DRAFT 19/01/2023 incorporating subgroup feedback

GUIDANCE FOR CALCULATING LAWA AIR QUALITY TRENDS

**Objective**

To provide consistency with Stats NZ method for annual of reporting 10-year air quality trends on LAWA. For more information on StatsNZ PM indicators see <https://www.stats.govt.nz/indicators/pm10-concentrations/>.

The 10-year trend period for LAWA reporting in 2023 is calendar years from 2013 to 2022 inclusive.

**Proposed method for councils**

Estimate 10-year air quality trends using the Theil-Sen method as implemented in the R Openair package. This non-parametric method works well with air quality data – as common features of air quality datasets don’t meet the underling statistical assumptions required for linear regression. For more information see <https://bookdown.org/david_carslaw/openair/theil-sen.html>. The Theil-Sen method estimates trends based monthly averages which may b calculated from hourly, daily, or higher time resolution data, as well as working directly with monthly averages. To streamline data handling it is recommended to use 1-hour averages as your starting point.

In future, StatsNZ will be updating their air quality trends methodology to align with IPCC trends classification, ie likely (~~66~~% 67 to 90%), very likely (90 to 100%) or indeterminate (<67%). LAWA already uses this framework for their water quality trends: <https://www.lawa.org.nz/learn/factsheets/calculating-water-quality-trends-in-rivers-and-lakes/>. In the interim we propose to use the existing StatsNZ approach.

**Steps for calculating LAWA trends**

1. Data cleaning and quality checks
2. Averaging and data completeness criteria
3. Prepare data file for analysis in Openair (R package)
4. Use Theil-Sen function in Openair (R package) to calculate 10-year trend
5. Fill out spreadsheet (MahiTahi) with direction of trend for each site for manual upload to LAWA
6. Data cleaning and quality checks

* Data cleaning and QA checks performed in accordance with NEMS Air quality: continuous particulate monitoring 2022.
* Remove all 1-hour averages <= -9 ug/m3 (AS3580.19 data processing standard). Note do not replace these negative data points with a zero value.

1. Averaging and data completeness criteria

* All base averaging periods (1-hour, 24-hr and annual) should have a least 75% data capture to be ‘complete’ (or ‘valid’ as per the NES definition) and therefore suitable for use in trends analysis.
* If you are only working with monthly data resolution the data completeness criteria is also 75% (ie, based on complete 24-hour averages, missing days/total number of complete days in month >= 75%).
* Round 24-hour averages to resolution of 0.1 ug/m3 (NEMS, 2022).
* Using StatsNZ criteria, a ten-year trend for a site and pollutant can be calculated where there are at least six years of complete data within the 10-year period (ie, six years with at least 75% valid 24-hour averages). Years with less than 75% data capture can still be used in the TheilSen 10-year trend calculations but there are some further data processing requirements discussed in 3.
* Note where there has been a change of monitoring instrument during the 10-year period that may introduce a step change or other non-environmental variation into the data record, apply a post data correction factor based on co-location data as described in the NEMS. It may be preferable to ‘backcast’ the data using the colocation correction factor so that the monitoring data going forward from the new instrument can be reported daily on LAWA and the long-term trend re-calculated based on ‘equivalence’ to the new method.
* Note it is not good practice to combine data record where a site has been re-located during the trend calculation period, unless monitoring at old and new sites for a year/winter/period of interest showed similar concentrations.

1. Prepare data file for analysis in Openair (R package)

* Wrangle your data into a file structure that works easily with Openair, a csv file is straightforward, but other file formats can be used (eg text or excel files or files downloaded from webservices) <https://bookdown.org/david_carslaw/openair/openair-package.html>
* You can use 1-hour data, 24-hour data or 1-month data. Note when using Openair the name of the date/time field must be date (note lower case).
* If you are exporting 1-hour averages directly from Hilltop then you can specify a gap tolerance of 15 minutes that will mean only complete hours (with 75% data capture are outputted). Alternatively, you can export data at finer resolution (eg 10 min) and use the time averaging threshold option (data.thresh=75) in Openair.
* The format of the values in the date field is critical and this can be complicated with issues such as time zones and daylight savings time. The convention is for air quality data to be recorded as NZST but there are also instances where NZDT is useful, for example when looking at diurnal profiles to match with activity data. See Attachment x for date/time formatting in R.
* Another issue to be aware of whether your data are timestamped at the beginning or end of the time period by your data acquisition system. Openair has some time averaging quirks that may or may not sit well with your data.  Hilltop convention is to timestamp a value at the end of the period, eg, a 24-hour average for the 31st December will have a time stamp 1 Jan 00:00:00.  Note Hill Laboratories are working on changing the data averaging convention for air quality data to avoid the issue of data being date/time stamped for the previous averaging period.
* TheilSen by default, calculates monthly averages.  The monthly averages won’t be entirely correct if your data timestamped for the end of the period, ie, it will take the last day of the previous month and leave out the last day of the month.   You may want to back shift your data (in this example by 1 day) to get the correct averaging. A straightforward solution is start with 1-hour averages exported from Hilltop and then subtract 3600 seconds when importing into Openair. This shifts the timestamp back by 1-hour, so all subsequent calculated daily and monthly averages are assigned to the right time window.

Import (mydata, correct.time=-3600)

* Note Openair uses date/time stamp from 00:00 to 23:00. Where your data is in a different format such as 01:00 to 24:00, use the correct.time -3600 as above to fix.
* Where you are including data for years that have less than 75% data capture, the deseason = TRUE argument will impute values for the missing months. However, where there is a continuous gap of more than 4 months, these imputed missing monthly values should be removed based on advice from StatsNZ who found unexpected values for these imputed months.

1. Use Theil-Sen function in Openair (R package) to calculate 10-year trend

* When using TheilSen, make sure deseason = TRUE so that values for any missing months are estimated. Set the data.thresh = 75 to ensure only months with 75% data capture are used (note base data needs to be NZST for this to work correctly). Sample code and output below.

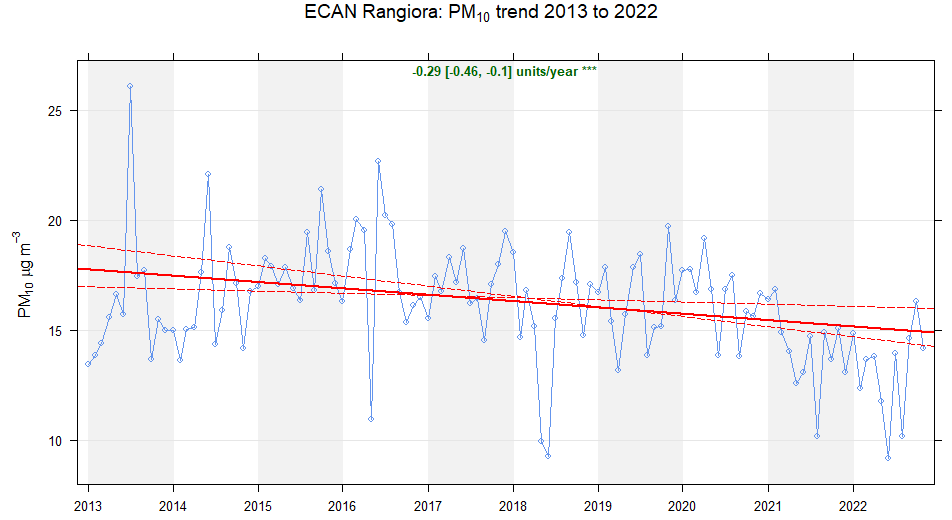
[TheilSen](https://davidcarslaw.github.io/openair/reference/TheilSen.html)(mydata, pollutant = "PM10",

ylab = "PM10 (ug/m3)",

data.thresh = 75 ,

deseason = TRUE,

date.format = "%Y")



* The trend is reported in pollutant units (ug/m3) per year with at least one star \* indicating that the trend is significantly (different from zero trend) at the 0.05 level (ie, 95% confidence level). Where there is a + the trend is significant at the ~~0.01~~ 0.1? level (ie, 90% CL) and is therefore doesn’t meet the LAWA criteria for a trend.
* The sign of the trend value shows the direction of the trend, ie a positive value is an increasing trend (air quality degrading) which is called a ‘downward’ trend in LAWA, whereas a negative value is a decreasing trend (air quality improving), called an ‘upward’ trend in LAWA.
* The values in square brackets next to the trend estimate is the 95% confidence interval in the slope of the trend. Where there is a trend, this interval does not include zero.

There are courses on Openair available through CASANZ (<https://www.casanz.org.au/trainingevents/course-offerings/>) or if you have any questions and need help contact:

xxxx

1. **Fill out Mahi Tahi**

**Attachment x for date/time formatting in R.**

When importing data into R it is important to know how the date-time is represented in your original data, especially in terms of time zone.

<https://bookdown.org/david_carslaw/openair/openair-package.html#reading-and-formatting-dates-and-times>

Most air quality and meteorological data around the world tends to be in GMT/UTC (Greenwich Mean Time/Universal Coordinated Time) or a fixed offset from GMT/UTC ie, not in local time where hours can be missing or duplicated. The national NEMS requires PM data to be recorded in New Zealand Standard Time (NZST) at the end of the averaging period. NZST = UTC + 12 hours.

Where your data is date/time stamped in NZDT it is relatively straight forward to import into Openair as it is assumed that the input data are in GMT (UTC) format and in particular there is no consideration of daylight saving time ie, where in the input data set an hour is missing in spring and duplicated in autumn.

If your data is in NZDT you can post-correct it to NZST by using the lubridate package.

For example:

datetime\_string <- "20230113T140000"

library(lubridate) *#* *load package*

*# format in NZST time zone*

t\_nzst <- ymd\_hms(datetime\_string, tz = "Etc/GMT-12")

[1] "2023-01-13 14:00:00 +12"

*# Convert to NZDT*

t\_nzdt <- with\_tz(t\_nzst, "NZ")

[1] "2023-01-13 15:00:00 NZDT"

?Date and time in separate columns