Two-Stage Hurdle Models for Outage Forecasting

Shourya Bose

Tomas Kaljevic

Yu Zhang

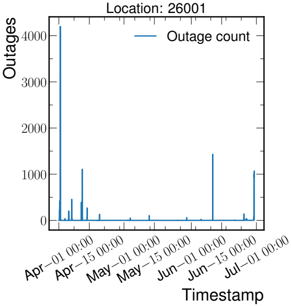
Team #41 TomaKalj
Department of Electrical and Computer Engineering
University of California, Santa Cruz

2025 INFORMS Data Mining Society Data Challenge

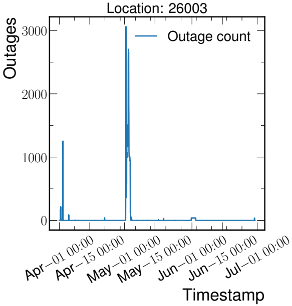
Outline

Competition Dataset: Outages (1/2)

Initial observation: Outage counts are **zero-inflated** and **bursty**.



Alcona County

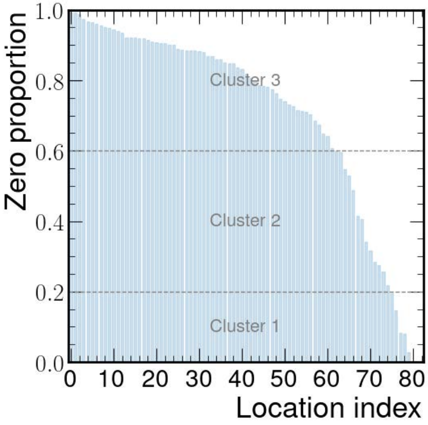


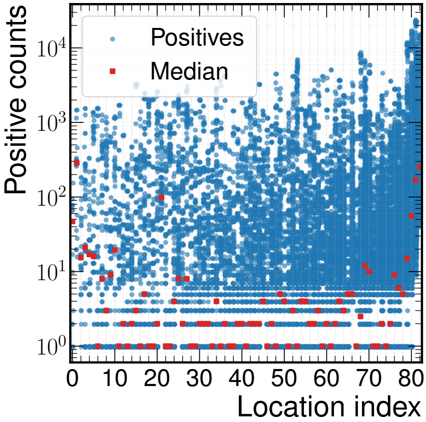
Alger County

Mean	Median	Standard Deviation	Min	Max
45.2	0	452.3	0	23346

Competition Dataset: Outages (2/2)

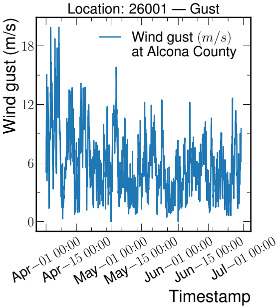
Initial observation: Outage counts are **zero-inflated** and **bursty**.



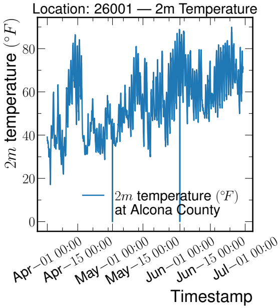


Competition Dataset: Weather

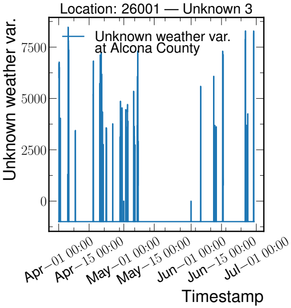
There are 109 weather features.



Variable `gust`



Variable `t2m`

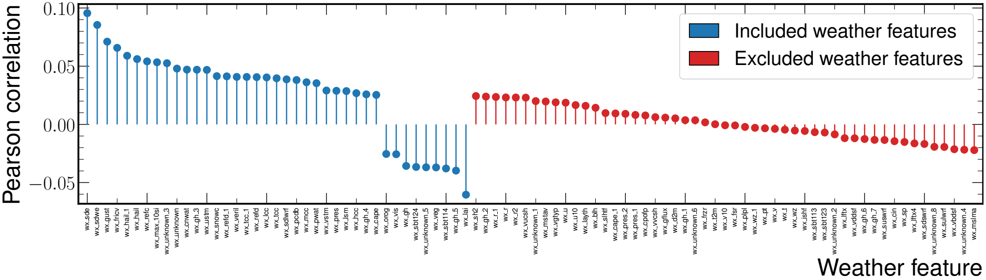


Variable `unknown_3`

Not all weather variables are identified (e.g. unknown_3).

Informative Weather Feature Selection

From 109 weather features, we select 39 with the highest Pearson correlation to outage counts across locations; these (in blue) are termed **informative weather features**.



Pearson correlation coefficient of two time series s_1, s_2:

$$r_{s_1 s_2} = \frac{\sum_{t=1}^{T}(s_{1,t} - \bar{s}_1)(s_{2,t} - \bar{s}_2)}{\sqrt{\sum_{t=1}^{T}(s_{1,t} - \bar{s}_1)^2}\sqrt{\sum_{t=1}^{T}(s_{2,t} - \bar{s}_2)^2}}$$

Motivation of the Two-Stage Model

Popular models face the challenges:

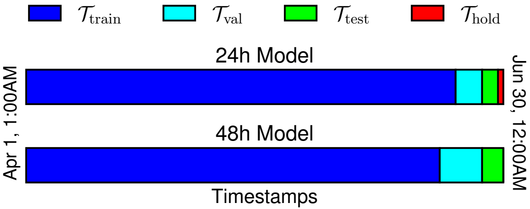
- Zero-inflated nature precludes *LSTM*.
- Small dataset size precludes *transformers*.

Key ideas of our two-stage model:

- Predict *probability* and *count* of outages separately.
- Use **gradient boosted decision trees (GBDT)** for data efficiency.
- Train separate *autoregressive models* for 24h and 48h ahead forecasting.

Dataset Splits

We split the dataset into **train**, **validation**, **test**, and **holdout** sets across all counties.



Let L be the lookback (context) and T the lookahead (rollout) length,

$$|\mathcal{T}_{\mathrm{val}}| = L + mT \ \ (m = 5), \quad |\mathcal{T}_{\mathrm{test}}| = L + T, \quad |\mathcal{T}_{\mathrm{hold}}| = \max\{48 - T, 0\}.$$

Holdout set: Ensures that the test sets' beginnings are aligned for $24h$ and $48h$ models.
Competition task: Forecast outages over a 24-hour and 48-hour interval after Jun 30, 12:00AM; denoted by $\mathcal{T}^{(24)}_{\mathrm{forecast}}$ and $\mathcal{T}^{(48)}_{\mathrm{forecast}}$.

Feature Engineering

For county i and timestamp t, the target is the outage count $o_{i,t}$, while the input vector $\mathbf{x}_{i,t}$ concatenates the following features:

- i: county index
- $o_{i,t-1:t-L}$: outages over L past timestamps
- κ_i: county i's average outage level, i.e. $\frac{1}{|\mathcal{T}_{\text{train}}|} \sum_{t \in \mathcal{T}_{\text{train}}} o_{i,t}$
- $\mathbf{h}_{i,t-1}$: informative weather feature on the previous timestamp
- $\Delta \mathbf{h}_{i,t-1} = \mathbf{h}_{i,t-1} - \mathbf{h}_{i,t-2}$: first-order difference of the informative weather feature
- $\boldsymbol{\tau}_t$: $\left(\sin\left(\frac{\texttt{HoD}(t)}{24}\right), \cos\left(\frac{\texttt{HoD}(t)}{24}\right), \sin\left(\frac{\texttt{DoW}(t)}{7}\right), \cos\left(\frac{\texttt{DoW}(t)}{7}\right)\right)$
- $o_{i,t-1} \times \mathbf{h}_{i,t-1}^{(\text{top})}$: interaction term with top-5 variance weather features.

First Stage: Probability Prediction

- The first stage model is a classifier that outputs raw probabilities:

$$p_{\mathsf{cls}}^{\mathsf{raw}} := \Pr\left[o_{i,t} > 0 \mid \mathbf{x}_{i,t}\right] = M_{\mathsf{cls}}^{\mathsf{raw}}(\mathbf{x}_{i,t}),$$

which is implemented as a **gradient-boosted decision tree** (GBDT).

- Once training is done, we perform *isotonic calibration* on the validation set:

$$g^* = \operatorname*{arg\,min}_{g \text{ non-decreasing}} \sum_{i=1}^{83} \sum_{t \in \mathcal{T}_{\mathsf{val}}} \left(\mathbb{1}_{\{o_{i,t}>0\}} - g(p_{\mathsf{cls}}^{\mathsf{raw}})\right)^2$$

- It recalibrates the model towards the *empirical frequency* of the validation dataset, which **prevents overfitting** on the training data.

Solution of Isotonic Calibration

- The solution g^* is found via the pool-adjacent-violators algorithm (PAVA). The final calibrated first-stage model is given as:

$$M_{\text{cls}}(\mathbf{x}_{i,t}) = g^*\left(p_{\text{cls}}^{\text{raw}}\right) = g^*\left(M_{\text{cls}}^{\text{raw}}(\mathbf{x}_{i,t})\right) \in [0,1]$$

- The calibration maintains the invariant condition:

$$M_{\text{cls}}^{\text{raw}}(\mathbf{x}_{i,t}) \geq M_{\text{cls}}^{\text{raw}}(\mathbf{x}_{i',t'}) \implies M_{\text{cls}}(\mathbf{x}_{i,t}) \geq M_{\text{cls}}(\mathbf{x}_{i',t'}),\ \forall (i,t),(i',t')$$

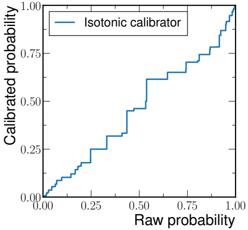
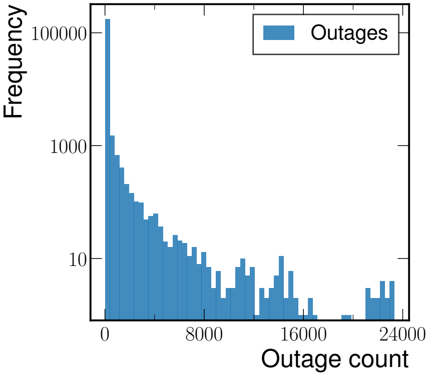


Figure: The calibration curve obtained from our model trained on 24h rollouts.

Second Stage: Count Prediction (1/2)

Observation: A spike at zero then exponentially distributed outage counts.



Tweedie distribution models semi-continuous data with a point mass at zero and a continuous positive tail.

Second Stage: Count Prediction (2/2)

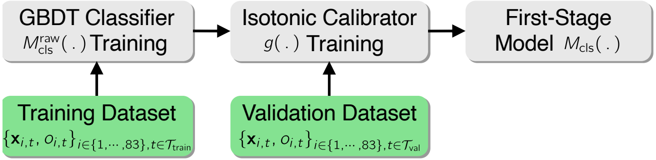
The **Tweedie distribution** belongs to the exponential dispersion family and can represent several well-known distributions depending on its *power parameter* p:

Power p	**Distribution Type**	**Support of Random Variable**
0	Normal	$\mathbb{R}$
1	Poisson	$\{0, 1, 2, \dots\}$
$1 < p < 2$	Compound Poisson–Gamma	$\{0\} \cup (0, \infty)$
2	Gamma	$(0, \infty)$
3	Inverse Gaussian	$(0, \infty)$

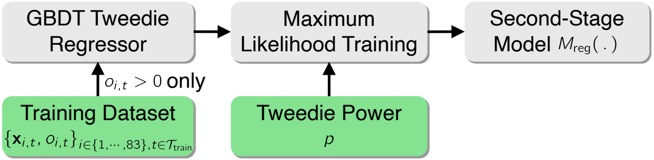
Implementation: We use a GBDT with a Tweedie head which learns the mean μ by maximizing log-likelihood. Thus, the second-stage model yields $\mu_{i,t} = M_{\text{reg}}(\mathbf{x}_{i,t})$.

Training and Inference Pipeline

Training for the first-stage model:



Training for the second-stage model:

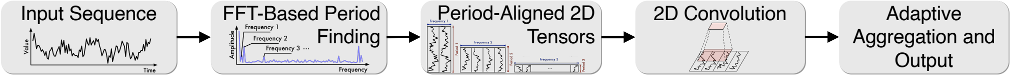


For a new point $\mathbf{x}_{i,t}$ inference entails $\boxed{\hat{o}_{i,t} = M_{\text{cls}}(\mathbf{x}_{i,t}) \times M_{\text{reg}}(\mathbf{x}_{i,t})}$.

Weather Forecasting

We leverage **TimesNet** model[1] to forecast weather feature vectors $\mathbf{h}_{i,t}$ for $t \in \mathcal{T}_{\text{forecast}}^{(24)}$ and $\mathcal{T}_{\text{forecast}}^{(48)}$.

- **Operating principle:**



- **Training and inference:** A global model trained on *all* weather features autoregressively with a lookback of 120*h* and one-shot rollout of 48*h* in the future.

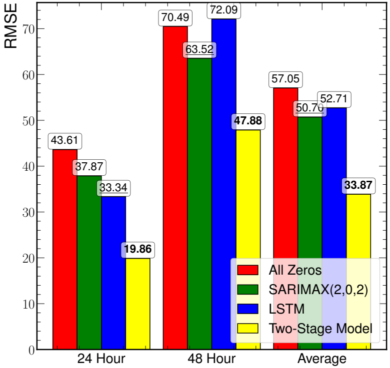
[1] "TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis", Wu *et. al.*, ICLR 2023.

Implementation Details[2]

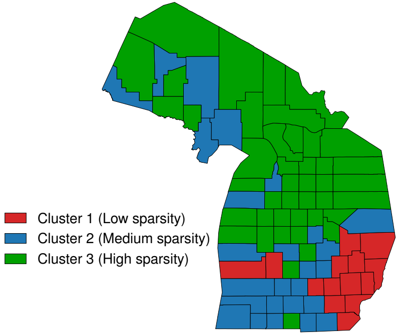
- LightGBM for both stages' GBDT models; Scikit-learn's `IsotonicRegression` for calibration. PyTorch used for training the weather model.
- We use Tweedie power $p = 1.55$ for training M_{reg}.
- Competing methods:
  - **LSTM:** add dense head and sparsity-based clustering; domain knowledge based feature selection; cyclic encoding of HoD and DoW; 48*h* lookback.
  - **SARIMAX:** $(p, d, q) = (2, 0, 2)$.
  - **All-zeros baseline**

[2] Our code: `https://github.com/shourya01/hurdle_model_outage_forecasting`

Test Set Results



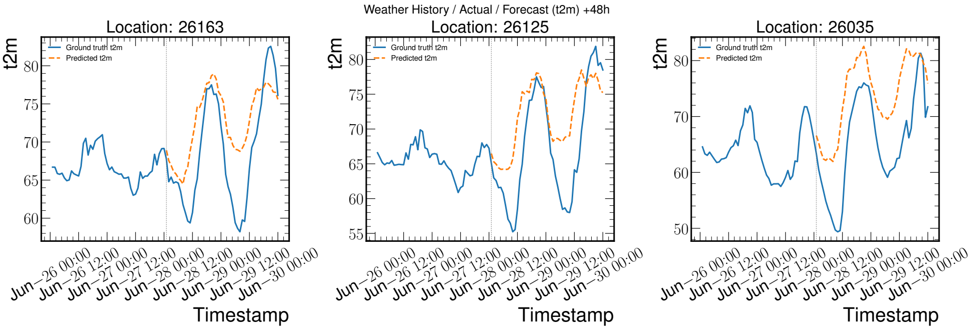
RMSEs on our test set.



Sparsity-based clusters of Michigan counties for LSTM.

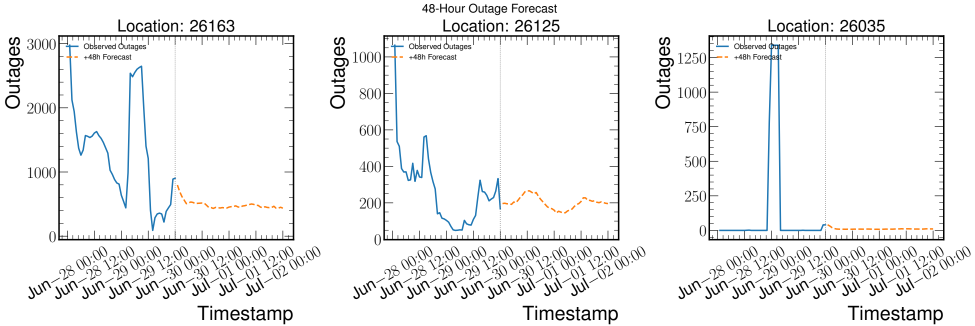
Plots (1/2)

Weather forecasting using TimesNet:



Plots (2/2)

48h forecasted outages (final submission): leaderboard RMSE of 189



Future Directions

- Explore **time series foundation models** (e.g. TimesFM) to explore weather information more efficiently.
- Apply deep learning architectures for training with **data augmentation**.

Thank You!