

Data to be used (MTLOG 0100).

Renzo Miguel Caballero Rosas
Renzo.CaballeroRosas@kaust.edu.sa
CaballeroRenzo@hotmail.com
CaballeroRen@gmail.com

January 27, 2020

Data information:

Normalization: $P_{max} = 1474 \text{ MW}$, $T = 24 \text{ h}$. Most modified days were removed, we will work with 255 days. We separate all the data into two sets, a training set, and a testing set.

We create three tables, all the data, the testing, and the training data. Each table is saved in .mat and .csv formats. All the names are:

- MTLOG_0100_and_Real_24h_Training_Data.mat.
- MTLOG_0100_and_Real_24h_Testing_Data.mat.
- MTLOG_0100_and_Real_24h_Complete_Data.mat.
- Table_Training_Complete.csv.
- Table_Testing_Complete.csv.
- Table_Complete.csv.

The .mat data has a trivial format if it is opened with MATLAB. The .csv data has all the daily information in each row; each row has:

- First column (**column 1**): **Date**.
- Next **145** columns (**from column 2 to column 146**): **Time: Normalized domain of each path**.
- Next **145** columns (**from column 147 to column 291**): **Normalized Interpolated Forecast**.
- Next **144** columns (**from column 292 to column 435**): **Normalized Forecast Interpolated Derivative**.
- Next **145** columns (**from column 436 to column 580**): **Normalized Interpolated Real Production from UTE**.
- Next **145** columns (**from column 581 to column 725**): **Normalized Real Production from ADME**.
- Next **145** columns (**from column 726 to column 870**): **Normalized Forecast Error (Real ADME minus Forecast)**.
- Next **144** columns (**from column 871 to column 1014**): **Normalized Forecast Error Transitions**.
- Next **145** columns (**from column 1015 to column 1159**): **Lamperti Transform of the Normalized Forecast Error**.
- Next **144** columns (**from column 1160 to column 1303**): **Lamperti Transform Transitions**.

Notice that the forecast derivative and the error transition can not be calculated at the final point.

Seasonality effect:

To guaranty an homogeneous year, we study the Mean Absolute Error (MAE) for the forecast for each day. In Figure (1), we can see that there are no big variations along the year.

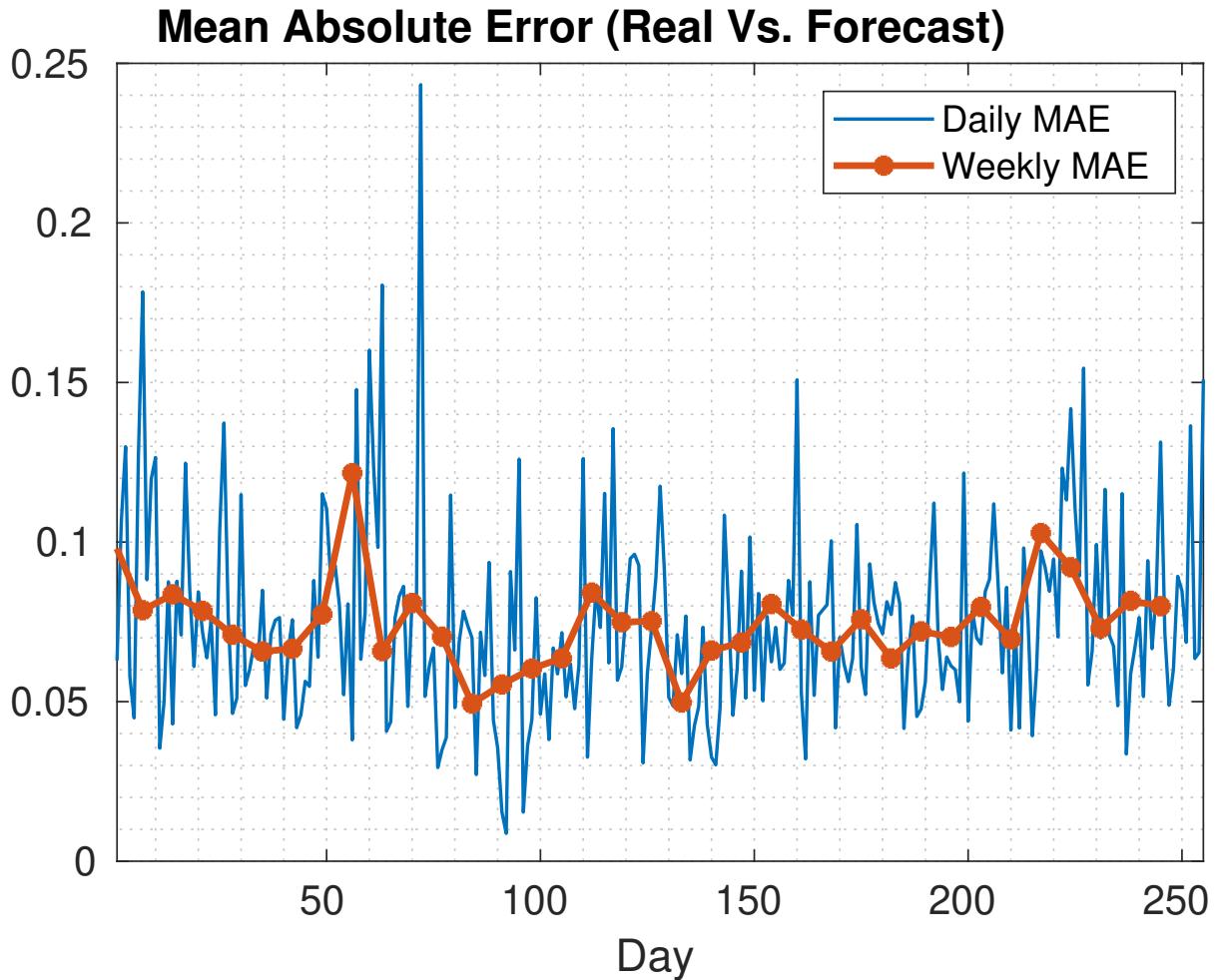


Figure 1: Daily MAE computed along the 255 days that will be used in the training and testing.

To compute the vector that we are piloting, we realize the operation

$$\hat{V}(j) = \frac{1}{145} \sum_{i=1}^{145} |V(i, j)| \quad \text{where } j \in \{1, \dots, 255\}.$$

Recall that $V(i, j)$ is the normalized error between the ADME real production and the UTE forecast at time i and for the day j .

Hourly effect:

We want to see the error throughout the day. We compute the Mean Absolute Error (MAE) for the forecast for each measurement during the day. In Figure (2), we can see that the error is greater in the mornings. Notice that as we advance in the time, the forecast becomes less accurate.

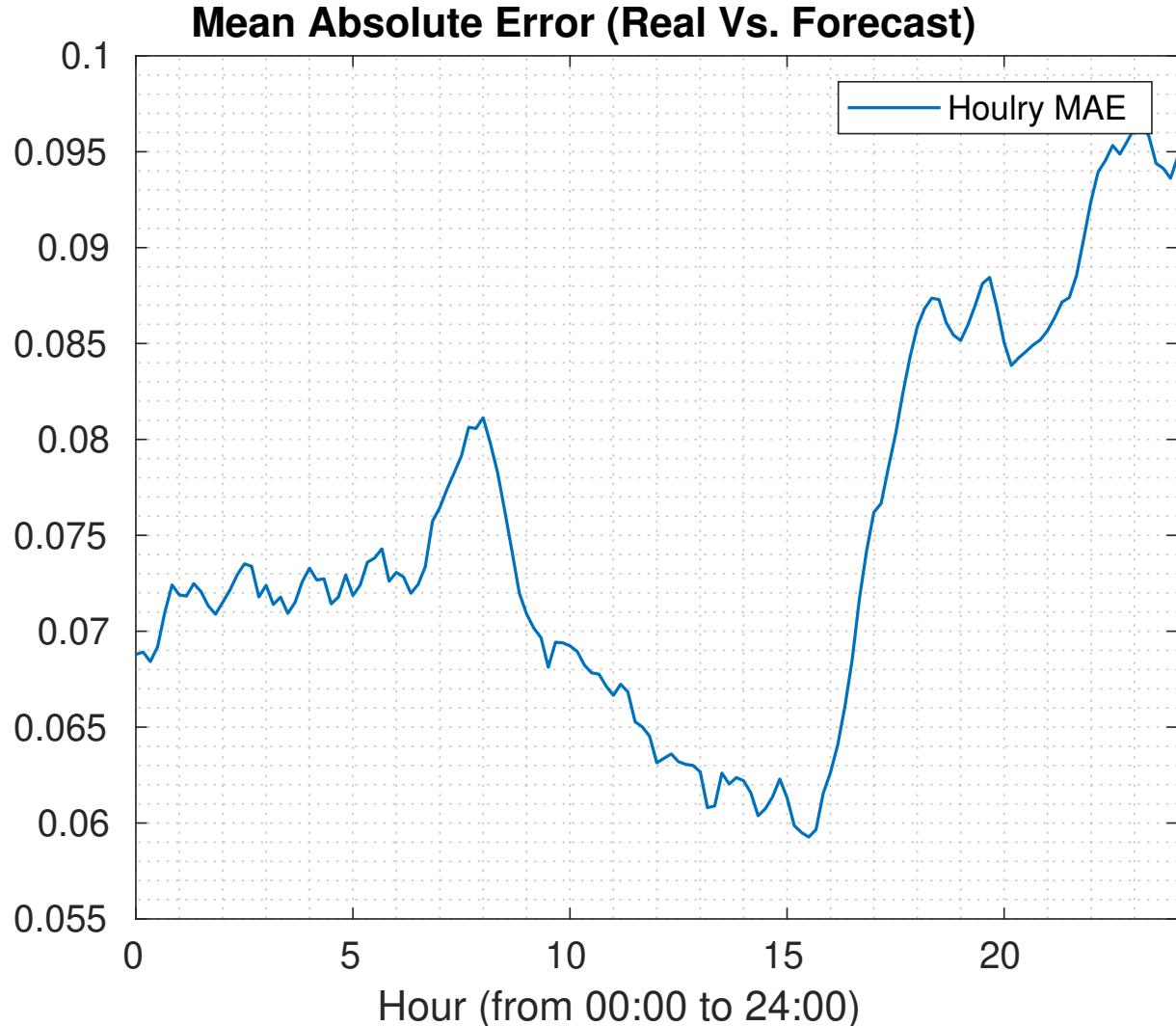


Figure 2: MAE along the day.

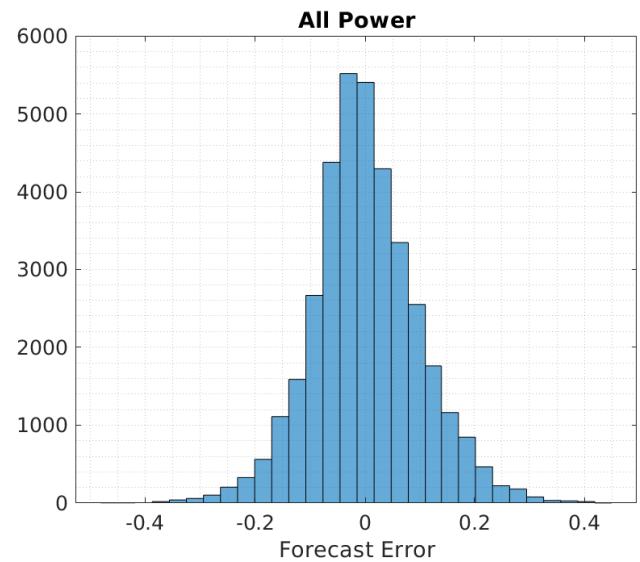
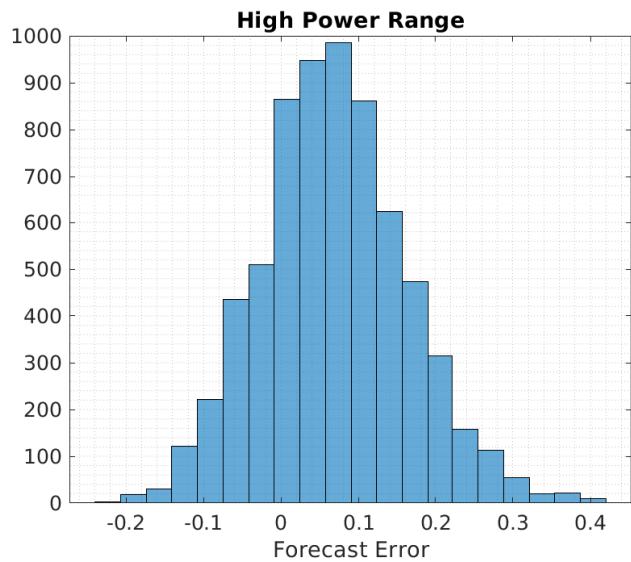
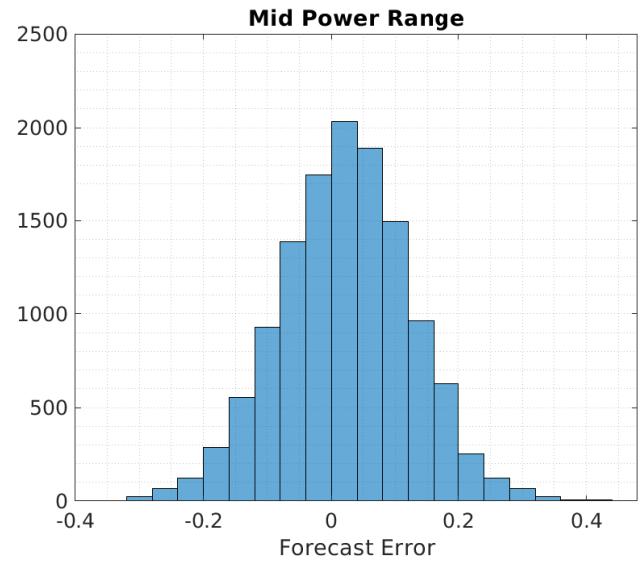
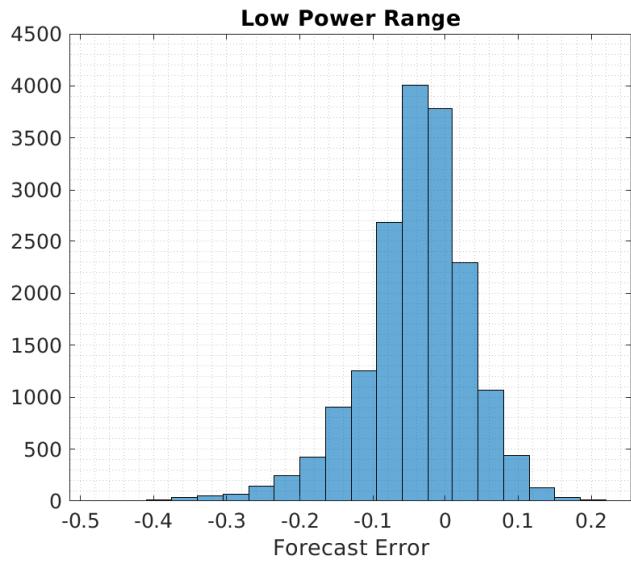
To compute the vector that we are piloting, we realize the operation

$$\hat{V}(i) = \frac{1}{255} \sum_{j=1}^{255} |V(i, j)| \quad \text{where } i \in \{1, \dots, 145\}.$$

Recall that $V(i, j)$ is the normalized error between the ADME real production and the UTE forecast at time i and for the day j .

We can see that the error depends on the time. For this reason, when we create the batches, we need to sample days and no random transitions. Otherwise, we may choose no representative samples.

Forecast Error:

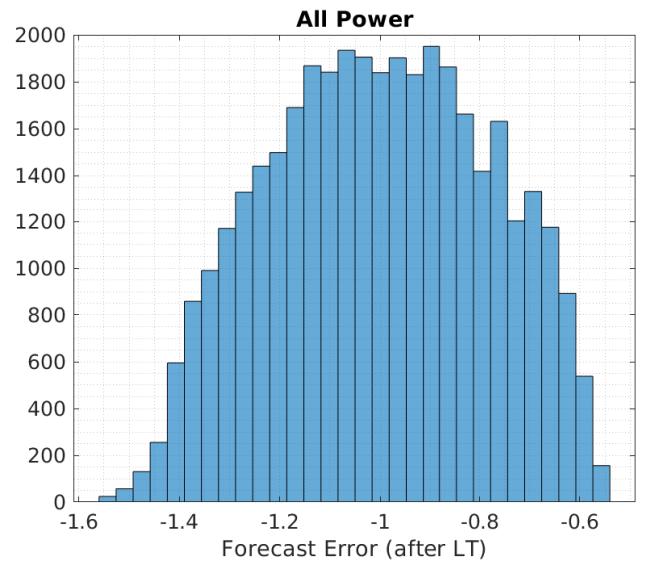
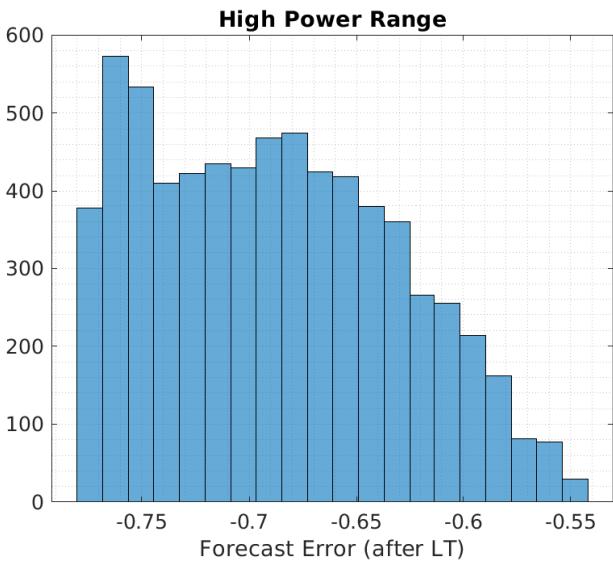
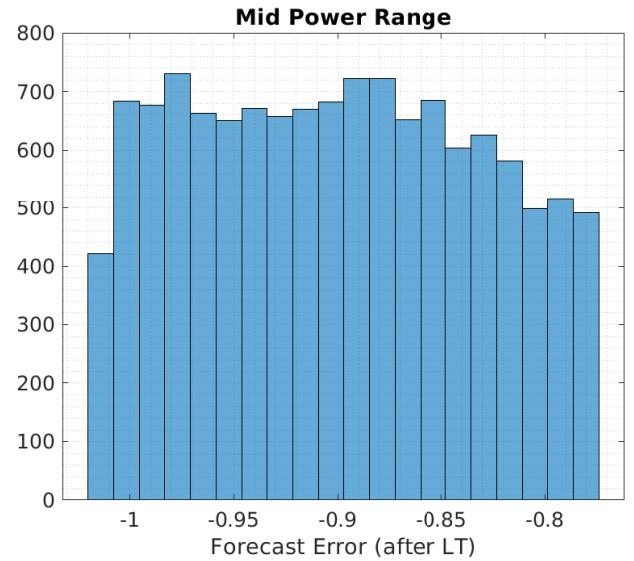
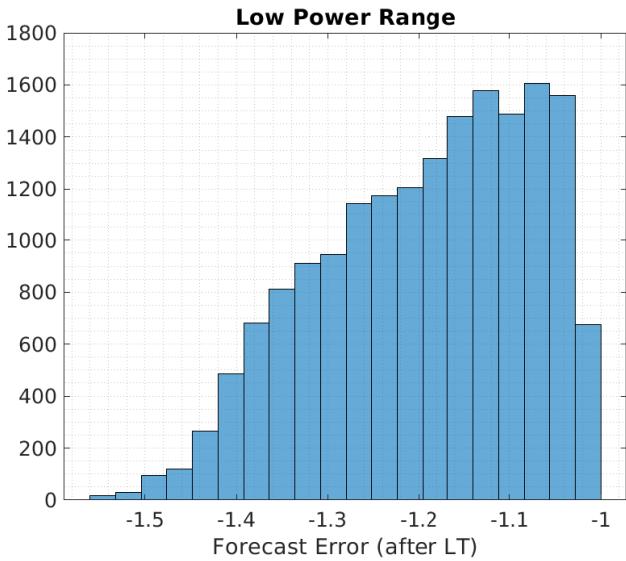


LOW : When the real production is in the interval $[0, 0.3)$.

MID : When the real production is in the interval $[0.3, 0.6)$.

HIGH : When the real production is in the interval $[0.6, 1]$.

Forecast Error after Lamperti Transform:



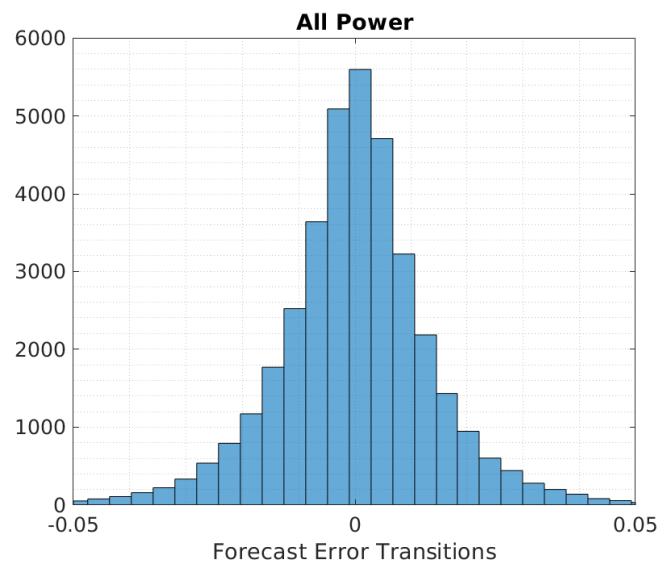
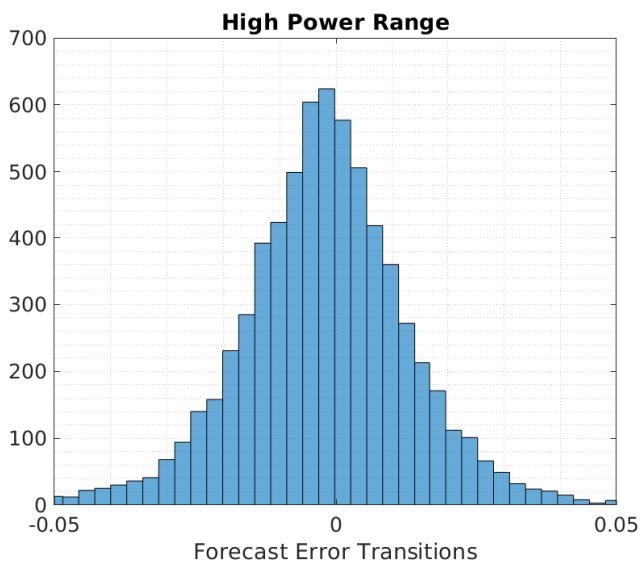
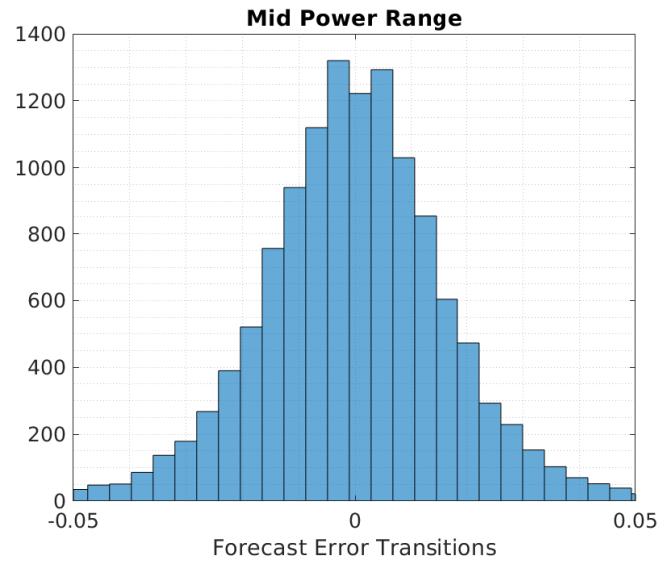
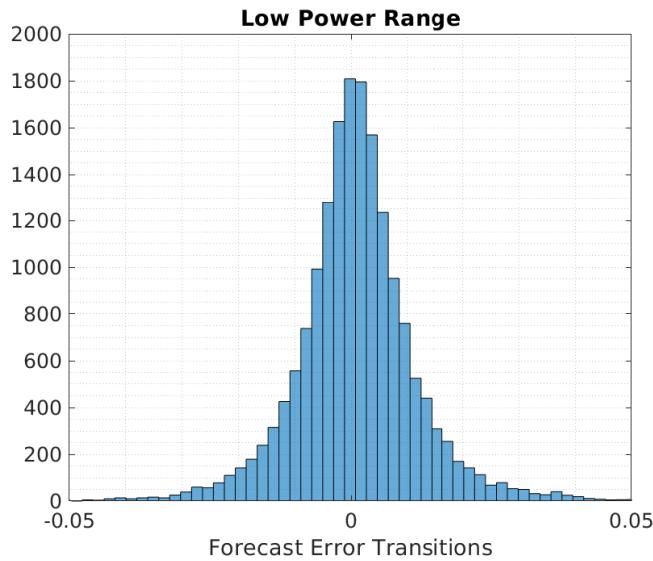
The Lamperti transform in this case is

$$Z^{(i)}(t) = \arcsin \left(\frac{V^{(i)}(t) + p^{(i)}(t)}{2} - 1 \right),$$

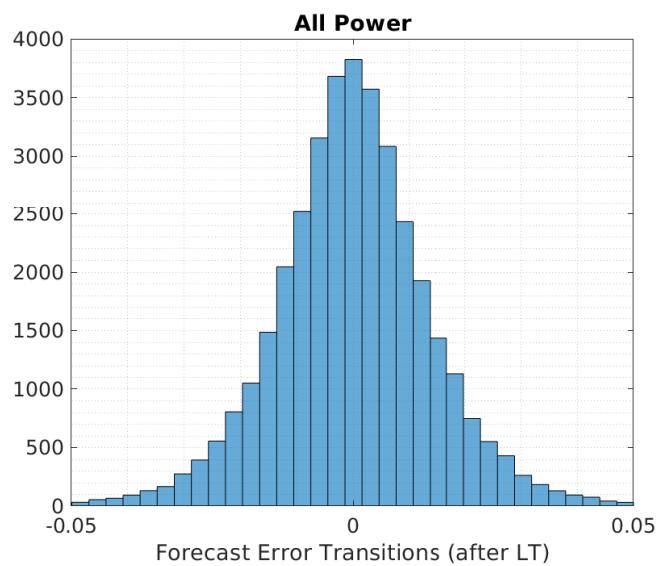
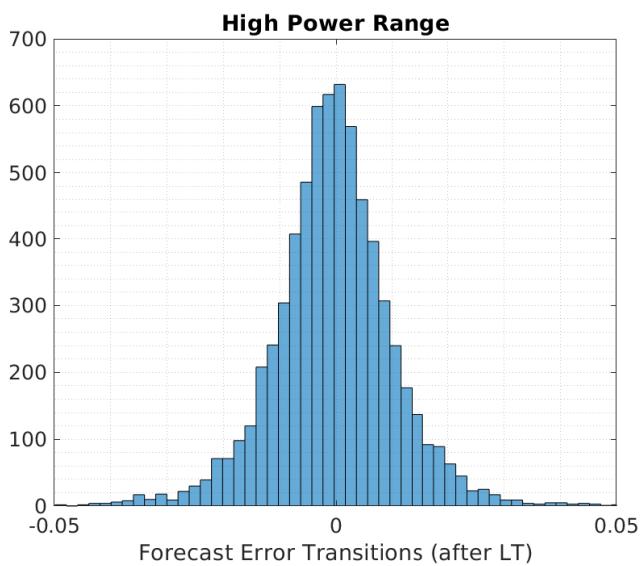
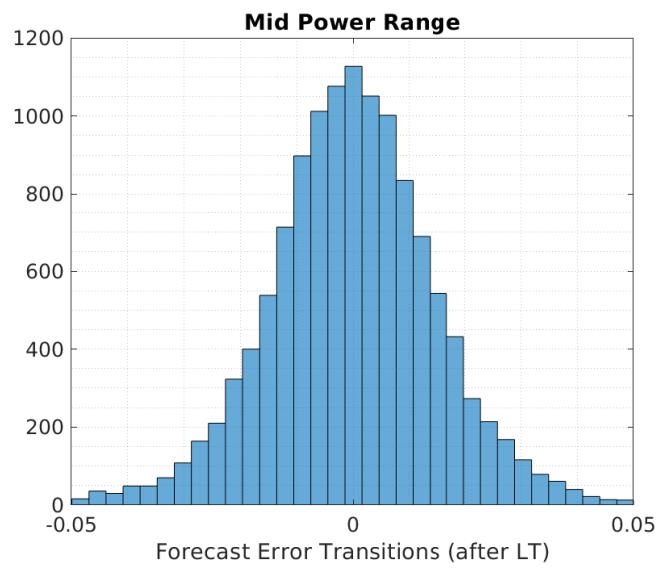
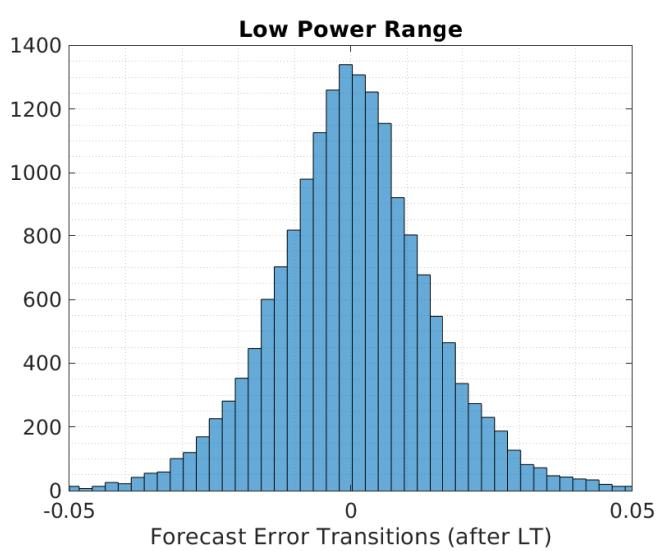
where the index i represents the path number.

Recall that $V^{(i)}(t) = X^{(i)}(t) - p^{(i)}(t)$, there $X^{(i)}(t), p^{(i)}(t) \in [0, 1]$ for all $i \in \{1, \dots, 255\}$ and $t \in [0, 1]$. Then, $Z^{(i)}(t) \in \arcsin([-1, -\frac{1}{2}])$ for the same index and time domains.

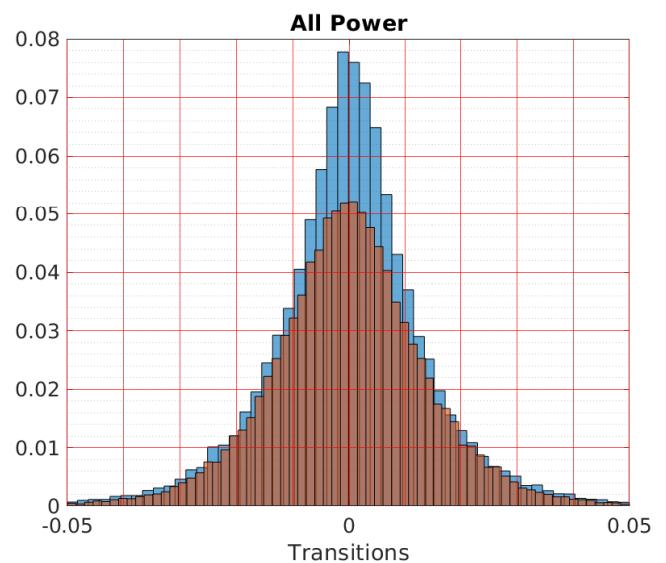
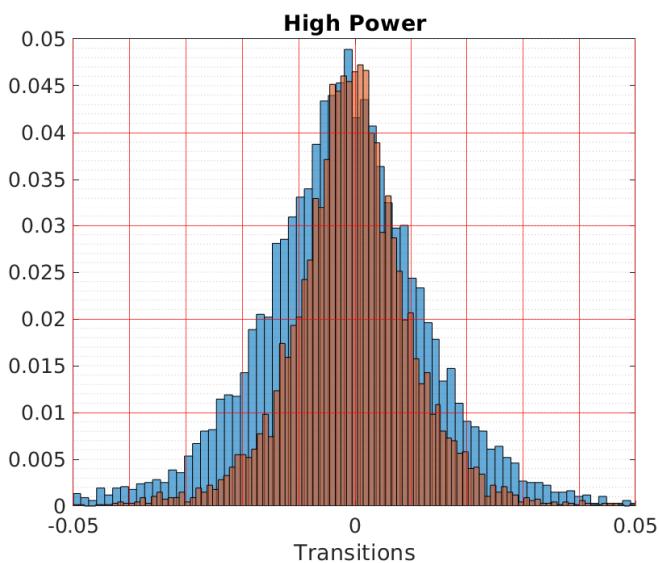
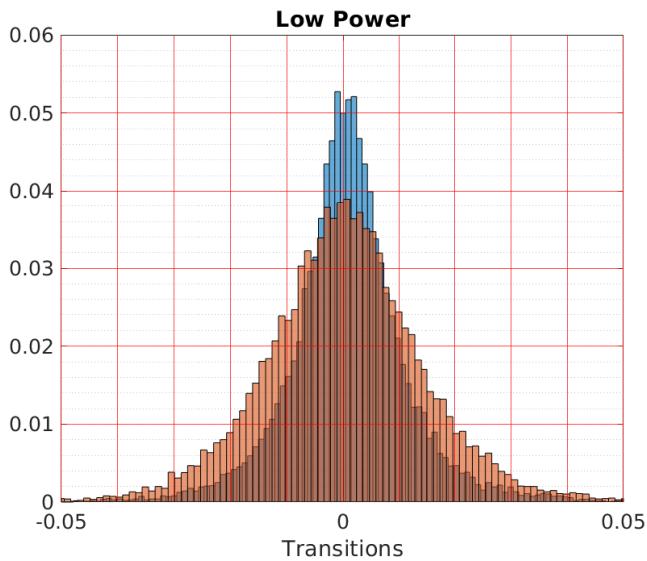
Forecast Error Transitions:



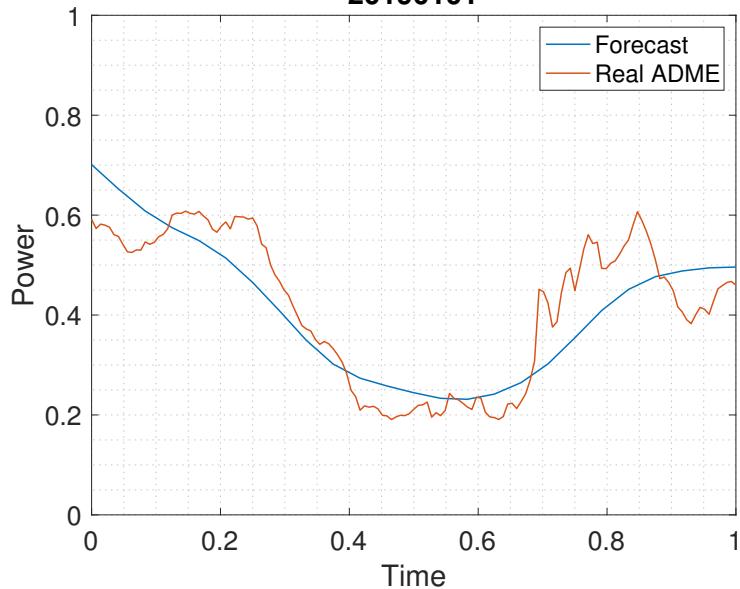
Lamperti Transform Transitions:



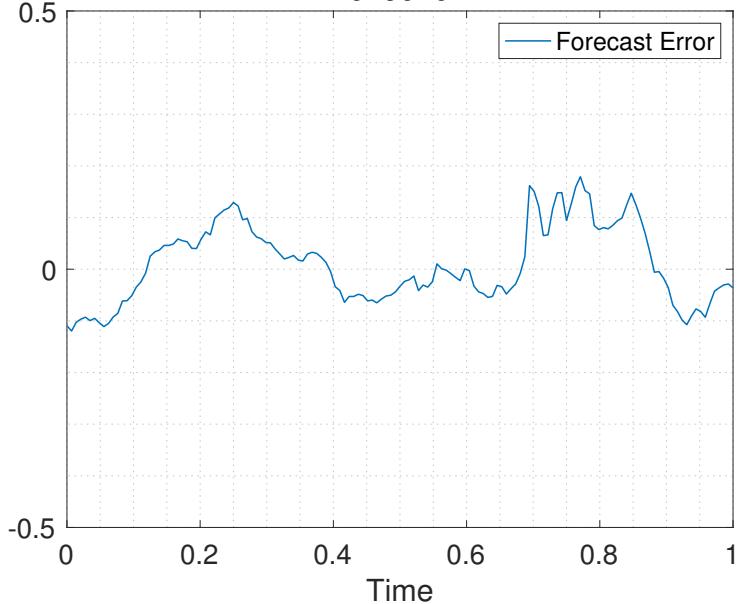
Transitions Comparison:



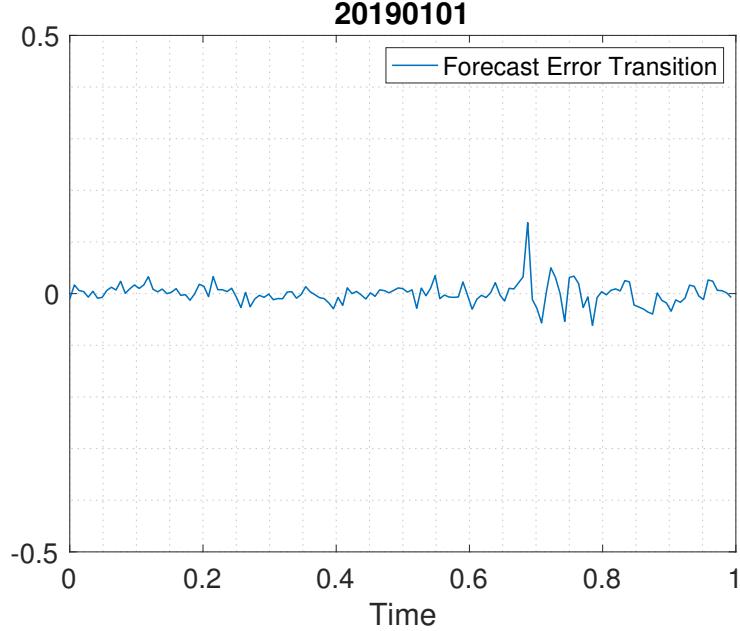
20190101



20190101



20190101



20190101

