

Prediction of forest fire ignition probabilities in time with survival analysis

Campania (Italy) and the Mediterranean area

Bachelor Thesis



Prediction of forest fire ignition probabilities in time with survival analysis
Campania (Italy) and the Mediterranean area

Bachelor Thesis

May, 2025

By

Hans Christian Zareh Lausten-Thomsen

Copyright: Reproduction of this publication in whole or in part must include the customary bibliographic citation, including author attribution, report title, etc.

Cover photo: Vibeke Hempler, 2012

Approval

This thesis has been prepared for 15 weeks at the Technical University of Denmark, DTU, in partial fulfilment for the degree Bachelor of Artificial Intelligence and Data, BSc Artificial Intelligence and Data.

It is assumed that the reader has a basic knowledge in the areas of statistics.

Hans Christian Zareh Lausten-Thomsen - s224174

.....
Signature

.....
Date

Abstract

Wildfire ignition is inherently a time-to-event process influenced by dynamic environmental and human factors. In this thesis, we frame ignition risk as a survival analysis problem and compare three modeling approaches: Andersen-Gill Cox proportional hazards (AG Cox), Random Survival Forests (RSF), and DeepSurv using data from two regions (Campania, Italy and the broader Mediterranean) under both spatial and temporal validation schemes. We develop time varying features such as days since last fire, seasonal covariates, and distance weighted neighborhood summaries, and thoroughly prevent information leakage through data preprocessing. In Campania, DeepSurv achieved the highest concordance index, suggesting potential gains from nonlinear modeling, while RSF provided the most stable performance when models in the Mediterranean proved challenging to train. AG Cox offered clear interpretability but yielded lower predictive accuracy. Integrated Brier scores and calibration analyses reveal modest improvements over the AG baseline, with spatial validation exposing substantial geographic variability and temporal validation yielding tighter, though possibly overconfident, uncertainty estimates. We discuss limitations including data sparsity, feature encoding ambiguity, and the absence of explicit fuel-recovery dynamics. Finally, we highlight that future work could utilize satellite derived vegetation recovery, explore convolutional neural networks to capture more nuanced spatial interactions, and apply transfer learning to improve cross-region performance. Our results demonstrate that dynamic survival analysis can deliver nuanced forecasts of ignition risk, supporting proactive, location- and time specific wildfire management.

Contents

Preface	ii
1 Introduction	1
1.1 Background	1
1.2 Research Motivation	2
1.3 Problem Statement	2
1.4 Objectives	3
1.5 Research Questions	3
1.6 Thesis Structure	3
2 Data	5
2.1 Raw Data and Study Regions	5
2.2 Feature Engineering and Preprocessing	6
2.3 Descriptive Statistics	6
3 Method	9
3.1 Study Area and Data Collection	9
3.2 Data Preprocessing, Splitting Strategy, and Leakage Avoidance	11
3.3 Models	12
3.4 Evaluation Metrics	16
4 Results	19
4.1 Model Performance Overview	19
4.2 Campania Results	19
4.3 Mediterranean Results	22
5 Discussion	27
5.1 Model Performance and Statistical Significance	27
5.2 Feature Importance Across Models and Regions	28
5.3 Contour and graph plots	31
5.4 Limitations and Future Work	33
6 Conclusion	37
Bibliography	39
A Appendix	41

1 Introduction

1.1 Background

Climate change has accelerated the frequency and severity of wildfires, even in regions that previously experienced few, if any, ignition events.[1]. Rising temperatures, longer droughts, and different precipitation patterns have created conditions more suitable to frequent and intense wildfires across Europe. Previously unaffected areas of Central and Northern Europe are experiencing an increase in fire size and frequency [1, 2].

Studies indicate that these fires not only threaten biodiversity and ecosystems but also pose significant risks to infrastructure, air quality, and human health. Wildfires release substantial amounts of carbon dioxide, worsening climate change. For example, in 2020, wildfires in California emitted 91 million metric tons of CO₂, surpassing the state's annual emissions from their power production [3]. Additionally, post-fire landscapes become more vulnerable to soil erosion and water contamination, with elevated nitrogen and phosphorus levels leading to harmful algal growth, affecting aquatic life and water quality [3, 4].

The economic impact of wildfires is also profound. The 2017 North Bay fires in California, for instance, resulted in approximately \$10 billion in insured losses [5]. More recently, the 2025 Los Angeles wildfires have been estimated to cause damages between \$135 billion and \$150 billion, highlighting the escalating financial toll of such natural disasters [6].

Global Relevance and Sustainability Goals

The increasing threat posed by wildfires aligns with several of the United Nations Sustainable Development Goals (SDGs). This research contributes directly to **SDG 13 (Climate Action)** by aiming to inform about potential hazards that can improve our ability to adapt to and mitigate the effects of climate-driven fire risk. This can enable earlier interventions and reducing the likelihood of large scale ignition events, which also supports **SDG 15 (Life on Land)**, helping to protect ecosystems, biodiversity, and natural resources. Additionally, since wildfires affect urban wildland interfaces, this work aligns with **SDG 11 (Sustainable Cities and Communities)** by informing risk management and emergency planning about potential risks. Through data driven innovation in environmental modeling, it also contributes to the idea of **SDG 9 (Industry, Innovation and Infrastructure)** by adapting technological approaches to natural disasters.

The Need for Time-Dependent Wildfire Modeling

Despite the increasing frequency and severity of wildfires, many traditional models used for wildfire prediction rely on machine learning approaches that often fail to incorporate time-dependent factors in ignition risk assessment. While methods such as Random Forest and Neural Networks can effectively classify fire-prone regions, they struggle to model the changing probability of ignition over time, which is critical for implementing proactive fire prevention strategies [7].

Ignition risk is inherently dynamic. It shifts in response to daily weather conditions, long-term climatic trends, and even human activity patterns. Accurately predicting when and where an ignition is likely to occur is crucial for enabling interventions such as fuel reduction, controlled burns, or firebreak construction before conditions become extreme. However, many traditional models only provide snapshot predictions rather than a continuous risk assessment over time [8, 9].

Recent advancements in survival analysis techniques provide an alternative by modeling fire ignition probability as a time-to-event process rather than a static classification problem. The Cox Proportional Hazards (Cox PH) model has been effectively applied to dynamically assess wildfire ignition likelihood under evolving climatic and fuel conditions, outperforming different parametric survival models in capturing the temporal dynamics of fire occurrence [10]. Additionally, survival analysis has been employed to study the duration and control time of ongoing wildfires, demonstrating its broader applicability in wildfire management scenarios beyond ignition modeling alone [11].

By using time-to-event modeling, survival analysis enables dynamic assessment of wildfire ignition risk by incorporating key factors such as the Fire Weather Index (FWI), fuel type, and human activity. Unlike static classification models, survival analysis accounts for how risk evolves over time, offering a more nuanced framework for predicting ignition under changing environmental conditions. This temporal perspective can improve both early warning systems and long-term fire management strategies [11].

Additionally, a hybrid approach combining machine learning with survival models has been suggested, where ML techniques can be used for feature selection or pre-classification, while survival analysis provides a temporal perspective on ignition risk dynamics [7]. This combination allows for a spatially and temporally adaptive fire risk assessment, ensuring that prevention strategies remain effective as conditions change. Although this requires that the set of input variables remains consistent over time, something we struggled with during development, as frequent changes to our features (due to issues like data leakage) limited our ability to implement and evaluate such a hybrid approach effectively.

1.2 Research Motivation

Given the increasing frequency, severity, and unpredictability of wildfire ignitions, there is an urgent need for models that go beyond static classification to incorporate time dependent changes in ignition risk. Traditional machine learning methods, while effective at identifying fire-prone areas, often overlook the evolving nature of environmental and human factors.

This study is motivated by the potential of survival analysis to address this gap. By framing ignition as a time-to-event process, survival models provide a dynamic risk assessment that supports early interventions such as fuel load reduction or firebreak construction. This approach supports broader wildfire management efforts by providing time sensitive risk assessments that can inform where and when preventive measures may be needed.

1.3 Problem Statement

Despite the growing threat of wildfires, most existing ignition risk models treat the problem statically, neglecting the time-varying nature of fire hazards. This thesis addresses that limitation by modeling wildfire ignition as a time-to-event process using survival analysis.

We focus on three survival modeling techniques: the Andersen-Gill extension of the Cox Proportional Hazards model (which explicitly accommodates recurrent events), Random Survival Forests (RSF), and DeepSurv. While only AG Cox is formally structured for recurrent event analysis[12], RSF and DeepSurv were trained on interval-level data that includes repeated fire events, allowing them to learn recurrence related patterns indirectly, particularly through features such as time since last fire and neighbor fire risk, which reflect hazard buildup over time. All models support censored observations and can incorporate both static and time-varying covariates[12, 13, 14] if carefully handled. We evaluate their

performance on wildfire data from Campania and the broader Mediterranean region, under both spatial and temporal validation schemes.

1.4 Objectives

This research aims to:

- Assess the effectiveness of AG Cox, RSF, and DeepSurv in modeling wildfire ignition as a time-to-event process.
- Evaluate model performance using spatial and temporal validation.
- Identify key environmental and anthropogenic variables influencing ignition risk.
- Contribute to proactive fire prevention through dynamic ignition risk estimation.

1.5 Research Questions

To address the challenges outlined in the problem statement and achieve the objectives above, this thesis investigates the following research questions:

1. **Modeling Time-Dependent Risk:** How effectively can survival analysis techniques specifically AG Cox, RSF, and DeepSurv model wildfire ignition as a dynamic, time-to-event process?
2. **Model Performance Comparison:** How do survival models compare to one another in terms of predictive accuracy, calibration, and uncertainty across spatial and temporal validation schemes?
3. **Predictive Drivers:** Which environmental, meteorological, and anthropogenic variables most significantly influence ignition risk over time?
4. **Spatio-Temporal Robustness:** To what extent do the models generalize across different geographical regions (Campania vs. broader Mediterranean), and what are the spatial implications of their performance?

1.6 Thesis Structure

The remainder of this thesis is structured as follows:

- **Chapter 2** reviews the raw data that were used
- **Chapter 3** describes the preprocessing steps of the data and the methodological framework that were used.
- **Chapter 4** reports and compares model results, with a focus on predictive performance, calibration, and feature importance.
- **Chapter 5** discusses the implications of the results for wildfire risk and outlines potential future research directions.
- **Chapter 6** concludes the thesis with a summary of key findings and contributions.

2 Data

2.1 Raw Data and Study Regions

This study uses spatio-temporal wildfire data for two regions:

- **Campania region (Southern Italy)** – divided into 650 grid cells of 5×5 km.
- **Broader Mediterranean region** – divided into 17,421 grid cells of 10×10 km.

Each grid cell is a spatial analysis unit where environmental, anthropogenic, and meteorological features are assigned over time.

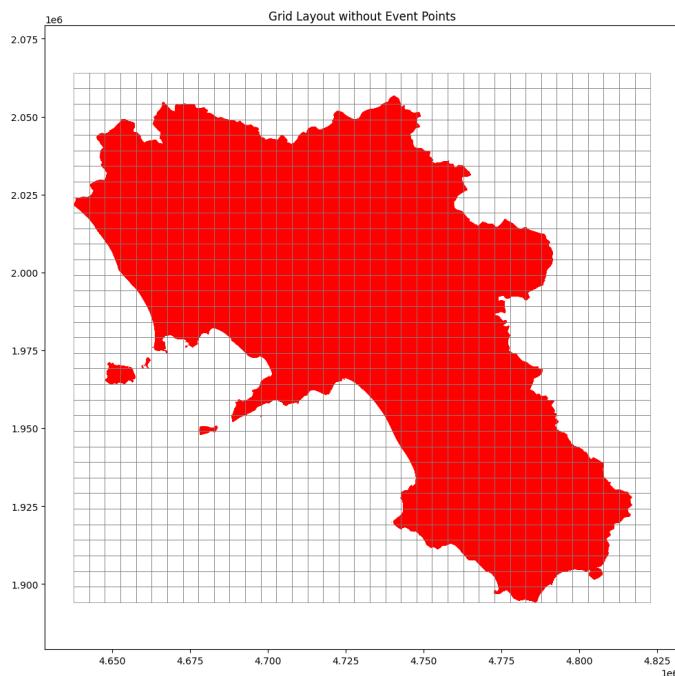


Figure 2.1: The 5×5 km gridded layout of the Campania region. This spatial discretization is used for assigning events and features to fixed locations.

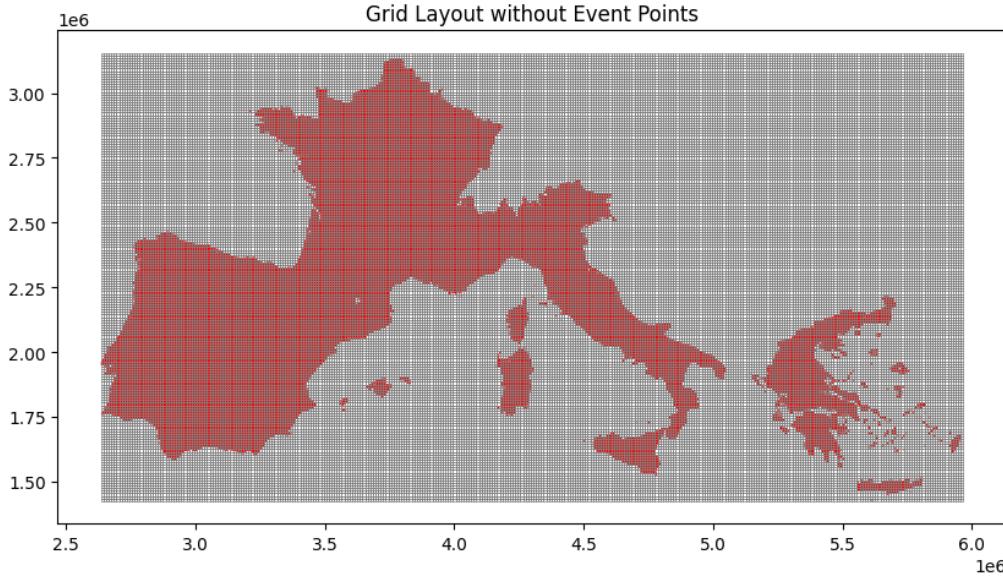


Figure 2.2: Gridded Mediterranean 10 km cells (area is large so lines may be blurred)

	Campania	Mediterranean
Number of grids	650	17,421
Number of fire grids	511	4,704
Censoring rate	78.6%	27%
Total fires	10,190	22,170

Table 2.1: Summary of grid and fire statistics for both datasets

2.2 Feature Engineering and Preprocessing

Each grid cell is observed over time and split into intervals, which serve as the modeling units for time-to-fire prediction. For each interval, we record:

- Whether a fire occurred at the end of the interval,
- Meteorological conditions (smoothed via rolling windows),
- Land cover, elevation, and slope,
- Human activity features (road/powerline density, urban cover),
- Fire history metrics (e.g., `days_since_last_fire`, `no_prior_fire`).

To avoid overfitting and information leakage, fire history features are constructed with appropriate censoring (details in the Method chapter). Meteorological variables use 7-day rolling means or 20-day rolling sums to smooth out noise and reduce contamination from the fire event itself.

Neighbor-based features (e.g., `nbr_fire_risk`, `nbr_forest`) were calculated using past events within a fixed spatial and temporal buffer. Their detailed definitions and leakage prevention mechanisms are provided in the method chapter.

2.3 Descriptive Statistics

We provide summary statistics of the covariates across both regions in Table 2.2.

Covariate	Med Mean	Med Std.	Camp Mean	Camp Std.
temperature (°C)	17.5607	8.7383	22.7175	7.5993
wind_speed ($m s^{-1}$)	11.3791	7.6201	9.5990	2.8805
drought20 (precip. sum, mm)	0.0310	0.0366	0.0390	0.0475
humidity (%)	57.5636	21.5025	51.2666	18.5279
mean_elev (m)	494.9513	421.4658	411.6725	257.7668
max_elevat (m)	—	—	833.6031	370.7459
slope30 (ratio)	0.1938	0.2328	0.0624	0.0794
road_den	785.0193	911.8829	19454.8453	11715.8758
pline_den	0.00028	0.00017	6504.3054	7692.6447
urban (fraction)	0.0458	0.0913	0.0458	0.0573
agricultur (fraction)	0.3155	0.2754	0.2758	0.1885
shrubland (fraction)	0.1996	0.2111	0.0714	0.0830
grassland (fraction)	0.0975	0.1410	0.0322	0.0595
forest (fraction)	0.2924	0.2165	0.3176	0.2278
other (fraction)	0.0279	0.0870	0.0285	0.0601
night_ligh (fraction)	—	—	55.1917	81.5405
duration (days)	1824.0463	1758.9924	219.0481	493.8219
days_since_last_fire (days)	265.5440	714.3772	323.2668	582.7483

Table 2.2: Combined Summary Statistics of Key Covariates (time-varying variables are rolled and some features are unscaled in this format)

3 Method

3.1 Study Area and Data Collection

3.1.1 Study Regions and Wildfire Event Structuring

This study focuses on wildfire ignitions in the Campania region of southern Italy and a larger Mediterranean area analyzed separately. Both data sets have analysis units in the form of spatially separated grids of 5×5 km for the smaller Campania region and 10×10 km for the larger Mediterranean area. As mentioned there are a multitude of static variables for both data sets and also time varying features from 2007 - 2019 for the Campania region and 2008 - 2023 for the larger Mediterranean set.

Wildfire ignition events from the respective datasets were compiled into an event dataframe, with each ignition treated as a separate event. Since each grid cell could experience multiple fires over time, the data were structured to accommodate multiple events per cell. For grid cells that did not experience any fires, we generated random intervals: three for the Campania dataset and two for the Mediterranean dataset.

To prevent the models from learning that all non-fire intervals necessarily correspond to 'fire-free' grids, we also added random non-fire intervals between actual fire events in cells that did experience fires. Additional features were computed, the most notable being `no_prior_fire` (a placeholder for unknown past information), `days_since_last_fire`, and `duration`.

Since no information about fire occurrences prior to the observation period is available, we introduced `no_prior_fire` to flag observations as left-censored. This feature requires careful handling: when noisy intervals are artificially introduced into non-fire grids, a correlation can emerge between `no_prior_fire` and the absence of fire, but only if the difference between the number of fire-free grids and those experiencing fire is substantial.

The feature was used in the Campania dataset, where $\frac{511}{650}$ grids experienced at least one fire. In this case, the first recorded fire in a grid would still be flagged with `no_prior_fire`, but a fire would occur, maintaining the feature's validity. However, in the Mediterranean dataset, only $\frac{4704}{17421}$ grids experienced a fire. Because this proportion is significantly lower, including `no_prior_fire` in the models without proper penalization would lead to information leakage, as a large subset of `no_prior_fire` observations would strongly correlate with fire not occurring. In the end, we chose to omit the `no_prior_fire` feature entirely, as it was difficult to determine whether it carried genuine predictive information or introduced data leakage. To avoid this ambiguity, we opted to exclude it from all final model runs.

3.1.2 Engineered Temporal Features and Compatibility

`days_since_last_fire` was introduced to capture ignition risk immediately following a fire, but required careful handling. Since the models do not accept negative values, this feature was initialized as 1 for grid cells with no recorded prior fires. After the first fire event, the value of `days_since_last_fire` was calculated as the number of days since the most recent ignition.

This setup often resulted in `days_since_last_fire` and `duration` being equal, as both reflected the length of time since the last event. To avoid overrepresenting one outcome class (fire vs. no fire) in either of these features, additional data alignment was performed to ensure balance between event types.

The duration variable itself was defined as the difference between the start and stop time of each interval, representing the period during which a grid cell was considered at risk. This formulation was necessary to ensure compatibility with the Andersen-Gill extension of the Cox model for recurrent events [15]. All event times were measured in days, and observations were right-censored if no ignition occurred by the end of the study period.

3.1.3 Meteorological Feature Rolling and Normalization

A seven-day rolling window for wind_speed, temperature, and humidity, and a 20-day rolling sum for precipitation were chosen because it is critical to capture the causal environment that drives fire risk and to avoid day-to-day noise. We want to avoid simultaneous bias, if we simply use, for example, temperature on the day of the event, it could be contaminated by the fire itself. Furthermore, we aim to model the “buildup risk” of meteorological effects, reducing the impact of short-term spikes and improving generalization.

To capture seasonal effects, we encoded the month of the year as cyclical features sin_month and cos_month, which represent seasonal trends in fire occurrence (e.g., higher risk in summer) without introducing artificial discontinuities between December and January. All continuous covariates were normalized (scaled) to ensure they are on comparable scales for model training.

To capture human influence, since it has been shown that human activity, such as the presence of roads [16], can increase access to risk-prone areas, we included variables such as urban and powerline density to represent human activity and population exposure.

3.1.4 Spatial Dependency Features

To capture spatial dependencies, we engineered a set of neighbor-based features around each fire event. All such features are computed strictly from the training history by using only events whose centroids lie within a 10 km radius (5 km for Campania) and whose end times are at least 10 days before the event begins. In particular:

For each event i , we let

$$\mathcal{N}_i = \{ j \neq i \mid \|c_i - c_j\| \leq R, t_j^{\text{stop}} < t_i^{\text{start}} - \Delta \},$$

where c_i, c_j are the centroids of events i and j , $R = 10$ km is the search radius, and $\Delta = 10$ days is the temporal lag.

For neighbor fire risk we combine all fire events in \mathcal{N}_i and divide by their combined exposure time

$$\text{nbr_fire_risk}_i = \frac{\sum_{j \in \mathcal{N}_i} \mathbb{I}(\text{fire}_j)}{\sum_{j \in \mathcal{N}_i} \text{duration}_j}$$

where $\mathbb{I}(\text{fire}_j) = 1$ if event j was a fire and duration_j is its length in days.

We define

$$w_{ij} = \frac{1}{\|c_i - c_j\| + \varepsilon}, \quad \varepsilon > 0 \text{ small to avoid division by zero.}$$

Then for each static feature $X \in \{\text{forest}, \text{road_den}, \text{pline_den}\}$:

$$\text{nbr_}X_i = \frac{\sum_{j \in \mathcal{N}_i} w_{ij} X_j}{\sum_{j \in \mathcal{N}_i} w_{ij}}.$$

For the spatial blocking split we do a 'ring' neighbor featuring to prevent spatial leakage. We draw an inner radius of $R_{\text{in}} = 11\text{km}$ and an outer radius of $R_{\text{out}} = 20\text{km}$ with the same 10Δ

$$\mathcal{N}_i^{\text{block}} = \{ j \mid R_{\text{in}} \leq \|c_i - c_j\| \leq R_{\text{out}}, t_j^{\text{stop}} < t_i^{\text{start}} - \Delta \}.$$

Otherwise the same risk and aggregation formulas apply as above.

Test events with $\mathcal{N}_i = \emptyset$ (meaning no qualifying neighbors within 10 km and 10 days) are assigned:

$$\text{nbr_fire_risk}_i = 0$$

$$\text{nbr_road}_i, \text{nbr_forest}_i, \text{nbr_powerline}_i = \text{median of their training set values}$$

The idea is that we want to capture the risk around a fire cell, but our data is structured so that if a fire occurs on the border of different grids we assign it each a unique fire per grids that way we want to keep a buffer zone in terms of days lagging for these fires to not leak information.

This is done for both data sets but with different radius, so for the Campania dataset we are looking at a radius of 6-10 km. Keep in mind that not all grids have a similar structure due to coastlines etc.

3.2 Data Preprocessing, Splitting Strategy, and Leakage Avoidance

To realistically evaluate model performance and prevent information leakage, we applied a combination of **temporal** and **spatial blocking** strategies, alongside careful preprocessing of input features.

3.2.1 Temporal Blocking

Temporal blocking simulates the goal of forecasting by training models on earlier years and evaluating them on later, unseen years. For instance, we reserved the most recent years of wildfire data for testing and used prior years for training. This ensures that no information from future fire seasons leaks into model fitting, providing a realistic estimate of how well the model generalizes to future conditions. Note that doing temporal splits required us to do further data preprocessing.

3.2.2 Spatial Blocking

To address spatial autocorrelation, we implemented spatial blocking. Random train/test splits risk placing test points close to training data, allowing the model to 'recognize' locations and inflate performance. Instead, we divided the study area into spatially distinct blocks and assigned entire blocks to either training or testing. A buffer zone ensured that no training fire lies within a certain radius of a test fire.

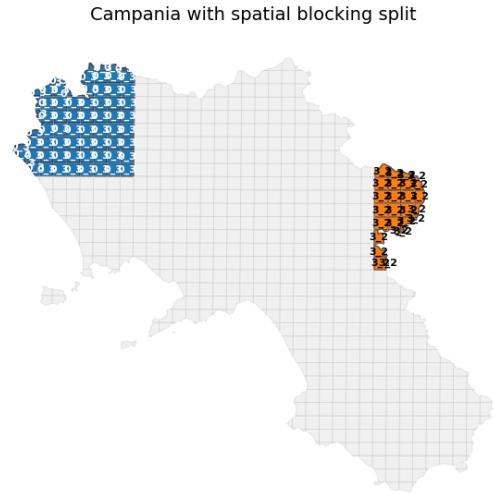


Figure 3.1: Spatial blocking in Campania: blue = training, orange = testing (multiple folds used).

Figure 3.1 illustrates one spatial fold in Campania, with blue and orange representing training and testing blocks, respectively. Multiple blocks per fold were used to ensure diversity.

Leakage-aware neighbor features. As mentioned all neighbor-based covariates are computed exclusively from events that ended prior to the start of each interval, and respecting the current split. Under temporal validation we only use training-period events; under spatial blocking we restrict neighbors to an annulus outside the test block (e.g. 11–20 km away) so that no test-block information leaks into the features. This still raises questions that will be discussed later.

3.2.3 Data Structures

Each grid cell is split into time intervals representing periods of fire risk, with engineered features derived from environmental conditions, human activity, and fire history. Monthly weather metrics were summarized using rolling windows and assigned per grid or station.

id	month	temp_mean	temp_median	temp_q25	temp_q75	...
37297.0	1					
37297.0	2					
:	:					
110367.0	12					

Table 3.1: Rolling monthly weather metric summaries by station ID

Similarly, an event-level dataframe tracks ignition intervals and corresponding features:

3.3 Models

3.3.1 Anderson-Gill Cox Model

The Andersen-Gill (AG) extension of the Cox proportional hazards model casts recurrent events as a multivariate counting process framework [12]. For each grid cell i , we define

$$N_i(t) = \text{number of fire events in cell } i \text{ by time } t,$$

id	start time	stop time	fire occurred	feature 1	feature 2	...
37297.0	0	780	0			
37297.0	780	x	1			
:	:	:	:			
110367.0						

Table 3.2: Event dataframe with intervals and fire indicators

and let

$$Y_i(t) = \begin{cases} 1, & \text{if cell } i \text{ is at risk at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

In our setup every cell is continuously used throughout the study, so

$$Y_i(t) = 1 \quad \forall t \in [0, T],$$

where T is the last time point at which data was collected for a grid cell.

We observe for each subject the counting process $N_i(t)$ and covariate history $X_i(t)$.

The AG model[12] specifies the intensity (hazard) of the process as

$$\lambda_i(t | X_i(t)) = Y_i(t) \lambda_0(t) \exp(\beta^\top X_i(t)), \quad (3.1)$$

where

- $\lambda_0(t)$ is an unspecified baseline hazard function,
- β is a vector of regression coefficients (increase or decrease in risk)
- $X_i(t)$ includes both time fixed covariates (e.g. static land cover) and time varying features (e.g. temperature).

Partial likelihood. We let t_{i1}, t_{i2}, \dots be the ordered event times for grid i . The AG partial likelihood over all observed jumps is defined as

$$L_p(\beta) = \prod_i \prod_{k: t_{ik} \leq T} \frac{\exp(\beta^\top X_i(t_{ik}))}{\sum_j Y_j(t_{ik}) \exp(\beta^\top X_j(t_{ik}))}$$

Here T is the stop time of the study. We then take the log of $L_p(\beta)$ and pick the β that makes it as large as possible that gives us the estimator $\hat{\beta}$. So given that a fire occurs at time t_{ik} what is the probability that it happened in grid cell i .[12]

Robust variance and clustering. Because fire events in the same grid cell can influence one another, we calculate standard errors that are robust to clustering by cell ID. This means that we group all observations from each cell together when estimating the variance, which corrects for the fact that multiple fires in one cell aren't independent.

Implementation. We split each cell's observation period into intervals $[start_{ik}, stop_{ik}]$ corresponding to each event or censoring. Time varying covariates (e.g. neighbor features) are computed at the start of each interval, and then we fit (3.1) using `lifelines.CoxPHFitter`

with the `cluster_col` option. This gives $\hat{\beta}$ along with confidence intervals that account for the fact that multiple events in the same cell are correlated.

This framework allows us to estimate how both environmental drivers and recent neighborhood fire history together influence the instantaneous hazard of fire occurrence in each spatial unit.

Hyperparameter tuning. To balance fit against overfitting, we selected the CoxPH penalization strength by comparing training and validation concordance indices across the same five-fold spatial GroupKFold. For each candidate penalizer ($10^{-3}, 10^{-2}, 10^{-1}, 1$), we performed a multiple fold spatial cross validation each fold training on 12-13 blocks and testing on the remaining 3-4 blocks and computed the concordance index on both the training and held out blocks. We then selected the penalizer that minimized the average train-test C-index gap across all folds. A penalizer of 0.01 achieved the smallest train-test concordance discrepancy and was applied to the following models.

3.3.2 Random Survival Forest (RSF)

Random Survival Forests (RSFs) extend decision tree ensembles to survival analysis by directly modeling time-to-event data with censoring [13]. In contrast to Cox models, Random Survival Forests do not rely on the proportional hazards assumption or a predefined relationship between covariates and hazard. This flexibility allows them to model complex, non-linear interactions across high-dimensional spatiotemporal data [13].

Setup and data structure. Each spatial cell is split into intervals over its observation history. For each interval, we define:

- T_i : the duration of the interval,
- δ_i : a binary indicator (1 if fire occurred, 0 otherwise),
- X_i : a covariate vector including environmental variables (e.g. temperature, vegetation, elevation), temporal features (e.g. month), and neighbor-based features.

Model training. We use the `RandomSurvivalForest` implementation from `sksurv` to train models on the derived survival data. For the Mediterranean dataset, models are trained with 100 trees, a maximum feature fraction of 10%, a minimum of 30 samples to split a node, and 40 samples required in each terminal node. For Campania, we use between 300 and 400 trees, a more conservative maximum feature fraction of 5%, and smaller node sizes (minimum 10 to split and 10 per terminal node). All covariates are scaled using MinMax normalization prior to training.

To address class imbalance, intervals with only one day since the last fire are down-weighted, as they are frequent and mostly correspond to non-fire observations.

Prediction. Given a new observation, the RSF predicts its survival function $\hat{S}_i(t)$ by averaging the Kaplan–Meier curves from the terminal nodes across the forest. From this, we extract:

- Survival probability at future time horizons,
- A fire risk score (via cumulative hazard or $1 - \hat{S}_i(t)$),
- Time-specific metrics for calibration and discrimination.

3.3.3 DeepSurv Neural Cox Model (Campania Only)

DeepSurv is a nonlinear extension of the Cox proportional hazards model implemented as a feedforward neural network [14]. It replaces the linear predictor $\beta^\top X$ in Cox models with a learned neural network, giving a flexible modeling of feature interactions while still optimizing the partial likelihood of a fire.

Scope. We used DeepSurv only for the Campania dataset. On the Mediterranean data, early results showed poor generalization, so we didn't pursue it further. Although the Mediterranean set is larger overall, fire events are much sparser relative to the number of grids. That lower signal density may have made DeepSurv harder to train effectively, especially given its sensitivity to tuning. In contrast, RSF handled this imbalance better. But the full reason wasn't properly established.

Splitting strategy. To ensure robust evaluation, we assessed model performance under both:

- **Spatial blocking:** dividing the study area into non-overlapping spatial blocks to limit geographic leakage,
- **Temporal blocking:** training on earlier years and testing on later ones to simulate true forecasting.

Features and Preprocessing. Inputs matched those used for AG and RSF: static landscape covariates (forest, roads, powerlines, slope, elevation), smoothed meteorological variables, fire history metrics, and neighbor based features. Interaction and squared terms (e.g. temperature_sq, temperature \times forest cover) were included, and all inputs were scaled to $[0, 1]$.

Model architecture. The DeepSurv network consisted of three hidden layers with 128, 64, and 32 neurons, ReLU activations, and a dropout rate of 0.3. We used the Adam optimizer with a learning rate of 1×10^{-4} and weight decay for regularization. Training was performed with early stopping based on validation fold performance.

Model Evaluation. All three models (AG Cox, RSF, and DeepSurv) were assessed under the same spatial and temporal validation frameworks described in earlier sections. Specifically, we performed:

- *Spatial cross-validation:* multiple fold GroupKFold on the 4 \times 4 grid (5x5 for mediterranean), withholding 3-4 blocks per fold;
- *Temporal cross-validation:* rolling forecasts by training on years t_1 and testing on year $t_1 + t_2$.

For each held out fold, we computed:

1. **Concordance index (C-index)** - Antolini's time-dependent C-index to measure ranking performance.
2. **Brier score** - time-dependent Brier scores with 200 day intervals starting from 100 days up until 2000 days (or end of the max observed duration), with the Integrated Brier Score (IBS) as an overall calibration/ranking summary.
3. **Calibration plots** - visual comparisons of model-predicted survival probabilities against the actual (Kaplan-Meier) outcomes across risk groups, showing how closely the predictions match reality.

Reported results are the average and/or standard deviation of each metric across all spatial folds. This purpose is to ensure a fair comparison of model discrimination, calibration, and overall predictive accuracy.

3.4 Evaluation Metrics

These metrics were computed using libraries designed for survival analysis and applied consistently across all spatial and temporal validation folds.

3.4.1 Concordance Index (C-index)

The Concordance Index (C-index) evaluates how well a survival model ranks individuals by their risk of experiencing an event. For any two comparable instances, the prediction is concordant if the model assigns a higher predicted risk (i.e., shorter survival time) to the individual who actually experiences the event earlier. A C-index of 0.5 indicates random performance, while a value near 1.0 reflects strong discriminative ability[17].

For Cox and RSF models, risk scores were derived from the cumulative hazard function or $1 - \hat{S}(t)$, while for DeepSurv, scores were taken directly from the network output (log-hazard). Computation was performed using `sksurv.metrics.concordance_index_censored`.

3.4.2 Brier Score

The Brier Score evaluates the accuracy of predicted survival probabilities at specific time horizons. It is defined as the squared difference between the predicted survival probability and the observed event status. A lower Brier score indicates better calibration and discrimination.

We computed time-dependent Brier scores using `sksurv.metrics.brier_score`. To summarize model performance across the time axis, we calculated the Integrated Brier Score (IBS) by averaging Brier scores across all time points.

3.4.3 Calibration Plots

We assessed calibration by grouping test samples into quantile based risk bins according to their predicted survival probability $\hat{S}(t)$. For each bin, we compared the mean predicted survival probability with the Kaplan-Meier survival estimate. Plots were generated at fixed time horizons (e.g., 100, 500, 1000 days) to reveal any consistent over- or under estimation of risk.

3.4.4 Survival and Hazard Estimation

In addition to summary metrics like the C-index and Brier Score, we compute survival probabilities and cumulative hazards for each observation to analyze risk over time and across regions. This form the basis for the survival maps, contour plots, and other evaluations in our results.

For the AG Cox and RSF models, survival probabilities $\hat{S}_i(t)$ are computed from the cumulative hazard functions using the relationship $\hat{S}_i(t) = \exp(-\hat{H}_i(t))$, where $\hat{H}_i(t)$ is the estimated cumulative hazard at time t .[13, 12] DeepSurv generates a log-risk score for each observation, which reflects how much more or less likely it is to experience an event compared to others. This score adjusts the baseline hazard learned from the training data, and the adjusted hazard is then used to compute the survival probability[14].

These model-derived survival functions allow us to evaluate the probability that a fire has not occurred by a given time, and form a basis for visual analyses comparing risk across space, time, and environmental conditions.

3.4.5 Limitations

Addressing Model Limitations

- **AG-Cox:** Assumes proportional hazards and linear covariate effects which is sensitive to outliers in the data. It can also be sensitive to multicollinearity, which can inflate variance of coefficient estimates [12, 18].
- **RSF:** May produce miscalibrated probabilities and overfit with deep trees or correlated predictors that can lead to miscalibrated survival curves. Also it has limited interpretability without post hoc (permutation importance) methods. Further also requires more memory usage than the other models [13, 19].
- **DeepSurv:** Requires large datasets and can make training unstable if event rates are low. Furthermore, the black-box nature makes it hard to extract clear, causal insights or validate feature effects [14].

4 Results

4.1 Model Performance Overview

We begin by comparing the test C-index performance of all models across both datasets and split strategies.

Model	AG Cox	RSF	DeepSurv
Campania (Spatial)	0.800	0.814	0.831
Campania (Temporal)	0.838	0.865	0.903
Mediterranean (Spatial)	0.811	0.863	—
Mediterranean (Temporal)	0.869	0.905	—

Table 4.1: Concordance index (C-index) for each model under spatial and temporal validation.

Region	Type	Event Rate (%)	Baseline Brier score
Campania	Spatial	54.9	0.248
Campania	Temporal	52.6	0.249
Mediterranean	Spatial	30.9	0.213
Mediterranean	Temporal	24.9	0.187

Table 4.2: Baseline brier scores for each individual model based on their respective event rates

4.2 Campania Results

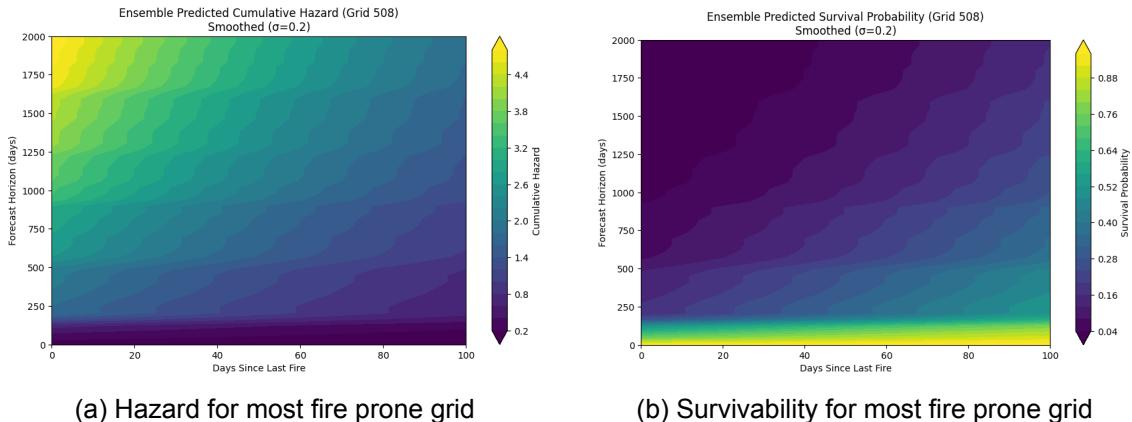


Figure 4.1: Hazard function (left) and survival function (right) predicted by the AG Cox model for the most fire-prone grid cell in Campania. These illustrate the model's estimated risk dynamics over time in a high-risk area under median monthly weather conditions. (Spatial)

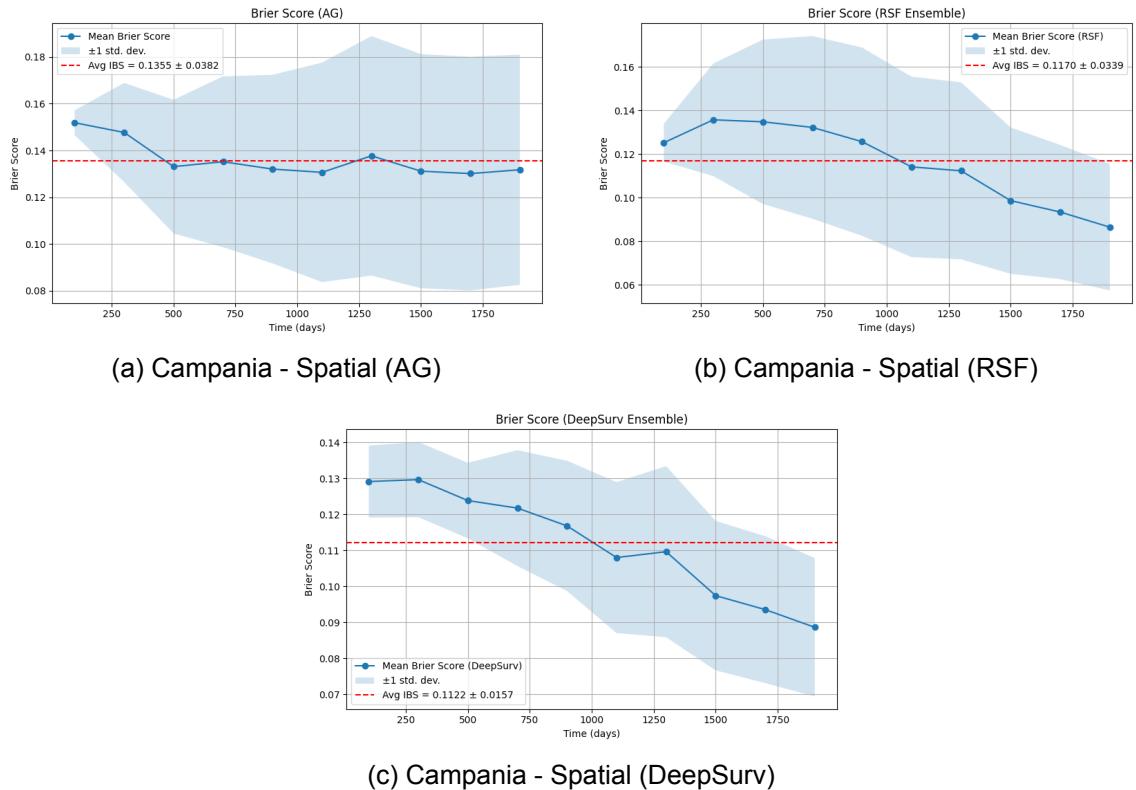


Figure 4.2: Brier score curves for the Campania spatial blocking models. Lower values indicate better predictive calibration.

Model	AG Cox	RSF	DeepSurv
Integrated Brier Score (Spatial)	0.1355 ± 0.0382	0.1170 ± 0.0339	0.1120 ± 0.0170
Integrated Brier Score (Temporal)	0.1347 ± 0.0049	0.1113 ± 0.0046	0.0886 ± 0.0034

Table 4.3: Brier scores for each model under spatial and temporal validation. (Campania)

Comparison	ΔC (95% CI)	Shapiro-Wilk		Paired t-Test		Wilcoxon	
		W	p	t	p	W	p
RSF vs AG	0.018 [0.011, 0.024]	0.951	0.739	4.990	0.0025	0.0	0.0156
DeepSurv vs AG	0.028 [0.024, 0.033]	0.963	0.843	11.124	<0.0001	0.0	0.0156
RSF vs DeepSurv	-0.011 [-0.016, -0.007]	0.978	0.948	-4.782	0.0031	0.0	0.0156

Table 4.4: Statistical test results and 95% confidence intervals for ΔC comparing model concordance indices across five spatial folds.

Comparison	Mean ΔC	95% CI
RSF vs AG	0.025	[0.019, 0.030]
DeepSurv vs AG	0.067	[0.061, 0.073]
RSF vs DeepSurv	-0.042	[-0.048, -0.037]

Table 4.5: Bootstrap C-Index Differences and 95% Confidence Intervals for Temporal Split

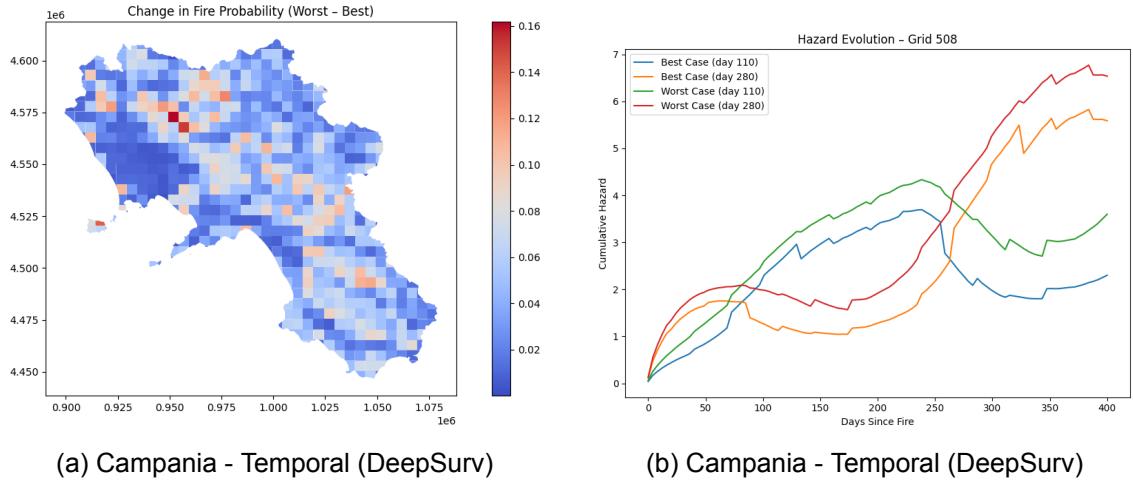
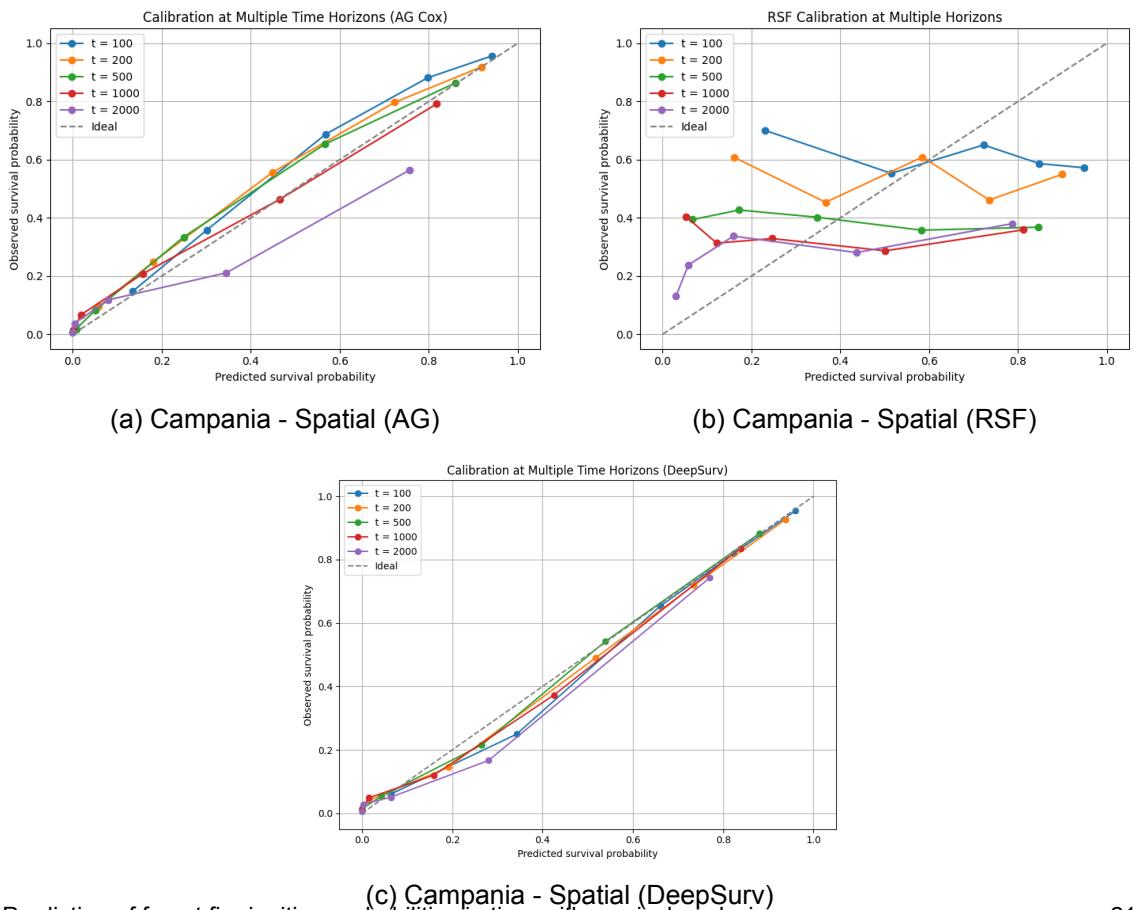


Figure 4.3: (a) Fire probability map of Campania given we are in August, it has been 2 months since last fire and we are looking at a time horizon of 60 days. (b) Hazard evolution over time if we experience a fire at day 110 (blue, green) vs day 280 (orange, red) of the calendar year. This is with worst and best case scenario for the weather metrics.



Prediction of forest fire ignition probabilities in time with survival analysis

Figure 4.4: Calibration plot in time horizons for the models. The closer the graphs are to 45 degrees the better performance over time.

4.2.1 Features

Variable	Model 1 Z-Score (Spatial)	Model 2 Z-Score (Temporal)	Mean Z
days_since_last_fire	-18.30	-22.25	20.28
drought20	-16.26	-18.17	17.22
neighbor_fire_risk	14.08	15.37	14.73
humidity	-11.56	-10.90	11.23
forest_combined	8.80	4.88	6.84
temperature	6.93	7.35	7.14
agri	5.83	5.90	5.87
temperature_sq	5.50	3.48	4.49
temp_forest_interaction	-4.89	-5.35	5.12
mean_eleva	-3.96	-4.22	4.09

Table 4.6: Top 10 Most Influential Variables Across Both Campania AG Models (Ranked by Mean Absolute Z-Score)

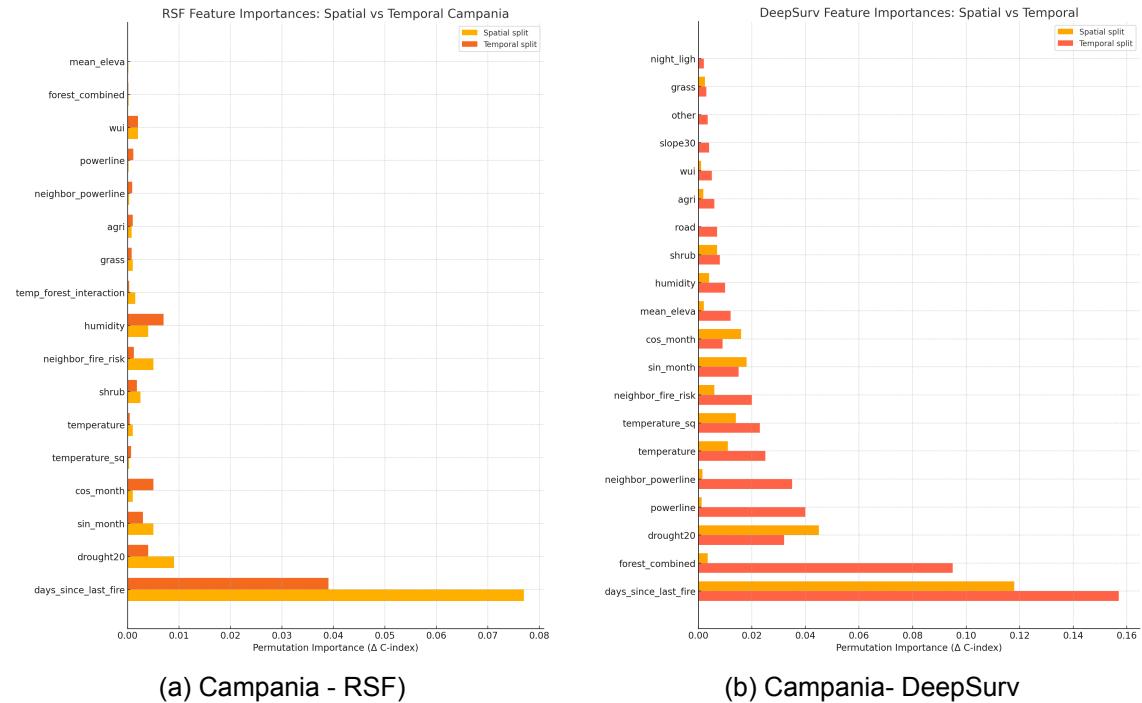


Figure 4.5: RSF and DeepSurv feature importances over Campania spatial and temporal split. Orange is Spatial and red is Temporal

4.3 Mediterranean Results

Model	AG Cox	RSF
Integrated Brier Score (Spatial)	0.1480 ± 0.0742	0.1154 ± 0.0610
Integrated Brier Score (Temporal)	0.1052 ± 0.0013	0.0657 ± 0.0007

Table 4.7: Concordance index (C-index) for each model under spatial and temporal validation.

Test	Statistic	p-value	95% CI
Shapiro-Wilk normality	$W = 0.927$	0.558	—
Paired t -test	$t = 5.352$	0.0031	—
Wilcoxon signed-rank	$W = 0.000$	0.0312	—
Bootstrap (RSF - AG ΔC)	$\Delta C = 0.055$	—	[0.036, 0.073]

Table 4.8: Statistical comparison of RSF vs. AG on the Mediterranean spatial split.

Test	Statistic	p-value	95% CI
Shapiro-Wilk normality	$W = 0.999$	0.853	—
One-sample t -test	$t = 921.884$	0.0000	—
Wilcoxon signed-rank	$W = 0.000$	0.0000	—
Bootstrap (RSF - CoxPH ΔC)	$\Delta C = 0.027$	—	[0.026, 0.029]

Table 4.9: Statistical comparison of RSF vs. CoxPH on the temporal split.

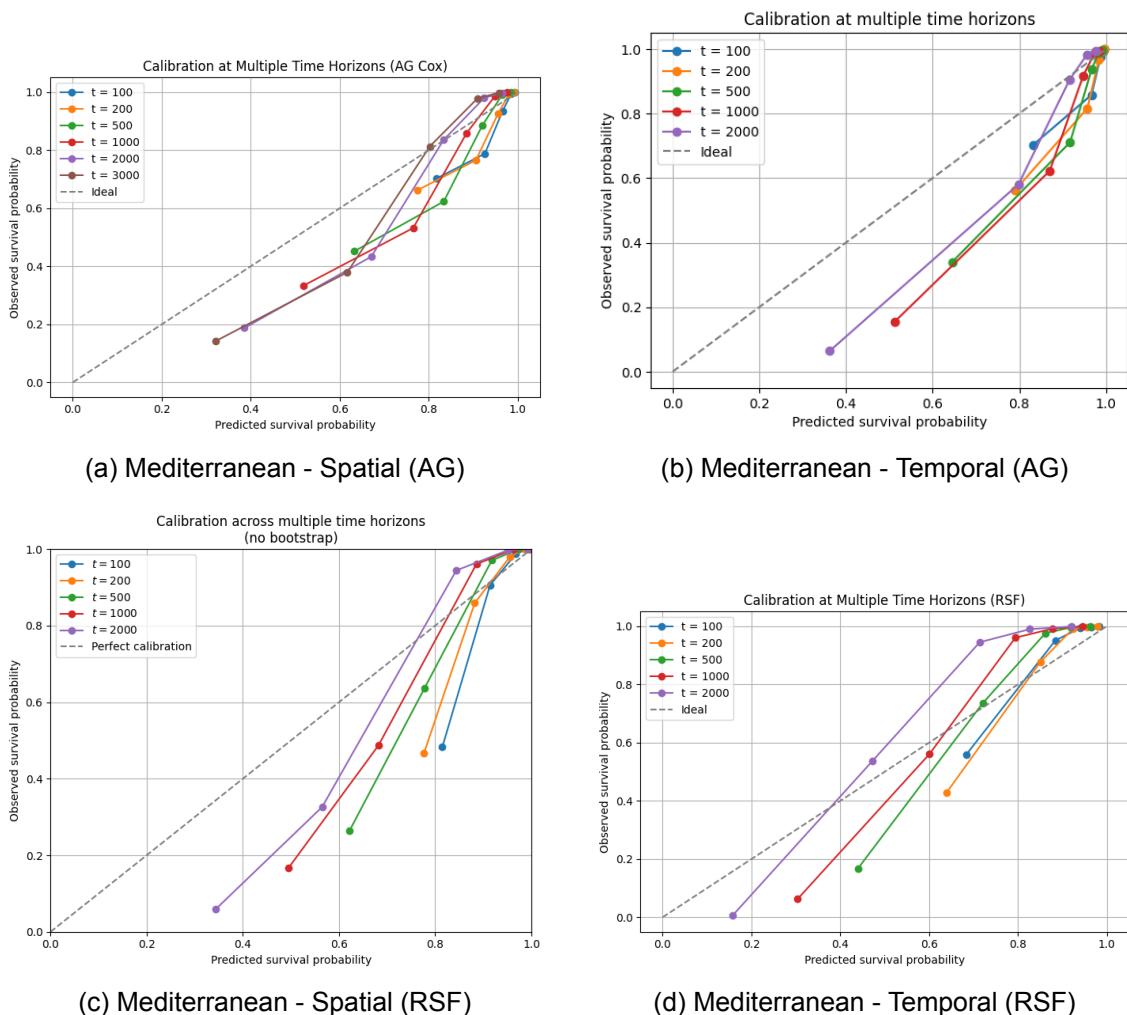


Figure 4.7: Calibration plot in time horizons for the models. The closer the graphs are to 45 degrees the better performance over time.

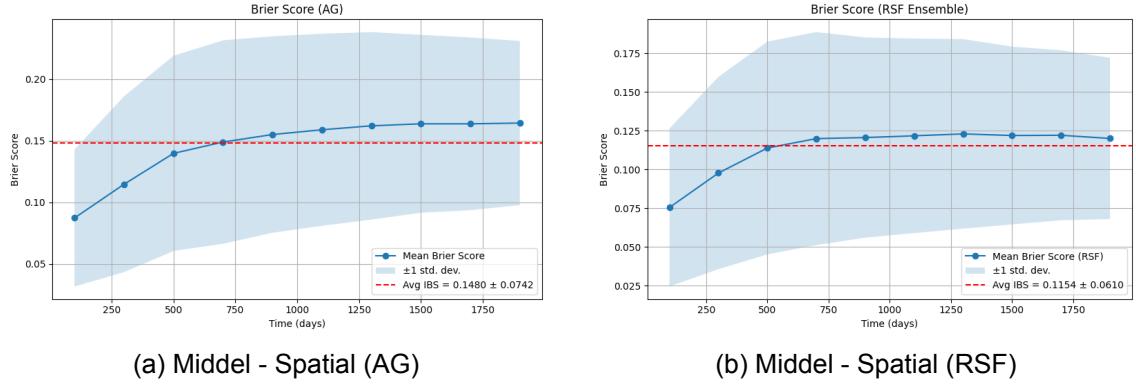


Figure 4.6: Brier score curves for the Mediterranean spatial blocking models. Lower values indicate better predictive calibration.

4.3.1 Features

Variable	Model 1 Z-Score (spatial)	Model 2 Z-Score (temporal)	Mean Z
humidity	-35.24	-27.15	31.20
nbr_fire_risk	24.93	21.86	23.40
shrubland	19.26	17.44	18.35
slope30	19.10	14.75	16.93
pline_den	12.94	15.79	14.36
temperatur	11.46	5.23	8.35
sin_month	8.13	-18.11	13.12
drought_20	-8.70	-7.05	7.88
days_since_last_fire	-11.54	-7.66	9.60
grassland	16.09	5.46	10.78

Table 4.10: Top 10 Most Influential Variables Across Both AG Models (Ranked by Mean Absolute Z-Score) in Mediterranean

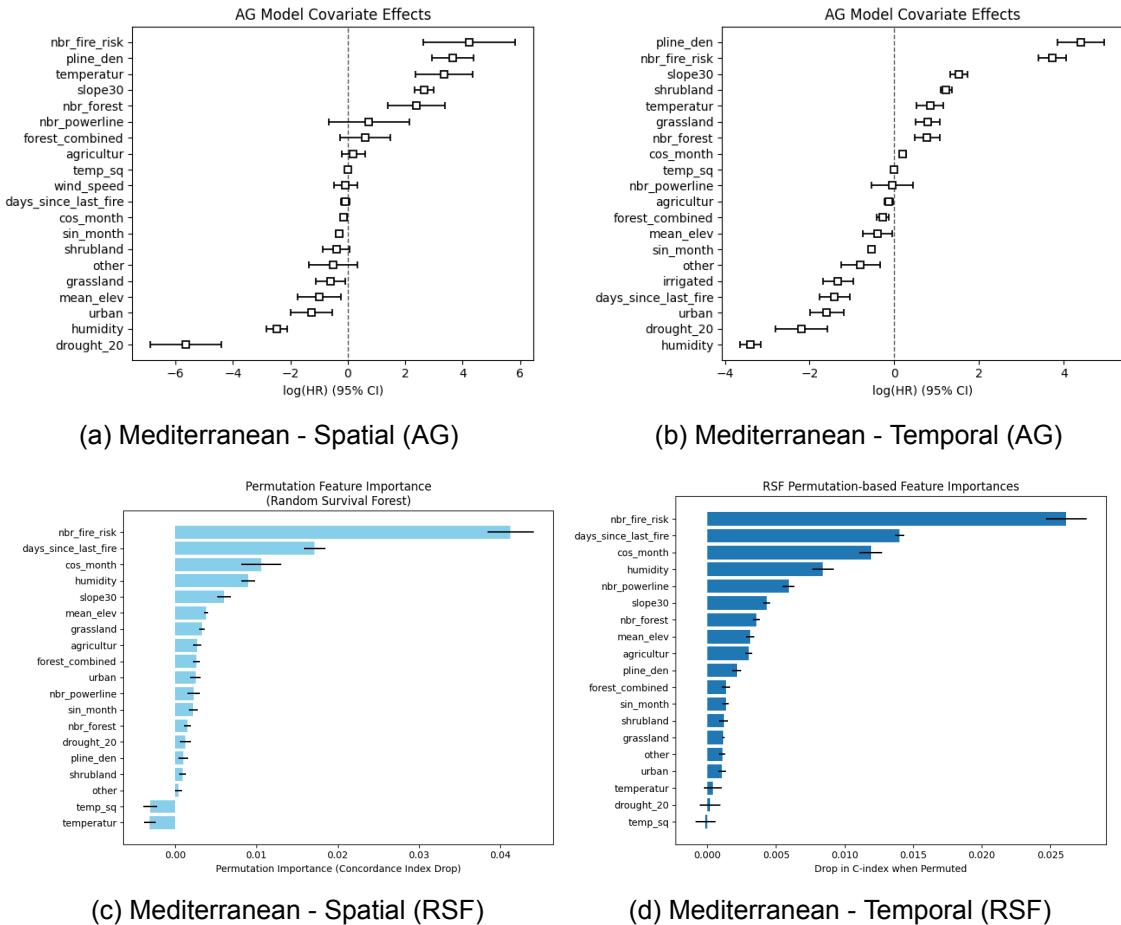


Figure 4.8: Covariate effect between a temporal split and a spatial blocking split.

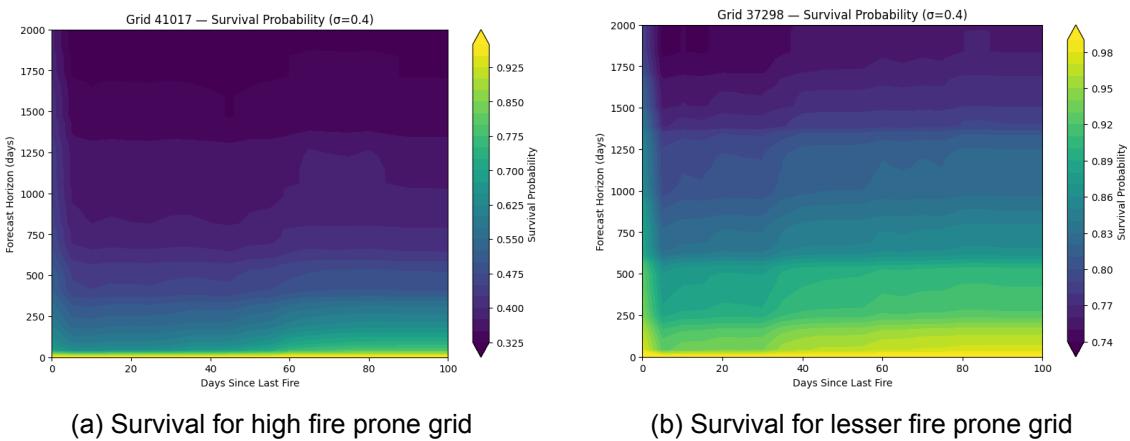
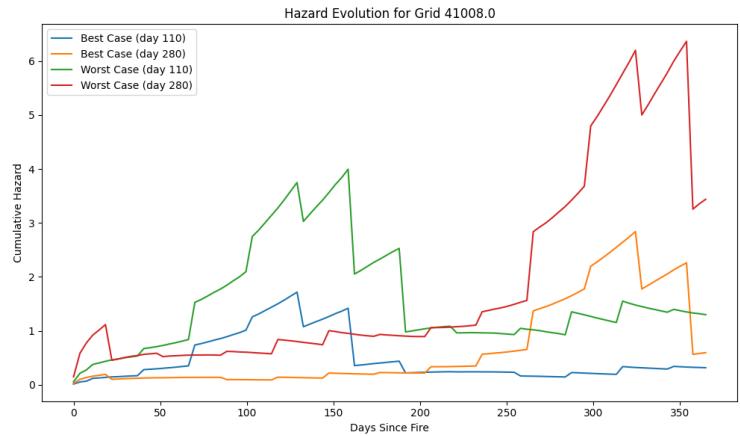
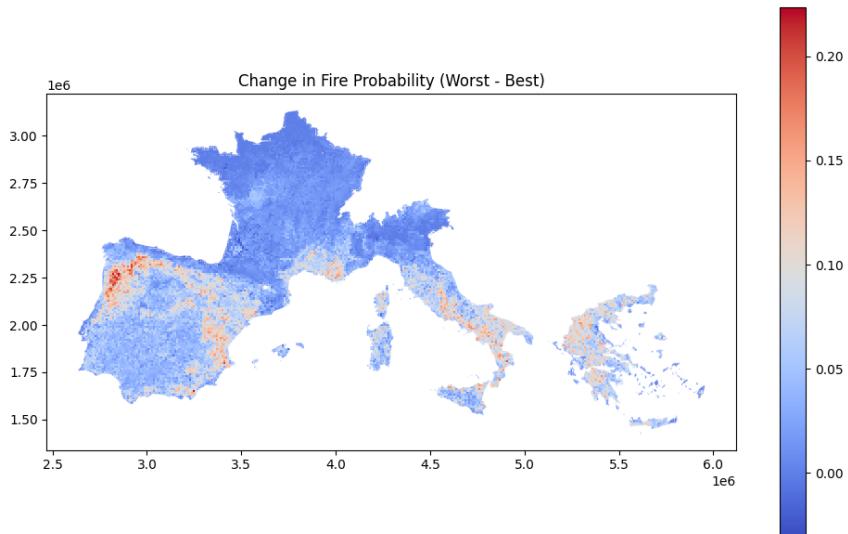


Figure 4.9: Survival rate for a high risk grid cell (left) and survival rate for a low risk grid cell (right) in the Mediterranean area. (Temporal - RSF)



(a) Mediterranean - Temporal (AG)



(b) Mediterranean - Temporal (RSF)

Figure 4.10: Modelled best weather conditions and worst conditions to see increase in fire risk at highest risk period (August) and the following 60 days given a fire occurred 2 months ago.

5 Discussion

In this study, we compared three modelling approaches: Andersen-Gill Cox (AG), Random Survival Forest (RSF), and DeepSurv on two different validation schemes (spatial and temporal blocking) and two regions (Campania and the broader Mediterranean). Our main findings can be summarized as:

5.1 Model Performance and Statistical Significance

Across both spatial and temporal splits in Campania, DeepSurv achieved the highest concordance index (C-index), followed by RSF and AG Cox (Table A.5). Under the spatial validation, DeepSurv outperformed AG Cox with a mean $\Delta C = 0.028$ (95% CI [0.024, 0.033]), while RSF also showed a significant improvement over AG ($\Delta C = 0.018$, 95% CI [0.011, 0.024]). Although RSF underperformed DeepSurv by $\Delta C = -0.011$ (95% CI [-0.016, -0.007]), all comparisons were statistically significant (Table 4.4). Paired t-tests and Wilcoxon signed-rank tests yielded $p < 0.01$ and $p < 0.016$ across comparisons, and Shapiro-Wilk tests confirmed approximate normality in the bootstrap ΔC distributions.

Under the temporal split, performance differences were even more pronounced. RSF again outperformed AG Cox by a mean $\Delta C = 0.025$ (95% CI [0.019, 0.030]), while DeepSurv showed a markedly larger improvement over AG Cox with $\Delta C = 0.067$ (95% CI [0.061, 0.073]). Crucially, DeepSurv also significantly outperformed RSF under the temporal setting ($\Delta C = -0.042$, 95% CI [-0.048, -0.037]), establishing a clear advantage for DeepSurv in temporal generalization (Table 4.5).

In the Mediterranean region, RSF consistently outperformed AG Cox under both spatial ($\Delta C = 0.055$, 95% CI [0.036, 0.073]) and temporal validation ($\Delta C = 0.027$, 95% CI [0.026, 0.029]) (Tables 4.8, 4.9). These differences were statistically significant ($p < 0.05$ across all tests), highlighting RSF's ability to capture complex ignition dynamics that proportional-hazards might not capture[13].

Since the temporal split involves only a single train-test separation, this limits the statistical confidence of our results. Although we attempted to mitigate this by applying bootstrapping on our test data, the lack of multiple temporal folds still constrains the robustness of the evaluation. To increase a better estimate more models should've been trained on different temporal splits, however they would've had to be carefully handled since training on 70% and testing on 30% might not really be that significantly different from doing an 80% 20% split since a lot of the same test and training would still fall into the same pool. However a type of cross-validation like time-series specific splits would be beneficial to reduce uncertainty in the temporal setting.

5.1.1 Brier scores - Campania

Table 4.3 summarizes the Brier scores for the three models: AG Cox, RSF, and DeepSurv evaluated under both spatial and temporal validation schemes. Overall, differences in mean performance are modest, with Brier scores generally ranging from 0.11 to 0.14. Both RSF and DeepSurv slightly outperform AG Cox, though the margins are relatively small except for DeepSurv under temporal validation.

Under the spatial blocking approach, which involved a 7 fold cross-validation across spatially distinct regions, the standard deviations of the Brier scores are notably high. For instance, AG Cox shows a mean Brier score of 0.1355 with a standard deviation of 0.0382,

while RSF and DeepSurv achieve lower means (0.1170 and 0.1120, respectively) with slightly reduced but still substantial variability. This uncertainty reflects meaningful spatial variety in model performance, some regions are clearly more predictable than others. These results highlight that evaluating models in geographically different regions is crucial for understanding generalization in real world applications.

In contrast, under temporal validation where models are trained on earlier years and tested on future observations, the mean Brier scores for AG Cox and RSF remain similar to their spatial counterparts. However, DeepSurv achieves a substantially lower score (0.0886 ± 0.0034), indicating better calibration and sharper absolute risk estimation in this setting. While this performance is promising, the tight uncertainty bands stem from bootstrapping on a single temporal split and should not be overinterpreted as evidence of greater model stability. Instead, they reflect the constraints of the evaluation framework, which may underestimate the true uncertainty that would emerge from using multiple or rolling temporal splits.

Taken together, these findings suggest that while RSF and DeepSurv offer modest improvements over AG Cox in terms of predictive accuracy, the spatial blocking results point to substantial geographic variability in performance. This variability must be accounted for when considering model implementation across diverse landscapes, as performance can differ significantly depending on where the model is applied.

5.1.2 Brier Scores - Mediterranean

The integrated Brier scores in Table A.6 show a pattern consistent with our observations from the Campania region. Both AG Cox and RSF models achieve lower Brier scores under temporal validation compared to spatial validation, indicating improved predictive accuracy when tested on chronologically held-out data. Notably, RSF again outperforms AG Cox across both validation schemes, mirroring its relative performance in Campania. A key observation is the difference in uncertainty: the spatial validation yields much larger standard deviations (e.g., ± 0.0742 for AG Cox and ± 0.0610 for RSF), reflecting substantial fold-to-fold variability across the Mediterranean region. This likely stems from the larger geographic area and greater environmental variation compared to Campania. In contrast, the temporal Brier scores exhibit extremely small uncertainty intervals, particularly for RSF, which shows a standard deviation of just 0.0007. This reflects the same trend seen in Campania, that temporal validation yields more stable but potentially overconfident evaluations due to the reliance on a single train-test split. As with the previous case, the tight uncertainty bands should not be interpreted as evidence of genuine model stability, but rather as a limitation of the evaluation framework.

5.2 Feature Importance Across Models and Regions

Campania

Feature importance analyses via permutation scores for RSF and DeepSurv (Figure 4.5) and Z-scores for AG Cox (Table 4.6) in Campania consistently highlight `days_since_last_fire` as the most influential predictor across all models in Campania. This reinforces the role of recent fire history for ignition risk, though it may also reflect information leakage if not carefully handled. Notably, its importance was especially pronounced under temporal validation for DeepSurv but more prominent in the spatial for RSF, suggesting some models may have learned grid specific fire recurrence patterns.

In earlier iterations, DeepSurv appeared to assign disproportionately high importance to `temperature_sq` under temporal validation, which was later traced to a feature scaling error. After correcting this, the feature remained moderately important, more than in RSF,

but no longer dominated the rankings as before. This outcome underscores DeepSurv's capacity to capture non-linear meteorological effects, although now in a more interpretable manner. In contrast, RSF maintained its emphasis on recurrence and seasonality related features, consistently assigning lower importance to meteorological variables.

These results suggest more nuanced shifts in feature emphasis across validation strategies. In AG Cox, meteorological variables such as `humidity` and `temperature` were consistently influential across both spatial and temporal validations, indicating a strong linear relationship between these features and ignition risk. For DeepSurv, there was a mixed shift: some weather related features like `temperature` and `temperature_sq` gained importance under temporal validation, while others like `humidity` slightly declined. Additionally, the importance of landscape variables such as `forest_combined` and `neighbor_powerline` increased, suggesting that DeepSurv may capture more complex environmental interactions over time.

In contrast, RSF remained relatively stable across both validation schemes, continuing to emphasize static and recurrent features such as `days_since_last_fire`, `drought20`, and `cos_month`. Meteorological features had consistently lower importance in RSF, suggesting that it may encode their effects indirectly through interactions or threshold-based splits. These trends reflect how model architecture and validation strategy influence the perceived relevance of different covariates.

Mediterranean

In the Mediterranean region, feature importance patterns diverged somewhat from those observed in Campania, reflecting broader spatial variety. For AG Cox (Table 4.10), `humidity` stood out as the most influential variable across both spatial and temporal models, with extremely high negative Z-scores, highlighting its strong linear association with ignition risk. Other top features included `nbr_fire_risk`, `shrubland`, `slope30`, and `pline_den`, indicating that both vegetation type, terrain characteristics and human factors played a critical role in the model's hazard function.

Interestingly, `days_since_last_fire`, which was the dominant variable in Campania, was less influential here, though still significant. This could be due to either reduced recurrence regularity or lower signal-to-noise ratio in the broader Mediterranean region, where fire return intervals may vary more drastically. In the RSF models (Figure 4.84.8), both spatial and temporal permutations reaffirmed the importance of `nbr_fire_risk` and `days_since_last_fire`, but with a notable drop in meteorological feature influence. While `humidity` retained moderate importance, variables like `temperature` and `drought_20` remained consistently low in RSF despite their prominence in AG Cox suggesting that RSF may capture these effects indirectly through interactions with static or contextual features.

Furthermore, `slope30` and `mean_elev` had greater relevance in the Mediterranean RSF models compared to Campania, especially under spatial validation. This aligns with the more rugged and topographically diverse nature of the Mediterranean dataset, where terrain driven climates may model fire risk more strongly.

However, interpreting importance in RSF and DeepSurv remains challenging. Permutation scores may underestimate the influence of correlated variables or hidden interactions[20], especially in neural networks like DeepSurv where effects can be distributed across layers. For example, weak individual importance for `temperature` in RSF may hide its indirect role via feature combinations or split thresholds. Permutation based methods offer some insight, but may miss non-linear effects and future work should consider using e.g. partial dependence plots to clarify model reasoning[21].

5.2.1 Regional Variability in Feature Importance

The differences in feature importance between Campania and the broader Mediterranean underscore the role of regional context in shaping fire dynamics. While Campania displayed a strong reliance on recent fire history and temporal drought indicators (e.g., `days_since_last_fire`, `drought_20`), the Mediterranean models showed greater diversity in top-ranked features, including terrain and vegetation attributes such as `slope30`, `shrubland` and `humidity`, as well as region specific risk indicators like `nbr_fire_risk`, which likely helped distinguish consistently hazardous areas from lower-risk zones.

This shift likely reflects the broader spatial heterogeneity and ecological diversity of the Mediterranean region, where ignition risk cannot be attributed to a single dominant process like recurrence. Instead, multiple interacting factors contribute to hazard accumulation, and their relative importance varies more across space.

Importantly, model-specific differences in how features are ranked highlight the impact of modeling assumptions. AG Cox tended to amplify linear relationships (e.g., with `humidity`), while RSF and DeepSurv appeared more sensitive to context dependent interactions. For example, RSF may downplay meteorological variables individually but still use them effectively via threshold based rules involving topography or vegetation.

Together, these findings suggest that feature relevance is not fixed but rather conditional on both the geographic setting and the modeling approach.

5.2.2 Interrelated Covariates and Multicollinearity

Although feature importance rankings provide important insights, their interpretation is complicated by multicollinearity among predictors. Many of the covariates such as temperature, drought, humidity, and vegetation cover are interrelated due to shared seasonal or climatic cycles (some features are also correlated e.g. forest and forest combined). In linear models like AG Cox, this multicollinearity can inflate the variance of coefficient estimates[22], making it difficult to reliably interpret the individual contribution of each feature.

Even in non-linear models like RSF and DeepSurv, highly correlated inputs can impact model training. While these models are more robust to such effects and can capture complex interactions, redundancy among covariates may reduce effective learning capacity or lead to instability in permutation based importance measures. This effect is likely amplified in the Mediterranean setting, where the larger spatial area results in greater variability within covariates such as forest cover, slope, and shrubland leading to more overlap in the information they provide. By examining our correlation and VIF tables (A.10,A.11), we see several near-perfect relationships that we did not explicitly address in our model work. Future work could explore dimensionality reduction or feature decorrelation techniques to improve robustness and interpretability.

5.2.3 Implications for Interpretability and Management

Many of the top-ranked features align with established environmental drivers of wildfire ignition, such as fuel dryness and prior fire activity[9]. This alignment lends both interpretability and credibility to the model outputs, strengthening their potential for operational usage. For instance, areas with consistently high `neighbor_fire_risk` could be prioritized for fuel management or early interventions, while features like `humidity`, which consistently influences risk, can support short-term fire weather alerts.

Crucially, the relative importance of features varied by region and model, underscoring the need for context aware interpretation. Understanding how environmental, geographic, and temporal factors influence model behavior is essential for translating predictions into effective management strategies.

Overall, these results highlight that effective wildfire prediction depends not only on model architecture, but also on careful feature engineering and contextual interpretation.

5.3 Contour and graph plots

5.3.1 Campania

Looking at the first contour plot in our results 4.1 we see an extremely low survival rate for longer durations, and if we compare it to the calibration plot for the same model4.4a we see that the model is too pessimistic, it predicts lower survival rates in the periods from [100 : 500] days however when we look at periods in the 2000 day range it is too optimistic. If we compare to the more stable DeepSurv calibration plot 4.4c and its respective hazard and survival plot (see appendixA.1) We see that it is more optimistic in its predictions regarding survival in the most fire prone grid. However they collectively rate this as a high risk grid, which is true since this grid has 146 recorded fires in the span of 13 years.

Interestingly, the RSF calibration curve (Figure 4.4b) differs strongly from the 45° “ideal” line and instead shows a pronounced zig zag pattern. One reason for the shape of the RSF calibration curve could how the model makes predictions. Each survival tree in the forest gives a rough, step estimate of survival over time. When all the trees are combined, the result is still not a smooth prediction and can only take on a few specific values [13]. This may have caused the predicted survival probabilities to “jump” rather than follow a smooth line, which makes the calibration curve bounce above and below the ideal 45° line. Interestingly, the RSF model for the temporal split didn’t behave this way A.2b which is weird since it was modeled in the same way besides train and test input.

Fire Probability Map

Our spatial ‘worst vs. best’ analysis (figure 4.3a) highlights how much local fire probabilities can swing depending on seasonal extremes. Under the ‘worst case’ combination of low humidity (25th percentile), severe drought (25th percentile drought20) and high temperature (75th percentile), many grid cells in Campania show increases in predicted fire probability of 0.0-0.25 compared to the ‘best case’ scenario (75th percentile humidity, 75th percentile drought20, 25th percentile temperature). The deep reds on the map mark the areas most sensitive to these climatic effects, whereas the near-blue cells are relatively less affected by topography, vegetation cover or do not experience the necessary dry and warm conditions.

Hazard Evolution Plot

Focusing on a single representative cell (ID 508) in Figure 4.3b, we see that “best-case” conditions (blue/orange curves) produce a slowly rising cumulative hazard that peaks around 3-4 hazard-units before leveling off, reflecting how cooler, moister months reduce the likelihood of ignition. In contrast, “worst-case” conditions (green/red curves) drive the hazard sharply upward, exceeding 5-6 units by late season demonstrating how the combination of heat, dryness, and drought accelerates risk accumulation. Moreover, if a fire occurs early in the season (day 110, April), it “resets” fuel loads and moisture conditions so that risk remains relatively low when the landscape reaches its peak flammability around August. In contrast, a fire late in the season means there is longer recovery time, so hazard by the following August is significantly higher. Finally, in this historically fire-prone cell, longer intervals without fire allow risk to build up over time due to potential fuel accumulation even if this pattern is not directly reflected in the model’s learned coefficients (e.g. days_since_last_fire).

Future work could integrate long term climate projections to forecast how the red zones on our probability map might expand or intensify under warming scenarios.

5.3.2 Mediterranean

In figure 4.9, we compare two contrasting Mediterranean settings: a “high-risk” grid cell (ID 41008, among the top five most fire-prone) on the left, and a “lower-risk” cell (ID 37298) on the right. In the high-risk cell, even immediately after a fire (Days Since Fire ≈ 0), the one-year survival probability is already below 0.6, and by two years it drops to the 0.3-0.4 range. In contrast, the low-risk cell maintains survival probabilities above 0.7 at one year and stays around 0.6-0.7 even after two years.

These differences are closely aligned with the key features identified by the RSF permutation based importance analysis (Figure 4.8). The dominant drivers `neighbor_fire_risk`, `days_since_last_fire`, `cos_month`, and `humidity` highlight the roles of local fire history and seasonal moisture. The monthly climate profiles show that grid 41008 tends to experience lower humidity and higher temperatures across the fire season (A.4), leading to faster reaccumulation of hazard after a fire. Static landscape features such as `slope30` and `mean_elev` also contribute, but climatic extremes like `drought20` and `temperature` rank relatively low in importance once these stronger effects are included in the model.

Both contour plots also display the recurrent “waves” of seasonality, each annual return of the fire season carves out a horizontal band of lower survival probability. These bands become more distinct as time since last fire increases (moving rightward), forming layered seasonal dips. In the high-risk cell, these dips are deeper and broader, reflecting stronger sensitivity to both fire history and seasonal moisture, especially humidity.

These patterns again highlight the need to adapt fire management to local fire history and seasonal sensitivity some areas require fast intervention after a single fire, while others allow longer recovery without escalating hazard.

Hazard Evolution Plot

In the hazard evolution plot 4.10a we again compare two ignition points, one early in the season (calendar day 110) and one later (day 280) under both best and worst-case environmental conditions. Fires igniting later in the season for this grid generally result in a more rapid accumulation of hazard (red line) while remaining low throughout the remainder of the season while early fires in the season (green/blue line) still peaks around the fire season in August - September. We would assume that the fire risk would be lower for the early season fire (day 110) but although there may have been a fire that doesn't mean that all the fuel in that cell has burned. What also is shown is that if we have a late season fire the risk throughout the early season remains relatively low. Note that in our overlay plots (4.10a, 4.3b), we reset the timeline at the moment of ignition, so t=0 represents the point immediately after a fire. At this time, the cumulative hazard is by definition zero, as no risk has yet been accumulated since the reset. This does not imply that the instantaneous hazard (the immediate risk of a new ignition) is zero, only that we have not yet integrated any hazard over time. This reset allows us to isolate and analyze post-fire risk dynamics more clearly. The low starting value also reflects how features like `days_since_last_fire` influence risk (e.g. when 1), which may further influence model predictions immediately after a burn.

Fire Probability Map

The change in fire probability 4.10b effectively captures the regional sensitivity to different environmental conditions. Warmer colors suggests these areas are particularly vulnerable to changes in climate changes. Such zones are of high risk. Worth noting is that these zones may experience larger temperature increases and longer droughts compared to other regions. factors that, beyond the baseline grid characteristics, further elevate their fire risk. The blue areas show minimal difference in fire probability, this doesn't mean that

these are not high risk areas (see A.5) however it just means that tuning on the weather metrics do not yield that much of a difference for these areas.

Calibration

Looking at the calibration 4.7 we see that our models are fairly optimistic in general, having lower observed survival probability than the predicted (besides 4.7b). The fairly consistent optimism, can maybe be explained by the sparsity of grids that experience fire or not, meaning that the margin for grids that survive a long time is really high and therefore the models might recognize that we generally survive a long time if we look at the region as a whole.

5.3.3 Impact of Ambiguous Temporal Features on Calibration

Although our models achieve high concordance indices and favorable Brier scores, the calibration plots reveal a more nuanced picture. Specifically, we observe that while the models can effectively rank risk, distinguishing between higher and lower risk areas, they struggle with the absolute magnitude of that risk. One possible explanation for this discrepancy lies in the treatment of the `days_since_last_fire` feature.

In our data setup, `days_since_last_fire` was assigned a value of 1 in two very different contexts: (1) for intervals immediately following a recorded fire (representing genuine post-fire risk), and (2) for intervals with no prior fire history, where the value was simply a placeholder due to lack of information. While this approach avoided negative or missing values, it introduced a confounding effect where the same numeric value could signal either high risk or complete uncertainty.

This conflation may have led the models to fit well in terms of ranking (which is what the C-index captures), while still producing biased risk calibrations, as the absolute predicted probabilities were influenced by a noisy or ambiguous feature. In effect, the models might “know” which cells are relatively more or less risky, but struggle to assign an accurate probability of fire occurrence when important temporal information is unreliable or non-informative.

5.4 Limitations and Future Work

5.4.1 Feature Engineering and Labeling Challenges

One of the main challenges in our modeling was the treatment of grid cells with no recorded fire history. While a feature like `no_prior_fire` could indicate absence of previous events, it introduced complications in handling the associated variable `days_since_last_fire`. Setting it to NaN was not viable due to model constraints, and using an arbitrarily large numbers like 10000 risked overfitting, information leaking or skewing calibration in another direction. We attempted to reduce this discrepancy manually, but doing so introduced tradeoffs with feature balance and noise, complicating model behavior. While `days_since_last_fire` captures broad temporal hazard trends, it cannot fully account for recovery such as vegetation regrowth, fuel accumulation, or soil moisture. These processes unfold and depend on local conditions, meaning this may oversimplify real-world ignition dynamics especially in areas with multiple consecutive fires.

5.4.2 Fire Extent and Land Cover Limitations

Although we had access to fire size per event, we lacked detailed data on how much of each grid cell’s fuel material was actually consumed. Therefore estimating post-fire hazard was not fully optimized, especially without a mechanism to dynamically update land cover or fuel availability. Without this, our hazard predictions could not fully reflect changes to the landscape caused by previous fires.

5.4.3 Data Quality and Reporting Bias

Another important limitation concerns the completeness and consistency of the fire occurrence data. The number of recorded fire events in the Campania dataset is disproportionately high relative to the Mediterranean dataset (even though Campania also is in that set), which may reflect better reporting infrastructure in certain areas rather than a true difference in fire incidence. This introduces potential bias: regions with underreported fires could appear deceptively low-risk to the model. Additionally, fires may be logged on incorrect dates due to late detection, especially in remote areas. In such cases, rolling averages on time-varying covariates (e.g., temperature or drought metrics) help smooth over these discrepancies, but they do not fully eliminate the underlying uncertainty. This limitation highlights the need for caution when interpreting model outputs across regions with variable data quality.

5.4.4 Spatiotemporal Data Leakage and Feature Correlation

Many adjacent grid cells shared identical temperature and humidity readings due to coarse sampled weather data. This creates small forms of data leakage, particularly when simultaneous fires occur across neighboring cells. Moreover, constructing neighborhood-based features (like `nbr_fire_risk`) was complicated by potential information leakage from the future. Our workaround restricting neighborhood calculations to the training set limited information use but also reduced performance in spatial splits. Future work for neighboring cell could involve CNN's to learn neighboring more complex patterns that can capture effects we didn't with the distance weighting.

5.4.5 Model Sensitivity to Feature Selection

Model performance varied across validation splits depending on feature inclusion. Removing noisy features sometimes had no effect or even slightly degraded performance due to complex model interactions. For instance, a feature that marginally improved the C-index might worsen the Brier score. Even sometimes keeping the temperature showed drop in c-index performance 4.8c. The only consistently removed feature was wind speed, which failed to improve any model across regions. This may have been solved by training different Random Forest models to get best feature selection on each of the 10 models, but we ran into time constraints.

5.4.6 Limitations of Temporal Evaluation

As mentioned before, the bootstrapping offered a way to estimate variability in the temporal split, its credibility is limited by the use of only one split. Ideally, we would have evaluated temporal generalization using multiple, time specific cross-validation folds. However, designing such splits is time consuming and requires careful data handling to avoid overlap in covariate space which we didn't have enough time for again.

5.4.7 Metric and Model Limitations

Attempts to compute dynamic AUCs [23] using `sksurv.metric.cumulative_dynamic_auc` were unsuccessful due to unresolved technical errors. Similarly, DeepSurv models trained on the Mediterranean data displayed erratic behavior, despite using the same framework as for Campania. As a result, we were unable to incorporate additional evaluation metrics to broaden our model comparison.

5.4.8 Generalization and Transferability

A key limitation of our current setup is that each region Campania and the broader Mediterranean is modeled and validated largely in isolation. While our models generalize well within each domain, we have not yet tested their robustness across regions with different environmental, climatic, and topographic profiles. One promising future direction would

be a cross regional generalization test training on the Mediterranean region while holding out Campania entirely for testing (also the location out from the Mediterranean set). This would provide a stronger evaluation of how well our models transfer to geographically different but climatically similar areas. However, this approach is currently hindered by differences in grid resolution across datasets, requiring substantial preprocessing of features.

Another direction to explore is the use of spatial-temporal blocking, where we would train on a specific geographic block over a given time period (e.g., years 0-x) and test on a different block over a subsequent period (e.g., years x-y). This setup would better mimic real-world usage scenarios, where models trained on past data are applied to new regions and future events. This would require a large and consistent dataset, one reason Campania alone may be insufficient, but maybe it could be done on the Mediterranean set.

Finally, transferability could be improved by incorporating ecological dynamics. For instance, estimating post-fire fuel recovery using vegetation indices like NDVI or EVI, or satellite-derived soil moisture data [24] could help better capture how ignition risk evolves as landscapes recover. Integrating these into time-to-event frameworks is a challenge, but would significantly enhance model realism, particularly in regions where vegetation recovery rates strongly influence future fire risk.

6 Conclusion

In this thesis, we have framed wildfire ignition as a dynamic survival problem and benchmarked three modeling approaches: Andersen-Gill Cox (AG Cox), Random Survival Forests (RSF), and DeepSurv, across two distinct regions (Campania and the Mediterranean area) and two validation strategies (spatial blocking and temporal splitting). Our evaluation in Campania showed that DeepSurv achieved the highest concordance index, suggesting it might capture nonlinear interactions in meteorological variables. In the broader Mediterranean setting, DeepSurv models proved unstable and could not be reliably trained, so RSF appeared as the most consistent approach, though this reflects model feasibility rather than definitive superiority. AG Cox still had strong interpretability but lagged a bit behind in predictive accuracy in terms of event ranking.

Calibration and Brier score analysis revealed that all models offer modest improvements over the classical Cox baseline, with DeepSurv showing the greatest gain under temporal validation in Campania. Spatial validation, however, exhibited substantial fold-to-fold variability, highlighting geographic variety in ignition drivers. Temporal validation produced tighter uncertainty bands that may overstate confidence when only a single split is used.

Across models and regions, `days_since_last_fire` consistently emerged as one of the most important features, although its interpretation is complicated by how fire history was encoded. AG Cox and DeepSurv placed greater emphasis on meteorological and landscape variables, such as `humidity`, `temperature`, and `shrubland`, while RSF prioritized recurrence-related features, especially under spatial validation. However, RSF's lower attribution to weather-related features may stem from the limitations of permutation based importance, which can underestimate nonlinear effects or threshold-driven interactions.

Practically, the survival-analysis framework provides continuously updating hazard estimates that pinpoint not only where but also when ignition risk peaks. These dynamic forecasts can guide proactive interventions, fuel-load handling, firebreak construction, and resource allocation in alignment with SDG 13 (Climate Action) and SDG 15 (Life on Land).

Several limitations hinder our findings. Data sparsity and high censoring rates in no fire cells limit confidence in certain areas. The simple “time since fire” feature cannot fully capture complex fuel recovery processes, and our feature encoding blurred two clearly different temporal states. Moreover, models were evaluated in isolation per region, so true regional transferability remains untested.

For future work, improving temporal features with satellite derived indicators of vegetation regrowth and soil moisture would improve post fire risk modeling. Integrating spatial learning components (e.g. convolutional neural nets) within survival frameworks could capture neighborhood effects better. Using transfer learning or domain adaptation could improve model performance when applied to new regions. Finally, close collaboration with fire-management agencies to incorporate controlled burn schedules and suppression activities would facilitate operational deployment of these survival models. However evaluating the threshold for acceptable risk is not within the scope for our models.

To summarize, this thesis demonstrates that dynamic survival analysis offers a powerful, flexible way for forecasting wildfire ignitions. AG Cox delivers clear interpretability, RSF

offers reliable generalization across regions, and DeepSurv achieved the highest predictive accuracy. This together enables more precise, proactive fire-risk management over both space and time.

Bibliography

- [1] Jesús San-Miguel-Ayanz et al. *Forest Fires in Europe, Middle East and North Africa* 2023. 2024. DOI: 10.2760/8027062.
- [2] Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2022: Impacts, Adaptation and Vulnerability*. IPCC Sixth Assessment Report, Working Group II, 2022. URL: <https://www.ipcc.ch/report/ar6/wg2/>.
- [3] Earth.Org. *The Environmental Impact of Wildfires*. 2023. URL: <https://earth.org/environmental-impact-of-wildfires/>.
- [4] Kevin D. Bladon et al. "Wildfire and the Future of Water Supply". In: *Environmental Science & Technology* 48.16 (2014), pp. 1411–1416. URL: <https://pubs.acs.org/doi/10.1021/es500130g>.
- [5] Bay Area Council Economic Institute. *The True Cost of Wildfires*. 2021. URL: <https://www.bayareaeconomy.org/report/the-true-cost-of-wildfires/>.
- [6] MarketWatch. *L.A. wildfires have caused more than \$135 billion in economic losses – and counting*. 2025. URL: <https://www.marketwatch.com/story/la-wildfires-have-caused-more-than-50-billion-of-economic-loss-and-counting-f57662eb>.
- [7] Clara Ochoa, Avi Bar-Massada, and Emilio Chuvieco. "A European-scale analysis reveals the complex roles of anthropogenic and climatic factors in driving the initiation of large wildfires". In: *Science of the Total Environment* 917 (2024), p. 170443. DOI: 10.1016/j.scitotenv.2024.170443. URL: <https://doi.org/10.1016/j.scitotenv.2024.170443>.
- [8] Mahsa Salehi et al. "Dynamic and Robust Wildfire Risk Prediction System: An Unsupervised Approach". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, CA, USA: Association for Computing Machinery, 2016, pp. 245–254. DOI: 10.1145/2939672.2939685. URL: <https://doi.org/10.1145/2939672.2939685>.
- [9] Piyush Jain et al. "Observed increases in extreme fire weather driven by atmospheric humidity and temperature". In: *Nature Climate Change* 12 (2021), pp. 63–70. URL: <https://www.nature.com/articles/s41558-021-01224-1>.
- [10] Douglas G. Woolford et al. "Characterizing temporal changes in forest fire ignitions: Looking for climate change signals in a region of the Canadian boreal forest". In: *Environmetrics* 21.5 (2010), pp. 789–800. DOI: 10.1002/env.1067.
- [11] Pier-Olivier Tremblay, Thierry Duchesne, and Steven G. Cumming. "Survival analysis and classification methods for forest fire size". In: *PLOS ONE* 13.1 (2018), e0189860. DOI: 10.1371/journal.pone.0189860. URL: <https://doi.org/10.1371/journal.pone.0189860>.
- [12] Per Kragh Andersen and Richard D. Gill. "Cox's Regression Model for Counting Processes: A Large Sample Study". In: *Annals of Statistics* 10.4 (1982), pp. 1100–1120. DOI: 10.1214/aos/1176345976. URL: <https://projecteuclid.org/journals/annals-of-statistics/volume-10/issue-4/Coxs-Regression-Model-for-Counting-Processes--A-Large-Sample/10.1214/aos/1176345976.full>.
- [13] Hemant Ishwaran et al. "Random survival forests". In: *The Annals of Applied Statistics* 2.3 (2008), pp. 841–860. DOI: 10.1214/08-AOAS169. URL: <https://doi.org/10.1214/08-AOAS169>.
- [14] Jared Katzman et al. "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network". In: *BMC Medical Research*

- Methodology* 18.1 (2018), p. 24. DOI: 10.1186/s12874-018-0482-1. URL: <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-018-0482-1>.
- [15] Ann-Kathrin Ozga, Meinhard Kieser, and Geraldine Rauch. “A systematic comparison of recurrent event models for application to composite endpoints”. In: *BMC Medical Research Methodology* 18.1 (2018), p. 2. DOI: 10.1186/s12874-017-0462-x. URL: <https://doi.org/10.1186/s12874-017-0462-x>.
- [16] Aleksandra Kolanek, Mariusz Szymanowski, and Andrzej Raczyk. “Human Activity Affects Forest Fires: The Impact of Anthropogenic Factors on the Density of Forest Fires in Poland”. In: *Forests* 12.6 (2021), p. 728. DOI: 10.3390/f12060728. URL: <https://doi.org/10.3390/f12060728>.
- [17] Frank E. Harrell et al. “Evaluating the Yield of Medical Tests”. In: *JAMA* 247.18 (1982), pp. 2543–2546. DOI: 10.1001/jama.1982.03320430047030.
- [18] UCLA Institute for Digital Research and Education. *What is multicollinearity?* <https://stats.idre.ucla.edu/other/mult-pkg/faq/general/what-is-multicollinearity/>. Accessed: 2025-05-13. n.d.
- [19] Andreas Altmann et al. “Permutation importance: a corrected feature importance measure”. In: *Bioinformatics* 26.10 (2010), pp. 1340–1347. DOI: 10.1093/bioinformatics/btq134. URL: <https://doi.org/10.1093/bioinformatics/btq134>.
- [20] Giles Hooker, Lucas Mentch, and Siyu Zhou. “Unrestricted Permutation Forces Extrapolation: Variable Importance Requires at Least One More Model, or There Is No Free Variable Importance”. In: *Statistics and Computing* 31.6 (2021), pp. 1–16. DOI: 10.1007/s11222-021-10057-z.
- [21] Christoph Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. <https://christophm.github.io/interpretable-ml-book/pdp.html>. Leanpub, 2022.
- [22] B. T. Babalola and W. B. Yahya. “Effects of Collinearity on Cox Proportional Hazard Model with Time Dependent Coefficients: A Simulation Study”. In: *Journal of Biostatistics and Epidemiology* 5.2 (Feb. 2020), p. 2348. DOI: 10.18502/jbe.v5i2.2348. URL: <https://doi.org/10.18502/jbe.v5i2.2348>.
- [23] Sebastian Pölsterl. “scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn”. In: *Journal of Machine Learning Research* 21.212 (2020), pp. 1–6. URL: <http://jmlr.org/papers/v21/20-729.html>.
- [24] Matthias Forkel et al. “Post-fire recovery of vegetation in boreal and temperate ecosystems in North America estimated from vegetation optical depth and related indices”. In: *Biogeosciences* 19.13 (2022), pp. 3317–3336. DOI: 10.5194/bg-19-3317-2022. URL: <https://doi.org/10.5194/bg-19-3317-2022>.

A Appendix

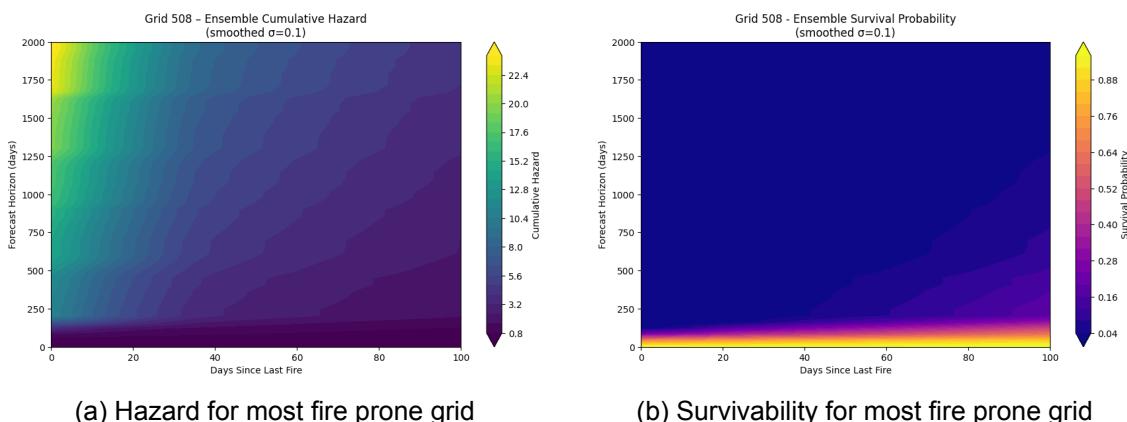
GitHub Repository

The code developed and used for the analyses presented in this thesis is publicly available at:

https://github.com/WikeMarzowsky/Wilfire_survival_analysis

Table A.1: Statistical Test Results on Bootstrap ΔC Distributions (Temporal Split)

Comparison	Shapiro-Wilk		Paired t-Test		Wilcoxon	
	<i>W</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>W</i>	<i>p</i>
RSF vs AG	0.999	0.842	259.95	< 0.0001	0.0	< 0.0001
DeepSurv vs AG	0.999	0.780	654.63	< 0.0001	0.0	< 0.0001
RSF vs DeepSurv	0.999	0.724	-452.23	< 0.0001	0.0	< 0.0001



(a) Hazard for most fire prone grid

(b) Survivability for most fire prone grid

Figure A.1: Hazard plot (left) and survival plot (right) predicted by the DeepSurv model for the most fire-prone grid cell in Campania. These illustrate the model's estimated risk dynamics over time in a high-risk area under median monthly weather conditions. (Spatial)

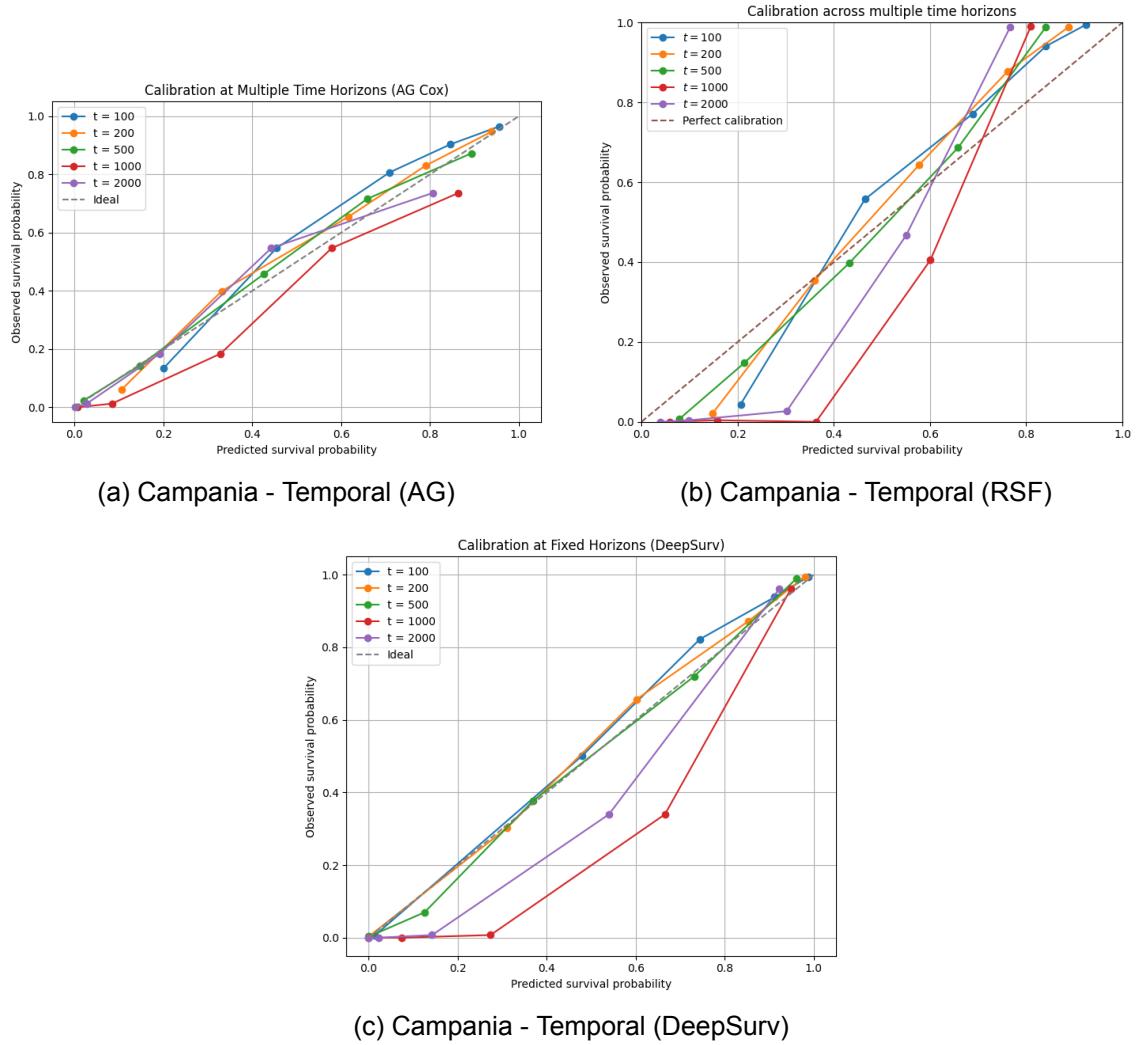


Figure A.2: Calibration plot (temporal) in time horizons for the models. The closer the graphs are to 45 degrees the better performance over time.

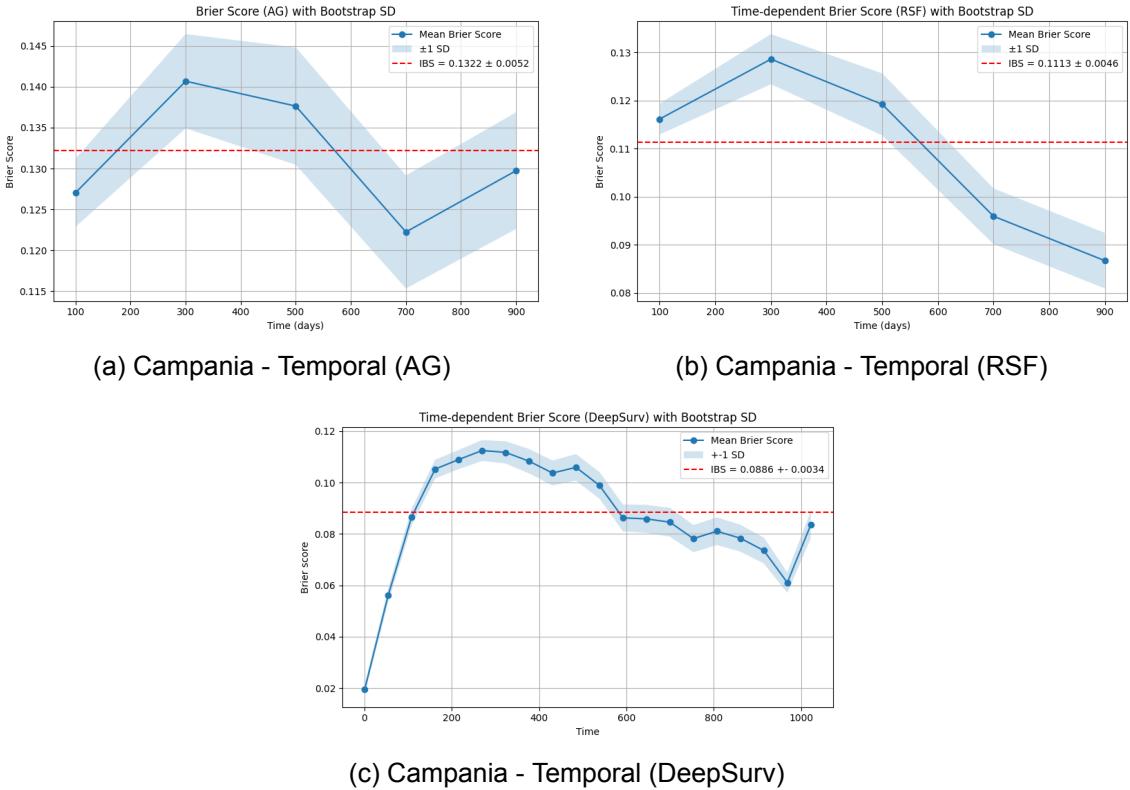


Figure A.3: Brier score curves for the Campania temporal models. Lower values indicate better predictive calibration.

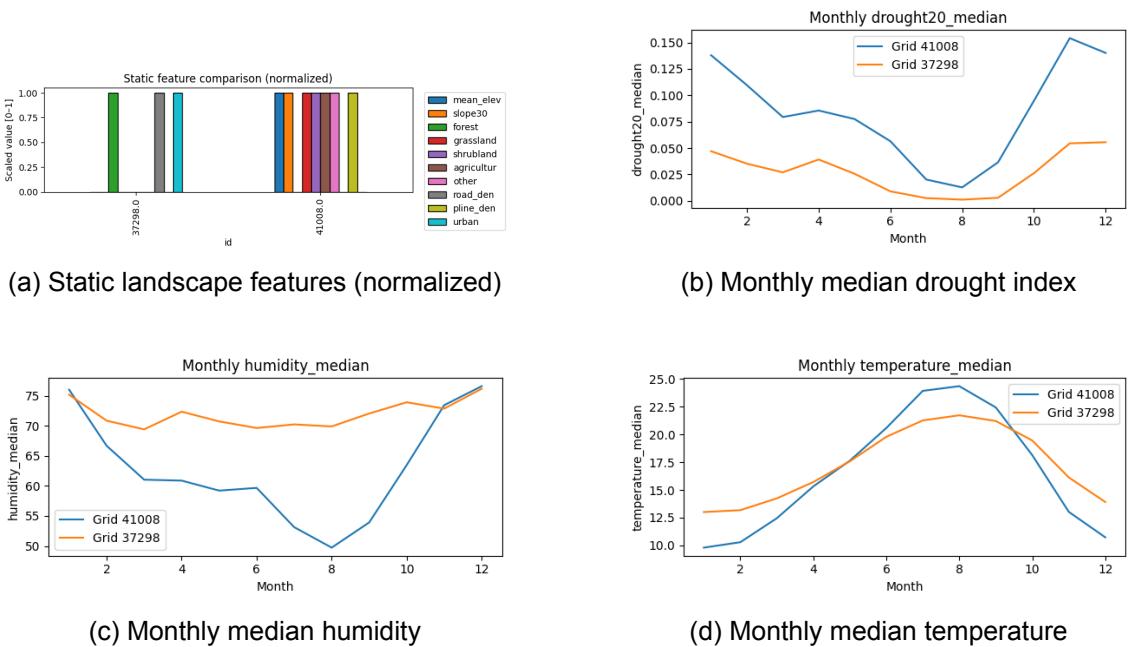


Figure A.4: Comparison between grid cell 41008 (high-risk) and 37298 (lower-risk) in the Mediterranean region. (a) Normalized static landscape features; Median monthly climatologies of (b) drought index, (c) humidity, and (d) temperature.

Model	AG Cox	RSF	DeepSurv
Campania (Spatial)	0.814	0.851	0.860
Campania (Temporal)	0.819	0.864	—
Mediterranean (Spatial)	0.865	0.905	—
Mediterranean (Temporal)	0.896	0.937	—

Table A.2: Training concordance index (C-index) for each model under spatial and temporal validation.

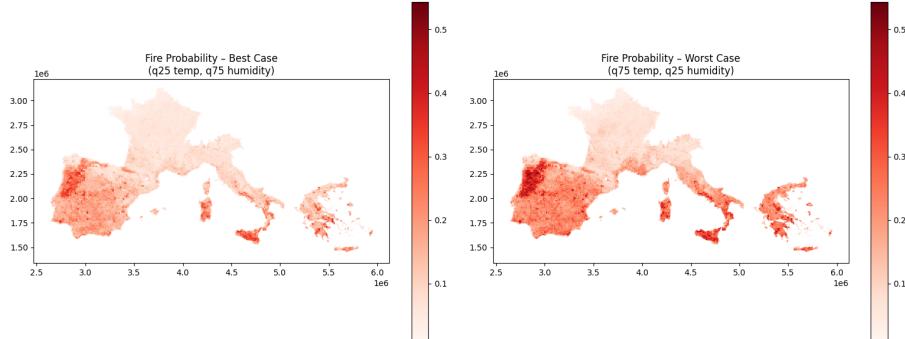


Figure A.5: Worst best scenario without overlay Mediterranean

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	P	-log2(p)
temperature	1.33	3.78	0.19	0.95	1.71	2.60	5.51	0.00	6.93	<0.005	37.81
drought20	-5.12	0.01	0.32	-5.74	-4.51	0.00	0.01	0.00	-16.26	<0.005	195.03
forest_combined	1.51	4.53	0.17	1.17	1.85	3.23	6.33	0.00	8.80	<0.005	59.33
grass	1.83	6.23	0.41	1.02	2.64	2.77	14.05	0.00	4.41	<0.005	16.59
shrub	1.63	5.12	0.35	0.95	2.31	2.60	10.10	0.00	4.72	<0.005	18.67
forest	1.51	4.53	0.17	1.17	1.85	3.23	6.33	0.00	8.80	<0.005	59.33
neighbor_fire_risk	2.52	12.42	0.18	2.17	2.87	8.74	17.64	0.00	14.08	<0.005	147.05
neighbor_powerline	0.05	1.05	0.08	-0.11	0.21	0.89	1.23	0.00	0.58	0.56	0.84
road	1.16	3.20	0.23	0.70	1.62	2.02	5.06	0.00	4.97	<0.005	20.55
powerline	0.05	1.05	0.08	-0.11	0.21	0.89	1.23	0.00	0.58	0.56	0.84
cos_month	0.08	1.08	0.05	-0.02	0.17	0.98	1.18	0.00	1.62	0.11	3.24
sin_month	-0.22	0.80	0.05	-0.32	-0.13	0.73	0.88	0.00	-4.65	<0.005	18.18
temperature_sq	0.90	2.47	0.16	0.58	1.23	1.79	3.41	0.00	5.50	<0.005	24.62
temp_forest_interaction	-2.12	0.12	0.43	-2.96	-1.27	0.05	0.28	0.00	-4.89	<0.005	19.92
days_since_last_fire	-6.53	0.00	0.36	-7.23	-5.83	0.00	0.00	0.00	-18.30	<0.005	246.20
mean_eleva	-0.85	0.43	0.21	-1.26	-0.43	0.28	0.65	0.00	-3.96	<0.005	13.72
agri	0.95	2.58	0.16	0.63	1.26	1.88	3.54	0.00	5.83	<0.005	27.46
humidity	-13.89	0.00	1.20	-16.25	-11.53	0.00	0.00	0.00	-11.56	<0.005	100.19
slope30	0.73	2.08	0.34	0.07	1.40	1.07	4.05	0.00	2.15	0.03	5.00
night_ligh	-0.41	0.66	0.37	-1.13	0.31	0.32	1.36	0.00	-1.12	0.26	1.92
other	1.35	3.85	0.51	0.35	2.34	1.42	10.42	0.00	2.65	0.01	6.95
wui	1.05	2.85	0.36	0.34	1.75	1.41	5.76	0.00	2.91	<0.005	8.10

Figure A.6: Feature table for Campania AG spatial

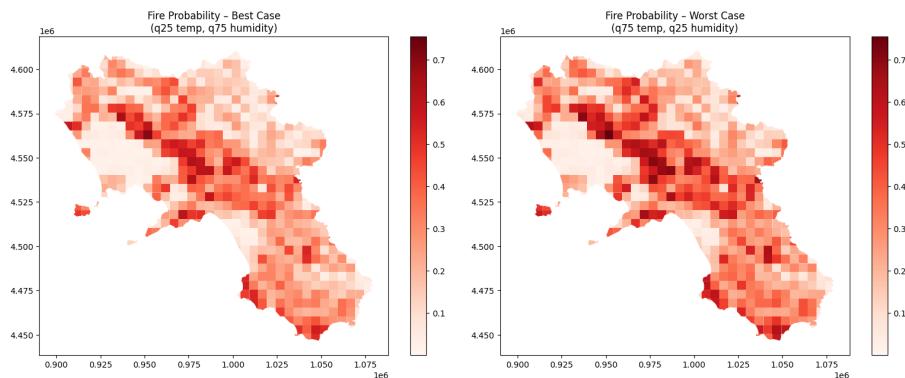


Figure A.7: Worst best scenario without overlay Campania

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
temperatur	2.09	8.10	0.14	1.81	2.37	6.12	10.71	0.00	14.65	<0.005	159.08
wind_speed	-0.02	0.98	0.11	-0.23	0.19	0.79	1.20	0.00	-0.22	0.82	0.28
drought_20	-2.03	0.13	0.24	-2.49	-1.57	0.08	0.21	0.00	-8.57	<0.005	56.46
forest_combined	0.22	1.24	0.11	0.00	0.44	1.00	1.55	0.00	1.96	0.05	4.33
nbr_fire_risk	3.60	36.74	0.15	3.31	3.89	27.52	49.05	0.00	24.44	<0.005	435.83
nbr_forest	0.81	2.24	0.14	0.53	1.08	1.71	2.95	0.00	5.79	<0.005	27.10
nbr_powerline	0.24	1.27	0.24	-0.22	0.70	0.80	2.02	0.00	1.01	0.31	1.68
cos_month	0.10	1.11	0.02	0.05	0.15	1.06	1.16	0.00	4.24	<0.005	15.47
sin_month	0.33	1.39	0.03	0.27	0.39	1.31	1.48	0.00	10.54	<0.005	83.86
temp_sq	0.00	1.00	0.00	0.00	0.00	1.00	1.00	0.00	4.25	<0.005	15.51
days_since_last_fire	-0.41	0.66	0.03	-0.47	-0.35	0.62	0.70	0.00	-13.14	<0.005	128.54
mean_elev	0.00	1.00	0.14	-0.27	0.27	0.77	1.31	0.00	0.01	0.99	0.01
slope30	1.63	5.13	0.10	1.44	1.82	4.24	6.20	0.00	16.86	<0.005	209.53
humidity	-3.14	0.04	0.09	-3.32	-2.96	0.04	0.05	0.00	-34.34	<0.005	855.91
pline_den	3.53	34.05	0.27	3.00	4.05	20.16	57.50	0.00	13.19	<0.005	129.59
agricultur	0.32	1.38	0.07	0.18	0.47	1.20	1.60	0.00	4.35	<0.005	16.17
shrubland	1.18	3.25	0.09	1.00	1.35	2.73	3.87	0.00	13.21	<0.005	129.87
grassland	1.47	4.37	0.14	1.19	1.76	3.29	5.79	0.00	10.22	<0.005	79.11
urban	-1.39	0.25	0.20	-1.78	-1.00	0.17	0.37	0.00	-6.93	<0.005	37.82
other	-0.87	0.42	0.26	-1.37	-0.36	0.26	0.69	0.00	-3.39	<0.005	10.48

Figure A.8: Feature table for Mediterranean AG spatial

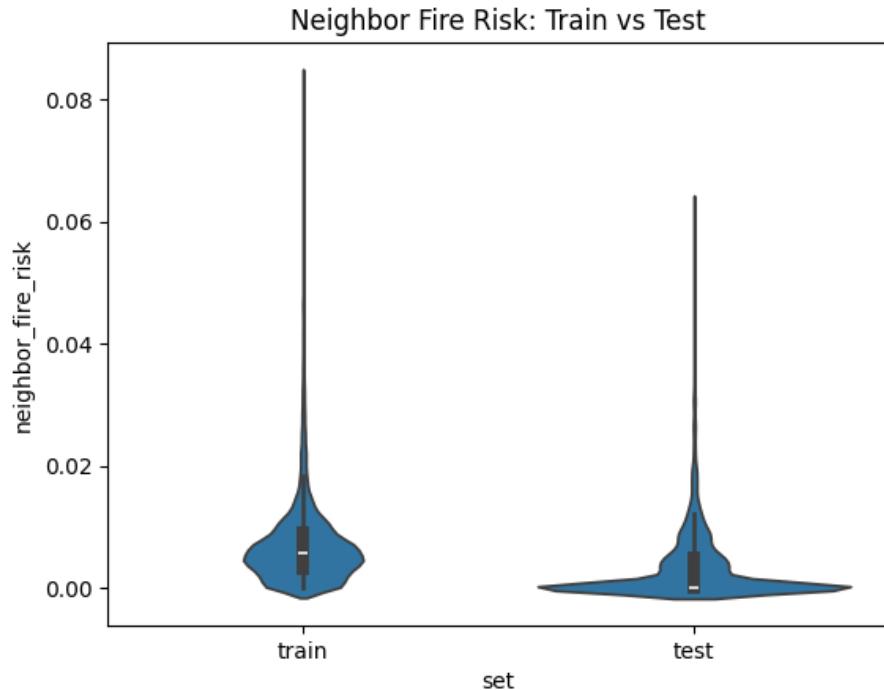


Figure A.9: Distribution of nbr fire risk for Campania spatial split

Feature	VIF
temperatur	99.0988
forest_combined	45.5467
temp_sq	44.0784
nbr_forest	28.0574
humidity	19.7327
agricultur	11.6322
nbr_powerline	11.1166
shrubland	10.1554
pline_den	7.4775
mean_elev	4.7324
cos_month	3.7065
grassland	3.2479
slope30	3.1923
urban	2.3833
drought_20	2.1951
sin_month	2.0861
nbr_fire_risk	1.9820
other	1.7427
days_since_last_fire	1.1796

Table A.3: Variance Inflation Factors for predictors in the Mediterranean dataset.

Feature	VIF
forest_combined	∞
road	∞
forest	∞
neighbor_forest	∞
neighbor_road	∞
temperature	186.4449
temperature_sq	112.5586
temp_forest_interaction	24.3221
mean_eleva	15.6372
neighbor_ERI	10.7645
humidity	6.7259
agri	6.2989
drought20	2.7605
neighbor_fire_risk	2.6318
sin_month	2.4823
shrub	2.1171
neighbor_powerline	2.0918
grass	1.6142
days_since_last_fire	1.3608

Table A.4: Variance Inflation Factors for predictors in the Campania dataset.

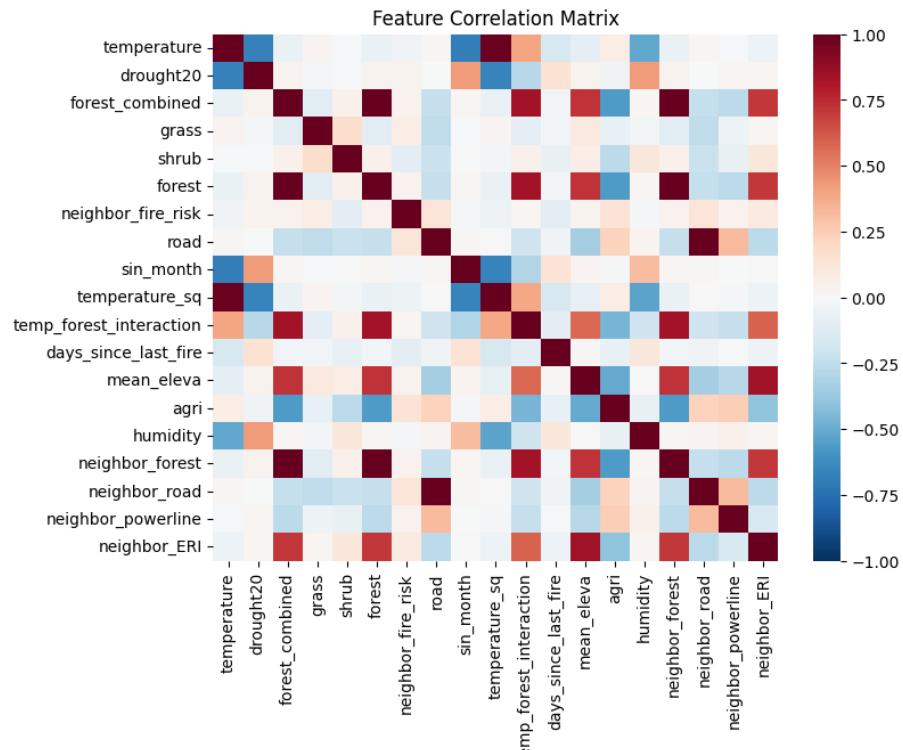


Figure A.10: Feature correlation matrix campania

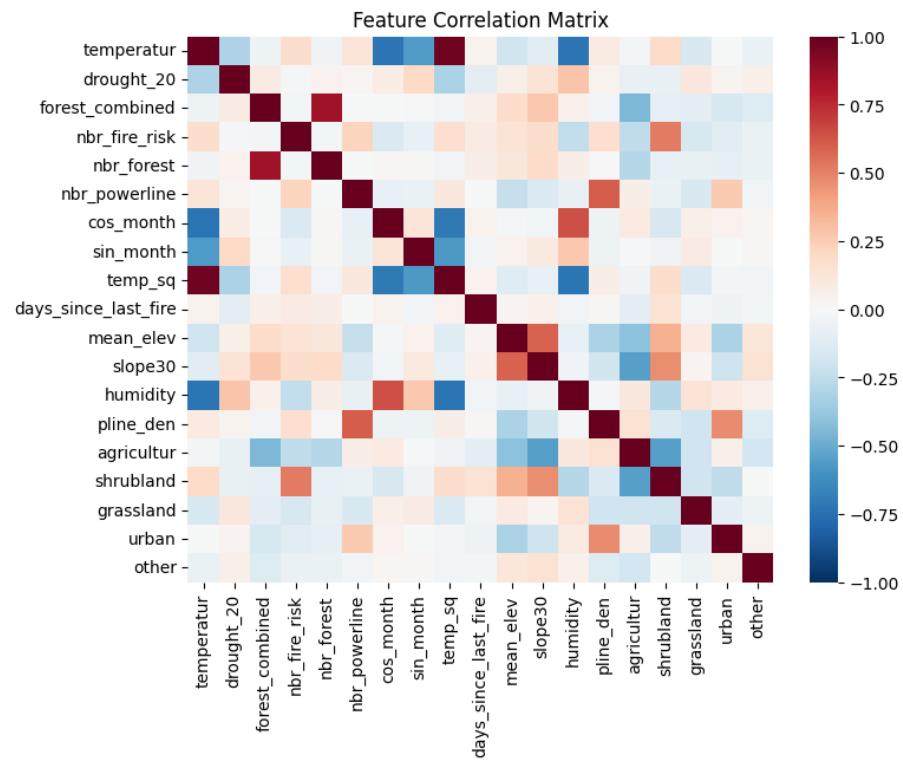


Figure A.11: Feature correlation matrix Mediterranean

Model	AG Cox	RSF	DeepSurv
Campania (Spatial)	0.800	0.814	0.831
Campania (Temporal)	0.838	0.865	0.903
Mediterranean (Spatial)	0.799	0.829	—
Mediterranean (Temporal)	0.859	0.879	—

Table A.5: Concordance index (C-index) for each model under spatial and temporal validation corrected for Mediterranean

Model	AG Cox	RSF
Integrated Brier Score (Spatial)	0.1602 ± 0.0755	0.1411 ± 0.0660
Integrated Brier Score (Temporal)	0.1167 ± 0.0012	0.0746 ± 0.0006

Table A.6: Concordance index (C-index) for each model under spatial and temporal validation corrected for Mediterranean.

Technical
University of
Denmark

Brovej, Building 118
2800 Kgs. Lyngby
Tlf. 4525 1700

www.dtu.dk