GMP-AE-4

Climate Informed Benefit-Cost Analysis Refinement for Green Mountain Power

A Methodology Document by Rhizome

Introduction

Green Mountain Power (GMP) engaged Rhizome in developing a methodology for climate-informing and refining the benefit-cost analyses (BCA) for its resiliency investments. In this document, Rhizome has summarized the methodology it has developed for using historic weather data, downscaled climate projections, and GMP data on outages and assets towards this purpose.

Climate and Weather Data

Historical Weather Data

We use daily aggregations of the ERA5 Land reanalysis product (source) dated from 1980 to 2026 to observe historic trends in temperature, precipitation and wind. We obtained raw data as a 0.1x0.1 degree gridded product (approximately $11KM^2$) and conducted spatial joins of the appropriate variables to feeders as well as producing an aggregated summary across Green Mountain Power's service territory.

Projected Climate Data

Initially, we assessed 3 data providers (NEXGDDP-CMIP6,[1] MACAv2-METDATA,[2] CRCM5-CMIP6[3]) and determined CRCM5-CMIP6 to provide the best balance of recency, spatial resolution and capture of extreme weather events. CRCM5-CMIP6 is a dynamically downscaled climate product, meaning that it incorporates additional local variability through landcover, topography and specific weather conditions.

In our analysis we operationalized all available ensemble members (CanESM5, CNRM-ESM2-1, MPI-ESM-1-2-LR, NorESM-MM) across SSPs (1-2.6, 2-4.45, 3-7.0, 5-8.5*) to compute a credible range of expected values under different warming scenarios. Native hourly data was aggregated to a daily cadence by applying temporal aggregations to derive daily-observed minimums, maximums and means. *(SSP 5-8.5 is only available for CanESM5)

As the CRCM5 data is natively available in polar-coordinates, we undertook regridding using the ESMF software package to reproject to a cartesian latitude/longitude grid. This was followed by the same spatial joining and aggregation process we used for the historic weather product.

After our preliminary analysis, we applied Quantile Delta Mapping (QDM) to bias-correct future climate projections using our historical weather product as the reference. QDM adjusts the distribution of each climate variable so that projected values are consistent with the statistical

---

[1] Source
[2] Source
[3] Source

characteristics of our observed data, even when the underlying climate model is trained on a different observational dataset. We selected QDM because it preserves the structure of future changes, particularly weather extremes, more effectively than alternative bias-correction methods. For an overview on QDM and its application, please refer to Cannon 2015 (source) and Gergel 2024 (source).

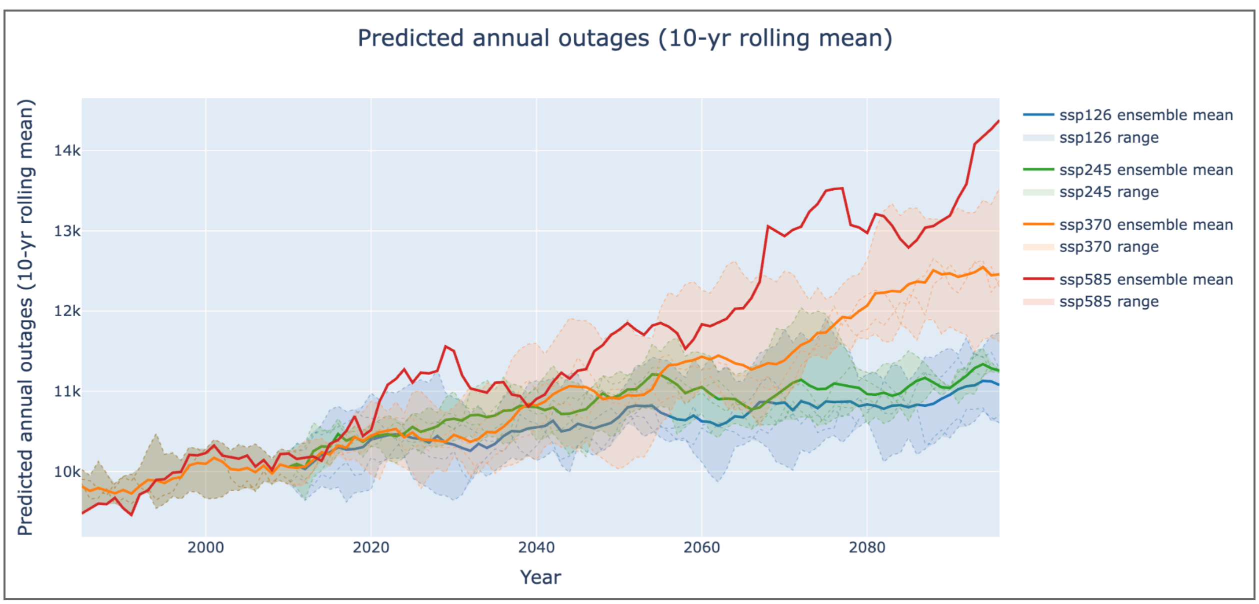
Land Attributes

In addition to the weather and climate data analyzed, we also derived feeder-level metrics for approximate canopy cover (source), average canopy height (source), and fractional landcover classifications (source).

Outage Model

Final Deliverable: Feeder-Level Outage Projections

These are in **feeder_projections.csv**. For each year, these take into account the climate projections +/- 5 years on either side of that year. Each has all feeder-level data fed through a QDM to ensure alignment between historical reanalysis data used in the outage model and projected climate data. In the file, the “min” represents the lowest rolling average of the global circulation models in the ensemble. The “max” represents the highest, and the mean represents the average across the rolling averages.



System-level chart showing the spread across different GCMs within a given scenario.

These bands reflect a similar approach as seen below at a system level. As a sanity check, final system-level counts for 2020-2030 are within 5% of the observed 2020-2024 (10,642 modeled vs 11,094 observed).

Input Data

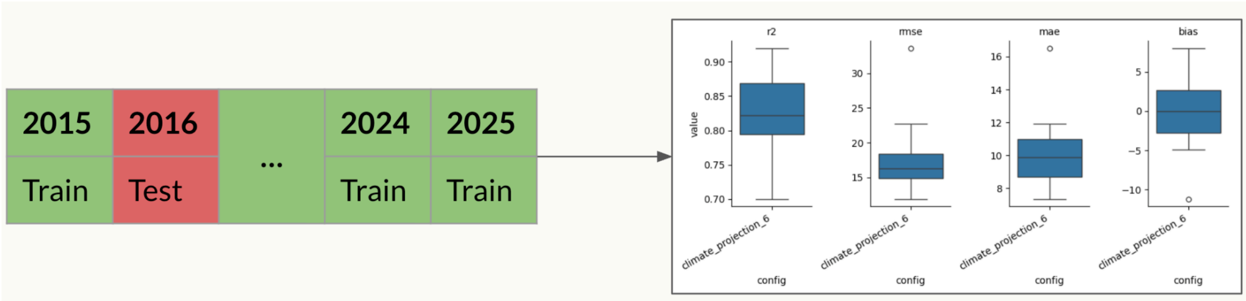
The core of our modeling effort is feeder-level aggregates of GMP's outage and asset data, blended together with (1) satellite-derived descriptions of tree cover and land cover type, (2) historical weather reanalysis data, and (3) projected climate data. We use these data to create the following feeder-level feature sets for every feeder in GMP's territory. The models are trained across all feeders in GMP's service territory, and the projections are provided across the 10 feeders of interest for the BCA.

<table>
<tr><th>Data Type</th><th>Source</th><th>Sample Features per Circuit</th></tr>
<tr><td>Outages</td><td>Green Mountain Power</td><td># Outage Events</td></tr>
<tr><td>Poles, Transformers, Conductors</td><td>Green Mountain Power</td><td>Summary Statistics as Available</td></tr>
<tr><td>Land</td><td>Google Dynamic World, GFCC, WRI + Meta</td><td>Canopy height, canopy coverage, % in different land cover classes</td></tr>
<tr><td>Weather</td><td>ERA5-Land</td><td rowspan="2">Min/max temperature, total precipitation, mean and max wind speed</td></tr>
<tr><td>Climate</td><td>CRCM5</td></tr>
</table>

Features for each feeder in the GMP service territory.

Methods & Validation

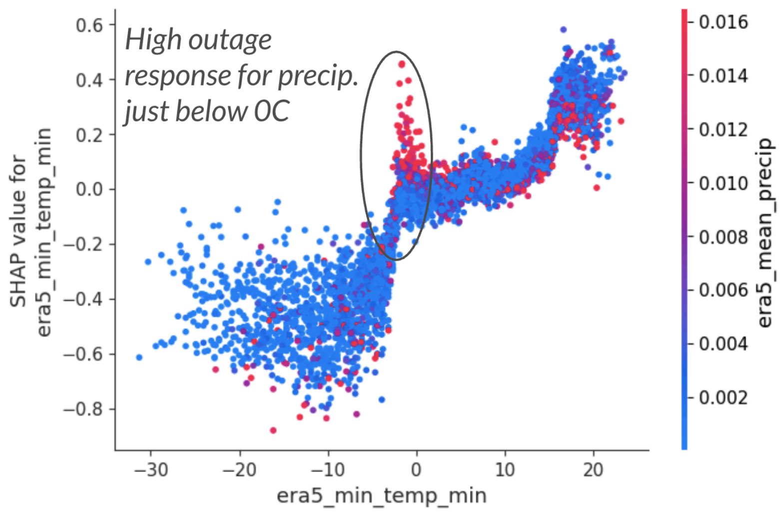
We use these input data to run a set of experiments in predicting annual outage counts at a feeder level. We test a variety of model architectures, in particular Ridge Regressions, Random Forests, and boosted trees implemented in XGBoost. Our primary mode of evaluation involves a "leave one year out" cross validation: We iteratively hold out one year of data and test how well we predict feeder-level outages on that unseen year, for each year. We verify this way that our results are relatively stable across years, and that our results capture 80-90% of the variation in yearly outages across feeders. The XGBoost models proved most performant and were the basis of our downstream projections.



Sample fold for "leave one year out" cross validation (left) with validation metrics across feeder-years (right).

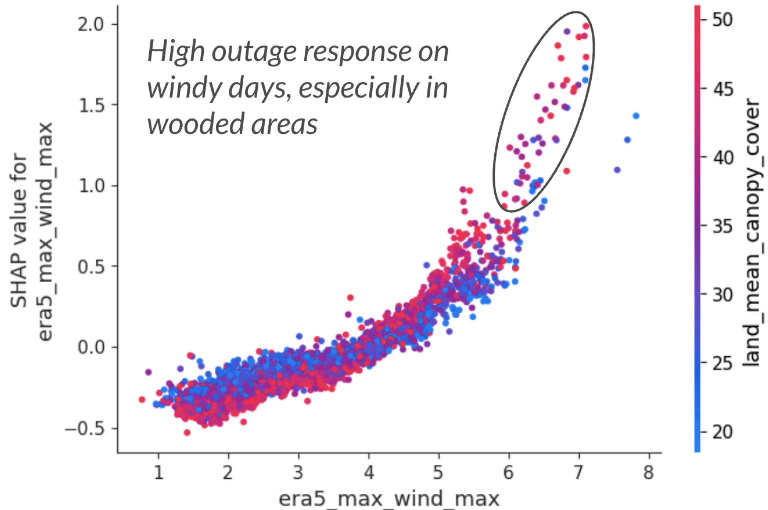
We also make sure that the model is picking up weather-outage trends that will be crucial in projecting outages into the future: once we train our models using past data, we run inference on each of the individual climate models. Here we use a combination of Shapley values and feature importance plots.

When we look at the relationship between minimum temperature and outages, controlling for precipitation, we see evidence of a spike in outages just below 0C on days with precipitation.



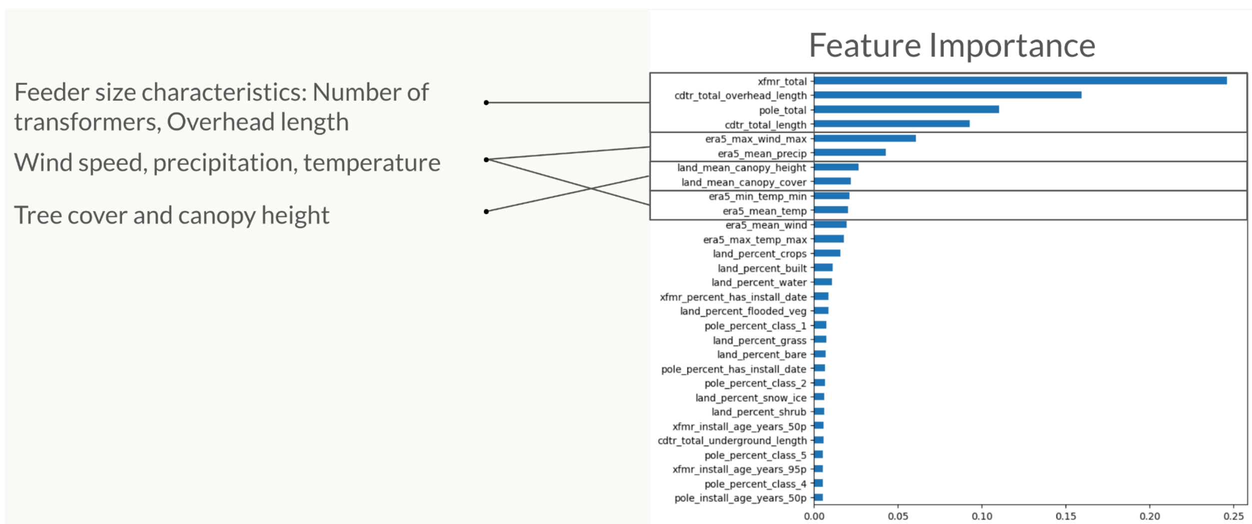
Effect of cold temperatures on outages, controlling for precipitation. Units normalized to reflected outages per day.

We see a more straightforward, upward picture when we look at the relationship between wind speed and outages: a near-monotonic increase in outage risk, potentially mediated by vegetation.



Effect of wind on outages, controlling for tree cover. Units normalized to reflect outages per 1 day.

When we look at feature importance more broadly, we see that among the most important predictors are feeder size, followed by weather and tree cover characteristics.



Feature importance for the outage model.

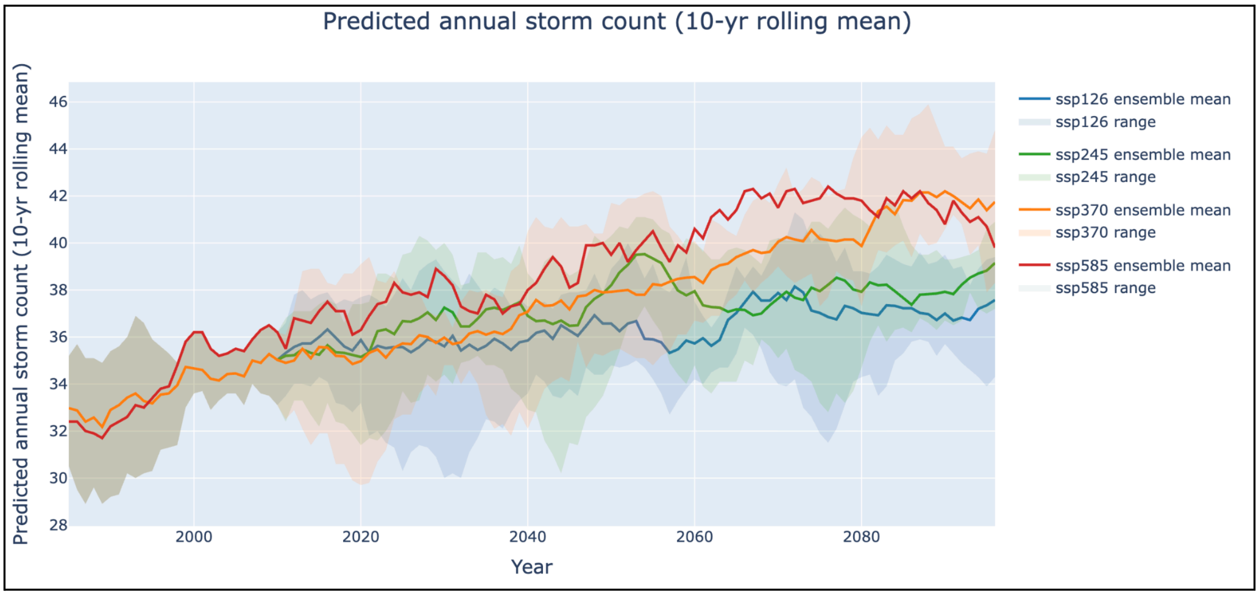
We use these models across different global circulation models to quantify the model uncertainty and across different future scenarios to quantify the driver uncertainty.

Storm Model

Primary Deliverable: System-Level Storm Projections

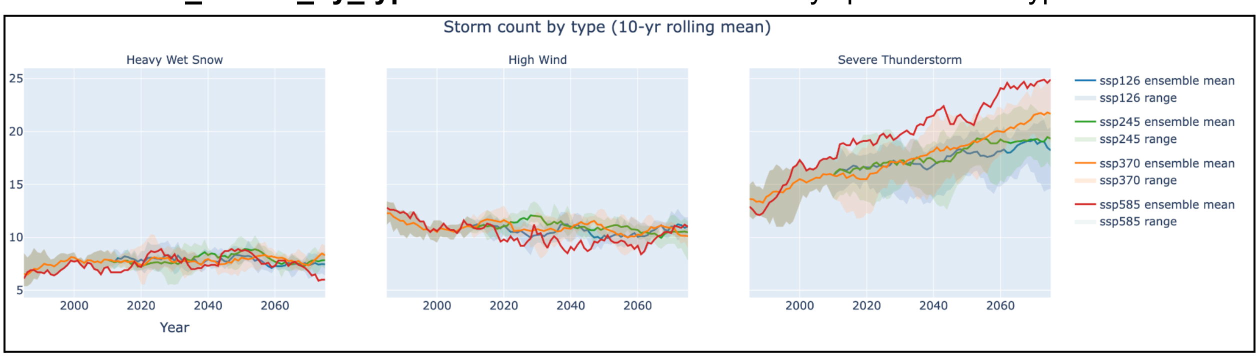
There are two relevant files here: **all_storm_counts.csv** and **storm_counts_by_type.csv**. Both use the same +/-5 year windowing strategy as above, on all storms.

The file **all_storm_counts.csv** reflects the total number of predicted Heavy Wet Snow, High Wind, and Severe Thunderstorm storms from the two-stage classifier. Again, the min and max are the high and low ensemble members.



Storm types overall.

The file **storm_counts_by_type.csv** reflects the breakdown by specific storm type.



Storms by type.

Data

We use the ERA5 Land data, in conjunction with the major and minor storm data provided by GMP, to train our storm models. For each day, we create a range of features that describe the weather over each of the previous 4 days. The features include statistical summaries of min temperature, max temperature, max wind, mean wind, and precipitation. Note that snow here is not included directly in the climate projections model, but we validate that the model can identify snow and winter storms through the combination of temperature, precipitation, and wind.

Our final training dataset has a combination of (1) a label for whether the day is in a storm, (2) the particular type and severity of the storm, and (3) statistical summaries of the weather as described above. For the positive samples, we include the weather sample at the end of a three day window period from the start of the storm period noted in the data provided by GMP. In order to verify the validity of this window, we cross-referenced 11 of the recorded storm periods in GMP's storm dataset that were longer than 72 hours, against records of the storms in NOAA's Storm Events Database (source).[4] Even among the more severe storms, the most severe weather periods appeared to be within the first 72 hours of the storm periods.

We create equivalent datasets for inference, for each of the global circulation models in the CRCM5 dataset. As above, we use the spread across GCMs to capture the *model uncertainty* in the projections of future climate, and we use the spread across scenarios to capture the scenario uncertainty between different levels of emissions.

Methods & Validation

Our storm model consists of two stages:

1. A weather severity model to identify the occurrence of potentially damaging storms in future climate data
2. A multiclass classifier to determine which type of storm is occurring

We again run our predictions using a variant of the "leave one year out" cross validation used in the outage model. Again, we experimented with three different models: a logistic regression (standard linear classifier), an XGBoost model (highly performant on tabular data but prone to overfitting), and a Catboost model (also boosted trees, but known to be slightly more robust to outliers and overfitting). We found that the Catboost model had the most stable out of sample predictions across years, and comparable or better performance to the XGBoost implementation. This was the one that we proceeded with.

In the first stage, we filter storms based on a weather severity cutoff to make sure we are capturing roughly 80-90% of the storms in the database. In order to capture the majority of storms, we also pick about 3 "near-storm" periods for each storm noted in the dataset. These "near-storm" periods have similar weather conditions as the storms but do not officially show up in the storm database provided by GMP. Though these are not explicitly trained on cost data, we

[4] These storms are STORM14I, STORM18S, STORM23H, STORM17O, STORM23E, STORM22Q, STORM22R, STORM19K, STORM24F, STORM18F, STORM14H.

show below that the first-stage severity scores, normalized to percentiles, are well correlated with observed historical storm costs.

In the second stage, we create confusion matrices to validate that these models are robustly distinguishing storm styles. We find that the correct, modeled classification is stronger for more severe storms, and the highest rates of “confusion” involve storms that are classified in the dataset as “High Wind”.

pred	**Heavy Wet Snow**	**High Wind**	**Severe Thunderstorm**	**Storm-Like**
true				
Heavy Wet Snow	5	0	0	1
High Wind	1	5	0	2
Severe Thunderstorm	0	1	3	1

Confusion matrix for True Major Storms. 5 of the 6 Major Wet Snow Storms are properly identified out of sample.

pred	**Heavy Wet Snow**	**High Wind**	**Ice Storm**	**Severe Thunderstorm**	**Storm-Like**
true					
Heavy Wet Snow	18	7	2	1	7
High Wind	8	30	1	3	20
Ice Storm	3	1	0	0	0
Severe Thunderstorm	0	3	0	36	18

Confusion matrix for Observed Minor Storms. 28 of 35 Minor Wet Snow Storms are identified as storms, though 7 are classified as High Wind storms instead of Heavy Wet Snow.