CL:AIRE technical bulletin

TB 21
(July 2019)

CL:AIRE technical bulletins describe specific techniques, practices and methodologies relevant to sites in the UK. This bulletin describes the GroundWater Spatiotemporal Data Analysis Tool (GWSDAT) which can be used to analyse and report trends in groundwater quality monitoring data.

The GroundWater Spatiotemporal Data Analysis Tool (GWSDAT) for Groundwater Quality Analyses

1. INTRODUCTION

Groundwater quality monitoring is carried out at sites where contamination of groundwater resources has occurred, or at sites where the potential consequences of contamination are high (Fouillac *et al.*, 2009). In both cases, the monitoring strategy assesses the spatial pattern of a certain environmental parameter (i.e. the contamination plume) as well as the trend through time (Loaiciga *et al.*, 1992). The collection of groundwater quality data requires installation of groundwater monitoring wells, periodic water sampling and laboratory analyses. All of which come at considerable cost and effort.

Following data collection, the groundwater quality data is analysed to assess risks to receptors or verify the absence of new releases and often needs to be presented to regulators in a clear and understandable manner. The most basic method of evaluating groundwater quality data and trends involves investigating the time-series of groundwater constituent concentrations independently. Time-series analyses with spatial mapping typically involve fitting a concentration trend surface (i.e. Kriging) to evaluate spatial patterns (Gaus *et al.*, 2003). It is still common practice when analysing groundwater quality data to independently apply spatial modelling techniques to episodic monitoring events (Ricker, 2008) or apply a single spatial model to a data set which has been consolidated over a time period (Aziz *et al.*, 2003). Alternatively, spatiotemporal modelling uses groundwater quality data from different time steps, which has been shown to be very effective to obtain more information from a groundwater dataset when compared with spatial analyses alone (Evers *et al.*, 2015, McLean *et al.*, 2019).

Whilst there are several groundwater data analysis applications, the most sophisticated tend to be designed for large-scale and long-term groundwater monitoring networks (Aziz *et al.*, 2003; Cameron, 2004). Similarly, whilst Geographic Information System (GIS) applications (e.g. ArcGIS, QGIS) have excellent visualisation tools for geographical interpretation they also have a high initial data setup cost, operator competence requirements, and perhaps surprisingly, only limited geostatistical modelling techniques.

In order to have an easy to install and freely available tool, the GroundWater Spatiotemporal Data Analysis Tool (GWSDAT) was developed as a collaboration between Shell Global Solutions and Glasgow University. GWSDAT is a user-friendly, open source software tool used to analyse and report trends in groundwater quality monitoring data. This technical bulletin summarises software architecture, installation requirements, application and functionalities, and future developments.

2. SOFTWARE ARCHITECTURE

The statistical engine used to perform geostatistical modelling and display graphical output is the open source statistical programming language R (R Development Core Team, 2012). The R project is used across a wide range of disciplines and has been adopted by the environmental sciences community (Carslaw and Ropkins, 2012). Members of the R community contribute statistical routines and functionality to this collaborative project by means of an open standardised package structure. An overview of available R packages, including GWSDAT, for hydrological applications can be viewed here: https://cran.r-project.org/web/views/Hydrology.html.

2.1 Installation Instructions

There are several options for installing and using GWSDAT:

1. **Excel Add-in Interface:** GWSDAT was historically designed principally to integrate with Microsoft Excel, a software application routinely used by environmental engineers for storing and analysing environmental data. The user entry point to GWSDAT is a custom-built Excel Add-in menu. Full instructions can be found at www.claire.co.uk/GWSDAT.
2. **Directly from R:** Install R together with the GWSDAT package – good option for experienced R programmers. See https://cran.r-project.org/web/packages/GWSDAT/index.html.
3. **Online version**: GWSDAT v3.0 uses the R web browser graphical user interface tool Shiny: http://shiny.rstudio.com, enabling a web-based implementation. See www.gwsdat.net – this includes access to the online version hosted by the School of Mathematics and Statistics at the University of Glasgow together with background supporting information.

Options 1 and 2 are local installations, i.e. no data is sent over the internet, and so may be the safer option if using sensitive data.

3. USING GWSDAT

GWSDAT's main functionalities include trend analyses, data smoothing, spatiotemporal smoothing, and determination of contamination plume characteristics. Groundwater quality observation data for individual monitoring points can be fitted to a linear or log-linear regression model, where the significance of a trend can be determined using the Mann-Kendall approach which is widely used for trend detection in groundwater and surface water studies (Bouza-Deaño *et al.*, 2008; Wahlin and Grimvall, 2010). In addition to these monotonic regression methods, groundwater solute concentration (contour) maps are developed using a spatiotemporal solute concentration smoother derived from a Penalized-Splines nonparametric regression. The simultaneous statistical smoothing over space and time generally provides a more accurate, consistent illustration of a contaminant plume and solute distribution when compared against contour maps of individual sampling rounds. In general, the spatiotemporally-smoothed plume will be less biased by missing sampling rounds or missing data within sample rounds. GWSDAT calculates a number of quantitative plume metrics (based on Ricker (2008)) describing the plume temporal behaviour: the plume's contaminant mass, average concentration, footprint area, and location of the centre of mass, all based on a spatial integration of the plume solute concentrations. Further details on each of the methodologies (including descriptions of R packages used, but excluding the plume metrics which comprise the latest addition), can be found in Jones *et al.* (2014).

3.1 Data Input

Data entry is via a standardised Microsoft Excel input template or directly in the on-line environment (Figure 1). When using the online version of GWSDAT, the template can be used to enter or copy data and can then be uploaded. Groundwater solute concentration data is entered for different constituents in different wells as a function of time. Input in this table can also comprise 'non-detects' (which can be specified at either the detection limit or half the detection limit), groundwater elevation data (which can be used for drawing additional groundwater contours), and light non-aqueous phase liquid (LNAPL) thickness (which can be automatically replaced by groundwater solute concentrations for interpolation purposes. This will be either the maximum groundwater solute concentration observed in the dataset or calculated effective aqueous solubility of the solute). The second table contains the coordinates for monitoring wells and the third table can be used to specify the location of a base map (in ArcGIS shapefile format).

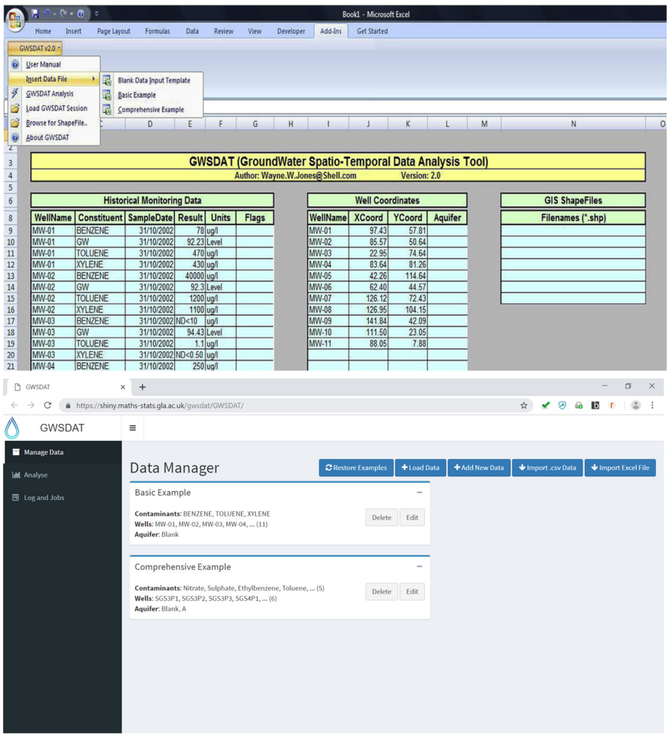

Figure 1. GWSDAT input sheet (top) and on-line GWSDAT data manager (lower) where template can be downloaded and uploaded.

3.2 Results and Reporting

The main output of GWSDAT consists of the screen shown in Figure 2 which has 3 main tab sheets (highlighted with A to C in Figure 2). The concentration map (panel A) shows a spatiotemporal model for a given time-slice (that was selected in panel A), including locations of monitoring wells and observed solute concentrations, and with optional groundwater elevations and LNAPL thicknesses. The plume boundary contour is illustrated relative to a specified background or regulatory compliance concentration value. If the boundary contour is closed (i.e. the entire plume is captured by the spatiotemporal model), plume mass per metre aquifer thickness and plume area will be calculated. In panel B the user can select which constituent to plot and whether to plot solute concentration trends, groundwater levels and/or LNAPL thickness. The trend and indicator threshold matrix (panel C) provides a fast way to determine concentration trends in the monitoring network for a given time slice (green = declining concentrations, white = stable, red = increasing). When

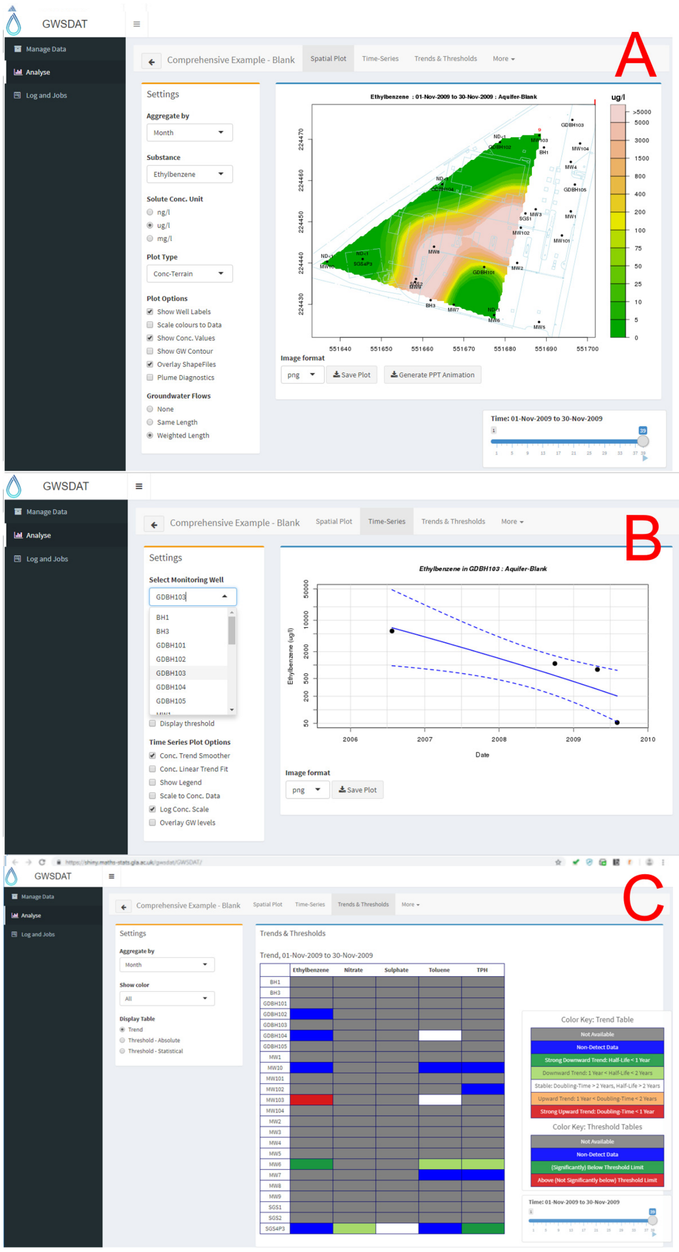

Figure 2. GWSDAT output screens from the on-line GWSDAT version.

using the matrix in 'threshold mode', the user can enter water quality threshold values and screen the data at the given time period against those thresholds (if the 'statistical threshold' option is used, the 95%-percentile of the data is screened against the threshold criteria).

GWSDAT can generate different reports in different formats. The plots from the output screen can be directly exported as 'jpeg', 'postscript' or 'pdf' files. It is also possible to export a sequence of plots capturing different time-slices (both graphs from a single well or spatiotemporal maps) and directly import these into Microsoft Word or PowerPoint. A complete series of graphs can be exported using the well reporting feature. This generates a matrix of graphs (one for each well) in which a selection of constituents can be plotted. Lastly, summary graphs can be exported plotting the plume metrics based on the method by Ricker (2008) for any given constituent over time (Figure 3). This feature in particular, often in combination with a movie-like presentation of the contour plots in PowerPoint, rapidly creates a comprehensive view of plume behaviour.

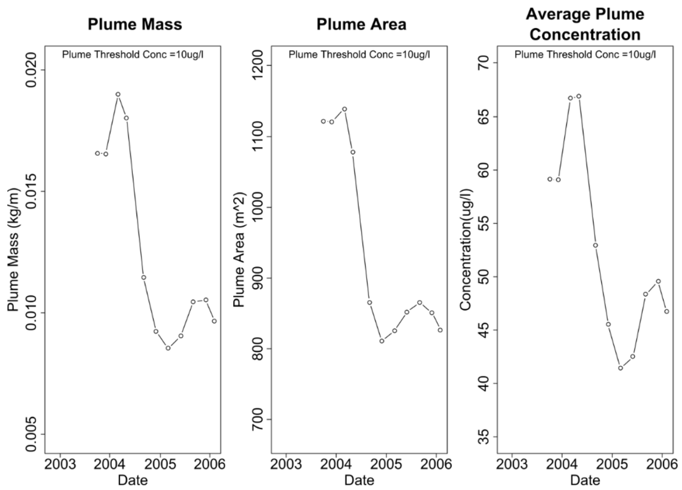

Figure 3. Example of the summary output of plume metrics: plume mass (left), plume area (mid), average concentration (right).

The functionality of the Excel GWSDAT version is almost identical to the on-line version with some small differences. For a more elaborate explanation of the Excel version, reference is made to (Jones *et al.*, 2015) and the supplementary information.

4. DISCUSSION

GWSDAT has been used operationally in the monitoring and assessment of Shell's global downstream assets (e.g. refineries, terminals, fuel stations) for over 9 years. Graphical output generated from GWSDAT is routinely included in reports submitted to environmental regulators. The cost savings, according to one survey of environmental engineers, run into millions of US$ and have been achieved through the following benefits:

- Rapid interpretation of complex data sets for both small and large groundwater monitoring networks.
- Earlier identification of new spills or off-site migration.
- Reduced reliance on engineered remediation through increased use of monitored natural attenuation remedies, where groundwater data analysis supports its effectiveness.
- Earlier closeout of sites in needless long-term monitoring and/or remediation.
- Simplified preparation of groundwater monitoring reports.

The use of open-source software in developing GWSDAT was important because it allowed complete transparency so that users and environmental regulators could see the code and techniques, such that they were open to full scrutiny and continuous improvement. Thanks to its open source foundation, GWSDAT has now become a globally-adopted software application in the environmental monitoring industry. GWSDAT has already been downloaded over 10,000 times and is used by many engineers and scientists, environmental regulators, students and researchers around the world. The expectation is that the environmental monitoring community will participate in its continued development and contribute to extending its capabilities. To this end, the University of Glasgow has taken GWSDAT one step further and developed the online version: www.gwsdat.net.

5. FUTURE DEVELOPMENTS

The major area for future development of GWSDAT is the addition of monitoring network optimisation tools. The primary objective of these tools will be to help users design monitoring location and sampling strategies to extract hydrogeological information with the required precision whilst minimising monitoring effort.

McLean *et al.* (2019) have shown that a similar or higher amount of information can be obtained with fewer observations when using a single spatiotemporal model rather than the conventional approach of applying independent spatial models (e.g. Kriging) to separate sampling events. This means that spatiotemporal methods can achieve the same level of performance but with fewer data points. GWSDAT users are already benefitting from this information extraction efficiency because it is based on such a spatiotemporal model.

McLean *et al.* (2019), together with body of research described in McLean (2018) will form the theoretical basis for the development of applied network optimisation tools in GWSDAT. The tools will include:

- Optimal Sampling Schedules: Recommendation on which monitoring wells should be sampled at different sampling events to maximise information extraction efficiency.
- Well Redundancy Analysis: Analyse which, if any, wells within the monitoring network can be removed or ignored without a significant loss of information.
- Well Sufficiency Analysis: Evaluate and identify locations for the construction of additional wells within the existing monitoring network to reduce uncertainty.

6. CONCLUSIONS

GWSDAT is a user-friendly, open source software tool used to analyse groundwater quality monitoring data. GWSDAT has a statistical engine which uses the statistical programming language R.

technical bulletin

GWSDAT can be used with either Microsoft Excel or a web browser as a front end. The software can be applied to analyse dissolved-phase concentration and trends in LNAPL thickness, plume dynamics (flow direction and transport) and attenuation rates for individual chemical constituents. The software can handle large and small data sets with multiple monitoring locations, variable sampling events and differing chemical constituents. GWSDAT has been used extensively in the assessment of soil and groundwater risks and the performance of active remediation systems at numerous assets around the world, including retail and manufacturing sites.

REFERENCES

- Aziz, J.J., Ling, M., Rifai, H.S., Newell, C.J. and Gonzales, J.R. 2003. MAROS: A decision support system for optimizing monitoring plans. Groundwater. 41(3), 355-367. https://doi.org/10.1111/j.1745-6584.2003.tb02394.x.
- Bouza-Deaño, R., Ternero-Rodríguez, M. and Fernández-Espinosa, A.J. 2008. Trend study and assessment of surface water quality in the Ebro River (Spain). Journal of Hydrology. 361(3-4), 227-239. https://doi.org/10.1016/j.jhydrol.2008.07.048.
- Cameron, K. 2004. Better optimization of long-term monitoring networks. Bioremediation Journal. 8(3-4), 89-107. http://doi.org/10.1080/10889860490887464.
- Carslaw, D.C., Ropkins, K. 2012. openair - an R package for air quality data analysis. Environmental Modelling & Software. 27-28 (0), 52-61. https://doi.org/10.1016/j.envsoft.2011.09.008.
- Evers, L., Molinari, D.A., Bowman, A.W., Jones, W.R., Spence, M.J., 2015. Efficient and automatic methods for flexible regression on spatiotemporal data, with applications to groundwater monitoring. Environmetrics. 26(5), 431-441. https://doi.org/10.1002/env.2347.
- Fouillac, A.M., Grath, J. and Ward, R. 2009. Groundwater Monitoring, Wiley. https://www.wiley.com/en-gb/Groundwater+Monitoring-p-9780470778098.
- Gaus, I., Kinniburgh, D.G., Talbot, J.C. and Webster, R. 2003. Geostatistical analysis of arsenic concentration in groundwater in Bangladesh using disjunctive kriging. Environmental Geology. 44(8), 939-948. https://doi.org/10.1007/s00254-003-0837-7.
- Jones, W.R., Spence, M.J., Bowman, A.W., Evers, L. and Molinari, D.A. 2014. A software tool for the spatiotemporal analysis and reporting of groundwater monitoring data. Environmental Modelling & Software. 55(0), 242-249. https://doi.org/10.1016/j.envsoft.2014.01.020.
- Jones, W.R., Spence, M.J. and Bonte, M. 2015. Analyzing groundwater quality data and contamination plumes with GWSDAT. Groundwater. 53(4), 513-514. https://doi.org/10.1111/gwat.12340.
- Loaiciga, H.A., Charbeneau, R.J., Everett, L.G., Fogg, G.E., Hobbs, B.F. and Rouhani, S. 1992. Review of ground-water quality monitoring network design. Journal of Hydraulic Engineering. 118(1), 11-37. https://doi.org/10.1061/(ASCE)0733-9429(1992)118:1(11).
- McLean, M.I., L. Evers, A.W. Bowman, M. Bonte & W.R. Jones. 2019. Statistical modelling of groundwater contamination monitoring data: A comparison of spatial and spatiotemporal methods. Science of The Total Environment. 652, 1339-1346. https://doi.org/10.1016/j.scitotenv.2018.10.231.
- McLean, M.I. 2018. Spatiotemporal models for the analysis and optimisation of groundwater quality monitoring networks. PhD thesis, University of Glasgow, Glasgow. http://theses.gla.ac.uk/38975/1/2018McLeanPhD.pdf.
- R Development Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL. http://www.R-project.org.
- Ricker, J.A. 2008. A practical method to evaluate ground water contaminant plume stability. Groundwater Monitoring & Remediation. 28(4), 85-94. https://doi.org/10.1111/j.1745-6592.2008.00215.x.
- Wahlin, K. and Grimvall, A. 2010. Roadmap for assessing regional trends in groundwater quality. Environmental Monitoring and Assessment. 165(1-4), 217-231. https://doi.org/10.1007/s10661-009-0940-7.

For further information please contact the authors: Wayne R Jones, Shell Global Solutions (UK) Ltd (Wayne.W.Jones@shell.com), Matthijs Bonte, Shell Global Solutions International B.V. (Matthijs.Bonte@shell.com) and Karina Cady, Shell International Ltd (Karina.Cady@shell.com).