Weather Datasets for Power System Planning

Problems with Existing Datasets

Dr. Justin Sharp
Technical Leader

ESIG Markets and Meteorology Workshop
June 2024. Salt Lake City, Utah

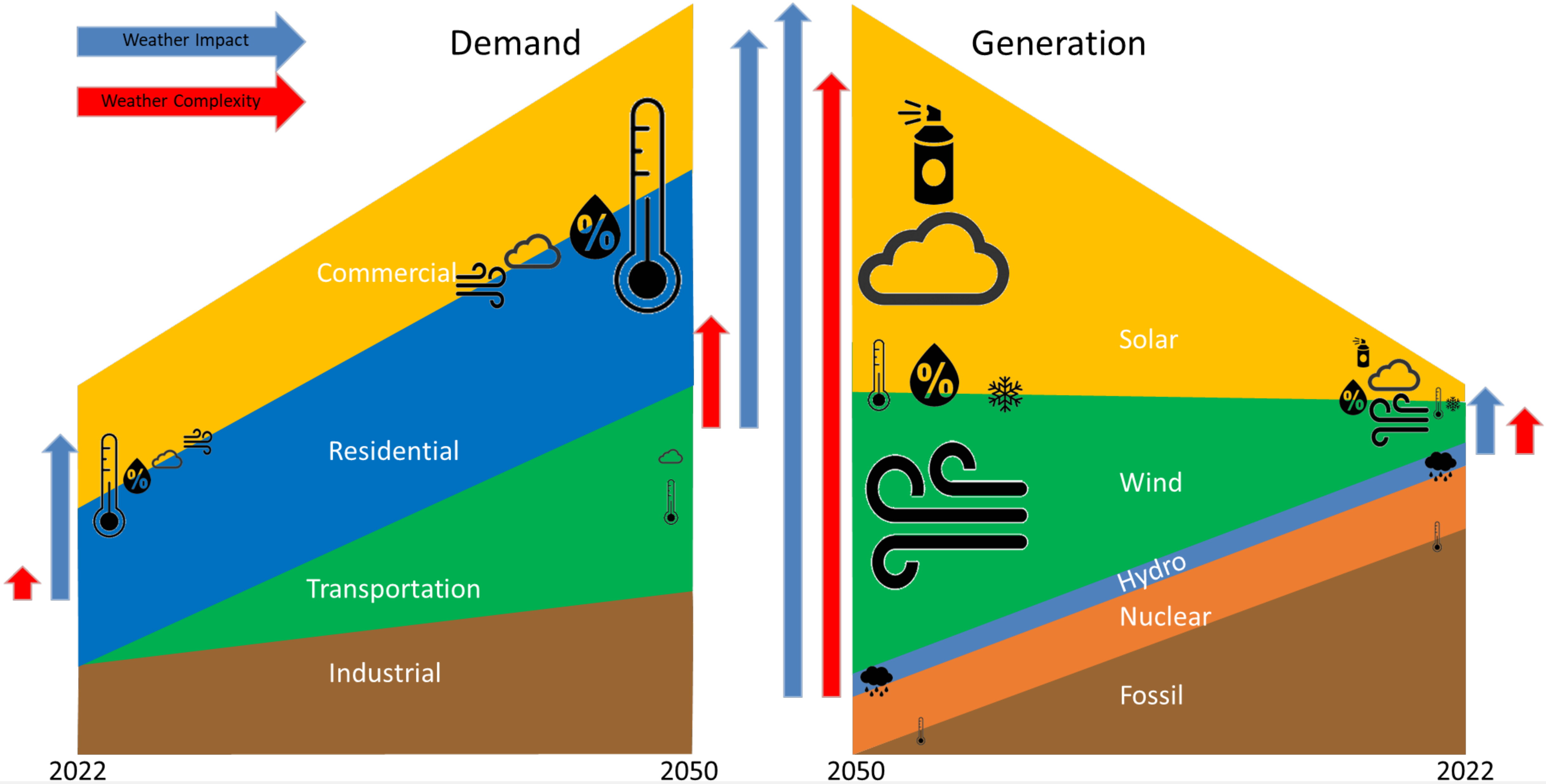
Support and content provided by:

ESIG ENERGY SYSTEMS INTEGRATION GROUP

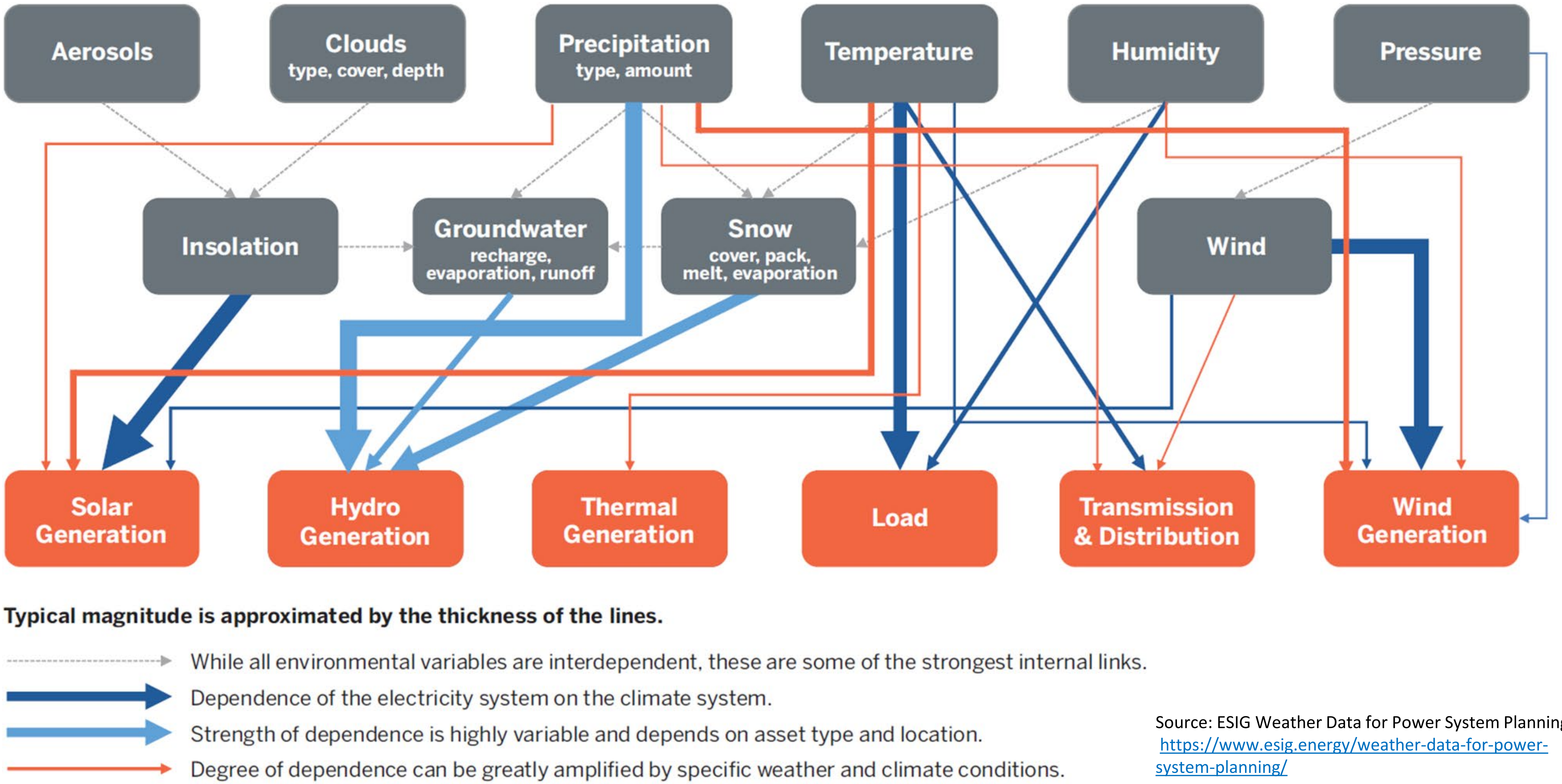
GridLAB

Sharply Focused Understanding Renewable Resources

The Evolving Energy – Weather Nexus



Electricity System Weather-Dependence



Source: ESIG Weather Data for Power System Planning
https://www.esig.energy/weather-data-for-power-system-planning/

Attributes of Quality Representative Datasets

Physical Requirements

- Representative of Actual Weather
- Coincident, Physically Consistent Weather Variables
- Sufficient Resolution (<=2 km, <=15 min)
- Chronological

Relevance Requirements

- Covers Multiple Decades with Consistent Method
- Validated and Uncertainty Quantified
- Regularly Extended
- Periodically Refreshed

Usability Requirements

- Transparently Documented
- Publicly Available and Easily Accessible
- Expertly Curated

The Main Attributes of Time Series Data Necessary to Meet General Power System Modeling Needs

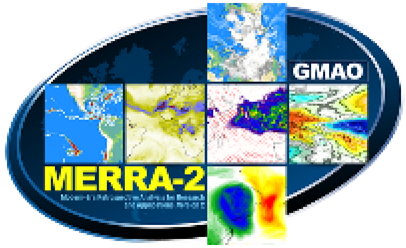
Source: ESIG Weather Data for Power System Planning https://www.esig.energy/weather-data-for-power-system-planning/

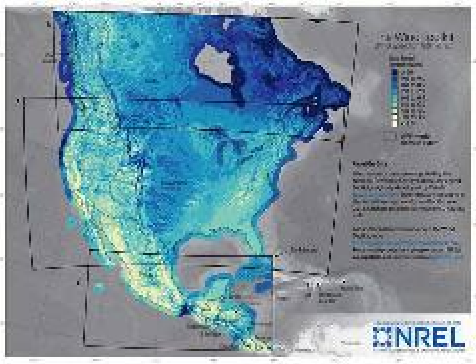
Including the necessary variables	Include the necessary variables at sufficient spatio-temporal resolution and accuracy to reflect actual conditions that define the generation potential at current and future wind/solar sites and temperature at load centers
Covering multiple decades with ongoing extension	Cover multiple decades with consistent methodology and be extended on an ongoing basis to capture the most recent conditions and allow climate trends to be identified
Coincident and physically consistent	Are coincident and physically consistent, in space and time, across weather variables
Validated	Are validated against real conditions with uncertainty quantified
Documented	Are documented transparently and in detail, including limitations and a guide for usage
Periodically refreshed	Are periodically refreshed to account for scientific and technological advancements
Available and accessible	Publicly available, expertly curated, and easily accessible

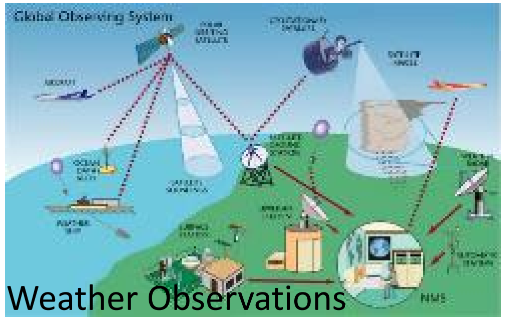
Gaps in Today's Sources

ERA5

Copernicus
Europe's eyes on Earth







All datasets have at least one of the following gaps:

- Insufficient spatial or temporal resolution
- Insufficient time history
- Insufficient validation/uncertainty quantification
- Distributions don't match reality, especially for extreme events
- Not coincident or physically consistent
- Archaic or not extended to present date
- Non-static modeling platforms
- Care required for use not well advertised!

Why it matters:

- Inaccurate weather => poor wind and solar generation estimates. Sometimes, VERY poor.
- Load estimation is more difficult
- Other concurrent system conditions are poorly estimated

EPRI

A Closer Look at the Data Available Today

- Observations:
  - Closest representation of truth
  - Too sparse, and requires rigorous QC
  - And/Or Proprietary
- ERA5 (~30 km) (and MERRA2, ~60 km):
  - Longest, most complete consistent time series
  - Easy to use
  - Too low resolution for generation estimates
- WIND TK (2 km, 5 min/hourly):
  - Resolves most physical phenomena
  - Includes forecast database
  - Some temporal seams
  - Outdated model, esp. not great for solar
  - Only 2007-2013 using same set up. 2014 available using different configuration.

- NSRDB (4 km, 30 min):
  - Based on satellite observations and a physics based model
  - Continuous and consistent since 1998
  - Not originally designed for integration studies
  - Non-solar fields are misleading interpolations of MERRA2
- HRRR (3 km, 15 min):
  - Resolves most physical phenomena
  - Data from operational forecast archive
  - Model configuration inconsistent in time
- Data from proprietary models:
  - Opaque and often unscientific in basis

Common Issue: Lack of validation and examination of use case applicability

EPRI

The Data We Have Today

The data currently available to the sector (on left) is not adequate for the tasks at hand. No single dataset meets all the needs. Mixing and matching causes physical consistency issues.

	Spatial Resolution	Temporal Resolution	Length	Continuously Extended	Correct Variables/ Levels	Coincident and Coherent	Validated/Uncertainty Quantified for Power System Use	Detailed Documentation	Future-Proofed	Availability/ Ease of Access	Curation and Advice	Region Covered
MERRA-2[a]	~60 km	60 min	1980–present	Yes	Yes/No	Yes	No		Probably		Basic	Global
ERA5[b]	~30 km	60 min	1940–present	Yes	Yes/No	Yes	Some		Yes		Good	Global
HRRR[c]	3 km	15 min	2014–present	Yes	Yes/No	Yes/No	No		Unideal		Basic	U.S.
WIND Toolkit[d]	2 km	5 min	2007–2014	No	Yes/Yes	Yes	Yes		No		Basic	Various
WTK-LED[e]	2 km/4 km	5 min	3 year/ 20 year	No	Yes/Yes	Yes	Not yet	Not yet	No	Unknown, dataset not yet available		Various
NSRDB[f]	4 km/ 60 km	30 min	1998–present	Yes	Yes/No	Solar only	Yes		Yes		Basic	Most of globe
CERRA[g]	11 km/5.5 km	60 min	1980–present		No/Yes	No solar	Yes		Possibly		Basic	Europe
CONUS404[h]	4 km	60 min/ 15 min (precip)	1980–2020	No	Unknown/ Probably	Yes	Not the intended use					Continental U.S.
BARRA[i]	12 km/ 1.5 km	60 min	1990–2019	No	Yes/ Probably	Yes				Fee-based		Australia/ New Zealand
Public Observing Networks[j]	Non-uniform, variable density	1 hr or less	Variable	Yes	Yes/No	Mostly	Varies. Not for power systems	Varies	Usually	Usually easy	Varies	Global
Renewable Energy Project Data[k]	Non-uniform, variable density	Usually minutes	Variable but rarely more than 10 years	Varies	Yes/ Usually	Yes	Usually	Varies, but usually poor	Varies	Usually poor	Usually none	Very limited
Proprietary Statistically Derived VRE Shapes[l]	Non-uniform, variable density	Usually hourly	Variable. Rarely reliable long records.	Varies	Usually incomplete	No	Partial	See note	No		None	Very limited

■ Fully Met ■ Close to Being Met ■ Partially Met ■ Met in a Very Limited Way ■ Not Met at All ■ Not Enough Info. for Determination

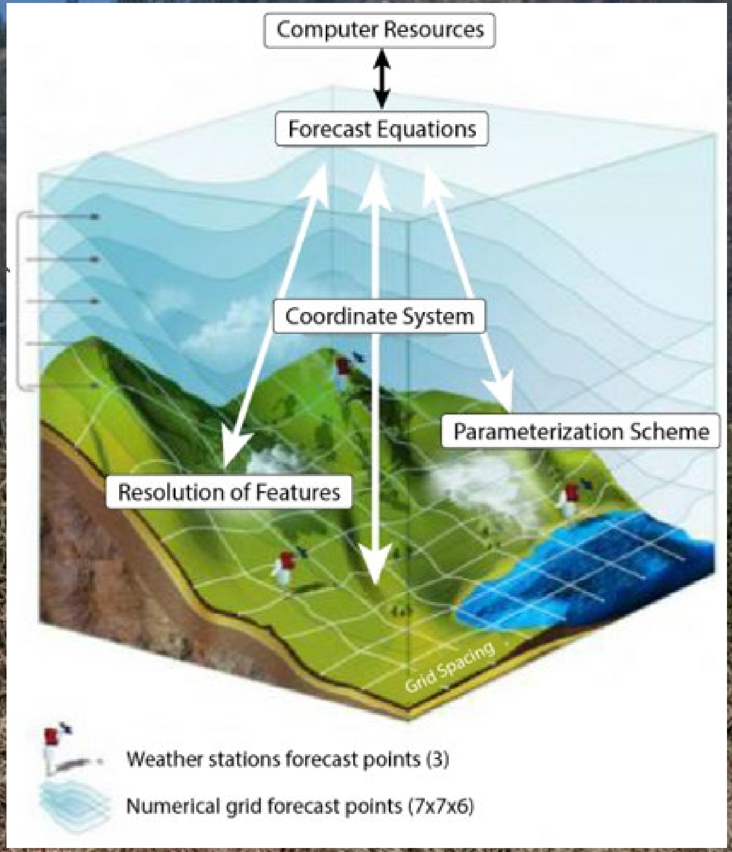
TABLE 2

Summary of Current Power System Modeling Weather Input Data Sources

Summary of the most applicable datasets globally that are (or can be) used to provide weather inputs for power system analysis tasks, especially for providing estimate of site-level generation, and concurrent weather-driven load and generation outage risks. The degree to which the needs of each column heading are met is estimated with color coding. See documentation for each dataset for all details. Footnotes on next page. P76, main report.

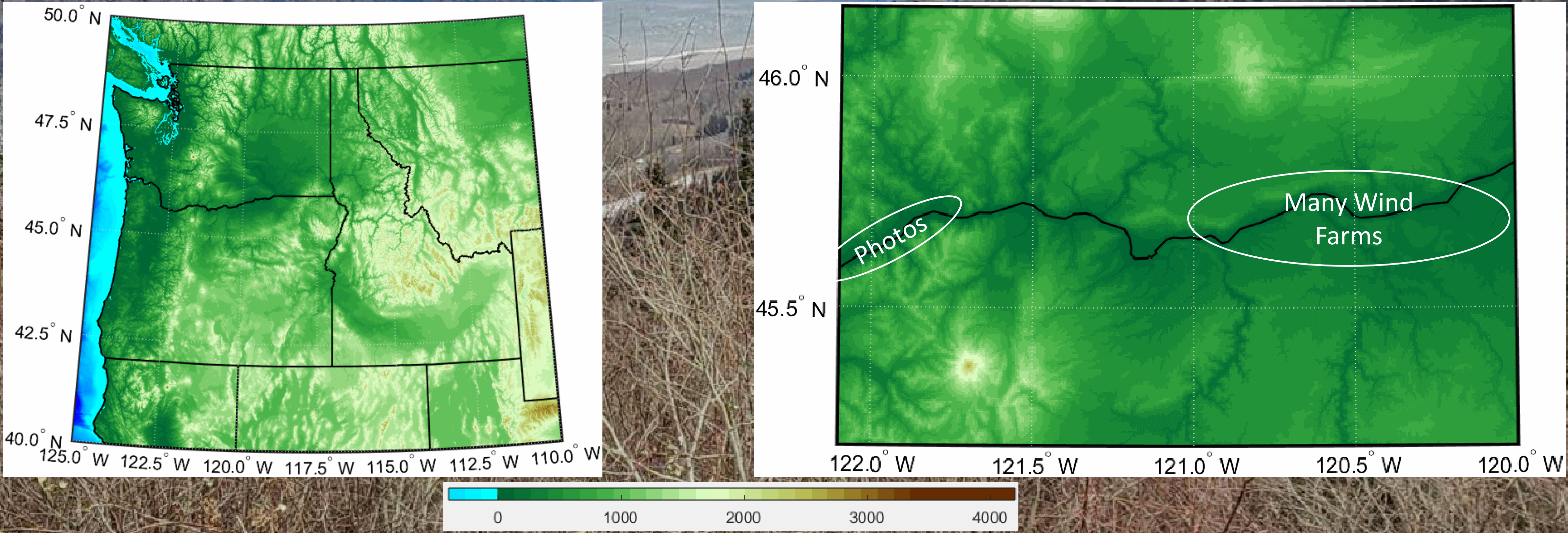
Source: Energy Systems Integration Group

Understanding Grid Spacing (Resolution)



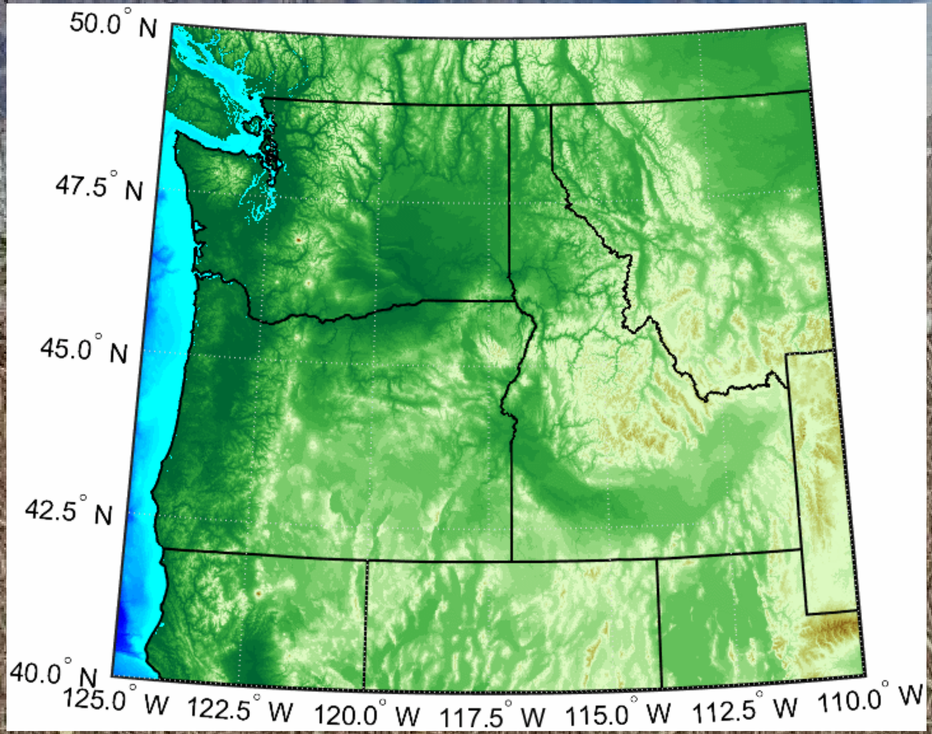
Understanding Grid Spacing (Resolution)

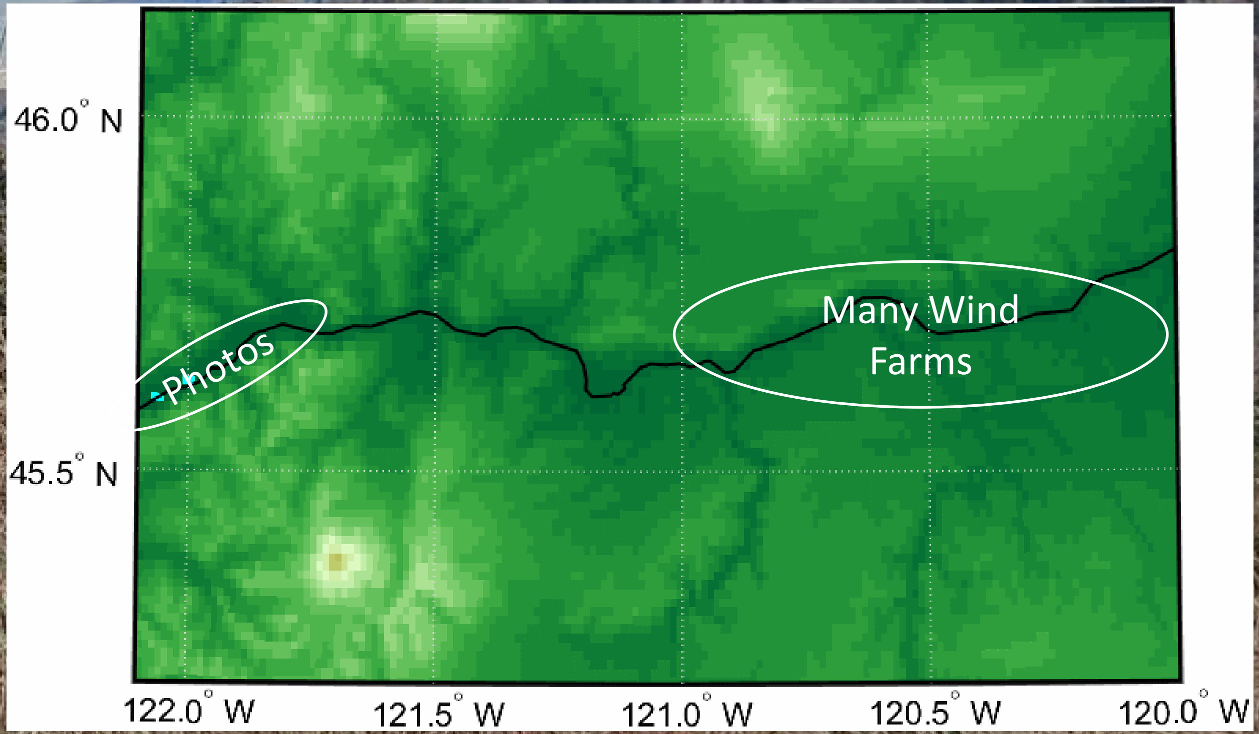
USGS 15 arcsecond topography



Understanding Grid Spacing (Resolution)

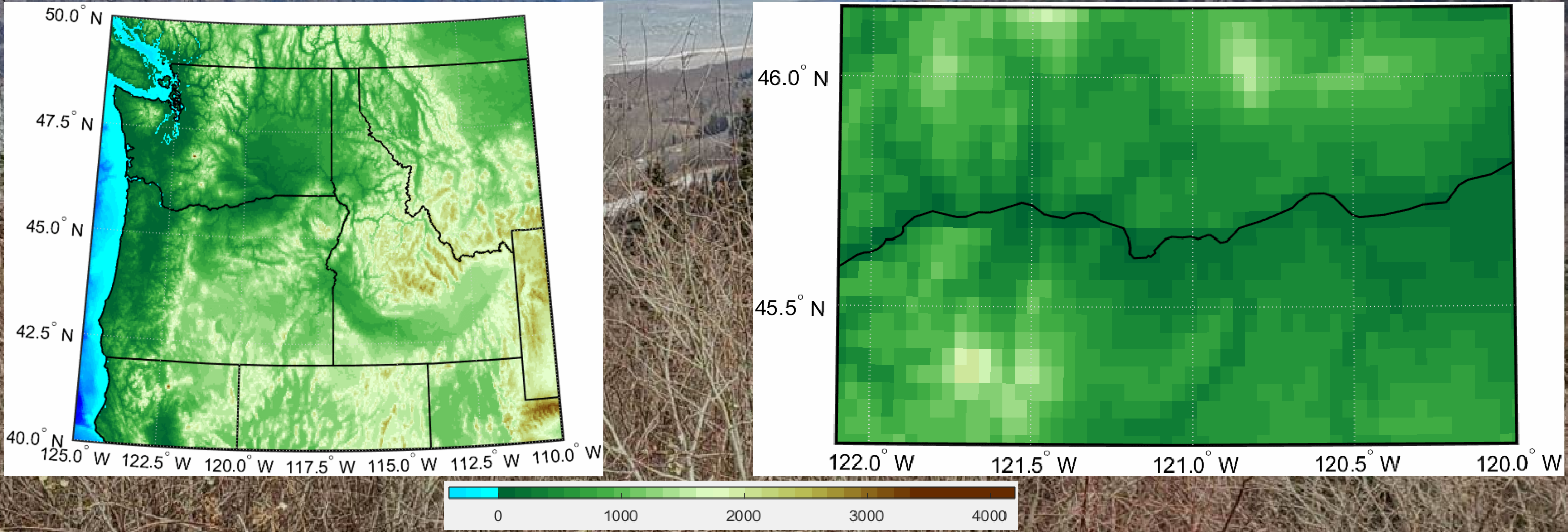
1 km Grid Spacing (WINDTK is 2 km)





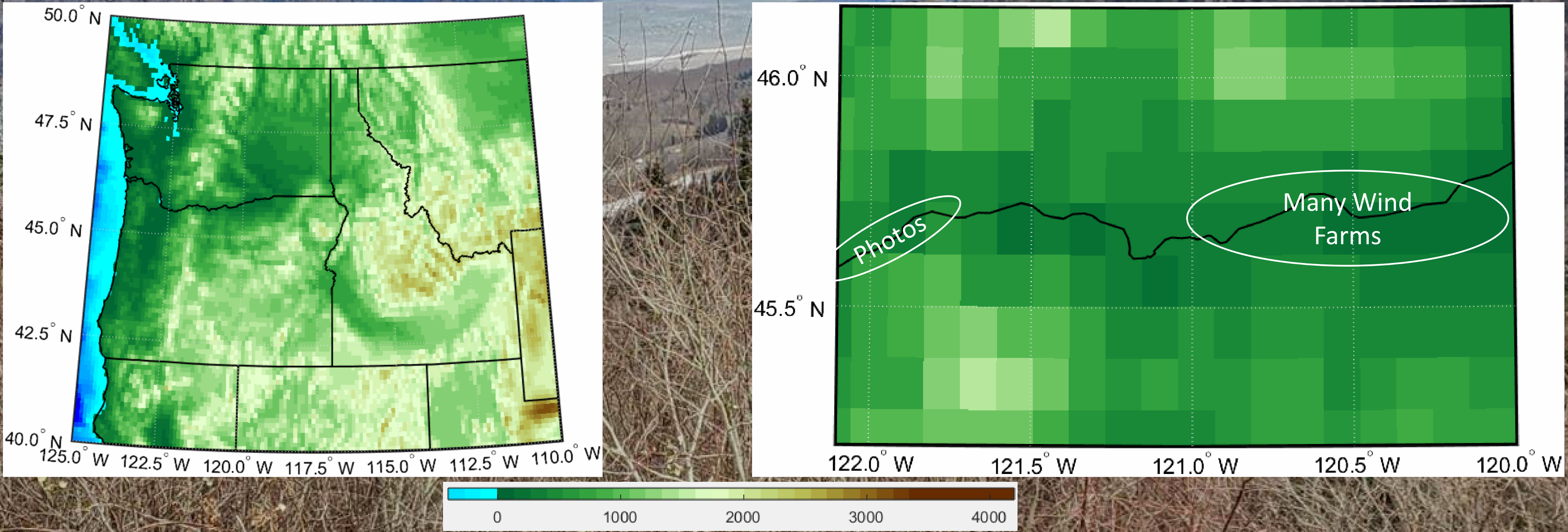
Understanding Grid Spacing (Resolution)

3 km Grid Spacing (~HRRR)

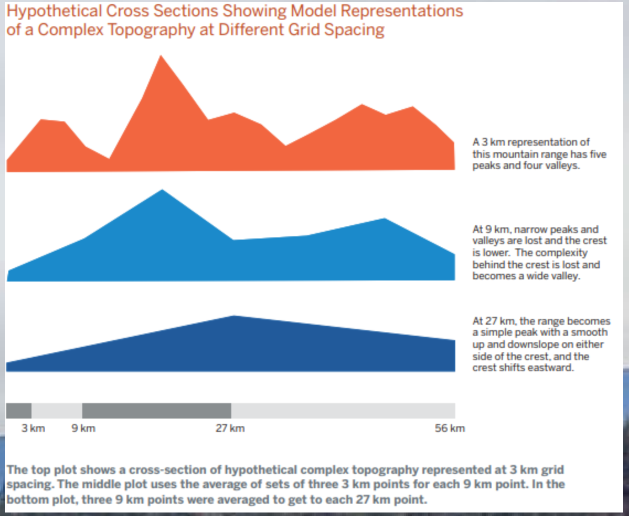


Understanding Grid Spacing (Resolution)

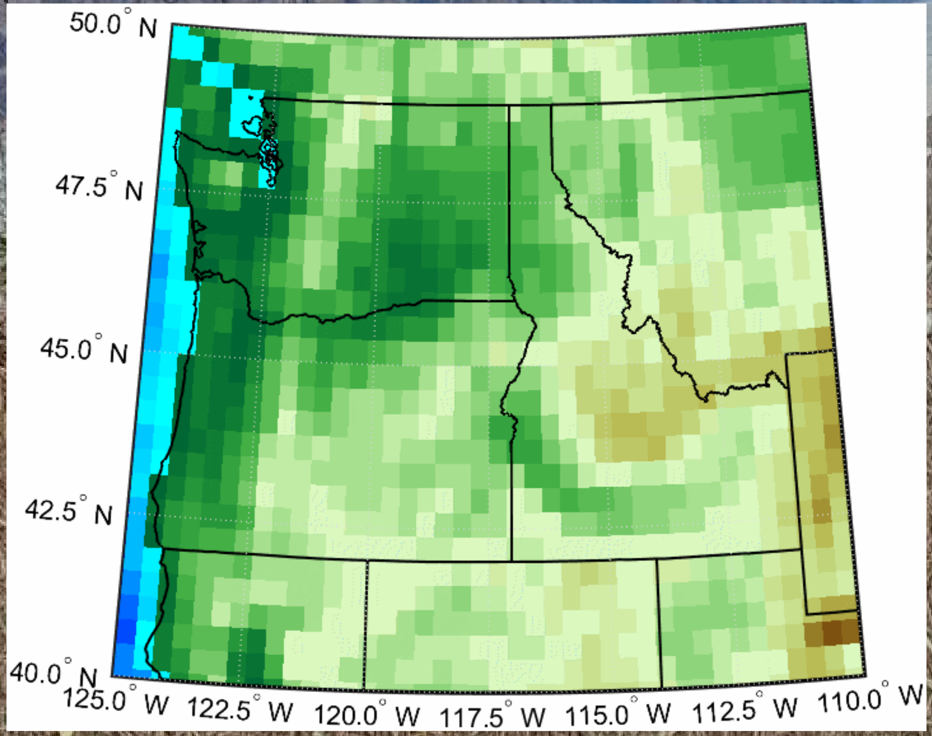
9 km Grid Spacing (~ECMWF)

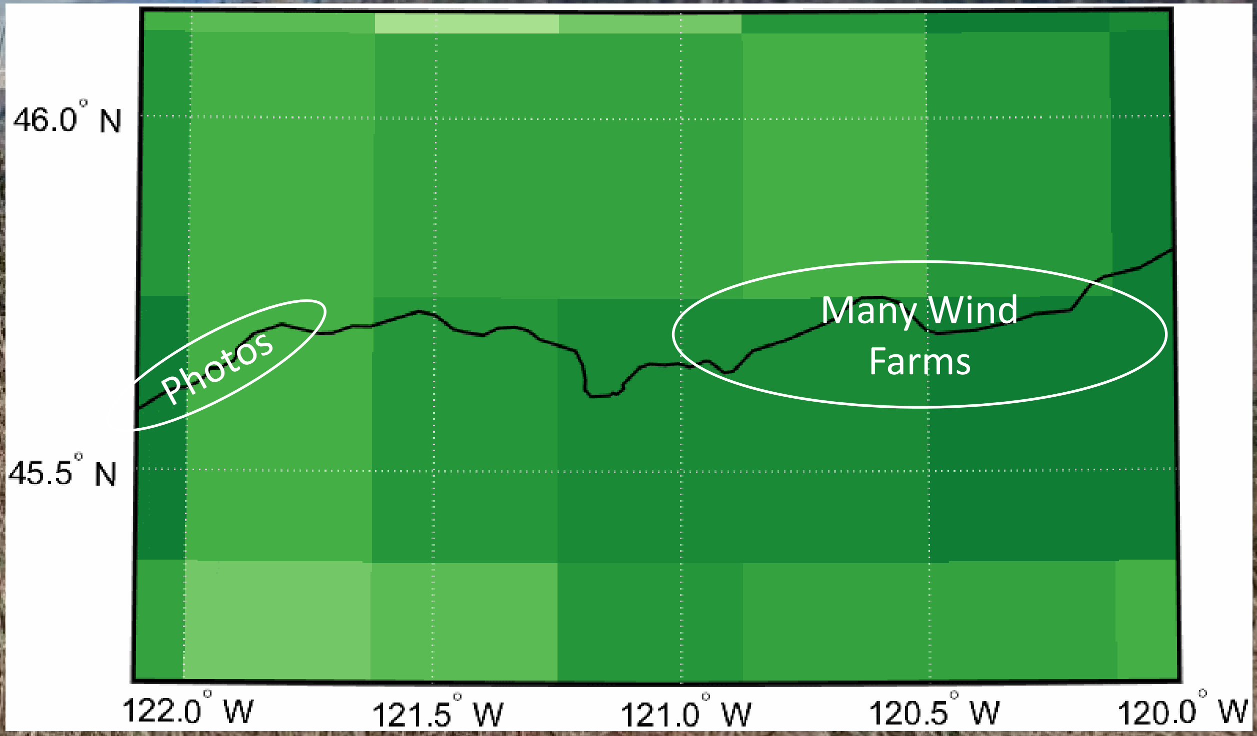


Understanding Grid Spacing (Resolution)

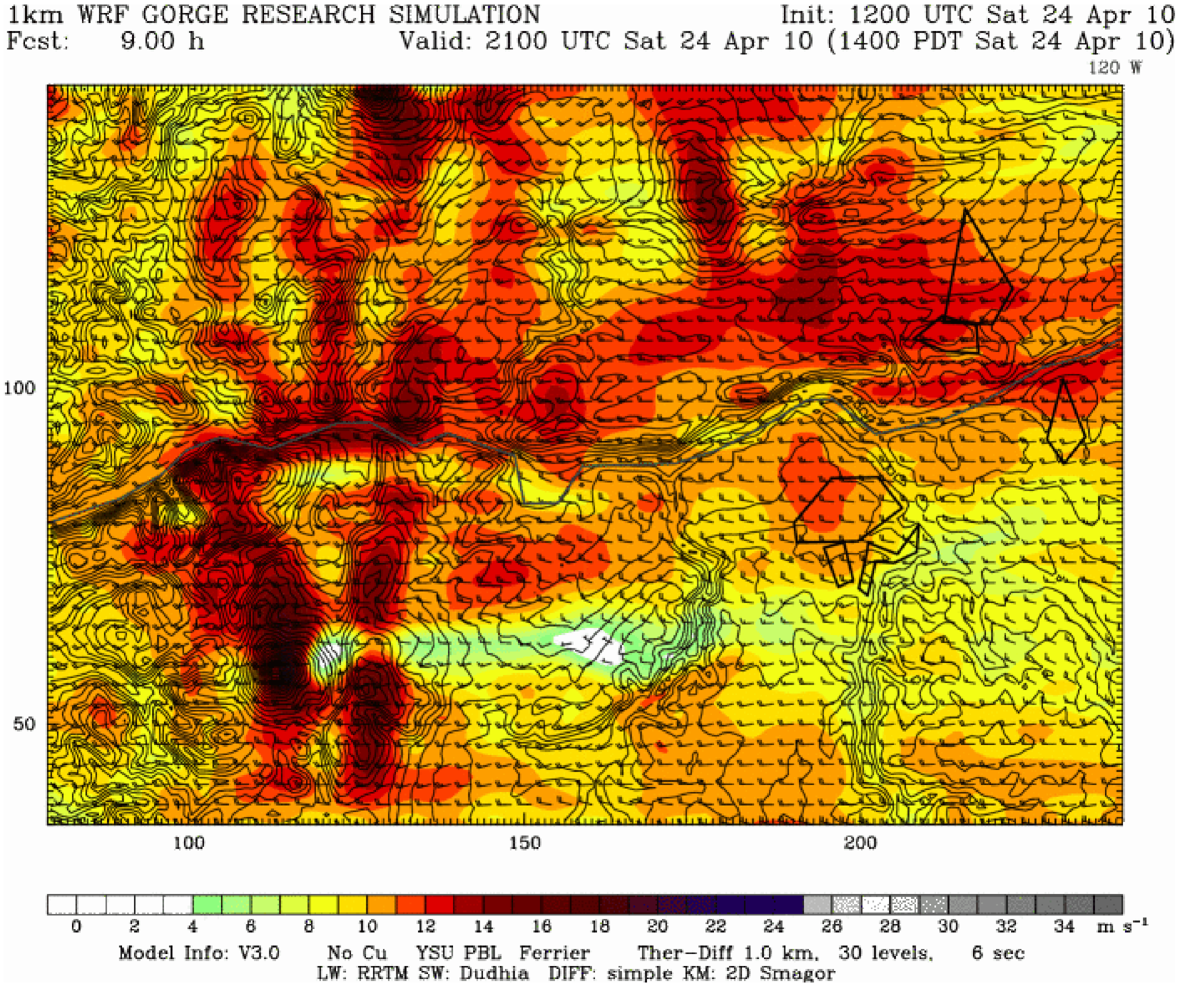


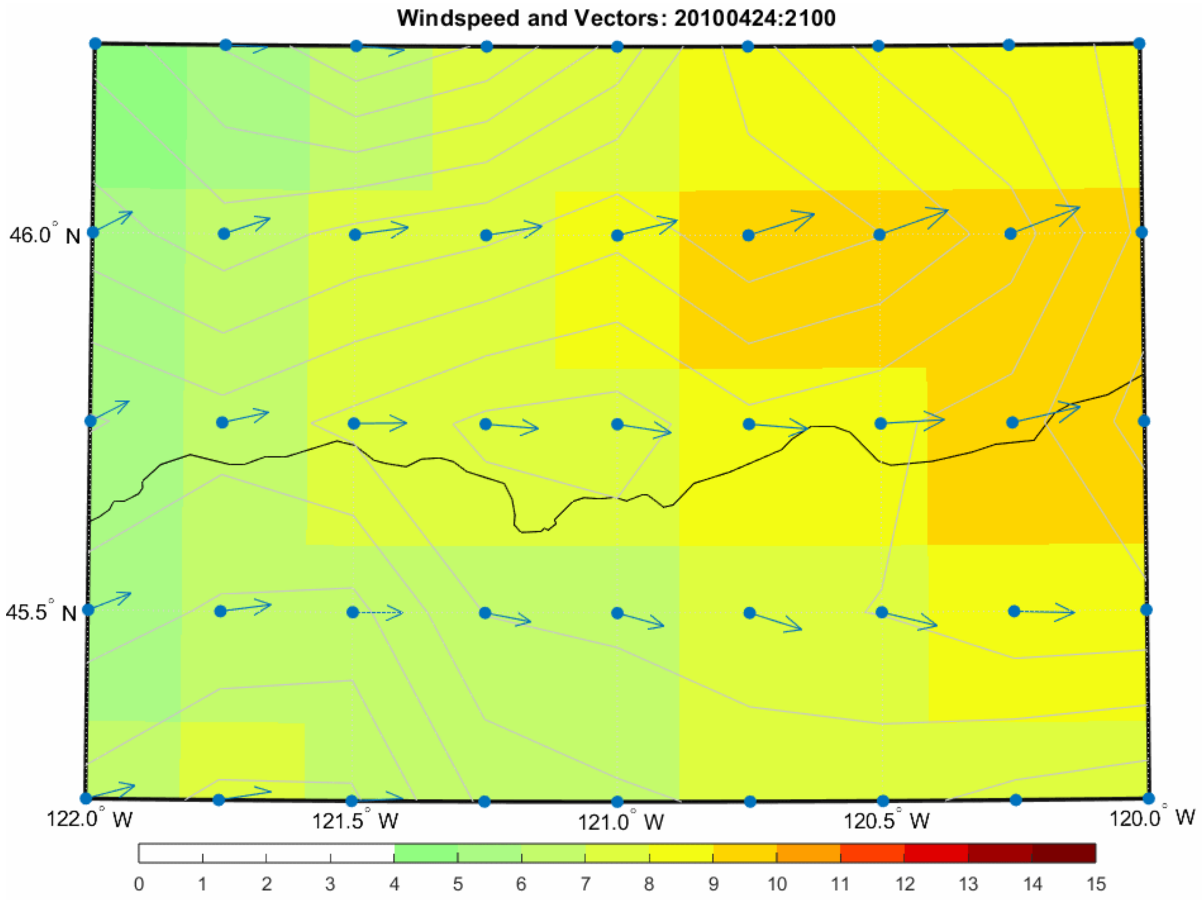
30 km Grid Spacing (~ERA5)





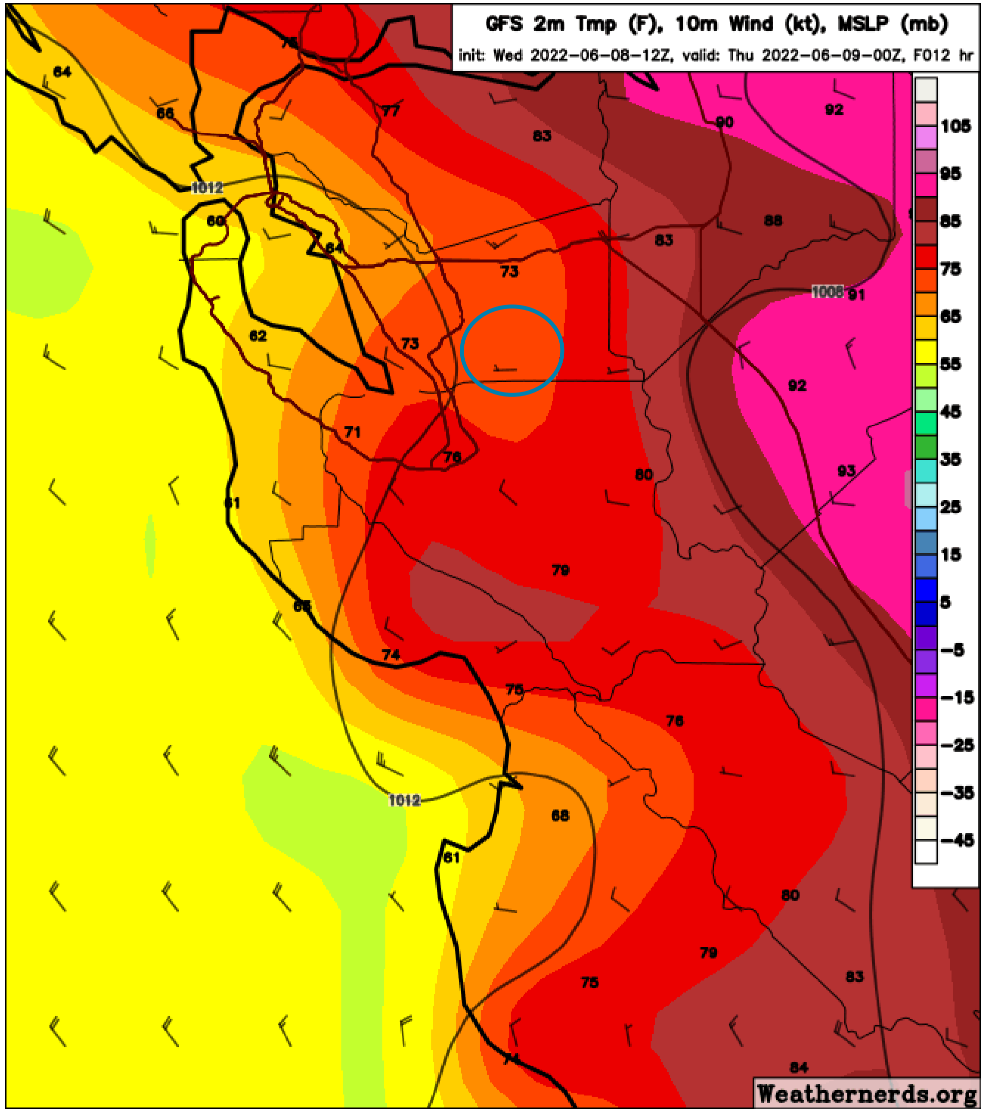
An Example of Resolution Impacts





Weather datasets For Power Systems Should…

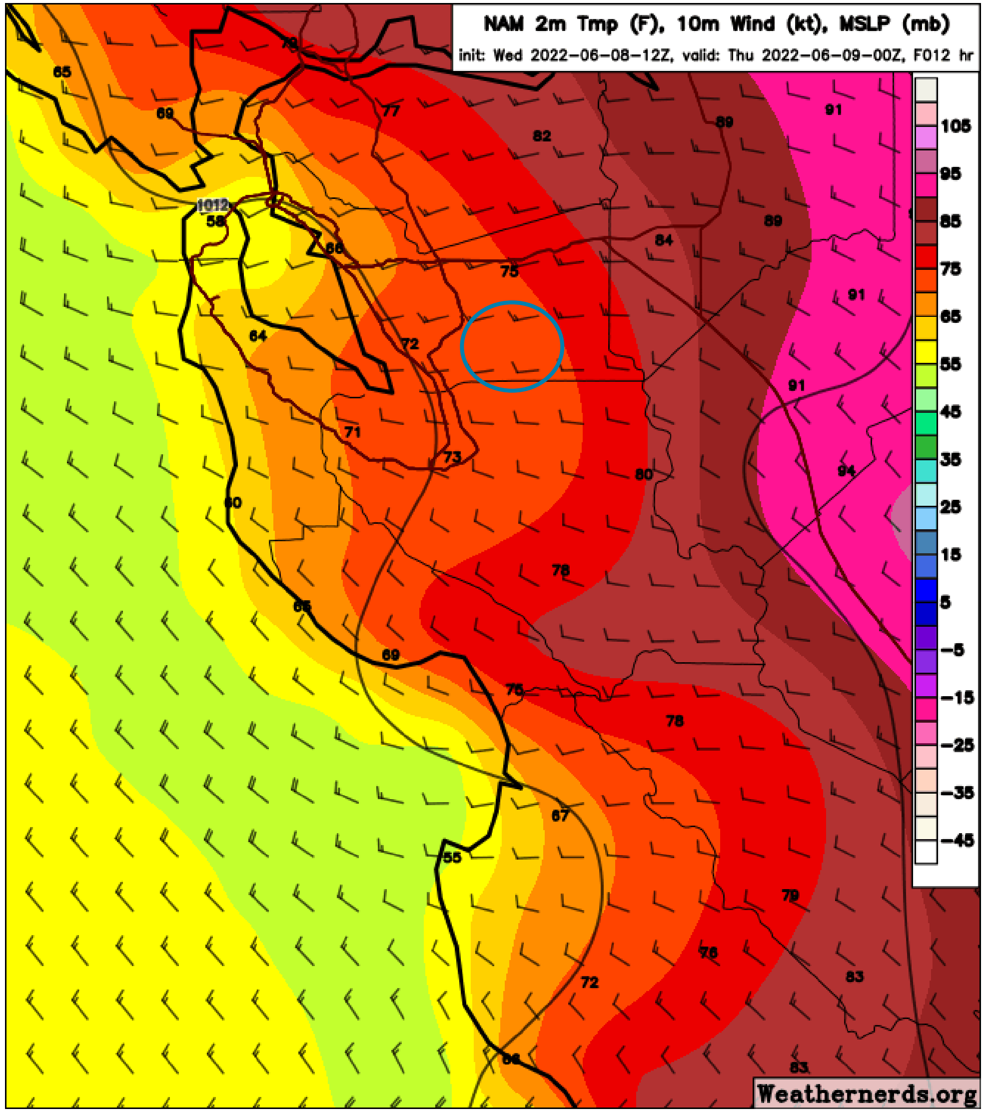
- …have spatial and temporal scales relevant to the system being modeled. See example on right.
  - Accurately capture the resource drivers and their variability
  - Capture the uncertainty in forecasts of the resource drivers
  - Do the same for drivers of system load
- …



Not All Data Is Created Equal. Model Data ≠ Observations

Weather datasets For Power Systems Should…

- …have spatial and temporal scales relevant to the system being modeled. See example on right.
  - Accurately capture the resource drivers and their variability
  - Capture the uncertainty in forecasts of the resource drivers
  - Do the same for drivers of system load
- …be concurrent and physically consistency
  - All variables represent the same time chronology and be dynamically consistent
  - How important is chronology and dynamic consistency?
    - NSRDB and WINDTK, and example on the right provide synchronous variables but they may lack dynamic consistency
- …

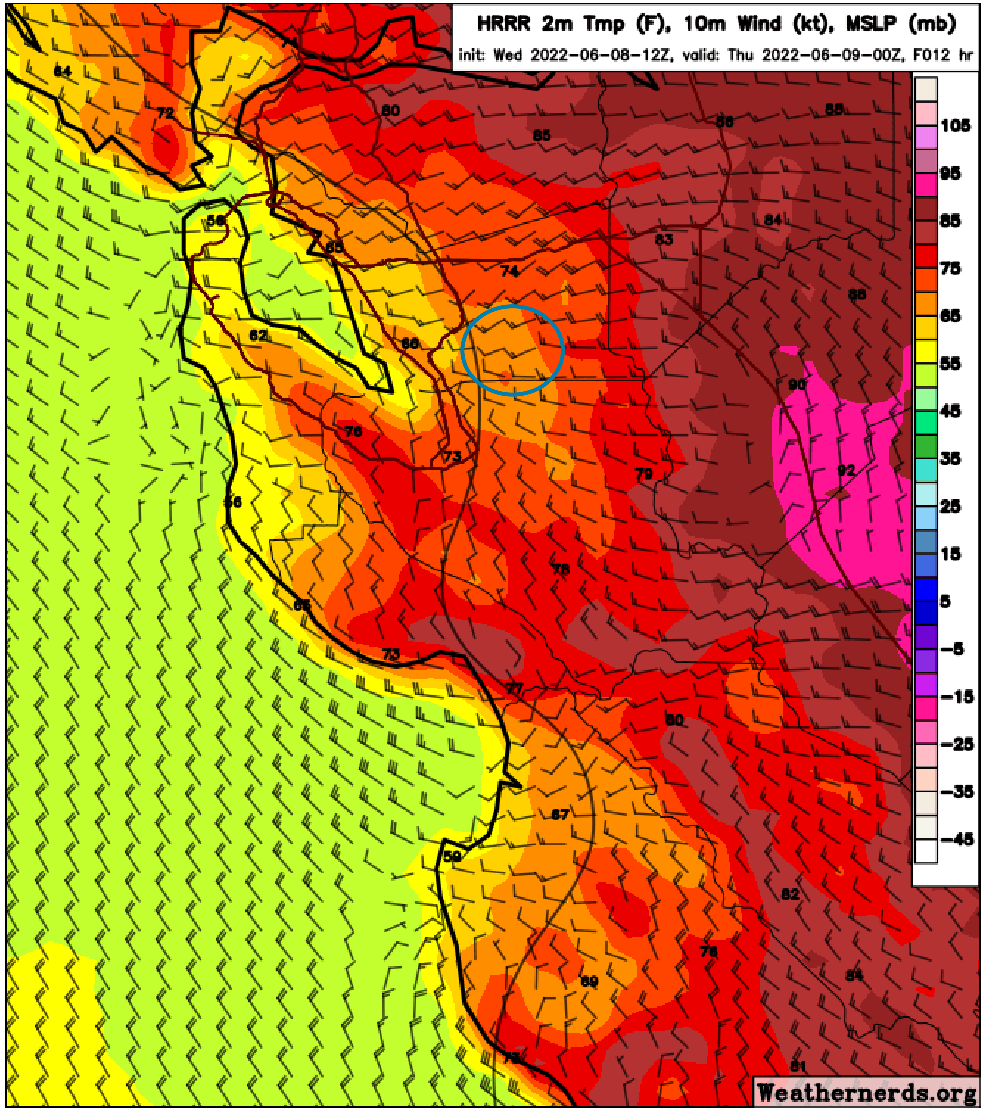


Not All Data Is Created Equal. Model Data ≠ Observations

EPRI

Weather datasets For Power Systems Should…

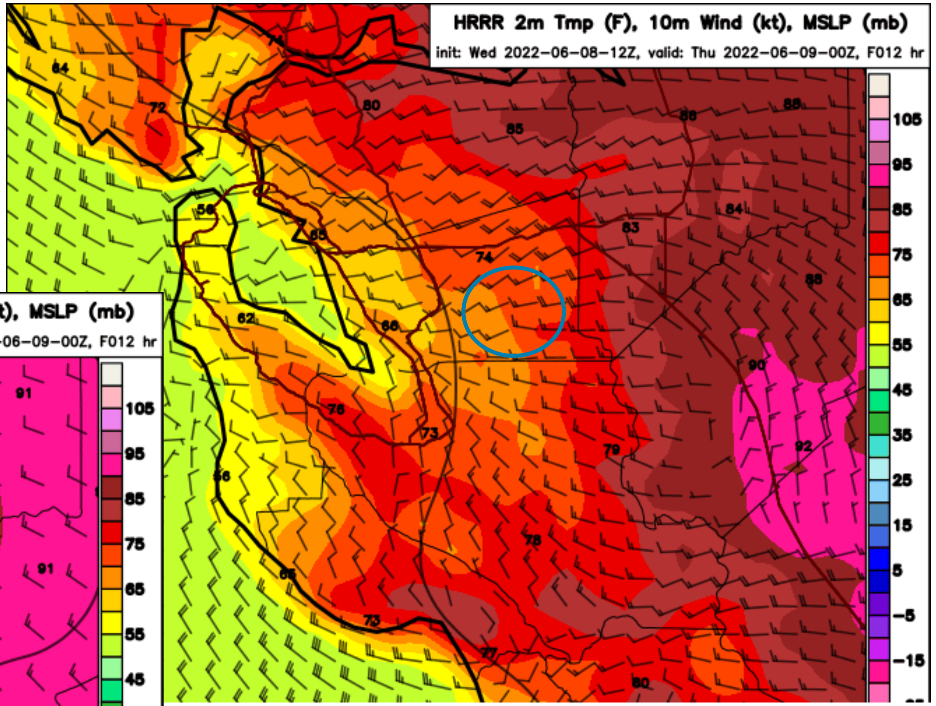
- …have spatial and temporal scales relevant to the system being modeled. See example on right.
  - Accurately capture the resource drivers and their variability
  - Capture the uncertainty in forecasts of the resource drivers
  - Do the same for drivers of system load
- …be concurrent and physically consistency
  - All variables represent the same time chronology and be dynamically consistent
  - How important is chronology and dynamic consistency?
    - NSRDB and WINDTK, and example on the right provide synchronous variables but they may lack dynamic consistency
- …provide a 30+yr, updated time history with a consistent methodology that minimize biases, trends, and “artifacts”

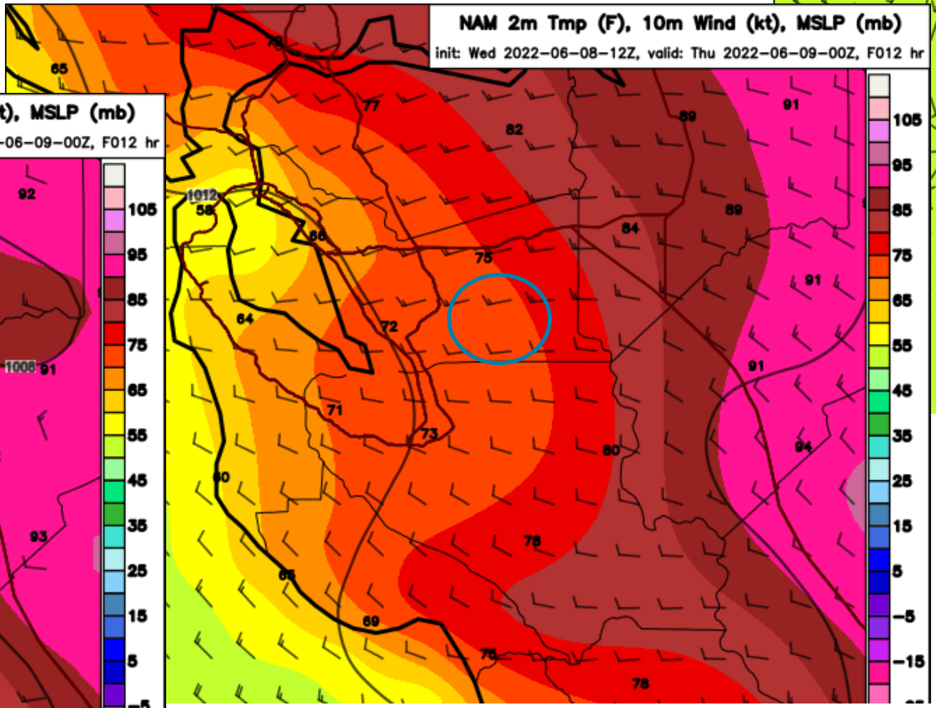


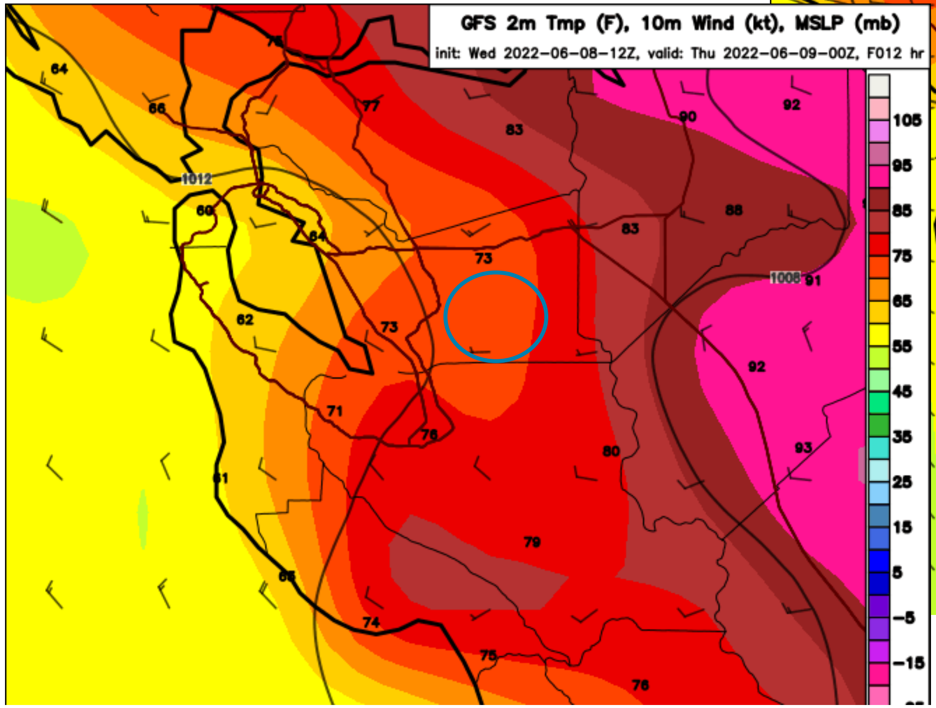
Not All Data Is Created Equal. Model Data ≠ Observations

Weather datasets For Power Systems Should…

- …be validated and have uncertainty quantified.
- Have you validated the data you use?
- Which of these representations is right? Most useful?
  - They can't all be right!







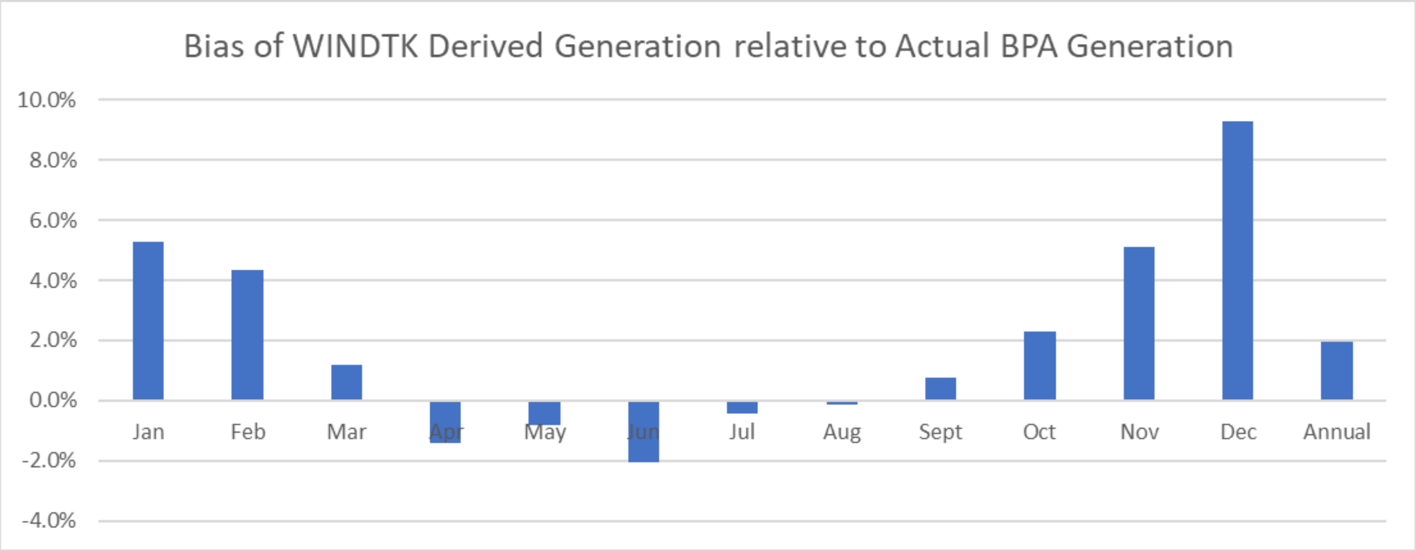
- 50's, 60's, 70's or 80's?
- Light, moderate, strong wind?
- Cloudy or sunny?

Not All Data Is Created Equal. Model Data ≠ Observations

(Use Case Specific) Validation and Uncertainty Quantification

- We must validate according to the use case. E.g. For RA, the distributions, and especially the tails, matter more than the averages
- The distribution of coincident tail events MUST be close to reality
- Example here:
  - WINDTK data in the BPA area
  - Wind resource in BPA BA is notoriously difficult to predict with NWP => WFIP2 Project
    - Complex terrain that needs a minimum of 1.33 km resolution to resolve
    - Stable boundary layer issues in the wintertime. => Low wind AND high load

These biased low wind speed events frequently coincide with high load events due to regional mesoscale meteorology



Tail event deviations can be >7x.
e.g. BA wide generation of 3% and model-based estimates of 23%!

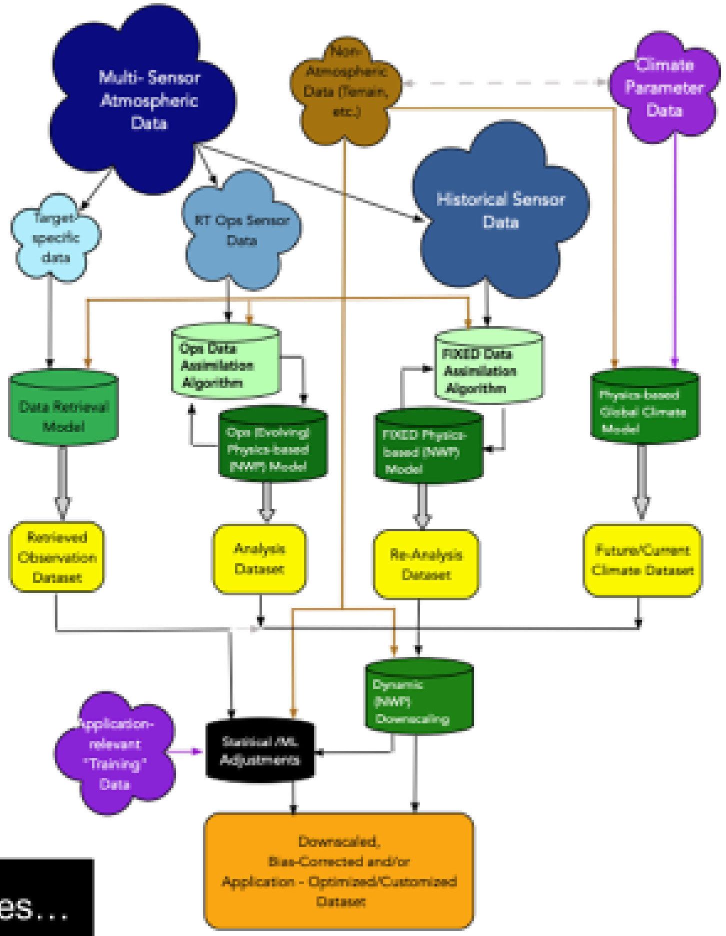
Low Hanging Fruit for Validation and Uncertainty Quantification

- Comprehensive industry wide data transparency and sharing is required: Met., generation, and availability data
  - Little proprietary value per site but a tremendous untapped asset if made public across all generators
- This will enable validation and UQ of synthetic datasets which is imperative for valid application. Ground truth data is also key to the model improvement process
- ERCOT is leading the way. Others should follow ASAP
  - This might require legislation/regulation.

User Knowledge/Education

Overview of the Current World of Datasets for Power System Planning

- ❑ **Wide Range of Methods to Construct Datasets**
  - **A few fundamental types of approaches**
  - Enormous number of significant variations within types
- ❑ **Therefore: Wide Range of Datasets Exist**
  - **Typically have very different attributes depending on how they were constructed**
  - Consistency of data attributes (e.g. spatial/temporal correlations) between datasets should not be assumed
  - **Critical need to evaluate comparative performance on parameters/scales important to specific applications**



Let's examine the key attributes of the fundamental types of approaches…

Slide courtesy of:

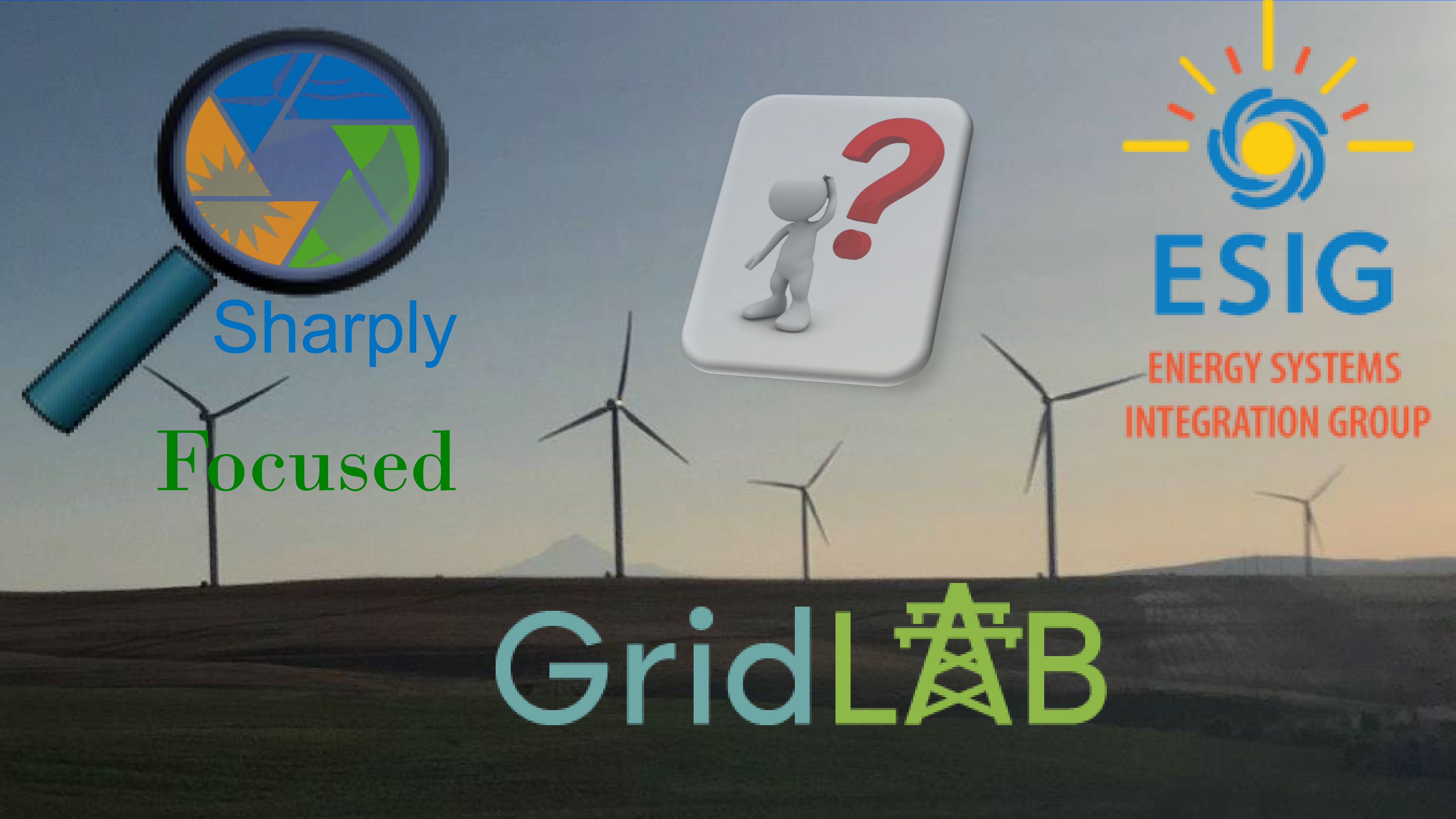
Dr. John W. Zack
Principal
MESO, Inc.
Troy, NY

MESO ATMOSPHERIC RESEARCH FORECASTING • SIMULATIONS

Used with permission.

It's Complicated!!!

EPRI


Sharply
Focused
ESIG
ENERGY SYSTEMS
INTEGRATION GROUP
GridLAB