Load forecasting for ERCOT

Todd Hay

November 3, 2016

–

1 Overview

In most of Texas the electrical grid is controlled by the Electric Reliability Council of Texas (ERCOT). The area of Texas served by ERCOT is further divided into eight weather zones. Figure 1 shows the load in each weather zone at 14:00 on 2/23/2012 with a visualization tool that we built.

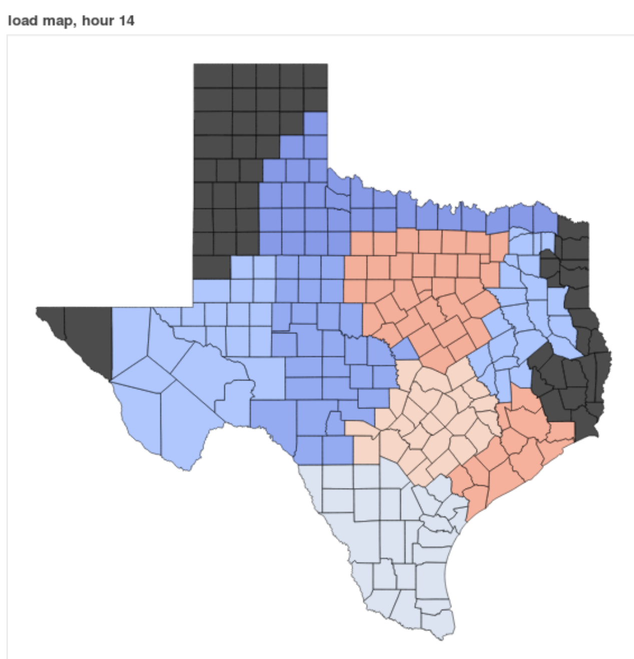

Figure 1: Load at 14:00 on 2/23/2012. The eight weather zones are different colors (indicating load). The black areas are counties in Texas not served by ERCOT. At this particular time the load ranges from 700 MW (blue) to 10,000 MW (red).

One pressing problem for ERCOT and other utilities throughout the country is the production of accurate electrical load forecasts. ERCOT produces day ahead load forecasts and provides historical hourly load data which can be downloaded. Figure 2 shows the actual load along with the current day (I believe this is one hour ahead), and day ahead forecasts produced by their models.

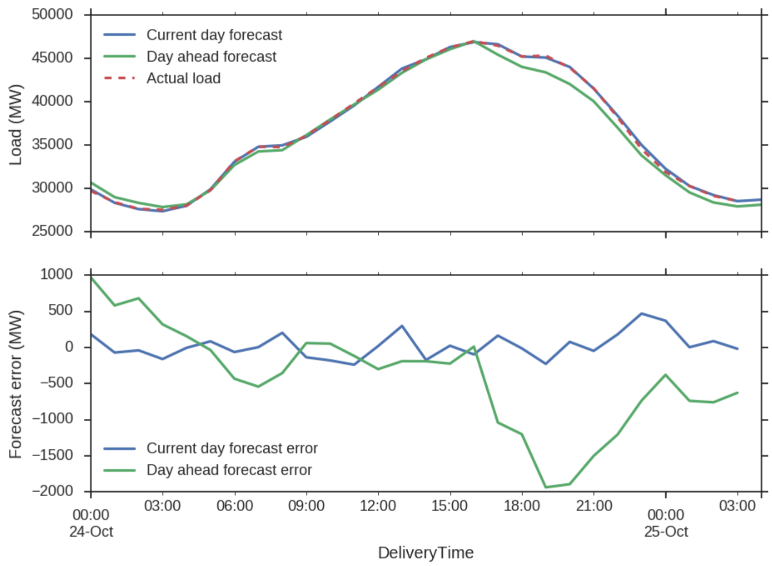

Figure 2: Top plot: Actual (dashed red) load and forecasts produced on the same day (blue) and the previous day (green). Bottom plot: Forecast error.

Load forecasting is crucial for ERCOT because bringing emergency power sources online is expensive. Accurate load forecasts allow ERCOT to plan for cheaper, but less agile, power plants to be brought online in time to generate sufficient power for the forecast load levels.

Figures 3 to 5 show the load on the ERCOT grid for all of 2015 (Fig. 3), and for August 2015 (Fig. 4) and February 2015 (Fig. 5). I've scraped load data from 2010 to the present from the ERCOT website and it's clear that the load is non-stationary (it generally increases each year as Texas' population booms) and has several seasonal components (year, week, day etc.).

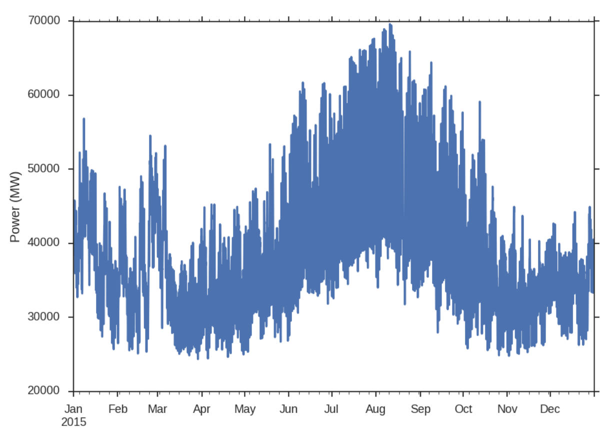

Figure 3: Load in 2015. Note the seasonal dependence.

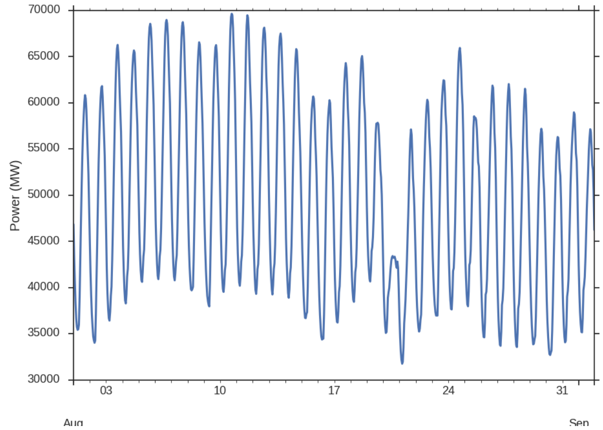

Figure 4: Load in August 2015. Note the strong diurnal cycle.

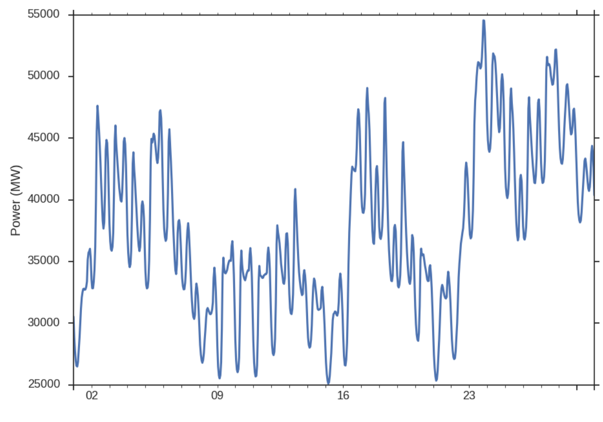

Figure 5: Load in February 2015. Something strange is happening here.

2 Forecasting using historical load alone

I've done some preliminary work with a SARIMAX model in the statsmodels package for python. Figure 6 shows the actual and forecast loads for a $ARIMA(2,0,1) \times (0,1,2)_{24}$ model for hour-ahead forecasts. Figure 7 shows the same model doing dynamic forecasts starting on April 1st. Note that the forecast error grows a lot after this point.

There's still a lot that could be done here. A few things that come to mind are:

- Do a grid search on the ARMA and seasonal parameters for the model. I believe there's an "auto.arima" function in R that does this.
- Are there ways to include multiple seasonal components? So far I've just tried 24 to take care of the diurnal component.
- So far I haven't included any exogenous factors in this model. We might want to include weather etc
- Are there ways to predict N timesteps into the future with SARIMAX instead of just one timestep (other than the "dynamic" model, above)? To be useful to ERCOT we need to predict several days in

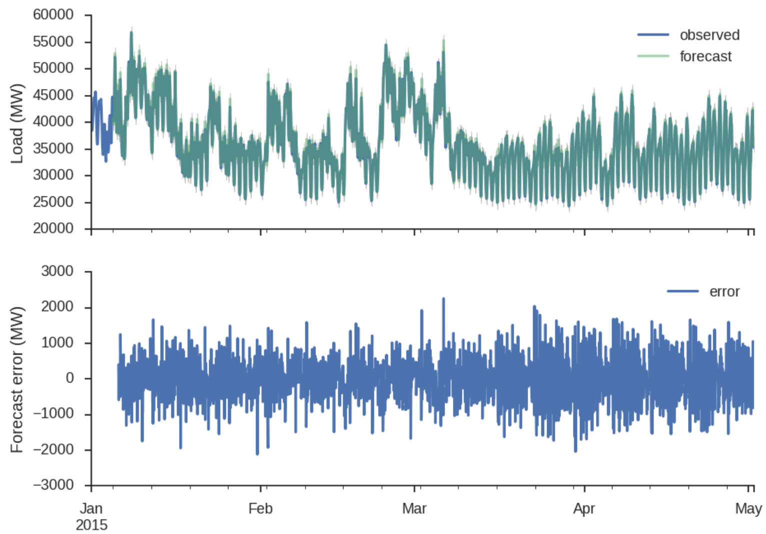

Figure 6: Hour-ahead ERCOT load forecast.

the future.

3 Exogenous factors and data sources

We expect that weather and the buisiness cycle (workdays vs weekends) to have an influence on load. ERCOT includes weather in their modeling and forecasts. We might also think about demographic (population, unemployment rate etc) and economic data. One challenge here is that demographic and economic data is available on a much coarser time scale than the hourly load data. Maybe there are some proxies that we could use?

I have built some Long Short-Term Memory neural network models which incorporate hourly weather measurements but results with this model are generally worse than with the SARIMA model above. However, there are even more "knobs" to turn with the neural network so I believe these results may improve.

3.1 Weather data

I fetch historical surface weather station data from NOAA and then interpolate the station data onto Texas counties to feed into the model. For example Fig. 8 shows the dew point distribution throughout the state for the loads shown in Fig. 1. In addition to dew point, temperature and wind speed (Texas has a significant wind power presence) were also collected.

Weather forecasts will likely be valuable as well for load forecasting. The aggregation site forecast.io provides an API but I haven't played much with it.

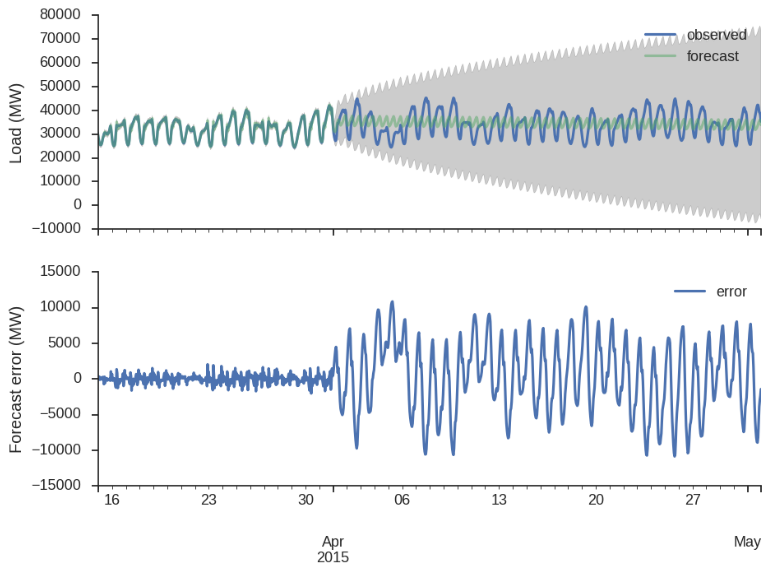

Figure 7: Dynamic ERCOT load forecast. Grey areas show the 95% confidence interval.

3.2 Demographic and economic data

There is population and economic data for each Texas county at the Texas Association of Counties website, though only available on an annual or biannual basis. Are there proxies we could use that have better temporal resolution? This is clearly a large driver of load levels so it would be good to include.

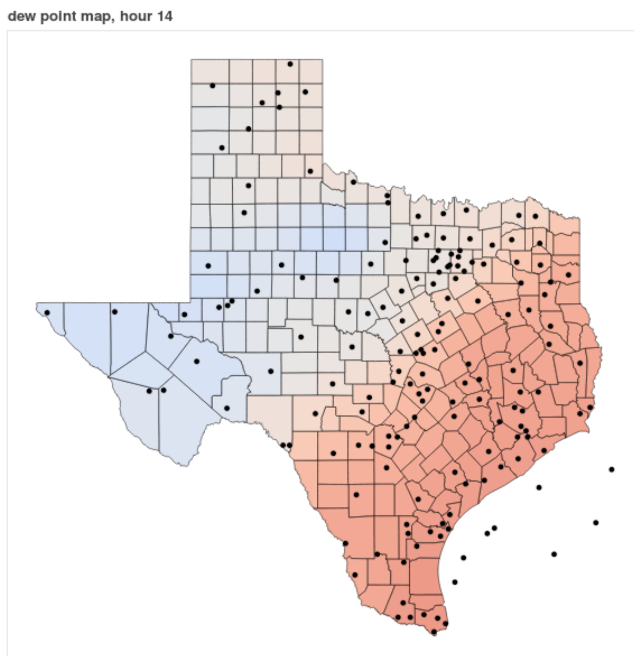

Figure 8: Temperature at 14:00 on 2/23/2012 by county. Dew points range from 50 deg. F (blue) to 70 deg. F (red). Black dots show the locations of weather stations that raw data was collected from. These raw measurements were then interpolated onto each county so the uncertainty varies by location (some counties have several weather stations while others have none).