Reply to comments of Reviewer 1, Sharon Gourdji

We want to thank Sharon Gourdji for the review of our manuscript and her helpful and inspiring comments and suggestions for improvement. Our replies are marked in blue.

This is an excellent study, which is very well-written and clear and with very useful implications for atmospheric inverse modeling, particularly in urban areas. A few small questions and concerns for clarification should be addressed before final publication:

- To use the VSI approach, does one also need an inventory containing the vertical height profiles of all point source emissions? This would be great to have, but in practice, this may currently exist in Europe only. (For example, I don't believe that the Vulcan product for the USA contains height of emissions sources now, nor other products like FFDAS or ODIAC.)

  Yes, that is correct. The VSI approach is based on vertical height profiles of the point source emissions. Ideally, one would have an inventory for the effective emission heights of all individual point sources (in the surroundings of the measurement site), which consider the actual stack heights plus plume rise. In our study we have used source sector-specific (average) emission height profiles from TNO, which are representative for Europe. While checking the Vulcan, FFDAS or ODIAC inventories, we didn't find similar emission height profiles for the USA so far. However, applying the TNO height profiles also in the USA could already lead to a first improvement compared to releasing all point source emissions from ground. But we agree that such an inventory containing the vertical point source emission heights for the whole globe would be important to have. We have added a sentence to our conclusions in the revised manuscript (lines 583ff).

- What are the additional computational requirements of the VSI relative to the SSI approach? Also, how would one go about creating a footprint from a single tower with a mix of the VSI approach for nearby point sources and the SSI for farther-away emissions sources? How would one do that practically with the WRF-STILT framework?

  In STILT, the trajectory information of the released particles is saved in RData files. In the SSI approach the number of particles below the half of the boundary layer height is counted in each grid cell and is then used to weight the surface fluxes in the respective grid cell. In the VSI approach we used vertical emission height profiles with seven different height intervals (see Fig. 3b in our manuscript). This means that in the VSI approach the number of particles in seven (instead of one) height intervals must be counted, which results in additional computational costs. That's why we recommend using the VSI approach only for nearby point sources (where it matters). For this, one could first set the nearby point sources to zero and calculate the contributions from the area sources and the farther-away point sources with the standard SSI approach. Then, one could use our R script provided in https://doi.org/10.5281/zenodo.5911518 to calculate separately the volume source influences for the seven height intervals in the near field of the station to get the contributions from the nearby point sources.

- I was left wondering what are the relative impacts of mixing assumptions versus PBL height errors when using night-time measurements. Could you include a small theoretical example to demonstrate the impact of realistic mixing height errors with the VSI approach and nighttime observations?

  The SSI approach assumes that the air masses within the bottom half of the PBL height (h_{PBLH}) are well mixed. Thus, the surface emissions are weighted with $1/\left(\frac{1}{2}h_{PBLH}\right)$. If the nocturnal PBL heights, and so the strength of the mixing, are overestimated (as it is the case for the ECMWF-derived mixing heights in Gerbig et al., 2008) this would result in an underestimation of the contribution from those surface emissions. For daytime situations, Gerbig et al. (2008) showed the propagated impact of those mixing height uncertainties on the modelled CO_2 mixing ratios by introducing (besides the turbulent winds) a second stochastic process, which rescales the footprint (i.e. the sensitivity to surface fluxes) and thus considers the mixing height uncertainty. Their ansatz would be more difficult for nighttime situations, which have much larger mixing heights biases.

  The VSI approach assumes well-mixed conditions in each of the seven fixed height intervals of the TNO vertical emission height profiles. But it also depends on the mixing to be correct. Imagine a nighttime situation with a too shallow modelled PBL height. This would imply that mixing is insufficient, and the tracer increments are overestimated within the PBL. If a power plant plume is within the PBL, also the VSI approach will yield too large power plant CO_2 contributions. Thus, also the VSI approach suffers from an incorrect representation of the PBL height.

Other small comments:

- Abstract, line 28: "to fall below 0.1 ppm" à during day or nighttime or both?

Thank you for this hint! This corresponds to situations with PBL heights smaller than 500 m (see Fig. 7c in the manuscript). We have specified it in the revised manuscript (line 30).

- Page 3, line 61: "nighttime situations showed a relative bias of more than 50%" -> in which direction is this bias?

This positive bias means that the ECMWF-derived mixing heights are larger than the mixing heights estimated from radiosonde data. We have clarified this in the revised manuscript (lines 64f).

- Is 100 particles enough for this study? I assume you would get the same results using 500 particles or more, but it might be worth a small check for sensitivity here.

Thank your for pointing us to this sensitivity study. To check this, we calculated the ffCO_2 contributions from the 12 pseudo power plants for the 30 m high receptor site Heidelberg by having released 500 particles. To save computational power, we considered only one month (January) in 2019. Fig. 1 shows the mean VSI minus SSI contribution differences (as Fig. 7c in the manuscript) for January 2019. The differences between 100 and 500 released particles are usually only very small. So, we can argue that 500 released particles would not change the overall picture and 100 particles are enough.

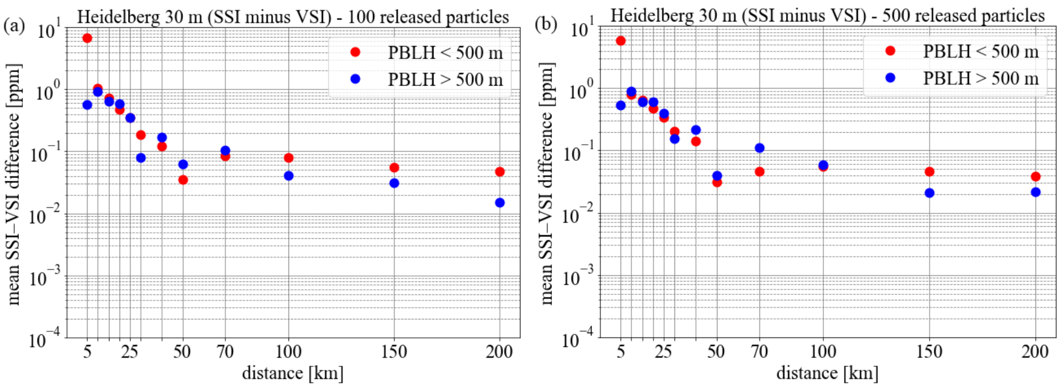


Figure 1: Mean SSI minus VSI difference in ffCO_2 contributions from pseudo power plants, which were placed at distances between 5 and 200 km from the observation site Heidelberg at 30 m. The time period is January 2019. Panel (a) shows the results if 100 particles are released each hour and panel (b) shows the results for 500 released particles per hour. For further details, please refer to the caption of Fig. 7 in the manuscript.

- Figure 1: This is a nice map, although it's a bit hard to see the country outlines and the actual distance from point sources to measurement locations. Consider additionally including a histogram or barplot of distance to nearest point source(s) for each measurement location? To what extent do existing measurement locations follow the ICOS recommendations to stay at least 40 km away from strong anthropogenic sources? (And how did ICOS derive this recommendation in the first place?)

The ICOS recommendation (https://doi.org/10.18160/GK28-2188) of 40 km distance between the station and strong anthropogenic sources was chosen to "ensure that observations can be represented in atmospheric transport models with spatial resolution of around 10-20 km". As can be seen in Fig. 1 in our manuscript, not all ICOS stations fulfill this recommendation. In those cases, “a footprint and representativeness study should be performed”. We have included a table, which sums up the point source emissions in a 50 km x 50 km box around the most affected ICOS stations in Fig. 1 in our revised manuscript.

- Page 3, line 61: “a relative standard deviation of about 40%” in mixing height, or errors in mixing height? Also, please clarify for following sentence.

  This means the relative standard deviation of the difference between the ECMWF-derived mixing heights $z_i(ECMWF)$ and the radiosonde estimates $z_i(RS)$, i.e. $\frac{std(z_i(ECMWF)-z_i(RS))}{\langle z_i(RS)\rangle}$. We have clarified this in the revised manuscript (line 63).

- Page 4, lines 62-64: if the uncertainty in daytime mixing height translates into uncertainties of ~3 ppm and 30% of the simulated biogenic signal during summer, what does this tell you about nighttime uncertainties? Just complete the thought here. Also, in reference to the previous comment, this article develops a better approach to dealing with mixing assumptions in STILT but doesn’t address or improve mixing height errors. So, what is the relative impact of these two types of errors on both daytime and nighttime measurements?

  The authors of this cited study (Gerbig et al., 2008) only investigated the propagation of uncertainties in mixing heights into mixing ratios during daytime situations. They state that nighttime mixing heights have much larger uncertainties and biases, which makes them much more difficult to consider in STILT. Nevertheless, we would expect much larger uncertainties in the mixing ratios during nighttime.

  We want to focus in our study on the difference between the mixing assumptions in the SSI and VSI approach and the improvements, which can be achieved when just using the optimized mixing assumptions of the VSI approach instead of the standard SSI approach. These improvements of the VSI approach can be seen for example in Fig. 4 and 7 in our manuscript, which also distinguish between daytime and nighttime situations.

- Page 5, lines 88-95: this is a great explanation for why the ability to use nighttime observations in inversions would be very useful and is a prime rationale for your study. I suggest adding a statement to this effect in the abstract about why this work would be very helpful for other researchers for the reasons laid out here.

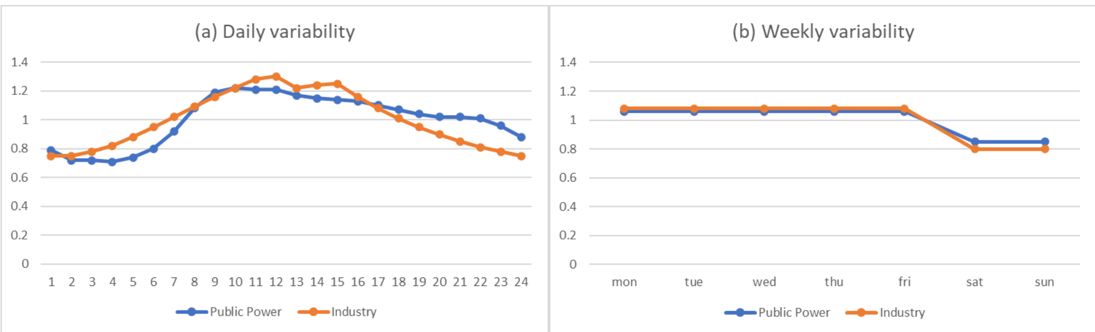
  Thank you for this suggestion! We have added a sentence in the abstract (lines 24f).

- Page 7, line 141: please spell out what TNO stands for, for those not familiar. In general, it might be nice to describe this inventory in a bit more detail for non-European audiences, especially because you are relying on the height profiles in this inventory to implement your VSI approach. Also, for the differing spatial resolutions between Germany and the rest of Europe, is this how it's produced in Super et al, 2020, or do you aggregate emissions yourself for the purposes of this study?

  TNO stands for the Netherlands Organisation for Applied Scientific Research. We have added a bit more information about the TNO inventories in the revised manuscript (lines 164ff). There are two inventories with different horizontal resolutions available: one high-resolution inventory (1/60° x 1/120°) for Germany and its surroundings and one low-resolution inventory (0.1° x 0.05°) for most of Europe. In our study, we have just nested these two inventories.

- Page 9, lines 189-191: How would time-varying emissions affect these TNO height profiles (e.g. with some emission sources starting and stopping again)? Also, do the TNO height profiles shown in Figure 3b represent sector-specific averages? Or are heights included for individual point source locations as well?

  We considered in our study time-varying point source emissions. For this we used the source sector-specific diurnal, weekly and seasonal temporal profiles from TNO (see Fig. 2 below). Yes, the height profiles shown in Fig. 3b in the manuscript represent sector-specific averages, which should be representative for Europe. Unfortunately, there are no stack heights for individual point source locations available. We have added a few sentences in our revised manuscript (lines 226ff).



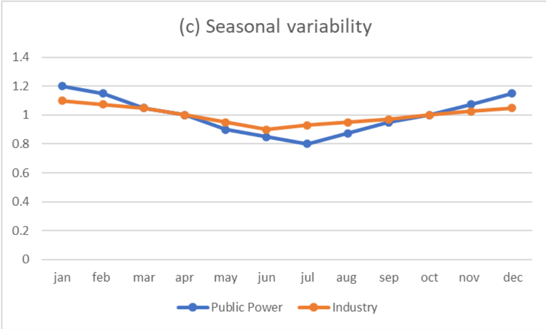


Figure 2: TNO temporal profiles for the public power (blue) and industry (orange) sectors: (a) Daily variability, (b) weekly variability and (c) seasonal variability.

- Figure 7f: it is nice to have a consistent y-scale with the subplot above (7c), but it's a bit confusing with the arrows and negative values. Consider changing the y-scale to include negative values for both.

  We agree, that Fig. 7f is a bit confusing with the arrows, which indicate the negative values. However, we decided to have a logarithmic y-scale, so that the (small) differences in the contributions from the farther-away point sources are easier to see. And we think that interrupting the logarithmic y-scale and showing the negative values in a linear scale would maybe be more confusing.

- Page 11, lines 247-250: is there a physical reason why these errors would be lower in summer than in winter? I think this could be interesting for the reader.

  We suppose that the larger errors in winter originate from synoptic events with suppressed atmospheric mixing. These synoptic events occur mainly in winter and last typically several days. This leads in winter also to afternoon situations with suppressed atmospheric mixing and worse model performance. We have included this information in our revised manuscript (lines 292ff).

- Page 16, lines 349-357: it's a bit hard to follow the argument here. For example, the statement "However, the power plant within a 5 km radius yields lower ffCO2 contributions during stable PBLH < 500 m conditions than during PBLH > 500 m situations" à is this referring to the VSI approach? And the opposite is true for the SSI approach? It sounds like it from the statement in the next paragraph that "the SSI approach simulates on average almost 5 ppm larger ffCO2 contributions than the VSI approach for the closest power plant during stable conditions." This is just for the 30-m tower, correct, and not the 200-m tower? Also, the possible explanation mentioned for the VSI behavior, is this in the model, in reality or both?

  Yes, the statement "However, the power plant within a 5 km radius yields lower $ffCO_2$ contributions during stable PBLH < 500 m conditions than during PBLH > 500 m situations" refers to the VSI approach. The suppressed mixing during PBLH < 500 m situations hinders the power plant emissions to get mixed down to the measurement site at 30 m height within the time the air mass needs to cover the 5 km distance.

In the SSI approach, the ffCO_2 contributions from the 5 km distant power plant are much larger during stable PBLH < 500 m conditions compared to PBLH > 500 m conditions, since during PBLH < 500 m conditions all emissions from the 5 km distant power plant are mixed into a very shallow layer, which is smaller than 250 m.

These two circumstances cause the 5 ppm SSI minus VSI ffCO_2 contribution difference for the closest 5 km distant power plant. These conclusions refer all to the 30 m high measurement site. For the 200 m high site, the ffCO_2 contributions from the 5 km distant power plant are less different in the SSI compared to the VSI approach (here the SSI approach leads to about 0.6 ppm larger ffCO_2 contributions than the VSI approach).

The explanation about the behaviour of the VSI approach refers to the model world. It is difficult to assess whether this explanation is true for the reality, too. However, we could show that the VSI approach leads to a better agreement with our measurements during stable conditions (see Fig. 4 in the manuscript).

We have tried to make our argumentation clearer in the revised manuscript (lines 401ff).

- Page 16, line 362: what are the typical inlet heights for ICOS tower stations?

  The inlet heights of ICOS towers range between about 30 and 250 m. Most of the towers have more than one inlet height. We have added that to the revised manuscript (line 423).

- Page 20, line 461: "inaccurate representation"

  Thank you! We have corrected it (line 528).

- References: please use better indentation to distinguish each reference.

  We have applied the indentations.

Reply to comments of Reviewer 2, Anonymous Referee

We want to thank the anonymous referee for the review of our manuscript and the helpful and inspiring comments and suggestions for improvement. Our replies are marked in blue.

General

The manuscript "Effects of point source emission heights in WRF–STILT: a step towards exploiting nocturnal observations in models" by Fabian Maier and co-workers describes the use of vertical emission profiles for point sources in the time-reversed application of the Lagrangian particle dispersion model WRF-STILT. The authors convincingly show that ignoring vertical emission profiles and assuming surface emissions only, as done in many applications of LPDMs, may lead to serious biases for sites influenced by elevated point sources. The study is an important contribution for regional-scale inverse modelling of greenhouse gas emissions as it directly address shortcomings that can easily be remedied without major modifications on the transport description in LPDMs. The manuscript is well organized and written, methods and results are presented in an appropriate manner. Some minor concerns and considerations remain that I would like the authors to consider in a revised version of the manuscript.

Major comment

The way the introduction (L84-95) and section 4.2 states the problem of using nighttime observations tends to suggest that including vertical emission profiles ('volume source approach') alone may enable modelers to use such observations in inverse modelling studies. However, an important prerequisite, and this is only mentioned rather weakly and hidden (e.g., citation of Geels et al, 2007), is the models ability to realistically reproduce nighttime stable boundary layers and the erosion of these stable layers in the morning hours. Analysis of simulated diurnal cycles and, where possible, vertical gradients against observations are inevitable before assimilating nighttime observations. This fact should be highlighted with more emphasize (introduction, section 4.2, and conclusion). The vertical emission profiles will not solve anything if, for example, the nighttime stable layers are only formed to weakly.

We fully agree, the model's ability to simulate nocturnal stable boundary layers appropriately is the important prerequisite for incorporating nighttime observations in inverse modelling studies. We could, however, show that in case of an urban site the poor model performance during night is not only caused by the limitations in modelling nocturnal boundary layers but also by an incorrect representation of point source emissions. Thus, considering vertical emission profiles for point source emissions is an important (first) step for using nighttime observations in inversion studies. However, this is not the only step to go for exploiting nighttime observations in models. To realistically reproduce nocturnal boundary layers would be a further (next) step as this might be more difficult than considering vertical point source emission height profiles. We have emphasized this more in the revised manuscript (lines 109f, 479ff, 581ff).

Minor comments

L55 & : This is specific for STILT. Other LPDMs (for example NAME, FLEXPART) use fixed sampling heights that do not vary with the boundary layer height.

Thank you for this remark. We have clarified this in the revised manuscript (line 57).

L66f: There is another issue with point sources in time-reversed LPDM simulations. The source sensitivities (footprints) are usually stored on a horizontal grid with limited resolution. This adds to model uncertainties as well, since the limited resolution of the footprints may lead to false attribution of point source emissions in cases where a higher resolution footprint may actually have missed the point source. Since STILT is using an adaptive output grid that becomes coarser with distance to the receptor location, this problem may be more important for distant sources, but also for near sources and an inappropriate output resolution false attribution may happen. I think this issue deserves mentioning at this point.

We fully agree, this point is missing in our manuscript. We have included it in the revised version (lines 69ff).

L89-93: Another important point is that the average daytime footprint will differ significantly from the average nighttime footprint. Especially for tall towers the nighttime footprint is usually larger, sampling more distant sources, whereas the daytime (convective) footprint is often dominated by more local sources. Similar to point 2 this may lead to sampling of different source mixtures. The use of nighttime data would certainly extent the 'field of view' of tall tower sites in any inverse modelling study. One requisite is however that the diurnal cycle of boundary layer heights and mixing are captured correctly in the LPDM (point source representation or not; see main comment above).

Thank you for this additional important point; we have added it in our revised manuscript (lines 100ff).

L119-121: Could these point sources be highlighted in the map? Maybe panel b should be zoomed even further, in order to clearly see the location of these four sources relative to the site.

We have zoomed Fig. 2b a little bit further and labelled the four closest point sources around Heidelberg in the revised manuscript.

Figure 2: I find the depiction of model/emission domains a bit confusing. There seem to be two different resolutions and domains for WRF and TNO. However, the figure

somehow can be read as if there are 3 WRF domains. Maybe just indicate the higher resolution nests on the left (yellow and black rectangles, as is, but label them only with TNO 1km and WRF 2 km, respectively). Then produce a high resolution zoom that is smaller than the WRF high resolution domain in order to show the nearby point sources (see last comment).

We have implemented your suggestions and have given an additional explanation in the caption of Fig. 2 in our revised manuscript.

L138: One hundred released air parcels per hour seems to be very small. How can one statistically resolve any vertical gradients with these? The VSI approach requires five different layers as applied here, the lowest two with a thickness of only 100 m. How can you be sure that one hundred air parcels can robustly represent any vertical gradient in such thin layers? The previous h_pbl/2 method may have allowed for such small air parcel numbers because no vertical gradient below h_pbl/2 had to be represented. The improved results with the VSI approach seem to justify the small number of air parcels, but they may merely result from improved separation between stable PBL and lower free troposphere at night. The lack of improvements during daytime (from SSI to VSI) may indicate that residence time gradients during the day are not well represented by the limited number of air parcels.

We agree, a larger number of released particles would increase the statistics and may also lead to a better resolution of the vertical gradients, especially during situations with large planetary boundary layer heights (PBLHs). To investigate the effect of a larger number of released particles, we calculated the $ffCO_2$ contributions from the 12 pseudo power plants for the 30 m high receptor site Heidelberg by having released 500 particles. To save computational power, we considered only one month – January – in 2019. Fig. 3 shows the mean VSI minus SSI contribution differences (as Fig. 7c in the manuscript) for January 2019. The differences between 100 and 500 released particles are usually only very small. During situations with large boundary layer heights (PBLH > 500 m), the mean SSI minus VSI difference in the $ffCO_2$ contributions from the pseudo power plants differs no more than 0.08 ppm for the two cases with the 100 and 500 released particles, respectively. Although the increased number of released particles can lead to some deviations, it should not change the overall picture, even for situations with large mixing heights. Based on these findings, we only released 100 particles to save computational power. We have added a note about this in the revised manuscript (lines 160ff).

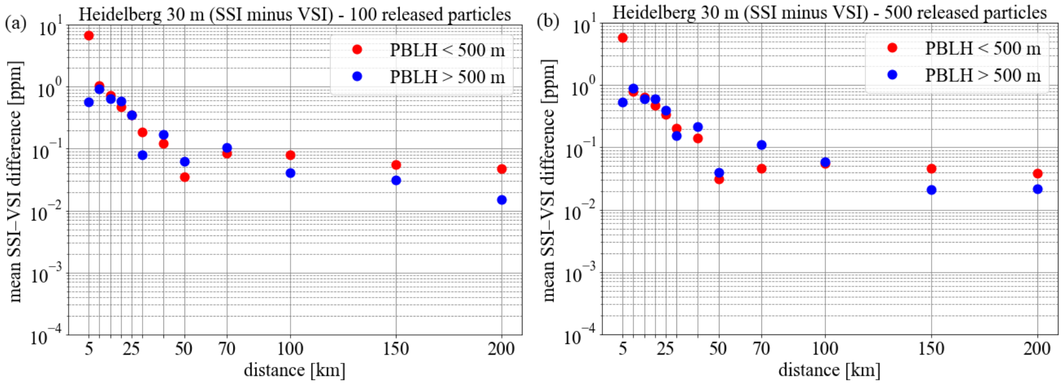


Figure 3: Mean SSI minus VSI difference in $ffCO_2$ contributions from pseudo power plants, which were placed at distances between 5 and 200 km from the observation site Heidelberg at 30 m. The time period is January 2019. Panel (a) shows the results if 100 particles are released each hour and panel (b) shows the results for 500 released particles per hour. For further details, please refer to the caption of Fig. 7 in the manuscript.

L139: Considering the outer WRF domain, this backward integration time seems to be rather short. What is the reason for the selected 3 days? How frequently do particles remain within the domain after 72 hours?

This is true, the rather short backward integration time of only 3 days results in 50.3 % of the trajectories remaining within the domain in 2019. Fig. 4 shows the distribution of the 365*24*100 trajectories' endpoints in 2019. The trajectories with endpoints within the domain end mainly over Central and Western Europe and over the Atlantic Ocean. To estimate the influence of the backward integration time, we calculated 10 days-backtrajectories for a winter and summer month in 2019 and compared the modelled SSI $ffCO_2$ concentrations with those obtained with the 3 days-backtrajectories. Here, we used the SSI approach, since we have shown that it leads to similar far-field (well-mixed) $ffCO_2$ contributions as the VSI approach. In January 2019, when 28.9 % of the 3 days-backtrajectories end within the domain, the 10 days-backtrajectories lead in the median to 0.09 ppm larger $ffCO_2$ contributions than the 3 days-backtrajectories. In July 2019, 65.7 % of the 3 days-backtrajectories end within the domain and the median $ffCO_2$ difference between 10 days- and 3 days-backtrajectories is 0.22 ppm. Thus, the usage of 3 days-backtrajectories instead of 10 days-backtrajectories results in a slight underestimation of the modelled $ffCO_2$ concentrations (Fig.4 in the manuscript), which affects the SSI and VSI approach similarly. However, considering the potential bias introduced by assuming that Mace Head is a representative $ffCO_2$ background for the whole domain border (i.e., ignoring for example a continental background for trajectories ending at the eastern domain border) this 10 days- vs. 3 days-backtrajectories bias seems to be rather small. To save computational power for the high-resolution simulations we therefore used a backward integration time of only 3 days. We have included a note about this sensitivity study in the revised manuscript (lines 160ff).

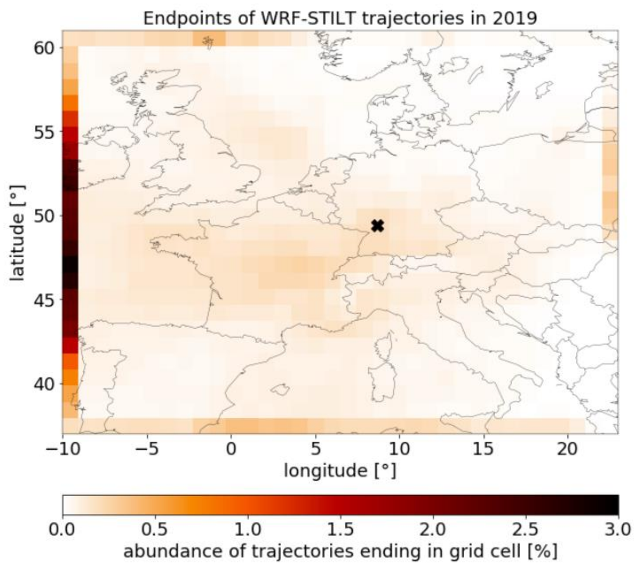


Figure 4: Abundance distribution for the endpoints of the WRF-STILT 3 days-backtrajectories from Heidelberg in 2019. At each hour 100 particles were released in STILT. If the trajectory leaves the shown domain, its endpoint is defined as the grid cell where the trajectory leaves the domain the first time. The station Heidelberg (HEI) is indicated with a black cross.

L139: There is no information here about the output resolution of STILT. Was this identical to the input resolution of WRF (2 km) or to that of the TNO emissions (1 km). As far as I know STILT output resolution varies with distance from the release location. What was the typical output resolution at distances covered by the synthetic source experiment?

Yes, the STILT output resolution is dynamically increased with distance from the release location. In the near-field of the release point the STILT footprint has a resolution of 1km (equal to the resolution of the TNO emissions). Depending on the spread of the individual back-trajectories with time the footprint resolution is reduced. The maximum coarsening factor for the resolution is 32. Thus, during situations with a large spread of the individual back-trajectories the footprint resolution can be 32 times lower at the distant pseudo power plants compared to the closest power plants. This can also be seen in Fig. 6 in the manuscript, which shows the aggregated hourly footprints in 2019.

L157: As mentioned before: other LPDMs use a fixed sampling heights in the order of 50 m to 100 m. A smaller sampling height actually assures that the assumption of instant vertical mixing is met. However, it also may require the use of larger particle ensembles in order to sufficiently represent particle distributions in more shallow layers.

We have made it clearer in the revised manuscript, that this h_pbl/2 mixing height assumption is specific to STILT (line 183).

L173: This may need further explanation. Wouldn't h' depend on may factors like wind speed, stability, etc. Is this 'effective mixing depth' used in any LPDM? Maybe this should be discussed in the previous section when STILT's sampling height is introduced. It seems to be unrelated to the 'volume source' approach.

The concept of the effective mixing depth h' has been introduced in Fasoli et al. (2018) by using STILT. Often the advective timescale is too short, so that emissions in the so-called hyper near field (HNF) of the receptor site are not fully mixed within the bottom half of the boundary layer. Since this is not considered in STILT, contributions from HNF emissions are underestimated in STILT. For this, the authors introduced the effective mixing depth h', which increases in the HNF with distance from the receptor site. Outside the HNF, h' is equal to the half of the PBLH. The growth of the effective emission depth h' depends on the meteorological conditions, i.e. the standard deviation in vertical velocities and the Lagrangian decorrelation timescale, which determines if the movement of the released particles in STILT behaves like a random walk or like advection by mean wind (Lin et al., 2003). We have shifted this discussion about h' to the previous section and added a bit more information in the revised manuscript (lines 194ff).

Section 2.2.2: This being GMD, I have a technical question: How are the height profiles implemented in STILT? Are these fixed levels according to the TNO suggestions or is the user able to specify them.

The code is such, that the heights of the individual layers are fixed throughout the whole model domain, and we assume one emission height profile for all point sources of the same source sector. However, the user can change the value of these layer heights and it is also possible to add more (fixed) layers accordingly.

L214f: This baseline assumption is most likely not very accurate for the eastern domain border and easterly advection. A note of caution should be added here.

This is certainly true; we have added a note of caution in our revised manuscript (lines 255ff).

L244 ("The standard deviation ..."): What does this aim at? Put observed variability and unexplained variability into perspective? Shouldn't this be done by simply giving the coefficient of determination of linear regression or by an analysis of variance?

We calculated the standard deviation to give a measure of the scattering between observed and modelled ffCO_2 (assuming a Gaussian distribution). Since the standard deviation is dependent on the magnitude of the observed ffCO_2 concentration, it might not be meaningful here. We agree that it is better to have a measure that gives a relation between observed variability and unexplained variability to be independent of the magnitude of the observed ffCO_2 concentrations. For this, we have calculated the proposed coefficient of determination of linear regression in the revised manuscript (lines 286ff, 318ff).

L243ff: The use of RMSD may be a bit misleading as it contains the bias as well. A bias-corrected (centered) RMSD (CRMSD) would allow analysing if the representation of variability beyond the bias was improved or not (CRMSD = sqrt(RMSD^2 – BIAS^2)). A quick check of the values in Fig4 suggests that the VSI approach mostly improves the bias but not the representation of variability.

We think that the RMSD is an appropriate measure here since it includes the bias between observed and modelled $ffCO_2$. The SSI vs. VSI comparison for nighttime samples nicely shows that, both, the bias and so the RMSD is reduced in case of the VSI approach. To check the CRMSD is a good idea. During nighttime it turns out, that the SSI approach leads to a CRMSD of 4.2 ppm (RMSD = 6.3 ppm) and the VSI approach leads to a CRMSD of 3.4 ppm (RMSD = 3.4 ppm). Whereas the RMSD is reduced by 46 % in the VSI approach compared to the SSI approach, the CRMSD is only reduced by 19 %. We have added this analysis to the revised manuscript (lines 320ff).

Figure 4, caption: 'standard error of the mean': Not clear which mean this refers to. That of the observations? Why give the standard error? The standard deviation could be more easily compared to RMSD.

This standard error of the mean refers to the standard deviation divided by sqrt(N), with N being the number of samples. After assuring ourselves that the observed minus modelled $ffCO_2$ differences are normally distributed in a reasonable manner we decided to show this standard error as an uncertainty of the mean bias. The prerequisite for dividing by sqrt(N) is that each two-week integral samples the identical meteorological situation. This is obviously not true. But we argue that the model tries to consider the different meteorological situations, so that we can compare the different samples with each other. That's why we find that this standard error should be an appropriate measure for the error of the mean bias.

Figure 6: Are these surface footprints or obtained from the VSI approach for a specific height?

We used the SSI approach to calculate these footprints since we just wanted to know the main catchment area of Heidelberg for the placement of the pseudo power plants. We have specified this in the caption of Fig. 6 and in the revised text (line 358).

L315: Which times does this refer to? Hours of release at the receptor or hours of the backward calculation? One particle back-trajectory seems to be a rather poor sample size for a LPDM. How can this be justified.

You are right, this is a bit confusing. We selected at the observation site Heidelberg the hours with the volume source influence matrix for a height range between 0 and 1106 m a.g.l. being larger than zero in all of the 12 pseudo power plant grid cells. This was done to have at the end for each of the 12 pseudo power plants an identical number of events (with nonzero contributions) for which we can calculate the mean SSI minus VSI contribution difference. This makes the data points in Fig. 7 for the different power plants more comparable among each other. We have tried to make this a bit clearer in the revised manuscript (lines 366ff).

L318: Are these the PBLH regimes at the time of arrival at the receptor? They may not be representative for the whole transport duration and the considered domain. Travelling times from the furthest power plant were probably larger than 8 hours. So arriving at nighttime in Heidelberg could mean that the power plant plume was still well-mixed over a large boundary layer during the previous day. Could be one reason why differences between the two PBL regimes seem to become smaller again for the largest power plant distance.

Yes, the PBLH regimes are at the time of arrival in Heidelberg. This is an interesting observation; we have included it in our revised manuscript (lines 418ff).

Figure 7: Wouldn't it make sense to show the differences as relative differences (e.g. 2(A-B)/(A+B))? The different logarithmic axis (compared to the individual SSI and VSI plots) make it difficult to judge if the differences are important. It would also allow for a more detailed discussion in the text.

We have changed the different logarithmic axes, so that all panels of Fig. 7 have the same logarithmic axis. However, we decided to show the absolute differences in Fig. 7c instead of relative differences. The main goal of Fig. 7c is to reveal for which power plants the VSI approach should be used instead of the SSI approach. We think that it is in the case of absolute differences more intuitive to assess if the difference (in ppm) between VSI and SSI approach is relevant and matters. Moreover, this difference scales linearly with the emission strength of the power plants. So, one can directly estimate the ppm difference between SSI and VSI approach for different power plant emissions and decide, which approach should be used. In case of relative differences, this difference must first be compared to Fig. 7a and 7b to assess if the SSI minus VSI difference matters for the contributions of a certain power plant. However, we also have included some discussion in relative terms in our revised manuscript (lines 426ff, see our answer to your comment after the next comment).

L334 and elsewhere: Since sub-panels of Figure 7 are labeled, please refer to them as a,b,c in the text as well.

Thanks for this remark. We have changed this in the revised manuscript.

L363f ('The SSI approach ...'): I don't see this. In Fig 7f (SSI-VSI) most differences are positive, exceptions being power plants at 15 and 20 km, but those are not the closest. Discussion in relative terms would be helpful. See comment on Figure 7.

Here, we wanted to say that the SSI-VSI differences for nearby power plants are smaller in the case of the 200 m air inlet compared to the 30 m air inlet. Take for example a look at the closest 5 km distant power plant during PBLH < 500 m situations. In case of the 30 m air inlet, we get an SSI-VSI $ffCO_2$ difference of 4.9 ppm and in case of the 200 m air inlet this SSI-VSI $ffCO_2$ difference is reduced to 0.6 ppm. This means that the SSI contribution in case of the 30 m air inlet is 17.4 times larger than the VSI contribution. In case of the 200 m air inlet the SSI contribution from the closest power plant is only 1.8 times larger during PBLH < 500 m situations. We have tried to make this point clearer in our revised manuscript (lines 425ff).

L386 ('quite low'): Please quantify and explain why these would be smaller. Uncertainty type 3 seems clear since the location is known. But especially uncertainty type 4 could be an important factor when it comes to the current study.

In general, the locations and annual emissions of point sources are well known and can be found in databases. That's the reason why Super et al. (2020) excluded them from the Monte-Carlo simulations. However, their temporal profile has much larger uncertainties, which we can't quantify. That's why we performed the sensitivity study in Fig. 8 of the manuscript. See further explanations in our answer to your next comment.

L394: Connected to the previous comment. Is this true? What is the diurnal profile that is applied to large point sources in the present simulations? This may well be different for different point sources. Can you be sure that the nearby point sources mentioned in section 2.1 can be represented well by the applied diurnal profile? A day/night difference of 50 % seems likely for a combined heat and power plant that can react relatively quickly to energy demands.

We used the source sector specific diurnal profiles from TNO (see Fig. 5 for the diurnal profiles for the most important energy and industry point sources). We agree that these profiles can have large uncertainties, especially when applied to individual point sources. An advantage could be that we analyse two-week integrated samples, which have seen different point sources depending on the meteorological situations. Thus, we are maybe less influenced by one specific point source and get a more mixed signal from different point sources. However, this would not help if the diurnal (day vs. night) emission profile is wrong for all point sources. We wanted to investigate this in Fig. 8 in the manuscript. It shows that the nocturnal point source emissions would have to be decreased by 70 % to get a similar performance (in terms of RMSD between modelled and observed $ffCO_2$) for the SSI and VSI approach. Since the averaged daily point source emissions seem to be relatively well known, this would mean that the point source emissions have to be significantly increased during the day so that the 70 % reduction during night is compensated. Since we already have applied a diurnal profile with nocturnal emissions being 24 % (energy) and 34 % (industry), respectively, lower than during the afternoon, the day/night difference would be much more than 50 %. We consider such large deviations as highly unlikely.

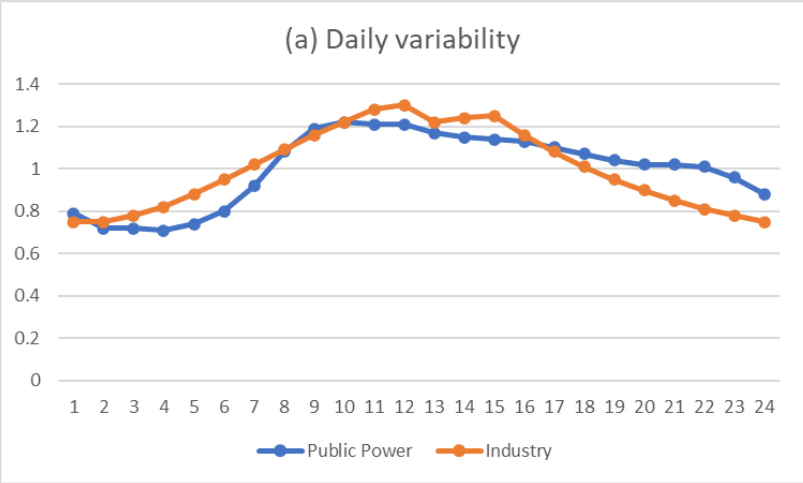


Figure 5: TNO diurnal profile for public power (blue) and industry (orange) emissions.

L461ff: I am not sure if this is the right location for this paragraph. Could also go to the introduction. It certainly does not fit the section title. Maybe also in the context of the Brunner et al 2019 publication L87, which also highlights the importance of representing point sources correctly in Eulerian models.

In this paragraph we report on Super et al. (2017), who found that point sources within a radius of 10 km distance should be modelled with a Gaussian plume model. This result fits quite well to the conclusions of our pseudo power plant experiment for the virtual 200 m air inlet, which suggests that power plants within a radius of also about 10 km should be modelled with the VSI approach. That's why we think that it is appropriate to mention the results from Super et al. (2017) here. However, we fully agree, that this does not fit the section title. For this, we have reformulated the section title to "Representation of nearby point source emissions in models" in the revised manuscript.

Code availability: It is not clear to me if the modifications of the 'volume source approach' have been made available in the latest STILT version and how they can be activated. A few more technical details on how to use the approach and from which version these are available would be appreciated.

We have uploaded the R script to calculate the $ffCO_2$ contributions from point sources together with the used trajectory information and the TNO point source emissions in the surroundings of Heidelberg to https://doi.org/10.5281/zenodo.5911518. With this R script the SSI and VSI $ffCO_2$ contributions from nearby point sources can be calculated off-line. It uses the calculated trajectory information and the TNO point source emissions, which are also available there. If one wants to calculate trajectories for different locations, the full STILT model has to be downloaded (http://stilt-model.org/). We used the revision number 747 of the STILT repository. We have added this information to the code availability section (lines 589ff).