Wildfire modeling with point process and conformal prediction

Chen Xu*[1] and Daniel A. Zuniga Vazquez†[2]Rui Yao‡[2]Feng Qiu§[2]Yao Xie¶[1]

[1]Industrial and Systems Engineering, Georgia Institute of Technology
[2]Energy Systems Division, Argonne National Laboratory

Abstract

Due to significant societal and environmental impacts, obtaining a more informed understanding of wildfire activities is always important. This work uses historical data to focus on wildfire pattern recognition, prediction, and subsequent uncertainty quantification. We propose an interpretable and flexible marked spatio-temporal point process model to accomplish the tasks and adopt recent advances in time-series conformal prediction. Through extensive real-data experiments, we demonstrate the effectiveness of our methods against competing baselines.

1 Introduction

In this paper, we specifically want to address the following three research questions for a more informed understanding of wildfires. (i) Predict *binary* fire occurrences at different location and time, given only *one-class* historical observations and available features. Challenges arise as data have a complex spatio-temporal dependency. (ii) Explore interaction among different geographic regions and feature contributions in interpretable ways. Importantly, building classifiers alone is insufficient. (iii) Quantify uncertainty in multi-class fire size predictions by constructing prediction sets containing true fire size variables with high probability. Doing so is largely unexplored for such spatio-temporal data.

We tackle the questions and challenges above through the following contributions:

- We propose a marked spatio-temporal model inspired by the Hawkes process [Hawkes, 1971] for modeling wildfire occurrences. The model only requires one-class fire occurrences as training data to save data storage and can make binary predictions under dynamic thresholds at any geographic location when the training data are irregularly distributed in space. In addition, our model interpretably quantifies interactions among different regions and assesses the feature contribution to wildfire hazards explicitly or implicitly. Lastly, the model parameters are efficiently estimated to high accuracy using an alternating convex optimization approach, in contrast to the more expensive expectation-maximization method [Reinhart, 2018].

- We build prediction sets for multi-class fire size classifiers using a time-series conformal prediction method. The sets intuitively quantify uncertainty in the point predictions, where

*Emails: cxu310@gatech.edu †dzunigavazquez@anl.gov ‡ryao@anl.gov §fqiu@anl.gov ¶yao.xie@isye.gatech.edu

large sets indicate more significant uncertainty. The method is computationally efficient and suitable for any multi-class classifier.

- Through extensive real-data experiments, we verify our models' competitive or better performances against other baseline methods.

1.1 Literature review

Wildfire prediction and modeling is an essential procedure for analyzing the occurrence of wildfire events. There have many indices, such as the burning index [Schoenberg et al., 2007] and the fire danger index [Sanabria et al., 2013] for general awareness of fire risks. Despite their popularity, these indices often fail to account for events' interactions. Meanwhile, regression-based approaches [Jain et al., 2020] are more flexible and often yield satisfactory prediction results. However, their performance can be sensitive to the number of available observations per location and thus not applicable under arbitrary spatial granularity with a fixed amount of training data. Lastly, stochastic point-process models [Koh et al., 2021] have been leveraged to examine the conditional fire risk given past data and allow a deeper understanding of the underlying stochastic mechanism. However, how to convert these probabilistic predictions into more intuitive ones (e.g., categorical prediction) for forestry managers and utility owners remains a challenge.

Since our proposed fire occurrence model is based on the Hawkes process, we briefly survey existing methods in a broader context. Initially proposed in [Hawkes, 1971], the Hawkes process is a stochastic temporal point-process model for rates of events conditioning on historical ones. There have been many extensions that take into account spatial interactions [Gabriel and Diggle, 2009] and influences by additional features (i.e., marks) [Scargle, 2004]. Neural-network-based Hawkes process models [Mei and Eisner, 2017a] have also been proposed for greater model expressiveness. Despite their emerging popularity and flexibility, how to make a prediction based on rate estimates and comparisons against predictive models have been less well studied.

We briefly survey conformal prediction, the primary tool used for constructing prediction sets that quantify uncertainty in fire size classification. Originated in the seminal work [Shafer and Vovk, 2008], conformal prediction has gained wide popularity for uncertainty quantification [Zeni et al., 2020]. It is particularly appealing as the methods are distribution-free, model-agnostic, and easily implementable. The only assumption is that observations are exchangeable (e.g., i.i.d.). On a high level, conformal prediction methods assign non-conformity scores to potential outcomes of the response variable. The outcomes that have small non-conformity scores are included in the prediction set. Many methods follow this logic with promising results [Romano et al., 2020, Xu and Xie, 2021]. More recently, there have also been works that relax the exchangeability assumption [Tibshirani et al., 2019, Barber et al., 2022], but time-series conformal prediction methods are still limited, and their applications to wildfire predictions remain largely unexplored.

2 Method

2.1 Marked spatio-temporal Hawkes process

Suppose we have a sequence of n spatial-temporal-contextual observations (e.g., fire incidents), where each observation consists of time, location, static location-specific marks, and dynamic marks:

$$x_i = (t_i, u_i, z_i, m_i),\ i = 1, \ldots, n$$

In particular, we require $t_i \in [0, T]\ \forall i, t_i < t_{i+1}$, $u_i \in \{1, \ldots, K\}$ for K locations[1], $z_i \in \mathcal{Z} \subset \mathbb{R}^q$, and $m_i \in \mathcal{M} \subset \mathbb{R}^p$. Common examples of the static marks z_i include the road type and facility

[1] For example, for fire data, we discretize the total space into K disjoint regions

infrastructure, which are fixed in a given location k. Hence, each static mark can be equivalently denoted as z_k. Meanwhile, dynamic marks m_i such as current weather and season only depend on time and location, so that they can be denoted as m_{tk}.

We model these event data using a marked spatio-temporal Hawkes processes. In particular, given the σ-algebra $\mathcal{H}_t$ that denotes all historical fire occurrence before time t, the conditional intensity function of the Hawkes process is the probability of an event occurring at time t and location k, with current mark $m := \{m_{tk}, z_k\}$:

$$\lambda(t, k, m|\mathcal{H}_t) = \lim_{\Delta t, \Delta u \to 0} \frac{\mathbb{E}\left[N([t, t+\Delta t) \times B(u, \Delta u) \times B(m, \Delta m) \mid \mathcal{H}_t\right]}{\Delta t \times B(u, \Delta u) \times B(m, \Delta m)}, \tag{1}$$

where $B(a, \Delta a)$ is a ball centered at a with radius Δ_a and N is the counting measure. The ball around u, which is a discrete integer, is around the actual geographic location in longitude and latitude which u represents.

For notation simplicity, we drop $\mathcal{H}_t$ in (1) from now on. A common way to express (1) is

$$\lambda(t, k, m) = \lambda_g(t, k) f(m|t, k) = (f(k) + \sum_{j:t_j<t} f(u_j, k, t_j, t)) f(m_{tk}, z_k), \tag{2}$$

which factors the conditional intensity into product of ground process $\lambda_g(t, k)$ and conditional density $f(m|t, k)$. In (2), $f(k)$ is the baseline intensity, $f(u_j, k, t_j, t)$ measures spatial and temporal influence from event happening at t_j in u_j till current time t, and $f(m_{tk}, z_k)$ measures the influence of marks on current density. In practice, we may assume minimal contribution from fire incidents long ago in the past, so that we can also truncate $\sum_{j:t_j<t}$ to most recent events.

In general, functions $f(k)$, $f(u_j, k, t_j, t)$, and $f(m_{tk}, z_k)$ can take many forms [Xu and Schoenberg, 2011, Mei and Eisner, 2017b]. Such choices often depend on the application of interest. For computation simplicity and model interpretability, In this work, we parametrize $\lambda_g(t, k)$ in (2) as

$$f(k) = \delta_k, \quad f(u_j, k, t_j, t) = \alpha_{u_j,k} \beta e^{-\beta(t-t_j)}, \tag{3}$$

where $\widetilde{\beta} = \{\delta, A, \beta\}$, $\delta = \{\delta_k\}_{k=1}^K, A = [\alpha_{i,j}]_{i,j=1}^K$is the set of parameters with dimension $\kappa = K+K^2+1$. As we shall see, parametrizations in (3) are interpretable and yield a convex optimization problem that can be efficiently solved.

The choice of $f(m_{tk}, z_k)$ is more flexible. We consider two approaches in the experiments:

$$f(m_{tk}, z_k) = d^T m_{tk} + \theta^T z_k \qquad (\texttt{ExplicitSTHawkes}) \tag{4}$$

$$f(m_{tk}, z_k) = \hat{f}(m_{tk}, z_k) \qquad (\texttt{ImplicitSTHawkes}), \tag{5}$$

where (4) (resp. (5)) explicitly (resp. implicitly) parametrizes the effect of marks on fire risk; together with (3), we hence name this parametrization of $\lambda(t, k, m)$ `ExplicitSTHawkes` (resp. `ImplicitSTHawkes`). Note that `ExplicitSTHawkes` based on (4) is more interpretable because each estimate has a specific physical meaning, which as we will show, can also be accurately estimated via convex optimization. On the other hand, `ImplicitSTHawkes` is more flexible due to the choice of feature extractors and the form of input features. We will compare both formulations in experiments.

2.2 Parameter estimation and prediction

Based on [Reinhart, 2018], we can derive and simplify the log-likelihood of $x_1, \ldots, x_n$ as follows:

$$\begin{aligned} l(\widetilde{\beta}) &= \sum_{i=1}^{n} \log(\lambda_g(t_i, u_i)) + \sum_{i=1}^{n} \log(f(m_i|t_i, u_i)) - \sum_{k=1}^{K} \int_0^T \lambda_g(\tau, k) d\tau \\ &= \sum_{i=1}^{n} \log(\delta_{u_i} + \sum_{j:t_j<t_i} \alpha_{u_j,u_i} \beta e^{-\beta(t_i-t_j)}) + \sum_{i=1}^{n} \log(f(m_i|t_i, u_i)) - \sum_{k=1}^{K} T\delta_k - \sum_{i=1}^{n} (\sum_{k=1}^{K} \alpha_{u_i,k})(1 - e^{-\beta(T-t_i)}). \end{aligned} \tag{6}$$

Algorithm 1 Location-wise Dynamic Threshold Selection

Require: Estimates $\{\hat{\lambda}(t, k, m)\}_{t=1}^{T}$, Hyperparameters $\{\tau_{k,\min}, \tau_{k,\max}, \eta_k, \delta_k, a_{1k}, a_{2k}\}$, and true anomalies $\{Y_{tk} \in \{-1, 1\}\}_{t=1}^{T}$, revealed individually after each prediction.

Ensure: Decision thresholds $\{\tau_{tk}\}_{t=1}^{T}$, anomaly estimates $\{\hat{Y}_{tk}\}_{t=1}^{T}$ where -1 (resp. 1) indicates normal (resp. anomaly)

Define $\Pi(x) := \underset{\tau \in [\tau_{k,\min}, \tau_{k,\max}]}{\arg\min} (\tau - x)^2$.
$\tau_{1k} := \tau_{k,\min}$ and $\hat{Y}_{1k} = 1$ if $\hat{\lambda}(1, k, m) > \tau_{1k}$.
if $\hat{Y}_{1k} \neq Y_{1k}$ **then**
    $\tau_{2k} := \max(\Pi(\tau_{1k} + \eta_k \hat{Y}_{1k}, \hat{\lambda}(1, k, m)/a_{1k})$
end if
for $t = 2, \ldots, T$ **do**
    $\Delta_{tk} := |(\hat{\lambda}(t, k, m) - \hat{\lambda}(t-1, k, m))/\hat{\lambda}(t-1, k, m)|$
    if $\Delta_{tk} \geq \delta_k$ **and** $\hat{\lambda}(t, k, m) > \tau_{tk}$ **then**
        $\hat{Y}_{tk} = 1$
        if $\hat{Y}_{tk} \neq Y_{tk}$ **then**
            $\tau_{tk} := \max(\Pi(\tau_{t-1,k} + \eta_k \hat{Y}_{tk}), \hat{\lambda}(t-1, k, m)/a_{1k})$
        end if
    end if
    if $\hat{\lambda}(t, k, m) \leq \hat{\lambda}(t-1, k, m)/a_{2k}$ **then**
        $\tau_{tk} := \hat{\lambda}(t, k, m)$.
    end if
end for

Note that the effect of marks de-couples from the rest. Thus, when using `ImplicitSTHawkes` based on (5), we first fit a feature extractor on the past mark data and then use maximum likelihood estimation to estimate the rest parameters. On the other hand, for `ExplicitSTHawkes`, we directly substitute (4) into (6) and solve the following maximization problem, which also includes l_1 regularizations on parameters θ and d for sparsity and feature selection. Note $\widetilde{\beta} := \{\delta, \theta, A, \beta, d\}$ contains all model parameters.

$$\min_{\widetilde{\beta}} \quad -l(\widetilde{\beta}) + \|\theta\|_1 + \|d\|_1 \tag{7}$$

$$\text{s.t.} \quad \alpha_{i,j} = 0 \text{ if } |i - j| \geq \tau \tag{8}$$

$$\|\delta\|_2 \leq 1, \|\theta\|_2 \leq 1, \|A\|_2 \leq 1, \|d\|_2 \leq 1 \tag{9}$$

$$\beta \geq 0, \delta \geq 0 \tag{10}$$

We briefly explain the intention of including constraints (8)—(10). Constraint (8) introduces sparsity in the interaction matrix, reduce the total number of parameters in the model for computationally efficiency. Constraint (9) ensures the objective (7) is bounded and is reasonable since the rate $\lambda(t, k, m)$ is typically very small. Constraint (10) is introduced since baseline rates (i.e. δ) and interaction propagation over time (i.e. β) are non-negative.

In addition, note that the log-likelihood $l(\widetilde{\beta})$ is concave in all parameters except the scalar β. Thus, we can solve the optimization problem in (7) iteratively and efficiently using convex optimization solvers [Diamond and Boyd, 2016] to high numerical accuracy. The alternating approach empirically terminates in a very small number of iterations (e.g., three) and each iteration only takes a few seconds to minutes, depending on problem size. Hence, it is computationally friendly.

Lastly, Algorithm 1 constructs dynamic thresholds based on the value of rate estimates $\hat{\lambda}(t, k, m)$ to predict the binary outcomes—whether or not there is a fire incident on day t with mark m at

location k. This is motivated by the hedging algorithm [Raginsky et al., 2012, Algorithm 4]. In particular, we observe that rate estimates $\hat{\lambda}(t, k, m)$ have clear seasonality (e.g., sharp drop from summer to fall and sharp rise from spring to summer), while fire incidents often occur when rate estimates suddenly increase on certain days. Figure 2a illustrates performance of our model based on the observations above.

2.3 Confident fire size classification

Besides predicting when and where fire occurs, fire size prediction is also desirable and valuable —knowing the possible fire sizes can better inform people of potential losses brought by such disasters and plan accordingly. More formally, this is a classification problem where given features $X_t \in \mathbb{R}^p$, one builds a mutli-class classifier that outputs $\hat{Y}_t \in \{1, \ldots, C\}$ as the point prediction out of C classes. However, point predictions are often insufficient in such settings—there are inherent uncertainties in these predictions due to randomness in data or the training of classifiers. Incorrect point predictions in such settings are often consequential—for example, if most people who might be affected by fire hazards believe in the point prediction of a small fire incident and thus ignore proper precaution, the actual occurrence of a large fire incident can be disastrous.

Therefore, a confident fire size prediction that quantifies point predictions' uncertainties is essential. One way for uncertainty quantification in classification is the construction of *prediction sets* around $\hat{Y}_t$ that contain actual observations Y_t with high probability before its realization. Formally, given a significance level α, we construct a *prediction set* $\hat{C}_t(X_t|\alpha) \subset \{1, \ldots, C\}$ such that

$$\mathbb{P}(Y_t \in \hat{C}_t(X_t|\alpha)) \geq 1 - \alpha. \tag{11}$$

A set satisfying (11) thus *confidently* predicts the actual fire size Y_t with high probability. Note that a trivial construction that always satisfies (11) is $\hat{C}_t(X_t|\alpha) = \{1, \ldots, C\}$, so we also want the prediction set to be as small as possible. Constructing such sets is challenging because fire incidents are highly correlated and non-stationary, and classifiers can be very complex (e.g., neural network classifiers).

To build prediction sets that satisfy (11) in practice, we produce uncertainty sets using recent advances in conformal prediction. In particular, we adopt the `ERAPS` algorithm in [Xu and Xie, 2022], as it can build prediction sets for time-series data and works for arbitrarily complex classifiers. On a high level, `ERAPS` assigns *non-conformity scores* to each possible fire size and includes in the prediction set fire sizes whose non-conformity scores are small compared to past ones.

3 Real-data experiments

We apply methods in the previous section to analyze 2014-2019 fire data in California, United States. The experiment is organized as follows. Section 3.1 starts with a small region to compare `ExplicitSTHawkes` with competing baselines. Section 3.2 presents results on a large region to demonstrate the scalability of our Hawkes process model. Section 3.3 builds prediction sets for fire size classification using `ERAPS`.

We briefly describe the dataset and evaluation metrics. The dataset contains 2014–2019 fire incident data collected by the California Public Utilities Commission.[2] A total of 3191 incidents are recorded. Each incident is provided with additional information, which we call *marks* from now on. Marks can be classified as discrete/continuous and dynamic/static. All discrete marks are one-hot encoded to be utilized in the model. Static marks do not change at a given location (e.g., existing vegetation types). Dynamic marks include meteorological season, fire threat zone, fire potential index, large fire probability, etc. Out of all dynamic continuous marks, we select

[2] California Public Utilities Commission (CPUC), available at https://www.cpuc.ca.gov/wildfires/

the temperature, relative humidity, fire probability index (FPI), large fire probability (LFP), and pressure for prediction. We interpolate missing entries of each feature using the spline function with degree 5. Each feature is then standardized to have unit variance and zero mean and further scaled to lie within the interval $[0, 1]$ so that estimated parameters for different marks are on the same scale. The unit for time t is in days, allowing fractional values as the exact hour and minutes are recorded when fires occurred. Regarding evaluation metrics, we use the F_1 score for assessment, which is a standard metric for classification when data are imbalanced—note that the number of no occurrence of fire incidents (denoted as 0) significantly outweighs the other (denoted as 1).

3.1 A small-scale example

We first focus on a small region because the distribution of fire incidents within the region and the performance of our model can be visualized clearly. The model is trained with incidents between 2014 and 2017 and examined on data in 2018. There were 238 fire occurrences in 2014-2017 and 70 in 2018. Upon consulting domain experts, we set the sides of discretized cells to be 0.24-degrees in both longitude and latitude directions so that a total of 36 non-overlapping cells cover the region. Figure 1 visualizes both the training and validation data, from which it is clear that the validation data have a much less number of actual fires; only a few grids have wildfires that occurred near them.

Estimated parameters. We interpret the estimated parameters of `ExplicitSTHawkes`. Figure 1 examines the location-location interaction parameters α_{ij}, which is zero if centroids of two cells exceeds 4×0.24 degrees. Values of α_{ij} greater than (resp. less than) 0 indicate inducive (resp. inhibitory) effect from nearby and past events. The distribution of interaction effects closely aligns with the actual data in Figure 1. Table 1 shows the estimated parameters for marks, whose magnitude indicates feature importance. Recall θ (resp. d) measures the effect of static (resp. dynamic) marks on fire risks—higher magnitude of estimates contribute more significantly to the growth of fire risk. Noticeably, the top two features in d (excluding summer) are also factors in defining the *Fire Danger Index*, which is one of the most commonly used indexes for fire hazard monitoring[3]. Therefore, the model estimates are physically meaningful.

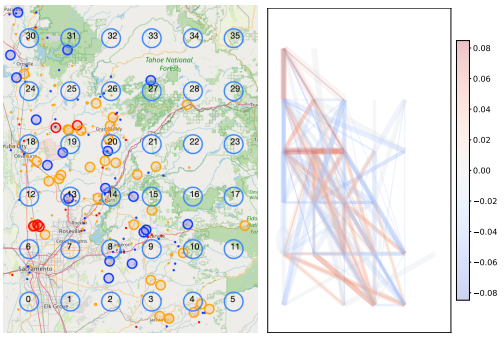


Figure 1: Visualize data and grid (left) and the estimated interaction parameters (right). Interaction parameters well model the distribution of fire incidents.

In addition, we can perform counterfactual analyses using the estimated parameters: suppose a decision-maker wants to know the increase in risk when an external condition changes from A to B (e.g., Fire tier zone shift, changes in vegetation types, etc.). Then, the change in risk at a certain location and time is $\Delta(A, B) := \lambda(t, k, B) - \lambda(t, k, A)$. Similar analyses can be performed for a change in location from k to k_1. Such analyses can help one better study the effect of different factors on fire risks, making risk management more effective.

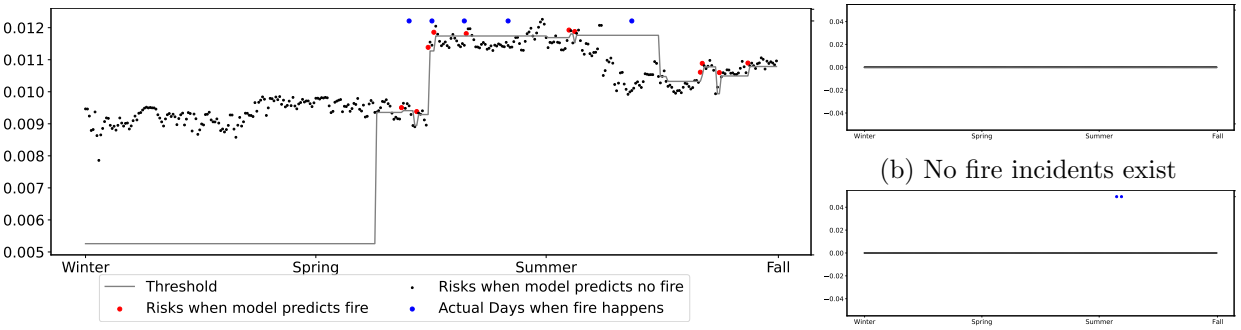
Prediction results. Figure 2 shows the trajectory of rate prediction by `ExplicitSTHawkes` on top of days of actual incidents, where the dynamic thresholds are described in Algorithm 1. We have several takeaways. First, sharp increases in predicted fire risks tend to occur near true fire events. Second, the sparsity of fire occurrence at different locations can significantly impact prediction—patterns of fire occurrences in Figure 2b (resp. Figure 2c) are the easiest (resp. most challenging) for our model to accurately. We suspect this difference occurs likely because of the difficulty in capturing

[3] http://learnline.cdu.edu.au/units/env207/fundamentals/weather.html

	Top 3 Largest Estimates			Bottom 3 Smallest Estimates		
θ estimate	0.301	0.231	0.184	0.046	0.024	0.008
θ feature name	Fire Tier1	Fire Tier2	Fire Tier3	PHYS=Developed-Roads	PHYS=Conifer	PHYS=Developed
d estimate	0.57	0.472	0.46	0.217	0.117	0.02
d feature name	Summer	Temperature	Relative Humidity	LFP	Spring	Winter

Table 1: Estimated static and dynamic mark parameters in `ExplicitSTHawkes`. Larger magnitude indicates more contribution of the feature to fire hazards.

distribution shift in the data, which is reflected in the emergence of fire occurrence at a specific location with no previous fire incidents. In practice, one may better adapt to distribution shifts and handle data non-stationarity by refitting the model parameters.



(a) A few fire incidents exist

(b) No fire incidents exist

(c) Few fire incidents exists

Figure 2: Predicted fire risks and fire incidents on top of actual incidents and dynamic thresholds. The pattern of fire incidents directly relates to the difficulty in prediction: it is the easiest (resp. hardest) when zero (resp. few) fire incidents exist, as shown in Figure 2b (resp. Figure 2c). When a non-trivial number of fire incidents exist as in Figure 2a, the prediction by `ExplicitSTHawkes` can closely match that of the actual data.

We also compare `ExplicitSTHawkes` with several one-class classification baselines. We choose isolation forest [Liu et al., 2008], one-class SVM [Chang and Lin, 2011], local outlier factor [Breunig et al., 2000], and elliptic envelope [Rousseeuw and van Driessen, 1999] due to their popularity and generality. These classifiers use the same data as `ExplicitSTHawkes`, including static and dynamic marks. Figure 3 visualizes the histograms of F_1 scores by each method and the map of precision, recall, and F_1 score by `ExplicitSTHawkes`. The top row shows that `ExplicitSTHawkes` outperforms competing methods by yielding less zero F_1 scores and more one F_1 scores. Note that zero (resp. one) F_1 scores appear at locations that are the easiest (resp. hardest) to predict discussed earlier. In addition, `ExplicitSTHawkes` can yield non-trivial fractional F_1 scores at other locations by capturing a decent number of true positives. Nevertheless, our model also yields many zero F_1 scores because the task is inherently challenging: it makes 365 daily predictions at each of 36 locations, in a total of 13140 predictions, when there are only 70 actual fire occurrences across all 36 locations. The bottom row in Figure 3 plots F_1 score, recall, and precision. Clearly, the pattern of such metrics matches the difficulty of prediction at each grid.

3.2 Large-scale occurrence modeling

We now show that our point-process models are scalable to a large region with much more fire incidents. The setup and hyperparameter choices are the same as Section 3.1. There are a total of 2011 fire occurrences in this region, comprising 63% total wildfire incidents in California from 2014

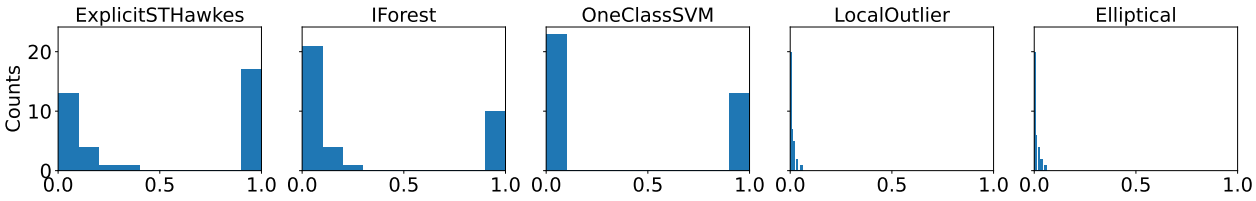


Figure 3: Small-region data, comparison across methods. Histograms of F_1 scores over all locations show that our `ExplicitSTHawkes` outperform other methods by yielding fewer zero F_1 scores, a moderate number of fractional F_1 scores, and more one F_1 scores.

to 2019. Figure 4 visualizes fire incidents within the region on map, which is discretized into 453 grids with side lengths equal to 0.24 degree. We remove regions that lie inside the ocean. Most grids have no fire in the 5-year horizon since fires seem to cluster near the coastal line with large populations.

Figure 5 compares `ImplicitSTHawkes` (under one-class SVM), `ExplicitSTHawkes`, IForest, and OneClassSVM. `ImplicitSTHawkes` performs better than both the `ExplicitSTHawkes` and the isolation forest by yielding more non-zero F_1 scores and a large number of F_1 scores being one. Due to its flexible feature extractor, the `ImplicitSTHawkes` is also competitive against the one-class SVM; importantly, it yields more F_1 scores between zero and one, making it more informative than the one-class SVM on certain grids with a few fire incidents. Hence, `ImplicitSTHawkes` maintains improved performance than other models even if the number of grids significantly increases. Meanwhile, the bottom row visualizes grid-wise performance by `ImplicitSTHawkes`, in which most grids with zero or near zero F_1 scores likely experience a distribution shift during prediction in 2019.

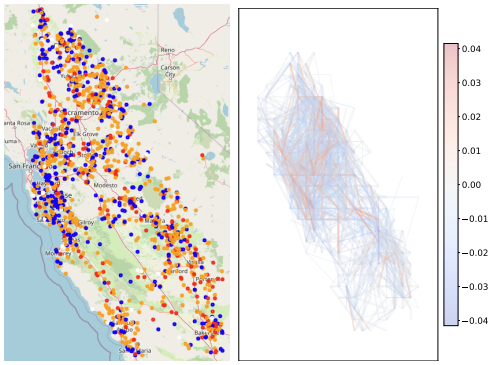


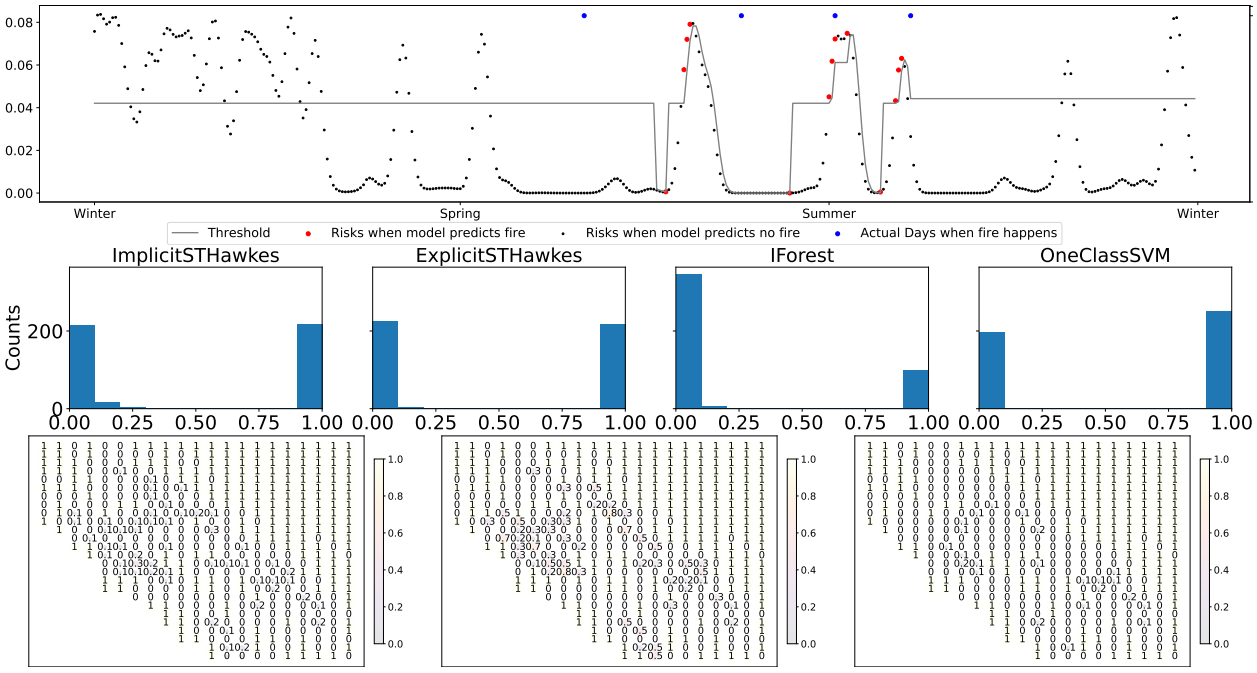
Figure 4: Data visualization. Left figure shows fire events colored by season, and right figure visualizes the interaction matrix parameters α_{ij}.

3.3 Fire size prediction sets

We show that prediction sets by `ERAPS` maintain desired coverage defined in (11) Data in 2014-2018 are training data, and data in 2019 are test data, where there are a total of five possible sizes. Besides the dynamic marks, each feature X_{tk} includes the location information (i.e., longitude and latitude), fire threat zone, and the existing vegetation type. Both the random forest classifier (RF) and the neural network classifier (NN) are used as prediction algorithms; their setup is the same as those in [Xu and Xie, 2022]. Figure 6 shows marginal coverage under both classifiers, where we also compare `ERAPS` against a competing method `SRAPS` [Angelopoulos et al., 2020]. The details of `SRAPS` are described in [Xu and Xie, 2022, Algorithm 1]. We have two findings. First, `ERAPS` performs very similarly under both classifiers and always maintains $1-\alpha$ coverage, whereas `SRAPS` tends to lose coverage at different values of α. Thus, `ERAPS` is more robust and consistent in terms of coverage. Second, both methods return prediction sets with almost the same sizes, but `ERAPS` performs better due to its ability to maintain near $1-\alpha$ coverage.

4 Conclusion

In this work, we leverage one-class wildfire occurrence data to understand wildfire activities better. We first focus on wildfire pattern recognition (e.g., feature importance and regional interactions) and occurrence prediction through the proposed marked spatio-temporal point process model. This model is efficient in data storage and parameter estimation, interpretable regarding model formulation, and accurate for predicting occurrences against competing baselines. We then quantify uncertainties in wildfire size prediction using recent time-series conformal predictions methods, which form prediction sets that contain actual response variables with high probability.



(a) F_1 score (b) Recall (c) Precision

Figure 5: Large-region data. The top row visualizes the trajectory of predicted risks by `ImplicitSTHawkes` at one grid. The middle row compares the histograms of F_1 score under various method. The leftmost `ImplicitSTHawkes` has the most number of non-zero F_1 scores with many being 1. The bottom row visualizes F_1 score, recall and precision of `ImplicitSTHawkes` on the terrain map. Overall, `ImplicitSTHawkes` yields the best performance among all models.

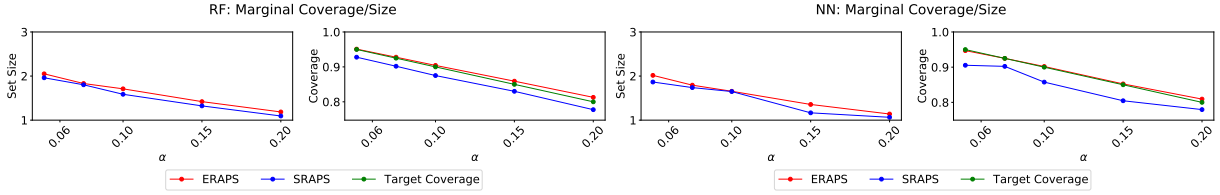


Figure 6: Marginal coverage and width of `ERAPS` and `SRAPS` under the random forest Classifer and the neural network classifier.

References

A. Angelopoulos, Stephen Bates, J. Malik, and Michael I. Jordan. Uncertainty sets for image classifiers using conformal prediction. *ArXiv*, abs/2009.14193, 2020.

Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal prediction beyond exchangeability. 2022.

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *SIGMOD '00*, 2000.

Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2:27:1–27:27, 2011.

Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

Edith Gabriel and Peter John Diggle. Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63, 2009.

Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58:83–90, 1971.

Piyush Jain, Sean C. P. Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D. Flannigan. A review of machine learning applications in wildfire science and management. *ArXiv*, abs/2003.00646, 2020.

Jonathan Koh, Franccois Pimont, Jean luc Dupuy, and Thomas Opitz. Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. 2021.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422, 2008.

Hongyuan Mei and Jason Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *NIPS*, 2017a.

Hongyuan Mei and Jason Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *NIPS*, 2017b.

Maxim Raginsky, Rebecca M Willett, Corinne Horn, Jorge Silva, and Roummel F Marcia. Sequential anomaly detection in the presence of noise and limited feedback. *IEEE Transactions on Information Theory*, 58(8):5544–5562, 2012.

Alex Reinhart. A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, 33(3), Aug 2018. ISSN 0883-4237. doi: 10.1214/17-sts629. URL `http://dx.doi.org/10.1214/17-STS629`.

Yaniv Romano, M. Sesia, and E. Candès. Classification with valid and adaptive coverage. *arXiv: Methodology*, 2020.

Peter J. Rousseeuw and Katrien van Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999.

L. A. Sanabria, X. Qin, J. Li, R. P. Cechet, and Christopher Lucas. Spatial interpolation of mcarthur's forest fire danger index across australia: Observational study. *Environ. Model. Softw.*, 50:37–50, 2013.

Jeffrey D. Scargle. An introduction to the theory of point processes, vol. i: Elementary theory and methods. *Technometrics*, 46:257 – 257, 2004.

Frederic Paik Schoenberg, Chien-Hsun Chang, Jon E. Keeley, Jamie Pompa, J. D. Woods, and Haiyong Xu. A critical assessment of the burning index in los angeles county, california. 2007.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9 (Mar):371–421, 2008.

Ryan J Tibshirani, Rina Foygel Barber, Emmanuel Candes, and Aaditya Ramdas. Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems*, pages 2530–2540, 2019.

Chen Xu and Yao Xie. Conformal anomaly detection on spatio-temporal observations with missing data. 2021.

Chen Xu and Yao Xie. Conformal prediction set for time-series, 2022. URL `https://arxiv.org/abs/2206.07851`.

Haiyong Xu and F. Schoenberg. Point process modeling of wildfire hazard in los angeles county, california. *The Annals of Applied Statistics*, 5:684–704, 2011.

Gianluca Zeni, Matteo Fontana, and Simone Vantini. Conformal prediction: a unified review of theory and new challenges. *ArXiv*, abs/2005.07972, 2020.