

# Grid Frequency Analysis

*Michael Glavind Bülow Pedersen*

## Research Objectives

1. Download data from an API and a Web server.
2. Store and handle a large dataset (31,536,001 observations per year).
3. Handle timeseries data that cross daylight saving.
4. Create beautiful, informative graphs.
5. Craft a piece of reproducible research.
6. Make a rStudio presentation showing the results.
7. Answer research questions:
  - is the European power grid frequency exactly 50 Hz on average?
  - what differences in grid frequency quality can be observed between the Nordic and Continental electricity grids?
  - how many occurrences of extreme events was there in the sample?

## Introduction

The European electricity grid consists of several distinct frequency areas. One frequency area is the Nordic grid which covers Norway, Finland, Sweden and Eastern Denmark. Another frequency area is the Continental grid which covers countries such as Germany, France, Spain, Portugal, Italy, Slovenia, Czech Republic, Switzerland, Slovakia, Poland, and all the way to the border of Turkey...

To keep a power grid stable, it is necessary for the grid frequency to move within a narrow band around a predetermined value. For the European grid the grid frequency 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 TSO's have contracted participants to withdraw power from the grid.

In the Continental grid extreme events are characterized by frequencies  $\pm$  100 mHz. Especially the low frequencies bound is critical because below -100 mHz all primary reserves have been exhausted and the grid cannot cope with additional loss of generation. Below -150 mHz there is a serious risk of power plants dropping out, because of the low frequency. On the other hand, at +200 mHz solar panels are also at risk of dropping out potentially leading to volatile situations..

It has been agreed to run the power system at exactly 50 Hz on average, so that systems can keep track of the approximate time by measuring grid frequency. Because accumulated short-term frequency deviations can move the average above 50 Hz on one day it is necessary to correct the grid frequency by setting a lower reference frequency point on subsequent days. This should be observable as effects in the daily average frequency measurements. Because of this correction, it should be possible to set a watch by measuring the grid frequency (give or take about a minute). We will see whether this correction does indeed occur and if the clock drift on average is zero.

## Getting the data

This section gets nitty gritty on how to retrieve frequency data. For the continent it is straightforward, because the french 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 actual 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 observations 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.

## 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 NA. 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 figured out 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"
```

## 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.

## Analysis

### 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.

```
# 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
```

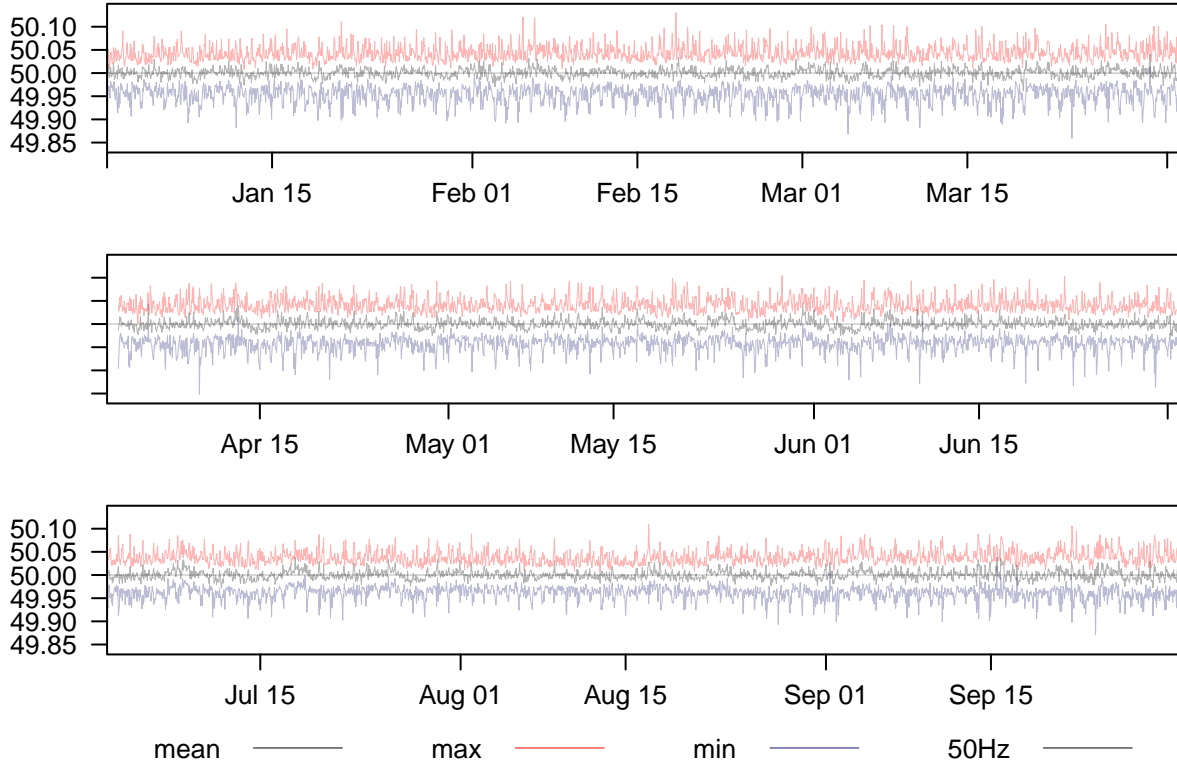
```
# Nordic:  
head(FN,15)
```

```
##                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
```

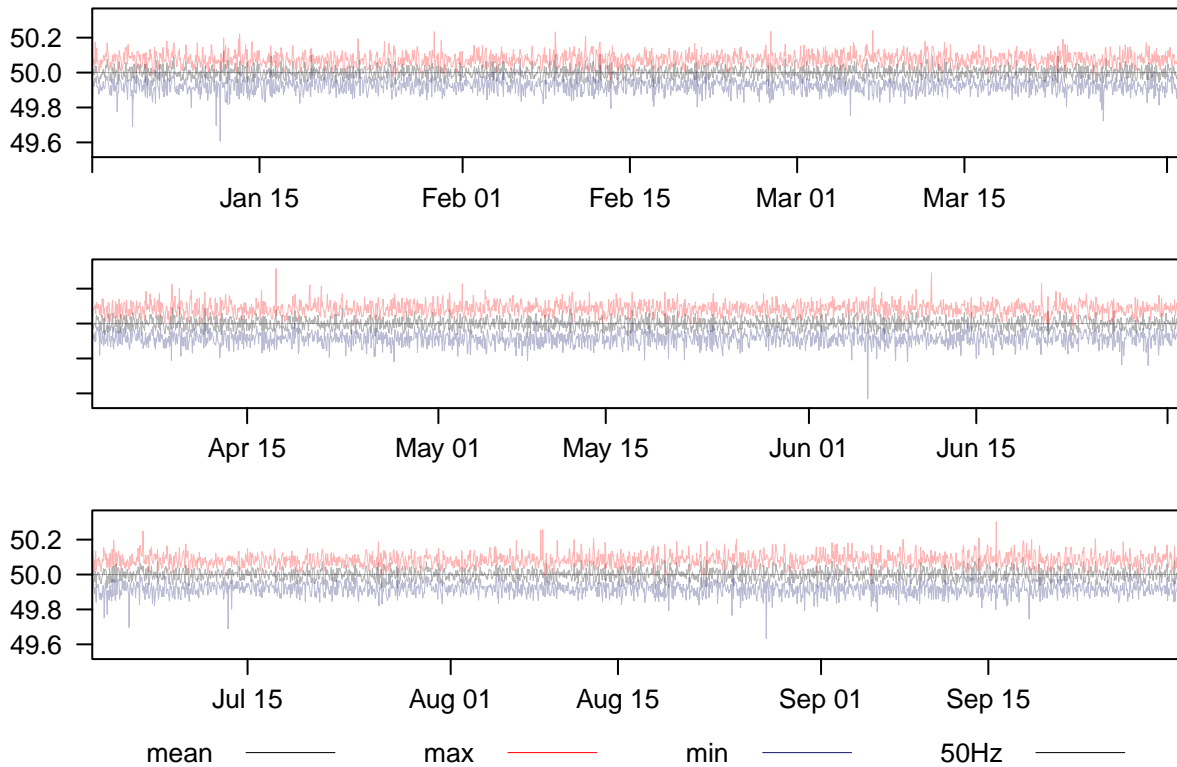
## Exploring the data

To get an overview, we first plot the average, maximum and minimum grid frequency for each hour of the test sample. From the two graphs it is evident that the Nordic grid shows significantly more extreme values of grid frequency. This might be because of the more frequent measurements in the Nordic grid data, but is also expected because smaller grids have less capacity to absorb power imbalances.

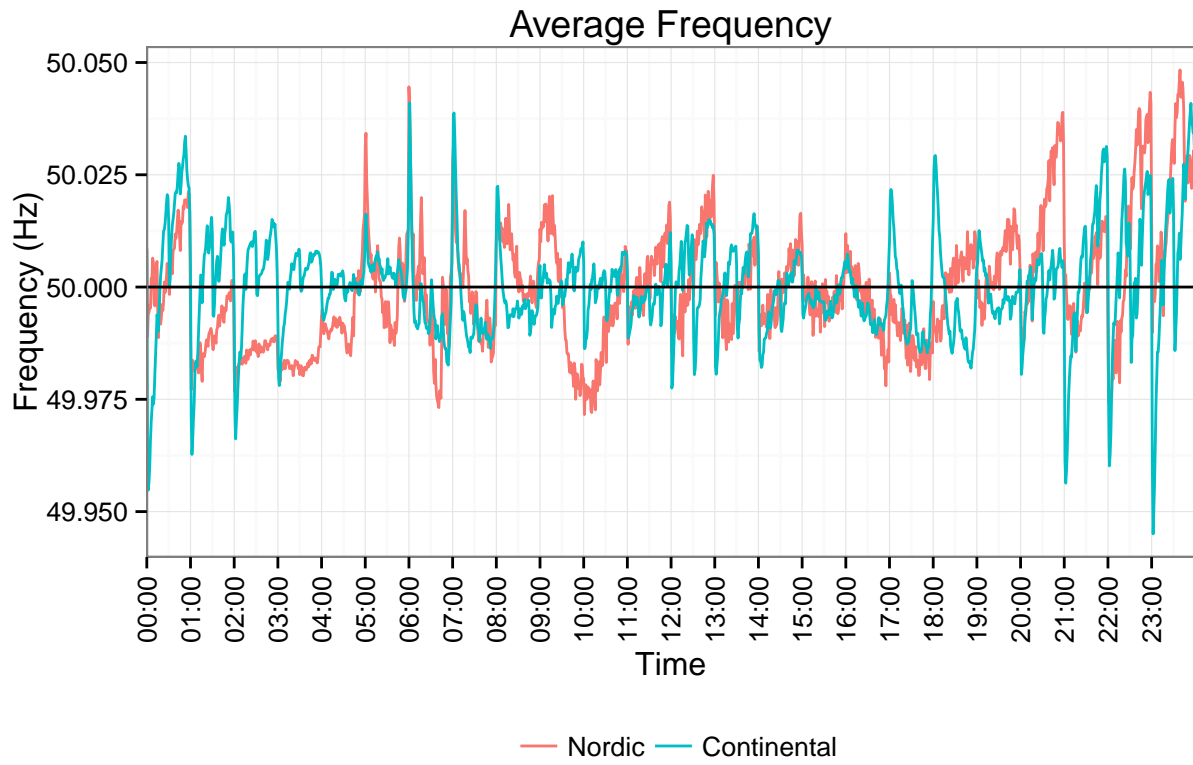
## Continent



## Nordic



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



## Frequency drift

