OR/18/052 Methods

Lapworth, D J, Crane, E J, Stuart, M E, Talbot, J C, Besien, T, and Civil, W. 2018. Micro-organic contaminants in groundwater in England: summary results from the Environment Agency LC-MS and GC-MS screening data. *British Geological Survey Internal Report*, OR/18/052.

Data sources

The Environment Agency has a statutory responsibility for monitoring the quality of groundwater in England. As part of their monitoring programme, samples for micro-organic compounds are collected and analysed in response to Water Framework Directive and Groundwater Directive requirements, and for State of the Environment reporting. Most of the monitoring points used in this report are from the Environment Agency's Groundwater Quality Monitoring Network. These monitoring points were selected to represent the regional quality of the groundwater. However, there are also a relatively small number of other sites included in the results and some of these could be point source monitoring points. These point source monitoring points could account for some of the maximum concentration values presented in later in this report.

Data we have been provided with (*see* Data collation and processing) have come from analyses undertaken by the National Laboratory Service. Targeted GC-MS and LC-MS screening for organic substances was carried out following sample pre-concentration. For the GC-MS method a double liquid-liquid extraction was employed, using acid-neutral dichloromethane, to extract non-polar substances. Oasis® HLB cartridges were used for SPE for the LC-MS method, elution was done using 0.1% Formic acid in methanol/acetonitrile (1:1).

The GC-MS target based (multi-residue) screening method allowed for almost all GC-amenable pesticides as well as hundreds of other organic contaminants to be identified from a single sample, incorporating over 850 substances and including both volatile organic compounds (VOCs) and semi-volatile organic compounds (SVOCs). Chemicals could be identified at concentrations as low as 0.01 µg/L using deconvolution reporting software (DRS).

A LC-MS (Q-TOF) method (e.g. Batt and Aga, 2005[1]) was used to screen for polar organic compounds in each sample. Target compounds for quantification have been analysed in a blank and at a concentration of 0.1 µg/L, the response factor obtained is used to create a single point calibration curve. Estimate of concentration is based on quant ion response and response of the internal standard. Quantification limits are compound specific and are typically between 0.001–0.1 µg/L for the vast majority of compounds. Target compound identification is made by retention time, accurate mass and by isotope distribution patterns (mass, ratio, spacing). The combined results contribute to an overall match score for each substance.

Data collation and processing

GC-MS and LC-MS data provided to BGS by the EA represents all available groundwater samples analysed as part of EA monitoring activities and the vast majority of data come from the EAs national groundwater quality monitoring network. Data from one-off investigations could also be included and may in some cases could account for high detection frequencies and concentrations, although this has not been investigated as part of this study. This target screening data provides surveillance

data on a broad range of substances and complements the dedicated and much smaller EA analytical suites used for monitoring and regulatory purposes (e.g. pesticides and TPs).

GC-MS data

Existing bgs database

BGS was first given access to Environment Agency GC-MS semi-quantitative screening data in 2010. The initial raw dataset contained 17 694 records with monitoring sites in both England and Wales. Additional data was provided in 2012, with an interpretation reported by Manamsa et al., 2016[2]. Data processing included both automated (using queries) and manual data separation (e.g. to extract data where multiple records were presented concatenated into a single record) and cleaning (e.g. correcting CAS numbers). The cleaned data is stored in a Microsoft Access database.

Additional raw GC-MS data

The most recent GC-MS data (England only) were provided in October 2017 and appended to the existing BGS database. The recent raw data has fewer data formatting errors than earlier tranches. Data processing was undertaken as described in [Workflow to prepare the GC-MS and LC-MS datasets](#).

The current version of the BGS database contains 27 283 records (each of which has a unique sample site + sample date + determined), taken from 7473 samples, and 2465 sites. The BGS database now contains data from samples which were analysed between June 2009 and July 2017.

In general, only detected values were reported, however, 134 records contain a '<' symbol in the 'less than' field: in these cases records were converted to below Limit of Detection (LOD), i.e. distinguishable from the blank, by recording '<LOD' in the 'value' field.

Three records had a blank 'value' field and were excluded from our data analysis.

GC-MS data from sites in Wales

The BGS database includes data from 62 sites in Wales. There are 213 records (unique sample site + sample date + determinand) for Wales (excluding records for the sulphur compound S8). One of the 213 records reports a 'less than' value. This data dates from prior to the formation of Natural Resources Wales in April 2013 (sample dates range from August 2009 to July 2012), and are therefore only a subset of data that exists for Wales.

The Wales data as described above was included in the statistical analysis but excluded from spatial plots. As they comprise less than 1% of the database their inclusion will not have a noticeable impact on the results of the statistical summaries.

LC-MS data

The LC-MS semi-quantitative screen has been introduced more recently and this is the first time that BGS has undertaken an analysis of this data. The data was provided to BGS in October 2017 and covers results between reported from April 2014 to October 2017. There are 4,089 records from 249 samples collected from 109 sites (all in England). No '<' values were recorded in the 'less than' field. This method was introduced as a trial in a number of EA Areas and has not been used across the EA groundwater quality-monitoring network in the same way as the GC-MS screen and therefore only data for a subset of monitoring sites is available to date.

Workflow to prepare the GC-MS and LC-MS datasets

The workflow used to prepare the datasets for statistical analysis and spatial plotting is summarised in Figure 2.1. The steps involved were:

1. Additional data (more recent GC-MS and all LC-MS) provided by the Environment Agency for this study was appended to the existing BGS database (in Microsoft Access), using a query to exclude duplicate records.
2. Data cleaning:

    - Manual correction of a small number of records was undertaken where data formatting errors were identified
    - Records where the value field (concentration) was blank were excluded
    - CAS numbers were reformatted where necessary to make them consistent (e.g. removing hyphens)
    - Records reporting concentrations of the sulphur compounds S8 (CAS number 10544500; cyclooctasulphur) and S6 (CAS number 13798237; hexathiane) were excluded, as they are not organic compounds

3. The remaining records in the database should be positive detections, however, some records reported '<' in the 'less than' field and others had a reported value (concentration) of '0': these were assigned <LOD value for data analysis and reporting purposes
4. The top 50 compounds (ranked by number of detections) were determined for the two screening methods (GC-MS and LC-MS)
5. The data for the top 50 compounds was then exported to Excel. The exported data consisted of a concentration for each (unique site + sampling date) and the dataset was then manually completed by assigning a <LOD value to empty cells (as this is a screening method, all determinands were by default analysed for in each sample).
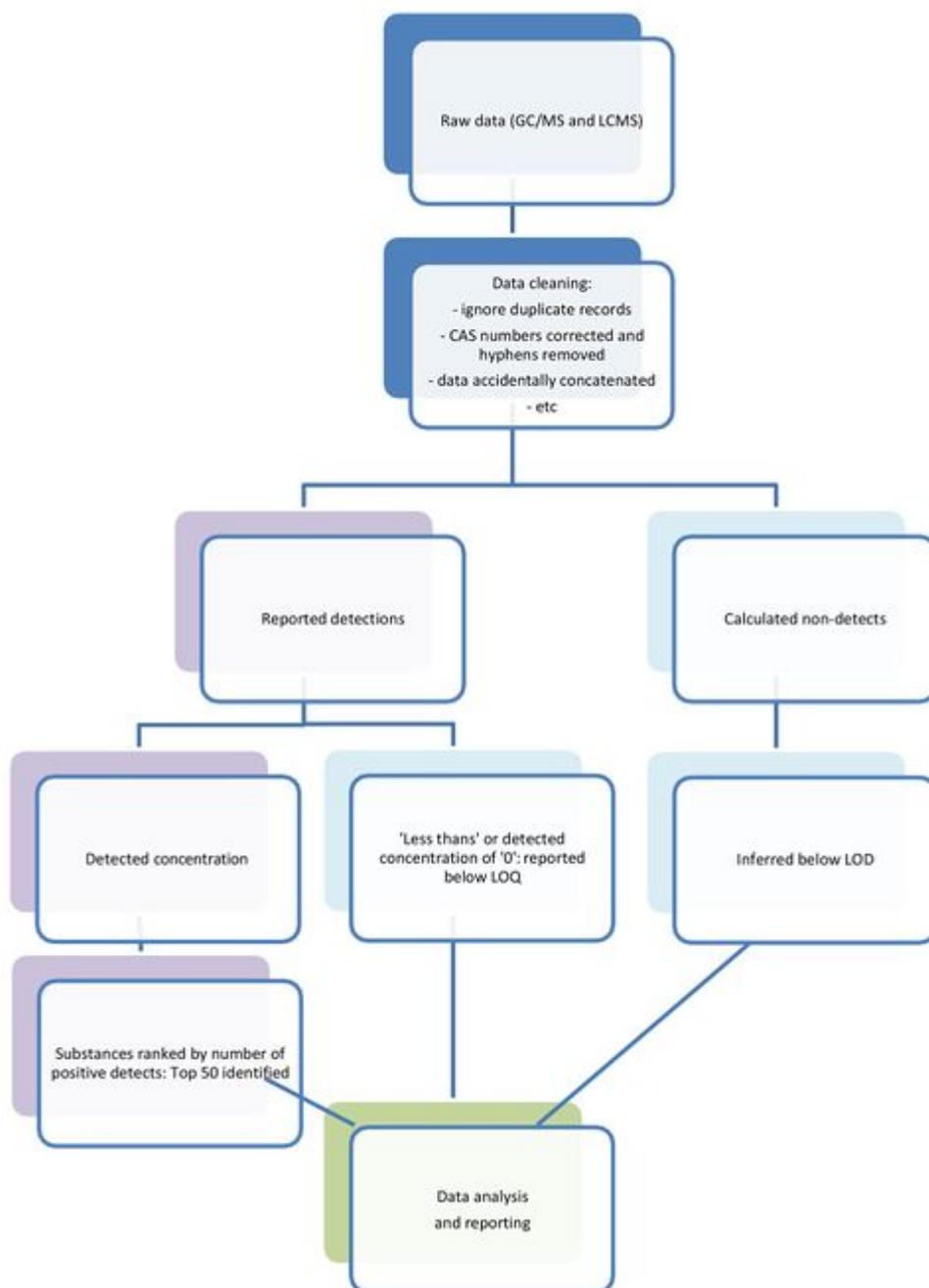
Figure 2.1   Summary workflow diagram.

Calculating non-detects

In order to calculate the proportion of samples in which a positive detection of a compound was made, it is necessary to calculate non-detects. As these are (mostly) not reported in the Environment Agency database, they have to be calculated. Where any compound has been detected at a site on a date by the analytical method of interest (GC-MS or LC-MS), all other compounds that could be detected by that analytical method have either been detected (value in the database) or (if no value in the database) are by inference non-detects.

It is possible that samples were analysed which had no positive detects of any determinands; if this is the case there would be no record of them or the non-detects in the database.

Statistical methods for data sets with non-detects

Due to the high proportion of censored results, i.e. those reported below the method LOD, i.e. reported as '< LOD', for each compound, summary statistics were computed using the R statistical package 'NADA: Non-detects and Data Analysis for Environmental Data' (Helsel, 2012[3]; R Core Team, 2018[4]; Lee, 2017[5]). Substitution methods, such as replacing non-detects with half the quantification limit or '0', are not recommended for calculation summary statistics. Two commonly used methods for estimating summary statistics (mean, median, quartiles) are maximum likelihood estimation (MLE) and robust regression on order statistics (ROS). Only where an adequate proportion of results was reported > LOD (detailed in Table 2.1), summary statistics such as the mean were computed. For large data sets (n≥50) with 50–80% of data censored the MLE is recommended, while ROS is recommended for smaller data sets or where fewer than 50% of data are censored in large data sets (Helsel, 2012[3]). No method is recommended when censoring is higher than 80%. These recommendations were applied to the calculation of summary statistics for this report.

Table 2.1   Recommended methods for estimating summary statistics with censored data.

Sample size	Percent of data censored		
	<50%	50–80%	>80%
n<50	ROS	ROS	censoring too high
n≥50	ROS	MLE	censoring too high

Grouping compounds

Compounds which have been detected by either of the analytical methods here have been classed into broad groupings to make the visualisation and description of the data more understandable. These classes are set out in Table 2.2. For many compounds, this process is straightforward with for example arable herbicides classed as pesticides, as are their transformation products (TPs) and chlorinated solvents classed as halogenated solvents.

Table 2.2   Compound use categories and short codes.

Code	Use category	Members
Pest	Pesticide	Herbicides, insecticides, fungicides and their TPs
HSol	Halogenated solvent	Chlorinated solvents, trihalomethanes (THM)s
PAH	Polyaromatic hydrocarbon	Napthalene, anthracene etc
Plast	Plasticiser	Phthalates, BPA
PPCL	Pharmaceutical, Personal Care Product, Lifestyle	Pharmaceutical, including veterinary, PCPs, e.g. UV absorbers, antimicrobials and insect repellent, caffeine
Indu	Industrial	Non-halogenated solvents, other industrial and PFAS
Ster	Sterols	e.g. cholesterol, squalene — are also naturally occurring in some cases and can have high background concentrations in some situations

Pharmaceuticals are grouped together with other compounds in a broad category which is likely to

enter the environment through wastewater, such as personal care products (PCPs) and caffeine. This category also includes veterinary compounds.

Assigning other compounds to a single class is difficult as they have multiple applications. UV absorbers may have both industrial and cosmetic uses, insecticides targeted at livestock or pet pests may also be classed as veterinary compounds (e.g. imidacloprid or fipronol), and industrial compounds can have a range of uses, including as plasticisers.

The use of some compounds has changed over time. Atrazine and simazine have been withdrawn in the UK for many uses since 2003 but still occur in groundwater. PFOS is another very good example: it currently has only a few restricted industrial uses, but it previously had a wide variety of uses including firefighting foams and as an impregnating agent in a number of products such as carpets, furniture, paper, textiles and leather.

This approach, while having such limitations as described above, makes the data interpretation more accessible.

References

1. ↑ Batt, A L, and Aga, D S. 2005. Simultaneous analysis of multiple classes of antibiotics by ion trap LC-MS/MS for assessing surface water and groundwater contamination. Analytical chemistry, 77(9), 2940–2947.
2. ↑ MANAMSA, K, CRANE, E, STUART, M, TALBOT, J, LAPWORTH, D, and HART, A. 2016. A national-scale assessment of micro-organic contaminants in groundwater of England and Wales. Science of the Total Environment, 568, 712–726.
3. ↑ 3.0 3.1 HELSEL, D. 2012. Statistics for censored environmental data using Minitab ® and R, Second Edition, John Wiley & Sons
4. ↑ R CORE TEAM (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
5. ↑ LEE, R L. 2017. NADA: Nondetects and Data Analysis for Environmental Data. R package version 1.6–1. https://CRAN.R-project.org/package=NADA

Retrieved from 'http://earthwise.bgs.ac.uk/index.php?title=OR/18/052_Methods&oldid=43861'
Category:

- OR/18/052 Micro-organic contaminants in groundwater in England: summary results from the Environment Agency LC-MS and GC-MS screening data

Navigation menu

- This page was last modified on 29 November 2019, at 10:43.