115, Intermountain Forest & Range Experiment Station, Forest Service, US.
- Shwartz-Ziv, R., Armon, A., 2022. Tabular data: Deep learning is not all you need. *Inf. Fusion* 81, 84–90.
- Syphard, A.D., Keeley, J.E., 2015. Location, timing and extent of wildfire vary by cause of ignition. *Int. J. Wildland Fire* 24 (1), 37–47.
- US Forest Service Research & Development, 2025. WindNinja: a computer program that computes spatially varying wind fields for wildland fire application. URL: <https://research.fs.usda.gov/firelab/products/dataandtools/windninja>. Release Date June 2 2025; Last updated August 8 2025.
- Wadhvani, R., Zhang, X., Li, Y., Sutherland, D., Moinuddin, K., Huang, X., 2025. Integrating deep learning with physics-based model for predicting grassfire spread: R. Wadhvani et al. *J. For. Res.* 36 (1), 140.
- Yu, W., Ghosh, A., Finn, T.S., Arcucci, R., Bocquet, M., Cheng, S., 2025. A probabilistic approach to wildfire spread prediction using a denoising diffusion surrogate model. arXiv preprint arXiv:2507.00761.
- Zagarrá, R., Laneri, K., Kolton, A.B., 2024. Infection fronts in randomly varying transmission-rate media. *Phys. Rev. E* 110 (3), <http://dx.doi.org/10.1103/physreve.110.034308>.