**Workflow for data analysis**

**Use Master Script to source all the other scripts in correct order!**

**Logger**

* Read in data and create list
  + Separate scripts for separate datasets
* Start time correction
  + Set all loggers to same start time (same interval)
    - Interpolate data to 1 min intervals and
    - Reduce again to 10 min values
* QAQC
  + Use calculated offset (reference thermometer) to correct data
    - Use offset from test in the lab and subtract the offset from temperature data
    - Offset for logger 33 is missing -> check that
  + Tidy up data regarding spikes
    - Set a threshold for a rise in temperature that are regarded as spikes and therefore set to NA
      * For water: threshold of 2.5°C/10min, remove 2h of data (negative drop -5°C)
      * For air: threshold of 5°C/10min, remove 30mins of data
  + Outlier removal: remove all values outside 2,5 IQR
* Plots
  + plot the tidy data in pairs (water, settlement, vegetation)
  + plot overviews for all water logger/ vegetation logger/ sealed area logger
  + plot the water temp together with air temp to see warming/cooling effect
* Split data into day and night datasets (creates separate lists)
  + Two hours per day for dawn are removed
  + Works for every dataset
* Plot the day and night datasets and save to file
  + Add the sunrise and sunset as vertical lines to the plots
  + Add description and type of location to plots
* Time series analysis:
  + Decompose the time series (plot seasonality, trend and noise)
  + Test if time series is stationary
  + Test for normality and, subsequently do ttest or Wilcoxon significance test for green vs grey infrastructure
    - Calculate median for all grey and all green ts and do ttest (or wilcox test)
  + Do SiZer test to test significance between green and grey, Plot SiZer map
  + Seasonal ARIMA
* Statistics
  + Get mean, median and standard deviation for daily, nightly and 24h data
  + Plot statistics
  + Hypothesis 1:
    - Significance test for GI/SI
    - Linear model and correlation for site parameters
  + Hypothesis 2: Influence BI
    - Integrate potential cooling/warming
    - Merge Aasee WOL logger and compare for day/night data with VL logger Haus Kump
* Map
  + Isarithmic map with interpolation of points through inverse path distance weighing

Use extra script for data in 30 min intervals (September to November)

* start time correction for 30 min values
* Set threshold for spikes
  + Water: 5°C/30 min (remove two hours of data)
  + Air:10°C/30 min (remove 30 min of data)

**Supplementary weather data**

* Windspeed, -direction (FMO – DWD)
  + mean wind rose
  + relation wind speed to air temperature
* Temperature (reference temp) (FMO – DWD)
* Shortwave radiation (GeoDach)

**Netatmo data**

* Merge both the temperature data and the metadata
* Perform QAQC in 4 steps 🡪 reduced stations to 22
  + Level A: Filter out inconsistent metadata, timestamps and use only stations with 80% data per day/per month
  + Level B: Filter for monthly average and standard deviation of daily min temp (reg. reference temp.)
  + Level C: Filter out stations with radiative errors (and single values with rad errors)
  + Level D: Filter out outliers
* Plot
  + On map
  + On map with logger
  + Overview plot for all stations

**To Do – Data analysis:**

* Create (working heatmap) -> check out other possibilities
* Tidy up split script with new dawn/dusk values
* Plot stats (add legend to grid plot)
* Download cloudiness, atmospheric stability data?
* Do something with ULB water logger (fell dry)
* Map with points colored for green/grey/blue Logger

**To Do – Literature**

* Split weather effects in Citavi
* Read paper about wind effects (again)

**Ideas/considerations Aasee**

* Effect of shallowness
* Effect surface area
* Trees around Aasee #check how many NAs were added to data
* Fountains for fish -> suppl evaporation