The Journal of Open Source Software

AQEval: R code for the analysis of discrete change in Air Quality time-series

Karl Ropkins [1], **Anthony Walker** [2], **Ian Philips** [1], **Christopher E. Rushton** [1], **Tony Clark** [2], **and James E. Tate** [1]

1 Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK **2** Joint Air Quality Unit, Department for Transport & Department for Environment, Food and Rural Affairs, Marsham Street London, SW1P 4DF, UK

DOI: 10.21105/joss.08839

Software

- Review
- Repository
- Archive

Editor: Gabriele Bozzola

Reviewers:

- @eldemet
- @meenakshi-kushwaha

Submitted: 02 June 2025
Published: 31 January 2026

Summary

`AQEval` (Air Quality Evaluation) is an `R` package for the routine investigation of discrete changes in air quality time-series. The main functions, `quantBreakPoints` and `quantBreakSegments`, use a three-step method to find possible 'points-of-change', test these and quantify the most likely 'points-of-change' or 'regions' about them. Other key functions build on these to provide a workflow to measure smaller changes and/or changes in more complex environments.

Statement of Need

Authorities responsible for air quality management are typically required to implement and evaluate air quality interventions they adopt (Bradley et al., 2019). These interventions are often costly, disruptive and unpopular (Glazener & Khreis, 2019), and the inherent variability monitoring data hinders impact assessments (Grange & Carslaw, 2019; Jones et al., 2012; Kelly et al., 2011; Pearce et al., 2011). Various methods have been developed to investigate discrete changes in a wide range of time-series (see e.g. Reeves et al., 2007; Truong et al., 2020) and several `R` (R Core Team & others, 2025) packages have been developed for their use, e.g. `bcp` (Erdman & Emerson, 2008), `changepoint` (Killick & Eckley, 2014), `segmented` (Muggeo & others, 2008), and `strucchange` (Zeileis et al., 2002). Some have even been applied to air quality time-series (see e.g. Carslaw et al. (2006), Carslaw & Carslaw (2007)). However, many of those tasked with air quality policy assessment, although highly skilled in a wide range of monitoring activities, are unlikely to be able to dedicate sufficient time and resources to the development of in-house expertise in such specialist analyses. `AQEval` was developed to address this skill gap and published as an open research tool. It aligns the inputs and outputs of a number of statistical methods to provide a one-package option for anyone interested in using `R` to routinely investigate change in air quality data.

Sources and Related Outputs

- The `AQEval` project website is at https://karlropkins.github.io/AQEval/;
- Release version on CRAN, the developers' version and code on GitHub, where issues or change requests can also be posted;
- An extended version of this paper, including code for Figures 2 and 3, and;
- Other related outputs include Ropkins & Tate (2021), Ropkins et al (2022), and work in UK Clear Air Zone (CAZ) impact assessment reports Report archive.

Analytical Rationale

The main steps of the Break-Point/Segment (BP/S) analysis are:

1. Breaks-points are determined using the `strucchange` methods of Zeileis and colleagues (Zeileis et al., 2002, 2003). Here, a rolling-window approach is applied: a first subset (time-series window TW_0 in Figure 1a) is selected and linear regression modelled; then the window is advanced (TW_1 in Figure 1a) and a second model built, and so on through the time-series; and, finally, likely points-of-change are assigned by comparing the F-Stat scores of sequential models.
2. In addition to the standard Bayesian Information Criterion (BIC) testing used by `strucchange`, `AQEval` also checks that all individual break-points are statistically valid ($p<0.05$), and down-scores less likely combinations.
3. Finally, the `segmented` methods of Muggeo and colleagues (Muggeo, 2003, 2017; Muggeo & others, 2008) are used to determine regions-of-change about break-points. Here, the confidence intervals for the selected break-points are used as start points, and segments assigned by random walk testing about these points as shown in Figure 1b.

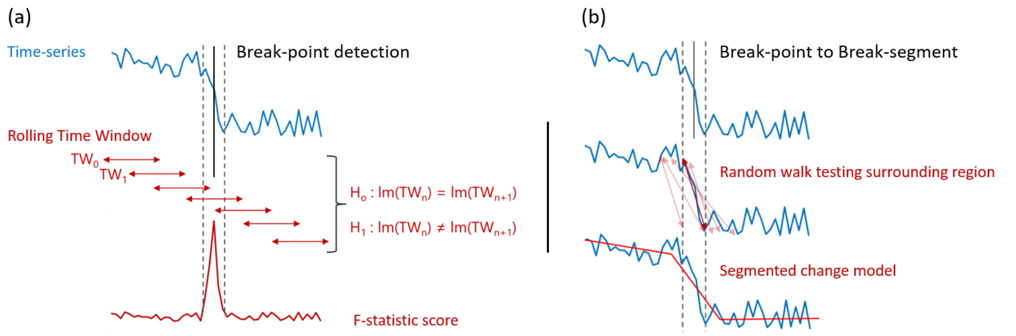

Figure 1: Break-point/segment scheme: (a) break-point, and (b) segment modelling about break-points.

Figure 2 shows the break-point/segment analysis of an NO_2 time-series from a roadside site where a change event (*ca.* 25 $\mu g.m^{-3}$; 31%) is detected between 2003-01-11 and 2003-02-19.

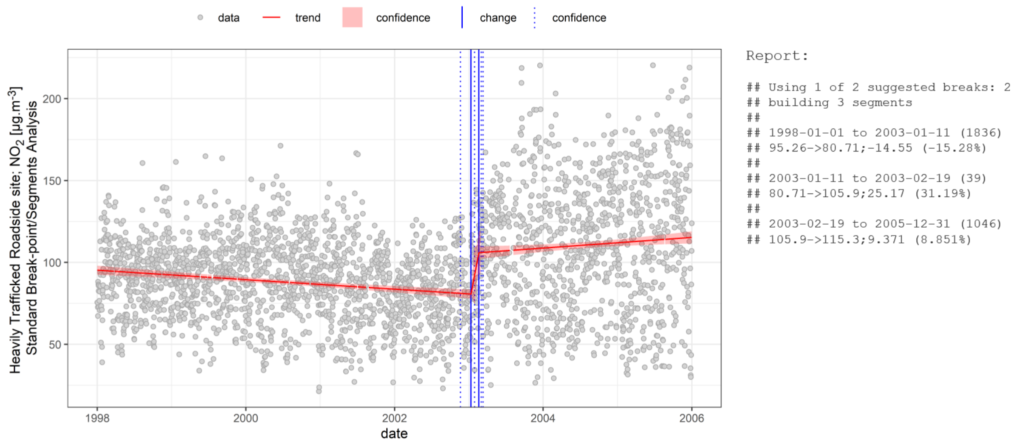

Figure 2: Standard AQEval break-point/segment analysis (graphical output and report) of an NO_2 1998-2005 time-series from a heavily trafficked roadside in the UK.

In some cases changes are small or local air inputs are complex, and time-series may require additional pre-processing to successfully isolate obscured change-events. For these, `AQEval` uses Generalized Additive Models (GAMs) (using `mgcv` methods, Wood, 2017, 2025) to subtract associated variance, by default:

$$[pollutant] = s_1(day\ of\ year) + s_2(hour\ of\ day) + te_1(wind\ speed, wind\ direction)$$

$$[pollutant]_{isolated} = ([pollutant] - [pollutant]_{predicted}) + mean(pollutant)$$

where the investigated pollutant concentration, *[pollutant]*, is modelled as a function of *day of year*, *hour of day* and *wind speed* and *direction* using a combination of spline (s_1 and s_2) and tensor (te_1) fit-terms, and the unmodelled component, $[pollutant]_{isolated}$, is estimated as the mean-centred residual of this model.

Figure 3a shows the break-point analysis of NO_2 from a nearby but less heavily trafficked site where seasonality dominates the time-series, and Figure 3b shows the smaller (*ca.* 6.6 μg.m^{-3}; 13%) underlying change-event observed at a similar time to the large change observed at the more heavily trafficked site in Figure 2 using signal isolation and then break-point/segment analysis (2002-09-09 to 2002-12-21 compared with 2003-01-11 and 2003-02-19).

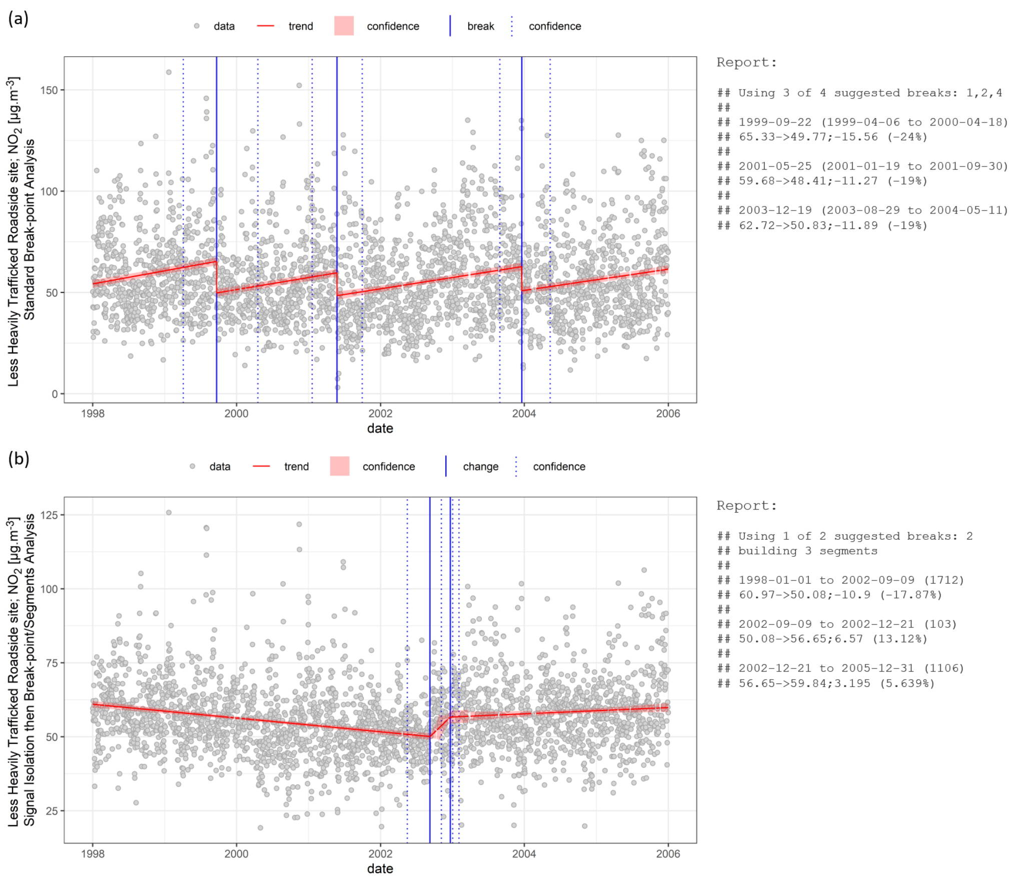

Figure 3: AQEval analysis of NO_2 1998-2005 time-series from a roadside site where: (a) standard (ambient air) break-point analysis exhibits a near-regular distribution of breaks typical of a site dominated by seasonal factors; and, (b) an underlying change-event is revealed using signal isolation and then break-point/segment analysis.

This default correction can also be modified to include other potential confounders, e.g. other frequency terms (e.g. day of week and/or week of year), background contributions (as local variance associated with trends at near-by site not affected by the investigated change), or proxies for other local contributors (e.g. other meteorological parameters like air temperature, markers for other sources, etc).

Data Requirements and Main Functions

As many air quality professionals already use the `R` package `openair` (Carslaw & Ropkins, 2012; Ropkins & Carslaw, 2012) for more conventional analysis and data visualisation, `AQEval` uses `openair` coding syntax and data structures. Most importantly, time-series to be analysed should be provide `data.frame` (or similar), also containing a paired column of `POSIX*` class time-stamps named `date` similar to packaged example data, `aq.data`.

Main Functions:

```
# to quantify Break-Points (as in Figure 3a)
quantBreakPoints(
  data,                 # data source, typically an openair-friendly data.frame
  pollutant,            # name of column containing the time-series to analysed
  h,                    # rolling window size, as proportion of time-series length
  ...                   # other arguments
)

# to quantify Break-Segments (as in Figure 2 and 3b)
quantBreakSegments(
  data,                 # as above but fits trends to regions about break-points
  pollutant,            # (see section 3 in Analytical Rationale above)
  h,
  ...
)

# For signal isolation (as used for Figure 3b)
isolateContribution(
  data,
  pollutant,
  ...                   # other arguments allow user to modify the isolation method
)                       # (see package documentation for further details)

# findBreakPoints and testBreakPoints can also be used to fine-tune BP/S models
```

Acknowledgements

Initial `AQEval` development was funded by the UK Department for Environment, Food and Rural Affairs (Defra). The authors gratefully acknowledge contributions from colleagues at University of Leeds, Defra and IPSOS Mori, and internal review by the Defra/Department for Transport Joint Air Quality Unit (JAQU) Technical Independent Review Panel (T-IRP). The authors also gratefully acknowledge the work of the R core team and collaborators in developing and maintaining the open-source statistical language R and associated packages http://www.r-project.org/, and the help, input and advice of editors and reviewers at JOSS.

The views and opinions expressed herein by the authors are their own and do not necessarily reflect those of UK Government or any agency thereof.

References

Bradley, N., Dobney, A., Exley, K., Aldridge, J., Craswell, A., Dimitroulopoulou, S., Hodgson, G., Izon-Cooper, L., Mitchem, L., Mitsakou, C., & others. (2019). *Review of interventions to improve outdoor air quality and public health*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/937341/Principal_interventions_for_local_authorities-air_quality_public_health.pdf

Carslaw, D. C., & Carslaw, N. (2007). Detecting and characterising small changes in urban nitrogen dioxide concentrations. *Atmospheric Environment*, *41*(22), 4723–4733. https://doi.org/10.1016/j.atmosenv.2007.03.034

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

Carslaw, D. C., Ropkins, K., & Bell, M. C. (2006). Change-point detection of gaseous and particulate traffic-related pollutants at a roadside location. *Environmental Science & Technology*, *40*(22), 6912–6918. https://doi.org/10.1016/j.scitotenv.2018.10.344

Erdman, C., & Emerson, J. W. (2008). Bcp: An r package for performing a bayesian analysis of change point problems. *Journal of Statistical Software*, *23*, 1–13. https://doi.org/10.18637/jss.v023.i03

Glazener, A., & Khreis, H. (2019). Transforming our cities: Best practices towards clean air and active transportation. *Current Environmental Health Reports*, *6*, 22–37. https://doi.org/10.1007/s40572-019-0228-1

Grange, S. K., & Carslaw, D. C. (2019). Using meteorological normalisation to detect interventions in air quality time series. *Science of the Total Environment*, *653*, 578–588. https://doi.org/10.1016/j.scitotenv.2018.10.344

Jones, A. M., Harrison, R. M., Barratt, B., & Fuller, G. (2012). A large reduction in airborne particle number concentrations at the time of the introduction of "sulphur free" diesel and the london low emission zone. *Atmospheric Environment*, *50*, 129–138. https://doi.org/10.1016/j.atmosenv.2011.12.050

Kelly, F., Anderson, H. R., Armstrong, B., Atkinson, R., Barratt, B., Beevers, S., Derwent, D., Green, D., Mudway, I., Wilkinson, P., & others. (2011). The impact of the congestion charging scheme on air quality in london. Part 1. Emissions modeling and analysis of air pollution measurements. *Research Report (Health Effects Institute)*, *155*, 5–71. https://europepmc.org/article/MED/21830496

Killick, R., & Eckley, I. (2014). Changepoint: An r package for changepoint analysis. *Journal of Statistical Software*, *58*, 1–19. https://doi.org/10.18637/jss.v058.i03

Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, *22*(19), 3055–3071. https://doi.org/10.1002/sim.1545

Muggeo, V. M. (2017). Interval estimation for the breakpoint in segmented regression: A smoothed score-based approach. *Australian & New Zealand Journal of Statistics*, *59*(3), 311–322. https://doi.org/10.1111/anzs.12200

Muggeo, V. M., & others. (2008). Segmented: An r package to fit regression models with broken-line relationships. *R News*, *8*(1), 20–25. https://journal.r-project.org/articles/RN-2008-004/RN-2008-004.pdf

Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J., & Tapper, N. J. (2011). Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmospheric Environment*, *45*(6), 1328–1336. https://doi.org/10.1016/j.atmosenv.2010.11.051

R Core Team, R., & others. (2025). *R: A language and environment for statistical computing*. https://doi.org/10.32614/R.manuals

Reeves, J., Chen, J., Wang, X. L., Lund, R., & Lu, Q. Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, *46*(6), 900–915. https://doi.org/10.32614/RJ-2012-003

Ropkins, K., & Carslaw, D. C. (2012). Openair: Data analysis tools for the air quality community. *The R Journal*, *4*(1), 20–29. https://doi.org/10.32614/RJ-2012-003

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, *167*, 107299. https://doi.org/10.1016/j.sigpro.2019.107299

Wood, S. N. (2017). *Generalized additive models: An introduction with r*. chapman; hall/CRC. https://doi.org/10.1201/9781315370279

Wood, S. N. (2025). *Mgcv: Mixed GAM computation vehicle with automatic smoothness estimation. R package version 1.9-3*. https://doi.org/10.32614/CRAN.package.mgcv

Zeileis, A., Kleiber, C., Krämer, W., & Hornik, K. (2003). Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis*, *44*(1-2), 109–123. https://doi.org/10.1016/S0167-9473(03)00030-6

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2002). Strucchange: An r package for testing for structural change in linear regression models. *Journal of Statistical Software*, *7*, 1–38. https://doi.org/10.18637/jss.v007.i02