Climate hand over

**Files and Folders**

The home folder for the project is on the shared drive **N:\Exception\_ LTMN Working Data\01\_Raw\_Data\\_All\_sites\Climate\\_All\_sites\_collation\_up\_to\_2020**

In this folder there is:

* The R scripts for the QA checks (please use these ones, not the ones on TRIM, as I slightly updated them) - **Gross Error Checks R Code Jun 2020, Suspect Checks R Code Jun 2020, Tidying Up R Code Jun 2020.**
* Instructions for collating and QA-ing the climate data - **Climate Data Collation Instructions.**
* Data template for the source comparison of datasets – I was using this earlier in the project to run comparisons between on-site AWSs, MO rep data, and MO modelled data, and probably isn’t useful any more - **LTMN\_AWS\_Source\_Comparison\_template.**
* Data templates for the raw datasets and QA’d datasets. The QA scripts require a one-tab CSV file, and spit out a one-tab CSV file, thus the need for two templates: the input data comes from the raw\_data\_template, and the output data is pasted as a tab in the QA\_data\_template. Full instructions in the Climate Data Collation Instructions. **LTMN\_AWS\_raw\_data\_template\_2020, LTMN\_AWS\_QA\_data\_template\_2020.**
* A folder for each site, in which there is all the raw datasets that relate to that site (N.B. many datasets hold data for more than one site. Where this occurs, a version of that dataset appears in each of the relevant site folders). Where data has been collated into a source\_comparison\_template, or a raw\_data\_template, or a MO\_raw\_AWS\_data\_recompiled sheet, these are also saved in the relevant site folder.

Additionally, on TRIM there are some useful documents in **Research and evidence - LTMN Data - LTMN Data Project Support – Weather**, notably:

* **LTMN\_weather\_data\_tracker**, tracks what data exists for each site (from which source, over which time period, and where on TRIM – though the latter doesn’t matter so much now all the raw data is all copied into the site folders on the N drive). It also lists which variables are measured by each data source in the second tab.
* **Climate\_data\_progress\_tracker**, tracks which sites have reached which stage of processing.
* There is also my **LTMN Climate Planning Note 09012020** from early in the year, but plans have changed since writing this, so it’s no longer any good as a guide.

**Current status**

**Assembling data**

* We are still awaiting a reply from Malham Tarn for their AWS data since 02/12/17 (CC’d you in on the most recent chase-up to Frances Graham).
* We are also still waiting for data from Dark Peak. I’ll forward you the email exchange thus far, but it has all got rather confusing.
* We did discuss getting representative data for Downton Gorge, as we discontinued the AWS contract in 2018. We haven’t requested this yet.
* Since Oct 2019 we haven’t been getting representative data for Linton-on-Ouse (which stands in for the Lower Derwent Valley site). And since Mar 2020 we haven’t had representative data for Walney Island (which stands in for North Walney).
* All other data has been assembled and saved in the relevant site folders on the N drive.

**Collating Data**

* All sites with a non-COSMOS AWS have an initial collation in the raw\_data\_template (apart from Dark Peak and Malham Tarn). The sites with a green fill in column F on the Progress Tracker spreadsheet have a complete timeline. However, of these Bure, Derbyshire, and Finglandrigg (maybe – see below) have MO data from the original data we were sent, which needs swapping out for the Version 1/0 data we were sent recently.
* Dersingham, Fenn’s, Ingleborough, Kielderhead, the Lizard, Martin Down, Mottey Meadows, North Solent, Old Winchester Hill, Roudsea, Wyre Forest and East Dartmoor have all the Campbell’s data in place up to the present day, with a gap left for the MO data.
* Ainsdale, Downton, Lullington, Stiperstones and Thursley never had an MO contract on their AWSs, so they are complete and ready to QA.
* There is some doubt about the source of Finglandrigg’s data from 2013, as the raw data does not look like MO format, and the way it’s saved in many two-month-long files is odd.
* East Dartmoor may also need some more investigation, as there seem to be two streams of data that cover the same time span (see weather data tracker).
* We decided to use the MO representative AWS (Great Dun Fell 2) rather than the NE AWS at Cross Fell as it’s actually closer to the LTMN site. I have collated a raw dataset for this, BUT it is MO data, so is presumably subject to the same QA issue we’ve been battling.
* See columns G, H, I, J in the Progress Tracker for which datasets have had their wind and solar rad units checked and converted. I usually just make an extra tab with a pivot table and graph of those variables to see if they remain the same unit over time (you’ll see the extra tabs on some of the raw datasets that have been checked). Campbell’s record wind speed in m/s, so it all needs converting. MO wind data is usually (always?) in knots. Campbell’s record solar rad in w/m2, but MO supply it in KJ/hour/m2, so that needs converting.
* I have an inkling it might be worth running a final duplicate datetime row check on all the Version 1/0 MO datasets before you paste them into the main timeline, but this is *slow*. The best way, I think, is the one Ruth suggested: select the datetime column > Home tab, Conditional Formatting > Colour Scales > More Rules > Format only unique or duplicate values (choose duplicate) > click format > select a highlighting colour. Then you just have to scroll through looking for highlighted cells (which is the slow part)
* We never quite decided whether to treat the data that was salvaged from the gap between the first Campbell’s contract and the MO contract as Campbell’s or MO data. It’s currently listed as MO on the weather data tracker, but in many of the raw datasets I’ve left some of the SOURCE column blank. You’ll need to check for blanks in this column and fill them in before you run the QA.
* When all the above is sorted, and we have the new wind direction script from Christoph, all non-COSMOS AWS datasets should be ready to QA. Exact instructions for this are in the Collation Instructions document.

**Still to do**

Beyond the collation and QA stage, here are some steps I’ve thought of over the months which might be worth doing:

* The QA process doesn’t currently include a rainfall stuck check, probably because you can have long periods of 0 rainfall. It might be worth using the homepage graph to check manually for periods of time where the rain gauge seems to have stuck, and adding a QA flag/code to that effect.
* Likewise it might be worth adding a QA flag/code for net radiation, which fails where solar radiation fails.
* On that note, I haven’t looked into how net radiation is calculated from solar radiation – I hope we have the right data to do it.
* Do we want to indicate where there are gaps in the timeline somehow? Pivot tables can be used to highlight where a day has fewer than 24 records, but unless you create an extra column explicitly for months, you would not be able to check for missing days using them.

**What that leaves**

* Once all of the above is resolved, LTMN sites with a non-COSMOS AWS should be well on their way to having a single published AWS dataset. That leaves –
* 14 sites from which we will still be getting regular Campbell’s AWS data. This will have to be added to the site datasets as we get it, and then the whole thing republished as an updated version.
* 23 sites that don’t have a non-COSMOS AWS installed, including 15 sites which *never* had a non-COSMOS AWS installed (see weather data tracker), and which will be reliant on either representative or modelled data, should we want to publish a climate dataset for them.
* Representative data from MO. This is presumably subject to the same QA problems we’ve had with the AWS data. I believe the MO has already said they will not be able to supply the Version 1 or 0 for representative data. Cross Fell is currently using representative MO data.
* Modelled data from MO. This is daily, so cannot be QA’d with our current process. The QA process it’s undergone already is unknown, and it’s freely available on HADUK, so maybe we shouldn’t be publishing this anyway?
* COSMOS AWS data is also partly daily (it was hourly in the earlier years), so cannot be QA’d with our current process. Are we still getting data from these AWSs? I don’t think I’ve seen anything since November 2019. I collated Chobham Common as an experiment. Note they record their radiation in MJ, so it needs conversion.
* Burnham Beeches starts daily, but becomes hourly in 2002 (with some odd, irregular times to begin with), so the early part of the dataset can’t be QA’d.