

NASA

Sensitivity of simulated wildfire spread to parameter uncertainty and spatial resolution of terrain and fuel maps

An ELMFIRE perturbed parameter ensemble of the 2020 Creek Fire, California

Authors:

MASON LEE¹, DR. NANCY KIANG², DR. ROBERT FEILD², DR. KEREN MEZUMAN², DR. MARCUS VAN LIER-WALQUI²

¹Brown University

²NASA Goddard Institute for Space Studies

Abstract

Contents

1	Introduction	2
2	Research Questions	3
3	Methods	3
4	Experimental Design	7
5	Results	11
6	Appendix	11

List of Figures

1	This figure contrasts two columns, each corresponding to increasing maximum spotting distances within the ELMFIRE wildfire simulation. The role of crown fire initiation in spotting dynamics is pronounced, with sensitivity closely linked to ELMFIRE's grid scale. Notably, at a 30-meter grid, a significant surge in fire spread becomes evident, highlighting the impact of spatial resolution.	12
2	(top subplot) This figure presents the analysis of IOU score distribution within clusters obtained through PCA-KMeans. A 9-cluster solution was used, and each subplot displays the IOU score histogram along with a Gaussian Mixture model fit, revealing distinct patterns in IOU score frequencies across the clusters. (bottom subplot) The PCA transformation with 17 components is computed for dimensionality reduction, followed by an analysis of the "elbow" point in the inertia versus number of clusters plot, helping to determine an optimal number of clusters for subsequent analysis.	13
3	This figure showcases a comparative analysis of fire behavior parameters, contrasting scenarios with and without spotting enabled. Histograms of selected parameters are presented for both cases, where blue represents 'ENABLE_SPOTTING = TRUE' and red represents 'ENABLE_SPOTTING = FALSE'. The analysis provides insights into how parameter distributions differ when spotting is enabled or disabled in fire simulations.	14
4	The figure consists of 3 subplots that explore the relationship between specific parameters and the 'Wall clock time' of a simulation. The first subplot presents a linear regression analysis for 'Fire volume (ac-ft)', showing a red line depicting the fitted linear model. The second subplot utilizes a logarithmic regression approach to analyze the impact of the 'Number of Simulated Embers' on 'Wall clock time', with a red curve illustrating the fitted log regression. The third subplot showcases the application of a logarithmic fit to model the correlation between 'Spotting Distance' and 'Total Fire Area'. The red curve represents the fitted logarithmic function, while a marked point signifies the location of maximum curvature, indicating a key threshold in the relationship.	15
5	Description for decreasing_resolution.	15
6	Description for fbfm40_creek.	15
7	Description for gam-splines.	16
8	Description for interesting_runs.	16
9	The depicted figure unveils discernible groupings denoting analogous fire behaviors in the parameter space. Within this portrayal, the top 100 best-performing runs are symbolized by 'x' markers, with color gradations corresponding to IOU metrics. This specific subset of the PCA component space reveals intriguing clusters that exhibit certain patterns, although their complete distinction remains somewhat obscured. The figure offers a preliminary insight into potential relationships and trends within the data, inviting further investigation into the underlying factors driving fire behavior variations.	17

10	On the left, Principal Component Analysis (PCA) was executed on a 17-component parameter space, and the resulting plot showcases the first two PCA components in relation to each other. When color-coded by the presence of enabled spotting, the plot compellingly delineates distinct clusters, underscoring spotting's pivotal role as a primary influencer within the parameter space. The right pane provides insight into the eigenvalue distribution of the 17 PCA components. The initial component stands out with the highest eigenvalue, followed by an exponential decline in subsequent components. This figure not only sheds light on the parameter space structure but also accentuates the dominance of spotting as a significant determinant in the analysis.	18
11	The top 20 wildfire simulations ranked by Intersection over Union (IOU) metrics, accompanied by the VIIRS fire perimeters validated on the fourth day (September 9, 2020, 12:00). The experiment IDs are cross-referenced with the "best20,uns.csv" dataset, housing optimal parameters for the Creek Fire.	
12	Figure 1 displays the VIIRS instrument data from the Suomi NPP satellite over the first four simulation days, alongside results from the ELMFIRE spread model. The figure includes raw VIIRS hotspots shown as black dots, VIIRS fire perimeters based on Chen et al. (2020), and NIROPS perimeters where available. This illustration effectively showcases the integration of VIIRS data and ELMFIRE modeling for analyzing wildfire progression	20
13	Description for wind_comp_analysis.	20

List of Tables

1	Spotting Parameters	5
2	Environmental Variables	6
3	Experimental Design Parameter Bounds	9

1 Introduction

Wildfires remain one of the most destructive and complex natural phenomena, having significant impacts on ecology, climate, and human livelihood. Understanding their dynamics and drivers is critical for both predicting fire behavior and managing fire risk. In this context, we explore various facets of fire dynamics and emphasize the case study of the Creek Fire.

1.1 The *pyrE* Global Fire Module

The interactive global fire module, *pyrE* (v1.0), developed by NASA's Goddard Institute for Space Studies, is an essential tool for understanding the spread of wildfires [?]. It enables the simulation of natural and anthropogenic drivers, feedbacks, and interactions of open fires, with a focus on the reactive gases and particles affecting the Earth's radiation balance. Recent results reveal a bias in *pyrE*'s fire emissions, underestimating them by 32%–42%, thereby influencing atmospheric composition studies and necessitating further refinement.

1.2 Vegetation and Wildfires

The coupling of *pyrE* with the vegetation and soil module, Ent, offers a way to explore the significance of vegetation characteristics beyond mere functional types. It allows for the modeling of leaf biophysics, allometry, canopy structure, and phenology. These details enable a more nuanced understanding of how meteorological drivers, terrain, and fuel load affect fire spread, burn area, and plant mortality, even at the global climate model's coarse spatial scales.

1.3 Understanding Mega-Fire Drivers

The drivers of large plume-driven mega-fires require intricate modeling, taking into account the interactivity and sensitivity of various parameters. For instance, pyrocumulonimbus clouds (pyroCBs) and their intricate physics must be accurately simulated. The necessity to design models that super-parameterize complex effects like terrain slope and internally generated wind illustrates the challenge of encapsulating fire physics within a global model.

1.4 The Creek Fire: A Case Study

The Creek Fire, starting on September 4, 2020, in California, exemplifies the destructiveness of plume-driven fires, burning 379,895 acres and becoming the fifth-largest wildfire in modern California history ⁷. With over \$193 million spent in fire suppression costs and massive property damage, this fire presents a compelling case for study.

1.5 Critical Review of Stephens et al.

The work of Stephens et al. 2022 ⁷ brought focus to tree mortality's significant impact on the Creek Fire behavior but faced limitations in reproducing the plume-driven spread accurately. Their effort underlines the complex interplay between various factors influencing wildfire dynamics.

2 Research Questions

This study aims to answer the following questions:

- What are the essential drivers for high-resolution fire modeling, and how do they interact? Notes: scaling physics for the GCM, impact of resolution.
- How can the NASA GCM incorporate detailed vegetation architecture and detailed fire physics to more accurately simulate fire behavior?
- How can the insights from the Creek Fire contribute to understanding large plume-driven fires?
- What lessons can be drawn from the Stephens et al. study, and what are the next steps in accurately modeling wildfires?

3 Methods

3.1 Wildfire Model: ELMFIRE

ELMFIRE is a fire behavior model that leverages Eulerian level set techniques to numerically trace the evolution of fire fronts on a computational grid. Its core methodology involves solving hyperbolic partial differential equations to establish the position of the fire front. The system derives fire spread rates by integrating the Rothermel surface fire spread model, Cruz's crown fire model, and Richards' elliptical dimensions. It calculates the spread rate in diverse directions around the fire perimeter by considering the varying elliptical dimensions at different points. The approach factors in wind and slope effects, estimating the length-to-width ratio for each elliptical "wavelet" segment of the fire front. ELMFIRE also links crown and surface fire behavior by modifying the wind factor and establishing a connection between the crown fire model and the surface fire spread formulation.

3.1.1 Spread Rate Formulation

- **Basics of Eulerian level set methods:** Eulerian level set methods, including that in ELMFIRE, involve numerically solving the following hyperbolic partial differential equation for the scalar variable ϕ :

$$\frac{\partial \phi}{\partial t} + V \cdot \nabla \phi = 0$$

ϕ has no physical meaning except that the $\phi = 0$ isopleth corresponds to the fire front. This provides a convenient way to track a curved surface, such as a fire front, on a regular grid. ELMFIRE integrates the governing PDE using a narrow-band formulation with a second-order Runge-Kutta method and superbee flux limiters to prevent numerical oscillations.

3.2 ELMFIRE Simulations: Computational Methods and Analysis

- **Simulation Environment:** NASA Local Server (Hammer): Parallelization across cores was implemented using GNU parallel and the Message Passing Interface (MPI), enabling dedicated multi-threading for efficient computation. HPC System (DISCOVER): We used the supercomputer High-Performance Computing (HPC) system called DISCOVER. Parallel processing was executed with the SLURM job manager to optimize performance.

- **Software and Languages:** Geospatial analysis was conducted using QGIS and Python. The wildfire simulation core was constructed in Fortran, with command-line operations handled through BASH script.

3.3 Spotting Physics and Parameters

Spotting is a critical mode of fire propagation, particularly under conditions of high temperature, dryness, and wind. The following parameters and their interrelationships define the behavior of spotting within the modeling system.

3.3.1 Spotting Configuration

The SPOTTING namelist group outlines the configuration of spotting and can be enabled or disabled with the ENABLE_SPOTTING parameter. A sample spotting configuration is as follows:

```
&SPOTTING
ENABLE_SPOTTING = .TRUE.
CROWN_FIRE_SPOTTING_PERCENT = 1.0
...
NMEMBERS_MAX = 1
PIGN = 100.0
/
```

3.3.2 Spotting Distance

Spotting distance is modeled as a lognormal distribution where the mean and standard deviation are determined semi-empirically as functions of ambient wind speed and fireline intensity. This distance is controlled by several parameters, including:

- **MEAN_SPOTTING_DIST:** Mean spotting distance
- **SPOT_FLIN_EXP:** Fireline intensity exponent
- **SPOT_WS_EXP:** Wind speed exponent
- **NORMALIZED_SPOTTING_DIST_VARIANCE:** Variance of the normalized spotting distance

These parameters can be expressed as:

$$\text{Mean spotting distance}(\mu) = a \times I^b \times W^c \quad (1)$$

$$\text{Variance of spotting distance}(\sigma^2) = d \times I^b \times W^c \quad (2)$$

3.3.3 Spotting Mechanism

The spotting mechanism takes into account several additional parameters, including:

- **CROWN_FIRE_SPOTTING_PERCENT:** Controls spotting initiation from passive/active crown fire pixels.
- **NMEMBERS_MIN** and **NMEMBERS_MAX:** Constrain the number of embers generated.
- **PIGN:** Probability that an ember initiates a spot fire.

3.3.4 Surface Fire Spotting

Surface fire spotting can be controlled through the following parameters:

- **ENABLE_SURFACE_FIRE_SPOTTING:** Enable/disable spotting from surface fire.
- **SURFACE_FIRE_SPOTTING_PERCENT:** Controlled by fuel model number.
- **CRITICAL_SPOTTING_FIRELINE_INTENSITY:** Intensity below which surface fire spotting does not occur.

3.3.5 Stochastic Spotting and Calibration

Stochastic spotting can be enabled by setting STOCHASTIC_SPOTTING = .TRUE. This allows for random generation of parameters within user-specified ranges, providing more realistic modeling of fire progression. For fire hindcasts, these parameters can be automatically adjusted to optimize agreement between modeled and observed fire perimeters.

3.4 Variable Definitions

A comprehensive list of variable definitions and their default values, ranges, and explanations is provided in Table 1.

Table 1: Spotting Parameters

Variable Name	Default Value	Explanation
ENABLE_SPOTTING	.TRUE.	Enable/disable spotting fire propagation.
ENABLE_SURFACE_FIRE_SPOTTING	.TRUE.	Enable/disable spotting from surface fire.
SURFACE_FIRE_SPOTTING_PERCENT(:)	1.0	Controls the percentage of surface fire that can cause spotting, based on the fuel model number.
CRITICAL_SPOTTING_FIRELINE_INTENSITY	2000.0	Fireline intensity (in kW/m) below which surface fire spotting does not occur.
SPOTTING_TYPE	N/A	Not explicitly explained, assumed to relate to the type of spotting model used.
SPOTTING_DISTRIBUTION_TYPE	'LOGNORMAL'	The type of distribution spotting distance follows, currently only lognormal is supported.
MEAN_SPOTTING_DIST_MIN	5	Min mean spotting distance, related to fireline intensity and wind speed.
MEAN_SPOTTING_DIST_MAX	15	Max mean spotting distance, related to fireline intensity and wind speed.
NORMALIZED_SPOTTING_DIST_VARIANCE_MIN		Min variance of the normalized spotting distance ~ fireline intensity and wind speed.
NORMALIZED_SPOTTING_DIST_VARIANCE_MAX		Max variance of the normalized spotting distance ~ fireline intensity and wind speed.
SPOT_WS_EXP_LO	0.4	Low bounds of the empirical wind speed exponent in the spotting distance formula.
SPOT_WS_EXP_HI	0.7	High bounds of the empirical wind speed exponent in the spotting distance formula.
SPOT_FLIN_EXP_LO	0.2	Low bounds of the empirical fireline intensity exponent in the spotting distance formula.
SPOT_FLIN_EXP_HI	0.4	High bounds of the empirical fireline intensity exponent in the spotting distance formula.
NMEMBERS_MIN	1	Min bounds of the number of embers generated from a crown fire pixel that can cause spotting.
NMEMBERS_MAX_LO	1	Low bounds of the number of embers generated from a crown fire pixel that can cause spotting.
NMEMBERS_MAX_HI	1	High bounds of the number of embers generated from a crown fire pixel that can cause spotting.
GLOBAL_SURFACE_FIRE_SPOTTING_PERCENT_MIN	0.2	Min global surface fire spotting percentage that can cause spotting, applied to all fuel models.
GLOBAL_SURFACE_FIRE_SPOTTING_PERCENT_MAX	0.8	Max global surface fire spotting percentage that can cause spotting, applied to all fuel models.
CROWN_FIRE_SPOTTING_PERCENT_MIN	0.2	Min percentage of crown fire that can cause spotting.
CROWN_FIRE_SPOTTING_PERCENT_MAX	0.8	Max percentage of crown fire that can cause spotting.
PIGN_MIN	100.0	Min probability in percent that an ember, once landed, initiates a spot fire.
PIGN_MAX	100.0	Max probability in percent that an ember, once landed, initiates a spot fire.

3.5 Environmental Variables in Rothermel Fire Spread Model

3.5.1 Influence of Environmental Variables

The Rothermel fire spread model incorporates various environmental variables to simulate fire behavior. These variables, along with associated equations, are described below:

- **Fuel Moisture Content (FMC):** Fuel moisture contents including $M1$, $M10$, $M100$, MLH , and MLW are key parameters. They influence the reaction intensity (IR), and are usually determined from field observations or remote sensing data.

Table 2: Environmental Variables

Variable	Description	Units
ADJ	Spread rate adjustment factor	-
CBD	Canopy bulk density	kg/m ³
CBH	Canopy base height	m
CC	Canopy cover	-
CH	Canopy height	m
FMC	Foliar moisture content	-
M1	1-hour fuel moisture	-
M10	10-hour fuel moisture	-
M100	100-hour fuel moisture	-
MLH	Live herbaceous fuel moisture	-
MLW	Live woody fuel moisture	-
WAF	Wind adjustment factor	-
WD	Wind direction	deg
WS	Wind speed	mph
DEM	Digital Elevation Map	Feet
ASP	Aspect is the orientation of slope, measured clockwise in degrees	degrees

- **Wind Factors:** Wind adjustment factor (WAF), wind direction (WD), and wind speed (WS) are used in the calculation of the wind factor ϕ_w in the spread rate equation:

$$Vs; Rothermel = \frac{IR\xi(1 + \phi_w + \phi_s)}{\rho_b \epsilon Qig}$$

- **Topographical Variables:** DEM (Digital Elevation Map) and Aspect (ASP) are used to calculate the slope factor ϕ_s and the orientation of slope. They influence both the direction and rate of fire spread as shown in Eq. (14) and Eq. (15).
- **Canopy Variables:** Canopy parameters like CBD (Canopy bulk density), CBH (Canopy base height), CC (Canopy cover), and CH (Canopy height) are essential in determining the critical fireline intensity (I0) for torching:

$$I0 = (0.01 \cdot CBH \cdot (460 + 26 \cdot M))^{3/2}$$

- **Additional Factors:** Other factors like spread rate adjustment factor (ADJ) and wind azimuth θ_w further refine the model's predictions and accommodate user-defined adjustments.

3.6 Application in Surface Fire Spread Rate

Rothermel's model offers a detailed understanding of surface fire spread rate, focusing on the direction and magnitude of fire spread in various terrains. By leveraging the above variables, the model can accurately simulate complex fire behaviors.

3.6.1 Surface Fire Acceleration

Equation (12) shows how the spread rate increases from zero to equilibrium, considering the time of ignition (t_{ign}) and an acceleration time constant (τ_{accel}):

$$\frac{Vs}{Vs; Rothermel} = 1 - \exp \frac{-t_{ign}}{\tau_{accel}}$$

3.6.2 Spread Rate Normal to Fire Front

Equations (13), (14), and (15) take into account the wind and slope factors to calculate the magnitude of surface fire spread rate at different angles. The derivation considers the vector representation of wind-induced spread rate and the dot product with fire front's unit normal vector.

4 Experimental Design

4.1 Sampling Methods for Parameter Space Exploration

4.1.1 Latin Hypercube Sampling (LHS)

Latin Hypercube Sampling is a stratified, multidimensional sampling method, which has been used to efficiently explore the model's parameter space.

- **Design Principle:** LHS divides each parameter's range into N equally probable intervals. A value is randomly selected from each interval, and the N values from each parameter are paired randomly with N values from other parameters. This ensures a comprehensive coverage of the parameter space.
- **Benefits:** LHS reduces sampling bias and improves the representativeness of the sampled points. It is particularly useful for high-dimensional problems where traditional sampling may be inefficient.
- **Application:** In our study, LHS was applied to perform 600 runs, allowing for a robust exploration of parameter combinations without the need for an exhaustive search.

4.1.2 Sobol Sequence Sampling

Sobol Sequence Sampling is a method for generating quasi-random low-discrepancy sequences.

- **Design Principle:** Sobol sequences utilize base-2 van der Corput sequences to create point sets that are uniformly distributed. They are generated using direction numbers and ensure that the points are well-spaced over the entire range.
- **Benefits:** Sobol sequences offer a more uniform distribution of points compared to random sampling, leading to a more efficient estimation of integrals or exploring parameter spaces.
- **Application:** In our research, 3700 runs were performed using Sobol Sequence Sampling. This method was chosen to enhance the precision of the simulation by ensuring a better distribution of sampled points across the parameter space.

4.2 Computational Resources

- **NASA Local Server (Hammer):** Parallelization across cores was implemented using GNU parallel and the Message Passing Interface (MPI), enabling dedicated multi-threading for efficient computation.
- **HPC System (DISCOVER):** We utilized a supercomputer High-Performance Computing (HPC) system called DISCOVER. Parallel processing was executed with the SLURM job manager to optimize performance.
- **Software and Languages:** Geospatial analysis was conducted using QGIS and Python. The wildfire simulation core was constructed in Fortran, with command-line operations handled through BASH script.

4.3 Remote Sensing Datasets for Fire Spread Validation

4.3.1 Fire Perimeter Data

- **SUOMI NPP VIIRS:** The Visible Infrared Imaging Radiometer Suite (VIIRS) instrument offers a 375m resolution and operates in a geostationary orbit, providing half-daily observations. This frequency ensures timely updates, critical for effective monitoring and response to rapidly evolving wildfires.
- **NIROPS (National Infrared Operations):** NIROPS generates detailed fire perimeters, utilizing infrared imaging onboard aircraft. These precise insights into the fire size and spread support targeted response strategies. Typically, daily perimeters are available into the second day of an incident, aiding immediate assessment.

- **Chen et al. 2020 VIIRS-based Fire Perimeters:** This method converts raw VIIRS hotspots into fire perimeters using a clustering algorithm, with updates twice daily. The transformation from raw data to perimeters enhances the accuracy of mapping the actual fire extent, which is pivotal for both real-time control and post-incident analysis.

4.3.2 Fuel Data

- **LANDFIRE 2020:** Incorporates 40 Scott and Burgan Fire Behavior Fuel Models (FBFM40) to represent the complex distributions of fuel loading among various surface fuel components (live and dead), size classes, and fuel types. This data is essential for modeling the potential fire behavior.

4.3.3 Wind and Weather Data

- **RTMA:** RTMA (Real-Time Mesoscale Analysis) offers spatiotemporal data at 2.5 km resolution, updated hourly. This continuous monitoring supports precise modeling of the fire's response to changing weather conditions.
- **Wind Ninja:** Wind Ninja employs NOAA's High-Resolution Rapid Refresh (HRRR) model to calculate winds at a microscale level, down to resolutions as fine as 100 meters. This takes into account the terrain, land cover, and atmospheric stability, providing a nuanced understanding of local wind patterns that directly influence fire spread.

Importance of Perimeter Data and Frequent Observations Perimeter data, as opposed to raw hotspots, provide a more accurate depiction of the fire's boundary and dynamic changes. These defined edges are vital for monitoring control efforts, predicting future spread, and historical analysis. Frequent observations, as exemplified by VIIRS, allow for near real-time tracking, facilitating effective management, and intervention, which is especially crucial during the initial stages of a wildfire incident.

4.4 Fire Behavior Characteristics

4.4.1 Intersection Over Union (IOU) Coefficient

The Intersection Over Union (IOU) coefficient, often used in deep learning computer vision tasks, is a measure to quantify the similarity between two shapes. In the context of fire spread analysis, it is utilized to compare the actual and predicted fire perimeters. The IOU coefficient is defined as:

$$\text{IOU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{A \cap B}{A \cup B} \quad (3)$$

Where A and B are the areas of the actual and predicted fire perimeters, respectively. The IOU's relevance to fire spread modeling is rooted in its ability to offer a quantitative and intuitive measure of the overlap between real and simulated fires ?. This enables more precise assessments of model accuracy, aiding in validation and calibration efforts.

4.4.2 Frequency and Metrics of Calculation

- **IOU Calculation:** IOU coefficients were calculated twice daily, using Chen et al. 2020 perimeters, to continuously monitor the model's performance in predicting fire spread.
- **Fire Size Tracking:** In parallel with IOU, the fire size was also tracked to provide complementary insights into the dynamic behavior of the fire.
- **Additional Metrics:** The study also monitored the Wall-Clock time, Crown Fire Area, Fire Volume, and Total Fire Area, Fire Line Intensity, and Fire Accelerations, encompassing a comprehensive understanding of various facets of fire behavior and spread.

4.4.3 Resampling and Remapping Inputs

In our study, the Geospatial Data Abstraction Library (GDAL) was utilized to resample and remap inputs into different resolutions. We ran ELMFIRE simulations at increasingly coarse resolution to understand the effects of GCM-scale physics and environmental drivers.

4.4.4 WindNinja Wind Data

WindNinja is a computer program specifically designed to compute spatially-varying wind fields in complex terrains. The program is known for its:

- Fast simulation times
- Low computational requirements
- Capability to run at high spatial resolution
- Adjustments for terrain and some thermal energy parameterizations
- Assumptions like no moisture/latent heat effects, steady-state condition, and optional conservation of momentum

WindNinja's process involves building a 3-D model of the air, prescribing an initial wind field, and computing the final mass/momenta-conserving wind field.

4.4.5 LHS + Sobol Fire Behavior Parameters

The following table enumerates the specific parameters, ranges, and additional factors employed in our study. The following parameter bounds were used in both the Sobol and LHS sampling. :

Table 3: Experimental Design Parameter Bounds

Parameter	Range
Surface Fire Spotting Percent	0.0 - 1.0%
Critical Spotting Fireline Intensity	0 - 6000 kW/m
Mean Spotting Distance Max	15 - 2000 m
Spot Wind Speed Exponent High	0.1 - 1.0
Spot Fireline Intensity Exponent High	0.1 - 1.0
Max Number of Embers High	1-6 Embers
Global Surface Fire Spotting Percent Max	0.1 - 1.0%
Crown Fire Spotting Percent Max	0.1 - 1.0%
Probability of Ignition Max	1 - 100.0%
Crown Fire Model	0 - 1.0 Boolean
Wind Speed	0 - 20 m/s
Wind Direction	-360 - 360 deg
Canopy Cover	±0.10 fraction
Crown Base Height	±5.0 m
Canopy Height	±5.0 m
Canopy Bulk Density	±0.10 kg/m ³
1-Hour Fuel Moisture	±0.10 moisture%
10-Hour Fuel Moisture	±0.10 moisture%
1000-Hour Fuel Moisture	±0.10 moisture%
Additional Factors	
Input Resolution	30m, 500m, 1000m, 6km, 0.25 deg, 0.5 deg
Input Wind	Wind Ninja, RTMA
Special Cases	Flat Terrain, Avg Wind

4.5 Statistical Analysis for Understanding Parameter Importance

4.5.1 ANOVA (Analysis of Variance)

ANOVA is a statistical method used to analyze the differences among group means by comparing the variability between groups to the variability within groups. It helps in making parsimonious models by identifying significant factors.

$$F = \frac{\text{Between-group variability}}{\text{Within-group variability}} \quad (4)$$

4.5.2 Generalized Additive Models (GAMs)

Introduction and Equations GAMs provide a generalized linear model structure that allows for the nonlinear relationship between dependent and independent variables. The model is given by:

$$g(E(y)) = \alpha + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) \quad (5)$$

where g is the link function, f_i are smooth functions, and x_i are predictors.

AICC Scores AICC (Corrected Akaike Information Criterion) scores provide a means to compare models with different numbers of parameters, penalizing for overfitting:

$$\text{AICC} = -2 \log(\mathcal{L}) + 2k + \frac{2k(k+1)}{n-k-1} \quad (6)$$

where \mathcal{L} is the likelihood, k is the number of parameters, and n is the sample size.

Interaction Effects In our experiments with Generalized Additive Models, both linear and spline basis functions were utilized. The pyGAM library in Python was employed ?, with 20 splines for modeling. Interaction terms were implemented in pyGAM using tensor products, drawing intuition and fire science to select a small set of interaction terms. ANOVA was subsequently conducted on the transformed GAM variables to find which parameter sets explained the most variance. The interactions in this study were identified as:

- Expect these are not independent: cc:cbd, ch:cbd, cbh:cbd, ch:cbh, cc:ch:cbd
- Expect these are interactive: GLOBAL_SURFACE_FIRE_SPOTTING_PERCENT_MAX:cc, GLOBAL_SURFACE_FIRE_SPOTTING_PERCENT_MAX:crown, CROWN_FIRE_SPOTTING_PERCENT_MAX:cc, CROWN_FIRE_SPOTTING_PERCENT_MAX:ch

4.6 Sobol's Sensitivity Analysis

Sobol's method is used to perform global sensitivity analysis, providing a means to quantify the contribution of each input variable to the variance of the output. The method handles non-linearities by decomposing the total variance into components. The decomposition is described as follows:

4.6.1 Decomposition

The model output, Y , is represented as a sum of functions of the input variables:

$$Y = f_0 + \sum_{i=1}^n f_i(X_i) + \sum_{i < j} f_{ij}(X_i, X_j) + \dots + f_{1,2,\dots,n}(X_1, X_2, \dots, X_n) \quad (7)$$

where:

- f_0 is the constant term,
- f_i are the first-order terms,
- f_{ij} are the second-order interaction terms, and so on.

4.6.2 Sensitivity Indices

- **First-order index (S_i):** Measures the contribution of an individual variable X_i to the variance of Y , ignoring its interactions with other variables. Calculated as:

$$S_i = \frac{V(f_i(X_i))}{V(Y)} \quad (8)$$

- **Second-order index (S_{ij}):** Captures the interaction between variables X_i and X_j . Calculated as:

$$S_{ij} = \frac{V(f_{ij}(X_i, X_j))}{V(Y)} - S_i - S_j \quad (9)$$

- **Total-effect index (S_{Ti}):** Measures the total effect of variable X_i on the output, considering all interactions. Calculated as:

$$S_{Ti} = 1 - \frac{V(Y|X_{-i})}{V(Y)} \quad (10)$$

where X_{-i} represents all variables except X_i .

These indices provide a complete picture of the sensitivity of the model, taking into account not only the main effects but also higher-order interactions. They allow for a comprehensive understanding of the non-linear relationships between input variables and the model output.

Sobol's method is applied to understand the relative importance of input parameters. It doesn't account for aleatory uncertainty or interactions beyond pairwise interactions. Sobol relies on a Sobol Sequence sampling strategy, unlike LHS (Latin Hypercube Sampling) which focuses on uniformly covering the parameter space. We derived the 1st order, 2nd order, and total interactions from this study. The sensitivity analysis specifically focused on the total fire area, rather than the Intersection Over Union (IOU) of the fire, providing insights into the dominant factors influencing the overall extent of the fire spread.

4.6.3 Statistical Correlations

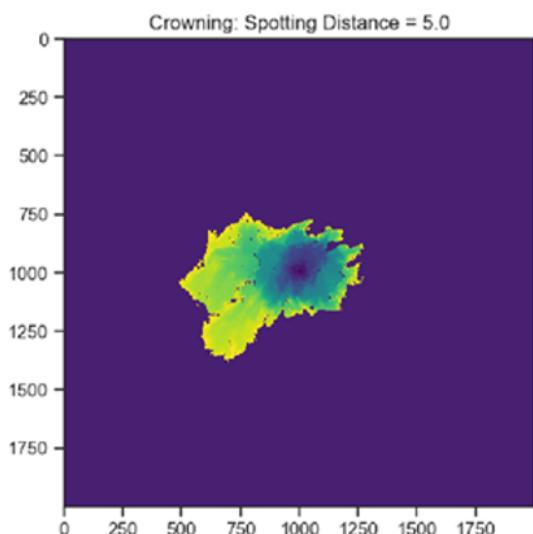
Different correlation measures like Pearson correlation, Chi-Square, and Spearman Sensitivity were used to study linear and nonlinear associations between variables.

- **Correlation:** Measures linear relationship
- **Chi-Square:** Tests for independence between categorical variables
- **Spearman Sensitivity:** Assesses monotonic relationship

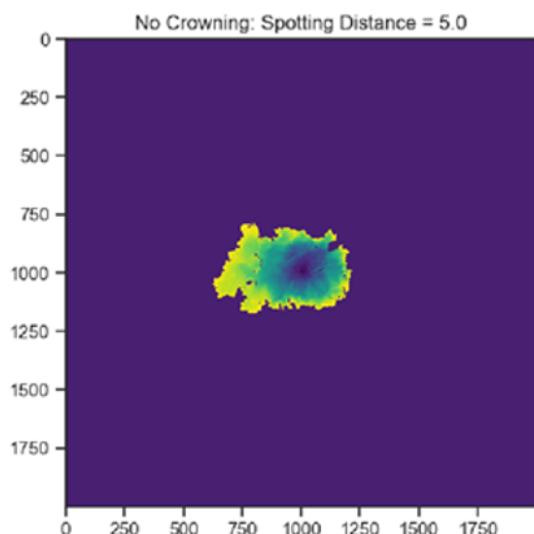
5 Results

6 Appendix

Crowning (On)

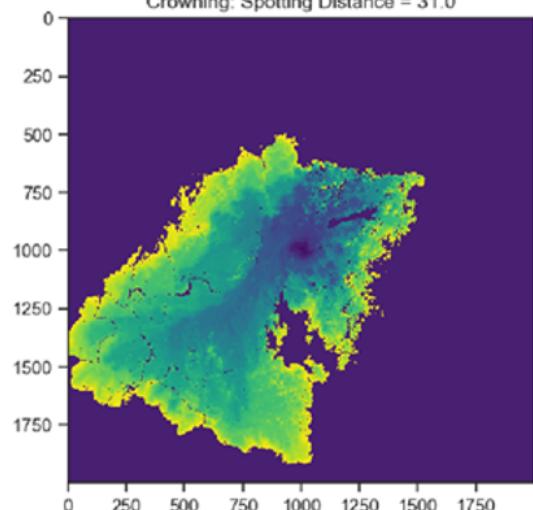


Crowning (Off)

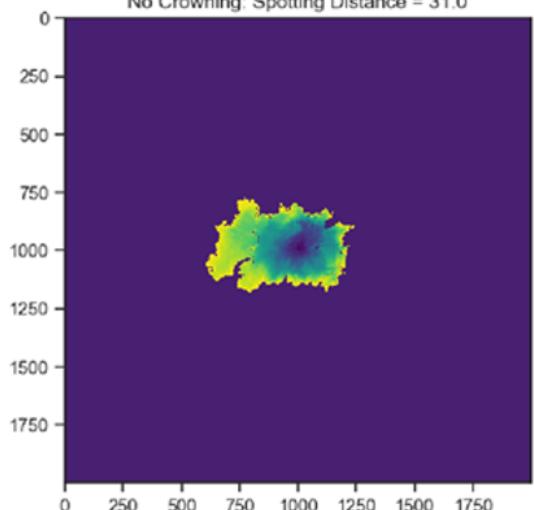


5 Meter Spotting

Crowning: Spotting Distance = 31.0

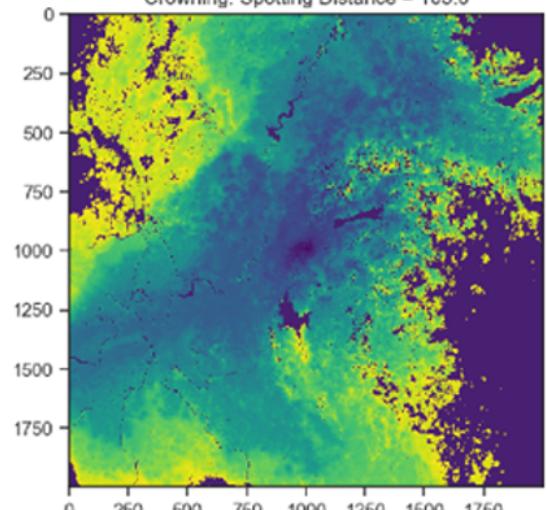


No Crowning: Spotting Distance = 31.0

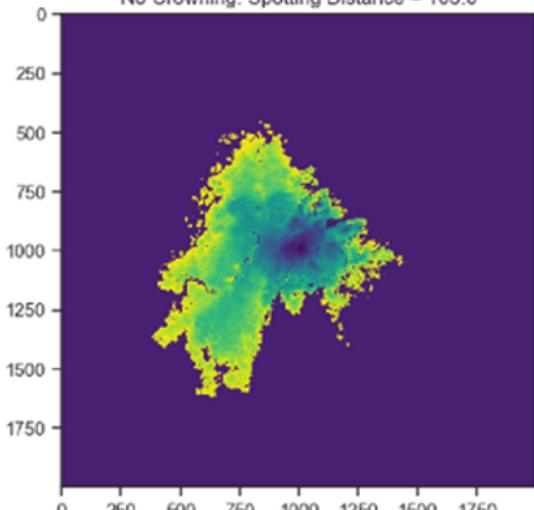


31 Meter Spotting

Crowning: Spotting Distance = 105.0

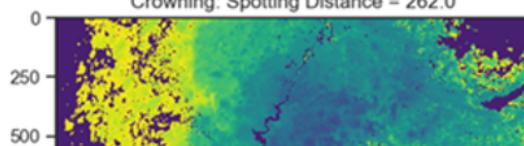


No Crowning: Spotting Distance = 105.0

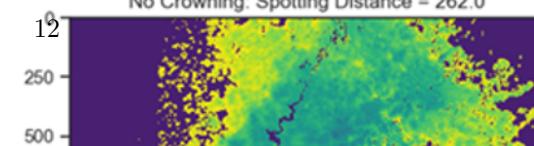


105 Meter Spotting

Crowning: Spotting Distance = 262.0

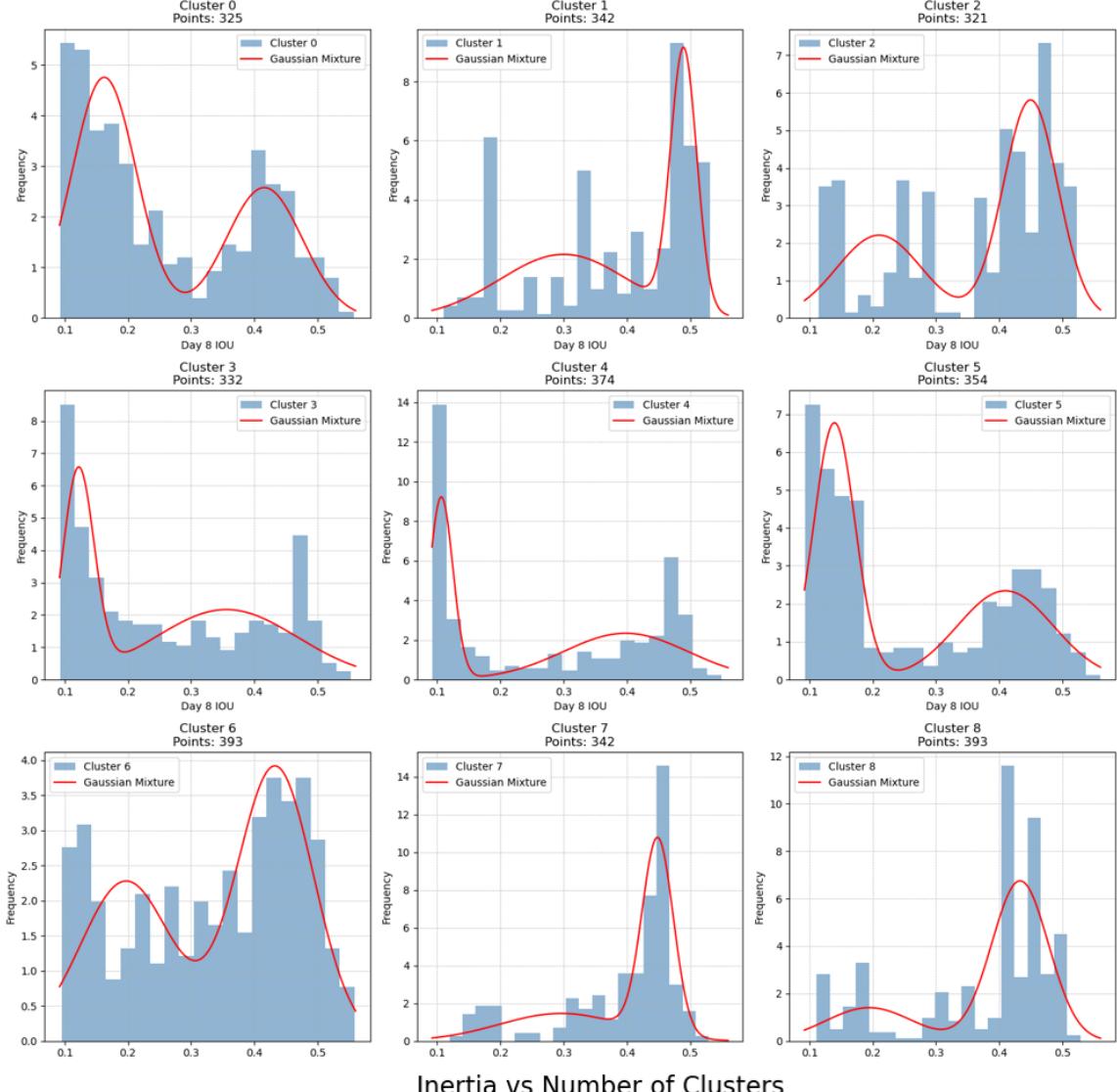


No Crowning: Spotting Distance = 262.0



262 Meter Spotting

IOU Score Frequency within PCA-KMeans Clusters



Inertia vs Number of Clusters

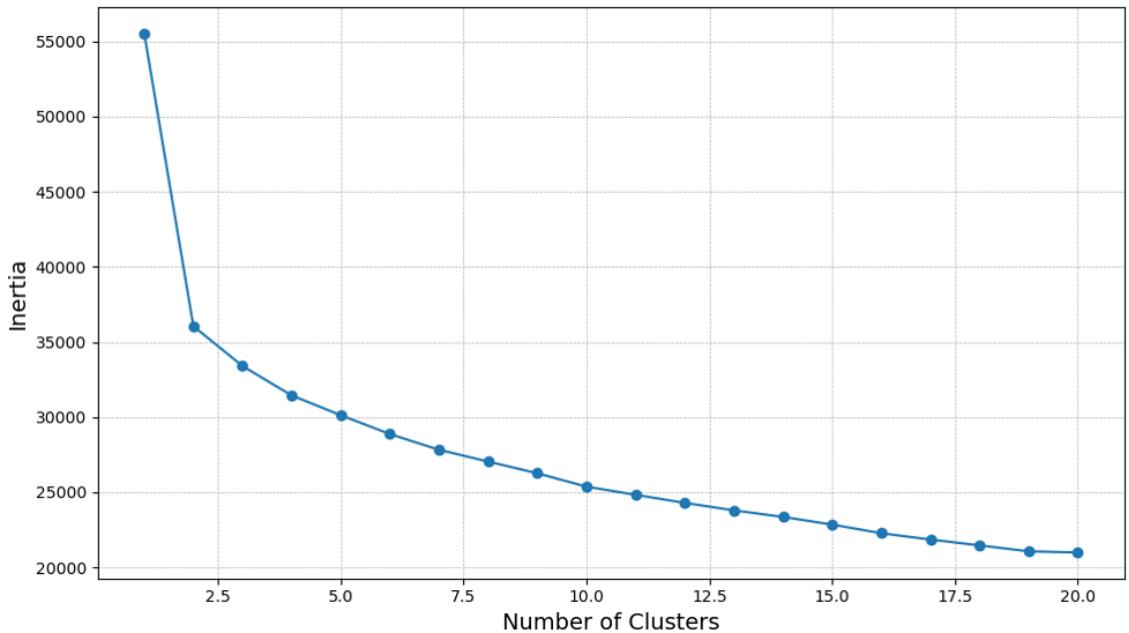


Figure 2: (top subplot) This figure presents the analysis of IOU score distribution within clusters obtained through PCA-KMeans. A 9-cluster solution was used, and each subplot displays the IOU score histogram along with a Gaussian Mixture model fit, revealing distinct patterns in IOU score frequencies across the clusters. (bottom subplot) The PCA transformation with 17 components is computed for dimensionality reduction, followed by an analysis of the "elbow" point in the inertia versus number of clusters plot, helping to determine an optimal number of clusters for subsequent analysis.

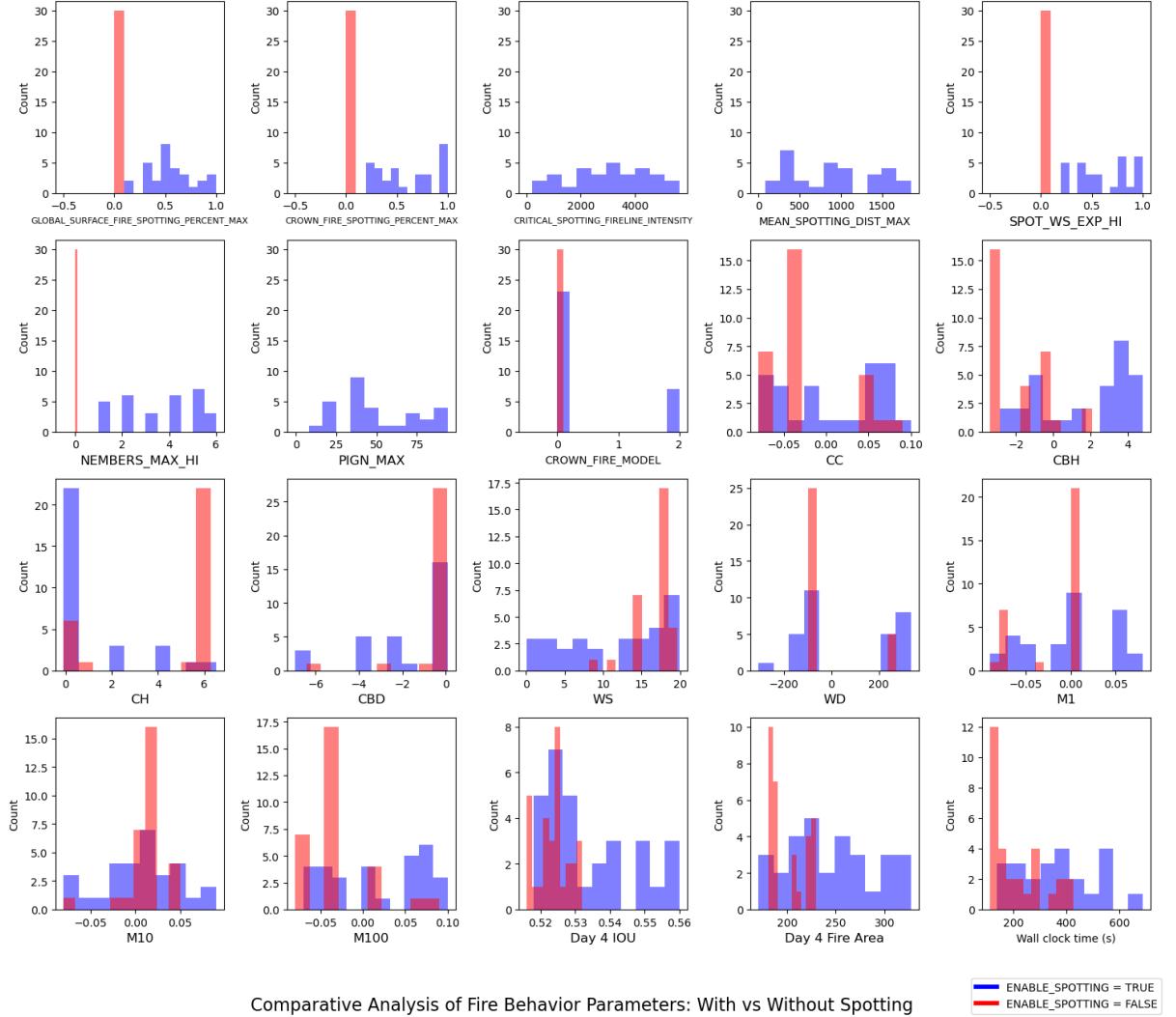


Figure 3: This figure showcases a comparative analysis of fire behavior parameters, contrasting scenarios with and without spotting enabled. Histograms of selected parameters are presented for both cases, where blue represents 'ENABLE_SPOTTING = TRUE' and red represents 'ENABLE_SPOTTING = FALSE'. The analysis provides insights into how parameter distributions differ when spotting is enabled or disabled in fire simulations.

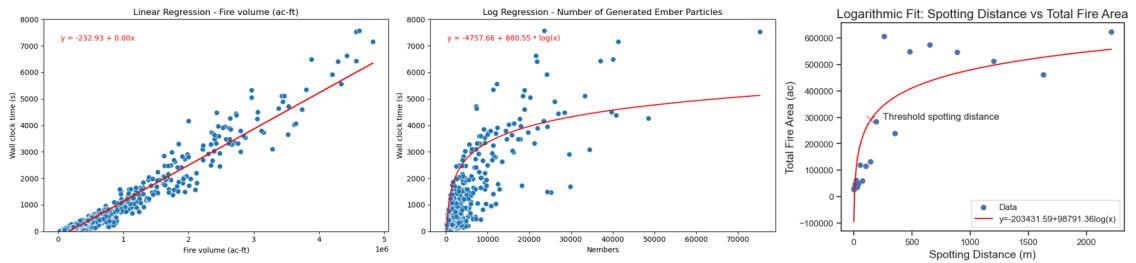


Figure 4: The figure consists of 3 subplots that explore the relationship between specific parameters and the 'Wall clock time' of a simulation. The first subplot presents a linear regression analysis for 'Fire volume (ac-ft)', showing a red line depicting the fitted linear model. The second subplot utilizes a logarithmic regression approach to analyze the impact of the 'Number of Simulated Embers' on 'Wall clock time', with a red curve illustrating the fitted log regression. The third subplot showcases the application of a logarithmic fit to model the correlation between 'Spotting Distance' and 'Total Fire Area'. The red curve represents the fitted logarithmic function, while a marked point signifies the location of maximum curvature, indicating a key threshold in the relationship.

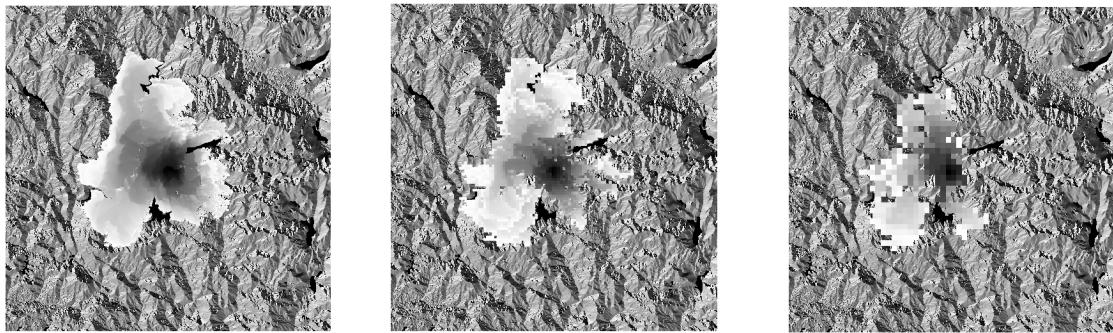


Figure 5: Description for decreasing_resolution.

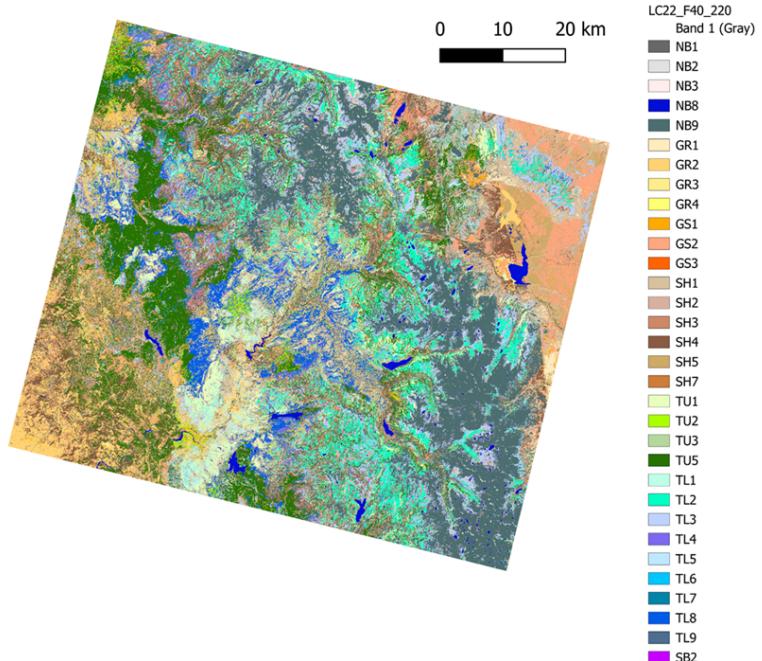


Figure 6: Description for fbfm40_creek.

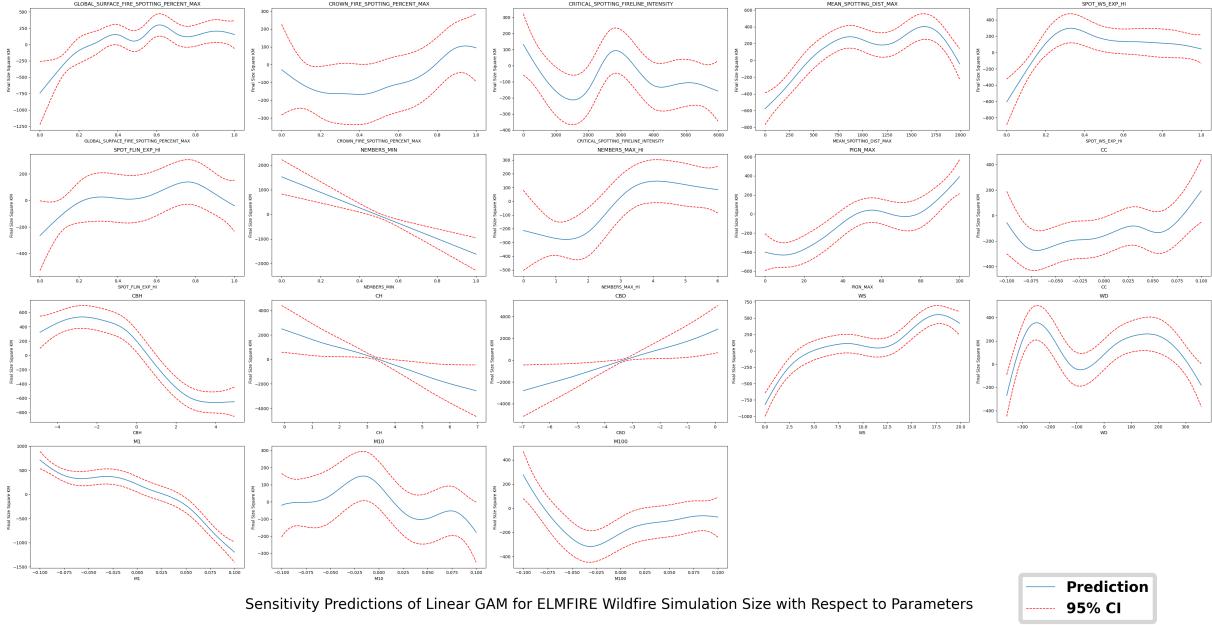


Figure 7: Description for gam-splines.

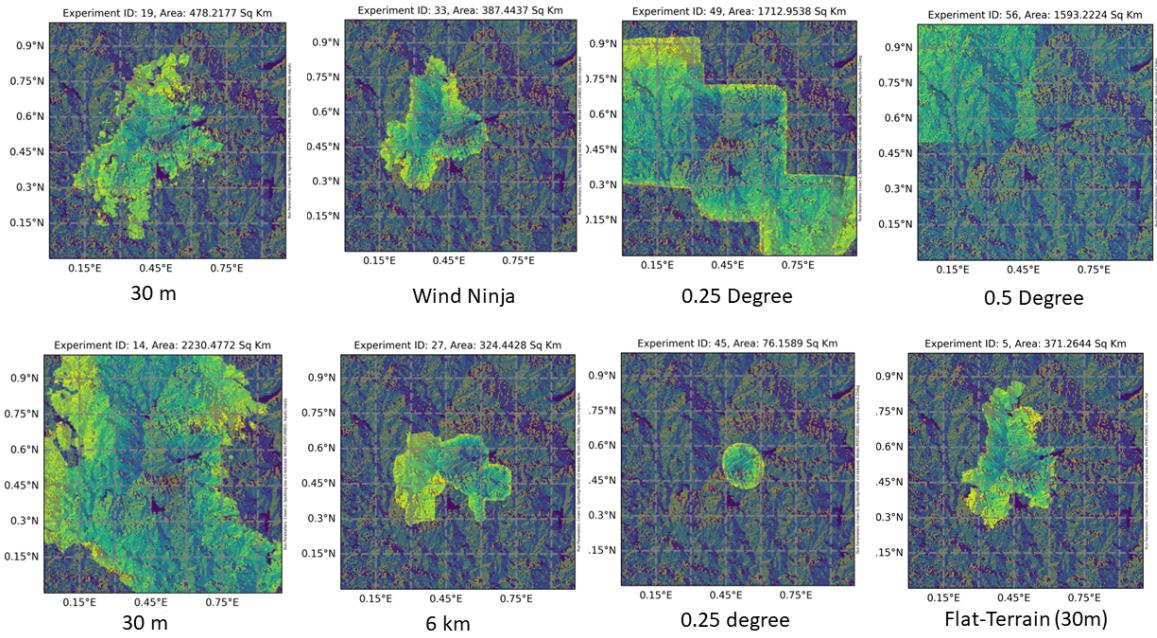


Figure 8: Description for interesting_runs.

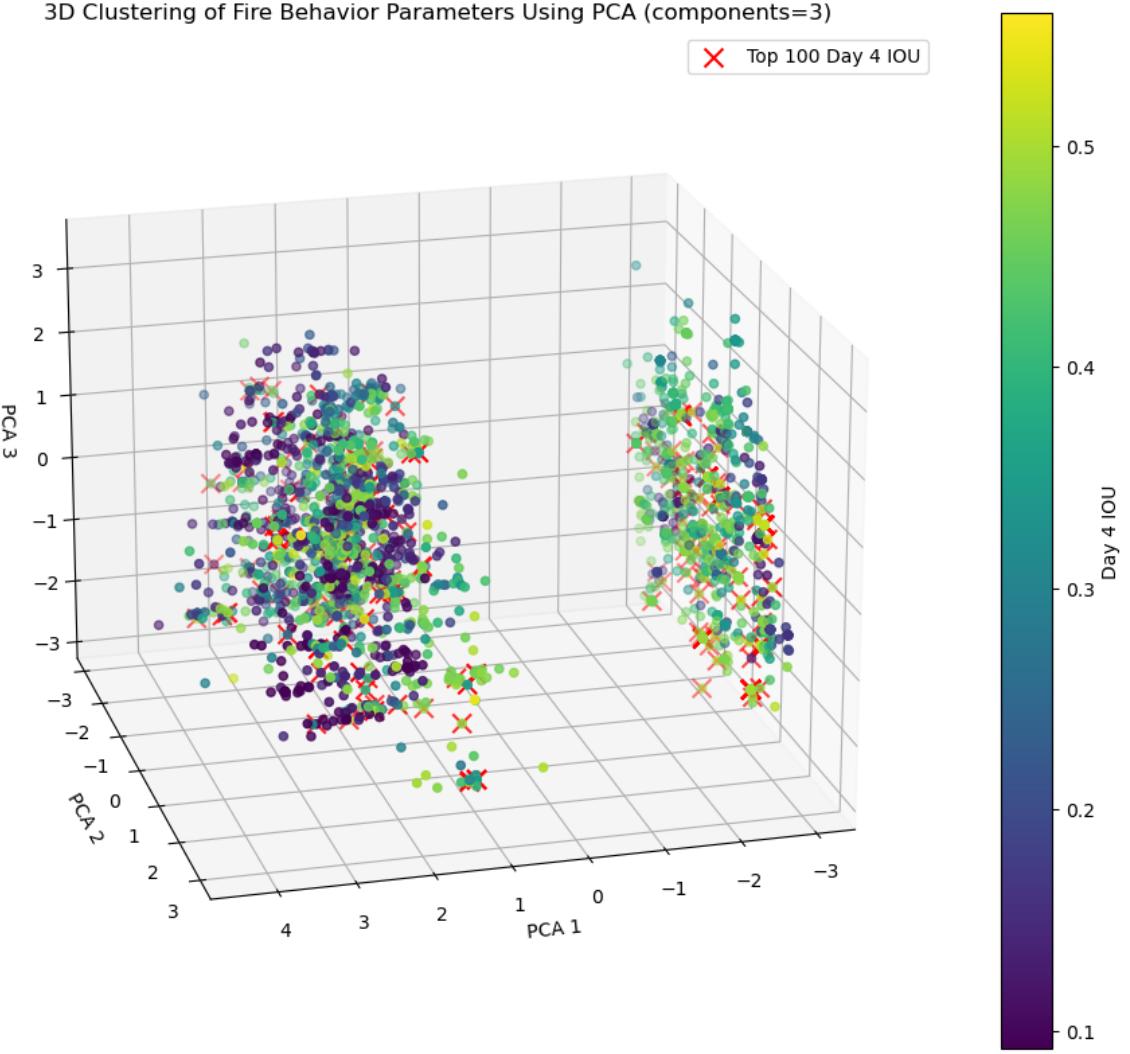


Figure 9: The depicted figure unveils discernible groupings denoting analogous fire behaviors in the parameter space. Within this portrayal, the top 100 best-performing runs are symbolized by 'x' markers, with color gradations corresponding to IOU metrics. This specific subset of the PCA component space reveals intriguing clusters that exhibit certain patterns, although their complete distinction remains somewhat obscured. The figure offers a preliminary insight into potential relationships and trends within the data, inviting further investigation into the underlying factors driving fire behavior variations.

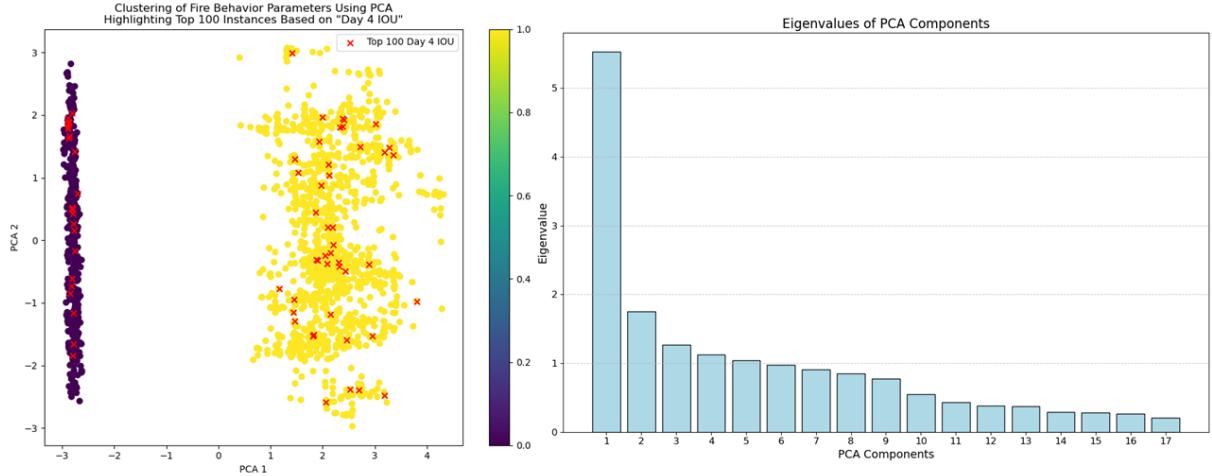


Figure 10: On the left, Principal Component Analysis (PCA) was executed on a 17-component parameter space, and the resulting plot showcases the first two PCA components in relation to each other. When color-coded by the presence of enabled spotting, the plot compellingly delineates distinct clusters, underscoring spotting's pivotal role as a primary influencer within the parameter space. The right pane provides insight into the eigenvalue distribution of the 17 PCA components. The initial component stands out with the highest eigenvalue, followed by an exponential decline in subsequent components. This figure not only sheds light on the parameter space structure but also accentuates the dominance of spotting as a significant determinant in the analysis.

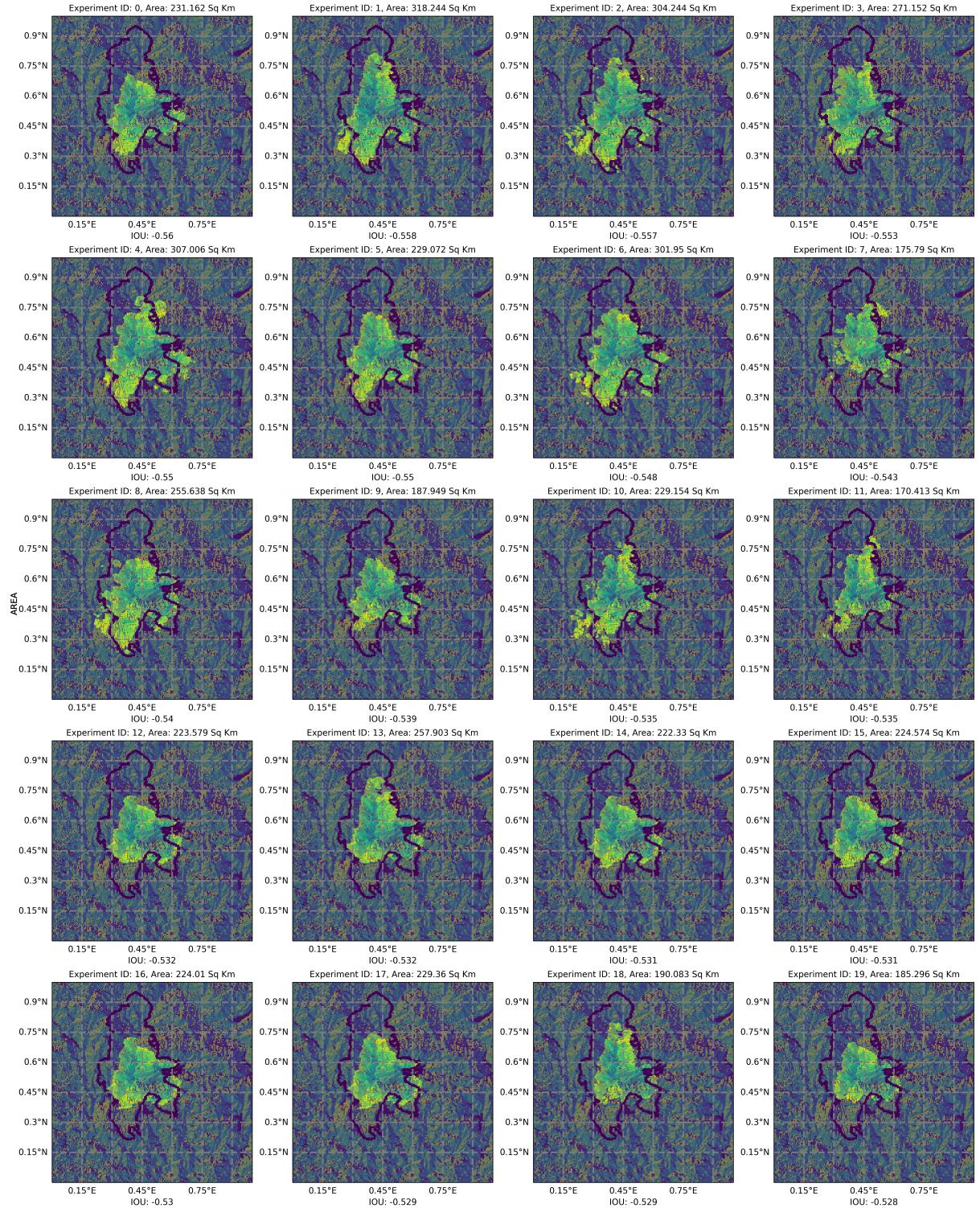


Figure 11: The top 20 wildfire simulations ranked by Intersection over Union (IOU) metrics, accompanied by the VIIRS fire perimeters validated on the fourth day (September 9, 2020, 12:00). The experiment IDs are cross-referenced with the "best20runs.csv" dataset, housing optimal parameters for the CreekFire. This figure underscores the efficacy of the selected

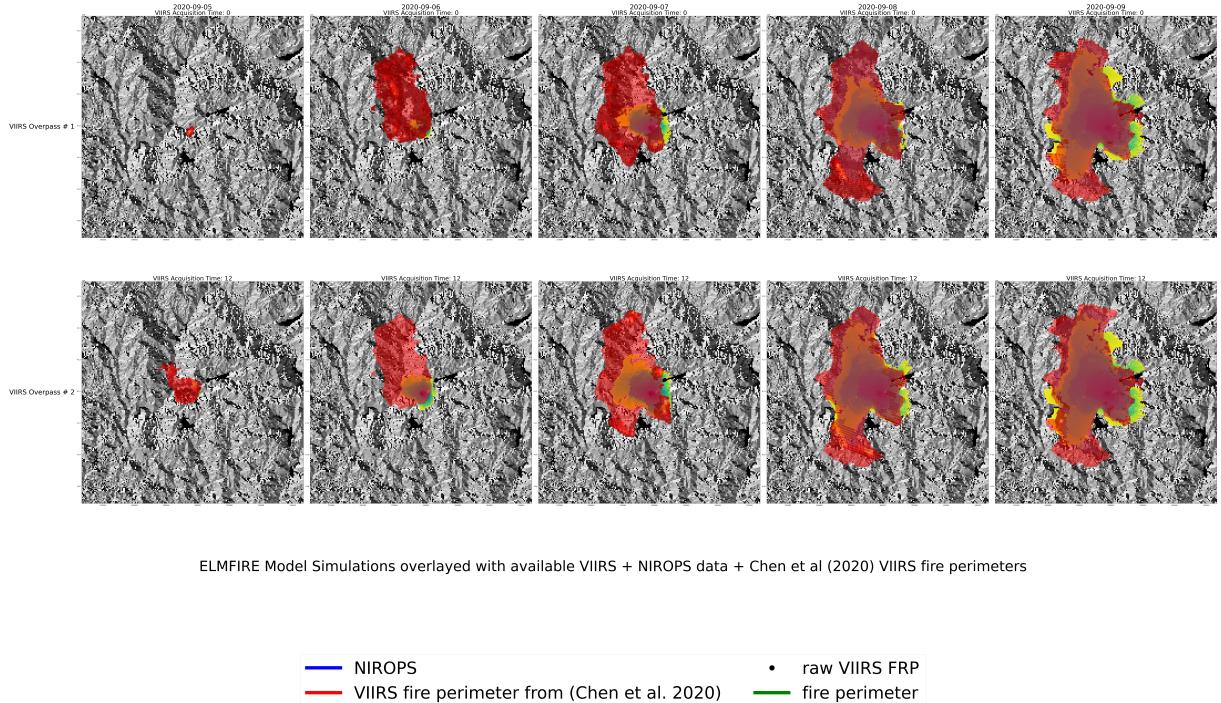


Figure 12: Figure 1 displays the VIIRS instrument data from the Suomi NPP satellite over the first four simulation days, alongside results from the ELMFIRE spread model. The figure includes raw VIIRS hotspots shown as black dots, VIIRS fire perimeters based on Chen et al. (2020), and NIROPS perimeters where available. This illustration effectively showcases the integration of VIIRS data and ELMFIRE modeling for analyzing wildfire progression

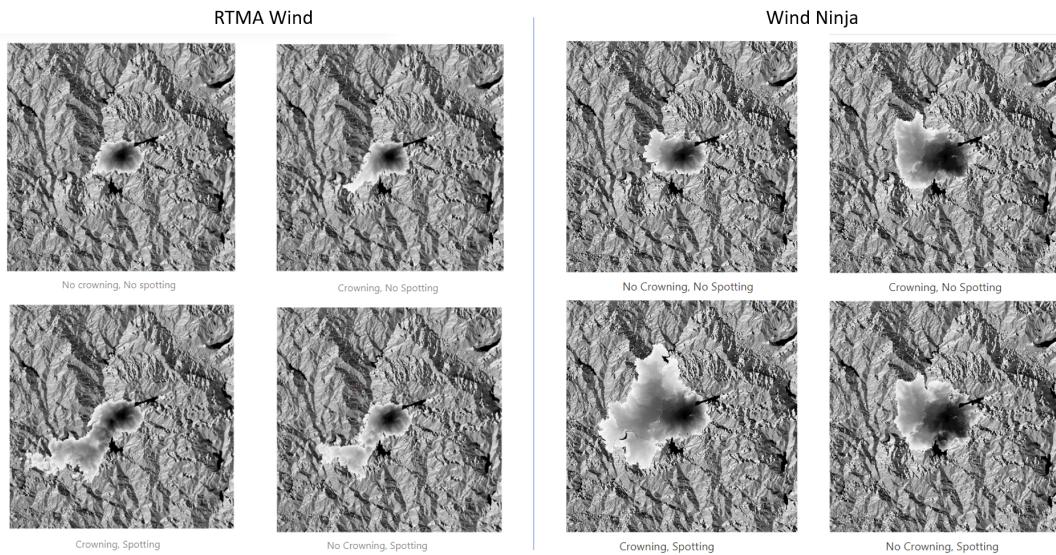


Figure 13: Description for wind_comp_analysis.