Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2018-1187-RC1, 2019

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Estimates of sub-national methane emissions from inversion modelling" *by* Sarah Connors et al.

Anonymous Referee #1

Received and published: 9 January 2019

1 Overview:

Review of "*Estimates of sub-national methane emissions from inversion modelling*" by Connors *et al.*

Connors *et al.* present an analysis from a year of methane measurements at 4 sites in East Anglia. They describe the 4 new instruments that are mounted on churches or other tall towers. These instruments are then used to constrain methane emissions in between June 2013 and May 2014. The description of the network is generally good and the figures are all reasonably well made. However the inversion portion of the manuscript needs quite a bit of work. I have some major concerns with the seemingly unsubstantiated choices, poor description, and potential overfitting. I think there is

Printer-friendly version

Discussion paper

interesting data here that should eventually be published, but there are some major issues that need to be dealt with first.

2 Major Comments

The section on inversion modelling (Section 2.2) could use a major re-write to describe what was actually done and justify choices made.

2.1 Over-fitting?

A **major** drawback with a least squares cost function is over-fitting. This is precisely why most inversions use a regularization or a prior. I don't see any discussion of over-fitting. How do the authors combat over-fitting?

I strongly suspect this is why they find a 'dipole effect' (see Minor Comment #1).

2.2 Simulated Annealing

2.2.1 Use of Least-Squares and Simulated Annealing?

A least squares cost function (also known as a maximum likelihood estimate), as the authors use, has a closed form solution for the optimal solution. Using the author's notation, the optimal solution ($\hat{\mathbf{x}}$) would be:

$$\hat{\mathbf{x}} = \left(\mathbf{K}^T \Sigma_\epsilon^{-1} \mathbf{K}\right)^{-1} \mathbf{K}^T \mathbf{y}'. \quad (1)$$

where $\mathbf{K}$ is the dilution matrix, Σ_ϵ is the error covariance matrix, $\mathbf{y}'$ is the vector of observations after removing the background concentrations: $\mathbf{y}' := \mathbf{y} - \mathbf{b}$.

There are four cases I could envision using something like Simulated Annealing:

1. if the system is non-linear (e.g., if $\mathbf{K}$ is a function of x)
2. to regularize the solution (e.g., by coarsening $\mathbf{K}$ as part of the inversion)
3. there are additional constraints being applied (e.g., non-negativity)
4. for computational expediency

In the case of additional constraints (such as a non-negativity constraint), it seems that bounded optimization would be preferable to a stochastic method such as simulated annealing. Gradient-based methods like L-BFGS-B ("https://en.wikipedia.org/wiki/Limited-memory_BFGS#L-BFGS-B") tend to be much faster for convex optimization problems such as this one. A stochastic method like Simulated Annealing would probably be better for a non-convex optimization problem.

2.2.2 Error statistics from simulated annealing

Regarding the use of Simulated Annealing, it's unclear to me why the authors chose to use a technique like simulated annealing here. Simulated annealing is an optimization technique that is quite efficient, so it works rather well in high-dimensional problems. However, the samples obtained from Simulated Annealing are inconsistent with the true posterior statistics (the uncertainties will be smaller that the true uncertainties). So reporting error statistics from Simulated Annealing strikes me as dubious at best.

Using something like an adaptive MCMC or a reversible jump MCMC (rjMCMC) as some of the co-authors here have previously done seems far superior to simulated annealing.

Printer-friendly version

Discussion paper

Related to this, it's entirely unclear why Manning *et al.* (2003) would be the citation for simulated annealing on Page 8, Line 26. This is not a technique or method that Manning developed. For example, the Machine Learning textbook from Christopher Bishop (Bishop, 2007) would be a much more appropriate citation for someone interested in how Simulated Annealing works.

2.3 Section 2.2.3: A priori emission estimates

It's not clear why this section is necessary at all. As mentioned above, if the system is linear (the dilution matrix is predetermined and does not change) and the authors choose a least-squares cost function then the solution can be directly computed. There would be no need for a "random, non-negative emission field". How is this being used? Is this the initial starting point for the simulated annealing? If so, then it should be referred to as such and probably not be discussed in a 2-sentence section on prior emissions.

2.4 Regarding the resolution of the solution grid

There has been an abundance of work looking at how to define a multi-scale state vector. The textbook from Rodgers (2000) talks about this. Bocquet *et al.* (2011; QJRMS) is entirely devoted to this topic. Other work like Turner & Jacob, (2015; ACP), Lunt *et al.*, (2016; GMD), Henne *et al.* (2016; ACP), and Bousserez & Henze, (2017; QJRMS) also talk at length about how to construct this multiscale formalism.

Briefly, by coarsening (or restricting) the grid you are applying a hard constraint on the inversion. Basically, sub-elements are no longer allowed to vary independently. This is a form of regularization. In the most extreme case you could coarsen to a single state vector element. The degree of coarsening can change the problem from an under-determined problem to an over-determined problem. This ties back to an earlier

Printer-friendly version

Discussion paper

question, how do you deal with potential over-fitting? Was there any cross-validation done?

At the bare minimum, Bocquet *et al.* (2011) should be cited.

2.5 Computing the Baseline

This section needs more work. Computing the baseline is a non-trivial task for any regional inversion and many studies are devoted entirely to estimating the baseline. At the bare minimum, the authors should provide some justification for their choice of the 18^{th} percentile. Other studies such as Henne *et al.* (2016; ACP) provided an extensive analysis on the choice of background including precomputing it (as Connors *et al.* have done) vs jointly estimating it as part of the inversion.

2.6 Evaluation of meteorology?

I didn't see any mention of evaluating the meteorology. This is a crucial step in atmospheric inversions using real measurements that seems to be missing.

2.7 Figure 3 seems odd

The dilution matrix as Connors *et al.* refer to it (also commonly known as the footprint matrix, transport operator, Jacobian, etc.) looks pretty Gaussian. Is there no dominant wind pattern at Haddenham? I would usually expect some dominant wind pattern (i.e., more sensitivity to the upwind region). A wind rose showing that, indeed, the winds are roughly uniformly represented from all sectors here would be useful. Otherwise, turning this into a 3 panel figure (current figure as a large column on the left and two subpanels on the right column) with two illustrative examples of dilution matrices from

two days. I would expect the illustrative days to show strong sensitivity upwind of the site for that day.

Basically, I'm curious if this was computed correctly.

2.8 Agreement with NAEI:

The authors mention good agreement with NAEI (to within ∼5%). However, Figure 6 looks strikingly different. Is this because they've further coarsened the emissions before this comparison? I don't see how they are getting a 5

3 Minor Comments

3.1 Dipoles and overfitting in the solution

The authors discuss a 'dipole effect' (e.g., Page 11, Line 21) in the inversion results. These are common in solutions with overfitting. The inversion is attempting to fit a high value at a measurement site, so it inflates the emissions to a very large value at that one location and then compensates by reducing the emissions in a neighboring grid cell where the observations have weak constraints. Basically, this is what happens:

- Location A: concentration too low -> increase emissions in just this location
- Location B: no constraint on concentration, domain wide emissions too high because of Location A -> reduce emissions

This is combatted in most inversion systems by having a prior or regularization that includes some off-diagonal relationships (e.g., emissions from Location A and Location B should be weakly correlated).

3.2 Suggestion for title

It would be nice if the authors specified *where* the work was done in the title. I'd suggest adding "in East Anglia" or "in the United Kindgom" (or something to that effect). Maybe something like: *"Estimates of sub-national methane emissions from inversion modelling in East Anglia"*

3.3 Proof of concept

On Page 3, Line 1 the authors motivate the work as a "Proof of Concept" that inversion schemes can work at sub-national scales. Although this has been shown numerous times in the past. Examples include Scot Miller's 2013 PNAS paper for methane in the US, Stephan Henne's 2016 ACP paper for methane in Switzerland, work from Kathryn McKain on methane emissions in Boston, work from Ken Davis' group on urban inversion modelling in Indianapolis for CO_2, and work from Thomas Lauvaux on CO_2 at urban scales. So I don't find a "Proof of Concept" to be a particularly compelling motivation.

The work is definitely interesting, but I don't think this should be a major motivation for the reader.

3.4 Table 1:

The instrument acronyms are rather confusing in Table 1. I would remove "UCAM" and "UEA" from the table and instead add a different column that lists the sampling rate. Alternatively, the authors could have footnotes under that table that explain the acronyms. As it stands, the reader needs to scan the text to try and figure out what the acronyms mean.

3.5 “Pseudo-observations”:

I would avoid using the term “pseudo-observations” because it sounds like the authors are doing a synthetic-data study (i.e., an OSSE). “Simulated”, “modelled”, or “predicted” concentrations would be preferable.

3.6 Equation #1:

Do not present an equation in this form. Use this equation to introduce your nomenclature for later. Something like this would be preferable:

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{b} \tag{2}$$

where $\mathbf{y}$ is an $n \times 1$ vector of concentrations (units: g m^{-3}), $\mathbf{K}$ is an $n \times m$ dilution matrix (units: s m^{-1}), $\mathbf{x}$ is an $m \times 1$ vector of gridded emissions (units: g s^{-1} m^{-2}), and $\mathbf{b}$ is an $n \times 1$ vector of concentrations upwind of the modelling domain (units: g m^{-3}).

3.7 Table 3:

“Simulated annealing” is not a cost function type. The cost function is a least-squares or maximum likelihood estimate.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-1187, 2018.

Printer-friendly version

Discussion paper