NYSolarCast: Utility and Distributed Solar Power Forecasting for New York State

7th Conference on Energy & Meteorology (ICEM 2023) ◆ Padova, Italy ◆ 29 Jun 2023

Jared A. Lee[1], Susan Dettling[1], Julia Pearson[1], Thomas Brummet[1], and David P. Larson[2]

[1] National Center for Atmospheric Research (NCAR)
[2] Electric Power Research Institute (EPRI)

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

Motivation

- New York State (NYS) Clean Energy Standard
  - 70% renewable energy (RE) generation by 2030
  - 100% RE generation by 2040
- Much more solar energy must be deployed across NYS
  - Highly variable, weather-driven resource
  - Challenge for grid balancing & stability
- Accurate forecasting is increasingly critical for electric utilities and independent system operators like NYISO
  - Nowcast/intra-day forecasts
  - Day-ahead forecasts
  - Utility-scale photovoltaic (UPV) plants
  - Distributed PV (DPV) sites
- Multi-phase project to build a solar power forecasting system for NYS
  - Funded by NYSERDA & NYPA
  - Research team: EPRI, BNL, NCAR, U of Albany

Image: Department of Energy, PV Magazine

NEW YORK STATE OF OPPORTUNITY. | NY Power Authority

NEW YORK STATE OF OPPORTUNITY. | NYSERDA

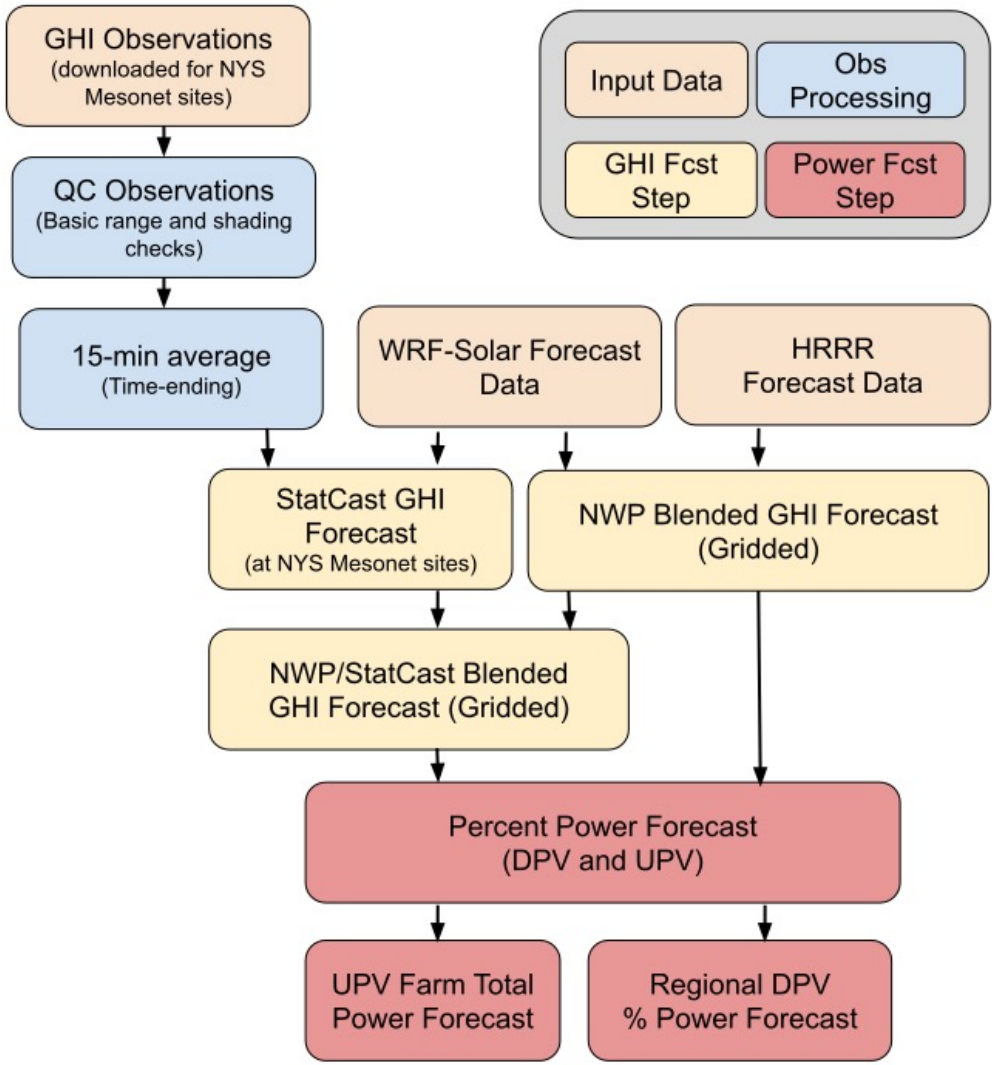
EPRI ELECTRIC POWER RESEARCH INSTITUTE

Brookhaven National Laboratory

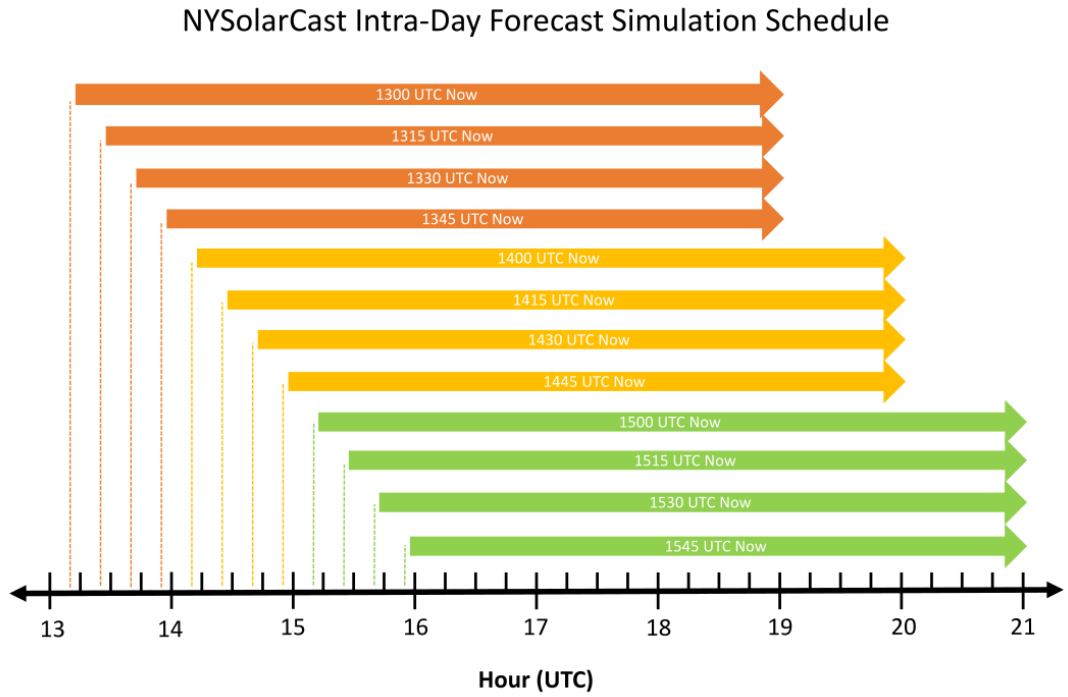
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

UNIVERSITY AT ALBANY State University of New York

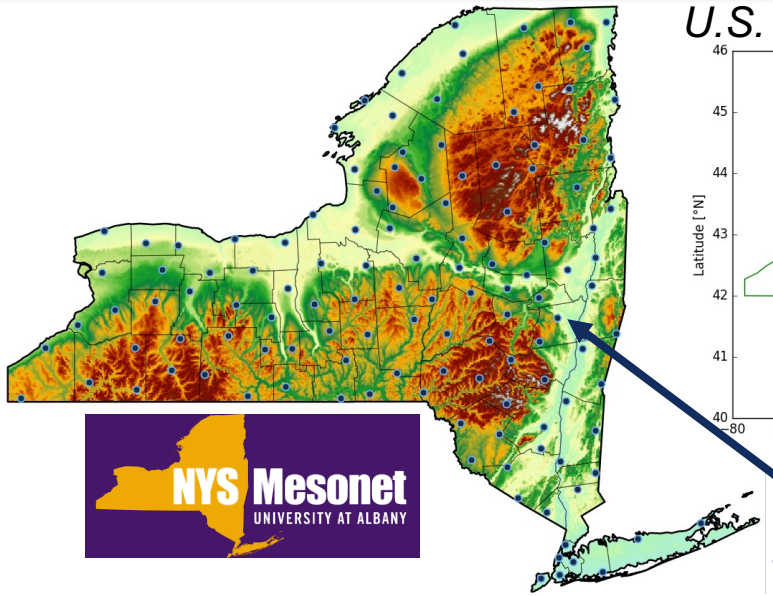
NYSolarCast System Design & Schedule



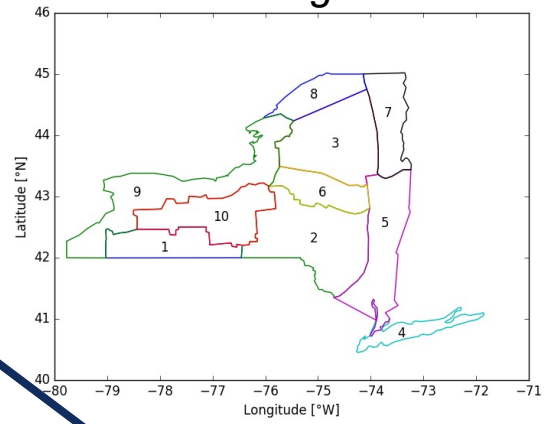
- Intra-day
  - Forecasts issued every 15 min 1115–1900 UTC
  - 15-min frequency for GHI and UPV forecasts
  - 1-h averages for DPV forecasts
- Day-ahead
  - Forecasts issued once daily at 0600 UTC



New York State Mesonet (NYSM)



U.S. Climate Regions in NYS



Voorheesville NYSM site:

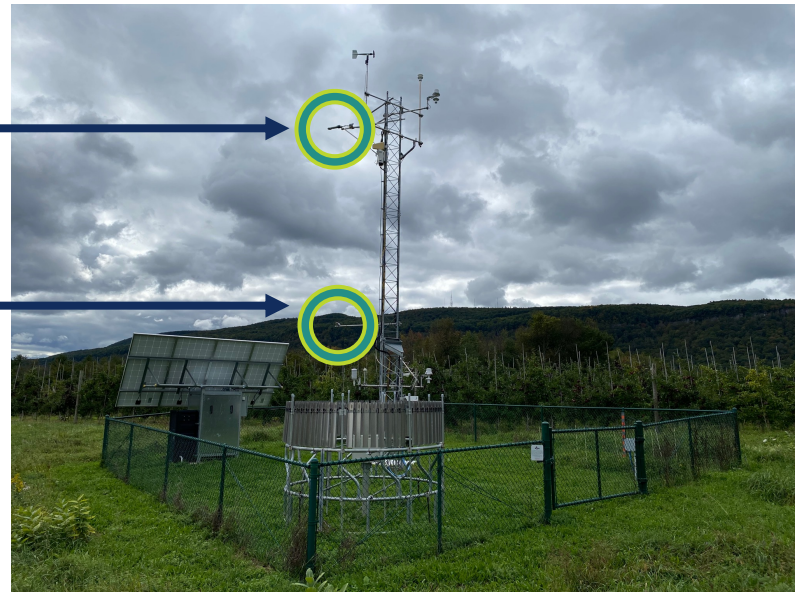
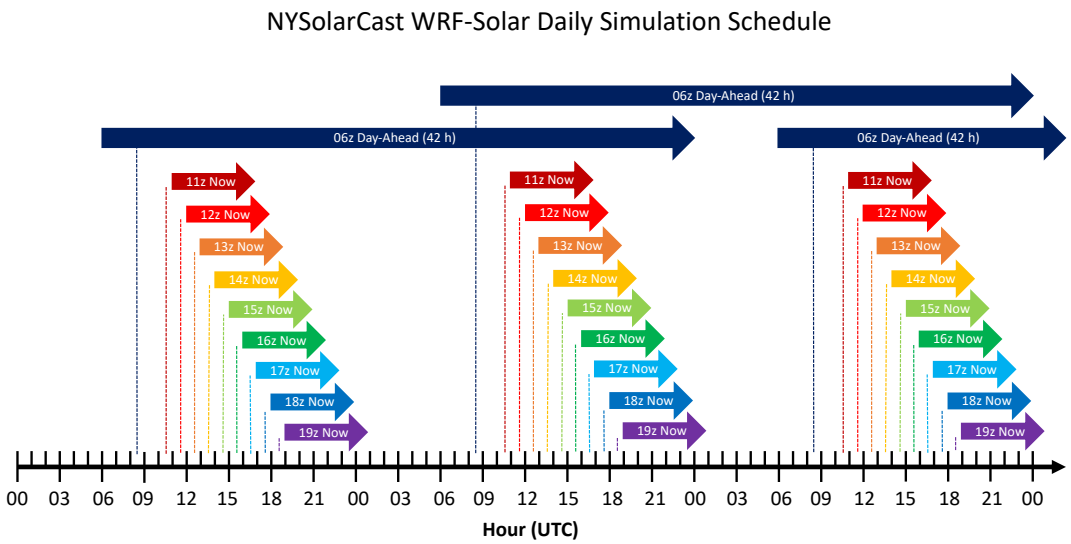


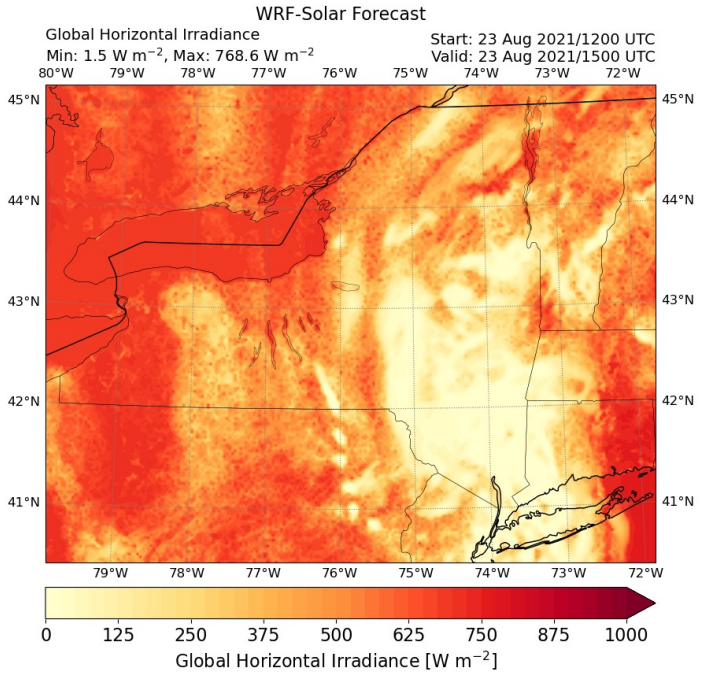
Photo by Jared Lee

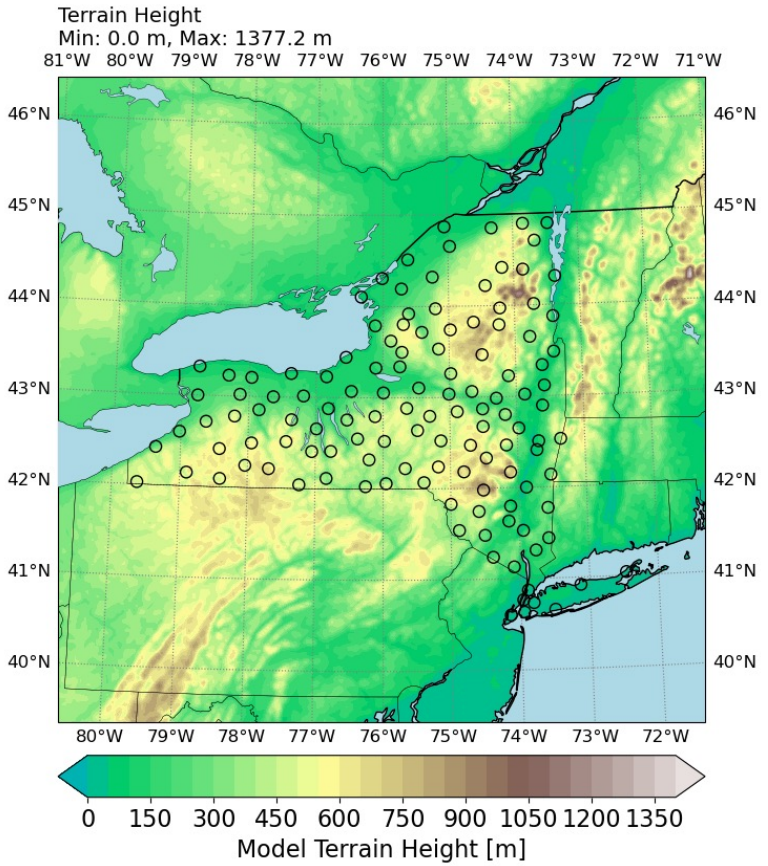
- Historical and real-time data from all 126 Standard NYSM stations, 1 Jan 2018–31 Aug 2022
  - All atmospheric data, incl. GHI, temperature, wind, humidity, etc.
  - Averaged into 15-min time-ending values for use in NYSolarCast
  - LI-COR LI-200R and LI-200RX pyranometers
- Instances of shaded or snow/ice-covered pyranometers were found, confirmed by U Albany, and excluded from training and validation datasets
- LI-COR pyranometers at the Standard sites can have bias/calibration issues, but these are not uniform network-wide
  - NYSM team is working to address this issue
  - Occasional updates to calibrations
  - Sensors periodically replaced
- 17 of these Standard Network sites are also NYSM Flux Network sites with high-quality Kipp & Zonen CNR4 pyranometers

WRF-Solar & HRRR

- Extended history of WRF-Solar® reforecasts over NYS for training for machine learning models
  - 15 Jul 2018 – 31 Aug 2022, using WRF v4.2
  - **Intra-day**: Out to 6 h, initialized hourly 11z–19z from 2-h old HRRR
  - **Day-ahead**: Out to 42 h, initialized once daily at 06z from 06z HRRR
  - Several valuable 2-D solar diagnostics in standard output
  - 15-min, 3-km gridded output for both Nowcast & Day-ahead cycles
- Operational HRRR also downloaded and re-gridded to WRF-Solar grid to provide a blended NWP background





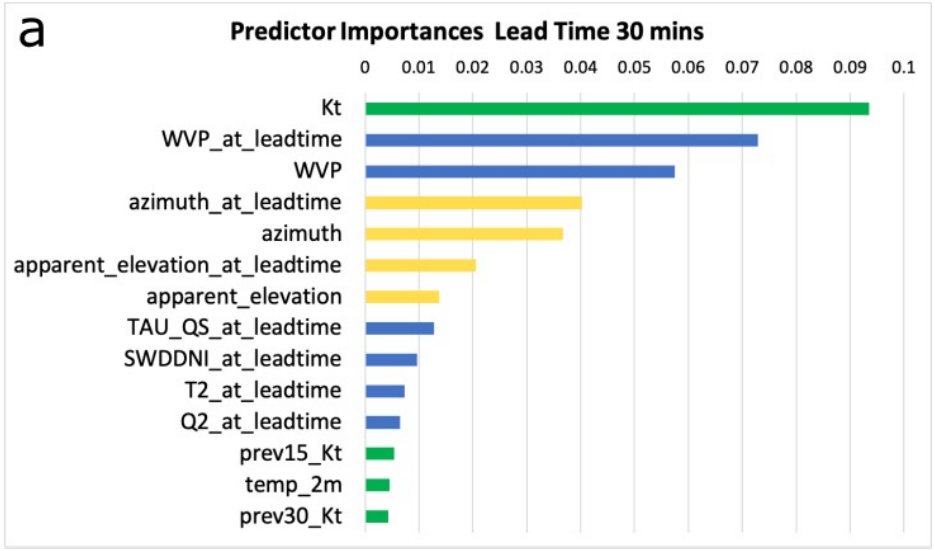


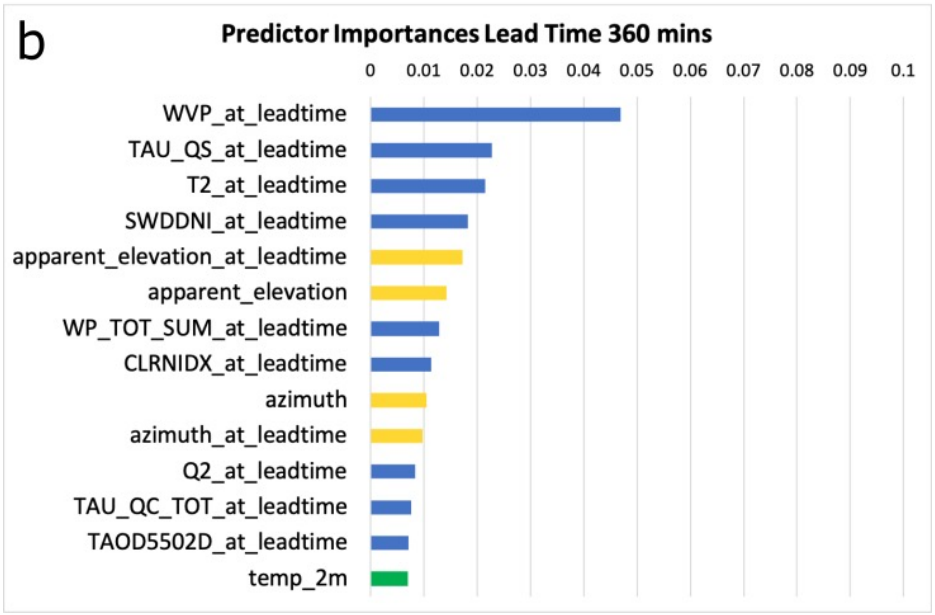
Above: WRF-Solar domain (265x265) with NYS Mesonet stations included

Left: Sample WRF-Solar GHI forecast

StatCast: Blending NWP Models and GHI Observations from NYSM

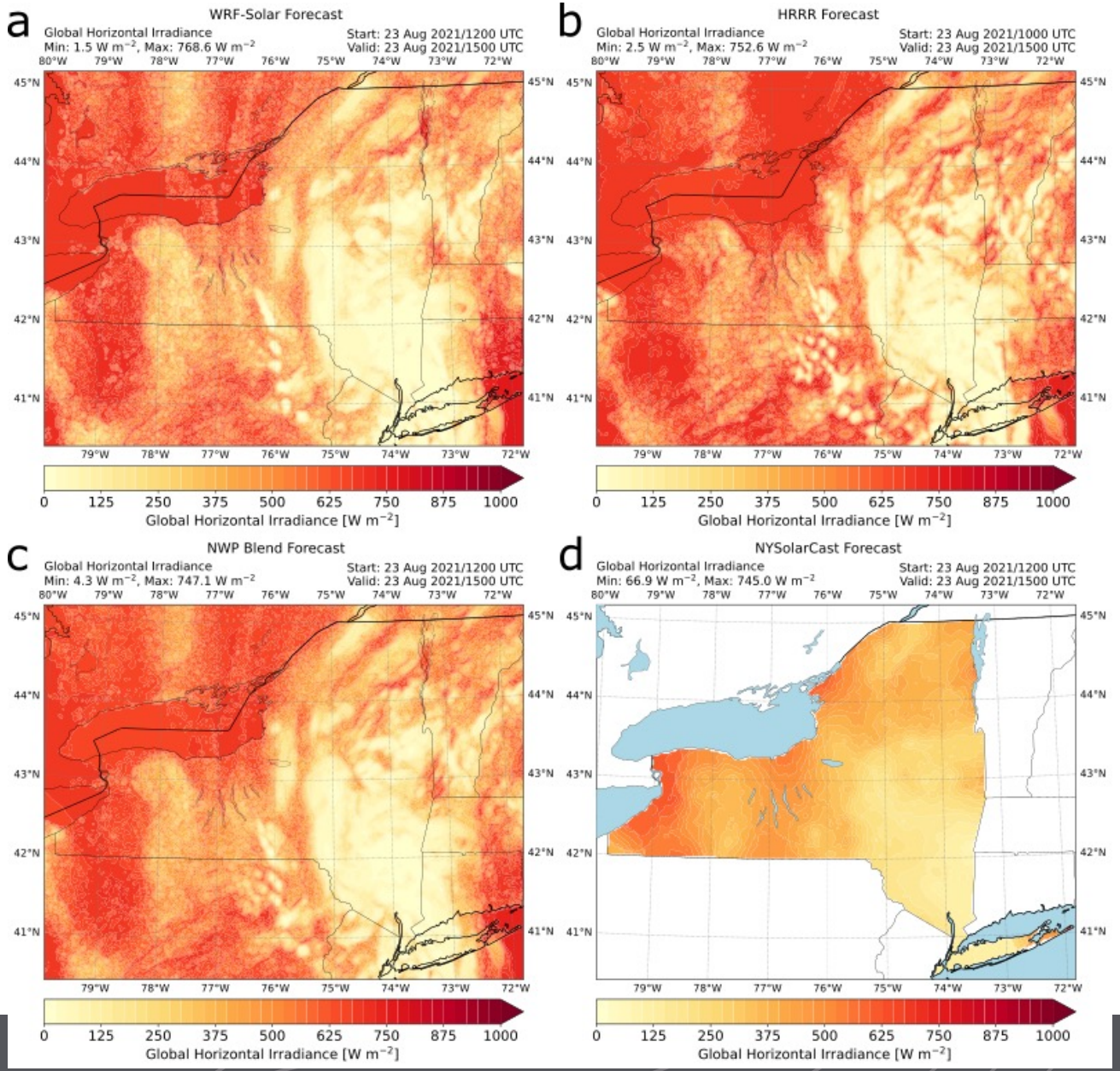
- StatCast is a statistical forecasting model developed by NCAR, and has been applied to wind/solar forecasting previously
- StatCast uses the Cubist machine learning (ML) algorithm
  - Rule-based decision trees
  - Separate model for each lead time
  - Single model for all of NYS
- Cubist predictand: Clearness index (Kt)
  - Removes strong diurnal trend in GHI
  - Can easily be converted back to GHI
- Cubist predictors for each lead time model include:
  - WRF-Solar variables
  - Past 45 min of observed Kt at NYSM sites
  - Known solar angles
- Training period: 15 Jul 2018–30 Apr 2021
- Validation period: 1 May 2021–30 Apr 2022





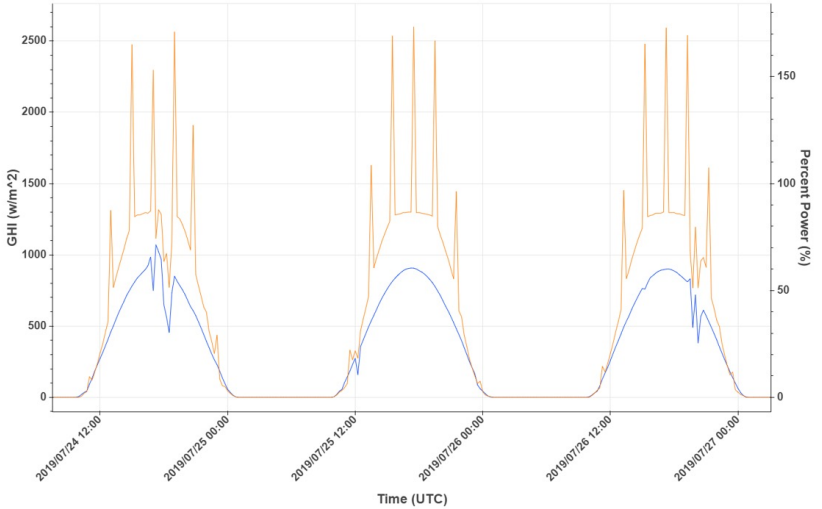
Blending StatCast & NWP

- NYSM sites mapped to WRF-Solar grid points
- Kt is converted back to GHI at NYSM sites
- Initial radius of influence (ROI) of 40 km
- Each grid point in NYS is a weighted average of forecast GHI, with weights inversely proportional to distance to nearest NYSM site (“intermediate product”)
- WRF-Solar (panel a) & HRRR (panel b) are blended together as a background forecast (currently a 50/50 blend, panel c)
- StatCast blends intermediate product with NWP blended GHI to generate **final gridded GHI product (panel d)**
- StatCast weights (linear in between these points):
  - 100% at 0 km, 90% at 30 km, 0% at 40 km from nearest NYSM site
  - 100% from 0–3 h, 50% from 5.5–6 h

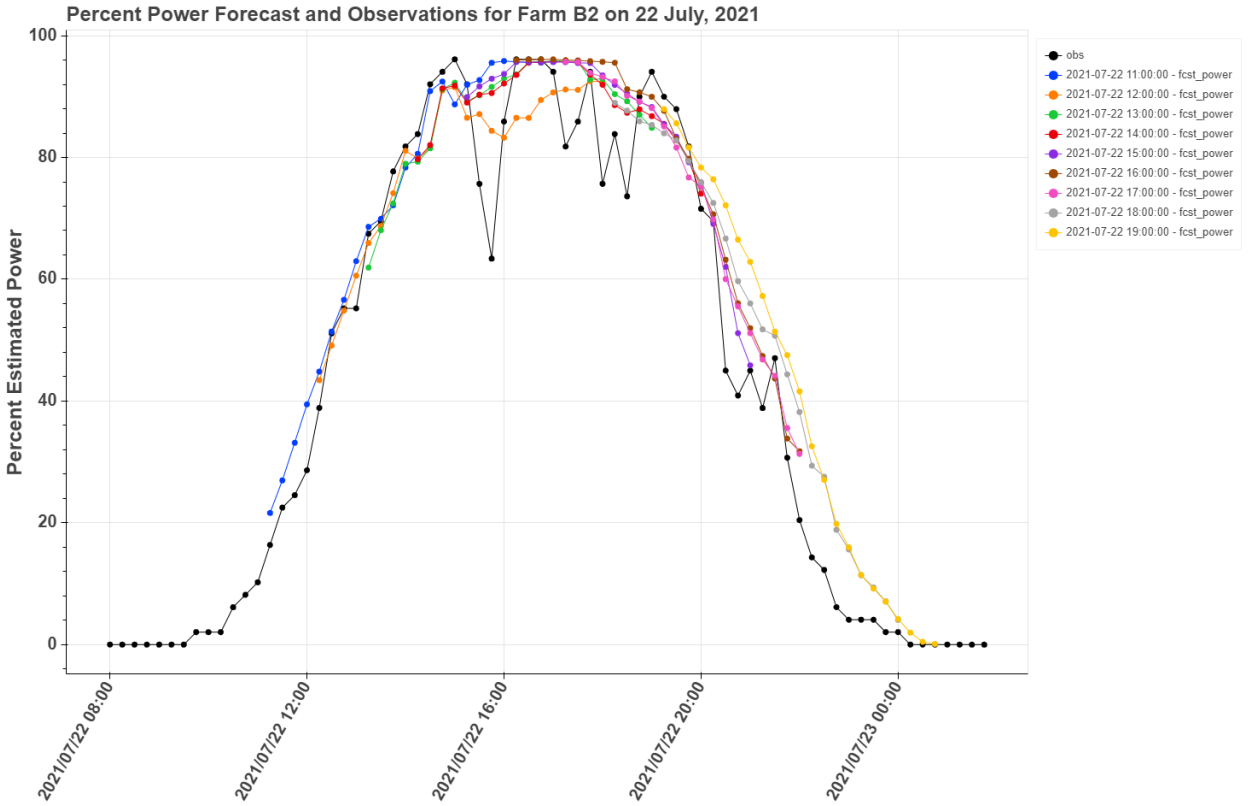


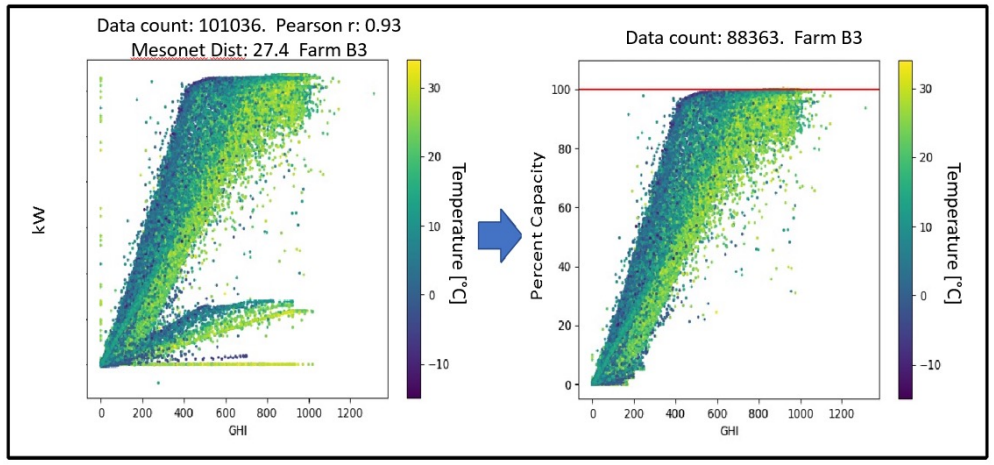
Percent Power Forecasts — UPV Farms

- Power production (in kW) and GHI data from several PV farms in NYS on monthly basis
  - Provider A: 4 farms, training data to Apr 2021
  - Provider B: 6 farms, training data to Apr 2021
  - Varying start dates, varying QC issues
  - Some farms are curtailed daily, some aren't
- Rescaled to % capacity, set to P99.9 of obs
- Cubist used to generate % power models from GHI forecasts every 15 min at 15-min res.



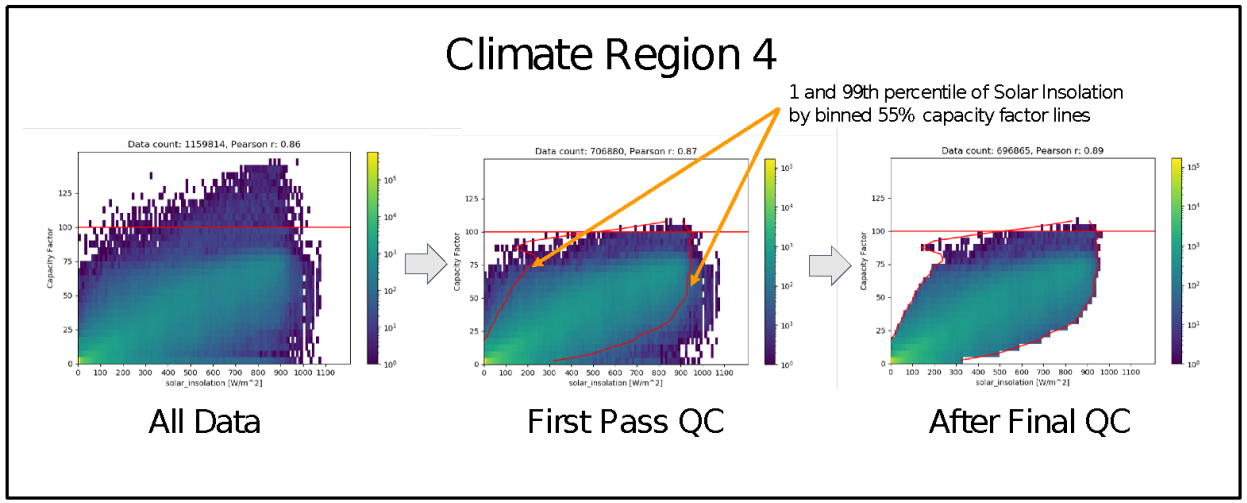
GHI and % power output for one PV farm.
Real data is often messy and QC is crucial!

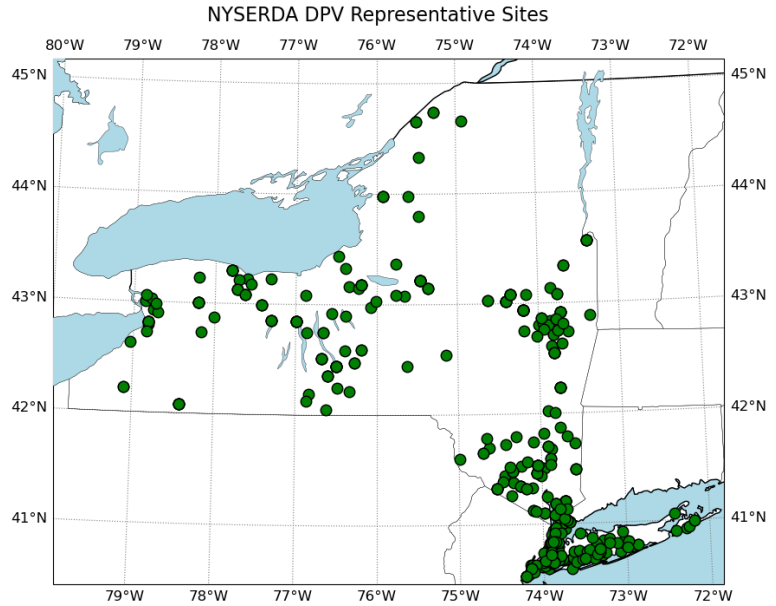




Percent Power Forecasts — DPV Sites

- NYSERDA has a database of over 101,000 DPV installations (nameplate capacity, lat, lon, ZIP)
- NYSERDA has 1-hourly DPV production data from almost 500 “representative sites”
  - Start/end dates & data quality vary by site
  - Training 1 Jan 2018–1 Apr 2020 when available
  - Most of these sites are within 10–15 km of the nearest NYSM station, all within 30 km
- NYSM obs (GHI, 2-m T, 2-m RH) converted to 1-hourly time-ending averages





The 436 NYSERDA DPV representative sites that were not completely rejected by QC

- Single statewide distributed % power Cubist model
  - U.S. Climate Regions in NYS is a variable for Cubist
- The data is messy—additional QC beyond NYSERDA’s QC is needed, e.g.:
  - Pearson r of capacity factor (CF) & GHI < 0.75?
  - P99 CF > 100% or < 50% of nameplate capacity?
  - GHI > 1200 W/m^2? CF > 150%?
  - CF = 0 and GHI > 240 W/m^2 (20% of 1200 W/m^2)?
  - GHI and CF both 0 or missing/NaN values for either?

DPV Forecast Aggregation

- % power forecasts produced for every grid point in NYS
- For each distributed PV site, % power at nearest grid point multiplied by nameplate capacity
- All sites' forecasted total power then aggregated regionally
- NYSolarCast currently configured to aggregate by regional (3-digit) ZIP code
- Could also use NYISO load zones, counties, or other useful regions of interest
- NYSolarCast framework is flexible for any aggregation — simply assign each grid point to a zone/region in a config file
- Note: Unresolved large mismatches of total DPV capacity in NYISO load zones between NYISO & NYSERDA databases

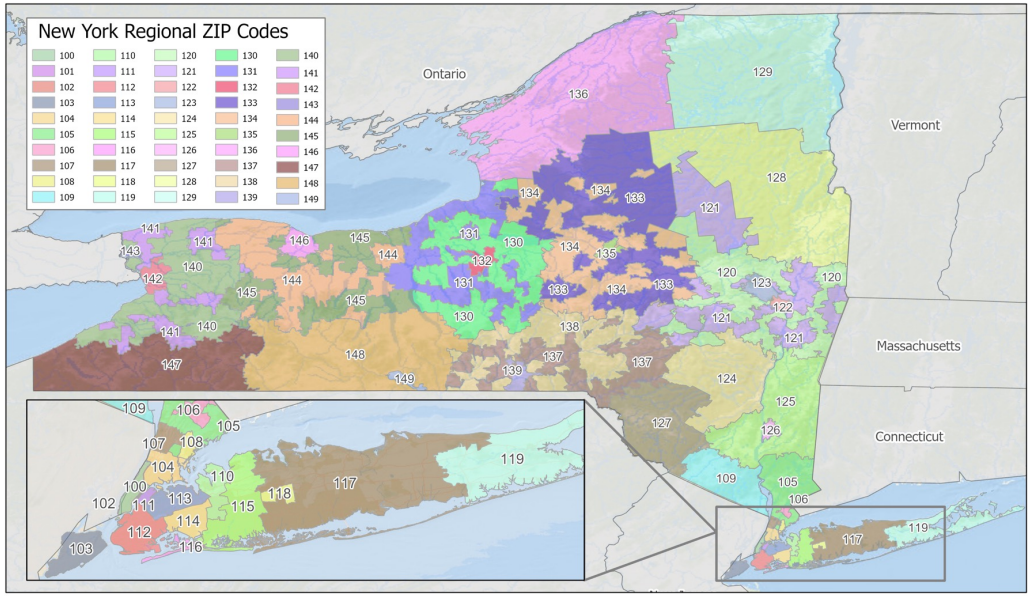
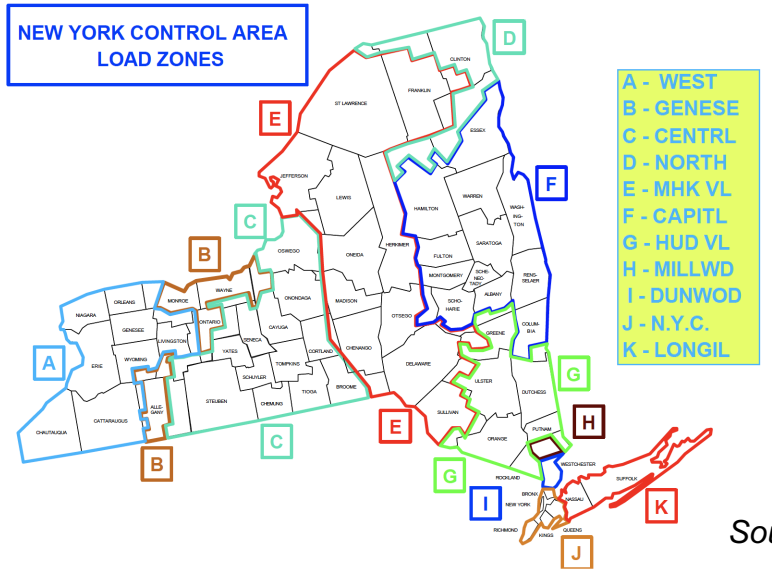
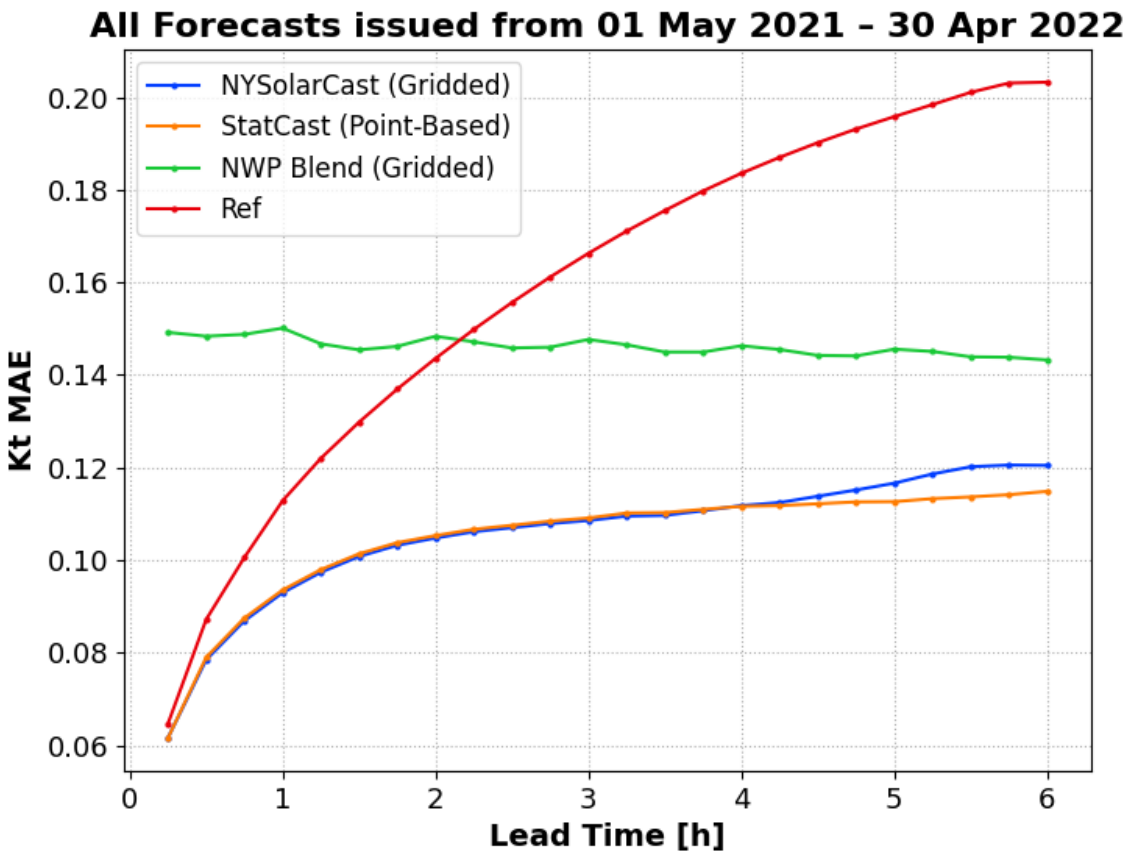


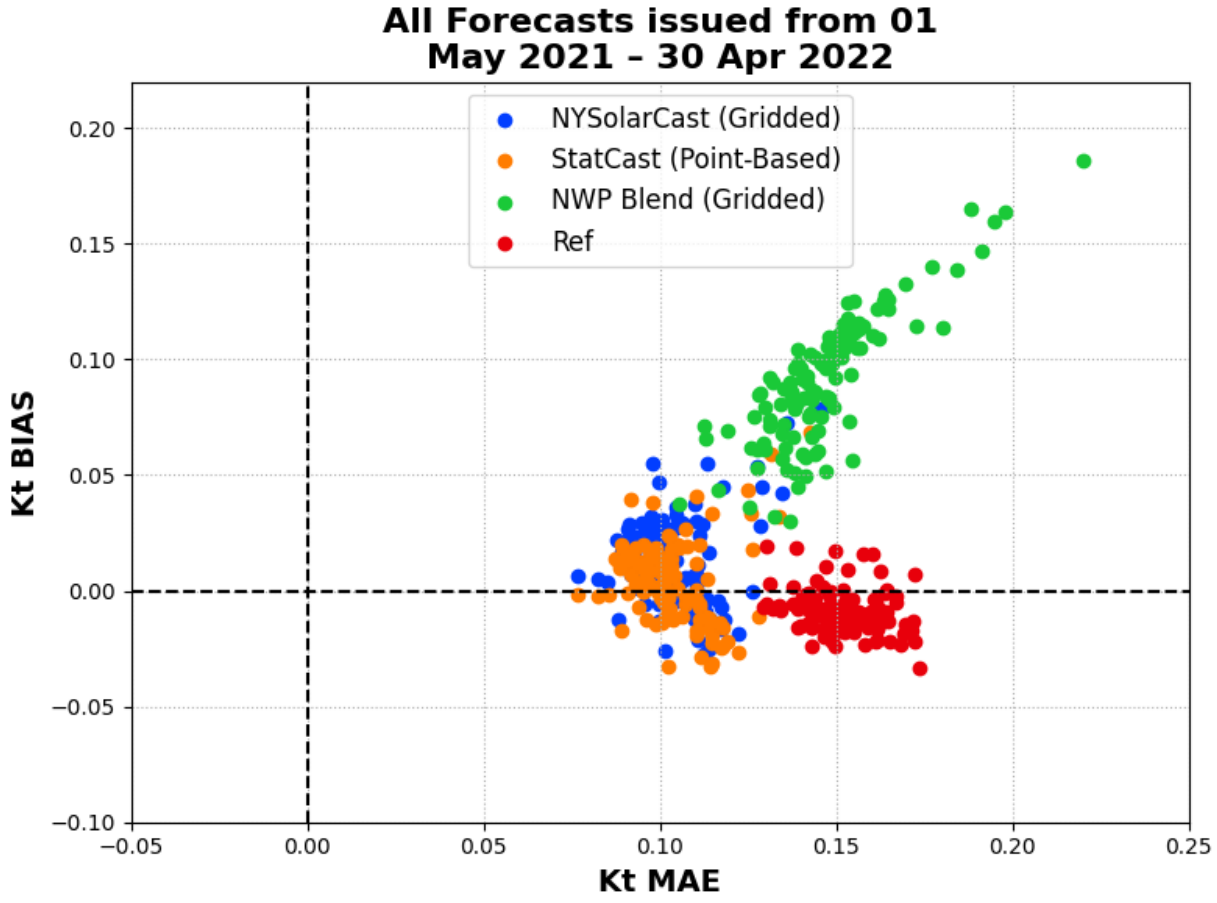
Image: Jennifer Boehnert (NCAR)



Source: NYISO

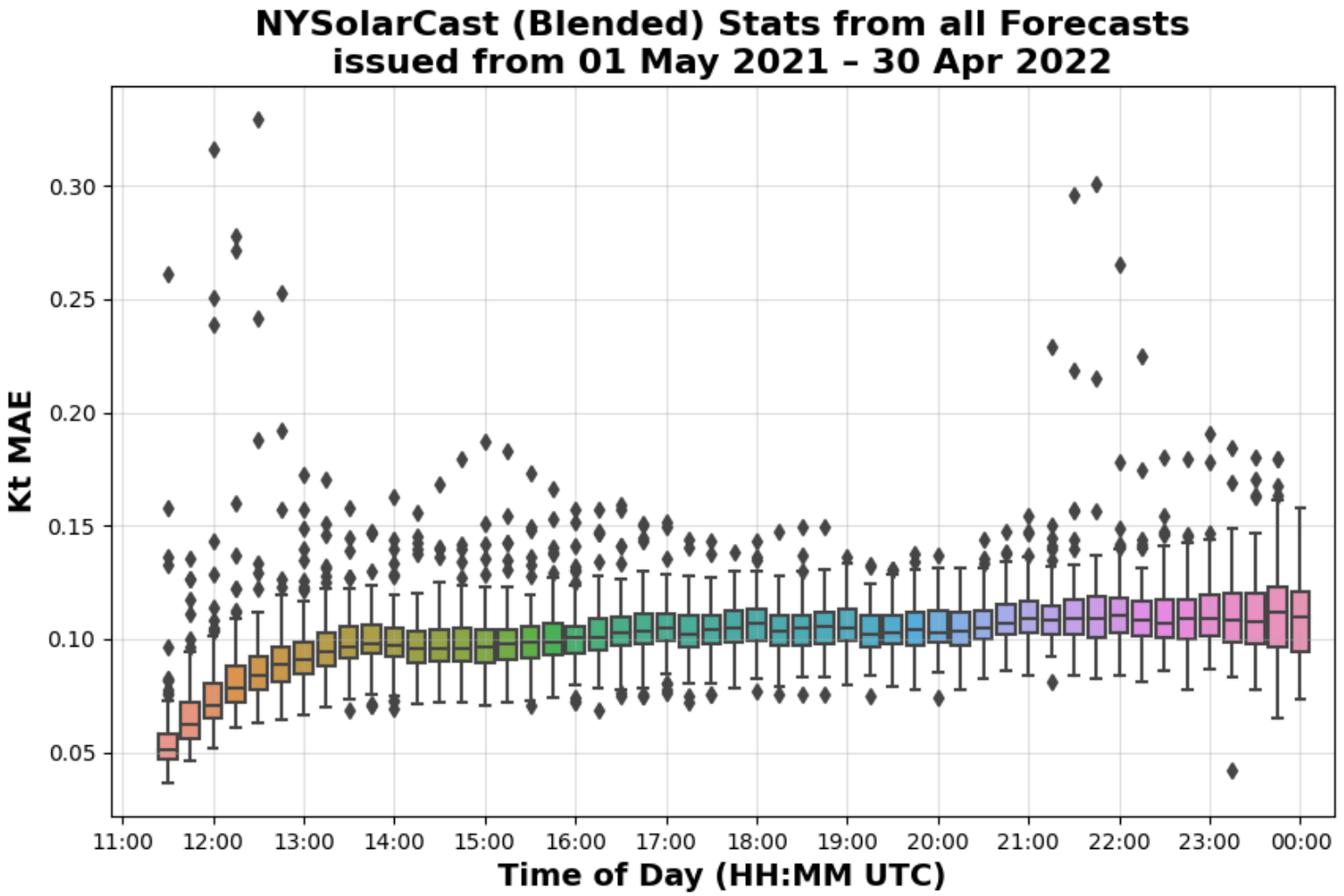
Kt Validation at NYSM Sites





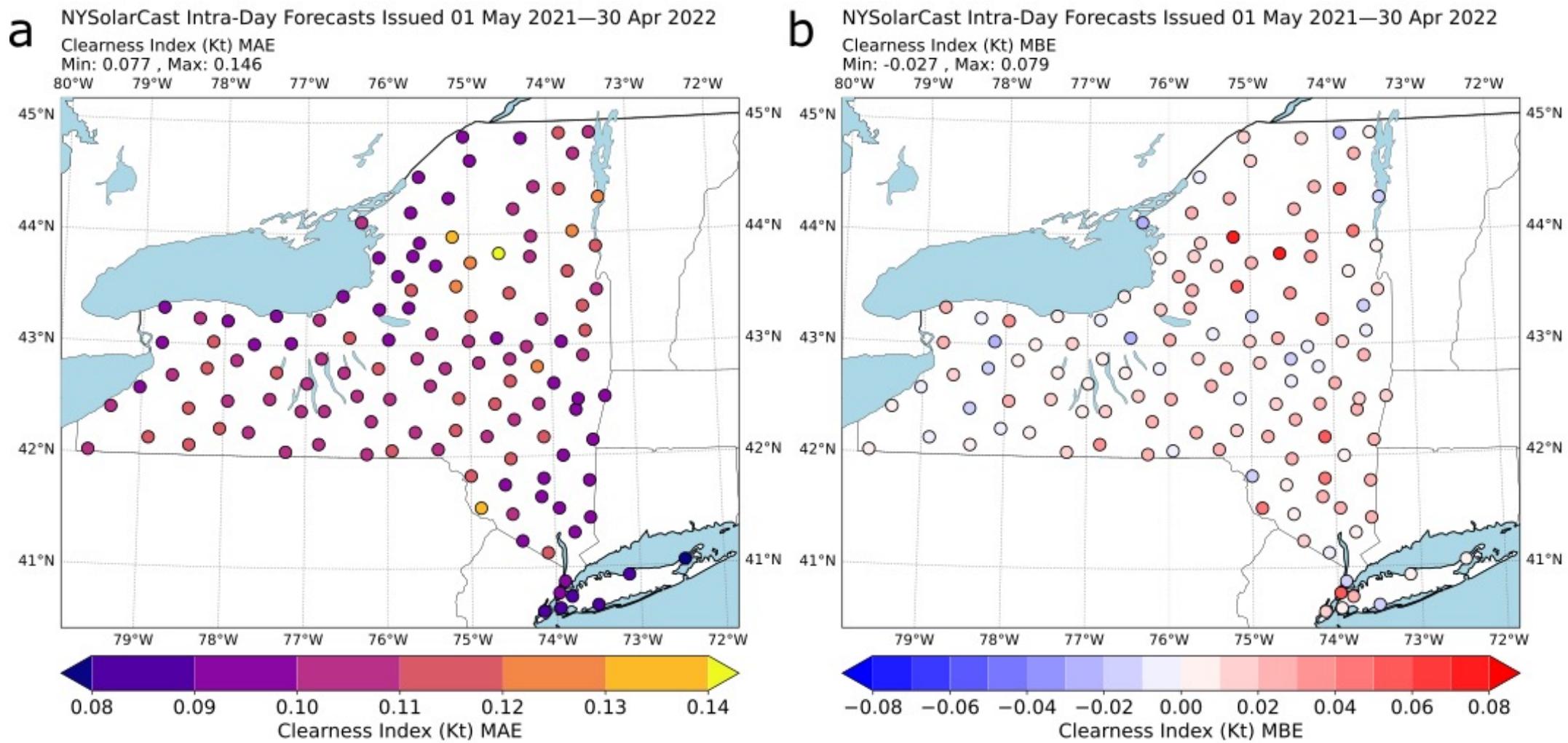
- NYSolarCast & StatCast identical at NYSM for first 3.5 h, then NYSolarCast relaxes toward NWP Blend
- NYSolarCast better MAE than smart persistence and NWP Blend at all lead times and nearly all sites
- NWP Blend better MAE than smart persistence after 2 h, and slowly declines with lead time

Kt Validation at NYSM Sites



- Kt MAE fairly consistent at most sites at all times of day
- A few outlier sites with high MAE in early morning and late afternoon (additional shading issues??)

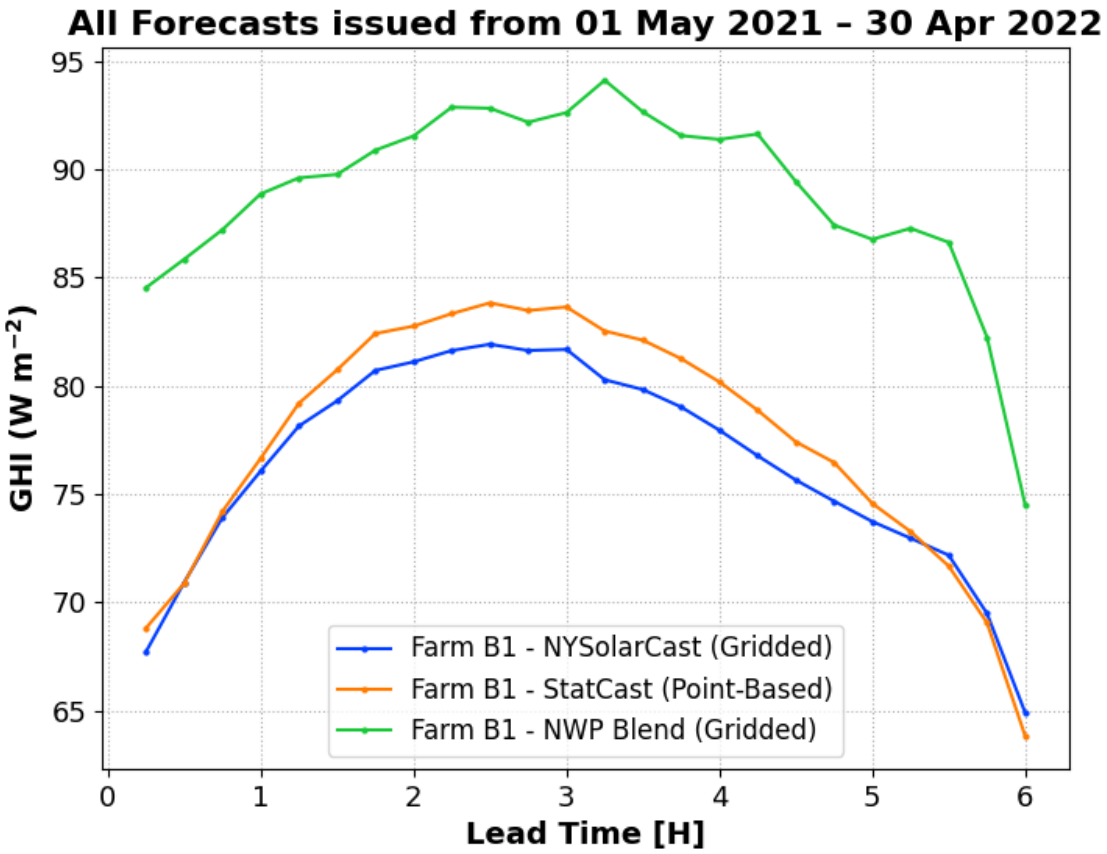
Kt Validation at NYSM Sites

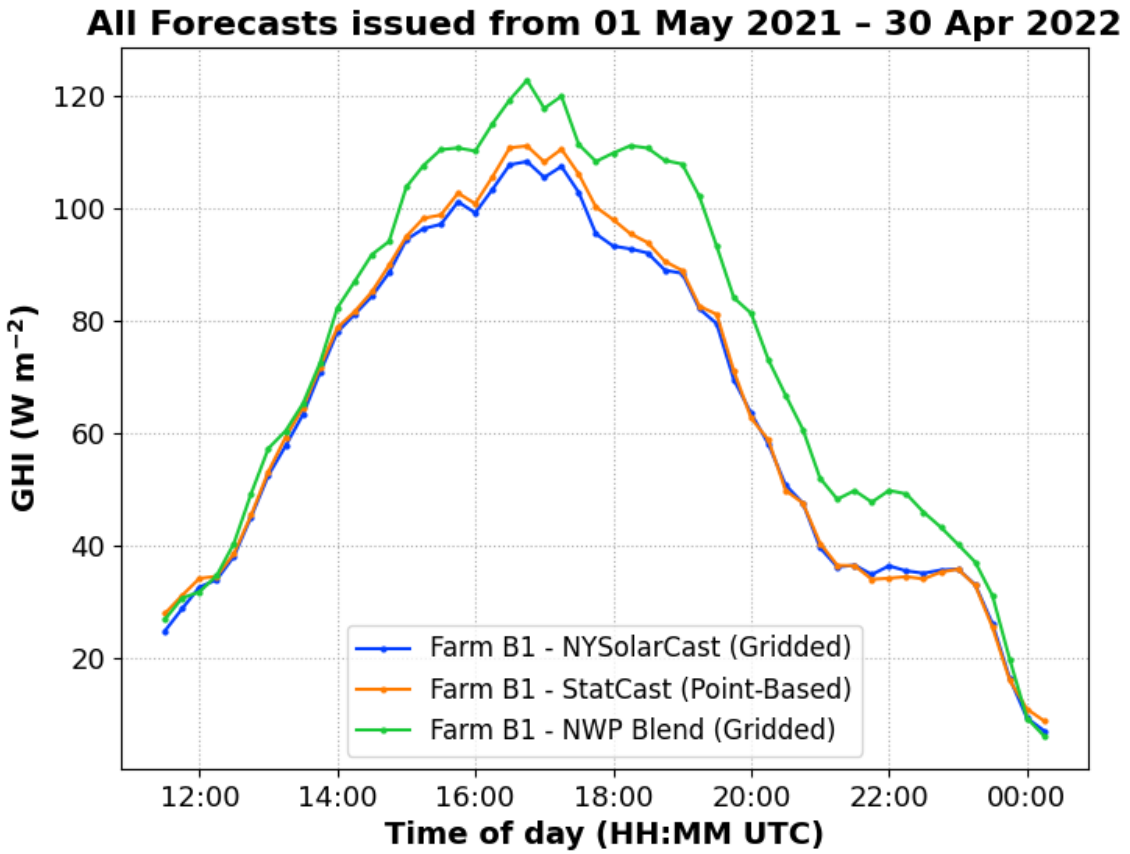


- Highest MAE mostly in Adirondacks, Upper Hudson Valley
- Possible additional snow/shading impacts at some sites?

- Near-zero Kt MBE at most sites
- Highest MBE sites mostly correspond with highest MAE sites

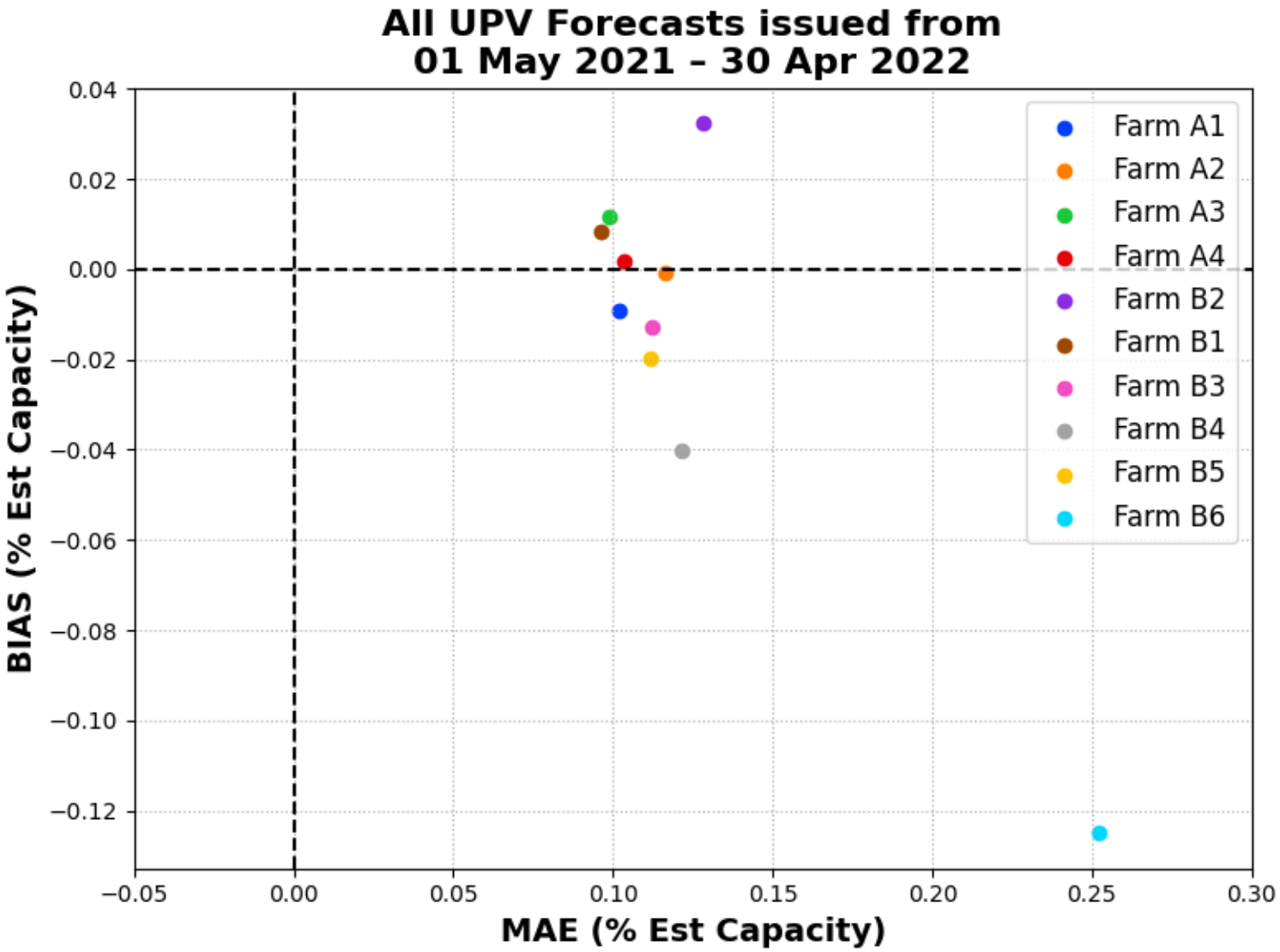
GHI Validation at UPV Sites





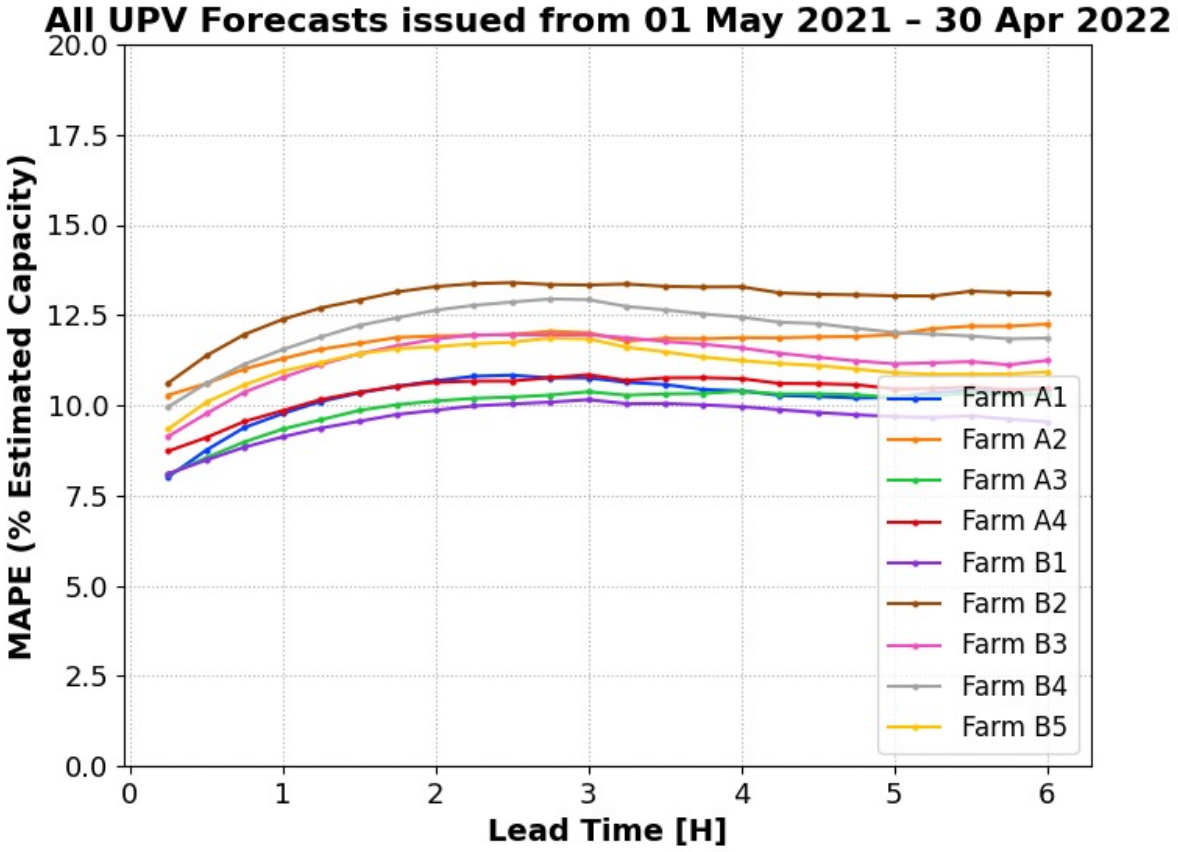
- **NYSolarCast GHI at the UPV farm is nearly always better than the StatCast GHI at the nearest NYSM site**
- If real-time obs from UPV farms are unavailable, NYSolarCast still adds value using nearby weather stations
- The forecast would be even better with access to real-time UPV farm obs (GHI, temperature, power)

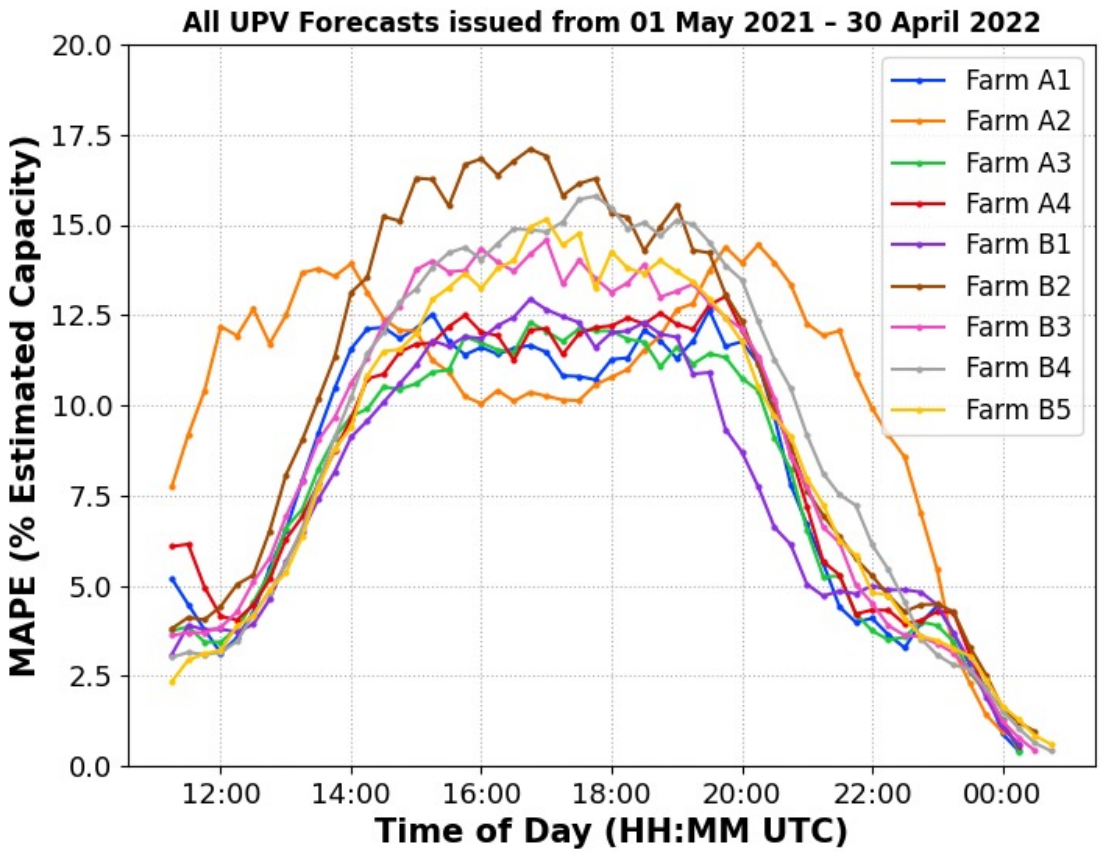
Power Validation at UPV Farms



- 7 of 10 farms have both an MPE of -2 to +2% and MAPE of 9–12%
- Farm B6 excluded from future plots
- Farm B6 outlier status attributed to much shorter training period than other farms and several months of missing data during this 1-year validation period

Power Validation at UPV Farms

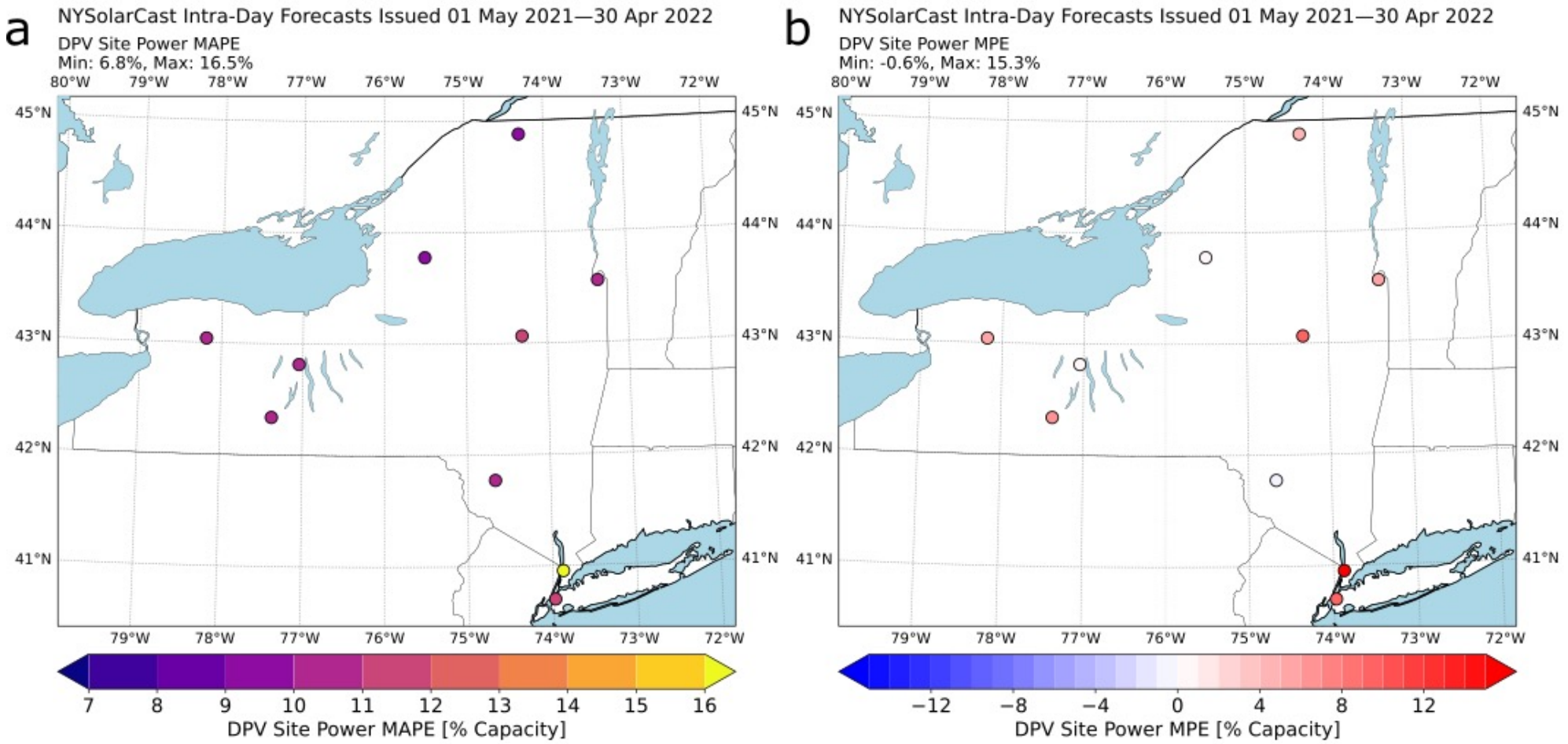




- Power MAPE for all UPV farms fairly constant as a function of lead time, between 8–13%
- Levels off after ~3 hours

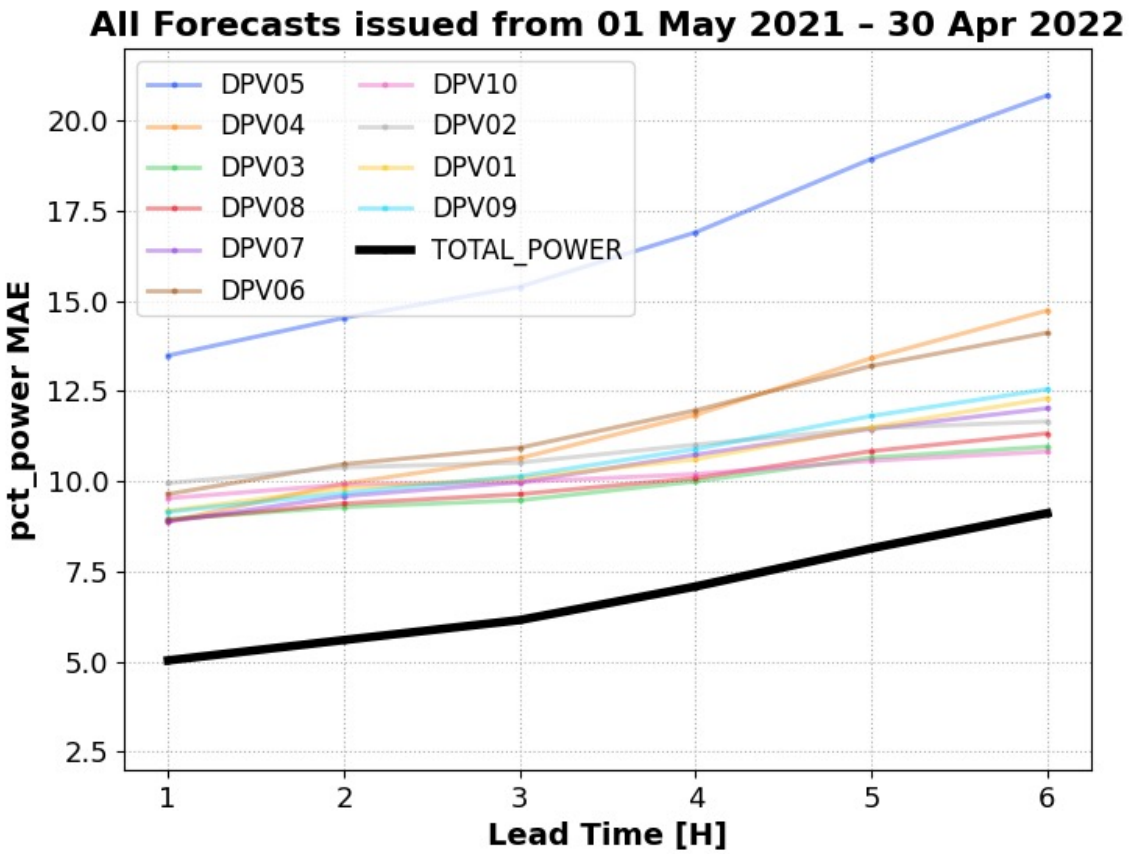
- Power MAPE for most UPV farms generally follows the diurnal GHI curve
- Farm A1, Farm A2, and Farm A4 all exhibit a duck curve, and production data indicates they are likely overbuilt

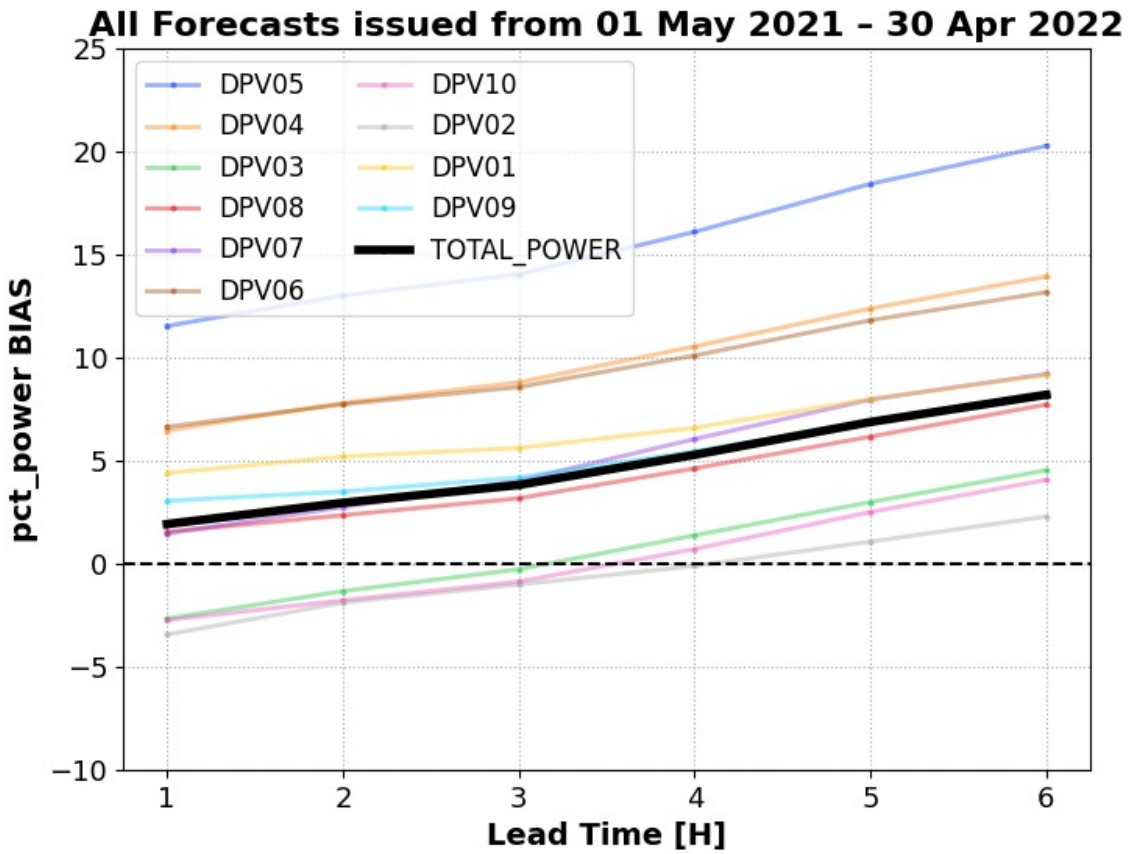
Power Validation at DPV Sites



- These 10 DPV sites have a range of sizes (783 kW–2.90 MW), tilt angles (10°–30°), and azimuth angles (141°–200°)
- 9 of the 10 sites have overall MAPE 6.8%–11.6%, and overall MPE –0.6% to 9.8%
- One outlier site just north of NYC has MAPE 16.5% and MPE 15.3% — data averaging/DST issues??

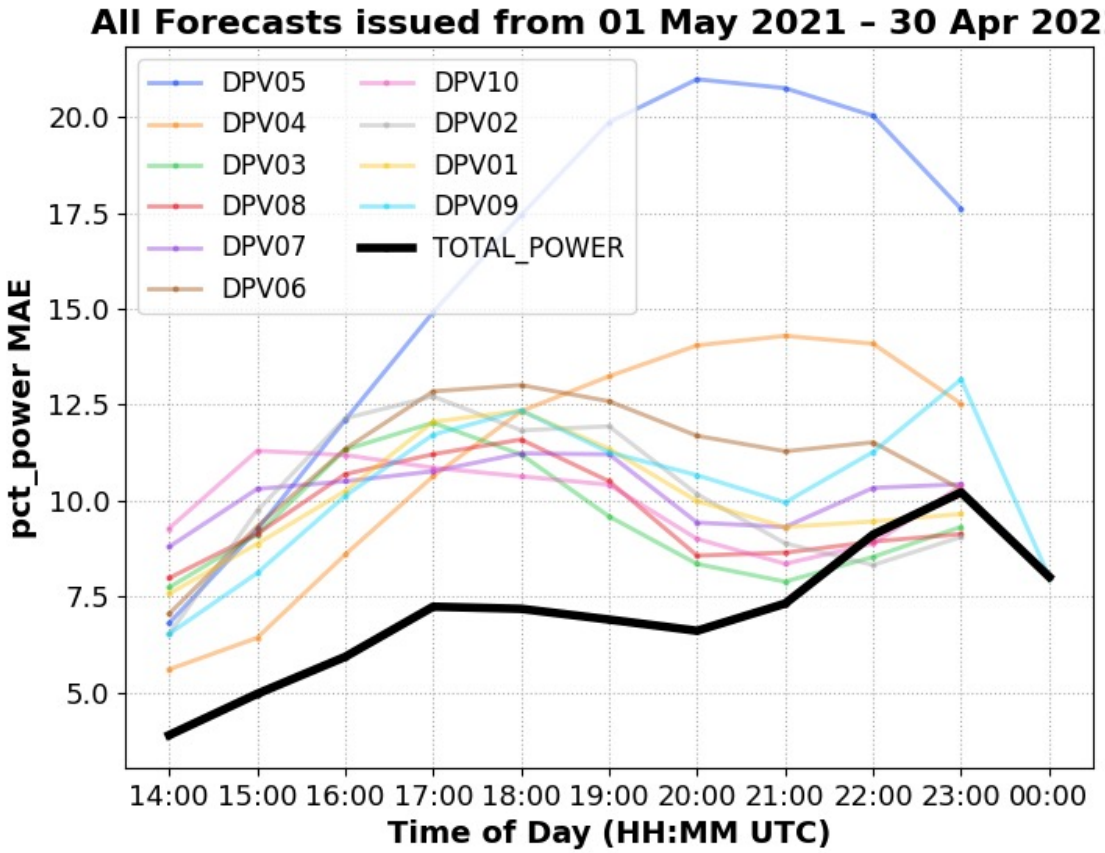
Power Validation at DPV Sites

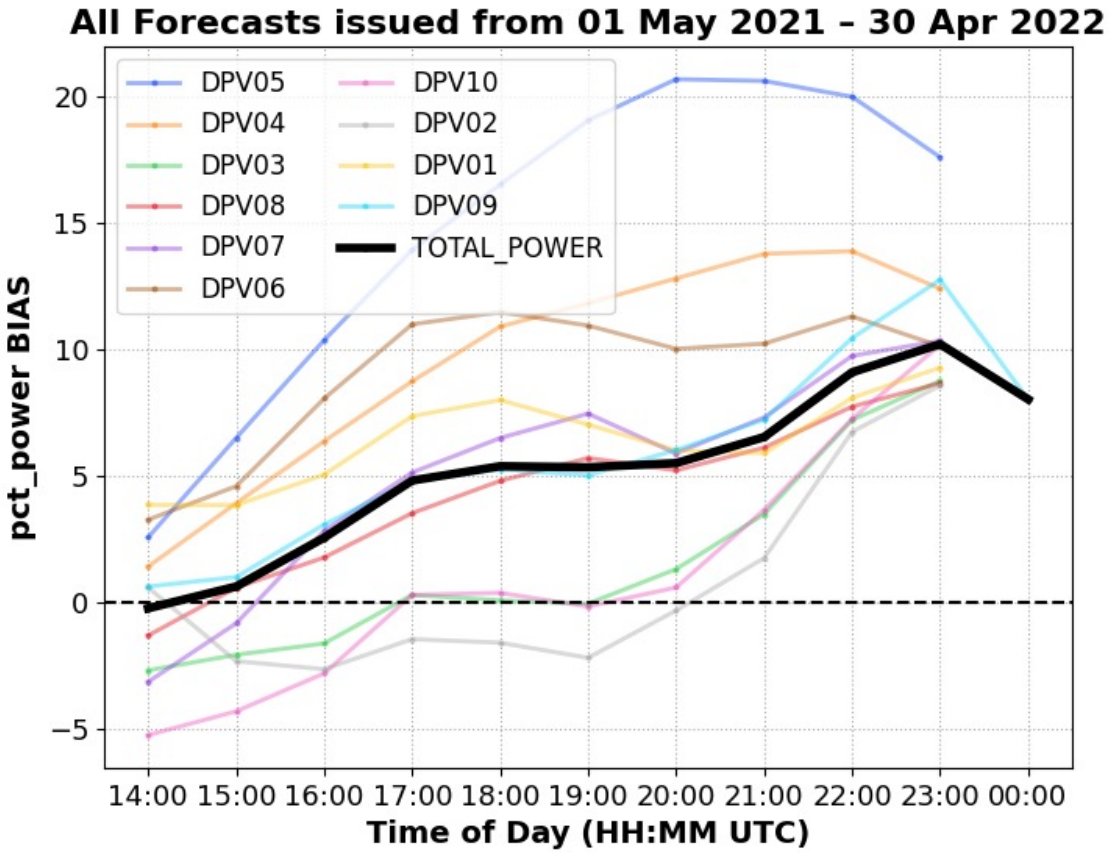




- Aggregating these 10 sites across NYS together yields a lower MAE than any individual site, generally smaller MBE
- Aggregation over regions helps “cancel out” some of these differences in tilt & azimuth angle, shading, etc.

Power Validation at DPV Sites





- Aggregating these 10 sites across NYS together yields a lower MAE than any individual site, generally smaller MBE
- Aggregation over regions helps “cancel out” some of these differences in tilt & azimuth angle, shading, etc.

Conclusions and Future Work

- We developed NYSolarCast to predict solar power in NYS — applicable in other areas!
  - Entirely open source software (https://github.com/NCAR/NYSolarCast_delivery)
  - Predicts GHI on a 3-km grid across NYS every 15 min out to 6 h, and at 06 UTC daily for day-ahead
  - Predicts 15-min % power capacity at select utility-scale PV farms
  - Predicts 1-hourly % power capacity for distributed PV aggregated to regions
- Real-time NYS Mesonet data is critical to NYSolarCast system, especially in the absence of real-time data from UPV farms or DPV sites
- NYSolarCast beats both smart persistence & NWP blend at all intra-day lead & valid times
  - A few sites with larger errors may have additional shading or snow cover issues not flagged in QC
- NYSolarCast GHI at UPV farms is nearly always better than StatCast GHI at nearest NYS Mesonet station
  - Especially valuable when real-time data from UPV farms is not available
- Most UPV farms with MPE –2% to +2%, MAPE 9–12% over 1-year validation period
- NYSolarCast aggregated DPV MAPE < 10%, MBE < 7% for all times of day but late PM
- Future: Improve day-ahead forecasts via dynamic NWP blend weighting, bias correction
- Manuscript in final preparation, submitting to *Solar Energy* in coming weeks

Thanks for listening!

Photos: ©2019 Jared Lee, Shagaya Renewable Energy Park, Kuwait

jaredlee@ucar.edu

Questions? Interested in using/expanding/improving NYSolarCast? Please email me!