Solar radiation ramping events modeling using spatio-temporal point processes

Chen Xu*[1], Minghe Zhang*[1], Andy Sun[2], Feng Qiu[3], and Yao Xie[1]

[1]H. Milton Stewart School of Industrial and Systems Engineering (ISyE), Georgia Institute of Technology
[2]Sloan School of Management and MIT Energy Initiative, Massachusetts Institute of Technology
[3]Argonne National Lab

Abstract

Modeling and predicting solar ramping events, which have complex spatio-temporal dependencies, are critical for improving situational awareness of solar power systems. To tackle these challenges, we adopt a novel spatio-temporal categorical point process model, which intuitively and effectively addresses correlation and interaction among ramping events. We demonstrate the interpretability and predictive power of our model in real-data experiments.

1 Introduction

Modeling ramping events in solar power systems is essential. Ramping events [Florita et al., 2013, Cui et al., 2015, Kamath, 2010] are one-bit information that represents abnormal events in sequential observations. They can be interpreted as abrupt slope increases or decreases in power generations and typically occur under extreme weather (e.g., rainstorm or hurricane) [Rocchetta et al., 2015]. The power system is vulnerable to such events, which forces the affected units to shut down. Therefore, predicting these events is valuable for distribution operators to take necessary precautions and reduce restoration costs.

In general, the ramping event is also challenging to predict because of (i) spatio-temporal correlation, (ii) non-stationarity, and (iii) computational efficiency in online prediction. This paper addresses these challenges through two contributions. First, we present a spatio-temporal categorical point process originally proposed by Juditsky et al. [2020]. The model can flexibly capture the spatio-temporal correlations and interactions among binary or categorical ramping events without assuming time-decaying influence. The model parameters are efficiently estimated using convex optimization with theoretical guarantees. In addition, the model can efficiently make online probabilistic ramping event predictions at any location and time. Second, we propose dynamic decision thresholds to address non-stationary when making online predictions of a future ramping event. These dynamic thresholds show improved performance over static ones.

Literature. There have been many works on solar ramping event modeling. A line of work started by Florita et al. [2013] adopts the Swinging Door Algorithm (SDA) commonly used in data compression, which does not build statistical models. Later, Cui et al. [2015] and Cui et al. [2017] propose an optimized SDA using dynamic programming. However, these works do not have theoretical guarantees and consider no spatio-temporal correlations. Among recent works, Abuella and Chowdhury [2019] uses ensemble-based probabilistic forecasts but needs access to many features besides historical ramping events. In addition, limited data can hamper performance. Meanwhile, Zhu et al. [2019] proposes a forewarning method using a credal network and imprecise Dirichlet model to study power change by meteorological fluctuation. However, the authors only provide probabilistic forecasts without online categorical predictions. As a result, dynamic solar ramping event modeling remains a great challenge, especially under strong spatio-temporal correlation.

We can also view ramping event modeling as a type of anomaly detection. Research on anomaly detection in the time series analysis area has also been active for decades, which could be traced back to Grubbs [1969]'s work in the 1960s. There are many conventional works on the topic, such as Aggarwal [2015], Akoglu et al. [2015], Gupta et al. [2013]. Also, many works use modern deep learning frameworks to deal with more complex patterns such as Hawkins et al. [2002] and more. However, none of these works have a

*Contributed equally.

flexible anomaly detection model under strong spatio-temporal correlation for non-stationary data. Lastly, Xu and Xie [2021] uses novel conformal prediction techniques for anomaly detection in spatio-temporal data, but it is yet unclear how the method can be applied to categorical responses.

2 Model and Prediction

Notation. Observed data are denoted as $\boldsymbol{\omega}^T = \{\omega_{tk}, 1 \le t < T, 1 \le k \le K\}$ for T time steps at K locations, where $\omega_{tk} \in \{0, 1, \ldots, M\}$ denotes the state of the ramping event out of $M+1$ possible states (0 means it is a non-ramping event). Meanwhile, given a memory window d, define $\boldsymbol{\omega}_{t-d}^t = \{\omega_{sk}, t-d \le s < t, 1 \le k \le K\}$ as all ramping data in the past d time units from all location.

Model. We now define the spatio-temporal probabilistic model for ramping events. For each location k, we associate an array of *birthrate* parameters $\bar{\beta}_k = \{\beta_k(p), 0 \le p \le M\}$, and for every pair of locations k, l and every $s \in \{1, \ldots, d\}$, an array of *interaction* parameters $\bar{\beta}_{k\ell}^s = \{\beta_{k\ell}^s(p, q), 0 \le p \le M, 0 \le q \le M\}$. Thus, for each ramping state $p \in \{1, \ldots, M\}$, the conditional probability of ω_{tk} being in state $p = 0, \ldots, M$ on $\boldsymbol{\omega}_{t-d}^t$ is written as

$$\mathbb{P}[\omega_{tk} = p | \boldsymbol{\omega}_{t-d}^t] = \beta_k(p) + \sum_{s=1}^{d} \sum_{\ell=1}^{K} \beta_{k\ell}^s \left(p, \omega_{(t-s)\ell}\right), \tag{1}$$

This conditional probability model explicitly captures the dependence of event at time t and location k on events from all locations in the past d days. The set of model parameters are $\boldsymbol{\beta} = \left\{\bar{\beta}_k, \bar{\beta}_{k\ell}^s : 1 \le s \le d, 1 \le k, \ell \le K\right\}$. Define the number of parameters $\kappa := K(M+1) + dK^2(M+1)^2$, then $\boldsymbol{\beta} \in \mathbb{R}^\kappa$, subject to (1) lies within $[0, 1]$.

Estimation. We introduce both the least-square (LS) method and the maximum likelihood (ML) estimation technique for model estimation.

We obtain the LS estimate $\hat{\boldsymbol{\beta}}^{\mathrm{LS}}(\boldsymbol{\omega}^N)$ by minimizing the following optimization problem objective:

$$\frac{1}{2N} \sum_{t=1}^{N} \sum_{k=1}^{K} \left\| \left(\bar{\beta}_k + \sum_{s=1}^{d} \sum_{\ell=1}^{K} \bar{\beta}_{k\ell}^s \left(\omega_{(t-s)\ell}\right) \right) - \bar{\omega}_{tk} \right\|_2^2 \tag{2}$$

In (2), $\bar{\omega}_{tk} := e_{\omega_{tk}} \in \mathbb{R}^{M+1}$ is the vector of ramping states, whose i^{th} entry is 1 if ω_{tk} is in state $p = 0, \ldots, M$. This vector encodes the true state of the data point at time t and location k. Also, $\bar{\beta}_k = \{\beta_k(p) \in \mathbb{R}, p \in \{0, ..., M\}\} \in \mathbb{R}^{M+1}$ and $\bar{\beta}_{k\ell}^s\left(\omega_{(t-s)\ell}\right) = \{\beta_{k\ell}^s\left(p, \omega_{(t-s)\ell}\right) \in \mathbb{R}, p \in \{0, ..., M\}\} \in \mathbb{R}^{M+1}$. Moreover, note this optimization problem has a strongly convex objective function (e.g. the least-square cost function), so it can be solved efficiently in polynomial time.

Following [Juditsky et al., 2020], we can obtain the ML estimate $\hat{\boldsymbol{\beta}}^{\mathrm{ML}}(\boldsymbol{\omega}^N)$ by minimizing the following objective:

$$-\frac{1}{N} \sum_{t=1}^{N} \sum_{k=1}^{K} \ln \left(\beta_k(\omega_{tk}) + \sum_{s=1}^{d} \sum_{l=1}^{K} \beta_{kl}^s(\omega_{tk}, \omega_{(t-s)l}) \right), \tag{3}$$

The objective function is convex since it resembles the likelihood function for a generalized linear model (GLM) with Bernoulli link functions. Thus it can be solved efficiently by convex optimization algorithms as we did for LS estimates.

Remark 1 (Computational Complexity). The total number of parameters κ depend on K (i.e., number of location) and M (i.e., number of states) quadratically. Therefore, even for convex solvers with polynomial complexity, the complexity of solving the problem can be high when these numbers are large. In such a case, sparsity can be imposed in this model to speed up computation. For example, we may assume that a pair of locations do not influence each other if the distance between them exceeds a threshold without losing model convexity.

Prediction. We first make sequential prediction of future conditional probabilities using estimated parameters according to equation (1). Then, probabilities are converted to states using abnormal thresholds $\{\tau_{tk}(p) : 1 \le p \le M\}$. These thresholds can be defined sequentially and dynamically or be fixed in advance.

We propose the following dynamic $\tau_{tk}(p)$ for sequential prediction. Denote w as the window of past w time units of observations. Then, the dynamic thresholds $\tau_{tk}(p), p = 1, \ldots, M$ have the form:

$$\begin{aligned} \tau_{tk}(p) =& \alpha_p \frac{\sum_{i=1}^{w} \hat{p}_{(t-i)k}(p) \omega_{(t-i)k}(p)}{\sum_{i=1}^{w} \omega_{(t-i)k}(p)} + \\ & (1-\alpha_p) \frac{\sum_{i=1}^{w} \hat{p}_{(t-i)k}(p)(1 - \omega_{(t-i)k}(p))}{w - \sum_{i=1}^{w} \omega_{(t-i)k}(p)}. \end{aligned} \tag{4}$$

Thresholds (4) uses the history of the ramping events. If α_p is close to 0, the threshold is likely small since the latter summand in (4) tends to have a larger denominator than its numerator. If there is no ramping event with the given state in the past w days, we let the dynamic threshold be the static threshold (e.g.

$\tau_{tk}(p) = \tau_k(p)$ for all t and $\tau_k(p)$ is found via cross-validation) to avoid 0/0.

3 Theory

The performance guarantee for $\hat{\boldsymbol{\beta}}^{\mathrm{LS}} = \hat{\boldsymbol{\beta}}^{\mathrm{LS}}(\boldsymbol{\omega}^N)$ is proven by Juditsky et al. [2020]. We state our contribution for the ML estimates $\hat{\boldsymbol{\beta}}^{\mathrm{ML}} = \hat{\boldsymbol{\beta}}^{\mathrm{ML}}(\boldsymbol{\omega}^N)$.

First, the gradient $F_{\omega^N}(x)$ of the ML objective (3) is

$$F_{\omega^N}(\boldsymbol{\beta}) = \frac{1}{N}\sum_{t=1}^{N} \eta\left(\omega_{t-d}^{t-1}\right)\theta\left(\eta^T\left(\omega_{t-d}^{t-1}\right)x, \omega_t\right),$$

with

$$\eta^\top(\boldsymbol{\omega}_{t-d}^{t}) = [I_K, I_K \otimes \mathrm{vec}(\boldsymbol{\omega}_{t-d}^{t-1})^\top] \in \mathbb{R}^{K\times\kappa}. \quad (5)$$

$$\begin{aligned}&\theta(z,\omega)\\ &= -\sum_{k=1}^{K}\left[\sum_{p=0}^{M}\frac{[w]_{kp}}{[z]_{kp}}e^{kp} - \frac{1-\sum_{p=0}^{M}[w]_{kp}}{1-\sum_{p=0}^{M}[z]_{kp}}\sum_{p=0}^{M}e^{kp}.\right]\end{aligned}$$

For $\xi_t := \eta\left(\omega_{t-d}^{t-1}\right)\theta\left(\eta^T\left(\omega_{t-d}^{t-1}\right)\beta, \omega_t\right)$, $\mathbb{E}[\xi_t] = 0$. Therefore, $\{\xi_t\}$ forms a martingale-difference sequence and $F_{\omega^N}(x) = \frac{1}{N}\sum_{t=1}^{N}\xi_t$. $F_{\omega^N}(x)$ will be abbreviated as $F(x)$ from now on. To make sure $F(\boldsymbol{\beta})$ is continuous, we also require that $\boldsymbol{\beta}$ leads (1) to be within $[\rho, 1-\rho]$ for a pre-specified $\rho \in (0,1)$

Lemma 3.1. *Assume $||\xi_t||_\infty \le \Theta$ for all t. For all $\epsilon \in (0,1)$ vector $F(\boldsymbol{\beta})$ satisfies*

$$\mathrm{Prob}_{\omega^N}\left\{\|F(\boldsymbol{\beta})\|_\infty \ge \Theta\sqrt{\frac{2\ln(2\kappa/\epsilon)}{N}}\right\} \le \epsilon.\rho \in (0,1)$$

Proof. We first provide a simple bound on $||\xi_t||_\infty$ and then link Θ to $||\xi_t||_\infty$ under ML with identity-link.

First,

$$\begin{aligned}&||\xi_t||_\infty\\ =&||\eta\left(\omega_{t-d}^{t-1}\right)\theta\left(\eta^T\left(\omega_{t-d}^{t-1}\right)\beta, \omega_t\right)||_\infty\\ \overset{(i)}{\le}&||\theta\left(\eta^T\left(\omega_{t-d}^{t-1}\right)\beta, \omega_t\right)||_\infty\\ \overset{(ii)}{=}&||\sum_{k=1}^{K}\left[\sum_{p=0}^{M}\frac{[w]_{kp}}{[z]_{kp}}e^{kp} - \frac{1-\sum_{p=0}^{M}[w]_{kp}}{1-\sum_{p=0}^{M}[z]_{kp}}\sum_{p=0}^{M}e^{kp}\right]||_\infty\\ \overset{(iii)}{\le}&||\sum_{k=1}^{K}\sum_{p=0}^{M}\frac{[w]_{kp}}{[z]_{kp}}e^{kp}||_\infty\\ \overset{(iv)}{\le}&||\sum_{k=1}^{K}\sum_{p=0}^{M}\frac{1}{\rho}e^{kp}||_\infty \overset{(v)}{=} 1/\rho,\end{aligned}$$

where (i) holds by the definition of η, (ii) holds under the simpler notation $z := \eta^T\beta, \omega := \omega_t$, (iii) holds as $1 \ge \sum_{p=0}^{M}[w]_{kp}$ and $\sum_{p=0}^{M}[z]_{kp}$, (iv) holds as $0 \le [\omega]_{kp} \le 1$ and $[z]_{kp} \ge \rho$ by definition, and (v) holds as $\sum_{k=1}^{K}\sum_{p=0}^{M}e^{kp}$ is the vector of all 1's in $\mathbf{R}^{K(M+1)}$.

Also, denoting by $\mathbf{E}_{|\omega^t}$ the conditional expectation given ω^t and $i = 1,\ldots,\kappa$, we have

$$\begin{aligned}&\mathbf{E}_{\omega^{t+1}}\left\{\exp\left\{\sum_{s=1}^{t+1}\gamma\left[\xi_s\right]_i\right\}\right\}\\ \le&\mathbf{E}_{\omega^t}\left\{\exp\left\{\gamma\sum_{s=1}^{t}\left[\xi_s\right]_i\right\}\exp\left\{\gamma^2/2\rho^2\right\}\right\},\end{aligned}$$

which holds due to the Hoeffding's inequality and, from the above bound that the conditional, ω^t given, distribution of $[\xi_{t+1}]_i$ is zero mean and supported on $[-1/\rho, 1/\rho]$. By induction, we have

$$\mathbf{E}_{\omega^N}\left\{\exp\left\{\gamma\left[\sum_{t=1}^{N}\xi_t\right]_i\right\}\right\} \le \exp\left\{N\gamma^2/2\rho^2\right\}.$$

Now using Chernoff bound, we have

$$\begin{aligned}&\mathrm{Prob}_{\omega^N}\left\{[F(\boldsymbol{\beta})]_i > \theta\right\}\\ =&\mathrm{Prob}_{\omega^N}\left\{\frac{1}{N}\left[\sum_{t=1}^{N}\xi_t\right]_i > \theta\right\}\\ \le&\exp\{-\mu\theta\}\mathbf{E}_{\omega^N}\left\{\exp\left\{\mu\frac{1}{N}\left[\sum_{t=1}^{N}\xi_t\right]_i\right\}\right\}\\ \le&\exp\left\{-\mu\theta + \frac{\mu^2}{2\tilde{N}}\right\},\end{aligned}$$

where $\tilde{N} = N\rho^2$. The union bound and further simplifications yields the desired result □

For our main theorem below, define $A[\boldsymbol{\omega}^n] := \frac{1}{N}\sum_{t=1}^{N}\eta(\boldsymbol{\omega}_{t-d}^{t})\eta^\top(\boldsymbol{\omega}_{t-d}^{t})$ for η in (5) and the condition number given $A \in \mathbb{R}^{\kappa\times\kappa}, A \succ 0$:

$$\theta_p[A] := \max\{\theta \ge 0 : g^\top A g \ge \theta||g||_p^2, \forall g \in \mathbb{R}^\kappa, p \in [1,\infty]\}. \quad (6)$$

Theorem 3.2 (Bounding ℓ_p error of ML estimate)**.** *For every $\epsilon \in (0,1)$, every $\boldsymbol{\omega}^N$, and any $p \in [1,\infty]$, we have with probability at least $1-\epsilon$ that*

$$||\hat{\boldsymbol{\beta}}^{\mathrm{ML}} - \boldsymbol{\beta}||_p \le \frac{(1-\rho)^2}{\rho}\sqrt{\frac{2\ln(2\kappa/\epsilon)}{N}}/\sqrt{\theta_p[A[\boldsymbol{\omega}^N]]\theta_1[A[\boldsymbol{\omega}^N]]}.$$

Proof. For simplicity, let $\hat{\boldsymbol{\beta}} := \hat{\boldsymbol{\beta}}^{\mathrm{ML}}$. Since $F(x)$ is the gradient of the convex and continuously differentiable log-likelihood $L_{\omega^N}(x)$, $F(x)$ is continuous and

monotone. Thus, standard results in monotone variational inequality yields $\left\langle F(\boldsymbol{\beta}) - F(\hat{\boldsymbol{\beta}}), \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \right\rangle \geq 0$ and $F(\hat{\boldsymbol{\beta}}) = 0$. In fact, we can prove a stronger inequality: define $\tilde{F}(\eta^T x)$ where $\eta\tilde{F}(\eta^T x) = F(x)$ with $\eta\eta^T = \frac{1}{N}\sum_{t=1}^{N} \eta(\omega_{t-d}^{t-1})\eta^T(\omega_{t-d}^{t-1})$. Then

$$\left\langle \tilde{F}(\eta^T\boldsymbol{\beta}) - \tilde{F}(\eta^T\hat{\boldsymbol{\beta}}), \eta^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\rangle \geq (1-\rho)^{-2}||\eta^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})||_2^2.$$

Now, by the earlier definition of $\theta_p(A) = \theta_p(\eta\eta^T)$, we can see that $||\eta^T(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})||_2^2 = (\hat{\boldsymbol{\beta}}^T - \boldsymbol{\beta}^T)A(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}) \geq \sqrt{\theta_1(A)\theta_p(A)}||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_1||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_p$.

As a result,

$$\begin{aligned}
&||F(\boldsymbol{\beta})||_\infty||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_1 \\
\geq &\left\langle F(\boldsymbol{\beta}) - F(\hat{\boldsymbol{\beta}}), \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \right\rangle \\
= &\left\langle \tilde{F}(\eta^T\boldsymbol{\beta}) - \tilde{F}(\eta^T\hat{\boldsymbol{\beta}}), \eta^T(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\rangle \\
\geq &(1-\rho)^{-2}||\eta^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})||_2^2 \\
\geq &\sqrt{\theta_1(A)\theta_p(A)}||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_1||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_p
\end{aligned}$$

We finish the proof by cancelling $||\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}||_1$ from the the final inequality and using Lemma 3.1 to bound $||F(\boldsymbol{\beta})||_\infty$. □

Remark 2 (Numerical Computation). The quantity $\theta_p[A[\boldsymbol{\omega}^N]]$ in (6) is readily computable when $p = 2$ or ∞: when $p = 2$, it is the minimum eigenvalue and when $p = \infty$, it is $\min_{1\leq i\leq\kappa}\{x^T Ax : \|x\|_\infty \leq 1, x_i = 1\}$, which is the minimum of κ efficiently computable quantities. Solving for $p = 1$ is hard in general, but the discussion in [Juditsky et al., 2020] obtains a bound tight within the factor $\pi/2$ [Nesterov, 1998].

Furthermore, we can expect that the minimum eigenvalue of $A[\boldsymbol{\omega}^n]$ will be of order 1 with high probability, so that the bound using Theorem 3.2 goes to 0 as $N \to \infty$ at the rate $O(1/\sqrt{N})$. As a result, $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. Lastly, the bound on the right-hand side in Theorem 3.2 is fully data-driven and computable given historical ramping events.

4 Real-data study

We focus on modeling single-state ramping events (e.g. $\omega_{tk} \in \{0, 1\}$) for visualization simplicity. Two-state ramping events occur when the current radiation is too high or too low compared to the past.

The dataset comes from the NSRDB[1]. We collect a non-uniform set of 10 downtown data across different cities[2] in California. Data in 2017 are used for parameter estimation and data in 2018 are for sequential prediction.

We adopt standard performance metrics for classification, including precision, recall, and F1 score. Specifically, we consider a predicted state (i.e., thresholding the probabilistic prediction) to be a true positive if it aligns with the actual state. Because ramping events are extreme cases in data (e.g., positive cases in data are rare), we do not use the ROC curve (true positive rate versus false-positive rate) in our setting.

4.1 Parameter estimation results

We first visualize the estimated birthrate and interaction parameters in Figure 1, which provide interpretable insights into correlation among locations. We then overlay these parameters on terrain maps in Figure 2 to illustrate certain spatio-temporary patterns uncovered by our point-process model.

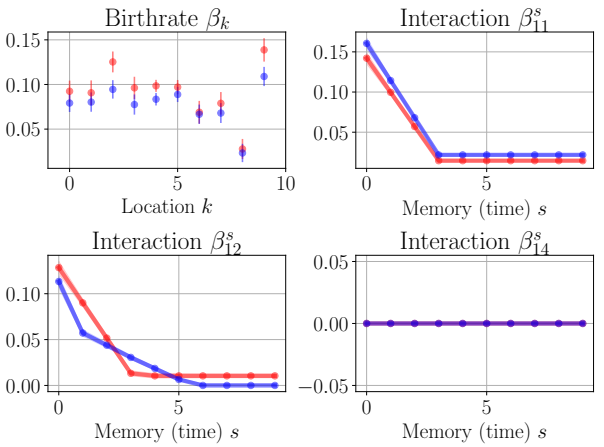

Figure 1: Recovered birthrate parameters at different locations and interaction parameters over time. The blue curve shows ML recoveries, and the red curve shows LS recoveries. The shaded areas are 95% confidence intervals under bootstrap. Birthrates are similar in magnitude across different locations, and interactions all decay very fast over time.

(a) Recovered spatio-temporal influence: Figure 1 shows the recovered single-state birthrate parameters β_k over different locations by ML and LS, and shows a selected number of interaction parameters β^s_{kl} over time. Bootstrap confidence intervals are plotted around the estimates. The results on birthrates

[1]Dataset available at https://nsrdb.nrel.gov/

[2]Complete list of cities: Fremont, Milpitas, Mountain View, North San Jose, Palo Alto, Redwood City, San Mateo, Santa Clara, South San Jose, Sunnyvale

show the existence of heterogeneity among locations. We also find that estimates of the interaction parameters between different sensors decay fast, indicating that location-to-location influences do not persist over time. In practice, this indicates that ramping events that happened several times ago at any location cannot noticeably influence the current ramping probabilities. Lastly, estimates recovered by ML and LS generally have very similar magnitude.

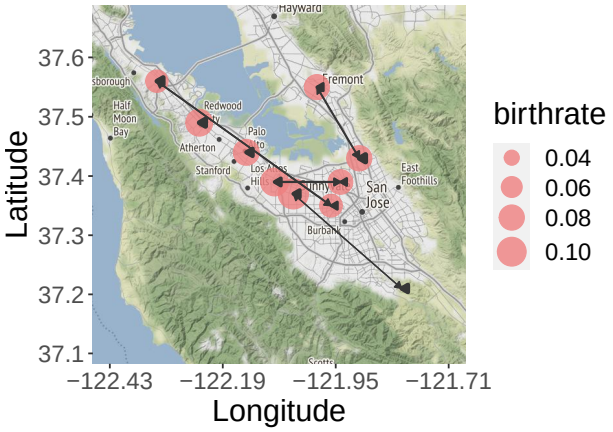

Figure 2: Visualization of ML birthrate (red circles) and interactions (black lines) recovery on terrain map. We do not show parameter recovery for $s > 1$ because the magnitude decays very fast with the same orientations.

(b) Visualize interaction parameters on terrain map: We plot these parameters as graphs overlaid on a terrain map to visualize the influences. We only show the ML estimates because LS estimates are similar. The vertices' size and edges' width are proportional to the magnitude of corresponding recovered parameters. Results in Figure 2 are interpretable and unveil previously unknown connections among the locations. In particular, locations close to each other have similar birthrates, and there are strong influences from a city onto itself. In addition, influences mainly flow towards the southeast, with large magnitudes even if cities are far away (e.g., San Mateo to Santa Clara and Sunnyvale to South San Jose).

4.2 Prediction performance:

We use aforementioned techniques to make sequential prediction in Palo Alto (one representative example in North California) and evaluate the metrics. Table 1 compares the performance of our model on Atlanta data with two additional baselines: the logistic regression and the linear regression. Our method yields significantly higher F_1 scores.

Table 1: Baseline comparison for single-state model: we compare the F_1 scores of our method (LS or ML) with two baselines methods, where ours yield significantly higher scores. Static τ means that the decision thresholds are held constant at a fixed location during prediction.

	LS	ML	Linear Regression	Logistic Regression
Static τ	0.96	**0.97**	0.64	0.67
Dynamic τ	0.88	**0.96**	0.65	0.67

To better visualize the estimated probabilities' trajectories, we plot the probability estimates, dynamic thresholds, and the prediction intervals in Figure 3. We also use the bootstrap confidence interval for $\boldsymbol{\beta}$, which was shown in Figure 1, to compute the confidence interval for p_{tk}. The prediction intervals at 95% confidence level concentrate closely around the estimates, even if the Bonferroni correction was used. Based on the figure, dynamic thresholds and probability estimates (red/blue dots) have similar rise-and-fall patterns. At the same time, the trajectory of probabilities also highly correlates with the actual ramping events (black dots). Such high correlation enables accurate prediction. It is clear that dynamic thresholds yield decision boundaries that help distinguish the ramping events.

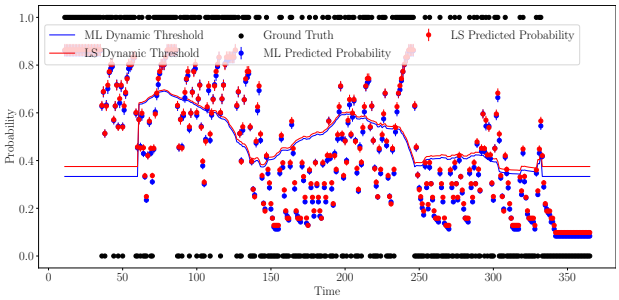

Figure 3: Prediction intervals for online point prediction of probabilities for single-state ramping events using LS (red dots) and ML (blue dots), compared with true ramping events (black dots). Dynamic thresholds well separate the set of predicted ramping vs. normal probabilities.

Remark 3 (Beyond single state). We remark that the above model has been applied to ramping events with more than two states (e.g., high abnormal state, normal state, and low abnormal state). The results are not shown due to space limitation. Nevertheless, the predicted probabilities well follow the dynamics in the ramping events as we observed in Figure 3. In addition, we examined the robustness of our model on data in different seasons, where the overall predictive

performance remains satisfactory.

5 Conclusion and future works

In this paper, we introduce a new framework for modeling spatio-temporal abnormal events in solar radiation, where model parameters can be efficiently estimated with strong performance guarantee. We apply the method to various cities in the US to show that the model is flexible and robust, yielding physically meaningful results for interpretation. We believe the proposed model is a general framework for other spatio-temporal data modeling tasks.

There are a few possible extensions. In terms of data, we can consider other definitions of ramping events and more categories per event, which may include the severity of each ramping event. We can also include link functions and other regularization techniques for improved performances during modeling.

References

Mohamed Abuella and Badrul Chowdhury. Forecasting of solar power ramp events: A post-processing approach. *Renewable Energy*, 133: 1380 – 1392, 2019. ISSN 0960-1481. doi: https://doi.org/10.1016/j.renene.2018.09.005. URL http://www.sciencedirect.com/science/article/pii/S0960148118310681.

Charu C Aggarwal. Outlier analysis. In *Data mining*, pages 237–263. Springer, 2015.

Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery*, 29 (3):626–688, 2015.

M. Cui, J. Zhang, A. R. Florita, B. Hodge, D. Ke, and Y. Sun. An optimized swinging door algorithm for wind power ramp event detection. In *2015 IEEE Power Energy Society General Meeting*, pages 1–5, 2015.

Mingjian Cui, Jie Zhang, Anthony R Florita, Bri-Mathias Hodge, Deping Ke, and Yuanzhang Sun. An optimized swinging door algorithm for identifying wind ramping events. *IEEE Transactions on Sustainable Energy*, 7(1):150–162, 2015.

Mingjian Cui, Jie Zhang, Cong Feng, Anthony R. Florita, Yuanzhang Sun, and Bri Mathias Hodge. Characterizing and analyzing ramping events in wind power, solar power, load, and netload. *Renewable Energy*, 111, 4 2017. ISSN 0960-1481. doi: 10.1016/j.renene.2017.04.005.

A. Florita, B. Hodge, and K. Orwig. Identifying wind and solar ramping events. In *2013 IEEE Green Technologies Conference (GreenTech)*, pages 147–152, 2013.

Anthony Florita, Bri-Mathias Hodge, and Kirsten Orwig. Identifying wind and solar ramping events. In *2013 IEEE Green Technologies Conference (GreenTech)*, pages 147–152. IEEE, 2013.

Frank E Grubbs. Procedures for detecting outlying observations in samples. *Technometrics*, 11(1):1–21, 1969.

Manish Gupta, Jing Gao, Charu C Aggarwal, and Jiawei Han. Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and data Engineering*, 26(9):2250–2267, 2013.

Simon Hawkins, Hongxing He, Graham Williams, and Rohan Baxter. Outlier detection using replicator neural networks. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 170–180. Springer, 2002.

Anatoli Juditsky, Arkadi Nemirovski, Liyan Xie, and Yao Xie. Convex recovery of marked spatio-temporal point processes. *arXiv preprint arXiv:2003.12935*, 2020.

Chandrika Kamath. Understanding wind ramp events through analysis of historical data. In *IEEE PES T&D 2010*, pages 1–6. IEEE, 2010.

Yu Nesterov. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization Methods and Software*, 9(1-3):141–160, 1998. doi: 10.1080/10556789808805690. URL https://doi.org/10.1080/10556789808805690.

Roberto Rocchetta, Yanfu Li, and Enrico Zio. Risk assessment and risk-cost optimization of distributed power generation systems considering extreme weather conditions. *Reliability Engineering & System Safety*, 136:47–61, 2015.

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

W. Zhu, L. Zhang, M. Yang, and B. Wang. Solar power ramp event forewarning with limited historical observations. In *2019 IEEE/IAS 55th Industrial and Commercial Power Systems Technical Conference (I CPS)*, pages 1–8, 2019.