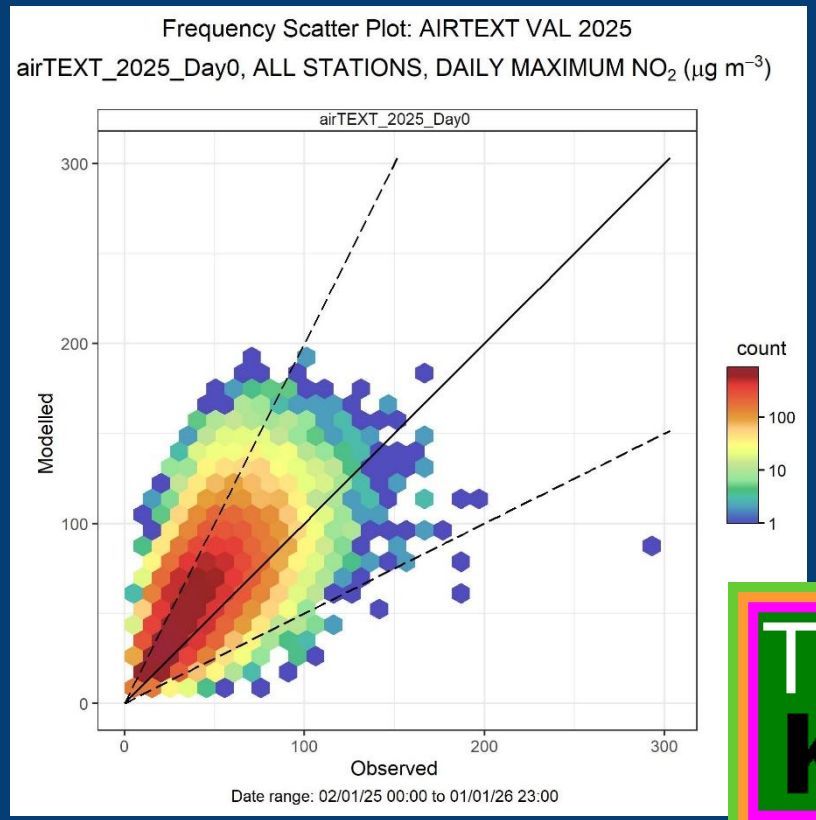
Introduction to the Model Evaluation Toolkit



Amy Stidworthy

Model Evaluation Workshop: Thresholds & Toolkit

UKHSA Training Centre

11th June 2026

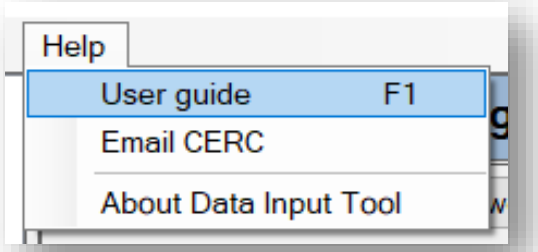
ADMLC

CERC

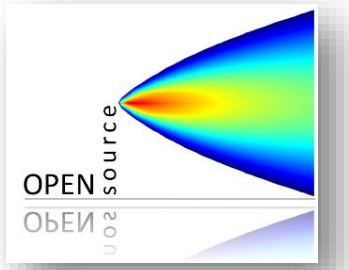
Overview

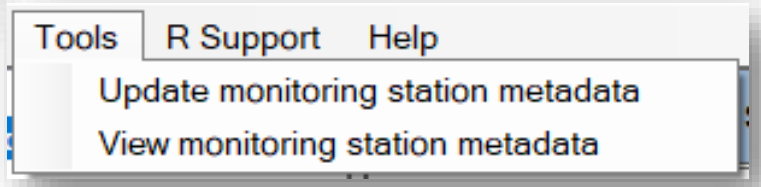
https://www.cerc.co.uk/environmental-software/model-evaluation-toolkit.html

- Developed by CERC – **free** to use
  - Developed under MAQS-Health and EU FP7 PASODOBLE projects
- Complete with comprehensive User Guide
- Easy evaluation of air quality models against observed data
- Offering wide range of graphs and statistics
- Including many openair functions
- User-friendly Windows interface
- Simple access to online air pollution measurement data









Carslaw, D.C. and K. Ropkins, (2012). **openair — an R package for air quality data analysis**. Environmental Modelling & Software. Volume 27-28, pp. 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008

History

Original motivation: needed a tool to facilitate the assessment of **regional** and **local air quality forecast** performance, building on best practice and pre-existing tools (EU FP7 PASODOBLE, 2010-2013)

PASODOBLE literature review identified the **openair** package of air pollution analysis tools and graphics, which is why the Toolkit is written in R

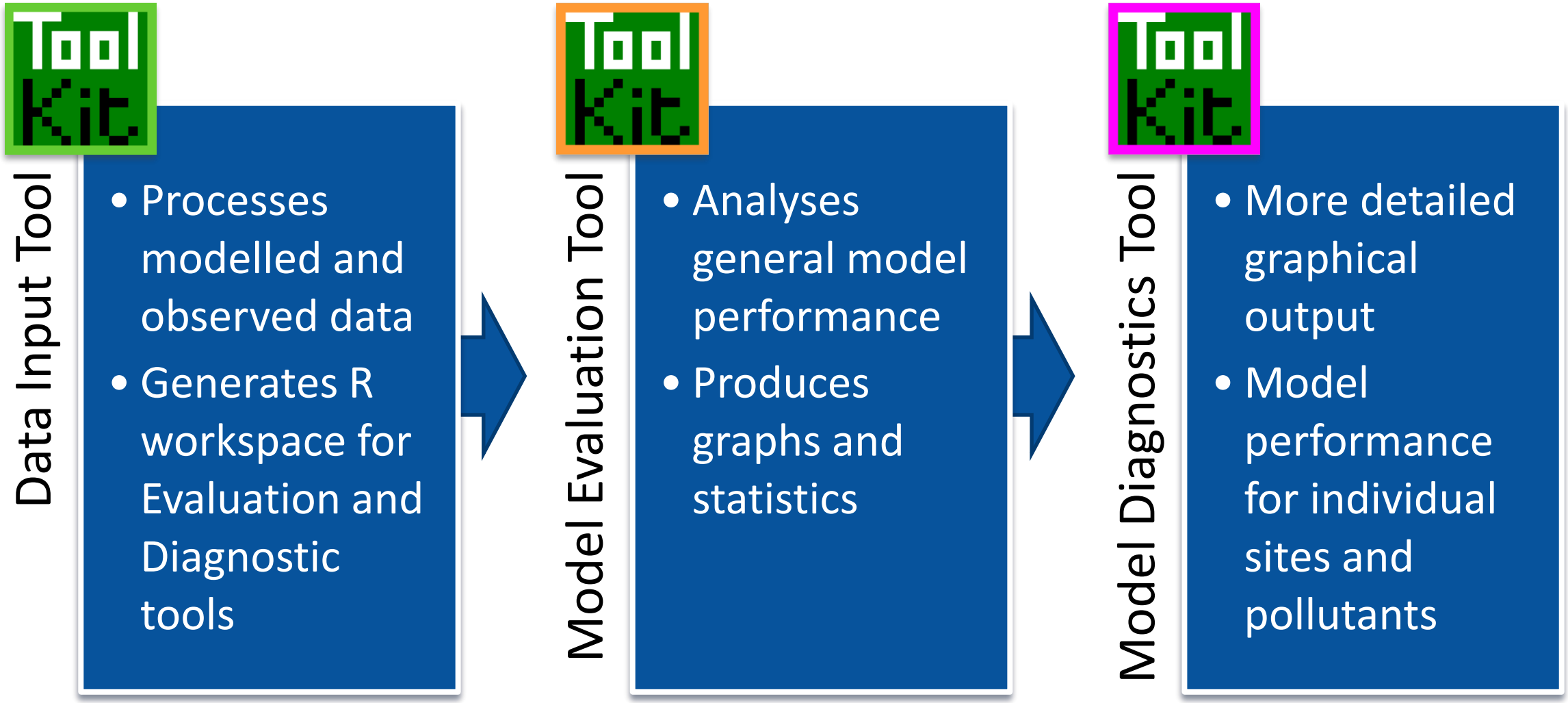
MAQS-Health project (2019-2022): significantly expanded range of supported modelled data formats, and range of measurement networks available

Used in the UK modelling inter-comparison exercise that CERC conducted for Defra (2021-2022)

ADMLC Model Evaluation Study (2023-2026): Added more statistical output

Stidworthy, A., Carruthers, D., Stocker, J., Balis, D., Katragkou, E. and Kukkonen, J., (2013). Myair toolkit for model evaluation. 15th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, Madrid, Spain. Available at https://www.harmo.org/Conferences/Proceedings/_Madrid/publishedSections/H15-186.pdf

Model Evaluation Toolkit



Model Evaluation Toolkit – updates for ADMLC Study



Model Evaluation Tool

- Analyses general model performance
- Produces graphs and statistics

- The **Toolkit** was used to generate the results for the ADMLC Model Evaluation Study
- The **Model Evaluation Tool** was updated for the project to output additional metrics identified in the literature review, for testing
- A new version of the **Toolkit** has been released (v5.4) to coincide with this workshop, which includes the metrics recommended in the Study report

Toolkit input: Modelled data

Range of compatible **modelled data formats**:

ADMS, MAQS, CMAQ, WRF-CHEM, EMEP, CAMx, CHIMERE, CAMS

Flexible **averaging time**:

n-hourly, daily, weekly, monthly, yearly

Flexible **averaging statistic**:

Mean, maximum, rolling mean

Supports **species mapping**

(e.g. combining modelled NO and NO_2 to compare with NO_x measurements)

Accounts for **time zone differences** between modelled and observed data

Provide as a **single file** or **multiple files** with a template path

Multiple modelled datasets can be assessed

Toolkit input: Observed data

Two options:

1. **Provide file(s)** of observed data in plain text format
   - Flexible averaging time: n-hourly / daily / weekly / monthly / yearly data
   - Flexible averaging statistic: mean, maximum, rolling mean

and/or

2. **Automatically access** hourly air pollution observations from 7 monitoring networks

Network	Coverage	Description
AURN	UK	UK Automatic Urban and Rural Network
SAQN	Scotland	Air Quality in Scotland network
WAQN	Wales	Welsh Air Quality Network
NI	Northern Ireland	Northern Ireland Air Network
AQE	England	Air Quality England network
KCL	UK	Networks operated by the Environmental Research Group (ERG) at Imperial College London (formerly at Kings College London), including the Londonair network
EUROPE	Europe	EEA's database of European monitoring sites

Model Evaluation Tool analysis settings

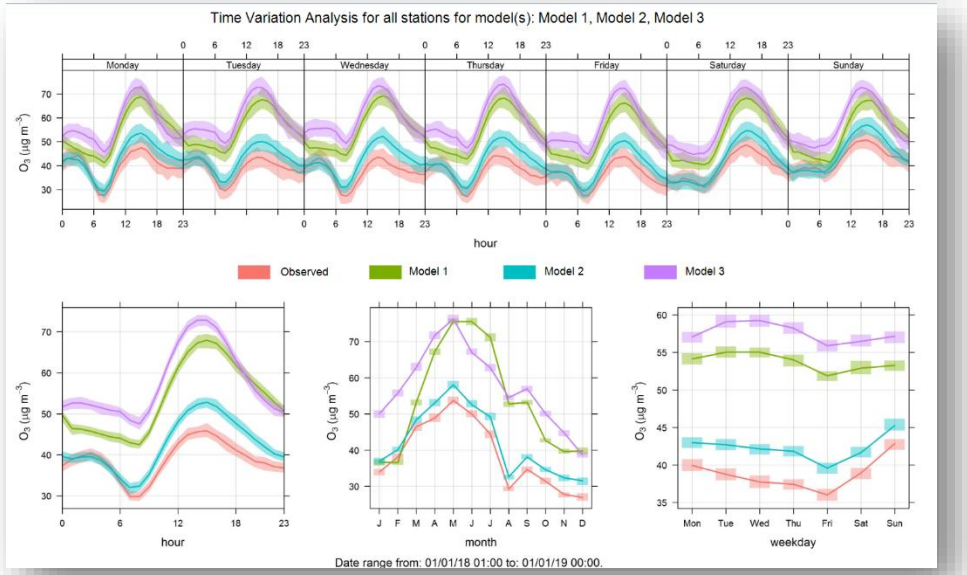
Select data to analyse

- Select data by modelled dataset, monitoring network, station type, station, and/or pollutant
- Add an 'ensemble' modelled dataset if required – the median of the modelled datasets selected
- Add composite pollutants if required, e.g. NO_2/NO_X, NO_2+O_3 etc – can be very powerful for diagnosing the cause of model error
- Select a date range if required

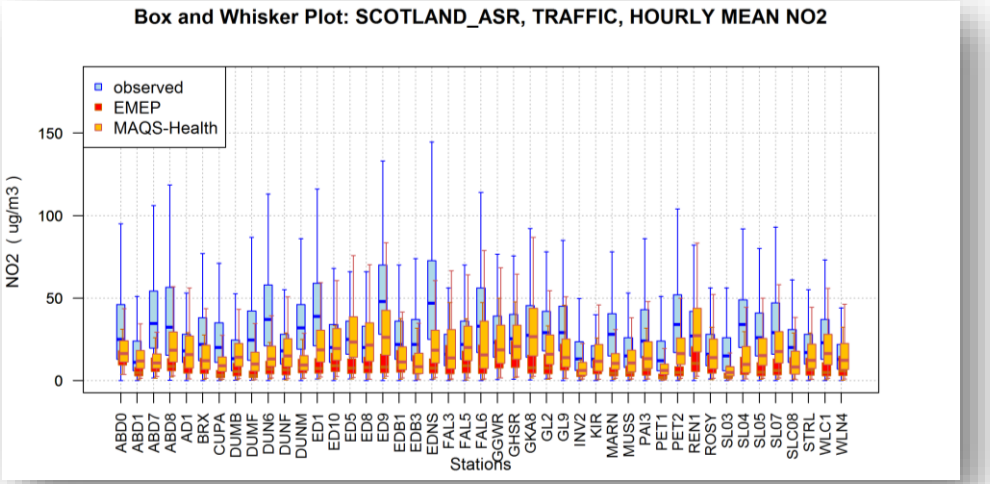
Grouping and filtering of results

- Powerful feature
- Group results by station, station type, model or pollutant
- Filter results by pollutant, model and/or station type
- Statistical output produced according to grouping/filtering setting
- Some graph types not compatible with some grouping/filtering options

Example: grouped by model, filtered by pollutant

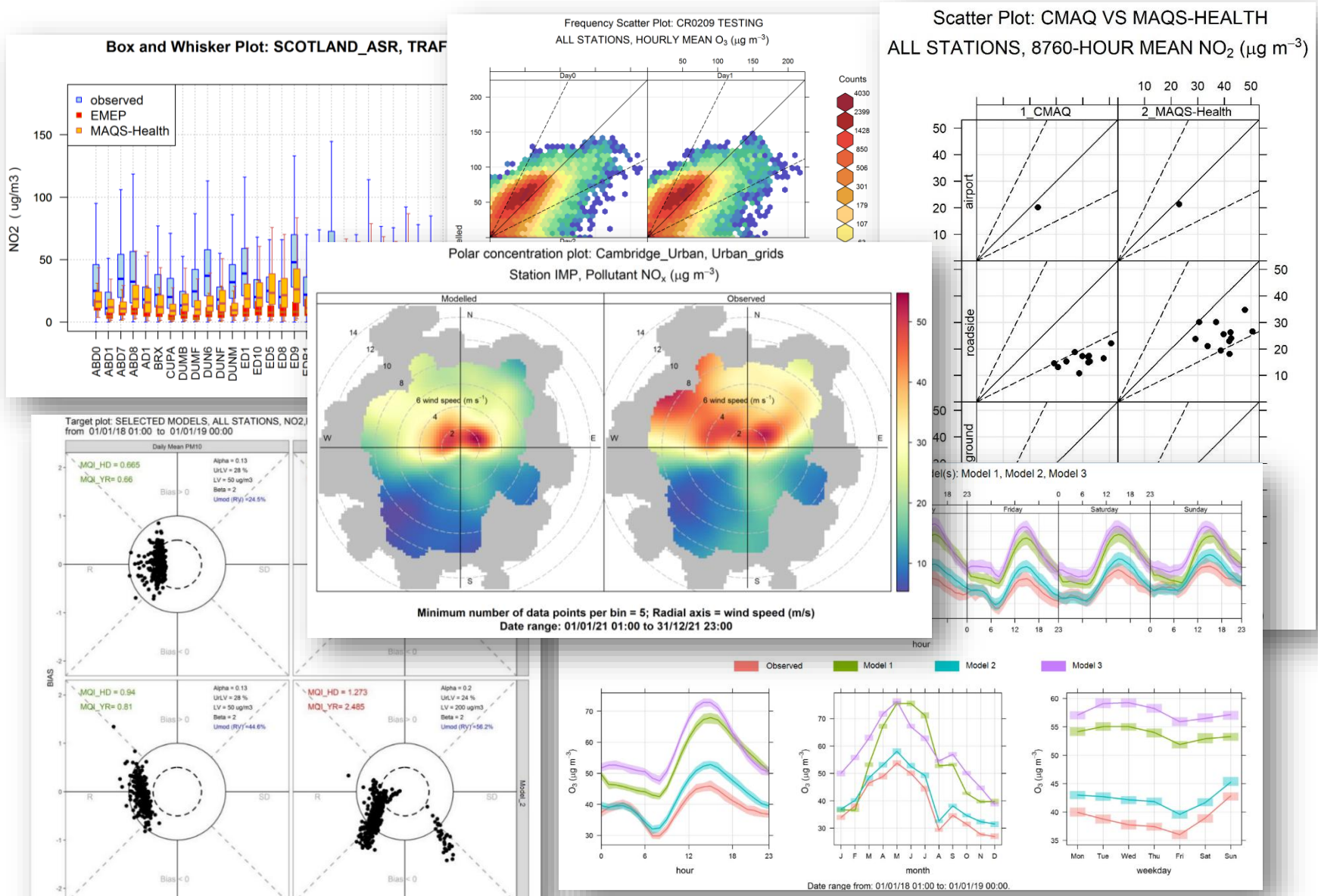


Example: grouped by station, filtered by model, pollutant and station type



Toolkit output: Graphs

- Box plot
- Scatter plot (conventional, frequency and statistical)
- Time variation plot
- FAIRMODE Target plot, accounts for measurement uncertainty
- Polar plot (requires met data)
- Quantile-Quantile plot
- Time plot, for specific time period
- Forecast index accuracy
- Forecast performance metrics
- Odds ratio skill score



ADMLC CERC

Toolkit output: Statistics

- CSV output

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	station	pollutant	model	output.avg.time	num.valid.	obs.mean	mod.mean	SDO	SDM	MB	NMSE	R	R2	Fac2	Fb
2	all	no2	Urban	hourly mean	8172	16.28	14.57	11.61	10.26	-1.71	0.29	0.73	0.53	0.82	-0.11
3	all	nox	Urban	hourly mean	8172	28.16	26.12	25.60	22.31	-2.04	0.40	0.76	0.57	0.80	-0.08

- User defined grouping and filtering

	A	B	C	D	E	F
1	station	pollutant	model	type	output.avg.time	num.valid
2	abd	no2	MAQS	urban background	annual mean	1
3	abd0	no2	MAQS	roadside	annual mean	1
4	abd1	no2	MAQS	roadside	annual mean	1
5	abd3	no2	MAQS	roadside	annual mean	1
6	abd8	no2				

	A	B	C	D	E
1	type	pollutant	model	output.avg.time	num.valid
2	industrial	pm2.5	MAQS	annual mean	14
3	roadside	pm2.5	MAQS	annual mean	64
4	rural background	pm2.5	MAQS	annual mean	5
5	urban background	pm2.5	MAQS	annual mean	42

- Also including:
  - Openair modStats: COE, IOA
  - Ozone: AOT40 (crops/forests)
  - FAIRMODE Target plot: MQI, RMSUM
  - Spatial: AFN, AFP, FMS, 1DMOE, 2DMOE
  - Forecasting: POD, FAR, AQ Index

For full details of available stats see the **Toolkit User Guide**

Link with ADMLC report recommendations

https://www.cerc.co.uk/environmental-software/model-evaluation-toolkit.html

- The Toolkit can be used to:
  - carry out a ‘sanity check’ on model outputs: scatter plots can be useful for showing if model outputs have the right magnitude, and box and whisker plots are useful for assessing data ranges
  - produce a wide range of dimensional and non-dimensional statistical metrics, and choose the most appropriate metrics for your application
  - diagnose sources of error using a range of graphical outputs (e.g. time variation plots, polar plots) and MSE components (systematic and un-systematic)
  - carry out consistent inter-comparisons between models or between different modelled datasets from the same model



The report recommended making the version of the Model Evaluation Toolkit used for the ADMLC project publicly available: this is **version 5.4**, released **June 2026**.

Any questions?