Authors' responses to the reviews of "Bayesian atmospheric tomography for detection and quantification of methane emissions: Application to data from the 2015 Ginninderra release experiment"

(https://doi.org/10.5194/amt-2019-124)

1 Reviewer 1

Overall, I think this article is well thought-out, well written, and addresses an important problem in source quantification. I have just a handful of suggestions and ideas for the authors.

We thank the reviewer for taking the time to read and comment on the manuscript, and for suggesting improvements to the paper where needed.

1.1 Broader ideas and suggestions

I think it can be difficult to figure out which Copernicus journal is the best fit for an article like this one. I have also seen several articles on inverse methods in Geoscientific Model Development (GMD), another Copernicus journal. I dont think it would be worth switching this article over to GMD at this juncture, but I do think it would be helpful to get guidance from the Copernicus editors on which journal they would recommend for future articles on inverse modeling.

The ICDC10 Special Issue is an inter-journal issue which includes, amongst others, AMT. We believe the results in this manuscript are of relevance to the measurement-techniques community, since we consider the use of various instruments, in isolation and in combination, for detecting and quantifying methane emissions. We defer any decision on whether another journal in the issue is deemed a better fit to the AMT and Special Issue editors.

I would be curious to see how the model results would change if you excluded observations collected when wind speeds are low. Many existing articles simply exlude observations that are collected when wind speeds are low, and the authors point out that this decision and the wind speed cutoff are subjective. I would argue that the choice of weighting function for the low wind speed observations is also somewhat subjective.

The reviewer is correct in that both procedures are somewhat subjective. However, there are several advantages to downweighting the observations rather than excluding them. First, as we

show next, while low wind speed observations are somewhat unreliable, they still contain information of use and which could be important when optimising parameters within the Gaussian plume dispersion model and when doing flux inversion. Second, the way in which we downweight is motivated by transport-model considerations. In this sense we view downweighting as less subjective than excluding the observations entirely. We elaborate further in our reply to the next comment.

I would be curious if these low wind speed observations contributes anything to the emissions estimate relative to the traditional approach of exlcluding those observations. I think it could be interesting to run a case with these observations excluded (if and only if it is not a lot of work to do).

We have run the inversion using the EC tower data, during both release periods, and excluding all observations with wind speeds below 1 m s^{-1} (which is also the value we use as a threshold in our downweighting scheme). The left panel of Fig. 1 in this response shows the original 95% posterior credible intervals obtained when downweighting observations with low wind speeds, and the right panel of Fig. 1 in this response shows the 95% posterior credible intervals when excluding the observations. There is only a small change between the results during the 5.8 g min^{-1} release period. However, in the case of the 5.0 g min^{-1} release period, there is a notable difference between the two posterior distributions. This is probably because there are far fewer observations during this release period than during the 5.8 g min^{-1} release period. For the 5.8 g min^{-1} release period, 2967 observations were included in the inversion which downweighted low wind speed observations, and 2460 observations were used when low wind speed observations were excluded. During the 5.0 g min^{-1} release period, 475 observations were used when low wind speed observations were downweighted, compared to 339 observations when low wind speed observations were excluded. The smaller amount of information available in the 5.0 g min^{-1} release period, along with so many unknown parameters to optimise, appears to have had a deleterious effect on the results.

We have also run the inversions for the two approaches with the threshold increased to 1.5 m s^{-1}. The results obtained using downweighting are shown in the left panel of Fig. 2 in this response. We see that the 95% posterior credible intervals for both release-rate periods have shifted by a small amount closer to the true release rate. The results for when data are discarded completely are shown in the right panel of Fig. 2 in this response. The 95% posterior credible interval is slightly wider than that of the downweighting method for the 5.8 g min^{-1} release period. In both cases, the posterior distributions do not seem to be overly sensitive to the choice of threshold.

However, the Markov chains for the 5.0 g min^{-1} release rate period and 1.5 m s^{-1} threshold did not converge when ω_y and ω_z were treated as unknown and data were completely discarded, probably due to the lower amount of available observations. The results shown are therefore conditional on these parameters being fixed to the maximum a-posteriori estimates obtained when downweighting. This lack of convergence gives us more reason to believe that downweighting is preferable to discarding observations entirely.

We have now modified the first paragraph of Sect. 3.2 in the manuscript to indicate that (i) in separate studies we have seen that our posterior inferences are not overly sensitive to the choice of threshold used and that (ii) downweighting instead of discarding is particularly important when there are not many observations at high wind speed available for inference.

I think it could be useful to include a greater discussion of when and how this statistical model could be applied to other inverse problems. The Ginninderra

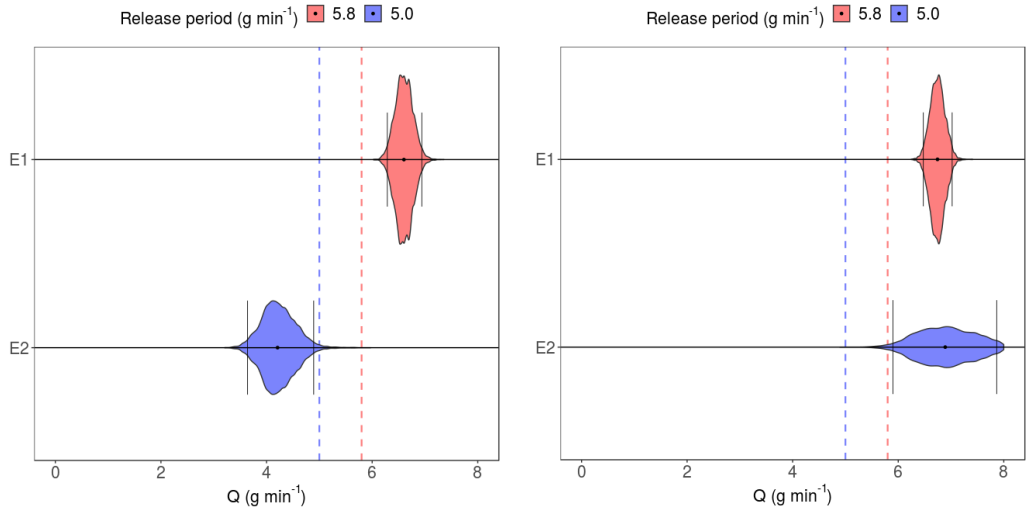

Figure 1: Left: posterior empirical distributions of the emission rate Q in g min^{-1}, for the EC towers (E), for each release-rate period (1 and 2) during the Ginninderra experiment. Observations with wind speeds below 1 m s^{-1} have been downweighted. The 5.8 g min^{-1} release-rate period is shown in red (E1), while the 5.0 g min^{-1} release-rate period is shown in blue (E2). The vertical dashed lines denote the respective true emission rates, the black dot represents the median estimate, and the vertical black bars represent the upper and lower limits of the 95% posterior credible interval. Right: same as left, however observations with wind speeds below 1 m s^{-1} have been removed entirely.

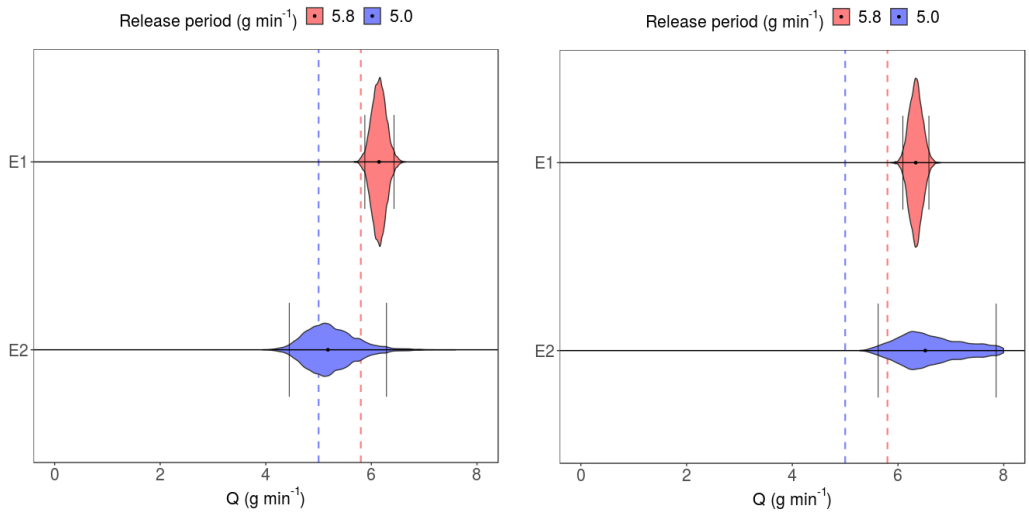

Figure 2: Same as Fig. 1, however in the left panel observations with wind speeds below 1.5 m s^{-1} have been downweighted, and in the right panel observations with wind speeds below 1.5 m s^{-1} have been removed entirely.

Controlled Release experiment is a relatively simple and controlled emissions set up, and most real world source quantification problems are likely to be more complicated. For example, there may be multiple leaky natural gas wells in a study area instead of a single point source. I think it could be helpful to lay out for the reader what types of problems you think this model would be well-suited for.

The methods developed in this study are ideal for quantifying local-scale leaks from industrial facilities or from the subsurface (e.g., well heads, buried pipelines or gas leakage up geological fractures and faults) where a surface leak has been detected but needs to be quantified. It can be used where physical access to the source location is limited, e.g., gas bubbling from a creek or where measurement is hazardous. Depending on the circumstance, detection of leakage can take many different forms, from visible bubble detection, optical gas imaging, handheld sniffers, noise detection, helicopters equipped with lasers, drones equipped with gas sensors, to monitoring die-off in vegetation using remote sensing techniques. Surface leakage typically expresses as small, concentrated hotspots if sourced from the subsurface (Feitz et al., 2014; Forde et al., 2019), for which the quantification approach outlined in this article ideally suited. Equipment placement can be optimised around the leakage site (i.e., prevailing upwind/downwind) for optimal quantification.

In most applications the number of sources, nor the source location, is known. As such, the framework we construct should be seen as a foundational building block that needs to be extended appropriately for each specific application. For example, if the source location is not known, then source localisation can be incorporated into the Bayesian framework as discussed by Humphries et al. (2012). If there are multiple possible sites, and these locations are not known, then the framework needs to be further extended to incorporate multiple Gaussian plume models (one for each site), and joint localisation/inversion will be required. While these extensions are straightforward both mathematically and computationally, in practice they are unlikely to be effective for detection of leakage over large spatial scales. Gas fields or geological storage sites can cover areas of tens to hundreds of square kilometres. Unless there is a high density of sensors ($\sim$ 100 m scale) (Jenkins et al., 2016; van Leeuwen et al., 2013), the sensitivity of detection will be poor (Luhar et al., 2014; Wilson et al., 2014). It is however relatively straightforward to effectively extend the methodology to when the emission is from an area, rather than a point source.

We now provide this discussion in Sect. 7 of the manuscript.

1.2 Specific line-by-line suggestions

Abstract, line 3: Could you be more specific with the term "misspecifications"? What kind of misspecifications appear at each stage of the inversion processing chain? I think some people in the inverse modeling community may also feel offended by the phrasing of this statement, especially given the very broad scope of this statement. Instead of phrasing the abstract this way, it could be more effective to explain why certain inverse problems are so challenging.

We did not intend to cause any offence: The term "model misspecification" has a technical meaning in statistical science, and refers to the fact that all statistical models (especially simple ones, like the one we are considering) are only approximations to the truth. For example, Rao (1971) states that model "[m]isspecification can arise either because of omission of a variable specified by the truth, [...] or because of inclusion of a variable not specified by the truth[.]" In addition to the statistical model being misspecified, it is also likely that the Gaussian plume

model is misspecified, in the sense that it does not (and can not) precisely capture the true dispersion of methane. To avoid ambiguity we have now used the word 'assumptions' instead of 'misspecifications' in the abstract.

Abstract, line 6: The word tomography is not common in the atmospheric inverse modeling community. I think I know what you mean, but it could be worth clarifying somewhere in the manuscript.

Atmospheric tomography was originally used by Humphries et al. (2012) due to its similarity to "tomographic imaging", where the aim is to detect and quantify the strength of sources in a medical context. We use this term to specify which problem of flux inversion (which often refers to the general problem of estimating fluxes over a region) we are concerned about in this manuscript. We have included the following sentence briefly describing the atmospheric tomography technique in the second last paragraph of Sect. 1 of the manuscript to clarify:

Atmospheric tomography, a term inspired from medical imaging, combines data from a collection of measurement sites with Bayesian inversion to detect and quantify emissions.

Abstract, line 9: Perhaps inverse modeling would be better than inversion?

Throughout the manuscript we use "inversion" synonymously with "inference", which we see as distinct from the "model" (or "inverse model"), which comprises the set of equations on which our inferences are based. We thus intentionally refrain from using the term "inverse model" in this context.

Page 2, line 9: Consider moving the clause at the beginning of the sentence to the end of the sentence. I think this change might make the sentence flow better.

This sentence has now been re-worded.

Page 2, line 29-31: This sentence has a relatively complex structure that makes it a bit difficult to read. You might consider simplifying this sentence or splitting it into two sentences for easier readability.

We now broke the sentence into two smaller sentences such that one key point is made in each sentence, as suggested. The first notes that online calibration of model parameters is not often done, while the second suggests a reason for this.

Page 3, line 4-5: I dont quite follow the point being made in this sentence. I know that it is important to evaluate atmospheric transport as part of inverse modeling, but I am not sure how that task is naturally part of classical inverse theory.

We apologise for the lack of clarity in this sentence. Evaluating the transport model is important for inverse modelling, but actually it is "calibrating the model from observations" that fits squarely within the Bayesian framework. The sentence has been rephrased to make this clearer.

Page 5, line 9: What is a Pasquill stability class? I am not positive that your reader will know this term.

The Pasquill stability classes are a collection of stability categories into which each observation can be placed. The categorisation of an observation is based on the values of the meteorological parameters at the time (for example, wind speed, wind direction, air temperature, air pressure), and how these values then affect the horizontal and vertical dispersion of the atmospheric particles. The more the particles are expected to spread in the horizontal and/or vertical

directions for a particular observation, the more unstable the Pasquill stability class associated with that observation. More details on the classes are given in Pasquill (1961), referenced in Sect. 3.1 of the manuscript. We have now added a note in Sect. 1, paragraph 3 that both the stability classes and the Monin-Obhukov length are described later in the manuscript. Further, in Sect. 3.1, we have now included a short description of the Pasquill stability class along with the reference to Pasquill (1961).

Page 6, line 8: What is a Monin-Obukhov length?

The Monin-Obukhov length is the theoretical height at which turbulence is produced by buoyancy and mechanical forces in equal amounts (Sienfeld and Pandis, 2006, Chapter 16). Hence, it is a continuous measure of stability. We have included a brief explanation of the L-value in the manuscript in Sect. 3.1.

Figure 2: What message should readers take away from this figure? It could be useful to include a one-sentence takeaway message in the caption.

The following sentence has been added highlighting the takeaway message:

Of the three, the best agreement between predicted and observed values occurs when σ_{y_i,k_i} is scaled by 2.5.

Fig. 3: The directional errors are difficult to see in this figure. Consider making the arrow directions larger.

Thank you, the size of the arrows has now been increased.

Pg 11, line 2: Is a half-normal distribution the same as a truncated normal distribution?

The half-normal distribution is a special case of the truncated normal distribution. If a normal distribution with mean zero is truncated from below at zero, then it is a half-normal distribution. We have clarified this in Sect. 4.2 of the manuscript.

Figure 5: There is a lot going on in this figure. It could be helpful to the reader to state the main takeaway message of this figure somewhere within the caption.

Thank you, the following sentence has been added highlighting the takeaway message:

We can recover a reasonable range of estimates for the emission rate, with no 95% posterior credible interval being far from the true emission rate. Further, we see that the posterior emission rate credible intervals move towards zero when the source is inactive, as desired.

2 Reviewer 2

This study investigates the performance of an inverse modelling algorithm for estimating greenhouse gas emissions from a single point source using data from a controlled release experiment. Inverse modelling methods are widely used to quantify emissions addressing a large range of scales. The advantage of the small scale investigated here is that true emission can be known, allowing a direct evaluation of the inversion performance. Usually this is not possible and the performance can only be evaluated indirectly. The outcome is rather sobering, emphasizing the difficulty to obtain reliable emission estimates despite the favorable availability of data from different types of instruments. An attempt is made to identify the most important factors limiting the performance.

We thank the reviewer for taking the time to read and comment on the manuscript, and for suggesting improvements to the paper where needed.

As explained in further detail below, it is unclear what we learn from this study that was not already known before. In part this is because the link is missing with earlier work, and how the performance that is achieved here compares with what was done in the past. It remains unclear also how well the optimized model is fitting the data and what would be needed to further improve the results. Further efforts in these directions will be needed to make this work publishable.

We thank the reviewer for the review. There are several concerns raised in the introduction to the review, namely concerning

(i) Clarity on what we learn on the study;

(ii) How this compares to what is done now and in the past;

(iii) The quality of the fits; and

(iv) What could be done to improve the results.

Issues (i)–(iv) appear in the "General Comments" and "Specific Comments" below, and are addressed in our responses there.

2.1 General comments

This study is following up on the study of Feitz et al (2018), in which not only the measurement techniques are described in detail but also different methods are used for emission quantification. That study is referenced, but it remains unclear how the method in this study relates to what was done before, and how the results compare.

This comment concerns issue (ii) raised in the introduction to this review. This study builds upon the atmospheric tomography technique described in Section 2.4.2 of Feitz et al. (2018). The method has been refined in the current study through the "online calibration" of parameters within the atmospheric transport model. Further, in the previous study, atmospheric tomography was only used on the Boreal lasers, and the methodology was tailored to suit those specific instruments. The methodology we present accounts for model error and instrument-specific bias, and is applicable to both point- and path-measurements. It is relatively robust

to the type of instrument, with only the decision to include ω_y as an unknown parameter dependent on whether the observation is a point or path measurement.

We have now re-worded the second last paragraph in Sect. 1 of the manuscript to better detail the connection of this study with the previous study.

Besides the Ginninderra release experiment, other similar studies have been conducted in the past. To keep track of progress, and make sure the recommendations of those studies are taken into consideration it is important to make a closer connection to them and compare the performance that is achieved here.

This comment concerns issue (ii) raised in the introduction to this review. The reviewer is correct that other similar studies have been carried out, and we cite two of these in the manuscript: that of Humphries et al. (2012) and that of Luhar et al. (2014). A more comprehensive list of controlled release experiments and related studies has now been included in the second paragraph of Sect. 1. In addition, we have now included a paragraph detailing the connections of our work to those of Ars et al. (2017); Humphries et al. (2012); Luhar et al. (2014) in Sect. 7 of the manuscript:

Our work is closely connected to other atmospheric tomography techniques, but with some small, significant, differences. Luhar et al. (2014) used a backward Lagrangian particle model to simulate the trajectories of methane and carbon dioxide backwards in time to localise the source and estimate the emission rates. Their approach yielded good quality estimates for the methane emission rates, but highly uncertain estimates for the carbon dioxide emission rates and source location parameters. Twenty-three runs of the Lagrangian model required approximately one hour of computing time, and therefore their framework becomes problematic with thousands of observations as we have in our study. More pertinently, online calibration of the atmospheric-transport model would be virtually impossible without the construction and use of a surrogate model or emulator (e.g., Harvey et al., 2018). In the study of Humphries et al. (2012), carbon dioxide and nitrous oxide emission rates and source locations were estimated relatively well. We do not consider the localisation problem, but otherwise extend their method to handle various instrument types and a number of extra levels of uncertainty. The case in our sensitivity analysis in which we fix $\omega_y = \omega_z = 1$ yields a model that is structurally very similar to that of Humphries et al. (2012); we see from our results that having this hard constraint is not a tenable assumption in practice. Our work also has close connections with that of Ars et al. (2017) where the Pasquill stability class for an observation is chosen from a subset of appropriate stability classes, based on the best fit of model predicted values to observed values. While this may help fit the Gaussian plume dispersion model to the data, it does not take into account the uncertainty arising from stability-class choice. Further, if all plume model standard deviations are off by a factor of two or more, there is a distinct possibility that no stability class yields a good fit. Online calibration of these standard deviations is needed to account for lack-of-fit arising from the the inherently simple Gaussian plume model.

This study arrives at the expected outcome that the performance of the inversion improves when the stability parameters of the Gaussian plume model are optimized. However, the finding that the performance of the OSSE is so much better than the results of the experiment using real data, despite using realistic settings in the OSSE, points to a significant remaining problem with the model. Given the simplicity of the Gaussian plume formulation this may not come as a surprise. Nevertheless, some further analysis of fit residuals is needed to find out what prevents the inversion from finding the right answer. Could it be as simple as the assumed

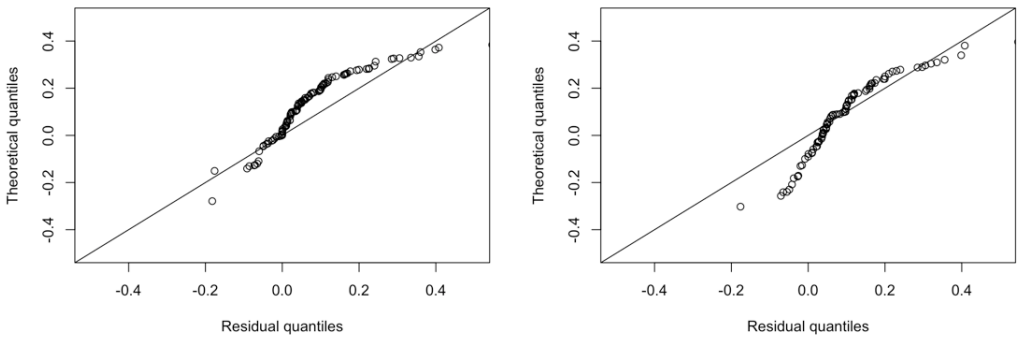

Figure 3: Q-Q plots of the residual (observed enhancements minus the Gaussian plume dispersion model predicted values). All data is from EC.A, without wind speeds below 1 m s^{-1}, as these are not assumed to have constant variance. The left panel are observations from stability class A, and the right panel are observations from stability class F. Deviations away from the solid line indicate quantile deviations from the Gaussian distribution of same mean and variance.

wind direction or speed being wrong? I didnt find back an exact specification of the information that was used for that.

This comment concerns issue (iii) raised in the introduction to this review. It is important to note that the OSSE solely aims to verify that, under ideal conditions and no model misspecification, the inferential algorithm performs as it should. Indeed, one should not expect that results from real data are as good as those obtained in an OSSE. For example, it is likely that the Gaussian plume model predicted output is not valid for every observation we consider. The most we can hope for is a framework that is relatively robust to model misspecification, and this seems to be the case since our posterior inferences corroborate quite well with the truth.

The reviewer suggests that the wind data might be wrong. In the Ginninderra experiment weather information was recorded on site at EC.A, which was equipped with a number of meteorological instruments, as detailed toward the end of the first paragraph in Sect. 2 of our manuscript. In a separate study we had assumed that the wind speeds were uncertain. This dramatically increased the computational time required for MCMC, but had a negligible effect on the results. However, we have reason to believe that the wind speed and wind direction data are of good quality; we give more details in our response to the reviewer's final Specific Comment in this document.

Figure 3 in this response shows a Q-Q plot (a plot of theoretical quantiles against the observed quantiles) of the residuals obtained during the 5.8 g min^{-1} release-rate period, for EC.A and stability classes A and F, and excluding observations with wind speeds below 1 m s^{-1} (this is necessary because the variance is modelled to be wind-speed dependent below this threshold). The residuals suggest that there is a mild deviation away from our Gaussianity assumption. However, with real data, such mild deviations are to be expected, and the deviations do not appear large enough to be of concern. We have now noted that we have looked at the Q-Q plots and observed mild deviations from Gaussianity in the residuals in Sect. 5.2 of the manuscript (third paragraph).

How appropriate is the use of a Gaussian plume model in this experiment?

As mentioned in Sect. 1, paragraph 4 of our manuscript, the Gaussian plume model is known to

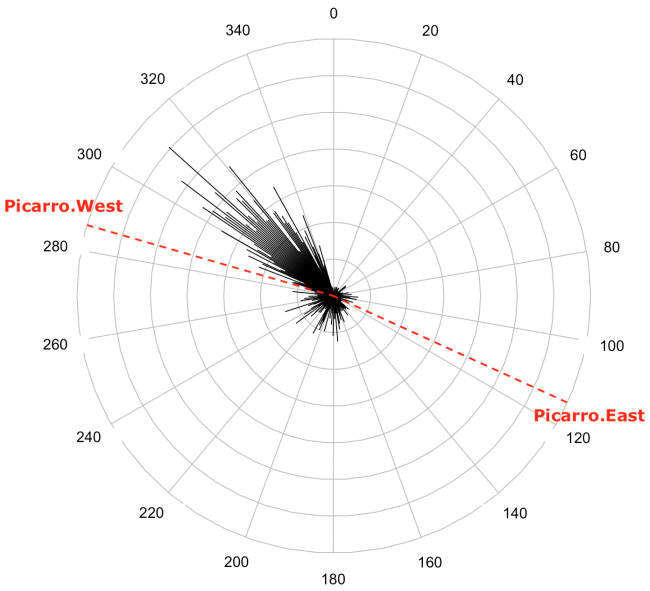

Figure 4: Radial histogram showing the direction *from* which the wind is blowing during the 5.8 g min^{-1} release-rate period.

work well over short distances/small areas, such as the area used in the Ginninderra experiment (e.g., see Riddick et al., 2017). Moreover, the methodology we propose is aimed at applications where near-real time leak detection is important. In order to do this, computationally fast methods are necessary; an attraction of the Gaussian plume model is that it is very quick to evaluate and calibrate.

Different measurement techniques are compared, but there is very limited discussion on the best technique for inferring emissions. What would be the recommendation for monitoring leaks?

This comment concerns issue (i) raised in the introduction to this review. We see in our inversion results that the wind direction plays a large role in our ability to obtain reasonable estimates for the emission rate. For example, Fig. 4 in this response shows that the dominant wind directions during the 5.8 g min^{-1} release-rate period range from approximately 300 to 340 degrees east of north. This suggests that very little of the released methane would have blown over Picarro.West, for example, while much of it would have blown over Picarro.East. Indeed, when the inversion is run on each of these instruments separately during this release-rate period, we see that a reasonable range of emission rate estimates is produced using Picarro.East. On the other hand, running the inversion using only Picarro.West returns the prior distribution set on Q, meaning these observations are not at all informative of the emission rate.

Therefore having more (less expensive) instruments set up to cover many more possible wind directions appears to serve better than having only one or two more expensive instruments. That is, unless the wind direction is favourable during a good portion of the experimental time frame, better coverage of the experimental field using less expensive/accurate instruments is likely to outperform a single, very expensive/accurate instrument when it comes to detecting and quantifying the emission rate from a point source. If one is limited to using a small number of instruments, then path measurements are more suitable than point measurements, since they are able to 'capture' a larger range of wind directions. These points have now been added to Section 7 of the manuscript (penultimate paragraph).

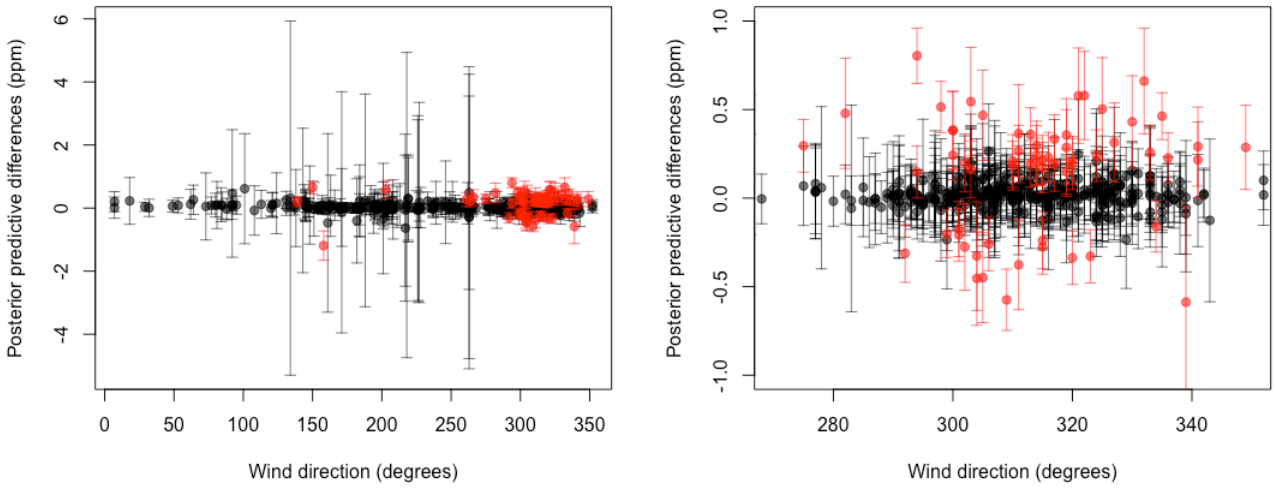

Figure 5: Left: Empirical posterior predictive intervals (mean ± 1SD) for the differences between Gaussian plume dispersion model predicted values and observed enhancements, using the 20% of EC Tower data during the 5.8 g min^{-1} release-rate period, plotted by wind direction. Those intervals in red do not contain zero. Right: The same plot zoomed in on wind directions between 280 and 360 degrees.

It would also be an option to use data that are not used in the inversion for evaluating the optimized concentrations.

This comment concerns issue (iii) raised in the introduction to this review. Controlled release experiments are unique in that the source location and emission rate are known. Since our inferential target in this work is the emission rate, and we can compare our inferences directly to the true value, we do not provide an analysis on the optimised concentrations in the manuscript. However, in response to this comment, we have now re-run the inversion on 80% (randomly selected) of the observations from the EC Towers during the 5.8 g min^{-1} release-rate period. We then sampled 100 values for ω_y, ω_z, $\boldsymbol{\tau}$, and Q from the MCMC trace plots, and used these to compute empirical posterior predictive intervals (mean ± 1SD) for the difference between Gaussian plume model predicted concentrations and the observed enhancements on the remaining 20% of EC Tower data during this release-rate period (Gelman et al., 2014, Chapter 6). Figure 5 shows these posterior predictive intervals plotted by wind direction. Those intervals which do not contain zero (which would suggest a poor fit) are plotted in red, while those that do contain zero are plotted in black. Note how the intervals widely vary in size, due to the heteroscedasticity introduced in the measurement-error model that accounts for wind speed and stability class. Of the 594 intervals, 526 (≈ 89%) contain zero. When the intervals are extended to ± 2SD, 578 (≈ 97%) contain zero. Considering the practical difficulty of exactly quantifying uncertainty in these applications, these values compare reasonably well to what we would expect theoretically (68% and 95%, respectively).

2.2 Specific comments

Page 1: Introduction section: Here I was missing some information about the specific application of inverse modelling that is studied here, among the large range of applications discussed in literature. Special for this case is the small scale of the experiment and the known location of the emission source. Usually this is not the case, raising the question for which kind of application this would be relevant (you might argue that there are easier methods to monitor emissions when the source

is known).

This comment concerns issue (i) raised in the introduction to this review. Please see our reply to Reviewer 1's comment on this same point on Pg. 4 of this response.

Page 6, line 22: '. . . serves to scale outputs vertically . . . ' In the end I understood that 'vertically' referred to the y-axis in figure 2. What is meant is that Q scales the concentration enhancements. Please rephrase to make this clearer.

We have now rephrased the sentence to indicate that by scaling we mean multiplying by a constant factor.

Page 7, line 18: I'm missing an explanation of the logic here. Please clarify the reason for quantifying the statistics of 1/U.

From Eq. (1) in the manuscript we see that, when conditioned on all other values, the Gaussian plume dispersion model is proportional to the inverse of the wind speed, and hence, the variance of a predicted concentration is proportional to the variance of the inverse of the wind speed. We use this relationship to establish a wind-speed dependent variance for the concentrations. We have expanded the explanation in Sect. 3.2 to make this clearer.

Page 9, line 7: '. . . graphically in Fig. 3 . . . ' Here the dependent variables of the optimization are introduced, but not explained. Here the meaning of tau and the omega's should be explained explicitly, and that in addition to these variables Q is estimated from the data.

These variables are explained later on in the text while the diagram is there to guide the reader through the subsequent sections. We have now given a brief definition of the symbols upon their first appearance in Sect. 4, and noted the section numbers in which they are each described in more detail.

Page 10, line 2: '. . . and one to the stability class . . . '. The model error contribution to e_i is not just the stability class.

This statement is not claiming that the model error contribution is just the stability class, rather that the uncertainty that we are going to associate with each model prediction is stability-class (and instrument-type) dependent. The consequence of this decision is that we cannot use a single variance parameter to explain model-output discrepancy. This is needed since the plume model behaviour changes considerably with stability class. There is also other model uncertainty that we incorporate through the use of prior distributions on ω_y and ω_z. These also help to describe the transport model uncertainty.

Page 10, line 3: "(variance)" does not correspond to "(increases)". This is a good example why this grammar style is better avoided.

We apologise for the lack of clarity. We have now restructured this sentence.

Page 10: What justifies a windspeed independent error for windspeeds $>$ 1 m/s?

The Gaussian plume dispersion model is known to be unreliable for low wind speeds. The decision of what constitutes a low wind speed is somewhat subjective and application (scale) dependent, but we see that our results are not too sensitive to the choice of threshold. Further, discarding data at low wind speeds may make the inferential problem harder since this results in less data with which to calibrate the transport model. Please see our reply to Reviewer 1 on

Pg. 2 and Figs. 1 and 2 in this response for more details.

Page 11, line 7: Is a 'point mass at zero' not just a 'point emission of zero'. If so then please change to avoid confusion.

The "spike and slab" distribution is a type of mixed distribution, consisting of a "spike" in the probability density at zero, and another (usually continuous) distribution. In our case, we can place a spike and slab prior distribution on the emission rate by setting the half-normal distribution to be the "slab", and a dirac mass at zero to be the spike. This same prior distribution can be employed irrespective of whether the physical emission source is a single point or an area/region. More details of this is given below in response to Reviewer 2's comment regarding the 95% posterior credible intervals not containing zero when the source is switched off.

Page 11, line 21: What is the relevance of the distribution of the inverse of the square root of the precision parameter? For a precision it would be straightforward to relate it to the numbers that are given in the same sentence. However, what is done here is more complicated for a reason that I dont see.

The precision of the measurement error is the inverse of the squared of its standard deviation. The standard deviation is on the same scale, and has the same units, as the concentrations. Thus it is more interpretable than the precision. Precisions or variances are however useful when doing MCMC as they allow for conjugate priors to be used, which simplifies and potentially speeds up the sampling process. This is the reason why we tune our prior distributions on the standard deviation rather than directly on the precision, but use precision within the MCMC framework. We now explicitly define the precision as the inverse variance (as per its conventional definition in statistics) before Eq. (3).

Page 11, line 7: 'While addressing . . . zero emission rate'. Looking at equation 4, I don't see why it precludes zero as it is in the interval where the half-normal prior applies.

This comment concerns issue (iii) raised in the introduction to this review. The lower bound on the support of the half-normal distribution is zero, which means that by construction it is not possible for a 95% posterior credible interval to contain zero. As such, 2.5% of the posterior estimates must lie between zero and the lower limit of the 95% posterior interval, thus precluding zero in the interval.

In the manuscript we suggested that the use of a "spike and slab" prior may be a solution to this problem. We give the mathematical details required to incorporate the spike and slab prior in our framework in an Appendix to this reply.

We have now run four additional inversions using the spike-and-slab prior distribution: one for each release rate using observations collected when the methane source was switched on, and one for each release rate using observations collected when the methane source was switched off. These inversions were run using the EC Tower data in all instances. The results are shown in Fig. 6 of this response. The plots are the same as those given in the Results section of the manuscript, however to accommodate the "spike" portion of the distribution, there is now a circle plotted at zero. The radius of this circle is proportional to the posterior probability that the source on/off variable, Z_Q, is equal to zero. Here the results show no change in the posterior credible intervals for the case when the source is switched on, however when the source was switched off, the results now clearly suggest that the emission rate is zero, giving a posterior probability that $Z_Q = 0$ of more than 0.99 for both release-rate periods. The spike and slab

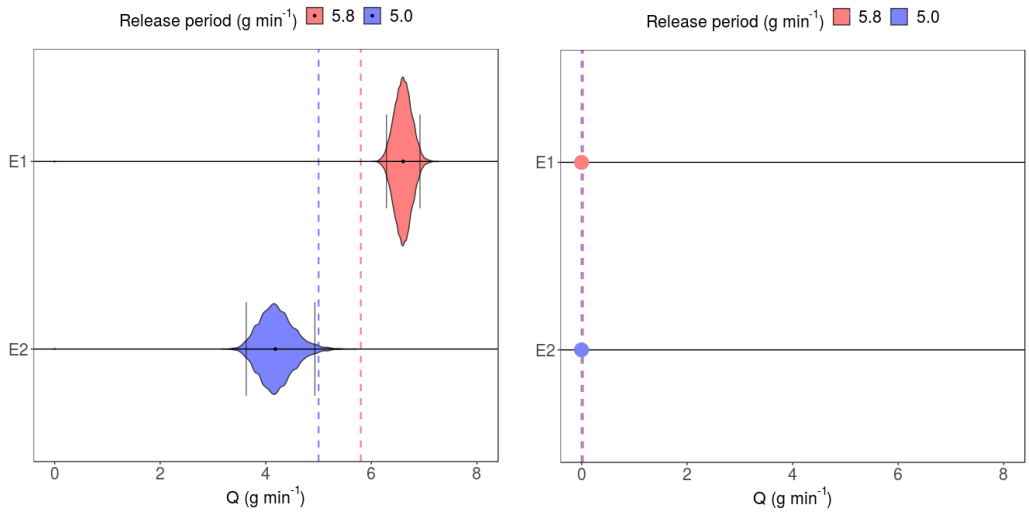

Figure 6: Left: 95% posterior credible intervals for the emission rate Q when the methane source is switched on using the EC Tower data, and in each release rate period: The 5.8 g min^{-1} interval is shown in red, and the 5.0 g min^{-1} interval is shown in blue. The vertical dashed lines represent the true emission rate in each case. The radius of the circle at $Q = 0$ is proportional to the posterior probability that the source on/off variable, Z_Q, is equal to zero. Right: Same as left but when the methane source is switched off.

prior thus can give a posterior inference of zero emission rate, as desired. We caution that what we show here is a (promising, but) preliminary result given the allowed time frame for the response, and will be investigated in more depth in future work.

Page 14, figure 5: It is explained in the text why the method precludes zero as a solution to the inverse problem (see my earlier comment on that). However, I had expected the estimates to be much closer to zero when emissions are switched off. The likely reason is not the zero condition, but the accuracy at which the background concentration is accounted for. It makes me wonder why the background is not fitted as an additional unknown parameter.

This comment concerns issue (iv) raised in the introduction to this review. We have seen in previous works (e.g., Zammit-Mangion et al., 2015), as well as our own experimentation with background estimation, that the 5th-percentile (by instrument type) is a relatively robust estimator for the background concentration. While the implementation of the spike and slab distribution above appears to have resolved the issue of the 95% posterior credible intervals not containing a zero emission rate, the background can be fitted as a collection of unknown parameters (one for each instrument type) if needed/desired, within the Bayesian framework. In Sect. 4.1 we now refer to Ganesan et al. (2015) who shows how this can be done in a Bayesian setting.

Page 15, line 23: why are wind speed and direction assumed to be known?

Due to the small study region and our own instruments recording the weather "on site" throughout the experiment, we assume that the wind speed and direction measurements are of good quality. Wind speed and direction were measured at EC.A using both 2D and 3D instruments. The 3D measurements were obtained from a Campbell CSAT3 sonic anemometer, and the 2D measurements from a Gill WindSonic (Both sonic anemometers were using factory calibration). As part of data quality control, horizontal wind speed and wind direction data from the two instruments were compared, with no arising issues. Wind directions were determined by man-

ually aligning the sonic anemometers so that the reference direction was true north. Before averaging over five-minute intervals, the 3D sonic was logged at 10 Hz, and the 2D sonic at 1 Hz. We have now added more information about how the wind speed and wind direction were measured to Sect. 2 of the manuscript (first paragraph).

2.3 Technical Corrections

Page 3, line 6: 'Houweling' instead of 'Houwelling'

Thank you, this error has been corrected.

Appendix

In this appendix we give the mathematical details for the inclusion of the spike-and-slab distribution into our inversion framework. First, let Z_Q be an indicator variable identifying whether the source is switched off ($Z_Q = 0$) or on ($Z_Q = 1$). Then, we can write the observed concentrations as

$$C_i = Z_Q Q \tilde{C}_i,$$

where

$$\tilde{C}_i = \frac{1}{2\pi U_i \tilde{\sigma}_{y_i,k_i} \tilde{\sigma}_{z_i,k_i}} \exp\left(-\frac{y_i^2}{2\tilde{\sigma}^2_{y_i,k_i}}\right)\left[\exp\left(-\frac{(z_i-H)^2}{2\tilde{\sigma}^2_{z_i,k_i}}\right) + \exp\left(-\frac{(z_i+H)^2}{2\tilde{\sigma}^2_{z_i,k_i}}\right)\right].$$

The spike and slab prior distribution is defined through our prior distributions on Q and Z_Q as follows:

$$p(Q) = \begin{cases} \frac{\sqrt{2}}{\sigma_Q\sqrt{\pi}} \exp\left(-\frac{Q^2}{2\sigma_Q^2}\right), & Q \in [0,\infty) \\ 0, & \text{otherwise,} \end{cases}$$

$$p(Z_Q) = \begin{cases} \pi_Z, & Z_Q = 1, \\ 1-\pi_Z, & Z_Q = 0, \\ 0, & \text{otherwise,} \end{cases}$$

where π_Q is the *prior* probability of a source being on. For clarity, let $\boldsymbol{\psi} \equiv \{\boldsymbol{\tau}, \omega_y, \omega_z, \boldsymbol{U}, H, \boldsymbol{\Theta}\}$. The likelihood function now is

$$p(\boldsymbol{Y} \mid Q, Z_Q, \boldsymbol{\psi}) = \prod_{i=1}^{N} \frac{\sqrt{\tau_{m_i}}}{\sqrt{2\pi}} \exp\left(\frac{-\tau_{m_i}(Y_i - Z_Q Q \tilde{C}_i)^2}{2}\right).$$

The full conditional distribution for Q, given by

$$\begin{aligned} p(Q \mid \boldsymbol{Y}, Z_Q, \boldsymbol{\psi}) &\propto p(Q) p(\boldsymbol{Y} \mid Q, Z_Q, \boldsymbol{\psi}) \\ &\propto \exp\left(\frac{-Q^2}{2\sigma_Q^2} - \frac{\sum_{i=1}^{N} \tau_{m_i}(Y_i - Z_Q Q \tilde{C}_i)^2}{2}\right) \prod_{i=1}^{N} \frac{\sqrt{\tau_{m_i}}}{\sqrt{2\pi}}, \end{aligned}$$

has no known form, so we use a Metropolis proposal to sample from this distribution. That is, at the tth iteration of our Gibbs sampler we propose $Q^* \sim \text{Gau}(Q^{t-1}, (\sigma^*)^2)$. where σ^* is the

standard deviation of the proposal distribution for Q. Then, assuming Q is sampled at the end of every Gibbs-sampler iteration, we compute the acceptance ratio via

$$\rho = \frac{p(Q^*)p(\boldsymbol{Y} \mid Q^*, Z_Q^t, \boldsymbol{\tau}^t, \omega_y^t, \omega_z^t, \boldsymbol{U}, H, \boldsymbol{\Theta})}{p(Q^{t-1})p(\boldsymbol{Y} \mid Q^{t-1}, Z_Q^t, \boldsymbol{\tau}^t, \omega_y^t, \omega_z^t, \boldsymbol{U}, H, \boldsymbol{\Theta})},$$

and accept the proposed Q^* with probability

$$\alpha = \min\{\rho, 1\}.$$

That is, we draw $u \sim \text{Unif}[0,1]$, and set $Q^t = Q^*$ if $\alpha \geq u$. Otherwise we set $Q^t = Q^{t-1}$.

The full conditional distribution for Z_Q is given by

$$\begin{aligned} p(Z_Q \mid \boldsymbol{Y}, Q, \boldsymbol{\psi}) &= \frac{p(Z_Q)p(\boldsymbol{Y} \mid Q, Z_Q, \boldsymbol{\psi})}{p(\boldsymbol{Y} \mid Q, \boldsymbol{\psi})} \\ &= \frac{p(Z_Q)p(\boldsymbol{Y} \mid Q, Z_Q, \boldsymbol{\psi})}{p(Z_Q = 0)p(\boldsymbol{Y} \mid Q, Z_Q = 0, \boldsymbol{\psi}) + p(Z_Q = 1)p(\boldsymbol{Y} \mid Q, Z_Q = 1, \boldsymbol{\psi})}. \end{aligned}$$

Now, for $Z_Q = 1$,

$$\begin{aligned} p(Z_Q = 1)p(\boldsymbol{Y} \mid Q, Z_Q = 1, \boldsymbol{\psi}) &= \pi_Z \exp\left(\frac{-\sum_{i=1}^N \tau_{m_i}(Y_i - Q\tilde{C}_i)^2}{2}\right) \prod_{i=1}^N \frac{\sqrt{\tau_{m_i}}}{\sqrt{2\pi}} \\ &= P_1, \end{aligned}$$

and, for $Z_Q = 0$,

$$\begin{aligned} p(Z_Q = 0)p(\boldsymbol{Y} \mid Q, Z_Q = 0, \boldsymbol{\psi}) &= (1 - \pi_Z) \exp\left(\frac{-\sum_{i=1}^N \tau_{m_i} Y_i^2}{2}\right) \prod_{i=1}^N \frac{\sqrt{\tau_{m_i}}}{\sqrt{2\pi}} \\ &= P_0. \end{aligned}$$

Finally, we therefore have that

$$\begin{aligned} p(Z_Q \mid \boldsymbol{Y}, Q, \boldsymbol{\psi}) &= \begin{cases} \frac{P_1}{P_0+P_1}, & Z_Q = 1, \\ \frac{P_0}{P_0+P_1}, & Z_Q = 0, \end{cases} \\ &= \begin{cases} \pi_Z^*, & Z_Q = 1, \\ 1 - \pi_Z^*, & Z_Q = 0, \end{cases} \\ &\sim \text{Bern}(\pi_Z^*), \end{aligned}$$

where π_Z^* is the *posterior* probability that the source is switched on. This distribution can be sampled from directly without the need for a Metropolis step in our Gibbs sampler.

List of Changes to 'Bayesian atmospheric tomography for detection and quantification of methane emissions: Application to data from the 2015 Ginninderra release experiment'

Authors

- 'Naghandhi' has been corrected to 'Negandhi'.

Abstract

- We have used the word 'assumptions' instead of 'misspecifications' in the abstract.

Section 1

- The first sentence of the second paragraph has been re-worded.
- A more comprehensive list of controlled release experiments and related studies has now been included in the second paragraph.
- We have added a note in the last sentence of paragraph 3 that both the stability classes and the Monin-Obhukov length are described later in the manuscript.
- We have broken the sentence starting on page 2, line 33 into two smaller sentences such that one key point is made in each sentence, as suggested. The first notes that online calibration of model parameters is not often done, while the second suggests a reason for this.
- We have rephrased the first sentence of paragraph 6 to make the point being addressed clearer.
- We have included a sentence briefly describing the atmospheric tomography technique in the second last paragraph.
- We have now re-worded the second last paragraph to better detail the connection of this study with the previous Ginninderra study.

Section 2

- More information has been added about how the wind speed and wind direction were measured.

Section 3

- In Sect. 3.1, we have included a short description of the Pasquill stability class along with the reference to Pasquill (1961).
- We have included a brief explanation of the L-value in Sect. 3.1.
- A sentence has been added to the caption of Fig. 2 highlighting the takeaway message of the figure.
- We have rephrased the sentence starting on page 7, line 6 to indicate that by scaling we mean multiplying by a constant factor.
- We have expanded the explanation for quantifying the statistics of $1/U_i$ in the first paragraph of Sect. 3.2 to increase clarity.
- We have modified the first paragraph of Sect. 3.2 in the manuscript to indicate that (i) in separate studies we have seen that our posterior inferences are not overly sensitive to the choice of wind speed cutoff used and that (ii) downweighting observations with low wind speeds instead of discarding them is particularly important when there are not many observations at high wind speed available for inference.

Section 4

- We have given a brief definition of $\boldsymbol{\tau}, \omega_y, \omega_z$, and Q upon their first appearance in the first paragraph, and noted the section numbers in which they are each described in more detail.
- The size of the arrows in Fig. 3 has been increased.
- In Sect. 4.1 we refer to Ganesan et al. (2015) who shows how background estimates can be fitted as unknown parameters in a Bayesian setting.
- We have restructured the last sentence of the second paragraph of Sect. 4.1 to increase clarity.
- We explicitly define the precision as the inverse variance (as per its conventional definition in statistics) in Sect. 4.1, before Eq. (3).
- We have noted that the half-normal distribution is a Gaussian distribution with mean zero, truncated from below at zero in the first sentence of Sect. 4.2 of the manuscript.

Section 5

- A sentence has been added to the caption of Fig. 5 highlighting the takeaway message of the figure.
- We have noted that we have looked at the Q-Q plots and observed mild deviations from Gaussianity in the residuals in the third paragraph of Sect. 5.2 of the manuscript.

Section 7

- We have provided some discussion on the envisioned applications of our method.

- We have now included a paragraph detailing the connections of our work to other previous, similar studies.
- Recommendations regarding experimental design have been added to the second last paragraph.

Author Contributions

- 'ND' has changed to 'NMD'.

Acknowledgements

- 'ND' has changed to 'NMD'.

References

- The spelling of 'Houweling' has been corrected from 'Houwelling' in Houweling et al. (2017).
- Additional references have been added: Ars et al. (2017); Etheridge et al. (2011); Feitz et al. (2014); Flesch et al. (2004); Forde et al. (2019); Ganesan et al. (2015); Harvey et al. (2018); Jenkins et al. (2016); Lewicki and Hilley (2009); Loh et al. (2009); van Leeuwen et al. (2013); Wilson et al. (2014)

Bayesian atmospheric tomography for detection and quantification of methane emissions: Application to data from the 2015 Ginninderra release experiment

Laura Cartwright[1], Andrew Zammit-Mangion[1], Sangeeta Bhatia[2, 3], Ivan Schroder[4], Frances Phillips[5], Trevor Coates[6], Karita Negandhi[7, 8], Travis Naylor[5], Martin Kennedy[7], Steve Zegelin[9], Nick Wokker[10], Nicholas M. Deutscher[5], and Andrew Feitz[4]

[1]School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, Australia
[2]Centre for Research in Mathematics, Western Sydney University, Parramatta, Australia
[3]School of Public Health, Imperial College London, London, UK (current address)
[4]Geoscience Australia, Canberra, Australia
[5]Centre for Atmospheric Chemistry, School of Earth, Atmosphere and Life Sciences, University of Wollongong, Wollongong, Australia
[6]School of Agriculture and Food, University of Melbourne, Melbourne, Australia
[7]Department of Earth and Planetary Sciences, Macquarie University, Sydney, Australia
[8]Office of Environment and Heritage, Parramatta, Australia (current address)
[9]CSIRO Oceans and Atmosphere, Canberra, Australia
[10]Department of Industry, Innovation and Science, Canberra, Australia

Correspondence: Laura Cartwright (lcartwri@uow.edu.au)

Abstract. Detection and quantification of greenhouse-gas emissions is important for both compliance and environment conservation. However, despite several decades of active research, it remains predominantly an open problem, largely due to model errors and ~~misspecifications~~ assumptions that appear at each stage of the inversion processing chain. In 2015, a controlled-release experiment headed by Geoscience Australia was carried out at the Ginninderra Controlled Release Facility, and a variety of instruments and methods were employed for quantifying the release rates of methane and carbon dioxide from a point source. This paper proposes a fully Bayesian approach to atmospheric tomography for inferring the methane emission rate of this point source using data collected during the experiment from both point- and path-sampling instruments. The Bayesian framework is designed to account for uncertainty in the parametrisations of measurements, the meteorological data, and the atmospheric model itself when doing inversion using Markov chain Monte Carlo (MCMC). We apply our framework to all instrument groups using measurements from two release-rate periods. We show that the inversion framework is robust to instrument type and meteorological conditions. The worst median emission-rate estimate we obtain from all the inversions is within 36% of the true value, while the worst posterior 95% credible interval has a limit within 11% of the true value.

1 Introduction

Methane (CH_4) is an important transition fuel for decarbonisation of the global energy system (International Energy Agency, 2017). As countries increase the renewable energy mix into their existing electricity networks, CH_4 can firm up network

stability and supply (International Energy Agency, 2017; Jenkins et al., 2018). Utilisation of biogas or natural gas with carbon capture and storage offers a lower cost pathway to achieve deep decarbonisation targets (Sepulveda et al., 2018). One of the disadvantages of CH_4, however, is that its global warming potential is much greater than that of carbon dioxide (CO_2), so that only a few percent of losses of CH_4 into the atmosphere can negate any climate-change mitigation advantages from reducing conventional coal-fired power production (Kinnon et al., 2018). For this reason, it is critical that losses of CH_4 along the supply chain are accurately accounted for to ensure public confidence in climate-change mitigation benefits of switching to natural gas. Unfortunately, while several types of instrumentation are available to aid the detection and estimation of fugitive emissions, harnessing acquired data for reliable emission detection and quantification remains a notoriously difficult problem.

~~In an effort to evaluate the performance of various instruments and algorithms for estimating the emission rate under controlled conditions~~Several controlled release experiments of CH_4 and CO_2 have been conducted in order to improve techniques for estimating greenhouse gas emissions (Flesch et al., 2004; Lewicki and Hilley, 2009; Loh et al., 2009; Etheridge et al., 2011; Humphries et al., 2012; van Leeuwen et al., 2013; Luhar et al., 2014; Jenkins et al., 2016; Ars et al., 2017). Building on this body of work,

in 2015 a CH_4 and CO_2 controlled-release experiment was held at the Ginninderra Controlled Release Facility in Canberra, Australia (Feitz et al., 2018). This large multidisciplinary, multi-institutional blind-release trial (i.e., the participants did not know the true release rate) simultaneously assessed eight different CH_4 emission-rate estimation techniques, using data from both mobile and stationary instrumentation. These eight techniques included tracer ratio techniques; backwards Lagrangian stochastic modelling; forward Lagrangian stochastic modelling; Lagrangian stochastic footprint modelling; and atmospheric tomography techniques. A full description of the methods and results is given in Feitz et al. (2018).

Every group involved in the analysis presented in Feitz et al. (2018) used a unique combination of instrumentation and estimation technique when carrying out the analysis, making it hard to establish the respective merits (or otherwise) of the employed techniques from the inversion results. Nonetheless, an interesting observation from the study is that none of the eight techniques deployed during the blind release trial had a leakage uncertainty range (95% interval) that included the true emission rate, while some estimates (including one obtained using atmospheric tomography) were factors of 2 or more off from the true value. Given that atmospheric methane concentration and meteorological instrument measurement uncertainty is generally low for each of the different approaches, it suggests that the techniques that were used did not adequately account for the variability of atmospheric measurements or the uncertainty introduced through parametrisation of atmospheric mixing conditions (e.g., Monin-Obukhov lengths and/or Pasquill stability classes; see Sect. 3.1) and atmospheric dispersion/transport model uncertainty.

A number of studies have highlighted the importance of atmospheric-model error in estimating emission rates or fluxes (e.g., Chevallier et al., 2010; Basu et al., 2018). For example, Peylin et al. (2002) showed that flux estimates are sensitive to the chosen spatio-temporal resolution of the fluxes and the chosen transport model. Uncertainty in the meteorological fields driving the transport model is also known to play a big role (e.g., Miller et al., 2015). While ensemble inversions are frequently used to highlight the sensitivity of the results to atmospheric models and meteorological fields, learning unknown parameters associated with transport *concurrently* with the emission rate is not often done~~,~~. This is largely due to the computational

implications of such an approach. Key here are the use of surrogate models (or emulators) to obtain simplified transport representations. For example, Lucas et al. (2017) use decision/regression trees as a surrogate for FLEXPART-WRF, which allows for quick simulation at various parameter settings that can in turn be used to make inference. For the Ginninderra data we employ the more traditional Gaussian plume model, which can be seen as a surrogate for a full-blown transport model. While known to work well in the small domain (an area of approximately 100×100 m) setting we consider (e.g., Riddick et al., 2017), importantly this plume model is quick to simulate from, giving us the opportunity to calibrate it while estimating the emission release rate (e.g., Borysiewicz et al., 2012). We see that "online plume-model calibration" is not only feasible, but crucial for obtaining accurate emission-rate estimates with our data.

~~Transport model calibration fits squarely in~~ The transport model plays an important role in inverse modelling. Calibration of the transport model from observations can be done within the classic inverse theory framework of Tarantola (2005)~~, which is~~. This framework is in turn seated within a Bayesian paradigm, ~~and~~ which underpins several of the inversion systems in place today; see, for example, ~~Humphries et al. (2012); Hirst et al. (2013); Ganesan et al. (2014); Luhar et al. (2014); ?); White et al. (2018)~~Flesch et al. (2004); Humphries et al. (2012); Hirst et al. (2013); Ganesan et al. (2014); Luhar et al. (2014); Houweling et al. (2017); White et al. (2018). Inference in such cases is often done using sampling techniques such as Markov chain Monte Carlo (MCMC) or importance sampling (Rajaona et al., 2015). Quick evaluation of the transport/dispersion model (or surrogate) is crucial when repeatedly evaluating it within an MCMC framework; the Gaussian plume model is hence a popular choice in these frameworks (e.g., Jones et al., 2016; Wang et al., 2017). MCMC is also our method of choice for Bayesian atmospheric tomography, because it allows relatively easy computation of posterior distributions of parameters that are deeply nested within a hierarchical model. It is also ideally suited for the case of point-source emissions, where the dimensionality of the latent space is low (unlike, for example, when doing regional emission quantification).

Atmospheric tomography, a term inspired from medical imaging, combines data from a collection of measurement sites with Bayesian inversion to detect and quantify emissions. The primary contribution of this article is an extension of the ~~typical Bayesian~~ atmospheric tomography technique ~~designed for detection and estimation of emissions from point sources~~described in Sect. 2.4.2 of Feitz et al. (2018). In Feitz et al. (2018), atmospheric tomography was only used on one type of instrument and did not account for uncertainty in the transport model. The technique we propose accounts for uncertainty in our data, in our process models, and in our parameters~~,~~; is applicable to both point and path-sampling instruments; and takes into account instrument-specific bias. Inference is made on all unknown parameters using MCMC ~~; a by-product of the model we employ is the ability to calibrate parameters that appear in the transport model, and to propagate~~ and uncertainty in the ~~model~~ transport-model parameters are propagated to our posterior inferences on the release rate. We demonstrate the efficacy and utility of the unifying Bayesian framework on data from point- and path-referenced instruments used in the Ginninderra experiment. A secondary contribution is the curated provision of a data set containing a large portion of the Ginninderra data at a five-minute resolution, which we hope will serve as a resource for other researchers to validate their own emission-rate estimation techniques on. The data and scripts required to reproduce the results in this article are available from https://github.com/Lcartwright94/BayesianAT.

The remainder of the article is organised as follows. Section 2 gives an overview of the experimental setup and the data collected during the 2015 Ginninderra experiment. Section 3 describes the atmospheric transport model used, while Sect. 4 details the hierarchical model we employ and the Bayesian methodology we develop for emission-rate estimation. Section 5 gives the results from application of our Bayesian atmospheric tomography technique on the Ginninderra data. Section 6 examines how our results would change if certain components in our model (e.g., relating to the plume model) are (erroneously) assumed fixed and known. Section 7 concludes.

2 The 2015 Ginninderra release experiment

A full description of the experimental setup, measurement techniques and quantification methods used in the 2015 Ginninderra release experiment are given in Feitz et al. (2018). Briefly, CH_4 (together with CO_2 and ~~N_2O~~nitrous oxide), was released from a small chamber located in a fallow agricultural field from 23 April to 12 June 2015, and 23 to 24 June 2015. A variety of CH_4 sensors were placed around the release chamber. The measurement data considered in this study were obtained from two Picarro G2201-i analysers (positioned in the predominant upwind (NW) and downwind (SE) location of the release chamber, labelled Picarro.West and Picarro.East, respectively), four eddy covariance (EC) towers equipped with Li-COR 7700 open path CH_4 sensors (labelled EC.A, EC.C, EC.D, and EC.E, respectively), two scanning FTIR analysers with four retro-reflectors terminating six measurement paths (labelled P1 to P6, respectively), and a scanning GasFinder 2 Boreal laser with seven reflectors forming seven measurement paths (labelled R1 to R7, respectively); see the left panel of Fig. 1. Meteorological data was collected from EC.A equipped with a Vaisala HMP50 relative humidity and temperature sensor, a CSI ~~CSAT3 sonic anemometer, a CSI~~ EC150 CO_2-H_2O sensor, a Li-COR 7700 CH_4 sensor, a Kipp and Zonen CNR4 radiometer, a CSI CSAT3 sonic anemometer, and a Gill WindSonic ~~2D sonic anemometer .~~ anemometer. Wind speed and wind direction were measured by the CSAT3 sonic anemometer and the Gill WindSonic anemometer. As part of data quality control, horizontal wind speed and wind direction data from the two instruments were compared, with no arising issues. Both sonic anemometers were using factory calibration. Wind directions were determined by manually aligning the sonic anemometers so that the reference direction was true north. Data from CSAT3 sonic anemometer was logged at 10 Hz, and data from the Gill WindSonic anemometer at 1 Hz.

The gases were released at a height of 0.3 m, and the standard CH_4 release rate was 5.8 g min^{-1}, limited mostly to daylight hours. On brief occasions, the CH_4 release rate varied between 2.9 and 20 g min^{-1} to enable testing of mobile CH_4 sensor platforms. Towards the end of the experiment (8 to 12 June), the CH_4 release rate was decreased from 5.8 to 5.0 g min^{-1} and the setup for the Boreal laser measurements was modified with the number of retro-reflectors and paths reduced to six (labelled R8 to R13, respectively; see the right panel of Fig. 1). The location of all other CH_4 sensors did not change over the duration of the experiment. The Picarro analysers were not deployed until 21 May, and the CH_4 release rate on 23 and 24 June was constantly varied. Hence, in this article we only consider data between 21 May and 12 June, excluding 26 and 27 May where the release rate was also constantly varied.

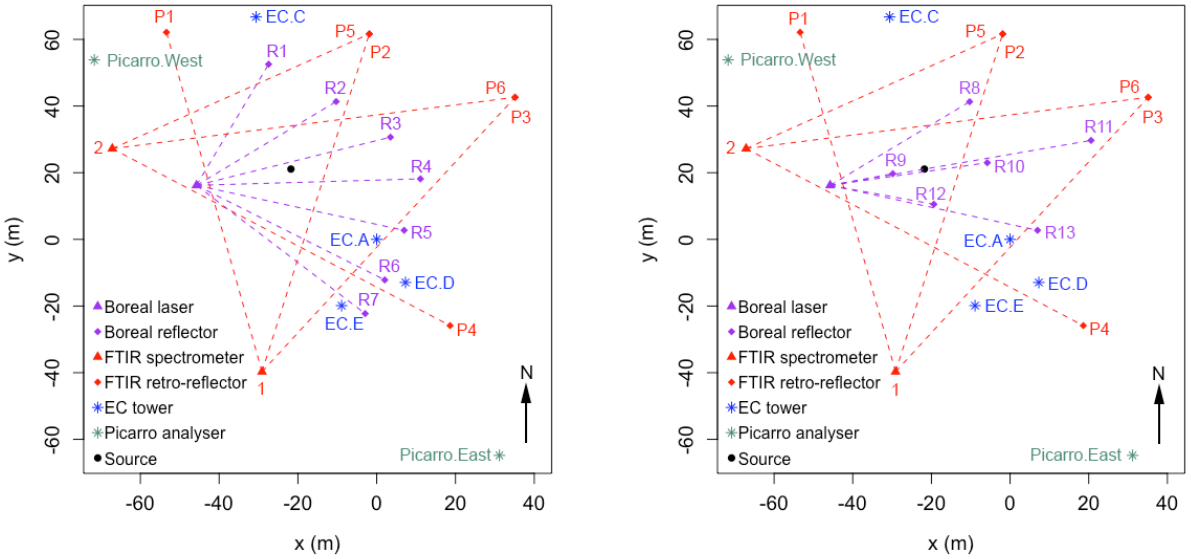

Figure 1. Left: Layout of instruments in the 2015 Ginninderra release experiment between 21 May and 7 June 2015. Right: Layout of instruments between 8 and 12 June 2015. R1 to R13 are the paths formed between the Boreal laser and reflectors; P1 to P6 are the paths formed between the FTIR spectrometers and retro-reflectors; EC.A to EC.E are the EC towers; and Picarro.East and Picarro.West are the Picarro analysers. All coordinates are relative to EC.A, which is situated at the origin.

The data set used to obtain the results presented in Sect. 5 was compiled by pooling together the separate meteorological and concentration data sets used in the Ginninderra experiment. A common resolution of five minutes was chosen, that is, all measurements of concentration and meteorological variables were averaged over a regular set of five-minute intervals. Measured CH_4 concentrations were then matched with corresponding meteorological measurements by time and placed into long-table format, with each row corresponding to a unique data point. For path measurements, two extra columns were used to denote the end-point coordinates of the paths.

Initial pre-processing was carried out to provide a complete data set without outliers. First, data containing missing values considered critical for emission-rate estimation (in particular, air temperature, air pressure, wind speed, and wind direction) were removed from the data set. Second, data points corresponding to upwind measurements that were more than three median absolute deviations away from the instrument's median upwind measured concentration, were determined to be outliers, and hence removed. A point measurement was classified as upwind if the angle subtended from the source by a line joining the instrument location to the plume centreline was more than $45°$. A path measurement was classified as upwind if the angles subtended at every point along the path were more than $45°$.

3 Transport modelling

In this section we detail the plume model employed, and how it is used to supply model-predicted concentrations for the path measurements.

3.1 Gaussian plume dispersion modelling

As outlined in Sect. 1, we use a transport model that is simply parameterised, and easy to evaluate, so that it can be calibrated online. One of the simplest models that works well on the short distances we consider is the Gaussian plume dispersion model (e.g., Wark et al., 1998, Chapter 4). Here the true emission rate is denoted by Q in g s^{-1}, the height of the CH_4 point source by H in m, and the total number of observations by N. The classic Gaussian plume model is given by

$$C(x_i, y_i, z_i, Q, U_i, H, \boldsymbol{\theta}_{k_i}) = \frac{Q}{2\pi U_i \sigma_{y_i,k_i} \sigma_{z_i,k_i}} \exp\left(-\frac{y_i^2}{2\sigma_{y_i,k_i}^2}\right)\left[\exp\left(-\frac{(z_i-H)^2}{2\sigma_{z_i,k_i}^2}\right) + \exp\left(-\frac{(z_i+H)^2}{2\sigma_{z_i,k_i}^2}\right)\right], \quad (1)$$

where C is the *model-predicted* concentration in g m^{-3} of CH_4 at a single spatial point (x_i, y_i, z_i) in m along the direction of the plume corresponding to the ith measurement; U_i is the wind speed associated with the ith measurement in m s^{-1}; $k_i \in \{A, B, C, D, E, F\}$ represents the Pasquill stability class ~~(Pasquill, 1961)~~ (a categorisation reflective of the expected level of horizontal and/or vertical spread of the atmospheric particles after emission; see Pasquill, 1961) associated with the ith measurement; and $\boldsymbol{\theta}_{k_i}$ are plume-specific parameters used the construct the standard deviations σ_{z_i,k_i} and σ_{y_i,k_i}. These standard deviations of the plume in the vertical and horizontal directions are given by

$$\sigma_{z_i,k_i} = a_{k_i} x_i^{b_{k_i}},$$
$$\sigma_{y_i,k_i} = 0.4651 x_i \tan(\nu_i),$$

respectively, where $\nu_i = 0.01745\,(c_{k_i} - d_{k_i} \ln(x_i/1000))$. Note that the coefficients $a_{k_i}, b_{k_i}, c_{k_i}, d_{k_i}$ correspond to the ith measurement and depend on the stability class associated with that measurement, $k_i \in \{A, B, C, D, E, F\}$. Values for these coefficients by stability class are given in Wark et al. (1998, Chapter 4) and shown here in Table 1 for completeness. We collect the plume-specific parameters in $\boldsymbol{\theta}_{k_i} \equiv (a_{k_i}, b_{k_i}, c_{k_i}, d_{k_i})'$ where $'$ denotes the transpose operator.

The stability class to which an observation is allocated is classically based on (i) the Monin-Obukhov length (the theoretical height at which turbulence is produced by buoyancy and mechanical forces in equal amounts; see Sienfeld and Pandis (2006), Chapter 16), and (ii) an effective roughness length. The Monin-Obukhov length (L-value) is given by $L = -u_*^3 \bar{\xi}_v / (qg(\overline{w^* \xi_v^*})_s)$ (Jacobson, 2005, Chapter 8), where u_* is the frictional velocity, $\bar{\xi}_v$ is the mean virtual potential temperature, $(\overline{w^* \xi_v^*})_s$ is the surface virtual potential temperature flux, q is the von Kármán constant, and g is the acceleration due to gravity. In our case we used WindTrax (http://www.thunderbeachscientific.com/windtrax.html) to determine the L-value for each observation; we provide the L-values with the compiled data. We set the effective roughness length $z_0 = 0.01$ m, corresponding to a relatively flat area, with short or no grass, and minimal buildings/trees/other obstacles; see Sienfeld and Pandis (2006, Chapter 16) and

Table 1. Stability classes to which observations within the Ginninderra experiment are allocated, and the corresponding values of $a_{k_i}, b_{k_i}, c_{k_i}$, and d_{k_i} used to construct the horizontal (σ_{y_i,k_i}), and vertical (σ_{z_i,k_i}) standard deviations of the plume when x_i is in m.

Stability Class (k_i)	**Stability Condition**	a_{k_i}	b_{k_i}	c_{k_i}	d_{k_i}
A	Extremely unstable	0.17993	0.94470	24.167	2.5334
B	Moderately unstable	0.14506	0.93198	18.333	1.8096
C	Slightly unstable	0.11025	0.91465	12.500	1.0857
D	Neutral	0.084739	0.86974	8.3330	0.72382
E	Slightly stable	0.075005	0.83660	6.2500	0.54287
F	Moderately stable	0.054370	0.81558	4.1667	0.36191

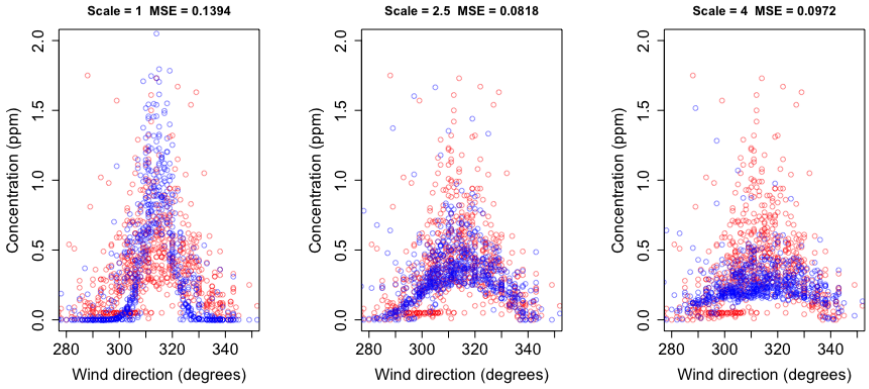

Figure 2. Predicted (blue) and observed (red) enhancements in ppm at EC.A between May 21 and June 7 2015 when scaling σ_{y_i,k_i} by 1, 2.5, and 4, respectively. The mean-squared errors (MSE) between the observed and the predicted enhancements are also shown. Of the three, the best agreement between predicted and observed values occurs when σ_{y_i,k_i} is scaled by 2.5.

World Meteorological Organisation (2008, Chapter 5). This is a suitable choice for the Ginninderra site. We used the results of Golder (1972) to allocate a stability class to each observation based on the L-values provided by WindTrax and $z_0 = 0.01$ m.

The coefficients typically used for each stability class could be off by a factor of 2 or more (Wark et al., 1998, Chapter 4). To show that this is also the case with our categorisation scheme, in Fig. 2 we show the Gaussian-plume-model predicted outputs together with the observed data enhancements (see Sect. 4.1) at one of our measurement locations (namely, EC.A) between 21 May and 7 June, when scaling σ_{y_i,k_i} by 1, 2.5, and 4, respectively. Clearly, with no scaling the predicted plume is too narrow, while with a scaling of 4 is too broad. A scaling of 2.5 gives good agreement. Importantly, since in Eq. (1) Q only serves to scale the ~~outputs vertically~~predicted concentrations (i.e., make them larger or smaller by a constant factor), it is apparent that this plume scaling factor is identifiable, in the sense that we can learn it from the data *while* estimating the emission rate (provided the source is active). Online plume-model calibration fits naturally within the MCMC framework discussed in Sect. 4.

3.2 Low wind speeds

It is well known that the Gaussian plume model is less accurate for low wind speeds (e.g., Turner, 1994, Chapter 2). One reason for this is that the wind-speed U_i is in the denominator of the scaling coefficient of Eq. (1); hence, the plume model prediction becomes very sensitive to U_i as it tends towards zero. This is problematic as U_i, although often assumed known, is an average calculated from noisy measurements taken over some time span (in our case five minutes), and is thus itself noisy. From Eq. (1) we see that, when conditioned on all other parameters, the variance of C_i is proportional to the variance of the inverse of U_i, which can be very large for small U_i. Instead of removing data at low wind speeds as is often done (e.g., Feitz et al., 2018), we analyse the theoretical relationship between the variance of the inverse wind speed and ~~its inverse, and~~ C_i. We then use this relationship to discount low wind-speed model-predictions in the Bayesian framework ~~. This approach relieves the analyst of having~~ in a principled manner. While the analyst still needs to choose a ~~hard cutoff wind-speed when pre-processing the data~~cutoff below which to model this relationship, in separate studies we found that our inferences are not particularly sensitive to the chosen cutoff. Moreover, we found that downweighting instead of excluding was necessary for making inference when not many observations associated with high wind speeds were available.

Each wind speed U_i is an average of a number of wind speeds (say n_{U_i}) recorded over five minutes. Therefore U_i is a sample mean, and thus an unbiased estimator of the true (population) mean wind speed, say μ_{U_i}, over this time interval. By the Central Limit Theorem, $\sqrt{n_{U_i}}(U_i - \mu_{U_i}) \xrightarrow{\mathcal{D}} \text{Gau}(0, \sigma^2_{U_i})$, where $\mathcal{D}$ implies convergence in distribution, $\text{Gau}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2, and $\sigma^2_{U_i}$ is the variance of the wind speeds over the ith time interval, which was derived from the raw (disaggregated) data. We can then use the delta method (e.g., Casella and Berger, 2002, Chapter 5) to deduce that

$$\sqrt{n_{U_i}}\left(\frac{1}{U_i} - \frac{1}{\mu_{U_i}}\right) \xrightarrow{\mathcal{D}} \text{Gau}\left(0, \left(\frac{\mathrm{d}}{\mathrm{d}\mu_{U_i}}\left(\frac{1}{\mu_{U_i}}\right)\right)^2 \sigma^2_{U_i}\right).$$

Hence, the variance of $1/U_i$ is approximately

$$\left(\frac{\mathrm{d}}{\mathrm{d}\mu_{U_i}}\left(\frac{1}{\mu_{U_i}}\right)\right)^2 \sigma^2_{U_i} = \frac{1}{\mu^4_{U_i}}\sigma^2_{U_i} \propto \frac{1}{\mu^4_{U_i}}.$$

Therefore, conditional on all other terms in Eq. (1), the variance of the model-predicted concentrations increases as a quartic of the true inverse wind speed. This is important, as it means that model predictions at low wind speeds~~(e.g., ~~, say less than 1 m s^{-1}~~)~~, could be highly uncertain; we show a way of handling this uncertainty when we detail the Bayesian inversion model in Sect. 4.

3.3 Predicted concentrations for point and path measurements

The plume model given by Eq. (1) sets the x-axis as its centreline and the CH_4 source at the origin. The predicted plume-model concentration at a physical location $(\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)$ is thus found by first applying a spatial shift and time-dependent rotation (by

wind direction) to $(\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)$ in order to obtain (x_i, y_i, z_i), which is then used to compute a model-predicted concentration (conditional on Q, U_i, H, and $\boldsymbol{\theta}_{k_i}$). Conversion to ppm is done via the ideal gas law.

Let C_i be a model-predicted concentration ($i = 1, 2, \ldots, N$). If C_i corresponds to a point measurement, then one need only evaluate Eq. (1) at the transformed point-measurement location to obtain a predicted concentration. If C_i corresponds to a path measurement, however, it represents an average of concentrations along the path. Denote the transformed end points of the straight-line path in the horizontal plane as $(x_{i,1}, y_{i,1})$ and $(x_{i,2}, y_{i,2})$, respectively. The line between the given points in the horizontal plane can be parametrised by $\boldsymbol{\rho}_i(s) = (\rho_{x,i}(s), \rho_{y,i}(s))'$, where

$$\rho_{x,i}(s) = s x_{i,2} + (1-s) x_{i,1}, \quad s \in [0,1],$$
$$\rho_{y,i}(s) = s y_{i,2} + (1-s) y_{i,1}, \quad s \in [0,1],$$

so that

$$C_i = \frac{1}{T_i} \int_0^1 C(\rho_{x,i}(s), \rho_{y,i}(s), z_i, Q, U_i, H, \boldsymbol{\theta}_{k_i}) \left\| \frac{\mathrm{d}\boldsymbol{\rho}_i(s)}{\mathrm{d}s} \right\| \mathrm{d}s, \tag{2}$$

where $T_i \in \mathbb{R}^+$ is the path length and $\|\cdot\|$ is the standard Euclidean norm. In our case, $T_i = \left\| \frac{\mathrm{d}\boldsymbol{\rho}_i(s)}{\mathrm{d}s} \right\|$, which is not a function of s, and so Eq. (2) simplifies to $C_i = \int_0^1 C(\rho_{x,i}(s), \rho_{y,i}(s), z_i, Q, U_i, H, \boldsymbol{\theta}_{k_i})\, \mathrm{d}s$. This integral can be approximated numerically over a fine partitioning of J segments $P = \{[s_0, s_1], [s_1, s_2], \ldots, [s_{J-1}, s_J]\}$, where $0 = s_0 < s_1 < \cdots < s_{J-1} < s_J = 1$. Then

$$C_i \approx \sum_{j=1}^{J} C(\rho_{x,i}(s_j^*), \rho_{y.i}(s_j^*), z_i, Q, U_i, H, \boldsymbol{\theta}_{k_i}) \Delta_s,$$

where $\Delta_s \equiv s_j - s_{j-1} = 1/J$ and $s_j^* = \frac{s_j + s_{j-1}}{2}$. In our experiments we set $J = 100$.

4 Bayesian atmospheric tomography

We are ultimately interested in obtaining a range of plausible values for the emission rate, Q, *a posteriori*, (i.e., after we have observed some data). In this section we present a hierarchical statistical model that relates Q to the observed concentrations via the Gaussian plume model. Although Q itself is univariate, the model contains several other unknown parameters that capture our uncertainty about the phyiscal and the measurement processes; inferences on these parameters and Q are made simultaneously. For ease of exposition we adopt the terminology of Berliner (1996) to describe the model, which we also summarise graphically in Fig. 3. The top layer in the hierarchy is the data model (the model for the observations, $\boldsymbol{Y}$, Sect. 4.1), the middle layer is the process model (the model for Q, Sect. 4.2), and the bottom layer is the parameter model (the unknown parameters not of direct interest, $\boldsymbol{\tau}$, ω_y, ω_z, Sect. 4.3). In Sect. 4.4 we outline the MCMC strategy we use to make inference with the model.

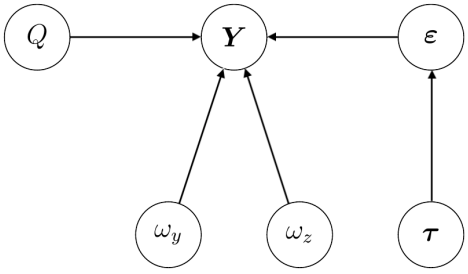

Figure 3. Directed acyclic graph showing the conditional dependence relationships between the data (enhancements) $\boldsymbol{Y}$ and the error components $\boldsymbol{\varepsilon}$ (Sect. 4.1), the emission rate Q (Sect. 4.2), and the unknown parameters $\boldsymbol{\tau}, \omega_y$, and ω_z (Sect. 4.3).

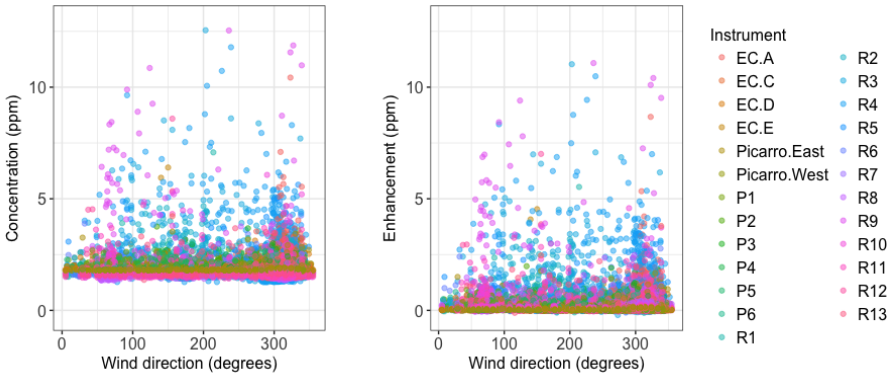

Figure 4. Left: Raw averaged concentrations, plotted by instrument and against wind direction. Right: Enhancements obtained by subtracting off the background and instrument-specific bias.

4.1 The data model

Let $\widetilde{\boldsymbol{Y}} \equiv (\widetilde{Y}_1, \widetilde{Y}_2, \ldots, \widetilde{Y}_N)'$ denote the measured concentrations averaged over five-minute intervals. We model each of these averaged measurements as $\widetilde{Y}_i = C_i + X_i + \varepsilon_i$, where C_i is the ith Gaussian plume-predicted concentration, X_i is the sum of the ith CH_4 background concentration and instrument-specific bias, and ε_i denotes the random error associated with the ith observed CH_4 concentration. ~~As in Zammit-Mangion et al. (2015)~~ The background concentration and bias can be explicitly modelled and predicted (Ganesan et al., 2015). Here, as in Zammit-Mangion et al. (2015), we estimate X_i as the 5th percentile of all the measurements from the instrument associated with the ith measurement. Figure 4 compares the raw averaged concentrations to those corrected for background and instrument-specific bias, which we term *enhancements*, when plotted against wind direction (in degrees East of North).

Now, let $\boldsymbol{Y} \equiv (Y_1, Y_2, \ldots, Y_N)'$ denote the enhancements. It is straightforward to verify that

$$\begin{aligned} Y_i &= \widetilde{Y}_i - X_i \\ &= C_i + \varepsilon_i, \quad i = 1, \ldots, N. \end{aligned}$$

Therefore, Y_i is made up of two main components of variability: the Gaussian plume-predicted concentration and a random error term. We assume that the $\varepsilon_i, i = 1, \ldots, N$, are Gaussian and independent, but that they are not identically distributed. Specifically, ε_i contains two components of variation, one pertaining to the error characteristics of the instrument, and one to the stability class with which we have categorised the measurement. Recall also from Sect. 3.2 that we model the variance of the ~~precision (variance) of the wind speed decreases (increases) as~~ predicted concentrations to be proportional to a quartic of the ~~wind speed~~true mean inverse wind speed for $U_i < 1$ m s^{-1}.

First, we capture instrument-specific measurement error characteristics and stability-condition-specific variation by introducing an auxiliary variable m_i ($m_i = 1, 2, \ldots, M$), where M is the total number of unique combinations of stability class and instrument type, and consider M different precision (i.e., inverse variance) parameters $\{\tau_{m_i}\}$ that need to be estimated, one for each combination. Second, we take the influence of low wind speeds into account by assuming that the precision of ε_i is τ_{m_i} multiplied by $\hat{U}_i$, where, for $U_i > 0$,

$$\hat{U}_i = \begin{cases} U_i^4 & 0 < U_i < 1, \\ 1 & U_i \geq 1, \end{cases} \tag{3}$$

which encapsulates our prior belief that observed model-measurement mismatch variability at low wind speeds (in this case under 1 m s^{-1}) are dominated by the low wind speed.

Putting these two components together, we have that, conditional on the instrument type and stability class encoded in m_i,

$$\varepsilon_i \mid m_i \sim \text{Gau}(0, 1/(\hat{U}_i \tau_{m_i})), \quad i = 1, \ldots, N.$$

We detail the prior distribution for τ_{m_i} in Sect. 4.3.1.

4.2 The process model

The process of interest in this application is the emission rate, Q, which we assume is constant. Since in this application $Q \geq 0$, we model it using a half-normal prior distribution (a Gaussian distribution with mean zero truncated from below at zero),

$$p(Q) = \begin{cases} \frac{\sqrt{2}}{\sigma_Q \sqrt{\pi}} \exp\left(-\frac{Q^2}{2\sigma_Q^2}\right), & Q \in [0, \infty) \\ 0 & \text{otherwise}, \end{cases} \tag{4}$$

with a standard deviation parameter, σ_Q, which is known and fixed. In our case we fixed σ_Q to 1.5 g s^{-1} (90 g min^{-1}) which results in a relatively uninformative prior distribution.

While addressing nonnegativity, half-normal priors do not contain a point mass at zero, and thus do not encode a prior belief that there is a possibility of having exactly a zero emission rate. As a consequence, a posterior estimate or even a

credible interval that includes zero is not possible. A spike-and-slab distribution (Mitchell and Beauchamp, 1998) consisting of a diffuse uniform distribution with a point-mass at zero, could be alternatively used at the cost of a slightly more complex model.

4.3 The parameter model

Our parameter model is divided into two parts, one pertaining to the precision parameters $\{\tau_{m_i}\}$ in the random-error component in the data model, and the other to the standard-deviations in the Gaussian-plume dispersion models which, as shown in ~~Section 3.2~~Sect. 3.1, are also uncertain.

4.3.1 The precision parameters

For conjugacy with the Gaussian likelihood, we model each τ_{m_i} using a Gamma prior distribution, with shape parameter α, and rate parameter β:

$$p(\tau_{m_i}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau_{m_i}^{\alpha-1} e^{-\beta\tau_{m_i}}, \quad i = 1, \ldots, N.$$

In our application we set $\alpha = 1.058$ and $\beta = 0.621$. These values were chosen through quantile-matching, such that the 1st and 99th percentiles of the distribution of $1/\sqrt{\tau_{m_i}}$ are approximately 0.35 ppm and 6.5 ppm, respectively (giving a mode close to 0.7 ppm). Values for these percentiles were selected based on prior exploratory data analysis of the measurements that were taken upwind of the source.

4.3.2 The Gaussian plume model parameters

From separate studies into the reliability of the model values for σ_{y_i,k_i} and σ_{z_i,k_i}, briefly discussed in Sect. ~~3~~3.1, we concluded that these parameters could indeed be off by factors of two or more and that, if they are off, they are so by similar amounts for each stability class. These factors correspond to vertical shifts of the Pasquill stability curves when plotted on a log-log scale (e.g., Wark et al., 1998, Chapter 4). We thus replaced σ_{y_i,k_i} and σ_{z_i,k_i} in Eq. (1) with $\tilde{\sigma}_{y_i,k_i}$ and $\tilde{\sigma}_{z_i,k_i}$, respectively, where

$$\tilde{\sigma}_{y_i,k_i} \equiv \omega_y \sigma_{y_i,k_i} \qquad \text{and} \qquad \tilde{\sigma}_{z_i,k_i} \equiv \omega_z \sigma_{z_i,k_i},$$

and $\omega_y, \omega_z \in \mathbb{R}^+$ are scaling parameters for σ_{y_i,k_i} and σ_{z_i,k_i}, respectively (Borysiewicz et al., 2012).

We use Gamma prior distributions for ω_y and ω_z. In our application we set the shape parameters equal to 1.6084, and the rate parameters equal to 0.7361. These parameters give approximate 1st and 99th percentiles of 0.1 and 8, respectively, and a mode close to 1 (representative of no scalar influence on σ_{y_i,k_i} or σ_{z_i,k_i}). This reflects our prior belief that the standard deviations could be up to an order of magnitude off from those derived using classical Pasquill stability-class theory.

4.4 Bayesian inference

Let $\boldsymbol{Y} \equiv (Y_1, Y_2, \ldots, Y_N)'$ denote the N observed enhancements. Similarly, define $\boldsymbol{U} \equiv (U_1, U_2, \ldots, U_N)'$ and $\boldsymbol{\Theta} \equiv (\boldsymbol{\theta}_{k_1}, \boldsymbol{\theta}_{k_2}, \ldots, \boldsymbol{\theta}_{k_N})'$. Further, let $\boldsymbol{\tau} \equiv (\tau_1, \tau_2, \ldots, \tau_M)'$ be the M parameters associated with each combination of instrument type and stability class. The posterior distribution of the emission rate Q is then given by

$$
\begin{aligned}
p(Q \mid \boldsymbol{Y}, \boldsymbol{U}, H, \boldsymbol{\Theta}) &\propto \int_0^\infty \int_0^\infty \int_{\mathbb{R}^{M+}} p(\boldsymbol{Y}, \boldsymbol{\tau}, \omega_y, \omega_z \mid Q, \boldsymbol{U}, H, \boldsymbol{\Theta}) p(Q) \, \mathrm{d}\boldsymbol{\tau} \, \mathrm{d}\omega_y \, \mathrm{d}\omega_z \\
&= p(Q) \int_0^\infty \int_0^\infty \int_{\mathbb{R}^{M+}} p(\boldsymbol{Y} \mid Q, \boldsymbol{\tau}, \omega_y, \omega_z, \boldsymbol{U}, H, \boldsymbol{\Theta}) p(\boldsymbol{\tau}) p(\omega_y) p(\omega_z) \, \mathrm{d}\boldsymbol{\tau} \, \mathrm{d}\omega_y \, \mathrm{d}\omega_z,
\end{aligned}
$$

where $p(Q)$ is given by Eq. (4) and $p(\boldsymbol{Y} \mid Q, \boldsymbol{\tau}, \omega_y, \omega_z, \boldsymbol{U}, H, \boldsymbol{\Theta})$ is the likelihood, which is Gaussian.

Computation of the the posterior distribution $p(Q \mid \boldsymbol{Y}, \boldsymbol{U}, H, \boldsymbol{\Theta})$ involves a high-dimensional integral that is analytically intractable. We therefore use MCMC, specifically a Gibbs sampler, to obtain samples from the posterior distributions of Q, $\boldsymbol{\tau}$, ω_y, and ω_z (see Gelman et al., 2013, for a comprehensive introduction to MCMC). The Gibbs sampler samples each parameter one at a time from their respective full conditional distributions, where conditioning is done using the most recent samples of all other parameters.

In the case of $\boldsymbol{\tau}$, use of Gamma prior distributions leads to full conditional distributions that are also Gamma. Hence, sampling $\boldsymbol{\tau}$ is straightforward. However, the prior distributions on the other parameters are not conjugate priors, and hence the full conditional distributions for each of these are not available in closed form. We therefore use standard Metropolis-within-Gibbs to sample from these conditional distributions, with Gaussian proposals and adaptive scaling during the early stages of the MCMC algorithm. Specifically, for each parameter, the standard deviation of the proposal was increased or decreased as appropriate whenever the acceptance rate fell below 10% or exceeded 80%.

5 Results and Discussion

5.1 Observing system simulation experiment

In this section we discuss results from applying our model to simulated data in an observing system simulation experiment (OSSE). To mimic the conditions in the real experiment, we simulated enhancements using the actual Boreal and EC instrument locations, meteorological observations from the Ginninderra data, and realistic variances for the random-error components. We considered the two release-rate periods separately, using a 6 g min^{-1} emission rate in the first, and a 12 g min^{-1} emission rate in the second. As in the real experiment, the first Boreal laser/reflector setup (seven paths) was used in the first release-rate period, while the second setup (six paths) was used in the second release-rate period; the EC tower locations were kept constant for both periods. We set the precisions $\tau_{m_i} = 4$ for $m_i = 1, \ldots, M$, and the scaling factors $\omega_y = \omega_z = 2$ to assess the

Table 2. Posterior median emission rates in g min^{-1}, and the posterior 95% credible intervals of the emission rate in g min^{-1}, ω_y, and ω_z from the OSSE. Results shown are from simulated data corresponding to the Boreals (B) and EC towers (E) when the emission rate is 6 g min^{-1} (E1 and B1) and 12 g min^{-1} (E2 and B2).

	Median Q	Q	ω_y	ω_z
B1	6.0718	$(5.7847, 6.3511)$	$(0.17978, 7.1675)$	$(1.8871, 2.1378)$
E1	6.0369	$(5.5695, 6.5300)$	$(1.7510, 2.1051)$	$(1.7382, 2.2167)$
B2	12.122	$(11.820, 12.409)$	$(0.17451, 6.7320)$	$(1.9066, 2.0295)$
E2	11.756	$(10.884, 12.761)$	$(1.7785, 2.0533)$	$(1.7810, 2.2151)$

algorithm's ability to calibrate the plume on-line. Following data simulation, we used MCMC to generate 60000 samples, left out 20000 of these as burn-in, and used a thinning factor of 10. Adaptation of the Metropolis samplers was only done during burn-in. Convergence was assessed through visual inspection of the MCMC trace plots.

We made inference on Q, as well as all other parameters in the model, for the Boreal- and EC-simulated data and the two emission rate settings. Table 2 shows the posterior median emission rates, the 95% posterior credible intervals for the emission rate, as well as the intervals for the plume standard-deviation scaling parameters ω_y, and ω_z. In all cases, we see that the true (simulated) emission rate is captured within our posterior credible intervals, and that the median estimates are very close to the true values. Interestingly, we see that while the plume-scaling coefficients have been accurately recovered in most cases, the posterior uncertainty over ω_y for the Boreals is very wide. This suggests that ω_y might not be identifiable for path measurements, possibly because the averaging effect of the line integral renders the measured concentration insensitive to a specific plume width in the horizontal direction.

5.2 Application to the Ginninderra data set

In this section we discuss results from applying our model to enhancements from the compiled Ginninderra data. We considered several settings. In the first setting, we estimated the emission rate separately for each of the four instrument types, and for each release-rate period (5.8 g min^{-1} and 5.0 g min^{-1}) when the source was active. In addition, for each release-rate period we estimated the emission rate for all the instruments combined, yielding a total of ten inversion results. In the second setting we estimated the emission rate for the same ten cases, but for periods when the source was switched off. In the third setting we again considered the same ten cases, but using only measurements that were taken when upwind of the source. These three settings serve to demonstrate how our inferences adapt to the various settings one might encounter in the field. In particular, online plume calibration is almost impossible in the latter two settings, and we expect this to result in large posterior uncertainties on the scaling coefficients, and also the emission rate in the third setting. In the second setting downwind measurements are present. Therefore, while online plume calibration is again almost impossible since there is no active source, the absence of

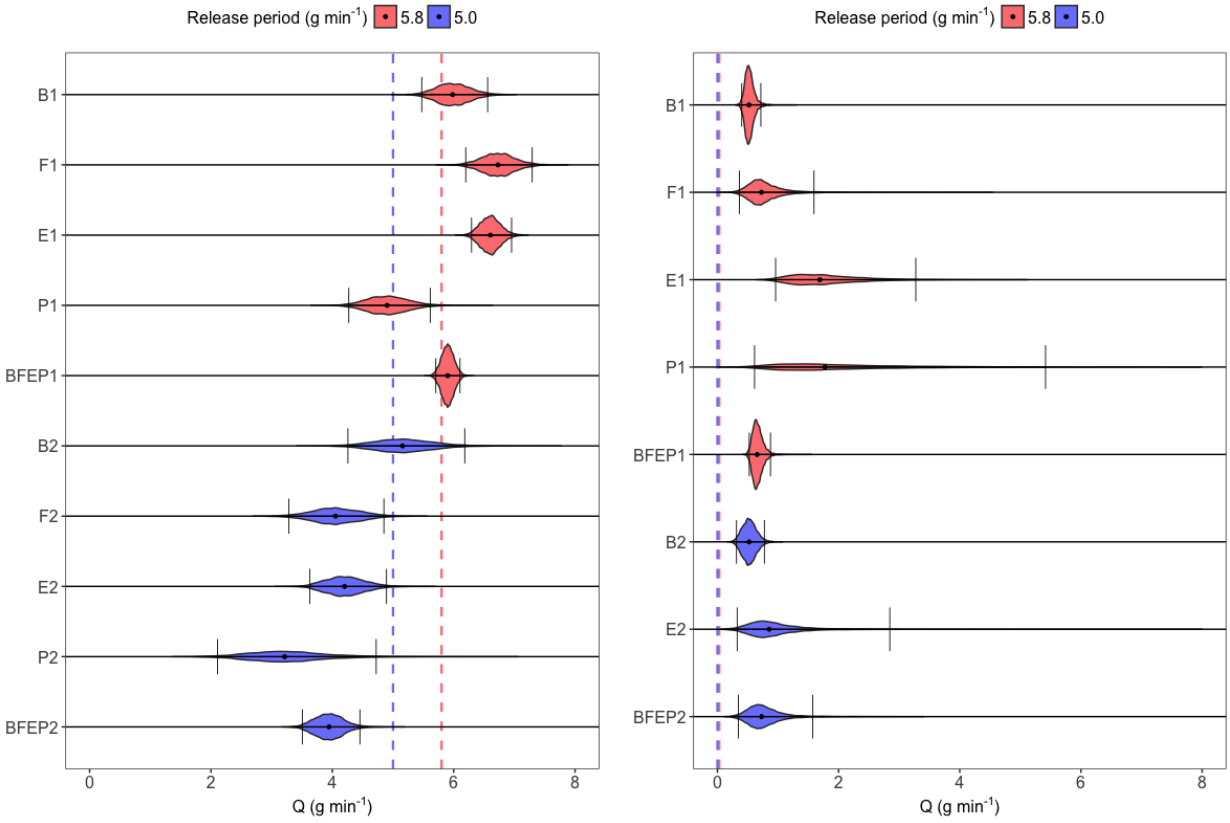

Figure 5. Left: posterior empirical distributions of the emission rate Q in g min^{-1}, for the Boreals (B), FTIRs (F), EC towers (E), Picarro analysers (P), and the ensemble of all instruments (BFEP), for each release-rate period (1 and 2) during the Ginninderra experiment. The 5.8 g min^{-1} release-rate period is shown in red (B1, F1, E1, P1, and BFEP1), while the 5.0 g min^{-1} release-rate period is shown in blue (B2, F2, E2, P2, and BFEP2). The vertical dashed lines denote the respective true emission rates, the black ~~dot represents~~ dots denote the median ~~estimate~~estimates, and the black vertical bars ~~represent~~ denote the upper and lower limits of the 95% posterior credible ~~interval~~intervals. Right: same as the left panel but showing results obtained using measurements taken when the methane point source was inactive. We can recover a reasonable range of estimates for the emission rate, with no 95% posterior credible interval being far from the true emission rate. Further, we see that the posterior emission rate credible intervals move towards zero when the source is inactive, as desired.

a source $(Q = 0)$ should be reflected in our posterior inferences (recall, however, that use of a half-normal prior distribution precludes the possibility of a zero emission rate being estimated; see Sect. 4.2).

As in the OSSE, we generated 60000 MCMC samples, left out 20000 of these as burn-in, and used a thinning factor of 10. In line with what we observed in the OSSE, our initial results showed that, more often than not, ω_y is not identifiable (leading to wide posterior distributions and poor MCMC mixing) when attempting to estimate the emission rate with the source switched

on with *path* measurements. We therefore chose to fix $\omega_y = 1$ (but not ω_z) for path measurements, and this choice is reflected in all the results discussed below.

The left panel of Fig. 5 summarises our results for Q in the first setting (both upwind and downwind measurements with the source switched on); full results are given in the first ten rows of Table A1. While our posterior inferences are reflective of the true underlying emission rate, unlike in the OSSE we see that with the real data the true values were not always captured within our 95% posterior credible intervals. This suggests that there are other important factors at play (e.g., with the meteorological data such as ambient temperature or wind direction, that we assume are fixed and known) that are not (or not fully) accounted for in our model. ~~Despite this~~A close inspection of the residuals at EC.A revealed mild deviations from our Gaussianity assumption, while posterior predictive distributions on left-out EC tower data in a re-analysis revealed coverage probabilities that are slightly too large. Nevertheless, our worst-case scenario, obtained with the combination of all instruments in the 5.0 g min^{-1} release-rate period, had an interval limit which was only 0.55 g min^{-1} (approximately 11%) off from the true value, while all posterior medians were within 36% of the true value (within 22% if one ignored results from the Picarro analysers during the 5.0 g min^{-1} release-rate period). This is encouraging because a single, common inference method was used to obtain the inferences from data at a common temporal resolution – no manual instrument-specific tuning was carried out. The approach thus seems relatively robust to instrument type; in Sect. 6 we show this is no longer the case once certain components in our model are assumed fixed and known.

The first ten rows of Table A1 also show the 95% posterior credible intervals for ω_y and ω_z. None of the obtained credible intervals for ω_y contain 1, and the results corroborate the conclusion from our exploratory data analysis in Sect. 3.1 that a plausible value for ω_y is about 2 or 3. This result lends credence to our ability to calibrate the Pasquill stability-class curves corresponding to σ_{y_i,k_i} while estimating the emission rate with point measurements. There was less agreement on ω_z in the inversions, suggesting that something more complex than a simple scaling is required (or that the model used for σ_{z_i,k_i} is, in this case, inappropriate) for calibrating the Pasquill stability-class curves corresponding to σ_{z_i,k_i}. Nonetheless, in Sect. 6 we show that our emission-rate estimates from point measurements were relatively less sensitive to the assumption $\omega_z = 1$ than to the assumption $\omega_y = 1$.

The right panel in Fig. 5 summarises our results for Q in the second setting (both upwind and downwind measurements with the source switched off), while full results are given in the second set of ten rows in Table A1. Recall from Sect. 4.2 that due to the choice of prior over Q (a half-normal distribution), it is not possible for the 95% credible interval to include zero. Clearly, however, the intervals for Q are close to zero and are suggestive of a small emission rate. As expected, the plume standard-deviation scaling parameters are not well-constrained in this setting when the source is off: Narrow credible intervals on the emission rate here are only possible when the measurement is largely insensitive to the plume shape. This is indeed the case for the Boreal paths, some of which pass very close to the source. With other instrument configurations, uncertainty in the plume scalings dominates. In some cases (FTIRs and Picarro analysers in the 5.0 g min^{-1} release-rate period) our MCMC algorithm did not converge after the 60000 samples; these results are thus omitted from Fig. 5 and Table A1.

The bottom ten rows in Table A1 give full results in the third setting (upwind measurements only with the source switched on). In this setting the 95% posterior credible intervals produced for the emission rates are very wide (most with a range of over

100 g min^{-1}), as are those produced for ω_y and ω_z: Our posterior distributions are largely uninformative. This was expected since upwind measurements contain no information on both the emission rate *and* the plume model parameters. These results from upwind measurements serve as verification, and confirm that we are indeed relying on useful information from downwind measurements when making inference on the emission rate and other parameters that appear within our model.

6 Sensitivity of results to model components

As detailed throughout Sect. 4, the Bayesian model we employ contains many parameters that are updated using MCMC. A natural question to ask is whether all these parameters do need to be updated, and what the effects on the emission rate inferences are when instead some of these are assumed fixed and known. Specifically, we are interested in seeing what happens when: (i) considering only one single precision parameter τ for all of the data regardless of stability class and/or instrument group; (ii) considering one τ_{m_i} per instrument group only; (iii) not accounting for plume-model variability in low wind speeds (i.e., setting $\hat{\boldsymbol{U}} = \mathbf{1}$); (iv) not updating ω_y when using point measurements; (v) not updating ω_z; and (vi) not updating both ω_y and ω_z when using point measurements. The 95% credible intervals for Q in g min^{-1} for all these settings and for each of the 10 groupings considered in Sect. 5 are given in Table A2.

Grouping the precision parameters $\{\tau_{m_i}\}$ parameters by instrument only (instead of by instrument *and* stability class) had a slightly negative impact on the emission-rate estimates obtained during the second release-rate period, but less so during the first release-rate period. Assuming (and fixing) $\omega_z = 1$ for both the point and path measurements also did not have a serious impact on the emission-rate estimates. Note that this does not mean that these components are not relevant in the general model – for example, from our estimates of ω_z in Table A1 we see $\omega_z = 1$ would be a plausible choice for this experiment if one opted to fix ω_z (while $\omega_y = 1$ would not be).

On the other hand several components in our model appear to be crucial to obtaining reasonable emission-rate estimates. Using a single precision parameter to capture all observed variability due to measurement error and the stability-class categorisation clearly had a negative impact on our emission-rate estimates. Similarly, assuming the variability of the measurements is independent of wind speed when doing inversion resulted in 95% posterior credible intervals on the emission rate that are considerably shifted in the negative direction. A similar observation was made by ~~(Feitz et al., 2018, p. 207)~~ Feitz et al. (2018, p. 207) when analysing data from the Boreal lasers. There, observations with wind speeds below 1.5 m s^{-1} were removed to mitigate this effect. ~~An attractive feature of the proposed model in this article is that the "soft" de-weighting of measurements at low wind speeds eliminates the need to determine a hard cutoff value when doing the analysis.~~

The scaling factor ω_y is clearly also crucial for obtaining emission-rate estimates of practical significance for point measurements, with the ensuing emission-rate estimates often being off by nearly a factor of two when $\omega_y = 1$ is assumed. As expected, the width of the credible intervals on the emission rate decreased substantially when $\omega_y = \omega_z = 1$ was assumed, indicating that ω_y and ω_z play a big role in quantifying uncertainty on the emission rate. Therefore, as noted in other studies discussed in Sect. 1, incorporating uncertainty in the transport model by treating parameters within the model itself as uncertain (note that

this is different from adding another component of variability in the data model, as is often done) is likely to have a positive impact on emission-rate estimates and uncertainty quantification.

7 Conclusions

In this article we have proposed a fully Bayesian model for atmospheric tomography that takes into account uncertainty in the data measurement process, the physical processes, and parameters appearing in the transport model, when estimating the emission rate. We see that the model is robust to different instrument types and configurations, and provides useful inferences on the emission rate and the plume dispersion model used. When applied to the Ginninderra data using a variety of instruments in different release-rate periods, we obtain 95% posterior credible intervals on the emission rate that either encapsulate the true emission rate, or that have a limit which is no more than 11% from the true value.

The methods developed in this study are ideal for quantifying local-scale leaks from industrial facilities or from the subsurface (e.g., well heads, buried pipelines or gas leakage up geological fractures and faults) where a surface leak has been detected but needs to be quantified. It can be used where physical access to the source location is limited, e.g., gas bubbling from a creek or where measurement is hazardous. Depending on the circumstance, detection of leakage can take many different forms, from visible bubble detection, optical gas imaging, handheld sniffers, noise detection, helicopters equipped with lasers, drones equipped with gas sensors, to monitoring die-off in vegetation using remote sensing techniques. Surface leakage typically expresses as small, concentrated hotspots if sourced from the subsurface (Feitz et al., 2014; Forde et al., 2019), for which the quantification approach outlined in this article ideally suited. Equipment placement can be optimised around the leakage site (i.e., prevailing upwind/downwind) for optimal quantification.

In most applications the number of sources, nor the source location, is known. As such, the framework we construct should be seen as a foundational building block that needs to be extended appropriately for each specific application. For example, if the source location is not known, then source localisation can be incorporated into the Bayesian framework as discussed by Humphries et al. (2012). If there are multiple possible sites, and these locations are not known, then the framework needs to be further extended to incorporate multiple Gaussian plume models (one for each site), and joint localisation/inversion will be required. While these extensions are straightforward both mathematically and computationally, in practice they are unlikely to be effective for detection of leakage over large spatial scales. Gas fields or geological storage sites can cover areas of tens to hundreds of square kilometres. Unless there is a high density of sensors ($\approx$ 100 m scale, van Leeuwen et al., 2013; Jenkins et al., 2016), the sensitivity of detection will be poor (Wilson et al., 2014; Luhar et al., 2014). It is however relatively straightforward to effectively extend the methodology to when the emission is from an area, rather than a point source.

Our work is closely connected to other atmospheric tomography techniques, but with some small, significant, differences. Luhar et al. (2014) used a backward Lagrangian particle model to simulate the trajectories of methane and carbon dioxide backwards in time to localise the source and estimate the emission rates. Their approach yielded good quality estimates for the methane emission rates, but highly uncertain estimates for the carbon dioxide emission rates and source location parameters.

Twenty-three runs of the Lagrangian model required approximately one hour of computing time, and therefore their framework becomes problematic with thousands of observations as we have in our study. More pertinently, online calibration of the atmospheric-transport model would be virtually impossible without the construction and use of a surrogate model or emulator (e.g., Harvey et al., 2018). In the study of Humphries et al. (2012), carbon dioxide and nitrous oxide emission rates and source locations were estimated relatively well. We do not consider the localisation problem, but otherwise extend their method to handle various instrument types and a number of extra levels of uncertainty. The case in our sensitivity analysis in which we fix $\omega_y = \omega_z = 1$ yields a model that is structurally very similar to that of Humphries et al. (2012); we see from our results that having this hard constraint is not a tenable assumption in practice. Our work also has close connections with that of Ars et al. (2017) where the Pasquill stability class for an observation is chosen from a subset of appropriate stability classes, based on the best fit of model predicted values to observed values. While this may help fit the Gaussian plume dispersion model to the data, it does not take into account the uncertainty arising from stability-class choice. Further, if all plume model standard deviations are off by a factor of two or more, there is a distinct possibility that no stability class yields a good fit. Online calibration of these standard deviations is needed to account for lack-of-fit arising from the the inherently simple Gaussian plume model.

Our results provide interesting insights into the design and monitoring of sensor networks for detecting and quantifying methane emissions. For example, our sensitivity analysis in Sect. 6 showed that estimates using the two Picarro analysers were particularly sensitive to assumptions made on the model plume parameters. Moreover, when uncertainty on these parameters was considered, the release-rate estimates from these instruments tended to be uncertain. This is despite the Picarro analysers being among the more accurate and expensive instruments used in the study. Uncertainty in our experiment is, as is often the case, dominated by that in the transport model. Hence, the number of instruments used, the proximity of the instruments to the source, and their configuration around the source, appear to be more important design criteria than instrument accuracy when the inferential target is emission-rate quantification of a point source. In particular, having more (less expensive) instruments set up to cover many more possible wind directions is better than having only one or two more expensive instruments with which to monitor emissions. If one is limited to using a small number of instruments, then those giving path measurements are preferrable to those giving point measurements, as the former will be able to 'capture' a larger range of wind directions. Our results also provide insight on the transport model used. For example, close inspection of our posterior inferences for $\boldsymbol{\tau}$ indicated that across all instrument groups and for both release-rate periods, the model-data mismatch was much lower for the more neutral stability classes C and D, than for the more stable/unstable classes A and F.

The fully Bayesian framework we adopt is adaptable to various scenarios. We envision, for example, that source localisation (e.g., Humphries et al., 2012; Hirst et al., 2013) could be done in tandem with plume-model calibration within an inversion framework, provided several instruments in suitable configurations (as in the Ginninderra experiment) are available. Future work will also investigate how uncertainty in other meteorological variables such as wind-direction, as well as the stability-class categorisation adopted (possibly via z_0) could be incorporated within the model.

Code and data availability. Software code and data are available at https://github.com/Lcartwright94/BayesianAT.

Appendix A: Full results

Table A1. Posterior median emission rate in g min^{-1}, and the posterior 95% credible intervals for the emission rate in g min^{-1}, ω_y, and ω_z, for the Boreals (B), FTIRs (F), EC towers (E), Picarro analysers (P), and an ensemble of all instruments (BFEP), for each release-rate period (5.8 g min^{-1} (1), and 5.0 g min^{-1} (2)) under various settings. Dashes correspond to parameters that were not updated via MCMC. Results for which MCMC did not converge are marked as NA.

Setting	Group	Median Q	Q	ω_y	ω_z
Source on (Upwinds & Downwinds)	B1	5.9833	$(5.4733, 6.5593)$	—	$(3.2062, 4.2104)$
	F1	6.7301	$(6.1985, 7.2937)$	—	$(1.4347, 1.8164)$
	E1	6.6048	$(6.2942, 6.9537)$	$(2.4946, 2.7848)$	$(1.0868, 1.1954)$
	P1	4.9028	$(4.2710, 5.6136)$	$(2.6065, 3.6707)$	$(0.41664, 0.64341)$
	BFEP1	5.9008	$(5.7050, 6.1038)$	$(2.3360, 2.5640)$	$(1.1944, 1.2989)$
	B2	5.1552	$(4.2571, 6.1820)$	—	$(0.84608, 1.1288)$
	F2	4.0525	$(3.2838, 4.8497)$	—	$(0.66723, 1.0944)$
	E2	4.2017	$(3.6297, 4.8923)$	$(1.4899, 2.1671)$	$(0.90941, 1.0981)$
	P2	3.2135	$(2.1071, 4.7236)$	$(2.0250, 5.2798)$	$(0.34677, 0.63648)$
	BFEP2	3.9455	$(3.5054, 4.4543)$	$(1.7138, 2.5325)$	$(0.97964, 1.1437)$
Source off (Upwinds & Downwinds)	B1	0.52073	$(0.40106, 0.71608)$	—	$(1.3051, 5.0262)$
	F1	0.72641	$(0.36438, 1.5935)$	—	$(1.2565, 9.0531)$
	E1	1.6906	$(0.95997, 3.2742)$	$(10.768, 21.971)$	$(3.1036, 11.826)$
	P1	1.7798	$(0.61237, 5.6367)$	$(3.3985, 13.853)$	$(0.31311, 7.3589)$
	BFEP1	0.65416	$(0.52512, 0.87510)$	$(7.0545, 12.789)$	$(2.2381, 5.3166)$
	B2	0.52202	$(0.31479, 0.77494)$	—	$(0.84995, 1.5319)$
	F2	NA	NA	—	NA
	E2	0.85549	$(0.32681, 3.3683)$	$(2.3136, 11.371)$	$(0.50337, 7.9746)$
	P2	NA	NA	NA	NA
	BFEP2	0.72846	$(0.34557, 1.5735)$	$(2.7823, 9.5185)$	$(0.97971, 7.1704)$
Source on (Upwinds only)	B1	62.452	$(2.7445, 206.22)$	—	$(0.16883, 6.8461)$
	F1	61.651	$(3.2040, 207.05)$	—	$(0.17249, 6.5361)$
	E1	16.136	$(7.5484, 41.030)$	$(4.5921, 6.9288)$	$(1.2708, 8.7488)$
	P1	22.913	$(2.0052, 168.70)$	$(0.15931, 9.2038)$	$(0.23560, 7.7829)$
	BFEP1	15.723	$(7.1673, 39.188)$	$(4.7485, 6.9568)$	$(1.4798, 8.8789)$
	B2	88.353	$(5.6799, 244.48)$	—	$(0.27891, 6.0868)$
	F2	58.568	$(2.7772, 192.74)$	—	$(0.18217, 6.4708)$
	E2	39.728	$(3.2357, 180.33)$	$(0.23683, 5.7448)$	$(0.19650, 6.8680)$
	P2	42.996	$(1.9088, 185.75)$	$(0.13023, 5.2626)$	$(0.18403, 6.9436)$
	BFEP2	37.909	$(2.9071, 186.65)$	$(0.22048, 5.4364)$	$(0.28261, 7.1149)$

Table A2. Posterior 95% credible intervals for the emission rates in g min^{-1} for the Boreals (B), FTIRs (F), EC towers (E) , Picarro analysers (P), and an ensemble of all instruments (BFEP), for each release-rate period (5.8 g min^{-1} (1), and 5.0 g min^{-1} (2)), and for various alterations to the model as detailed in Sect. 6. Dashes correspond to redundant case (e.g., $\omega_y = 1$ was assumed for all path measurements in the full model).

Group	Full model	Assuming $\tau_{m_i} = \tau$ for $m_i = 1, \ldots, M$	Assuming $\{\tau_{m_i}\}$ are only instrument-group dependent	Assuming $\hat{\boldsymbol{U}} = \mathbf{1}$
B1	$(5.4733, 6.5593)$	—	$(4.7238, 5.6727)$	$(2.6092, 3.1975)$
F1	$(6.1985, 7.2937)$	—	$(5.9526, 7.1190)$	$(3.6482, 4.7116)$
E1	$(6.2942, 6.9537)$	—	$(6.2062, 7.0047)$	$(5.4894, 5.9759)$
P1	$(4.2710, 5.6136)$	—	$(4.8748, 6.1139)$	$(2.9868, 3.9053)$
BFEP1	$(5.7050, 6.1038)$	$(4.7252, 5.2433)$	$(5.8133, 6.2731)$	$(3.4032, 3.6424)$
B2	$(4.2571, 6.1820)$	—	$(4.0863, 6.4436)$	$(2.5337, 3.5319)$
F2	$(3.2838, 4.8497)$	—	$(2.7180, 4.2555)$	$(1.4055, 2.1349)$
E2	$(3.6297, 4.8923)$	—	$(3.2692, 9.4560)$	$(3.1329, 4.1516)$
P2	$(2.1071, 4.7236)$	—	$(1.6784, 4.7147)$	$(1.8451, 3.0813)$
BFEP2	$(3.5054, 4.4543)$	$(2.5283, 3.4837)$	$(2.3224, 3.2790)$	$(1.9421, 2.4779)$
Group		Assuming $\omega_y = 1$	Assuming $\omega_z = 1$	Assuming $\omega_y = \omega_z = 1$
B1		—	$(4.0341, 4.7974)$	—
F1		—	$(5.4152, 6.3851)$	—
E1		$(3.3635, 3.7084)$	$(6.8646, 7.5289)$	$(3.6129, 3.8937)$
P1		$(2.0142, 2.5424)$	$(5.8880, 7.5225)$	$(2.6043, 3.4691)$
BFEP1		$(3.6176, 3.8644)$	$(5.9888, 6.3946)$	$(3.8251, 4.0726)$
B2		—	$(4.3543, 5.7021)$	—
F2		—	$(3.2608, 4.7770)$	—
E2		$(2.6442, 3.5321)$	$(3.6982, 4.7588)$	$(2.7116, 3.3605)$
P2		$(0.93638, 5.2326)$	$(1.8757, 4.2556)$	$(0.91678, 1.9052)$
BFEP2		$(2.4699, 3.0227)$	$(3.6202, 4.5213)$	$(2.5319, 3.0744)$

Author contributions. LC compiled the data with the help of all authors and ran all the analyses. LC and AZM conducted the research. AF conceptualised and supervised the study. LC, AZM, and AF wrote the manuscript. SB conducted an initial investigation using a simplified version of the proposed model. IS, FP, TC, KN, TN, MK, SZ, NW and ~~ND~~ NMD acquired the field data for the study. All authors discussed the results and commented on the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. LC acknowledges support of the Australian Government Research Training Program Scholarship. AZM is supported by an Australian Research Council (ARC) Discovery Early Career Research Award (DECRA), DE180100203. LC, AZM, and AF would like to acknowledge APR.Intern for facilitating the first five months of this modelling study. ~~ND~~ NMD is supported by an ARC Future Fellowship, FT180100327. All authors thank Gareth Davies for reviewing an earlier version of this manuscript. The Ginninderra field site was supported by the Australian Government through the Carbon Capture and Storage – Implementation budget measure. The authors also acknowledge funding for the research provided by the Australian Government through the CRC program and support from the CO2CRC. The National Geosequestration Laboratory is thanked for making the two Picarro instruments available for the study. We would like to thank Phil Dunbar and his staff (CSIRO Plant Industry) for maintaining the site and Dale Hughes (CSIRO) for his assistance with maintenance of the CSIRO EC tower. The authors also wish to acknowledge the assistance of Field Engineering Services at Geoscience Australia. Geoscience Australia and the Western Sydney University team would like to acknowledge Charles Jenkins (CSIRO) for early discussions about the atmospheric tomography line technique and the Australian Mathematical Sciences Institute. The University of Wollongong wishes to acknowledge Joel Wilson, Maximilien Desservettaz, and Ruhi Humphries for their assistance in the site operations. AF and IS publish with the permission of the CEO, Geoscience Australia.

References

Ars, S., Broquet, G., Kwok, C. Y., Roustan, Y., Wu, L., Arzoumanian, E., and Bousquet, P.: Statistical atmospheric inversion of local gas emissions by coupling the tracer release technique and local-scale transport modelling: a test case with controlled methane emissions, Atmospheric Measurement Techniques, 10, 5017–5037, 2017.

Basu, S., Baker, D. F., Chevallier, F., Patra, P., Liu, J., and Miller, J.: The impact of transport model differences on CO_2 surface flux estimates from OCO-2 retrievals of column average CO_2, Atmospheric Chemistry and Physics, 18, 7189–7215, 2018.

Berliner, L. M.: Hierarchical Bayesian time series models, in: Maximum Entropy and Bayesian Methods, edited by Hanson, K. M. and Silver, R. N., pp. 15–22, Springer, New York, NY, 1996.

Borysiewicz, M., Wawrzynczak, A., and Kopka, P.: Stochastic algorithm for estimation of the model's unknown parameters via Bayesian inference, Proceedings of the Federated Conference on Computer Science and Information Systems, Wroclaw, Poland, 501–508, 2012.

Casella, G. and Berger, R. L.: Statistical Inference, Duxbury Press, Pacific Grove, CA, 2nd edn., 2002.

Chevallier, F., Feng, L., Bösch, H., I. Palmer, P., and Rayner, P.: On the impact of transport model errors for the estimation of CO_2 surface fluxes from GOSAT observations, Geophysical Research Letters, 37, L21803, 2010.

Etheridge, D., Luhar, A., Loh, Z., Leunning, R., Spencer, D. Steele, P., Zegelin, S., Allison, C., Krummel, P., Leist, M., and van der Schoot, M.: Atmospheric monitoring of the CO2CRC Otway Project and lessons for large scale CO_2 storage projects, Energy Proceedia, 4, 3666–3675, 2011.

Feitz, A., Leamon, G., Jenkins, C., Jones, D. G., Moreira, A., Bressan, L., Melo, C., Dobeck, L. M., Repasky, K., and Spangler, L. H.: Looking for leakage or monitoring for public assurance?, Energy Proceedia, 63, 3881–3890, 2014.

Feitz, A., Schroder, I., Phillips, F., Coates, T., Negandhi, K., Day, S., Luhar, A., Bhatia, S., Edwards, G., Hrabar, S., Hernandez, E., Wood, B., Naylor, T., Kennedy, M., Hamilton, M., Hatch, M., Malos, J., Kochanek, M., Reid, P., Wilson, J., Deutscher, N., Zegelin, S., Vincent, R., White, S., Ong, C., George, S., Maas, P., Towner, S., and Griffith, D.: The Ginninderra CH_4 and CO_2 release experiment: an evaluation of gas detection and quantification techniques, International Journal of Greenhouse Gas Control, 70, 202–224, 2018.

Flesch, T. K., Wilson, J. D., Harper, L. A., Crenna, B. P., and Sharpe, R. R.: Deducing ground-to-air emissions from observed trace gas concentrations: A field trial, Journal of Applied Meteorology, 43, 487–502, 2004.

Forde, O. N., Mayer, K. U., and Hunkeler, D.: Identification, spatial extent and distribution of fugitive gas migration on the well pad scale, Science of the Total Environment, 652, 356–366, 2019.

Ganesan, A. L., Rigby, M., Zammit-Mangion, A., J. Manning, A., Prinn, R., Fraser, P., Harth, C. M., Kim, K.-R., B. Krummel, P., Li, S., Mühle, J., O'Doherty, S., Park, S., K. Salameh, P., P. Steele, L., and Weiss, R.: Characterization of uncertainties in atmospheric trace gas inversions using hierarchical Bayesian methods, Atmospheric Chemistry and Physics, 14, 3855–3864, 2014.

Ganesan, A. L., Manning, A. J., Grant, A., Young, D., Oram, D. E., Sturges, W. T., Moncrieff, J. B., and O'Doherty, S.: Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network, Atmospheric Chemistry and Physics, 15, 6393–6406, 2015.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian Data Analysis, Chapman & Hall/CRC Press, Boca Raton, FL, third edn., 2013.

Golder, D.: Relations among stability parameters in the surface layer, Boundary-Layer Meteorology, 3, 47–58, 1972.

Harvey, N. J., Huntley, N., Dacre, H. F., Goldstein, M., Thomson, D., and Webster, H.: Multi-level emulation of a volcanic ash transport and dispersion model to quantify sensitivity to uncertain parameters., Natural Hazards and Earth System Sciences, 18, 41–63, 2018.

Hirst, B., Jonathan, P., del Cueto, F. G., Randell, D., and Kosut, O.: Locating and quantifying gas emission sources using remotely obtained concentration data, Atmospheric Environment, 74, 141–158, 2013.

~~Houwelling~~

Houweling, S., Bergamaschi, P., Chevallier, F., Heimann, M., Kaminski, T., Krol, M., Michalak, A. M., and Patra, P.: Global inverse modeling of CH_4 sources and sinks: an overview of methods, Atmospheric Chemistry and Physics, 17, 235–256, 2017.

Humphries, R., Jenkins, C., Leuning, R., Zegelin, S., Griffith, D., Caldow, C., Berko, H., and Feitz, A.: Atmospheric tomography: a Bayesian inversion technique for determining the rate and location of fugitive emissions, Environmental Science and Technology, 46, 1739–1746, 2012.

International Energy Agency: Energy Technology Perspectives 2017, OECD/IEA, Paris, France, 2017.

Jacobson, M. Z.: Fundamentals of Atmospheric Modeling, Cambridge University Press, New York, NY, 2nd edn., 2005.

Jenkins, C., Kuske, T., and Zegelin, S.: Simple and effective atmospheric monitoring for CO_2 leakage, International Journal of Greenhouse Gas Control, 46, 158–174, 2016.

Jenkins, J. D., Luke, M., and Thernstrom, S.: Getting to zero carbon emissions in the electric power sector, Joule, 2, 2487–2510, 2018.

Jones, M., Goldstein, M., Jonathan, P., and Randell, D.: Bayes linear analysis for Bayesian optimal experimental design, Journal of Statistical Planning and Inference, 171, 115–129, 2016.

Kinnon, M. A. M., Brouwer, J., and Samuelsen, S.: The role of natural gas and its infrastructure in mitigating greenhouse gas emissions, improving regional air quality, and renewable resource integration, Progress in Energy and Combustion Science, 64, 62–92, 2018.

Lewicki, J. L. and Hilley, G. E.: Eddy covariance mapping and quantification of surface CO_2 leakage fluxes, Geophysical Research Letters, 36, L21 802, 2009.

Loh, Z. M., Leuning, R., Zegelin, S. J., Etheridge, D. M., Bai, M., Naylor, T., and Griffith, D.: Testing Lagrangian atmospheric dispersion modelling to monitor CO_2 and CH_4 leakage from geosequestration, Atmospheric Environment, 43, 2602–2611, 2009.

Lucas, D. D., Simpson, M., Cameron-Smith, P., and Baskett, R. L.: Bayesian inverse modeling of the atmospheric transport and emissions of a controlled tracer release from a nuclear power plant, Atmospheric Chemistry and Physics, 17, 13 521–13 543, 2017.

Luhar, A. K., Etheridge, D. M., Leuning, R., Loh, Z. M., Jenkins, C. R., and Yee, E.: Locating and quantifying greenhouse gas emissions at a geological CO_2 storage site using atmospheric modeling and measurements, Journal of Geophysical Research: Atmospheres, 119, 10 959–10 979, 2014.

Miller, S. M., Hayek, M. N., Andrews, A. E., Fung, I., and Liu, J.: Biases in atmospheric CO_2 estimates from correlated meteorology modeling errors, Atmospheric Chemistry and Physics, 15, 2903–3914, 2015.

Mitchell, T. J. and Beauchamp, J. J.: Bayesian variable selection in linear regression, Journal of the American Statistical Association, 83, 1023–1032, 1998.

Pasquill, F.: The estimation of the dispersion of wind-borne material, The Meteorological Magazine, 90, 33–49, 1961.

Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P.: Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO_2 data, Journal of Geophysical Research: Atmospheres, 107, ACH 5–1–ACH 5–25, 2002.

Rajaona, H., Septier, F., Armand, P., Delignon, Y., Olry, C., Albergel, A., and Moussafir, J.: An adaptive Bayesian inference algorithm to estimate the parameters of a hazardous atmospheric release, Atmospheric Environment, 122, 748–762, 2015.

Riddick, S. N., Connors, S., Robinson, A. D., Manning, A. J., Jones, P. S. D., Lowry, D., Nisbet, E., Skelton, R. L., Allen, G., Pitt, J., and Harris, N. R. P.: Estimating the size of a methane emission point source at different scales: from local to landscape, Atmospheric Chemistry and Physics, 17, 7839–7851, 2017.

Sepulveda, N. A., Jenkins, J. D., de Sisternes, F. J., and Lester, R. K.: The role of firm low-carbon electricity resources in deep decarbonisation of power generation, Joule, 2, 2403–2420, 2018.

Sienfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change, John Wiley & Sons, Hoboken, NJ, 2nd edn., 2006.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, siam, Philadelphia, PA, 2005.

Turner, B.: Workbook of Atmospheric Dispersion Estimates, Lewis Publishers, Boca Raton, FL, 2nd edn., 1994.

van Leeuwen, C., Hensen, A., and Meijer, H. A. J.: Leak detection of CO_2 pipelines with simple atmospheric CO_2 sensors for carbon capture and storage, International Journal of Greenhouse Gas Control, 19, 420–431, 2013.

Wang, Y., Huang, H., Huang, L., and Ristic, B.: Evaluation of Bayesian source estimation methods: A comparison of likelihood functions and distance measures, Atmospheric Environment, 152, 519–530, 2017.

Wark, K., Warner, C. F., and Davis, W. T.: Air Pollution: Its Origin and Control, Addison Wesley Longman, Menlo Park, CA, 1998.

White, E. D., Rigby, M., Lunt, M. F., Ganesan, A. L., Manning, A. J., O'Doherty, S., Stavert, A. R., Stanley, K. M., Williams, M., Smallman, T. L., Comyn-Platt, E., Levy, P., Ramonet, M., and Palmer, P. I.: Quantifying the UK's carbon dioxide flux: an atmospheric inverse modelling approach using a regional measurement network, Atmos. Chem. and Phys. Discussions, in review, 2018.

Wilson, P., Feitz, A., Jenkins, C., Berko, H., Loh, Z., Luhar, A., Hibberd, M., Spencer, D., and Etheridge, D.: Sensitivity of CO_2 leak detection using a single atmospheric station, Energy Proceedia, 63, 3907–3914, 2014.

World Meteorological Organisation: Guide to Meteorological Instruments and Methods of Observation, https://library.wmo.int/pmb_ged/wmo_8_en-2012.pdf, 2008.

Zammit-Mangion, A., Cressie, N., Ganesan, A. L., O'Doherty, S., and Manning, A. J.: Spatio-temporal bivariate statistical models for atmospheric trace-gas inversion, Chemometrics and Intelligent Laboratory Systems, 15, 227–241, 2015.