

Wind Farms Conclusions

This document attempts to summarize all the observations from the exploratory data analysis.

Notations

J	Number of assets
$d = 1, \dots, D; h$	Days, Hours
C_i	Capacity for Asset i
$\mathbf{g}_d = (g_{d,h})$	Actual Production MWh
$\mathbf{f}_d = (f_{d,h})$	forecasted Production MWh
$\tilde{\beta}_{d,h}$	prediction bias
i	assets indices
$\alpha_{d,h}$	actuals fraction $\in [0, 1]$
$\gamma_{d,h}$	forecast fraction $\in [0, 1]$
$\beta_{d,h}$	bias fraction
$\tau_{d,h}$	kendall's tau
r^2	corelation coefficient
σ	normalized standard deviation of bias
$\tilde{\sigma}$	unnormalized standard deviation of bias
S	the set of all assets

- Days are indexed by $d = 1, 2, \dots, 365$, and hours are indexed by $h = 1, 2, \dots, 24$
- $\mathbf{g}_{d,h}, \mathbf{f}_{d,h}, \alpha_{d,h}, \beta_{d,h}$ are vectors of Dimension J Where

$$\alpha_{d,h} \in [0, 1] = \frac{\mathbf{g}_{d,h}}{C} \quad \text{and} \quad \beta_{d,h} \in [0, 1] = \frac{\mathbf{f}_{d,h}}{C}$$

- The bias fraction is then

$$\beta_{d,h} = \gamma_{d,h} - \alpha_{d,h}$$

Bias

Bias at hourly interval

For a given asset, let

$$\beta_h = \frac{1}{365} \sum_{d=1}^{365} \beta_{d,h}$$

and β_h is the prediction bias for a given hour of the day. For each $h \in [1, 24]$, we plot out β_h in a violin plot and a box plot.

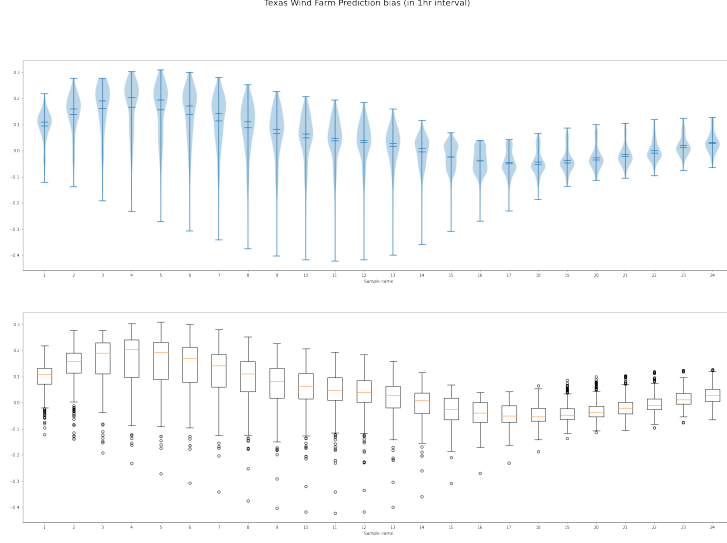


Figure 1:

It was found that, overall in the morning the prediction bias tend to be positive, with negatively skewed distribution. In the afternoon and evening the prediction bias tend to be negative, with less skewed distribution. With closer analysis, the negative skewness in the morning is brought by **Canadian Breaks Wind** and **Desert Sky repower**.

Bias at 24 Hours interval

We are also interested in the geographic distribution of the prediction bias.

For a given asset, let

$$\beta = \frac{1}{365 \times 24} \sum_{d=1}^{365} \sum_{h=1}^{24} \beta_{d,h}$$

and β is also a vector of length J . Plotting β on the map of Texas,

As observed from the scatter plot, the bias, across all hours and all year, is relatively higher for Southern and central wind farms and is lower for Northern and Western windfarms.

Standard Deviation of Bias

We are also interested in the standard deviation of the Bias. Let

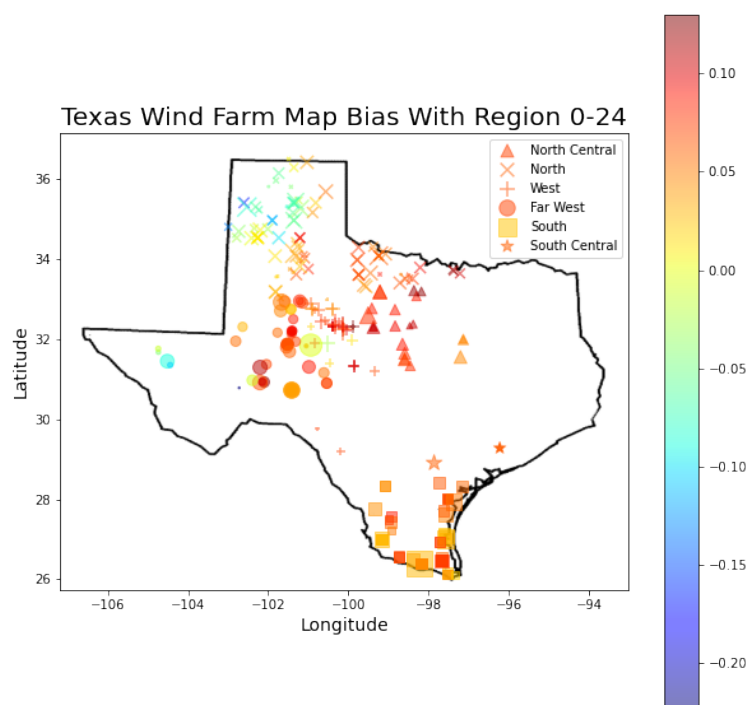


Figure 2:

$$\sigma_h^\beta = \sqrt{\frac{\sum_{d=1}^{365} (\beta_{d,h} - \bar{\beta}_h)^2}{365}}$$

where $\bar{\beta}_h = \frac{1}{365} \sum_{d=1}^{365} \beta_{d,h}$ is the mean of the bias at a given hour across all year. For each $h \in [1, 24]$, we plot out σ_h^β in a violin plot and a box plot.

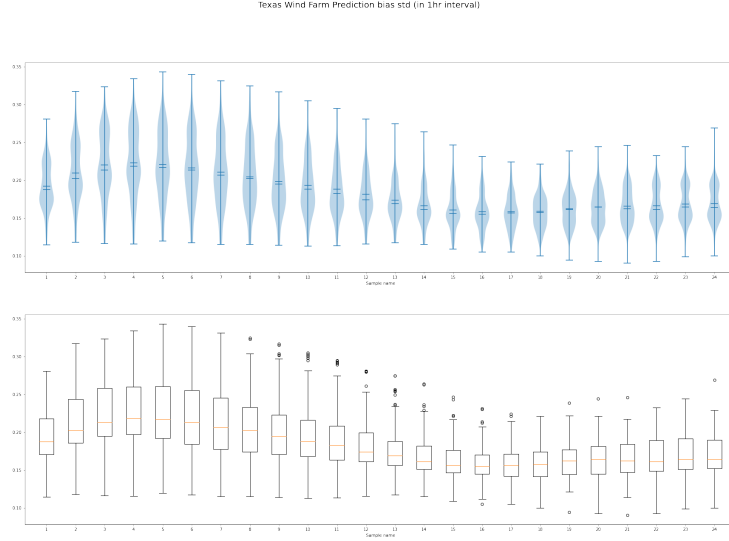


Figure 3:

It was found that, from the above violin/box plot, the mean and median of the standard deviation of bias tend to be stable across 24 hours. There are some outliers during sunlight time and few outliers at night. It was also found that the standard deviation of bias tend to be larger in the early morning comparing to the afternoon.

Standard Deviation of Bias at 24 hour interval

We are also interested in the geographic distribution of the standard deviation of bias.

For a given asset, let

$$\sigma^\beta = \sqrt{\frac{\sum_{h=1}^{24} \sum_{d=1}^{365} (\beta_{d,h} - \bar{\beta})^2}{365}}$$

where $\bar{\beta} = \frac{1}{365 \times 24} \sum_{h=1}^{24} \sum_{d=1}^{365} \beta_{d,h}$ is the mean of the bias across all year and all hours. σ is also a vector of length J . Plotting σ on the map of Texas, we have

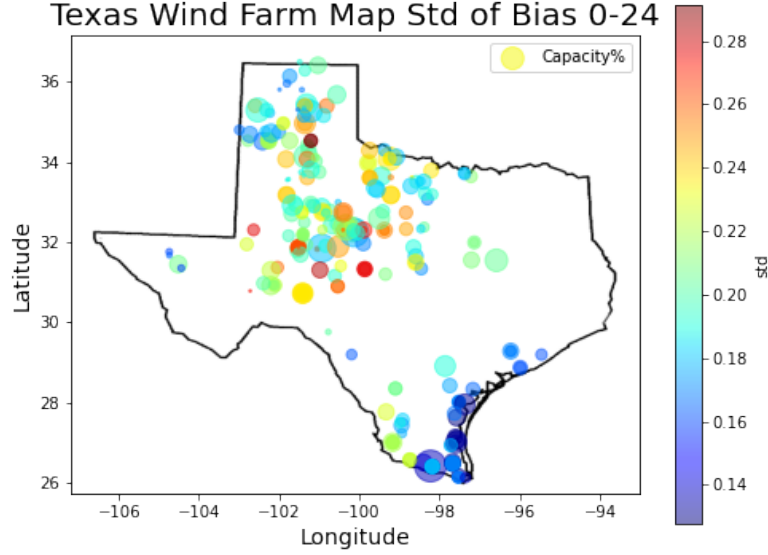


Figure 4:

From this scatter plot, it was found that the standard deviation of bias is relatively lower for the Southern assets.

kendall's Coefficient for Each Region

We are also interested in the ordinal correlation of σ between each asset across hours. Since there are a total of 6 regions in Texas and S is the entire set of assets. Let

$$S = \{S_1, S_2, S_3, S_4, S_5, S_6\}$$

Where $S_i, i \in \{1, 2, 3, 4, 5, 6\}$ is the set of assets in a particular region. And let

$$\tau_i^{h+1} = \frac{2}{|S_i|(|S_i| - 1)} \sum_{m < n} \text{sgn}(\sigma_m^{i,h} - \sigma_n^{i,h}) \text{sgn}(\sigma_m^{i,h+1} - \sigma_n^{i,h+1})$$

So τ_i^{h+1} is the Kendall correlation coefficient for asset S_i during hour interval

$[h, h + 1]$. Then for $h \in [1, 23]$ and for $i \in \{1, 2, 3, 4, 5, 6\}$ we plot out the kendall correlation for each region across 23 hour intervals,

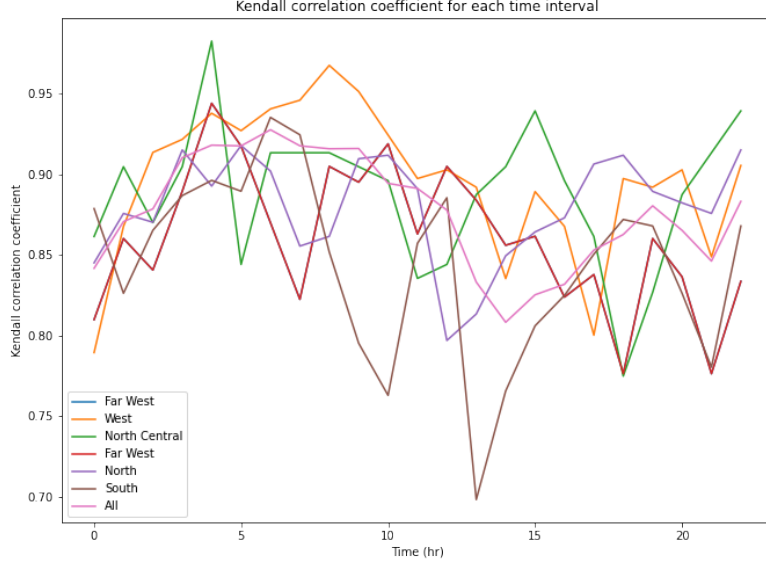


Figure 5:

As seen from the plot, overall, the kendall's τ tend to be lower in the afternoon comparing to the mornings. It was also found that the assets in the South behave unusual comparing to other regions, its kendall's τ is lowest at hour 13.

Capacity vs. std of unnormalized bias

We are also interested in the correlation between an asset's capacity and the standard deviation of unnormalized bias. Let

$$\tilde{\sigma} = \sqrt{\frac{\sum_{h=1}^{24} \sum_{d=1}^{365} (\tilde{\beta}_{d,h} - \bar{\tilde{\beta}})^2}{365}}$$

Plotting Capacity C on the x-axis and $\tilde{\sigma}$ on the y axis, we calculate the correlation coefficient r^2

$$r = \frac{\sum_{i=1}^J (\sigma_i - \bar{\sigma}) (C_i - \bar{C})}{\sqrt{\sum_{i=1}^J (\sigma_i - \bar{\sigma})^2 \sum_{i=1}^J (C_i - \bar{C})^2}}$$

where $\bar{\sigma} = \frac{1}{J} \sum_{i=0}^J \sigma_i$ and $\bar{C} = \frac{1}{J} \sum_{i=0}^J C_i$

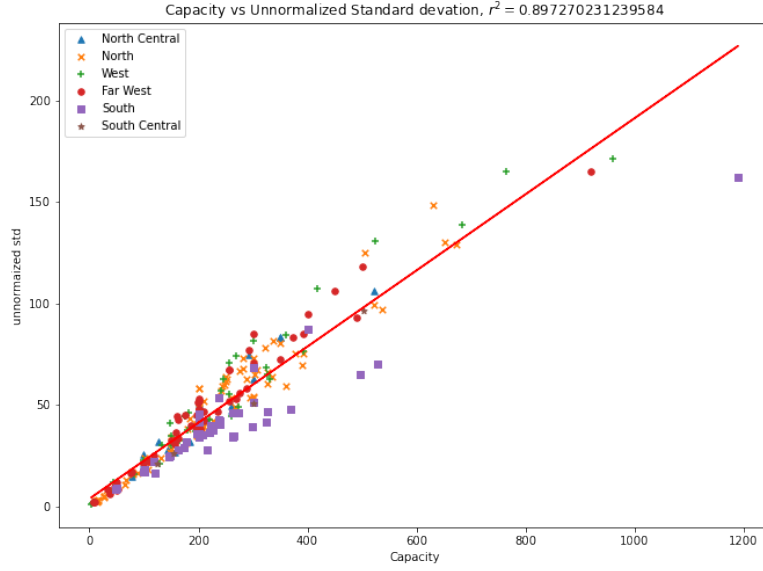


Figure 6:

It was calculated that $r^2 = 0.897$ indicating a relatively strong linear relationship. However, from the scatter plot we can see that the linear relationship is strong for $C < 450$. It was also observed that for $C > 450$ There is a clearly non linear pattern. Besides, the assets in the south has lower std/capacity ratio comparing to other regions.

References