Referee comment

anonymous author

1 General comments

The study *Recovery of sparse urban greenhouse gas emissions* addresses the problem of localizing and quantifying greenhouse gas emissions in urban settings. The authors compare different approaches to solve the inverse problem that arises when an unknown emission field is estimated by local (column) observations of trace gas enhancements. The reconstruction quality of the methods is analyzed in several synthetic and idealized settings. The theoretical study can be considered as a preparational work for current projects to monitor trace gas emissions from cities.

The content of the study is presented in a structured form and the clear language is easy to follow. In some places the description of the experimental setup is incomplete.

The study uses a simplified atmospheric transport model in an idealized setting to create a test environment for inverse methods. Modern inverse modeling approaches, sparse reconstruction and sparse reconstruction in the wavelet domain, are compared to the standard method applied in most environmental applications, Bayesian inversion with Gaussian prior. Even though the introduced methods are well studied in the inverse problem and compressed sensing communities, these methods have only been applied a few times to environmental problems. Despite some shortcomings in the evaluation, the presentation of their potential is welcome.

Code and figures are uploaded as supplementary material. Some instructions and comments are included to run the code. However, some scripts require input not shared by the authors.

In total, an interesting study that includes new modeling approaches in this setup. The experimental setup and the evaluation leave room for improvement. I recomment the publication after considering the following comments.

2 Specific comments

Inverse problem

The authors make a well structured introduction of the inverse problem and their solution methods. Though many readers will be familiar with some inverse problem, the methods used are rather unknown in the atmospheric research community. The introduction is therefore instructive. However, underdetermination is only one characteristic difficulty of inverse problems. Another typical property not discussed in detail in the manuscript is the sensitivity of the estimate to (noisy) data. I recommend to mention this property in the abstract (line 3) and include a small discussion in Sect. 2.1.

It could also be instructive to show a row of the sensing matrix as a footprint (as an addition to Fig. 2), though I assume that the currently shown matrix is only for illustration purposes and to small to hold a meaningful footprint.

Sparse reconstruction

Bayesian inversion with Gaussian prior is commonly applied to inverse problems in environmental studies. It is therefore the correct method to compare against. The method introduced by the authors, sparse reconstruction, has only been applied in atmospheric studies a few times. References are provided. A relation between least squares, Bayesian inversion with Gaussian prior and sparse reconstruction is nicely presented, but there are a few points to address:

The Bayesian inversion approach calculates a posterior distribution. Often, this posterior distribution is assessed by its maximum a posteriori solution as a best estimate and the (co-)variances as uncertainties. In the Gaussian case the interpretation of the posterior distribution is completely described by these parameters. Variational methods, Eq. (8), mainly focus on the best estimate by assuming certain properties via the penalty function, e.g. smoothness or sparsity. The penalty functions $R(\cdot)$ can be chosen more freely. Particularly, C_2 can be more general than a correlation matrix (cp. line 112). Some choices may be difficult to formulate as a Bayesian prior or the analysis of the posterior may become too difficult. However, the Bayesian equivalent to sparse reconstruction uses a Laplacian prior. This being said, I cannot completely agree with some formulations in the text, e.g. 'sparse reconstruction [...] does not require a prior emission field' (line 5). In this sense, neither does l_2-regularization. My feeling is that both inverse modeling interpretations are mixed and I recommend to review the text for imprecise formulations.

As a side note: I think the l_0-norm and the l_p-norm should also be defined for readers new to the topic (lines 133 and 167). Actually, I found the l_0-norm in Table 1, but the reader is not guided there in the first reading.

Implementation of the methods

The authors reduce the dimensionality of the optimization problems by removing insensitive grid cells and setting their values to zero. While a physical interpretation of insensitive grid cells is given, an explanation why this preprocessing step is justified is missing (around line 247). In the standard domain the emissions on insensitive grid cells should be estimated as zero for both approaches. However, in the wavelet domain, I suppose that these emissions could (slightly) deviate as they may be parameterized by some wavelets with nonzero coefficients. Are the wavelets created prior to the removal or on the reduced system? In the latter case, the removal may influence the wavelet setup.

Experimental setup
During the first reading, I was not aware that most scenarios, e.g. Sect. 4.2, are evaluated without any noise on the measurements. The setup in each experiment should be clearly pointed out. Also, I am not sure which wind fields are used in the scenarios. Are they artificially generated? Is the wind uniform? I found some incomplete information in the Appendix.

In general, the setup in all experiments is idealized. It seems that most scenarios are evaluated with noiseless data. In the noisy case only measurement noise is included. Transport errors, uncertainties from the background and temporal variability of emissions are not considered. Also, the wind variation may be less ideal, assuming they are artifical, in an applied scenario. Aren't the results overly optimistic for any application?

In a theoretical study some simplification may be reasonable, but to me the methods lack a realistic test. For example, experiment 4.2.3 aims at discovering emissions not included in a good prior, e.g. an inventory. Small deviations from a good prior emission field should only produce small deviations in the observed enhancements. Are these detectable in the noisy case (cp. Sect. 4.2.5), particularly when including other sources of uncertainty (transport, background, etc.)?

Also, it seems that many results from the noiseless case do not transfer to the noisy evaluation, e.g. the wavelet approach performs less convincing.

In the noisy case, relative noise related to the measurement uncertainty and observed enhancements for the city of Indianapolis (not further evaluated) are considered. As measurement uncertainties are identical and expected trace gas enhancements can be calculated for each city, I wonder why a fixed SNR is used for all cities.

Analysis
While the discussion of compressed sensing is interesting in general, I am not convinced that this debate is that helpful. The options to design the sensing matrix in an optimal way for sparse methods are rather limited. As observations from the same location under similar wind conditions tend to have similar footprints, variety of measurement information can only come from changing wind conditions and different measurement locations. The authors find that even with full wind direction coverage the reconstruction criteria cannot be met, even in the noiseless case (also cp. App. Lines 443).

What is meant by Line 444: Some parameterization of the emissions as in the wavelet approach may create different dependencies, but the footprints will always create some structure to the sensing matrix. Do you have an example?

Then, I doubt that the results of Fig. 5 should be interpreted with respect to the coherence, since the coherence measures the maximum similarity between columns. An identity matrix extended by the copy of one column has a terrible coherence of 1, but is the perfect sensing matrix for all but the two underdetermined parameters. Therefore, the coherence may be a measure for the maximum error - I do not know - but the results in Fig. 5 are probably better explained by a sensitivity analysis.

An explanation for the increase of the relative error of regularized least squares is missing. I also wonder why the results for sparse reconstruction in the wavelet domain is not included. Does it show similar behaviour to SR?

Results

In Fig. 4 it is a bit surprising to see that regularized least squares produce a somewhat sparse emission field, particularly in the outskirts. To me it looks as oscillations (negative emissions are not shown) originating from a regularization parameter λ (Eq. 8) that is too small. This would also explain the large number of negative emissions. Many classical parameter choice rules reduce the model-data mismatch to a factor greater than 1 times noise level, before instabilities are introduced to the estimate. The same issue may apply to the sparse reconstruction approaches. Though the sparsity is increased with sparse reconstruction, the solution is not really sparse (only 10% or 20%) as discussed in Table 2. As it seems that this scenario considers noiseless data, forcing equality between model and data may be too strict or tolerances in the optimization routine could be fine tuned. In the noisy case, the authors also seem to observe overfitting (line 361).

It would be interesting to see the l_2-errors for the reconstructions in Fig. 4.

In general, it is also a bit surprising to see relative errors much larger than 1, e.g. in Fig. 8, for a stable inversion. What is the interpretation?

In Fig. 8, scenario (b) uses the highest emission resolution with varying number of observations. Shouldn't the results from scenario (a) at 1 $km \times km$ line up at $\frac{m}{n} = 0.75$ in panel (b)? Maybe, I missunderstand.

Code

Thanks for including the code. Useful comments and instructions are provided. Inputs, except the inventories, are available for download. Maybe pseudo-inventories (with a warning in the code) could be created to make all codes executable. Overall, great effort to make the programming approaches available.

3 Technical corrections

Line 155: '... make good estimates of ...', estimation is the process of making an estimate

Line 255 and others: 'good compressible' and 'not good compressible', a better formulation should be found, e.g. 'compressible' and 'incompressible' (define what is meant by incompressible)

Line 234: 'sensing matrix matrix A', delete one 'matrix'

Line 243: '... sensitivity is beneath a certain threshold', 'below' works better

Line 332: '... descent ...', should be 'decent'

The code is more clear if variable x_l2 is used in the l_2-case in file *optimizeL2_noise.m*.