1. The QA code is not written by me. I have only written one line (861)

(180 - abs(abs(values[1] - values[2]) - 180) > 100)

Which makes sure the two angles don’t have a difference in angle of larger than 100 degrees. It is written like this to account for changes such as 1 degree and 359 which are actually very close.

2. The weather data analysis mostly consists of bringing together the bits of data in various files from various sources. It is often presented in excel or csv files in different ways, so I have a set of functions for all of these general ways. These take the data in and output them in a standard format with most of the important weather metrics.

from\_mod\_collate()

from\_raw()

from\_camp2()

from\_cosmos\_collate()

from\_cosmos\_month\_excel()

from\_cosmos\_month\_csv3()

from\_aws()

from\_aws\_processed2()

from\_rep()

Lots of the different input data files are in the folder Data and these functions have been used in the script “collate\_data.R”. You can see there how they are used. There may be more ways in which the data is stored for other sites that I haven’t looked at but hopefully they will follow a similar pattern.

In collate\_data, the first section deals with collating the data for each site and turning it all into the same form. The convert\_daily function is used to take only the aggregate daily values for rain\_sum, temp\_max and temp\_min as these are the metrics in the modelled data. At the end of this section I take in the modelled data from MO. We have it here given to us at a single point on the 1km by 1km grid of point they calculate for. MO have calculated the nearest point and given us data up until the middle of 2019 here. The data for more recent than that is available online, but it is in grid form and for each month separately.

The second section creates a df\_comb which is a combination of collected site data with modelled data for that site. Use the sitecode to choose which site from the modelled data. At the moment, only one is uncommented but you can change this around to look at different sites. I then process that data by making a column for the collected data where it is present and replacing it with modelled where it is missing. And rolling averages for temperature. I also calculate gaps where the data is missing.

The third section visualises the data. I have split the timeseries into seasonal – a regular pattern which occurs each year, trend – the overall changes over long periods of time, and random – the fluctuations beyond that such as a particular wet month or dry week. These are all displayed with collected vs modelled and the time which there is no data for collected is covered with a blue box.

The fourth section tries to quantify it all. DTWDistance is used to give a value of how similar the two lines are. It is important to include accommodation for slight differences in the timing of events. i.e. if the modelled and collected predict a day of heavy rain but they predict subsequent days, the effect is the same. There are a bunch of other metrics such as days below freezing temp or days with over 10mm rain, but the data may need to be scaled for this to be valuable.

3. The weather data

It’s a bit of a mess, it comes from different places, those different places have changed how it is presented and then its been processed to different amounts. The sources are

Campbell’s first contract

Met office contract (AWS)

Campbell’s second contract

Burnham beeches local council

COSMOS (CEH data stations near or on our sites)

Met office representative stations

Met office modelled data

Often the old data has been collated to some degree so Campbell’s first contract I think is consistently collected into the same format. I have chosen sites which use all of these sources so hopefully the functions I have provided will load the data in every case and change it to a consistent format. I chose to use only metrics which are reasonably consistently collected but there are plenty more and these will need to be included for publication.