Geosci. Model Dev. Discuss., 7, C801–C810, 2014
www.geosci-model-dev-discuss.net/7/C801/2014/


Geoscientific Model Development Discussions
Open Access

# *Interactive comment on* "A multiresolution spatial parameterization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions" *by* J. Ray et al.

**J. Ray et al.**

jairay@sandia.gov

Received and published: 6 June 2014

*The reviewer states: "The authors obtain spatial sparsity fractions of about 80% for representing Vulcan emissions using wavelets. While these compression rates may, at first, seem phenomenal, they are not surprising given that most of the gridcells in the region of interest in Fig. (1b) do not contain significant levels of fossil fuel CO2 emissions. One alternative and naïve approach to wavelets would be to consider only those gridcells, or some aggregated set of gridcells (e.g. 4×4), that contain emissions above a specified level (e.g. >1% of the max). Another approach could be to prescribe spatial basis functions that have areas proportional to population (i.e. small areas for*



*large metropolitan regions, and large areas for rural regions). I surmise that these naïve approaches would also lead to large sparsity fractions or reductions in dimension. To better illustrate the strengths of their wavelet approach, I recommend that the authors devise a naïve metric of sparsity and compare and contrast their numbers to this metric."*

**Response:** The reviewer is correct inasmuch that there are other sparse representations of ffCO2 emissions e.g., simply retaining grid-cells above a threshold. However, what a collection of selected grid-cells will not provide is a random field model, i.e. a systematic way of constructing a field based on independent model parameters. The discussion in Section 3 (selection of wavelets) is not about sparsity per se but about a sparse random field model for ffCO2 emissions. A random field model could be constructed from the selected grid-cells by requiring that they be related to others in some fashion. One such relation could be to constrain the values of the grid-cells using a spatial variogram. This would result in a Gaussian random field, which are used when estimating biospheric CO2 fluxes (and which, as we described in Sec. 1, will not work for ffCO2 emissions). Alternatively, one could use a multi-resolution approach for the spatial model, using wavelets to develop spatial structures in a continuous-level-of-detail manner.

A sparse collection of grid-cells, with an inversion method treating each as an independent parameter, would not be very useful – measurements at sensors are not very sensitive to them individually. In contrast, the MsRF allows one to systematically perturb the weights of wavelets whose supports span the entire domain (which will have a substantial impact globally on the sensors) down to the fine-scale ones (2 degree x 2 degree supports), with very local impact. Ensuring that the MsRF is sparse (low-dimensional) is thus very significant in an inversion setting.

In our paper, we had assumed that simply using the term "random field model" would convey our intention of constructing fields in a systematic manner (for use inside an optimization algorithm, when solving an inverse problem). We will add this clarification

- our explicit interest in constructing fields systematically and ensuring that we can do so using as few free parameters as possible – in the beginning of Sec. 3.

*The reviewer states: "On page 1300, lines 26-28, the authors note that the deterministic nature of their presented method is a drawback. Without quantified confidence intervals and uncertainties, it is difficult to ascertain the significance of the inversion results (e.g. as shown in Fig. 7). The authors should run additional inversion tests that vary $\epsilon_2$, $\epsilon_3$, and other relevant parameters, and then report on the sensitivity of their results to these variations. Furthermore, the manuscript should contain a discussion of the errors described in items 3 and 4 below."*

**Response:** The main source of uncertainty in our inversions is the sparse set of measurement towers and their locations (far from the source of ffCO2 emissions). In a realistic (not synthetic data) inversion, the boundary conditions will also contribute their uncertainty, as will the "error" model for the data-model discrepancy. The numerical parameters mentioned above are not significant sources of errors.

The two parameters, $\epsilon_2$ and $\epsilon_3$ are numerical tolerances that are set, as is common in numerical studies, by reducing them till the results (wavelet coefficient estimates) become insensitive to their values (variations drop below 1.0e-6). Selection of the correct setting for $M_{cs}$ is more involved and is described in Ray et al, (2013); it involved increasing $M_{cs}$ till the error in ffCO2 estimates fellow below a threshold (10%).

We will add the rationale for the values of the numerical tolerances, and the citations to Sec. 5. In the new "Discussion" section, we will enumerate the sources of uncertainties whose impacts our deterministic method does not quantify.

*The reviewer states: "Underreporting is a known and persistent bias in using inventory-based estimates for monitoring anthropogenic emissions. The authors should describe what happens to this important source of error when using nightlights and BUA as spatial proxies for inventories in their wavelet representation. Does this error become*



*confounded with separate errors in the proxies and can it be attributed to the inventory post-inversion? In a similar vein, are there errors in the proxies (e.g. clouds obscuring nightlights) that become confounded with the inventory in the wavelet representation?"*

**Response:** Underreporting will have no effect on the accuracy of the estimation. Underreporting of country-level emissions lead to a smaller value of $c$ in Eq. 4, and a smaller $f_{pr}$. However, $f_{pr}$ serves as normalization for wavelet weights $\mathbf{w}$. A smaller $c$ will not lead to any differences in the wavelet weights *relative to each other*, and consequently, have no impact on the minimization of the L1 norm. The normalizing prior, of course, has no impact when evaluating the $(Y - G'w')$ constraint in Eq. 10.

Errors in nightlights and BUA can lead to smeared reconstructions. Errors in these proxies lead to an erroneous selection of wavelets in the MsRF. If we omit a fine-scale wavelet from the MsRF in a region with high ffCO2 emissions, it will be captured using a coarser wavelet that covers the region with the ffCO2 source (leading to a smeared source). If we select a wavelet in a region without significant ffCO2 emissions, the sparse reconstruction method will simply set its weight to zero. Complications can occur if an erroneously chosen wavelet has (1) a single sensor sited in its support AND (2) is far from all other sensors. The measurements at a sensor are immensely sensitive to emissions/fluxes in its vicinity. In an inversion, in this particular case, the measurements at the sensor could be attributed entirely to the emissions modeled by the erroneously chosen wavelet. Occlusion of nightlights by clouds is not a major issue since these proxies are annually averaged quantities.

We will add this in the new "Discussion" section of the paper.

*The reviewer states: "The inversions are performed assuming a perfect atmospheric model. In reality, atmospheric models contain biases and other imperfections that can severely limit the ability to invert for regional scale surface emissions. The authors should describe how model imperfections could be included the inversion (e.g. as an extra term in Eq. 5) and how they might be confounded with other errors in their sparse* *wavelet representation."*

: Transport model errors are not unique to estimation of ffCO2 emissions; they are also encountered in the estimation of biospheric CO2 fluxes too, and are addressed by using observation-specific model-data error variances (the diagonal terms in Re, in Eq. 6). We have cited papers by Chatterjee et al, (2012) and Gourdji et al, (2012), where their calculation has been described in detail. Adapting these methods to our ffCO2 emission problem is outside the scope of the paper. Observation-specific model-data error variances would result in the rescaling of the constraint in Eq. 10. We will add this to the "Discussion" section.

*The reviewer states "The inversion results for the U.S. shown in Fig. (7a) exhibit pronounced seasonality, with small error reductions during periods 7 and 27, and large error reductions offset by 2-3 months during periods 15 and 35. The time dependence of the inversion suggests the presence of multiple time scales of interest that do not seem to be represented in the inversion demonstration. Although the wavelet coefficients in Eq. 7 vary with time (i.e. they contain index k), the wavelets themselves do not (i.e. do not contain index k). Are the spatial distributions of the nightlight and BUA proxies fixed for the year? If so, would introducing time-varying spatial distributions of these proxies reduce this seasonality? Please respond and include appropriate discussion in the manuscript."*

**Response:** The errors in the estimation are due to changes in the wind direction, which blow the ffCO2 away from the measurement locations. Having a new MsRF for each k would imply that ffCO2 emitting regions change significantly on a weekly (or seasonal) basis, to the point that one needs to select new wavelets. This is unlikely. More practically, nightlights and BUA maps are computed as annual averages to remove the effect of cloud cover, holidays etc. on nightlight radiances, and it would be difficult to construct time-varying prior emissions, from a feasibility point of view.

We do, of course, allow for multiple timescales; our emissions are estimated at 8-day



resolution, which is sufficiently fine to capture any seasonal changes e.g., change of wind patterns, and we perform inversions for 360 days.

*The reviewer states: "In a comment related to item 5, fossil fuel emissions also vary over multiple time scales (daily, weekly, monthly, and yearly). Although the manuscript adequately describes the various spatial scales (and "spatial" is specified in the title), the discussion of multiple time scales is haphazard. I recommend including this discussion in the manuscript and describing how the sparse wavelet technique can (or cannot) be extended to capture multiple time scales. Making a clearer distinction between multiple time and space scales will also be helpful."*

**Response:** We are somewhat confused by the comment of haphazard modeling of time-scales of ffCO2 emissions, since our paper contains none. Emissions are estimated over 8-day periods, and each period is considered to be independent of the others. The MsRF is designed to address the non-stationary spatial nature of ffCO2 emissions. We have not addressed the temporal modeling of ffCO2 emissions because of the reason below.

Changing (seasonal) wind patterns (which blow ffCO2 away from our limited set of measurement sites) pose one of the stiffest challenges to accurate emission estimation, and is the issue that we have investigated here. The primary variation of ffCO2 emissions, as represented by the Vulcan dataset, is diurnal (approximately 2x, with an afternoon peak). A much smaller spatial variation occurs, comparatively very slowly, when emissions shift from the hot south to the cold north in winter. In this paper, we have focused on investigating whether the MsRF is useful for estimation purposes, given the seasonal nature of wind patterns. Seasonal processes are adequately resolved at weekly resolution. The seasonal variation of ffCO2 emissions occur at a timescale far longer than the time taken by ffCO2 to be transported across the US (roughly two weeks) and consequently we have not had to impose any kind of temporal correlation in the emissions to obtain our ffCO2 estimates.

We will add, in the new Discussion section, why we have not attempted to address the multiscale temporal nature of ffCO2 emissions. Sophisticated temporal modeling of ffCO2 emissions, to simultaneously capture both the diurnal and seasonal variations, could probably be performed using non-stationary correlation functions. Our MsRF was not designed to do this, and we conjecture that sparse reconstruction and wavelets would likely be overkill.

There is a possibility that the reviewer may have been misled by our use of the term "non-stationary". We do not use it to mean time-variant or unsteady. Rather, we use it in the statistical sense. For a stationary function, statistical summaries (means, variances etc.) computed within a moving window defined inside the support of the function would remain the same. More practically, a function that can be characterized by a single time/length scale e.g., a sine wave, is stationary, whereas another, displaying different scales in different parts of its support, is not. Consequently, a time-series of smoothly varying temperature can be stationary, whereas the porosity field of a block of dry soil may not.

*The reviewer states: "The manuscript attributes inversion differences to differences between EDGAR and Vulcan emissions. The authors should also compute and report the raw differences between these two emissions inventories before they are used in the inversion demo."*

**Response:** We do not ascribe the estimation error to differences between EDGAR and Vulcan. The primary source of estimation error is the lack of informative measurements (too few towers, and winds tend to blow ffCO2 emissions away from them). The differences between EDGAR and Vulcan are a small source of error. We will mention these errors (between annually averaged Vulcan and EDGAR) in Sec. 5.

*The reviewer states: "The synthetic observations used in the inversion, which are first introduced on page 1291 and later discussed on page 1295, should be described more clearly and in more detail. Were the elements of the sensitivity matrix **H** generated for*



*another problem and adapted for this manuscript or were they computed specifically for this paper? As a numerical verification test, do the sensitivities multiplied by the Vulcan fluxes equal the concentrations obtained from a single forward simulation using Vulcan (i.e. does y equal Hf as given in Eq. 5)? More information about the WRF setup would also be useful (What lateral boundary conditions were used to generate the winds? What physics packages options were used? and so on)."*

**Response:** The **H** matrices used in this paper were generated for two previous studies (Gourdji et al, 2010; Gourdji et al, 2012), which describe in great detail the gridding and the WRF settings used to construct them. We cite the papers on Pg 1291:21, and fail to see what repeating the same details would contribute to our paper. However, we will update the paper to explicitly mention the references where details on the calculation of H (meshes, models and settings) can be obtained.

The reviewer also asks "do the sensitivities multiplied by the Vulcan fluxes equal the concentrations obtained from a single forward simulation using Vulcan (i.e. does y equal Hf as given in Eq. 5)?" We presume he/she means "if the sensitivities multiplied by the estimated (NOT Vulcan) fluxes equal the concentrations obtained from a single forward simulation using Vulcan?" Yes, they do. This is equivalent to asking if the estimated fluxes reproduce the observations (y). The agreement is plotted for a few measurement towers in Fig. 9 (right).

*The reviewer states: "The authors analyze and display (Fig. 3) the statistics of non-zero wavelet coefficients. To help with visualization, it may also be useful to display maps of a few of the major features obtained from the wavelet decomposition."*

**Response:** We thank the reviewer for this excellent suggestion. We will do so in the revised version of the paper.

*The reviewer states: "On page 1288, line 13, the authors incorrectly associate static sources with emissions from highways. While it is true that highways are fixed, the traffic flow along them is not. CO2 emissions from traffic are usually categorized as* *mobile and non-stationary."*

**Response:** The reviewer is correct that emissions from highways are classified as mobile. Averaged over time, they appear as line sources whose strengths vary along the line. However, they do not move, and can be captured by the same set of wavelets. Hence we called them "static" sources.

We will add this clarification as a footnote in Sec. 3.2.

*The reviewer has some suggestions regarding rewording of some figure captions, in the interest of clarity*

**Response:** We will make the suggested changes

*The reviewer states: "Please add "et al" to the Friedlingstein reference on pages 1278 and 1303. Also, according to recent work (see Fig. 1a in Regnier et al, doi:10.1038/ngeo1830), fossil fuel emissions are not the largest net carbon flux at the atmosphere-surface interface. Please revise the second sentence in the Introduction accordingly."*

**Response:** Thank you for pointing out the error in the Friedlingstein reference. We are a little confused by the Regnier reference. Fig 1a therein clearly shows that fossil-fuel emissions are indeed the largest NET carbon exchange between land and atmosphere.

**References**

[Chatterjee et al, 2012] Chatterjee, A., Michalak, A. M., Mueller, K. L., Anderson, J. L., and Yadav, V.: Towards reliable ensemble Kalman filter estimates, J. Geophys. Res.-Atmos., 117, D22306, doi:10.1029/2012JD018176, 2012.

[Gourdji et al, 2010] Gourdji, S.M., Hirsch, A.I., Mueller, K.L., Yadav, V., Andrews, A.E., and Michalak, A.M.,: Regional-scale geostatistical inverse modeling of North America CO2 fluxes: a synthetic data study", Atmospheric Chemistry and Physics, 10:6551-6167, 2010.



[Gourdji et al, 2012] Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, doi:10.5194/bg-9-457-2012, 2012.

[Ray et al, 2013] Ray, J., Lee, J., Lefantzi, S., Yadav, V., Michalak, A. M., Bloemen-Waanders, B., and McKenna, S. A.: A multiresolution spatial parametrization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions, SAND Report SAND2013-2919, Sandia National Laboratories, Livermore, CA 94551-0969, available at: http://www.sandia.gov/ jairay (last access: 3 February 2014), 2013.

Interactive comment on Geosci. Model Dev. Discuss., 7, 1277, 2014.