R as a statistical engine for a water quality trend analysis web-service

Paul Rustomji[1,*], Brent Henderson[2], Katie Mills[1], Bai Qifeng[1], Peter Fitch[1]

1. CSIRO Land and Water, GPO Box 1666, Canberra, A.C.T., 2601, AUSTRALIA
2. CSIRO Mathematical and Information Sciences, ANU Campus, Acton, A.C.T., 2601, AUSTRALIA
*Contact author: paul.rustomji@csiro.au

Keywords: Water quality, web service, trend analysis, Sweave

The range of environmental characteristics (climatic, hydrologic etc.) being routinely monitored is growing rapidly with time and increasingly these observations are being made available on-line. One technology that has an important role to play in analysing this growing data volume (not least in the field of hydrology) is web-service technology (1; 3). A web service is an interface that describes a collection of operations that are network-accessible through standardised XML messaging (4). Here, a web service that performs trend analyses of water quality data, using R as the statistical engine for analysis and visualisation, is presented.

Microsoft .NET is used to construct the web service. For each analysis a text file containing user choices regarding method selection etc. and a data file (`*.csv` format, based on data retrieved either from on-line data bases or stored locally), are generated by the client application. R is invoked with an initial `trend_analysis.r` script using the `Rscript.exe` command with the data and parameter files supplied as arguments. This script in turn calls individual sub-routines (depending on parameters specified via the parameter file). Each sub-routine comprises a **Sweave** file which, when `Sweave`'d produces both a LaTeX output file and PDF graphs. The LaTeX output files (`*.tex` format) and graphs are compiled using PDFLaTeX to produce a downloadable PDF report. Current trend analysis methods supported include generalised additive models incorporating a non-linear spline trend term (5), linear models and a non-parametric Seasonal Kendall's Tau slope estimate (2). Finally, a multi-site wrapper routine using GoogleMaps tiles downloaded on the fly (based on **RGoogleMaps**) provides a spatial map of water quality trends.

The advantages of this approach include:

1. Making the statistical power, flexibility and graphing capabilities of R available to a larger audience. Essentially a web-service approach provides access to high level statistical analyses without any requirements for direct R programming capacity on behalf of the user.

2. The ability to reference R objects in the report using the Sweave `Sexpr{}` function.

3. The ability to include interpretative statements in the report tailored to the statistical results (using the LaTeX **ifthenelse** package in conjunction with `Sexpr{}` statements).

4. Harnessing the typesetting capabilities of the TeX engine to produce a high quality PDF report.

5. Accessing the mapping capabilities of GoogleMaps (an extant web-service).

6. Internet-wide accessibility and the ability for this web-service to be called by other users and/or applications (e.g. from Microsoft Excel).

The web service can be accessed at `http://wron.net.au/WebApps/WQSARPortal/Home.aspx` through either a browser based web application or a downloadable Add-In for Microsoft Excel.

References

[1] J. L. Goodall, J. S. Horsburgh, T. L. Whiteaker, D. R. Maidment, and I. Zaslavsky, *A first approach to web services for the National Water Information System*, Environmental Modelling and Software **23** (2008), 404–411.

[2] B. Henderson and R. Morton, *Strategies for the estimation and detection of trends in water quality*, Tech. report, CSIRO Mathematical and Information Sciences Technical Report 07/70, Canberra, 2008.

[3] J. S. Horsburgh, D. G. Tarboton, M. Piasecki, D. R. Maidment, I. Zaslavsky, D. Valentine, and T. Whitenack, *An integrated system for publishing environmental observations data*, Environmental Modelling & Software **24** (2009), 879–888.

[4] H. Kreger, *Web Services Conceptual Architecture (WCSA 1.0)*, Tech. report, IBM Software Group, Somers, New York, 2001.

[5] R. Morton and B. Henderson, *Estimation of non-linear trends in water quality: An improved approach using generalized additive models*, Water Resources Research **44** (2008), 7420.