Response to reviewers

A sparse reconstruction method for the estimation of multiresolution emission fields via atmospheric inversion

Ray, Lee, Yadav, Lefantzi, Michalak and van Bloemen Waanders, GMDD 7:5623-5659, 2014.

Response to Reviewer No. 3

We thank the reviewer for his/her suggestions and comments. Our responses to his/her suggestions are below.

Overall response: At the very outset we would like to clarify that our study is focused on investigating the algorithmic aspects of a sparse reconstruction method (based on Stagewise Orthogonal Matching Pursuit, StOMP) for estimating rough emission fields, such as that of fossil-fuel CO2 (ffCO2). A sparse reconstruction method is necessary since the spatial parameterization for rough fields tends to be high dimensional (many parameters). The parameters that can be estimated depend on the information content of the observation data, which can change with time/season. The method is not customized to a particular tracer, measurement network or a transport model. Customization to a tracer occurs when we choose a spatial parameterization (a wavelet-based random field model in this study) for use with our sparse reconstruction method. It also occurs when we choose an observational dataset. The method can accommodate prior information on the field being estimated, but only uses its spatial pattern; thus, by design, it is insensitive to under/overestimation of the emissions in the prior information.

The paper investigates which formulations of the inverse problem do and do not work, and explains why. It develops a metric (mutual coherence) to quantify the information content in the observations collected by our measurement network. It finds the information content lacking, which motivates the need to introduce prior information into the inverse problem. We then identify a way to do so; the obvious/intuitive ways do not work. We also show how a wavelet-based field model, designed for modeling fields in rectangular geometries, can be used to estimate emission fields in an irregular region $\mathcal{R}$ (the Lower 48 states of the US). Finally, we show how StOMP can be extended to enforce non-negativity on the estimates. Sparse reconstruction methods are typically not used in atmospheric inversions.

Our motivation to develop this method arose from a need to construct and/or validate gridded inventories of ffCO2. Fortunately, many gridded ffCO2 inventories are available and a wavelet-based spatial parameterization also exists. We demonstrated the method in an idealized, synthetic-data inversion. The idealizations include: (1) assuming ffCO2 to be a radiocarbon-like tracer and ignoring interference by biospheric CO2 which can make ffCO2 estimation impossible except in winter (see Shiga et al., [2014]); (2) using a model-data mismatch ε that is smaller than the one used in real-data inversions and (3) assuming the same distribution for ε for all towers (i.e., ignoring transport model errors). These idealizations allowed us to explore issues related to the algorithm and formulations in a relatively “clean” setting. We also use an observational dataset collected from a measurement network that was sited with biospheric CO2 fluxes, not ffCO2, in mind (the towers are usually far from locations with high ffCO2 emissions); a network for ffCO2 does not currently exist.

Due to these idealizations adopted in our test, we do not claim that the method can be used to estimate ffCO2 emissions fields in a realistic setting using measurement techniques and infrastructure that are currently available (or could be in the near future). At the very least, our method has to be extended to include the estimation of biospheric fluxes as well as larger and tower-dependent model – data mismatches. This is a substantial body of work and outside the scope of this study. In order to check how accurate the estimates would be, we would have to conduct an OSSE (Observational System Simulation Experiment) or design an ideal network. Our tests also provide no information on the best method to collect information on estimation of ffCO2 emissions over regional scales (tower, airplane transects etc.).

We check our inversion method using the following metrics:

1. As part of our algorithmic development, we modify StOMP to incorporate prior information to improve estimates. We check whether it indeed does so, since the information content of the observations are found to be poor.
2. The aim of sparse reconstruction is to estimate parameters that are supported by data (usually large spatial patterns in the emission field) and remove the details that are not. We check whether this "sparsification" characteristic of the algorithm is still present after including prior information.
3. Our method restricts emission fields in an irregular region $\mathcal{R}$ (while using a wavelet-based model); this incurs a computational cost that can be limited by a user-defined setting. We check if the behavior of the algorithm provides a principled way of computing that setting (e.g., if improvement of results shows a "diminishing returns" behavior with the computational cost).

Note that in this study we do not use the accuracy of the estimated emission fields as a metric for evaluating our inversion method. This is because accuracy of estimation is determined primarily by two factors (once we have specified a model – data mismatch): (1) the suitability of the spatial parameterization for the rough fields being estimated and (2) the information content of the observational dataset. In our previous paper [Ray et al, 2014] we fixed the observational data and used the accuracy of the emission estimates to gauge the quality of the spatial parameterization. The converse procedure – fixing the spatial parameterization and varying the quantity of observational data – is not a very useful direction for investigation, for our StOMP-based method, because of the following reasons:

1. The estimation accuracy of StOMP, as the quantity of observational data is varied, has been investigated in Donoho et al, [2012]
2. If the aim is to obtain a very accurate reconstruction of the ffCO2 field (when we have full discretion to design an ideal observation network/technique), then we are limited only by what the spatial parameterization can capture. As reported in our previous paper [Ray et al, 2014], the spatial parameterization with 1023 wavelet coefficients (parameters) has a relative error of 10% at the 1-degree resolution; this would be recovered (modulo the small model – data mismatch) in case of an ideal network. If we retain all wavelet coefficients that can be described on a 1-degree mesh in the spatial parameterization (4096 coefficients), the reconstruction will be perfect (modulo the model – data mismatch).

Atmospheric inversion could be a way of estimating/verifying self-reported ffCO2 emissions in countries where the uncertainty is high. The uncertainty in emission reports from China is estimated to be 15-20% [Andres et al, 2012], though studies based on the TRACE-P campaign proposed a 54% revision of inventory estimates for 2000 [Suntharalingam et al, 2004] (it was officially revised upwards by 23% between 2006-2007). Other countries have larger variations. These uncertainties affect inventories, but do not affect our inversion method (see paragraph 1). Even if our variable of interest were to be total emissions over a region (nation or province), estimating a spatially variable emission field before spatially aggregating it reduces the aggregation error. However in order to do this, a measurement infrastructure designed with ffCO2 in mind is a requirement. Its size will be determined by whether we are interested estimating total national emissions or we seek fine scale details.

In addition, as mentioned above, our method can be used with other tracers provided we have a spatial parameterization for them.

The introduction section in our paper does not describe the idealized nature of our tests or the limits/caveats on the conclusions that can be drawn from them. It also does not describe the reasoning behind the metrics that were adopted for evaluating our algorithm. We will add them in the revised paper.

General comments

1. The reviewer states (Para 2) "However, it is a big pity to me that the authors did not choose a good test case to prove the practicability of the method developed. As also pointed out by another RC, the setup for their synthetic fossil fuel CO2 (ffCO2) emission inverse problem is too far from the reality. Because of that, I had a difficulty in discussing the feasibility of this method in a fossil fuel emission estimate. In my opinion, what the authors supported by the synthetic experiment is the replication of "non-negative emission fields" in a linear inverse problem, not ffCO2 emission fields. Thus, I would suggest to rework on the synthetic experiment and/or reword some of the text in the manuscript depending on what the authors would like to claim by this manuscript. In the following two sections, I'm trying to discuss several major concerns. Initially, I was listing points I wanted to discuss, but I learned that most of them have been raised by RCs for the author's previous paper (Ray et al. 2014).

http://www.geosci-model-dev-discuss.net/7/1277/2014/gmdd-7-1277-2014-AR1.pdf (last access: Nov 20, 2014)

*So I decided to make use of the *pdf as a reference. I understand that the *pdf was for discussions for Ray et al. (2014), but I think this is still fair to do as in anyways I would raise the exactly same questions/concerns and we probably don't want to replicate the same conversations presented in the *pdf."*

The aim of the paper was to present a sparse reconstruction method for estimating rough emission fields. We demonstrated it on an idealized test case, as described in the "Overall response" above. It is a methodological first step towards estimating ffCO2 emissions in a realistic scenario. There are limits to the conclusions that can be drawn from the idealized test, regarding the usefulness of the method in estimating ffCO2 emissions using existing measurement infrastructure. These were not mentioned in the paper, and we will correct this shortcoming, as described in the "Overall response" above.

Section called "ffCO2 does not seem to be the best emission to be used for the synthetic study"

2. The reviewer states, in para 1, "As acknowledged in Ray et al. (2014), ffCO2 emissions is very difficult to estimate given the existing observation network and also data collected there. The recent paper Shiga et al. (2014) (I see two of the authors are listed) also confirmed that as well. Shiga et al. (2014) also pointed out that the winter month have a better luck in estimating (detecting) ffCO2, but it would be very difficult the rest of the month. Probably this is something the authors should acknowledge in the manuscript."

We agree that we should have cited Shiga et al., [2014], about the limits of estimating ffCO2 emissions. See "Overall response" above for details.

3. The reviewer states, at the end of paragraph 1, "I also strongly feel that the authors need to convince us of the use of tower CO2 for ffCO2 estimation. Or since it was a synthetics study, the authors could come up with an ideal tower network for a ffCO2 emission inverse problem."

We agree with the reviewer that the tower network used in this study will not result in very accurate ffCO2 emission estimates. As described in the "Overall response" above, we use these not-very-informative measurements to uncover numerical characteristics of our method.

The term "ideal tower network" above introduces an ambiguity. If the term means that one is allowed to have as many towers as needed, then the test is not very useful (see "Overall response"). If, on the other hand, the term implies a network with *M* towers ideally sited for ffCO2 measurement, then one needs to perform a network design first, which in turn requires an ability to solve the (emission estimation) inverse problem, the topic of our paper. Demonstrating the algorithm inside a network optimization loop is outside the scope of the paper.

4. The reviewer points out, in the beginning of paragraph 2, "As we see in the manuscript, the use of existing tower network made the synthetic inversion setup far off from the reality. For example, the error assigned to the radiocarbon-like tracer was 0.1 ppm. This is too small compared to an actual C14 case as pointed out by the review discussion for Ray et al. (2014). It was acknowledged in Ray et al. (2014), but this small error was used again in this manuscript without any note."

This was an oversight. We will update the manuscript to reflect our reason for such a small value of ε. The reason is the same as it was in Ray et al., [2014]– it isolates the effect of the spatial parameterization and the inversion algorithm on the estimates.

5. The reviewer states, in the middle of paragraph 2, "The authors mentioned the use of Carbon Monoxide (CO) in their response to defend the use of the 0.1 ppm error. I would not say no, but it is not clear to me how CO data from the existing tower data would help us to estimate ffCO2 emissions that well. If the authors meant to say the use of satellite-driven CO, that would become another problem as an inversion setup needs to be modified (also, data number would dramatically increase). My point here is that it is very questionable that such an ideal tower-based observational data become available sometime in near future, although I definitely think we push forward to make that happen as a community (in my opinion). The reality is we *don't see any plan to expand the tower network to look at fossil fuel emissions (as far as I know). Perhaps I would be a little bit more convinced if the authors would have used observation from Orbiting Carbon Observatory 2 (OCO2) and/or Greenhouse Gases Observing Satellite (GOSAT) (although I don't know how to derive ffCO2 contribution in XCO2 data)."*

We agree that there does not seem to be any movement to design a network for ffCO2 and such low errors are not currently possible. However, these were idealizations that we adopted to demonstrate that we could estimate rough/multiresolution fields, and to isolate the effects of the algorithm and the spatial model on the estimated emissions. We do not claim that the method can be used with the current measurement infrastructure or transport models.

6. The reviewer points, in the last third of paragraph 2, "Also, the emission imposed by Vulcan does not keep the nature of fossil fuel emission fields because of the simplification (averaging). Fossil fuel emissions are not constant over 8 days. We do sometime estimate monthly and/or weekly natural fluxes via inversion, but we do usually have diurnal cycle in forward modeling. Because of points I made above, I thought this manuscript is misleading when I read the abstract saying "We demonstrate the method on the estimation of ffCO2 emissions".

We agree that we have approximated the true emissions (Vulcan) by averaging over 8-day periods. We will update our manuscript to state that the method is demonstrated on an idealized ffCO2 estimation problem. We will also state the idealizations upfront. We will also point out in the paragraph where we describe how the observations are generated that the averaging over 8 days removes the diurnal variation on ffCO2 emissions.

7. The reviewer states, in paragraph 3, "I would suggest to the authors to choose a different application to design a synthetic inversion. It is fair to claim that this manuscript is focusing on a method development. But the synthetic experiment needs to be reasonably close to the reality, especially if a particular target is assumed. If the authors want to stick to the ffCO2 inverse problem, the authors should design more realistic experiment, prove the feasibility of the method and show the reproducibility of the truth. Maybe one thing the authors easily could do is to reword "fossil fuel emissions" to "non- negative emission fields". After proving the reproducibility of the true field, the authors could discuss a possible application to fossil fuel emission estimation (like people do when they propose a numerical scheme). This correction should be fair enough to support the development of the inversion scheme for multi-resolution fields in a linear inverse problem."

Our paper is an algorithmic one, with an idealized test case of estimating ffCO2 emissions. We will reword the paper to use non-negative emission fields where we can. However, the test case, no matter how idealized, is connected to ffCO2 since we use Vulcan, EDGAR and a spatial parameterization for ffCO2. Consequently, we have retained the use of ffCO2, rather than the unwieldy "non-negative emission fields" when describing the results in Sec. 4.

In the Conclusions section, we will discuss the advances that need to be made in order to enable the use of our method in a real-data inversion.

In the section named "What is the real utility of this method in ffCO2 study"?

8. The reviewer states, in lines 1-9 "Apart from the synthetic study presented for a moment, I would like to discuss about the utility of the method presented in the manuscript (which I *thought a major weakness of this manuscript). Again, I understand this manuscript is trying to advance the methodological aspect of future ffCO2 estimation method using atmospheric measurements (which I think it is great). I'm grad to see the authors mentioned current issues in ffCO2 inventory/modeling. But I was a little bit disappointed because this work is not really responsive to those issues. For instance, even in the highly-idealized condition, the method does not give an accurate answer. Given by the nature of the method, this method could only offer a good approximate of ffCO2 fields, but not an accurate field. I thought this is a critical shortcoming."*

The aim of the paper was to present a sparse reconstruction method for estimating rough emission fields such as ffCO2. The idealized test case that we employed presents limits on the conclusions that can be drawn from it; we will update our manuscript to reflect them (see "Overall response" above). Also, reconstruction accuracy, when we have full discretion to design the measurement network in our idealized test case, is not a useful metric for judging our method (see "Overall response" and Response no. 3 above).

9. The reviewer states, in lines 9-13, "The authors claimed that the error in their estimation is around 5%. But the 5% is not the same thing as 5% two sigma uncertainty for reported national emissions. In real world, we need to deal with ffCO2 that has diurnal cycle while other strong signals from biosphere are present (and often we have a difficulty in decent angle them)."

The reviewer does not specify where in the text we state 5% as the estimation error. The only instance that we can find is on Pg 5646, line 15. The error is evaluated using the output from Stage I of our algorithm, which computes ffCO2 fluxes *before* non-negativity enforcement. It is also not globally aggregated over $\mathcal{R}$ (the figure stated is root mean square relative error of the 1-degree flux field with respect to Vulcan); the 3-5% error margin that the reviewer states is the uncertainty in total emissions reported by the US. We believe that the comparison is being made between two different things.

On a broader note, we fail to understand why the estimation accuracy of a guessed flux field, before non-negativity enforcement, should matter, especially in such an idealized inversion test case. We have computed and plotted reconstruction errors (after non-negativity enforcement!) in Fig. 2, and they do vary between 2-15%, but only to discriminate between inversion formulations. Nowhere do we claim that the reconstruction errors obtained in our idealized test case are indicative of what could be achieved in a realistic setting.

10. The reviewer states, in lines 13-26, "In addition, the method is assuming a 1x1 degree field where we still don't resolve most of sectoral emission activities (at this spatial resolution, emission spatial proxy such as population and nightlights works pretty well). If I need a ffCO2 emission field, say for this year 2013 for US at a 1x1 degree, I would just project Vulcan emission data (or CDIAC 1x1 degree map) using fuel consumption data (I see this is essentially a very, very crude version (also w/o observation) of what the authors have done in this manuscript). In this way, I don't get any update on the emission spatial extent, but at least I could expect a very good estimate of the annual total emission. Given that, I'm not clear about what the method presented in this manuscript could offer to us. It seems to me that it is very difficult to expect the method to yield a very accurate emission field or update inventories. What we ultimately expect is to be able to tell the discrepancy between what is reported (or calculated) and what is measured using atmospheric observation (e.g. MRV). But it is not clear to me if the method could offer the opportunity, as it seems to be difficult for atmosphere to tell *a subtle difference in emissions especially at an aggregated 1x1 degree. Of course these do not need to be fully addressed, but I would be very curious how the authors envision to approach to the goal from where we are. I think these kind of discussion should be done to figure out what we really need to do, prior to getting into the technical details."*

The reviewer's method for updating inventories is applicable when one has an accurate inventory to start with and good fuel usage statistics. As we point out in the "Overall response", there are many large ffCO2 emitters with significant uncertainties in their national-level emission. They are better targets for atmospheric inversion (though, unfortunately, their tower networks are even sparser than the one in North America, and are also not sited with ffCO2 in mind).

We are surprised by the reviewer's statement: "*But it is not clear to me if the method could offer the opportunity as it seems to be difficult for atmosphere to tell a subtle difference in emissions especially at an aggregated 1x1 degree"* - we have performed no tests whatsoever in our paper that investigate how large a change in an inventory we can detect using atmospheric inversion. The magnitude of the change that we could detect would be conditional on the observational dataset that we are provided.

11. The reviewer states, on line 31, "I thought the authors could easily work on more simple inverse problem for non-negative fields using an ideal observation network (and/or satellites) to estimate accurate emissions. I would thus suggest to rework the text and the synthetic experiment to be more responsive to ffCO2 issues if the authors would like to address a ffCO2 issues."

We did consider using a simplified inverse problem. In order to demonstrate that our method could identify and remove a large number of parameters from an emission field model and restrict emissions to an irregular region $\mathcal{R}$, the problem had to be about estimating a 2D field (for a 1D field, these are trivial).

We first considered a simplified 2D problem, with a tower network that would allow us to estimate most of the parameters of the emission field model, and generate accurate solutions. However, in such a data-rich/informative setting, the impact of including the prior would be very muted (Approaches A and C would give very similar solutions). Such a test would not allow us to verify whether our modification of StOMP, to include prior information, had been performed correctly and was contributing information to the inversion (the prior information would be overwhelmed by the observations). Thus we could test only two out of the three features of our new algorithm (non-negativity and restricting emissions in $\mathcal{R}$ and not inclusion of prior information).

Consequently, we adopted a 2D problem where the observations were somewhat informative about the emission field, but still allowed the prior information to be reflected in the solution of the inverse problem. However, we retained some simplifications – no interference by biospheric CO2 fluxes, as well as a small and (statistically) similar model-data mismatch at the towers. Further, the reconstruction errors have a known lower bound (see "Overall response").

12. P5624, L13: The reviewer states "This synthetic inversion was implemented in the highly-idealized synthetic world. Although the Vulcan data was used to create a truth field, this *synthetic study is not fair enough to support the implementation of a fossil fuel emission estimation (see my general comment). This needs to be rephrased. Otherwise this could easily give an wrong impression to the audience of GMD."*

We agree that the idealized test case imposes limits/caveats on the conclusions that can be drawn from it. We will update the manuscript to make it clear. See "Overall response" above.

13. P5624, L17: The reviewer states, "This is certainly an improvement achieved by this paper. But in my opinion, if the authors meant to develop a method to estimate FFCO2, the first thing the authors should try is to get an accurate emission estimate. Even if a method is very computationally heavy, people would be happy to use it to get an accurate estimate. As a study for FFCO2 estimation, this is not very appealing (at least, to me)." He/she states this in the context of our 10x reduction in the computational time needed to enforce emissions in an irregularly shaped region.

We agree with the referee that accuracy is of paramount importance. But our novelty is in Step 1 that essentially generates an imperfect "guess" of the emission field; it does not affect the accuracy of the final estimates. The novelty reduces the time our algorithm spends to generate the guess. Generating the guess quickly has some practical benefits e.g., being able to scale up to large problems. It would be remiss of us not to point out this feature, especially in an algorithmic paper.

14. P5625, L16: The reviewer states "I think the important thing here is how accurate the estimate is." The context is as follows: We describe that spatial parameterizations (or models) of emission fields have too many independent parameters to be estimated from the available atmospheric observations, and thus have to be simplified. Our method performs the simplification based on the observations, and if the information content of the observations varies in time, our method can construct models with variable complexity.

We agree with the reviewer that inversion accuracy should be the guiding principle. Given an observational dataset, the best accuracy is obtained when (1) the spatial parameterization is sufficiently low dimensional to not overfit the observations and (2) sufficiently high dimensional to capture all the structures in the emission field being estimated that are supported by the data. This optimal dimensionality is not known, *a priori,* and our sparse reconstruction method is one way to efficiently identify it, simplify the spatial parameterization and fit it to data in one integrated mathematical formulation.

15. Pg 5626, L11: The reviewer states "I don't see the method developed here is going towards this goal.", the goal being developing spatially resolved estimates of ffCO2 emissions.

We beg to differ with the reviewer's comment that the method we have described is irrelevant to the estimation of ffCO2 emission fields; see "Overall response" above. The arguments that the reviewer has presented show that the current measurement network, along with current transport models and measurement methodologies (inaccurate and expensive radiocarbon observations) will not allow the estimation of ffCO2 emissions fields, but we do not claim that we could do so.

16. P5626, L21: The reviewer states "Since this method is based on the parameterization, it would be difficult to achieve an emission estimate that is accurate enough to updated or improve an emission inventory. This limitation needs to be acknowledged."

We agree. We will update our manuscript with this caveat, at the end of the paragraph.

17. P5626, L26: The reviewer states "It is unclear to me how this method could offer a method to give an accurate emission estimate using atmospheric measurements." He/she states this in the context of our discussion where we say that satellite and airplane transect do not offer a scalable way of updating inventories.

See Response no. 15 above.

18. Pg 5627, L19: The reviewer states "I would like to learn the benefit this method could offer as opposed to the use of an gridded inventory plus fuel statistics (see my general comment)." He/she makes this remark in the context of our statement that "inventories being updated can serve as very informative priors and reconstruction methods could profitably use them".

The reviewer's method for updating inventories is applicable when one has an accurate inventory to start with and good fuel usage statistics. As we point out in the "Overall response", there are many large ffCO2 emitters with significant uncertainties in their national-level emission. They may be better targets for atmospheric inversion, though unfortunately, their networks are even sparser than in North America, and also were not sited with ffCO2 in mind. Further, atmospheric inversions can also uncover uncertainties in inventories; see Brioude et al., [2012].

19. Pg 5628, L20: The reviewer asks "So here, the authors assumed that biosphere fluxes is perfect? Or ffCO2 can be solved as a separate problem?"

In our idealized test case, we have assumed that ffCO2 can be solved as a separate problem.

*20. Pg 5630 L11 and L20: The reviewer states "How would you specify Q here?" (Line 11, Q being the prior covariance). He/she also asks "Given the use of **f - fpr,** would you be able to "correct" an inventory using atmospheric observation?"*

We assume that the reviewer means whether we could correct an inventory given the use of (**f – fpr**) in Eq. 1, since **f** would be affected by under-reporting in the inventory **fpr** unless there was overwhelming atmospheric evidence against it.

We are a little confused by the reviewer's comment, since it implies that we solve Eq. 1 to obtain our emissions estimates. We do *not*. Eq. 1 is based on a multiGaussian assumption (alternatively, a Gaussian random field) for the field being estimated, and we clearly state L18-L23 that this model is not very convenient for expressing rough emission fields like those seen for ffCO2. Consequently we do not define Q, nor do we use (f – fpr) in our inversion. Instead, we use a multiscale random field model developed in our previous paper, which is sufficiently flexible to accommodate rough fields such as ffCO2 emissions.

Our inversion scheme is described in Sec. 3.2 "Posing and solving the inverse problem". It is quite different from Eq. 1. It is the sparse reconstruction method that we have developed in this paper.

21. Pg 5640, L1: The reviewer states "The synthetic observation is really a key in this study. This should be fully described. It is really misleading because CO2 concentration is actually a radiocarbon-like tracer, not normal CO2 concentration usually measured."

We understand the reviewer wants us to make it clear, upfront, that this is a synthetic data study, with all the simplifications and idealizations that have been adopted.

We agree. We will state clearly in the Introduction that the paper develops a method for estimating rough, non-negative emission fields and is tested with an *idealized, synthetic-data* study. ffCO2 estimation is merely the test case, and we will use a multiscale random field model that was developed for rough ffCO2 emission fields in a previous paper.

22. Pg 5640, L2: The reviewer states "Vulcan is averaged. This should be considered when you evaluate the numerical accuracy. I assume the authors did it right, but I was actually not very sure of it by this manuscript."

Yes, we average Vulcan up to 1 degree resolution in space and 8-days in time. We use this averaged Vulcan as our "truth" emissions, and use it to generate the observational data.

23. Pg. 5640, L7: The reviewer states "Again, this CO2 is a synthetic radiocarbon-like tracer."

We understand that the reviewer wants us to make clear that ffCO2 behaves as a synthetic radiocarbon-like tracer.

We agree. We will explicitly state so.

24. Pg. 5640, L7: The link to Ray et al, 2013 is dead

We will specify an alternate website from where the technical report can be downloaded.

25. Pg5640, L22 The reviewer asks "So your "truth" is a 8-day averaged Vulcan. Correct? Did you use the averaged Vulcan for creating synthetic CO2 data or hourly Vulcan data? I want to make sure."

We used the 8-day-averaged Vulcan emissions to generate observations. It is our "truth" emission field. The text on L25 is badly phrased and we will make it clear.

26. Pg 5640, L24: The reviewer states, "So the afternoon selection is not applied". This statement is made in the context where we discuss how we obtain the sensitivity of 3-hourly concentration observations to 8-day-averaged emissions by summing up 3-hourly sensitivities

Correct. We use all the observations, and not just the afternoon data.

27. Pg 5641, L2: The reviewer states "This is too small. Also, no transport uncertainties?" He/she states this in the context of the data – model mismatch that we use (0.1 ppmv).

We agree that the model –data mismatch is too small to be realistic. This issue was also raised in our previous paper, and we provide the same explanation.

The primary problem with using a realistic value of ε (1 ppmv rather than 0.1 ppmv) is the placement of the measurement towers – they are far from the sources of ffCO2 emissions leading to measurements that are around 2 ppmv. Adding a noise with variance of 1 ppmv makes them unusable. A true test of our method, in realistic conditions, would require a sensor network sited with ffCO2 emissions in mind. Consequently, we have chosen an idealized scenario and focused on the algorithmic aspects of the problem.

We did mention in the paper that we chose ε to be 0.1 ppmv to construct an idealized scenario within which we could test the quality of the proposed numerical scheme. We will update the manuscript to be more explicit about the reason why.

28. Pg 5645, L3 The reviewer states "Spell out BAO and MAP. I assume they are not used for the synthetic inversion. Correct? Why were those two sites selected? I saw this in a pessimistic way as the model is showing almost perfect fit to the observation while the emission estimate is still not perfect."

Thank you for catching this error. BAO = Boulder Atmospheric Observatory, Colorado; MAP = Mary's Peak, Oregon. We will update our manuscript with this information.

BAO and MAP are not being used as a measure of the accuracy of the inverse problem (as an out-of-sample test). They were used in the estimation of the emission field. In Sec 4.3, we check that the algorithm satisfies some of the necessary conditions: (1) after estimation, the emission field should be able to closely approximate the observations and (2) since the measurements are too un-informative to estimate all the parameters in the spatial model for the emission field, the sparse reconstruction method should be able to identify the "un-estimate-able" ones and set them to zero. In Fig. 5, we show that these necessary conditions are met.

The reviewer is correct in remarking that we seem to reproduce observations well, but the estimated emission field is far from the truth (Fig. 2 and 3). This is a consequence of the roughness of the emission field and the fact that measurement sensors are very sensitive to emissions in their vicinity. The spatial model allows rough fields (within limits) that can accommodate localized sources (i.e., set some wavelet coefficients in the spatial model to non-zero values) in the vicinity of the measurement tower. This feature of rough emission fields can be used to match observations very well, if the improvement in the match overcomes the penalty that the sparse reconstruction imposes due to the consequent increase in the complexity of the spatial model. This often happens if the surrounding towers are too far away, or upstream in the wind direction, to provide a constraining influence on what may be an erroneous emission source. The spatial parameterization can also be leveraged to prevent the introduction of such localized sources and consequently requires careful design. This was discussed in our previous paper. Note that if the emission fields are smooth and modeled as multiGaussian fields, such localized sources in the vicinity of the tower cannot be accommodated i.e., in the absence of informative observations, the spatial model imposes a much tighter constraint than our multiscale random field model.

The accuracy of the inversion is thus very dependent on the spatial model and consequently inversion accuracy was discussed in our previous paper. There, we quantified the inversion accuracy entirely in terms of the estimated emission field. Since inverse problems can support non-unique solutions, we use the ability to closely reproduce tower observations as a necessary condition, but not a proxy, for estimation accuracy. We will update our manuscript to (1) clarify that the accuracy of the inversion should be judged using Fig. 2 and 3, as well as the results in the previous paper and (2) Fig. 5 should be seen as a means of verifying that certain necessary conditions that the sparse reconstruction scheme should meet are actually being satisfied.

We will also change the title of the section to "Numerical consistency and computational efficiency". It is a better description of the results presented there.

29. Pg. 5645, L1: The reviewer states "Again, I think the accuracy of estimation needs to be discussed first if ffCO2 emission estimation is the ultimate target of this project."

The reviewer makes this statement at the beginning of a section titled "Numerical accuracy and computational efficiency. While we agree that accuracy of estimation should be the guiding principle, we are a little confused by what he/she implies.

- If the reviewer wants us to discuss the accuracy of the emission fields estimated using our idealized tests, then see Ray et al., [2014], "Overall response" above which explains the reason for them being in our previous paper and Response no. 28.
- If the reviewer wants us to discuss the implications of these results on the accuracy of a real-data inversion, see "Overall response".
- If the reviewer wants us to discuss the accuracy that our method could achieve using *M* optimally sited towers, see Response no. 3 above.

30. Pg 5645, L13: The reviewer states "The authors should defend that why those small minor feature can be ignored while a very high accuracy is required to improve and/or update inventories."

The use of "unimportant" was a poor choice of words and we will remove it.

What we meant was that in the test case (and real life), the measurement network is unlikely to be sufficiently dense to estimate all the fine scale details of the emission field via atmospheric inversions, and it is better to remove them lest they fit themselves to noise in the data. In our case, we leverage the explicit separation of scales in the multiscale random field model to set these details to zero and simplify the emission model, if doing so does not decrease the quality of the model fit below a threshold. This discrimination is performed in our sparse reconstruction method.

31. Pg 5645, L28: The reviewer states "We should not have a trade off especially in this highly-idealized world. Again, we need an accurate estimate to address an issue in ffCO2 study." This is said in the context of the trade-off between computational efficiency (M_{cs}) and estimated fluxes ($\mathcal{F}_{\mathcal{R}}$ and $\mathcal{F}_{\mathcal{R}'}$) in an idealized computation.

As clearly mentioned in the same paragraph, the choice of M_{cs} only affects Step 1, where we develop an approximation of the emission field. The approximation is an "emission" field that is mostly, but not exclusively, non-negative, and is mostly limited to the region $\mathcal{R}$ where non-negative emissions are desired. These emissions are referred to using f_k or $\mathcal{F}$. The approximated "emission" field is processed in Step II of our algorithm to enforce non-negativity; we refer to the non-negative emissions as E. In both steps, we iteratively update the $\mathcal{F}$ or E field to minimize the difference between tower observations and model predictions.

Setting M_{cs} to small number results in a poor approximation and a large number of iterations in Step II. A large M_{cs} spends a long time in Step I, but Step II is quick. The algorithm converges and stops when a certain accuracy metric is met in Step II; the estimated emission field's accuracy is independent of the quality of the approximation that is introduced into Step II via M_{cs}. However, the computational time is not.

We found that increasing M_{cs} (and computational cost) quickly reached a point of diminishing returns, with respect to the quality of the guess that was introduced into Step II. That provided us with a principled way of limiting the time spent in generating a guess.

32. Pg 5646, L15: The reviewer states "This means the method doesn't get both national total and spatial distribution right." This is said in the context of the guesses of emissions being prepared in Step I, prior to non-negativity imposition in Step II.

The reviewer is correct that the guesses of emissions prepared in Step I, and before non-negativity enforcement in Step II, do not get the national total and spatial distribution correctly.

33. Pg. 5647, L16. The reviewer states "Again, this needs to be rephrased." This is said in the context of our statement "we have demonstrated our method on the estimation of ffCO2 emissions ..."

We agree. We have demonstrated estimation of non-negative, spatially rough emissions, in an idealized setting.

34. Pg. 5648, L9: The reviewer states "Two drawbacks are acknowledged in Ray et al. (2014) (See P1917, 1: Need for a good fossil fuel tracer, 2: uncertainty as also mentioned here), but the first one is not acknowledged."

We agree. We will make this correction.

35. Pg 5649, the reviewer points out we misspelt Candés

Fixed.

36. Pg 5652, L2: The reviewer points out a dead link

Fixed.

37. Pg. 5655, Fig 2. The reviewer states "Compared to Table 2 from Asefi-Najafabady et al. (2014), the correlation is not very good. The FFDAS (Nightlight+Population) scores a correlation of 0.86. This is a part of the reason I don't see a utility of this method. Especially this is a synthetic study, the author should seek the way to get much better score." This is said in the context of two bottom-up inventories, FFDAS v2 and Vulcan, having a very good agreement (spatial correlation) between themselves, far better than the agreement between Vulcan and our inverse problem solution.

Vulcan and FFDAS share many of the underlying data sources. Both FFDAS and Vulcan use the same self-reported emissions and similar data on population density etc. Further, as the authors in [Asefi-Najafabady et al, 2014] clearly point out, the high spatial correlation is due to the presence of pointwise power plant emissions – "FFDASv2 and Vulcan have nearly identical power plant emission data, and the total correlation is driven by the near-perfect agreement for this sector". It is not surprising that the two inventories match so well.

We fail to understand why the reviewer thinks that the agreement between the two is a good way of setting performance/quality thresholds for a technique that would provide an independent check on such inventories.

[Andres et al, 2012] R. J. Andres et al, "A synthesis of carbon dioxide emissions from fossil-fuel combustion", Biogeosciences, 9, 1845-1871, 2012. doi: 10.5194/bg-9-1845-2012.

[Asefi-Najafbady et al, 2014] S. Asefi-Najafbady et al, "A multiyear, global gridded fossil fuel CO2 emission data product: Evaluation and analysis of results", Journal of Geophysical Research: Atmospheres, doi:10.1002/2013JD021296, 2014.

[Brioude et al, 2012] J. Brioude, et al, "A new inversion method to calculate emission inventories without a prior at mesoscale: Application to the anthropogenic CO2 emission from Houston, Texas", Journal of Geophysical Research, 117, D05312, doi:10.1029/2011jD016918, 2012.

[Donoho et al, 2013] D. Donoho et al, "Sparse solutions of underdetermined linear equations by Stagewise Orthogonal Matching Pursuit", IEEE Transactions on Information Theory, 58, 1094-1121, 2012.

[Ray et al, 2014] J. Ray et al, "A multiresolution spatial parameterization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions", Geoscientific Model Development, Geosci. Model Dev., 7, 1901-1918, 2014. doi:10.5194/gmd-7-1901-2014.

[Shiga et al, 2014] Y. P. Shiga et al, "Detecting fossil fuel emissions patterns from sub-continental regions using North American in-situ CO2 measurements", Geophysical Research Letters, 41(12):4381-4388, 2014.

[Suntharalingam et al, 2004] P. Suntharalingam et al, "Improved quantification of Chinese carbon fluxes using CO2/CO correlations in Asian outflows", Journal of Geophysical Research, 109, D18S18, 2004. doi:10.1029/2003JD004362.