Dear reviewer 1,

We are very grateful for your comments and suggestions, which have helped to improve our manuscript significantly. We have revised the manuscript accordingly, and the changes can be found in the track-changes file. The following is a point to point response to your comments and suggestions. Corresponding changes in the manuscript are also made available below at the appropriate places, if applicable.

Sincerely,

Benjamin Zanger and Jia Chen on behalf of all co-authors.

Anonymous Reviewer #1:

Line 57: What is a 3rd-level wavelet transform? If emissions on a N x N grid are subjected to Haar transform, one gets a transform hierarchy of log2(N) levels. In a 3rd-level transform, do you ignore all levels finer than 3 (which would give a very rough emission field) or do you only keep level 3?

The first level of the (Haar) wavelet transform decomposes the emission grid into high frequency components and low frequency components. The second level then acts on the low frequency part of the first level and decomposes it into high frequency components and low frequency components. A 3rd level wavelet transform therefore consists of the high frequencies of the first 3 levels and the remaining low frequency part of the 3rd level. For visualization, we refer to Fig. 1.1 in Mallat (1999) (a free version of the first 3 chapters can be found here: https://www.di.ens.fr/~mallat/papiers/WaveletTourChap1-2-3.pdf).
Thus, we do not ignore any levels, we only stop decomposing the low frequency part further.
We added a reference to Mallat (1999) in the text and referred to Fig. 1.1 for a visualization of a 3rd level wavelet transform:

Lines 183 ff:	*For an introduction to wavelets, we refer to Graps (1995) and Mallat (1999). For the visualization of a 3rd level DWT, see Fig. 1.1 in Mallat (1999).*

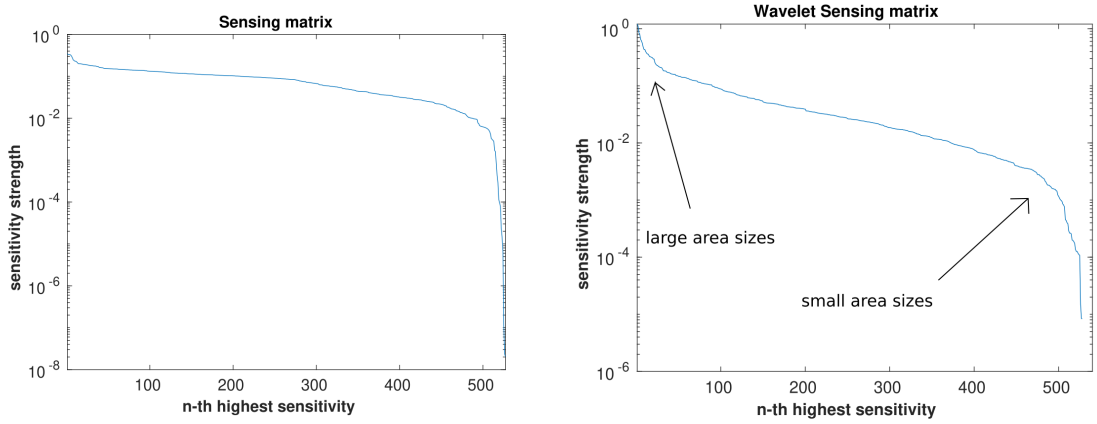
Line 390: "SR is good at localization and SR-in-wavelet-domain is not robust to noise". I see that empirically from the results, but can you explain why? Does it have anything to do with the 3rd-level transform? Or Haar wavelets, which are oscillatory? Or because you do not impose non-negativity on the estimated emissions?

Thank you for this question!

The reason why SR works well for localizing sources is as follows:

1. SR prefers sparse over smooth solutions. For the spatial domain, this means that preferably a single strong emitter is picked, while LS, which prefers smooth solutions, prefers to pick multiple small emitters to account for emissions. Therefore, SR is more capable of detecting point sources.
2. Compressed sensing (CS) property guarantees us the right answer for the emission location of the highest emitters. Even though the compressed sensing condition is not satisfied all the time, we still can have a “high probability” for the guarantees of CS for emission distributions which are similar/close to the true emissions.

Secondly, let us elaborate on the robustness of SR DWT towards noise. Please note, that the coefficients of the wavelet transformation have a huge difference in contribution to the measurement. To show this, we plotted the sensitivities of the measurement to the coefficients, given by the sum of the contributions of a coefficient i to each measurement j, $\sum_j \alpha_{ji}$, where α_{ji} is the entry of the j-th row and i-th column of the sensing matrix:



While for the sensing matrix in the spatial domain (left side), those sensitivities do not differ too much in magnitude (except for the right outermost sensitivities), the sensitivities of the wavelet sensing matrix possess much higher differences. Furthermore, the large sensitivities belong to coefficients of large spatial areas, while the low sensitivities belong to small spatial areas.

This huge difference in sensitivity values is not ideal for fulfilling CS conditions properties, especially for CS condition properties needed for the noisy case, which are more strict than those in the noiseless case.

Furthermore, coefficients which have a lower contribution to the measurement are also more prone to noise, so that in the noisy case there will be a higher estimation error in those coefficients.

We also want to highlight Ray's approach (Ray et al., 2015) to make the SR DWT more robust to noise. In their paper, they used prior knowledge so that all estimates are in the same order of magnitude; hence bringing all sensitivities to the same order of magnitude as well. They showed that this improves the performance of the estimation.
However, we chose not to do this, because the prior knowledge is not accessible for unknown emitters.

One could try other wavelets (other levels), but according to the explanation we provided above, this would not solve the problem. We think that other domain transformations than wavelets could improve the situation. However, it is not clear which domain transformations are better suited, since a transformation always changes sparsity of the signal as well as the properties of the resulting sensing matrix. It would be interesting to perform a systematic study of other domain transforms in a future study.

In Sec 4.2.2, 4.2.3 and 4.2.4, what are the wind coverage in the cities?

In Sec. 4.2.2, 4.2.3 and 4.2.4, the same wind fields are used. These wind fields have a wind coverage of about 143°. We added this information to the main text:

Line 246:	*In our work, we use artificial wind data. These wind fields used, except in Sec. 4.2.1, possess a wind coverage of ~143°. All of the footprints generated by those wind fields are available at https://doi.org/10.5281/zenodo.5901298 (Zanger et al., 2022a).*

Please keep in mind that this number is not strictly comparable to the wind coverage used in Sec 4.2.1, because in Sec 4.2.1 a static Gaussian plume model is used, while the other sections use a non-static Gaussian plume model. Please also see the details why we are using a static model for Sec 4.2.1 in the Appendix B:

Lines 460 ff:	*For Sec. 4.2.1, we use a static Gaussian plume model to generate footprints for different wind coverages. This allows us to ignore additional effects from the dynamical case. For example, in the dynamic case the change in the wind direction produces spiral looking footprints. These footprints have a different shape compared to the footprints generated using a static Gaussian plume model . By using a static model, the footprints are more predictable and systematic, which makes the results of different wind coverages more comparable to each other.*

Are the wind velocities the same or are the footprints stretched to cover the cities which are of different sizes?

The footprints are stretched and the velocities change. Please see Appendix B:

Lines 457 ff:	*The footprints are then scaled to the domain size of the city. Because of the scaling, the wind speed and diffusion changes for each city. Nevertheless, we think that this approach makes the reconstruction performances for different cities more comparable than using different footprints for every city.*

Fig 7 and Fig. D1: The x-axis has bins called "0 - 0.4%" ; the label on the x-axis is "Highest emission in inventory". What does it mean? Do you rank-order the gridded emissions, and bin them in unequal bins? How were the bin cutoffs decided? Should that be cast in terms of percentiles of the gridded emissions? Or is it the percentiles of their ranks? Some explanation of what this axis is, in the text or captions, would be helpful

Thank you for this comment. We added an explanation in the text:

Lines 331 ff:	*The x-axes show bins of emission grids ranked by their emission strength. For example, the first bin (0 - 0.4%) contains the first 0.4% highest emission grids while the last bin in the lower panels of Fig. 7 contains emission grids which are among the 36.2 - 53.2% highest emission grids. Since the emissions are highly compressible (first highest emissions have a huge contribution to the total emissions) we use a logarithmic scale for the bins.*

The English in Sec 3, 4 and 5 needs to be improved. [...]

We contacted the language center at TUM and a lecturer there (native speaker) has gone through our text. The language of our paper has been improved. Please see the track-change version to see our changes.

The is a lot of use of the emissions being "good compressible" or "well compressible". I think one could simply use "compressible"[...] Also, in the text and tables, one sees emissions "not good compressible" or "not well compressible". What about non-compressible?

Thank you for the suggestions. We adapted the wording to “good compressible” and “not good compressible”. We think that “non-compressible” would be too strict, since in fact those emissions are still compressible ($\sigma_{10\%}$ for Paris is 0.217).

The is also use of the terms "amount of measurements" and "less measurements". Measurements in this paper are counts. What about using "number of measurements" and "fewer measurements"?

Thank you for this comment, we adapted it in the paper.

Line 332: The authors talk about estimating the measurement counts needed to achieve "descent results". That is rather casual. What about "acceptable results"? Also, what would constitute an "acceptable result"? Please clarify

Thank you for this comment. We wanted to find a general relation between the number of measurements and the accuracy. We changed the wording accordingly:

Lines 352 ff:	*However, the question of how many measurements m are needed to produce results of a certain accuracy remains unanswered.*

Fig 7: The caption mentions "how many percents ...." This is hard to decipher. What about explaining the figure in the text where there is no shortage of space?

Thank you for this suggestion. We removed the explanation from the caption in Fig. 7 and improved the explanation of the qualitative and quantitative measure in the text (please see line 330 ff).

References:

Aster, R.C., Borchers, B. and Thurber, C.H., 2018. *Parameter estimation and inverse problems*. Elsevier.

Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics.

Ray, J., Lee, J., Yadav, V., Lefantzi, S., Michalak, A., and Bloemen Waanders, B. v., 2015. *A sparse reconstruction method for the estimation of multi-resolution emission fields via atmospheric inversion*. Geoscientific Model Development, 8, 1259–1273. https://doi.org/10.5194/gmd-8-1259-2015.

Ray, J., Yadav, V., Michalak, A. M., van Bloemen Waanders, B., and McKenna, S. A.: A multiresolution spatial parameterization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions, Geosci. Model Dev., 7, 1901–1918, https://doi.org/10.5194/gmd-7-1901-2014, 2014.

Hase, N., Miller, S. M., Maaß, P., Notholt, J., Palm, M., and Warneke, T.: Atmospheric inverse modeling via sparse reconstruction, Geosci. Model Dev., 10, 3695–3713, https://doi.org/10.5194/gmd-10-3695-2017, 2017.

Jones, T.S., Franklin, J.E., Chen, J., Dietrich, F., Hajny, K.D., Paetzold, J.C., Wenzel, A., Gately, C., Gottlieb, E., Parker, H. and Dubey, M., 2021. Assessing urban methane emissions using column-observing portable Fourier transform infrared (FTIR) spectrometers and a novel Bayesian inversion framework. *Atmospheric Chemistry and Physics*, *21*(17), pp.13131-13147.

Lauvaux, T., Miles, N. L., Deng, A., Richardson, S. J., Cambaliza, M. O., Davis, K. J., Gaudet, B., Gurney, K. R., Huang, J., O'Keefe, D., Song, Y., Karion, A., Oda, T., Patarasuk, R., Razlivanov, I., Sarmiento, D., Shepson, P., Sweeney, C., Turnbull, J., & Wu, K. (2016). High-resolution atmospheric inversion of urban CO_2 emissions during the dormant season of the Indianapolis Flux Experiment (INFLUX). *Journal of geophysical research. Atmospheres : JGR*, *121*(10), 5213–5236. https://doi.org/10.1002/2015JD024473.

Mallat, S., 1999. *A wavelet tour of signal processing*. Elsevier.