Exploring Environmental Factors and their Impact on Solar Power

Sandra Villamar, Zhenduo Wen, Zian Wang, Shixuan Wu, Zengrui Li

Agenda

- Introduction
- Initial Exploration of Variables
- Correlation to Power
- Extract Significant Variables
- Power Prediction

Introduction

Background

- Solar energy is a rapidly **growing** market.
- Predicting solar power output is **vital**.
  - Maintaining current solar plants
  - Deciding where to build next
- Power is normally modeled **with irradiance** data.
  - Measurements require specific sensors which are **costly in time and money**.
- We want to **predict power** using **only weather** and **location** features.

Objective

- Find the **weather** and **location features** that best **explain** solar **power** output.

Data

- **Power output + environmental factors** from **12 US sites** over **14 months** (2017 to 2018)
- 21K observations
- Features:

Location:
- Latitude
- Longitude

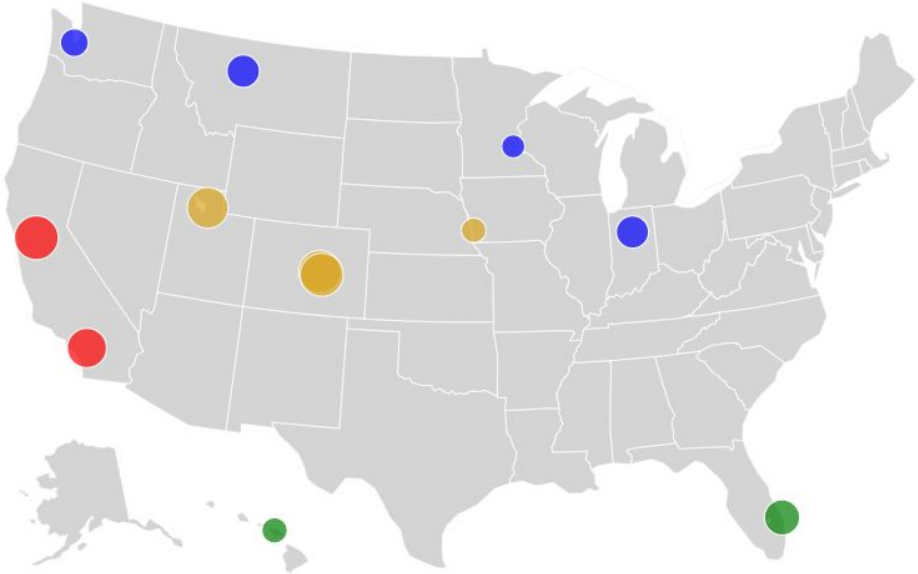
Time:
- Date
- Time sampled
- Season

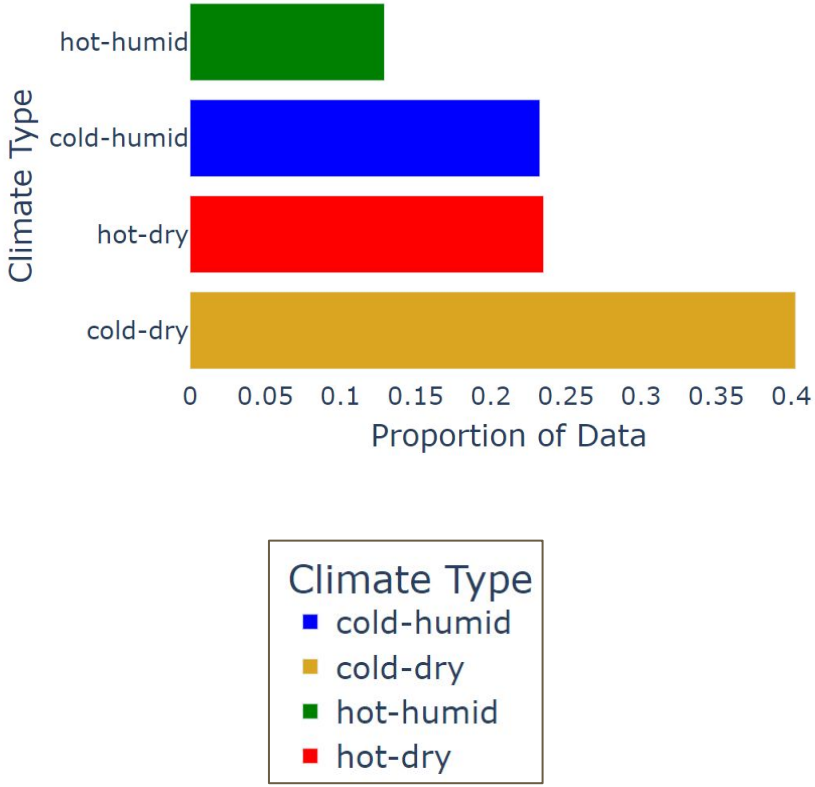
Weather:
- Altitude
- Humidity
- Ambient temperature
- Wind speed
- Visibility
- Pressure
- Cloud Ceiling

Source:
Williams, Jada; Wagner, Torrey (2019), “Northern Hemisphere Horizontal Photovoltaic Power Output Data for 12 Sites”, Mendeley Data, V5, doi: 10.17632/hfhwmn8w24.5

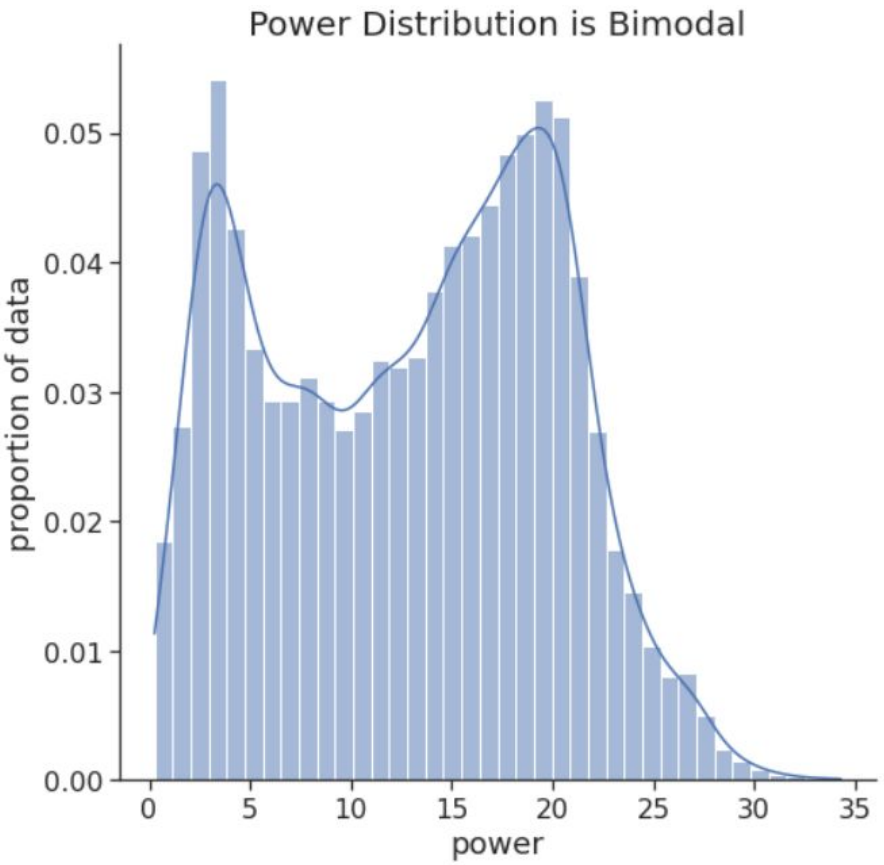
Initial Exploration of Variables

Location of Solar Plants



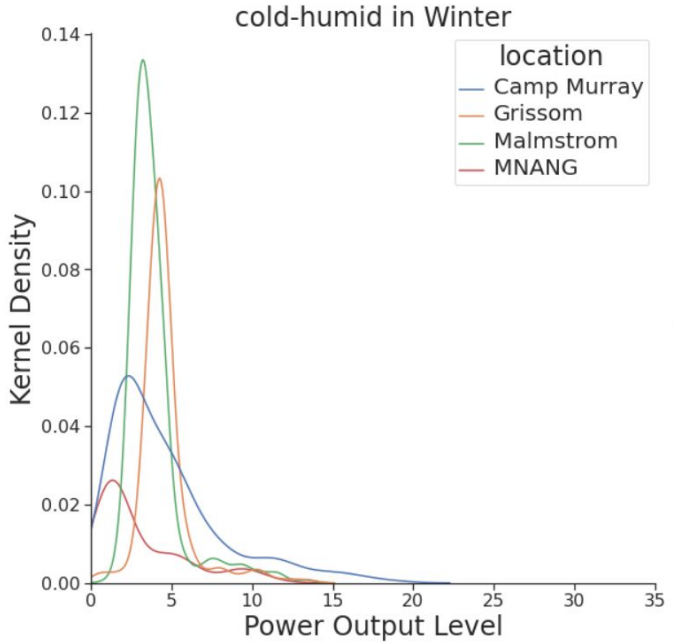


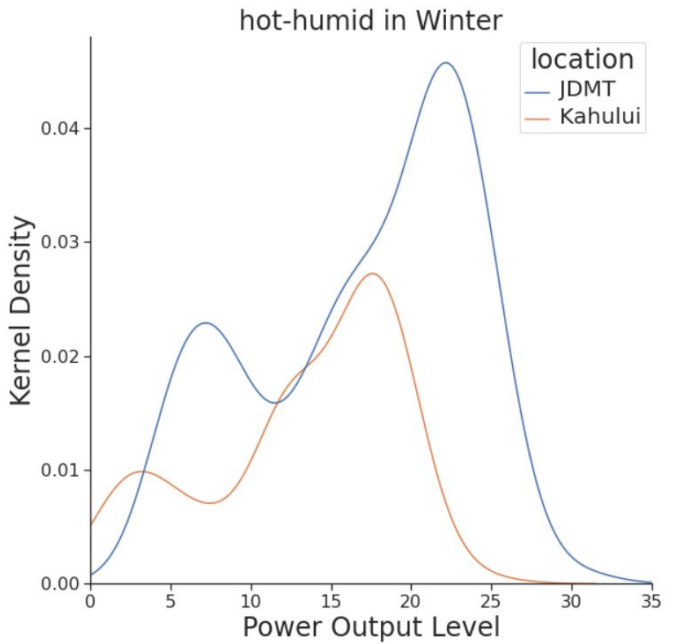
Power


Power Distribution is Bimodal
proportion of data
0.05
0.04
0.03
0.02
0.01
0.00
0
5
10
15
20
25
30
35
power

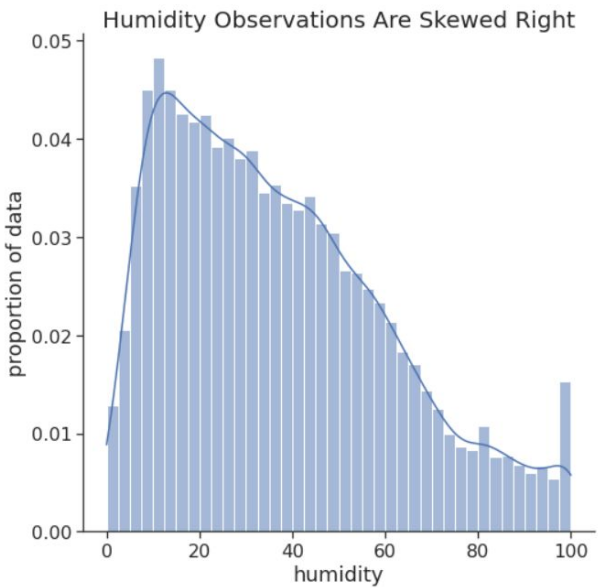
Why Classify by Climate?

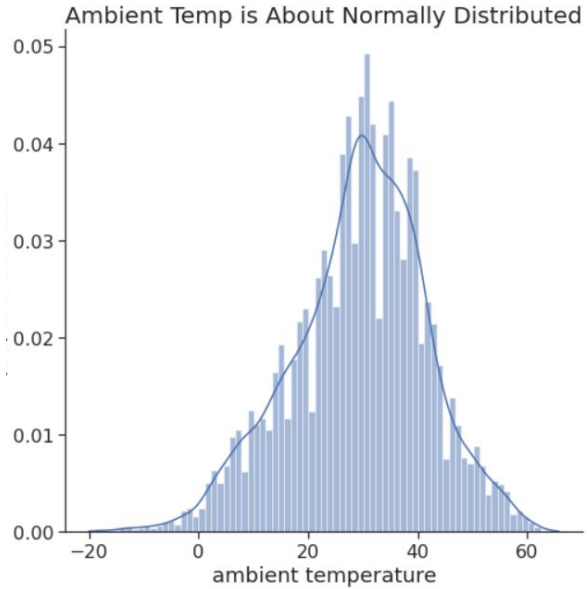
Locations in the Same Climate Have Similar Power Output

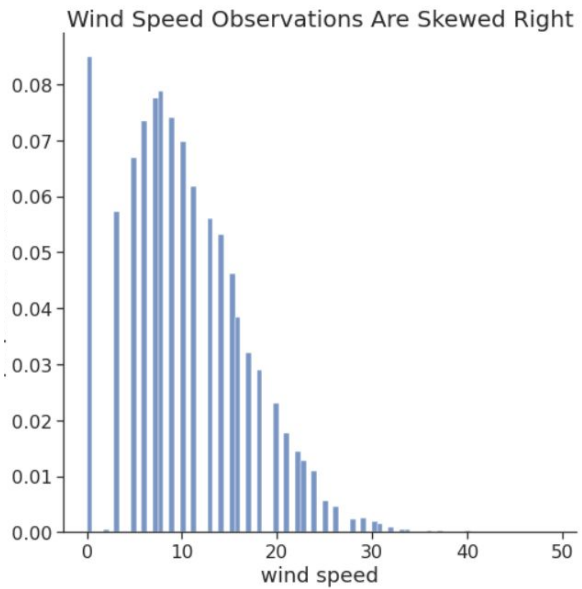




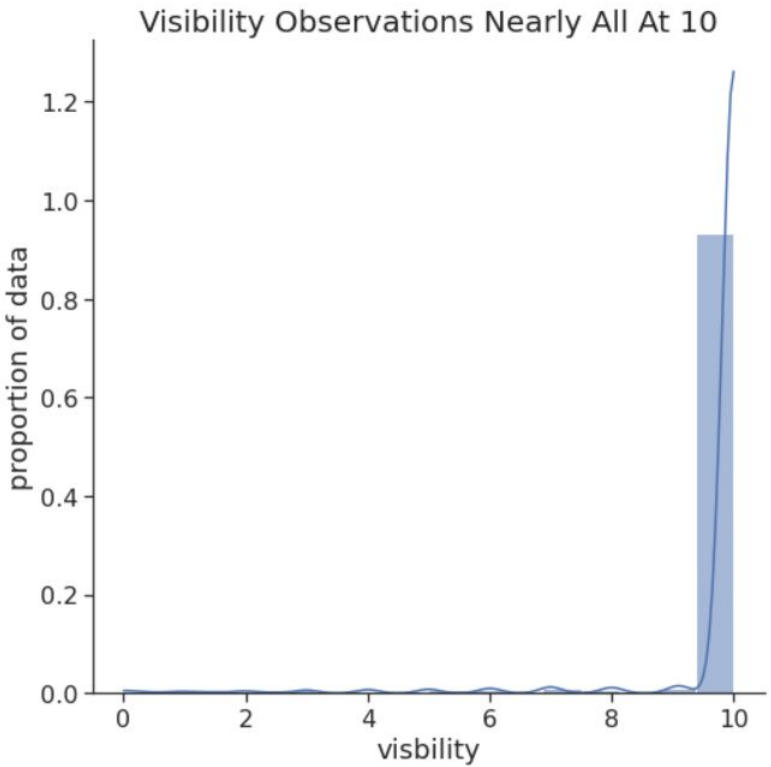
Humidity, Temperature, Wind Speed

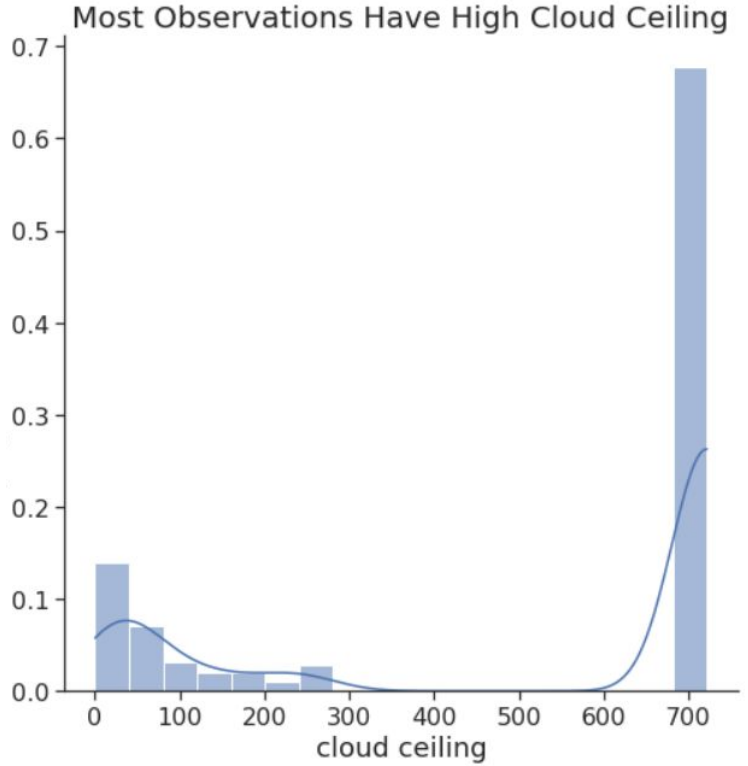






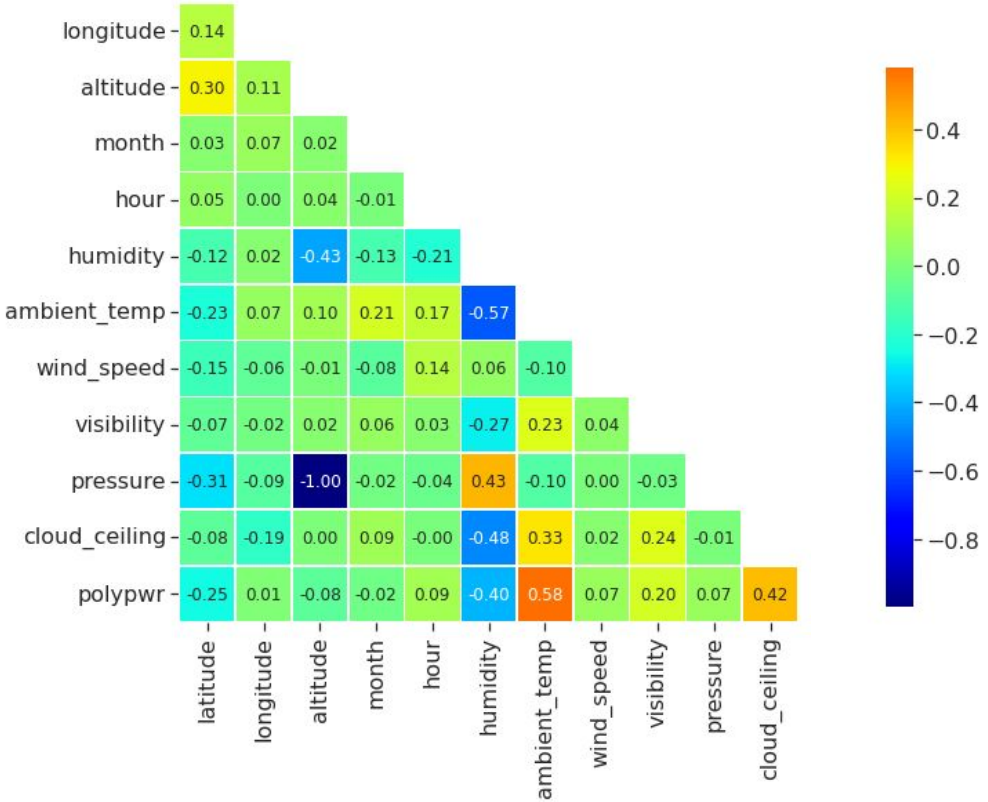
Visibility and Cloud Ceiling



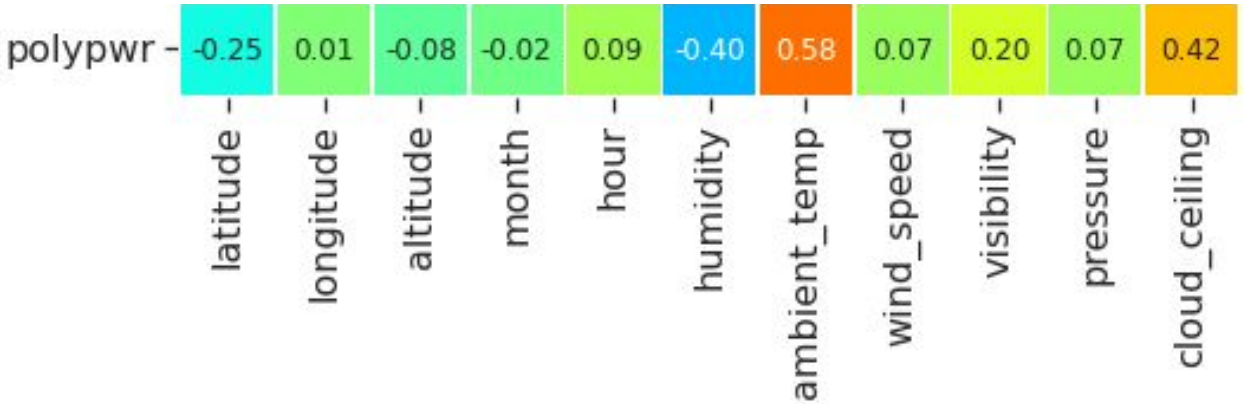


Correlation to Power

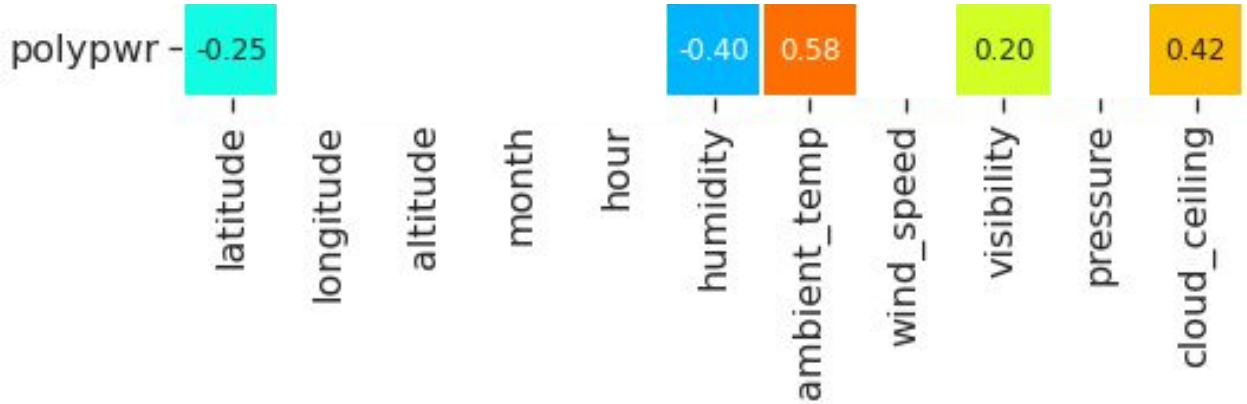
Correlation for All Sites



Correlation Between Variables



Correlation Between Variables

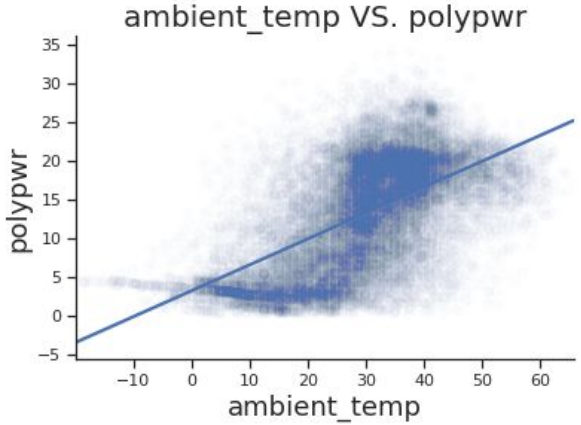


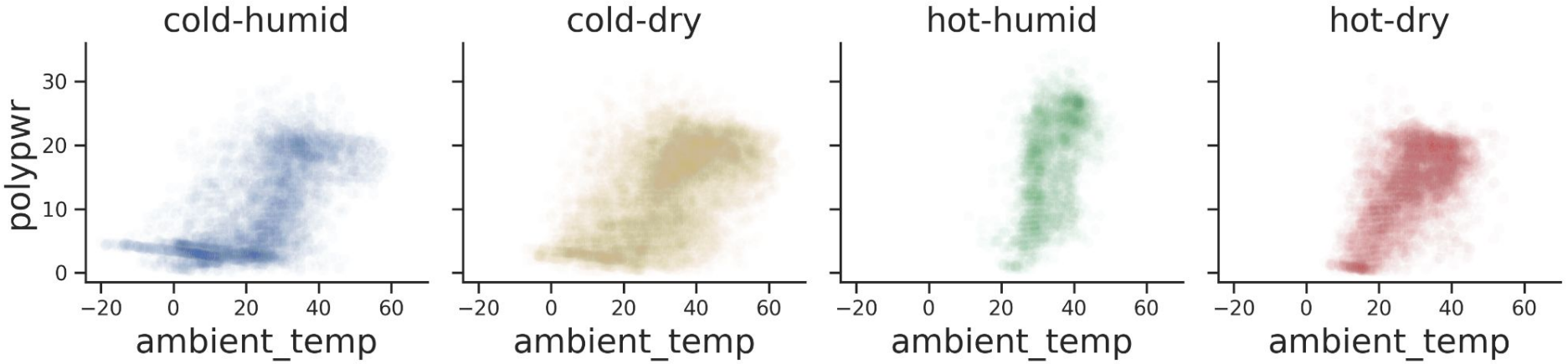
Extracting Variables that Explain Power

Variables that Explain Power

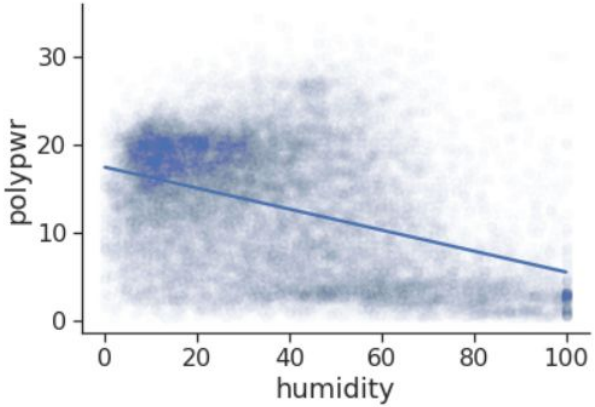
- Climate Type
- Temperature
- Humidity
- Cloud Ceiling
- Visibility

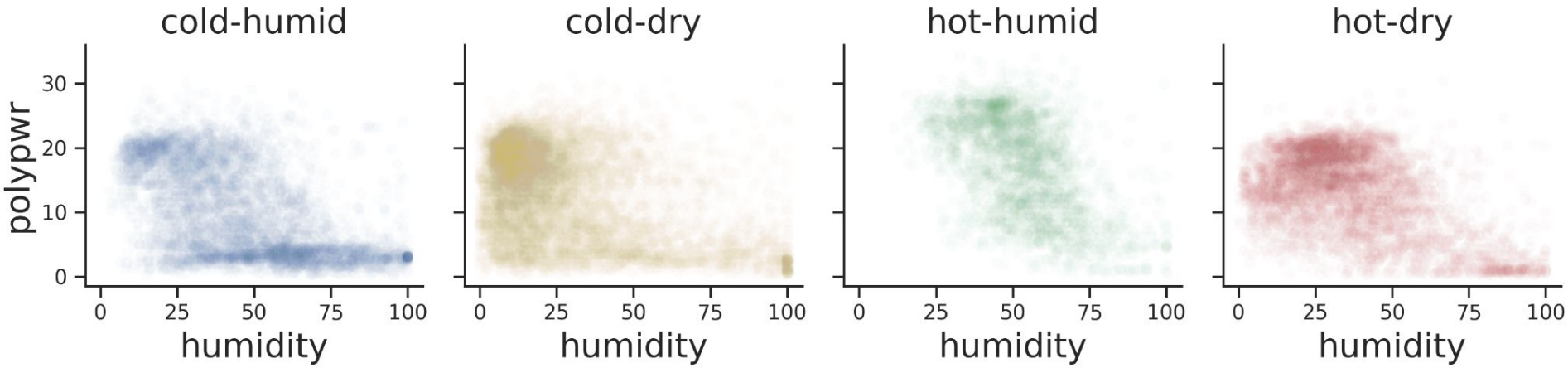
Temperature VS. Power



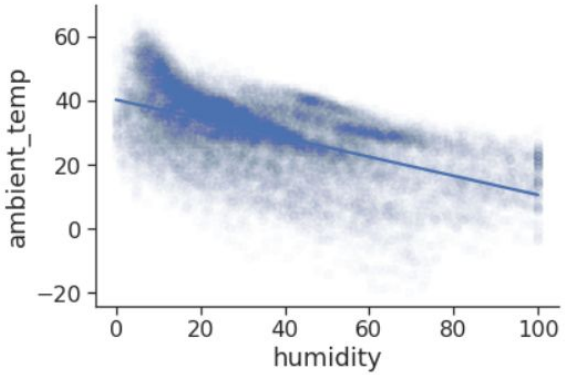


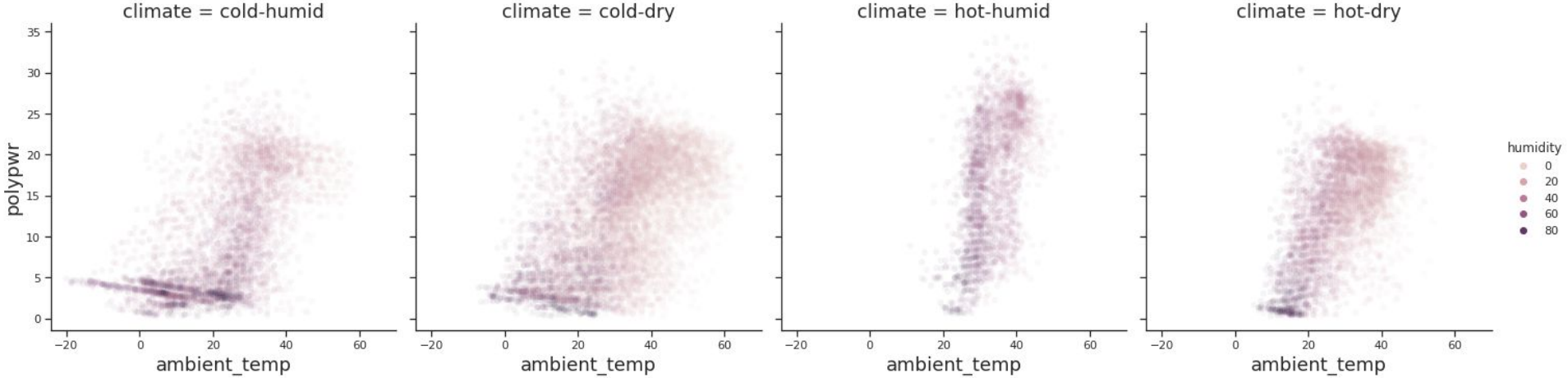
Humidity VS. Power



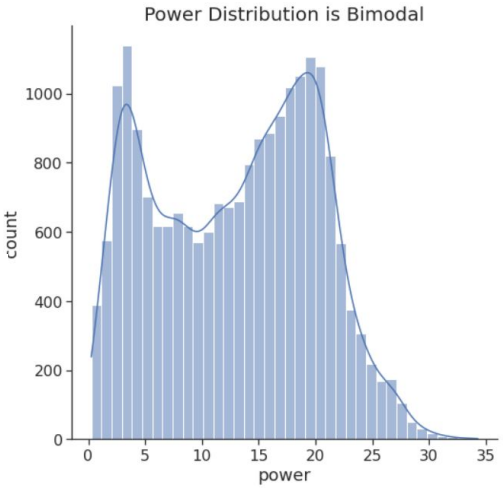


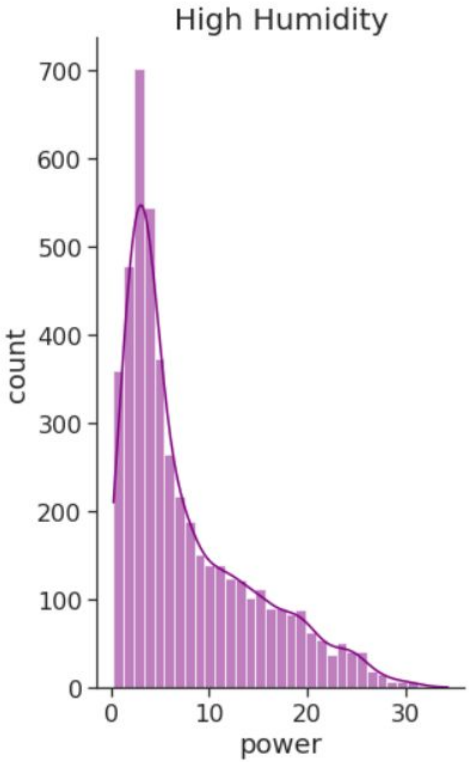
Why Consider both Temperature and Humidity?

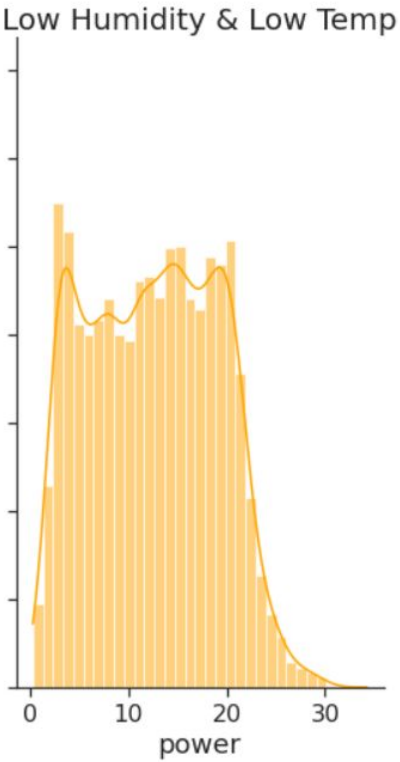


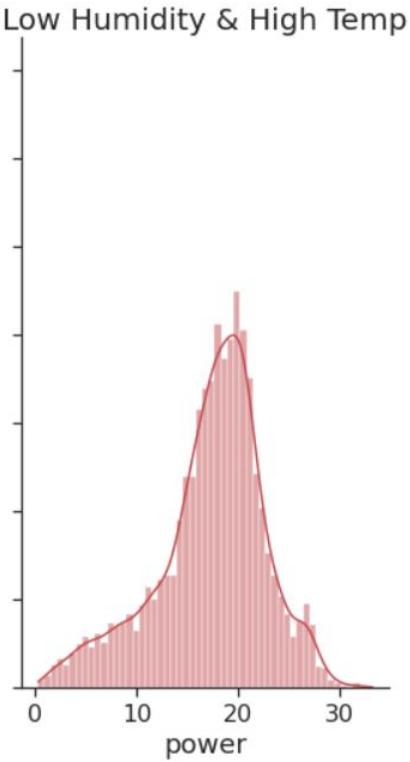


Why is Power Bimodal?





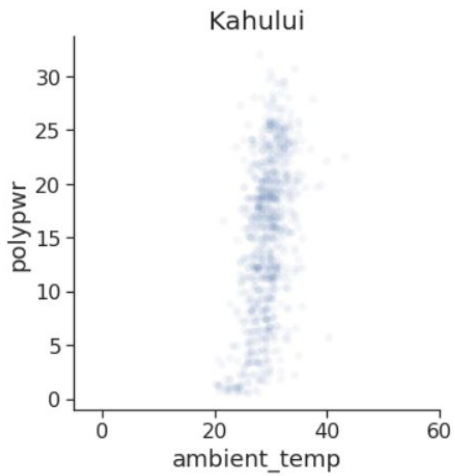


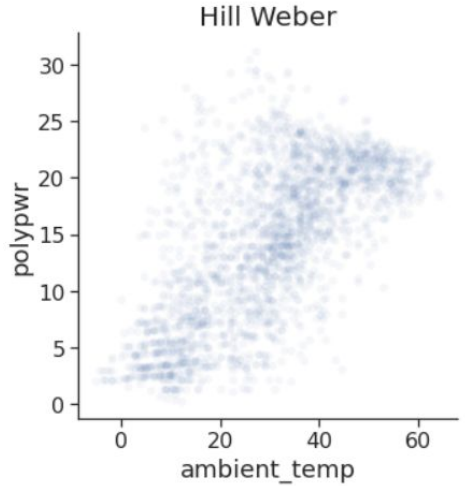


Case Study

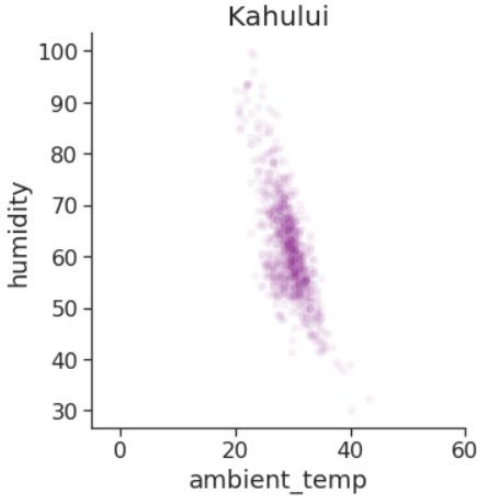
Kahului ‘s temperature remains rather **constant**.

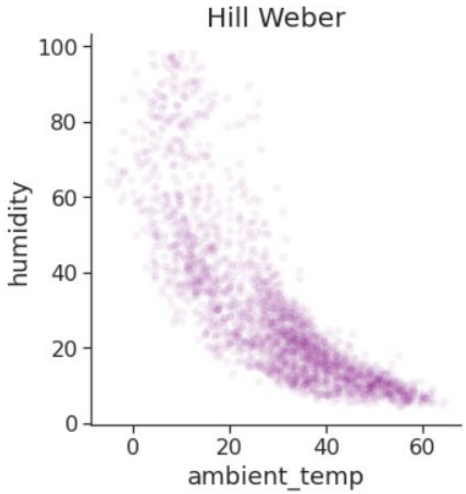
We can conclude that for Kahului, **temperate alone** is **not** a good indicator of **power**.



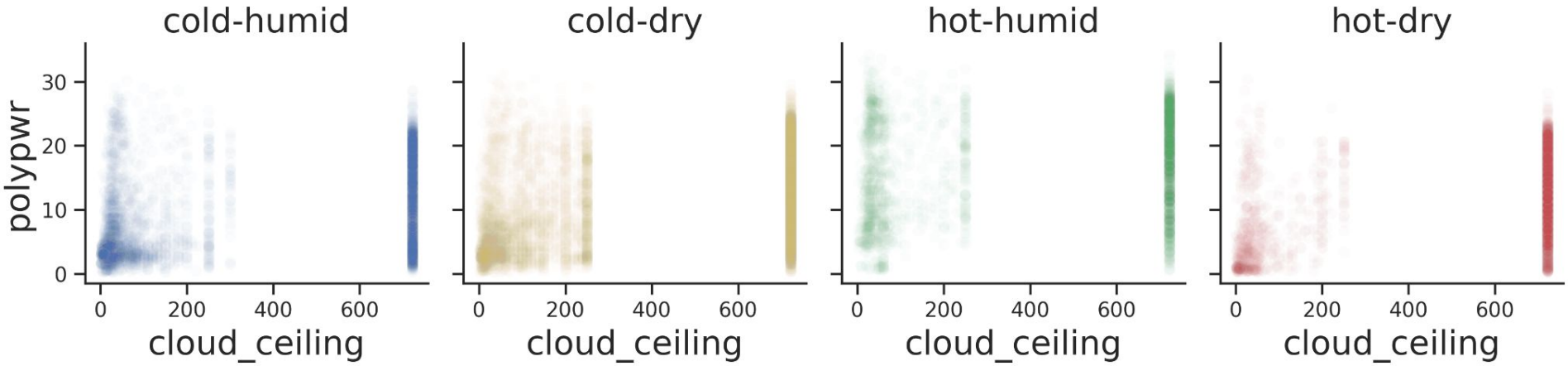


VS.

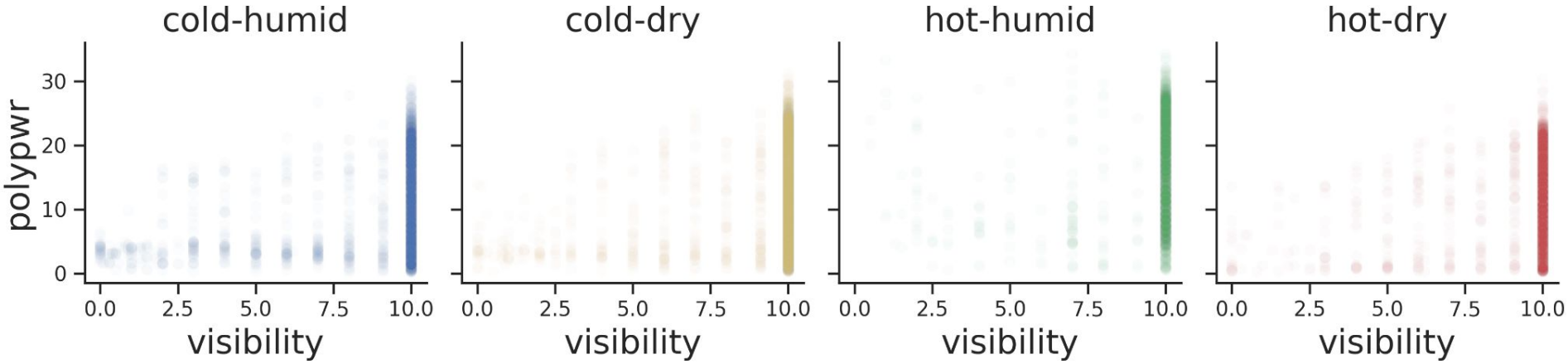




Cloud Ceiling VS. Power



Visibility VS. Power



Power Prediction

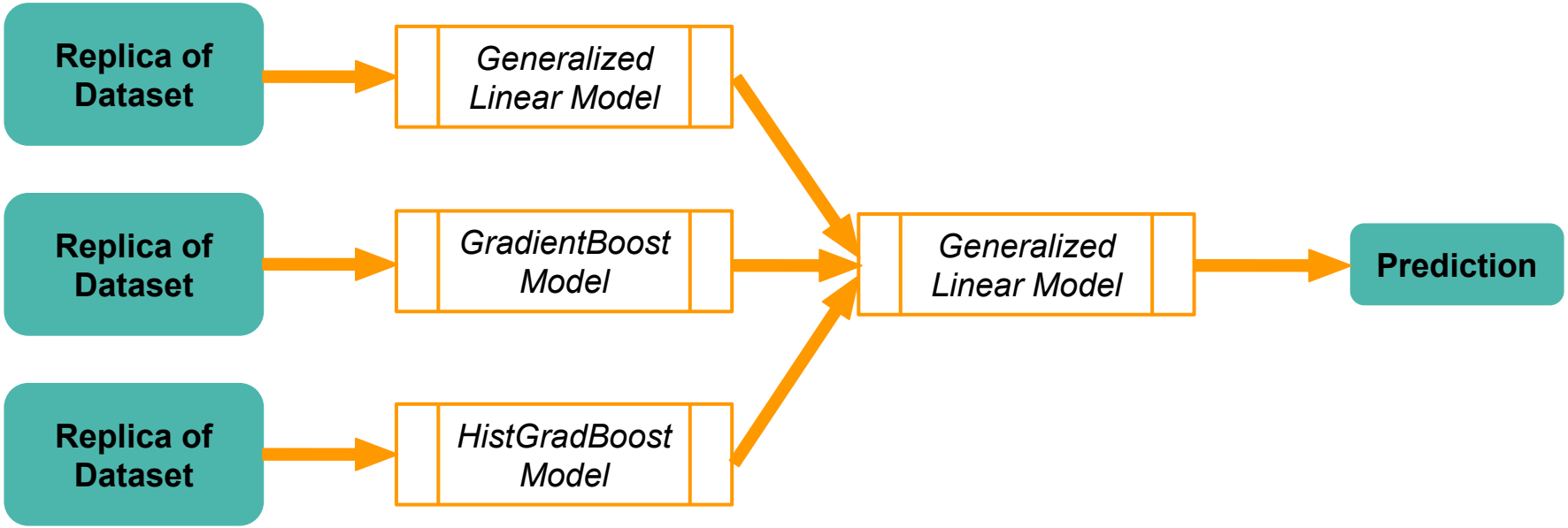
Model Performance for different climates

Optimal Model	*Climate Type*	*Training Set Size*	*RMSE*	*MAE*	*R^2*
GLM	**hot-dry**	4455	3.024	2.223	0.745
GLM	**cold-dry**	7630	4.316	3.211	0.599
HistGradBoosting Regression	**hot-humid**	2448	5.683	4.529	0.442
GradientBoosting Regression	**cold-humid**	4407	3.782	2.548	0.700

In total we have four climate types, we train the model for each climate type. The optimal model for each climate is listed above.

* GLM: Generalized Linear Model
  HistGradBoosting: Histogram Gradient Boosting Regression Model
  “Optimal model” means achieving lowest average of RMSE and MAE

Stacking Model



Trade-offs of Stacking Model

Model	*Climate*	*Training Set Size*	*RMSE*	*MAE*	*R^2*	*Training Time (in seconds)*
GLM	**hot-dry**	4455	3.024	2.223	0.745	0.4363
StackingReg	**hot-dry**	4455	3.029	2.188	0.743	8.8630
GLM	**cold-dry**	7630	4.316	3.211	0.599	0.5601
StackingReg	**cold-dry**	7630	4.249	3.142	0.611	15.6120
GradientBoosting	**cold-humid**	4407	3.782	2.548	0.700	1.3830
StackingReg	**cold-humid**	4407	3.768	2.521	0.711	9.3810
HistGradBoosting	**hot-humid**	2448	5.683	4.529	0.442	0.3631
StackingReg	**hot-humid**	2448	5.591	4.460	0.460535	6.2752

Conclusion

- In the absence of irradiance data, **environmental data** can be used to make **accurate predictions for solar power output**.
- **Which variables** best explain solar power output?
  - Climate type
  - Temperature
  - Humidity
  - Cloud ceiling
- Why is **power bimodal**?
  - High humidity → low power
  - Low humidity + high temperature → high power

Thank you!