

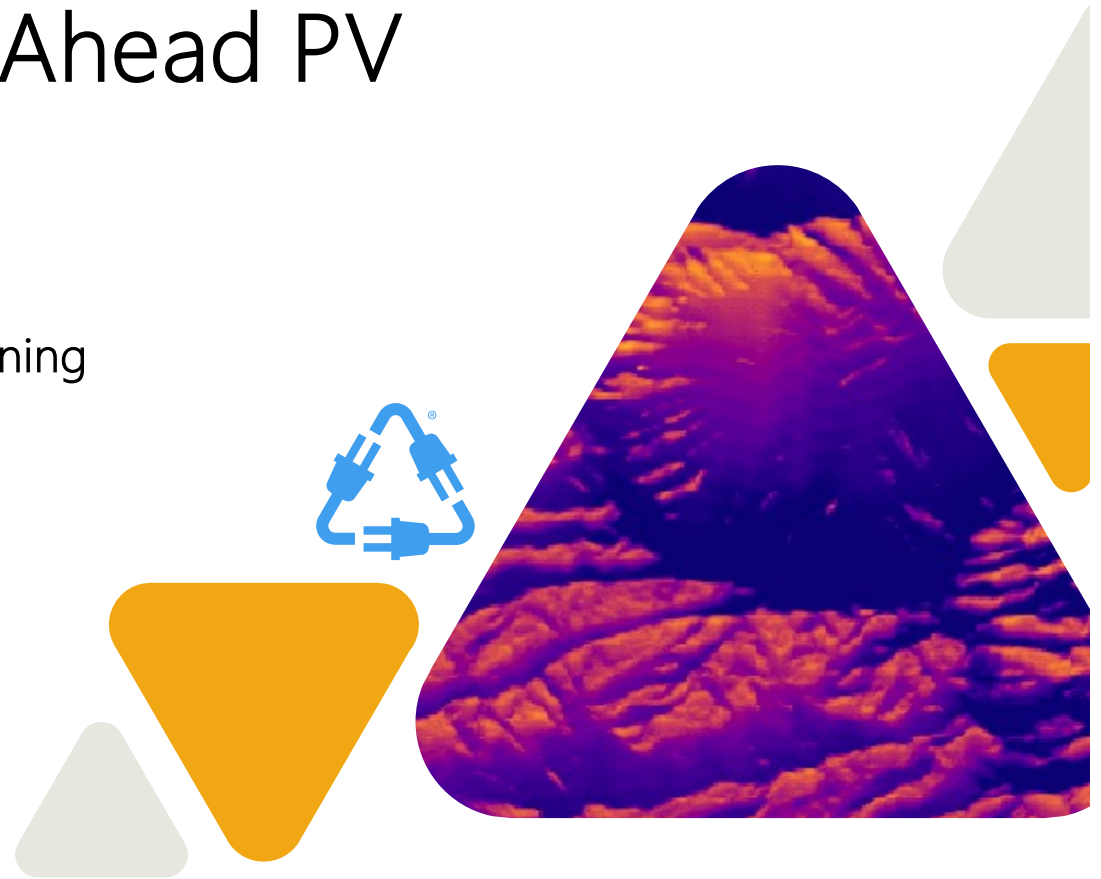
# Reducing Month-Ahead PV Yield Uncertainty

with ENSO-Informed Machine Learning

Marc Perez

May 14<sup>th</sup>, 2025

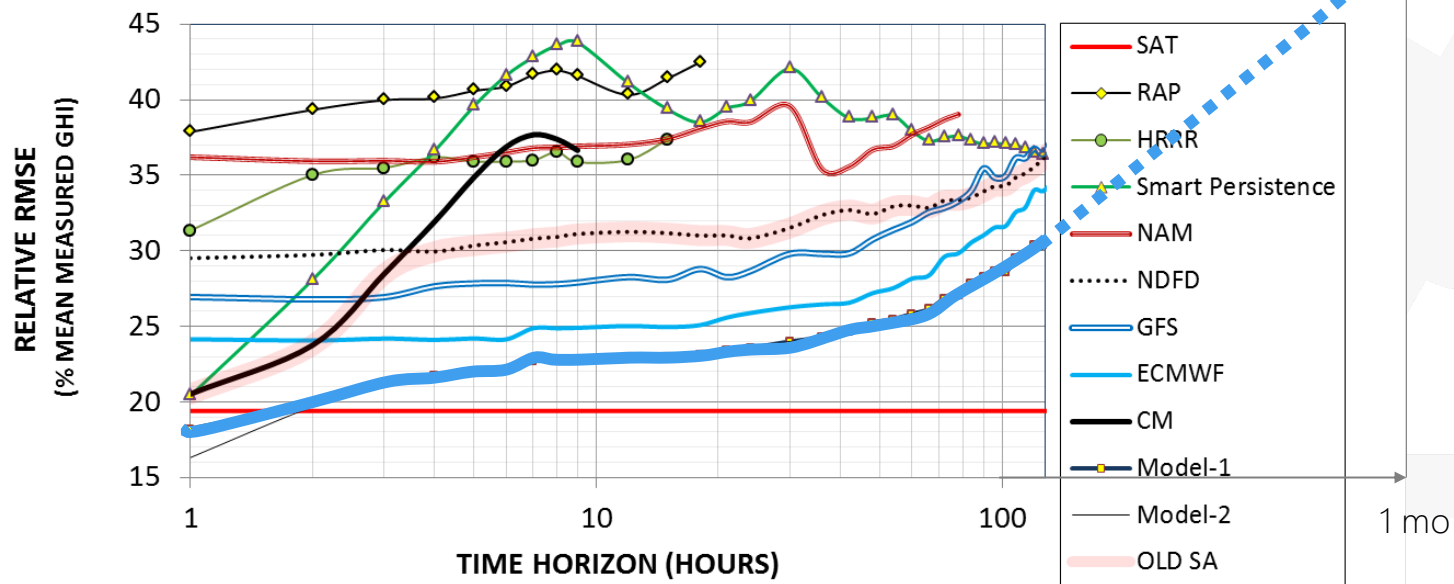
PVPMC '26, ABQ, NM



# Motivation

The sub-seasonal forecasting gap

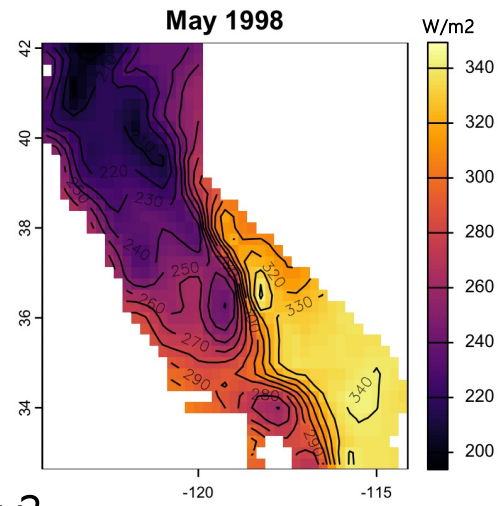
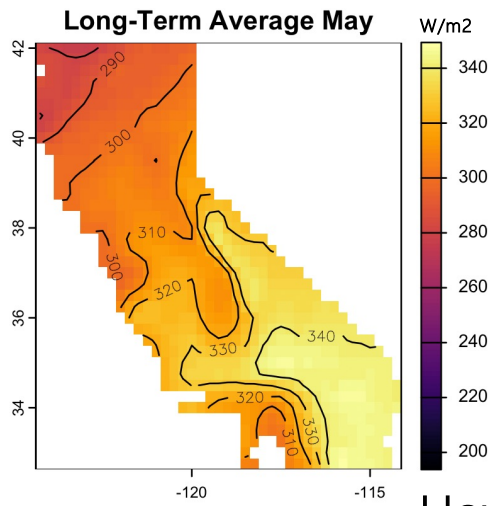
What do we do beyond the skill horizon of the NWP's?



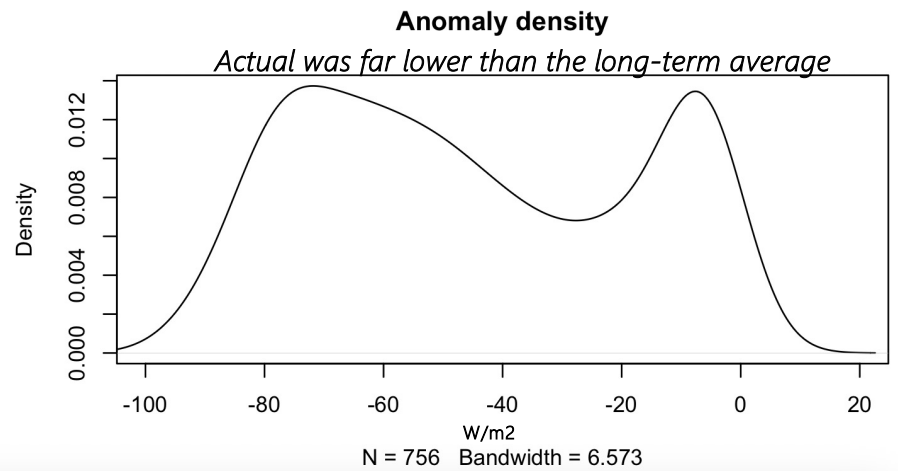
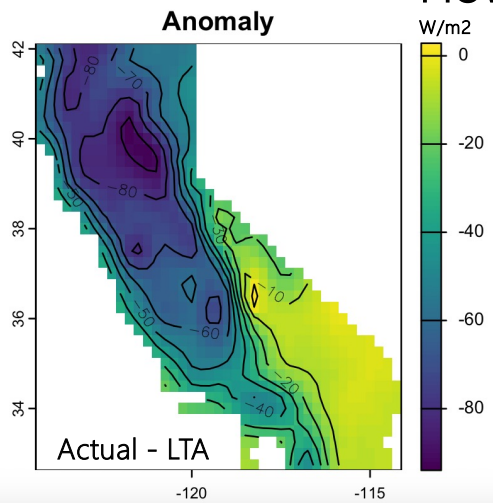
Climatology

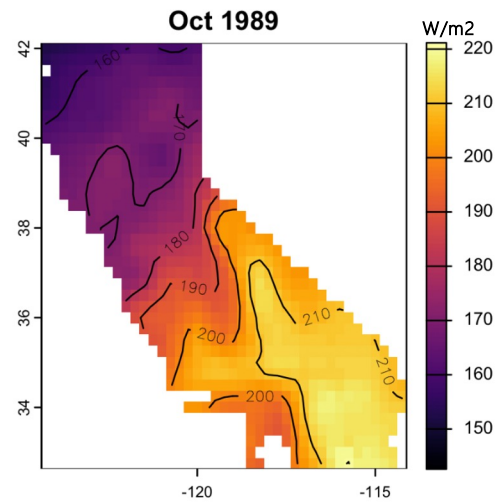
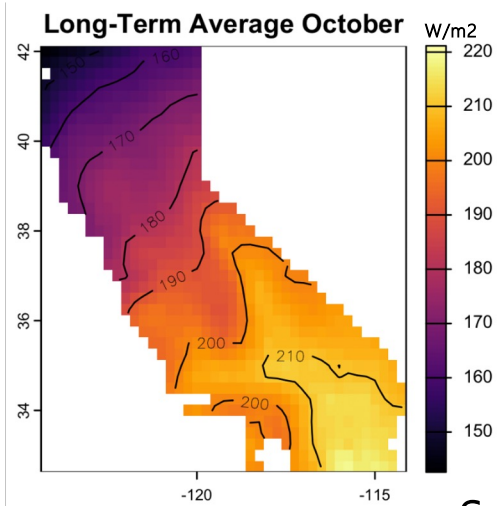


TMY, TGY, LTA

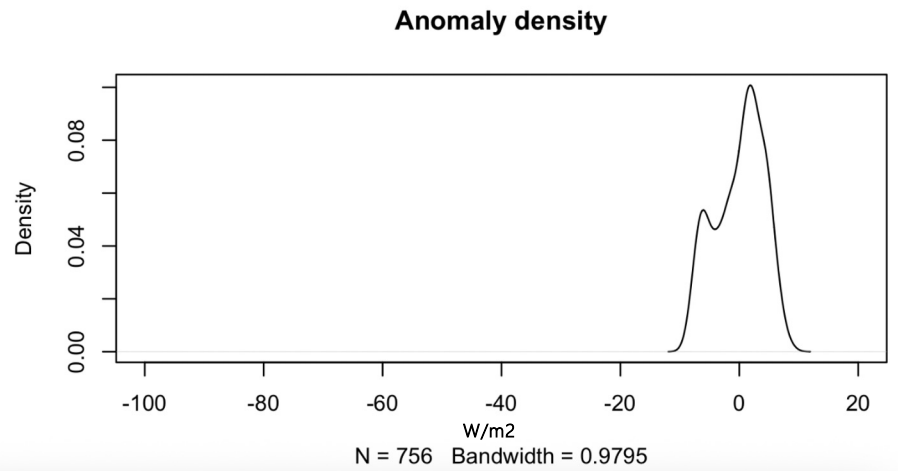
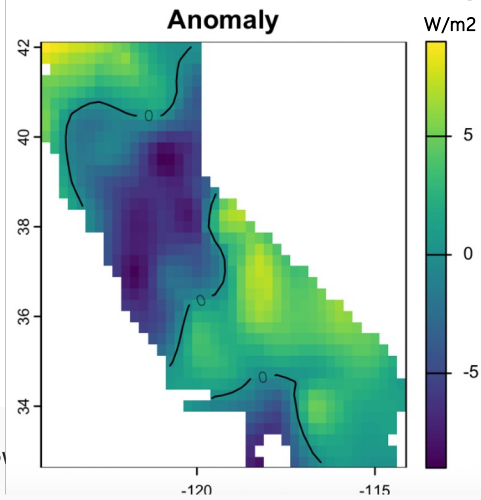


How good is climatology?

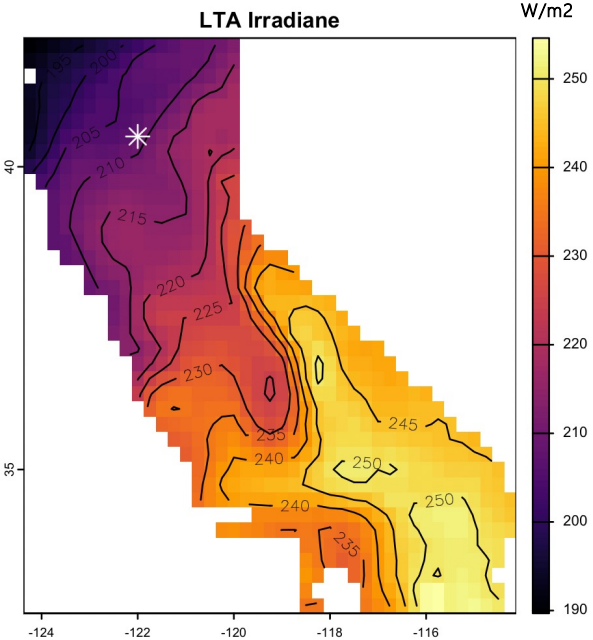




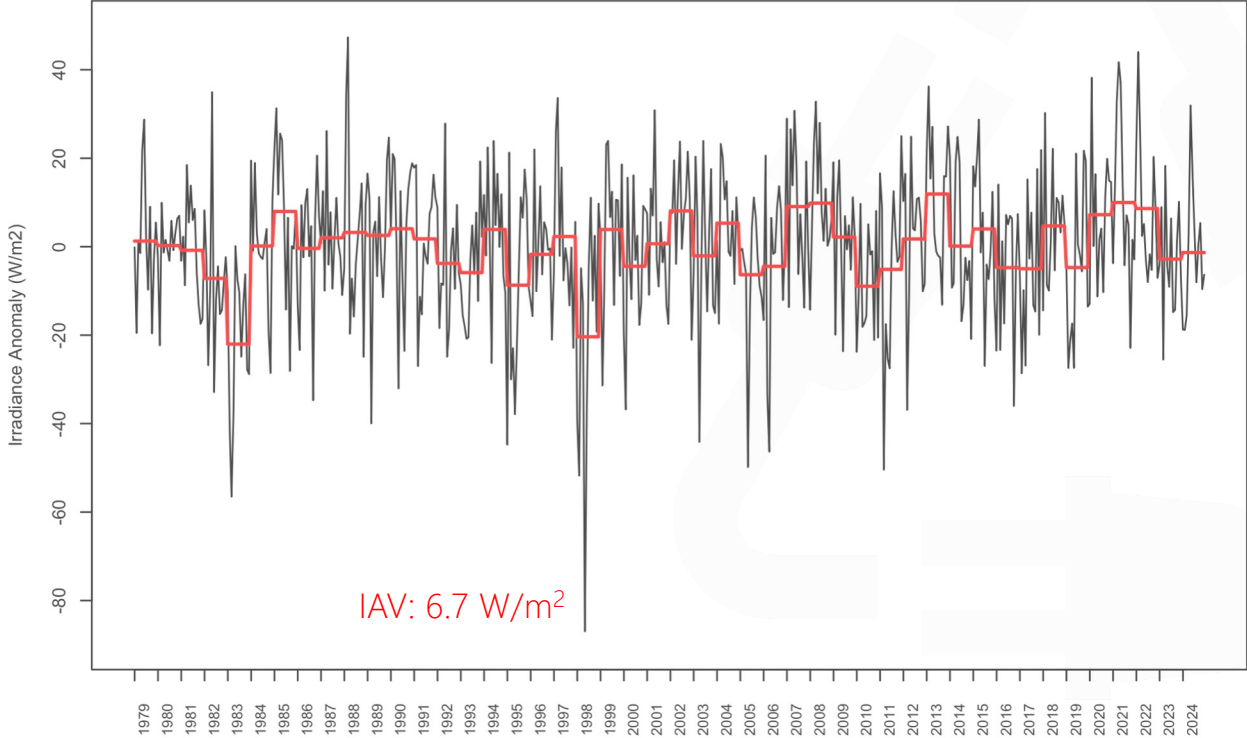
Some months it's better...

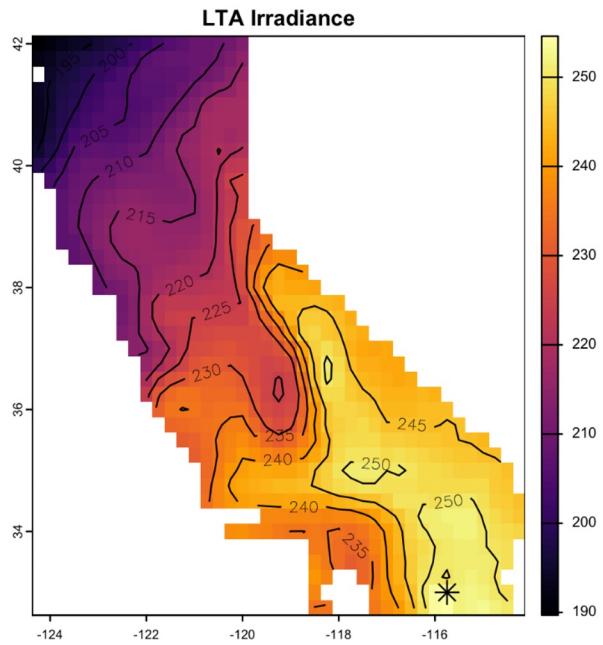


Some sites it's worse...

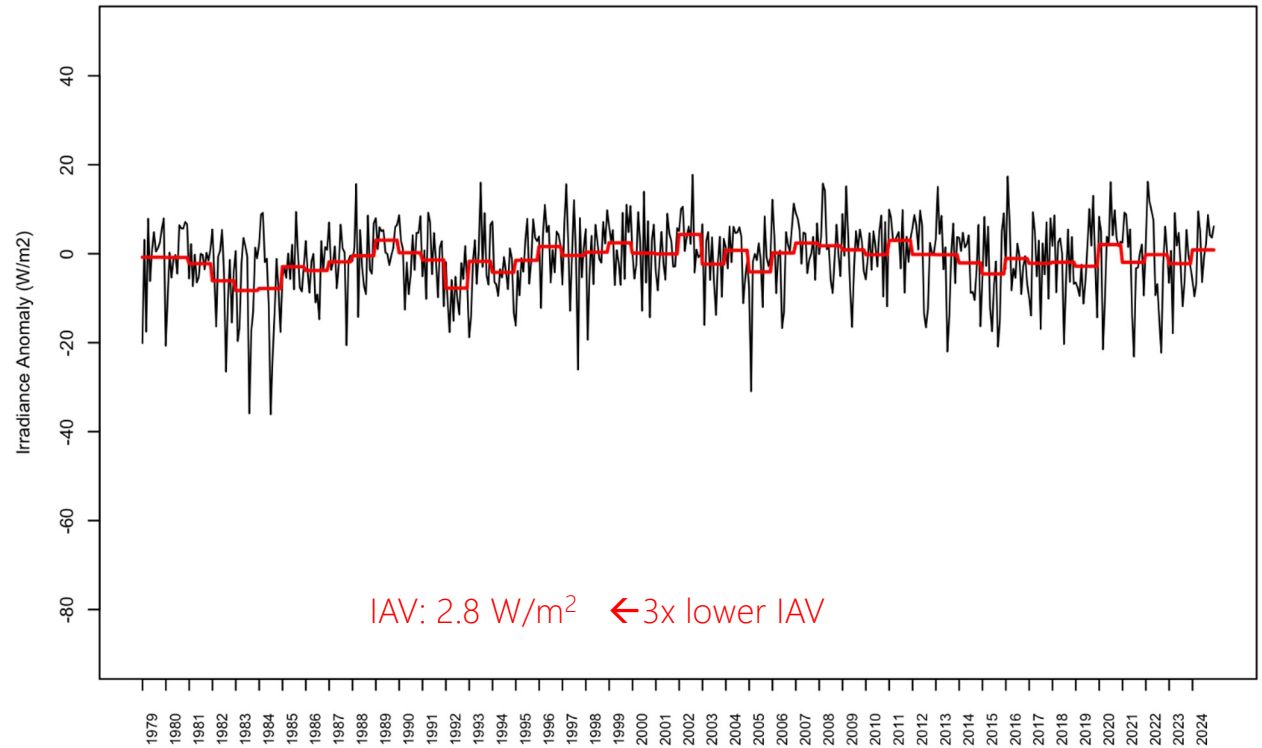


Monthly Anomaly vs. LTA





Monthly Anomaly vs. LTA



# What is the ENSO?

## El Niño–Southern Oscillation (ENSO)

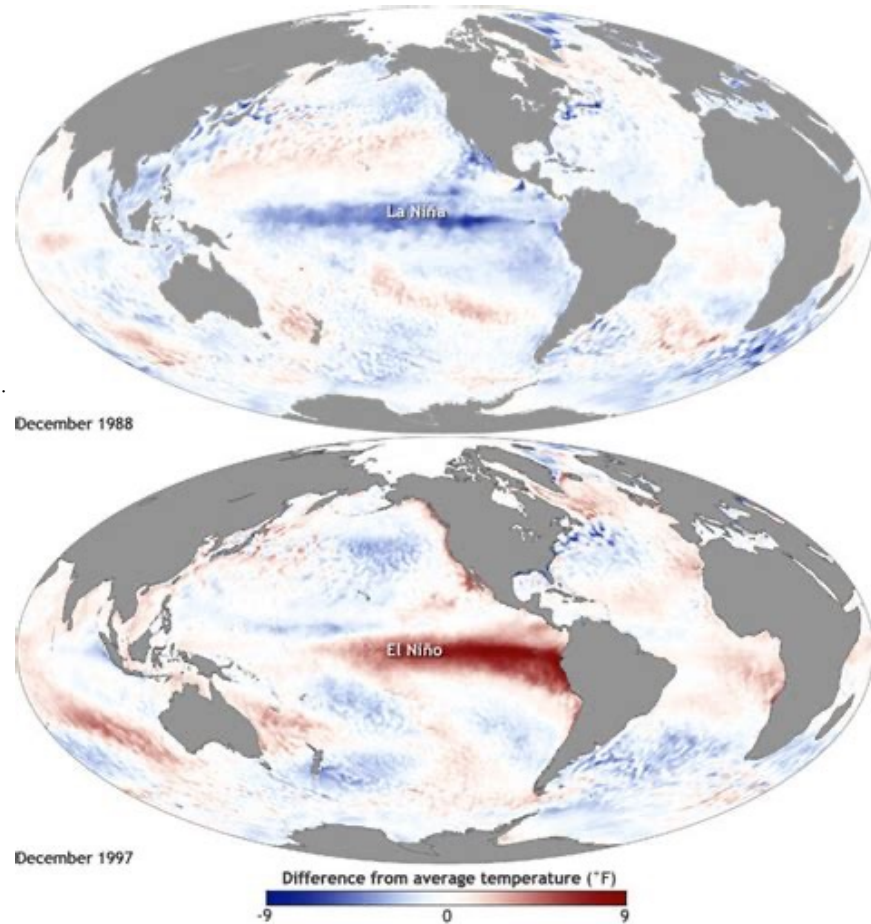
It's a natural climate pattern involving changes in sea surface temperatures and atmospheric pressure in the **tropical Pacific Ocean**. ENSO has two main phases:

### 1. El Niño –

- warmer-than-average ocean temperatures in the eastern Pacific.

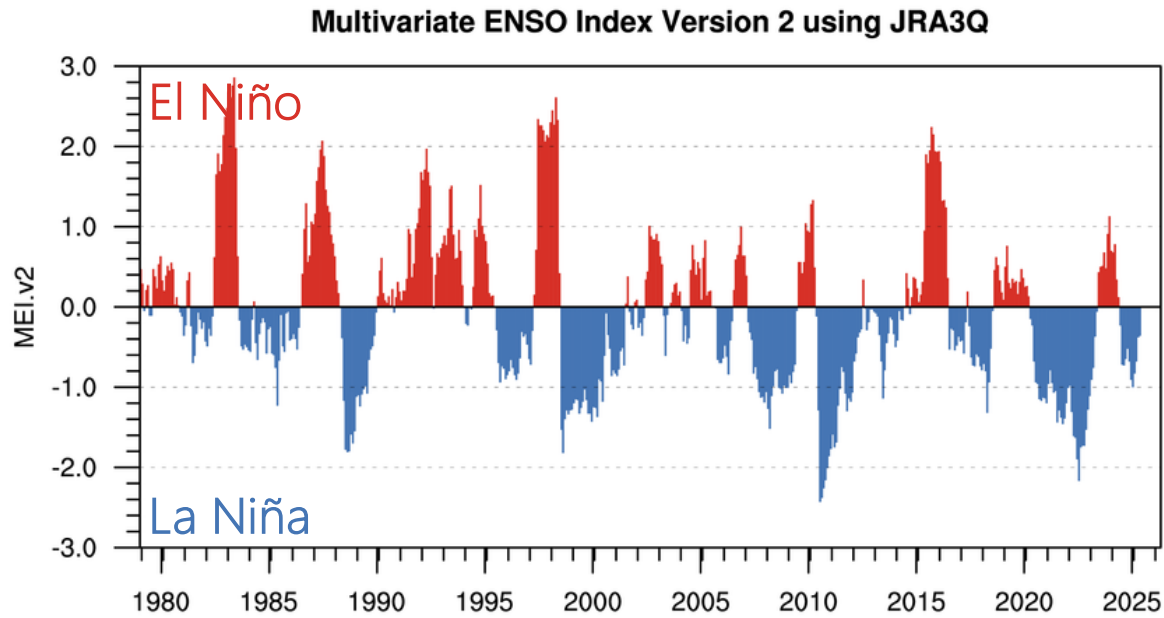
### 2. La Niña

- cooler-than-average ocean temperatures in the same region.
- *ENSO causes shifts in global circulation and regional cloudiness changes.*
- *Using the ENSO signal, can we better predict cloudiness (KT\*) anomalies and therefore yield at the month-ahead levels?*
- *We're translating a known, large-scale climate driver into a usable, site-specific forecast signal*



# What is the ENSO?

El Niño–Southern Oscillation (ENSO)



# But why does it matter?

## Month-Ahead Deviations Materially Impact:

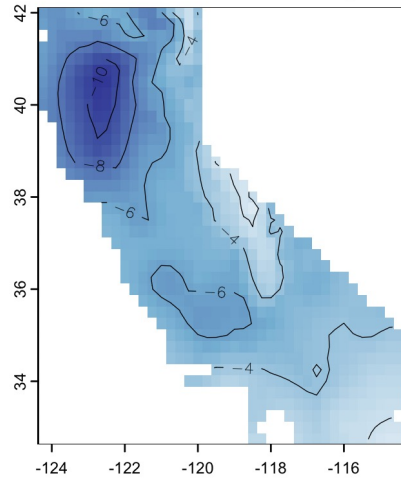
- PV Yield
- Operational Planning; maintenance scheduling
- Financing uncertainty: DSCR adherence
- Hedge positioning

...yet the MA window remains poorly addressed in operational PV modeling

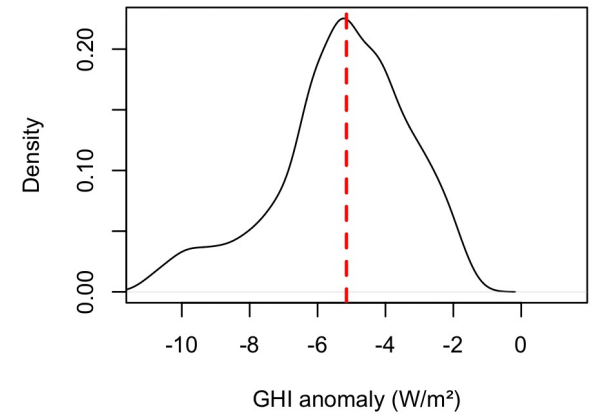
↳ So how can we improve this?

↳ Let's continue to focus on California, where the ENSO deviations in weather from L

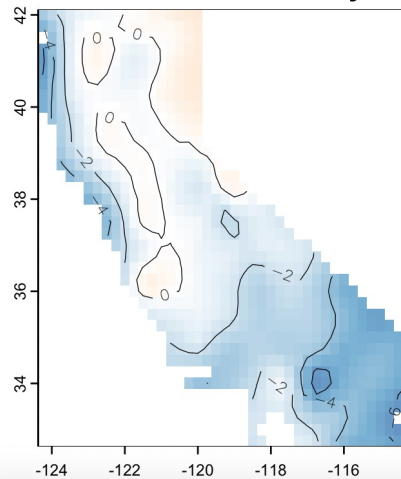
### El Niño GHI Anomaly



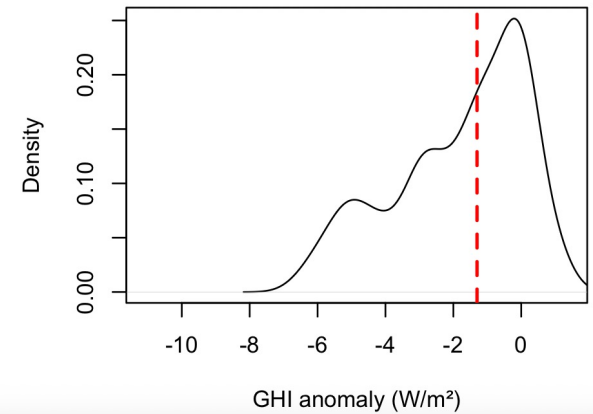
### El Niño GHI Anomaly Distribution



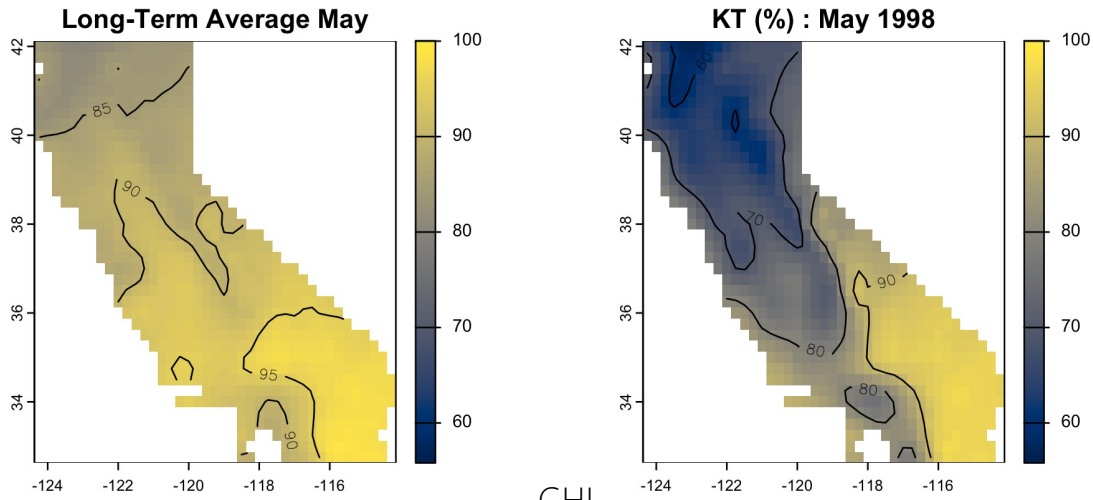
### La Niña GHI Anomaly



### La Niña GHI Anomaly Distribution

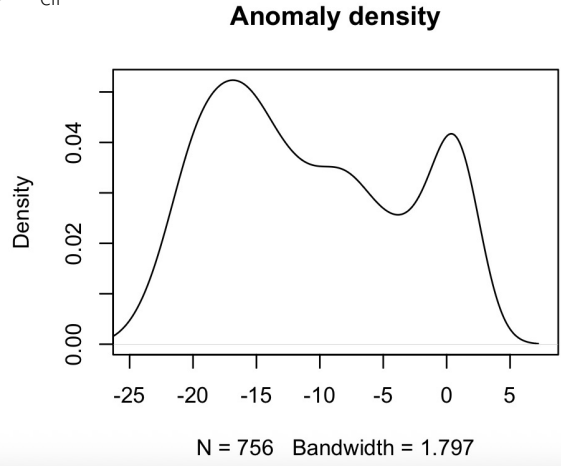
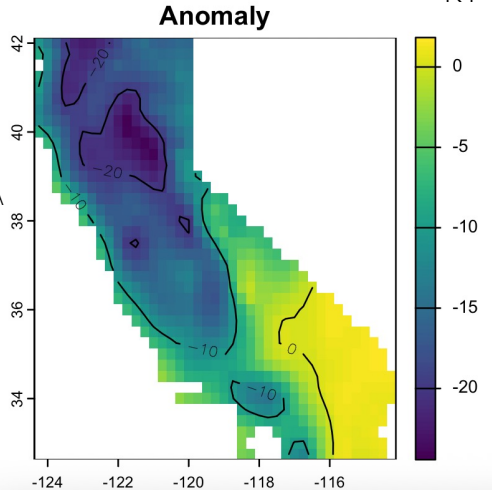


Let's build a model to predict  $KT^*$  anomaly using this signal



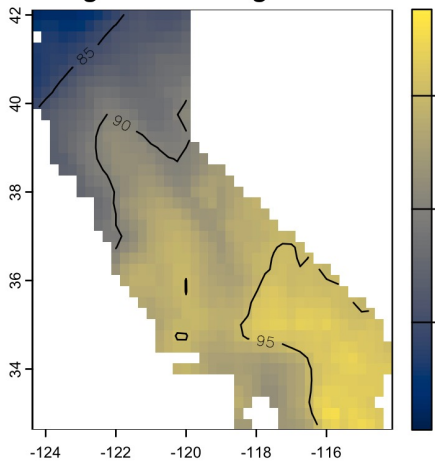
$$KT^* = \frac{GHI}{GHI_{clr}}$$

$$KT^*_{Anomaly} = KT^*_M - KT^*_{LTA}$$

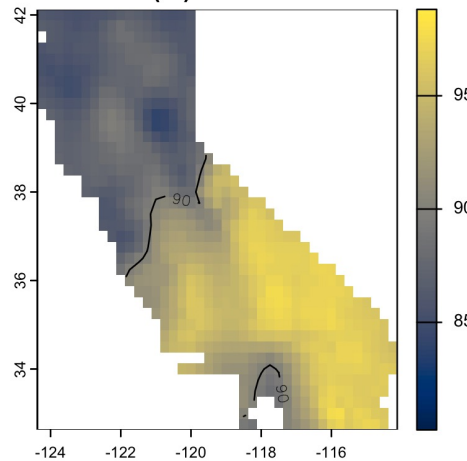


Here's another month: Sometimes the *Anomaly* is tighter

**Long-Term Average October**

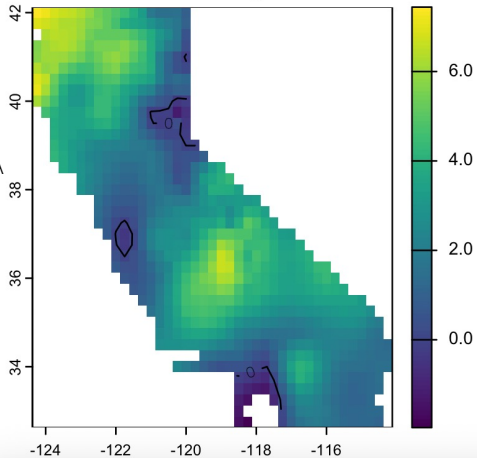


**KT (%) : Oct 1989**

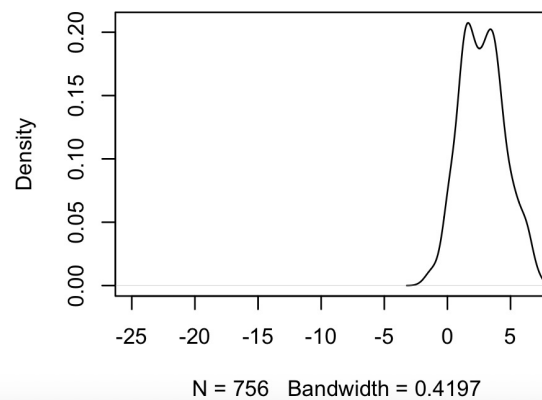


$$KT^* = \frac{GHI}{GHI_{clr}}$$

**Anomaly**

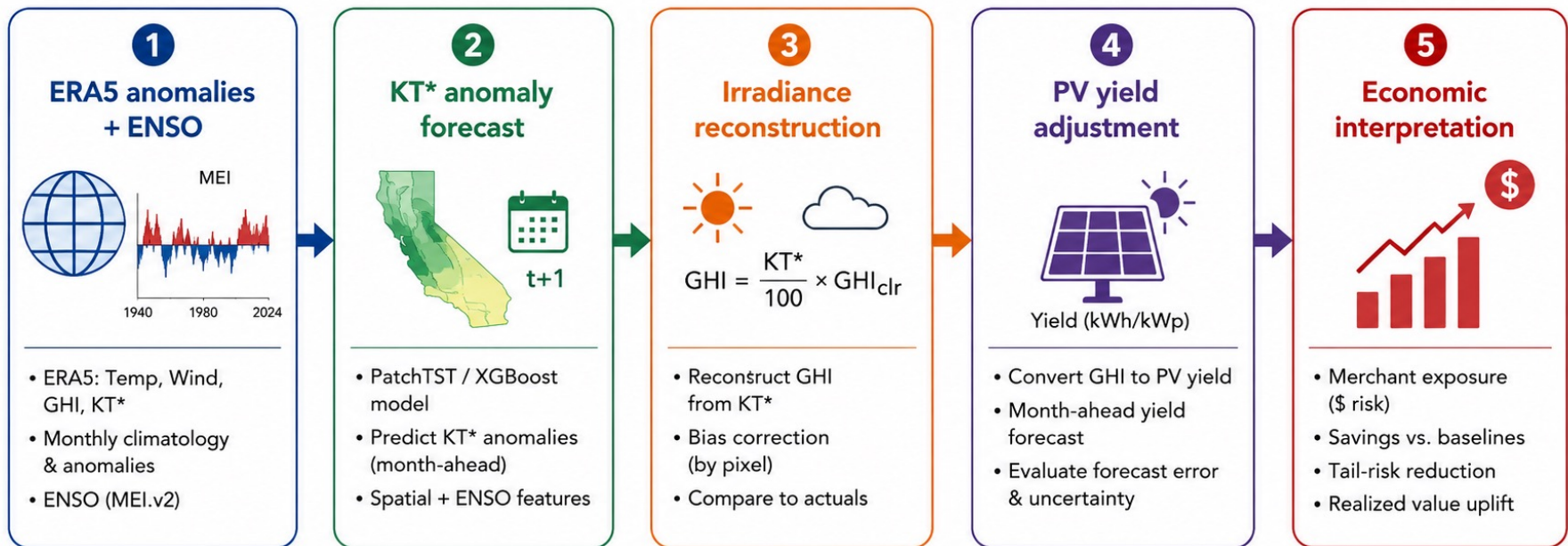


**Anomaly density**



$$KT^*_{Anomaly} = KT^*_M - KT^*_{LTA}$$

## How do we do it?

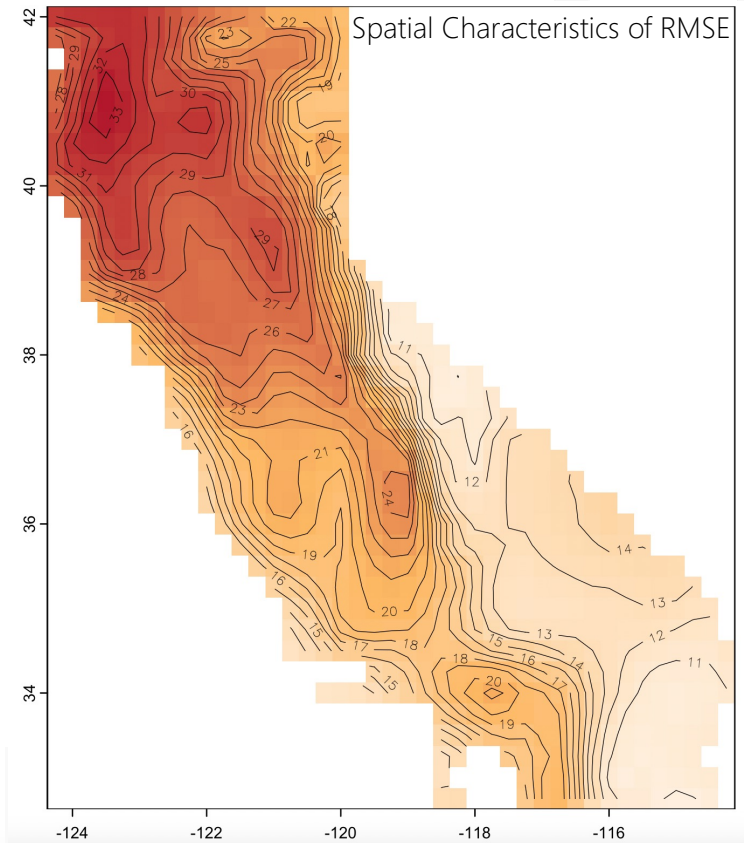
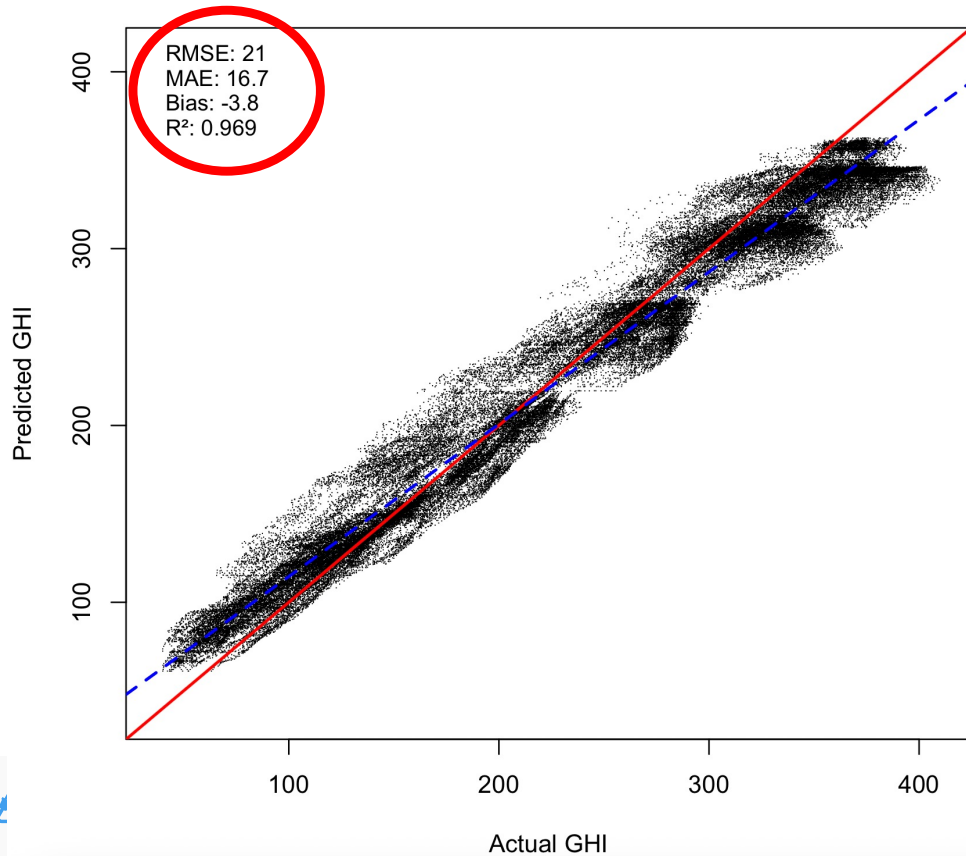


$$KT^*_{Prediction} = KT^*_{Anomaly,Pred} + KT^*_{LTA}$$

Let's establish some *baselines* first

↳ Typical Year (TMY)

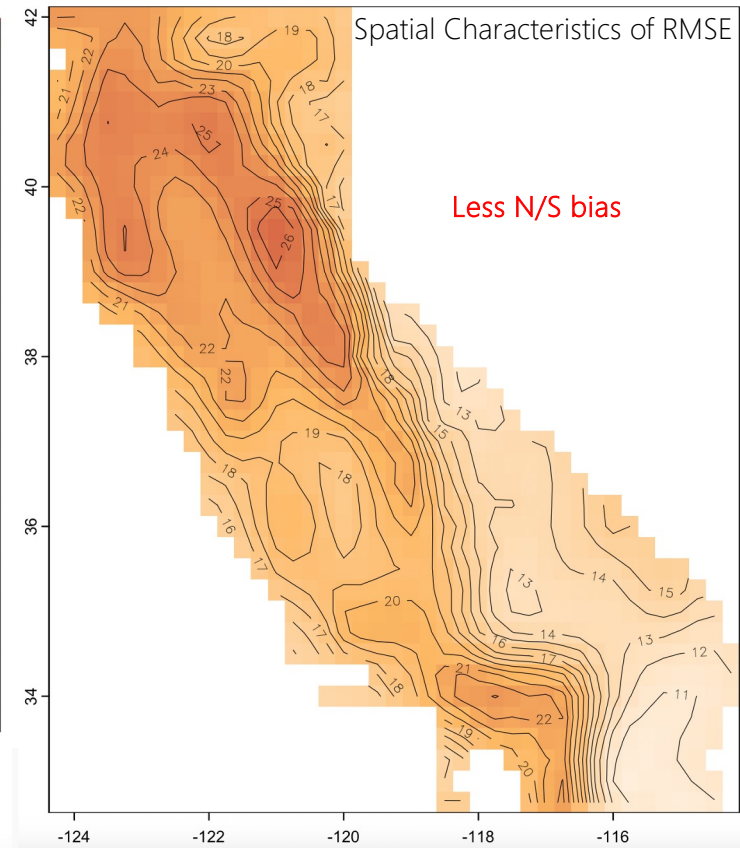
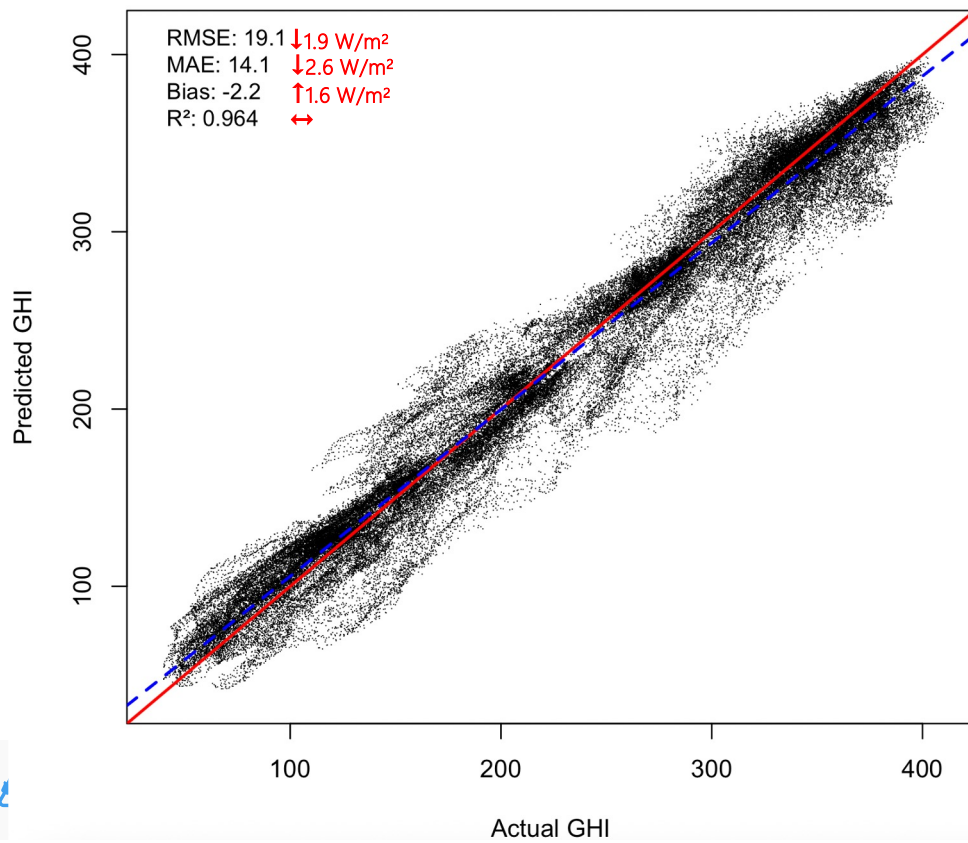
↳ Next month's GHI will be the typical GHI of next month



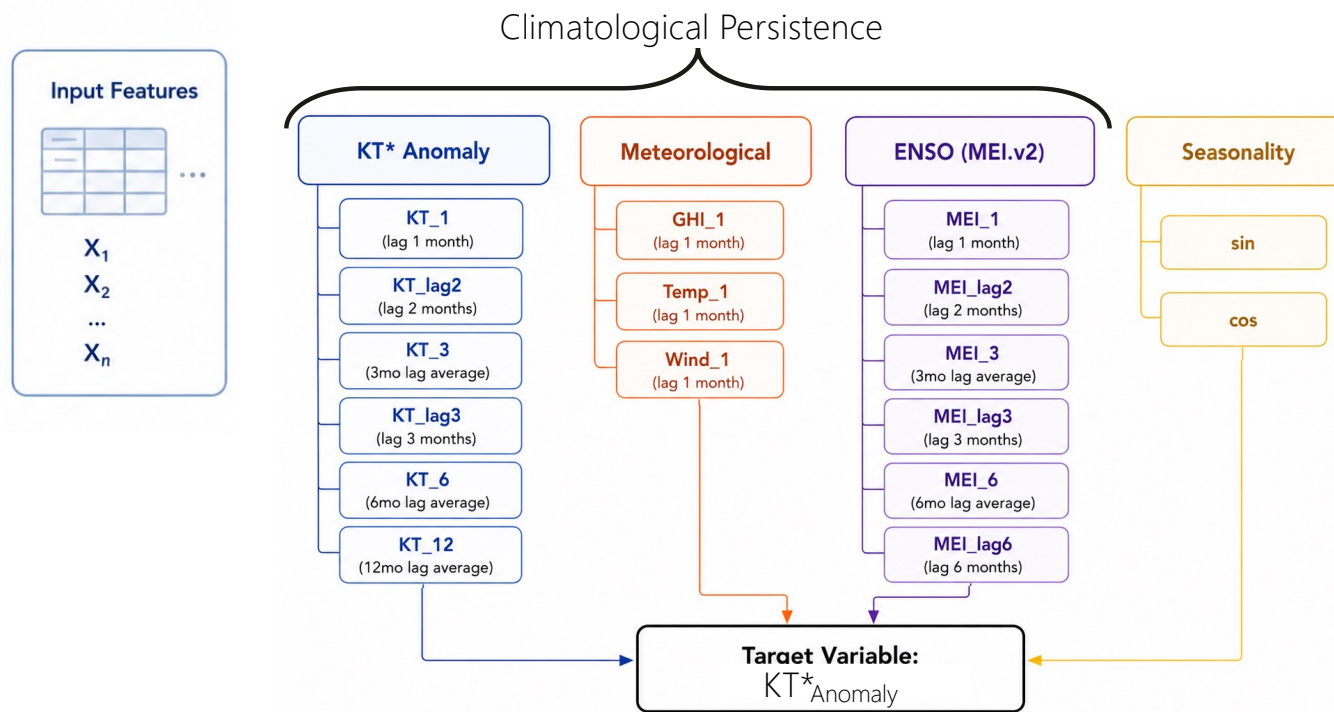
Let's establish some baselines first

↳ KT\* Anomaly Persistence

↳ Next month's anomaly in cloudiness will be the same as this month's

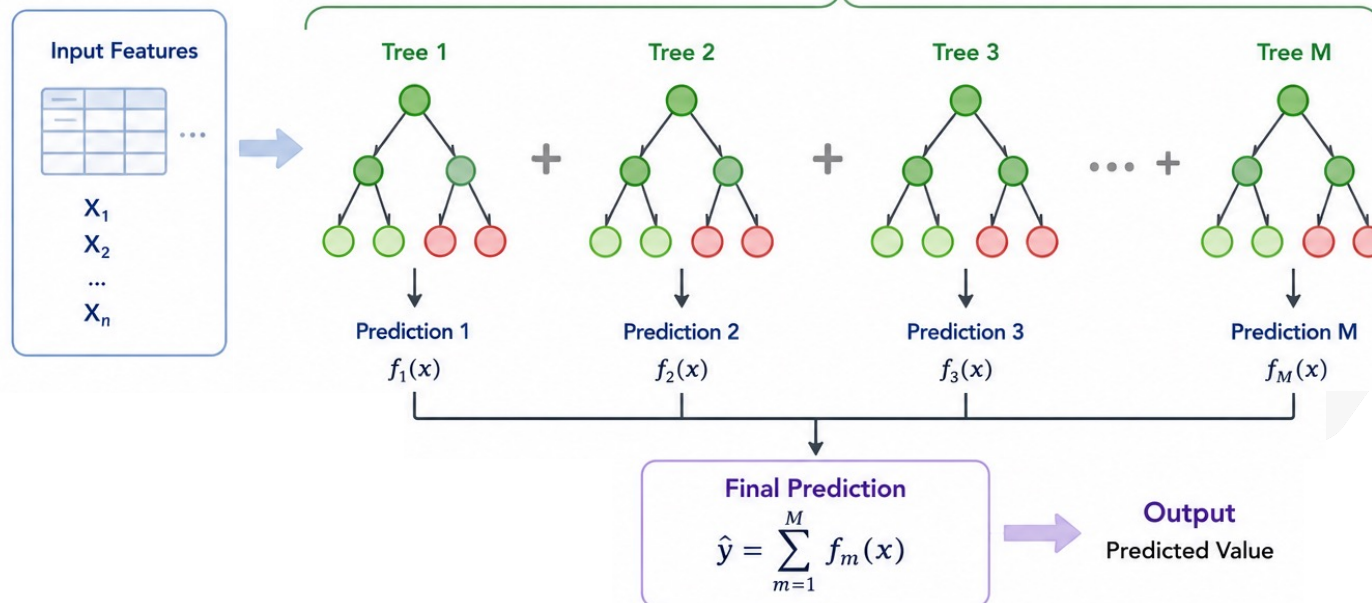


We train an XGB<sup>1</sup> model on readily available weather data



# How does XGB work?

## Ensemble of Decision Trees (Boosting)

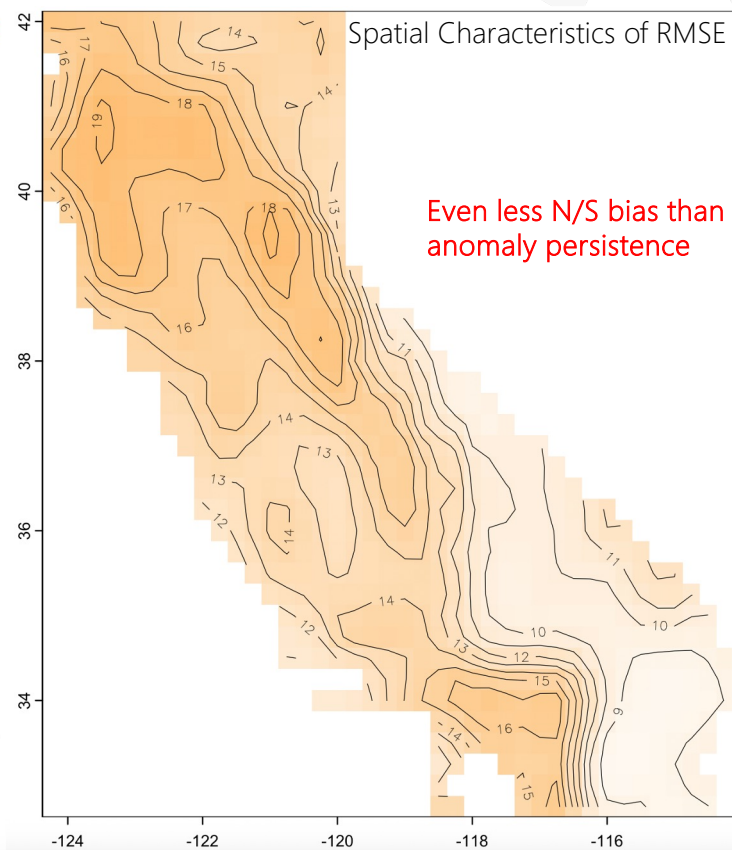
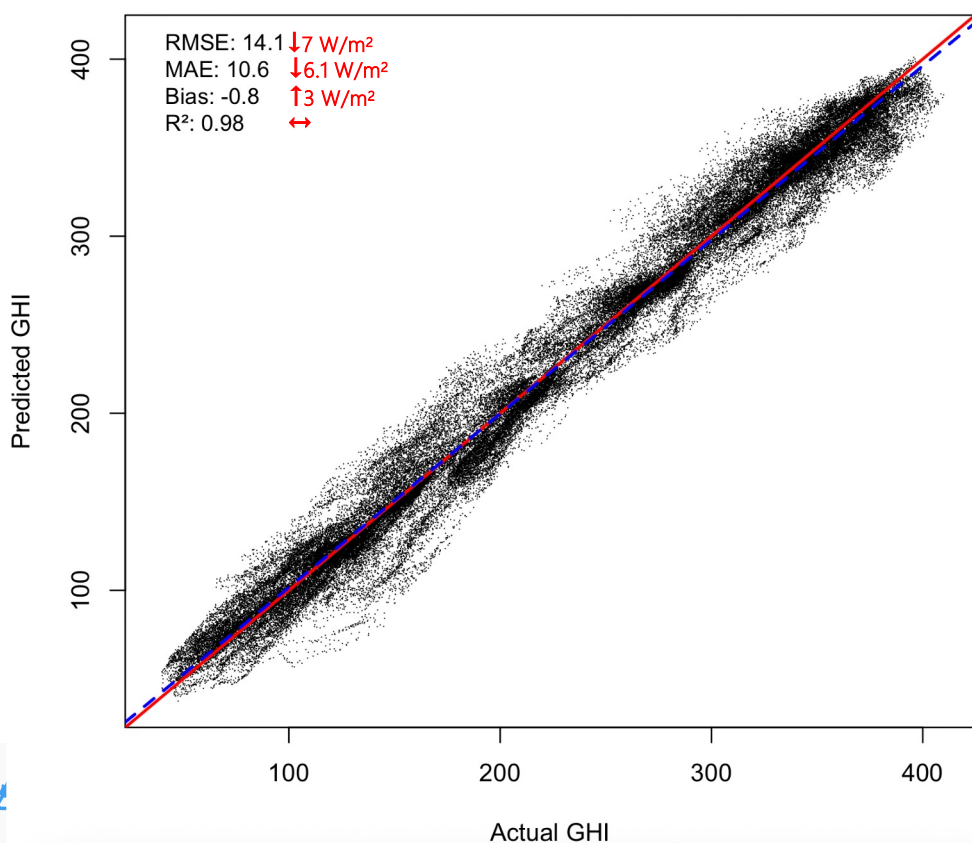


### How it works:

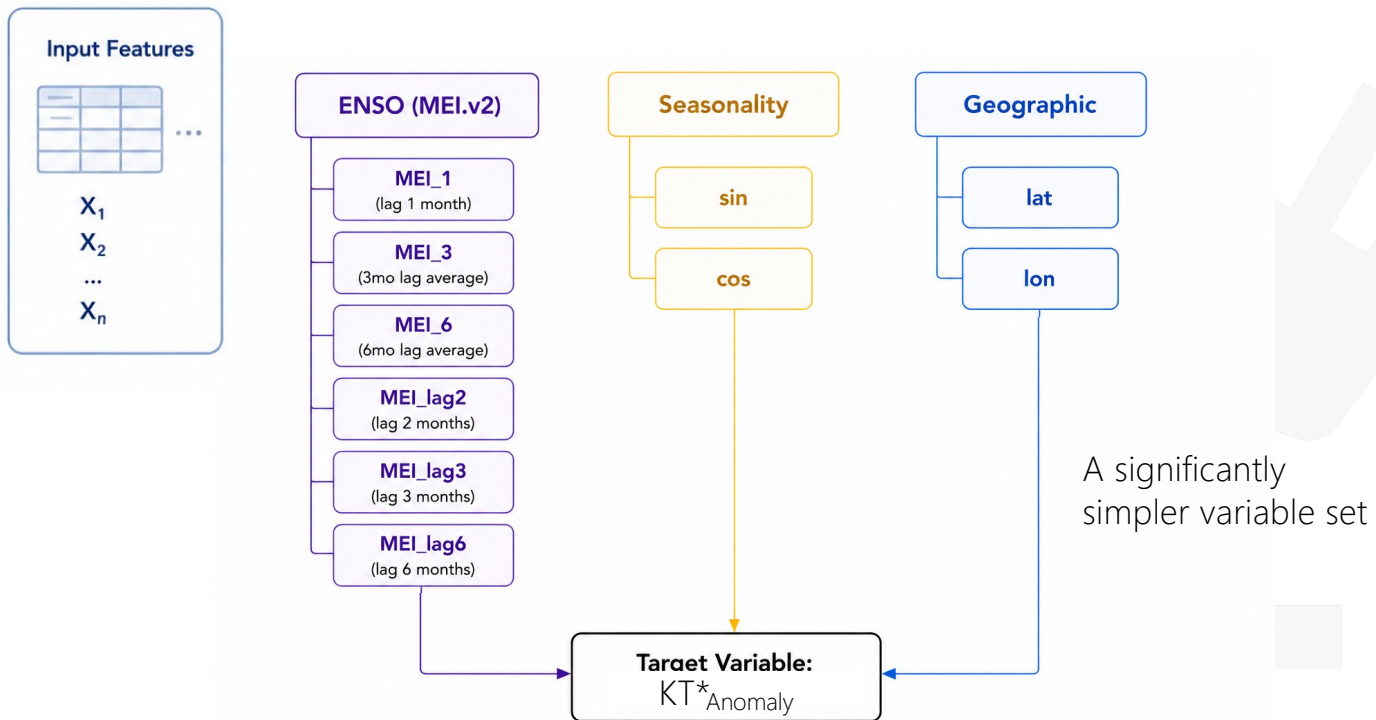
XGBoost builds an ensemble (forest) of decision trees sequentially. Each new tree is trained to correct the errors (residuals) of the previous ensemble by minimizing a loss function with gradient boosting. The final prediction is the sum of the predictions from all trees.

# How does XGB Perform?

↳ Seasonality + Multivariate Climatological Persistence as predictors

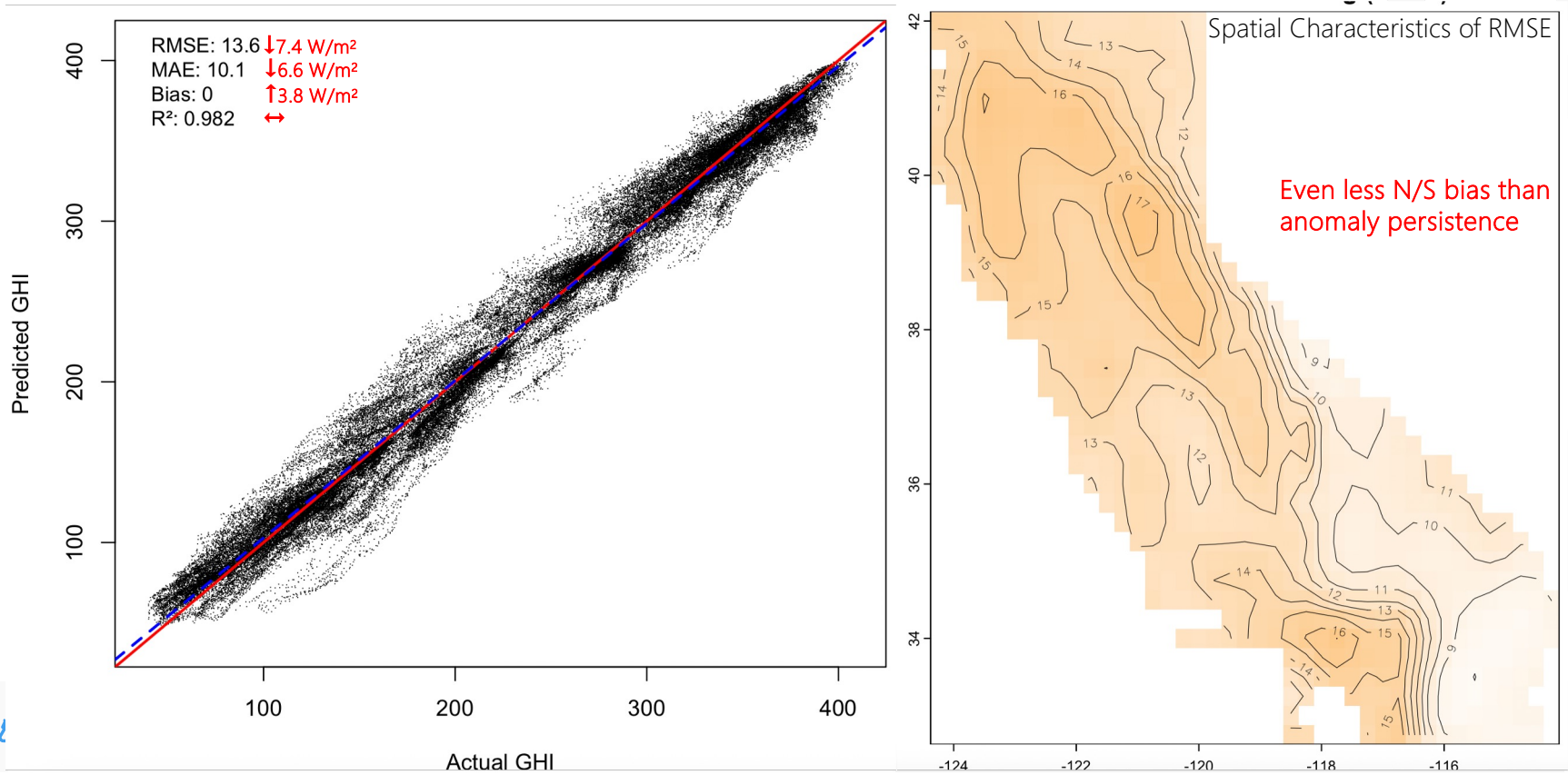


Let's try with *just* the MEI along with lat/long and seasonality



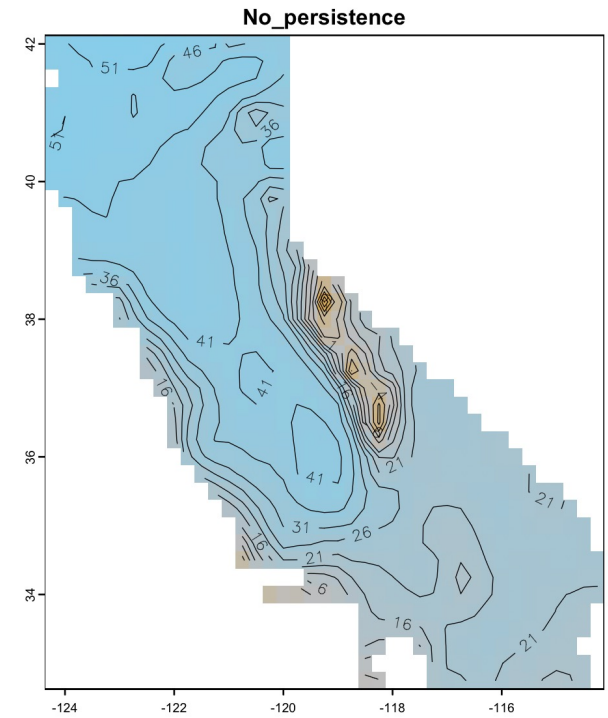
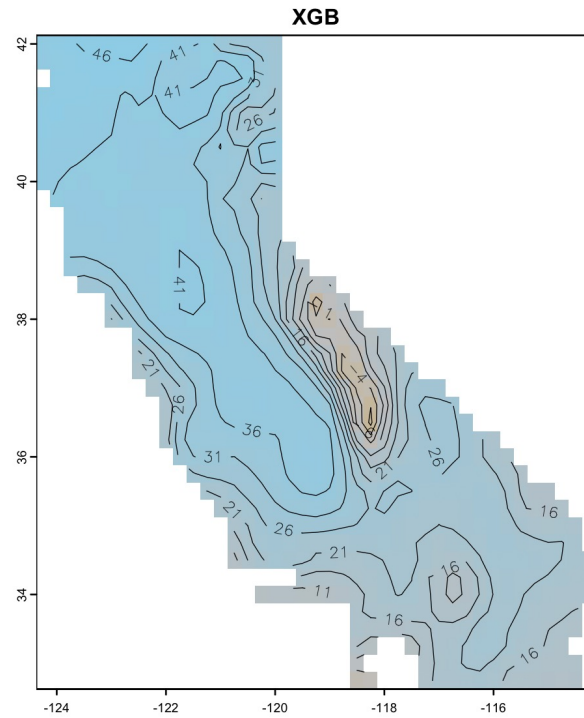
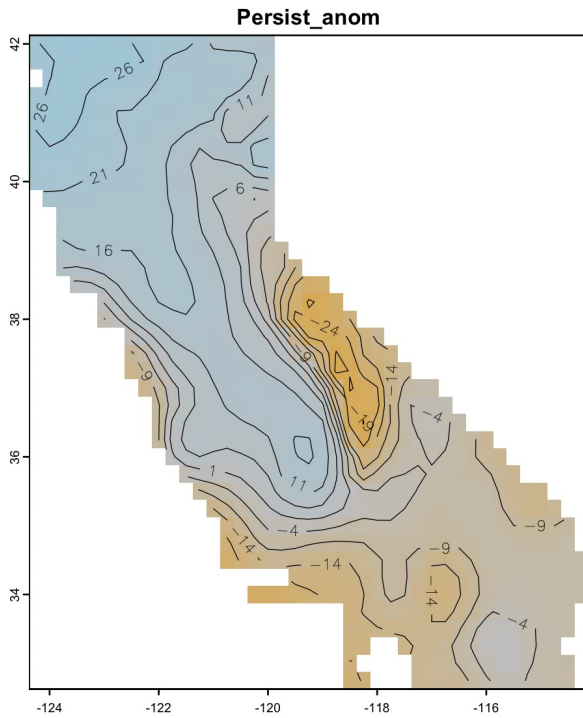
How does XGB Perform without a bunch of lagged weather?

↳ Our best model yet!



# Improvement of models versus TMY-as-prediction

None of the models do particularly well in the high sierras



Simpler Model with MEI as main predictor is best  
But degrades a bit relative to XGB in the Sierras  
Where there's lots of weather variation

We've significantly improved forecast skill at the MA horizon  
*Forecast skill only matters insofar as it reduces uncertainty in operational and financial decision-making.*  
*Let's quantify it:*



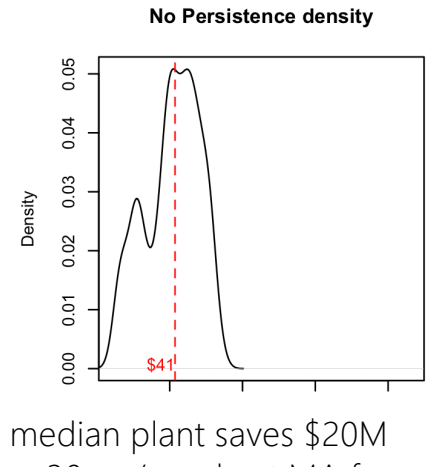
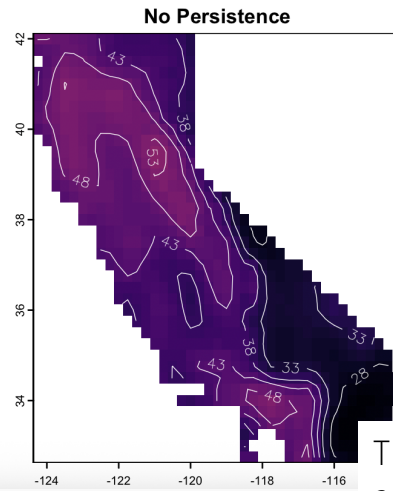
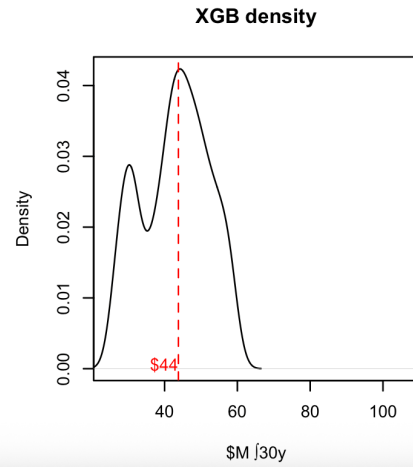
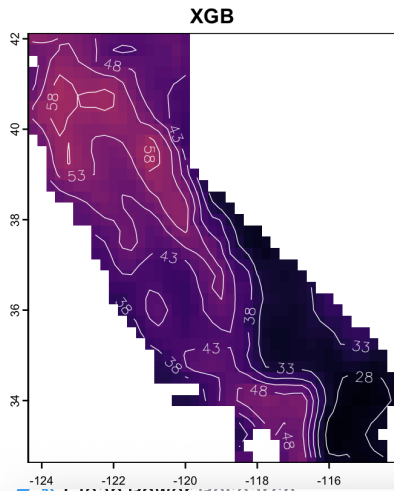
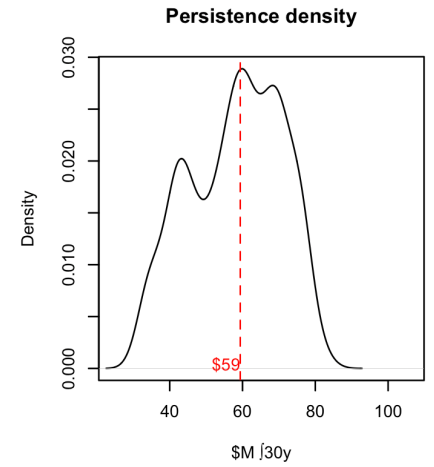
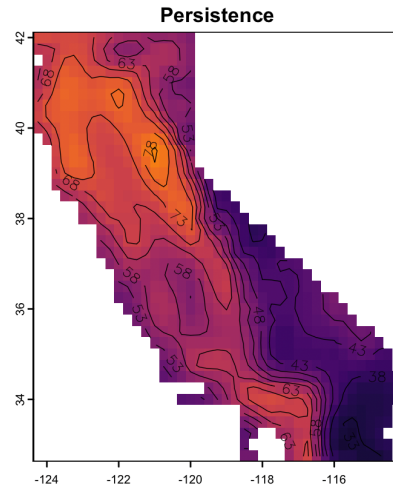
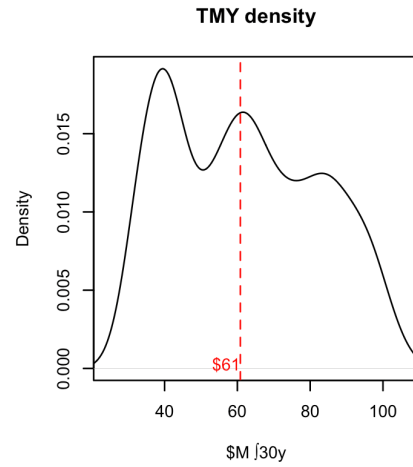
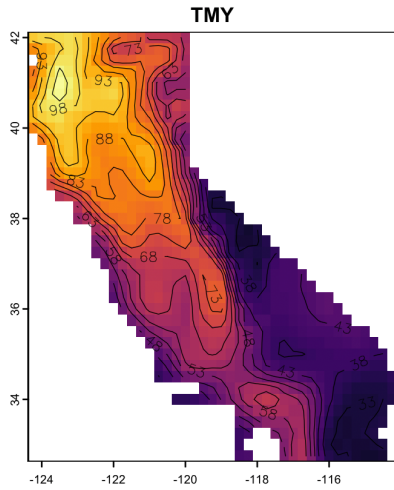
31% average improvement in MA GHI prediction relative to TMY



Yield  $\propto$  GHI at this timescale  $\therefore$  31% average improvement in MA yield prediction



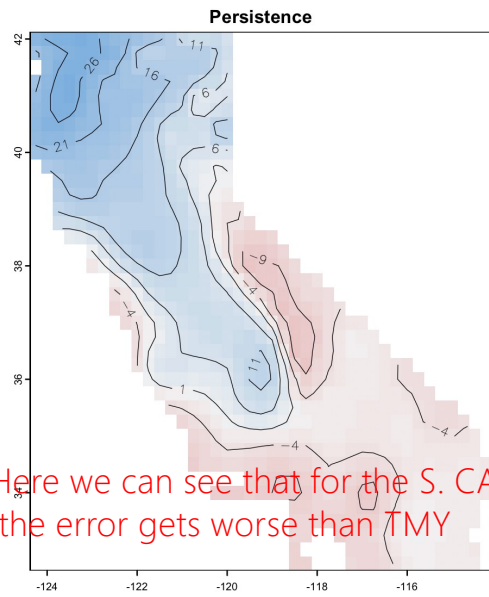
Merchant Exposure = Yield Forecast Error (MWh) X Merchant Power Price (\$/MWh)  
↳ Example for a 200 MW<sub>DC</sub> plant @ 75\$/MWh merchant price | 30y integration



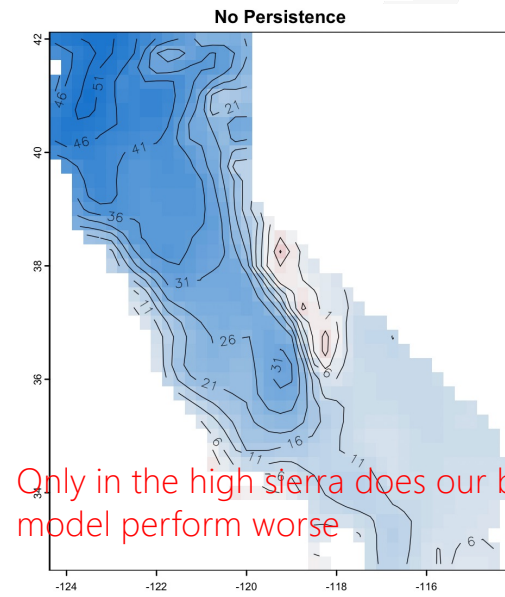
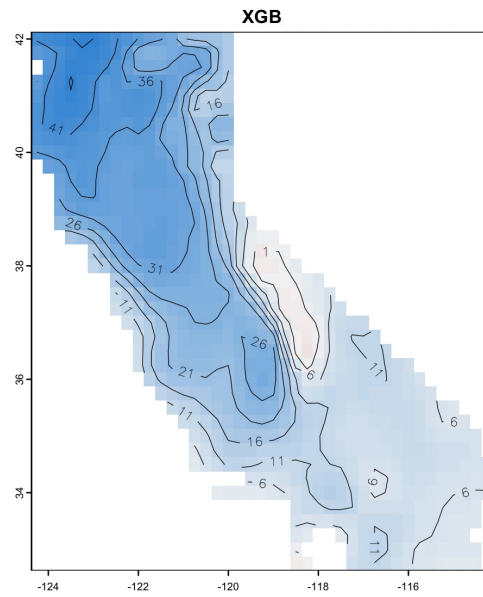
The median plant saves \$20M across 30y w/ our best MA forecast

## Relative Improvement vs. TMY

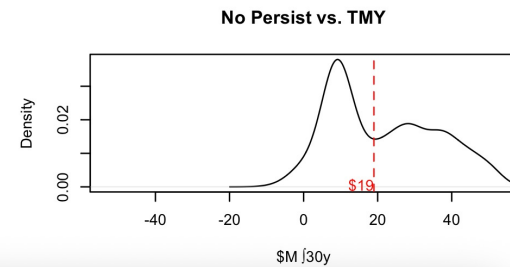
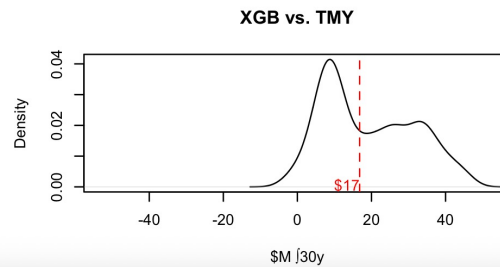
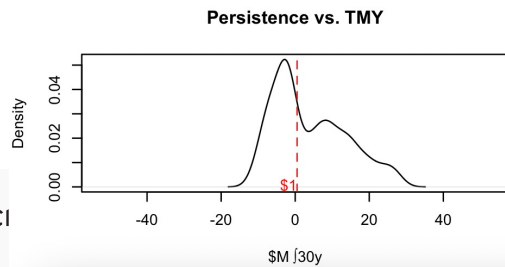
- ↳ Overprediction → Replacement energy purchases/ hedge shortfalls
- ↳ Underprediction → Replacement energy purchases/ hedge shortfalls



Here we can see that for the S. CA, the error gets worse than TMY



Only in the high sierra does our best model perform worse



CI

We've significantly improved forecast skill at the MA horizon

*Forecast skill only matters insofar as it reduces uncertainty in operational and financial decision-making.*



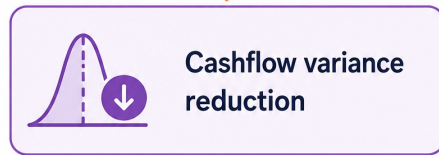
31% average improvement in MA GHI prediction relative to TMY



Yield  $\propto$  GHI at this timescale  $\therefore$  31% average improvement in MA yield prediction



Merchant Exposure = Yield Forecast Error (MWh) X Merchant Power Price (\$/MWh)

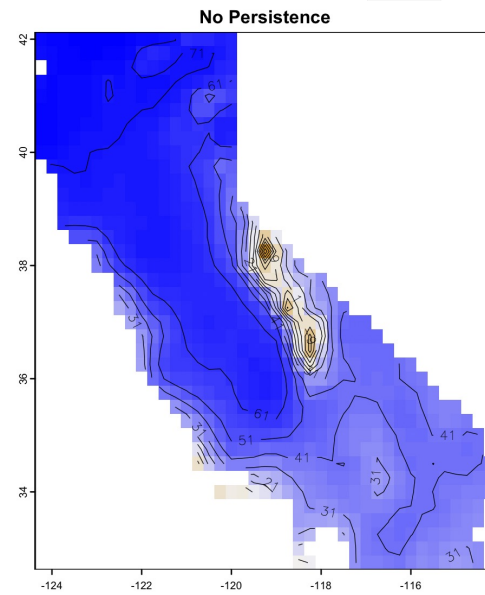
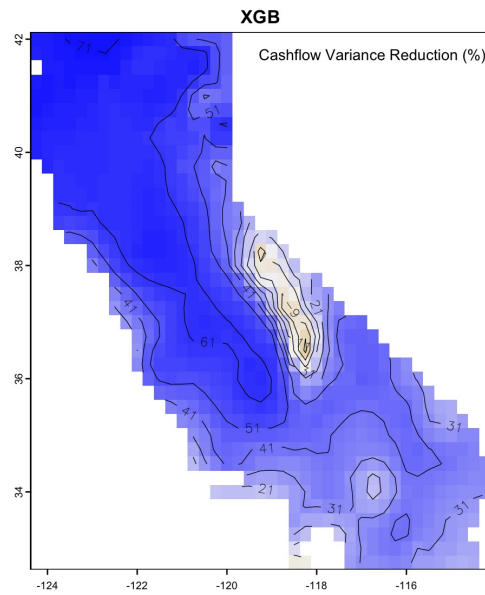
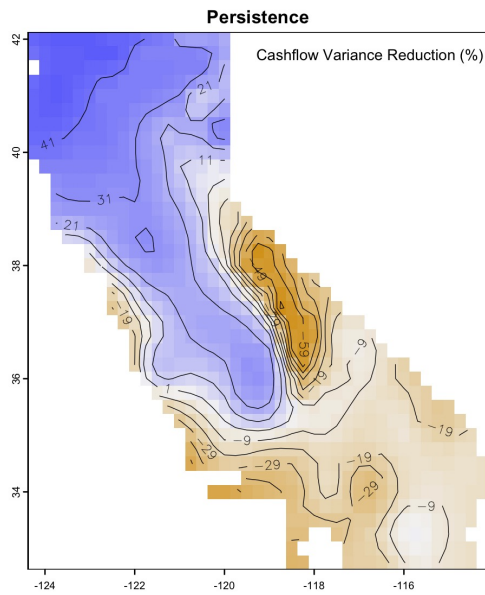


Variance Reduction =  $1 - (\text{RMSE}_{\text{model}}/\text{RMSE}_{\text{TMY}})^2$  : Variance  $\propto$  RMSE

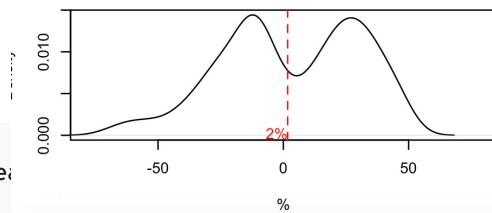
*\*fractional reduction in forecast error variance relative to climatology*

# CF Variance

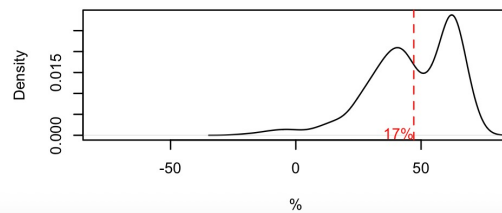
...resulting from forecast error remains about the same as TMY for persistence  
XGB reduces it by 17% on median No persistence, MEI-driven reduces it by 52% on median



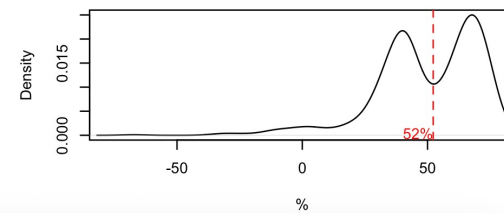
Persistence CF Var Red



XGB CF Var Red



No Persist CF Var Red



## Conclusions

- **Climate Signal as Predictor**— ENSO materially modulates irradiance across California, w/ strong spatial structure.
- **Forecast Skill** — Simple ML models using ENSO and seasonality significantly outperform LTA (TMY) as prediction.
- **Mountain Noise** — Forecast improvements  $\gg$  outside highly stochastic mountain microclimates (e.g., High Sierras).
- **Risk Reduction** — Reduced irradiance forecast  $\epsilon$  translates directly into lower merchant settlement exposure and reduced cashflow variance.
- **Economic Value** — Median modeled merchant exposure reductions across CA approach ~\$20M over 30 years for a representative 200 MW plant.
- **Commercial Relevance** — Results suggest actionable economic value for month-ahead probabilistic solar forecasting in merchant and hybridized PV portfolios.

# Reducing Month-Ahead PV Yield Uncertainty

with ENSO-Informed Machine Learning

Marc Perez

May 14<sup>th</sup>, 2025

PVPMC '26, ABQ, NM

