

Grid Frequency Analysis

Michael Glavind Bülow Pedersen

In this short project, I use some of the techniques from the excellent Coursera [Data Science Specialization](#) from John Hopkins to first download and then analyse two large datasets of the power grid frequency in EU.

The R code to generate this PDF and to download data from the external sources can be found at my [GitHub repository](#).

The rest of this document is structured as follows: Section 1 contains an introduction to the research problem. In Section 2 a description of data sources and examples of API return values are provided. In section 3 we do a short graphical analysis of the data. This allows us to answer the research questions, and in Section 4 we conclude.

Research Objectives

1. Download data from an API and a Web server using R.
2. Store and handle large datasets (up to 31,536,001 observations per year).
3. Handle timeseries data that cross daylight saving.
4. Create beautiful, informative graphs using both the ggplot2 and Lattice packages.
5. Generate this document in R using Knitr/Markdown.
6. Answer research questions:
 - What differences in grid frequency quality can be observed between the Nordic and Continental electricity grids?
 - Is the European grid frequency exactly 50 Hz on average?

1 Introduction

The European electricity grid consists of several distinct frequency areas. One of these frequency areas is the smaller Nordic grid, which covers Norway, Finland, Sweden, and Eastern Denmark. Another of these frequency areas is the larger Continental grid, which covers countries such as Germany, France, Spain, Portugal, Italy, Slovenia, Czech Republic, Switzerland, Slovakia, Poland, etc. This analysis will focus on comparing these two frequency areas.

To keep a power grid stable it is necessary for the grid frequency to move within a narrow band around a pre-determined frequency setpoint. For the European grid the setpoint has been fixed at a reference frequency of 50 Hz \pm 1mHz. Whenever the frequency moves below this level, Transmission

System Operators have contracted grid connected market participants to supply power to the grid at seconds notice. Similarly, if the frequency moves above 50 Hz TSOs have contracted participants to withdraw power from the grid.

In the Continental grid extreme events are characterized by frequencies ± 100 mHz. Especially the low frequency bound is critical because below -100 mHz all primary reserves have been exhausted and the grid has to cope with additional loss of generation without backup supply. Below -150 mHz there is a real risk of power plants dropping out, because of the low frequency. But an extreme positive frequency, can also become problematic for grid-connected entities. For example, at +200 mHz solar panels are at risk of dropping out to protect their electronics, leading to a disruption in supply.

Because of these issues, the grid frequency is tightly controlled by TSOs and any deviation is generally promptly dealt with.

In Europe it has furthermore been agreed to run the power system at exactly 50 Hz on average, so that systems can keep track of the approximate time by counting measuring the grid frequency. This way systems know that 50 fluctuations in the grid equal 1 second in realworld time.

But, because accumulated deviation can move away from the zero due to persistent grid frequency deviations, it is sometimes necessary to correct the grid frequency by setting a lower reference frequency point on subsequent days. In practice this is done by fixing the frequency setpoint at 49.99 Hz for 24 hours if the grid time is ahead of clock time and at 50.01 Hz if the grid time is behind clock time. When this correction is performed, the correction to the grid-time is 17.28 seconds/day.

The time correction should be observable in the daily average frequency, whenever the grid time deviates significantly from clock time, leading to autoregressive effects. We should also observe that the long-run average deviation from 50Hz is zero.

2 Getting the data

This sections gets nitty gritty on how to retrieve frequency data. For the Continental grid this is straightforward, because the French Transmission System Operator (TSO) RTE provides this data in nicely formatted csv-files. For the Nordic grid however, we have to retrieve the data from an undocumented api at the Norwegian TSO Statnett. This requires some probing and error correction in the code.

If data sources and data retrieval is not of interest, you may proceed to the analysis. We end up with two dataset for the first nine months of 2015.

For the Continental grid we have an observation for each 10 seconds leading to 2,099,160 observations. For the Nordic grid we have an observation each second leading to 20,991,600 observations.

In the sample period there was a summer Daylight Savings Time (DST) change in Hour 3 on 2015-03-29, where the hour does not exist in local time. Similarly there will be a winter DST change

in Hour 3 on 2015-10-25 (not included in sample), where there will be two Hour 3's.

The code to download from both sources can be found in “/code/getData.R”.

2.1 Nordic

We query the API of the [Norwegian TSO Statnett](#) to get 1 second measurements (31,536,000 observations per year).

The call to the API is <http://driftsdata.statnett.no/restapi/Frequency/BySecond?FromInTicks=FromTimeInMs&ToInTicks=ToTimeInMs>. This GET request returns 1 second measurements as well as startDate and endDate of the request in Local timezone. The inputs are in ms since epoch and therefore UTC.

Data retrieval is somewhat complicated by the fact that the return data do not have datetimes associated with each observation. This becomes a problem because there are missing observations, which are just excluded from the returned list. This happens for instance on 2015-01-07 15:33:00 to 15:34:00, where only 39 observations are returned. To get a complete dataset, we are therefore forced to download an hour, check if it is complete. If not, get each minute and check if complete. If not, get each second and check if complete and if not record a missing observation. This process is time consuming (one GET request for one hour takes appr. 0.3-0.5 sec.), especially when we have to loop across each minute and second. To get a full year the approximate runtime is on the order of multiple hours.

The current implementation sometimes result in duplicate observations being added to the data frame. I have not prioritized finding the place where duplication occurs, so for now duplicate values are just excluded afterwards using the unique() function.

The API can only return 1 hour of data (probably 3600 observations). So,

```
fromDate <- ymd_hms("2015-10-19 07:00:00")
toDate <- ymd_hms("2015-10-19 08:00:00")
```

returns data from “2015-10-19 09:00:00” to “2015-10-19 09:59:59”

API Examples

The following examples illustrate how Daylight Savings Time (DST) affects return values.

Before DST:

```
fromDate <- dmy_hms("28-03-2015 12:00:00")
toDate <- dmy_hms("28-03-2015 13:00:00")
```

“2015-03-28 13:00:00”

“2015-03-28 13:59:59”

After DST:

```
fromDate <- dmy_hms("29-03-2015 12:00:00")
toDate <- dmy_hms("29-03-2015 13:00:00")
```

"2015-03-29 14:00:00"

"2015-03-29 14:59:59"

Under DST (02:00-03:00 local time / 01:00-02:00 UTC does not exist):

```
fromDate <- dmy_hms("29-03-2015 00:00:00")
toDate <- dmy_hms("29-03-2015 01:00:00")
```

"2015-03-29 01:00:00"

"2015-03-29 01:59:59"

```
fromDate <- dmy_hms("29-03-2015 00:30:00")
toDate <- dmy_hms("29-03-2015 01:30:00")
```

"2015-03-29 01:30:00"

"2015-03-29 03:29:59"

```
fromDate <- dmy_hms("29-03-2015 01:00:00")
toDate <- dmy_hms("29-03-2015 02:00:00")
```

"2015-03-29 03:00:00"

"2015-03-29 03:59:59"

2.2 Continental

We can query the Website of the [French TSO RTE](#) to get 10 second measurements (3,153,600 observations per year).

The data is returned as a zip file containing one csv file with 10 second measurements for one month. The format of the date is local time (Europe/Paris), with NA for hour 3 (02:00:00-02:59:50) in spring DST.

There are also a number of missing values in this data set.

3 Analysis

3.1 Initial data validation

Firstly, we check the raw data for missing observations.

```
## Continental: missing 8825 observations ( 0.3742 %)
```

```
## Nordic: missing 1413 observations ( 0.0060 %)
```

The relatively high amount of missing observations in the Continental data stems from one whole April day missing.

Secondly, we print a sample of the raw data. Here we see that even though the Nordic data is recorded at each second, changes only occur every 8-10 secs. This might indicate that the actual sampling is done every 8-10 seconds and not every second leading to a loss in precision.

```
# Continental:  
head(FC,15)
```

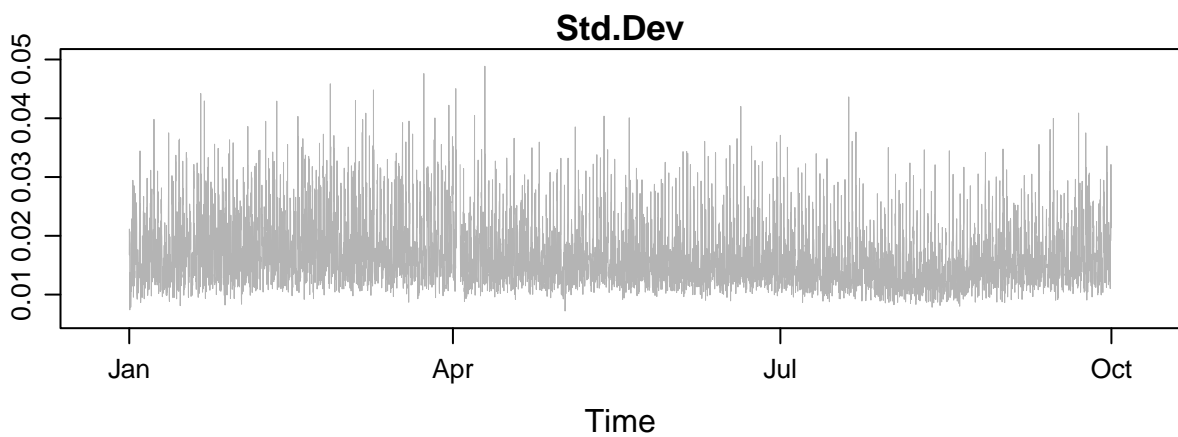
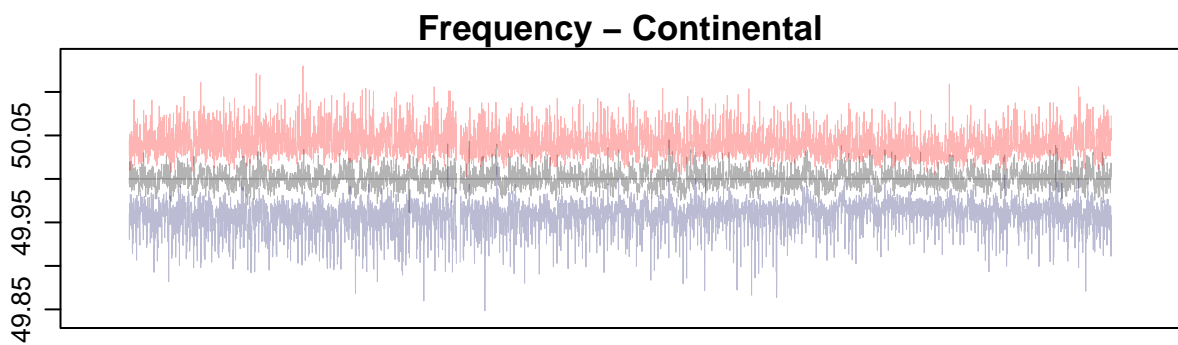
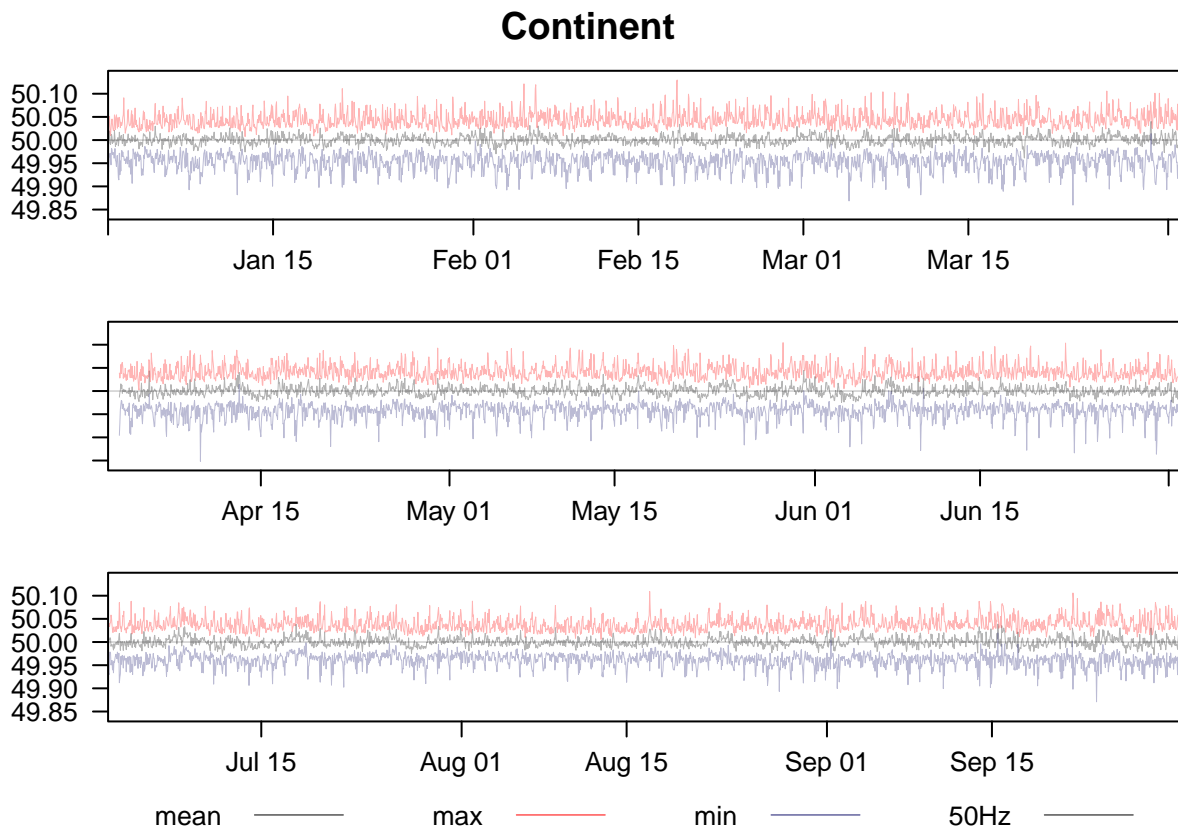
```
##                frequency  
## 2015-01-01 00:00:00    50.0184  
## 2015-01-01 00:00:10    50.0169  
## 2015-01-01 00:00:20    50.0141  
## 2015-01-01 00:00:30    50.0104  
## 2015-01-01 00:00:40    50.0025  
## 2015-01-01 00:00:50    50.0016  
## 2015-01-01 00:01:00    50.0014  
## 2015-01-01 00:01:10    50.0092  
## 2015-01-01 00:01:20    50.0023  
## 2015-01-01 00:01:30    50.0021  
## 2015-01-01 00:01:40    50.0009  
## 2015-01-01 00:01:50    49.9965  
## 2015-01-01 00:02:00    49.9853  
## 2015-01-01 00:02:10    49.9796  
## 2015-01-01 00:02:20    49.9749
```

```
# Nordic:  
head(FN,15)
```

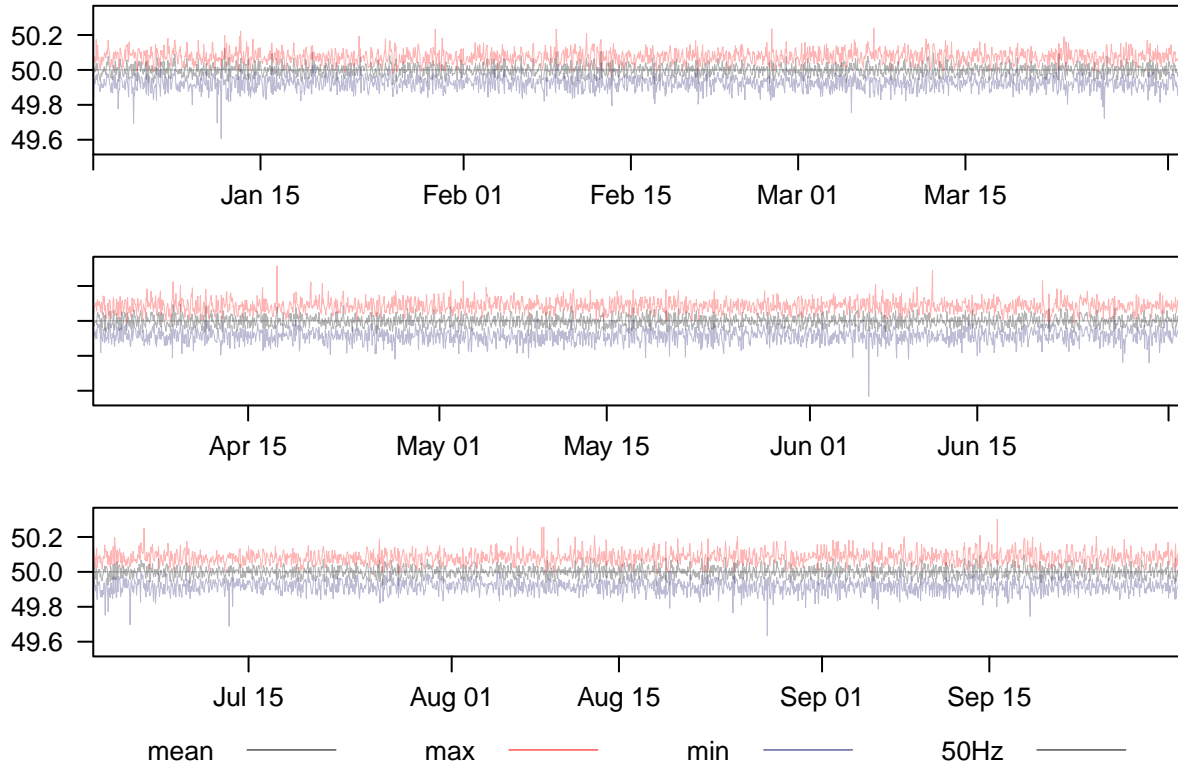
##	frequency
## 2015-01-01 00:00:00	49.994
## 2015-01-01 00:00:01	49.994
## 2015-01-01 00:00:02	49.994
## 2015-01-01 00:00:03	49.994
## 2015-01-01 00:00:04	49.994
## 2015-01-01 00:00:05	49.963
## 2015-01-01 00:00:06	49.963
## 2015-01-01 00:00:07	49.963
## 2015-01-01 00:00:08	49.963
## 2015-01-01 00:00:09	49.963
## 2015-01-01 00:00:10	49.963
## 2015-01-01 00:00:11	49.963
## 2015-01-01 00:00:12	49.963
## 2015-01-01 00:00:13	49.963
## 2015-01-01 00:00:14	49.949

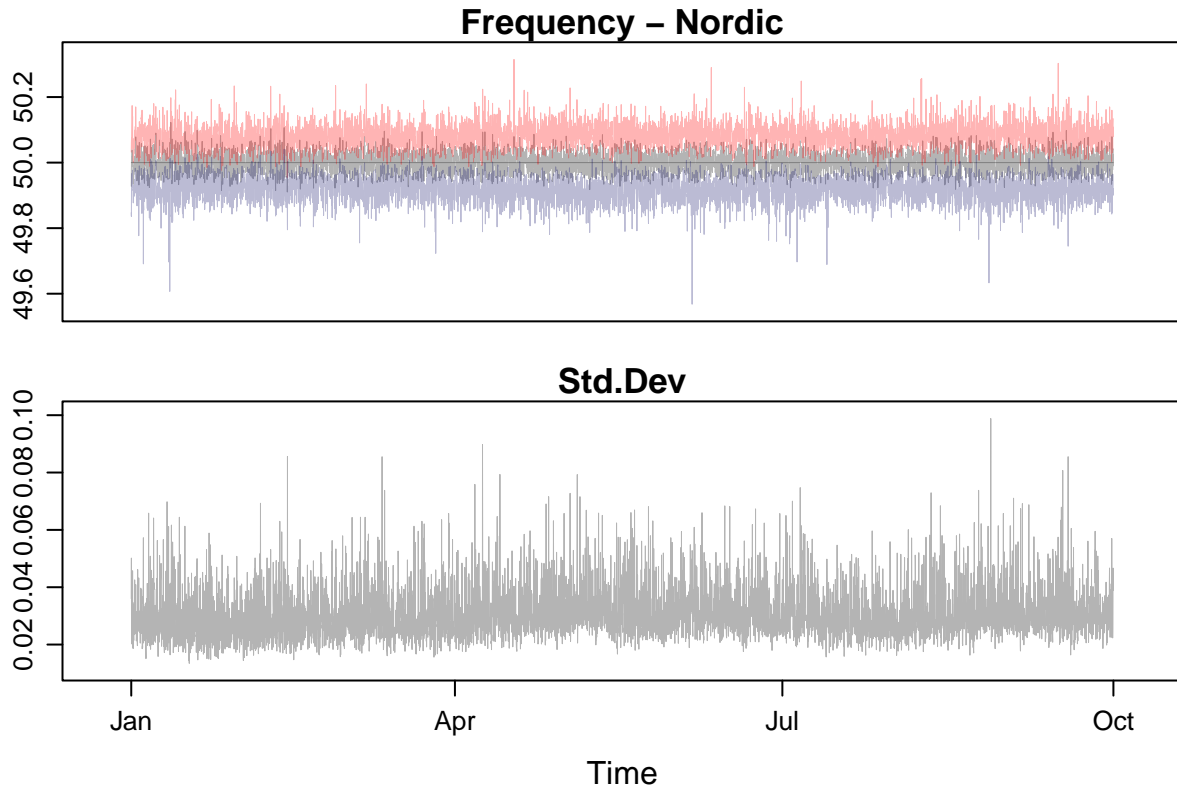
3.2 Exploring the data

To get an initial understanding of the data, we plot the average, maximum, and minimum grid frequency for each hour in the sample. From the two graphs it is evident that the Nordic grid frequency shows significantly more extreme values than the Continental grid and has a larger variation within hours. This might be because of the more frequent grid measurements in the raw Nordic data, but is also the expected result as a smaller grid, such as the Nordic, has less capacity to absorb power imbalances and therefore likely show larger frequency variations.



Nordic

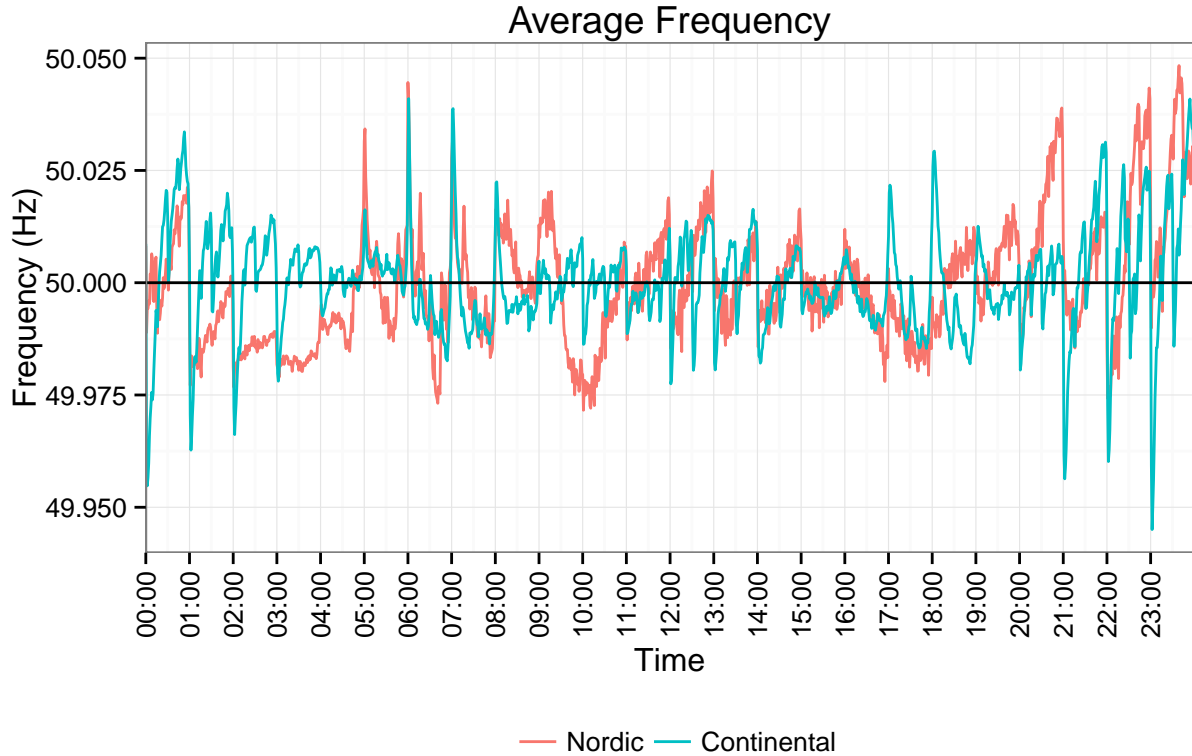




Next, we create a plot of the average grid frequency across the day.

Here we observe clear trends in the grid frequency with notable spikes around the changes in hour. This is likely a market design artifact that materializes because trading and power exchange scheduling between countries is typically done in hourly intervals (though during the last 5 years quarterly periods has also begun trading with some volume internally in Germany). So when the power flows change around the hour a mismatch the timing of changes in generation and consumption can affect the grid frequency.

The largest variation for the Continent is seen around evening/night when generation has to adapt to a falling consumption and in the morning where consumption picks up and generation has to follow. For the Nordic system, we largely see the same pattern.



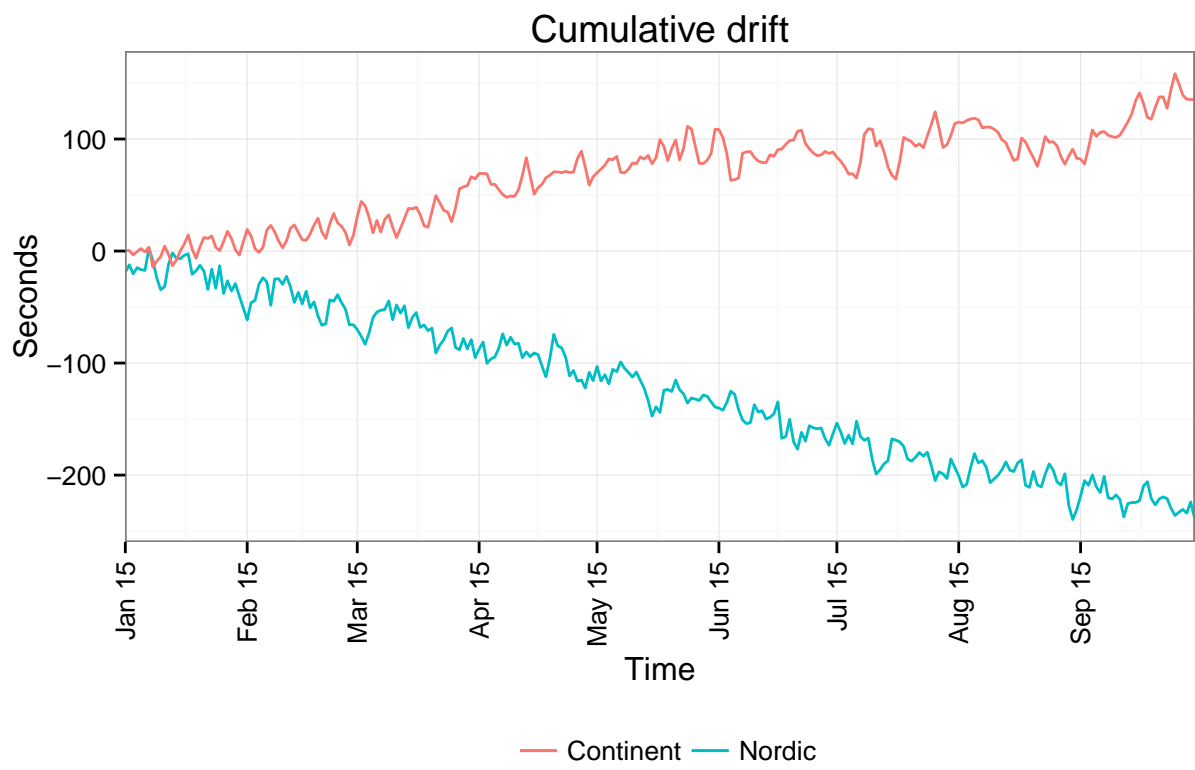
3.3 Frequency drift

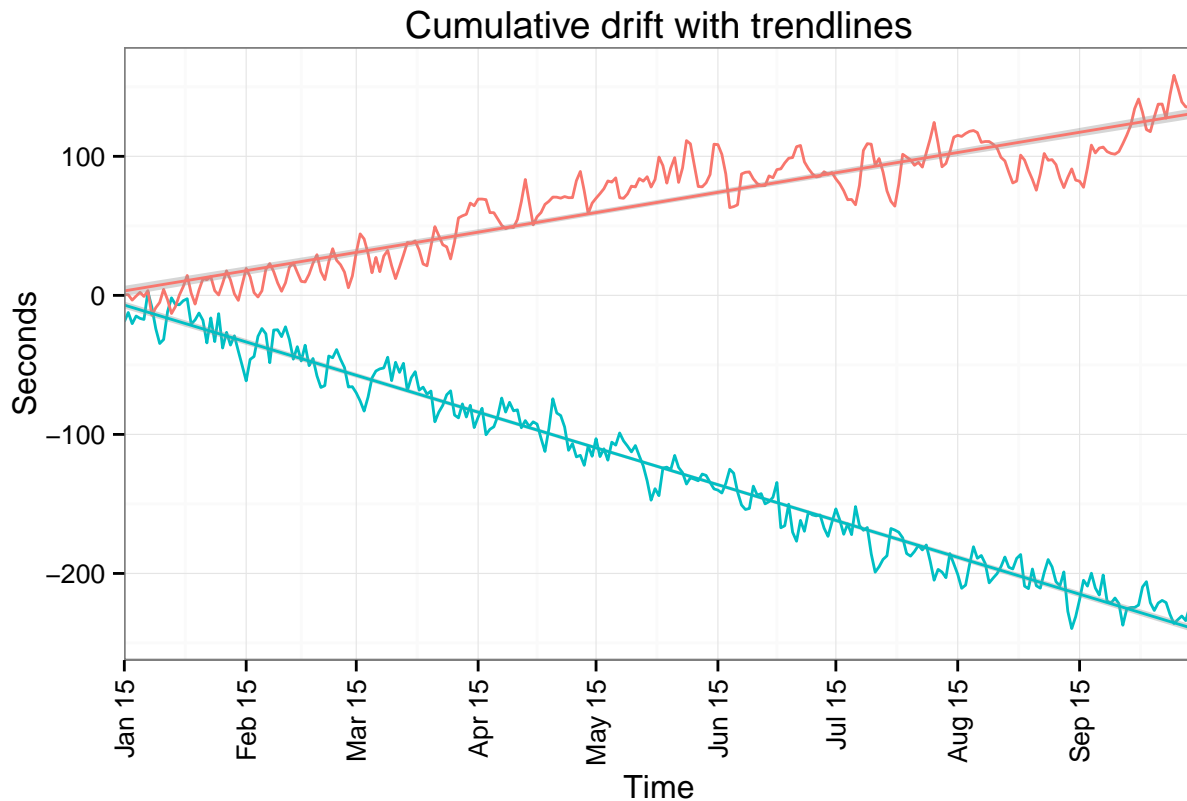
Last, we plot the clock drift in the two grids. This is the amount of seconds that a grid-run clock would deviate from an atomic clock during the period. We see that, contrary to the previous statements, the clock drift is not zero. In the Continental grid the clock would run around 140 seconds faster on 1. Oktober compared to the atomic clock, while the Nordic clock would run around 225 seconds slower.

On the homepage of [Swissgrid](#) we can see the current clock drift for the Continental grid (which is nowhere near as much off). This suggests that our data lacks the true precision needed to run an accurate grid clock.

This could either be due to the observations being point measurements of grid frequency (instead of providing the average over the last 1/10 seconds), it could be a small lack of precision in each measurement that adds up, or it could be due to shifts during the periods with missing data (the frequency during missing periods is assumed to be 50Hz exactly so there no change in clock drift is induced during missing periods).

Nonetheless, we do see clear reversals in the grid frequency. Fitting a straight line to the data points (to account for a constant drift in the series) we observe what looks like systematic reversals to this trend line.





4 Conclusion

In this short project, we have shown how to download frequency measurements at a 1/10 second resolution from the webpage of two European Transmission System Operators. The resulting data were then validated, cleaned, and put in a structured format. Then we used two of R's graphing packages to produce a couple of informative graphs which reveal the following insights:

- The Continental grid has a more stable frequency than the Nordic grid.
- The frequency quality is most strained around the change in hour and during ramping of production.
- The public data from the TSO's are not precise enough to serve as a grid clock (which should be possible for a proper frequency controlled clock that is directly connected to the grid), but we do observe reversals in the grid frequency which is consistent with an active frequency management aimed at keeping the long-run average frequency at 50 Hz.

Future Research / Links

For more information on the management of grid frequency check [ENTSOE](#).

In the “Network Code on Load-Frequency Control and Reserves” some commonly agreed measures of frequency quality are presented. It would be interesting to apply these quality measures on the downloaded grid data to more formally assert the quality of the grid frequency in the Continental and Nordic grids.