Air quality measurements

Jenny Powell

OCEANS AND ATMOSPHERE
www.csiro.au

I am involved with projects that make atmospheric observations of air pollutants including particulate matter (PM2.5), mercury and persistent organic pollutants using research grade instruments and portable low cost sensors.

My Data School FOCUS Project

My project involved converting most of the workflow I perform in excel into R, which allows much larger datasets to be quickly processed and visualised. I wrote scripts in R to:

1. perform daily qa/qc on two mercury analyers located at Cape Grim;
3. analyse air quality and met data (trends, regressions, pollution roses);
2. develop a filter to remove noise from low cost sensors; and
4. apply principle components analysis and clustering techniques to AQ data

My Digital Toolbox

Excel for developing and testing algorithms on a subset of data before coding in R Tidyverse for data import into tibbles for visualisation, QA/QC and analysis Ggplot – histogram (normality), boxplot (percentiles and outliers) Ggplot - geom point, geom line and facet wrap to view air quality data by week or month, Rmarkdown and knitr - generating monthly html or word documents of air quality and meteorological data Python - used for processing data from various instruments Lubridate – formatting of date/time data, Openair – pollution rose to visualise origin of pollution events; Openair - time averaging to pad out missing dates and vector averaging of wind speed and direction

My time was spent …

Learning the features of Rstudio and how to code in R (beginner programmer). Importing data into a tibble and checking data qulity and types (chr, dbl, int, factor, etc). Learning how to use various libraries for visualisation and statistical analysis.

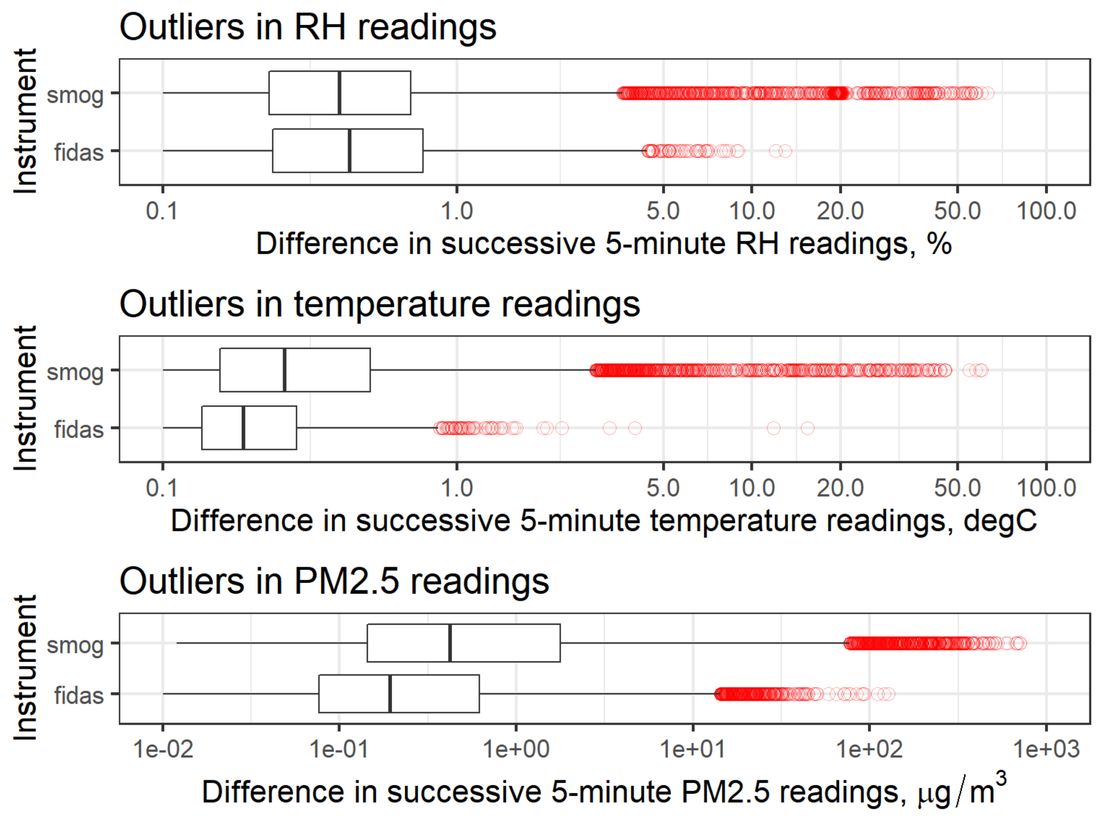
Next steps

1. Apply principle components analysis and clustering techniques to AQ data.,
2. Use r markdown document to generate monthly reports of mercury levels at Cape Grim.
3. Use Deming regression and/or Bland-Altman analysis on co-located instrument measurements to assess instrument performance.
4. Incorporate the low cost sensor filter algorithm to raw data prior to it being streamed on the web.

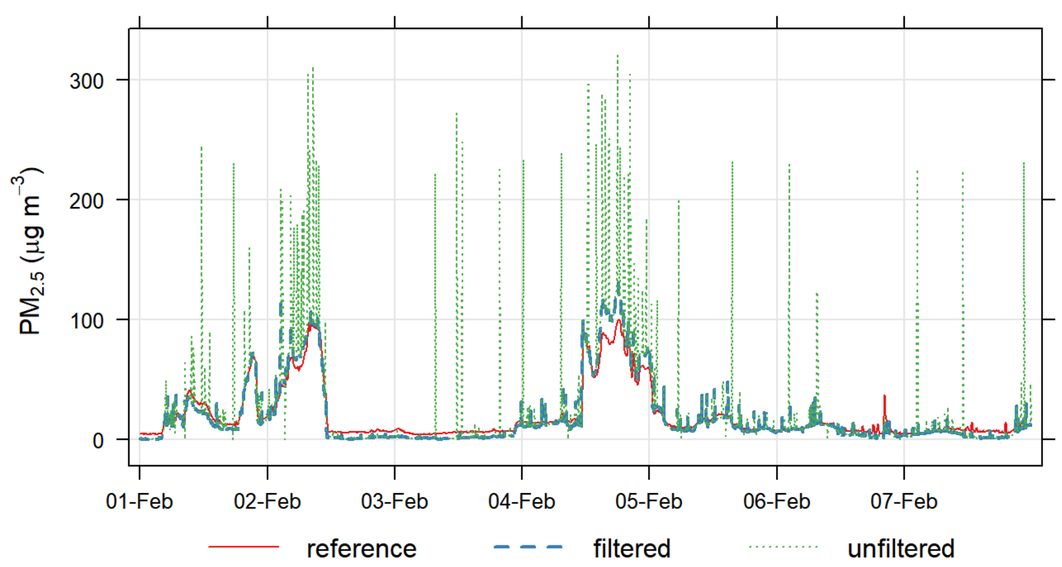
Table of criteria used to filter low sensor data determined from 95th percentiles

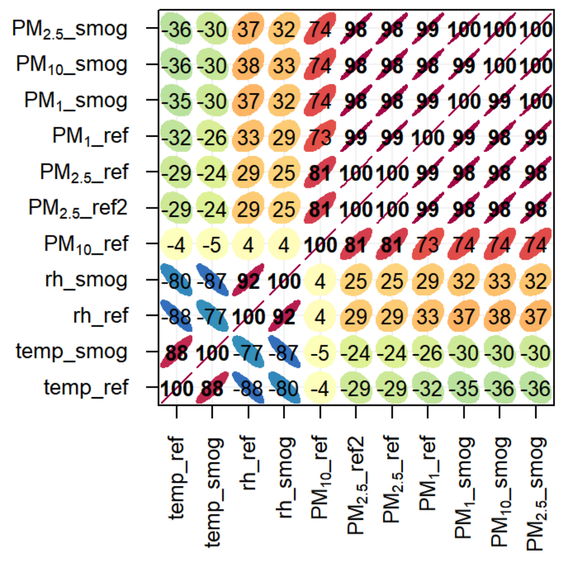
Percentile 5-minute successive difference	PM10, µg/m3	PM2.5, µg/m3	PM1, µg/m3	RH, %	Temperature, degC
5	0.0	0.0	0.0	0.0	0.0
25	0.2	0.1	0.1	0.1	0.1
50	0.8	0.5	0.3	0.3	0.1
75	4.5	3.0	1.3	0.7	0.3
95	61.2	33.0	19.8	20.0	10.3

Boxplots showing outliers between successive 5-minute readings



The plot below shows a comparison of 5-minute PM2.5 from a reference instrument versus unfiltered and filtered PM2.5 using the criteria shown in the table. Correlations between reference equipment and filtered low cost sensor data were strong, as shown in the correlation plot below.





MY DATA SCHOOL EXPERIENCE

Learning how to manipulate, visualise and analyse data in R were the core skills I gained from data school focus. The school provided useful on-line resources for learning R. The data school community and wider on-line community are great resources for helping to solve R coding issues.

Other skills learnt include using git and bit bucket to maintain version contol of data handling when working on group projects. Resources on research data planning, archiving and publishing data are useful for our team and I now have an understanding of how to upload data to the dap.

Meeting all the participants and facilitators in the first week was great to understand everyone's data challenges and to create a supportive learning network over duration of the course. Being able to participate in data school focus was a great opportunity to learn these very relevant skills - I highly recommend it.