Detailed Siting Enhancement of MISO High Penetration Wind, Solar and Storage:
Dataset Methods & Analysis

Prepared By:

Vibrant Clean Energy, LLC

Prepared For:

Midcontinent Independent System Operator

Table of Contents

1. VCE® Background

Vibrant Clean Energy, LLC (VCE®) is a Colorado company that has positioned itself as a world-class provider of renewable energy assessment and energy optimization studies. VCE®, since its beginnings, has focused on providing the analytical underpinning for the energy transition underway across the world. The team at VCE® have provided support to the private and public sectors enabling more intelligent implementation of energy resources onto the electricity grid.

The primary mission of VCE® is to provide clients with the least-cost pathways to fulfill their particular needs. The least-cost pathways can be benchmarked against sensitivities to assess the impacts of alternative options. VCE® has expertise on Renewable Energy (RE), Energy Efficiency (EE), electric/thermal energy storage, system integration, Electric Vehicles (EVs), sector electrification, natural gas markets, economics, software development, policies and regulations, and big-data analytics.

VCE® is led by founder and CEO Dr Christopher T M Clack, who has a background in mathematics, statistics and renewable energy modeling. He has been building energy grid integration models for almost a decade with a strong interest in agnostic cost co-optimization. All the models that Dr Clack has created are constructed from the ground up to incorporate high-resolution weather and load data. In a nutshell, the models are designed to deal with big data.

The flagship model is known as WIS:dom® [Weather-Informed energy Systems: for design, operation, and market optimization]. It is the successor to the C-OEM suite. It is the first, and only, commercially available combined capacity expansion and production cost model that can solve for the entire North American grid, while considering variable generation, at 5-minute 3-km resolution, transmission power flow, generator physical limitations, retirements, and yearly investment periods. It, also, simultaneously solves for electricity storage, electric demand, sector electrification, and fuel markets and supply.

2. Purpose of Study

Study Objectives:

- The creation of a consecutive three-year weather and power dataset for wind and solar generators at 3-km, 5-minute resolution across the entire Eastern Interconnection (EIC) footprint;
- Description of the process and metrics surrounding the dataset creation;
- Perform detailed modeling of generator, storage and transmission siting using the WIS:dom® optimization tool for the MISO and EIC electricity grids as they transition to extremely high penetration levels of variable renewable energy (VRE), electric vehicles (EV), distributed solar (DPV) and electric storage;
- Compute the potential impacts of transmission and storage on the siting of the VREs;
- Address the issues with extremely high levels of VREs within MISO footprint;
- Calculate the fuel burns in each region of MISO under all scenarios.

Study Approach:

1. Create a high-resolution wind and solar dataset that covers the entire contiguous United States. The dataset will consist of three chronological calendar years and has a granularity of 3-km grid spacing with 5-minute time steps. The input data will be of the highest quality available and the power algorithms will contain state-of-the-science methods.
2. Provide the full dataset to MISO for use in all their future work.
3. Document the processes and methods for MISO and its stakeholders (present document).
4. Incorporate the high-resolution power datasets into WIS:dom® for production cost analysis at 5-minute frequency.
5. Perform WIS:dom® optimizations on twenty scenarios (each with ten investment cycles to reach their final target) to create portraits of the evolution of the MISO and EIC electricity grid under deep penetration levels of VREs.
6. Compose a summary of the modeling results to determine salient features and observations that the modeling displays (the companion document to the present one).
7. Feed the VRE siting decisions from the WIS:dom® scenarios into the MISO MTEP process.

3. Overview of the Power Dataset

As the integration of VREs into the electricity grid grows, it is increasingly important to have accurate estimates of expected electricity generation from these sources. Accurate estimates of power output from renewable energy generators depend on two factors: (1) accurate forecasts of the meteorological variables that affect renewable energy generation and (2) realistic modeling of the renewable energy generators. VCE® strives to achieve both of these requirements for the present study, as will be described in the following sections.

VCE® provides a normalized power dataset for both wind and solar technologies for various weather years based on the National Oceanic and Atmospheric Administration's (NOAA) High Resolution Rapid Refresh (HRRR) weather forecast model. The power dataset is the best available estimate of what the synchronous wind and solar power profiles looked like across the contiguous United States (CONUS). These are provided on a calendar year basis, gridded spatially at 3km and temporally at five minutes. The calendar years originally provided to MISO were for 2014, 2015 and 2016. In late 2019, MISO requested two additional calendar years (2017 and 2018) to complete their data set for five chronological years from 2014 through 2018.

The input weather data is obtained from the NOAA HRRR weather forecast model, which is a specially configured version of Advanced Research WRF (ARW) model. The HRRR is a run hourly on a 3-km grid resolution and its domain covers the continental United States as well as portions of Canada and Mexico. Since its inception, the HRRR has undergone rapid and continuous improvement to its physical parameterization schemes, many of which have specifically targeted improved forecasts for the renewable energy sector. Through collaborative research efforts between Department of Energy (DOE) and NOAA, projects such as the Solar Forecast Improvement Project (James et al. 2015, Benjamin et al. 2016), the Wind Forecast Improvement Projects I and II (Wilczak et al. 2015, Shaw et al. 2019) were conducted to improve forecasts of meteorological quantities important for wind and solar energy forecasting.

Weather not only influences the production of renewable energy, but also impacts demand. Weather events like the "*Polar Vortex*" of 2014 can result in reduced (or zero) generation of energy from wind and solar generators, while increasing energy demand (for heating, increased frequency of EV charging, and so on). These events can last several days and it is important to model the chronological nature of these events on the energy grid to ensure reliability. In addition, it is important to perform the modeling using multiple weather years to account for the inter-annual variability of weather patterns. For example, the jet stream remained further north than usual in 2015. This brought well below normal mean wind speeds across the whole of the CONUA that came to be known as the "*wind drought of 2015*". Years like these can result in a severe deficit of wind energy production especially during summer nighttime periods (for example, lower intensity low level jets across the great plains) when demand for cooling will be high. Providing multiple years of the power dataset allows insight into how the weather changes year-on-year.

4. Wind Power Dataset Method

The amount of wind power produced from a wind turbine is proportional to the cube of the wind speed directed into the wind turbine, perpendicular to the rotor. As wind turbine hub heights have grown and the

rotor swept areas have increased, different sections of the wind turbine rotor are exposed to (sometimes significantly) different conditions of wind speed, temperature, air density, and precipitation. Therefore, it is very important to model the impact of the change of these meteorological conditions with height on wind power production. The impact of change in meteorological quantities with height on wind power production is modeled using the rotor equivalent formulations (see e.g. Clack et al. 2015 and Choukulkar et al. 2015). The details of the wind power calculations are described below.

The wind power dataset is created using the HRRR weather forecasts available from NOAA. The model outputs are available for the following forecast hours: 00 (also known as initialization or assimilation), 02, 06, and 12. The forecast hour 02 output has been found to be the most accurate (when compared to observations), and is used for the power calculations[1]. The following variables are used to create the wind power dataset:

(1) Horizontal components of wind (u, v)
(2) Pressure (P)
(3) Temperature (T)
(4) Specific humidity (spH)
(5) Geopotential height (GPT)
(6) Cloud-water mixing ratio (cwr)
(7) Rain-water mixing ratio (rwr)
(8) Wind gust at lowest level (WG)

The above variables are output by the model on three different vertical coordinates: (1) Pressure coordinates, (2) Terrain following sigma coordinates, (3) Hybrid vertical coordinates. The hybrid vertical coordinate was found to mitigate the small-scale noise found near steep terrain, while having better vertical resolution than the pressure coordinates. Figure 1 shows an example horizontal transect through complex terrain from a NOAA presentation. It can be seen that the vertical velocity fields are much more realistic in magnitude and less noisy in the hybrid coordinate.

---
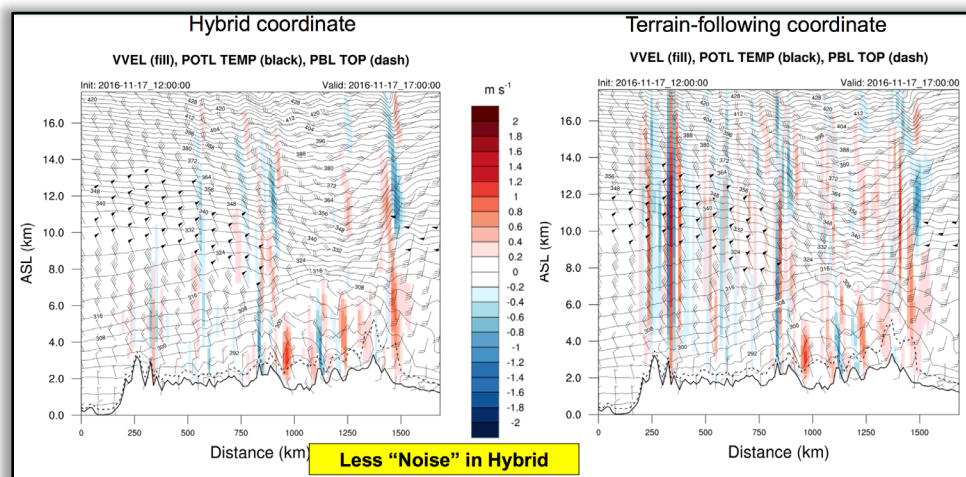
[1] During periods where forecast hour 02 is missing, forecast hour 00 is used and for periods where forecast hour 00 is also missing, we fall back to forecast hour 06.

The model outputs do not include density and needs to be calculated. Density is calculated using a modified formulation of the ideal gas law

$$\rho = \frac{P_o^{0.2854} P^{(1-0.2854)}}{R T_v},$$

(1)

where R is the specific gas constant (287.058 J.kg^{-1}.K^{-1} for dry air) and T_v is the virtual potential temperature, which is calculated using the formula

$$T_v = \frac{T}{\left(P/P_o\right)^\kappa} \left(1 + 0.61 spH - (cwr + rwr)\right),$$

(2)

where P_o is the standard pressure which is 10^5 Pa and κ is the Poisson constant given by

$$\kappa = 0.2854 * (1 - 0.24 * spH).$$

(3)

The modified formulation is used because we need to account for the buoyancy effects of change in temperature and pressure with height above the ground. This effect is captured by the virtual potential temperature calculated using Eq. (2). The horizontal wind speed components (u, v), density, temperature and cloud-water mixing ratio are now interpolated to heights 20 m above ground to 300 m above ground with 15 m vertical resolution. The wind gust outputs are used to calculate the gust factor expressed as a fraction of the mean wind speed.

Rotor Equivalent Calculations:

Potential wind turbine power generation is given by the kinetic energy flux through the wind turbine rotor layer. This general relationship is shown in Eq. (4)

$$P_w = \frac{1}{2} C_p \rho A U^3$$

(4)

where C_p is the coefficient of power (ratio of actual power generated to available power in the wind), ρ is the air density, A is area of the wind turbine rotor and U is the horizontal wind velocity component along the horizontal axis of the wind turbine rotor. The vertical component of the velocity does not contribute appreciably to wind power production as modern wind turbines use aerodynamic lift for propulsion. In addition, any drag-component due to vertical velocity is assumed to cancel out over the rotor swept area. Equation (4) is valid as long as the density and velocity do not change within the rotor swept area of the wind turbine. However, modern wind turbine rotors can span vertical extents of 100 m or more and the variables that impact wind power production can change substantially within this vertical extent (Figure 2).
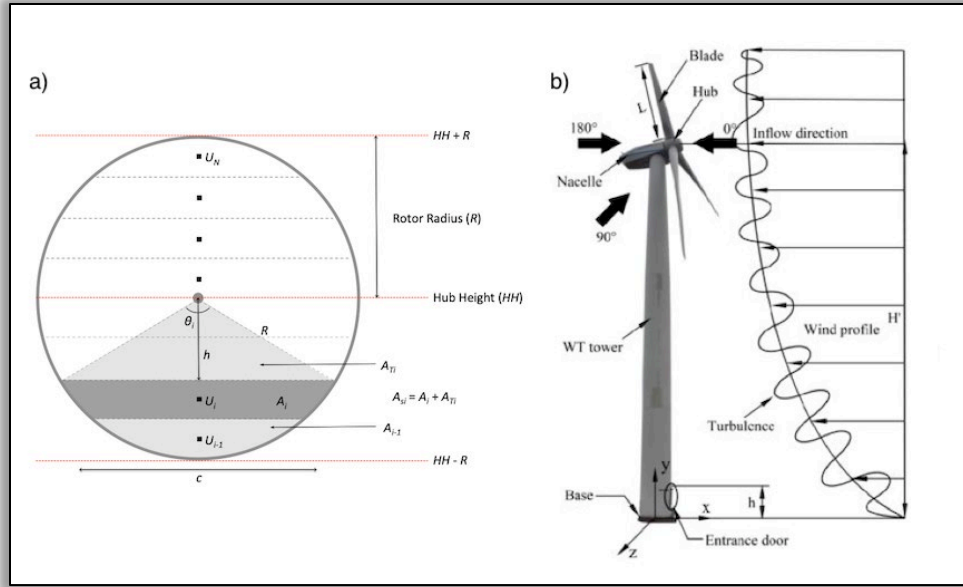
Figure 2: Schematic showing how the wind profile can change within the rotor layer and how the wind turbine rotor is divided to calculate the rotor equivalent variables.

The rotor equivalent formulation provides a more accurate estimate of wind power production by taking into account the vertical profile of the variables affecting wind power production. At its core, the rotor equivalent formulation calculates an area weighted mean of the various meteorological variables across the wind turbine rotor. The rotor area is divided into sections with respect to height equal to the vertical resolution of the dataset (15m). Then each variable value at a given height is area weighted by the portion of the wind turbine rotor it represents (equal to the vertical resolution) in order to estimate the equivalent effect of the vertical profile of that variable. Equations (5) and (6) show how the speed and density in Eq. (4) can be replaced by their rotor equivalent counterparts.

$$U_{eq} = \frac{1}{A}\sum_i \frac{u_i u_H + v_i v_H}{U_H} A_i \tag{5}$$

$$\rho_{eq} = \frac{1}{A}\sum_i \rho_i A_i \tag{6}$$

In Eq. (5), the effect of turbulence is neglected. The effect of turbulence can be included in the power calculations using Eq. (7) in Choukulkar et al. 2015 re-arranged here:

$$U_{eqT} = \frac{1}{A}\sum_i \frac{(u_i + u_i')(u_H + u_H') + (v_i + v_i')(v_H + v_H')}{[(u_H + u_H')^2 + (v_H + v_H')^2]^{1/2}} A_i \tag{7}$$

where $(.)'$ denotes tendency of that variable in a given time period (5-min in our case). Equation (7) shows that the effect of turbulence results in additional wind power being generated, which makes sense analytically as turbulence represents additional energy in the wind. However, actual wind turbine response to turbulence results in additional power generated at the lower end of the power curve (due to the additional energy) and under-performance at the higher end of the power curve due to the positive velocity fluctuations being damped by the wind turbine control (Wharton & Lundquist 2012). This effect of turbulence on power production due to wind turbine control can only be modelled through a full mechanical modeling of the wind turbine as is done by NREL's FAST software. In our analysis, this effect of

turbulence is neglected as it is found to be much smaller compared to the effect of wind speed and direction shear (Choukulkar et al. 2015).

The rotor equivalent technique allows us to take into account change in density with respect to height as well as wind speed and direction shear on wind turbine power potential calculations. The rotor equivalent formulation is also applied to the temperature and moisture information as shown in Eqs (8) and (9). The rotor equivalent temperature and moisture information is used to determine icing possibility within the wind turbine rotor.

$$T_{eq} = \frac{1}{A} \sum_i T_i A_i \qquad (8)$$

$$cwr_{eq} = \frac{1}{A} \sum_i cwr_i A_i. \qquad (9)$$

The rotor equivalent quantities are then linearly interpolated to 5-min intervals for each of the HRRR grid cells. The linear interpolation also covers any possible periods of data outages. These 5-min rotor equivalent quantities are used in the power calculations. In order to calculate actual power generation from the theoretical available power in the wind, a C_p curve is used. A C_p curve is the ratio of the actual electrical power generated for a given wind speed to the theoretical available power in the wind given by Eq. (10):

$$C_p = \frac{P(u)}{P_{wind}(u)}. \qquad (10)$$

As the wind interacts with the wind turbine, it loses momentum because it is used to drive the turbine blades and produce electricity. This loss is momentum from energy extraction results in the control volume of the wind passing through the turbine to expand downstream of the rotor to preserve continuity. Based on this, Betz (1927) calculated the theoretical maximum energy that can be extracted by a wind turbine. This theoretical maximum, called the Betz limit, is equal to 59.3% and is the maximum value a C_p can take.

The C_p curve varies for different types of wind turbines. The International Electrotechnical Commission (IEC) described four classes for wind turbines: Classes I, II, III and offshore. The shape of the C_p curve is defined not only by the physical limits on converting wind power to electricity, but also the control strategies employed by the wind turbine. The C_p curves for the various wind turbine classes are shown in Figure 3. The C_p curve for the various wind turbine classes are composited and fit with a 12th order polynomial, shown as bright blue line in Figure 3. This composite C_p curve is used to calculate actual power production. It is assumed that this derived C_p curve is valid for the rotor equivalent wind speed and for the observed changes in density.
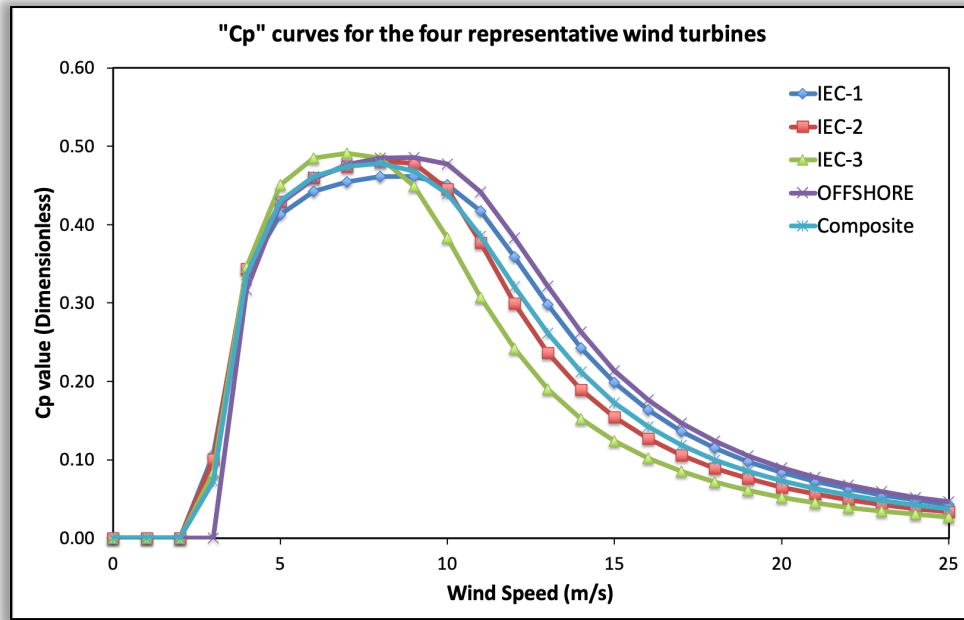
Figure 3: Coefficient of power curves for the various IEC class turbines.

WIS:dom® takes into account periods where generation may not be possible due to weather conditions. Normal operational temperatures for wind turbines are set to between -25°C and 45°C. In addition, potential for icing is also calculated. Icing is considered possible when temperatures are below -15°C and cloud-water mixing ratio is greater than zero. The periods with potential for icing or temperatures outside of normal operating conditions are set to zero power output. It is important to identify periods such as the above where generation will be limited or zero as these are usually correlated with periods of high energy demand. WIS:dom® then has to ensure that the demand during these periods will be met in some other way.

4.1. Updates to the Wind Power Methods (2019)

In order to further improve the accuracy of the wind power calculations, the methodology was updated to include the following two additional components:

a) The impact of turbulence on power generation
b) Wind turbine response to changes in air density

4.1.1. Introducing effect of turbulence

As described before, turbulence has a complicated impact on the wind power generation. At wind speeds closer to the cut in wind speed, presence of turbulence increases power generated from the turbine while at wind speeds closer to the rated power, turbulence reduces the power output from a wind turbine (Wharton and Lundquist 2012). The reason for this behavior is that near cut-in speed, the positive fluctuations due to turbulence are allowed to generate excess power, while the negative fluctuations do not have any effect as the turbine is not generating any power in that case. Near the rated wind speed, the positive fluctuations due to turbulence get damped out by the wind turbine control, while the negative fluctuations reduce power output and hence the net effect is a reduction in power output from the turbine.

Modelling this effect analytically is difficult as seen from Eq. (7) where presence of turbulence always results in increased power production. Therefore, instead of trying to model this effect analytically, it was decided to utilize the characteristics of the *Cp* curve to simulate the wind turbine control response. To do this, the model wind speed output needs to be perturbed in a manner that actual atmospheric turbulence would as shown in Eq. (11), known as the Reynold's decomposition:

$$u(t) = \bar{u}(t) + u'(t), \tag{11}$$

where, *u(t)* is the east-west component of wind speed including effect of turbulence at given timestep,
$\bar{u}(t)$ is the mean east-west component at a given timestep from the HRRR model,
$u'(t)$ is the random turbulence perturbation at that timestep.

The perturbations that need to be added to the model wind speed are estimated using the wind gust model output. The model estimate of wind gust represents a sudden, brief increase in peak wind speed (lasting less than 20 seconds) expected at a given timestep. It is now required to estimate the standard deviation of turbulence from this peak value. Assuming that the turbulence distribution is symmetric (skewness of 0), which is reasonable for horizontal velocity turbulence, and that it follows a standard normal distribution, the standard deviation can be estimated using Eq. (12):

$$\sigma_U = \frac{\left(U_{gust} - U_{eq}\right)}{4}, \tag{12}$$

where, U_{gust} is the model outputted wind gust at a given timestep,
U_{eq} is rotor equivalent wind speed from Eq. (5),
σ_U is the standard deviation of wind speed due to turbulence.

The reasoning used in Eq. (12) to calculate standard deviation is that since the gust is the peak wind speed observed, it is assumed to be a value in the 99.9936[th] percentile, which is four standard deviations from the

mean. The standard deviation of wind speed due to turbulence calculated using Eq. (12) is now used to calculate the random perturbation to the rotor equivalent wind speed at a given timestep using Eq. (13):

$$U_{eqT}(t) = U_{eq} + rand(0, \sigma_U),$$ (13)

where, U_{eqT} is the rotor equivalent wind speed including the effect of turbulence. The rest of the calculations proceed as described before.

4.1.2. Capturing turbine response to change in density

Modern wind turbines have control responses to maximize wind generation in presence of changing air densities. This control response is usually active close to the rated wind speed, but can also extend to *region 2* of the power curve. Figure 4 shows the turbine response in terms of the observed *Cp* values in response to changes in air density.

As seen from Figure 4 the *Cp* value is a function of both wind speed and density (top panel) and the changes in *Cp* values compared to the *Cp* value at standard density are highly non-linear (bottom panel). However, the change in *Cp* with respect to density at a given wind speed is linear, with the slope and intercept of this linear behavior changing at every wind speed. Thus, a model was constructed to predict the slope and intercept of the change in *Cp* at a given wind speed. This model allows to predict a "*correction*" to the *Cp* at standard density and given wind speed that will produce the correct *Cp* value at that specific wind speed and density.

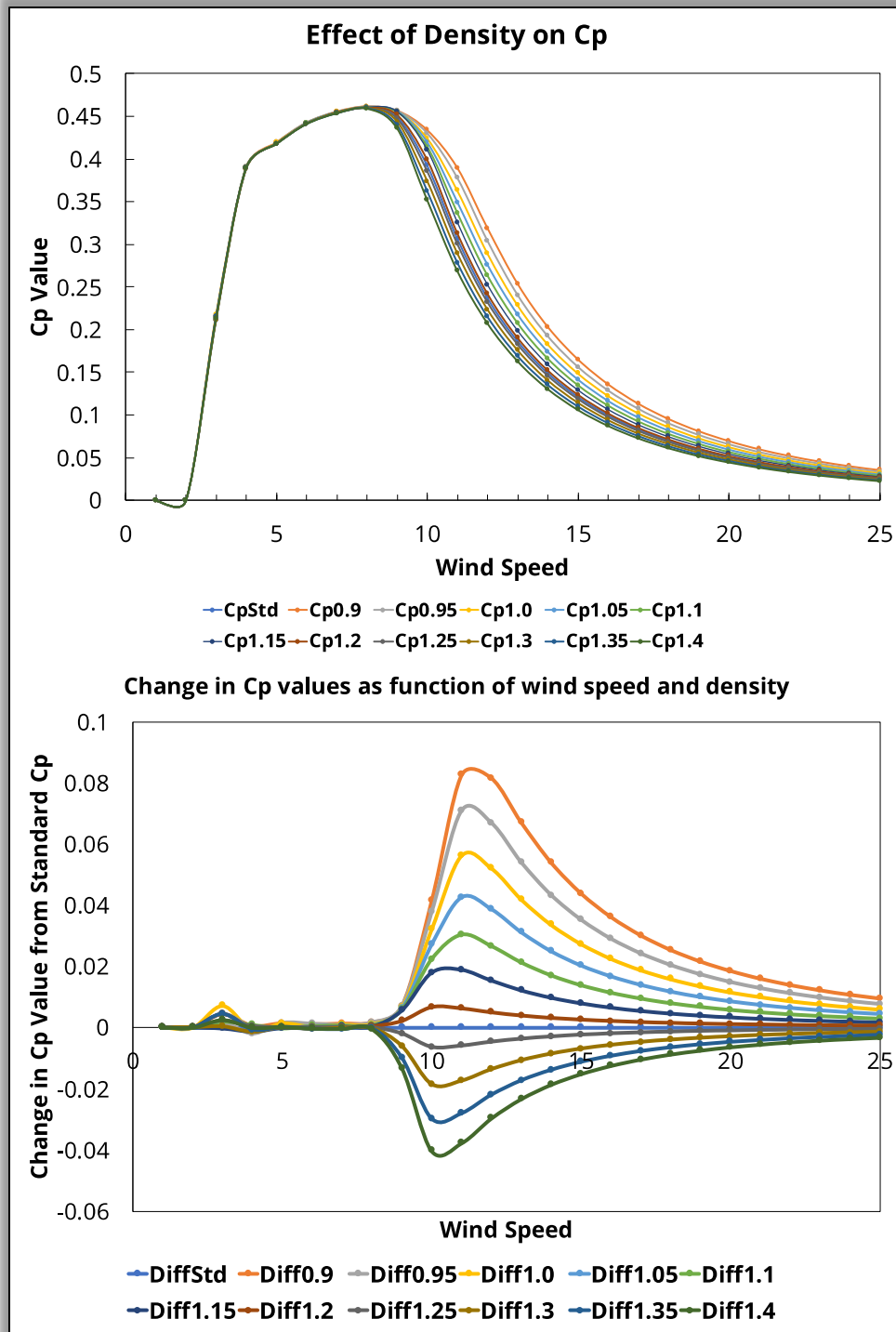Figure 4: Impact of density on Cp values for a 2.3 MW Siemens wind turbine.

Figure 5 shows comparisons of the above model predicted *Cp* values against the actual manufacturer supplied *Cp* values. It is seen that the model is able to predict the changes to the *Cp* values at various densities and wind speeds accurately. The comparison of the *Cp* values at various densities to the *Cp* values

at the standard density (1.225 kg/m^3) show that there can be differences of up 50% of the *Cp* value at a given wind speed. Hence it is very important to quantify the impact of density on the *Cp* values.
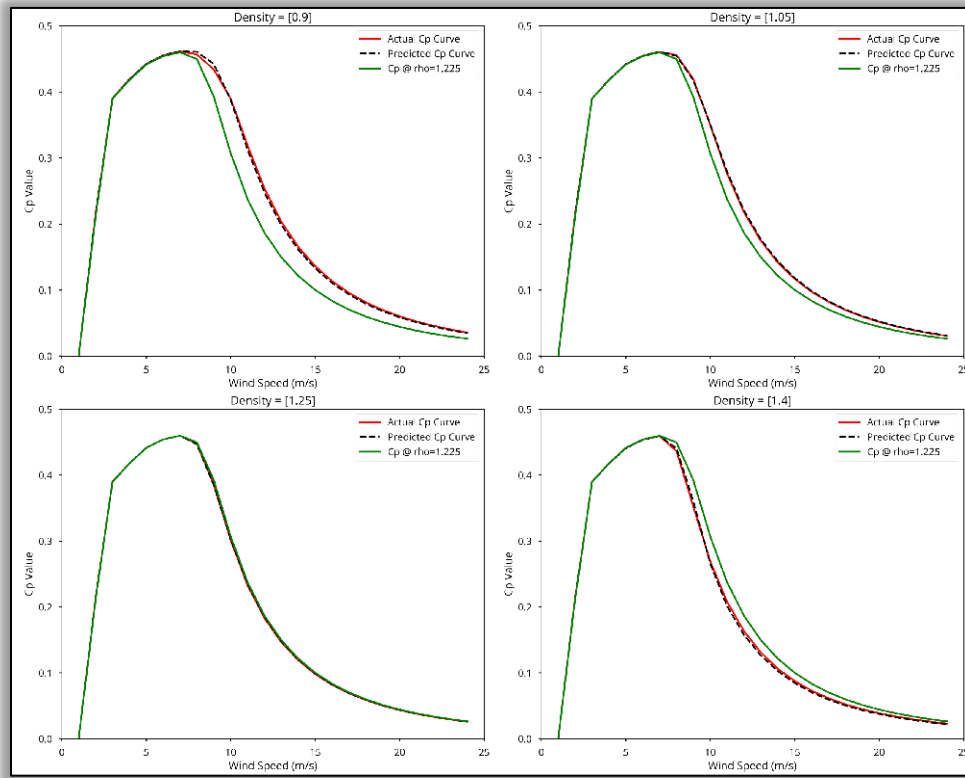

Figure 5: Comparison of model predicted Cp (black dashed line) values to the actual turbine Cp values (red solid line). The standard Cp value at density of 1.225 kg/m^3 is shown in solid green line.

4.2. Analysis of the Wind Power Dataset

The wind power calculations are performed for each HRRR cell (over 1.9 million cells in this case) for all the years required to run WIS:dom®. The WIS:dom® model is run on the same grid as the HRRR, however, only a subset of the HRRR cells are made available for wind plant development. The potential for wind development in MW at each HRRR cell is made available to WIS:dom®, which is used in determining whether wind generation gets built or not. The available wind capacity potential provided to WIS:dom® is shown in Figure 6(d). When choosing to build wind generation, WIS:dom® can choose the most optimal hub-height wind turbine to build. As seen in Figure 6 (a) and 6(c), higher hub-heights give higher wind power capacity factors. However, there are additional costs associated with building taller towers and wind turbines capable to withstanding higher wind loading. WIS:dom® takes these costs into account and determines the optimal hub-height at a given location. The optimal height is determined by evaluating whether the increased cost due to the higher tower height is offset by increased revenue or demand met from additional power generation at the higher hub-height. In this analysis it is assumed that the same turbine rotor is installed on taller towers. An important impact of this assumption is that as the hub-heights increases, the wind power capacity factors also increase due to the higher average wind speeds at increased heights above the ground. However, beyond a certain hub-height, wind power capacity factors start to decrease. This decrease in power capacity factors is due to increased wind speeds at higher hub-heights, the wind turbines are in the cut-off portion of the power curve more often. Therefore, to take full advantage of the increased wind resource at higher heights will require a redesign of the turbine rotor to operate in the higher wind speed regime.
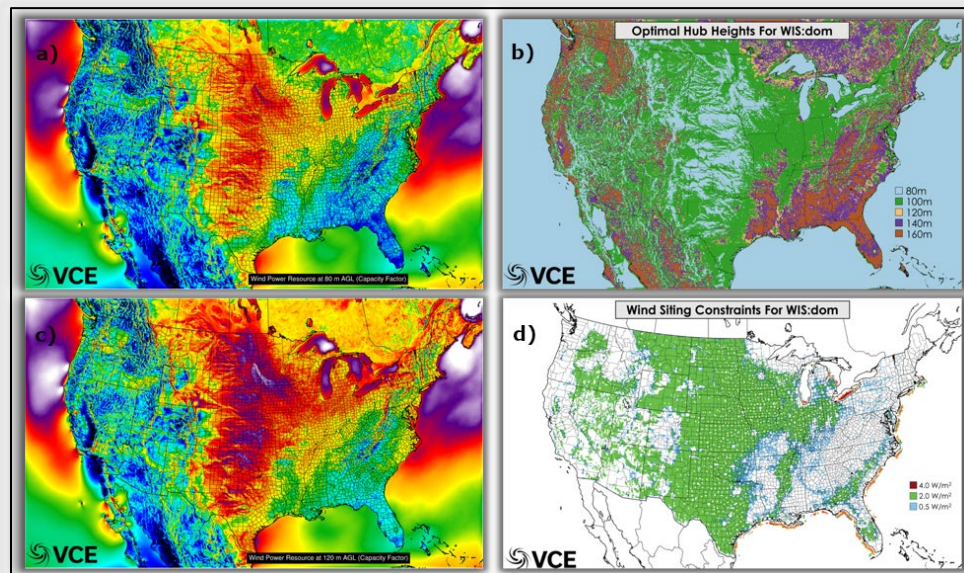
Figure 6: The wind power dataset. (a) Mean wind power capacity factor at for 80-m hub-height using data from year 2014 (b) Optimal hub-height for the CONUS (c) Mean wind power capacity factor for 120 m hub-height using data from year 2014 (d) Wind plant siting constraints for the CONUS

An important consideration in variable renewable energy development is the impact of inter-annual variability on power generation. Figure 7 shows annual average wind power capacity factors calculated for years 2014 (Figure 7a), 2015 (Figure 7b), 2016 (Figure 7c) and the average over the three years (Figure 7d). It can be seen that there are differences in the wind power capacity factors over the three years. The wind power capacity factors are seen to be lower over the entire CONUS in 2015, while in 2016 the region of high

wind power capacity factors over the Great Plains is seen to be narrowed as lower in magnitude compared to the average over the three years.
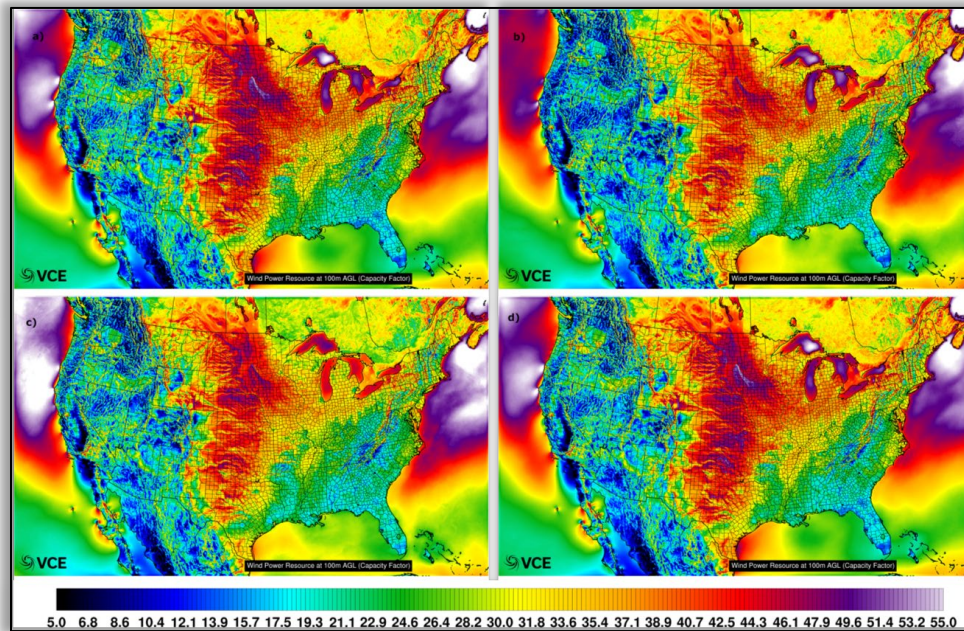

Figure 7: Average wind power capacity factors at 100m AGL for (a) year 2014 (b) year 2015 (c) year 2016 (d) average over years 2014, 2015 and 2016.

The differences in wind power capacity factor are better visualized by plotting the difference of the yearly average capacity factor from the three-year average as shown in Figure 8. It is observed that the year 2014 indeed produced higher than average wind power capacity factors over the CONUS. However, 2014 has lower than average wind power capacity factors offshore. The reason for year 2015 being called the year of "*wind drought*" is apparent from middle panel of Figure 8. The wind power capacity factors are seen to be lower by 2.5 to 3 percentage points (which is almost 10% reduction from the average capacity factors observed). The lower capacity factors are observed over the full CONUS except a few states in the mid-west such Minnesota, Michigan, Illinois. The year 2016 is characterized by lower than average wind power capacity factors over the MISO region, while the rest of CONUS shows wind power capacity factors closer to the three-year average.
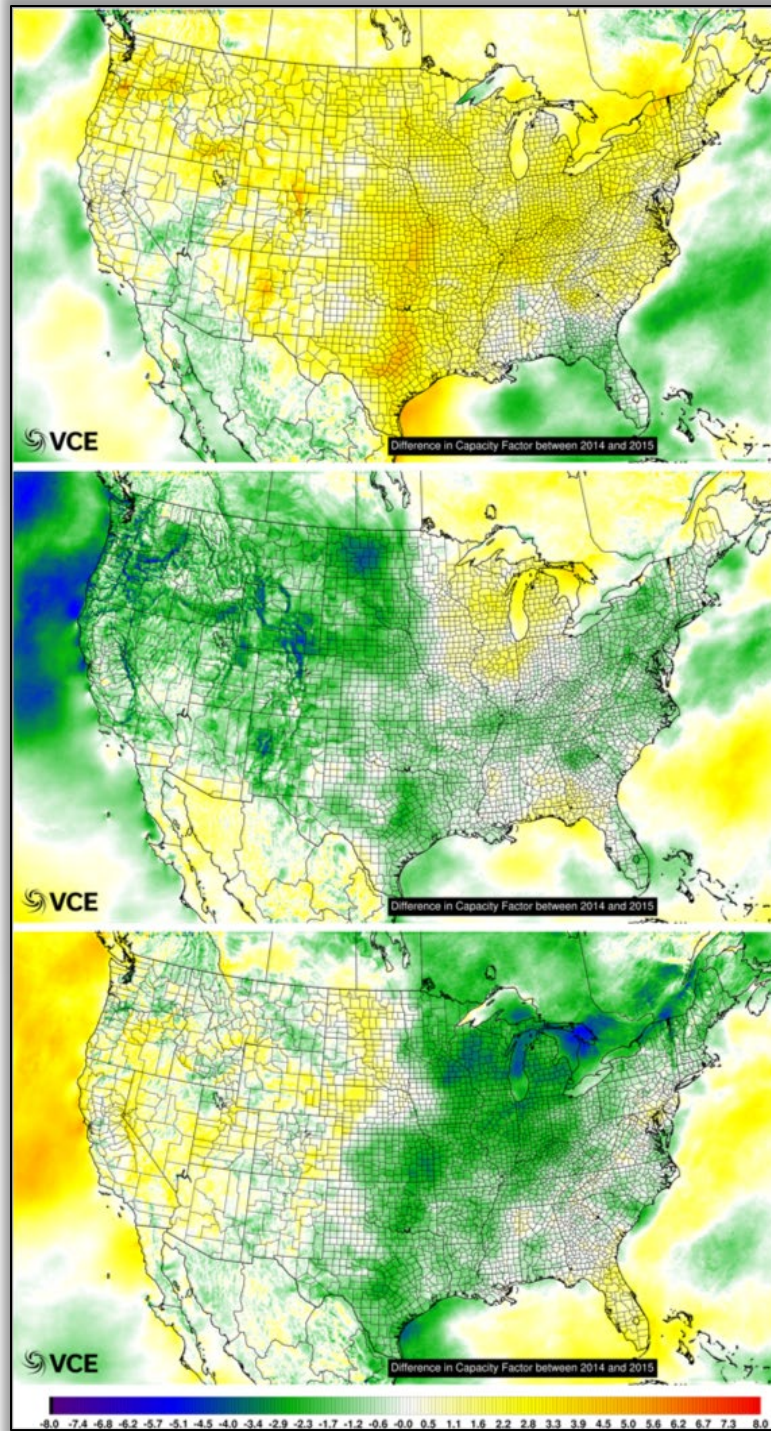
Figure 8: Percentage difference in annual capacity factor average compared to the three-year average.

The difference in annual average air density for years 2014, 2015 and 2016 from the three-year average is shown in Figure 9. It is observed from Figure 9 that air density was higher in 2014 by about 0.7% over a substantial portion of the CONUS. While in 2016, the air density was lower than the three-year average by about 0.5%, especially over the MISO region. In the year 2015, the annual average density differences were much smaller with the northern part of the CONUS showing about a 0.3% reduction. Since the change in

density has a linearly proportional impact on power generation, of the approximately 3 percentage point increase in wind power capacity factor observed in 2014, about a third was from increase in density. Similarly, of the approximately 2.5 percentage point reduction in wind power capacity factor observed in 2016, 0.5 percentage points came from reduction in air density. From the temperature data (not shown) it is observed that regions and periods of higher density were accompanied by lower than average temperatures and vice versa for periods of lower density. Therefore, in these three years variations in density accounted for approximately 20% to 33% of the inter-annual variability in wind power capacity factors. These results illustrate the importance of calculating the air density as accurately as possible which is done using Eqs (1) – (3) and incorporating the changes over the rotor swept area through the rotor equivalent calculations done using Eq. (6).

The scale of the inter-annual variability observed from these three years gives a sense about the challenges involved in capacity expansion planning when it comes to variable renewable energy. It is observed that many of the locations that saw lower capacity factors in year 2015 were locations where wind generation is concentrated; such as the great plains, Tehachapi pass and in the pacific northwest. Having several years of weather data in the WIS:dom® model ensures the model finds a solution where demand is met given these kinds of situations.

In addition to the ability to meet demand, a particular weather year can have an impact on the economics of the installed generation. For example, if a capacity expansion model is run using a weather year with higher than average wind power capacity factors (like 2014 for example), it can make wind generation looks more economically attractive than it is on average. This can result in skewed installation of generation resources, which will not only create difficulties in meeting demand during meteorologically different years, but also result in economic hardships for operators of those generation assets.
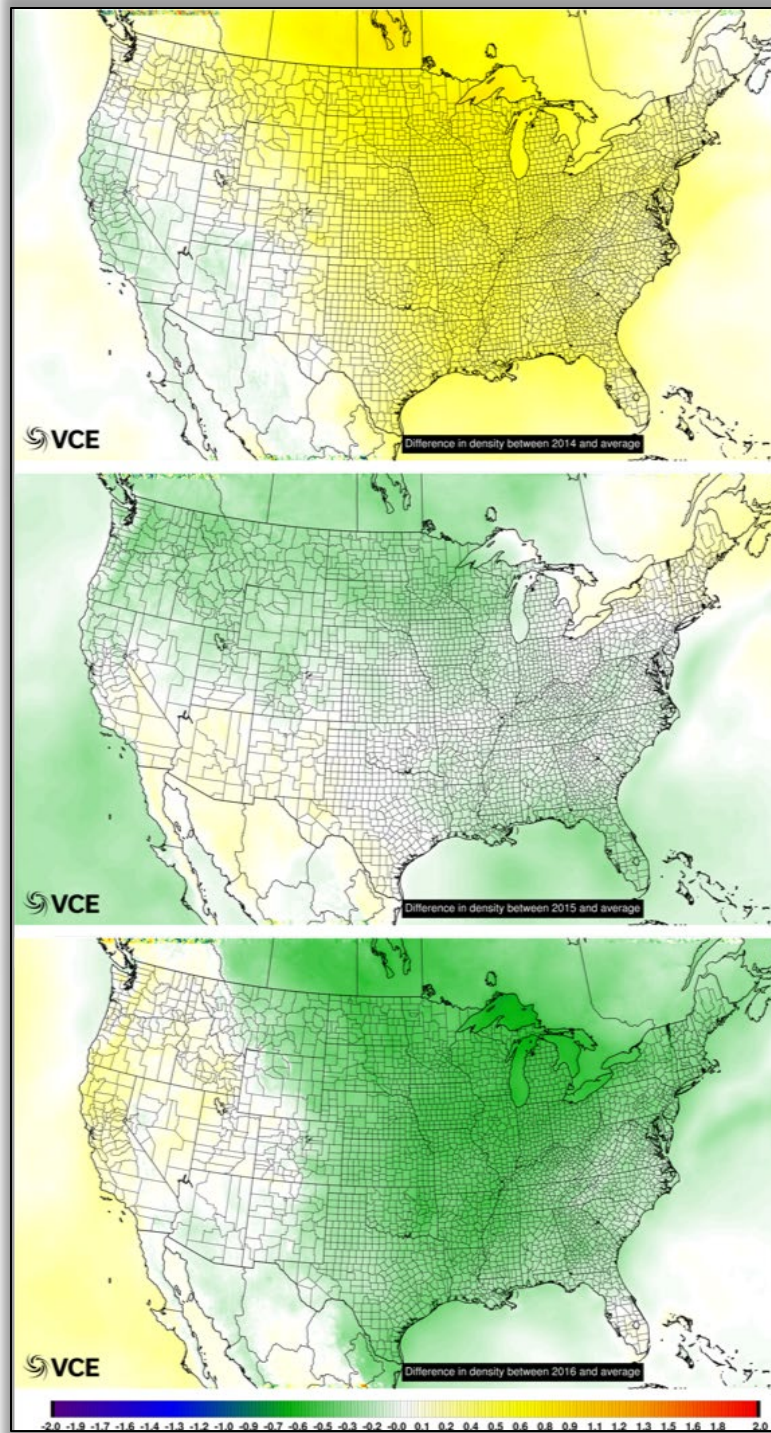
Figure 9: Percentage difference in annual average air density for year 2014 (top), 2015 (middle) and 2016 (bottom) compared to the three-year average.

Another way of looking at the inter-annual variability in a resource is to consider the Coefficient of Variation over different combinations of historical weather years. The Coefficient of Variation (CV) is a measure of relative variability. This is the standard deviation of a series divided by the mean of the same series. If the CV < 100%, the standard deviation of the series is less than the mean of the series. The CV is defined as:

$$CV = \frac{\sigma}{\mu} * \ 100\%$$

For all the weather years used for the study (2014-2018), CV was calculated for all the various combinations of years possible. Combinations are defined through the following equation:

$$C(n,r) = \frac{n!}{r!\,(n-r)!}$$

This is the formula to find the number of different combinations of n distinct objects taken *r* at a time. In this case, the objects, *n*, are 2014, 2015, 2016, 2017 and 2018. The number of times, *r*, is 1 year at a time, 2 years at a time and so on up to 5 years at a time.
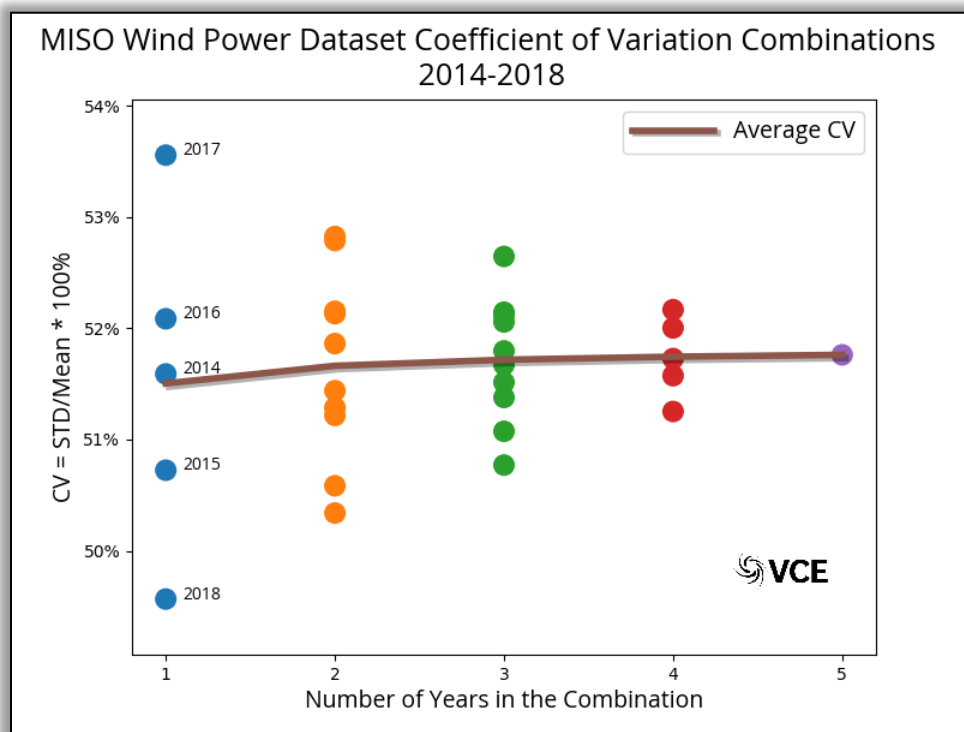

Figure 10: The Coefficient of Variation shown using 2014-2018 data over 1- to 5-year groupings for the Wind Resource across MISO. CV is shown for all combinations within each grouping.

In Figure 10, the CV is shown for all the various combinations of the 5 years of wind data available. The average CV of all events for each count, *r*, is also plotted. The average CV slowly increases with more years. Each year will bring new weather events to the dataset that were not experienced previously, thus increasing the variability over time. This average value should increase as more years are added, as it does here. When looking at the individual years, the year 2017 shows a higher extent of variability in relation to the mean. The average capacity factor from wind for that year is one of the lowest years. The standard deviation, though, is quite high. Combining these two together gives that higher extent of variability. Calendar year 2018 shows the lowest extent of variability in relation to the mean. The average capacity factors in 2018 for wind were highest from all the years available. This increase in the mean decreased the CV metric for this year even though standard deviation for 2018 was actually higher than that seen in 2017. In short, using

five chronological years provides modeling with much more variability than an average of those same five years.

4.3. Verification of Power Calculations

To validate the wind power calculations, VCE® obtained Security Constrained Economic Dispatch (SCED) data from Electric Reliability Council of Texas (ERCOT) for years 2017 and 2018[2]. Data from 109 wind farms in ERCOT is obtained for the comparison. The metadata of these wind farms is obtained from EIA and Seasonal Assessment of Resource Adequacy (SARA) reports. In addition, the shape of the wind farm and the locations of individual turbines is obtained from United States Geological Survey (USGS) United States Wind Turbine Database (USWTDB). Figure 11 shows the locations of wind farms and wind turbine locations found from the USWTDB dataset. The USWTDB dataset contains all the required metadata information to model the wind farm such as nameplate capacity, wind turbine types in the farm, hub-heights, rotor diameter and wind turbine rated capacity. The following section describes how wind farm and wind turbine information is used to model the power output from each wind farm.
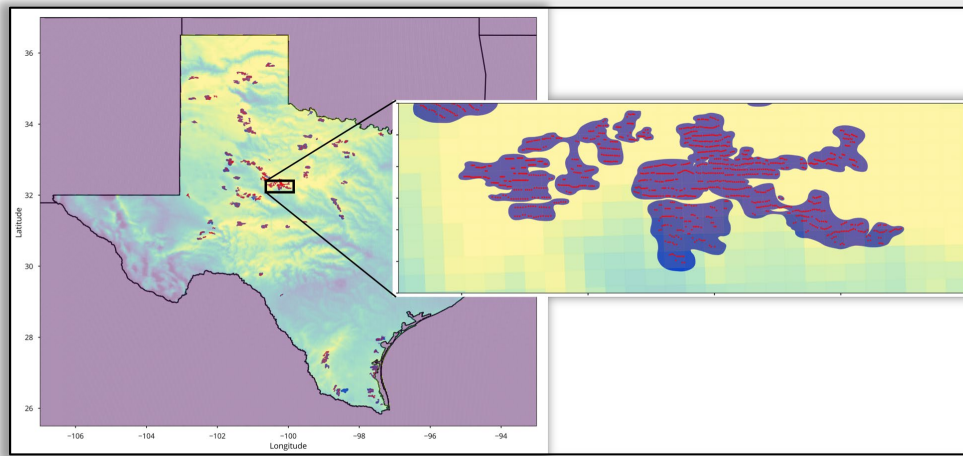

Figure 11: Wind turbine location information from the USWTDB dataset. Inset shows a zoomed in image of the area denoted by the solid black box. The background colors are average wind power capacity factors for 80 m hub-heights for years 2014 to 2016. Brighter colors indicate higher capacity factors.

There are many different open data sources available to find metadata on existing wind farms. Each source can have both large and small differences between them. There is no answer that is 100% accurate from any one data source. Having different sources is pertinent to achieving a more complete picture of what is out there for any generator technology, wind included. In ERCOT, three sources were used for analyzing and setting the metadata for all the utility scale wind farms under that ISO's umbrella. The sources included: Annual EIA 860 generator and plant data, the latest max capacity values from the SARA report, and finally the max power output observed from the Security SCED observation data reported by ERCOT for individual wind farms. The general process to align the data was as follows:

1. Align the actual wind farms reported from SARA, the EIA 860, and the SCED data was the first main task. Each source might have a slightly different count of plants or plants named differently.
2. SARA report capacities and maximum outputs from the SCED observation data might reveal that the EIA 860 annual numbers were out-of-date and usually needed to be adjusted upward.

---

[2] The SCED data obtained with the assistance of MAP®.

3. SARA report capacities might align well with EIA 860 numbers, but the SCED data might not reveal the same thing. This would be cause for further investigation into what could be going on for this farm.
4. In cases where all three sources aligned, there was much higher confidence in the final metadata decisions.
5. Cases where no resources aligned were rare. However, when they did happen, further investigation occurred into these sites to see if other sources (as an example, the USGS wind turbine location dataset) might provide any insight.

All of this work helped provide a backdrop to the representation of wind farms physically installed in the ERCOT system.

4.3.1.   Modeling the Wind Farm Generation

As described in the previous sections, the wind power capacity factors are available at 3-km horizontal resolution and 5-min time resolution for hub-heights of 80m, 100m, and 120m. For each wind farm, the wind turbine metadata is used to select the capacity factor profiles at the appropriate hub-height. For older wind farms where hub-heights of lower than 80m are present, a cubic function is used to de-rate the wind power capacity factors from 80 m level. If turbines of multiple hub-heights are present in a given farm, currently the farm is modeled using a capacity weighted average hub-height. Future work will model each wind turbine hub-height separately in order to get the most accurate results.

Next, the wind turbine type is used to select an appropriate power curve for that wind turbine type. Using a transfer function, the capacity factors from the IEC-3 power curve used for our calculations is converted the power curve of the turbine type present in the farm. If power curve information is not present then the IEC-3 power curve is used. Finally, it is checked that the sum of all the wind turbine rated capacities add up to the wind farm's nameplate capacity. If this is not so (which can happen due to missing wind turbine location information), then each wind turbine capacity is adjusted equally so that they add up to the wind farm nameplate capacity.

Now the turbine location information is used to retrieve the power capacity factors from HRRR cell the turbine is located in. Appropriate corrections to the capacity factor profiles are made as described above and then this generation information is saved. In a similar manner, all turbine generation profiles are created and finally added together to get the wind farm generation output.

4.3.2.   Comparisons of Wind Power Generation

The wind power generation profiles created as described above are validated by comparing against the SCED data obtained for 109 wind farms in ERCOT. First, the appropriate settlement point(s) for each wind farm is identified. Data from all the settlement points associated with a wind farm are summed to obtain total SCED generation from that wind farm. Initial filtering of the SCED data is performed to only keep data with "*Telemetered Status*" was "*ON*". The SCED data goes through a second, more rigorous, quality control where periods of obvious curtailment are removed. It is important to remove periods of curtailment as they cannot be simulated by the power calculation model and will result in over-estimating the model errors. The quality control for curtailed periods is performed as follows for 5-minutely data (some thresholds applied below would change for other granularities):

1. Scanning the data for any sharp spikes. The spike can be in the upward or downward direction. Any given timestamp was flagged if the power changed more than +/-20% of farm capacity and in opposite directions in the both the forward and backward direction in time. As an example, a suspect period would be flagged when looking back a single time step the power changed more than +20% of maximum capacity and looking forward a single time step the power dropped more than -20% of max capacity. This would also be flagged if it had been a spike in the downward direction. This threshold was used since the data was 5-minutely.
2. When power flatlines or hovers around +/- 1% of a value that is not within a certain range of max capacity or 0% capacity for more than two hours it is flagged as suspect. The offending value has to flatline or hover more than 5% below max capacity or more than 5% above zero.
3. Any sharp drop or rise to or from 0% capacity where the farm was at 0% capacity for more than two hours or the farm dropped to 0% capacity for more than two hours is suspect. For 5-minutely data, a 15% rise or drop in capacity would be flagged.
4. Any drop or rise from any point that was greater than 50% capacity in a 5-minute time frame is flagged as suspect.

Any of the above thresholds can be changed. Loosening the thresholds will release certain curtailment periods from being captured or flagged. Tightening the threshold will remove more periods of data that are actually physical weather events that create sharp ramps. Figure 12 shows a partial time-series of SCED generation that was quality controlled. It is observed that the obviously curtailed periods (dark blue line) are rejected while keep the rest of profile (light blue line) intact.
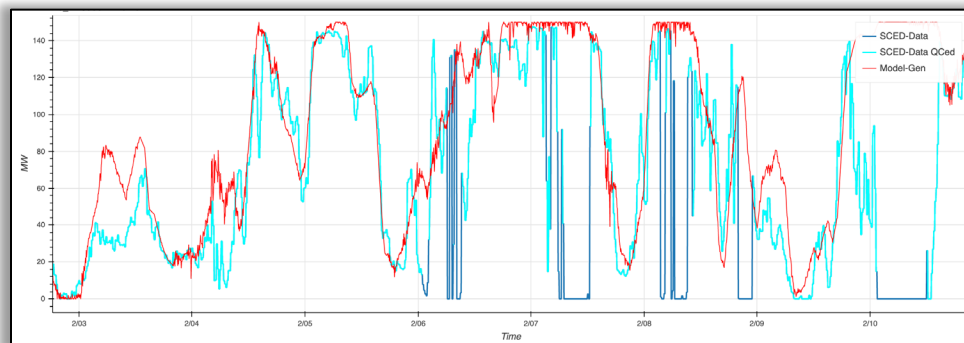
Figure 12: An example wind farm in ERCOT where model output, SCED Data and quality-controlled SCED data is shown. When dark blue can be seen is where data was flagged as potential curtailment through VCE® quality control processes.

The metrics used to estimate the quality of the modeled wind power profiles are bias, root mean square error (*RMSE*), correlation coefficient and coefficient of determination (R^2). Figure 13 shows an example time-series of the VCE® model wind generation output compared to the SCED data for the Amazon Wind Farm in ERCOT. The timeseries show that the model is able to capture the temporal variability in the power production accurately. This is due to a combination of accurate physical model of wind turbine operation, accurate wind forecasts from the HRRR and finally accurately determining the wind farm and wind turbine metadata to ensure that the wind generation model is working off correct details.

Figure 13: Comparison of VCE® model wind generation output (orange) with SCED data (blue) for the Amazon Wind Farm in ERCOT.

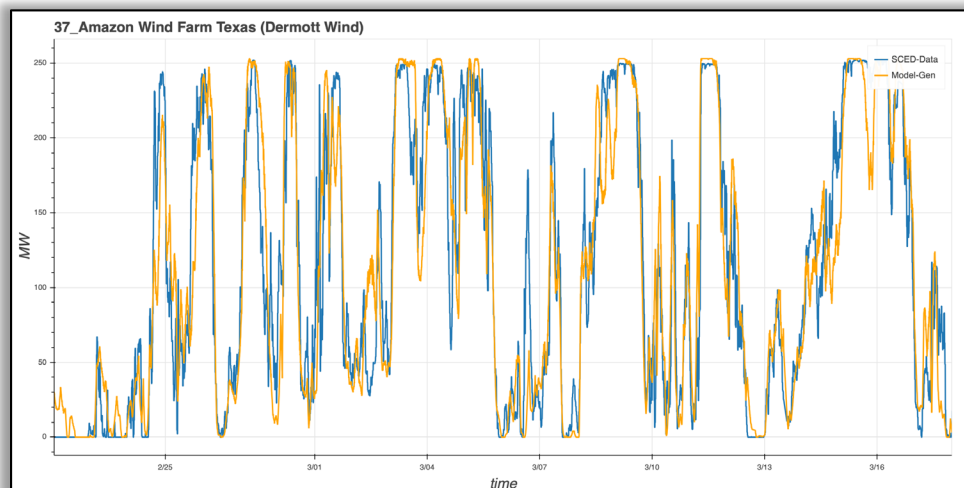It is observed that of the 109 wind farms for which SCED data is available, 100 wind farms have correlation coefficients of greater than 80%, while the remaining nine have correlation coefficients of greater than 70%. Figure 14 summarizes the comparison statistics from the wind farm validation. As seen from Figure 14, almost all wind farms show correlation coefficients greater than 80%. It is also observed that many wind farms are clustered around ratio of standard deviation of unity, which indicates that the model and the SCED data show similar level of temporal variability.

The main challenge in validation of wind farm generation output is removing periods of curtailment as these cannot be modeled by a wind farm generation model. Although it was attempted to remove as many periods of obvious curtailment as possible, many curtailed generation periods still remain. In addition, wind farm output is modified due to turbines being down for maintenance or disbanded (true in cases of many older wind farms). Since this information on wind farm maintenance schedules and turbine status is unavailable, differences in the modeled generation and SCED data arise. These differences are evident in Figure 14 from the centered RMS difference in many wind farms being on the order of 40% of the standard deviation. In addition, the curtailed periods artificially reduce the standard deviation of the SCED data and hence it appears that the model forecasts are over-predicting variability, which is probably not the case as the model output is expected to the smoother than the actual generation.

Figure 14: Taylor diagram summarizing the comparison metrics between VCE® model predicted wind power generation and SCED data.

Wind Farm Validation

Legend:

★ Reference
‑ ‑ Reference Std Dev

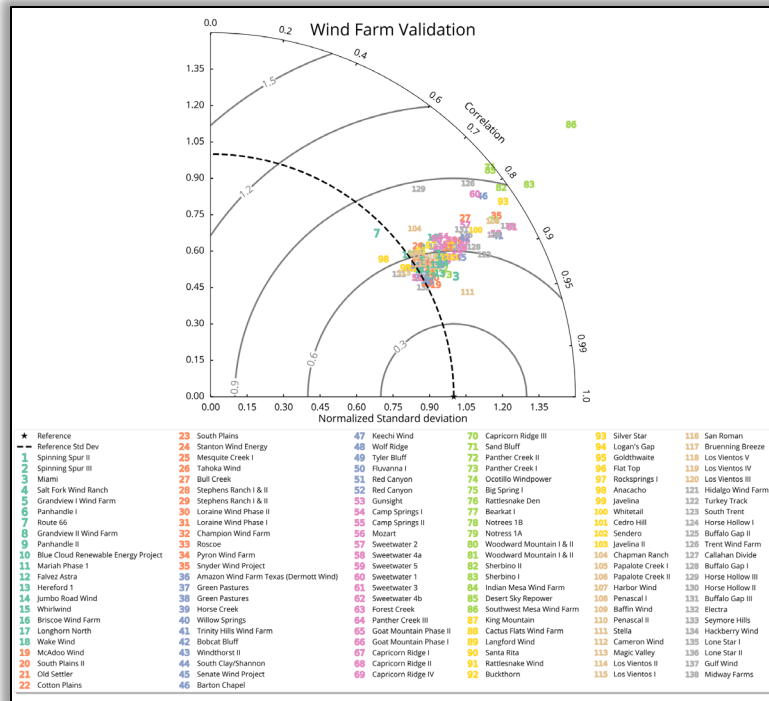| #	Name	#	Name	#	Name	#	Name	#	Name	#	Name
1	Spinning Spur II	23	South Plains	47	Keechi Wind	70	Capricorn Ridge III	93	Silver Star	116	San Roman
2	Spinning Spur III	24	Stanton Wind Energy	48	Wolf Ridge	71	Sand Bluff	94	Logan's Gap	117	Bruenning Breeze
3	Miami	25	Mesquite Creek I	49	Tyler Bluff	72	Panther Creek II	95	Goldthwaite	118	Los Vientos V
4	Salt Fork Wind Ranch	26	Tahoka Wind	50	Fluvanna I	73	Panther Creek I	96	Flat Top	119	Los Vientos IV
5	Grandview I Wind Farm	27	Bull Creek	51	Red Canyon	74	Ocotillo Windpower	97	Rocksprings I	120	Los Vientos III
6	Panhandle I	28	Stephens Ranch I & II	52	Red Canyon	75	Big Spring I	98	Anacacho	121	Hidalgo Wind Farm
7	Route 66	29	Stephens Ranch I & II	53	Gunsight	76	Rattlesnake Den	99	Javelina	122	Turkey Track
8	Grandview II Wind Farm	30	Loraine Wind Phase II	54	Camp Springs I	77	Boerkat I	100	Whitetail	123	South Trent
9	Panhandle II	31	Loraine Wind Phase I	55	Camp Springs II	78	Notress 1B	101	Cedro Hill	124	Horse Hollow I
10	Blue Cloud Renewable Energy Project	32	Champion Wind Farm	56	Mozart	79	Notress 1A	102	Sendero	125	Buffalo Gap II
11	Mariah Phase 1	33	Roscoe	57	Sweetwater 2	80	Woodward Mountain I & II	103	Javelina II	126	Trent Wind Farm
12	Falvez Astra	34	Pyron Wind Farm	58	Sweetwater 4a	81	Woodward Mountain I & II	104	Chapman Ranch	127	Callahan Divide
13	Hereford 1	35	Snyder Wind Project	59	Sweetwater 5	82	Sherbino II	105	Papalote Creek I	128	Buffalo Gap I
14	Jumbo Road Wind	36	Amazon Wind Farm Texas (Dermott Wind)	60	Sweetwater 1	83	Sherbino I	106	Papalote Creek II	129	Horse Hollow III
15	Whirlwind	37	Green Pastures	61	Sweetwater 3	84	Indian Mesa Wind Farm	107	Harbor Wind	130	Horse Hollow II
16	Briscoe Wind Farm	38	Green Pastures	62	Sweetwater 4b	85	Desert Sky Repower	108	Penascal I	131	Buffalo Gap III
17	Longhorn North	39	Horse Creek	63	Forest Creek	86	Southwest Mesa Wind Farm	109	Baffin Wind	132	Electra
18	Wake Wind	40	Willow Springs	64	Panther Creek III	87	King Mountain	110	Penascal II	133	Seymore Hills
19	McAdoo Wind	41	Trinity Hills Wind Farm	65	Goat Mountain Phase II	88	Cactus Flats Wind Farm	111	Stella	134	Hackberry Wind
20	South Plains II	42	Bobcat Bluff	66	Goat Mountain Phase I	89	Langford Wind	112	Cameron Wind	135	Lone Star I
21	Old Settler	43	Windthorst II	67	Capricorn Ridge I	90	Santa Rita	113	Magic Valley	136	Lone Star II
22	Cotton Plains	44	South Clay/Shannon	68	Capricorn Ridge II	91	Rattlesnake Wind	114	Los Vientos II	137	Gulf Wind
		45	Senate Wind Project	69	Capricorn Ridge IV	92	Buckthorn	115	Los Vientos I	138	Midway Farms
		46	Barton Chapel								

Figure 14: Taylor diagram summarizing the comparison metrics between VCE® model predicted wind power generation and SCED data.

To better understand how curtailment, turbine maintenance or disbandment can affect wind farm output, the model forecasted generation is compared against the SCED data for the Southwest Mesa Wind farm in ERCOT in Figure 15. As seen from Figure 15, the SCED data and the model forecasted generation are well correlated, but the SCED data never reaches the maximum power forecasted by the model. The main reason for this is that this wind farm is quite old and from visual inspection of Google Earth®, many turbines are broken down or feathered for maintenance. As a result, although the model is correctly simulating the wind generation from the farm, it shows a high RMS difference, but with a high correlation coefficient as seen in Figure 14.
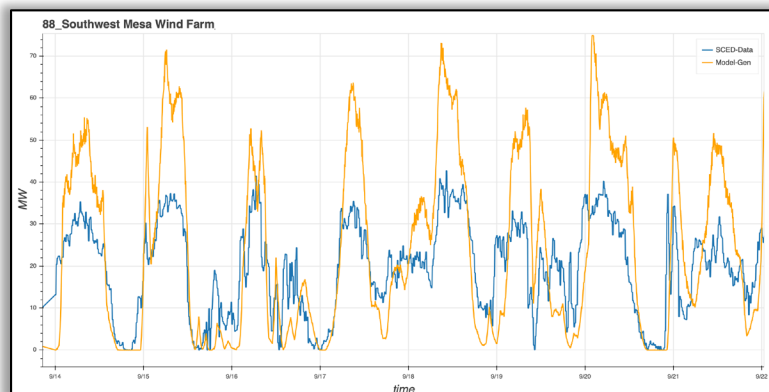
Figure 15: Comparison of VCE® model wind generation output (orange) with SCED data (blue) for the Southwest Mesa Wind Farm in ERCOT.

5. Solar Power Dataset Method

Calculation the solar PV power output requires accurate forecasts of Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DHI). These variables are then input into a PV cell power modeling algorithm. The components of the solar irradiances are related to each other by:

$$\text{GHI} = \text{DNI} * \cos(\text{sza}) + \text{DHI} \tag{14}$$

where *sza* is the solar zenith angle.

Numerical weather prediction models did not output forecasts of DNI and DHI until 2016. In addition, forecasts of DNI and DHI produced by the HRRR after 2016 have significant biases mainly due to improper representation of clouds. To obtain forecasts of DHI and DNI from model outputs for years before 2016 and correct for model biases for years after 2016, VCE® employs a linear multiple multivariate regression technique developed by Clack (2017). The variables used to create the solar power data are shown below.

From HRRR:
  (1) Downwelling shortwave (*SW*),
  (2) Downwelling longwave (*LW*),
  (3) 10-m wind speed (*Wind10m*),
  (4) 2-m temperature (*T2m*),
  (5) Direct normal irradiance (*DNI*) – 2016 onwards,
  (6) Diffuse horizontal irradiance (*DHI*) – 2016 onwards.

From GOES-east and GOES-west (for datasets before 2016 only):
  (7) Visible,
  (8) 4μm,
  (9) 11μm,
  (10) 13μm,
  (11) Water-vapor.

Calculated:
  (12) Direct normal irradiance at the top of the atmosphere (DNI_0),
  (13) Solar zenith angle (*sza*),
  (14) Solar azimuth angle (*azm*),
  (15) Hour-angle (*hrang*),
  (16) Declination angle (*dec*).

The satellite observations are not included starting in 2016 as these observations are already assimilated within the HRRR. To perform the regression procedure, we get observations of GHI, DNI and DHI from 15 ground-based radiation measurement sites (SURFRAD and SOLRAD) operated by the NOAA. The above variables are chosen as they are most likely to impact the amount of solar irradiance reaching the Earth's surface and its attenuation along the way. A substantial portion of the effort in creating the solar power dataset is spent on getting the data ready for regression. First, the required HRRR variables are extracted from the HRRR output files at 1-hour resolution. These HRRR variables are then linearly interpolated to 5-min intervals. For data gaps of longer than 1-hour, persistence is assumed and they are filled in with data from the same hour on the previously available day. Linear interpolation is carried out only sub-hourly.

Satellite measurements from GOES-east and GOES-west are used, which allows a stereoscopic observation of the cloud field. Each of the GOES satellite observations cover the full CONUS with observations available at 15-min time interval. The GOES satellites make measurements in 5-channels listed above. The measurements are in bit count which are converted to temperature (in Kelvin) using the formula (Clack 2017):

$$T = \frac{1}{2}(660 - B), \quad 0 \leq B \leq 176 \quad and$$
$$T = 418 - B, \quad 176 < B \leq 255. \tag{15}$$

The spatial resolution of the satellite data is 1-km for the visible channel and 4-km for the remaining channels (infrared and water vapor). Since the HRRR has a spatial resolution of 3-km, the satellite data are spatially interpolated on to the HRRR grid. This spatially interpolated satellite data is then linearly interpolated in time to 5-min intervals to match the interpolated HRRR output.

In addition to the variables obtained from the HRRR and satellite measurements, five additional variables (12 to 16 above) are calculated. The calculation of solar irradiance at the top of the atmosphere needs to take into account the eccentricity of Earth's orbit. The average DNI_0 at the top of the atmosphere is 1360.8 W m^{-2} and is denoted by I_o. The equation for the actual irradiance hitting the top of the atmosphere is given by:

$$DNI_0 = I_o \left(\frac{R_{avg}}{R}\right)^2 \tag{16}$$

where R_{avg} is the average Earth-Sun distance and R is the instantaneous Earth-Sun distance. The ratio of R_{avg} to R is given by the Fourier expansion in Eq. (14), which is accurate to 0.0001 (Spencer 1971):

$$\left(\frac{R_{avg}}{R}\right)^2 \approx 1.000110 + 0.034221\cos(\delta) + 0.00128\sin(\delta) + 0.000719\cos(2\delta) + 0.000077\sin(2\delta) \tag{17}$$

where, the day angle $\delta = \frac{2\pi d}{365.242}$ radians and d is the day of the year.

The declination angle is also given as a Fourier expansion in Eq. (18), which accurate to 0.0006 radians (Spencer 1971):

$$dec = 0.006918 - 0.399912\cos(\delta) + 0.070257\sin(\delta) - 0.006758\cos(2\delta) + 0.000907\sin(2\delta)$$
$$- 0.002697\cos(3\delta) + 0.00148\sin(3\delta) \tag{18}$$

The hour angle, given by Eq. (19) is defined as the number of degrees the Sun moves across the sky compared to local solar noon. The hour angle is zero at local solar noon, positive in the afternoon and negative in the morning:

$$hrang = 15 * (LST - 12), \tag{19}$$

where LST is the local solar time given by

$$LST = LT + \frac{TC}{60}, \tag{20}$$

where *LT* is the local time and *TC* is the time correction factor that accounts for the variation in the local solar time due to the range of longitudes within the same time zone, eccentricity of the Earth's orbit and Earth's axial tilt (to calculate the last two, the equation of time given by Eq. (23) is used). The time correction factor (*TC*) is calculated as

$$TC = 4 * (lon - LSTM) + EoT, \tag{21}$$

where, *LSTM* is the local standard time meridian is the reference meridian used for a particular time zone and is calculated using

$$LSTM = 15 * (LT - UTC) \tag{22}$$

and *EoT* is the equation of time, which is an empirically derived relationship that corrects for the eccentricity of the Earth's orbit and the Earth's axial tilt. The *EoT*, in radians, is given by a Fourier expansion that is accurate to 0.0025 radians or 35 seconds (Spencer 1971):

$$EoT = 0.000075 + 0.001868\cos(\delta) - 0.032077\sin(\delta) - 0.014615\cos(2\delta) - 0.040849\sin(2\delta). \tag{23}$$

Now the solar zenith angle can be calculated using Eq. (24)

$$\cos(sza) = \sin(lat)\sin(dec) + \cos(lat)\cos(dec)\cos(hrang), \tag{24}$$

where, the latitude (*lat*), declination angle (*dec*) and hour angle (*hrang*) are all in radians.

Once all the quantities required for the regression are collected, the regression can be performed. The regression is performed separately for the GHI and DNI for computational efficiency and DHI is calculated using Eq. (14) once the GHI and DNI estimates are known. The regression is represented mathematically as (Clack 2017)

$$Y_{n\times p} = X_{n\times(r+1)}\beta_{(r+1)\times p} + \varepsilon_{n\times p}, \tag{25}$$

where, $Y_{n\times p}$ are the endogenous variables (here the ground-based measurements of GHI, DNI and DHI), $X_{n\times(r+1)}$ are the exogenous variables (here the variables from the NWP model, satellite measurements and calculated variables), $\beta_{(r+1)\times p}$ are the regression coefficients, and $\varepsilon_{n\times p}$ are the measurement errors in the ground-based observations.

The ground-based observations of the irradiance components measured by the SURFRAD and SOLRAD sites are available at 1-min time resolution. These measurements are averaged to 5-min resolution to reduce measurement noise and helps reduce the discrepancy between a point measurement from the SOLRAD/SURFRAD sites and the grid-cell average from the HRRR model. The errors in the SURFRAD/SOLRAD observations are modelled as:

$$\varepsilon = 5 + 0.02 * (1 - \cos(sza)) + 0.01 * T2m + 0.005 * GHI. \tag{26}$$

The SURFRAD/SOLRAD measurements are known to have error bars of ± 5 W m^{-2} under ideal conditions. These errors get larger depending on various factors such as total irradiance, ambient temperature and solar zenith angle. The regression is performed using the advanced statistics package from IDL and analysis of variance (ANOVA) techniques are used to determine performance of the regression. The regression

coefficients that give the best performance are applied to the HRRR data to get irradiance estimates over the whole contiguous United States. Once the irradiance components are calculated, the power production from a photovoltaic panel can be estimated.

5.1. Estimating Power from Solar Photovoltaics

A solar photovoltaic (PV) cell converts radiation incident on its surface to electrical power. A PV cell utilizes both the direct (DNI) and diffuse (DHI) radiation to produce current and a voltage that determines the power generated by the cell, as shown in Figure 16,

$$P_s = V * I. \tag{27}$$
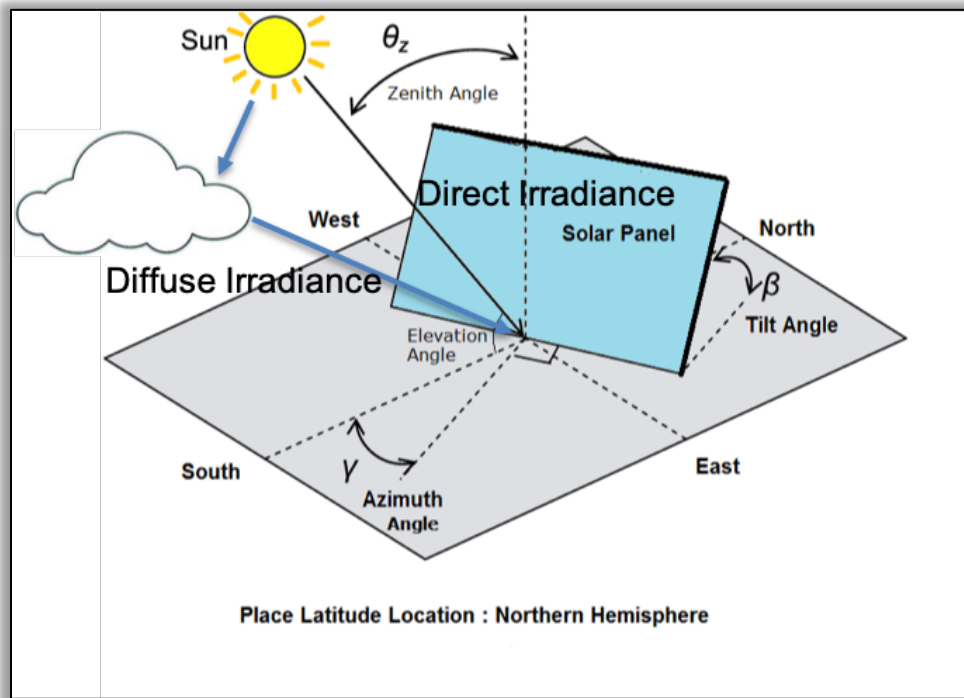
Figure 16: Schematic showing the direct and diffuse irradiance on a PV panel with respect to its tilt and azimuth orientation.

The power performance of a PV cell is a complex function of several environmental factors (such as ambient temperature, wind speed, incident irradiation) as well as the PV cell characteristics. These environmental factors interact non-linearly and make estimating the power output from a PV cell difficult. The power performance model used by VCE® is an empirically derived model developed by Sandia National Laboratory and described in King et al. 2004. To calculate the voltage and current induced in the PV cells, Eqs (11) to (20) from King et al. 2004 are used. These equations attempt to model the non-linear response of a PV cell as an interaction of several factors each having well defined, experimentally derived, relationships with the independent variables affecting PV cell performance.

To calculate the power produced from a PV cell requires being able to model the shape of the I-V curve of the PV panel accurately. The I-V curve of a PV cell shifts depending on the amount of incident radiation on the panel and ambient temperature. In order to replicate these effects accurately the King et al. 2004 paper models the voltage and current response separately using 3,300 measurements made over a range of clear and cloudy conditions, wide range of solar irradiance and module temperatures. The measured voltage values are first translated to a common temperature of 50°C to remove effects of temperature. The translated measurements of voltage and associated irradiance are regressed using Eqs (28) and (29) to find values of n, C_2 and C_3.

$$V_{oc} = V_{oco} + N_s\delta(T_c).\ln(E_e) + \beta_{Voc}E_e(T_c - T_o) \qquad (28)$$

$$V_{mp} = V_{mp0} - C_2N_s\delta T_c\ln(E_e) - C_3N_s(\delta T_c\ln(E_e))^2 - \beta_{vmp}E_e(T_c - T_o) \qquad (29)$$

where, V_{oc} is the open-circuit voltage, V_{oco} is the open-circuit voltage constant, V_{mp} is the voltage at maximum power, V_{mp0} is the constant for voltage in I-V curve, $\delta(T_c) = n\,k\,(T_c + 273.15)/q$ is the thermal voltage per cell at temperature T_c, n, c_2, c_3 constants for voltage formula, q is the elementary charge (1.60218e^{-19} coulomb), k is the Boltzmann's constant (1.38066e^{-23} J K^{-1}), N_s number of cells in series in a module's cell-string, T_o reference cell temperature, and $T_c = GHI * e^{a+b*WS} + T_{2m}$ is the cell temperature inside the module (with WS being the wind speed, T_{2m} the 2-m temperature, and a, b constants).

In a similar way, to determine the dependence of module current on incident irradiation, the current values are translated to a common temperature and regression coefficients, C_o, C_1, C_4, C_5, C_6, C_7, are determined using Eqs (30), (31) and (32).

$$I_{mp} = I_{mpo}\{C_0E_e + C_1E_e^2\}\{1 + \alpha_{Imp}(T_c - T_o)\}, \qquad (30)$$

$$I_x = I_{xo}\{C_4E_e + C_5E_e^2\}\{1 + \alpha_{Isc}(T_c - T_o)\}, \qquad (31)$$

$$I_{xx} = I_{xxo}\{C_6E_e + C_7E_e^2\}\{1 + \alpha_{Imp}(T_c - T_o)\}, \qquad (32)$$

where, I_{mp} is the current at maximum power, I_{mp0} is the constant for current in I-V curve, I_x is the current at module V = 0.5 V_{oc}, I_{xx} is the current at module V = 0.5 $(V_{oc} + V_{mp})$, I_{x0} is the constant for current in I-V curve, I_{xx0} is the constant for current in I-V curve, α_{Imp} is the normalized temperature coefficient for I_{mp}, α_{Isc} is the normalized temperature coefficient for I_{sc}, the short-circuit current.

In the above equations, E_e is the effective irradiance to which the PV cells in the module respond to and is given by

$$E_e = f_1 * SF * \left[\frac{E_bf_2 + f_d * DHI}{E_o}\right], \qquad (33)$$

where, $E_b = DNI * \cos(AOI)$, which is the beam component of the solar irradiance incident on module surface, E_o is the reference solar irradiance (1000 W/m^2), f_1 is the relation between solar spectral variation and short circuit current given by $f_1 = a_0 + a_1AM_a + a_2AM_a^2 + a_3AM_a^3 + a_4AM_a^4$, where a_0, a_1, a_2, a_3, a_4 are constants and AM_a is the absolute air-mass (dimensionless), SF is the soiling factor, f_2 is the relation between optical influences and solar angle-of-incidence ($f_2 = b_0 + b_1 * AOI + b_2 * AOI^2 + b_3 * AOI^3 + b_4 * AOI^4$, where b_0, b_1, b_2, b_3, b_4 are constants, $AOI = \cos\beta\cos\theta_z - \sin\beta\sin\theta_z\cos\gamma$ is the angle of incidence, where β is tilt angle of the panel with respect to the ground, θ_z is solar zenith angle, γ is the azimuth angle with respect to the north-south, and f_d is the relative response to diffuse versus beam irradiance).

The empirical functions $f_1(AM_a)$ and $f_2(AM_a)$ quantify the effect of solar spectral variation and optical influences on short-circuit current. These functions are determined from laboratory testing and account for systematic effects that occur during clear sky periods. Absolute airmass provides a relative measure of the path length solar radiation has to travel at a given solar zenith angle compared to a solar position of directly overhead.

The performance of a PV panel also depends on the module temperature as seen in Eqs (28) - (32). The thermal response of a PV cell can be modelled as

$$T_c = GHI * e^{a+b*WS} + T_{2m}. \tag{34}$$

The simple model for expected module temperature given by Eq. (34) has been shown to have accuracy of ± 5 °C, which results in uncertainty in power output of less than 3%.

The constants in the power generation model are obtained from De Soto et al. 2006 and the NREL System Advisory Model (SAM). It is assumed that the individual panels are placed far enough apart so as not to create any shadowing effects. The above formulation is used to calculate solar power production for the following technologies:
   (a) Fixed PV panel for various tilt angles (0°, 15°, 30°, 45°, latitude tilt),
   (b) One-axis tracking at latitude tilt,
   (c) Two-axis tracking.

Finally, the calculated power output is de-rated based on expected losses from wiring and soiling (4.5% loss), AC/DC conversion (3.3% loss) and presence of snow on the panels (assume no production if snow is present – for fixed panels at 0° and 15° elevation). WIS:dom® can update the magnitude of these losses to account for improved technology in the future or panel performance degradation with age.

The average solar power capacity factor over the contiguous United States for 2014, 2015, 2016 and the three-year average is shown in Figure 17. It is observed that the south-west region of the US has the best solar potential. Comparing the solar power capacity factors from Figure 17 and the wind power capacity factors from Figure 7, it is observed that locations of the best solar resource are broadly complementary to locations of the best wind resource. The south-west and the south-east have the best solar potential over the CONUS which as seen from Figure 7 are not always the best locations for wind power. However, several locations in the Great Plains are seen to have good wind and solar power capacity factors. It is also observed from Figure 17 that similar to the wind power capacity factors, the solar power capacity factors also show significant inter-annual variability. The year 2016 is seen to have significantly higher solar power capacity factors especially in the south-west.
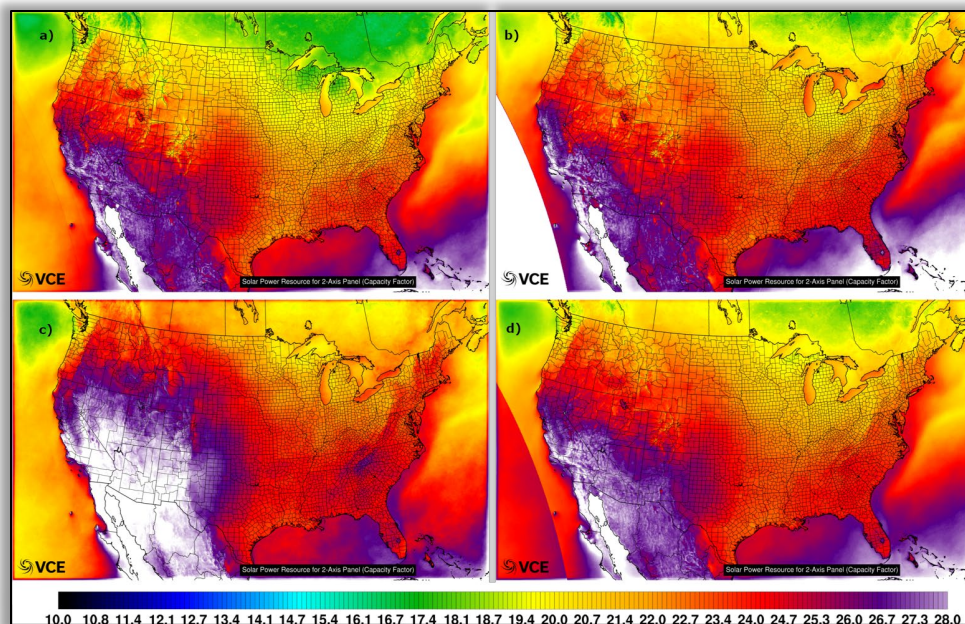
Figure 17: Annual average solar capacity factors for 2-Axis panels. (a) for year 2014, (b) for year 2015, (c) for year 2016 and (d) three-year average

In order to study the variability in the solar power capacity factors, the difference between the yearly average solar power capacity factors is calculated with respect to the three-year average as shown in Figure 18. It is seen from Figure 18 that the year 2014 had lower than average solar power capacity factors by 1.5 to 2 percentage points (a 5 – 10% reduction in expected solar power capacity factors depending on location). This reduction in solar power capacity factors in 2014 was driven in some part by the persistent polar vortex setup over much of the CONUS from January through March 2014. In 2016, the solar capacity factors were much higher compared to years 2014 and 2015 in spite of 2016 summer being the hottest on record at the time, which should reduce PV performance. However, the lower than normal moisture in the air ensured enough clear sky radiation for this year to produce higher than average solar power capacity factors. The increase was dominant over the western part of the CONUS as seen from the bottom panel of Figure 18. As further documented in Figure 18, the year 2015 had lower than average solar power capacity factors, especially over the western part of the CONUS. The lower than average solar power capacity factors combined with the lower than average wind power capacity factors in year 2015 (see Figure 7) make it a particularly difficult year for renewable energy generation. If capacity expansion planning was performed using weather data from year 2015 alone, it would vastly under-predict the utility of renewables to meet demand.
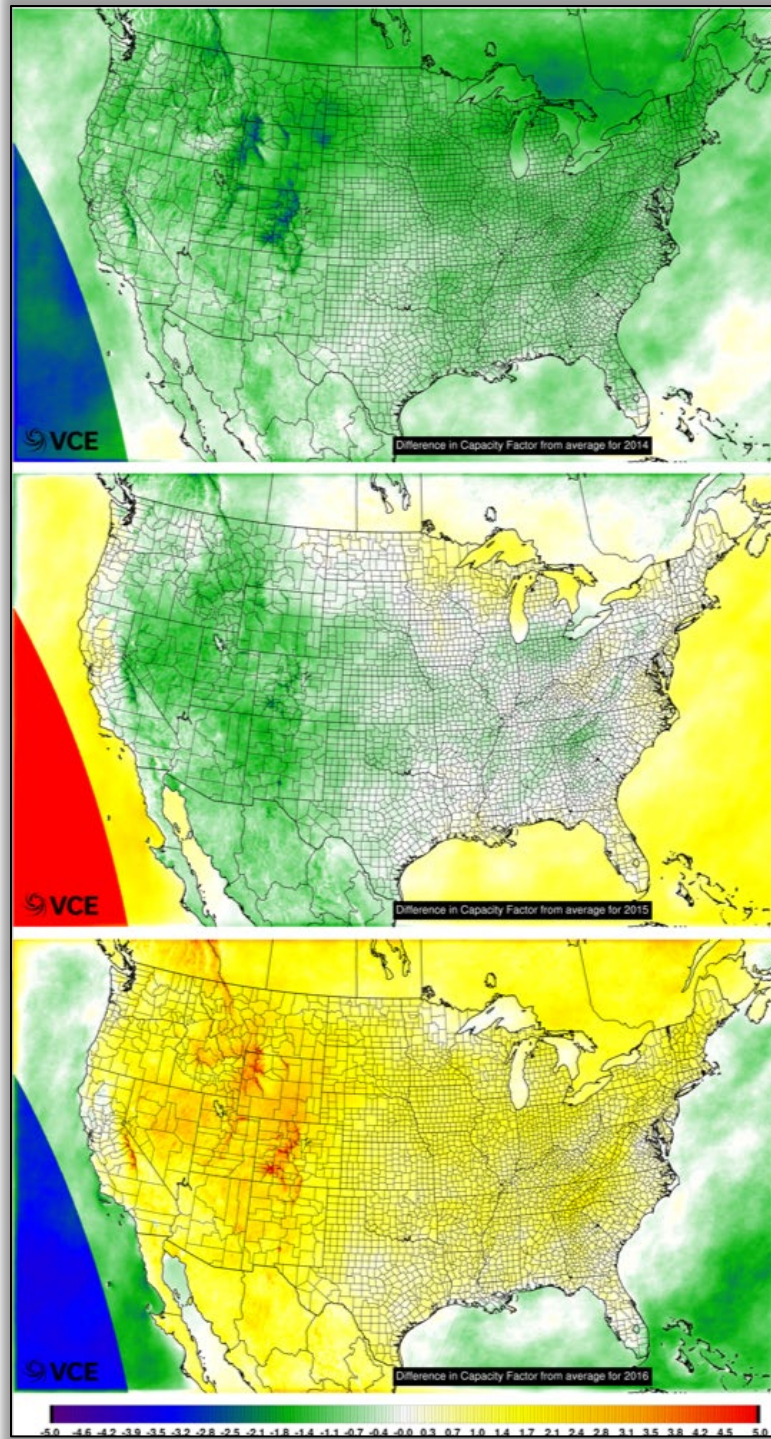
Figure 18: Difference in annual average capacity factor for a 2-axis PV panel compared to the three-year average.

The solar power capacity factors are calculated for the following technologies:
1. Rooftop Solar,
2. Fixed panels with a 0-degree tilt,
3. Fixed panels with a 15-degree tilt,

4. Fixed panels with a 30-degree tilt,
5. Fixed panels with a 45-degree tilt,
6. Fixed panels with a "latitude" degree tilt,
7. Single axis panels with a "latitude" degree tilt,
8. Dual axis panels.

The technologies for utility scale PV range from simplest and least cost (Fixed panels with 0-degree tilt) to most complex and highest cost (dual-axis panels). The fixed panels at 0-degree tilt will result in the lowest power capacity factors, while the dual-axis panels will result in the highest as they track the sun across the sky to ensure maximum possible power production (Figure 19a and 19c). WIS:dom® can determine using the weather data if the added complexity of the PV technologies is worth the additional cost in terms of increased power production. As seen from Figure 19b, Fixed panels at various elevation angles (with respect to the latitude of the geographic location) are the optimal choice for most of the CONUS with only the northern-most part of the country justified in using either single or dual axis tracking.
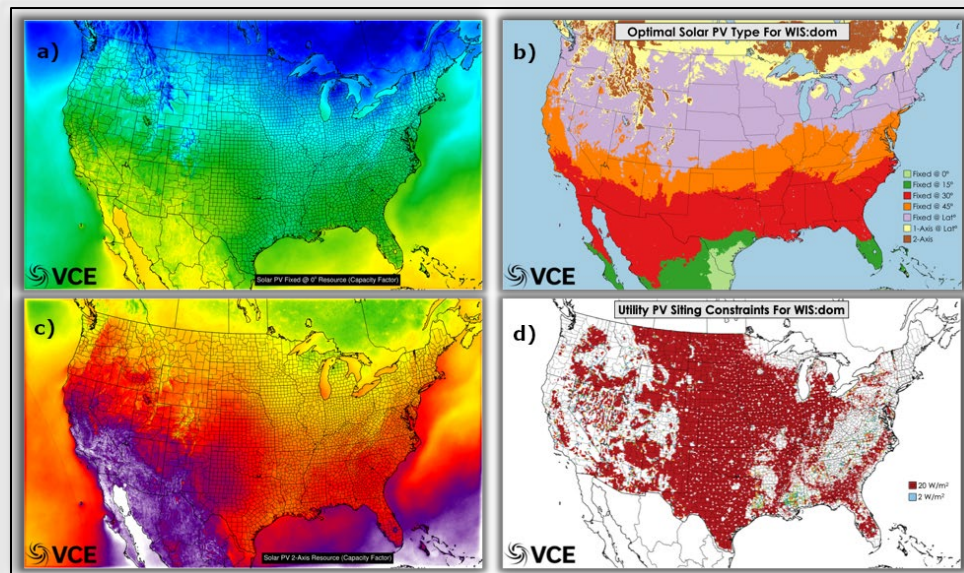
Figure 19: The solar power dataset. (a) Mean solar PV capacity factor for fixed panel at 0° elevation for year 2014 (b) Mean solar PV capacity factor for a two-axis tracking PV panel for year 2014 (c) Optimal PV panel type for the CONUS (d) Utility PV siting constraints for the CONUS

In Figure 20, the CV is shown for all the various combinations of the 5 years of Solar data available. The average CV of all events for each count, r, is also plotted. Each year will bring new weather events to the dataset that were not experienced previously, thus increasing the variability over time. This average value should increase as more years are added, as it does here. For Solar, the CV is > 100%. This shows that the standard deviation is larger than the mean of the resource.
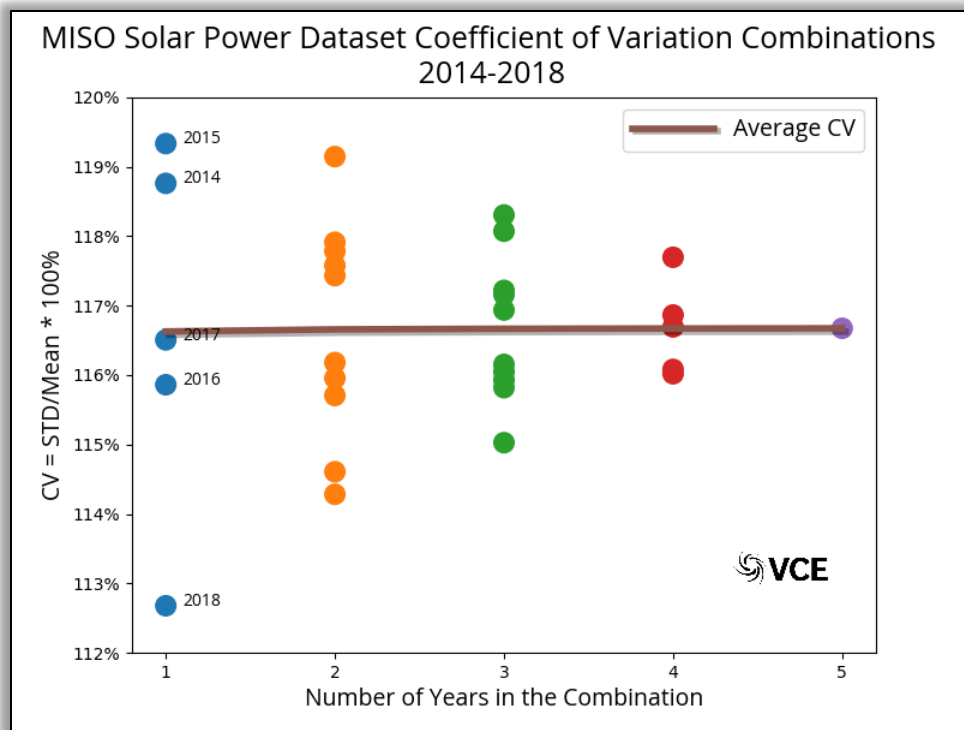
Figure 20: The Coefficient of Variation shown using 2014-2018 data over 1- to 5-year groupings for the Solar Resource across MISO. CV is shown for all combinations within each grouping.

When looking at the individual years, the year 2015 shows a higher extent of variability in relation to the mean. The average capacity factor from Solar for that year is one of the lowest years. The standard deviation though, is quite high. Combining these two together gives that higher extent of variability. Calendar year 2018 shows the lowest extent of variability in relation to the mean. The average capacity factors in 2018 for Solar were not the highest, but also not the lowest amongst all the weather years. The standard deviation was on the lower side and brought the CV down for that year. The solar dataset shows much lower inter-annual variability compared with the wind dataset.

6. Discussion on the Wind and Solar Datasets

VCE® restructures the spatial datasets discussed above and provides a time series product for every 3-km grid cell available across the CONUS. This allows for viewing the solar and wind power potential over an entire year, or in this case for MISO, every year for years 2014-2018. This is over 800,000 sites that can be viewed in this context across the CONUS.

The time-series analysis of the wind and solar resources bring out some interesting characteristics that can be leveraged by WIS:dom® to minimize cost through an optimal mix of these energy sources. In many regions of the US, wind and solar exhibit complementary behavior diurnally, solar being active during the day and wind picking up at night. This behavior is especially true in the great plains where the nocturnal low-level-jets start like clockwork during the stably-stratified nighttime periods. To demonstrate this complementary behavior, wind and solar power capacity factors from a site in Iowa are diurnally averaged over the year 2014. The resulting wind and solar power capacity factors from the diurnal average are shown in Figure 21a and demonstrate the complementary nature of the wind and solar resource. At this location, the correlation coefficient between the diurnally averaged wind and solar capacity factors over a year was found to be -95%. While the diurnal mean of the wind and solar power capacity factors show a smooth and almost perfectly anti-correlated behavior, the daily power capacity factors are much more variable (Figure 21b). The intra-day and the inter-day variability of wind and solar resource is the central challenge WIS:dom® has to solve to integrate renewables into the electricity system.
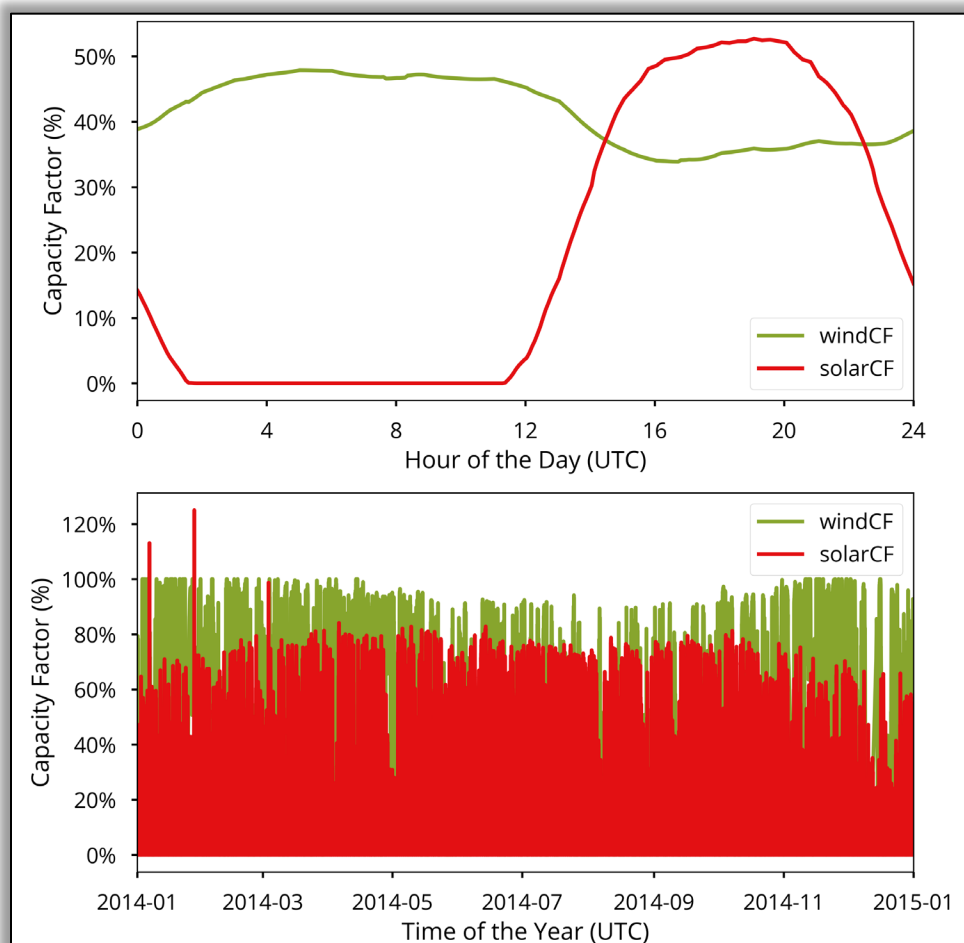
Figure 21: Comparison of wind and solar power capacity factors for a location in Iowa. (a) Diurnally averaged over year 2014 power capacity factors for two-axis tracking solar panel and 100-m hub height wind turbine. (b) 5-min power capacity factors from year 2014 for two-axis tracking solar panel and 100-m hub height wind turbine.

Another way to average out (apart from temporal averaging) the variability observed in the wind and solar resource is to spread the generation over larger geographic areas. This can be illustrated with the help of wind solar power capacity factors over the MISO region. A given location within the MISO footprint can have substantial intra-day and inter-day variability (as shown in Figure 21) due to local weather events such as scattered clouds or spatial variability in wind and temperature fields. However, as the power capacity factors are calculated over larger and larger geographical footprints, a significant portion of this variability gets averaged out and results in smoother profiles of wind and solar power capacity factors. This is illustrated in Figure 22a for one location within LRZ5 of MISO and the average solar power capacity factor over the whole of LRZ5. It is observed that the power capacity factor profiles for the LRZ5 average are much smoother than the individual location, although they could be lower in magnitude during some periods. Similar behavior is observed in the wind power capacity factors, where the intermittency due to wind lulls and gusts gets averaged out as the power capacity factor is calculated over a larger geographic footprint (Figure 22b).
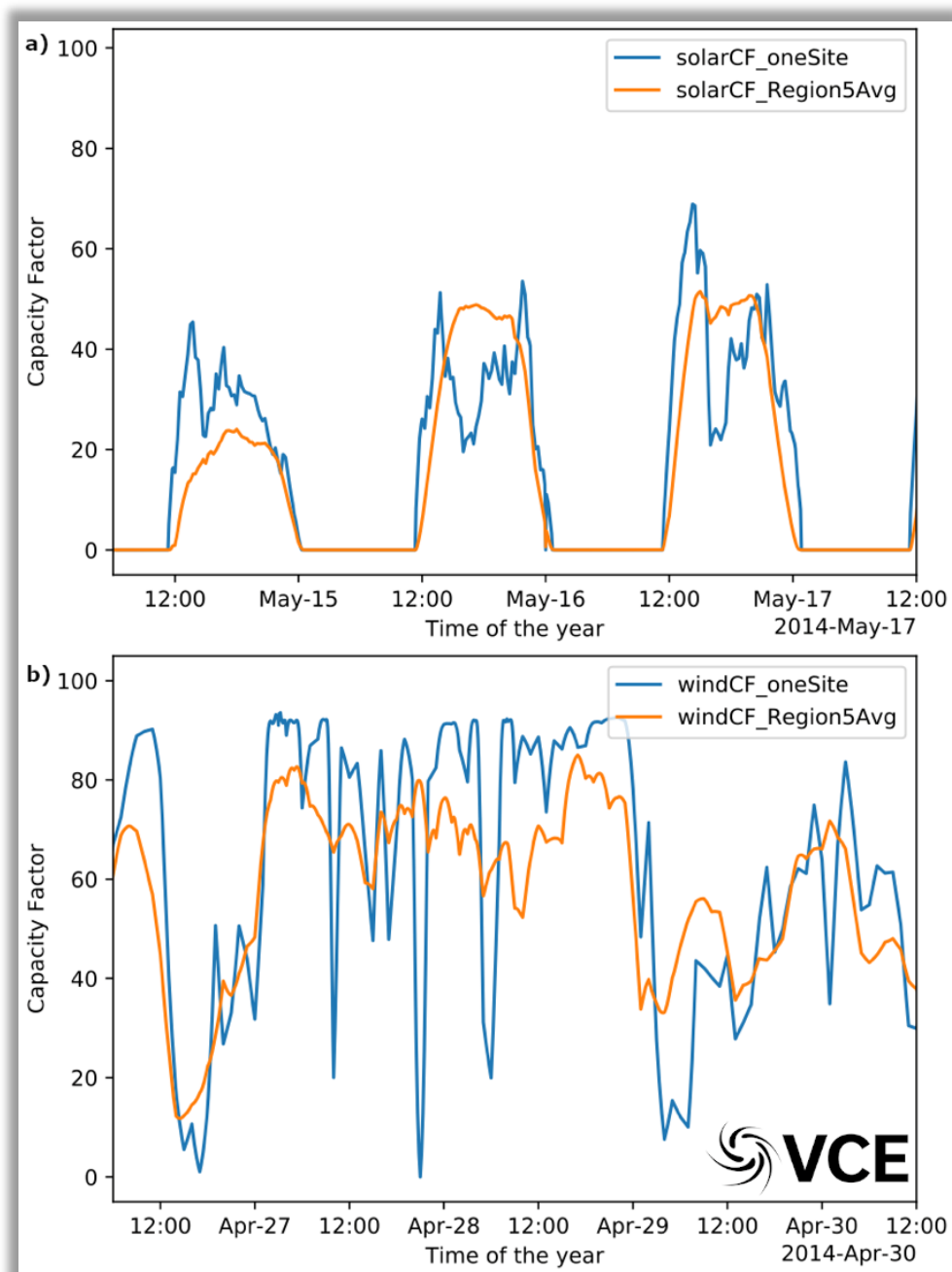
Figure 22: Comparison of solar power capacity factors from a location in region 5 of MISO versus the average solar capacity factor over the region 5. The times are in UTC.

Each individual MISO LRZ will still exhibit variability in the wind and solar power capacity factors. This variability within each MISO LRZ can be further reduced by increasing the geographic averaging over the whole MISO footprint. As shown in Figure 23, this further reduces the variability and results in much smoother profiles of wind and solar power capacity factors.
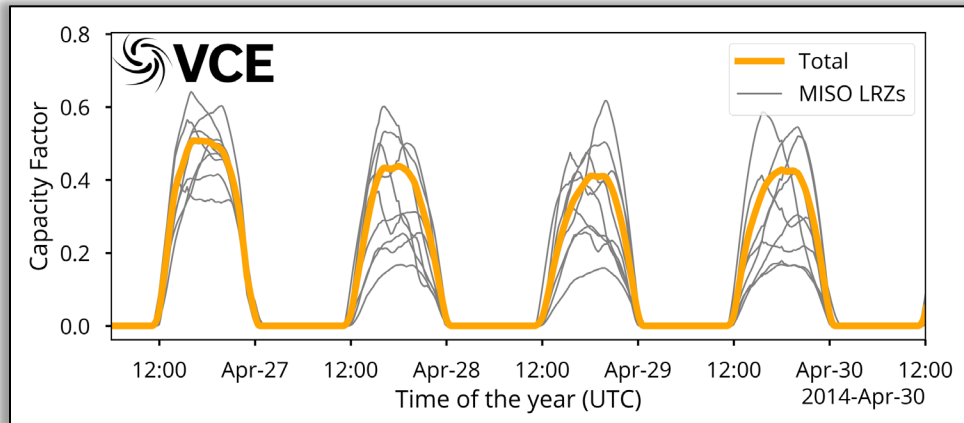
Figure 23: Comparison of the solar power capacity factors from the various MISO local resource zones with the solar power capacity factor over the full MISO footprint.

A third way to reduce variability in renewable energy generation is by combining generation from wind and solar. To demonstrate this, it is assumed that each HRRR grid cell within MISO has equal amount of wind and solar capacity installed. The coefficient of variability (CV) of the power capacity factors for wind, solar and wind+solar is now calculated over increasing geographic footprints. The geographic averaging starts from the central MISO LRZ (Region 5) and expands outwards symmetrically north-south until the full MISO footprint is covered. The CV of the power capacity factors for wind, solar and wind+solar over the increasing number of LRZs is shown in Figure 24.
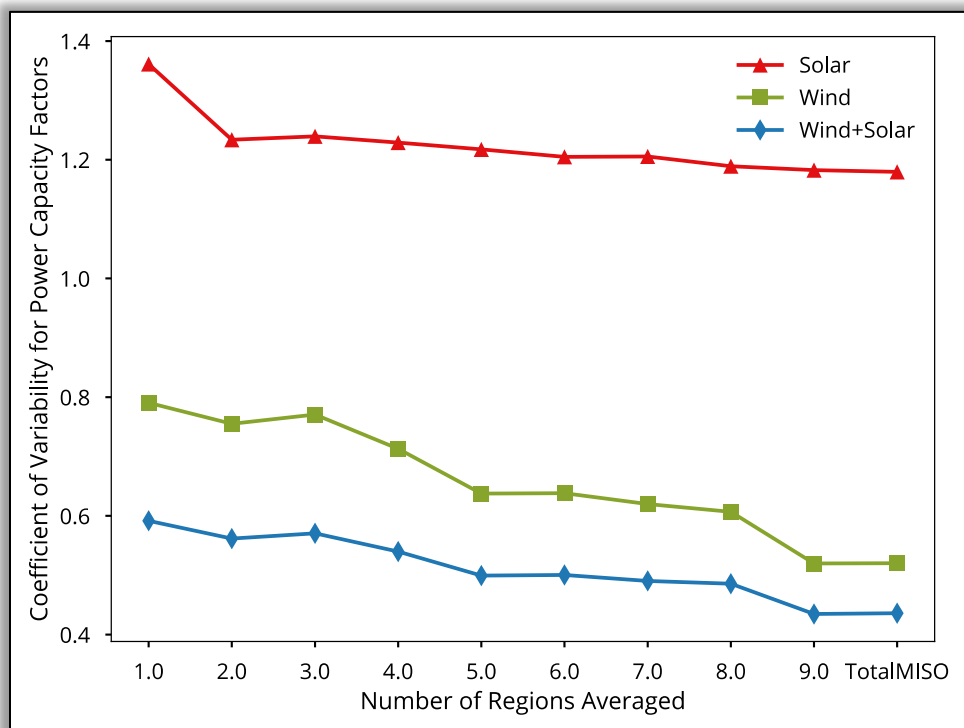
Figure 24: Coefficient of variability of power capacity factors of wind, solar and wind+solar calculated over various number of LRZs starting from the center (Region 5) and spreading outwards until the full MISO footprint is covered.

It is observed from Figure 24 that there is about a 30-percentage point reduction in variability in the wind power capacity factor as the geographic area under consideration increases from one region to the full MISO footprint. The variability in the solar power capacity factor starts out much higher compared to the variability in the wind power capacity factor due to the strong diurnal pattern in solar power. The reduction in variability in the solar power capacity factor due to increasing the geographic footprint is about 18-percentage points. While this is a significant reduction in variability, it would have been higher if the MISO region spanned a larger longitudinal extent, which would increase the amount of time solar insolation was available. Finally, it is observed that the coefficient of variability of combined wind and solar power capacity factors starts out being the lowest and decreases further to stay the lowest as the geographic area over which the CV is calculated is increased. The main reason the combined wind and solar CV is the lowest can be seen from Figure 21(a) where the diurnal variability of wind and solar generation gets reduced significantly once the two sources of generation are combined.

In addition to the above three ways to reduce the variability of the wind and solar generation, WIS:dom® can employ another technique to reduce variability, which is through strategically deploying wind and solar installation such that the installed generation weighted power capacity factors show a reduced variability. It is observed from Figure 25 that WIS:dom® does just that. Figure 25 shows a comparison of CV for the MISO regions, calculated as explained previously, with the CV of WIS:dom® installed generation weighted power capacity factor calculated for each investment period. For the scenarios modeled by WIS:dom®, the optimal capacity expansion occurs over ten investment periods, or iterations, to evolve the grid from its current state to the "goal" state. The model needs multiple investment periods because there are limitations to the amount of turnover of the resource mix that can happen in each iteration. Therefore, in the plots below, calculation of CV of the power capacity factors for the MISO region and for WIS:dom® installed capacity occurs in ten steps each. While each of the ten steps are not completely inter-comparable as WIS:dom® can choose to install generation at any location within the full MISO footprint, they show progress towards the end goal and the final step is completely inter-comparable.
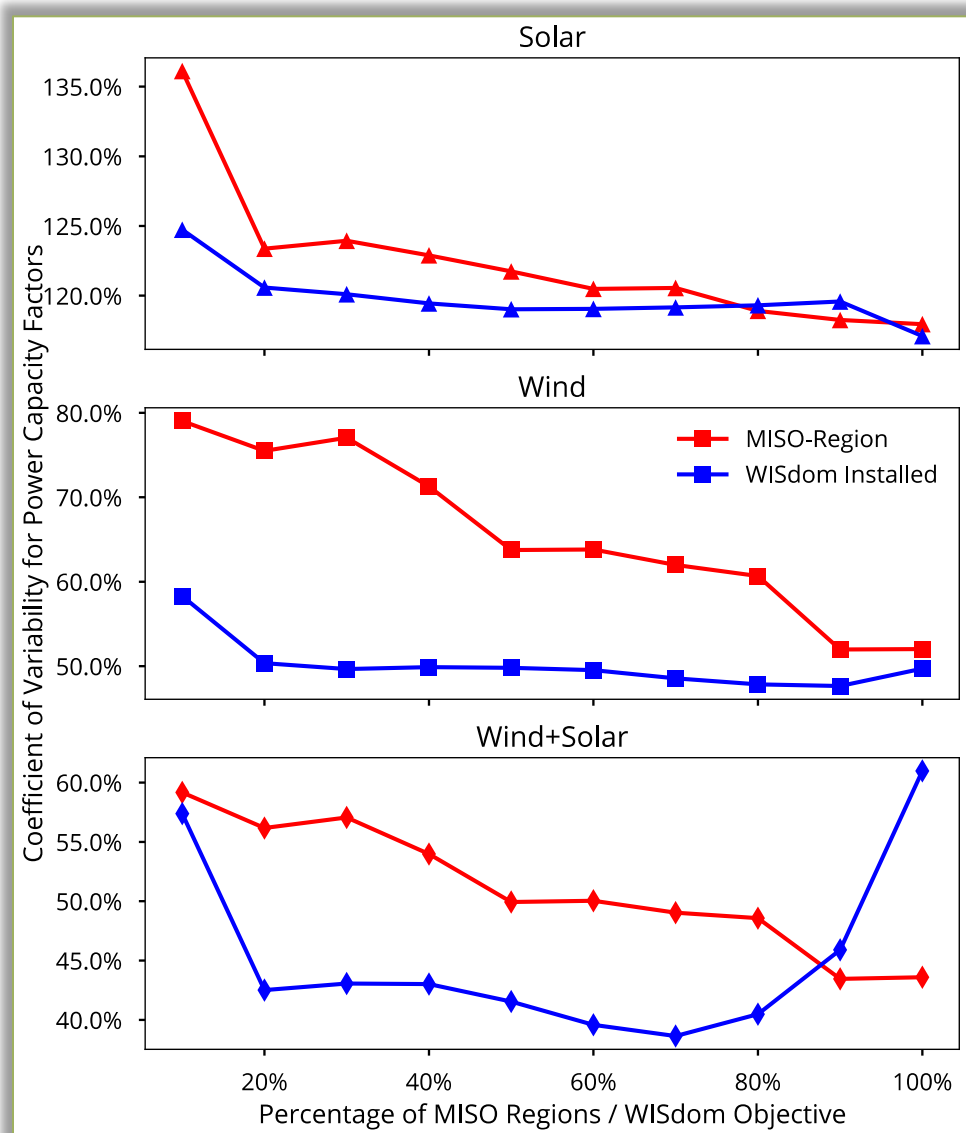
Figure 25: Comparison of coefficient of variability between wind/solar power capacity factors calculated for various coverages of the MISO region versus the WIS:dom installed capacity weighted power capacity factors calculated for increasing percentage of renewable energy penetration.

It is observed that the CV at the first step for installed capacity weighted power capacity factors for WIS:dom® scenarios are significantly lower than the CV calculated for the MISO region. The reason for this, as explained before is that WIS:dom® has the advantage of being able to build generation anywhere over the full MISO footprint. It is observed from Figure 25 (a) and (b) that the CV of WIS:dom® installed weighted power capacity factors quickly converges to the minimum possible value with departures from the minimum occurring only to meet a possibly difficult demand period.

Figure 25(c) shows an interesting feature of the CV for the combined wind+solar installed by WIS:dom®. It is observed that the coefficient of variability starts out low and reduces as the WIS:dom® run progresses. However, past 70% of the goal, the variability starts to pick and increases to about 12-percentage points above the starting value. The main reason for this is that as the model run progresses, WIS:dom® has to not

only reduce variability, but meet demand at lowest possible cost. The most effective way to do this, as it turns out, is by installing a lot of utility scale PV (at twice the rate of wind installation). The main effect of the increased utility PV is that the solar variability has a much higher weight on the combined wind+solar variability value, which results in an increase in the CV. It should be noted that the final CV value is still almost half the magnitude of solar only CV. This trend in the installed generation weighted power capacity factor illustrates the challenges involved in meeting 100% of the demand using renewables alone. Meeting 100% of the demand using only wind or only solar is even harder as seen from Figure 26.
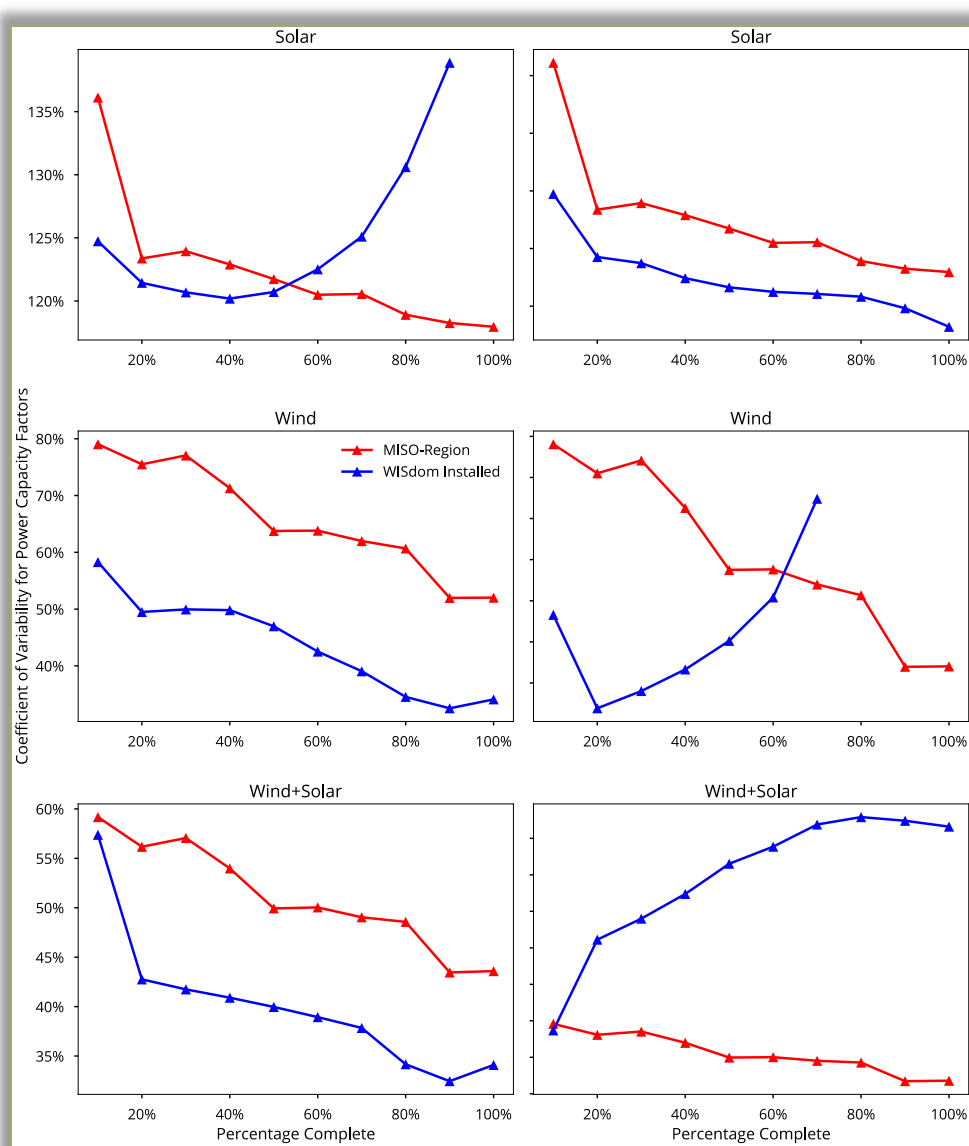
Figure 26: Coefficient of variability for scenarios of meeting demand through 100% wind (left) and 100 solar (right).

It is observed from Figure 26 that CV increases dramatically when WIS:dom® has to meet demand through either wind or solar alone. The increase in CV is significantly smaller when meeting demand through wind alone versus meeting demand through solar alone. The main reason for this being the MISO region has reasonably good wind resource which does not have as strong a diurnal trend as the solar resource.

Finally, when the whole Eastern Interconnect (EIC) is considered, WIS:dom® is able to site generation of wind and solar such that the CV value is further reduced. It can be seen from Figure 27 that the CV for WIS:dom® installed generation weighted solar capacity factor starts out being approximately 5% lower for the EIC region compared to MISO alone.
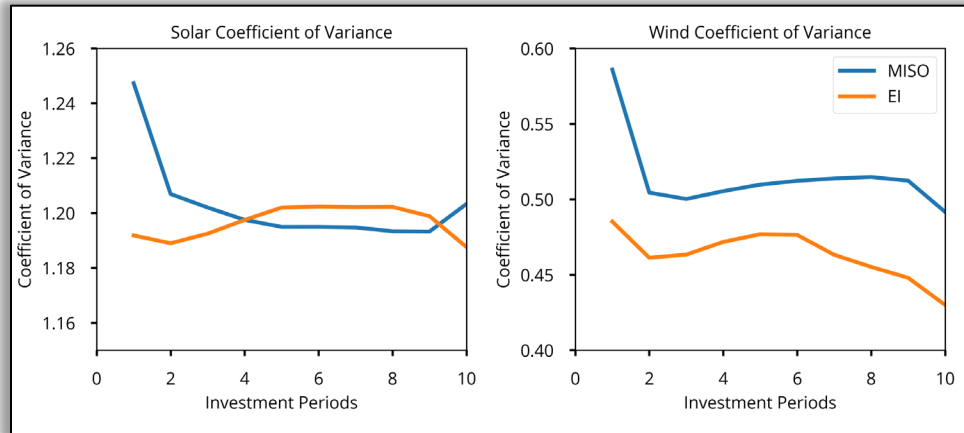

Figure 27: Comparison of coefficient of variance for the WIS:dom installed generation weighted wind and solar power capacity factors the MISO region versus the full EIC region.

It has to be noted that in all the above discussion, the impact of storage on variability is not considered. Therefore WIS:dom® can choose any of the above ways to handle the variability of renewable generation while meeting demand. In addition, it can build either storage or transmission to ensure energy is available where and when it is needed.

7. Supplying Demand Profiles Using VREs

Generation capacity is built on the basis of the maximum load that has to be met, which occurs only for a very small duration of time. This behavior of the load can be demonstrated using the load duration curve plotted using model output for the 100% through Wind and Solar scenario as shown in Figure 28. The supply duration curve for wind and solar are also shown in Figure 28. It is observed that in the initial investment period, wind and solar contribute only a small portion towards meeting the load. By the final investment period, wind and solar are required to meet 100% of the load and hence have substantial curtailment.

The combined wind and solar duration curves show some interesting characteristics. It is pretty much the same as the wind duration curve for the initial investment period as there is only a very small amount of solar on the grid. However, by the final investment period, there is significant contribution from solar. The combined wind and solar duration curve can be broken down into two parts. The portion of the curve to the left of the 50% mark on the x-axis runs parallel to the solar duration curve and dominated by solar. This behavior is expected as wind is usually weaker during the day and there is significantly more solar installed capacity compared to wind by the final investment period. To the right of the 50% mark on the x-axis we see that wind and solar work together to meet the load.
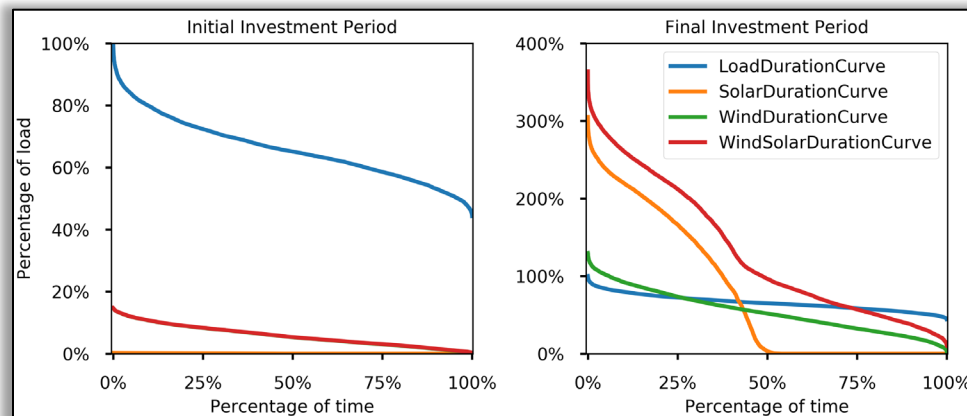
Figure 28: Duration curves as fraction of load for wind (green solid line), solar (orange solid line), load (blue solid line) and combined wind and solar (red solid line) for the initial and final investment periods.

It may seem from Figure 28 that there is significant curtailment of energy (mainly through solar). However, Figure 28 does not represent the energy that is sent to storage. To properly represent the stored portion of the energy, the duration curves are re-drawn with respect to the altered load, which is the demand plus the energy sent to storage. The duration curves with respect to altered load are shown in Figure 29.
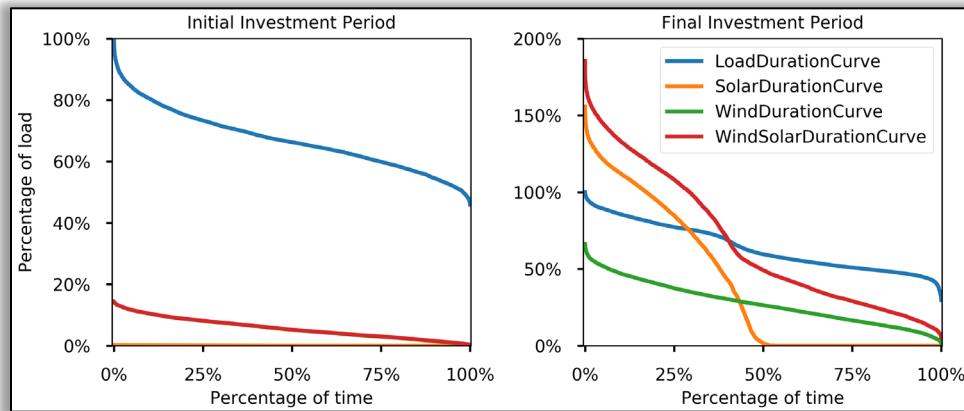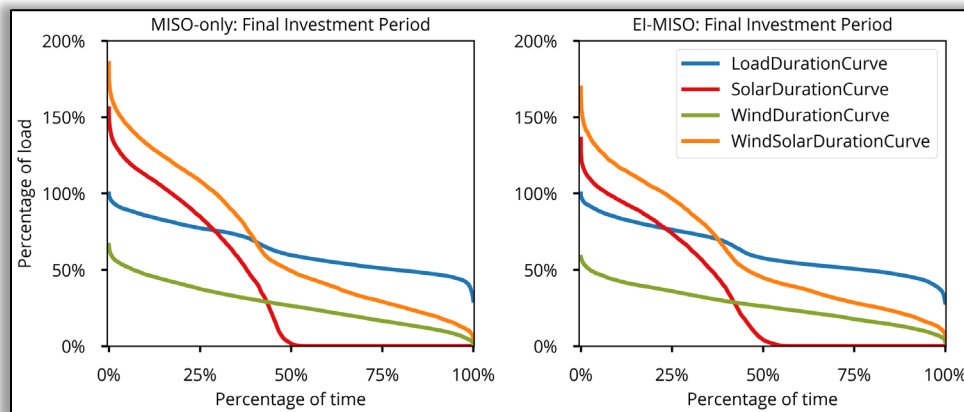
Figure 29: Duration curves as fraction of altered load for wind (green solid line), solar (orange solid line), altered load (blue solid line) and combined wind and solar (red solid line) for the initial and final investment periods.

As seen from Figure 29, there is no difference to the duration curves for the initial investment period as there is (almost) no storage on the grid. However, for the final investment period, there are substantial differences observed. The load duration curve has a hump in the middle portion which is from energy transferred to storage. The extremes of the altered load duration curve are not different to the original load duration curve. It is also observed that the magnitude of energy actually curtailed is now accurately represented. It is observed from Figure 29 that peak energy curtailment is about 75% of the altered load.

The duration curves for wind, solar and load as percentage of altered load for the MISO only region and combined MISO and EIC are shown in Figure 30. It is observed when the MISO+EI region is considered together, the curtailment percentage is reduced appreciably.

Figure 30: Duration curves as fraction of altered load for wind (green solid line), solar (orange solid line), altered load (blue solid line) and combined wind and solar (red solid line) for the MISO region only and combined MISO and the Eastern Interconnect.

An important consideration in planning generation is to ensure that it is available when it is needed most, during the peak demand periods. When a significant portion of the demand is met through VREs, it can help to "shave demand peaks" and thus increasing grid stability. It is useful to study the percentage of demand satisfied by a generation type in order to understand their relative importance in meeting load during the highest demand periods. The generation as a percentage of the original load and net load is plotted in Figure 31. The net load is defined as the original load minus the generation from other sources, DSM and storage flux. It is observed that at the final investment period, there is significant amount of over

generation from wind and solar. This over-generation does not capture the energy movement from and to storage and DSM. Therefore, a better way to understand the contribution of a generation resource is by plotting the generation as a percentage of net load as shown in bottom panel of Figure 31.
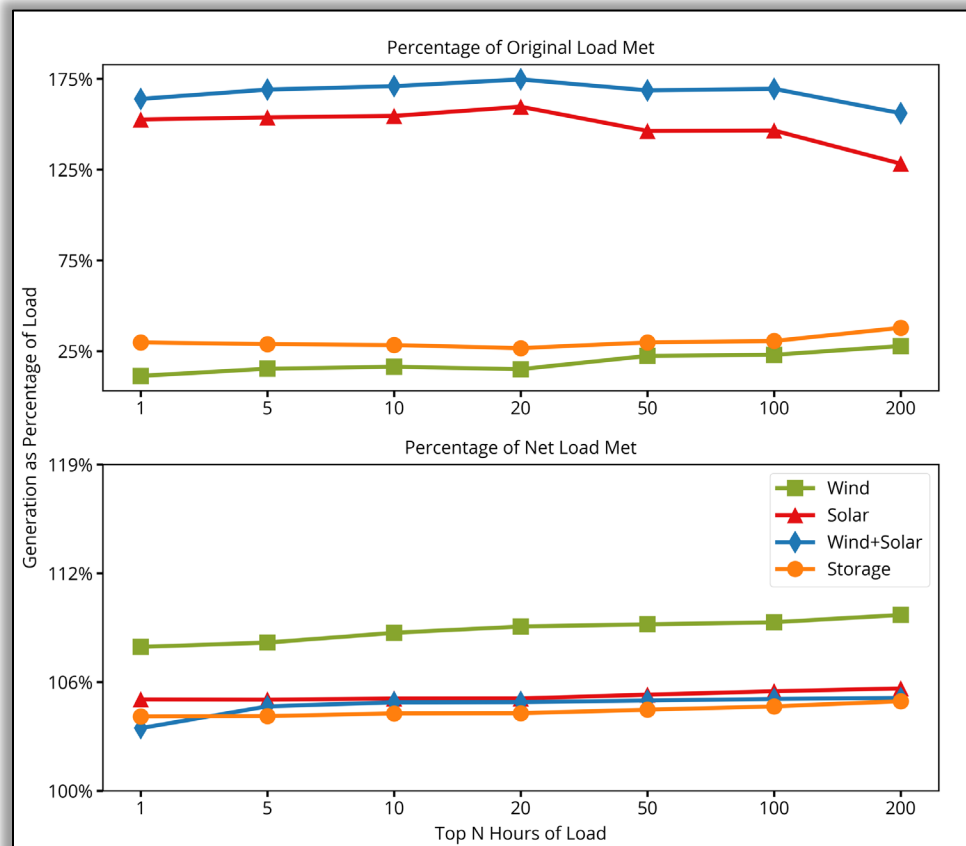

Figure 31: Generation as a percentage of load during the highest few demand hours

When the contribution of each resource during the highest demand periods is plotted as a percentage of the net load, the importance of each resource becomes much clearer. It is observed that both wind and solar play equally important roles in ensuring the demand is met. It should be noted that when considering net load, the actual time periods of highest net load are different when considering wind, solar, combined wind, solar and storage. This is due to the generation from wind+solar as percentage of net load being lower than similar metrics of wind or solar alone.

It is observed that wind, solar and storage is dispatched at more than 100% of the net load. The reason these dispatches are larger than 100% of the net load is due to the fact that there are losses associated with transmission, and round-trip efficiencies (for storage). Therefore, the additional percentage above 100% accounts for the inefficiencies associated with getting the energy to where the demand exists. Similar analysis is done for the full EIC region (Figure 32) and it is found that there was no significant difference in behavior of generation from the various sources. It is observed that in the full EIC region, wind dispatch is approximately 115% of net load. The reason for this is the larger distances wind needs to be moved to get to the load centers on the east coast.
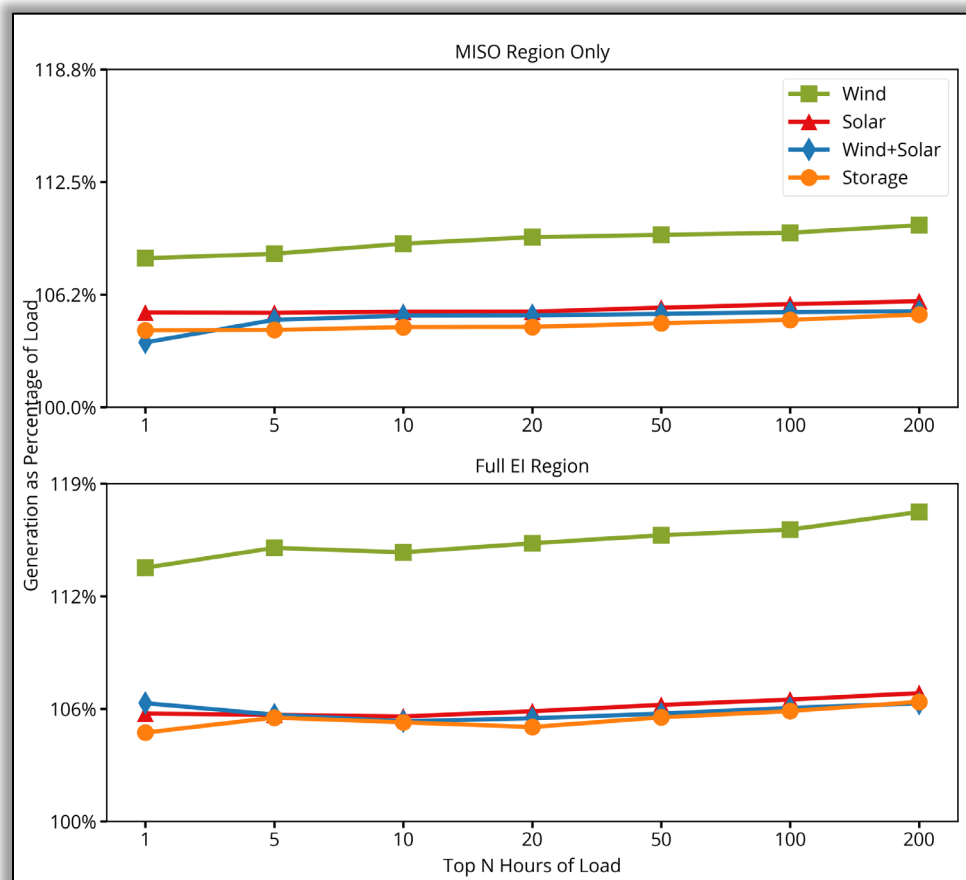
Figure 32: Generation as percentage of net load for MISO region only (top) and for the full EI region (bottom).

It is interesting to study the trend in contribution of generation from the various resources as a function of investment period. Figure 33 shows the percentage of load and net load met through various sources of generation during the highest ten periods of demand. As explained previously, when contribution from each source of generation is plotted, it gives an impression that wind and storage are not playing as important a role as solar during periods of highest demand. In addition, it conveys that a large amount of over generation (250%) is taking place to meet demand. However, the reality is a bit more nuanced. When contributions of generation from each resource are plotted as function of the net load, it is seen that each resource plays an important part in meeting the load. In addition, it becomes clear how storage starts to get deployed to meet periods of high demand increasingly towards the later investment periods and eventually being used to meet 100% of the net load at the final investment period.
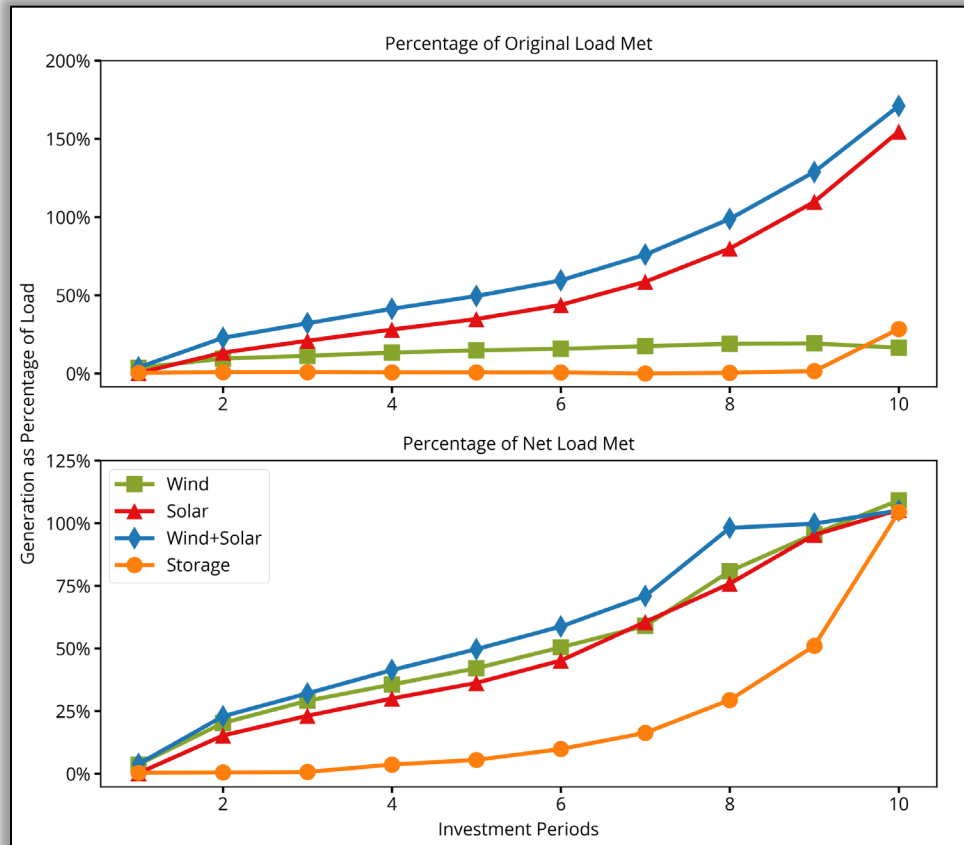
Figure 33: Generation as percentage of load (top) and net load (bottom) for the highest 10 hours of demand.

Figure 34 shows a comparison of generation contributions from each source for the MISO only scenario and the full EIC region. For brevity, the generation as percentage of net load is only shown for the two scenarios.
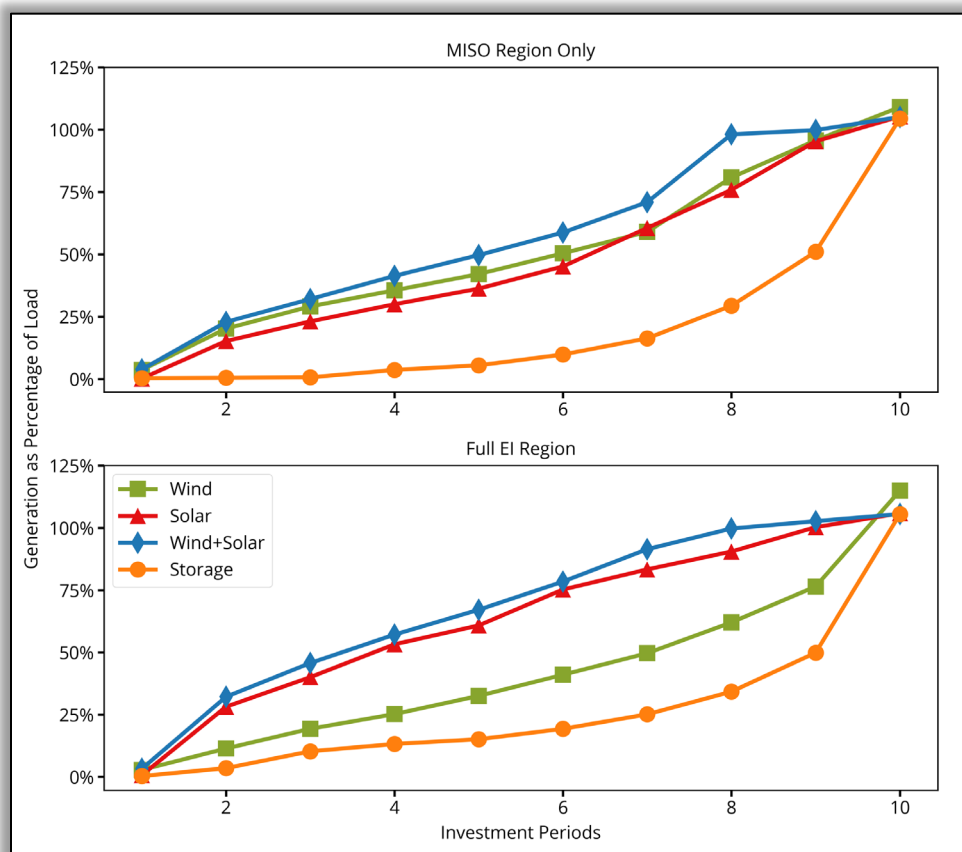
Figure 34: Generation as percentage of net load for MISO region only (top) and for the full EI region (bottom).

It is seen from Figure 34 that there are appreciable differences in resource mix to meet difficult demand periods between the MISO only case and the full EIC case. For the MISO only case, wind plays a more important role compared to the full EIC case. This is to be expected as MISO has substantially better wind resource on average versus the full EIC region. In addition, space for wind farm development is much more constrained in the eastern United States due to higher population densities. Therefore, storage is seen to come into play earlier as well as ramp up slightly faster compared to the MISO only case.

The contribution of a generator in meeting peak demand is known as resource adequacy. The fraction of a generator's nameplate capacity that can be counted towards resource adequacy is known as capacity credit. The capacity credit for a particular generator can be estimated using the duration curve method. The duration curve method estimates capacity credit as the reduction in highest peak net load hours relative to the highest load peak hours. The net load for a generator is calculated by subtracting the generation that went to meeting load from that generator from the full load. Using the duration curve method, the capacity credit per MW of installed capacity and incremental capacity credit per incremental MW of generation added for each technology can be determined as the model steps through towards the end goal.

Figure 35 shows the capacity credit per MW of installed generation and the incremental capacity credit per MW of generation added for wind, solar and wind+solar for the full EI region in the 100% wind+solar scenario. It is observed that as the penetration of renewables increases, the capacity credit for all renewable generation steadily decreases. There is almost a 30% drop in the solar capacity credit going from 10% penetration to 20% due to the fact that the installed solar generation increases by 3,800% in this step and it goes from being a negligible part of the generation mix to a more significant portion. Storage shows a

different trend compared to the renewable generation. It is seen that capacity credit of storage increases from the first step to the second, during which some firm generation such as coal, CCGT and CT are retired and hence increase the value of storage. From the second iteration to 40% of renewable penetration, the capacity credit from storage is seen to decrease, which is from a combination of wind and solar installations that meet demand. From 40% penetration level to 100% the value of storage starts to increase again as it starts being used to firm up the renewable generation.
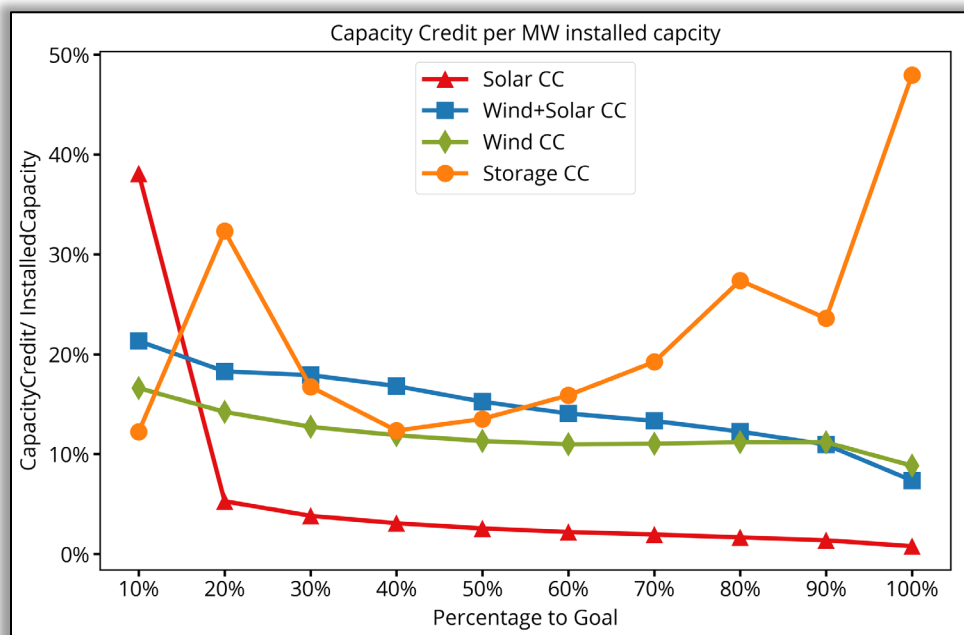
Figure 35: Capacity credit per MW of installed capacity for wind, solar and wind+solar, calculated over the full EI region for 100% wind+solar scenario.

Given the trend in capacity credit seen from Figure 35, it gives the impression that adding more renewables reduces their effectiveness on the grid. However, a different trend emerges when the capacity credit is calculated based on the altered load. Figure 36 shows the capacity credit for the various renewable generation penetrations calculated using the altered load duration curve. It is seen from Figure 36 that the solar capacity credit still drops from 10% penetration to 20%, but then starts to gradually increase when we go to 100% penetration. The capacity credit for wind+solar stays constant irrespective of the penetration level, which illustrates how WIS:dom® strategically deploys wind and solar to get maximum benefit with respect to meeting peak demand.
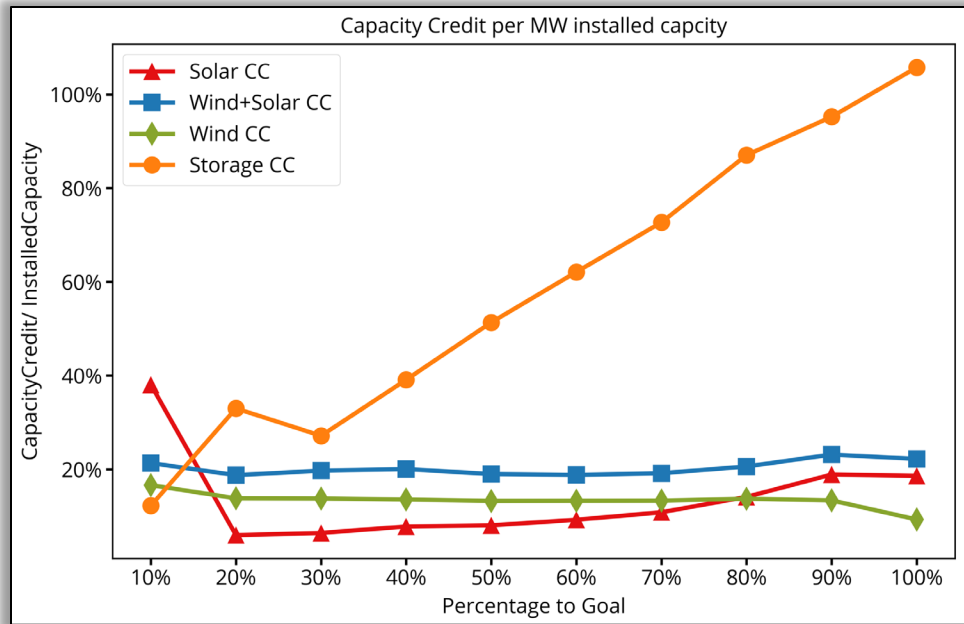
Figure 36: Capacity credit calculation from altered load.

It is observed that storage shows a different trend as well. Similar to Figure 35, the jump in capacity credit of storage is still observed going from penetration level of 10% to 20% from retirement of other firm generation. However, it is seen that storage capacity credit continues to grow linearly until it reaches 105% at 100% wind and solar penetration. The storage capacity is at 105% to account for transmission losses and round-trip efficiencies. Therefore, it is seen that as other forms of firm generation are retired, storage reaches its full capacity credit to firm up the generation from renewable sources.

The contribution of wind, solar and storage in meeting daily peak altered loads are presented in Figure 37. Figure 37 shows the fraction of the daily peak altered load satisfied by a generation type during the peak load hour over the course of the year and increasing renewable energy penetration. It is observed from Figure 37 that solar contributes the most during peak load hours during the summer period, where the peak load hours are usually during the day time. It is also observed that the contribution from solar at the peak load hour reaches its maximum value at 90% renewable penetration. At 100% renewable penetration, some of the peak demand is met by storage as it is more advantegeous to deploy compared to adding more solar.
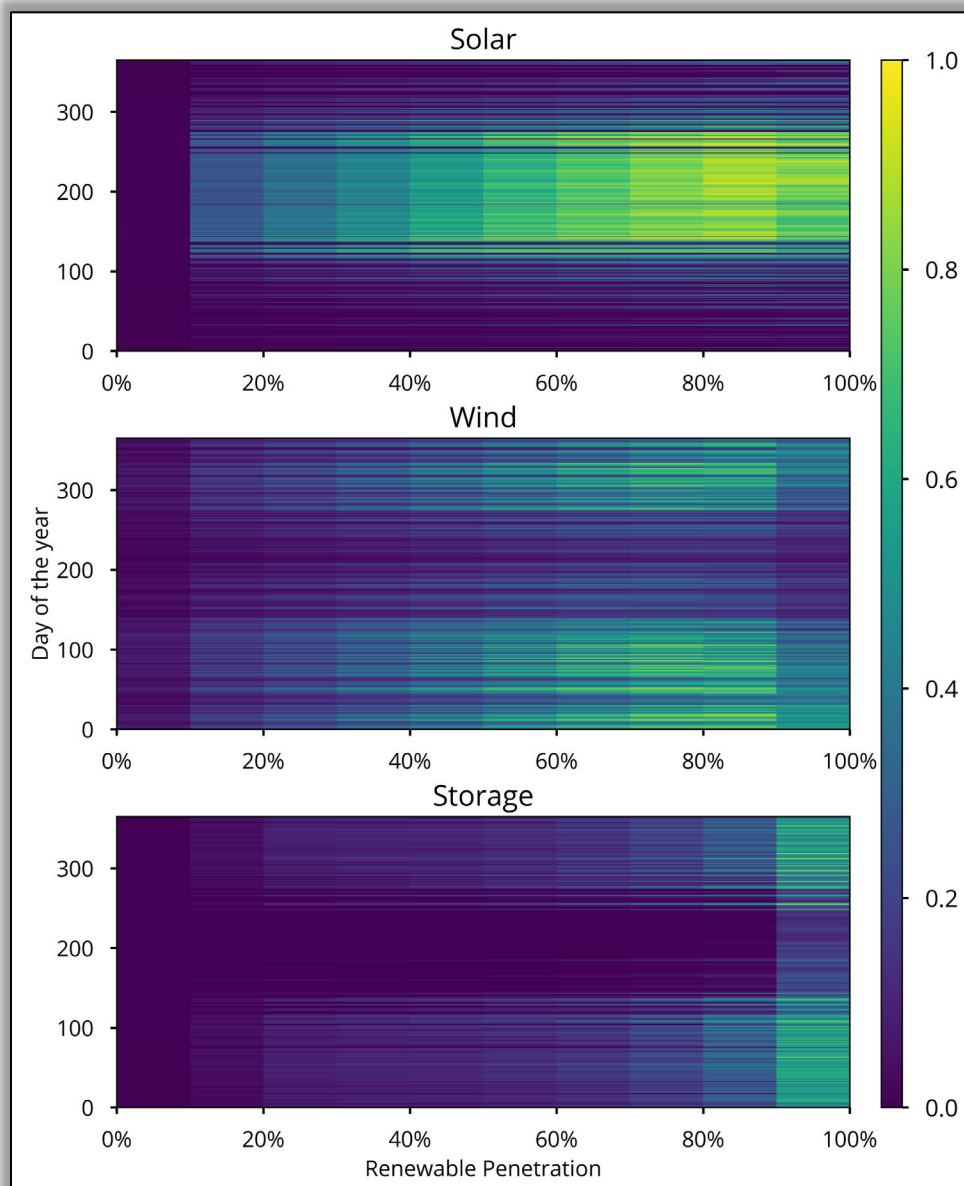
Figure 37: Generation from wind, solar and storage as a fraction of daily peak altered load over a one-year period and with increasing renewable penetration levels. The sum of fractional generation from wind, solar and storage at 100% penetration add up to 1 or greater as there is no-longer any other generation type.

Wind and storage dominate generation during peak load hours during the winter, late fall and early spring periods. The reason for this is that the peak load hours are shifted to evening periods during these seasons where generation from wind dominates. However, wind is unable to meet all the demand and uses storage to meet the rest of the peak load. Similar to solar, it is seen that wind reaches its maximum peak load fraction between 80% and 90% renewable penetration levels. Beyond this it is more advantageous to deploy storage to meet peak demand.

8. Wind and Solar Forecasts

In addition to understanding and having insight into wind and solar capabilities across the United State with the power datasets, it is also prudent to know what was forecasted historically as well. This can provide further insight into the behavior of wind and solar dispatch in real-time. VCE® has provided wind and solar forecasts to MISO for five years, 2014 through 2018. The Forecast Hours 02 and 06 were provided.

The forecasts are based on the NOAA HRRR weather forecast model. In the case of missing HRRR forecast output in the VCE® historical archives, the forecasts for the missing periods were created such that no missing data exists in the datasets. For wind, any missing gaps with a given year are filled using linear interpolation. For Solar, persistence is utilized where the corresponding hour of the previous day fills the missing value of the current day. The forecast data is available at hourly time intervals from the model runs made at the top of every hour.

The process for creating the forecasts is very similar to that of the power datasets overall. One exception is 2017 Forecasts, which is described in more detail below. The forecasts are originally processed spatially in the same way as the power datasets. This is reviewed in great detail earlier in this document. The forecasts are also "restructured" to provide temporal data for over 800,000 sites across the CONUS. As an example, a file for Site 814293, shows the running Hour 02 forecast for that day. A similar file exists for the Hour 06 forecast. These can be used to see changes in the forecast over time, but also in comparison to the power dataset or any observations known. For further explanation, Forecast Hour 02 is a forecast generated on January 27th 2015 0800 UTC that is valid for 10 UTC on the same day. Forecast Hour 06 is a forecast generated on January 27th 2015 0800 UTC that is valid for 14 UTC on the same day. All metrics shared below are for a 100-meter hub height for the wind and a single-axis panel for solar.

There are some differences in processing between the Forecasts and the power datasets. As we have improved our own processing over time, the more recent forecasts for 2017 and 2018 also have improvements in the final deliverables that were not yet available when 2014-2016 Forecasts were delivered. A high-level description of each year is provided here. Much of this will be pertinent for the forecast metrics provided further below in this report.

8.1. Forecasts Hour 02 and Hour 06 (2014 – 2016)

The main difference in methodology when creating these Forecasts for these years compared with the power dataset calculations is in the solar power forecast calculations. As described in the previous sections, before calculating the solar power capacity factors, the solar irradiance quantities are regressed with the ground-based observations to remove bias errors caused by inaccurate representation of the clouds in the model. For the forecast datasets of 2014-2016, a new regression is not performed. Instead the regression coefficients from the power calculations of the same year are used. This ensures that the regression only corrects for the model bias observed in the initialization and does not correct for the errors due to inaccurate model physics. This method ensures that the power forecasts retain the error characteristics expected from real-time forecast runs. This was updated in later Forecast years, however, as we will describe below.

Neither the Forecasts or power datasets for these three years have the following updates for the wind Resource:
1. Introduction of the effects of turbulence
2. Capturing turbine response to air density changes

Neither the Forecasts or power datasets for these three years have a Bi-Facial PV series in the Solar Resource options.

8.2. Forecast Hour 06 (2017)

For this year, the source VCE® used for historical HRRR Forecast files does not have actual Forecast Hour 02 or Forecast Hour 06 data to pull into VCE®'s historical archives. Our source did have Forecast Hour 00. This is what the Power Dataset is based off of for this year. Since other Forecast Hours were not available, a creation of these series was performed. As such, there will be some behavior in this year that is not as uniform or congruent with the Forecasts provided for other years. This will be apparent in many of the metrics shown below as well. To create the Forecasts for this year for both the Wind and Solar Resource, VCE® executed the following:

Wind:

We calculated the Percentage Error in 2018 between the Forecast Hour 06 and the Power Dataset and applied it to the Power Dataset in 2017. The Power Dataset for 2017 was based off of Hour 00, which was available historically:

$$Percentage\ Error = \frac{Forecast\ 2018_{Hour\ 06} - Power\ Dataset\ 2018}{Power\ Dataset\ 2018}$$

The Percentage Error is then applied to the Power Dataset of 2017 to Create Forecast Hour 06 of 2017.

$$Forecast\ 2017_{Hour\ 06} = Percentage\ Error * Power\ Dataset\ 2017 + Power\ Dataset\ 2017$$

In some situations, this method can create unrealistic results. To limit what the Percentage Error can do the following stipulations were applied:
1. When the Power Dataset 2018 == 0, the Percentage Error == 0.
    a. This will artificially lower the errors in 2017 since the 2017 Hour 06 Forecast will be equal to the Power Dataset for that year when the criterion is met. This artifact will be seen in the error metrics further down.
    b. Setting this stipulation, prevents division by 0.
2. When the Forecast 2018 for Hour 06 == 0, the Percentage Error becomes -1.0. Large negative biases can arise in the Forecast from this if a bound is not placed on the 2018 Forecast Hour 06. VCE® caps this behavior with the following:
    a. If (Forecast 2018 for Hour 06 == 0) AND (Percentage Error < -0.94), the Percentage Error == 0.
    b. This will also artificially lower the errors in 2017 since, once again, the 2017 Hour 06 Forecast will be the same as the Power Dataset.
    c. This improves the Magnitude Bias (defined below, but in short is the average Forecast divided by the average of the Power Dataset) of the Forecast Hour 06 in 2017. Without this bound, the Forecast Hour 06 becomes low biased compared to the Power Dataset. This is opposite to any other forecast year provided, where the forecast values for wind are generally higher than the power dataset. With this bound, the Magnitude Bias lines up with other Forecast years in terms of direction. The Magnitude Bias comes closer to 1.0 in 2017 where the average of the Forecast in equal to the average of the power dataset. This means the Magnitude Bias is also improved in 2017 due to lower forecast errors.
3. The previous two bullets innately reduce the Forecast error that can be seen in 2017. To further aim to limit the impact of the previous two bullets, VCE® averaged the Percentage Error before and after periods where the Power Dataset 2018 == 0. The same is done around periods when Forecast

2018 for Hour 06 == 0 and Percentage Error < -0.94. There are periods where the Power Dataset of 2018 are flatlined at zero for some time where this method will not make any changes. As such, errors for 2017 will still remain lower than 2018. However, during testing we saw this additional step bring in some more realistic error for about 3-4% of the year.

4. If left unbounded, the Percentage Error can become quite large. VCE® caps this calculation to +/- 0.95. This means that the error in the 2017 Forecast for Hour 06 cannot be greater or less than +/- 95% of the Power Dataset in 2017.
   a. This will also artificially lower the errors in 2017 as it is possible to have higher differences. However, this also limits periods where the Percentage Error can become quite large and also create some unrealistic differences.

5. For quality control:
   a. If the created Forecast Hour 06 > 100% (meaning 100% capacity factor), it is set to 100%.
   b. If the created Forecast Hour 06 < 0% (meaning 0% capacity factor), it is set to 0%.

There were several updates VCE® made to the 2018 Wind Forecasts. Those are described below in detail. These changes would innately reflect in the metrics of 2017 since 2018 performance is used to generate 2017 Forecasts.

Solar:

We calculated the Percentage Error in 2018 between the Forecast Hour 06 and the Power Dataset. It starts similar to wind but then some additional steps are taken for this resource. The Percentage Error is then applied to the Power Dataset of 2017 to Create Forecast Hour 06 of 2017.

In some situations, this method can create unrealistic results. To limit what the Percentage Error can do the following stipulations were applied:
1. If left unbounded, the Percentage Error can become quite large. VCE® caps this calculation to +/- 0.95. This means that the error in the 2017 Forecast for Hour 06 cannot be greater or less than +/- 95% of the Power Dataset in 2017.
   a. This will also artificially lower the errors in 2017 as it is possible to have higher differences. However, this also limits periods where the Percentage Error can become quite large and also create some unrealistic differences.

2. For quality control:
   a. If the created Forecast Hour 06 > 100% (meaning 100% capacity factor), it is set to 100%.
   b. If the created Forecast Hour 06 < 0% (meaning 0% capacity factor), it is set to 0%.

After the above steps, the resulting 2017 Forecast for Hour still has a very high bias. It creates power values that are not physically possible. To further create realistic values, VCE® applies the following:
1. A daily max time series is created from the Power Dataset of 2017. A fitted curve (spline) is then applied to this max time series. If, at any point throughout the year, the created spline is less than the max time series, the spline is snapped up the max time series. This is performed because if the spline was lower than the max time series, it would have artificially suppressed the values that could actually be observed.

2. If the calculated 2017 Forecast for Hour 06 is greater than 5% of the spline values, the spline values are used. This helped keep both some shape provided from 2018, but also allows application of the characteristics of the 2017 Power Dataset.

In 2018, VCE® updated how regression was applied with the ground-based observations to remove bias errors caused by inaccurate representation of the clouds in the model for the Solar Forecasts. In 2018, VCE® took a more realistic approach to this. In general, a forecast may not have all of the information and data used to create the regression coefficients in real-time. The forecast would have regression coefficients from a previous year available. The most similar year we had to 2018 for Solar was 2016 in terms of how observation data was used to improve the Solar Forecast and model cloud biases. VCE® applied the regression coefficients from 2016 to the Solar Forecast data in 2018. This created a slight bump up in error as seen further below in the metrics. However, this does create a more realistic dataset as to what would actually be available in real-time for energy balancing/trading. Since the 2017 Hour 06 Solar Forecasts were created from 2018 Percentage Errors, this new regression application will also be apparent in the behavior of the Solar Forecasts for 2017.

Lastly, a Bi-Facial Solar series is available in 2017. This was not the case for 2014-2016.

8.3. Forecast Hour 02 (2017)

VCE® updated how Forecast Hour 02 was created starting in the 2017 year. The Power Dataset is weighted heavily on Forecast Hour 02. From 2017 forward, missing data is far reduced in the VCE® HRRR data archives. A Forecast Hour 02 Forecast and the corresponding Power Dataset would be virtually the same. A new Forecast Hour 02 is now created as an interpolation between the Power Dataset and the Forecast Hour 06. For all intents and purposes, the new Forecast Hour 02 is likened to an "Hour 03" series.

Wind:

After we interpolate between the Power Dataset and the Forecast Hour 06, the following is applied:
1. We introduce some small turbulence here to the newly created Forecast. When the Power Calculations are zero, we do not add any in any turbulence. This creates events that are not realistic for an Hour 02 Forecast. This was performed based on the following.
    a. The previous historical performance in Wind between Hour 02 and Hour 06 and the Power Dataset was analyzed for 2014-2016. On average, the Root Mean Square Error (RMSE) was roughly around double for Hour 06 when compared to Hour 02. Turbulence was added to match standard deviation of errors observed in previous years with the following:

$$\sigma_{Turbulence} = \sqrt{(Goal\ Standard\ Deviation_{Hour\ 02\ Error})^2 - (Current\ Standard\ Deviation_{Hour\ 02\ Error})^2}$$

    b. The Current Standard Deviation in Hour 02 Error is what came straight from the interpolated Hour 02 Forecast.
    c. The Goal Standard Deviation in Hour 02 Error is what is needed to get to a Standard Deviation in Error for Hour 02 that would be about half of the Standard Deviation in Error for Hour 06.
    d. In general, we added at random about 0-4% in capacity throughout the year. We allowed slightly less magnitudes to be added in 2017 as opposed to 2018. This was due to the fact that errors were already artificially suppressed from the Forecast creation process described above for 2017.
2. For quality control:
    a. If the created Forecast Hour 02 > 100% (meaning 100% capacity factor), it is set to 100%.
    b. If the created Forecast Hour 02 < 0% (meaning 0% capacity factor), it is set to 0%.

Solar:

This is much more simple and straight-forward than what we do for wind. After we interpolate between the Power Dataset and the Forecast Hour 06, the following is applied:
1. For quality control:
    a. If the created Forecast Hour 02 > 100% (meaning 100% capacity factor), it is set to 100%.
    b. If the created Forecast Hour 02 < 0% (meaning 0% capacity factor), it is set to 0%.

8.4. Forecast Hour 06 (2018)

VCE® performed several updates this year to both the Wind and Solar Forecasts.

Wind:

Several improvements have been added to both the Power Datasets and the Forecasts regarding the Wind Resource. These methods were described in detail under the Power Datasets section of this report. At a high-level these were the following improvements:

1. Introduction of the effects of turbulence
2. Capturing turbine response to air density changes

Solar*:*

Several items were also updated on the Solar side for 2018. A new series, a Bi-Facial Solar Panel power series, is provided in the restructured data both for the Forecasts and the Power Calculations.

VCE® also updated how regression was applied with the ground-based observations to remove bias errors caused by inaccurate representation of the clouds in the model for the Solar Forecasts. In 2018, VCE® took a more realistic approach to this. In general, a forecast may not have all of the observation data/information used to create the regression coefficients in real-time. The forecast would have regression coefficients (known biases and observations differences) available from a previous year. The most similar year we had to 2018 for Solar was 2016 in terms of how observation data was used to improve the Solar Forecast and model cloud biases. VCE® applied the regression coefficients from 2016 to the Solar Forecast data in 2018. This created a slight bump up in error as seen below. However, this does create a more realistic dataset as to what would actually be available in real-time for energy balancing/trading. Since the 2017 Hour 06 Solar Forecasts were created from 2018 Percentage Errors, this new regression application will also be apparent in the behavior and the Solar Forecasts for 2017.

8.5. Forecast Hour 02 (2018)

VCE® updated how Forecast Hour 02 was created starting in the 2017 year. The Power Dataset is weighted heavily on Forecast Hour 02. From 2017 forward, missing data is far reduced in the VCE® HRRR data archives. A Forecast Hour 02 Forecast and the corresponding Power Dataset would be virtually the same. A new Forecast Hour 02 is now created as an interpolation between the Power Dataset and the Forecast Hour 06. For all intents and purposes, the new Forecast Hour 02 is likened to an "Hour 03" series.

Wind:

After we interpolate between the Power Dataset and the Forecast Hour 06, the following is applied:
1. We introduce some small turbulence here to the newly created Forecast. When the Power Calculations are zero, we do not add any in any turbulence. This creates events that are not realistic for an Hour 02 Forecast. This was performed based on the following.
    a. The previous historical performance in Wind between Hour 02 and Hour 06 and the Power Dataset was analyzed. On average, the Root Mean Square Error (RMSE) was roughly around double for Hour 06 when compared to Hour 02. Turbulence was added to match standard deviation of errors observed in previous years with the following:

$$\sigma_{Turbulence} = \sqrt{(Goal\ Standard\ Deviation_{Hour\ 02\ Error})^2 - (Current\ Standard\ Deviation_{Hour\ 02\ Error})^2}$$

    b. The Current Standard Deviation in Hour 02 Error is what came straight from the interpolated Hour 02 Forecast.
    c. The Goal Standard Deviation in Hour 02 Error is what is needed to get to a Standard Deviation in Error for Hour 02 that would be about half of the Standard Deviation in Error for Hour 06.
    d. In general, we added at random about 0-6% in capacity throughout the year. We allowed slightly less magnitudes to be added in 2017 as opposed to 2018. This was due to the fact that errors were already artificially suppressed from the Forecast creation process described above for 2017.
2. For quality control:
    a. If the created Forecast Hour 02 > 100% (meaning 100% capacity factor), it is set to 100%.
    b. If the created Forecast Hour 02 < 0% (meaning 0% capacity factor), it is set to 0%.

Solar:

This is much more simple and straight-forward than what we do for Wind. After we interpolate between the Power Dataset and the Forecast Hour 06, the following is applied:
1. For quality control:
    a. If the created Forecast Hour 02 > 100% (meaning 100% capacity factor), it is set to 100%.
    b. If the created Forecast Hour 02 < 0% (meaning 0% capacity factor), it is set to 0%.

8.6. General Year-on-Year Wind and Solar Resource Trends

Figure 38 shows the general daily trends of the power datasets and the two forecast hours available for comparison for 2014 – 2018 for both wind and solar. Wind was pulled from a 100-meter hub height. Solar is a single-axis tracking panel.
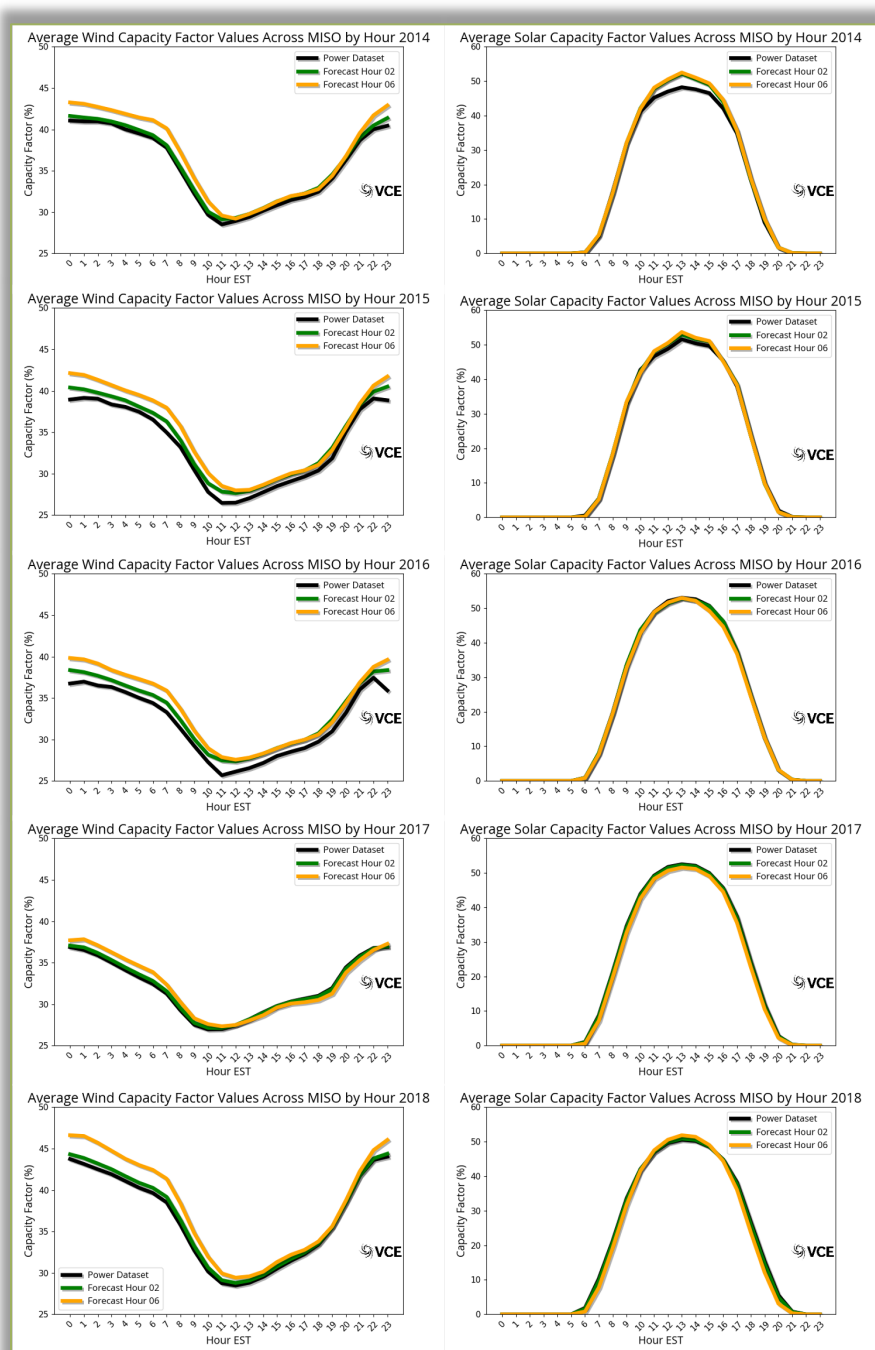
Figure 38: The average capacity (%) for Wind by hour across all of MISO (left) for 2014 to 2018 from top to bottom. The average capacity (%) for Solar by hour across all of MISO (right) for 2014 to 2018 from top to bottom. This is shown for the Power Dataset, Forecast Hour 02 and Forecast Hour 06.

Figure 38 shows the average shape of the wind and solar profiles throughout the day averaged across all the 3km grid points in the entire MISO region for five years. It is straight-forward to see the anti-correlated nature of the Wind and Solar resources across MISO.

Analyzing the Wind data set, several items jump out. The Wind power dataset peaks in the nighttime hours. The wind forecasts follow this shape as well, albeit, a high bias is apparent when compared to the Power

Dataset. The one exception is 2017 where the magnitude bias is lower the latter half of the day. This again reveals the behavior from the 2017 forecasts being created from the 2018 Forecast errors. Across the board for 2017, the magnitude bias is lower than 2018. Since the magnitude bias for 2018 was small on from 1300-2300 EST, the 2017 forecasts exhibit a slight negative bias there. It is observed that 2015, 2016 and 2017 were lower wind years. In 2014 and 2018, the nighttime peaks with capacity factors over 40%. The daytime average lull also does not dip as low.

Looking at the Solar dataset, the average maximum solar resource occurs for all years around 1300 EST. 2016 was a good year for solar production. In general, aside from 2014, the Solar resource peak does not change as much year-on-year like the Wind does.

Forecast Metrics, Evaluation and Analysis:

The forecasts VCE® created were also compared to the Power Dataset product using the following standard metrics. In this case, "F" represents either the Forecast Hour 02 series or the Forecast Hour 06 series. The "PD" term represents the Power Dataset. This was done for both Wind and Solar forecasts to show the performance of the forecasts against the Power Dataset.

Mean Absolute Error (MAE) – Shows the average magnitude of errors. A value closer to zero is better.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |F_i - PD_i| \quad * \quad 100\%$$

Root Mean Square Error (RMSE) – Similar to MAE and shows the average magnitude of errors. However, it is affected more by outliers. A value closer to zero is better.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (F_i - PD_i)^2} \quad * \quad 100\%$$

Magnitude (Multiplicative) Bias – Compares the average forecast magnitude to the average Power Dataset Magnitude. Here, if the value is greater than "1", the Forecast is generally higher than the Power Dataset.

$$Magnitude\ Bias = \frac{\frac{1}{N} \sum_{i=1}^{N} F_i}{\frac{1}{N} \sum_{i=1}^{N} PD_i}$$

Coefficient of Variation – This is a measure of relative variability. This is the standard deviation of a series divided by the mean of the same series.

$$CV = \frac{\sigma}{\mu} \quad * \quad 100\%$$

For the metric analysis shown below, the Wind and Solar capacity factors for each 3-km site within MISO were averaged together for all hours of each year provided to MISO (2014 - 2018). This is about 160,000 sites within the MISO footprint that were included. All five years were also further concatenated together to form the "All" series. This is to show a benchmark, where possible, on how individual years perform against everything altogether. This was done for both Forecast Hours and the Power Dataset for comparison. The following graphics are based off of these averages.
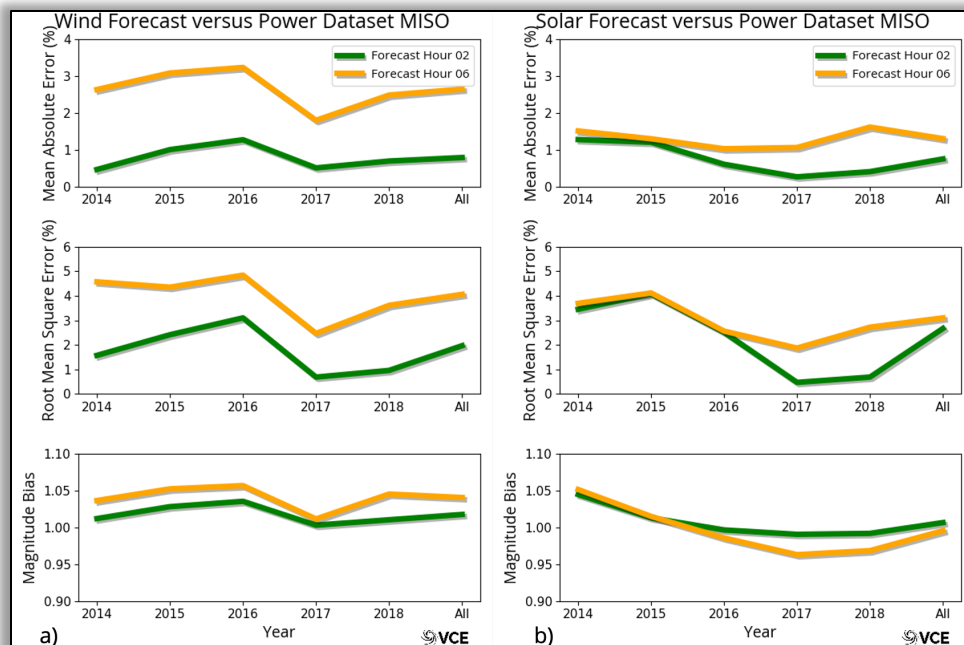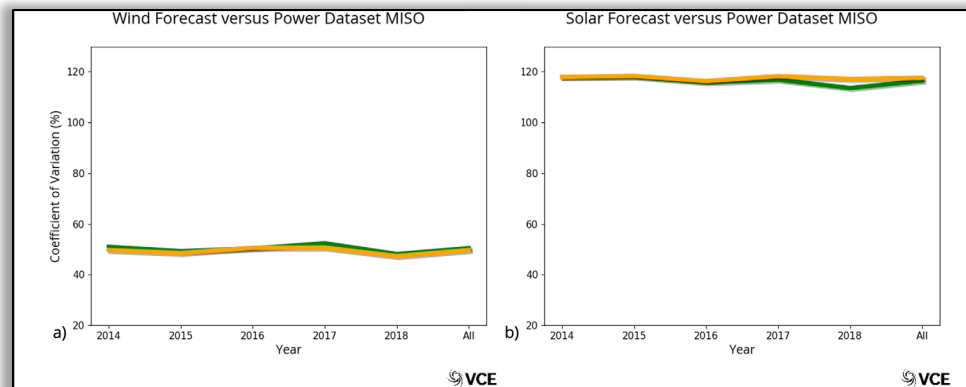
Figure 39: MAE, RMSE and Magnitude Bias for Forecast Hour 02 and Forecast Hour 06 for a) Wind and b) Solar for all of MISO. All five years are shown individually (2014 - 2018). The same metrics were applied to all five years concatenated together ("All").

Figure 39 shows several interesting items to point out. First, the error metric values for MAE and RMSE are quite low in general. Part of this can be attributed to the fact that these metrics are averages of all the 3km sites in MISO. Individual sites could show higher errors. The other piece to consider here is that both the Power Dataset and the Forecasts are based on the same HRRR model data as a foundation. Comparing to actual farm measured data will reveal a different error profile. The metrics for Forecast Hour 02 are generally better than Forecast Hour 06 as is expected.

Wind performance in MAE and RMSE generally improves marching forward in time from 2014-2018. Both error metrics dip down in 2017. This is due to the way the 2017 Forecasts Hours were created. The magnitude bias also improves in 2017, but that is an artificial artifact of the way the Forecasts were created. Across the board, the Forecasts do exhibit a high bias when compared to the Power Dataset. In all metrics here, Forecast Hour 02 outperforms Forecast Hour 06. The distance in performance between the two hours stays fairly consistent across all years.

For Solar, there were several changes made for the 2017 and 2018 forecasts that jump out in the metrics shown above. For this resource as well, the Forecasts 2017 were made based off of the 2018 Hour 06 Forecast and Power Dataset differences. Similar to Wind, MAE and RMSE are lower in 2017 and other years due to the nature of how they were created. In 2016, there was a model update to the HRRR from NOAA that year as well. Starting 2016, the components of irradiance were output from the HRRR model. In general, the HRRR model has seen updates since 2016 to further help improve renewable energy applications. Errors are lower this year, pointing out the improvements made. For 2018, MAE and RMSE do increase. The main reason for this is the way we applied regression coefficient corrections for that year. For 2014 through 2016, we calculated regression coefficients from observation data (a mix of surface and satellite data) to help improve the irradiance values coming from the HRRR forecasts. The same regression coefficients used for the Power Datasets, were used for the Forecasts those years. This is not entirely realistic for real-time understanding as much of this data may not be available in real-time for a forecast to use. In 2018, we used the regression coefficients from 2016 (the most similar year for the observation datasets). This still helps

improve the irradiance coming from the model. It also more accurately portrays what a forecast could be in real-time. In 2017 and 2018, the Forecast Hour 02 was created differently. Fundamentally, it was created as an interpolation between Forecast Hour 06 and the Power Dataset for each year. This was also the case for wind, but this change is more noticeable in the error metrics for Solar. The performance between Hour 02 and Hour 06 starts to diverge starting 2017. The magnitude bias from 2016 forward is less than 1.0. This means that the Forecasts are generally lower than the Power Datasets. Since 2016 exhibited a low forecast bias, this was also reflected in 2018, in part, due to the use of the regression coefficients from 2016. Since 2017 was created using tendencies from 2018, it also shows the characteristics of 2018 in the magnitude bias metric.

Figure 40: The CV for Forecast Hour 02 and Forecast Hour 06 for a) Wind and b) Solar for all of MISO. All five years are shown individually (2014 - 2018). The same metrics were applied to all five years together ("All").

The Coefficient of Variation (CV) is much higher for Solar than it is for Wind. For Solar, this metric is also greater than 100%. This means that the standard deviation for Solar capacity is generally higher than the mean. This is reversed for Wind, where the mean is generally larger than the standard deviation and the Coefficient of Variation is less than 100%. This speaks to the nature of Solar's diurnal variation. It can also point to Wind being less variable with less steep ramps. Wind can stay at max capacity for several days at a time for instance.

These same metrics were calculated within binned hours of the day to delve deeper into what was happening on average over any given day within MISO. The "All" series is not incorporated here since there are so many years to consider.
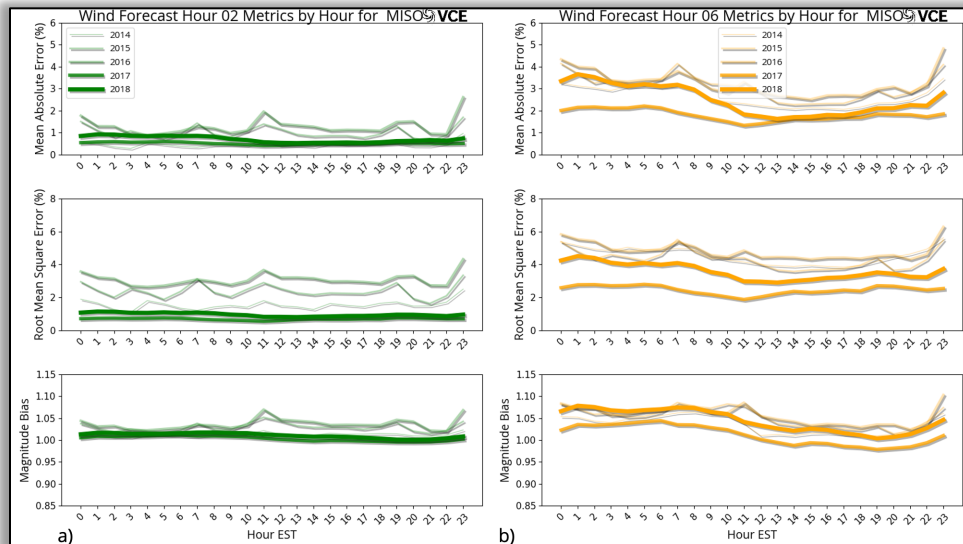
Figure 41: The MAE, RMSE and Magnitude Bias for Wind are shown for a) Forecast Hour 02 and b) Forecast Hour 06. This was calculated for each hour of the day for all of MISO. The years 2014 - 2018 are shown individually.

Looking at Wind, it is clear to see that Forecast Hour 06 errors remain higher than Forecast Hour 02 errors for the entire day both for MAE and RMSE. It is also shown that higher Wind forecast errors tend to occur during the night for Forecast Hour 06. This signature is not observed for Forecast Hour 02. The 2017 series show very low errors and performs the best in all hours for Forecast Hour 06. The reasoning for this was described previously. Year 2018 also shows improvements in error across almost all hours for MAE and RMSE. For Forecast Hour 02, the metrics for 2017 and 2018 are the best performers across the board. This speaks to the new way of creating and interpolating the Forecast Hour 02 those two years as well as the improvements to the forecast in 2018. RMSE especially is lower. Given the fact that Hour 02 is an interpolation between Hour 06 and the Power Dataset, large errors will inherently be capped. This will show up strongly in the RMSE signature as we see here.
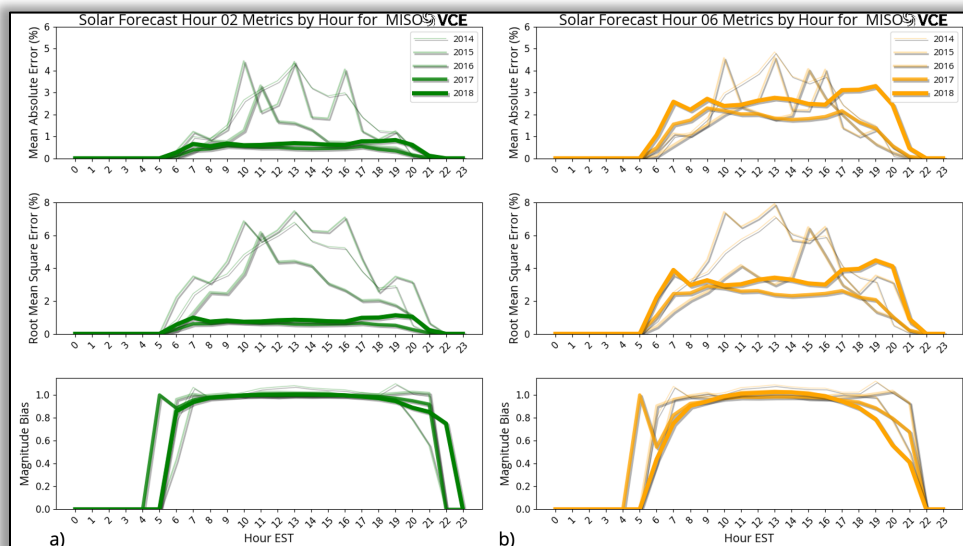

Figure 42: The MAE, RMSE and Magnitude Bias for Solar are shown for a) Forecast Hour 02 and b) Forecast Hour 06. This was calculated for each hour of the day for all of MISO. The years 2014 - 2018 are shown individually.

Looking at Solar Hour 06 Forecast, the errors in 2017 are generally less than the other years. 2017 has lower errors due to the nature of how those Forecasts were created. In 2018, higher forecast errors are apparent around sunrise and sunset. During the day, the 2018 Forecast does much better than other forecasts except for 2016 which is still the best performer. Errors are slightly higher in 2018 due to the fact that we updated how regression coefficients are applied to correct the cloud representation from the HRRR model.

For Solar Hour 02 Forecast, both 2017 and 2018 have very low forecast errors. The new interpolation method to create Hour 02 here really shows as errors are quite reduced in those years. Innately, Forecast errors are higher for Solar during the day.

To show the temporal effect on variability on this analysis, metrics were also calculated over 1-hour, 2-hour, 6-hour, 12-hour and 24-hour average periods. This innately smooths out errors with time. The goal is to show how much variability can change when metrics are produced over different averaging periods. Below is a graphic that shows the effects of this for all of the 3km sites within MISO.

From Figure 43, several items jump out. Once again, the MAE values for Forecast Hour 02 and Forecast Hour 06 for Solar are rather similar. This changes in 2017 and 2018 where we created Forecast Hour 02 differently. This is not the case for Wind. For Wind, Hour 02 and Hour 06 Forecasts differences are similar year-to-year. For both Wind and Solar, when temporal averages are increased the MAE for both Forecast Hour 02 and Forecast Hour 06 decreases for all years. This concept is quite obvious to realize. It is beneficial to see how much the MAE drops off with increasing time averages.

The Solar resource sees a sharper dropping Coefficient of Variance when more hours are averaged together for metrics. This in part is due to the inclusion of nighttime hours where there is no change in Solar Resource. It also is attributable to the entire diurnal pattern of Solar becoming more and more included. Ultimately, the standard deviation falls and the average of the series increases for Solar as more hours enter the average. For Wind, the CV does not change as drastically when more hours are averaged together. For Wind, the mean overpowers the standard deviation. Wind CV decreases with increasing time length because the standard deviation falls while the mean stays relatively constant.
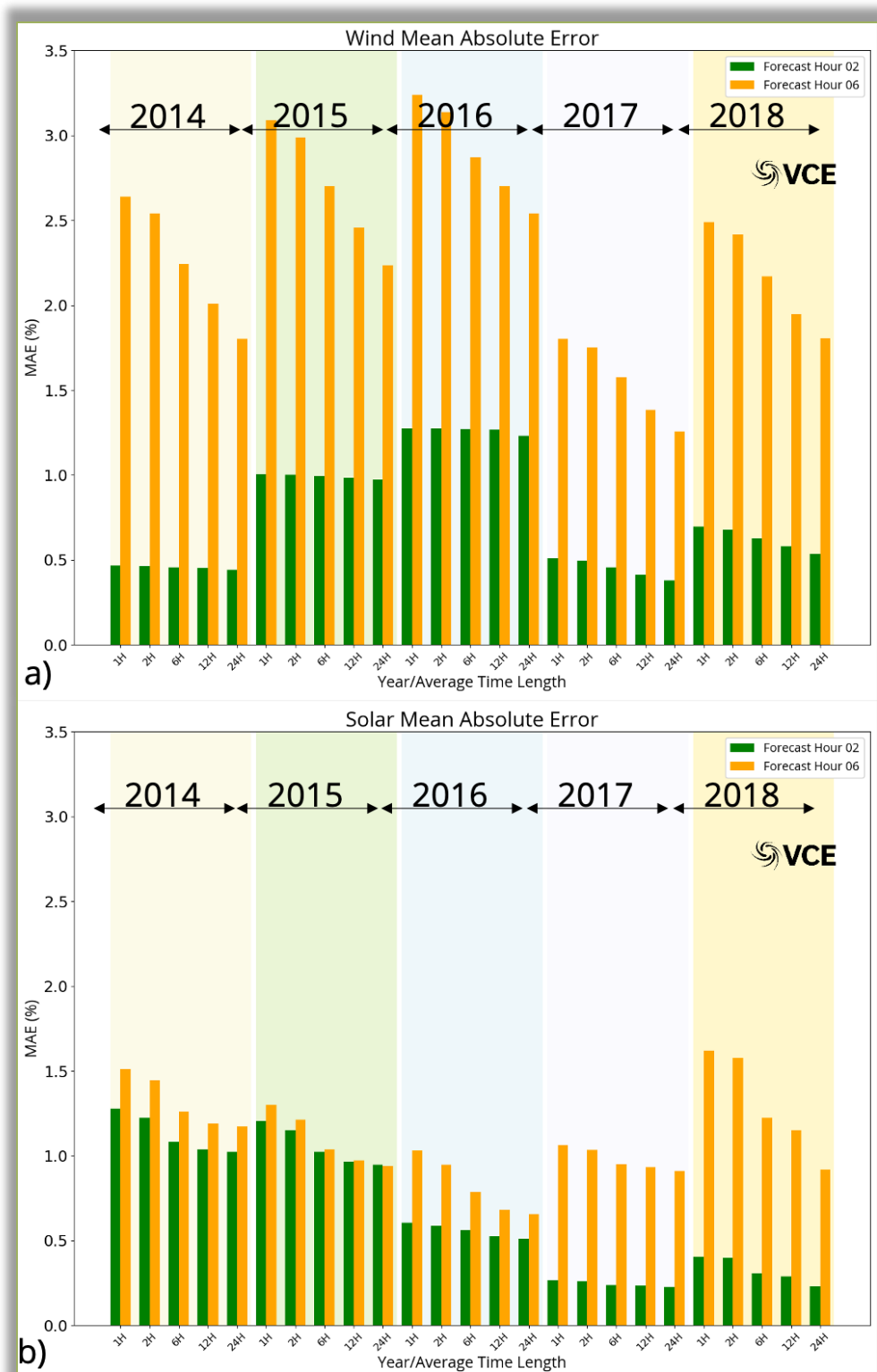
Figure 43: Mean Absolute Error display for hourly data (1H), data averaged over two hours (2H), data averaged over six hours (6H), data averaged over twelve hours (12H) and data averaged over twenty-four hours (24 H). This is shown for a) Wind and b) Solar for all of MISO. Years 2014 - 2018 are provided.
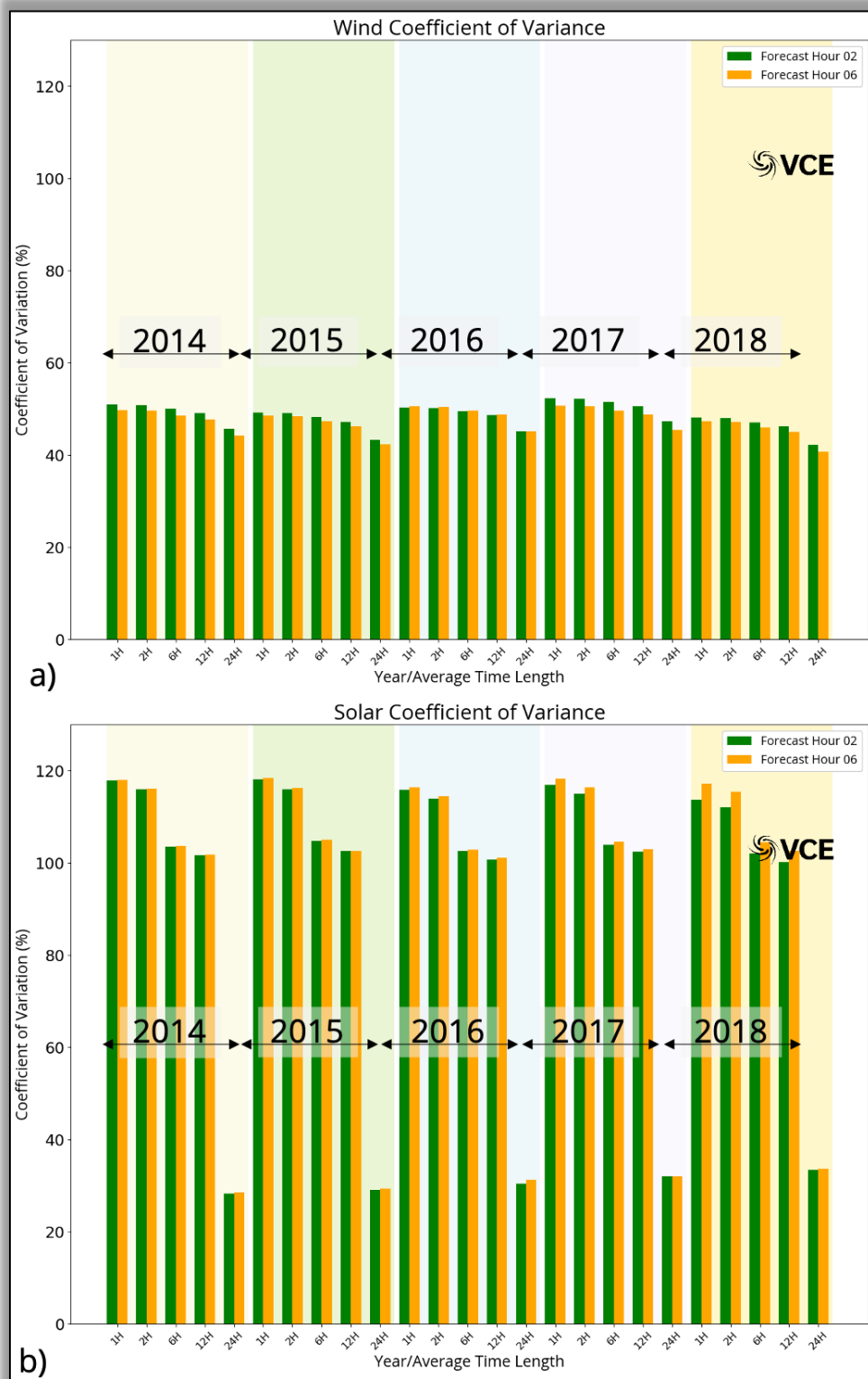
Figure 44: Coefficient of Variance display for hourly data (1H), data averaged over two hours (2H), data averaged over six hours (6H), data averaged over twelve hours (12H) and data averaged over twenty-four hours (24 H). This is show for a) Wind and b) Solar for all of MISO. Years 2014 - 2018 are provided.

8.7. VRE Ramp Analysis

VCE® also looked closer at the change in the Power Dataset and Forecasts every hour compared to the previous hour. This look at ramps can help pull out when the variables are changing the fastest on average across MISO.
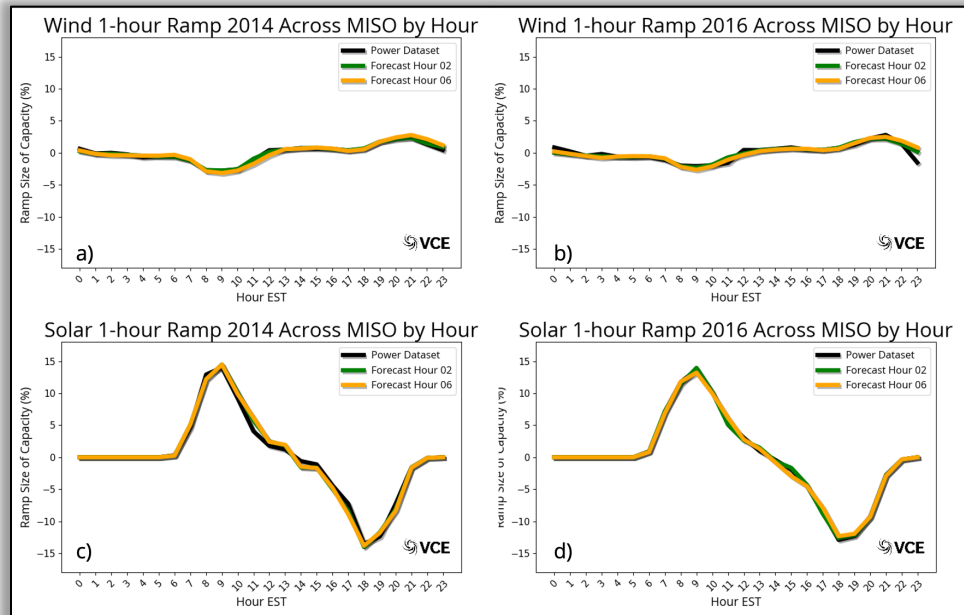
Figure 45: A look at the 1-hour change in resource capacity for a) Wind in 2014, b) Wind in 2016, c) Solar in 2014 and d) Solar in 2016 across all of MISO. These patterns hold for other years as well (2015, 2017 and 2018). Here a higher capacity factor Wind year (2014) and a lower capacity Wind year (2016) are shown. A higher capacity factor Solar year (2016) and a lower capacity Solar year (2014) are shown.

Interestingly on the 1-hour scale, the Solar ramp that can occur is far larger than that of Wind on average across all of MISO. This speaks to the nature of Wind where, during the same hour, one location could be experiencing a ramp up, while another a ramp down. Solar is more aligned with its daily sunrise/sunset ramps. Ramp sizes for Wind between 2014 and 2016 are similar. The same is true for Solar as well. Figure 45 shows that the ramp rates did not change much between the two years. This is said in context of Capacity Factor percentages. If more of a generator resource was added (additional wind farms for instance), the ramps would innately increase when looking at the MWs produced from a resource. The steepest ramps occur for Wind around 0900 EST for a ramp down and 2100 EST for a ramp up in wind capacity. For Solar, the morning ramp up is steepest around 0900 EST, while the evening ramp down is largest around 1800-1900 EST. Figure 45 also shows the general anti-correlated tendencies of Wind and Solar. When Wind is ramping down, on average Solar is ramping up and vice versa.

The frequency of strong ramps at different time scales is another way to determine the characteristics of the series analyzed. VCE® created histograms of the ramp sizes over several time scales. As an example, for a Power Dataset value at 01/01/2014 at Hour 15, a 1-hour ramp is the difference in potential generation subtracting Hour 14 from Hour 15. For the same time (01/01/2014 at Hour 15), a 6-hour ramp is the difference in potential generation subtracting Hour 09 from Hour 15. Essentially, this is a "*forward-looking*" ramp.
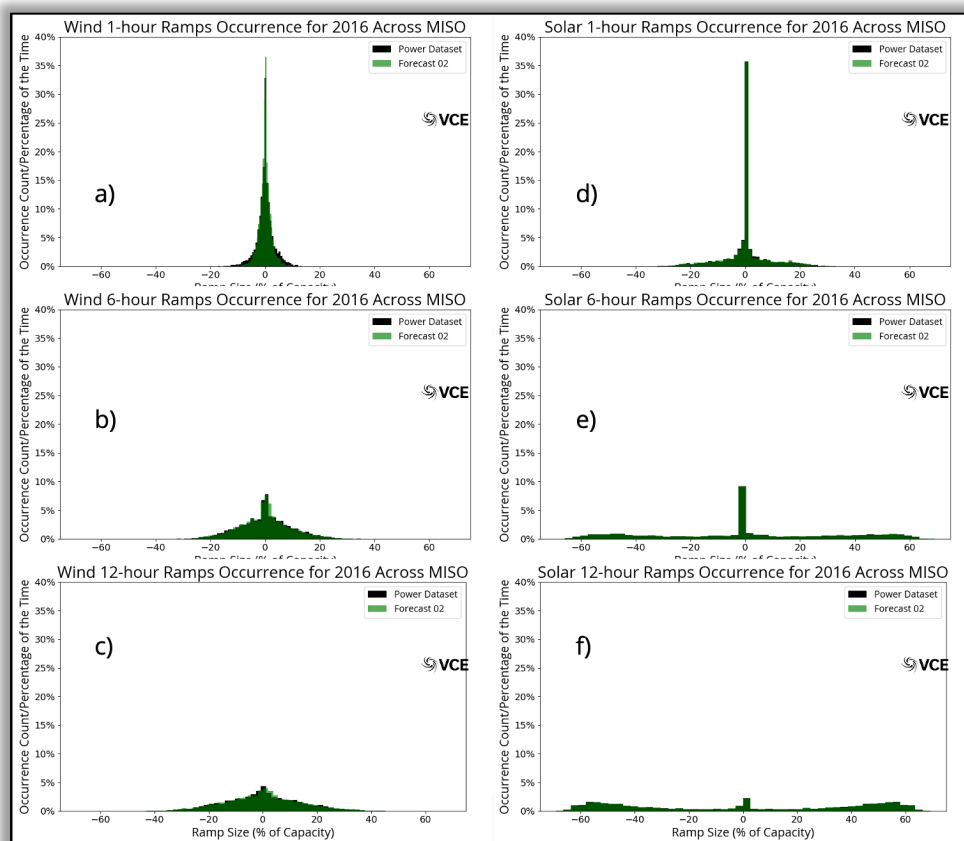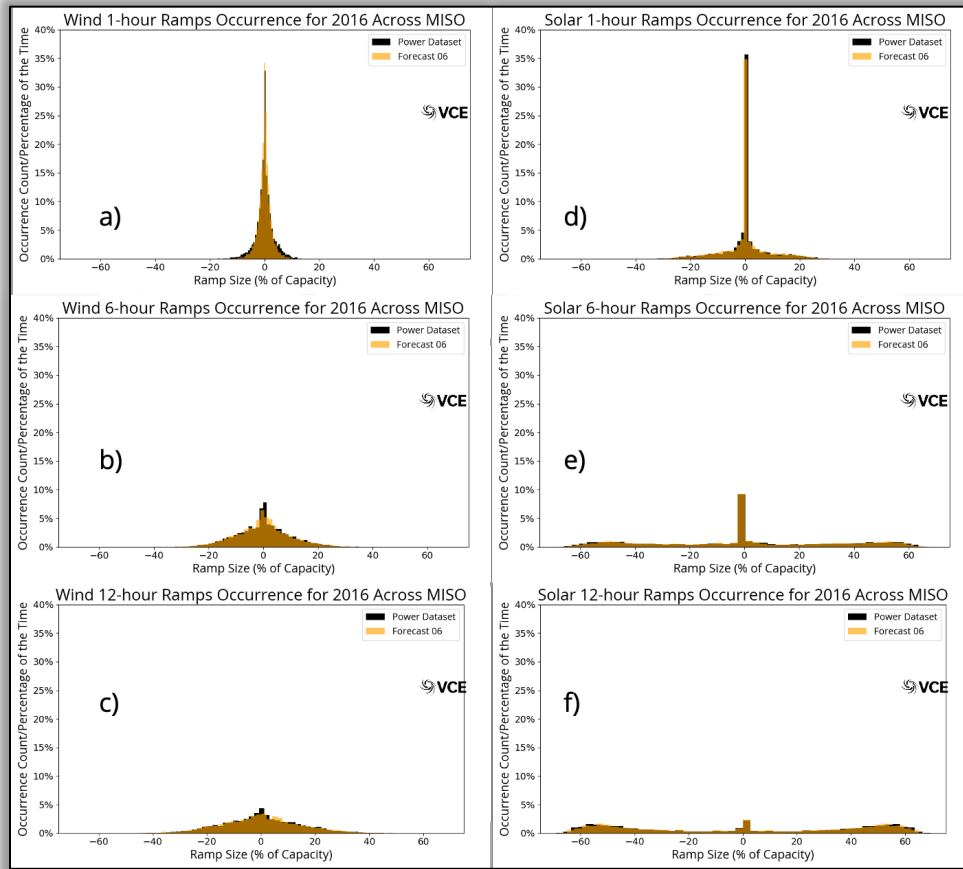
Figure 46: The ramp size (% of capacity) frequency distribution for a) 1-hour ramp length for Wind, b) 6-hour ramp length for Wind, c) 12-hour ramp length for Wind, d) 1-hour ramp length for Solar, e) 6-hour ramp length for Solar and f) 12-hour ramp length for Solar. The Forecast Hour 02 series is compared to the Power Dataset. This is for the year 2016.

Figures 46 and 47 reveal the difference in ramping properties of the Wind and Solar variables. Wind ramps are far less steep even at increased time lengths across all of MISO. Innately, this simply speaks to the diurnal pattern of solar with sunrise and sunset. Further it also outlines the challenges of supporting such steeps ramps at such short periods that come from solar. The Forecasts follow the distribution shape of the Power Dataset for Wind very well.

Figure 47: The ramp size (% of capacity) frequency distribution for a) 1-hour ramp length for Wind, b) 6-hour ramp length for Wind, c) 12-hour ramp length for Wind, d) 1-hour ramp length for Solar, e) 6-hour ramp length for Solar and f) 12-hour ramp length for Solar. The Forecast Hour 06 series is compared to the Power Dataset. This is for the year 2016.

8.8. Spatial Forecast Biases

VCE® also considered the spatial forecast errors. Figure 48, below, shows the percentage difference for all 3-km sites across the CONUS between Wind Forecast Hour 06 and the Power Dataset for 2014. If the difference is positive, the forecast is, on average, higher than the Power Dataset. It is observed that the forecast is high biased over the eastern half of the US, and also the majority of the MISO region for 2014. The other four years exhibit similar behavior.
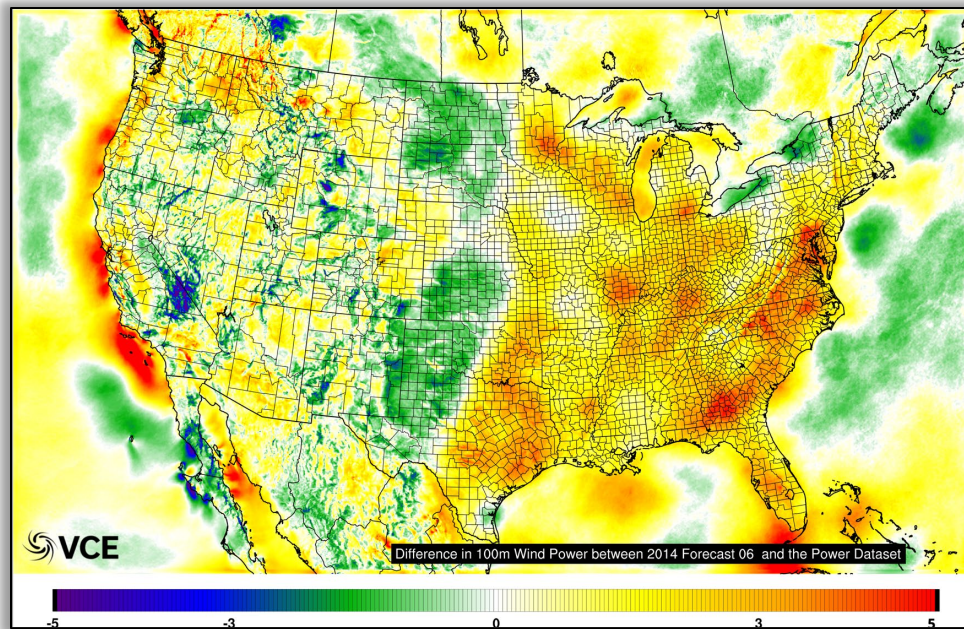

Figure 48: The percentage difference between the Wind Forecast Hour 06 and the Power Dataset for 2014 across the CONUS.

It should be noted that these forecast errors shown for both wind and solar in these sections will be different when comparing to observations. The Power Datasets are created with HRRR model data, thus, everything is based in model space.

9. Generator Input Dataset

VCE® processed the Energy Information Administration (EIA) annual data from 2017 to create the baseline input generator dataset for the United States. The Midcontinent ISO has a large geographic extent and contains approximately 195,000 MW of generation capacity. WIS:dom® has the ability to solve over such scales at 5-minute resolution for several years chronologically.

The generator input datasets are based off of the publicly available EIA 860 and EIA 923 data. The 2017 data is what was available for the present study. We go through several steps to align and aggregate technology types to the 3-km grid space to match with the HRRR weather datasets. In the process, we also analyze year-on-year changes. General trends show coal capacities falling with Natural Gas Combined Cycle growing. Wind, Solar and Storage plants are on the rise as well. This continues even into the newly released EIA 860 2018 data.

The following outlines the process VCE® undergoes to prepare the generator input datasets:
1. Data is merged between the EIA 860 and EIA 923 data.
2. Initial quality control is applied to the data.
3. Align the location of the generators to the nearest 3-km HRRR cell. This can be more difficult for generators on state boundaries as well as land/water boundaries. As such, extra time is given to ensure that the mapped generators are correct.
4. Aggregation of the generator types in each 3-km grid cell. As an example, if two separate coal plants are in the same grid cell, the capacity is summed for coal in that grid cell.
5. Spatial checks are performed to make sure the output aligns with the original data.
6. Final model input format produced. A county-level average of all generator types is also created.

VCE® is now also working with the Catalyst Cooperative (https://catalyst.coop/), a company with the goal to help the energy research community by processing major publicly available sources into a format that is organized and stream-lined to use. This is helping our processes become quicker and eventually more frequent on this input dataset.

1	Coal
2	Natural Gas Combined Cycle
3	Natural Gas Combustion Turbine
4	Storage
5	Nuclear
6	Hydroelectric
7	Onshore Wind
8	Offshore Wind
9	Residential Solar
10	Utility-scale Solar
11	Concentrated Solar Power
12	Geothermal
13	Biomass
14	Other Natural Gas
15	Other Generation
16	Natural Gas - CCS
17	Pumped Hydro Storage
18	Small Modular Reactors
19	Molten Salts

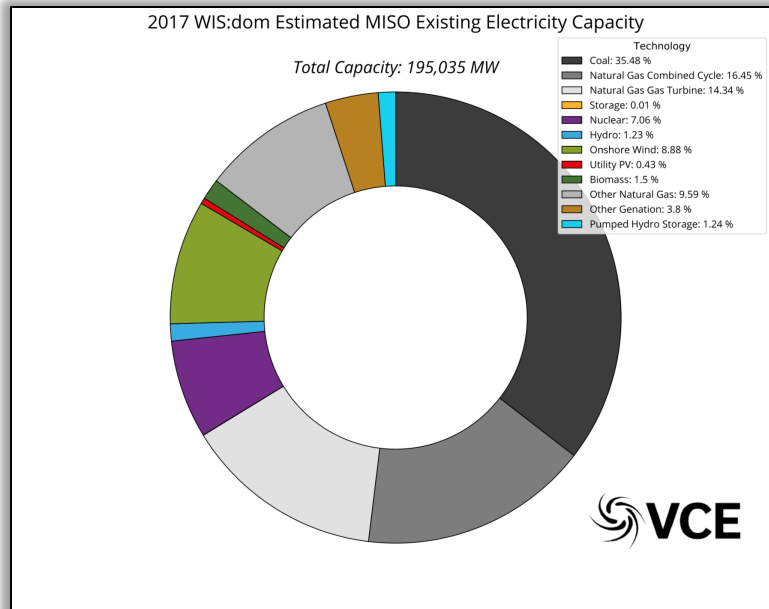Figure 49: The VCE® generator technology classes.

Figure 50: WIS:dom® estimated capacity share for the MISO. The total capacity modeled is 195,035 MW. This is based upon EIA 860 numbers from 2017.
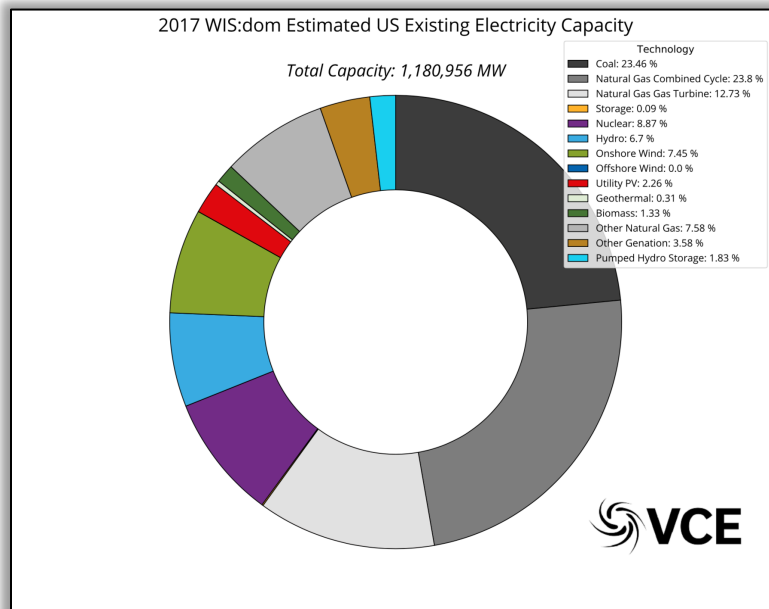

Figure 51: WIS:dom® estimated capacity share for the United States. The total capacity modeled is 1,180,956 MW. This is based upon EIA 860 numbers from 2017.
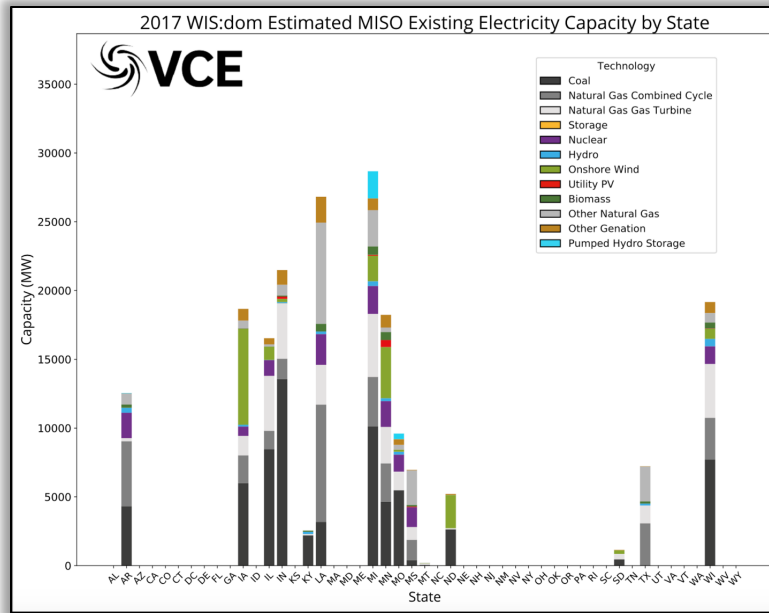
Figure 52: WIS:dom® estimated capacity share for the states within MISO. This is based upon EIA 860 numbers from 2017. This is only showing the generators under MISO in each state.

Figure 53 displays the geographic siting of all MISO generators. It shows a large count of hydro plants in the northern states of MISO (Wisconsin, Minnesota and Michigan). Most wind farms are present in Iowa and southwestern Minnesota. There are many wind farms taking advantage of the lake effect meteorology in Wisconsin and Michigan. North Dakota also hosts some larger wind farms. Most utility PV units are observed in Minnesota and Indiana.
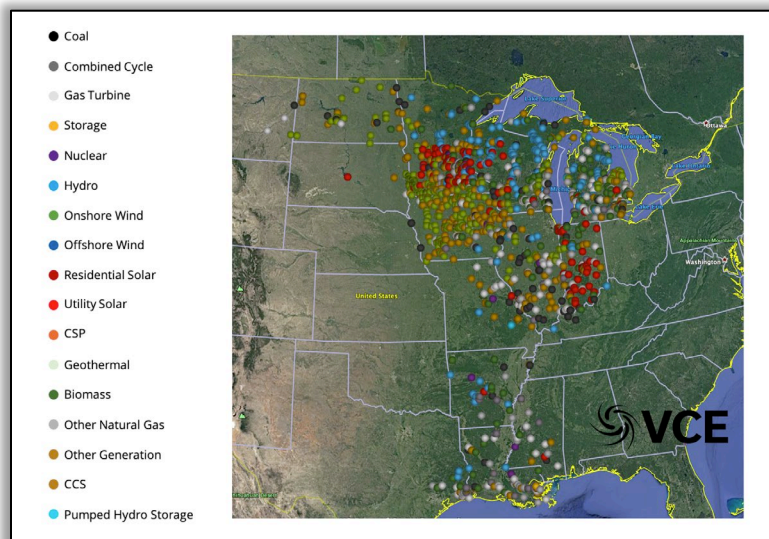

Figure 53: WIS:dom® estimated location of various technologies within MISO. This is based upon EIA 860 numbers from 2017.

10. Potential Input Dataset

VCE® undergoes a procedure to determine the siting potential of new generators as well. This ensures the WIS:dom® model has limitations on where it can build new generation. First and foremost, USGS land use information is utilized as a foundation within each 3- km grid cell to determine what uses the land currently has (see Figure 54 below).
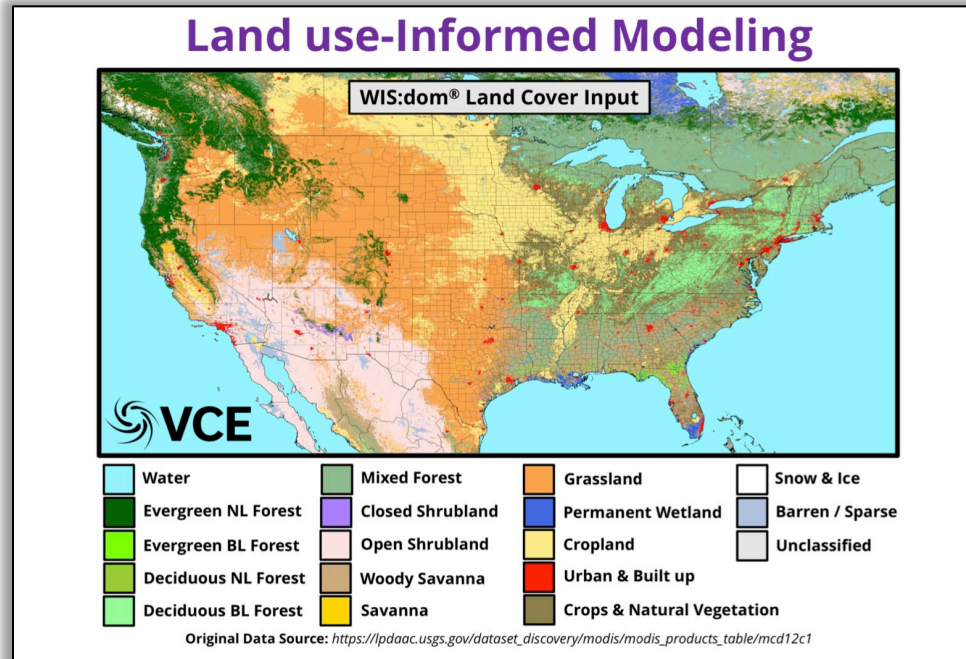

Figure 54: WIS:dom land use information providing properties of the land within each 3km HRRR grid cell.

The procedure to further quality control this information is as follows:

1. Remove all sites that are not on appropriate land-use categories.
2. Remove all sites that have protected species.
3. Remove all protected lands; such as national parks, forests, etc.
4. Compute the slope, direction and soil type and determine its applicability to VRE installations.
5. Determine the land cost multipliers based on ownership type.
6. Remove military and other government regions that are prohibited.
7. Avoid radar zones and shipping lanes.
8. Avoid migration pathways of birds and other species.

The above, along with the knowledge of what is already built within a HRRR cell from the Generator Input data provides WIS:dom® with the view of where it can actually build certain generators as well as certain technologies.