

A First Glance Analysis of German EEX Power Market

Jacob S Townson

May 17, 2018

Introduction

This paper covers my observations and analysis of the data given to me by Josef Spalenka of Genscape in order to determine my qualifications for a position at Genscape. The data used in this document is used in the German day-ahead power market called “EEX”. It contains the power prices, fuel prices, emission prices, heating degrees, and wind and solar forecasts for the next day.

My goals with this data are to first and foremost, understand what exactly is this data presenting us with. Then, I will make some code to read in this data to make it usable in R. Then, I will look to see if there are any interesting correlations and relationships to be found in the data. I will present my findings here, and make conclusions based off of them.

The Data

The data given to me was split into 7 files, and some excel files with more sheets than others. My goal here is to explain precisely what this data is telling us, then to explore what possible correlations and relationships we may be able to find.

Coal Future Data

In this section we will be exploring the file “coal_futures_historie_2013.xls”. This excel file has 8 sheets. The first tells us most of what we need to know, and that’s that coal seems to be phasing out from the German market, if it already hasn’t been now that we are well past 2013. The second sheet indicates similarly with it’s page full of zeros.

The rest of the sheets in this excel file show the 2013 data for specific contracts and their trade unit along with prices. These bits of data make me curious, as the previously mentioned sheets seem to indicate that there were no trades through the year, however these sheets show the monthly, quarterly, and yearly trade units with their settlement prices. I can only assume this means that these prices are the prices that the trade units cost, however no contracts for these units were made.

Since we can see that most of this data is just zeros, we will ignore it from here. However, I would like to understand more of what exactly these datasets are saying, especially the differences between the “Coal-Futures Total” sheets and the contract sheets afterwards. Maybe if offered the job, I can learn more about these things!

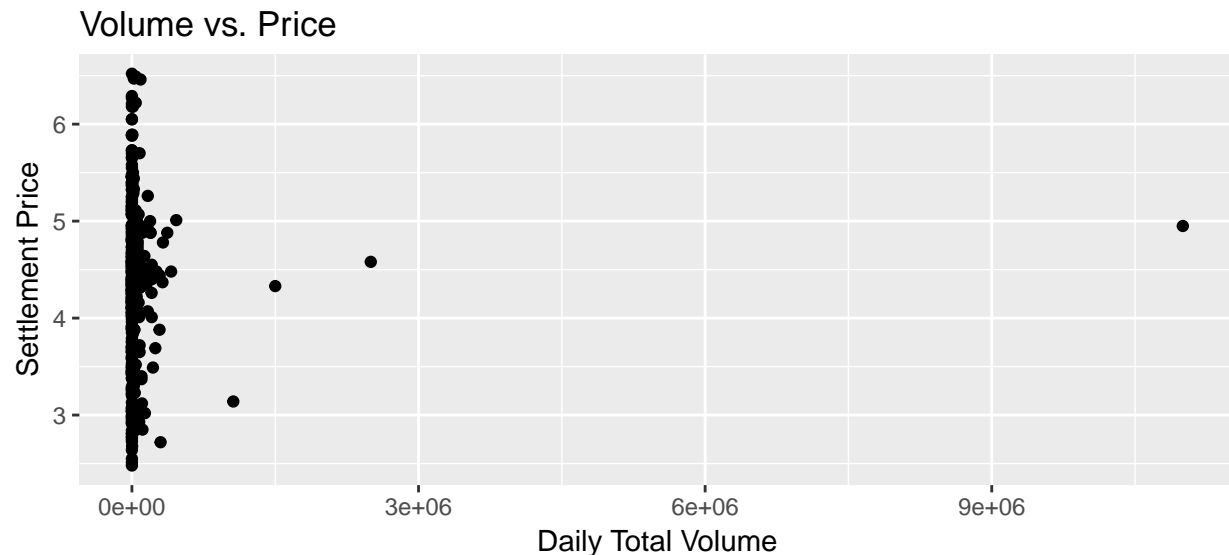
Emission History

This dataset (in the file “emission_spot_historie_2013.xls”) gives the emission prices based on specific contracts and their emission volumes. This excel file has 3 sheets, first explaining the contracts’ emission allowance, the next showing reductions, and the final showing the market auction information.

For the most part, this dataset is relatively clear in what it shows, however in the market auction sheet, we are given a column labeled “auction details”. For my purposes, this information will be ignored as without more former knowledge of the system, I will not be able to accurately try to analyze anything about it.

For curiosity's sake, let's quickly look at the emission allowance data, and compare the settlement price with the total volume of emissions. We will use the "emission_allowance" dataframe I created in R to do this efficiently.

```
em_price = emission_allowance$Settlement.Price.EUR.EUA
em_vol = emission_allowance$Daily.Total.Volume..incl..OTC.
em_cont = emission_allowance$Contract
em_trades = emission_allowance$Trades
qplot(em_vol,em_price, xlab = 'Daily Total Volume', ylab = 'Settlement Price',
      main = 'Volume vs. Price')
```



It's clear there are some outliers here, so for now, let's remove them and see what the plot looks like then.

```
allow = data.frame(em_price,em_vol,em_cont,em_trades)
allow = arrange(allow, desc(em_vol))
```

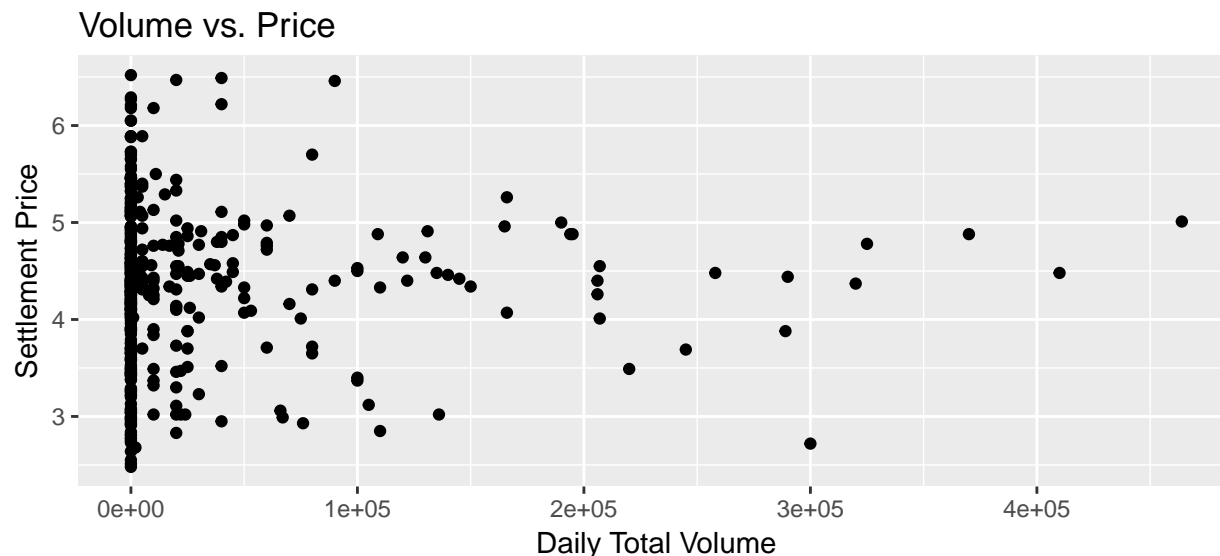
```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
head(allow)
```

```
##   em_price  em_vol em_cont em_trades
## 1    4.95 11000000    EUSP         11
## 2    4.58  2500000    EUSP          2
## 3    4.33  1500000    EUSP          2
## 4    3.14  1061000    EUSP         21
## 5    5.01   464000    EUSP          1
## 6    4.48   410000    EUSP         25
```

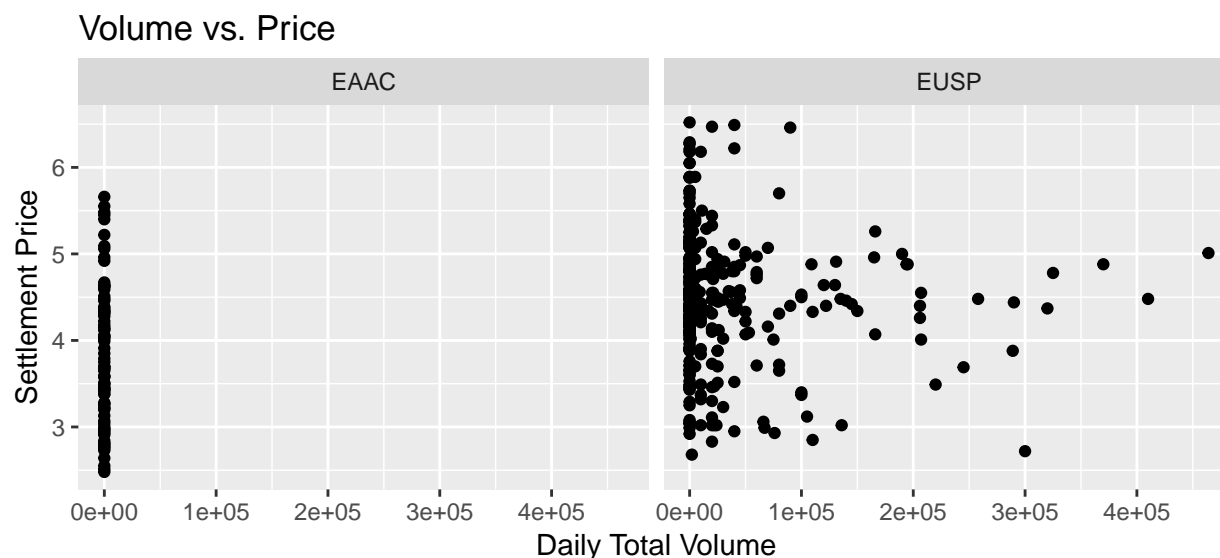
So from here we can see our outliers, and now we can remove them accordingly.

```
allow = filter(allow, em_vol < 1061000)
qplot(em_vol,em_price, data = allow, xlab = 'Daily Total Volume', ylab = 'Settlement Price',
      main = 'Volume vs. Price')
```



This plot still seems fairly noisy. Perhaps we should divide the information by contract to see if that has something to do with the noise in price. There are two types of contracts in this dataset, EAAC, and EUSP.

```
qplot(em_vol, em_price, data = allow, xlab = 'Daily Total Volume', ylab = 'Settlement Price',
      main = 'Volume vs. Price', facets = .~em_cont)
```



This seems to show just more noise, no correlation seems to show for this information. I think we can then conclude that the daily volume of emissions doesn't affect the settlement price. So let's move on!

Gas History

This dataset (found in the `gas_spot_historie_2013.xls`) is by far the most dense of any of them all. With 10 sheets, it has a lot of information to offer. The first of these sheets gives us the reference price for each market area for each day. The strange thing about this collection of data is all of the missing values. I have to wonder if this is possibly because there were no trades on these given days for these market areas, or if whoever was collecting this data was unable to get it. Either way, for time's purpose, we will simply ignore the empty values for now.

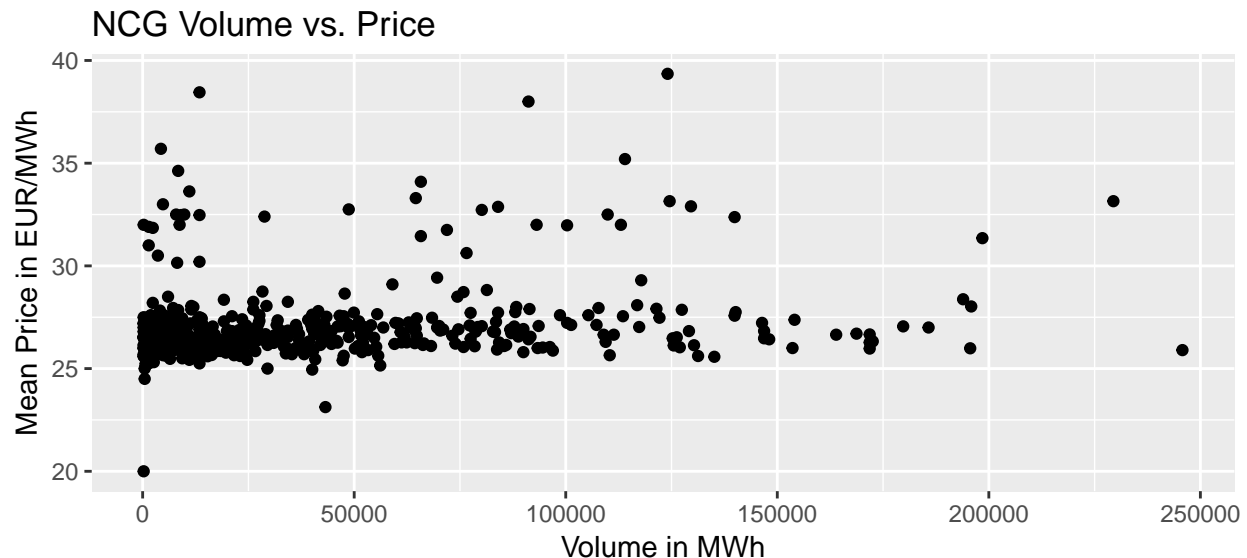
The following sheets give us data on what I have understood to be 3 European gas companies/traders, NCG (NetConnect Germany), GASPOOL, and TTF (Title Transfer Facility). For each one, we get the market information for the whole 2013 year. This has their number of trades, price information, and volume traded for every day.

Let's look and see if we can find any trends in this data! To keep things simple, let's start by looking at the ncg data, and if we find some interesting things, we'll dive down the rabbit hole. Otherwise, we'll move on. First things first, let's write a bit of code to simplify the low vs. high price problem in the data.

```
ncg.high = gas_market.ncg$High.Price.EUR.MWh
ncg.low = gas_market.ncg$Low.Price.EUR.MWh
ncg.mean_price = rowMeans(cbind(ncg.low,ncg.high))
```

Now that we have this mean price for each day, let's see if the volume vs. price plot shows us anything interesting.

```
ncg.trades = gas_market.ncg$Trades
ncg.vol = gas_market.ncg$Volume.MWh
ncg.day = gas_market.ncg$Trading.Date
NCG = data.frame(ncg.day,ncg.mean_price,ncg.vol,ncg.trades)
qplot(ncg.vol,ncg.mean_price, data = NCG, xlab = 'Volume in MWh', ylab = 'Mean Price in EUR/MWh', main =
```

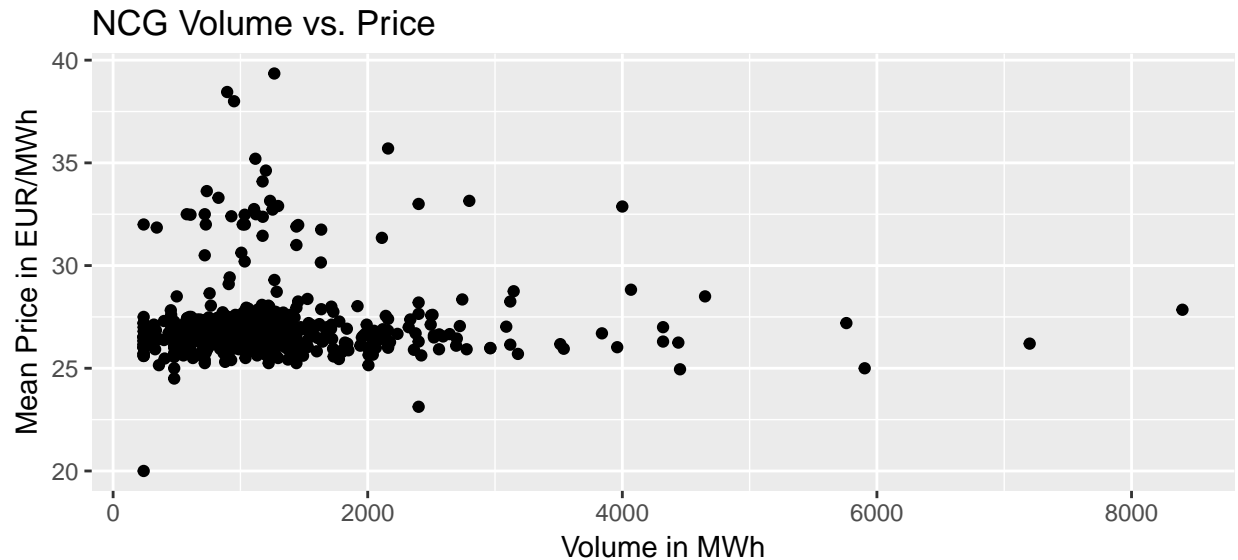


We seem to get a strange looking plot here. It almost looks as if no matter the volume, the price stays between 25 and 30 euros. This makes me wonder if possibly the number of trades has something to do with it. Let's take a look.

```
NCG = filter(NCG, ncg.trades > 0)
NCG = mutate(NCG, vol_trade = ncg.vol/ncg.trades)
```

What I have done in this code chunk is to first remove any zero trade values in the data, then make a new variable to show roughly the amount of volume per trade.

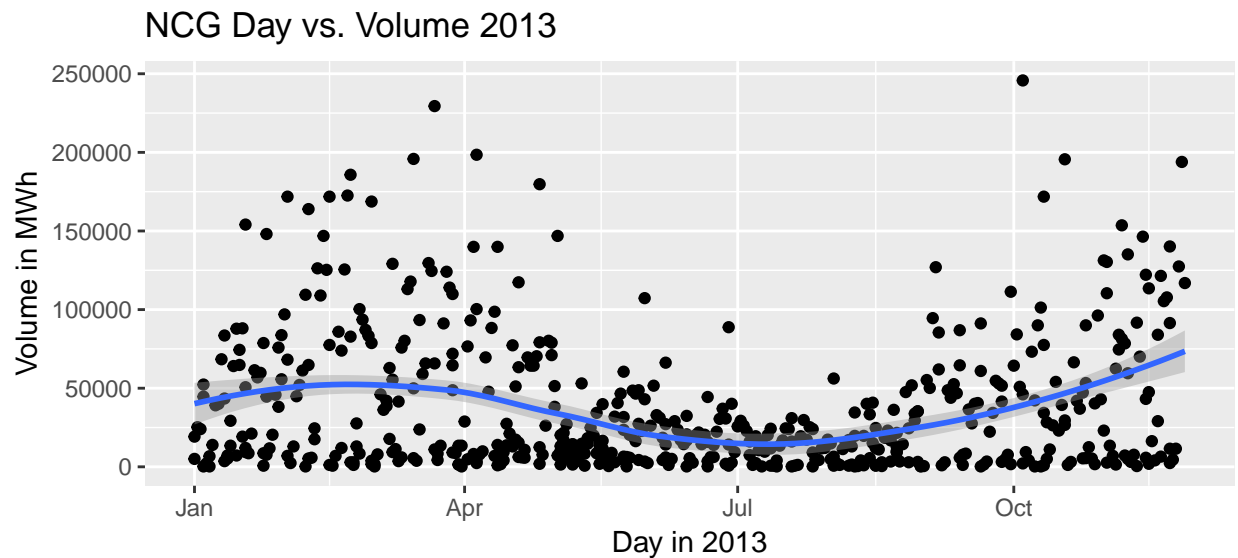
```
qplot(vol_trade,ncg.mean_price, data = NCG, xlab = 'Volume in MWh', ylab = 'Mean Price in EUR/MWh', main =
```



This didn't really clear much up for us. So seemingly the price in euros per MWh is not affected by the volume. Before we move on, let's take a quick look and see how it changes throughout the year though.

```
qplot(ncg.day, ncg.vol, data = NCG, xlab = 'Day in 2013', ylab = 'Volume in MWh', main = 'NCG Day vs. V
```

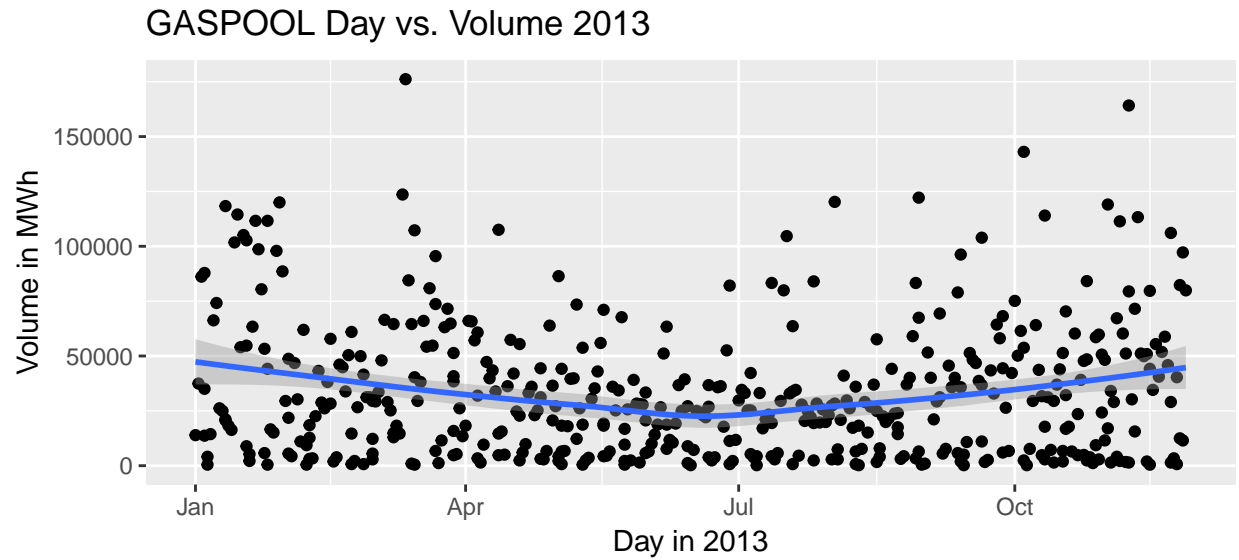
```
## `geom_smooth()` using method = 'loess'
```



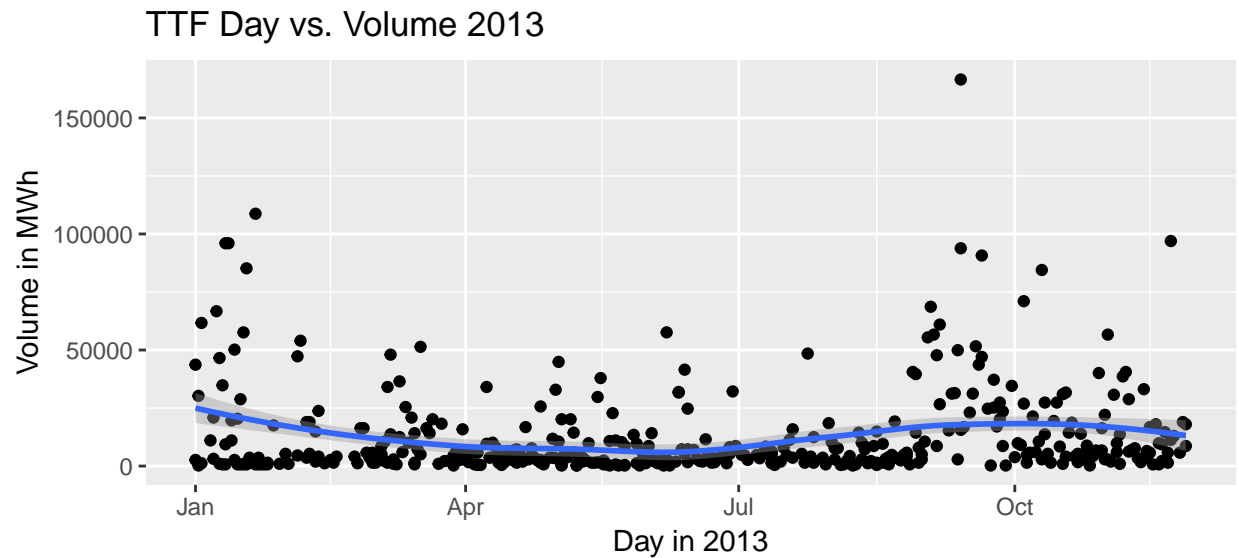
Here we see an interesting bit of information. There was a bit of a surge of volume traded right before April, and going into December. This could mean a number of things, from people needing to heat their houses in the winter months, to people driving and using fuel to travel during spring. Without more data, we can't be sure, but we can make some interesting assumptions!

If we take this work and apply it to the GASPOOL and TTF data, and look at the time vs. volume plot, this is what we find:

```
## `geom_smooth()` using method = 'loess'
```



```
## `geom_smooth()` using method = 'loess'
```



It's strange, the GASPOOL data doesn't seem to tell us much when it comes to the time of year other than there is a dip in average volume around July, but even that is slight. The TTF data shows us a little more, however it strangely looks almost like the exact opposite of the NCG data. Why this is, I can't say from this data alone, but it is interesting to note nonetheless.

Heating Degrees

Wind and Solar Forecasts

Power Prices (energy__spot)

Analysis

Conclusions

Appendix