Novel Forecast Metrics

2025-06-24

Will Hobbs, SCS R&D

Why are forecast error metrics important?

Forecasts are:

- Designed — By vendors
- Tuned
- Selected — By us

based (largely) on error metrics.

So, we could be “optimizing” incorrectly without the right metrics.

Forecast Error Metrics

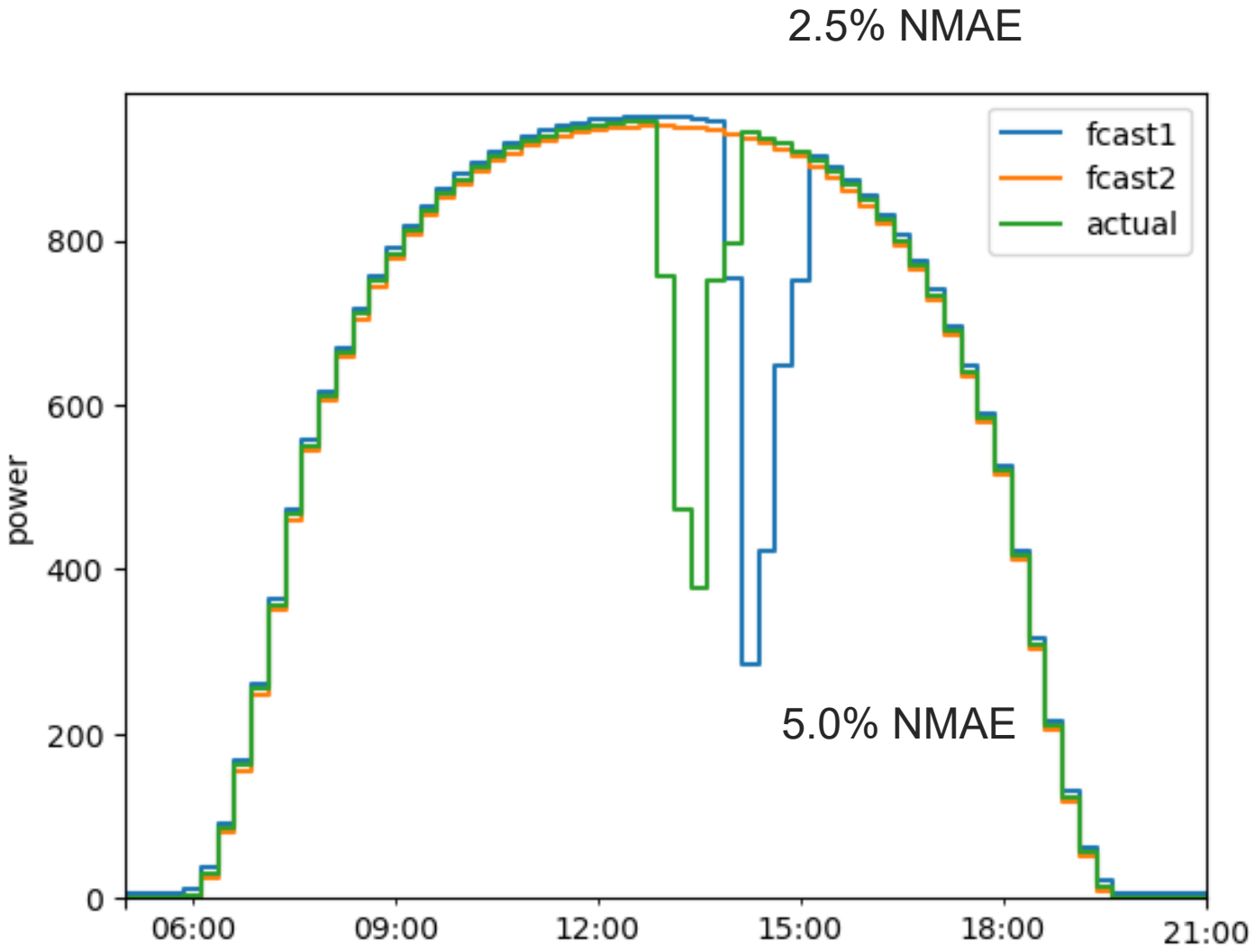
- Our best single(-ish) metric:
  - Mean Absolute Percentage Error (MAPE), Normalized Mean Absolute Error (NMAE, a.k.a Normalized MAPE)
- 2nd best:
  - RMSE?
- Note: MBE is definitely important, but should be so close to zero for a well-tuned forecast that you hope to not look at it much…

- Often run *hourly* (sometimes 15min or similar)

- Penalize large errors, so better forecasts are likely to be “smooth”

Quick illustration of NMAE “weakness”

- (more on this later)



- Fcast 1 gets ramp almost perfect, but timing is off
- Result is 2x higher NMAE
- Timing isn’t always critical in real operations

Goals and thoughts for today:

- **Novel forecast error metrics**
  - Solar, load, net load
  - Ramp-based, minimum solar generation, minimum reserves
  - Probabilistic, sliding windows
- **Novel forecast *products***
  - Example: Maybe we can't optimize for both NMAE and some ramp-based metric, so we could need a separate ramp rate forecast

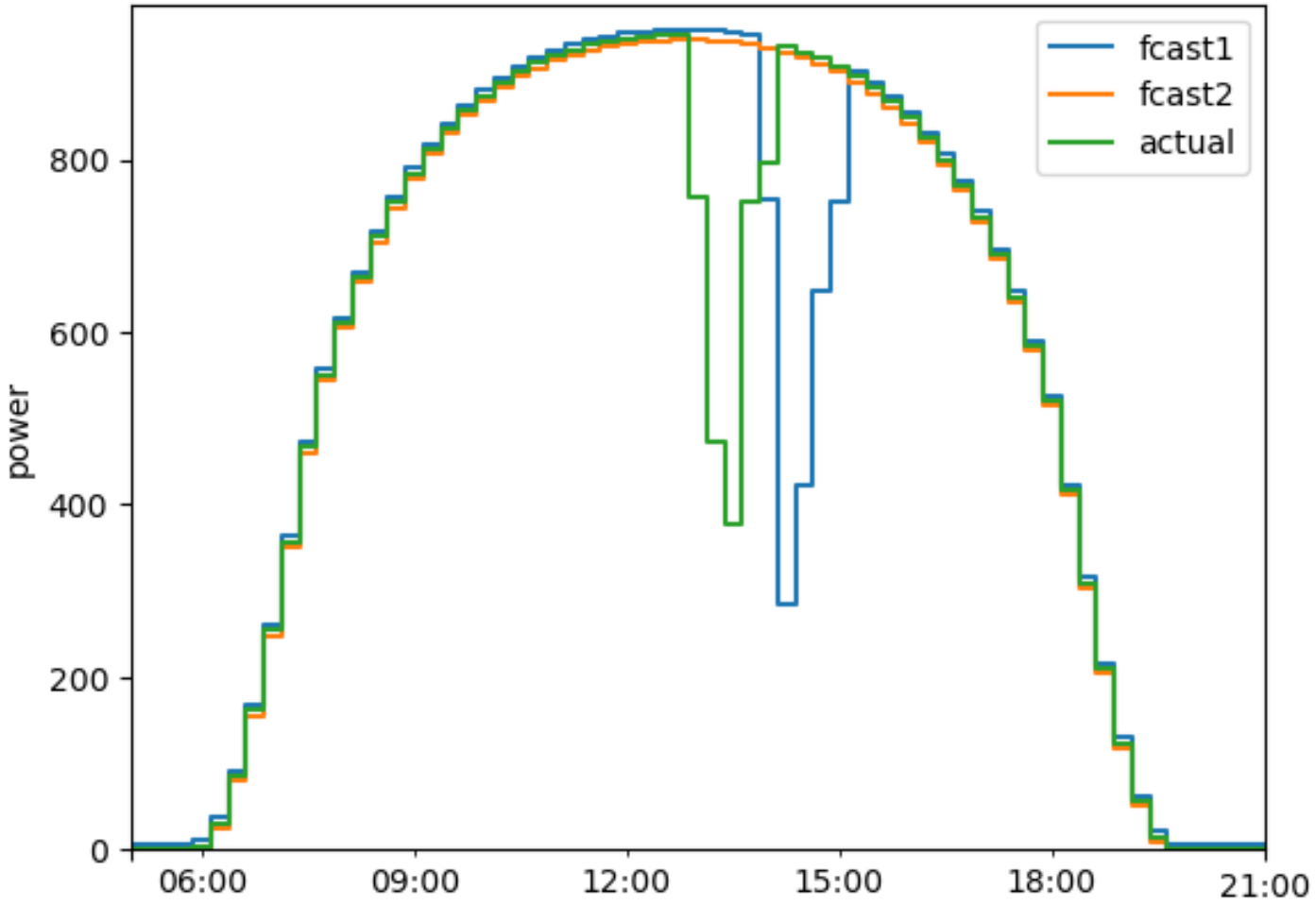
- Present to ESIG, get conversation going, settle on good standard metrics, and get better forecasts

Ramp example

Southern Company

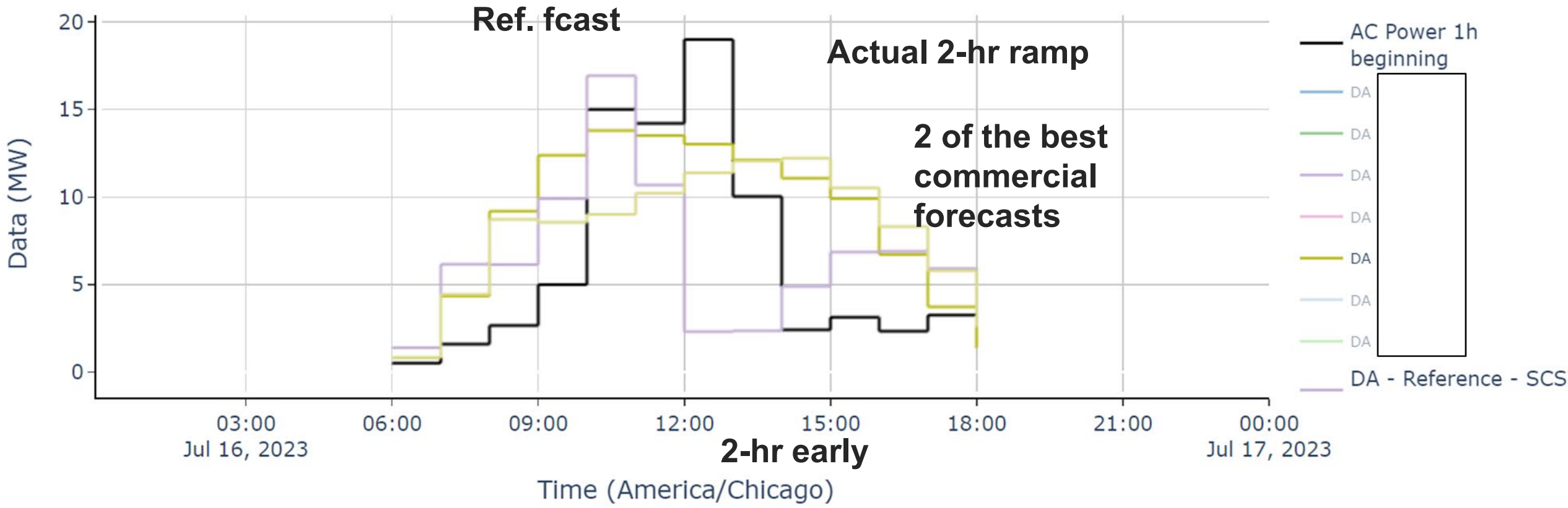
Forecast Error Metrics

- Normalized Mean Absolute Error (NMAE) penalizes large errors, so better forecasts are “smooth”



- Fcast 1 gets ramp almost perfect, but timing is off
- Result is 2x higher NMAE
- Timing isn’t always critical in real operations

Real Example from Trial, Single Site – DA

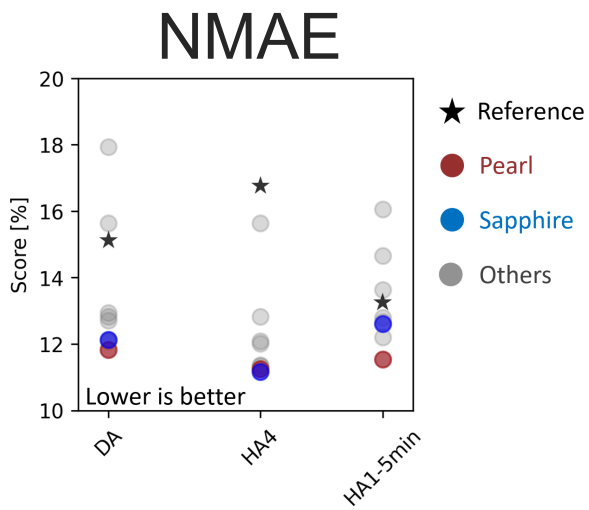


Tentative ramp metric

- General ideas are:
  - downward ramps in solar are worse than upward ramps
  - exact timing isn't critical
- On a 5-hr rolling interval (current hour, 2 before, 2 after), calculate the largest 2-hr downward ramp
  - For each forecast, calculate NMAE of 5-hr rolling largest 2-hr down-ramp (excluding ramps smaller than 1% of nameplate)

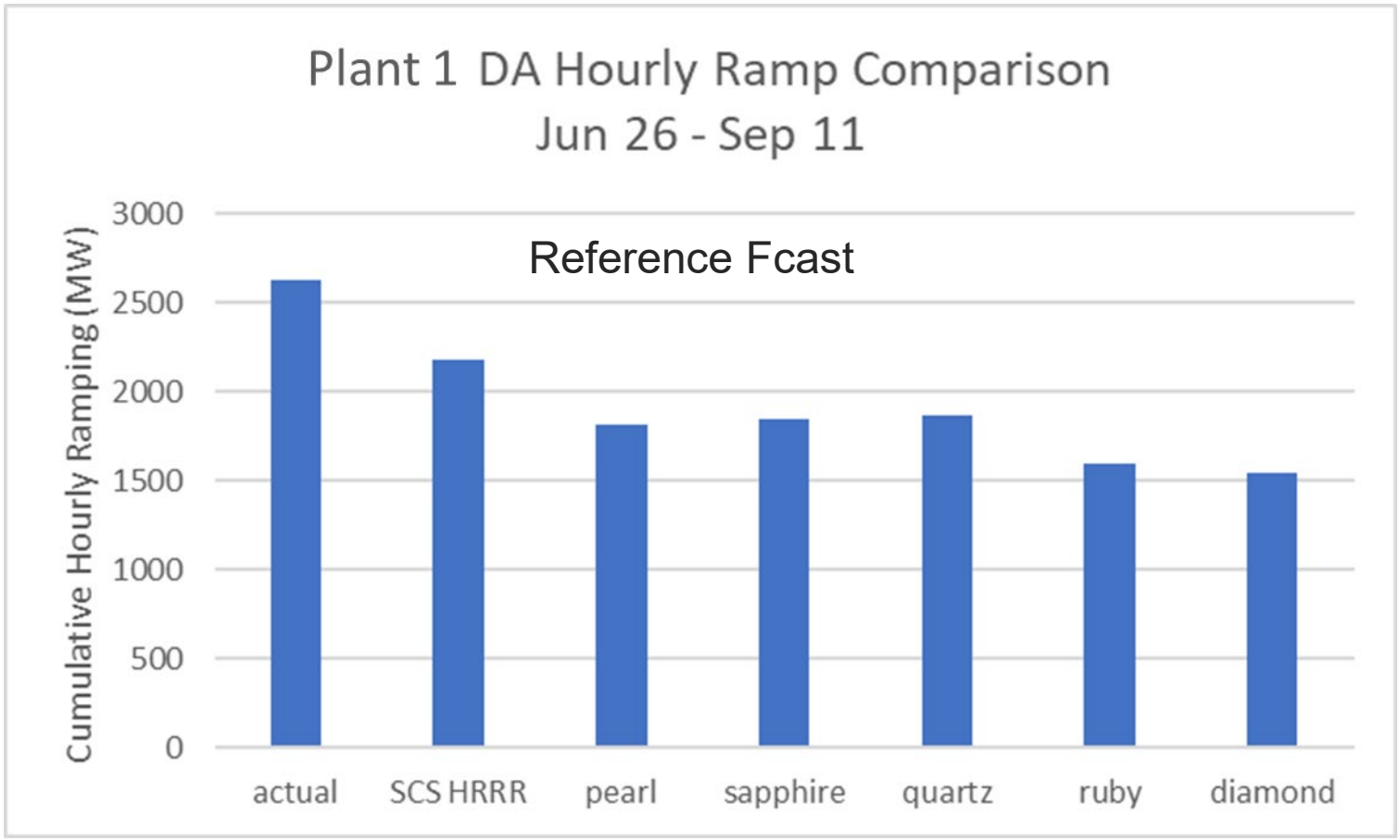
(lower is better)

Forecaster	DA NMAE, 5-hr max down ramp
Reference	14.6
Pearl	15.3
Sapphire	15.9



Ramp Statistics (1-hr)

- Another way of looking at ramps/variability



Minimum generation

Southern Company

Solar minimum gen

- Operators have been concerned when solar “drops”
- *How low will it go?*

- Forecast product (and metric) idea:
  - Minimum 5-min avg* total solar gen within a 1-, 2-, or 3-hour window
  - 95% confidence or similar
  - Need 2-hr notice
  - Could also use maximum in the morning
  - Window width increases with lead time

* They really care about *instantaneous*, not 5-min

More ramps

Intra-hour ramps

- 97^{th} percentile of 10-min ramps

Probabilistic hourly energy

Probabilistic metrics

- Prediction Interval Coverage Probability (PICP)
  - E.g., does the 95% chance forecast happen 95% of the time?

- Prediction Interval Normalized Average Width (PINAW)
  - E.g., now “sharp” or “narrow” is the forecast? Saying “*it will be somewhere between 0 and 100%*” isn’t helpful

- Others: CRPS, Brier Skill Score (compares to ref fcast), more…

Questions, Feedback?

whobbs@southernco.com