

<b>Institution: University of Glasgow</b>
<b>Unit of Assessment: 10 – Mathematical Sciences</b>
<b>Title of case study: Development of an innovative data analysis tool to monitor groundwater pollution and environmental impact</b>
<b>1. Summary of the impact</b>

With global demand for energy ever increasing, environmental impact has become a major priority for the oil industry. A collaboration between researchers at the University of Glasgow and Shell Global Solutions has developed GWSDAT (GroundWater Spatiotemporal Data Analysis Tool). This easy-to-use interactive software tool allows users to process and analyse groundwater pollution monitoring data efficiently, enabling Shell to respond quickly to detect and evaluate the effect of a leak or spill. Shell estimates that the savings gained by use of the monitoring tool exceed \$10m over the last three years. GWSDAT is currently being used by around 200 consultants across many countries (including the UK, US, Australia and South Africa) with potentially significant impacts on the environment worldwide.

<b>2. Underpinning research</b>
---------------------------------

‘Flexible regression’ has been a major statistical research theme at the University of Glasgow for over 20 years. The main aim of this research is to use observed data to estimate an underlying regression function, describing the relationship between a response and one or more covariates, without constraining the function to follow any particular shape, apart from requiring it to be smooth. Professor Adrian Bowman (Professor of Statistics since 1995) has made extensive contributions to this topic, including a book (Bowman & Azzalini, 1997) which provided methods for a wide variety of data structures and placed a strong emphasis on the inferences which could be drawn from real applications.

Since 2002, research has been directed towards spatiotemporal models for data which have been collected over both space and time. There are many environmental applications for models of this type. A particular example is the modeling of sulphur dioxide air pollution over Europe (Bowman et al., 2009) using data from an extensive network of monitoring stations. The underlying methodology represented one of the first developments of flexible regression in a spatiotemporal setting, allowing complex trends and interactions to be explored, and both theory and computational methods were described.

In 2006, Shell Global Solutions began to investigate the use of these methods in the context of groundwater monitoring, prompted in part by the appointment of a PhD student in Statistics from the University of Glasgow. The statistical and computing tools developed in Glasgow fed into the creation of new software, which gained extensive use within the company. The software tool allowed responsible staff to view and interpret the extensive groundwater data which is routinely collected, and the company quickly recognised the potential value of the tool within Shell operations. Version 1 of GroundWater Spatiotemporal Data Analysis Tool (GWSDAT) was released for use by Shell consultants in 2009.

With interest from Shell, discussions about how the methodology might be developed further were initiated. A PhD studentship, jointly funded by Shell and the University of Glasgow, was created and a student recruited in 2009. The subsequent research agenda placed strong emphasis on (i) the need for very efficient computational methods for large datasets and (ii) the production of very

## Impact case study (REF3b)

stable estimates which provided good predictions even in regions where the monitoring sites were sparsely located. Dr. Ludger Evers (Lecturer in Statistics, University of Glasgow, 2008-present) joined the research project in 2009, contributing extensive experience in computational and Bayesian methods. The ensuing research used B-spline bases to provide pollution surface estimates which can be represented mathematically in very compact forms, employing a fully Bayesian model to account for all the sources of uncertainty and using a combination of matrix techniques to provide very fast solutions. The combination of these methods allows pollution levels to be estimated in a fully automatic, and very stable, manner. The new model also gives access to useful measures of uncertainty, including credibility intervals and standard errors. These methods have been implemented in Shell software since 2012 but a paper describing the underlying methodology was deferred until full practical evaluation could take place and illustrative examples reported (Bowman et al., 2013).

In parallel with this line of research, a statistical software tool for interactive analysis and graphics is under development at the University of Glasgow (School of Mathematics and Statistics). The standard statistical computing environment in academic research and many areas of industry is 'R'. In 2005 Bowman began to lead the development of 'rpanel', an add-on package for R, which was created to make it as easy and efficient as possible for non-technical users to construct interactive software such as dynamic graphics (Bowman et al., 2006ab). Users can easily create buttons, sliders and other types of graphical control and can interact with plots in ways which considerably enhance exploration of datasets. The package is widely used and continues to be developed in Glasgow, with the most recent version released in 2013.

Shell Global Solutions used the 'rpanel' package to build the first version of GWSDAT in 2007-2009. The University of Glasgow team worked with the company to add functionality to 'rpanel' in response to specific requests. The resulting Shell software, GWSDAT incorporates the statistical methodology described above and uses the rpanel package to provide users with easy to use, interactive controls for reading data, fitting models and exploring the results in graphical form.

### 3. References to the research

- Bowman, A.W. & Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis*. Oxford University Press. [ISBN 9780198523963](#) [available from HEI]
- Bowman, A.W., Crawford, E., Alexander, G. & Bowman, R.W. (2006a). [rpanel: simple interactive controls for R functions using the tcltk package](#). *Journal of Statistical Software* 17, issue 9.\*
- Bowman, A.W., Crawford, E. Bowman, R.W. (2006b). rpanel: making graphs move with tcltk. *R News*; 6, issue 4. ([pdf](#); [software](#))
- Bowman, A.W., Giannitrapani, M., & Scott, E.M. (2009). Spatiotemporal smoothing and sulphur dioxide trends over Europe. *Applied Statistics*, 58, 5, 737-752. [doi:10.1111/j.1467-9876.2009.00671.x](#) \*
- Bowman, A.W., Evers, L., Molinari, D., Jones, W. & Spence, M.J. (2013). [Bayesian smoothing of spatiotemporal data with applications to groundwater monitoring](#). Submitted for publication. (arxiv:1310.7815) \*
- Jones, W., Spence, M.J., Bowman, A.W., Evers, L., Molinari, D. (2013) GWSDAT - Groundwater Spatiotemporal Data Analysis Tool. Submitted for publication. ([arxiv:1310.8158](#))

\* best indicators of research quality

Full details on the rpanel package can be found at;

- <http://cran.r-project.org/web/packages/rpanel/>
- <http://www.stats.gla.ac.uk/~adrian/rpanel/>

#### 4. Details of the impact

Shell Global Solutions (based in Chester, UK) undertakes a wide variety of projects where statistical and mathematical modelling is required. The GWSDAT software was developed by Shell, based on statistical methodology and interactive software tools from the University of Glasgow. GWSDAT was first released for use by Shell consultants in 2009 and the most recent version was released in 2012. A survey carried out by Shell in 2013 estimates there are currently around 200 users worldwide, from countries including, for example, the UK, USA, Australia and South Africa.

Shell is responsible for the operation of a huge number of installations, from small petrol stations to extensive refinery sites, and significant attention is paid to monitoring the surrounding environment to ensure the prevention of any adverse effects. Inadvertent release of soluble material can lead to pollution of groundwater, with a risk that contamination is transported beyond the site boundaries. It is standard practice to drill boreholes around installations, from which samples can be regularly drawn and analysed, in order to identify and monitor the presence of any pollutants.

The measurements from boreholes at a particular site generate a dataset which records pollutant levels at points in space and time, but there are issues with the practicalities and costs of sampling which restrict the number of measurements made. Also, the consultants responsible for interpreting these measurements generally have a strong scientific or engineering background but may not have extensive training in statistical methods. Using data from boreholes to construct reliable estimates of the underlying pollution patterns across (i) the whole site and (ii) the entire time period requires specialist statistical methods and concepts. However, referring results to others with the necessary expertise for statistical analysis would potentially lead to significant delays.

The GWSDAT directly addresses these needs. Shell has identified the benefits of the software to include:

- The early identification of rising contaminant concentration trends, leading to reduced response time in the event of leaks or spills. There has been more than one instance of unknown leaks being identified by the use of the software tool.
- Improved data transparency leading to better designed monitoring networks and more robust conceptual site models, thereby avoiding collection of redundant data.
- Clarity on the relationship between dissolved solute concentrations and groundwater elevation and flow direction, for improved plume control measures and fit-for-purpose remediation system design.
- Significantly reduced time and resource expenditure on the analysis and reporting of monitoring data, particularly in the case of smaller sites where insufficient data is available to justify the cost of using geographic information systems or transport simulations.
- Significant cost savings. A recent survey undertaken by the company indicates that the software package has led to savings in the region of \$10m in the last 3 years.

## Impact case study (REF3b)

Users view an Excel spreadsheet and explore, graphically and interactively, the data it contains. Importantly, the underlying statistical model constructs an estimate of the pollution surface across the entire spatial region of interest and a slider can be used to view how this changes over time. Different solutes can be selected or the data from particular wells viewed in greater detail, with estimates of trend superimposed. Large datasets (more than 50,000 rows) can be modelled in a matter of seconds. The added value and flexibility of the system has been very well received by users.

Key staff responsible for the GWSDAT software application within Shell commented that:

*Data analysis and reporting that previously required weeks of effort can now be completed automatically, at the click of a mouse. The GWSDAT spatiotemporal modelling research programme has yielded total savings for Shell in excess of \$10M over the past 3 years.*

Although developed specifically for use by Shell, the package is also being made available to external users, including environmental regulators. It is currently being submitted to an industry website from which it will be freely available to other industrial users. Shell also regularly publicises the package at national and international conferences.

### 5. Sources to corroborate the impact

#### Shell have presented GWSDAT at the following conferences:

- AquaConSoil 2013 (12th International UFZ-Deltares Conference on Groundwater-Soil-Systems and Water Resource Management), Barcelona, Spain, April 2013
- ENBIS (European Network for Business and Industrial Statistics), Coimbra, Portugal, Sep 2011
- UseR! (The R User Conference), Warwick, UK, Aug 2011
- NICOLE (Network for Industrially Contaminated Land in Europe) Network Conference, Copenhagen, Denmark, May 2011. Workshop on Innovative Site Characterisation Tools.
- European Geosciences Union (EGU), Vienna, Austria, Apr 2011
- Environmental Protection Agency National Tanks Conference, Boston, USA, Sep 2010
- UseR! (The R User Conference), Rennes, France, Jul 2009

#### Testimonials:

- Statistical Consultant, Shell Global Solutions UK (details of software application within Shell, confirmation of improved data analysis and savings)
- Environmental Geochemist, Shell Global Solutions UK (details of software application within Shell, confirmation of improved data analysis and savings)