https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

Quantifying methane emissions from Queensland's coal seam gas producing Surat Basin using inventory data and an efficient regional Bayesian inversion

Ashok K. Luhar[1], David M. Etheridge[1], Zoë M. Loh[1], Julie Noonan[1], Darren Spencer[1], Lisa Smith[2], and Cindy Ong[3]

[1]CSIRO Oceans and Atmosphere, Aspendale, Victoria 3195, Australia
[2]Katestone Environmental Pty. Ltd., Milton, QLD 4064, Australia
[3]CSIRO Energy, Kensington, WA 6152, Australia

Correspondence to: Ashok Luhar (Ashok.Luhar@csiro.au)

Abstract. Methane emissions across Queensland's Surat Basin, Australia, result from a mix of activities, including the production and processing of coal seam gas (CSG). We measured methane concentrations over 1.5 years from two monitoring stations established 80 km apart on either side of the main CSG belt located within a study area of 350 × 350 km^2. Coupling bottom-up inventory and inverse modelling approaches, we quantify methane emissions from this area. The inventory suggests that the total emission is 173 × 10^6 kg CH_4 yr^{-1}, with grazing cattle contributing about half of that, cattle feedlots ~ 25%, and CSG Processing ~ 8%. Using the inventory emissions in a forward regional transport model indicates that the above sources are significant contributors to methane at both monitors. However, the model underestimates approximately the highest 15% of the observed methane concentrations, suggesting underestimated or missing emissions. An efficient regional Bayesian inverse model is developed, incorporating an hourly source-receptor relationship based on a backward-in-time configuration of the forward regional transport model, a posterior sampling scheme, and the hourly methane observations. The inferred emissions obtained from one of the inverse model setups that uses a Gaussian prior whose averages are identical the gridded bottom-up inventory emissions across the domain with an uncertainty of 3% of the averages best describes the observed methane. Having only two stations is not adequate at sampling distant source areas of the study domain, and this necessitates a small prior uncertainty. This inverse setup yields a total emission that is very similar to the total inventory emission. However, in a subdomain covering the CSG development areas, the inferred emissions are 33% larger than those from the inventory.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

1 Introduction

Methane (CH_4) is a major greenhouse gas, with a global warming potential 28 times greater than carbon dioxide (CO_2) (over a 100-year period; IPCC, 2013). It is the second most important anthropogenic greenhouse gas after CO_2 in terms of radiative forcing. Globally averaged surface CH_4 concentrations have increased by almost 160% since pre-industrial times, from a level of 722 ppb to 1859 ppb in 2018 (WMO, 2018), and this increase has been largely due to changes in anthropogenic methane (e.g., IPCC, 2013). Compared to CO_2, the atmospheric lifetime of methane is relatively short (~ 10 years), which means that the climate impact of methane could diminish rapidly following mitigation actions that reduce its emissions. Being chemically reactive, methane also plays an important role as a precursor to tropospheric ozone, itself a greenhouse gas and an air pollutant. Thus, understanding and quantifying methane emissions at various scales is crucial to studying changes in atmospheric radiative forcing and air quality.

Globally, a top-down estimate over the period 2000-2012 suggests that agriculture and waste contribute to about 57% of the total anthropogenic methane emissions, followed by fossil fuels (gas, oil, coal mining and industry) at 32% (Saunois et al., 2016). However, a study using measurements of carbon-14 in methane recently showed that nearly all methane from fossil sources is anthropogenic, and that fossil fuel methane emissions may be underestimated by up to 40% (Hmiel et al., 2020). Significant CH_4 emissions from conventional and unconventional gas fields have been reported in the scientific literature (e.g., Brandt et al., 2014; Schneising et al., 2014; Alvarez et al., 2018).

In the Australian state of Queensland, since the mid-2000s there has been a rapid growth of the production of coal seam gas (CSG), which is virtually pure methane (Towler et al., 2016; DNRM, 2017). CSG, also known as coalbed methane, is classed as an unconventional natural gas, typically extracted from coal seams at depths of 200–1000 m. As of 2015-16, 96% of the gas production in Queensland was CSG, with most of it coming from the Surat Basin (78%, 21187 Mm^3) and the rest (18%, 4958 Mm^3) from the Bowen Basin (DNRM, 2017). With the sharp rise of CSG production, methane emissions from the Surat Basin are a focus, for example, through Australia's CSIRO Gas Industry Social and Environmental Research Alliance (GISERA) (https://gisera.csiro.au) research in Air Quality and Greenhouse Gas. The Surat Basin is predominantly rural, and methane sources other than CSG include agriculture and coal mining. CSG activities that lead to potential methane emissions include CSG wells, pumps, pipelines, vents, pneumatic controls, and produced water bodies (see Day et al., 2013).

The objective of the present paper is to quantify methane emissions from a region of 350 × 350 km^2 of Queensland's side of the Surat Basin (Figure 1, covering the area 148° 17' 43.4"–151° 49' 30.5" E, 25° 3' 48.8"–28° 5' 3.7" S) that encompasses the main CSG production and processing areas using both bottom-up and top-down techniques. The former involves deriving emissions through a compilation of sources and activity data and application of emission factors. We conducted concurrent in-situ atmospheric monitoring of methane during July 2015 – December 2016 at two locations 80 km from each other. The two stations were setup such that they were on either side of the broad present and projected CSG work area in the Surat Basin. These concentration data allow for an independent verification of the bottom-up inventory emissions by using

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

the latter in a forward mesoscale meteorological and transport model and comparing the predicted methane concentrations with the data.

A greater focus in the paper is on the formulation of an efficient top-down, or inverse, modelling methodology for regional scale (~ 100–1000 km), and its application to quantify CH_4 emissions in the Surat Basin. It combines a Bayesian inference approach, an hourly-averaged high-resolution backward-in-time construction of the forward mesoscale meteorological and transport model, and a posterior probability density function (PDF) sampling scheme. A method to correct for time-lag effects in the backward plume methodology is presented. The 1.5 years long hourly methane measurements from the two stations are combined in a Bayesian calculation to derive a top-down emission distribution. Methane background calculation and filtering methodologies are devised. Various Bayesian priors and their uncertainties, including the use of the bottom-up emissions to act as a prior, are tested. The inferred top-down CH_4 emissions are examined alongside the bottom-up inventory emissions for the whole study domain as well as a subdomain containing the CSG activities. We also compare the performance of the top-down emissions by comparing the modelled methane concentrations obtained using them with the observed concentrations. As far as we know, this work is the first in Australia to quantify regional CH_4 emissions through this top-down approach.

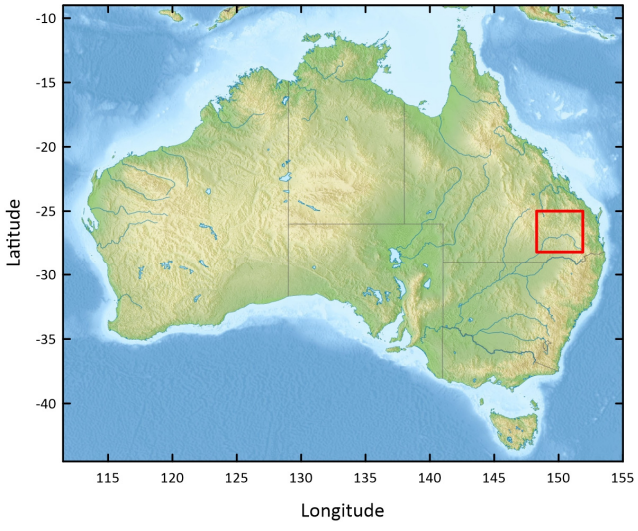


Figure 1. Map of Australia, showing the 350 × 350 km² study domain (red square) of Queensland's part of the Surat Basin. The base relief map is from https://www.mapsland.com/oceania/australia/large-relief-map-of-australia (used under Creative Commons Attribution-ShareAlike 3.0 Licence).

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

2 Monitoring and data filtering

We set up two monitoring stations, namely Ironbark (150° 14' 37.6" E, 27° 8' 6.6" S; 226.806 km east, 6995.596 km north MGA (Map Grid Australia), Zone 56) and Burncluith (150° 42' 5.4" E, 26° 34' 2.4" S; 271.051 km east, 7059.430 km north MGA, Zone 56), located about 80 km apart on two sides of the main coal seam gas belt of the Surat Basin (Figure 3). The selection of the site locations was partly based on a meteorological and dispersion modelling study (Day et al., 2015) that suggested that with the prevailing winds from the north-east and south-west quadrants, long-term continuous monitoring of greenhouse gas concentrations at these two locations would optimise the size and frequency of detection of methane emissions from the broader CSG source region without being unduly impacted by individual sources in the proximity of the measurement sites.

Continuous high frequency (~ 0.3 Hz) measurements of the concentrations of CH_4, CO_2 and water vapour (and also carbon monoxide (CO) at Burncluith) were made at the two sites for about three years with an overlapping period of 1.5 years (July 2015 to December 2016) using Picarro cavity ring down spectrometers (model G2301 at Ironbark, and G2401 at Burncluith) with inlets placed at a height of 10 m. The installations are described by Etheridge et al. (2016). Measured concentrations (strictly speaking, mole fractions in dry air, also volumetric mixing ratios) from each site can be exactly intercompared due to identical calibrations and measurement methodologies. The additional CO measurements at Burncluith are useful in detecting combustion sources of CO_2 and CH_4. Measurement accuracy was better than ± 0.1 ppm for CO_2 and ± 1 ppb for CH_4 (Etheridge et al., 2014). Concurrent meteorological observations included winds measured at 5.8 m AGL (above ground level) at Ironbark and at 7.6 m AGL at Burncluith using sonic anemometers.

The Burncluith station was located on a private farm and there were 30–40 cattle in the paddocks next to it. Occasionally, under suitable meteorological conditions with the cattle upwind of the inlet, the emissions from the local cattle caused one or many sharp peaks in the observed methane signal, typical of a nearby point source. We developed a method which removes these sharp, transient peaks but does not alter the underlying signals from the numerous, region-wide feedlots, grazing cattle or other sources. This filtering method is described in Supplement S1.1 and, for consistency, was also applied to the data from Ironbark, although local cattle are less in number and further away at this site.

Frequently, high methane concentrations at the two sites were observed at night under light wind stable conditions, particularly at Burncluith. In spite of being of much practical interest, however, light winds are difficult to represent in a mesoscale meteorological and transport model. The causes for that include inadequate physical understanding of light-wind processes, flow properties being very sensitive to local topography, and model resolution constraints (Luhar and Hurley, 2012). As a practical measure, we filtered out the nighttime sampling hours for light wind conditions, and this method is described in Supplement S1.2.

Methane emissions due to biomass burning are not part of the bottom-up inventory that we consider in the present modelling due to their being sporadic and highly unpredictable. Enhanced levels of CH_4 and CO were detected at Burncluith in the

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020


Atmospheric Chemistry and Physics Discussions

course of forest fires in the northern sector of Burncluith and wood-heater operations from the property located in the proximity of the monitoring station. The observed CO was used to filter out these occasional biomass burning events from the measured concentration time series, which is an approach similar to that used by Jeong et al. (2012). Details of the CO filter are given in Supplement S1.3.

The number of data hours after the filtering was 6432 for Ironbark and 4149 for Burncluith (cf. the original, valid number of data points of 10938 and 12660, respectively). Unless stated otherwise, the filtered CH_4 data were used for our analysis and modelling.

3 Bottom-up emission inventory

Activity data for the year 2015 were used to develop a bottom-up emission inventory for methane for the Surat Basin. The emission inventory covered a domain of 345 × 345 km^2 with a spatial resolution of 1 × 1 km^2. Standard methodologies were generally adopted with data from various State and Federal Government Departments (e.g. (National Pollutant Inventory (NPI), National Greenhouse and Energy Reporting (NGER), and National Resource Management (NRM)). The bottom-up inventory included the following fourteen emission sectors: (1) feedlots, (2) grazing cattle, (3) piggeries, (4) poultry farms, (5) power stations, (6) coal mining, (7) CSG processing, (8) CSG production, (9) domestic woodheating, (10) vehicular traffic, (11) land-fills, (12) sewage treatment plants, (13) river seepage, and (14) geological seepage. The first four can be grouped as agricultural activities. The inventory excluded CH_4 emissions from burning of biomass, land clearing, termites, ground-water wells (that were registered), wetlands, or fuel consumption and any material handling related to mining activities. Additional details pertaining to the bottom-up inventory compilation are given in Supplement S2.

Figure 2 presents the bottom-up inventory emissions attributed to the various sectors in the Surat Basin, with the total emissions being 173.2 × 10^6 kg CH_4 yr^{-1}. Grazing cattle has the largest contribution, followed by cattle feedlots and CSG processing. We use this emission inventory for our study duration, July 2015–December 2016, with the assumption that any emission changes from the year 2015 to 2016 were insignificant. It is also assumed that all emissions are invariant with time. Although diurnal and seasonal variations for some emissions, viz. wood-heating, traffic, and power plant, are available in the raw data used in the inventory, contributions from these emissions are amongst the smallest and, therefore, we averaged these emissions over the full year for the purpose of computational efficiency in the modelling conducted here.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

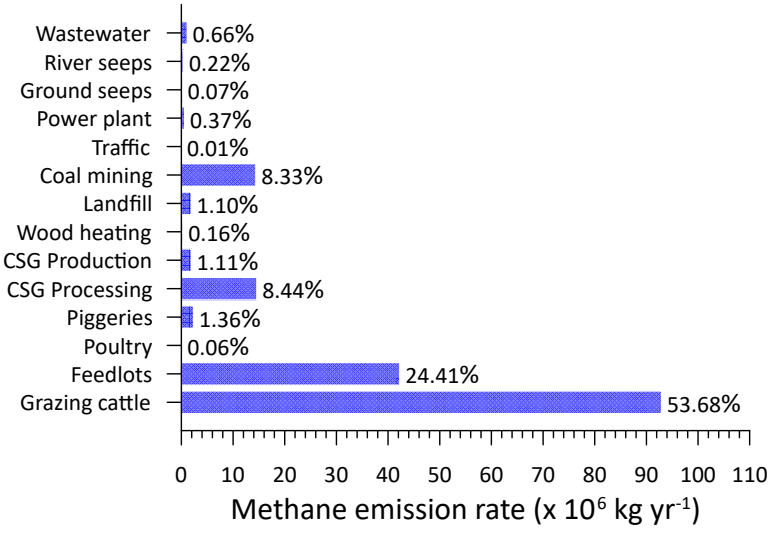


Figure 2. Bottom-up methane inventory emissions from the Surat Basin by sector/source; % of the total also shown. The total emission is 173.2 × 10^6 kg CH_4 yr^{-1}.

Figure 3a presents the distribution of inventory methane emissions (kg yr^{-1} $gridcell^{-1}$) at a grid resolution of 5 × 5 km^2 (69 × 69 grid points). There are localised sources as well as extensive, uniformly distributed source areas. The latter are emissions due to grazing cattle. These emissions are plotted in Figure 3b in which four different coloured areas are the so-called National Resource Management (NRM) regions. In each of these regions the available total number of grazing cattle was distributed uniformly, with the total number of grazing cattle in the study area being 1,086,059. There were 235 cattle feedlots and Figure 3c shows the distribution of their emissions. These are localised, but distributed throughout the region, with some located between the two monitoring stations. Two mining source areas are also located between the two monitoring stations (Figure 3d).

The CSG emissions are shown in Figure 3e (processing) and Figure 3f (production). The CSG production emissions are from wellhead (separators, wellhead control equipment, maintenance and leaks), combustion (flaring, well head pumps, backup generators, and diesel used by vehicles) and pipeline emissions (high point vents on produced water pipelines and pipeline control equipment) (Day et al., 2013). The CSG processing sources consist of processing facility emissions (control equipment, compressor venting, and gas conditioning units), combustion emissions (flaring, plant compressors, backup generators, and diesel used by vehicles), and collection and storage of water produced. Emissions from some of the CSG sources are continuous while others are intermittent (however, the inventory assumes all CSG emissions are time invariant). There were 5 CSG operators with 13 processing facilities and 4628 wells within the study domain. The well numbers

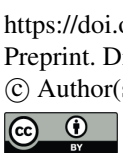
https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020


Atmospheric Chemistry and Physics Discussions

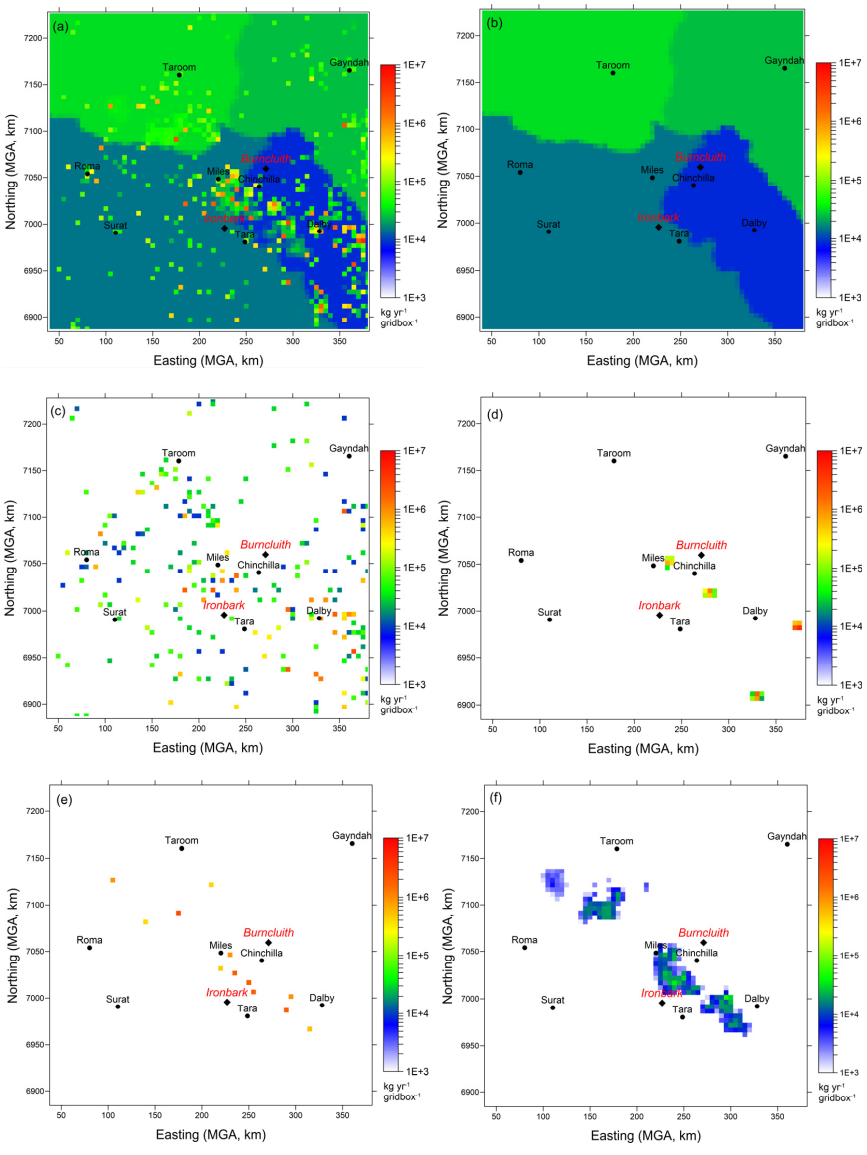
included CSG producing (~ 85%) as well as exploration/appraisal/capped wells. Because of insufficient information, methane emissions from two of the five operators are not part of the inventory, but it was established that these two operators only accounted for about 1.5% of emissions related to CSG activities (which include a total of 256 wells). The biggest contributor to the total CSG methane emissions was venting (88%) from processing, followed by flaring (8%) from both processing and production.

All major sources considered in the bottom-up emissions, namely grazing cattle, feedlots, CSG processing and production, and coal mining, have considerable uncertainty, which arises from uncertainty in both the activity data and emission factors, for example their potential temporal variation and how up to date they are with respect to the study period considered. It is difficult to calculate the uncertainty accurately, but a rough estimate of the upper bound of uncertainty in each of these source emissions is ± 50%.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020



Atmospheric Chemistry and Physics Discussions



170 **Figure 3. Bottom-up methane inventory emissions from the Surat Basin (kg CH_4 yr^{-1} $gridbox^{-1}$, the grid-box size is 5 × 5 km^2). Also shown are the Ironbark and Burncluith monitoring sites and some towns. (a) All emissions, and those due to (b) grazing cattle, (c) cattle feedlots, (d) coal mining, (e) CSG processing, and (f) CSG production.**

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

4 Modelling regional methane using the bottom-up inventory

We use the above inventory emissions in a (forward) regional meteorological and transport model and compare the modelled methane with the ambient measurements from the two sites.

4.1 Model setup

The prognostic mesoscale model used is The Air Pollution Model (TAPM vn4.0.4) developed by CSIRO, which has coupled meteorological and dispersion components and which is designed for applications ranging in scale from local to regional (~ < 1000 km) (Hurley et al., 2005; Hurley and Luhar, 2009). The model has previously been applied to a variety of dispersion problems, such as those reported by Luhar and Hurley (2003), Luhar et al. (2008), Luhar and Hurley (2012) and Matthaios et al. (2017).

TAPM can be used in a one-way nestable mode. The global databases input to the model include land use, terrain height, leaf-area index, synoptic-scale meteorological reanalyses, and sea-surface temperature (SST).

The meteorological module of TAPM predicts the finer-scale flow against a background of larger-scale meteorology supplied by the input synoptic-scale reanalyses (or forecasts). The dispersion module makes use of the predicted finer-scale meteorology and turbulence fields and comprises an Eulerian grid-based conservation equation for species concentration (Hurley et al., 2005).

TAPM uses the synoptic-scale meteorological reanalyses given for horizontal winds, moisture and temperature and available from the U.S. NCEP (National Centers for Environmental Prediction) every 6 hours at a spatial resolution of 2.5° × 2.5° on several levels.

We applied TAPM for the duration 1 July 2015 – 31 December 2016 by using two nested domains: 370 × 370 km^2 with grid resolution 5 × 5 km^2 and 1110 × 1110 km^2 with grid resolution 15 × 15 km^2. Both domains had 75 × 75 grid points and were centred on (150°4.5’ E, 26°35’ S), which is equivalent to 208.657 km east and 7056.383 km north in MGA. There were 25 vertical levels, of which the lowest four were 10 m, 25 m, 50 m and 100 m AGL.

The bottom-up inventory emissions lie within the inner model domain. In this model setup, each inventory emission grid cell (at 5 × 5 km^2) was considered as an area source, apart from the emissions from the power stations which were taken as point sources together with specification of their stack heights and plume-rise parameters. For computational efficiency, rather than considering all 14 emission categories plotted in Figure 2 as separate sources, we aggregated them into 9 sectors with each sector taken as a tracer source: Grazing cattle (Source 1); Feedlot, Piggeries and Poultry (Source 2); CSG Processing (Source 3); CSG Production (Source 4); Mining (Source 5); River seeps (Source 6); Domestic wood heating, Wastewater treatment and Motor vehicles (Source 7); Ground seeps and Landfill (Source 8); and Power stations (Source 9). The relative emissions (%) of the above nine Sources are 53.8, 25.8, 8.4, 1.1, 8.3, 0.21, 0.82, 1.2 and 0.37.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

4.2 Estimated background methane concentration

Since the simulated methane does not include the background levels that are representative of methane emissions located outside the bottom-up inventory, we devised a method for estimating hourly varying background CH_4 for each site involving concentrations under high atmospheric mixing conditions and the hourly standard deviation of concentration (see details in Supplement S3. The estimated background concentration can be either added to the simulated methane or subtracted from the observed methane.

Figure 4 presents (green line) the estimated time series of the background CH_4 concentration for Ironbark, showing a marked seasonal variation with a peak in September (early spring) and a minimum in February (late summer). To view the background variation with respect to the measured methane signal, we also present in Figure 4 as dot points the unfiltered hourly mean observations (clipped at 2100 ppb). The estimated background concentration time series for Burncluith looks very similar (not shown). The uncertainty in the background CH_4 is 3.6 ppb and 3.3 ppb for Ironbark and Burncluith, respectively. The difference between the estimated background at Ironbark and that at Burncluith (purple line in Figure 4) is small and within ± 5 ppb. Any difference between the two backgrounds could be due to different sites in the study area getting impacted by different out-of-domain emissions depending on the transport meteorology. On average, the background concentration at Ironbark is greater by 1 ppb, and the standard deviation of the difference is 1.4 ppb. We take the average of the two background time series to represent the regional hourly background CH_4 concentration, with an average uncertainty of 3.5 ppb.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

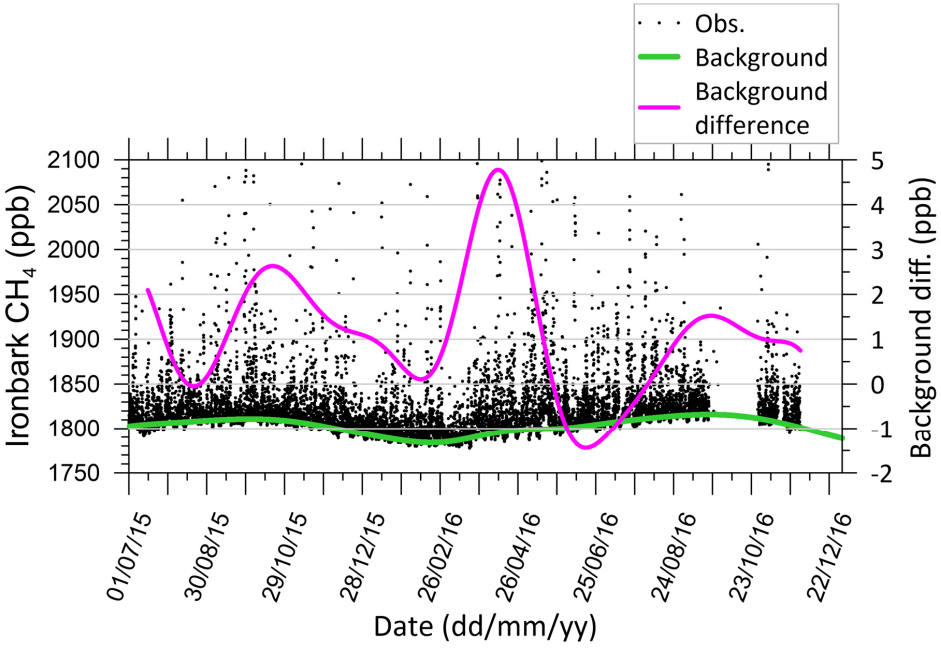


Figure 4. Estimated hourly-averaged background CH_4 concentration time series at Ironbark (green line), and the difference between the estimated backgrounds between Ironbark and Burncluith (purple line). The data points are the hourly mean measurements at Ironbark without any filtering (clipped at 2100 ppb to make the background concentration variation stand out better).

4.3 Model performance for meteorology

Accurate modelling of the flow field over our region of interest is important as it controls the atmospheric plume transport and dispersion which in turn influences the accuracy of prediction of CH_4, and conversely the accuracy of inferred emissions. The hourly-averaged winds extracted from the model output for the inner nest at a height of 10 m were compared with the observations from the two monitoring stations for the duration of the simulation, with the missing data hours not considered. The details of the model performance for meteorology is given in Supplement S4. At both sites, the measured winds were most frequent from the north-east sector, with those at Burncluith being generally weaker in strength than those at Ironbark. As judged from the correlation coefficient (r) and index of agreement (IOA) values, the performance of TAPM for wind speed and wind direction was comparable to that obtained in other TAPM modelling studies.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

4.4 Modelled methane compared to observations

The hourly-averaged methane concentrations simulated for individual 9 source categories were aggregated and added to the estimated background concentration to compare with the observed, filtered CH_4 concentrations.

The scatter plots in Figure 5 comparing the modelled and observed CH_4 at the two sites display a substantial degree of scatter, which is not unusual for atmospheric transport and diffusion models driven by predicted meteorology and using hourly-averaged concentrations paired in both time and space (e.g. Luhar et al., 2008). While the correlation coefficient values of 0.57 and 0.74 for Ironbark and Burncluith, respectively, imply a reasonable model prediction, it is clear that the modelled levels are generally lower than the observations, particularly the higher-end concentrations at Ironbark.

There could be various reasons for the differences between the modelled and observed methane, including uncertainty associated with the bottom-up emission inventory, its potential temporal variation, sources missing from the emission inventory, potential changes to the 2015 bottom-up inventory used here in the year 2016 (see Section 7.4), and uncertainty in the model's ability to fully represent the atmospheric processes within the study domain.

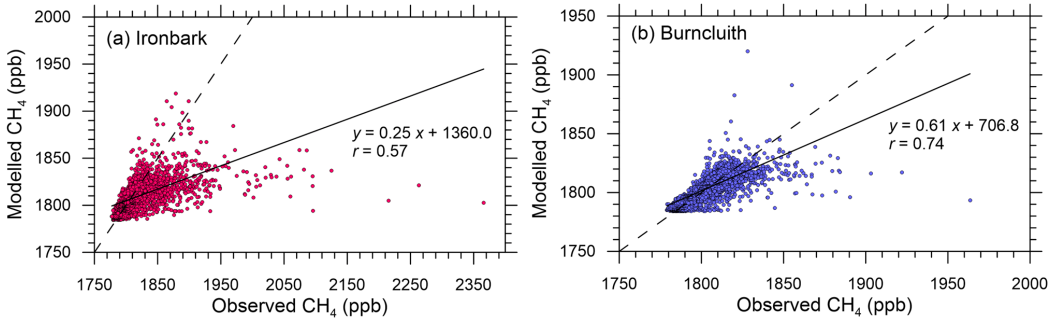


Figure 5. Hourly-averaged observed methane plotted against the simulated methane for the two monitoring stations. The solid line is the least-squares fit.

The comparison in Figure 5 involving hourly methane paired in time and space enables a simple, yet stringent, validation check of a transport model, especially one that is driven by turbulent flow fields predicted by a prognostic meteorological model instead of observations. A complementary but less stringent approach in validating air quality models is the quantile-quantile (q-q) plot, which is a graphical technique for testing "goodness of fit" between two distributions. In such a plot, typically, sorted modelled concentrations are plotted against sorted observed values (i.e. unpaired in time) at a monitoring location (e.g., Venkatram et al., 2001; Luhar and Hurley, 2003; http://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm). If the two sets come from a population with the same

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

distribution, the data points should fall approximately along the 1:1 line. The principal advantage of a q-q plot is that a "good fit" is easy to recognize, and various distributional aspects, such as shape, tail behaviour and outliers, can be simultaneously examined.

In the q-q plot in Figure 6 for Ironbark, the observed CH_4 distribution is modelled well for measurements < 1820 ppb, but for higher observed concentrations, which account for approximately 25% of the sample size, the modelled values are smaller For Burncluith, the q-q plot shows a substantially better model performance, with the model underestimation of higher-end (> 1820 ppb) methane observations, which is approximately 10% of the sample size, much reduced compared to Ironbark. Overall, TAPM is largely predicting the observed CH_4 distribution correctly, except for a relatively few higher-end concentrations.

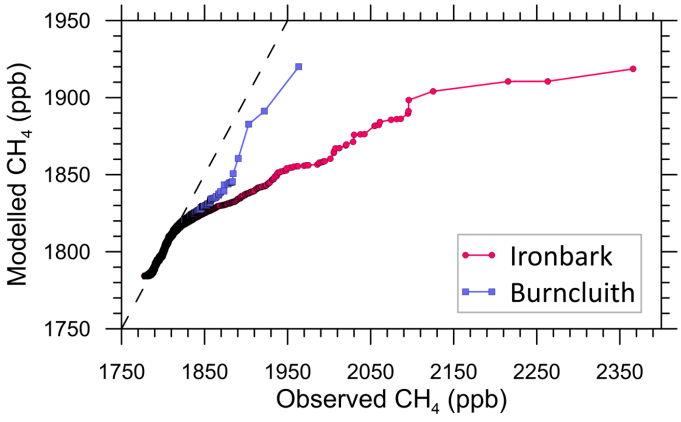


Figure 6. Q-q plot showing the sorted hourly-averaged observed CH_4 concentrations versus the sorted modelled ones at Ironbark and Burncluith. The line of perfect agreement (dashed line) is also shown.

4.5 Contribution to the modelled methane by various source categories

The top four source categories based on their contribution to the modelled CH_4 averaged over the study period at Ironbark were Source 1 (45%, Grazing cattle), Source 2 (25%, Feedlot, Piggeries and Poultry), Source 3 (19%, CSG Processing), and Source 5 (5.5%, Mining). These were the same at Burncluith, but with their respective contributions being 69%, 17%, 6.4% and 4.1%. The CSG Production (Source 4) contributions are 2.2% and 0.73%, respectively, at the two sites.

In contrast, the largest four contributors to the top 5% of the simulated methane at Ironbark turn out to be Source 3 (35%), Source 2 (27%), Source 1 (25%) and Source 5 (7%). These at Burncluith are Source 1 (28%), Source 2 (25%), Source 3

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

(22%) and Source 5 (13%). The CSG Production (Source 4) contributes 3.8% and 2.5%, respectively, at the two sites. The Source 2 grouping is dominated by Feedlots.

The CSG Processing (Source 3) emissions are localised near the two sites which result in methane spikes under favourable winds and thus contribute more to the higher-end modelled methane than to the overall average methane. In contrast, the simulation average methane is dominated by Sources 1 and 2 because concentration enhancements due to these sources occur under most wind conditions as a result of their very wide distribution across the region.

5 Regional top-down, or inverse, modelling for emission estimation

Given that the top-down emission inventory underestimates the observed methane in the Surat Basin, then one may ask what is the quantity and distribution of methane emissions that is implied by the methane concentration measurements at Ironbark and Burncluith? This is addressed by the inverse modelling approach for regional emissions formulated and applied below.

5.1 Bayesian inverse modelling approach

Our inverse model uses a Bayesian inference approach that incorporates, a source-receptor relationship, concentration measurements, and prior information on source parameters (i.e. source information obtained independently of the measurements) (Rao, 2007; Singh et al., 2015). The approach updates the source prior as concentration measurements are considered, and accounts for both model and observational uncertainties.

Several applications using the Bayesian approach have previously been conducted for methane source estimation, including those at local scale (Yee and Flesch, 2010; Luhar et al., 2014; Feitz et al., 2018) and regional scale (Jeong et al., 2012; Miller et al., 2014; Henne et al., 2016; Cui et al., 2017).

The approach hinges on Bayes' theorem (Jaynes, 2003):

$$p(\mathbf{q}|\mathbf{c}) = \frac{p(\mathbf{c}|\mathbf{q}) \cdot p(\mathbf{q})}{p(\mathbf{c})}, \quad (1)$$

where the *prior* PDF $p(\mathbf{q})$ reflects our knowledge of the source parameter vector $\mathbf{q}$ prior to receiving the concentration observations $\mathbf{c}$; $p(\mathbf{c}|\mathbf{q})$ is the *likelihood* function which is the probability of experiencing $\mathbf{c}$ for a given $\mathbf{q}$ and is typically obtained using a model-derived source-receptor linkage; the *posterior* $p(\mathbf{q}|\mathbf{c})$ relates to the update of $p(\mathbf{q})$ by its modulation by $p(\mathbf{c}|\mathbf{q})$ which contains the new information brought in by the concentration measurements $\mathbf{c}$; and $p(\mathbf{c})$ [$=\int p(\mathbf{c}|\mathbf{q})p(\mathbf{q})d\mathbf{q}$] is the *evidence* and is basically a normalisation constant in the present application (Yee and Flesch, 2010). The likelihood function, also termed the source-receptor relationship, is derived using a transport and dispersion model.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

It is assumed that the number of sources (N_s) and their locations $(\mathbf{x}_{s,1}, \ldots, \mathbf{x}_{s,j}, \ldots, \mathbf{x}_{s,N_S})$ where $\mathbf{x}_{s,1} \equiv (x_{s,1}, y_{s,1}, z_{s,1})$ are given *a priori* and the source emissions are non-zero. The emission rates of these sources are to be estimated, and these are represented by $\mathbf{q} \equiv (q_1, \ldots, q_j, \ldots, q_{N_S})$ with a total of N_S unknown emission rates. Assuming each source emission to be independent, the prior PDF can be written as:

$$p(\mathbf{q}) = \prod_{j=1}^{N_s} p(q_j). \tag{2}$$

Assuming that the model and measurement uncertainties are independent and distributed normally, the total likelihood of all $\mathbf{c}$ for a given hypothesis of $\mathbf{q}$ is calculated as (Yee, 2012)

$$p(\mathbf{c}|\mathbf{q}) = \prod_{i=1}^{N_m} \frac{1}{\sqrt{2\pi}\left(\sigma_i^2 + \sigma_{m,i}^2\right)^{1/2}} \exp\left\{-\frac{\left(c_{m,i}(\mathbf{q}) - c_i\right)^2}{2\left(\sigma_i^2 + \sigma_{m,i}^2\right)}\right\}, \tag{3}$$

$\mathbf{c} \equiv (c_1, \ldots, c_i, \ldots, c_{N_m})$, c_i is the observed concentration at i-th instant (time and location), $c_{m,i}$ is the corresponding modelled concentration for a given hypothesis of $\mathbf{q}$, σ_i is the independent measurement error, $\sigma_{m,i}$ is the independent model error, N_m is the number of concentration data (which can be time series from several independent monitors). $c_{m,i}$ for all hypotheses, or possible values, for $\mathbf{q}$ is calculated and used in constructing the likelihood distribution $p(\mathbf{c}|\mathbf{q})$. Hence the posterior PDF for a given source hypothesis $\mathbf{q}$ is calculated as:

$$p(\mathbf{q}|\mathbf{c}) = \frac{1}{Z_0} \prod_{j=1}^{N_s} p(q_j) \prod_{i=1}^{N_m} \frac{1}{\sqrt{2\pi}\left(\sigma_i^2 + \sigma_{m,i}^2\right)^{1/2}} \exp\left\{-\frac{\left(c_{m,i}(\mathbf{q}) - c_i\right)^2}{2\left(\sigma_i^2 + \sigma_{m,i}^2\right)}\right\}, \tag{4}$$

where Z_0 is equivalent to $p(\mathbf{c})$ and is essentially a normalisation constant. The posterior yields probabilities of all emission rates ($\mathbf{q}$) considered.

The total modelled concentration at a given location $\mathbf{x}_r$ and time is determined as

$$c_{m,i} = \sum_{j=1}^{N_s} c_{m,ij}. \tag{5}$$

Because methane is treated as a passive tracer, the concentration field simulated for one rate of emission can be scaled linearly for another without the need to re-run the model. Thus

$$c_{m,ij} = q_j \alpha_{ij}(\mathbf{x}_{s,j}, \mathbf{x}_{r,i}), \tag{6}$$

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

for each emission rate component of $\mathbf{q}$. The quantity $\alpha_{ij}(\mathbf{x}_{s,j}, \mathbf{x}_{r,i})$ is the source-receptor relationship or coupling coefficient and is equivalent to the modelled mean concentration at a given time and location $\mathbf{x}_{r,i}$ due to j-th source release at location $\mathbf{x}_{s,j}$ with a unit emission rate.

In Eq. (4), in the absence of an informative prior, a uniform prior PDF can be used with the given limits (q_{max}, q_{min})

$$p(q_j) = \frac{1}{q_{max,j} - q_{min,j}}, \tag{7}$$

with the probability being zero outside these bounds.

If the prior is Gaussian, then

$$p(q_j) = \frac{1}{\sqrt{2\pi}\,\sigma_{p,j}} \exp\left\{-\frac{(q_j - q_{p,j})^2}{2\sigma_{p,j}^2}\right\}, \tag{8}$$

where q_p and σ_p are the prior mean emission rate and its standard deviation, respectively.

High dimensionality of the posterior makes its direct computation and the subsequent integration (the 'brute-force' method) over the source-parameter space very expensive or perhaps even impossible. For Gaussian priors and uncertainties, the posterior can be solved for the mean and variance with their analytical matrix forms (Tarantola, 2005; Jeong et al., 2012). To make the inverse approach more generally applicable and efficient, we use a Markov chain Monte Carlo (MCMC) technique incorporating the Metropolis-Hastings algorithm to sample the posterior PDF (Tarantola, 2005; Yee, 2012). With MCMC, non-Gaussian priors or uncertainties, or parameters with known physical constraints can also be included (Miller et al., 2014). The normalization constant Z_0 in Eq. (4) need not be known before MCMC samples can be drawn from the posterior PDF. This ability to generate a sample without knowing this constant of proportionality (which is often extremely difficult to compute) is a major feature of MCMC algorithms (Luhar et al., 2014). The frequency distribution of the MCMC-generated samples represents the posterior.

The posterior PDF can be marginalized to obtain the mean emissions rate for each source as follows:

$$\bar{q}_j = \int q_j \, p(\mathbf{q}|\mathbf{c}) \, d\mathbf{q}, \tag{9}$$

and likewise, the variance can also be determined.

5.2 Construction of the hourly source-receptor relationship

In order to use hourly measurements, the source-receptor relationship needs to be calculated every hour for every source (real or potential) location and every monitor location using either forward or backward transport modelling (Rao, 2007). Generally speaking, if the number of source locations under consideration is greater than the number of receptor locations (as for the present case) then the backward approach is much more computationally efficient (Luhar et al., 2014).

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020


Atmospheric Chemistry and Physics Discussions

In the backward approach, source matter is tracked backwards in time from a monitor treated as a source. The value at a given point of the constructed backward concentration field is analogous to the magnitude of contribution made by an emitting source at that point to the true (i.e. forward) modelled concentration at the monitor. Hence, we can use a single backward source-receptor relationship distribution determined every hour to get the contribution made by each real or potential source located in the domain. This contrasts with the forward modelling approach in which each source location has to be considered as a unique, separate source and its dispersion computed for every hour. Essentially, the source-receptor relationship furnishes a way to chart the distribution of source potential within given geographical domain. However, it does not quantitatively allocate the real contribution of sources within the domain to the concentration levels detected at monitoring stations— this is done by the Bayesian inference (Eq. (4)).

One backward approach for regional scale is to use backward trajectories constructed by only using three-dimensional winds computed from a meteorological model (e.g., Cheng et al., 1993). However, such wind trajectories only represent advective transport and do not account for turbulent mixing which causes a plume to disperse as it travels in the atmosphere. If measurements given at a high temporal resolution, e.g. hourly averages, are to be used for inversion it is necessary that the influence of atmospheric flow and dispersion processes that occur at such scales is considered. This can only be properly done by simulating backward tracer plumes which considers both advection and turbulent mixing.

We modify TAPM to construct backward dispersing plumes. The Eulerian dispersion module in TAPM comprises a solution of the advection-diffusion equation for the ensemble mean concentration c, which for a passive species is (e.g. Yee et al., 2008):

$$\frac{\partial c}{\partial t} + \bar{\mathbf{u}}.\nabla c - \nabla.(\mathbf{K}\,\nabla c) = S, \tag{10}$$

in which the unknown turbulent flux terms are closed using the K-theory or gradient transport approach. The forcing term S represents species emissions. The elements of the eddy diffusivity tensor $\mathbf{K}$ are zero except along its main diagonal (K_x, K_y, K_z). The diffusion is assumed to be symmetric in the horizontal plane, so $K_x = K_y = K_H$ (say). K_H and K_z are determined using the modelled turbulent kinetic energy (TKE) and the TKE dissipation rate.

The vertical component $\bar{w}$ of the mean wind vector $\bar{\mathbf{u}}$ ($\equiv \bar{u}, \bar{v}, \bar{w}$) in Eq. (10) is determined by using the continuity equation after the mean horizontal wind velocity components ($\bar{u}, \bar{v}$) are calculated.

The Eulerian adjoint of Eq. (10) describes the backward evolution of a scalar field (c^*), and is also termed backward or retro plume, adjoint function, sensitivity function, or influence function, and is given as (Marchuk, 1995; Pudykiewicz, 1998; Hourdin and Talagrand, 2006; Yee et al., 2008)

$$-\frac{\partial c^*}{\partial t} - \bar{\mathbf{u}}.\nabla c^* - \nabla.(\mathbf{K}\nabla c^*) = M, \tag{11}$$

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020


Atmospheric Chemistry and Physics Discussions

where M is the forcing term representing the measurement distribution, which is treated as a source at the measurement (or receptor) location.

The implementation of Eq. (11) in TAPM is done through changes in the forward model code as follows. The meteorological and turbulence fields calculated by the model at every hour (not hourly-averaged) are stored for the full simulation period. The modelled horizontal components $(\bar{u}, \bar{v})$ of wind are reversed (i.e. by sign change). The (inverted) vertical wind component $(\bar{w})$ is then calculated by solving the continuity equation given the reversed horizontal wind components. The turbulence parameters remain the same. The diffusivities in the dispersion component are positive and do not have any correction for counter-gradient flux in the vertical, and, therefore, they were not modified for the backward mode. The two monitor locations were treated as separate 'sources' each having unit emission, and hourly-averaged plume dispersion fields due to these 'sources' was determined by running the TAPM dispersion module backwards in time for the entire simulation duration by using the reversed winds calculated previously. The meteorological and turbulence fields were linearly interpolated in time for dispersion calculations for model time steps lying between two successive hours. The resulting hourly-averaged backward concentration fields were used as the source-receptor relationship. Since we assume that all methane sources are located near the ground within the lowest model level (i.e. 10 m AGL), only the 10-m hourly source-receptor relationship was required.

One complexity with doing a backward dispersion calculation using one continuous release over the full simulation period over a large domain, as done here, is that the source-receptor field at a given hour is a superposition of plume footprints from the current hour as well as previous hours (typically up to 4–5 hours for the present domain size). So, there is a time history in the source-receptor field at a given time (whose influence becomes smaller and smaller as the distance between the source and the receptor becomes smaller, or the averaging time is increased, or when the winds are strong). However, this time history in a backward run corresponds to future hours in a forward run, so at a given hour there can be a time mismatch between the forward concentration at a grid point and the backward concentration at that point. One way to deal with this problem is to do a backward simulation for every hour separately; however, this is extremely expensive computationally. As a practical fix to this issue, at a particular backward travel hour (t) the plume travel time (t_r) from the release point (i.e. the monitor location) to a grid point ($\mathbf{x}$) is determined by releasing a second tracer (c_2) backwards from the monitor simultaneously with the main tracer ($c = c_1$) with the same tracer properties except that it decays exponentially with a decay rate of λ, so

$$c_2(\mathbf{x}, t) = c_1(\mathbf{x}, t) \exp(-\lambda t_r), \tag{12}$$

which gives

$$t_r = \frac{1}{\lambda} \ln\left[\frac{c_2(\mathbf{x}, t)}{c_1(\mathbf{x}, t)}\right]. \tag{13}$$

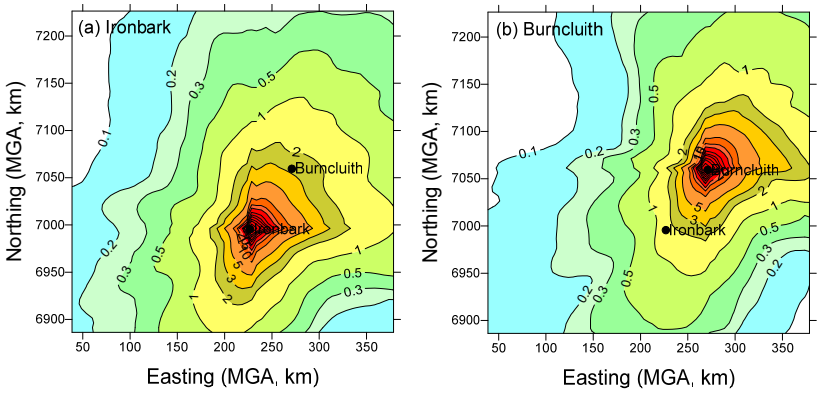
https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

405 The source-receptor value calculated at a grid point at a given backward travel hour (t_b) is then taken to correspond to $t_b + t_r$ (t_r rounded to an hour). The corresponding forward travel hour for a grid point is then equal to the total hours in a simulation period minus ($t_b + t_r$).

A modelled backward concentration field obtained for a unit emission rate is in essence the required source-receptor relationship which can be linearly scaled for any other emission rate. We use the hourly-averaged backward fields computed
410 at the lowest model level (i.e. 10 m AGL) for the innermost model domain in our inverse methodology.

As an example, Figure 7a presents the modelled backward concentration field (c/q, s m^{-3}) due to a unit point release (q = 1 g s^{-1}) averaged over all hourly fields over the simulation period for Ironbark. Essentially, the value at any point in Figure 7a is equivalent to the simulation-average forward model concentration value at this monitoring location if there were a source at that point with unit emission. Put differently, the backward concentration value at a given location represents the
415 probability (including both frequency and intensity) a source emission at that location adds to the concentration at the monitoring site. The backward field is mainly determined by flow the field across the domain and the separation between the receptor and the source.



420

Figure 7. Normalised modelled backward distribution of near-surface concentration (c/q, × 10^{-9} s m^{-3}), which is an average over the entire study period: (a) Ironbark, and (b) Burncluith.

It is apparent from Figure 7a that any sources located farther from the monitoring station would contribute less as plume
425 concentrations decrease with increasing distances, and vice versa. The directional distribution of the backward field is also a function of the distribution of regional winds which determine how often the receptor is downwind of a source (see wind

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

roses in Figure S3). The values in the south-east and north-west corners of the study domain are particularly low, so potential sources there would, on average, have low probability of being sampled at Ironbark.

The backward distribution for Burncluith (Figure 7b) is very similar, but since it is located north of Ironbark it would sample potential sources in the north-east better.

The two monitoring sites combined sample most part of the CSG sources in the domain (which was the prime objective of our monitoring).

5.3 Bayesian inversion setup

The source array of 69 × 69 used in the forward modelling above is too large a source number for the inverse methodology to explore all the source possibilities (i.e. hypotheses), even with use of the MCMC sampling, and, moreover, there is only a limited amount of information available from only two monitoring sites. Consequently, an array of 11 × 11 sources ($N_s = 121$, cell size ~ 31 × 31 km^2) is considered within the same model domain. The hourly source-receptor relationships calculated at 5 × 5 km^2 resolution for Ironbark and Burncluith were used. We assume that all source emissions are time invariant during a given simulation period – this allows the utilisation of all valid hourly concentrations available during that period in a single Bayesian calculation to determine the emission rates.

To reduce serial correlations in the sequence of MCMC samples drawn from the posterior using the Metropolis-Hastings algorithm, we only retained every 5th sample. The total number of useable samples was 21,000 for each source, of which the first 1,000 samples were discarded as "burn-in" samples. The selected samples were then used in the calculation of the source statistics.

6 Inversion using 'synthetic' concentration data

A 'synthetic' inverse run is first performed by using the simulated time series of concentration at Ironbark and Burncluith involving the bottom-up inventory (regridded to 11 × 11 sources, see Figure 8b) to investigate whether the inverse methodology is able to retrieve the bottom-up emissions and under what type of priors and their uncertainties. The given modelled (or synthetic) time series do not directly possess any background, instrumental, and model errors and, therefore, prove very useful in validating an inversion methodology. The results from an inversion of synthetic data can provide guidance about selection of the prior and its uncertainty specification for an inversion using real-world data.

6.1 Selection of the prior

Specifying the prior PDF $p(\mathbf{q})$ is an important step, even for the present synthetic case because we are still limited to the same degree of information available (i.e. the modelled concentration time series from only two sites), the number of

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

455 unknown sources to estimate, and the domain size as in the inversion case with the real concentration data considered subsequently.

An identical Gaussian $p(\mathbf{q})$ for each source with a mean methane emission rate q_p = 45.4 g s^{-1} (= 1.43 × 10^6 kg yr^{-1}) per source is specified. This mean value is essentially the total bottom-up emission from the domain divided by the number of sources (i.e. 121). It is observed that a large prescribed uncertainty in the prior biases the posterior PDF towards emission
460 rates that have high probability, and a small one biases it towards $p(\mathbf{q})$. As a balance between the two, it became evident that we needed to use a small value σ_p = 0.5% of the mean emission rate.

The uncertainty in the transport model and that in the synthetic concentrations need to be specified. Both are essentially zero for the synthetic case, but we use a nominal value of 5% of the mean for the former and 0.3 ppb for the latter. All hourly synthetic concentrations modelled for the full simulation period at the two sites (i.e. $N_m = 2 \times 13200$) were used in one
465 single Bayesian inversion to derive the emission rates.

6.2 Results for the synthetic case

The emission rates inferred by the inverse model are shown in Figure 8a, with the total emission being 162 × 10^6 kg yr^{-1}, a figure very similar to the bottom-up inventory total 173 × 10^6 kg yr^{-1}.

A comparison of Figure 8a with the bottom-up inventory (Figure 8b) indicates that the inverse model is able to simulate the
470 large emission rate in the region located just north of the Ironbark site. There is a strong inventory emission on the eastern domain boundary which the model does not replicate. A possible reason for this is that the two monitoring stations do not sample this source area sufficiently. Extra monitoring stations and/or separate, narrower priors for sources that make very small contributions to methane at the two sites would be needed to reduce the differences between Figure 8a and Figure 8b.

The synthetic case results suggest that the regional inverse model formulated is stable, feasible with MCMC, and credible as
475 evident from its getting the total emissions nearly right and replicating the largest emission area reasonably well with only a broad prior and two monitoring locations, but at the same time requiring a relatively small prior uncertainty. The synthetic case considered is an overly demanding case because the prior used is not very informative, compared to the real inversion cases considered in the next section in which the bottom-up inventory emissions allow the option of a better prior.

480

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

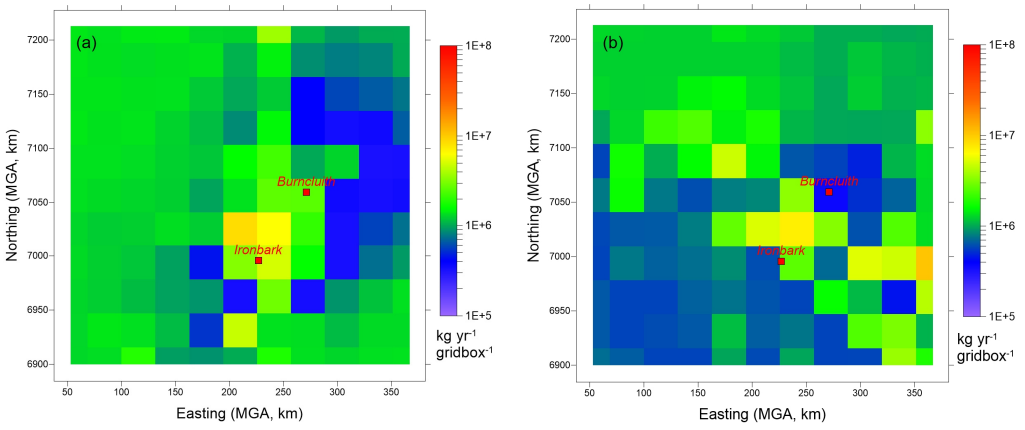


Figure 8. (a) Emission rates of CH_4 (kg yr^{-1} $gridcell^{-1}$) estimated by the synthetic inversion, and (b) bottom-up inventory emission rates.

7 Inversion using methane measurements

The same filtered methane observations as used in the forward transport modelling (so $N_m = 10581$) are used in one single Bayesian inverse run, with the uncertainty in the measurements σ = 3.5 ppb based on previous calculation and σ_m = 20% of the modelled concentration (Yee and Flesch, 2010; Luhar et al., 2014).

7.1 Priors and inferred emissions

Three cases involving different priors are considered.

7.1.1 Non informative prior (Case 1)

A case of virtually no prior, or uninformative prior, is first considered, in which the only constraint is that the emission rate for each source lies within the broad range 10–10,000 g s^{-1}, where the upper limit is nearly double the total domain-wide bottom-up inventory.

The inferred emissions (Figure 9a) between the two monitoring sites and around the centre of the region are qualitatively in accordance with the bottom-up inventory emissions (Figure 8b), but with larger magnitudes. In contrast, the inverse estimates in locations farther from these source areas are smaller than the inventory emissions. Remarkably, the total inferred emission with no prior is 162 × 10^6 kg yr^{-1} which compares well with the inventory total. The largest emission rate of about 1100 g s^{-1} in Figure 9a is about 10% of the upper bound of the specified prior range.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

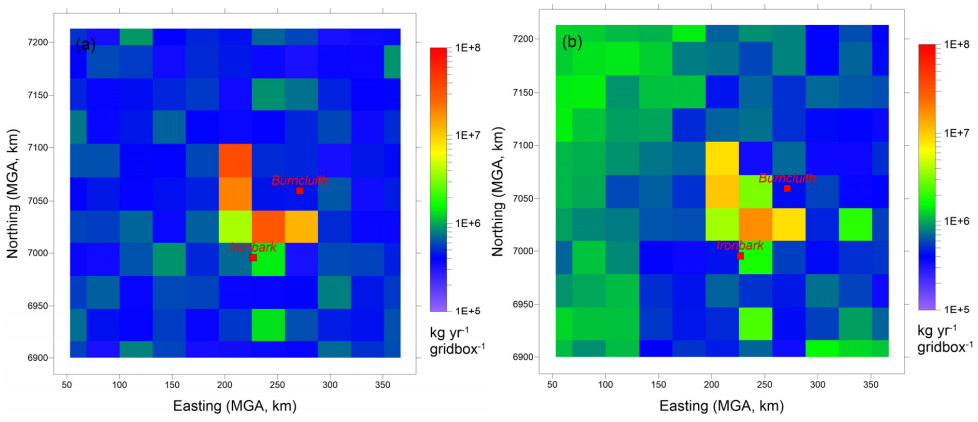


Figure 9. Emission rates of CH_4 (kg yr^{-1} $gridcell^{-1}$) estimated by the inversion: (a) with no prior (Case 1); and (b) with a Gaussian prior (Case 2).

7.1.2 Uniform Gaussian prior (Case 2)

Next, a more realistic prior PDF is specified with a Gaussian distribution having an identical mean of 45.4 g s^{-1} and σ_p = 10% of the mean, for each source. The mean is the same as that is used for the synthetic run.

The inferred emissions for this case shown in Figure 9b are qualitatively similar to Figure 9a; however, in the former the high emission sources are relatively less pronounced, with emissions from other source locations generally being larger. The total annual emission from the Surat Basin obtained using this inversion is 143 × 10^6 kg yr^{-1}.

7.1.3 Gaussian prior with the bottom-up inventory emissions (Case 3)

In this case, the inventory emissions shown in Figure 8b are taken as the mean values of a Gaussian prior for each source. As every source prior now has a more realistic specification of the mean value compared Case 2, the uncertainty in the prior is chosen to be smaller than that specified in Case 2.

The inferred emission rates in Figure 10a obtained for Case 3 with σ_p = 1% of the mean appear very similar to the inventory emission rates (Figure 8b). The fact that even the intense emission on the eastern boundary of the domain present in the inventory is mostly reproduced despite this area being not sampled relatively sufficiently by the two network locations means that the chosen prior with a very small uncertainty is somewhat too inflexible which forces the inversion towards a result that is very similar to the prior itself, thus in essence overriding the information inherent in the concentration observations.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

520

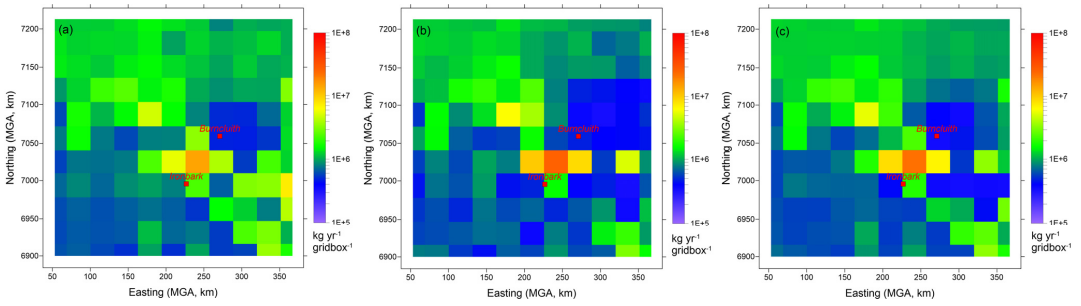


Figure 10. Emission rates of CH_4 (kg yr^{-1} $gridcell^{-1}$) estimated by the inversion with a Gaussian prior involving mean values equal to the bottom-up emissions (Figure 8b) and the standard deviation equal to (a) 1%, (b) 5% and (c) 3% of the mean values.

525

Figure 10b is obtained using the same inverse model setup as Figure 10a, except that the prior is relaxed somewhat by increasing σ_p to 5% of the mean. This leads to the source areas in the centre of the Surat Basin and those between Ironbark and Burncluith becoming more conspicuous. In contrast, the source areas near the eastern boundary of the domain nearly fade, with the concentration observations applying greater influence in areas where the source-receptor relationship, shown in Figure 7, is stronger. Clearly, the inversion is sensitive to σ_p, however, it is apparent that σ_p = 1% to 5% yields a reasonable trade-off between the benefit of the inversion approaching the prior in areas where the chances of the two monitoring stations detecting methane signal is small and simultaneously making sure that the selected prior would not unduly overrule the information supplied by the concentration measurements. Consequently, another inversion was performed for σ_p = 3%. The inferred emission from this run presented in Figure 10c in essence stand between the inferred emissions for σ_p = 1% and those for 5%. This Case 3 inversion with σ_p = 3% is our best estimate, which gives an annual total CH_4 emission of 166 × 10^6 kg yr^{-1}. The fine tuning of prior uncertainty also has some trial and error component driven by the need that the inferred emissions are able to describe the measured concentrations when used in a forward model simulation (see Section 7.2).

As noticed in the synthetic inversion case, and in Figure 10a and Figure 10b, a large prior uncertainty biases the inversion towards emission rates that have high probability, which may indicate that the number of monitoring stations is insufficient for the uncertainty in the prior to be relaxed.

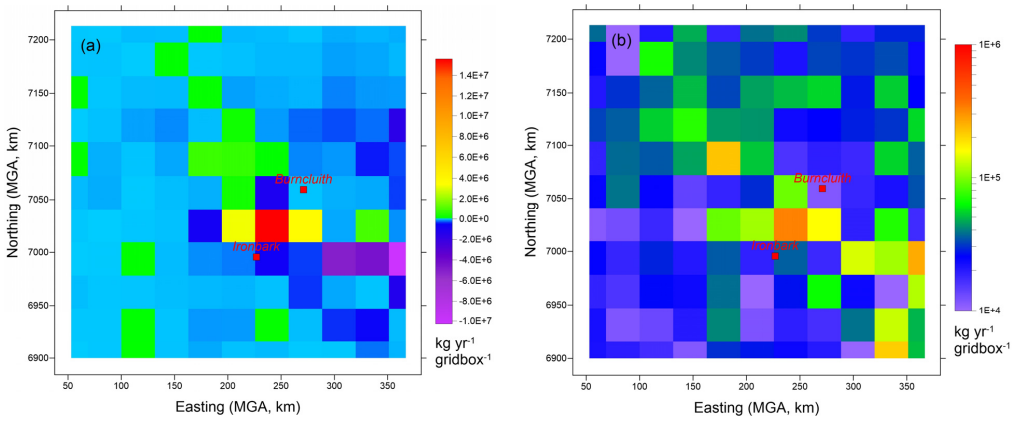
Figure 11 presents the difference between the inferred methane emissions given in Figure 10c and the bottom-up inventory emissions in Figure 8b. The largest difference is found for the grid box between Ironbark and Burncluith, with the inferred emissions (22.9 × 10^6 kg yr^{-1}) being larger by approximately a factor of three than the latter (7.3 × 10^6 kg yr^{-1}). The total

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

545 inventory emission for this source grid is controlled by CSG Processing (51%); feedlots, poultry and piggeries combined (32%); and CSG Production (6%) sectors.

The calculated standard deviation of the inferred emissions corresponding to the case shown in Figure 10c is presented in Figure 11b. These values are approximately 1-2 orders of magnitude lower than the mean emission rates in Figure 10c. In general, the standard deviations are larger for larger inferred emissions. One reason as to why these uncertainties in the
550 emission posterior are quite low, as discussed above, is the very small prescribed uncertainty that needs to be specified in the prior. Interestingly, the farthest grid point east of Ironbark has disproportionally large uncertainty compared to the mean emission rate in Figure 10c.



555

Figure 11. (a) Difference between the inferred methane emissions (Figure 10c) and the bottom-up inventory emissions (kg yr^{-1} gridbox^{-1}), and (b) standard deviation (uncertainty) of the inferred emissions (kg yr^{-1} gridbox^{-1}) presented in Figure 10c.

7.1.4 Sensitivity to background methane

Figure 4 shows that there is a slight difference in the estimated background CH_4 levels between the two monitoring
560 locations, with the Ironbark background methane larger by 1 ppb on average and the standard deviation of the background differences being 1.4 ppb, the latter is comparable to the background concentration uncertainty (= 3.5 ppb) considered in the inversion.

We conducted an inverse modelling sensitivity test with the same model setup as that for Figure 10c, except that instead of using the background times series that was averaged over the two sites we used the respective background timeseries for
565 these sites. The results were no different compared to Figure 10c, other than some insignificant changes in areas with low emissions. The annual emission total obtained was 164.8 × 10^6 kg yr^{-1}.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

7.2 Validation of the inverse emission estimates

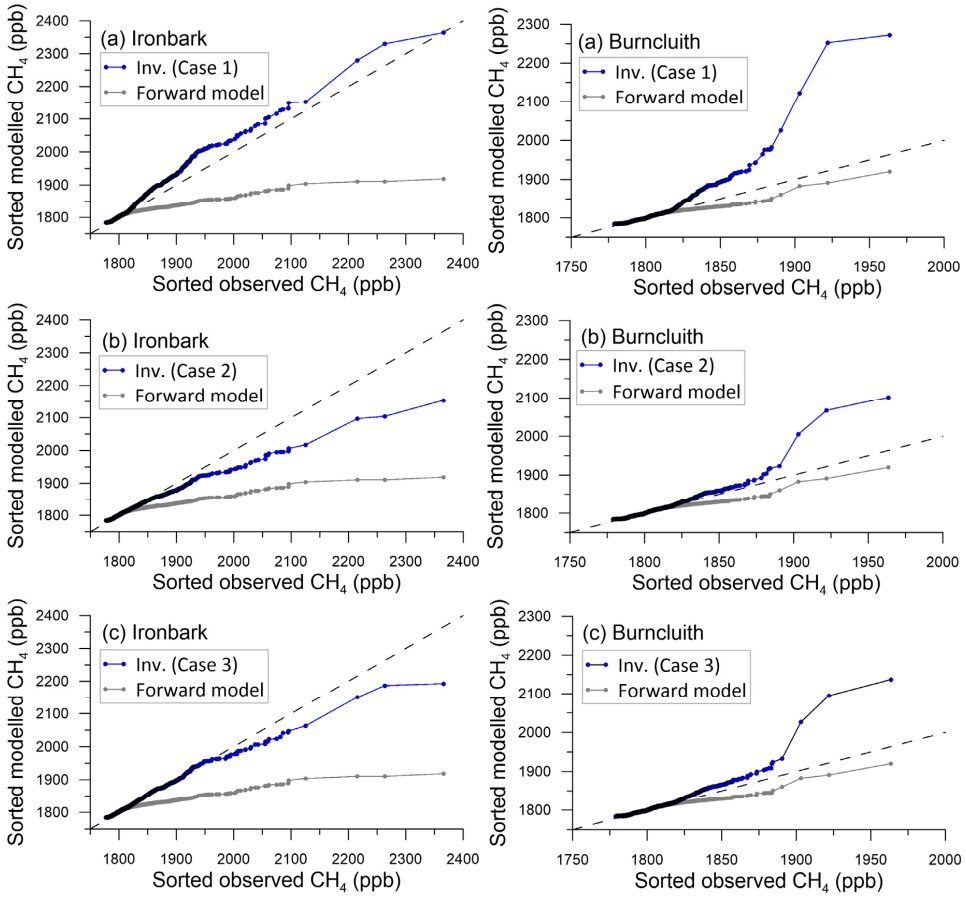
To examine to what extent the inferred emissions represent the methane concentration measurements compared to the bottom-up emissions, we conducted three separate forward transport model runs using the inferred emissions from the above inverse modelling Cases 1, 2, and 3 (i.e. Figure 9a, Figure 9b and Figure 10c, respectively).

Figure 12a presents q-q plots of the observed data against the modelled CH_4 computed using the Case 1 inferred emissions. There is an overestimation of methane at both monitoring stations for the higher-end concentrations, but the simulated CH_4 at Ironbark is much better reproduced than when using the bottom-up emissions (grey lines). For Burncluith, the overestimation is almost as large in magnitude as the underestimation obtained when the inventory emissions are used.

As demonstrated by Figure 12b, the Case 2 inferred emissions involving a proper, but still crude, prior lead to a significant improvement in the methane simulation, especially at Burncluith. As apparent from Figure 12c, further refinement in the prior in Case 3 yields emission estimates that further improve the simulation of methane, especially at Ironbark. With the exception of about 4 outlying data points at the higher-end of the concentration distribution, the Case 3 inversion corresponding Figure 12c leads to the best model reproduction of the measured CH_4 from the two monitoring sites. The underprediction seen when the inventory emissions are used is nearly eliminated.

Clearly, differences between the model and observations remain, and the possible causes for that include differences between the observed and modelled regional meteorology, only two monitoring sites within a relatively large study domain, the selected 11 × 11 source distribution representing the emissions in the domain being rather coarse, and potential temporal variation of source emissions.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020



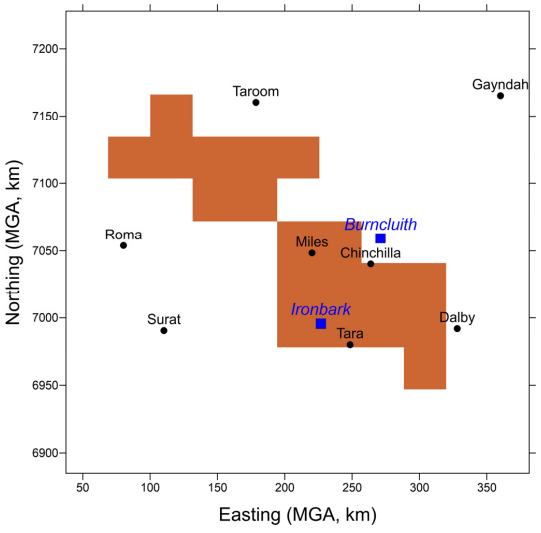
590 **Figure 12. Q-q plots showing the sorted hourly observed versus the sorted modelled CH_4 at the two monitoring stations. The modelled concentrations utilise emission estimates from (a) Case-1 inversion, (b) Case-2 inversion, and (c) Case-3 inversion (i.e. with 3% uncertainty in the prior). The forward model concentrations from Figure 6 predicted using the bottom-up emissions are shown as grey lines. Dashed line represents perfect agreement.**

7.3 Emissions from the CSG area

595 Given the focus on CSG activity related emissions in the Surat Basin, we compare the aggregate bottom-up and inferred emissions from the CSG areas, many of which are concentrated near and between the two monitoring stations. The subdomain that includes all the CSG sources in the study area is shown Figure 13, which is an area of about 18260 km^2 and covers 15% of the study domain. The CSG subdomain also contains emissions from other sectors (see Figure 3).

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions



600

Figure 13. A subdomain of the study area that corresponds to all the CSG source areas (shaded grid cells) in the bottom-up emission inventory. It consists of 19 gridcells (each with a source footprint of 31 × 31 km^2) and is considered for aggregating emission rates.

605

The total bottom-up inventory emissions from the CSG sub-domain is 47.7 × 10^6 kg yr^{-1} (cf. 173.2 × 10^6 kg yr^{-1} for the domain) whereas that obtained using the inversion (Case 3, Figure 10c) is 63.6 × 10^6 kg yr^{-1} (cf. 165.8 × 10^6 kg yr^{-1} for the domain) which is 33% larger than the former. The total bottom-up emission for this subdomain is dominated by CSG (34.7%, of which 30.6% is due to CSG Processing), followed by grazing cattle (29.9%), feedlots (23.5%) and coal mines
610 (7.7%), which together account for 95.8% of the emissions. Since the inverse methodology does not differentiate between source sectors, emissions from individual sectors cannot be inferred. Considering that the grazing cattle emissions are diffuse sources and thus not responsible for peaks in the measurements that dominate the inverse estimates, and since feedlots are scattered throughout the domain (Figure 3c) including the non-CSG areas from where there is no general inference of higher emissions, it is plausible that the increase in the inferred emissions would mainly correspond to CSG as the source sector.

615 A considerable portion of the CSG emissions is in the area between the two monitoring stations. The inferred emissions in this area are much greater than the corresponding bottom-up inventory emissions. This area also has significant coal mining emissions nearby (Figure 3d). It is possible that the methane emissions from a combination of these two source sectors are much larger than the inventory emissions.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020


Atmospheric Chemistry and Physics Discussions

Apart from the uncertainties associated with the bottom-up emissions, potential methane emissions from some sources, namely wetlands (the amount of which in the area is very limited; https://wetlandinfo.des.qld.gov.au), land clearing, termites, material handling and fuel usage related to mining activities, ground-water wells, and biomass burning are not part of the bottom-up emissions. In contrast, all CH_4 sources are implicitly represented in the inversions, apart from the biomass burning events which have been filtered using the CO filter. It is difficult to pinpoint which source sectors might be underrepresented in the bottom-up inventory without some kind of source discrimination, for instance, through the use of tracers such as the CH_4 isotopes.

7.4 Temporal variation of the inferred emission

In the previous inverse calculations, all filtered methane measurements obtained during July 2015–December 2016 were combined in one Bayesian calculation to derive a time invariant top-down emission distribution. Here we apply the inverse model with the Case 3 settings (as used for Figure 10c) to 3-monthly measurement blocks within the above period in order to examine potential temporal variation of the inferred emissions. Obviously, for a 3-monthly simulation the amount of concentration data supplied to the Bayesian inversion is much less than that for the full simulation. Figure 14a presents the 3-monthly variation of the inferred emissions as kg CH_4 yr^{-1} (bar plots), along with the (constant) bottom-up inventory emissions (red line) and the (constant) inferred emissions from Case 3 (blue line). The 3-monthly emission rates are within 165–180 kg yr^{-1} and are generally larger than when the full measurement duration is considered. This is because as the amount of information supplied to the inverse model reduces, the inferred emissions are not modulated to the same extent as that for the full period, and thus they tend to move closer to the bottom-up inventory which is used as a prior with a tight uncertainty. (Time-varying inventory emissions, if available, would act as a better prior, together with additional measurement sites).

Figure 14b is the same as Figure 14a but for the CSG subdomain. The 3-monthly inferred emissions lie between the bottom-up inventory value and the inferred value obtained when the measurements from the full period are used. Again, as in Figure 14a, 3-monthly inferred emissions push towards the inventory value as the information supplied to the inverse model reduces.

The temporal variations of inferred emissions in Figure 14a and Figure 14b are qualitatively similar. It is difficult to know whether these variations truly represent reality. Again, we conducted a forward TAPM run using the 3-monthly emissions from the above inversion, and the resulting q-q plots (red dots) are shown in Figure 15. The methane data at Burncluith are best described by these 3-monthly varying emissions compared to any other emission setup, but at Ironbark, these emissions underestimate the methane data (the inversion setup corresponding to Figure 12c best describes the Ironbark data). Additional measured parameters (e.g. tracers), as well as more monitoring stations and other additional data (e.g. about the prior), would be useful in further constraining the emissions.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

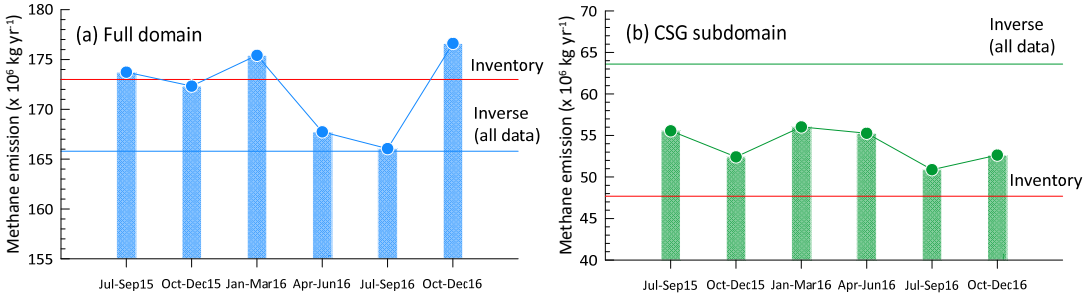


Figure 14. Temporal variation of the inferred emissions (bar plots) for (a) the full study domain and (b) the CSG subdomain. The constant bottom-up inventory emissions (red line) and the constant inverse/inferred emissions from Case 3 (Figure 10c) are also shown.

655

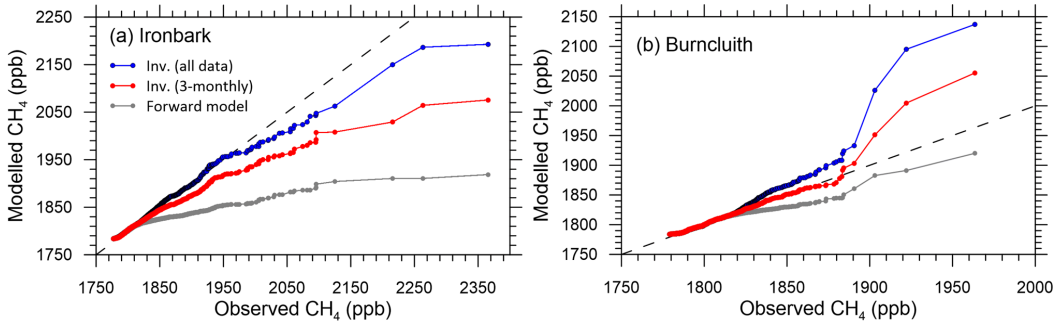


Figure 15. Q-q plots showing the sorted hourly observed versus the sorted modelled CH_4 at the two monitoring stations. The modelled values (blue dots) are predicted using the emissions from Case-3 inversion (with 3% uncertainty in the prior); the red dots are produced from 3-monthly inversions; and the forward model concentrations from Figure 6 predicted using the bottom-up
660 **emissions are shown as grey lines. Dashed line represents perfect agreement.**

Given the rapid CSG development in the Surat Basin, one may deduce that the 2016 CSG methane emissions were larger than the 2015 bottom-up emissions and, therefore, could potentially explain the top-down emissions in the CSG area being higher than the inventory emissions. Figure 16 shows that compared to July–December 2015, the total CSG produced was
665 higher by 32% during January–June 2016 and by 45% during July–December 2016 (which correlates with an increase in the number of CSG production wells in the area). However, notwithstanding the limited number of data available in deriving the top-down trend in Figure 14b and the 2015 bottom-up inventory used as the prior and assuming that the CSG area remains the same, this figure does not support any consistent increase in emissions from 2015 to 2016. As stated in Section 3, the main contributor to the total inventory CSG methane emissions was venting, followed by flaring. Figure 16 suggests that

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

670 although the amount of CSG produced steadily increased until 2017, the amount of venting/flaring does not show any such trend. As a matter of fact, overall, there is a decreasing trend in venting/flaring. Therefore, an increase in the amount of CSG produced may not necessarily mean that the methane emissions would have increased proportionally. Thus, the 33% higher top-down emission estimate from the CSG area compared to the inventory estimate cannot be explained in terms of the growth in the CSG production from 2015 to 2016. This also implies that the emissions from CSG may be more closely
675 related to practices in the industry than to the amount of CSG produced.

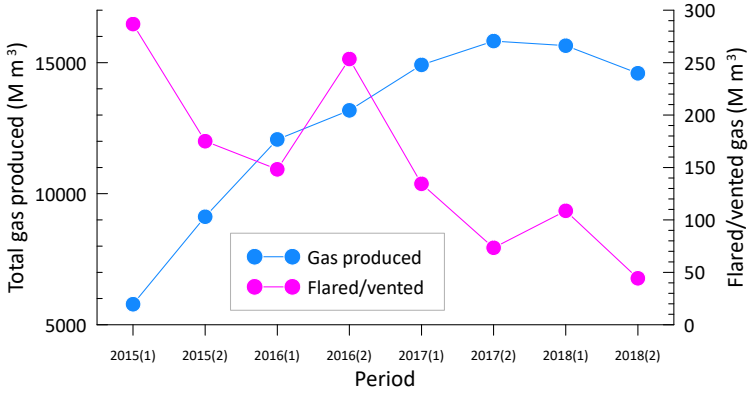


Figure 16. Six-monthly trends of the total CSG produced and the amount of flared/vented gas in the Surat Basin (data from https://www.data.qld.gov.au/dataset/petroleum-gas-production-and-reserve-statistics[1]).

680 **8 Conclusions**

This paper presents top-down and bottom-up quantification of gridded methane emissions from the CSG producing Surat Basin, an area of 350 × 350 km^2 in Queensland, Australia. The 2015 bottom-up emission inventory served as a very useful prior in our regional top-down methodology based on a Bayesian inference approach that utilised hourly-mean CH_4 concentrations monitored at the Ironbark and Burncluith stations for 1.5 years, hourly source-receptor relationship, and an
685 MCMC technique for posterior PDF sampling.

The largest contribution to the emissions in the bottom-up methane inventory is from grazing cattle (~50%), cattle feedlots (~25%), and CSG processing (~8%), with the aggregate emissions in the study area being approximately 173 × 10^6 kg CH_4

[1] This data file places the gas fields of Spring Gully and Peat within the Bowen Basin whereas in our bottom-inventory these are part of the Surat Basin. This is because of how the gas field zones and basin boundaries are defined. The gas fields included in our study are based on their geographic locations relative to the square study domain selected. Adding these two gas fields to the Surat Basin does not change the trends shown in Figure 16.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020


Atmospheric Chemistry and Physics Discussions

yr^{-1}. Although the forward transport modelling with the bottom-up emissions yielded a credible simulation of the suitably filtered observed methane concentrations, about 15% of the higher-end concentration observations were underestimated.

The top-down Bayesian inverse approach demonstrated that even when we do not specify an informative prior, the source signal inherent in the methane observations from only two sites constrains the total emission well. But, in contrast to the inventory emissions, the inferred emissions are more intensely located in the centre of the study region and less in regions that are farther. The importance of specifying a suitable prior in the Bayesian inference was apparent, with the bottom-up inventory proving very valuable for that purpose. Particularly, a Gaussian prior having mean values taken the same as the bottom-up emissions with an uncertainty equal to 3% of the mean yielded the best emission distribution, as evident from its performance in faithfully reproducing the measured methane concentration timeseries. This inverse setup yielded a domain-wide emission of 166×10^6 kg CH_4 yr^{-1} which is very slightly less than the one obtained from the bottom-up inventory. However, within a subdomain covering all the CSG source areas, the inferred emissions are 33% larger than those deduced from the bottom-up inventory. The dominant localised inventory emissions in this area are from CSG, followed by feedlots. Since feedlots are scattered throughout the domain including the non-CSG areas from where there is no inference of higher emissions, it is plausible that the increase in the inferred emissions would mainly correspond to CSG as the source sector.

The source-receptor relationship showed that having only two monitoring stations is inadequate for sampling distant source areas within the large study domain, especially areas in the south-east and north-west corners (the network design for the two monitoring stations mainly focused on the central CSG regions). Lengthening the measurement period to sample these areas better would not have helped because the wind climatology of the area is not likely to change considerably. When source areas are not sampled well, one may impose stricter priors that are more credible than the inferred emissions, or alternatively increase the number of stations. The former strategy is probably reflected in our use of a small uncertainty in the prior (i.e. 3% of the mean) for the best inversion case. A smaller prior uncertainty pushes the inversion more towards the prior itself with distant source areas not sampled sufficiently by the network sites looking like the prior distribution. A larger prior uncertainty results in the inversion moving towards higher emissions for sources that are close to the monitoring stations.

The inverse methodology could not distinguish between different source categories, mainly because the concentration of methane alone was monitored and not tracers specific to methane source types. To do source discrimination and attribution, monitoring of tracer species such as methane isotopes ($^{13}CH_4$, CH_3D and $^{14}CH_4$), or other hydrocarbons in cases where they are associated with the source gas, would prove useful when suitable sampling systems or instrumentation for field deployment become available.

Data availability

The data and model output included in this paper can be made available by contacting the corresponding author (Ashok Luhar: ashok.luhar@csiro.au).

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020


Atmospheric Chemistry and Physics Discussions

Author contributions

AKL performed the model development and application, analysed model output and data, and wrote the paper with contributions and comments from the co-authors. DME conducted the field study design, in-situ monitoring and data analysis, and GISERA project management. ZML conducted the in-situ monitoring, data collection, and data analysis. JN contributed to data processing. DS assisted with the monitoring sites, instrumentation and data collection. LS developed the bottom-up inventory. CO provided general information on methane sources.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This work was partially supported by CSIRO's Gas Industry Social and Environmental Research Alliance (GISERA) (https://gisera.csiro.au) and Active Integrated Matter (AIM) Future Science Platform (FSP) (https://research.csiro.au/aim). Mark Kitchen and Steve Zegelin provided valuable instrumental and technical support. The authors thank Peter Rayner, Martin Cope and Dimitri Lafleur for their helpful comments on this work, and Natalie Shaw for her assistance with the preparation of the bottom-up inventory for the Surat Basin. Damian Barrett advised on this work, and Stuart Day furnished insights into local source monitoring.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

735 **References**

Alvarez, R. A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., et al.: Assessment of methane emissions from the U.S. oil and gas supply chain, Science, 361(6398), 186–188, https://doi.org/10.1126/science.aar7204, 2018.

Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Pétron, G., Jordaan, S.M., et al.: Methane leaks from North
740 American natural gas systems, Science, 343(6172), 733–735, https://doi.org/10.1126/science.1247045, 2014.

Cheng, M.-D., Hopke, P. K., and Zeng, Y.: A receptor-oriented methodology for determining source regions of particulate sulfate at Dorset, Ontario, Journal of Geophysical Research, 98(D9), 16,839–16,849, https://doi.org/10.1029/92JD02622, 1993.

Cui, Y. Y., Brioude, J., Angevine, W.M., Peischl, J., McKeen, S. A., Kim, S.-W., et al.: Top-down estimate of methane
745 emissions in California using a mesoscale inverse modeling technique: The San Joaquin Valley, Journal of Geophysical Research, 122(6), 3686–3699, https://doi.org/10.1002/2016JD026398, 2017.

Day, S., Dell'Amico, M., Etheridge, D., Ong, C., Rodger, A., Sherman, B., and Barrett, D.: Characterisation of regional fluxes of methane in the Surat Basin, Queensland. Phase 1: A review and analysis of literature on methane detection and flux determination. CSIRO Australia report, 2013.

750 Day, S., Ong, C., Rodger, A., Etheridge, D., Hibberd, M., van Gorsel, E., et al.: Characterisation of regional fluxes of methane in the Surat Basin, Queensland: Phase 2: A pilot study of methodology to detect and quantify methane sources, CSIRO Australia report EP15369, 2015.

DNRM: Queensland's Petroleum and Coal Seam Gas 2015–16, Department of Natural Resources and Mines, Queensland Government, Australia, 8 pp., http://www.australiaminerals.gov.au/__data/assets/pdf_file/0003/47622/Queenslands-
755 petroleum-and-coal-seam-gas-2017.pdf, 2017.

Etheridge, D., Loh, Z., Schroder, I., Berko, H., Kuske, T., Allison, C., et al.: Metadata report: Arcturus atmospheric greenhouse gas monitoring. Record 2014/37, Geoscience Australia, Canberra, 2014.

Etheridge, D. M., Day, S., Hibberd, M. F., Luhar, A., Spencer, D. A., Loh, Z. M., et al.: Characterisation of regional fluxes of methane in the Surat Basin, Queensland: The continuous monitoring results – installation, commissioning and operation
760 of two field stations and preliminary data, Milestone 3.1 GISERA Greenhouse Gas Research – Phase 3, CSIRO Australia report, 2016.

Feitz, A., Schroder, I., Phillips, F., Coates, T., Neghandhi, K., Day, S., et al.: 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, https://doi.org/10.1016/j.ijggc.2017.11.018, 2018.

765 Henne, S., Brunner, D., Oney, B., Leuenberger, M., Eugster, W., Bamberger, I., et al.: Validation of the Swiss methane emission inventory by atmospheric observations and inverse modelling, Atmospheric Chemistry and Physics, 16(6), 3683–3710, https://doi.org/10.5194/acp-16-3683-2016, 2016.

https://doi.org/10.5194/acp-2020-337
Preprint. Discussion started: 15 May 2020

Atmospheric Chemistry and Physics Discussions

Hmiel, B., Petrenko, V. V., Dyonisius, M. N. et al.: Preindustrial $^{14}CH_4$ indicates greater anthropogenic fossil CH_4 emissions, Nature 578, 409–412, https://doi.org/10.1038/s41586-020-1991-8, 2020.

Hourdin, F. and Talagrand, O.: Eulerian backtracking of atmospheric tracers. I: Adjoint derivation and parametrization of subgrid-scale transport, Quarterly Journal of the Royal Meteorological Society, 132(615), 567–583, https://doi.org/10.1256/qj.03.198.A, 2006.

Hurley, P. J., Physick, W. L., and Luhar, A. K.: TAPM: a practical approach to prognostic meteorological and air pollution modelling, Environmental Modelling and Software, 20(6), 737–752, https://doi.org/10.1016/j.envsoft.2004.04.006, 2005.

Hurley, P. J. and Luhar, A. K.: Modelling the meteorology at the Cabauw tower for 2005, Boundary-Layer Meteorology, 132(1), 43–57, https://doi.org/10.1007/s10546-009-9384-4, 2009.

IPCC: Climate Change: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., 2013.

Jaynes, E. T.: Probability theory: The logic of science, Cambridge University Press, Cambridge, UK, 753 pp., 2003.

Jeong, S., Zhao, C. Andrews, A. E., Bianco, L., Wilczak, J. M., & Fischer, M. L.: Seasonal variation of CH_4 emissions from central California, Journal of Geophysical Research, 117, D11306, https://doi.org/10.1029/2011JD016896, 2012.

Luhar, A. K. and Hurley, P.: Evaluation of TAPM, a prognostic meteorological and air pollution model, using urban and rural point source data, Atmospheric Environment, 37(20), 2795–2810, https://doi.org/10.1016/S1352-2310(03)00204-8, 2003.

Luhar, A. K., Mitchell, R. M., Meyer, C. P., Qin, Y., Campbell, S., Gras, J. L., and Parry, D.: Biomass burning emissions over northern Australia constrained by aerosol measurements: II–Model validation, and impacts on air quality and radiative forcing, Atmospheric Environment, 42(7), 1647–1664, https://doi.org/10.1016/j.atmosenv.2007.12.040, 2008.

Luhar, A. K., & Hurley, P. J.: Application of a coupled prognostic model to turbulence and dispersion in light-wind stable conditions, with an analytical correction to vertically resolve concentrations near the surface, Atmospheric Environment, 51, 56–66, https://doi.org/10.1016/j.atmosenv.2012.01.046, 2012.

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(18), 10959–10979, https://doi.org/10.1002/2014JD021880, 2014.

Marchuk, G. I.: Adjoint equations and analysis of complex systems. Dordrecht, Netherlands: Springer Science. pp. 466, 1995.

Matthaios, V. N., Triantafyllou, A. G., and Koutrakis, P.: PM_{10} episodes in Greece: Local sources versus long-range transport—observations and model simulations. Journal of the Air and Waste Management Association, 67(1), 105–126, https://doi.org/10.1080/10962247.2016.1231146, 2017.

Miller, S. M., Michalak, A. M., and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, Geoscientific Model Development, 7(1), 303–315, https://doi.org/10.5194/gmd-7-303-2014, 2014.

Pudykiewicz, J. A.: Application of adjoint tracer transport equations for evaluating source parameters, Atmospheric Environment, 32(17), 3039–3050, https://doi.org/10.1016/S1352-2310(97)00480-9, 1998.

Rao, K. S.: Source estimation methods for atmospheric dispersion, Atmospheric Environment, 41(33), 6964–6973, https://doi.org/10.1016/j.atmosenv.2007.04.064, 2007.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., et al.: The global methane budget 2000–2012, Earth System Science Data, 8(2), 697–751, http://doi.org/10.5194/essd-8-697-2016, 2016.

Schneising, O., Burrows, J. P., Dickerson, R. R., Buchwitz, M., Reuter, M., and Bovensmann, H.: Remote sensing of fugitive methane emissions from oil and gas production in North American tight geologic formations, Earth's Future, *2*, 548–558, https://doi.org/10.1002/2014EF000265, 2014.

Singh, S. K., Sharan, M., and Issartel, J.-P.: Inverse modelling methods for identifying unknown releases in emergency scenarios: an overview, Int. J. Environment and Pollution, 57(1/2), 68–91, https://doi.org/10.1504/IJEP.2015.072121, 2015.

Tarantola, A.: Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics, Philadelphia, 342 p., 2005.

Towler, B., Firouzi, M., Underschultz, J., Rifkin, W., Garnett, A., Schultz, H., et al.: An overview of the coal seam gas developments in Queensland, Journal of Natural Gas Science and Engineering, 31, 249–271, https://doi.org/10.1016/j.jngse.2016.02.040, 2016.

Venkatram, A., Brode, R., Cimorelli, A., Lee, R., Paine, R., Perry, S., et al.: A complex terrain dispersion model for regulatory applications, Atmospheric Environment, 35(24), 4211–4221, https://doi.org/10.1016/S1352-2310(01)00186-8, 2001.

WMO: WMO Greenhouse Gas Bulletin, No. 14, 22 November 2018, https://library.wmo.int/doc_num.php?explnum_id=5455, ISSN 2078-0796, 2018.

Yee, E., Lien, F.-S., Keats, A., and D'Amours, R.: Bayesian inversion of concentration data: Source reconstruction in the adjoint representation of atmospheric diffusion, Journal of Wind Engineering and Industrial Aerodynamics, 96(10-11), 1805–1816, https://doi.org/10.1016/j.jweia.2008.02.024, 2008.

Yee, E. and Flesch, T. K.: Inference of emission rates from multiple sources using Bayesian probability theory, Journal of Environmental Monitoring, 12, 622–634, https://doi.org/10.1039/B916954G, 2010.

Yee, E.: Inverse dispersion for an unknown number of sources: Model selection and uncertainty analysis, ISRN Applied Mathematics, Article ID 465320, 20 pp., 2012.