|  |  |
| --- | --- |
| To: |  |
| Copy: |  |
| From: |  |
| Date: | Kjeller, |
| Ref.: |  |

Explanation of scripts and plots related to pairing of EEA and EMEP sites.

A matching of EEA and EMEP sites has been done with the aim of detecting EBAS stations not being reported to EEA.

EEA data has been extracted by the R routine saqgetr whereas the EBAS data has been extracted at NILU, using shell script commands directly to the database.

The file ‘missing\_emep\_<comp>\_<year>.csv contains the EBAS sites that are not found directly in EEA. This regards very few sites (4 sites for o3 in 2021): DK0010G, DK002G (both on Greenland), and GR0002R, IT0018R (both in the Mediterranean).

In addition, a number of EMEP sites are obviously wrongly matched by the automatic pairing as seen from the distance and correlation coeff. Furthermore, for several sites there are errors in the monitoring heights and other site characteristics as well as discrepancies in the actual data.

The file ‘distances\_<year>\_<comp>.csv’ contains all the matched pair of EEA and EBAS sites including the correlation coefficients between the time series taking into account that there could be errors in the time zones, i.e. that there is a time lag of ± 4 hrs in either of the data.

Table 1 shows an extract from the file ‘distances\_2021\_o3.csv’ showing only the paired sites with a distance > 5 km. It is thus likely that all the EBAS O3 sites in Table 1 are lacking in the EEA data



Table 1. Extract from the file ‘distances\_2021\_o3.csv’, showing only the sites where the automatic pairing of sites found a distance > 5 km

In addition, although a correct pairing of sites is found, there could be differences in the actual data values, as mentioned. In the automatic script, we produced plots of matched pairs of sites if the case of corr. coeff < 0.999. This turned out to be valid for a fairly large amount of pairs and revealed numerous differences between the EEA and EBAS data. Figure 1 - Figure 4 below show the two types of plots that were produced for all the pairs with r < 0.999. A zip file containing plots of all these EEA/EMEP station pairs is attached.

Differences in the time series are seen to be due to all kinds of reasons; missing data in only one of the timeseries, individual datapoints that are flagged out in EBAS and not in EEA etc. Figure 1- Figure 4 show some examples.

A screenshot of a graph

Description automatically generated

Figure 1. hourly time series as extracted from EBAS (blue) and EEA (red). The EBAS data have been multiplied with 1.1 to be able to see the actual data series. The site code, year, and month are given in the header. ‘dt’ means the time lag added to the EEA data to get the highest correlation coefficient that reflects uncertainties in the assumed time zone.

A grid of graph paper

Description automatically generated with medium confidence

Figure 2. EEA time series of the same station as in Figure 1 where the colour indicates the difference to the matched EMEP site (EEA- EBAS).

A screenshot of a graph

Description automatically generated

Figure 3. Same as Figure 1 for GB0048.

A screenshot of a graph

Description automatically generated

Figure 4. Same as Figure 2 for GB0048.