INFORMS 2026 Data Mining Society Data Challenge

Forecasting Infrastructure Resilience Under Extreme Weather

Competition Co-Chairs

Imtiaz Ahmed (imtiaz.ahmed@mail.wvu.edu), West Virginia University
Zimo Wang (zimo.wang@binghamton.edu), Binghamton University
Woody Zhu (shixianz@andrew.cmu.edu), Carnegie Mellon University
Jason Shafer (jay@poweroutage.us), PowerOutage.com

Problem Statement

Widespread wind events cause the majority of major power outages in the United States. When multiple large storm systems affect the same region in rapid succession, the impacts can compound: utilities may still be restoring service from the first event when the next arrives, crews and materials can become constrained, mutual-assistance resources may already be deployed, and damaged infrastructure may be more vulnerable to additional failures. These closely spaced events therefore represent a significant industry challenge for preparedness, operational coordination, and restoration response.

From March 13–19, 2026, two successive wind-driven storm systems swept across Indiana (IN), Ohio (OH), Pennsylvania (PA), and West Virginia (WV), producing widespread power outages. Together, these storms represented the largest weekly storm-related outage cycle in the United States in 2026 as of the publication of this challenge, with over 2 million customers losing power at some point during the storm cycle. Hourly county-level outage observations were collected from **PowerOutage.com**, a leading real-time outage monitoring platform that tracks customer-level outage counts at sub-county resolution across the United States, for the full event window. A two-day pre-event period covering March 11–12 was also included to support the construction of lagged and antecedent-condition features. Weather data are drawn from NOAA URMA and ERA5 reanalysis.

Storm-related power outages impose significant operational, economic, and societal burdens on utilities, emergency managers, communities, and affected residents. Predicting how outage severity may evolve over the coming hours and days can help utilities pre-position repair crews, request mutual assistance, allocate equipment and materials, and prioritise emergency response before conditions deteriorate.

This challenge asks participants to build models that make these forward-looking predictions using real-world storm and outage data, framed as a genuine spatiotemporal forecasting problem. Competitors will learn cross-county outage dynamics from 239 fully observed training counties and apply their models to 63 held-out test counties using only the first 72 hours of observed outage data combined with full weather information for the entire event window. The objective is not only to estimate where outages may become most severe, but also to better understand how impacts develop, persist, and compound across an interconnected regional event.

Dataset

The dataset was assembled from **PowerOutage.com** in combination with two meteorological sources and covers March 11–19, 2026 across 302 counties. The two-storm structure produces a two-wave outage signature that distinguishes naive persistence approaches from models that genuinely learn storm dynamics.

Outage data are provided at hourly, county-level resolution and include raw customer counts, derived state and flow variables, and the composite outage severity index (OSI). Weather data combine high-resolution URMA fields (2.5 km, observation-corrected, polygon-mean aggregated per county) for near-surface wind

and temperature, with ERA5 reanalysis fields (∼31 km) for atmospheric context including precipitation, cloud cover, boundary layer height, soil moisture, and radiation. The variable set includes:

- **Outage variables:** hourly customer outage counts, outage fraction, OSI state/flow components (P_t, N_t, D_t, R_t), composite OSI, and lag features at 1h, 3h, 6h, 24h, and 48h horizons.
- **URMA (2.5 km):** wind gust, sustained wind speed, wind direction, 2m temperature, and 2m dew-point.
- **ERA5 (∼31 km):** surface and sea-level pressure, boundary layer height, total and liquid precipitation, snowfall, snow depth, cloud cover (total/low/mid/high), shortwave radiation, relative humidity, soil moisture, and reference evapotranspiration.

The training set comprises 239 counties (51,624 rows) with complete observations for all 216 hours and all variables. The test set comprises 63 held-out counties (13,608 rows) with the following structure: weather variables are provided for all 216 hours; outage variables are provided only for March 11–13 (the first 72 hours) and are withheld for March 14–19. OSI forecast targets are withheld entirely. This structure mirrors real operational forecasting: a utility operator has access to weather forecasts for the full storm window but can only observe outages up to the present hour. The train/test split is county-level (not time-based), stratified by state and severity level. Complete variable documentation, the OSI formula, and data quality notes are provided in the data package released to registered participants following execution of the Non-Disclosure Agreement (NDA).

Prediction Task

Hourly OSI Forecasting. Using outage information from the March 11–13 observed window and weather information for the full March 11–19 period, predict the OSI at four future horizons for each hour in the March 14–19 prediction window:

Target	Horizon	Operational significance
`osi_target_t01h`	$t + 1$ hour	Immediate next-hour severity
`osi_target_t06h`	$t + 6$ hours	Crew pre-positioning window
`osi_target_t24h`	$t + 24$ hours	Next-day mutual aid planning
`osi_target_t48h`	$t + 48$ hours	Two-day emergency outlook

Predictions are submitted as one row per test county per hour for the March 14–19 prediction window (9,072 rows total; 63 counties × 144 hours) using the provided template. Rows where the target extends beyond March 19 carry no ground truth and should remain NaN in the submission.

Temporal causality rule. Outage variables are withheld for March 14–19 in the test file. Models must use only the March 11–13 observed outage history as outage-side inputs. Weather features for any timestamp may be used freely since they represent forecast information available to a real operator. **Submissions found to have used future outage values will be automatically disqualified**, regardless of predictive accuracy. Code review is part of the evaluation process.

Evaluation

Submissions will be evaluated on both predictive performance and methodological quality. Numerical performance on the held-out OSI targets will be assessed across all four horizons. The specific evaluation metrics and scoring details will be communicated to registered participants via the Google Drive folder. Finalists will be selected based on a combination of predictive accuracy and the rigour, novelty, and clarity of the modelling approach described in the written report. Detailed submission instructions will also be provided

in the data package upon registration.

Submission Requirements

Each submission must include three components. Detailed submission instructions, including file naming conventions, formatting requirements, and the submission portal, will be provided to registered participants in the Google Drive data package.

1. **Prediction file.** Predicted OSI values submitted using the provided template. Detailed instructions are included in the data package.
2. **Reproducible code.** Self-contained code that reproduces the submitted predictions when run end-to-end on the provided data files. Python or R are preferred. Submissions without reproducible code will not be considered for final rankings.
3. **Written report.** A report of no more than **6 pages** (excluding references) in PDF format. The report should describe the proposed methodology, key modelling decisions, and a critical discussion of results. This report forms a central part of the evaluation.

Registration and Data Access

Registration is open to all currently enrolled undergraduate and graduate students and to postdoctoral researchers. Teams of up to four members are permitted; interdisciplinary teams are encouraged. To register, please complete the registration form below with your name, institutional affiliation, email address, proposed team name, and the names and affiliations of any additional team members:

https://forms.gle/snHGp75SbKj1hfry5

Upon successful registration, participants will receive a Non-Disclosure Agreement (NDA) to sign electronically. Access to the competition data package (hosted on a shared Google Drive folder) is granted only after the signed NDA is returned. The data package includes the training and test data files, submission template, detailed submission instructions, and full documentation (problem description, OSI methodology, weather variable descriptions, and variable reference guide).

Timeline

Milestone	Date
Competition opens; data and templates available	July 2026
Registration deadline	August 31, 2026 (AOE)
Submission deadline (predictions, code, report)	September 25, 2026 (AOE)
Finalist notification	October 5, 2026
Finalist presentations	INFORMS 2026 Annual Meeting, San Francisco, CA

Prizes

The top three finalist teams, selected on the basis of both predictive performance and the quality of their methodology as described in the written report, will be invited to present their work in a special session of the **21st INFORMS Workshop on Data Mining and Decision Analytics (DMDA)** at the INFORMS

2026 Annual Meeting in San Francisco, CA. Presentations will be evaluated by a panel of judges drawn from academia and industry, including a representative from PowerOutage.com, who will determine the final ranking. Cash prizes and certificates of recognition will be awarded based on the judges' final assessment.

To be considered eligible for prizes and presentation, at least one member from each finalist team must be registered for the DMDA workshop.

Award	**Prize**
First Place	$700
Second Place	$400
Third Place	$250

All participating teams will receive a certificate of participation from the INFORMS Data Mining Society.

Contact

For questions regarding the competition, dataset, or submission process, please contact the competition co-chairs: Imtiaz Ahmed (imtiaz.ahmed@mail.wvu.edu), Zimo Wang (zimo.wang@binghamton.edu), or Woody Zhu (shixianz@andrew.cmu.edu).