Accepted Manuscript


Research papers

# Setting the baseline for shale gas – establishing effective sentinels for water quality impacts of unconventional hydrocarbon development

Fred Worrall, Andrew J. Wade, Richard J. Davies, Alwyn Hart

PII: S0022-1694(19)30157-X
DOI: https://doi.org/10.1016/j.jhydrol.2019.01.075
Reference: HYDROL 23482

To appear in: *Journal of Hydrology*

Received Date: 16 August 2018
Revised Date: 25 January 2019
Accepted Date: 29 January 2019

Please cite this article as: Worrall, F., Wade, A.J., Davies, R.J., Hart, A., Setting the baseline for shale gas – establishing effective sentinels for water quality impacts of unconventional hydrocarbon development, *Journal of Hydrology* (2019), doi: https://doi.org/10.1016/j.jhydrol.2019.01.075

# SETTING THE BASELINE FOR SHALE GAS – ESTABLISHING EFFECTIVE SENTINELS FOR WATER QUALITY IMPACTS OF UNCONVENTIONAL HYDROCARBON DEVELOPMENT

Fred Worrall[1*] , Andrew J. Wade[2], Richard J. Davies[3] and Alwyn Hart[4]

1. Department of Earth Sciences, Durham University, Science Labs, Durham DH1 3LE, UK.
2. Department of Geography and Environmental Science, University of Reading, Whiteknights, Reading, RG6 6AB, UK.
3. School of Natural and Environmental Sciences, Newcastle University, Newcastle, NE1 7RU, UK.
4. Environment Agency, Research Assessment and Evaluation, Sapphire East, Streetsbrook Road, Solihull, B91 1QT, UK

**ABSTRACT**

There is a need for the development of effective baselines against which the water quality impacts of industry in general, and shale gas extraction specifically, can be assessed. The salinity, and hence the specific conductance, of fluids associated with shale gas extraction is typically many times higher that of river water. The contrast between these two water types means that testing for salinity (specific conductance) could provide an ideal sentinel for detecting environmental impact of shale gas extraction. Here, Bayesian generalised linear modelling was used to predict specific conductance across English surface waters. The modelling used existing, spot-sampled data from 2005 to 2015 from 123 sites to assess

* Corresponding author: Fred.Worrall@durham.ac.uk; tel. no: +44 (0)191 334 2295; fax no: +44 (0)191 334 2301

whether this approach could predict variation for subsequent years or for a new site (data from 2002 to 2015). We show that the results were readily projected in to subsequent years for sites included in the initial analysis. The use of covariates (land-use, hydroclimatic and soil descriptors) did not prove useful in predicting specific conductance at further sites not previously included in the analysis. The extension of the approach to 6833 English river monitoring sites with 10 or more observations from more than one year over the period 2005 to 2015 showed that it was possible to reproduce the seasonal variation in river water specific conductance. The approach taken here shows that it is possible to use low-frequency but widespread monitoring data to predict natural variation at monitoring sites to give a probabilistic assessment of whether or not a pollution incident has occurred and the seasonal variation, expressed as uncertainty bounds around the observations, at a specific site has been exceeded.

**Keywords:** shale gas; Bayesian statistics; generalised linear modelling

## 1. Introduction

To assess and indeed demonstrate an impact of any activity, it is necessary to show, within a reasonable level of certainty, that the industry has changed an environmental state over and above either that which was true without the activity present or beyond some accepted minimum level of harm. The need for demonstrating impact or indeed the ability to confirm the absence of an impact means that a baseline, or pre-intervention control, needs to be established for comparison with subsequent observations. The United Kingdom has a nascent shale gas industry and, given experience from the United States shale gas industry, one concern is the impact upon water quality of ground and surface water (eg. Kahrilas et al., 2014; Vengosh et al., 2014). To reassure the public and ensure protection of the UK water resource it is important that techniques exist for the detection, identification and attribution of pollution for possible impacts of unconventional hydrocarbon resource development. A number of technologies are used for water quality monitoring and several have been proposed for rapid, even continuous monitoring to detect any the water quality impacts of shale gas developments (eg. $CH_4$ – Teasdale et al., 2014; Radium – Lagace et al., 2018; Barium and Sulphate - Niu et al., 2018; Strontium isotopes – Kohl et al., 2014). However, here we propose a sentinel approach in which a single key parameter can be used as a rapid and early warning. However, to be an effective and robust sentinel of change the parameter monitored should have four properties. Firstly, any water quality parameter should be a lead, and not a lag, indicator of change, i.e. it should occur at the beginning of any impact to provide early warning and so that mitigation could be rapidly deployed. Second, the parameter must be sufficiently sensitive having a high contrast with the normal or background activity and so that any change cannot be mistaken for background or natural variation. Thirdly, the parameter should show a high specificity for the activity of concern and not normally be associated with or mistaken for, other activities; i.e. in this case it should be specific to a shale gas industry and not to other industries for example, conventional hydrocarbon extraction. Finally, the measurement technology should be cheap and readily deployable so that it can be used widely used and provide a large sample size.

By far the greatest difference between the waters arising from a shale gas well pad (those waters could be the fracking fluid, the flowback water or the produced water), and surface waters is salinity or its associated determinands, eg. total dissolved solids (TDS) or electrical conductivity (in this study, specific conductance which is the electrical conductivity of water standardised to a fixed temperature). The salinity of flowback water and deep formation water, as determined by TDS is often greater than seawater let alone greater than the salinity of river waters. Rowan et al. (2011) reviewed the total dissolved solids (TDS) of shale gas flowback water from US shale gas formations and showed that the flowback fluids were between two thirds and 10 times the seawater TDS (log TDS of seawater < 4.6) and much larger still than freshwater TDS (log TDS of freshwater ~ 2.6). Equally, the salinity of fracking fluids is far higher than that of surface waters and so salinity can also be used as a parameter for detecting fracking fluids as well as flowback water in surface and groundwater. For example, the only shale gas well so far fracked in the UK was at Preese Hall in Lancashire (Environment Agency, 2011, as cited in Almond et al., 2014). In this case, the flowback fluid salinity was between 3 and 5 times higher that of seawater; in contract freshwater salinity is typically only 0.2% of seawater, i.e. only a 0.07% addition of such flowback water would cause a doubling of salinity in an English surface water. Yet rather than being expensive or requiring specialist equipment salinity, or specific conductance or TDS, are regularly and routinely measured in surface and ground waters and there are long term records of freshwater specific conductance measurements whereas there are no long term measurements across multiple sites of dissolved $CH_4$ (eg. Teasdale et al., 2013). These properties mean that salinity, and its allied measures specific conductance and TDS, make an ideal sentinel of change for detecting water quality impacts of a developing shale gas industry as it readily measured; shows a high contrast against a background of freshwater environments; is highly specific for shale gas development; and its high specificity and contrast with background mean that it could be a lead indicator of any incident. Furthermore, high salinity water from hydrocarbon exploitation has been observed to be a major cause of toxicity in exposed organisms (He et al., 2017; Blewett et al., 2017) and in the Canadian province of Alberta in 2015 there were 113 documented incidents of spills of flowback and produced water (Alessi et al., 2016).

However, although there are considerable numbers of measurements of specific conductance available, these measurements have not been collected for the purpose of creating a baseline against which impacts of a new industry can be judged. The Environment Agency have identified a range of statistical tools for use with monitoring data for specific sites and are currently trialling these at two sites in the north of England. However, there is no coherent and consistent means of handling existing data to make the assessment of any impact; a coherent method is needed for objectivity and transparency and therefore, this study proposes a new method to use existing specific conductance data to assess the impact of fracking on surface and groundwater quality based upon generalised linear modelling. This approach is entirely data driven and uses all the existing data without the need for the parameterisation required in physical models; it is flexible with respect to the distribution chosen to represent the specific conductance data; and can include existing factorial (eg. location) and covariate information (eg. river flow or land use). The model was developed within a Bayesian framework. The Bayesian framework means that the approach creates a structure whereby all information has some value, i.e. information from monitoring sites not in a catchment of interest help inform the distribution of data within the catchment of interest. Furthermore, new information can be directly added to update estimates; and all model outputs come with a probability which means that risk and uncertainty are considered at all stages. The approach creates a dynamic baseline for assessment of water quality effects of a shale gas industry. Such a baseline is dynamic in both time and space, i.e., generating a time series of expected results that would be different for different catchments. Estimated and predicted baseline results are both specific to a given location and develop over time in response to natural changes meaning that it will improve with ongoing monitoring at shale gas or other infrastructure sites. Therefore, the approach of this study was to construct a dynamic baseline for surface water specific conductance using Bayesian generalised linear modelling such the outputs of the model give a probability of an unusual event, i.e. a pollution incident. The approach used the extensive, low frequency (generally monthly) monitoring of specific conductance across English surface waters as this gave access to many years of data (data between 2002 and 2015 were used in this study) from many sites and rivers while including catchments where shale gas development is planned.

## 2. Methodology

### *2.1. Study sites*

The study initially used specific conductance data from the 123 Harmonised Monitoring Scheme sites across England (HMS - Bellamy and Wilkinson, 2001 – Fig. 1). HMS monitoring sites were selected for inclusion into the original monitoring programme if they were at the tidal limit of rivers with an average annual discharge greater than 2 $m^3s^{-1}$, or any tributaries with a mean annual discharge above 2 $m^3s^{-1}$ (Bellamy and Wilkinson, 2001). The specific conductance of natural waters increases with temperature. This study used data for specific conductance – specific conductance is the electrical conductivity of the water sample at a set temperature, in the case of this study 25 $^oC$. Records of specific conductance for HMS sites can be paired with records of either instantaneous or average daily flow for these sites. For the purpose of this study records from 2002 to 2015 were considered. Although the main study period for this study was the decade 2003 – 2014 as records from 2002 were used to construct prior information for the statistical model and for 2015 there were incomplete flow records available meaning that data for 2015 were used for testing and validating the models developed.

On the basis of the result from the HMS sites the study was extended to include all river sites in the England sampled between 2003 and 2015 where there were 10 or more samples with the measurements made in more than one year. The sampling constraints were included to ensure that interaction terms could be estimated and to limit the quantity of data to be analysed. Only measurements from routine river monitoring and not pollution incidents were considered.

*2.2. Bayesian generalised linear modelling*

The statistical modelling was based the Bayesian approach to generalised linear modelling. Each data point (specific conductance measurement - $\kappa$) is is assumed to be generated from a particular distribution in the exponential family of distributions, the mean, $\boldsymbol{\mu}$, of the distribution depends on the independent variables, **X**, through:

$$E(\kappa) = \mu = g(X\beta) \quad \text{(i)}$$

where $E(\kappa)$ is the expected value of $\kappa$ – the specific conductance; $X\beta$ is the linear predictor, a linear combination of unknown parameters $\beta$; and g is the link function. The link function is often defined by the choice of distribution and in this case a gamma distribution was chosen. A priori, a gamma distribution has a number of advantages over other distributions, firstly, it readily approximates normal, log normal, exponential and Weibull distributions.

This flexibility means that no adjustment for values close to the limit of detection is required. Second, the gamma distribution is only defined for positive numbers and so there is no possibility that physically impossible negative values would be predicted as would be case with a normal distribution. Evidence from high frequency sampling has supported the use of a gamma distribution (Worrall et al., 2015). However, to test the appropriateness of the use of a gamma distribution the analysis of the HMS data was repeated using Weibull, normal, log normal and exponential distributions.

The form of the gamma distribution is defined as $\Gamma(\alpha,\beta)$ where $\alpha$ is commonly known as the shape factor and $\beta$ is the rate factor, and:

$$E(x) = \frac{\alpha}{\beta} \qquad \text{(ii)}$$

$$\sigma^2 = \frac{\alpha}{\beta^2} \qquad \text{(iii)}$$

Linear predictors included factors and covariates. The factors considered in this study were Site, Month and Year. The Site factor is the difference between all the monitoring sites from the HMS for which specific conductance data were available – this factor had 123 levels one for each site. The Year factor had 12 levels for each year from 2003 to 2014. The Month factor had 12 levels one for each calendar month. The two-way interactions between factors were included.

The Bayesian approach was achieved by Markov Chain Monte Carlo (MCMC) simulation to estimate the posterior distribution of the specific conductance using WinBUGS version 14 (Lunn et al., 2013). The length of the MCMC chain was 30000 cycles after a 10000 burn in cycles with samples saved every 10 cycles and with 1 chain. Model fit was tested using a number of approaches. First, that the 95% credible interval for any factor does not include zero, this is henceforward referred to as being significantly different from zero at a probability of 95%. Second, that inclusion of the factor, interaction, or covariate caused the total model deviance to decrease, and third, that the inclusion of an additional factor, interaction or covariate decreased the deviance information criterion (DIC). It is generally true that inclusion of factors, interactions or covariates will decrease the total deviance of a model as the inclusion means greater degrees of freedom for fitting and so the DIC accounts for the inclusion of more fitting parameters against the additional fit of the model.

In the Bayesian analysis a weak uninformative Jeffrey prior distribution was used whereby the expected value was set as the mean of all specific conductance from the year 2002 and the standard deviation was set as 100 times the coefficient of variation of the dataset, i.e. the prior was centred on the expected value of the data and was almost uniform in distribution. Given the size of the dataset and its spatial and temporal coverage it was deemed unnecessary or reasonable to develop a stronger prior distribution.

*2.3. Covariate information*

Covariate information was defined and developed as for Worrall et al. (2014). The CEH Wallingford digital terrain model (Morris and Flavin, 1994) was used to calculate the catchment area to each monitoring point. The CEH digital terrain model has a 50 m grid interval and a 0.1 m altitude interval. Secondly, the dominant soil-type of each 1 km$^2$ grid square classified into one of three types (mineral, organo-mineral or organic soils) based upon the system of Hodgson (1997) using nationally-available data (Smith et al., 2007). In this classification system, peat soils are classed as organic soils. Thirdly, Land use for each 1 km$^2$ of England was classified into three land uses: arable, grass and urban from the June

Agricultural Census for 2004 (Defra, 2005). The June Agricultural Census also records the number of cattle and sheep in each 1 $km^2$ and so as to provide a single measure for livestock, the equivalent sheep per hectare were calculated based on published nitrogen export values (Johnes et al., 1996) which gives a ratio of 3.1 sheep per cow. The soil and land-use characteristics for each 1 $km^2$ were summed across the catchment to each of the monitoring points and the relative proportion of different soil and land-use properties was determined.

For each of the HMS catchments for which specific conductance data were available, hydrological characteristics were available from the UK's National River Flow Archive (www.ceh.ac.uk/data/nrfa/). The characteristics used were: the base flow index (BFI), the average actual evaporation (AET) and the average annual rainfall (SAAR). The average annual total river flow for each catchment was taken as the difference between average annual rainfall and the average actual evaporation for each catchment.

The river flow at the time of sampling was available from the HMS records and was paired with the specific conductance data. Flow data, even instantaneous flow data, will be co-linear with catchment area, i.e. river flows are more likely to be larger for larger catchments and so as an alternative approach, flow records for each site were converted to the percentile flow for that site.

All covariate information was tested for normality using the Anderson-Darling test (Anderson and Darling, 1952) and log-transformed if required. To understand the importance of covariates a simple sensitivity analysis was conducted whereby a 10% increase in the average value of each significant covariate was imposed and the change in the specific conductance noted.

*2.4. Model application*

The model was considered in two stages. Firstly, to predict the specific conductance at an HMS site, i.e. a monitoring site included in the analysis. In this case the model was developed including the Site factor but without those covariates that are specific to each site and therefore would be co-linear with the Site factor. Secondly, the model was applied to predict conductance at a non-HMS site whose monitoring records were available but because the monitoring site is not part of the HMS it was not included in the first stage analysis, ie. a site not included in the original Site factor. This second analysis, therefore could not include the Site factor and so this second analysis used Year and Month as factors but considered the entire range of covariates defined for the new site.

On the basis of the results of the above a subsequent analysis included all the English sites with 10 or more data over at least two years in the period 2003 to 2015. In this third analysis the Site, Year and Month factors were used and their two-way interactions also included.

Given outputs and fit of the model were developed to consider the impact of shale gas developments and so for application and comparison sites were chosen within the one of the developing shale gas basins of the UK. Both chosen sites were selected to be the nearest available to the development sites in the Vale of Pickering (Fig. 1). The first site is an HMS monitoring site on the River Derwent at Loftsome Bridge and was included in the 123 sites in the Site factor of the initial analysis. The predicted specific conductance at this site was compared to observed conductance and then predicted for the year 2015, i.e. the subsequent. The second site of application was to a site not in the HMS monitoring network and therefore not included in the first analysis with the Site factor. The site chosen was on the Costa Beck (Fig. 1), chosen because it the monitoring site nearest to the proposed shale gas extraction site.

The purpose of this study was to create a dynamic baseline against which any influx of highly saline waters from fracking operations could be detected, therefore, the real question is what volume of fracking fluid could this approach detect at a given probability. There has only been one fracking operation conducted in the UK at Preese Hall in Lancashire (Fig. 1) and the conductivity of flowback fluid from the Preese Hall well varied from 133730 and 150614 μS/cm (Broderick et al., 2011).

No salinity or total dissolved solids (TDS) is reported within the available databases but standard relationships between salinity and specific conductance exist (Weyl, 1964)

$$Salinity = 0.000004\kappa^2 + 0.53\kappa - 201 \qquad \text{(iv)}$$

Where Salinity is in mg/l. Equation (iv) was used to convert specific conductance to values, but it should be remembered that Equation (iv) was only defined for salinity > 1000 mg/l which is equivalent to a conductance of 2200 μS/cm.

## 3. Results

### *3.1. Model development*

Between 2003 and 2014 there were 14495 measurements of specific conductance at 123 sites across England which could be paired with flow records and matched with catchment characteristics. Preliminary examination of the data showed one site should be removed (River Weaver at Frodsham) as it regularly had specific conductance over 10000 μS/cm which was not seen at any other site – the high values could simply be due to the site being too close to the tidal limit. The distribution of all results shows a bimodal distribution with peaks at 200 μS/cm and at 550 μS/cm. Fitting single gamma distribution to all the data gives $\Gamma(2.2, 282)$ which gives an expected value of specific conductance, $E(\kappa)$ = 633.5 μS/cm, with the 95% interval being 95 to 1117 μS/cm and given a freshwater limit of 1000 mg/l salinity then 0.2% of conductivity measurements exceeded this limit. The fit of this single distribution represents a base case for the prediction of specific conductance at any one site against which it is possible to judge the benefit of more complex models.

The model using only known factors (Site, Month and Year) shows that all three factors were significant (where significance is as defined above that the 95% credible interval does not contain zero) and so to were the interactions of the three factors (Table 1). It should be noted that at this stage of modelling that the deviance for models fitted using normal, log normal, exponential and Weibull distributions each lead to tot total deviance > 200000, i.e. a gamma distribution provided the best-fit. The percentile flow, when included, was significant and showed that specific conductance decreased with increasing flow which is a dilution effect with new, more rainwater-like and lower conductivity water coming in with higher flows. The inclusion of the covariates decreased the credible interval and the deviance of the model, however, the DIC did not decrease suggesting that inclusion of this additional covariate may not be justified.

Given the inclusion of all the factors and the percentile flow covariate it is now reasonable to calculate and plot the expected value of the specific conductance ($\kappa$) for each site (Fig. 2). The expected value so calculated allows for the differences in sampling times and conditions. The values do show regional differences with the lowest values in the north and the west of England and the highest values in the east and centre of the country. These regional differences may reflect underlying geology or climate differences.

When catchment covariates were included the Site factor was removed. The best-fit model is detailed in Table 2 and shows that a range of catchment characteristics are not significant in the prediction of conductivity and these are: BFI, AET, and the area of organic soils. Amongst the significant terms by far the most important was the change in flow and as flow increases the specific conductance of river water decreases and the term in flow is very close to, but still significantly different from, $-Q^{1/4}$. However, it should be noted that flow is co-linear with catchment area and rainfall, i.e. flow increases with both increased average rainfall and catchment area. River water specific conductance decreases with increasing catchment size and increasing average rainfall. The effect of flow and rainfall can be ascribed to dilution from rainfall, however, the impact of increasing catchment area is less straight forward as it might be expected that increased catchment size in the UK means that increased influence of groundwater rather than rainwater but this term may be co-linear with the river flow. The most important of the soil terms was the area of organo-mineral soils and while increasing the area of the mineral soils leads to decreased conductivity the presence of organo-mineral soils increases river water conductivity. As for land-use, the area of grassland decreased the conductivity, while increasing urban area increased conductivity; urban areas are sources of salt from roads and wastewater inputs can also increase salinity. The map in Fig. 2 cannot show the catchment area contributing to each site but the significant covariates could help explain the pattern of expected values observed in Fig. 2. Relatively low expected values of $\kappa$ are observed in the north and west of England where rainfall is higher and river flows might also be expected to be higher. The pattern with respect to land use and soil type is more complex as mineral soils dominate to the east and south and so to do arable and urban land use, i.e. competing effects of soil and land use effects on the specific conductance.

When no covariates were included, the Month factor did show a significant seasonal cycle although only three months are significantly different from zero – October, November and December - and all three led to lower specific conductance. When the covariates were included then four months were significantly different from zero; during April and July the specific conductance was significantly higher than the annual mean, while for November and December the specific conductance was significantly lower. The month factor appears to follow river flow rather than following road salt applications which would peak in the winter months.

The Year factor was significant but for most years there is no significant difference from zero and only 2007 and 2008 showing significantly lower values and 2014 showing significantly higher values. The difference between levels of the Year factor are clearly explained by including covariates which when included showed that 2004, 2005, 2007, 2008 and 2012 all show significantly lower values and only 2013 showed significantly higher values. When Year was included as a covariate rather than a factor then there was a significant role for Year as a covariate with specific conductance increasing over the time period across all sites but only by 0.01 μS/cm/yr, i.e. although significantly different from zero the trend is very small compared to other changes due to the other covariates, factors, or interactions.

*3.2. Model Application*

First, the approach was applied to the River Derwent at Loftsome Bridge, a site included in the dataset for analysis. There were 151 observations of specific conductance at Loftsome

Bridge between 2002 and 2014, and the best-fit gamma distribution across all years and months gives E($\kappa$) = 544 $\mu$S/cm and 95% credible interval of 405 to 735 $\mu$S/cm. In comparison to the observations for 2014 at Loftsome Bridge (Fig. 3) shows that all but one observation is within the credible interval suggesting that this one observation could be considered as an unusual observation. When prediction at the included site was performed, prediction for specific conductance ($\kappa$) at Loftsome Bridge for 2015, i.e. for a site included in the analysis but for a year beyond that included in the data, then the observed data was within the predicted credible interval (Fig. 4) – note that there were only 9 measurements of $\kappa$ at Loftsome Bridge in 2015. Of course, as an alternative approach to assessing the performance of the modelling the predicted values of the expected value for Loftsome Bridge in 2014 between difference models with their varying inclusion of factors, interactions and to compare to prediction of the model for specific conductance (Table 3). The comparison of models shows that it is the inclusion of all three factors with their two-way interactions that brings the results to include those observed, but the further inclusion of covariates does not improve the model prediction.

Second, the model was applied to the site at Costa Beck, i.e. a site never included in the analysis. Over the period 2002 to 2015 there were 65 observations of specific conductance with an expected value of specific conductance, E(k) = 621 $\mu$S/cm and 95% credible interval of 568 to 684 $\mu$S/cm. The results show that the model overpredicts $\kappa$ (Fig. 5), of the 20 observations at Costa Beck measured 11 were within the range predicted but of the remaining 9 observations all were lower than predicted. So whereas the model approach works well for modelling and prediction at sites which are included in the original dataset any extension to other, not previously considered, sites was not as effective. Therefore, the study extended the application to all monitoring sites in England.

*3.3. Model of all English monitoring sites*

In total there were 6833 river monitoring sites which met the criteria (Fig. 6) and plotting the calculated expected values (E($\kappa$)) shows a tendency of increasing E($\kappa$) from west to east across England and perhaps also from north to south, but the largest values of E($\kappa$)) are not in the south east corner of England but in more central areas of England and especially rivers entering the Wash. This tendency across England perhaps follows gradients in climate from the wetter western and more mountainous areas of the west and north towards drier, lowland areas of eastern England. Furthermore, the tendency for higher E($\kappa$)) to eastern England also seems to follow geology with more permeable and younger geology occurring in east compared to the west. The map in Fig. 6 also shows that other potential sources of high salinity water are not important. For example, it might expected that urban conurbations with their high density of major roads, which would be salted in winter, would represents “hot spots” of specific conductance, but the major English conurbations are not visually obvious in Fig. 6. Furthermore, areas of the UK with worked salt deposits (Cheshire, north-west England) do not show up as “hot spots” of specific conductance in Fig. 6.

Application of the model from all English monitoring sites to the specific conductance data for Costa Beck shows that rather than a systematic overprediction the results now show only three observations were overpredicted but none were underpredicted (Fig. 7).

*3.4. Model sensitivity*

With respective to sensitivity then it is true for a volume of incoming high salinity water could be detected if:

$$\frac{Q_f}{Q_r} = \frac{(\kappa_r^{max} - \kappa_r)}{(\kappa_f - \kappa_r^{max})} \approx \frac{(\kappa_r^{max} - \kappa_r)}{\kappa_f} \quad \text{(iv)}$$

Where: $Q_x$ = the discharge due to the river (r) or from fracking (f) – $m^3$/day; $\kappa_x$ = specific conductance for the river (r) and for the fluid from the fracking operation (f) – μS/cm; and $\kappa_r^{max}$= the maximum specific conductance predicted for the river – μS/cm. Given that $\kappa_f$ >>> $\kappa_r^{max}$ the denominator simplifies. For the Preese Hall well flowback fluid and the river discharge recorded at Loftsome Bridge in 2014 shows that in this case there was a 95% probability of being able to detect as little as 272 $m^3$/day in February 2014 but this rose in wetter winter months to as high as 745 $m^3$/day (Fig. 8). The volume of fracturing fluid used varies depending on the shale-play, the operator, well depth, the number of fracturing stages and the length of the wells (Nicot and Scanlon, 2012). The European Parliament summarised the US literature on the volume of water required per well and found the volume ranged from 1500 to 45000 $m^3$ (Clancy et al., 2018), whilst Jiang et al. (2014) note that the average Marcellus well consumes 20000 $m^3$ (with a range from 6700 to 33000 $m^3$) of freshwater per well over its lifetime. The single well drilled in the UK at Preese Hall (Lancashire) required 8400 $m^3$ of water. Taylor et al. (2013) when considering the scenarios for the development of a UK shale gas industry considered the development of a 10-well pad of 10 laterals which would require 136000 $m^3$ of water per well. Initially it is likely that the water required will be trucked to the site rather than piped, thus requiring between 2856 and 7890 trucks over a 20 year period with truck movements concentrated in to the first two years at between 3.9 – 10.8 truck movements per day during phases of site development and production. Given the volume that a single truck can transport (30 $m^3$)

means that a site might need storage for approximately 600 $m^3$ of water, i.e. two days worth of truck movements at maximum predicted number of trucks. Therefore, the alternative question to ask is how small a river would need to be monitored in order to give a defined chance of detecting a leak or spill? Applying Equation (iv) to calculate $Q_r$ given the values of $\kappa_r$ for Loftsome Bridge in 2014 and the range of values of $\kappa_f$ observed for Preese Hall flowback fluid and a $Q_f$ of between 30 and 600 $m^3$/day means that for a 97.5% probability of detecting leaks with river flow of 0.6 and 1 $m^3$/s (Fig. 9). Given the catchment characteristics used as covariates in this study an average flow of 1 $m^3$/s would be true in the UK for catchments of less than 9 $km^2$.

The approach above assumes the water quality problem arises from an acute incident of spill or leakage to surface water and not a chronic seepage of contaminated fluids from depth to surface. Osborn et al. (2011) reported that contamination of shallow groundwater overlying the Marcellus shale resulted from poor well integrity in the shale gasfields, while Warner et al. (2014) reported no such contamination for shallow groundwater overlying the Fayetteville shale in Arkansas and Wilson et al. (2017) showed that contamination from the shale layers was extremely unlikely for the UK's Bowland shale.

## 4. Discussion

This study has developed a consistent and coherent approach to the use of conductivity monitoring data. The Bayesian approach uses all available data to predict distributions at sites of interest. For determinands with defined environmental quality standards (eg. water framework directive – EC Directive, 2000) individual results are viewed relative to these standards while for other determinands (eg. specific conductance) even such comparisons may not occur as no legal standard exists. Furthermore, the review period for water quality monitoring is not always clear, under an operators permit the operator should review continuously, i.e. data reviewed each time new data is produced and the regulator informed if there is an issue. The regulator in the UK may be asked to report at anytime to the Secretary of State at the highest government level, but how often this occurs is not clear. In the approach used here each datum can be viewed against a prediction that is based upon all available information and this can be viewed in a probabilistic framework, i.e. what is the probability that a new observation is exceptional and not what should be expected. In the case of used here measured specific conductance was judged against a predicted distribution as a means of testing whether an exceptional has or has not occurred. But equally we can use the predicted distribution to assess the probability that an environmental standard has been breached, for example in the case of specific conductance what would be the probability that the stream has a salinity > 1000 mg/l ($\kappa >$ 2270 $\mu$S/cm).

In effect this approach has built up a method to improve assessment at any one site. At the simplest level one could examine the distribution of observed data at any site and compare the latest observation with that distribution. But that would not be a fair comparison because a local interannual variation might mean that comparing one observation with data from all years would be inappropriate, i.e. there is a interannual trend at site which values in the current year would tend to be lower than those in a previous year; thus a distribution for the given year would be better than comparing with data from all years. Equally there could be expected to be an intra-annual cycle in values and so even grouping observation by year would be misleading as some months would naturally be expected to have higher values than others. So including a measure of intra-annual cycle (eg. month) would improve the distribution for comparison. But of course it is unlikely that there will be sufficient observations to give such a reasonable distribution for any month for any year and any one site or indeed enough observations for any site and so it would be if information from other sites could be drawn open: this then is what this approach has achieved. By using all available information the approach here estimate a distribution of observations for every month, for every year at each site. An analogous, non-Bayesian approach might be that of weighted regression analysis (Hirsch et al., 2010, 2015),

The approach could improve with the use of further covariates. The study has considered a range of covariates but in most cases covariates were surrogates for site information (eg. catchment area or land use). Within the HMS dataset it was possible to include river flow but this was not possible at all sites simply because in this dataset there are only 677 sites which are co-located with river flow gauging stations. However, as data has been chosen from water quality monitoring sites there would be other water quality parameters measured at these sites which may provide additional, covariate information. Specific conductance could be expected to co-vary with some cations and anions but equally the compositions of hydraulic fracking fluid may lead to use of other water quality parameters with a reasonably high degree of specificity for pollution incidents from unconventional hydrocarbon operations. Further, the analysis could become multi-dimensional, i.e. a further determinand could be to the analysis. Johnson et al. (2015) have suggested that sources of brine in areas of unconventional hydrocarbon extraction could be distinguished bu use of Cl/Br ratio; Sr isotopes or the ratio (Ba + Sr)/Mg. Indeed, Wilson and Van Briesen (2013) used Cl/Br ratios to detect shale gas fluids in surface waters of the Mononghela river in Pennsylvania. However, all three of these fail the criteria outlined in this study for a good being a good sentinel if for no other reason than they are not regularly measured.

The approach proposed here could be applied to the majority of data from water quality monitoring. Even in a focused network of monitoring sites such as may be used within the context of a developing shale gas industry there is no criteria for assessing whether pollution has or is occurring. For example, Krogulec and Sawicka (2015) discuss groundwater monitoring in Poland for the impacts of shale gas development but at no point suggest numbers of monitoring points or frequency of sampling. Niu et al. (2018) proposed a change point analysis upon water quality time series in streams from areas of unconventional hydrocarbon exploitation. Loomer et al. (2018) used a higher frequency sampling of groundwater in area of Canada to determine the appropriate sampling frequency for monitoring unconventional hydrocarbon exploitation. Austen et al. (2017) suggest that unconventional hydrocarbon operations in the Fayetteville Shale had no impact on surface water quality on the basis of trends solely recorded after the unconventional hydrocarbon well pads had been installed and did not formally compare to any control. Down et al. (2015) have published a baseline geochemical assessment of the Triassic basin of North Carolina, a prospective shale gas basin at the time of the study, however the study provides no suggestion as to how these results might be used to assess any impact of a shale gas industry. Alternatively, Werner et al. (2013), Darrah et al. (2014) and Hildenbrand et al. (2015) have provide extensive water quality surveys of Arkansas' Fayetteville shale; Marcellus shale and the Barnett shale of Texas respectively, but in each case the surveys were after shale gas had been exploited in the area for many years. However, Hildenbrand et al. (2016) did consider the change in groundwater quality with the development of unconventional hydrocarbon resources in the Permian Basin of Texas and the sampling started before shale gas had been extracted in the majority of the area.

The approach developed and tested provides a number of clear advances over the current situation:

i) This is a systematic transparent approach to analysing data and provides a probability, with uncertainty, as to the nature of any observed data. Thus in turn the probability that any pollution has, or has not, happened can be assessed.

ii) The approach makes use of all available information and so the approach gains value from the whole monitoring network, i.e. maximum information is gained from the current, past and ongoing monitoring. This approach, therefore, gives good value for the money invested in environmental monitoring.

iii) All risk assessment is actual a probability statement and the tools here use Bayesian approaches so all results will be a probability and with an uncertainty.

iv) The Bayesian framework means that the tool automatically updates and so contributes to the development of a dynamic baseline in time and space.

v) The approach proposed can be used to assess information content and informational efficiency of the current monitoring network monitoring.

In regions of especial interest or concern with respect to shale gas extraction it would be easy for industry or regulators to place a water quality sonde in a local waterway to produce quasi continuous records of water quality and especially conductivity. Indeed, conductivity is the most commonly measured water quality parameter on such sondes (Halliday et al., 2012). Unlike for spot sampling in-situ water quality sondes are subject to damage and vandalism and must be maintained and calibrated in-situ. Son et al. (2015, 2018) have proposed the use of in-situ water quality sondes down borehole in areas of active hydraulic fracturing in northern Colorado to monitor for pollution events. The problem of interpretation would be equally true for high frequency as for low frequency data obtained from spot sampling, i.e. a coherent framework for assessing the probability that a pollution event had or was happening would still be required and an expectation of what baseline conditions represent natural would still need to be constructed. The United States Environmental Protection Agency have developed a system for working with real-time, quasi-continuous data for the detection of pollution events (CANARY - USEPA 2012b). Quasi-continuous data could be readily incorporated into the approach presented here and analysis with the network of existing data providing informative prior information within the Bayesian framework proposed. Furthermore, such quasi-continuous records have been viewed by many authors as perfect information and so in comparison to results from less frequent spot sampling it would be possible to judge the value of perfect information relative to low frequency sampling (Worrall et al.,2013).

## 5. Conclusions

The study has developed a Bayesian generalised linear modelling approach to understanding specific conductance in English river waters. We could model specific conductance at river sites down to the natural variation at the monthly time step. The model could predict at sites included in the analysis but did not work well within the currently available covariates to predict at unknown sites. The model was extended to 6883 sites across England and this enabled our approach to predict a monthly distribution at any of these sites. The approach can be used to assess whether an observation is unusual against a regulatory standard or by predicting a distribution at each point of time at a point of interest the regulator could set their own criteria more appropriate for the local activity being monitored. The model shows that most rivers could readily absorb leaks of fracking fluids due to low volume of daily use on a single well pad. We propose that this approach could provide a coherent and consistent approach to analyzing water quality data while enhanced use of all available data.

**Acknowledgements**

The data for this study was all taken from the Environment Agency's WIMS database and all of which is available freely from the UK government. This research was carried out as part of the ReFINE research consortium led by Newcastle and Durham Universities. ReFINE has been funded by Ineos, Shell, Chevron, Total, GDF Suez, Centrica and NERC. We thank the ReFINE Independent Science Board for prioritising the research projects undertaken by ReFINE.

## References

Alessi, D.S., Zlofarghari, A., Kletke, S., Gehman, J., Allen, D.M., Goss G.G., 2016. Comparative analysis of hydraulic fracturing wastewater practices in unconventional shale development: water sourcing, treatment and disposal practices. Canadian Water Resources Journal 42(2), 122-137.

Almond, S., Clancy, S.A., Davies, R.J., Worrall, F., 2014. The flux of radionuclides in flowback fluid from shale gas exploitation. Environmental Science & Pollution Research 21(21), 12316-12324.

Anderson, T. W., Darling, D.A., 1952. Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. Annals of Mathematical Statistics 23, 193–212.

Austin, B.J., Scott, E., Massey, L., Evans-White, M.A., Entrekin, S., Haggard, B.E., 2017. Unconventional natural gas development did not result in detectable changes in water chemistry (within the South Fork Little Red River). Environmental Monitoring and Assessment 189(5), Art. No. 209

Bellamy, D., Wilkinson, P., 2001. OSPAR 98/3: an environmental turning point or a flawed decision? Marine Pollution Bulletin 49, 87-90.

Blewett, T.A., Delompre, P.L.M., He, Y.H., Folkerts, E.J., Flynn, S.L., Alessi, D.S., Goss, G.G., 2017. Sublethal and Reproductive Effects of Acute and Chronic Exposure to Flowback and Produced Water from Hydraulic Fracturing on the Water Flea Daphnia magna. Environmental Science & Technology 51(5), 3032-3039.

Broderick. J., Wood, R., Gilbert, P., Sharmina, M., Anderson, K., Footitt, A., Glynn, S., Nicholls, F., 2011. Shale gas: an updated assessment of environmental and climate change impacts. A report commissioned by The Co-operative and undertaken by researchers at the Tyndall Centre, University of Manchester.

Clancy, S.A., Worrall, F., Davies, R.J., Gluyas, J.G., 2018. The potential for spills and leaks of contaminated liquids from shale gas developments. Science of the Total Environment 626, 1463 – 1473.

Darrah, T.H., Vengosh, A., Jackson, R.B., Warner, N.R., Poreda, R.J., 2014. Noble gases identify the mechanisms of fugitive gas contamination in drinking-water wells overlying the Marcellus and Barnett Shales. Proceedings of the National Academy of Sciences of the United States of America 111(39), 14076-14081

DEFRA, 2005. Agriculture in the United Kingdom - 2004. Department of Environment, Food and Rural Affairs, HMSO, London, 2005.

Down, A., Schreglmann, K., Plata, D.L., Elsner, M., Warner, N.R., Vengosh,A., Moore, K., Coleman, D., Jackson, R.B., 2015. Pre-drilling background groundwater quality in the Deep River Triassic Basin of central North Carolina, USA. Applied Geochemistry 60, 3-13.

EC Directive, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal L 327 , 22/12/2000 P. 0001 – 0073

Environment Agency, 2011. Shale Gas. North West – Monitoring of Flow back water. webarchive.nationalarchives.gov.uk/20140328145127

Halliday, S. J., Wade, A.J., Skeffington, R.A., Neal, C., Reynolds, B., Rowland, P., Neal, M., Norris, D., 2012. An analysis of long-term trends, seasonality and short-term dynamics in water quality data from Plynlimon, Wales. Sc*ience of the Total Environment*, 434, 186-200.

He, Y., Flynn, S.L., Folkerts, E.J., Zhang, Y., Ruan, D.L., Alessi, D.S., Martin, J.W., Goss, G.G., 2017. Chemical and toxicological characterizations of hydraulic fracturing flowback and produced water. Water Research 114, 78-87.

Hildenbrand, Z.L., Carlton, D.D., Fontenot, B.E., Meik, J.M., Walton, J.L., Taylor, J.T., Thacker, J.B., Korlie, S., Shelor, C.P., Henderson, D., Kadjo, A.F., Roelke, C.E., Hudak, P.F., Burton, T., Rifai, H.S., Schug, K.A., 2015. A Comprehensive Analysis of Groundwater Quality in The Barnett Shale Region. Environmental Science & Technology 49(13), 8254-8262.

Hildenbrand, Z.L., Carlton, D.D., Fontenot, B.E., Meik, J.M., Walton, J.L., Thacker, J.B., Korlie, S., Shelor, C.P., Kadjo, A.F., Clark, A., Usenko, S., Hamilton, J.S., Mach, P.M., Verbeck, G.F., Hudak, P., Schug, K.A., 2016. emporal variation in groundwater quality in the Permian Basin of Texas, a region of increasing unconventional oil and gas development Science of the Total Environment 562, 906-913.

Hirsch, R.M., Moyer, D.L., Archfield, S.A., 2010. Weighted regression on time, discharge and season (WRTDS), with application to Chesapeake Bay river inputs. Journal of the American Water Resources Association 46(5), 857-880

Hirsch, R.M., Archfield, S.A., De Cicco, L.A., 2015. A bootstrap method for estimating uncertainty of water quality trends. Environmental Modelling & Software 73, 148-166.

Hodgson, J.M., 1997. Soil Survey Field Handbook: Describing and Sampling Soil Profiles. Soil survey Technical Monograph No. 5. Soil Survey and Land Research Centre, Silsoe. England.

Jiang, M., Hendrickson, C.T., VanBriesen, J.M., 2014. Life cycle water consumption and wastewater generation impacts of a Marcellus shale gas well. Environmental Science & Technology 48(4), 1911-1920.

Johnes, P., Moss, B., Phillips, G., 1996. The determination of total nitrogen and total phosphorus concentrations in freshwaters from land use, stock headage and population data: testing of a model for use in conservation and water quality management. Freshwater Biology 36, 451-473.

Johnson, J.D., Graney, J.R., Capo, R.C., Stewart, B.W., 2015. Identification and quantification of regional brine and road salt sources in watersheds along the New York/Pennsylvania border, USA. Applied Geochemistry 60, 37-50.

Kahrilas, G. A., Blotevogel, J., Stewart, P.S., Borch, T., 2014. Biocides in hydraulic fracturing fluids: A critical review of their usage, mobility, degradation, and toxicity. Environmental Science & Technology 49(1), 16-32.

Kohl, C.A.K., Capo, R.C., Stewart, B.W., Wall, A.J., Schroeder, K.T., Hammack, R.W., Guthrie, G.D., 2014. Strontium Isotopes Test Long-Term Zonal Isolation of Injected and Marcellus

Formation Water after Hydraulic Fracturing. Environmental Science & Technology 48(16), 9867-9873.

Krogulec, E., Sawicka, K., 2015. Groundwater protection in shale gas exploration areas - a Polish perspective. Episodes 38(1), 9-20.

Lagace, F., Foucher, D., Surette, C., Clarisse, O., 2018. Radium geochemical monitoring in well waters at regional and local scales: an environmental impact indicator-based approach. Chemosphere 205, 627-634.

Loomer, D.B., MacQuarrie, K.T.B., Al, T.A., Bragdon, I.K., Loomer, H.A., 2018. Temporal variability of dissolved methane and inorganic water chemistry in private well water in New Brunswick, Canada. Applied Geochemistry 94, 53-66.

Lunn, D., Jackson, C., Best, N., Thomas, A., Speigelhalter, D., 2013. The BUGS book – A practical introduction to Bayesian analysis. CRC Press, Abingdon, UK.

Morris, D.G., Flavin, R.W., 1994. Sub-set of UK 50m by 50m hydrological digital terrain model grids. NERC, Institute of Hydrology, Wallingford.

Nicot, J.P., Scanlon, B.R., 2012. Water use for shale-gas production in Texas, US. Environmental Science & Technology 46(6), 3580-3586.

Niu, X.Z., Wendt, A., Li, Z.H., Agarwal, A., Xue, L.Z., Gonzales, M., Brantley, S.L., 2018. Detecting the effects of coal mining, acid rain, and natural gas extraction in Appalachian basin streams in Pennsylvania (USA) through analysis of barium and sulfate concentrations. Environmental Geochemistry and Health 40(2), 865-885.

Osborn, S.G., Vengosh, A., Warner, N.R., Jackson, R.B., 2011. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. Proceedings of the National Academy of Sciences of the United States of America 108(20), 8172-8176

Rowan, E.L., et al. (2011). Radium Content of Oil- and Gas-Field Produced Waters in the Northern Appalachian Basin (USA): Summary and Discussion of Data. Scientific Investigations Report, US Geological Survey 2011-5135.

Son, J.H., Carlson, K.H., 2015. Real-time surrogate analysis for potential oil and gas contamination of drinking water resources. Applied Water Science 5(3), 283-289.

Son, J.H., Hanif, A., Dhanasekar, A., Carlson, K.H., 2018. Colorado Water Watch: real-time groundwater monitoring for possible contamination from oil and gas activities. Environmental Monitoring and Assessment 190(3), Art. No 138

Taylor, C., Lewis, D., Byles, D., 2013. Institute of Directors (Infrastructure for Business) report: Getting shale gas working. Institute of Director, London, UK.

Teasdale, C.J., Hall, J.A., Martin, J.P., Manning, D.A.C., 2014. Ground Gas Monitoring: Implications for Hydraulic Fracturing and $CO_2$ Storage. Environmental Science & Technology 48(23), 13610-13616.

United States Environmental Protection Agency, 2012. CANARY user's manual version 4.3.2. EPA 600/R-08/040B.

Vengosh, A., Jackson, R. B., Warner, N., Darrah, T. H., Kondash, A., 2014. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. Environmental Science & Technology 48(15), 8334-8348.

Warner, N.R., Kresse, T.M., Hays, P.D., Down, A., Karr, J.D., Jackson, R.B., Vengosh, A., 2013. Geochemical and isotopic variations in shallow groundwater in areas of the Fayetteville Shale development, north-central Arkansas. Applied Geochemistry 35, 207-220.

Weyl, P.K., 1964. On the change in electrical conductance of seawater with temperature. Limnology & Oceanography 9(1), 75–78.

Wilson, J.M., Van Briesen, J.M., 2013. Source Water Changes and Energy Extraction Activities in the Monongahela River, 2009-2012. Environmental Science & Technology 47(21), 12575-12582.

Wilson, M.P., Worrall, F., Davies, R.J., Hart, A., 2017. Shallow Aquifer Vulnerability From Subsurface Fluid Injection at a Proposed Shale Gas Hydraulic Fracturing Site. Water Resources Research 53(11), 9922-9940

Worrall, F., Howden, N.J.K., Burt, T.P., 2013. Assessment of sample frequency bias and precision in fluvial flux calculations – an improved low bias estimation method. Journal of Hydrology 503, 101-110.

Worrall, F., Burt, T.P., Howden, N.J.K., 2014. The fluvial flux of particulate organic matter from the UK: Quantifying in-stream losses and carbon sinks. Journal of Hydrology 519, 611-625.

Worrall, F., Howden, N.J.K., Burt, T.P., 2015. Understanding the diurnal cycle in fluvial dissolved organic carbon - The interplay of in-stream residence time, day length and organic matter turnover. Journal of Hydrology 523, 830-838.

Fig. 1. Location of the Harmonised monitoring scheme (HMS) sampling sites used in this study including the chosen sites within The Vale of Pickering (River Derwent at Loftsome Bridge; and Costa Beck) as well as the site at Preese Hall.

Fig. 2. Maps of: a) the expected mean (E($\kappa$)); b) the 97.5$^{th}$ percentile; and c) the 2.5$^{th}$ percentile of the specific conductance ($\kappa$).

Fig. 3. The comparison of the predicted and observed specific conductance for Loftsome Bridge (River Derwent) in 2014.

Fig. 4. The comparison of the predicted and observed specific conductance for Loftsome Bridge (River Derwent) in 2015.

Fig. 5. The comparison of the predicted and observed specific conductance for Costa Beck based upon model from HMS data.

Fig. 6. Maps of: a) All English stream and river water sites with sufficient data to be included in this study; and b) the expected mean (E($\kappa$)).

Fig. 7. The comparison of the predicted and observed specific conductance for Costa Beck using the model based upon data from all English monitoring sites.

Fig. 8. The detectable volume of fracking discharge (a leak of any of the possible high salinity fluid from the well pad) predicted at Loftsome Bridge.

Fig. 9. The flow required to detect a typical volume stored within a single well pad.

Table 1. The details of model fit with increasing introduction of factors, their interactions and inclusion of Year and percentile flow (%flow) as covariates.

| Factors | | | Interactions | Covariates | | Deviance | DIC |
|---|---|---|---|---|---|---|---|
| Site | Month | Year | | Year | Log(%flow) | | |
| Observed | | | | | | | |
| x | | | | | | 17772 | 17773 |
| x | x | | | | | 17690 | 17770 |
| x | x | | x | | | 17590 | 17773 |
| x | x | x | | | | 17650 | 17470 |
| x | x | x | x | | | 17373 | 17630 |
| x | x | x | x | | x | 17270 | 17530 |
| x | x | | x | x | x | 17200 | 15500 |

Table 2. The coefficient of those covariates found to be significant and the sensitivity of the prediction of specific conductance to a 10% increase in the average value.

| Covariate | Mean | 2.5% | 97.5% | Average | Sensitivity (μS/cm) |
|---|---|---|---|---|---|
| LogQ | -0.23 | -0.24 | -0.22 | 4.46 $m^3/s$ | -14.4 |
| Area | -0.00016 | -0.0002 | -0.00011 | 146 $km^2$ | -0.95 |
| Aver. rainfall | -0.0016 | 0.0018 | -0.0014 | 1369 mm | -8.7 |
| Mineral soil | -0.00016 | 0.0022 | 0.00009 | 28.2 $km^2$ | -0.18 |
| Organo-mineral soils | 0.0007 | 0.0046 | 0.00088 | 95.4 $km^2$ | 2.95 |
| Arable | 0.00029 | 0.00012 | 0.00047 | 10.4 $km^2$ | 0.12 |
| Grass | -0.0003 | -0.00047 | -0.00014 | 78.5 $km^2$ | -1.0 |
| Urban | 0.026 | 0.0022 | 0.003 | 5.5 $km^2$ | 0.6 |
| Constant | 6.02 | 5.97 | 6.07 | | |

Table 3. The application of the derived models to predict the distribution of specific conductance at Loftsome Bridge, River, Derwent, 2015.

| Factors | | | Interactions | Covariates | | Predicted | | |
|---|---|---|---|---|---|---|---|---|
| Site | Month | Year | | Year | Log(%flow) | Mean | 2.5% | 97.5% |
| | | | | | | 633 | 95 | 1117 |
| x | | | | | | 543 | 526 | 568 |
| x | x | | | | | 546 | 523 | 571 |
| x | x | | x | | | 545 | 474 | 629 |
| x | x | x | | | | 535 | 510 | 562 |
| x | x | x | x | | | 616 | 508 | 744 |
| x | x | x | x | | x | 617 | 513 | 739 |
| x | x | x | x | x | x | 612 | 510 | 732 |
| Observed | | | | | | 606 | 571 | 643 |

Research Highlights

1) There is a need for effective baselines against which impacts can be assessed
2) For the impacts of shale gas, specific conductance is an ideal quality sentinel
3) A Bayesian model was used to predict specific conductance for surface waters
4) The model built upon spot-sampled data from 2005 to 2015 for 6833 across England
5) The approach assesses how exceptional of any observation would be.