Predicting Uncertainty in Renewable Energy Growth: A Spatio-Temporal Conformal Framework

Wenbin Zhou[1], Shixiang Zhu[1] and Feng Qiu[2]

[1]*Carnegie Mellon University,* [2]*Argonne National Laboratory*

TL;DR: In this paper, we introduce a general algorithmic framework named *Spatio-Temporal Conformal Prediction (STCP)*, designed to generate prediction intervals for the future number of distributed energy resources (DERs) installations within each spatial subregion. These intervals account for uncertainties arising from neighboring subregions, providing robust references for utility decision-makers to undertake strategic grid planning at the spatio-temporal level. Theoretically, we demonstrate that the algorithm is both efficient (informative) and valid (reliable), ensuring that the prediction intervals are not only sharp but also asymptotically cover the true count with a specified coverage probability. Lastly, in collaboration with AES Indiana, we apply our algorithm to forecast the growth of photovoltaic installations in the Indianapolis metropolitan area through 2050, offering insights into the future energy landscape and uncovering potential growth patterns.

Introduction

Distributed energy resources (DERs), such as solar panels, are renewable energy resources associated with high unpredictability and high stakes.

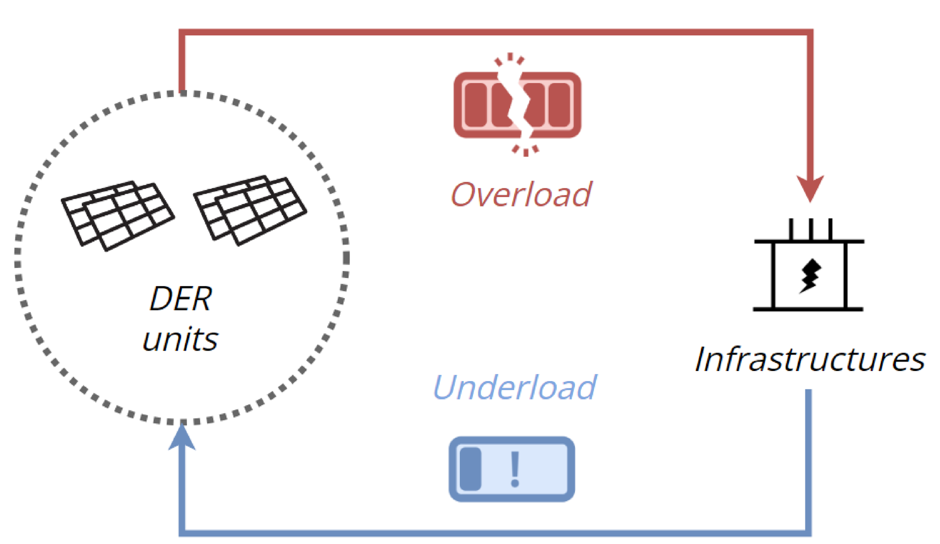

Figure: Insufficiently reinforced infrastructures with DER units may cause overload backfeed and underload normal feed to customers.

Spatio-temporal uncertainty quantification is a critical step in the decision-making pipeline of strategic grid planning & operations.

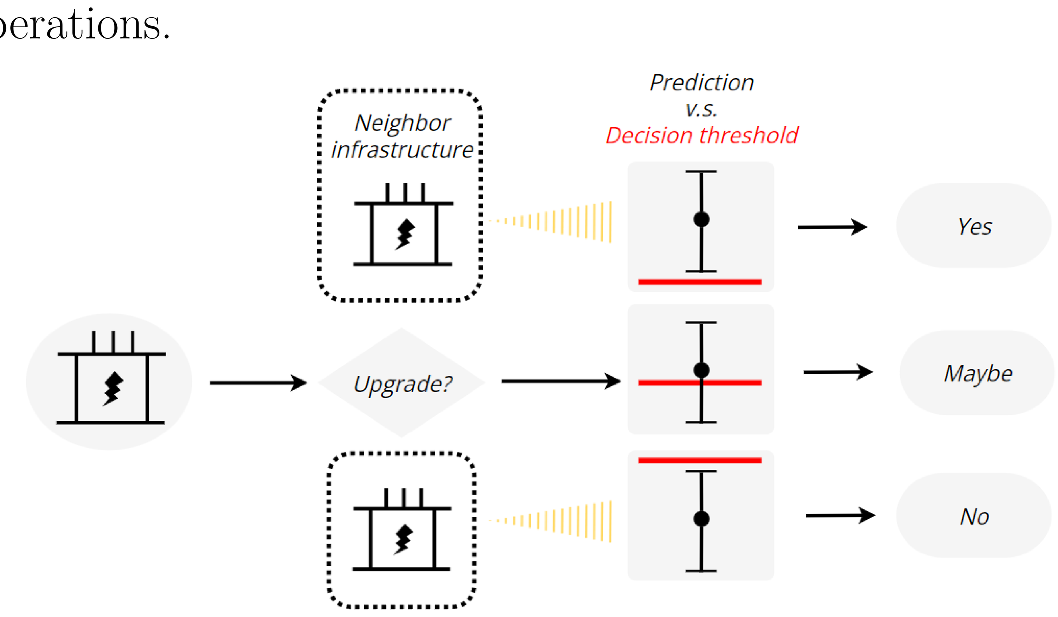

Figure: Example of a decision-making flow chart for substation upgrade.

Problem Setup

The goal is to predict and quantify the uncertainty of the growth of DERs across M subregions (*e.g.* substations). We want:

- **Reliable:** A total of M prediction intervals with probabilistic coverage guarantees,
- **Informative:** Appropriately account for underestimating & overestimating uncertainty across multiple subregions.

Mathematical formulation: Given some significance level α, we aim to obtain M prediction intervals $\{\widehat{C}_m(\alpha)\}_{m=1,\dots,M}$ such that for all $m = \{1,\dots,M\}$,

$$\mathbb{P}\left(\bigcap_{m'\in\mathcal{N}(m)} \{Y_{T,m'} \in \widehat{C}_{m'}(\alpha)\}\right) \geq 1-\alpha, \tag{1}$$

where $Y_{T,m'}$ is the prediction target (the future installation counts at m' of DER units), and $\mathcal{N}(m)$ denotes the neighborhood region of subregion m.

Highlights:

1. Tailors to practical spatio-temporal decision-making, which account for "neighborhoods" rather than individual or all infrastructures.
2. User can *flexibly* control for how "radical" or "conservative" the uncertainty quantification result is.
3. Avoids *curse of dimensionality* by exploiting low-rank (if available) neighborhood structure.

Theory Sketch

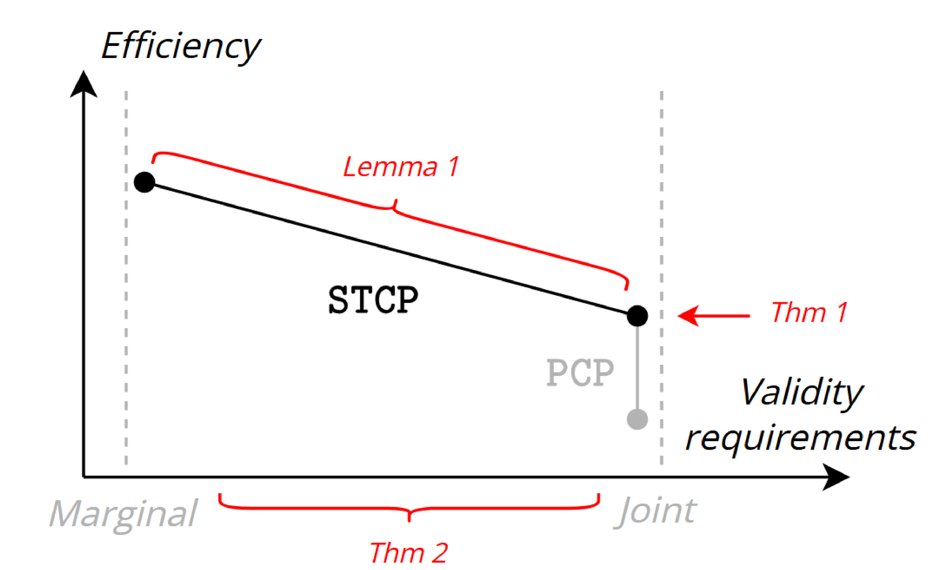

Figure: Diagrammatic representation of theoretical insights

Efficiency measures how informative the prediction interval is, which is characterized by its size.

Lemma 1: Smaller neighborhood specification leads to higher efficiency of `STCP`.

Theorem 1 (Relative efficiency): The worst-case efficiency of `STCP` is no worse than the best-case efficiency of a trivial modification of `PCP` under our setting.

- `PCP` is a baseline who cannot directly achieve (1).

Theorem 2 (Asymptotic validity): Objective (1) can be asymptotically approximated with large samples.

- Asymptotic is the best that we can achieve without placing assumptions on the temporal dependency of data.

Algorithm: Spatio-Temporal Conformal Prediction (STCP)

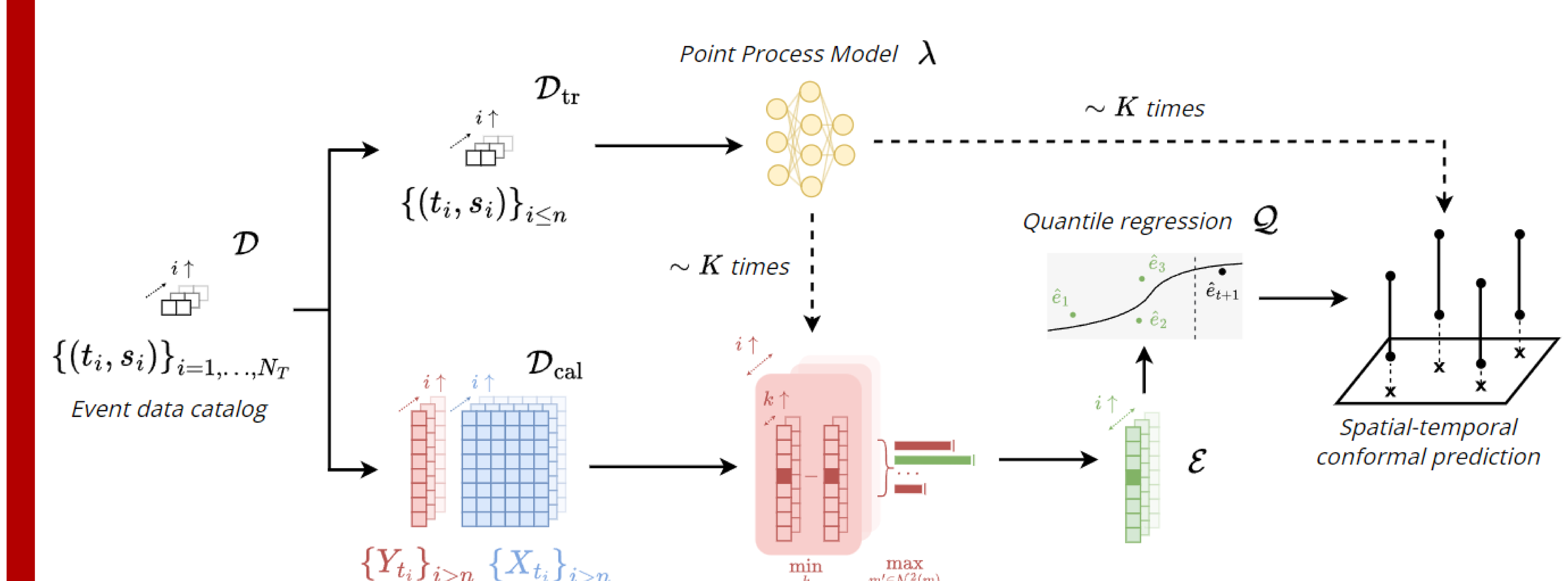

Figure: Illustration of our proposed algorithm. The training phase splits the dataset ($\mathcal{D}$), where the training data ($\mathcal{D}_{\text{tr}}$) is used to fit the model (λ), and then used with the calibration data ($\mathcal{D}_{\text{cal}}$) to calculate the set of neighbor-aware nonconformity scores ($\mathcal{E}$). Finally, Quantile regression ($\mathcal{Q}$) is fitted over $\mathcal{E}$ to predict the magnitude of the uncertainty and augment the base predictions of λ to produce the prediction intervals.

Case Study

Data description: In collaboration with AES Indiana, we collected 1,742 customer-level rooftop solar panel installation records, including all installations from 2010 to mid-June 2024 within the AES service territory.

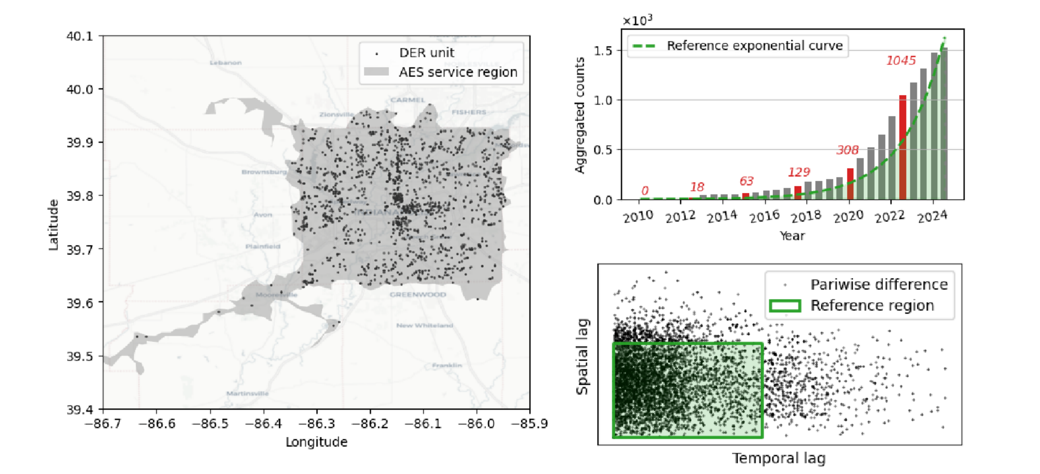

Figure: Exploratory data analysis and visualizations of the data

Analysis: We apply our `STCP` algorithm in a rolling manner to provide spatio-temporal forecasts of the total number of solar panel installations and their uncertainties through 2050.

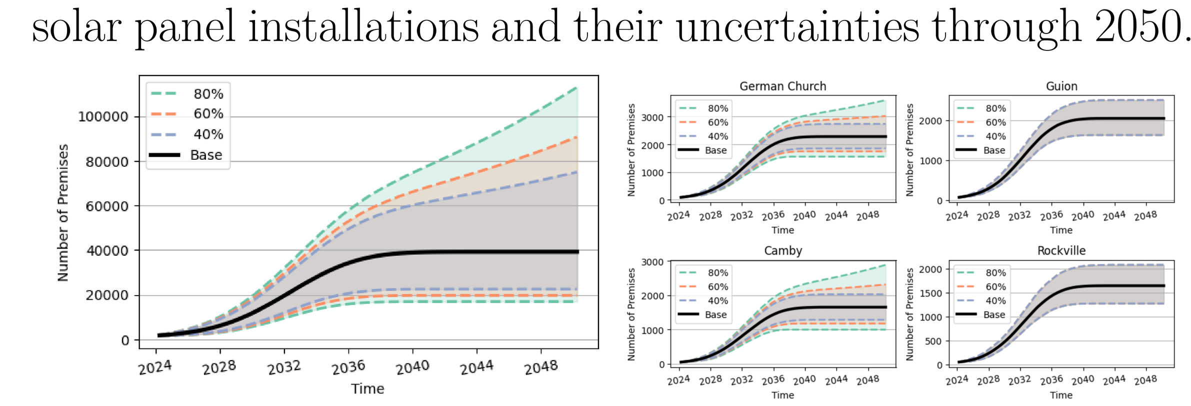

Figure: Temporal visualization. *Left*: aggregated. *Right*: top four substations with highest base prediction.

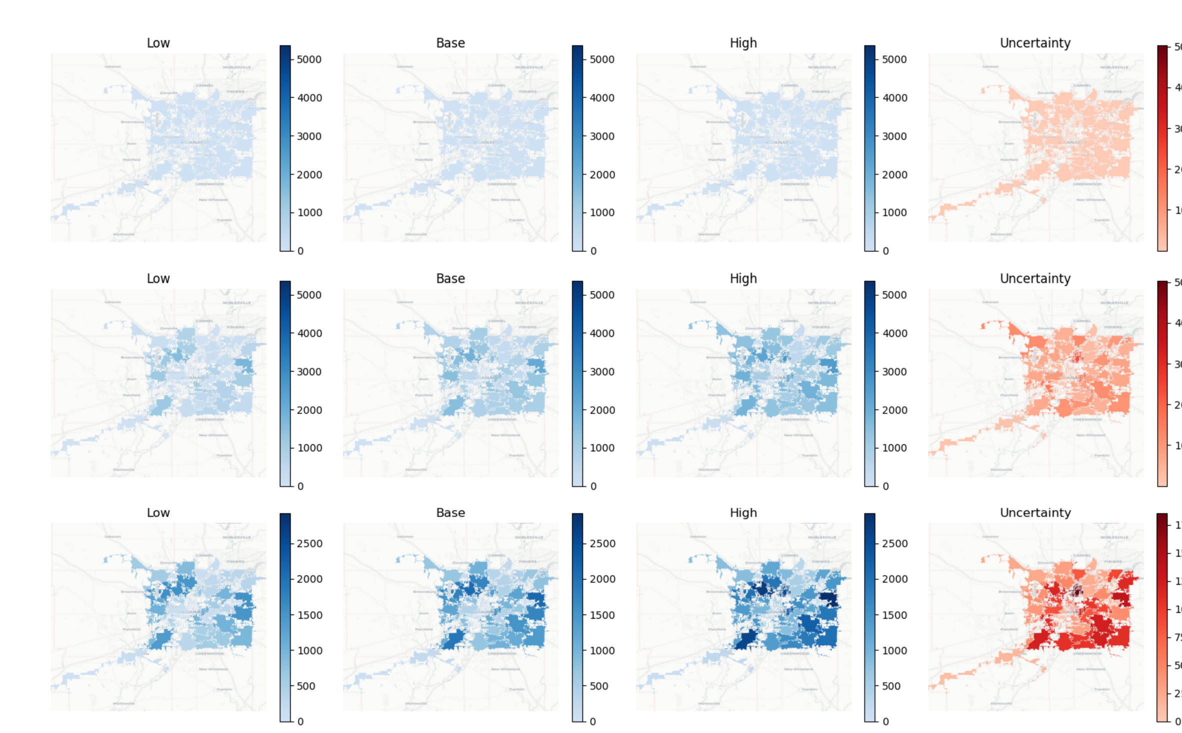

Figure: Spatial snapshots of the evolution of installation. *From left to right*: low, base, high, uncertainty. *From top to bottom*: Jan of 2024, 2037, 2050.

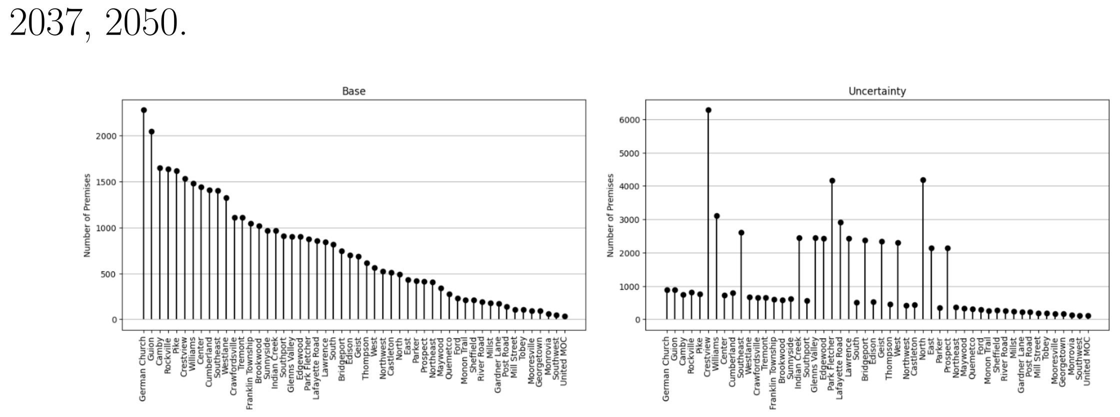

Figure: Base and uncertainty comparison across 52 substations.