A dynamic baseline for dissolved methane in English groundwater

M.P. Wilson, F. Worrall, R.J. Davies, A. Hart

PII: S0048-9697(19)34846-6
DOI: https://doi.org/10.1016/j.scitotenv.2019.134854
Reference: STOTEN 134854

To appear in: *Science of the Total Environment*

Received Date: 13 June 2019
Revised Date: 3 October 2019
Accepted Date: 4 October 2019

Please cite this article as: M.P. Wilson, F. Worrall, R.J. Davies, A. Hart, A dynamic baseline for dissolved methane in English groundwater, *Science of the Total Environment* (2019), doi: https://doi.org/10.1016/j.scitotenv.2019.134854

A dynamic baseline for dissolved methane in English groundwater

M. P. Wilson[a]*, F. Worrall[a], R. J. Davies[b], and A. Hart[c]

[a] Department of Earth Sciences, Durham University, Science Labs, Durham, DH1 3LE, UK

[b] School of Natural and Environmental Sciences, Newcastle University, Newcastle, NE1 7RU, UK

[c] Environment Agency, Research Assessment and Evaluation, Sapphire East, Streetsbrook Road, Solihull, B91 1QT, UK

* Corresponding author: Miles Wilson (miles.wilson@durham.ac.uk)

Abstract

Elevated dissolved methane (CH_4) concentrations in groundwater are an environmental concern associated with hydraulic fracturing for shale gas. Therefore, determining dissolved CH_4 baselines is important for detecting and understanding any potential environmental impacts. Such baselines should change in time and space to reflect ongoing environmental change and should be able to predict the probability that a change in dissolved CH_4 concentration has occurred. We considered four dissolved CH_4 concentration datasets of English groundwater using a Bayesian approach: two national datasets and two local datasets from shale gas exploration sites. The most sensitive national dataset (the previously published British Geological Survey CH_4 baseline) was used as a strong prior for a larger (2153 measurements compared to 439) but less sensitive (detection limit 1000 times higher) Environment Agency dataset. The use of the strong prior over a weak prior improved the precision of the Environment Agency dataset by 75%. The expected mean dissolved CH_4 concentration in English groundwater based on the Bayesian approach is 0.24 mg/l, with a 95% credible interval of 0.11 to 0.45 mg/l, and a Weibull distribution of W(0.35±0.01, 0.34±0.16). This indicates the amount of CH_4 degassing from English groundwater to the atmosphere equates to between 0.7 to 3.1 kt CH_4/year, with an expected value of 1.65 kt CH_4/year and a greenhouse gas warming potential of 40.3 kt CO_{2eq}/year. The two local monitoring datasets from shale gas exploration sites, in combination with the national datasets, show that dissolved CH_4 concentrations in English groundwater are generally low, but locations with concentrations greater than or equal to the widely used risk action level of 10.0 mg/l do exist. Statistical analyses of groundwater redox conditions at these locations suggest that it may be possible to identify other locations with dissolved CH_4 concentrations ≥10.0 mg/l using redox parameters such as Fe concentration.

Keywords

Aquifer, Bayesian, CH_4, Fracking, Hydraulic fracturing, Shale gas

1. Introduction

Methane (CH_4) dissolved in groundwater has no known health effects to humans or animals when consumed. However, gaseous CH_4 is known to be a potent greenhouse gas [*IPCC*, 2013], can pose as an asphyxiation hazard in confined spaces, and can create an explosion hazard when concentrations reach 5% by volume in air [*Hooker and Bannon*, 1993]. Because of these hazards and climate change effects, the degassing of CH_4 from groundwater, and therefore elevated dissolved CH_4 concentrations, have become a topic of environmental concern associated with the hydraulic fracturing (fracking) of unconventional reservoirs, such as shale, to extract commercial quantities of CH_4 [*Vengosh et al.*, 2013]. Concerns primarily stem from the United States of America (USA) where there has been considerable debate on whether elevated dissolved CH_4 concentrations, widely publicized by the "Gaslands" film, are naturally occurring or the result of stray gas from fracking operations [*Llewellyn et al.*, 2015; *Molofsky et al.*, 2013; 2016; *Osborn et al.*, 2011]. Further concerns may arise from historic industry-related incidents such as the CH_4 explosions at Abbeystead (in 1984) and Loscoe (in 1986) as a result of water pumping and landfill waste, respectively [*Exploration Consultants Ltd.*, 1986; *Williams and Aitkenhead*, 1991]. To assess and demonstrate that an industry has impacted an environment it is necessary to show, within a reasonable level of certainty, that an indicator of concern has changed environmental state over and above that which was true without the industrial activity present. Alternatively, that the industrial activity has changed an environmental state beyond some accepted minimum level of harm. Experience from the USA shale gas industry has highlighted that demonstrating the impact, or the ability to confirm the absence of an impact, of fracking on dissolved CH_4 concentrations in groundwater requires that a baseline, or pre-intervention control, needs to be established for subsequent observations.

England has a nascent shale gas industry and although a number of companies are developing plans to exploit shale gas reserves, only two sites have been hydraulically fractured to date. In 2012 the Royal Society and the Royal Academy of Engineering recognized the need to develop a dissolved CH_4 baseline prior to shale gas development [*Mair et al.*, 2012]. In 2017 the British Geological Survey (BGS) contributed to establishing this baseline by publishing dissolved CH_4 concentrations from a range of aquifers across England, Scotland and Wales [*Bell et al.*, 2017]. *Bell et al.* [2017] built on previous dissolved CH_4 sampling results [*Ó Dochartaigh et al.*, 2001; *Gooddy and Darling,* 2005; *Darling and Gooddy,* 2006] to increase the number of BGS sampling locations across Great Britain to 343 sites. Approximately 30% of the 343 sites comprised of private and Environment Agency of England (EA) monitoring sites. However, the EA have numerous other monitoring sites where dissolved CH_4 concentrations in groundwater have been measured and these data are publically available online from the year 2000. These additional dissolved CH_4 concentration measurements were not included in the BGS baseline because of their limited geographical coverage and considerably higher detection limit; and the subsequent difficulty in combining them with the newly acquired BGS CH_4 concentration data [*Rob Ward,* personal communication, 2018]. Therefore, there is a need to create methods which can bring together dissolved CH_4 datasets with different detection limits into a single coherent and consistent framework to make maximum use of all the data available while providing a probabilistic, and thus risk-based, assessment of any impact.

In this study generalised linear modeling within a Bayesian framework [*Gelman et al.*, 2004] is used to combine the BGS and EA dissolved CH_4 concentration data and develop a dynamic baseline for English groundwater. We also collate and incorporate dissolved CH_4 concentrations from local baseline monitoring at two shale gas exploration sites. This approach is entirely data driven without the need for the parameterisation required in physical models; it is flexible with respect to the distribution chosen to represent the data and can include factors (e.g. location) and covariate information. The Bayesian framework allows differing data sources to be used; one dataset being the prior information to inform the analysis of the next, with the result (the posterior) becoming the prior information for the subsequent dataset. In this way the data are brought together in one framework that can be updated and improved with time, i.e. it provides a dynamic baseline in time and space. Equally, it means that historic data, derived for a variety of purposes, or data from other locations, can be used to strengthen the current monitoring at the current locations of interest: in a Bayesian framework all information has value. Furthermore, all the outputs from a Bayesian analysis come as a probability distribution which means that risk can be assessed. Such a baseline can generate a time series of expected dissolved CH_4 concentrations and indicate the likelihood of an unusually elevated dissolved CH_4 concentration occurring. To further extend our analysis we also investigate whether groundwater redox conditions can be used to predict locations in England with elevated dissolved CH_4 concentrations. These redox conditions may be easier and less expensive to monitor than dissolved CH_4 and may thus better inform where and when more expensive dissolved CH_4 analysis is required.

2. Approach and Methodology

To combine four datasets of dissolved CH_4 concentrations in groundwater with differing detection limits a Bayesian generalised linear modeling approach was taken [*Gelman et al.*, 2004]. This approach uses the concentration distribution of one dataset to inform the sub-detection limit distribution of the other datasets. The output is also probabilistic and can therefore be used as a predictor of likely dissolved CH_4 concentrations. To investigate if groundwater redox conditions could be used to predict locations in England with elevated dissolved CH_4 concentrations, statistical analyses of binomial probability modeling and analysis of variance (ANOVA) were carried out on locations observed to have elevated dissolved CH_4 concentrations.

2.1. Dissolved CH_4 concentration datasets

2.1.1. BGS dataset

The BGS dissolved CH_4 concentration data (herein termed the BGS dataset) were downloaded from the supplementary material of *Bell et al.* [2017]. The dataset contains 439 dissolved CH_4 concentration measurements across 336 unique locations in Great Britain (Fig. 1, Table S1). The detection limit for the dataset was ~0.5 μg/l, although in some cases measured dissolved CH_4 concentrations were lower due to favourable gas chromatography conditions during batch analysis [*Bell et al.*, 2017].

2.1.2. EA dataset

Dissolved CH_4 concentrations (in mg/l) in English groundwater were downloaded from the EA Water Quality Archive [*EA,* 2019a] for the years 2000-2017, inclusive. Only those entries with identifiable grid references and location names were used - this dataset is hereafter referred to as the EA dataset (Table S2). Groundwater monitoring was undertaken by the EA for a range of purposes, most commonly as planned investigations and routine monitoring (Table S3). Measurements taken for statutory failures and unplanned reactive monitoring (pollution incidents) were not removed from the dataset because although certain water quality determinands would be expected to be elevated after a pollution incident, CH_4 is not classified as a groundwater pollutant and therefore dissolved CH_4

measurements may simply have been taken opportunistically when investigating the cause of the incident (we cannot prove otherwise given the available information). Waste monitoring measurements, which might show elevated CH_4 concentrations due to the biogenic breakdown of anthropogenic waste, were also retained in the dataset because these are still important for defining the current status of dissolved CH_4 in groundwater prior to potential widespread shale gas exploitation. The EA dataset contains 2153 dissolved CH_4 concentration measurements from 571 unique locations in England (Fig. 1; Table S2). Approximately 87% of the dissolved CH_4 concentration measurements and ~92% of the unique locations are located in the northwest of England (Cumbria, Lancashire, Greater Manchester, Merseyside and Cheshire) (Fig. 1; Table S4). This geographically focussed sampling is a result of increased dissolved CH_4 sampling in response to the 1984 Abbeystead explosion [*Rob Ward*, personal communication, 2018]. The source of the CH_4 at Abbeystead was attributed to the thermally-mature, organic-rich Bowland Shale [*Exploration Consultants Ltd.*, 1986] and a large area of northwest England is underlain by the Bowland Shale, which is the United Kingdom's (UK) largest prospective shale gas resource [*Andrews*, 2013].

2.1.3. Preston New Road and Kirby Misperton datasets

The law in the UK requires that 12 months of baseline monitoring of dissolved CH_4 concentrations in groundwater must occur at a shale gas exploration site prior to fracking [*Infrastructure Act*, 2015]. In England, at the time of this study, there are two sites (Preston New Road, Lancashire, and Kirby Misperton, North Yorkshire – Fig. 1) where groundwater has been monitored as local baselines prior to fracking. Dissolved CH_4 concentrations for the Preston New Road (PNR) site were compiled from the Cuadrilla Resources Ltd. ePortal [*Cuadrilla Resources Ltd.*, 2019]. This dataset, referred to hereafter as the PNR dataset, consisted of four boreholes at the well pad with monthly dissolved CH_4 concentration measurements in groundwater from two superficial formations (Glacial Till and Middle Sands) within each borehole (Table S5). Where multiple concentration measurements were taken within a month the mean concentration was used. The same approach was used for the Kirby Misperton (KM) site using the monitoring reports submitted to the EA [*EA*, 2019b]. The KM dataset consisted of 11 boreholes (five at the well pad and six at distances 1.7-3.1 km from the well pad) with monthly dissolved CH_4 concentration measurements in single formations: six within the Superficial Deposits/weathered Kimmeridge Clay, two within the un-weathered Kimmeridge Clay, two within the weathered/un-weathered Kimmeridge Clay/Corallian Group, and one within the Corallian Group (Table S6).

2.2. Bayesian generalised linear modeling

To create a dynamic baseline generalised linear modeling within a Bayesian framework [*Gelman et al.*, 2004] was used. Each data point (i.e. a dissolved CH_4 concentration measurement) was assumed to be generated from a particular distribution in the exponential family. *A priori*, four distributions were considered: normal, log-normal, gamma and Weibull. The latter three distributions are only defined for positive numbers and so there is no possibility that physically impossible negative concentrations would be predicted, as would be the case with a normal distribution. For the BGS and EA datasets only one factor was considered; the difference between all the monitoring locations from which dissolved CH_4 concentration data were available – this factor is henceforward known as the "Location factor". Specific environmental conditions at each location were unknown and the dates of the sampling were not even or consistent within and between the datasets and so no further factors could be included in the analysis. Note that the inclusion of a spatially varying location factor does incorporate unknown spatial variation in environmental conditions across the datasets. For the PNR and KM datasets the sampling at each location within the datasets was sufficient that the difference between the months of sampling could be tested – this factor had 12 levels, one for each month, and is henceforward referred to as "Month factor". Location factor was also considered for the PNR and KM datasets.

Model fit was tested using four approaches. Firstly, that the 95% credible interval (CI) for any factor did not include zero. This is henceforward referred to as being significantly different from zero at a probability of 95%. Secondly, that inclusion of the factor caused the total model deviance to decrease. Thirdly, 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 because inclusion allows 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. Fourthly, the fit of the generalised linear model was compared with the original datasets and the fit assessed using the root mean square error (RMSE).

Given the available datasets the analysis was constructed sequentially. Firstly, the four exponential family distributions were fitted to the BGS dataset. Because this was the first of the four datasets to be considered there were no prior datasets to inform this model. Therefore, a weak uninformative prior distribution was used. A Jeffreys prior [*Jeffreys*, 1946] was chosen as the form of the uninformative prior whereby the expected value was set as the mean of all the dissolved CH_4 concentrations in the BGS dataset. For the Jeffrey prior the standard deviation was set as 100 times the coefficient of variation of the dataset. In this way the Jeffrey prior distribution was centred on the expected value of the data and was almost uniform in distribution. The best-fit model for the BGS dataset was used as the prior distribution for the analysis of the EA dataset – this would represent a strong prior. Because sampling locations differed between the BGS and EA datasets it was not possible to use the BGS dataset as a prior for each location within the EA dataset. To demonstrate the effect of a strong versus weak prior the EA dataset was also analyzed using a Jeffrey prior and the result compared to that when the BGS dataset was used as a strong prior. The best-fit model for the EA dataset was then used as strong prior information for the PNR and KM datasets. In all the datasets those values which were below the analytical detection limit were treated as censored data; the exact value of the dissolved CH_4 concentration was not known but the observation could still provide information. Therefore, censored data were included in the analysis even with varying detection limits within and between the study datasets.

The Bayesian generalised linear modeling was achieved by Markov Chain Monte Carlo (MCMC) simulation to estimate the posterior distribution of the dissolved CH_4 concentrations using WinBUGS (Version 1.4) [*Lunn et al.*, 2000]. The length of the MCMC chain was 30000 cycles after 10000 burn in cycles with samples saved every 10 cycles and with one chain.

2.3. Dissolved CH_4 concentrations ≥10.0 mg/l

In the UK there are no regulatory limits regarding dissolved CH_4 concentrations in groundwater. However, the US Department of the Interior risk action level of 10.0 mg/l [*Eltschlager et al.*, 2001] provides a reference that may assist decision making in other countries [e.g. *Bell et al.*, 2019; *Humez et al.*, 2016; *Schloemer et al.*, 2016]. Consistent with previous studies we adopt 10.0 mg/l as an indicative risk action level in this study and accordingly identify locations with dissolved CH_4 concentrations ≥10.0 mg/l.

2.4. Groundwater redox conditions

Dissolved CH_4 concentrations are expected to show relationships with groundwater redox conditions. Reducing conditions (i.e. a lack of oxygen) can promote CH_4 production by methanogenic bacteria and preserve pre-existing biogenic and thermogenic CH_4 [*Molofsky et al.*, 2016]. Conversely, aerobic conditions may permit rapid consumption of CH_4 by methanotrophic bacteria [*Whittenbury et al.*, 1970]. Therefore, information on groundwater redox conditions could help predict locations with elevated dissolved CH_4 concentrations. For example, in the Appalachian Basin, USA, low nitrate (NO_3^-) and sulfate (SO_4^{2-}) concentrations were associated with higher dissolved CH_4 concentrations and offered strong predictive power when combined with other factors such as sodium (Na)-rich water and upland topography [*Molofsky et al.*, 2016]. To investigate if relationships between redox conditions and dissolved CH_4 concentrations exist in English groundwater, locations in the EA dataset with dissolved CH_4 concentrations ≥10.0 mg/l (henceforward referred to as elevated CH_4 locations) were paired with their nearest neighbouring location where dissolved CH_4 concentrations were <10.0 mg/l (henceforward referred to as low CH_4 locations). The following redox-linked groundwater parameters (henceforward termed "indicators") were extracted from the EA Water Quality Archive for the paired sites: ammonium (NH_4) concentration; biochemical oxygen demand (BOD) as 5 day Allylthiourea (ATU); chemical oxygen demand (COD) as O_2; dissolved oxygen (DO); iron (Fe) concentration; manganese (Mn) concentration; NO_3^- concentration; and SO_4^{2-} concentration. Additionally, calcium (Ca), magnesium (Mg), and Na concentrations were extracted to provide an indication of groundwater type and subsequently determine if the boreholes within a pair sampled the same groundwater body. As a further constraint on groundwater bodies at paired locations, boreholes from the EA dataset were manually matched with boreholes from the BGS Onshore GeoIndex [*BGS*, 2019] using geographic coordinates and location names. The borehole depths and aquifer designations recorded in the BGS Onshore GeoIndex were used as proxies for groundwater bodies sampled by the EA for dissolved CH_4 concentrations.

Two statistical methods were employed to investigate relationships between groundwater redox conditions and dissolved CH_4 concentrations. Firstly, binomial probability modeling was used to predict the probability of association between elevated CH_4 locations and levels of indicators. Dissolved CH_4 concentrations were classified using two levels, either "one or more CH_4 samples ≥10.0 mg/l" or "all CH_4 samples <10.0 mg/l", that is elevated or low, respectively. Multiple measured values for the different indicators at each location were mean-averaged to average out irregular timings between groundwater samples and any seasonal variations in groundwater chemistry. The mean-averaged indicators at each elevated CH_4 location were compared to the mean-averaged indicators at the corresponding paired low CH_4 location. A success was recorded when concentrations of NH_4, BOD, COD, Fe and Mn at locations with elevated CH_4 concentrations were higher than their concentration in the corresponding low CH_4 location. Successes for DO, NO_3^- and SO_4^{2-} were recorded when concentrations at elevated CH_4 locations were lower than their concentration in the comparative location. If indicator values were equal in a comparative location pair, this was not counted as a success. The binomial distribution was then used to assess the probability of the proportion of observed successes and failures. In effect the distribution of indicators between pairs of elevated and low CH_4 locations was treated as modeling a coin toss (binomial distribution). For example, if DO was always lower in the elevated CH_4 locations compared to the paired low CH_4 locations then this would have low probability of occurring by random chance and it could be reasonably concluded that elevated dissolved CH_4 was associated with low DO. By analogy tossing a coin and always coming up heads would be unlikely unless it was a biased coin.

The second statistical method used to investigate any relationships between dissolved CH_4 concentrations and indicators was ANOVA. One-way ANOVA was used where the single factor was the dissolved CH_4 concentration. The dissolved CH_4 concentrations were classified as for the binomial probability modeling, either elevated (≥10.0 mg/l) or low (<10.0 mg/l). Three types of one-way ANOVA were performed: indicators without covariates; indicators with single covariates; and indicators with multiple covariates. The covariates used were the remaining indicators. In this way the ANOVA was used to judge whether or not indicators were significantly different between locations with elevated or low dissolved CH_4 concentrations and, by including or not including covariates, to assess whether differences in indicators could be explained by the behaviour of other indicators. Prior to the ANOVA the normality of each indicator was tested using the Anderson-Darling test and indicator values were log-transformed if necessary and re-tested to confirm improved normality. Statistical significance was judged at the 95% probability of an indicator not having zero effect.

3. Results

3.1. *Dissolved CH_4 concentrations in England*

The BGS, EA, PNR and KM datasets are provided in the supplementary material (Tables S1, S2, S5 and S6). The EA dataset has approximately five times more dissolved CH_4 concentration measurements than the BGS dataset and on average there are approximately four measurements per unique location in the EA dataset, compared to mostly one measurement per unique location in the BGS dataset (Table 1). The spatial distributions of dissolved CH_4 concentrations from the four datasets are shown in Figure 2. The general detection limit for the BGS dataset (<0.5 µg/l) is three orders of magnitude lower than the EA dataset (<0.5 mg/l) (Fig. 3) and the detection limits for the PNR and KM datasets were <0.01 mg/l and <0.001 mg/l, respectively. The maximum dissolved CH_4 concentration measurement in the EA dataset was 25.9 mg/l, compared to 4.72 mg/l in the BGS dataset, and 36 concentration measurements (over 11 unique locations) were ≥10.0 mg/l (1.7% of the EA dataset) (Table 1). Nine of the 11 locations with a dissolved CH_4 concentration measurement ≥10.0 mg/l are located in the prospective area of the Bowland Shale, although no directional spatial trend is readily observable across England using the currently available data. Both national datasets indicate that dissolved CH_4 concentrations in English groundwater are generally below the risk action level of 10.0 mg/l.

The PNR dataset contains no dissolved CH_4 concentrations ≥10.0 mg/l in either the Glacial Till or Middle Sands across all four monitoring boreholes (Fig. 4a). The maximum concentration recorded at PNR was 5.75 mg/l in the Middles Sands of Borehole 1 in March 2018 (Table S5). Dissolved CH_4 concentrations at KM were more varied than at PNR but predominantly ranged from below detection limit (<0.001 mg/l) to 0.1 mg/l (Fig. 4b), although the maximum concentration recorded at KM was 67.1 mg/l in the Corallian Group of Borehole E in June 2016 (Table S6). However, this maximum concentration should not necessarily be considered exact because it lies outside the calibration limits of the analytical equipment, i.e. it should be considered as an indicative concentration only [*Envireau Water*, 2017]. Nevertheless, all monthly concentration results from Borehole E were ≥10.0 mg/l (Table S6). The higher dissolved CH_4 concentrations detected at KM compared to PNR may reflect the greater sampling depths at KM, for example samples from Borehole

E are from ~200 m below ground level and CH_4 solubility increases with depth. With the exception of Borehole E at KM, dissolved CH_4 concentrations at PNR and KM are not unusual compared to the BGS and EA datasets.

3.2. Bayesian generalised linear modeling

3.2.1. BGS and EA datasets

For the BGS dataset the deviances of the distributions were: normal (4242); log-normal (2527); gamma (2087); and Weibull (2076). The Weibull distribution had the lowest deviance and the generalised linear modeling was therefore taken forward using this distribution. The best-fit Weibull distribution was W(0.35±0.01, 0.34±0.16), where the Weibull function has the form W(*r*, *b*) where *r* is the shape factor and *b* is:

$$b = \lambda^{-r} \; for \; x > 0 \tag{i}$$

where λ is the scale factor. The expected mean, $E(x)$, and variance, $Var(x)$, were:

$$E(x) = b^{-1/r}\,\Gamma\left(1 + \tfrac{1}{r}\right) = 0.23\ mg/l \tag{ii}$$

$$Var(x) = \lambda^2\left[\Gamma\left(1 + \tfrac{2}{r}\right) - \left(\Gamma\left(1 + \tfrac{1}{r}\right)\right)^2\right] = 0.7\ mg/l \tag{iii}$$

The analysis of the BGS dataset provided the strong prior information for the EA dataset. When the strong prior was used for the EA dataset the deviance of the generalised linear model was 1601, compared to 1727 when the weak Jeffrey prior was used. When the predicted values from the weak and strong priors were compared the estimated values from the weak prior were 54% smaller than those predicted from the strong prior. The CI for each location was compared between the weak and strong priors; the CI for the strong prior was on average 58% of the CI for the weak prior (Fig. 5) – and this equates to an increase in equivalent sample size of 75%. With the strong prior the Location factor was significant and this could be used to calculate the expected value (the “average”) for each location within the dataset (Figure 6). The distribution of observed and predicted dissolved CH_4 concentrations are plotted in Figure 7. Based upon the EA dataset with the strong prior, then $E(x)$ = 0.24 mg/l and $Var(x)$ = 0.83 mg/l, with the 95% CI at 0.11 to 0.45 mg/l.

3.2.2. PNR and KM datasets

When using the weak Jeffreys prior for the PNR dataset the deviance was 661, which improved to 643 on use of the strong prior (the EA dataset with the BGS dataset used as a strong prior). The expected mean dissolved CH_4 concentration for the PNR dataset was 0.033 mg/l, with the 95% CI at 0.025 to 0.044 mg/l. The Month factor and its interaction with Location factor proved insignificant, indicating no significant seasonal cycle in the dissolved CH_4. The distribution of the data confirmed that the dissolved CH_4 concentrations of the Middle Sands in Borehole 1 were indeed significantly different from all the other boreholes (Fig. 8a). For the KM dataset the Month factor and its interactions also proved insignificant. Use of a strong prior made only a small difference to the deviance of the model fit, decreasing from 822 to 820. The expected mean dissolved CH_4 concentration for the KM dataset was 0.060 mg/l, with the 95% CI at 0.035 to 0.103 mg/l. The distribution of the data for the boreholes in the KM dataset showed that seven of the 11 boreholes were not significantly different from each other, but four boreholes (D, E, G1, and G3) did have significantly higher dissolved CH_4 concentrations (Fig. 8b). However, in this study we can give no particular explanation for this.

3.3. Relationships between elevated CH_4 concentrations and redox conditions

Thirty six dissolved CH_4 concentration measurements in the EA dataset had concentrations greater than or equal to the risk action level of 10.0 mg/l (Table 2). These 36 concentrations occurred across 11 unique locations and nine of these locations were situated in the prospective area of the Bowland

Shale, although no link to thermogenic gas migration from the Bowland Shale can be demonstrated based upon the available data. The percentage of dissolved CH_4 concentration measurements ≥10.0 mg/l from the total number of dissolved CH_4 concentration measurements taken at the 11 locations ranged from 4-100%.

3.3.1. Groundwater body sampling

Three of the 11 locations with dissolved CH_4 concentrations ≥10.0 mg/l showed clear similar relative concentrations of Ca, Mg and Na to their paired low CH_4 location, indicating the corresponding boreholes sampled the same groundwater bodies (Table S7). A further four paired locations also showed similar, although less distinct, relative proportions. However, sampling of the same groundwater body at three of these four paired locations was further supported by the BGS aquifer designations (i.e. aquifer designations were the same for both boreholes in a corresponding pair). We assumed that seven of the 11 pairs sample the same groundwater body within their pairs and therefore ran binomial probability modeling for both the seven and 11 pairs.

3.3.2. Binomial probability modeling

When all 11 paired locations were considered all eight indicators showed probabilities greater than 50% (Table 3), i.e. the observed proportion of successes to failures was more likely than not. However, taking a 95% probability as the significant probability then only NH_4, BOD, DO and Fe showed significant ratios for the elevated and low CH_4 locations. For BOD and DO the probability of success was 100.0%. However, BOD and DO had only one and four trials, respectively, reducing confidence in their results. For those redox indicators present at all 11 paired locations, NH_4 and Fe were the best predictors, both with probabilities of 96.7% (Table 3). When only the seven pairs that likely sampled the same groundwater bodies were considered all eight indicators again showed probabilities greater than 50% (Table 4). Only BOD and DO showed significant ratios for the elevated and low CH_4 locations (100.0%) but had only one and three trials, respectively, again reducing confidence in their results. For those redox indicators present at all seven paired locations Fe, Mn, and SO_4^{2-} were the best predictors, all with probabilities of 93.8% (Table 4).

3.3.3. *Analysis of variance*

Prior to ANOVA the Anderson-Darling test suggested that all indicator values should be log-transformed before subsequent analysis – no further transformation was necessary. One-way ANOVA without covariates showed that NH_4, COD, Fe, Mn, and NO_3^- showed statistically significant differences between elevated and low CH_4 locations (Table 5). All statistically significant factors, except NO_3^-, had probabilities <0.0005 of having zero effect. COD explained the highest variation in the data ($R^2 = 15.6\%$) and Mn explained the least variation ($R^2 = 6.3\%$). *A priori* it would be expected that NH_4, BOD, COD, Fe and Mn would be at higher concentrations in boreholes with elevated dissolved CH_4 concentrations, and conversely, that DO, NO_3^- and SO_4^{2-} would all be lower. When the change in mean indicator values going from low to elevated CH_4 locations were considered, COD, Fe and Mn followed the expected pattern (i.e. values increased in elevated CH_4 locations) but NH_4 and NO_3^- did not, despite showing the largest percentage change. This may mean that NH_4 and NO_3^- may not be useful indicators of elevated dissolved CH_4 concentrations. When covariates were considered only one combination improved the fit of the ANOVA and that was when NH_4 was considered with BOD and COD as covariates ($R^2 = 45.0\%$).

4. Discussion

4.1. *Comparison of EA dataset to other studies*

Since 2012 a number of published studies related to oil and gas industrial activities have provided measurements of dissolved CH_4 concentrations in groundwater (Table 6). These range spatially from those covering the 5180 km^2 Wattenberg oil and gas field, Colorado [*Li and Carlson*, 2014], to those which cover whole states and nations [e.g. *Bell et al.*, 2017]. Comparison of the EA dataset to other studies shows that there is little evidence that the EA data were biased towards locations with higher dissolved CH_4 concentrations (Table 6). The EA dataset contains the largest number of dissolved CH_4 concentration measurements of any of the published datasets to date and is also one of the largest in terms of unique sampling locations (Table 6). Despite having one of the highest detection limits (<0.5 mg/l) of the published dissolved CH_4 datasets, the EA dataset has one of the lowest mean dissolved CH_4 concentrations (0.24 mg/l using the Bayesian generalised linear modeling approach) and maximum CH_4 concentrations (25.9 mg/l) (Table 6). Furthermore, the percentage of dissolved CH_4 concentration measurements greater than or equal to the US Department of the Interior risk action level of 10.0 mg/l was also one of the lowest (1.7%); only surpassed by the datasets for Sullivan County (0.8%) [*Reese et al.*, 2014], New York State (0.0%) [*McPhillips et al.*, 2014] and Great Britain (0.0%) [*Bell et al.*, 2017]. These comparisons suggest that although dissolved CH_4 is ubiquitous in English groundwater, including areas with prospective shale gas resources, concentrations are generally low compared to other geographical areas and very rarely above concentrations requiring further investigation. This observation is in agreement with the conclusions of *Bell et al.* [*2017*]. However, the data presented in this study do also highlight that dissolved CH_4 concentrations in English groundwater can exceed 10.0 mg/l.

4.2. Dissolved CH_4 flux to atmosphere

The results of the Bayesian generalised linear modeling for the EA dataset suggest an average dissolved CH_4 concentration of 0.24 mg/l in English groundwater. The solubility of CH_4 in water is 31.0 mg/l at atmospheric pressure and 283 K, with solubility doubling for every 10 m increase in depth [*Bordeleau et al.*, 2018; *Humez et al.*, 2016]. However, the Henry's constant of CH_4 means that in equilibrium with the atmosphere a far lower concentration of dissolved CH_4 would be expected in groundwater:

$$[CH_4] = P_{CH_4} \times K_{HT} \tag{iv}$$

$$K_{HT} = K_{H298} exp\left(A\left(\frac{1}{T} - \frac{1}{298}\right)\right) \tag{v}$$

where T is temperature of the water (K), K_{HT} is Henry's law constant (mol/m^3/Pa), K_{H298} is Henry's law constant at 298 K (mol/m^3/Pa), $[CH_4]$ is dissolved concentration of CH_4 in water (mol/l), P_{CH_4} is the partial pressure of CH_4 in the atmosphere (Pa) and A is a constant.

Given $P_{CH4} = 1.81 \times 10^{-6} Pa$, $K_{H298} = 1.4 \times 10^{-5} mol/m^3/Pa$ and $A = 1600$ [*Sander*, 2015] then the $[CH_4]$ at equilibrium with air is $4.0 \times 10^{-9}\, mol/l\, (6.4 \times 10^{-5}\, mg/l)$ at 274 K to $2.8 \times 10^{-9}\, mol/l\, (4.4 \times 10^{-5}\, mg/l)$ at 293 K. Therefore, the expected mean concentration for dissolved CH_4 in English groundwater, based on this study, is in excess of that expected at equilibrium with the atmosphere. As a result we would expect almost complete degassing of the CH_4 to the atmosphere upon emergence of groundwater at the surface. The average annual recharge to the main aquifers in England is estimated to be $7 \times 10^9\, m^3/day$ [*EA,* 2007]. If it is assumed that groundwater levels are not changing substantially across England then this same volume must be discharged to surface waters. If it is also assumed that the CH_4 degasses as it emerges to the atmosphere, rather than degassing from the soil profile where it could readily be oxidised, then given the 95% CI of dissolved CH_4 concentrations predicted from this study (0.11 to 0.45 mg/l), then the amount of CH_4 degassing from groundwater would be 0.7 to 3.1 kt CH_4/year with an expected value of 1.65 kt CH_4/yr. This equates to greenhouse gas warming potential (GWP) of 17 to 74 kt CO_{2eq}/yr with an expected value of 40.3 kt CO_{2eq}/yr. The UK greenhouse gas inventory estimates a total UK CH_4 flux of 53500 kt CO_{2eq}/yr [*DECC*, 2016]. Therefore, CH_4 degassing from groundwater in England potentially contributes ~0.01% to the estimated CH_4 flux in the UK. This estimate is in good agreement with

Gooday and Darling [2005] who estimated groundwater contributes a maximum of 0.05% to all UK CH_4 emissions.

4.3. A dynamic baseline and predicting dissolved CH_4 concentrations

This study has developed a coherent and transparent approach for combining dissolved CH_4 datasets with different detection limits. The Bayesian approach employed uses all available data to predict distributions of dissolved CH_4 concentrations at national and local scales (e.g. around shale gas sites). Such distributions represent a baseline against which future observations, and especially in this case observations made after fracking operations and shale gas extraction have commenced, can be judged. An advantage of using the Bayesian approach is that the tool automatically updates with new data and so contributes to the development of a dynamic baseline in time and space. In addition, the approach gains value from the whole monitoring network, i.e. maximum information is gained from past and ongoing monitoring. Therefore, this approach gives good value for the money invested in environmental monitoring and can be used to assess information content and informational efficiency of the current monitoring network.

For groundwater determinands with defined environmental quality standards [e.g. *EC,* 2000], individual results from monitoring will be viewed relative to these standards and the probability that the standard is being exceeded assessed. For determinands such as dissolved CH_4 concentrations, no legal standard exists and such comparisons may not occur. Furthermore, the review period for environmental monitoring is not always clear. Under an operator's permit the operator should review continuously, i.e. dissolved CH_4 concentration data reviewed each time new data are obtained and the regulator informed if there is an issue. The environmental regulator in the UK (the EA in England) may be asked to report at any time to the Secretary of State. However, no fixed review period has currently been set for the UK. In the approach used in this study each new datum can be viewed against a prediction that is based upon all available data, Furthermore, a probability (with uncertainty) is calculated that a new observation is exceptional and not what should be expected, even for locations sampled only once. In the case used here a measurement of dissolved CH_4 concentration would be judged against the predicted distribution as a means of testing whether a CH_4 leak to groundwater has or has not occurred at a shale gas site.

In effect this study has built up a method to improve groundwater baseline 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. However, this would not be a fair comparison because a local variation might mean that comparing one observation with data from all years would be inappropriate, for example if there was an interannual trend at the site where values in the current year tend to be lower than those in a previous year. Under such circumstances a distribution for the given year would be better than comparing with data from all years. Of course it is unlikely that there will be sufficient observations at a site to give such a reasonable distribution for any month for any year. Therefore, it is preferable if information from other sites could be drawn upon, which is what this Bayesian approach has achieved. An analogous, non-Bayesian approach might be that of weighted regression analysis [*Hirsch et al.*, 2010; 2015].

Our approach could be improved with the use of covariates. This study has considered a range of possible groundwater redox parameters as potential indicators of elevated concentrations of dissolved CH_4. Using binomial probability modeling and ANOVA it was shown that high Fe concentration was a consistently successful indicator of elevated dissolved CH_4 concentrations in English groundwater. This contrasts with research in the Appalachian Basin, USA, which concluded no significant relationship existed between Fe and elevated CH_4 [*Molofsky et al.*, 2016]. Alternatively, *Molofsky et al.* [2016] found that low NO_3^- and SO_4^{2-} concentrations were associated with elevated CH_4 concentrations and offered strong predictive power when combined with other factors of Na-rich water and upland topography. Although our analysis has suggested high Fe concentration may be a good indicator of elevated CH_4 concentrations in English groundwater, more extensive studies using historic data and ongoing groundwater monitoring programmes are needed to investigate this and provide a relationship between concentrations of Fe and dissolved CH_4. Importantly, the redox parameters considered in this study do suggest the possibility that in the future, rather than directly monitoring dissolved CH_4 concentrations, areas with elevated dissolved CH_4 concentrations may be initially identified for further investigation by surveying more readily available and cheaper determinands such as Fe and NH_4. More expensive analysis measuring dissolved CH_4 concentrations and stable carbon isotopes could then be targeted to confirm elevated CH_4 concentrations and discriminate potential sources [e.g. *Teasdale et al.*, 2019].

5. Conclusions

This study has developed a dynamic baseline for dissolved CH_4 concentrations in English groundwater, from which the potential environmental impact of anthropogenic activities, such as shale gas exploitation, can be assessed. The Bayesian approach used can systematically and transparently bring together multiple monitoring datasets, including those with differing detection limits, and provide all results within a probabilistic framework with uncertainty. Furthermore, the results from previous analysis can be used as prior information for future analysis, leading to results that develop over time and space. This study used four different datasets of dissolved CH_4 concentrations in groundwater: two nationwide datasets and two local baseline monitoring datasets. By building on the most sensitive national dataset (the previously published British Geological Survey CH_4 baseline for Great Britain) a strong prior distribution was developed. This strong prior was used for the subsequent national dataset (that of the Environment Agency of England), which had approximately five times more dissolved CH_4 concentration measurements but a detection limit three orders of magnitude higher. When the strong prior was used over a weak prior for the subsequent dataset the precision improved by 75%. The expected mean dissolved CH_4 concentration in English groundwater based on the Bayesian approach is 0.24 mg/l, with a variance of 0.83 mg/l, a 95%

credible interval of 0.11 to 0.45 mg/l, and a Weibull distribution of W(0.35±0.01, 0.34±0.16). These results indicate that the amount of CH_4 degassing from English groundwater to the atmosphere equates to between 0.7 to 3.1 kt CH_4/year, with an expected value of 1.65 kt CH_4/year and a greenhouse gas warming potential of 40.3 kt CO_{2eq}/year. The two local baseline monitoring datasets were from the two current shale gas exploration sites in England. These sites, in combination with the national datasets, indicate that dissolved CH_4 concentrations in English groundwater are generally low, but locations with concentrations greater than or equal to the widely used risk action level of 10.0 mg/l do exist. Analyses of groundwater redox conditions at locations with dissolved CH_4 concentrations ≥10.0 mg/l suggest that it may be possible to identify other locations with elevated dissolved CH_4 concentrations using redox parameters such as Fe concentration.

Conflicts of interest

There are no conflicts of interest to declare.

Supplementary material

Appendix A (Tables S1-S7)

Acknowledgments

Miles Wilson is funded by a Durham Doctoral Studentship. Richard Davies, Fred Worrall, and Miles Wilson are part of the Researching Fracking (ReFINE) consortium which has been funded by the Natural Environment Research Council (UK), Total, Shell, Chevron, GDF Suez, Centrica and Ineos. The results are solely those of the authors. Water quality data were downloaded from the Environment Agency of England's publically available Water Quality Archive (downloaded 5th and 6th February

2018). Lisa Molofsky is thanked for providing data for Susquehanna County, USA. Rob Ward and Rachel Bell from the British Geological Survey are thanked for their constructive comments on an earlier version of the manuscript. We also thank an anonymous reviewer for their useful comments which helped improve the final manuscript.

References

Andrews, I. J. (2013). The Carboniferous Bowland Shale gas study: geology and resource estimation. British Geological Survey for Department of Energy & Climate Change, London, UK.

Bell, R. A., Darling, W. G., Ward, R. S., Basava-Reddi, L., Halwa, L., Manamsa, K., & Dochartaigh, B. Ó. (2017). A baseline survey of dissolved methane in aquifers of Great Britain. *Science of the Total Environment*, 601, 1803-1813.

BGS (British Geological Survey) (2019). British Geoindex Onshore. Available at: BGS Onshore GeoIndex (Last accessed: 05/03/2019).

Bordeleau, G., Rivard, C., Lavoie, D., Lefebvre, R., Malet, X., & Ladevèze, P. (2018). Geochemistry of groundwater in the Saint-Édouard area, Quebec, Canada, and its influence on the distribution of methane in shallow aquifers. *Applied geochemistry*, 89, 92-108.

Cuadrilla Resources Ltd. (2019). Cuadrilla ePortal. Available at: https://www.cuadrillaresourceseportal.com/jan2019.html (Last accessed: 05/03/2019).

Darling, W. G., & Gooddy, D. C. (2006). The hydrogeochemistry of methane: evidence from English groundwaters. *Chemical Geology*, 229(4), 293-312.

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*, 111(39), 14076-14081.

DECC (Department of Energy & Climate Change) (2016). 2014 UK greenhouse gas emissions, final figures: statistical release. National Statistics, Department of Energy & Climate Change, London, UK.

EA (Environment Agency) (2007). Underground, under threat – The state of groundwater in England and Wales. Environment Agency, Bristol, UK.

EA (Environment Agency) (2019a). Water Quality Archive. Available at: https://environment.data.gov.uk/water-quality/view/landing (Last accessed: 08/03/2019).

EA (Environment Agency) (2019b). Third Energy – Kirby Misperton (KM8 well) information page. Available at: https://consult.environment-agency.gov.uk/onshore-oil-and-gas/third-energy-kirby-misperton-information-page/ (Last accessed: 05/03/2019).

EC (European Council) (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 of the European Communities, 22(2000), L327.

Eltschlager, K. K., Hawkins, J. W., Ehler, W. C., & Baldassare, F. (2001). Technical measures for the investigation and mitigation of fugitive methane hazards in areas of coal mining. US Department of the Interior, Office of Surface Mining Reclamation and Enforcement.

Envireau Water (2017). Baseline Water Quality Data, April – June 2017, Kirby Misperton A Wellsite, North Yorkshire. Envireau Water for Third Energy UK Gas Limited, East Knapton, UK.

Exploration Consultants Ltd. (1986). Lancashire Conjunctive Use Scheme, Further Investigations: Wyresdale Tunnel, Final Geological/Geophysical Report. Exploration Consultants Limited for North West Water, UK.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). Bayesian Data Analysis (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gooddy, D. C., & Darling, W. G. (2005). The potential for methane emissions from groundwaters of the UK. *Science of the Total Environment*, 339(1-3), 117-126.

Hirsch, R. M., Moyer, D. L., & Archfield, S. A. (2010). Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs 1. *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 Modeling & Software*, 73, 148-166.

Hooker, P., & Bannon, M. (1993). Methane: its occurrence and hazards in construction. CIRIA Rep. 130.

Humez, P., Mayer, B., Ing, J., Nightingale, M., Becker, V., Kingston, A., Akbilgic, O. & Taylor, S. (2016). Occurrence and origin of methane in groundwater in Alberta (Canada): Gas geochemical and isotopic approaches. *Science of the Total Environment*, 541, 1253-1268.

Infrastructure Act (2015). Part 6 Energy, Onshore hydraulic fracturing: safeguards. Available at: http://www.legislation.gov.uk/ukpga/2015/7/contents/enacted (Last accessed: 05/03/2019).

IPCC (Intergovernmental Panel on Climate Change) (2013). Climate Change 2013: The Physical Science Basis.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007), 453-461.

Kappel, W. M., & Nystrom, E. A. (2012). Dissolved methane in New York groundwater. US Geological Survey Open-file report 2012-1162.

Carrier, M. A., Lefebvre, R., Rivard, C., Parent, M., Ballard, J. M., Benoît, N., Vigneault, H., Beaudry, C., Malet, X., Laurencelle, M. & Gosselin, J. S. (2013). Portrait des ressources en eau souterraine en Montérégie Est, Québec, Canada (No. R1433). INRS, Centre Eau Terre Environnement.

Li, H., & Carlson, K. H. (2014). Distribution and origin of groundwater methane in the Wattenberg oil and gas field of northern Colorado. *Environmental science & technology*, 48(3), 1484-1491.

Llewellyn, G. T., Dorman, F., Westland, J. L., Yoxtheimer, D., Grieve, P., Sowers, T., Humston-Fulmer, E. and Brantley, S. L. (2015). Evaluating a groundwater supply contamination incident attributed to Marcellus Shale gas development. *Proceedings of the National Academy of Sciences*, 112(20), 6325-6330.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-a Bayesian modeling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4), 325-337.

Mair, R., Bickle, M., Goodman, D., Koppelman, B., Roberts, J., Selley, R., Shipton, Z., Thomas, H., Walker, A., Woods, E. & Younger, P. (2012). Shale gas extraction in the UK: a review of hydraulic fracturing.

McIntosh, J. C., Grasby, S. E., Hamilton, S. M., & Osborn, S. G. (2014). Origin, distribution and hydrogeochemical controls on methane occurrences in shallow aquifers, southwestern Ontario, Canada. *Applied geochemistry*, 50, 37-52.

McPhillips, L. E., Creamer, A. E., Rahm, B. G., & Walter, M. T. (2014). Assessing dissolved methane patterns in central New York groundwater. *Journal of Hydrology: Regional Studies*, 1, 57-73.

Molofsky, L. J., Connor, J. A., McHugh, T. E., Richardson, S. D., Woroszylo, C., & Alvarez, P. J. (2016). Environmental factors associated with natural methane occurrence in the Appalachian Basin. *Groundwater*, 54(5), 656-668.

Molofsky, L. J., Connor, J. A., Wylie, A. S., Wagner, T., & Farhat, S. K. (2013). Evaluation of methane sources in groundwater in northeastern Pennsylvania. *Groundwater*, 51(3), 333-349.

Moritz, A., Helie, J. F., Pinti, D. L., Larocque, M., Barnetche, D., Retailleau, S., Lefebvre, R. & Gélinas, Y. (2015). Methane baseline concentrations and sources in shallow aquifers from the shale gas-prone region of the St. Lawrence Lowlands (Quebec, Canada). *Environmental science & technology*, 49(7), 4765-4771.

Ó Dochartaigh, B. E., Smedley, P. L., MacDonald, A. M., Darling, W. G., & Homoncik, S. (2011). Baseline Scotland: groundwater chemistry of the Carboniferous sedimentary aquifers of the Midland Valley. British Geological Survey Report (OR/11/021).

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*, 108(20), 8172-8176.

Reese, S. O., Neboga, V. V., Pelepko, S., Kosmer, W. J., & Beattie, S. (2013). Groundwater and petroleum resources of Sullivan County, Pennsylvania. Pennsylvania Geological Survey, 4th series, Water Resource Report 71.

Sander, R. (2015). Compilation of Henry's law constants (version 4.0) for water as solvent. *Atmospheric Chemistry & Physics*, 15(8), 4399–4981.

Schloemer, S., Elbracht, J., Blumenberg, M., & Illing, C. J. (2016). Distribution and origin of dissolved methane, ethane and propane in shallow groundwater of Lower Saxony, Germany. *Applied geochemistry*, 67, 118-132.

Teasdale, C. J., Hall, J. A., Martin, J. P., & Manning, D. A. (2019). Discriminating methane sources in ground gas emissions in NW England. *Quarterly Journal of Engineering Geology and Hydrogeology*, 52(1), 110-122.

Vengosh, A., Warner, N., Jackson, R., & Darrah, T. (2013). The effects of shale gas exploration and hydraulic fracturing on the quality of water resources in the United States. *Procedia Earth and Planetary Science*, 7, 863-866.

Whittenbury, R., Phillips, K. C., & Wilkinson, J. F. (1970). Enrichment, isolation and some properties of methane-utilizing bacteria. *Microbiology*, 61(2), 205-218.

Williams, G. M., & Aitkenhead, N. (1991). Lessons from Loscoe: the uncontrolled migration of landfill gas. *Quarterly Journal of Engineering Geology and Hydrogeology*, 24(2), 191-207.

A dynamic baseline for dissolved methane in English groundwater

M. P. Wilson[a]*, F. Worrall[a], R. J. Davies[b], and A. Hart[c]

[a] Department of Earth Sciences, Durham University, Science Labs, Durham, DH1 3LE, UK

[b] School of Natural and Environmental Sciences, Newcastle University, Newcastle, NE1 7RU, UK

[c] Environment Agency, Research Assessment and Evaluation, Sapphire East, Streetsbrook Road, Solihull, B91 1QT, UK

* Corresponding author: Miles Wilson (miles.wilson@durham.ac.uk)

Figure captions

Figure 1: Map of Great Britain showing the dissolved CH_4 concentration sampling locations of the BGS and EA datasets. Current shale gas exploration sites with local baseline monitoring of dissolved CH_4 concentrations and the prospective area of the Bowland Shale [*Andrews*, 2013] are also shown.

Figure 2: Dissolved CH_4 concentrations for the BGS, EA PNR and KM datasets (note prospective Bowland Shale is not shown in either inset map for data point clarity). Where more than one dissolved CH_4 concentration measurement is present at a single location, the maximum dissolved CH_4 concentration measured is presented. Monitoring borehole names at shale gas exploration sites correspond to names given in formal site documentation. Prospective area of the Bowland Shale from *Andrews* [2013].

Figure 3: Frequency distributions of dissolved CH_4 concentrations in the BGS and EA datasets. Dissolved CH_4 concentrations in the EA dataset below the general detection limit of 0.5 mg/l are assigned to the $0.1 \leq [CH4] < 1.0$ mg/l category, although their concentrations may be lower.

Figure 4: Dissolved CH_4 concentrations in groundwater from local baseline monitoring at the (a) Preston New Road and (b) Kirby Misperton shale gas exploration sites. Gaps show missing data.

Figure 5: Comparison of the credible intervals for the EA dataset using the strong and weak priors.

Figure 6: Predicted dissolved CH_4 concentrations for the EA dataset using the strong prior. Inset map shows an enlarged view of northwest England. Prospective area of the Bowland Shale from *Andrews* [2013].

Figure 7: Distributions of the modeled and observed dissolved CH_4 concentrations for the EA dataset. Note the x-axis has been limited to 4.0 mg/l even though the observed results extend beyond this.

Figure 8: Predicted dissolved CH_4 concentrations for groundwater monitoring boreholes at the (a) Preston New Road, and (b) Kirby Misperton shale gas exploration sites. Error bars show 95% credible intervals.

Figure 1

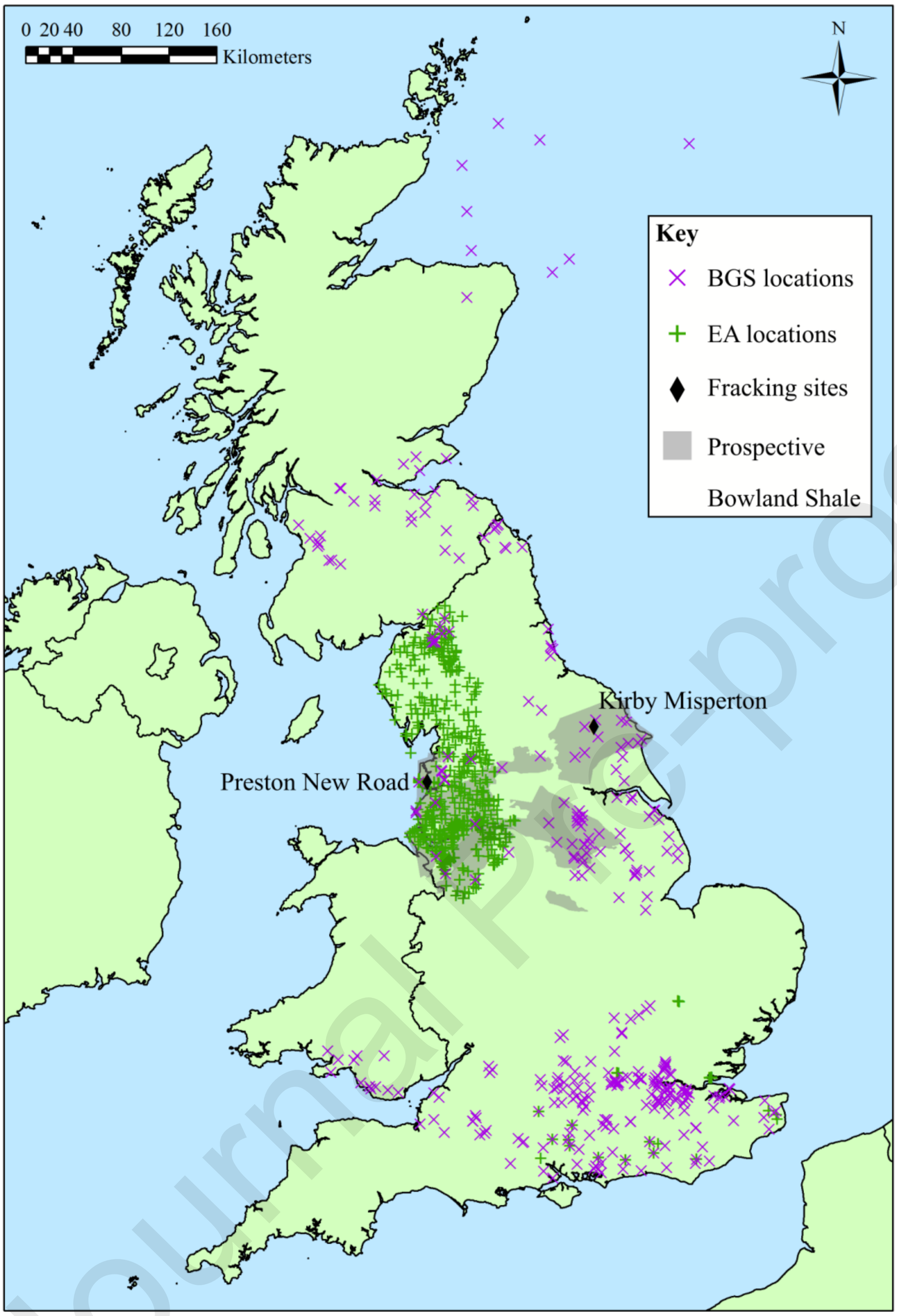

0 20 40 80 120 160
Kilometers
N
Key
BGS locations
EA locations
Fracking sites
Prospective
Bowland Shale
Kirby Misperton
Preston New Road

Figure 2

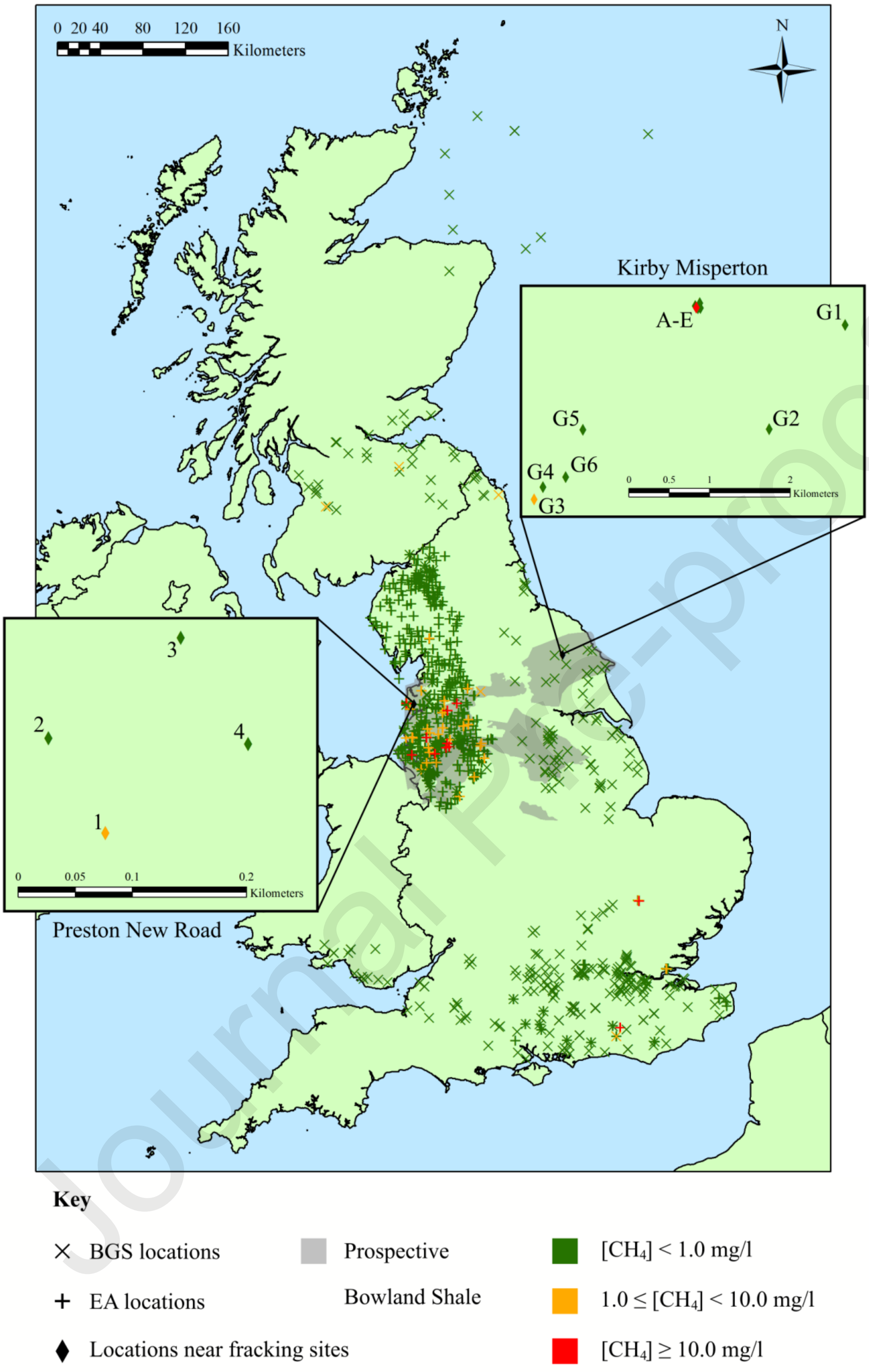

0 20 40 80 120 160
Kilometers
N
Kirby Misperton
A-E
G1
G5
G2
G4
G6
G3
0 0.5 1 2
Kilometers
3
2
4
1
0 0.05 0.1 0.2
Kilometers
Preston New Road
Key
× BGS locations
+ EA locations
♦ Locations near fracking sites
Prospective Bowland Shale
[CH$_4$] < 1.0 mg/l
1.0 ≤ [CH$_4$] < 10.0 mg/l
[CH$_4$] ≥ 10.0 mg/l

Figure 3

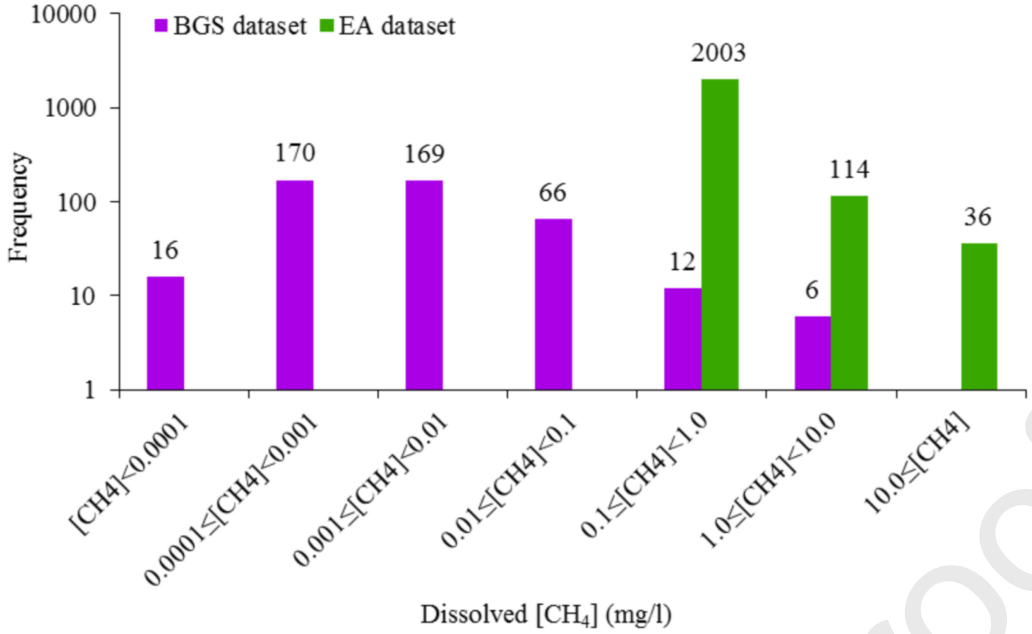

10000
1000
100
10
1
Frequency
BGS dataset
EA dataset
16
170
169
66
12
2003
6
114
36
[CH4]<0.0001
0.0001≤[CH4]<0.001
0.001≤[CH4]<0.01
0.01≤[CH4]<0.1
0.1≤[CH4]<1.0
1.0≤[CH4]<10.0
10.0≤[CH4]
Dissolved [CH4] (mg/l)

Figure 4

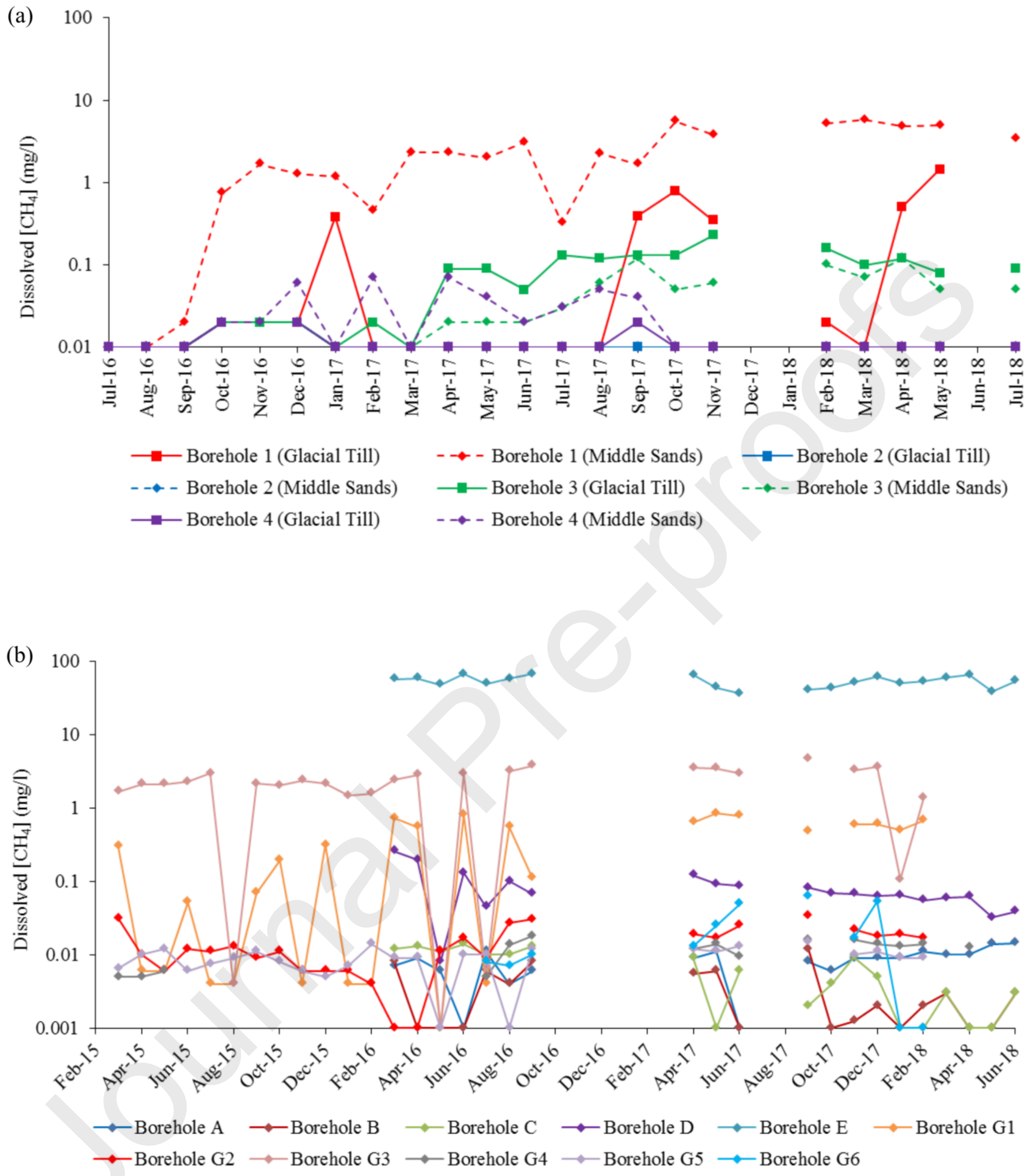

(a)
Dissolved [CH$_4$] (mg/l)
100
10
1
0.1
0.01
Jul-16
Aug-16
Sep-16
Oct-16
Nov-16
Dec-16
Jan-17
Feb-17
Mar-17
Apr-17
May-17
Jun-17
Jul-17
Aug-17
Sep-17
Oct-17
Nov-17
Dec-17
Jan-18
Feb-18
Mar-18
Apr-18
May-18
Jun-18
Jul-18
Borehole 1 (Glacial Till)
Borehole 1 (Middle Sands)
Borehole 2 (Glacial Till)
Borehole 2 (Middle Sands)
Borehole 3 (Glacial Till)
Borehole 3 (Middle Sands)
Borehole 4 (Glacial Till)
Borehole 4 (Middle Sands)
(b)
Dissolved [CH$_4$] (mg/l)
100
10
1
0.1
0.01
0.001
Feb-15
Apr-15
Jun-15
Aug-15
Oct-15
Dec-15
Feb-16
Apr-16
Jun-16
Aug-16
Oct-16
Dec-16
Feb-17
Apr-17
Jun-17
Aug-17
Oct-17
Dec-17
Feb-18
Apr-18
Jun-18
Borehole A
Borehole B
Borehole C
Borehole D
Borehole E
Borehole G1
Borehole G2
Borehole G3
Borehole G4
Borehole G5
Borehole G6

Figure 5

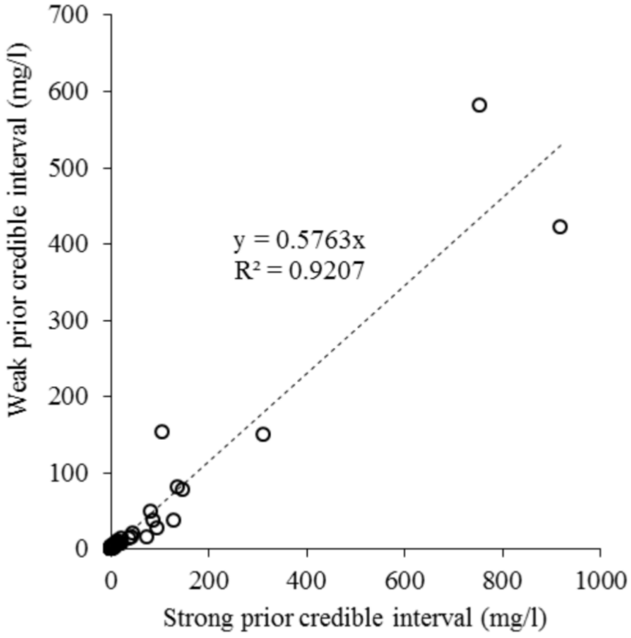

700
600
500
400
300
200
100
0
Weak prior credible interval (mg/l)
0
200
400
600
800
1000
Strong prior credible interval (mg/l)
y = 0.5763x
R² = 0.9207

Figure 6

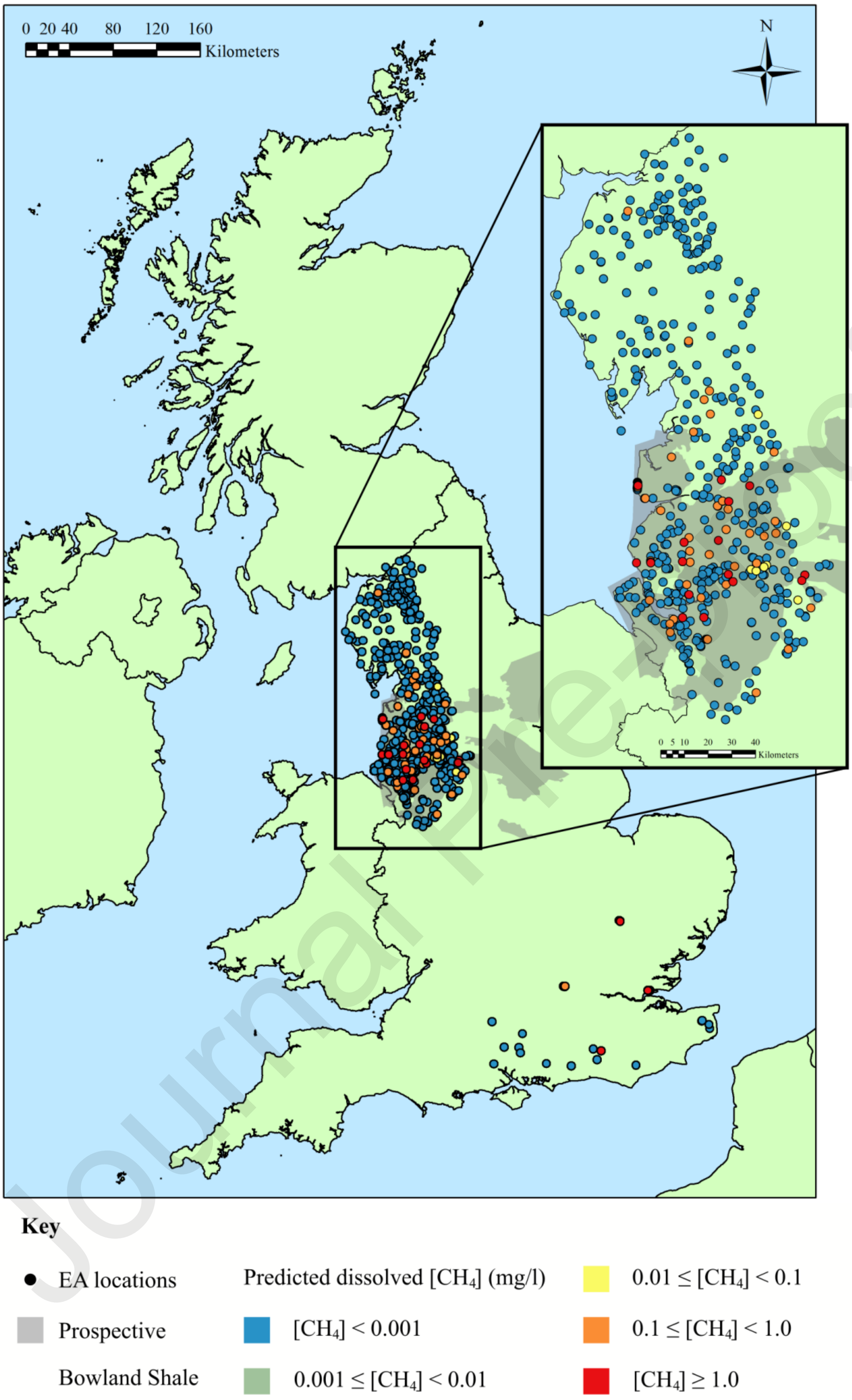

0 20 40 80 120 160
Kilometers
N
0 5 10 20 30 40
Kilometers
Key
● EA locations
Prospective
Bowland Shale
Predicted dissolved [CH4] (mg/l)
[CH4] < 0.001
0.001 ≤ [CH4] < 0.01
0.01 ≤ [CH4] < 0.1
0.1 ≤ [CH4] < 1.0
[CH4] ≥ 1.0

Figure 7

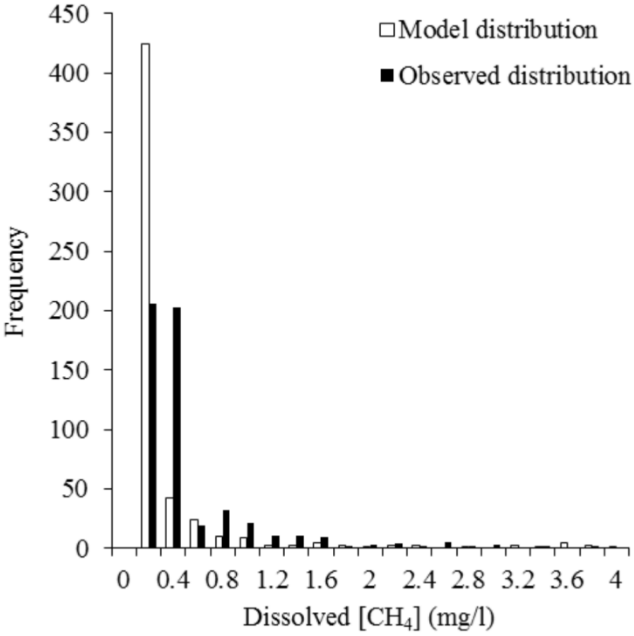
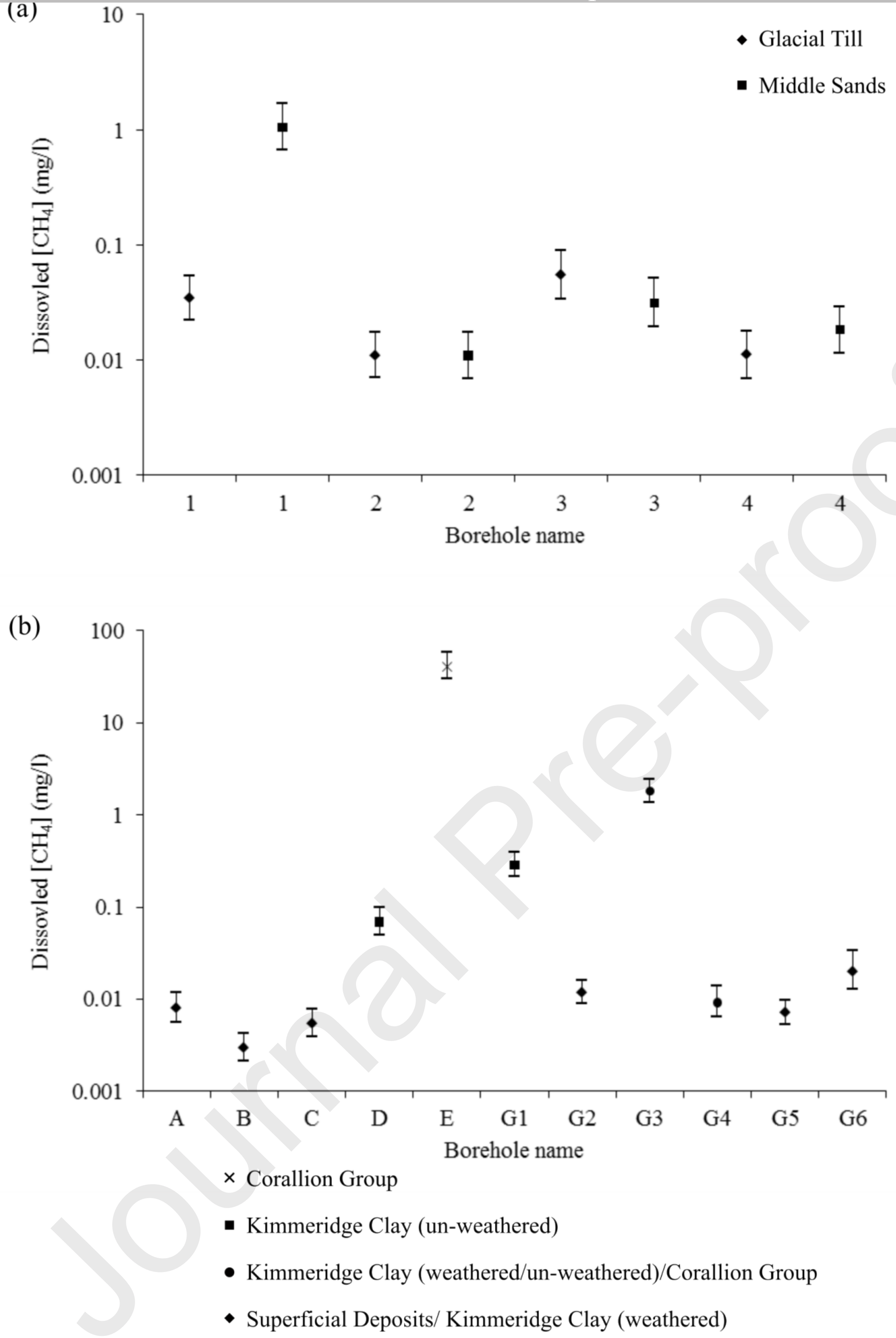

450
400
350
300
250
200
150
100
50
0
Frequency
□ Model distribution
■ Observed distribution
0
0.4
0.8
1.2
1.6
2
2.4
2.8
3.2
3.6
4
Dissolved [CH_4] (mg/l)

Figure 8


(a)
10
1
0.1
0.01
0.001
Dissovled [CH$_4$] (mg/l)
◆ Glacial Till
■ Middle Sands
1
1
2
2
3
3
4
4
Borehole name
(b)
100
10
1
0.1
0.01
0.001
Dissovled [CH$_4$] (mg/l)
A
B
C
D
E
G1
G2
G3
G4
G5
G6
Borehole name
× Corallion Group
■ Kimmeridge Clay (un-weathered)
● Kimmeridge Clay (weathered/un-weathered)/Corallion Group
◆ Superficial Deposits/ Kimmeridge Clay (weathered)

A dynamic baseline for dissolved methane in English groundwater

M. P. Wilson[a]*, F. Worrall[a], R. J. Davies[b], and A. Hart[c]

[a] Department of Earth Sciences, Durham University, Science Labs, Durham, DH1 3LE, UK

[b] School of Natural and Environmental Sciences, Newcastle University, Newcastle, NE1 7RU, UK

[c] Environment Agency, Research Assessment and Evaluation, Sapphire East, Streetsbrook Road, Solihull, B91 1QT, UK

* Corresponding author: Miles Wilson (miles.wilson@durham.ac.uk)

Tables

Table 1: Descriptive statistics of dissolved CH_4 concentration measurements in the BGS and EA datasets.

	BGS dataset	**EA dataset**
Number of dissolved CH_4 concentration measurements	439	2153
Number of unique sampling locations	336	571
Mean number of dissolved CH_4 measurements per unique location	1.3	3.8
Detection limit (mg/l)	<0.0005	<0.5
Maximum dissolved CH_4 concentration (mg/l)	4.72	25.9
Number of dissolved CH_4 measurements ≥10.0 mg/l	0	36

Table 2: Descriptive statistics of unique locations in the EA dataset where dissolved CH_4 concentrations have been measured ≥10.0 mg/l.

EA location name	**Geographical area**[a]	**Number of dissolved CH_4 measurements**	**Number of dissolved CH_4 measurements ≥10.0 mg/l**	**Maximum dissolved CH_4 concentration (mg/l)**
30006 ad-hoc @ Holiday Moss, Reeds Brow	GMMC	18	12	25.9
30335 ad-hoc @ Bromborough Dock North	GMMC	9	2	12.0
60301 ad-hoc @ Moss Side Farm, Rixton	GMMC	25	1	11.1
60565 ad-hoc @ Arpley Landfill Site, Arpley Meadows	GMMC	37	5	13.6
Blackburn Yarn Dyers ABH	CL	11	3	21.5
Blackpool Pleasure Beach No.1 borehole	CL	3	1	13.9
Brentwood Moss Nurseries Borehole	GMMC	10	4	13.5
Indigo Yarn Company Ltd. borehole	CL	1	1	10.0
Lower Stumble groundwater borehole	SSD	4	3	11.4
Sawston Inv.Spikefields BGS SCO4A 35M BH	CB	2	1	12.4
Whitegate Farm borehole	GMMC	3	3	14.3
Total		123	36	

[a] Geographical area codes are given in Table S4.

Table 3: Results of the binomial probability modeling for the 11 paired locations.

Indicator	**Number of independent trials**	**Number of successes (N)**	**Probability at most N successes (%)**
NH_4	11	8	96.7
BOD	1	1	100.0
COD	9	6	91.0
DO	4	4	100.0
Fe	11	8	96.7
Mn	11	7	88.7
NO_3^-	5	3	81.3
SO_4^{2-}	11	7	88.7

Table 4: Results of the binomial probability modeling for the seven paired locations sampling the inferred same groundwater bodies.

Indicator	**Number of independent trials**	**Number of successes (N)**	**Probability at most N successes (%)**
NH_4	7	4	77.3
BOD	1	1	100.0
COD	5	3	81.3
DO	3	3	100.0
Fe	7	5	93.8
Mn	7	5	93.8
NO_3^-	4	3	93.8
SO_4^{2-}	7	5	93.8

Table 5: Results of the ANOVA without covariates and values of the percentage change in mean dissolved CH_4 concentration when going from the low to elevated CH_4 locations (only given for those indicators where a significant difference has been demonstrated). P-values and R^2 values underlined in italics are those found to be significantly different from zero at the 95% probability.

Indicator	**P-value**	**R^2 (%)**	**% increase to elevated CH_4 location**
NH_4	*<0.0005*	*14.8*	*-90*
BOD	0.414	0.6	
COD	*<0.0005*	*15.6*	*28*
DO	0.720	0.1	
Fe	*<0.0005*	*13.0*	*22*
Mn	*<0.0005*	*6.3*	*17*
NO_3^-	*0.003*	*11.0*	*99*
SO_4^{2-}	0.620	0.1	

Table 6: Comparison of published baselines for dissolved CH_4 concentrations in groundwater.

Region and study	**Number of dissolved CH_4 concentration measurements**	**Number of unique locations**	**Detection limit (mg/l)**	**Minimum dissolved CH_4 concentration (mg/l)**	**Median dissolved CH_4 concentration (mg/l)^a**	**Mean dissolved CH_4 concentration (mg/l)^a**	**Maximum dissolved CH_4 concentration (mg/l)**	**Number and percentage of dissolved CH_4 concentration measurements ≥10.0 mg/l**

New York State, USA [*Kappel and Nystrom*, 2012]	239	239	0.001	<0.001	<1.0	Unknown	41.0	21 (8.8%)
Susquehanna County, USA [*Molofsky et al.*, 2013]	1701	1701	0.00005-0.026	<0.00005	0.00076	0.705	43.0	50 (2.9%)
Marcellus Shale, USA [*Darrah et al.*, 2014]	113	113	Unknown	<Detection limit	0.585[b]	8.040[b]	102.164[b]	32 (28.3%)
Barnett Shale, USA [*Darrah et al.*, 2014]	59	20	Unknown	0.716[b]	5.519[b]	13.242[b]	46.962[b]	23 (39.0%)
Sullivan County, USA [*Reese et al.*, 2014]	1882	1882	0.005-0.026	<0.005	0.149	Unknown	16.5	≤15[c] (0.8%)
Wattenberg field, Colorado, USA [*Li and Carlson*, 2014]	223	176	Unknown	<Detection limit	Unknown	4.0	37.1	36 (16.1%)

New York State, USA [*McPhillips et al.*, 2014]	113	113	0.01	0.002	0.007	0.464	8.26	0 (0.0%)
Southwestern Ontario, Canada [*McIntosh et al.*, 2014]	1010	862	1	0	Unknown	Unknown	248	150 (14.9%)
St. Lawrence Lowlands, Quebec, Canada [*Moritz et al.*, 2015]	138	130	0.0006	<0.0006	0.1	3.8	45.9	17 (12.3%)
Chaudière-Appalaches, Canada [*Lefebrve et al.*, 2015]	Unknown	74	Unknown	Unknown	0.2	5.1	31.2	Unknown
Lower Saxony, Germany [*Schloemer et al.*, 2016]	1043	1043	0.0000178[d]	0.0000178[d]	0.00183[d]	Unknown	44.0188[d]	60 (5.8%)
Alberta, Canada [*Humez et al.*, 2016]	179	186	3.333	<3.333	<3.333	5.232	42.9	28 (15.6%)

Great Britain, UK [*Bell et al.*, 2017]	439	336	0.00005-0.005	<0.00005	0.0012	0.0465	4.72	0 (0.0%)
Saint-Édouard, Quebec, Canada [*Bordeleau et al.*, 2018]	Unknown	48	0.006	<0.006	4.9	10.4	82.0	Unknown
EA dataset, England This study	2153	571	0.1-0.5	<0.1	<0.5	0.24[e]	25.9	36 (1.7%)

---

[a] May include concentrations below detection limit taken as the detection limit concentration, i.e. "<0.5 mg/l" was taken as "0.5 mg/l". [b] Calculated using a CH_4 density of 0.657 mg/cm^3 at 1 atm and 298 K. [c] Fifteen wells with CH_4 concentrations ≥7.0 mg/l. [d] Calculated using 0.71 mg/l = 1.0 ml/l at 293 K and 101325 Pa [*Schloemer et al.*, 2016]. [e] Expected mean from the Bayesian generalised linear modeling.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: