Response to Reviewer 1

We thank the reviewers for their constructive feedback on our manuscript (https://www.geosci-model-dev-discuss.net/gmd-2018-20/). The reviewers' comments are shown below in *italics* with our responses directly following.

Anonymous Referee #1

Fasoli et al report on developments made for the Stochastic Time-Inverted Lagrangian Transport Model (STILT). They added high-level functions to make simulations using the R language. They added code to make parallelized simulations. They introduce a new method to deal with near-field emissions which have not yet been homogeneously mixed within the boundary layer. Fasoli et al. further introduce a smoothing technique to estimate plume surface response functions that is compared to the existing approach. Finally, they show how their developments perform in an experiment in which CO2 measurements have been taken aboard a light-rail in the Salt Lake City metropolitan area.

In general I find the manuscript well written, developments and results are presented in a concise and understandable manner. The manuscript fits the scope and contains enough scientific content to warrant publication in GMD. I have a number of comments that I would like to see addressed before publication, hence I end up with "minor revisions".

General comments:

1) There is no mentioning of how this work and code repository relates to the original code repository and work at http://stilt-model.org / BGC Jena. None of the co-authors are from Jena or other developers of STILT. I can only assume that this development has been made in accordance and in agreement with the rest of the STILT developers, and that there are no licensing issues. Should be checked and stated explicitly.

This work is intended to serve as the future replacement for the current "stiltR" wrapper, distributed from the BGC Jena SVN repository. We have been working with the BGC Jena team and the fortran source code remains hosted at BGC Jena. Migrating this wrapper code to GitHub has already enabled significant collaborative development between groups and has led to implementing features beyond those described in this paper.

2) Model performance is assessed in a very qualitative manner ("looks better") and sometimes overly positive. I suggest authors consider more quantitative assessments, and take a step back before claiming (see below) e.g., that the model represents enhancements in (individual) roadways and intersections.

Thank you for the suggestion. We have made an effort to improve the assessments of the results to quantify the differences between methods. Please see the specific comments regarding changes to the manuscript.

*3) While developing a new method to deal with incompletely mixed sources close to the receptor, the fall back to limiting mixing to the crude 0.5*PBLH formulation. There is no physical basis for that, and I urge the authors to reconsider this artificial limitation. More on this below.*

See comment relating to p4 l7ff below.

Specific comments:

p2 l31ff It is unclear to me why 1-10km and 0.1-1 hr spatial and time scales should qualify as "hyper" near-field. Unless you show that "near-field" is a common term that refers to larger spatial or temporal scales I suggest removing the "hyper", as is is hyperbole.

We thank the reviewer for their comment as because it is important to clarify the naming conventions of domain length scales for readers. We chose the term “hyper near-field” as an extension of the definition of “near-field” in the foundational work of Lin et al 2003, which was “a domain extending over 10^2-10^3km”. To clarify this definition in the text, we have added the following statement:

> Previous work has defined the near-field domain as extending over 10^2 - 10^3 km (Lin et al., 2003).

p3 l26ff Can this be used with other queue managers apart from SLURM?

As of writing, SLURM is the only cluster job scheduler that has been implemented. SLURM is open source and utilized heavily by the high performance computing (HPC) systems at the University of Utah. Due to limited availability of HPC clusters, SLURM is the only job scheduler that has been validated. However, modifications to the project scaffolding described in this manuscript that facilitate parallel computation within single-node and SLURM-scheduled environments opens the doors to other queue managers as well. We encourage future collaboration with users who have access to these job schedulers and would be willing assist with testing development code on their systems. To clarify this in the text, we have added the following statement:

> While SLURM is the only cluster job scheduler that has been implemented to date, the open source code can be modified to run on systems managed by other job schedulers including TORQUE/OpenPBS, Sun Grid Engine, OpenLava, Load Sharing Facility, or Docker Swarm using methods described by Lang et al. (2017).

p4 l3ff Again, "hyper near-field" sounds very hyperbole and I suggest renaming it - you are talking about the region in which the well-mixed criterion does not hold.

See comment relating to p2 l31ff above.

*p4 l7ff Mixing to h = 0.5 * PBLH (p4 l7) is a crude assumption with no physical meaning - if you would wait long enough you would have mixing into 1.0 * PBLH (ignoring en- /detrainment) at the top of the BL.*

While we agree that the h = 0.5 PBLH mixing height serves as a crude assumption when used for vertically diluting surface fluxes, it has been extensively validated for the traditional "near-field" domain. This assumption was first introduced by Gerbig et al., 2003 (https://doi.org/10.1029/2003JD003770, Section 3.2. Depth of "Surface Layer") who performed a sensitivity study and found that "no significant change in the modeled vegetation signal was found" by varying the fraction of the PBL height considered between 0.1 and 1.0.

The goal of this manuscript is to simplify the model workflow and improve STILT's relevance to the HNF domain. Further, the more complex formulation based on turbulence theory is implemented as an optional feature which can be disabled by the user to replicate past simulations while taking advantage of the other improvements described in this manuscript. With this in mind, we retained the traditional h = 0.5 PBLH for the "near-field" domain for consistency with previous work.

The following text has been added:

> While the assumption that surface fluxes can instantaneously mix to h^* has been validated within the traditional near-field domain (Gerbig et al., 2003), this method underrepresents the influence of HNF fluxes on the tracer mole fraction arriving at the receptor.

p4 l15ff You then derive a more complex formulation based on turbulence theory, which you propose to be better. However in the end you use h = min(h', h), with h* the crude approximation (see above), which effectively stops dilution at 0.5 * PBLH. This seems wrong - why not dilute up to whatever your new formula gives you, maybe cap at 1.0 * PBLH? There is no reason in reality why emissions from the ground should not be mixed further up than half the PBLH.*

As you say, the formulation of h = min(h', h*) results in the use of the turbulence theory estimation until the traditional 0.5 PBLH is met. The model timestep at which h' is approximately equal to h* is the transition between the "hyper near-field" and "near-field" domains. Rather than a rigid definition of the "hyper near-field" length scale, setting the dilution depth to min(h', h*) allows the "hyper near-field" spatial domain to adapt to meteorological conditions while enabling a smooth transition between the two methods of vertically diluting surface fluxes.

To clarify this point in the manuscript, the following text has been added:

> The timestep at which h' ≈ h* signifies the transition between the HNF and traditional near-field domains. The formulation of h' grows the particle-specific dilution depth relative to local turbulence and enables the extent of the HNF domain to vary depending on receptor location, meteorological conditions, and local topography.. For surface based applications, h' grows to h* over 0.1 - 1 hr, affecting a spatial domain of 1 - 10 km adjacent to the receptor.

p4 l21ff: this spatial domain should be variable and strongly dependent on receptor location, topography and meteorology - this should be emphasized. In general, sensitivity studies on how this new formulation performs are required.

We agree that it the spatial domain of the HNF is highly variable and have updated the text to clarify. See comment relating to p4 l15ff above.

p4 l21ff: it should be mentioned how (whether?) this method will work with intermittent turbulence and nighttime (stable) conditions (see p7 l22 where you exclude nighttime values for such reasons).

We agree that we should emphasize that nighttime turbulence is an unsolved problem that is beyond the scope of this manuscript.

The following text was added to ~p9L10:

> We have defined the HNF using the effective vertical mixing depth of surface fluxes arriving at a receptor and shown this formulation to improve model agreement with observations. However, calculating the effective vertical mixing depth using turbulence variables σ_w and T_L does not extend well to stable nighttime conditions which remain a difficult problem (Holtslag et al., 2013) and a subject of future work.

p4 l28: this should be Figure 4.

Thanks for catching that mistake.

p5 l8ff: explain better: which particles go into the sigma calculations?

We have modified the equation notation and text to show that the sigma calculations are derived from the positions of all particles in the ensemble at each model timestep.

To clarify this in the text, we have modified the text following equation 3 to include:

> where σ_x^2 and σ_y^2 are Euclidean variances in the x and y positions of all particles in the ensemble at time t.

p5 l11: it would be helpful for the reader to see how b enters the two-dimensional Gaussian you are using for density estimation.

While we agree that is important to understand how the kernel bandwidth relates to the smoothing applied, visualizations and discussions regarding bandwidth selection and how it relates to bias-variance optimization can be found in many general descriptions of kernel density estimation.

To help the reader understand what effect the bandwidth has on the model output without requiring an additional visualization, we have added the following text to ~p5L17:

> As model time or total ensemble dispersion increase, the kernel bandwidths increase the amount of smoothing applied to each particle.

p5 l22ff: "improved" is based purely on visual aesthetics ("looks more similar!") - a quantitative measure would be very beneficial here.

Agreed. We have performed additional analysis and added a quantitative measure for the difference between the calculation methods and the ideal case.

The following text has been added to ~p6L1:

> The effects of varying the smoothing parameter (f = 1,2) are shown and errors are quantified using the difference of calculated grid cells from the idealized brute force case.
>
> For the typical case (N = 200 and f = 1), the kernel density estimator shows improved agreement with the brute force method (rmse = 5.60 * 10^{-4} ppm (umol^{-1} m^2 s)) compared to the traditional dynamic grid coarsening (rmse = 5.79 * 10^{-4} ppm (umol^{-1} m^2 s)), preserving a Gaussian plume adjacent to the receptor, a clustered area of high influence, and capturing split flow upstream. When the kernel bandwidths are doubled by increasing the smoothing parameter (f = 2), the footprint field becomes over-smoothed and becomes less similar with the brute force case (rmse = 5.66 * 10^{-4} ppm (umol^{-1} m^2 s)). In the extreme case using atypically few particles (N = 10), the dynamic grid coarsening method produces a footprint field dominated by noise from individual particles (rmse = 6.12 * 10^{-4} ppm (umol^{-1} m^2 s)). The kernel density estimator (f = 1) improves results but shows fragmentation further from the receptor (rmse = 5.75 * 10^{-4} ppm (umol^{-1} m^2 s)). In this case, the kernel density estimator smoothing parameter enables users to manually widen the plume reproduced in the footprint. Doubling the smoothing parameter (f = 2) improves similarity with the smaller particle ensemble (rmse = 5.70 * 10^{-4} ppm (umol^{-1} m^2 s)) and demonstrates how users can modify the kernel bandwidths to adapt the model to unique cases. While tracer mole fraction differences between the two footprint calculation methods vary depending upon the locations of footprint differences relative to sources, tracer mole fractions calculated using the kernel density estimator are more similar to the idealized brute force case.

p5 l27ff: "compensating" it might be, but only in the case here - just concede what f is: a fudge factor without physical mean.

We have changed the language to highlight the reviewer's comment at ~p6L12:

> In this case, the kernel density estimator smoothing parameter enables users to manually widen the plume reproduced in the footprint. Doubling the smoothing parameter (f = 2) improves similarity with the smaller particle ensemble (rmse = 5.70 * 10^{-4} ppm ($umol^{-1}$ m^2 s)) and demonstrates how users can modify the kernel bandwidths to adapt the model to unique cases.

p7 l9: "dilution correction" refers to the fudge factor f being set to 2? Explain!

We agree that using the terms "dilution correction" and "vertical mixing depth correction" interchangeably was confusing for readers. We have changed the text to "HNF vertical mixing depth" to be consistent with terms used in the methods description (Section 2.3).

The following text has been added to ~p7L14:

> We compute footprint fields using the legacy dynamic grid coarsening (LEG) algorithm as well as gaussian kernel density estimation with the HNF vertical mixing depth correction (GWD) and without the HNF vertical mixing depth correction (GND) to illustrate the differences between methods.

p7 l20ff: You are doing the right thing by ignoring nighttime values, but I suggest you still include the nighttime data in the plots to elucidate the magnitude of this problem - this is something that all model approaches have in common and it helps to remind people that comparing nighttime values is difficult and care needs to be taken.

While we agree that it is useful to compare nighttime modeled values between manuscripts, the light-rail measurement platform only operates during specific hours of the day. We have clarified this in the text.

The following text has been added to ~p6L14:

> The light-rail train typically operates between the hours of 05:00-23:00 Local Daylight Time (LDT) and only these hours were used in analyses.

p8 l7: there is no appreciable "evening rush hour" peak in this figure. Neither does the model "capture" it, as it is too low throughout the day compared to observations. Remove.

We agree that discussing an “evening rush hour peak” may be inaccurate in this context. The late afternoon increase in modeled CO_2 is the result of increased emissions from both anthropogenic inventories as well as meteorological factors. We have removed the text as the reviewer suggested.

p8 l1 and Figure 7 caption: there are no consistent enhancements in modelled CO2 concentrations visible in the bottom right plot that would conincide with the individual intersections shown. I disagree with the statement that the method captures these enhancements. Rephrase and state more careful what you actually can resolve.

We agree that it is important to not inflate the improvements and the language describing spatial resolution needs to be more clearly defined.

The following text was added to ~p7L18:

> This grid resolution was chosen to pair analyses with the 0.002º Hestia inventory and because 0.002º corresponds roughly with the size of a Salt Lake City block.

The following text was added to ~p8L21:

> The model generally produced mole fraction enhancements (ΔCO_2) for grid cells containing or downwind from major roadways (Fig. 7). However, modeling intersection scale enhancements would require finer grid spacing capable of resolving sub-city-block spatial scales that is not yet feasible given current constraints on inventories, meteorological data, and computing resources.

The following text has been added to the Fig. 7 caption:

> The model captures the overall urban-suburban-rural CO_2 mole fraction gradient (top) as well as localized enhancements near grid cells containing large emitters such as busy roads (bottom).

p8 l18ff: Careful to make sure that you are not mistaking increasing resolution with the "hyper" near-field approach described earlier - this last section just shows that higher spatial resolution can be beneficial. Might want to rephrase "fine-scale approach".

We agree that readers may confuse the language with the hyper near-field definition. The text “fine-scale” has been changed to “0.002º grid resolution”.

Figure 6: axis labels missing, should appear at least once for x and y

We have added the axis labels (Longitude and Latitude) to Figure 6 as recommended.

Figure 5 and 9: plot x axis from 0 to 24, add nighttime values (shade to make clear you don’t use them).

The hours that are not represented on the x-axis do not contain any data. See comment relating to p7 l20ff above for details.

Figures 4, 6 - 8: Background maps at least for Figures 7-8 seem to come from Google Earth, are you sure you have the license to use and publish them?

Google Maps and Google Earth permits use in periodicals (https://www.google.com/permissions/geoguidelines.html#maps-print) with proper attribution. However, it appears that the attributions were cropped out of several of the figures. The content was updated in accordance with Google's attribution guidelines (https://www.google.com/permissions/geoguidelines/attr-guide.html).

Response to Reviewer 2

We thank the reviewers for their constructive feedback on our manuscript (https://www.geosci-model-dev-discuss.net/gmd-2018-20/). The reviewers' comments are shown below in *italics* with our responses directly following.

Anonymous Referee #2

This work documents the workflow of STILT simulations and presents improved physical processes for fine-scale simulations. I appreciate the authors' efforts in addressing overdue problems for the community, in particular those who use STILT extensively. I hope that the authors continue updating their work through GitHub.

I can easily follow the method and think the paper is relatively well written given the conciseness in length.

Thanks for the positive comments. We hope that readers will agree.

I have some questions/concerns in the evaluation of the improved method. In current form, the authors do not characterize the errors, in particular in surface emissions. So it is hard to evaluate the results. The model evaluation is a key result in this study, and the authors need to describe how much they know (or pre- scribed) the errors in surface emissions (and others if prescribed) so that we can be sure that the better results from GWD are due to the improved schemes.

We have added a discussion regarding difficulties in estimating uncertainties in emissions inventories. Please see below for details.

Detailed Comments:

L13 - 21: STITL-R should be applicable to other tracer gases, not only CO2. The authors describe CO2 only, which seem to be strange. This is probably because the authors show an evaluation study using CO2, but this CO2 focus is limited.

STILT's applications certainly exceed only simulating atmospheric CO2. We attempt to describe the use of LPDMs and the STILT model (~p2L9, ~p2L20) using generalized language such as "atmospheric mole fractions", "pollutant concentrations", and model applicability to "observed emissions" and "surface fluxes". We use urban CO_2 as the primary motivation for several reasons: urban CO_2 cycling is the focus of a large and growing body of scientific literature that this model update will play a prominent role in, it allows for the use of novel CO_2 surface flux

inventories purpose-built for the study region (the Hestia model), and it applies well to the case study using the unique data available from the light-rail measurement system.

P2, L21: Need to cite older work about HYSPLIT.

We have added a citation for Draxler, R.R., and G.D. Hess, 1998.

P2, L28 - 29: Need to mention more recent work on city-scale or regional inversion work based on multiple receptors that uses STILT extensively. Literature review here does not represent a full range of the use of the traditional STILT, which I believe is import to for the reader to understand the context, and motivation for the new development.

We have added citations for McKain et al., 2012 and McKain et al., 2015 describing STILT modeling applications in Salt Lake City and Boston as well as Kort et al., 2013 describing STILT's use to assess measurement network design in Los Angeles.

P3, L6: Need to include the reference for R properly. Not doing so is irresponsible because without R this work is not possible.

Thank you for the suggestion. We have added the citation for the R software at ~p3L7.

P3, L20: For large-scale simulations, the users have applied other types of parallelizations in running STILT, e.g., running multiple jobs (each job may represent one receptor for a give period) at the same time taking advantage of high performance computing. The authors need to briefly mention what the difference between the old method and the one introduced here would be although the method described here seems to be similar to what users have been using. Is there a new concept here?

We recognize that we did not adequately describe past efforts to run parallel simulations. While the concept of executing batches of receptors across multiple jobs is not new, users have previously had to write and run separate scripts defining the receptors and relevant data inputs for each job which can require significant manual labor or develop their own methods for batch processing receptors. The manuscript formalizes methods for automatically executing the parallel batches of receptors, with receptor batches distributed between the parallel jobs and managed by the code itself rather than the user. The workflow presented, controlled with *run_stilt.r* and with output saved to simulation ID directories, remains the same for serial and parallel execution with only changing the setting for the number of parallel processes.

To clarify this point, the following text has been added to ~p3L28 :

> However, past methods for parallelizing simulations require users to manually define batches of receptors and relevant meteorological inputs in unique initialization scripts and submitting each script as a separate job to the scheduler. While increasing the

number of parallel threads decreases the size of each simulation batch, the requirements of the user become more complex.

We formalize methods for automatically distributing batches of receptors across many parallel threads managed by the model rather than the user.

P3, L27: Not all systems use SLURM although it is popular. Is there an option for a different job scheduling tool?

As of writing, SLURM is the only cluster job scheduler that has been implemented. SLURM is open source and utilized heavily by the high performance computing (HPC) systems at the University of Utah. Due to limited availability of HPC clusters, SLURM is the only job scheduler that has been validated. However, modifications to the project scaffolding described in this manuscript that facilitate parallel computation within single-node and SLURM-scheduled environments opens the doors to other queue managers as well. We encourage future collaboration with users who have access to these job schedulers and would be willing assist with testing development code on their systems. To clarify this in the text, we have added the following statement:

While SLURM is the only cluster job scheduler that has been implemented to date, the open source code can be modified to run on systems managed by other job schedulers including TORQUE/OpenPBS, Sun Grid Engine, OpenLava, Load Sharing Facility, or Docker Swarm using methods described by Lang et al. (2017).

P4. L4 - 22: In many cases, PBL heights from meteorological models (e.g., WRF) are directly used to represent z_pbl. The authors need to clarify this and describe more on the use of WRF PBL related to equations (1) and (2). For HNF simulations, WRF needs to be run at a similarly fine sale, which is really expensive? If not, what would be the impact on h = min(h',hˆ)?*

The formulation for the HNF vertical mixing depth adjustment h' is intended to fix systematically low footprints without needing to explicitly resolve z_{pbl} at HNF resolutions. It provides an estimate for the effective mixing depth based on homogeneous turbulence theory without requiring meteorological inputs (e.g. WRF) to be at a scale that explicitly defines the fine variations in PBL height within a city. However, the meteorological data are used outside of the HNF domain to calculate h* using a modified Richardson number method that has been extensively validated for the traditional "near-field" domain.

P5, L1-2: Reading this, my immediate thought was if this would require more simulation time to estimate the weighted influence. It would be nice to mention the cost.

Agreed. Calculating the footprint field using smoothing methods involves a cost tradeoff with a larger particle ensemble. While it is almost always less expensive to apply smoothing methods compared to calculating particle trajectories, quantifying the advantage is difficult. The cost to

calculate particle trajectories varies depending on model configuration, meteorological data source, the size of the meteorological domain, and the size of the ensemble while the cost to apply smoothing depends on the method and the spatial and temporal domain of the output footprint.

To clarify this point, the following text has been added to ~p5L1:

> Computing trajectories of large particle ensembles ($N > 10^4$) is computationally expensive. To lessen the cost of each simulation, footprint fields are often calculated from smaller particle ensembles by applying smoothing methods to compensate for the smaller ensemble size. These smoothing methods are less computationally expensive than calculating trajectories for a larger ensemble but vary in their ability to reproduce the robust footprint field of the large particle ensemble.

P6, L32: Should not include a paper in preparation.

Agreed. We removed the citation since the manuscript is still in preparation.

P7, L5: 24-h backward in time seems to be too short. How was the upstream boundary condition treated? I see a short description from L17. Boundary conditions are complex due to wind directions. Is the wind consistent from one direction? I would like to see a more description on this.

We find that particles exist within the footprint domain for 11 hours on average. The meteorological domain encompasses a larger area than the footprint domain and fluxes from outside of the footprint domain are assumed to be resolved by the background atmospheric signal described at p8L6.

To clarify this point, the following text has been added to ~p7l24:

> Urban development and expansion in the area surrounding SLC is limited by the mountainous topography surrounding the city and the Great Salt Lake which restrict the expansion of the city and suburbs. This confines large anthropogenic and biologic sources into a relatively small area surrounding the SLV and simplifies boundary conditions for SLV-centric modeling efforts. From each receptor, 24 h backward trajectories of 200 particle ensembles were calculated using meteorological fields from the HRRR model, available at an hourly interval with a 3 km grid resolution. On average, particles travel within the model domain for 11 h. Computation of the 33,608 particle trajectories and a single set of footprints completed in 5.5 hours utilizing 80 parallel threads across 5 nodes, each equipped with 64 GB of memory with two 8-core Intel XEON E5-2670 2.6 GHz processors. 6.7% of the simulations were not completed due to short-term outages in the HRRR data product.

P7, L30: Please use rˆ2 and state which method was used in calculating r. Pearson's method? How are these rˆ2 values statistically different? The simulations from GWD is distinguishably

from a different distribution from the other two so that we have more confidence in GWD? Note that in this evaluation, we want to clearly see better results from GWD. Right?

As recommended, we have modified the text to use r^2 instead of r to explain model variance and have clarified that it is based on Pearson's method.

While there is likely no statistical significance in the differences between GND and LEG for this case study, we show that GND agrees better with the physical "ideal" case and may give improvements that depend on the locations of differences between GND and LEG relative to the locations of surface fluxes. With the vertical dilution correction (GWD), the results agree more closely with measurements in both time and space.

P8: L1: I think this is probably the most important single statement in this paper. I would like to know how the authors determined the uncertainty in the surface fluxes. Without precise uncertainty characterization, the results are not reliable. What if the inventory is systematically low and GWD overestimated the mole fraction, which could be shown to be closer to the observations than the other two methods? I believe that the authors have considered this point, but I don't see the details here to the level that I can clearly see the outperformance of GWD. Also we need to note that the r^2 values are all low and similar to each other.

We agree that it is important to investigate uncertainty in inventory estimates. While we can show improvements to footprint smoothing algorithms using physically constrained "ideal" cases, uncertainty estimates within emissions inventories remains an unresolved question within the emission inventory scientific community. Developers of the Hestia inventory have documented that "a devoted effort is needed to generate uncertainty and propagate those uncertainties through the Hestia approach to provide an improved understanding of where results are more or less certain in space and time. This remains a high priority for future research" (Patarasuk et al., 2016) and determination of GHG fluxes and uncertainty bounds is one of the primary goals in the ongoing Indianapolis Flux Experiment (http://sites.psu.edu/influx/). Improvements to LPDMs can help future inverse modelling frameworks that would be better equipped to quantify uncertainties in flux inventories.

To further clarify this, we have added a discussion regarding the difficulties in assessing emission inventory uncertainties. Both of the inventories we discussed in the manuscript (Hestia and ODIAC) agree on the total emissions within the SLV domain which is evidence one inventory is not systematically lower than the other. However, mapping uncertainty to a moving receptor using two emissions inventories that encompass different spatial domains and allocate fluxes using different methods in time and space is a difficult question that requires more tools and analysis than are available in our present manuscript and should be the focus of future work.

The following text has been added to ~p7L20:

> Within the SLV domain where the inventories overlap, Hestia and ODIAC agree on the total anthropogenic emissions to within 1.5% during our study period. However, uncertainties of fluxes applied to our analyses are likely larger since the two inventories allocate fluxes differently in space and time. Further, only Hestia is used to represent the SLV whereas ODIAC is used outside of the SLV to account for regional-scale emissions. Uncertainties in inventory estimates are difficult to quantify in time and space and require a devoted effort within the emission inventory scientific community to propagate uncertainties through underlying assumptions within each inventory (Patarasuk et al., 2016; Lauvaux et al., 2016).

P8, L6: Please be more quantitative. It is not clear what has been reproduced. C3

We have changed the text to generalize that the model sees enhancements downwind from major roadways and introduced a caveat that better details the limitations regarding model resolution.

The following text was added to ~p8L21:

> The model generally produced mole fraction enhancements (ΔCO_2) for grid cells containing or downwind from major roadways (Fig. 7). However, modeling intersection scale enhancements would require finer grid spacing capable of resolving sub-city-block spatial scales that is not yet feasible given current constraints on inventories, meteorological data, and computing resources.

P8, L10 - 15: The simulated mole fractions are a combined result of transport and surface flux emissions. The authors, as mentioned, need to say how much we know about the surface emissions (used here) related to this discrepancy as well as the transport arguably improved from this work.

See comments relating to p8L1.

Simulating atmospheric tracer concentrations for spatially distributed receptors: updates to the Stochastic Time-Inverted Lagrangian Transport model's R interface (STILT-R version 2)

Benjamin Fasoli[1], John C. Lin[1], David R. Bowling[2], Logan Mitchell[1], and Daniel Mendoza[1,3]

[1]Department of Atmospheric Sciences, University of Utah, Salt Lake City, 84112, USA
[2]Department of Biology, University of Utah, Salt Lake City, 84112, USA
[3]Division of Pulmonary Medicine, School of Medicine, University of Utah, Salt Lake City, 84112, USA

Correspondence to: Benjamin Fasoli (b.fasoli@utah.edu)

Abstract. The Stochastic Time-Inverted Lagrangian Transport (STILT) model is comprised of a compiled Fortran executable that carries out advection and dispersion calculations as well as a higher level code layer for simulation control and user interaction, written in the open source data analysis language R. We introduce modifications to the STILT-R codebase with the aim to improve the model's applicability to fine-scale ($< 1\mathrm{km}$) trace gas measurement studies. The changes facilitate placement of spatially distributed receptors and provide high level methods for single and multi-node parallelism. We present a kernel density estimator to calculate influence footprints and demonstrate improvements over prior methods. Vertical dilution in the hyper near-field is calculated using the Lagrangian decorrelation timescale and vertical turbulence to approximate the effective mixing depth. This framework provides a central source repository to reduce code fragmentation between STILT user groups as well as a systematic, well documented workflow for users. We apply the modified STILT-R to light-rail measurements in Salt Lake City, Utah, United States and discuss how results from our analyses can inform future fine-scale measurement approaches and modeling efforts.

1 Introduction

Cities are the source of over 70% of global fossil-fuel carbon dioxide (CO_2) emissions (International Energy Agency, 2008; Hoornweg et al., 2012; Gurney et al., 2015), the largest anthropogenic forcing on climate change (Canadell et al., 2007). As governing bodies examine ways to address climate change, urban areas are appropriately a focus for emissions regulation. Atmospheric measurements (Duren and Miller, 2012; McKain et al., 2012) provide a top-down constraint for estimating urban carbon emissions, especially when combined with bottom-up information from fuel consumption statistics, traffic data, and building characteristics that result in highly resolved emission inventories (Gurney et al., 2009, 2012). However, traditional evaluation strategies for estimating CO_2 emissions that focus on quantifying regional scale (10^2 to 10^3 km) averages at coarse resolutions are unable to resolve urban areas beyond bulk estimates. Implementing and evaluating effective policies for emissions mitigation requires understanding where, when, and how emissions occur at a within-city scale.

Novel measurement strategies are emerging to help resolve fine-scale within-city trace gas concentrations, such as measurements made from trains, buses, and cars (Apte et al., 2017; Bush et al., 2015; Lee et al., 2017) as well as dense networks of inexpensive sensors (Shusterman et al., 2016; ?)(Shusterman et al., 2016; Turner et al., 2016). However, traditional atmospheric modeling tools were not designed for densely located and spatially distributed measurements. Simulating atmospheric transport for multiple locations over time often increases the number of simulations by factors of 10^1 to 10^3, necessitating the use of scalable parallel computing to best utilize available hardware and reduce total simulation time. To make use of recent measurement advances, modeling approaches must structure the model framework in ways that enable simulations to execute in parallel, adapt to finer spatial scales, and facilitate simulating atmospheric mixing ratios for locations distributed across space and time.

The link between measured atmospheric mole fractions and upstream surface fluxes is often established using Lagrangian particle dispersion models (LPDMs), popular tools for simulating atmospheric transport and dispersion in the Planetary Boundary Layer (PBL) (Lin, 2013). The LPDMs simulate transport of an ensemble of theoretical particles (representing air parcels) using a combination of mean winds interpolated from meteorological model fields with stochastic fluctuations representing turbulent motions introduced as a Markov process. This approach offers advantages over Eulerian methods by explicitly simulating transport trajectories and better representing atmospheric mixing, turbulent eddies, and convection (Lin, 2013). Particle motion can be simulated either forward in time from an emissions source or backward in time from a location of interest, referred to as the "receptor". The forward configuration is often used to simulate pollutant concentrations downstream from an emission source (Stohl et al., 2005) whereas backward simulations determine the source of observed emissions and quantify surface fluxes (McKain et al., 2012, 2015; Stein et al., 2015). As receptors are often greatly outnumbered by sources, significant computational savings are realized by applying LPDMs in the receptor-oriented configuration (Lin, 2013).

The Stochastic Time-Inverted Lagrangian Transport (STILT) model couples Lagrangian particle dispersion with the mean advection scheme from the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Stein et al., 2015)(Draxler and Hes STILT simulations are reversible in time (Lin et al., 2003), enable quantitative evaluation of transport error (Lin and Gerbig, 2005), and are closely coupled with the commonly used Weather Research and Forecasting mesoscale meteorological model (Nehrkorn et al., 2010), on which the High Resolution Rapid Refresh (HRRR) model is based (Sun et al., 2014). STILT is most commonly used to follow the backwards time evolution of a particle ensemble and calculate a receptor's footprint, a sensitivity matrix defining the upstream area that contributes to tracer mole fractions observed at the receptor. Footprints can be convolved with emissions inventories and an atmospheric background signal to calculate atmospheric mole fractions at the receptor, which is among the most common applications of the STILT model (Gerbig et al., 2003; Lin et al., 2004; Kort et al., 2008; Macatangay et al., 200

This paper discusses limitations within the existing STILT codebase and introduces an updated framework intended to improve the model's applicability to fine-scale spatially distributed measurement approaches. Previous work has defined the near-field domain as extending over 10^2 - 10^3 km (Lin et al., 2003). We introduce the hyper near-field (HNF) area, typically covering length scales of 1 - 10 km and time scales of 0.1 - 1 hr, from which surface fluxes are diluted to a fraction of the PBL height and thus more strongly influence the receptor. Parameterizations within the STILT model were originally intended for regional scales and require refinements to improve source-receptor relationships in the HNF. We also describe a footprint

calculation scheme using kernel density estimation, rescaling of the effective mixing depth for fluxes in the HNF, and methods for parallelizing simulations. The value of STILT as a tool for interpreting within-city CO_2 mole fractions is shown using an example of data collected on the roof of a train car on Salt Lake City (SLC), Utah's light-rail system. We discuss how results from our analyses can inform future measurement approaches and modeling efforts.

2 Modifications to the STILT model

2.1 Software enhancements

The R (R Core Team, 2017) component of the STILT model exists as a group of core functions used to track particle locations, calculate footprints, and apply surface flux grids. User groups have built upon these functions, adding scripts for common modeling workflows and additional functionality. Key components of the higher level functions remain unpublished and undocumented prior to this paper, including a description of methods used to aggregate the particle ensemble to calculate footprints. Here, we adopt a widely-used collaborative software development platform (GitHub) as a common source code repository that meets the needs of STILT users. This repository is built upon existing advection and dispersion calculations but has restructured and modernized the core functions used to interact with the model (Fig. 1).

A single script (*run_stilt.r*) defines model inputs such as receptor locations and meteorological fields, controls and executes the parallelized model, and outputs footprints. Footprints are saved in a netCDF format consistent with conventions for Climate and Forecast metadata (cfconventions.org), the standard for gridded model datasets by the University Corporation for Atmospheric Research (UCAR). This format is compatible with most popular data analysis software platforms and facilitates analysis of model output. The script *run_stilt.r* serves as the primary STILT interface, interacting with R functions which in turn call Fortran subroutines for the bulk of calculations and providing a systematic, well documented workflow for users.

2.2 Model parallelization

Executing simulations in parallel is essential to leverage the full capability of computing resources. STILT receptors are defined in a table of space (x, y, z) and time (t) coordinates enabling users to fix a receptor in space and model the time evolution of the influence field, distribute receptors across space and capture a snapshot at a single time, or distribute the receptors across both space and time. Since each STILT simulation is computationally independent, total simulation time can be reduced by distributing batches of simulations between parallel threads (Fig. 2). ~~As the~~ However, past methods for parallelizing simulations require users to manually define batches of receptors and relevant meteorological inputs in unique initialization scripts and submitting each script as a separate job to the scheduler. While increasing the number of parallel threads ~~increases,~~ decreases the size of each simulation batch~~decreases.~~ , the requirements of the user become more complex.

We formalize methods for automatically distributing batches of receptors across many parallel threads managed by the model rather than the user. Within-node parallelism is achieved through process forking, in which batches of receptors are allocated across multiple parallel threads on a single machine. Multi-node parallelism is accomplished by interfacing with the

Simple Linux Utility for Resource Management (SLURM), an open-source tool that provides the framework for interfacing with clusters of computer nodes (Jette and Grondona, 2003). ~~SLURM~~ While SLURM is the only cluster job scheduler that has been implemented to date, the open source code can be modified to run on systems managed by other job schedulers including TORQUE/OpenPBS, Sun Grid Engine, OpenLava, Load Sharing Facility, or Docker Swarm using methods described by Lang et al. (2017). SLURM allocates computational resources with low overhead and can be used to dispatch job arrays of STILT simulations to multiple nodes. SLURM is used to parallelize *between* nodes and process forking by the modified STILT framework is used to parallelize *within* nodes. Process forking can be used independently to execute parallel simulations on a single machine or combined with SLURM to parallelize simulations within each SLURM node. Provided that memory limits are not exceeded, these methods enable total simulation time to decrease linearly with available CPU cores.

2.3 Hyper ~~Near Field~~ Near-Field vertical mixing depth

The influence of surface fluxes on air arriving at the receptor depends upon vertical dilution within the atmospheric column. The STILT model determines the height of the boundary layer z_{pbl} using a modified Richardson number method (Vogelezang and Holtslag, 1996). In the original STILT model, surface fluxes are instantaneously diluted within an effective mixing depth of $h^* = 0.5 \cdot z_{pbl}$ for which the vertical mixing timescale is comparable to the model timestep for advection (Gerbig et al., 2003). As described ~~in~~ by Lin et al. (2003), an atmospheric column of height $h(x,y,t,p)$ is used to relate surface fluxes $F(x,y,t)$ to the mole fraction influence $S(x,y,t,p)$ for each particle p as

$$S(x,y,t,p) = \begin{cases} \frac{F(x,y,t)m_{air}}{h(x,y,t,p)\overline{\rho}(x,y,t,p)} & z \leq h \\ 0 & z > h \end{cases} \quad (1)$$

where $\overline{\rho}$ is the average air density below h and m_{air} is the molar mass of dry air. Thus, particles below h perceive surface fluxes diluted within an atmospheric column of depth h. However, the advective timescale is often too short for complete turbulent mixing of HNF fluxes to h before arrival at the receptor. ~~Thus~~While the assumption that surface fluxes can instantaneously mix to h^* has been validated within the traditional near-field domain (Gerbig et al., 2003), this method underrepresents the influence of HNF fluxes on the tracer mole fraction arriving at the receptor.

We apply a method of calculating the effective mixing depth in the HNF based on homogeneous turbulence theory, described by Taylor (1922). Each model time step component k of the HNF mixing depth h' at time t_k will be of the form

$$h'_k(p,t) = z_r + \int_0^{t_k} \sigma_w \sqrt{2T_L \left(t + T_L \left(e^{-\frac{t}{T_L}} - 1\right)\right)} dt \quad (2)$$

where z_r is the height above ground of the receptor, σ_w is the standard deviation in vertical velocities encountered by p during the integration timestep, and T_L is the Lagrangian decorrelation timescale. ~~This method grows the dilution depth over time relative to local turbulence.~~ Substitution of $h = \min(h', h^*)$ into Eq. (1) enhances the mole fraction influence of HNF sources

on the receptor (Fig. 3). The timestep at which $h' \approx h^*$ signifies the transition between the HNF and traditional near-field domains. The formulation of h' grows the particle-specific dilution depth relative to local turbulence and enables the extent of the HNF domain to vary depending on receptor location, meteorological conditions, and local topography. For surface based applications, h' grows to h^* over ~~roughly 10 minutes~~0.1 - 1 hr, affecting a spatial domain of ~~approximately~~ 1 - 10 km adjacent to the receptor.

2.4 Kernel density estimation of footprint field

Computing trajectories of large particle ensembles ($N > 10^4$) is computationally expensive. To lessen the cost of each simulation, footprint fields are often calculated from smaller particle ensembles by applying smoothing methods to compensate for the smaller ensemble size. These smoothing methods are less computationally expensive than calculating trajectories for a larger ensemble but vary in their ability to reproduce the robust footprint field of the large particle ensemble.

Prior to methods described in this section, STILT footprints have been calculated by accumulating the influence of particles over an averaging volume. To lessen grid noise from few particles spread throughout the grid, the spatial extent of the particle ensemble was used to dynamically coarsen the size of the averaging volume by a factor of 2 as the particle cloud spreads, first shown in Gerbig et al. (2003). However, at finer resolutions, this method results in excessive smoothing, removing information calculated by the advection and dispersion routines (Fig. ~~3~~4).

Here we introduce a kernel density estimator to spatially allocate the influence of particles to the footprint grid and show improvements over the prior method at fine spatial resolutions. This method distributes the influence of each particle using a Gaussian weighted spatial kernel centered over the particle's position. The size and intensity of the spatial kernels are defined by the kernel bandwidth, which is determined at each model time step using elapsed time and total dispersion of the stochastic ensemble as proxies for uncertainty in the locations of individual particles. As model time or total ensemble dispersion increase, the kernel bandwidths increase the amount of smoothing applied to each particle. Dispersion of the particle cloud at each time step is represented using a nondimensionalized standard deviation of particle locations ~~σ_d~~ $\sigma_d(t)$ given by

$$\sigma_d(t) = \sqrt{\sigma_x^2 + \sigma_y^2}\sqrt{\sigma_x^2(t) + \sigma_y^2(t)} \tag{3}$$

where ~~σ_x^2 and σ_y^2~~ $\sigma_x^2(t)$ and $\sigma_y^2(t)$ are Euclidean variances in ~~horizontal particle positions in degrees~~the x and y positions of all particles in the ensemble at time t. We find ~~σ_d~~ $\sigma_d(t)$ to agree with other ~~dispersion metrics for~~ metrics of dispersion within the particle ensemble, such as the average pairwise distance ($r^2 > 0.99$), with less computational expense. ~~Kernel bandwidths~~ The bandwidths of the gaussian smoothing kernels are then given by

$$b(t) = f\frac{0.06\sqrt{t\sigma_d}}{\cos\phi}\frac{0.06\sqrt{t \cdot \sigma_d(t)}}{\cos(\phi(t))} \tag{4}$$

where t is time elapsed in days, ~~ϕ is latitude~~ $\overline{\phi(t)}$ is mean latitude of the particle ensemble used for approximation for meridional grid convergence, and 0.06 is an empirically derived constant. f defaults to 1 and is provided as a user defined smoothing adjustment to enable manual manipulation of kernel sizing.

~~Next, we~~ We test the new footprint calculation methods against a brute force simulation with an atypically large particle ensemble size ($N = 10^5$) aggregated over a homogeneous grid (Fig. 4). This large simulation is computationally expensive but generates an idealized, physically constrained footprint without smoothing algorithms. The simulation receptor was positioned at a ~~Salt Lake City~~ SLC CO_2 measurement site on a summertime afternoon and particles were followed backward in time for 24 hours. We then demonstrate differences between the new kernel density estimator and the ~~old~~ traditional dynamic grid coarsening footprint calculation methods (Fig. 3) for a typical particle ensemble ($N = 200$) and for an extreme case with atypically few particles ($N = 10$). The effects of varying the smoothing parameter ($f = 1, 2$) are ~~also shown~~ shown and errors are quantified using the difference of calculated grid cells from the idealized brute force case.

For the typical case ($N = 200$ and $f = 1$), the kernel ~~method~~ density estimator shows improved agreement with the brute force method ($\text{rmse} = 5.60 \cdot 10^{-4} \text{ppm}\ \mu\text{mol}^{-1}\text{m}^2\text{s}$) compared to the traditional dynamic grid coarsening method ($\text{rmse} = 5.79 \cdot 10^{-4} \text{ppm}$ preserving a Gaussian plume adjacent to the receptor, a clustered area of high influence, and capturing split flow upstream. When the kernel bandwidths are doubled by increasing the smoothing parameter ($f = 2$), the footprint field becomes over-smoothed and ~~loses similarity~~ becomes less similar with the brute force case ($\text{rmse} = 5.66 \cdot 10^{-4} \text{ppm}\ \mu\text{mol}^{-1}\text{m}^2\text{s}$). In the extreme case using atypically few particles ($N = 10$), the dynamic grid coarsening method produces a footprint field dominated by noise from individual particles ($\text{rmse} = 6.12 \cdot 10^{-4} \text{ppm}\ \mu\text{mol}^{-1}\text{m}^2\text{s}$). The kernel density estimator ($f = 1$) improves results but shows fragmentation further from the receptor ($\text{rmse} = 5.75 \cdot 10^{-4} \text{ppm}\ \mu\text{mol}^{-1}\text{m}^2\text{s}$). In this case, the ~~scarcity of particles can be compensated for by increasing the~~ kernel density estimator smoothing parameter enables users to manually widen the plume reproduced in the footprint. Doubling the smoothing parameter ($f = 2$) improves similarity with the smaller particle ensemble ($\text{rmse} = 5.70 \cdot 10^{-4} \text{ppm}\ \mu\text{mol}^{-1}\text{m}^2\text{s}$) and demonstrates how users can modify the kernel bandwidths to adapt the model to unique cases. While tracer mole fraction differences between the two footprint calculation methods vary depending upon the locations of footprint differences relative to sources, tracer mole ~~fraction~~ fractions calculated using the kernel density estimator ~~improves similarity with~~ are more similar to the idealized brute force case.

3 Evaluation

3.1 ~~Salt Lake City~~ SLC light-rail measurements

We demonstrate these changes to STILT by comparing CO_2 mixing ratios simulated by the STILT model with corresponding measurements on-board an electric ~~Salt Lake City, Utah~~ SLC light-rail commuter train during July 2015. The Salt Lake Valley (SLV) is a 1,300 km^2 area encompassing ~~Salt Lake City~~ SLC and its surrounding suburbs, bounded by the Wasatch mountain range to the east, the Oquirrh mountain range to the west, the Traverse mountain range to the south, and the Great Salt Lake to the northwest. A light-rail train is equipped to measure high-frequency (1 Hz) CO_2 mole fractions in repeated transects of the SLV using a Los Gatos Research Ultraportable Greenhouse Gas Analyzer. CO_2 and CH_4 mole fractions are corrected for

water vapor dilution and spectrum broadening and are calibrated every hour using a compressed whole air tank with known tracer mole fractions traceable to World Meteorological Organization standards. The light-rail train typically operates between the hours of 05:00-23:00 Local Daylight Time (LDT) and only these hours were used in analyses. For details related to the measurement platform, refer to Mitchell et al. (in review).

The observations generally show higher CO_2 mole fractions in ~~Salt Lake City~~SLC's urban center and along the north-south oriented urbanized corridor centered in the SLV (Fig. 7), consistent with urban spatial CO_2 gradients observed in previous studies (Idso et al., 2001; Pataki et al., 2007). The lowest mole fractions were observed in the southwest corner of the SLV at the margin of recent suburban developments. At a finer scale, the high-frequency measurements show mole fraction enhancements near busy roads and intersections. Measured mole fractions are also consistently higher along a 3 km section of the light-rail track running along the center of a busy six-lane road.

3.2 Surface flux inventories

The Hestia bottom up anthropogenic CO_2 emissions inventory characterizes carbon fluxes by estimating emissions at the scale of individual buildings and roadways (Gurney et al., 2012). Hestia is available for a handful of U.S. cities including Indianapolis, Los Angeles, Baltimore/D.C., and ~~Salt Lake City~~SLC. Details pertaining to the Salt Lake County Hestia product are described by Patarasuk et al. (2016). For this simulation, Hestia anthropogenic CO_2 fluxes are aggregated hourly to a 0.002° grid (Fig. 6). However, the Hestia inventory only encompasses the SLV and requires the use of a larger scale anthropogenic emissions inventory to account for fluxes originating from outside of the SLV.

~~Anthropogenic CO_2 emissions outside of Salt Lake County are derived from the~~ We apply the $1\text{km}\times1\text{km}$ ODIAC ~~(Oda and Maksyutov, 2~~ ~~inventory . These flux estimates~~ (Oda et al., 2018) inventory to account for anthropogenic CO_2 emissions originating outside of the SLV. ODIAC is a globally available gridded dataset that uses power plant profiles and satellite-observed nightlights to spatially allocate estimates of total anthropogenic CO_2 emissions. The gridded ODIAC data are temporally allocated to hourly time steps using methods described by Nassar et al. (2013).

Within the SLV domain where the inventories overlap, Hestia and ODIAC agree on the total anthropogenic emissions to within 1.5% during our study period. However, uncertainties of fluxes applied to our analyses are likely larger since the two inventories allocate fluxes differently in space and time. Further, only Hestia is used to represent the SLV whereas ODIAC is used outside of the SLV to account for regional-scale emissions. Uncertainties in inventory estimates are difficult to quantify in time and space and require a devoted effort within the emission inventory scientific community to propagate uncertainties through underlying assumptions within each inventory (Patarasuk et al., 2016; Lauvaux et al., 2016).

The biological CO_2 inventory determines land surface types using the 2011 National Land Cover Database (Homer et al., 2007) and 1 m LIDAR derived discrete land cover classifications across the SLV . The link between land cover classification and CO_2 exchange is established using AmeriFlux eddy covariance data that provides a robust estimate of biologic fluxes from different vegetation types (Strong et al., 2011). A lookup table with independent axes for temperature, incoming shortwave radiation, and week of year is used to describe the relationship between land cover classification and biological fluxes (Strong et al., 2011) over a 0.01° grid. For details pertaining to the biological flux inventory, refer to Strong et al. (2011)~~and ?~~.

3.3 STILT configuration

STILT receptors are defined by averaging light-rail measurements hourly over a $0.002°$ grid (roughly 200 m at $\phi = 41°$ N). The grid resolution was chosen to pair analyses with the ~~Hestia inventory that has the finest spatial and temporal resolution of the flux inventories used~~$0.002°$ Hestia inventory and because $0.002°$ corresponds roughly with the size of a SLC block. This method results in 33,608 unique receptors for the month of July, 2015, necessitating the use of parallel simulations and fine scale footprint calculation included in the modified framework.

Urban development and expansion in the area surrounding SLC is limited by the mountainous topography surrounding the city and the Great Salt Lake which restrict the expansion of the city and suburbs. This confines large anthropogenic and biologic sources into a relatively small area surrounding the SLV and simplifies boundary conditions for SLV-centric modeling efforts. From each receptor, 24 h backward trajectories of 200 particle ensembles ~~are~~ were calculated using meteorological fields from the HRRR model, available at an hourly interval with a 3 km grid resolution. On average, particles travel within the model domain for 11 h. Computation of the 33,608 particle trajectories and a single set of footprints completed in 5.5 hours utilizing 80 parallel threads across 5 nodes, each equipped with 64 GB of memory with two 8-core Intel XEON E5-2670 2.6 GHz processors. 6.7% of the simulations were not completed due to short-term outages in the HRRR data product.

We compute footprint fields using the legacy dynamic grid coarsening (LEG) algorithm as well as gaussian kernel density estimation with the HNF ~~dilution~~ vertical mixing depth correction (GWD) and without the HNF ~~dilution~~ vertical mixing depth correction (GND) to illustrate the differences between methods. Further, these three methodologies are applied for two model domains, resulting in six different permutations of footprint fields for each receptor. A fine-scale $0.002°$ grid encompasses the SLV and is used to apply SLV anthropogenic emissions. A $0.01°$ grid covering a larger area of Northern Utah is used to apply biological fluxes and non-SLV anthropogenic emissions. ~~Computation of the 33,608 particle trajectories and a single set of footprints completed in 5.5 hours utilizing 80 parallel threads across 5 nodes, each equipped with 64 GB of memory with two 8-core Intel XEON E5-2670 2.6 GHz processors. 6.7% of simulations failed to complete due to short-term outages in the HRRR data product.~~

Footprints are convolved with anthropogenic and biological CO_2 fluxes and added to background CO_2 mole fractions that are representative of CO_2 mole fractions that have not been influenced by urban emissions (Mitchell et al., in press). The background mole fractions are taken from a nearby high elevation measurement site at Hidden Peak at the top of the Snowbird ski resort in the Wasatch Mountains (Stephens et al., 2011). We use a similar approach to prior studies (e.g. McKain et al. (2012)) and focus this analyses on the afternoon and early evening hours (12:00-19:00 ~~Local Daylight Time,~~ LDT) to lessen the influence of boundary layer development, nocturnal stratification of the boundary layer, and shallow turbulence on measured mole fractions that would not be represented in the 3 km resolution of the HRRR meteorological fields.

3.4 Results

Observed and simulated mole fractions are averaged by hour of day to generate mean diel cycles, shown in Fig. 5. Observations show elevated mole fractions at night and early morning, decreasing into the afternoon as convective mixing increases

(Mitchell et al., in press). All three of the simulated diel cycles derived from the different footprint algorithms systematically underestimate nighttime and early morning mole fractions, consistent with previous studies (Macatangay et al., 2008; McKain et al., 2015; Mallia et al., 2015; Lauvaux et al., 2016). However, during afternoon hours the simulated values track more closely with the observations, with GWD exhibiting closer correspondence than GND and LEG (Fig. 5).

~~Correlations~~ Explained variance over space between the time-averaged modeled and measured concentrations ~~are~~ is highest for GWD (~~$r = 0.52$~~Pearson's $r^2 = 0.27$) followed by GND (~~$r = 0.46$~~$r^2 = 0.21$) and lastly LEG (~~$r = 0.45$~~$r^2 = 0.20$). While we have demonstrated ~~GND~~ the gaussian kernel methods to compare favorably with idealized footprints calculated with brute force particle simulations, this analysis found modeled concentration differences between GND and LEG fall within the uncertainties in surface flux inventories. As GWD agrees most closely with observations over time and space, we focus on GWD for the remainder of analyses.

Footprints convolved with surface flux inventories (Fig. 6) show measurements made on the light-rail train to be highly sensitive to fluxes in the HNF domain. Spatially averaged model results capture the mole fraction gradient between the urban center and surrounding suburbs (Fig. 7). The lowest modeled mole fractions occurred in the southwest corner of the SLV, in agreement with measurements. The model ~~also reproduced~~ generally produced mole fraction enhancements (ΔCO_2) for grid cells containing or downwind from major roadways (Fig. ~~6)as well as the evening rush hour enhancement (Fig. 5)~~7). However, modeling intersection scale enhancements would require finer grid spacing capable of resolving sub-city-block spatial scales that is not yet feasible given current constraints on inventories, meteorological data, and computing resources. On average, we found the largest contributor to modeled CO_2 mole fractions is the SLV anthropogenic fluxes ($\Delta CO_2 = 4.18\text{ppm}$), followed by biological fluxes ($\Delta CO_2 = -0.89\text{ppm}$), and the smallest contribution is from non-SLV anthropogenic fluxes ($\Delta CO_2 = 0.37\text{ppm}$).

Key differences between modeled and measured mole fractions exist near HNF sources at the sub-grid scale (Fig. 7). While the model does capture localized mole fraction enhancements near busy roads and intersections ('I' in Fig. 7), measured mole fractions are systematically higher than corresponding model estimates in these areas. These results indicate that the light-rail measurement platform is sampling emissions prior to mixing with the surrounding air. This is evident along the section of light-rail track that shares the six-lane road with other vehicles ('R' in Fig. 7) on which large discrepancies between measurements and model estimates are regularly observed. By diluting emissions throughout a larger grid cell, the model predicts elevated mole fractions localized within and downwind from cells containing significant sources but does not fully capture the magnitude of enhancement resulting from the close proximity to the emissions source.

To demonstrate the benefits of the ~~fine-scale approach~~0.002° grid resolution, we use the above 0.002° grid as well as spatially degraded 0.01° and 0.1° grid resolutions ~~for~~ to convolve footprints and fluxes (Fig. 8) ~~to~~ and compare against light-rail measurements. Correlations over space between the time-averaged modeled and measured concentrations are highest for the 0.002° grid ($r = 0.52$). We find that degrading the resolution to 0.01° still captures the SLV-scale urban-suburban-rural mole fraction gradient but fails to resolve much of the roadway and intersection scale enhancements, resulting in a modest decrease in agreement with measurements ($r = 0.48$). However, degrading the resolution to 0.1° prevents the model from resolving much of the spatial mole fraction variation ($r = 0.25$). Further, evaluating the spatially-averaged concentration by hour of day

shows improved agreement with measurements among the finer grid resolutions (0.002°, 0.01°) over the more coarse 0.1° resolution (Fig. 9). While all three resolutions mimic the temporal pattern in the observed mole fraction enhancements due to the temporal variability assigned to emissions inventories, the finer resolutions better capture the mole fraction enhancements observed by the light-rail train in both time and space.

4 Summary and Conclusions

In this paper, we have introduced modifications to the STILT-R code that have improved the spatial averaging of the footprints and the model speed. These changes improve the functionality of the STILT model for applications investigating fine-scale patterns in urban emissions. We have defined the HNF using the effective vertical mixing depth of surface fluxes arriving at a receptor and shown this formulation to improve model agreement with observations. However, calculating the effective vertical mixing depth using turbulence variables σ_w and T_L does not extend well to stable nighttime conditions which remain a difficult problem (Holtslag et al., 2013) and a subject of future work.

Given the importance of footprints in the STILT workflow, a kernel density estimator was applied and shown to improve agreement with an idealized brute force method over prior methods. High level methods for single and multi-node parallelism were introduced in this distribution, significantly reducing total simulation time. We then applied STILT to simulate CO_2 mole fractions observed along a light-rail train in ~~Salt Lake City~~SLC, at high resolution and show that the model and observations track one another in terms of average spatial and temporal patterns during the afternoon period. However, key differences remain between modeled and measured mole fractions at night and in the proximity of HNF sources.

Results indicate that fine-scale inverse analyses will be sensitive to the proximity of observations to upwind sources. Modeling difficulties arise when emissions within the HNF are distributed throughout a larger grid cell that no longer reflects the close proximity of fluxes to the receptor. Fine-scale measurements and modeling approaches are useful for applications such as quantifying pollution exposure in different neighborhoods or locating large point source emitters. However, fluxes originating within the HNF domain can often dominate the modeled signal and error. Observation techniques that are strongly influenced by HNF fluxes such as trains and cars can have limited usefulness in larger scale applications for which LPDMs have previously been used, such as bulk flux estimates from urban areas. Measurements striving to quantify emissions or assess the validity of emissions inventories should seek to reduce the influence of HNF sources. Prioritizing measurement placement at the top of tall buildings or towers or at least 0.5 km from large sources such as busy roadways enables natural dilution of emitted species, reduces direct sampling of emissions, and improves model agreement with observations.

Code availability. STILT model source code and documentation can be obtained at https://uataq.github.io/stilt/. Development of the model is ongoing and updates will continue to become available through this repository. The precise version of the STILT-R model source code discussed within this manuscript is preserved at http://doi.org/10.5281/zenodo.1196561. Contributions are welcome and should be submitted via pull request. Issues should be reported to the integrated issue tracking system. Questions should be directed to the author.

Acknowledgements. This study was supported by NOAA Climate Program Office's Atmospheric Chemistry, Carbon Cycle, and Climate Program, Grant No. NA14OAR4310178. We are grateful to the STILT development community for the prior work that enabled these developments. We thank Douglas Catharine and Philip Dennison for their work assembling the Utah biological flux inventory used in analyses. The support and resources from the Center for High Performance Computing at the University of Utah are gratefully acknowledged. We thank NOAA's Air Resource Laboratory for the HRRR meteorological data. We are grateful to B. Stephens and NCAR's Regional Atmospheric Continuous CO_2 Network (RACCOON) for the high altitude CO_2 measurements used as the atmospheric background signal in analyses.

References

Apte, J. S., Messier, K. P., Gani, S., Brauer, M., Kirchstetter, T. W., Lunden, M. M., Marshall, J. D., Portier, C. J., Vermeulen, R. C., and Hamburg, S. P.: High-resolution air pollution mapping with Google street view cars: exploiting big data, Environmental Science and Technology, 51, 6999–7008, https://doi.org/10.1021/acs.est.7b00891, 2017.

Bush, S. E., Hopkins, F. M., Randerson, J. T., Lai, C.-T. T., and Ehleringer, J. R.: Design and application of a mobile ground-based observatory for continuous measurements of atmospheric trace gas and criteria pollutant species, Atmospheric Measurement Techniques, 8, 33–63, https://doi.org/10.5194/amtd-8-33-2015, http://www.atmos-meas-tech-discuss.net/8/33/2015/amtd-8-33-2015.html, 2015.

Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R. A., and Marland, G.: Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks., Proceedings of the National Academy of Sciences of the United States of America, 104, 18 866–70, https://doi.org/10.1073/pnas.0702737104, http://www.scopus.com/inward/record.url?eid=2-s2.0-36749099520{&}partnerID=tZOtx3y1, 2007.

Draxler, R. R. and Hess, G. D.: An Overview of the HYSPLIT_4 Modelling System for Trajectories, Dispersion, and Deposition, Australian Meteorological Magazine, 47, 295–308, 1998.

Duren, R. M. and Miller, C. E.: Measuring the carbon emissions of megacities, Nature Clim. Change, 2, 560–562, http://dx.doi.org/10.1038/nclimate1629, 2012.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, Journal of Geophysical Research: Atmospheres, https://doi.org/10.1029/2003JD003770, 2003.

Gurney, K. R., Mendoza, D., Zhou, Y., Fisher, M., Miller, C., Geethakumar, S., and De La Rue Dy Can, S.: High Resolution fossil fuel combustion emission fluxes for the United States, Enviromental Science and Technology, 2009.

Gurney, K. R., Razlivanov, I., Song, Y., Zhou, Y., Benes, B., and Abdul-Massih, M.: Quantification of fossil fuel CO_2 emissions on the building/street scale for a large U.S. city, Environmental Science & Technology, 46, 12 194–12 202, https://doi.org/10.1021/es3011282, http://dx.doi.org/10.1021/es3011282, 2012.

Gurney, K. R., Romero-Lankao, P., Seto, K. C., Hutyra, L. R., Duren, R. M., Kennedy, C., Grimm, N. B., Ehleringer, J. R., Marcotullio, P., Hughes, S., Pincetl, S., Chester, M. V., Runfola, D. M., Feddema, J. J., and Sperling, J.: Track urban emissions on a human scale, Nature, 525, 179–181, https://doi.org/10.1038/525179a, http://www.nature.com/news/climate-change-track-urban-emissions-on-a-human-scale-1.18311, 2015.

Holtslag, A. A., Svensson, G., Baas, P., Basu, S., Beare, B., Beljaars, A. C., Bosveld, F. C., Cuxart, J., Lindvall, J., Steeneveld, G. J., Tjernström, M., and Van De Wiel, B. J.: Stable atmospheric boundary layers and diurnal cycles: Challenges for weather and climate models, Bulletin of the American Meteorological Society, 94, 1691–1706, https://doi.org/10.1175/BAMS-D-11-00187.1, 2013.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J., and Wickham, J.: Completion of the 2001 National Land Cover Database for the conterminous United States, Photogrammetric Engineering & Remote Sensing, 73, 337–341, https://doi.org/citeulike-article-id:4035881, 2007.

Hoornweg, D., Freire, M., Lee, M. J., Bhada-Tata, P., and Yuen, B.: Cities and climate change : responding to an urgent agenda : Main report (English), Urban development series. Washington, DC: World Bank., 2, http://documents.worldbank.org/curated/en/613201468149671438/Main-report, 2012.

Idso, C. D., Idso, S. B., and Balling Jr., R. C.: An intensive two-week study of an urban CO_2 dome in Phoenix, Arizona, USA, Atmospheric Environment, 35, 995–1000, https://doi.org/10.1016/S1352-2310(00)00412-X, 2001.

International Energy Agency: World Energy Outlook 2008, Tech. Rep. 4, https://doi.org/10.1049/ep.1977.0180, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3378216{&}tool=pmcentrez{&}rendertype=abstract{%}5Cnhttp://www.smithschool.ox.ac.uk/wp-content/uploads/2010/06/Fatih-Birol.pdf, 2008.

Jette, M. and Grondona, M.: SLURM: Simple Linux Utility for Resource Management, ClusterWorld Conference and Expo CWCE, 2682, 44–60, https://doi.org/10.1007/10968987, http://www.springerlink.com/content/c4pgx63utdajtuwn/, 2003.

Kort, E. A., Eluszkiewicz, J., Stephens, B. B., Miller, J. B., Gerbig, C., Nehrkorn, T., Daube, B. C., Kaplan, J. O., Houweling, S., and Wofsy, S. C.: Emissions of CH_4 and N_2O over the United States and Canada based on a receptor-oriented modeling framework and COBRA-NA atmospheric observations, Geophysical Research Letters, 35, 1–5, https://doi.org/10.1029/2008GL034031, 2008.

Kort, E. A., Angevine, W. M., Duren, R., and Miller, C. E.: Surface observations for monitoring urban fossil fuel CO_2 emissions: Minimum site location requirements for the Los Angeles megacity, Journal of Geophysical Research Atmospheres, 118, 1–8, https://doi.org/10.1002/jgrd.50135, 2013.

Lang, M., Bischl, B., and Surmann, D.: batchtools: Tools for R to work on batch systems, The Journal of Open Source Software, 2, 135, https://doi.org/10.21105/joss.00135, https://www.theoj.org/joss-papers/joss.00135/10.21105.joss.00135.pdf{%}0Ahttp://joss.theoj.org/papers/10.21105/joss.00135, 2017.

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., and Wu, K.: High-resolution atmospheric inversion of urban CO_2 emissions during the dormant season of the Indianapolis flux experiment (INFLUX), Journal of Geophysical Research, 121, 5213–5236, https://doi.org/10.1002/2015JD024473, 2016.

Lee, J. K., Christen, A., Ketler, R., and Nesic, Z.: A mobile sensor network to map carbon dioxide emissions in urban environments, Atmospheric Measurement Techniques, 10, 645–665, https://doi.org/10.5194/amt-10-645-2017, 2017.

Lin, J. C.: Lagrangian modeling of the atmosphere, https://doi.org/10.1029/2012GM001376, http://dx.doi.org/10.1029/2012GM001376, 2013.

Lin, J. C. and Gerbig, C.: Accounting for the effect of transport errors on tracer inversions, Geophysical Research Letters, 32, 1–5, https://doi.org/10.1029/2004GL021127, 2005.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, Journal of Geophysical Research, 108, ACH 2–1–ACH 2–17, https://doi.org/10.1029/2002JD003161, 2003.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Grainger, C. A., Stephens, B. B., Bakwin, P. S., and Hollinger, D. Y.: Measuring fluxes of trace gases at regional scales by Lagrangian observations: Application to the CO2 Budget and Rectification Airborne (COBRA) study, Journal of Geophysical Research D: Atmospheres, 109, 1–23, https://doi.org/10.1029/2004JD004754, 2004.

Macatangay, R., Warneke, T., Gerbig, C., Körner, S., Ahmadov, R., Heimann, M., and Notholt, J.: A framework for comparing remotely sensed and in-situ CO_2 concentrations, Atmospheric Chemistry and Physics, 8, 2555–2568, https://doi.org/10.5194/acp-8-2555-2008, http://www.scopus.com/inward/record.url?eid=2-s2.0-43849111352{&}partnerID=tZOtx3y1, 2008.

Mallia, D. V., Lin, J. C., Urbanski, S., Ehleringer, J., and Nehrkorn, T.: Impacts of upstream wildfire emissions on CO , CO 2 , and PM 2 . 5 concentrations in Salt Lake City , Utah, Journal of Geophysical Research: Atmospheres, 120, 147–166, https://doi.org/10.1002/2014JD022472, 2015.

McKain, K., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J. R., and Stephens, B. B.: Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region., Proceedings of the National Academy of Sciences of the United States of America, 109, 8423–8, https://doi.org/10.1073/pnas.1116645109, http://www.pnas.org/cgi/content/long/109/22/8423, 2012.

McKain, K., Down, A., Raciti, S. M., Budney, J., Hutyra, L. R., Floerchinger, C., Herndon, S. C., Nehrkorn, T., Zahniser, M. S., Jackson, R. B., Phillips, N., and Wofsy, S. C.: Methane emissions from natural gas infrastructure and use in the urban region of Boston, Massachusetts., Proceedings of the National Academy of Sciences of the United States of America, 112, 1941–6, https://doi.org/10.1073/pnas.1416261112, http://www.pnas.org/content/112/7/1941.short, 2015.

Miller, S. M., Matross, D. M., Andrews, a. E., Millet, D. B., Longo, M., Gottlieb, E. W., Hirsch, a. I., Gerbig, C., Lin, J. C., Daube, B. C., Hudman, R. C., Dias, P. L. S., Chow, V. Y., and Wofsy, S. C.: Sources of carbon monoxide and formaldehyde in North America determined from high-resolution atmospheric data, Atmos. Chem. Phys., 8, 7673–7696, https://doi.org/10.5194/acp-8-7673-2008, http://www.atmos-chem-phys.net/8/7673/2008/, 2008.

Mitchell, L., Crosman, E., Jacques, A., Fasoli, B., Leclair-Marzolf, L., Horel, J., Bowling, D., Ehleringer, J., and Lin, J.: Continuous monitoring of trace gases and pollutants using a light rail public transit platform, in review.

Mitchell, L. E., Lin, J. C., Bowling, D. R., Pataki, D. E., Strong, C., Schauer, A. J., Bares, R., Bush, S. E., Stephens, B. B., Mendoza, D., Mallia, D., Holland, L., Gurney, K. R., and Ehleringer, J. R.: Long-term urban carbon dioxide observations reveal spatial and temporal dynamics related to urban characteristics and growth, Proceedings of the National Academy of Sciences, in press.

Nassar, R., Napier-Linton, L., Gurney, K. R., Andres, R. J., Oda, T., Vogel, F. R., and Deng, F.: Improving the temporal and spatial distribution of CO_2 emissions from global fossil fuel emission data sets, Journal of Geophysical Research: Atmospheres, 118, 917–933, https://doi.org/10.1029/2012JD018196, http://doi.wiley.com/10.1029/2012JD018196, 2013.

Nehrkorn, T., Eluszkiewicz, J., Wofsy, S. C., Lin, J. C., Gerbig, C., Longo, M., and Freitas, S.: Coupled weather research and forecasting-stochastic time-inverted lagrangian transport (WRF-STILT) model, Meteorology and Atmospheric Physics, 107, 51–64, https://doi.org/10.1007/s00703-010-0068-x, 2010.

Oda, T. and Maksyutov, S.: A very high-resolution (1km x 1 km) global fossil fuel CO_2 emission inventory derived using a point source database and satellite observations of nighttime lights, Atmospheric Chemistry and Physics, 11, 543–556, https://doi.org/10.5194/acp-11-543-2011, 2011.

Oda, T., Maksyutov, S., and Andres, R. J.: The Open-source Data Inventory for Anthropogenic CO_2, version 2016 (ODIAC2016): A global monthly fossil fuel CO_2 gridded emissions data product for tracer transport simulations and surface flux inversions, Earth System Science Data, 10, 87–107, https://doi.org/10.5194/essd-10-87-2018, 2018.

Pataki, D. E., Xu, T., Luo, Y. Q., and Ehleringer, J. R.: Inferring biogenic and anthropogenic carbon dioxide sources across an urban to rural gradient, Oecologia, 152, 307–322, https://doi.org/10.1007/s00442-006-0656-0, 2007.

Patarasuk, R., Gurney, K. R., O'Keeffe, D., Song, Y., Huang, J., Rao, P., Buchert, M., Lin, J. C., Mendoza, D., and Ehleringer, J. R.: Urban high-resolution fossil fuel CO_2 emissions quantification and exploration of emission drivers for potential policy applications, Urban Ecosystems, 19, 1013–1039, https://doi.org/10.1007/s11252-016-0553-1, 2016.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2017.

Shusterman, A. A., Teige, V., Turner, A. J., Newman, C., Kim, J., and Cohen, R. C.: The BErkeley Atmospheric CO_2 Observation Network: initial evaluation, Atmospheric Chemistry and Physics Discussions, pp. 1–23, https://doi.org/10.5194/acp-2016-530, http://www.atmos-chem-phys-discuss.net/acp-2016-530/, 2016.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System, https://doi.org/10.1175/BAMS-D-14-00110.1, 2015.

Stephens, B. B., Miles, N. L., Richardson, S. J., Watt, A. S., and Davis, K. J.: Atmospheric CO_2 monitoring with single-cell NDIR-based analyzers, Atmospheric Measurement Techniques, 4, 2737–2748, https://doi.org/10.5194/amt-4-2737-2011, 2011.

Stohl, a., Forster, C., Frank, A., Seibert, P., and Wotawa, G.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmospheric Chemistry and Physics Discussions, 5, 4739–4799, https://doi.org/10.5194/acpd-5-4739-2005, 2005.

Strong, C., Stwertka, C., Bowling, D. R., Stephens, B. B., and Ehleringer, J. R.: Urban carbon dioxide cycles within the Salt Lake Valley: A multiple-box model validated by observations, Journal of Geophysical Research Atmospheres, 116, 1–12, https://doi.org/10.1029/2011JD015693, 2011.

Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., Joe, P., Barker, D. M., Li, P. W., Golding, B., Xu, M., and Pinto, J.: Use of NWP for nowcasting convective precipitation: Recent progress and challenges, Bulletin of the American Meteorological Society, 95, 409–426, https://doi.org/10.1175/BAMS-D-11-00263.1, 2014.

Taylor, G. I.: Diffusion by continuous movements, Proceedings of the London Mathematical Society, s2-20, 196–212, https://doi.org/10.1112/plms/s2-20.1.196, 1922.

Turner, A. J., Shusterman, A. A., McDonald, B. C., Teige, V., Harley, R. A., and Cohen, R. C.: Network design for quantifying urban CO_2 emissions: Assessing trade-offs between precision and network density, Atmospheric Chemistry and Physics, https://doi.org/10.5194/acp-16-13465-2016, 2016.

Vogelezang, D. H. P. and Holtslag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height formulations, Boundary-Layer Meteorology, 81, 245–269, https://doi.org/10.1007/BF02430331, 1996.

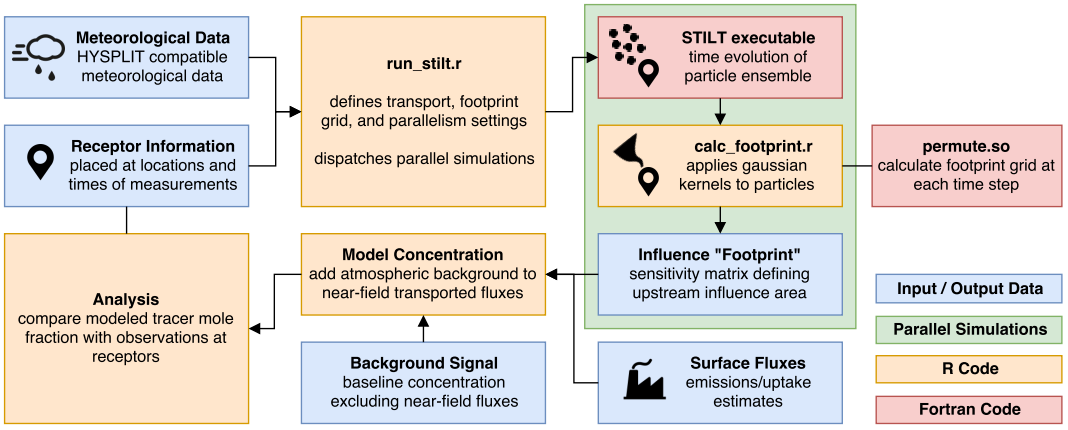

Figure 1. STILT workflow to model tracer mole fraction at a receptor. STILT advects particles and calculates the influence footprint for each receptor. Footprints are convolved with surface fluxes and an atmospheric background signal to model the tracer mole fraction.

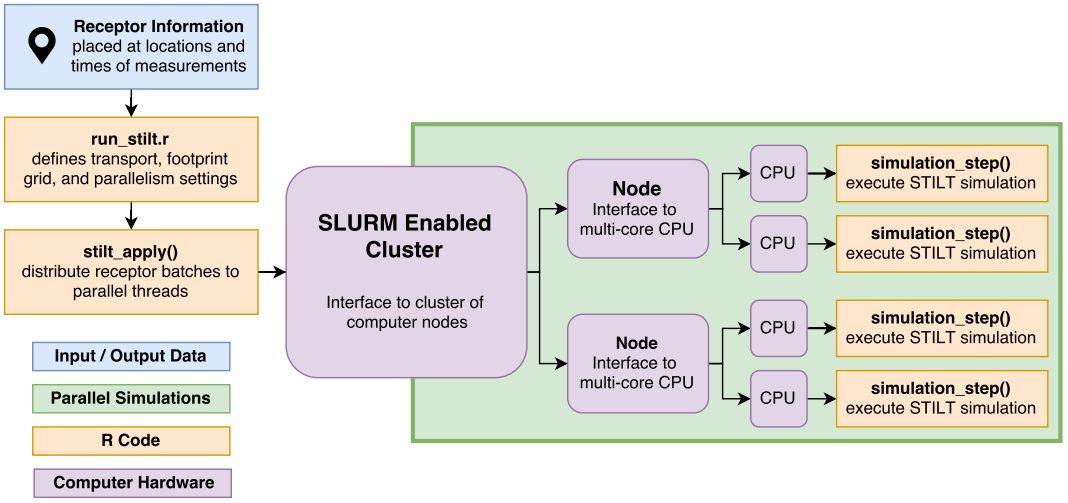

Figure 2. Receptor batches are distributed across parallel threads to enable multiple concurrent simulations. Provided memory limits are not exceeded, the total simulation time decreases linearly with the number of CPU cores available.

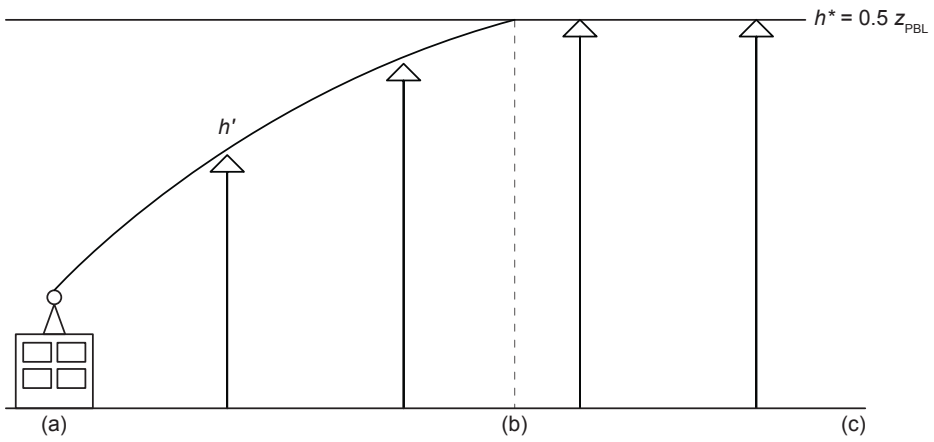

Figure 3. Growth of the effective mixing depth. From the receptor (a), surface fluxes are diluted within an atmospheric column depth of h' in the HNF until $h' = h^*$ (b), amplifying the contribution of Hyper ~~Near-Field~~ Near-Field (HNF) sources and sinks on the receptor. Once h' has reached h^*, surface fluxes are diluted to depth h^* until the end of the simulation (c).

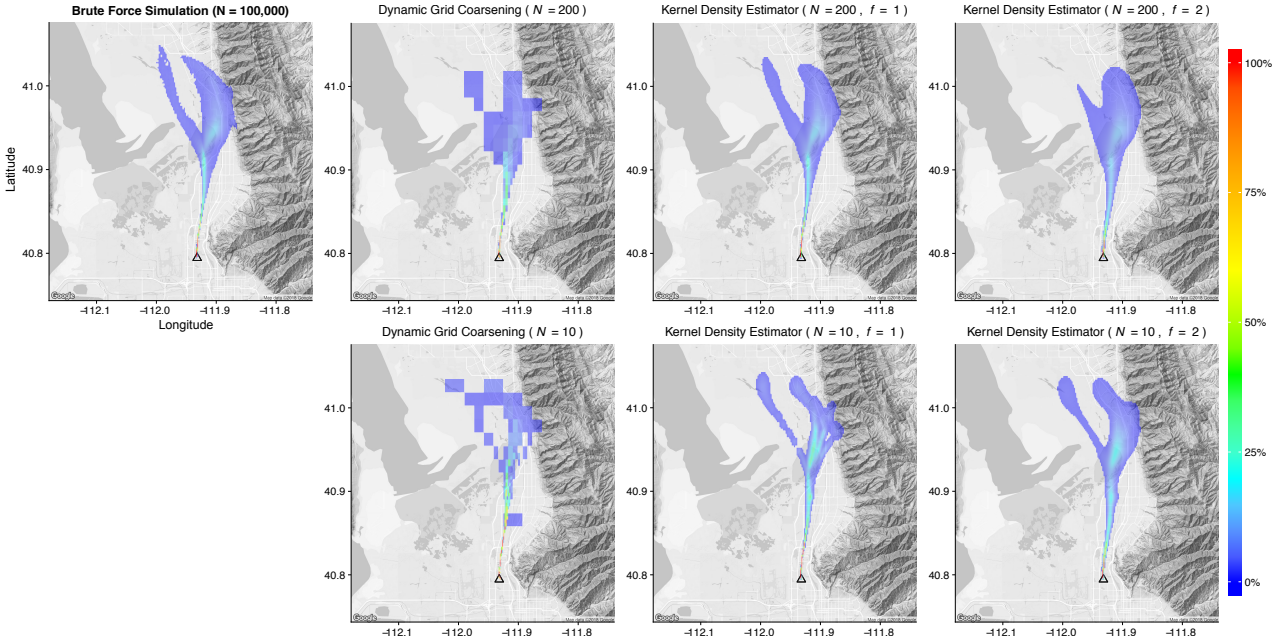

Figure 4. Comparison of footprint calculation methods. Simulating a large number of particles ($N = 10^5$) and gridding by location (top left) gives a physically constrained expectation for the footprint. Using subsets of 200 particles (top) and 10 particles (bottom), the kernel density estimator demonstrates considerable improvements over the traditional dynamic grid coarsening. Modifying the kernel bandwidths ($f = 2$) can improve results in uncommon cases, such as the 10 particle ensemble.

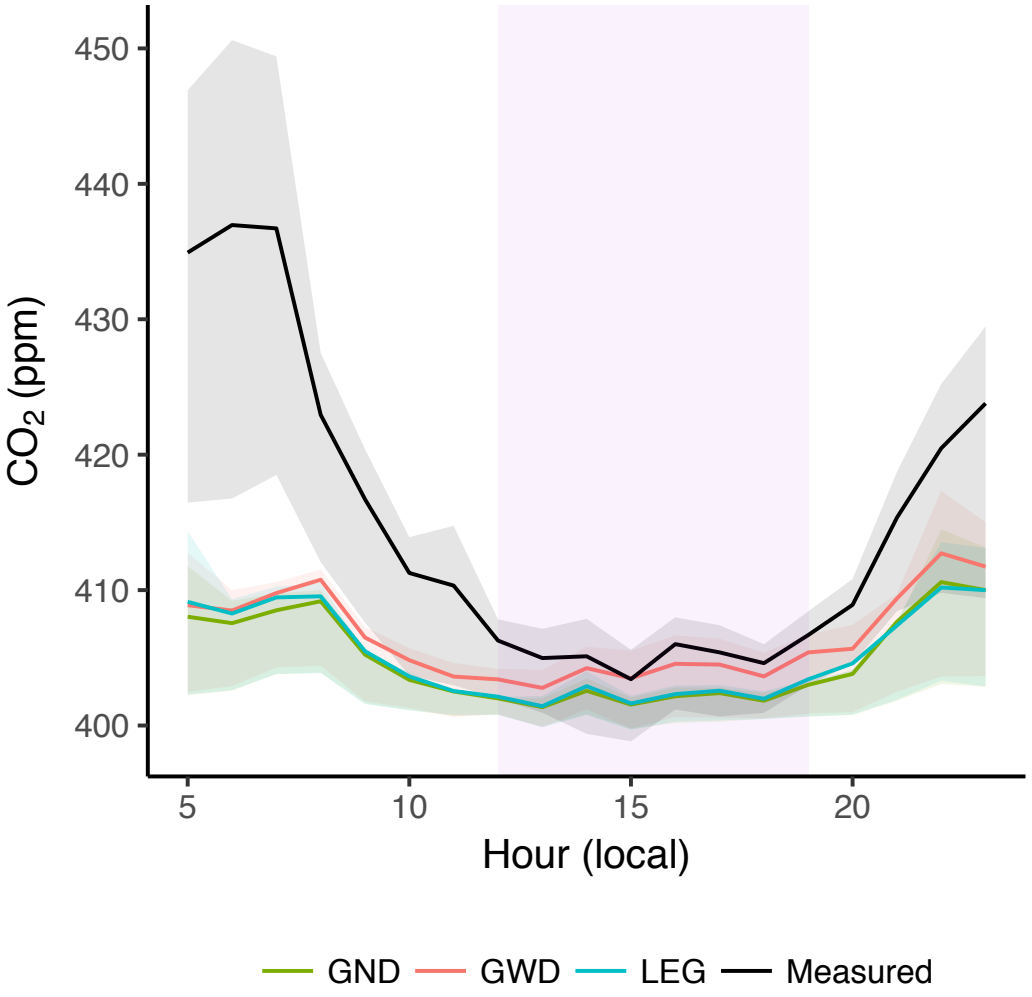

Figure 5. Mole fractions of various footprint calculation methods by hour of day during light-rail operating hours from early morning to late evening. Purple shading indicates afternoon hours (05:00-23:00 Local Daylight Time) used for analyses. Solid lines represent the mean and shading represents the interquartile range. Mole fractions modeled using Gaussian kernel calculated footprints with correction for HNF dilution depth (GWD) modeled mole fractions agree most closely with measurements, with underestimation attributed to sub-grid scale sampling of emissions sources.

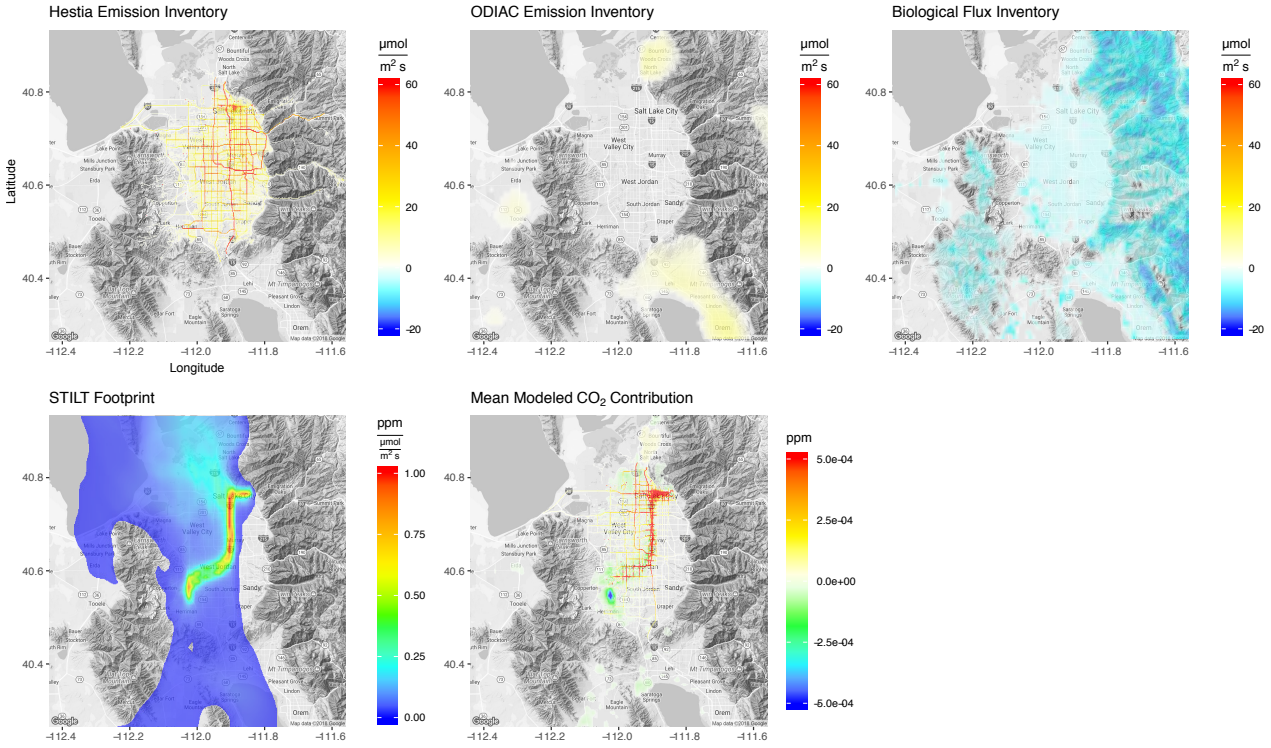

Figure 6. July 2015 afternoon Salt Lake Valley (SLV) Hestia-derived and non-SLV ODIAC-derived anthropogenic CO_2 emissions, biological fluxes, and average STILT footprint. The anthropogenic and biological flux inventories convolved with the footprints give the contribution of near-field fluxes to measured mole fractions in ppm. The light-rail train is highly sensitive to HNF emissions sources and is strongly influenced by large roadways and agriculture adjacent to the line.

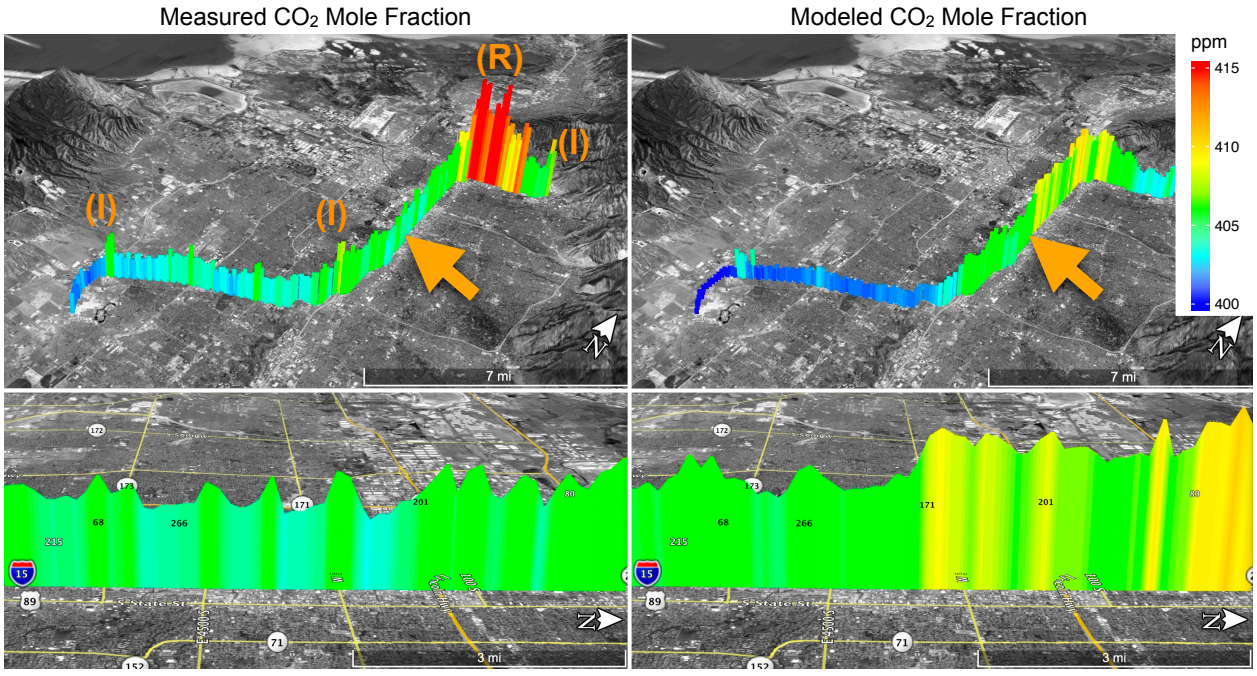

Figure 7. Key differences between measured and modeled tracer mole fraction occur near HNF sources, including passing large roadways and intersections (I) and where the light-rail track is shared by other vehicles on the roadway (R). Orange arrow indicates viewpoint of bottom panels. The model captures the overall urban-suburban-rural CO_2 mole fraction gradient (top) as well as localized enhancements near ~~many~~ grid cells containing large emitters such as busy roads (bottom). Map data: Google, Landsat / Copernicus.

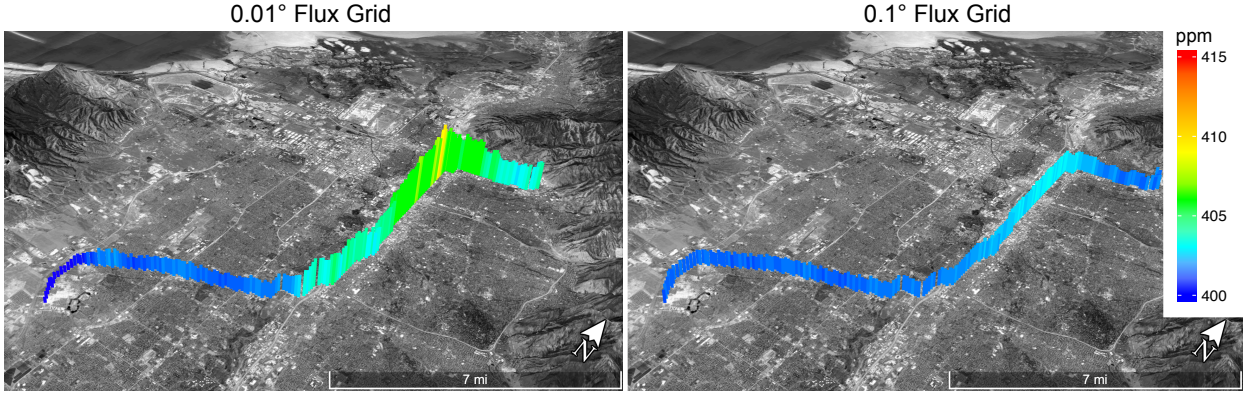

Figure 8. Spatially degraded flux and footprint grids to 0.01° and 0.1° resolutions demonstrates the advantages of the fine-scale, 0.002° grid (Fig. 7). While the 0.01° resolution (left) retains the CO_2 mole fraction enhancements for the SLV-scale urban-suburban-rural gradient, the 0.1° resolution (right) fails to resolve the locations and magnitudes of observed CO_2 variations. Map data: Google, Landsat / Copernicus.

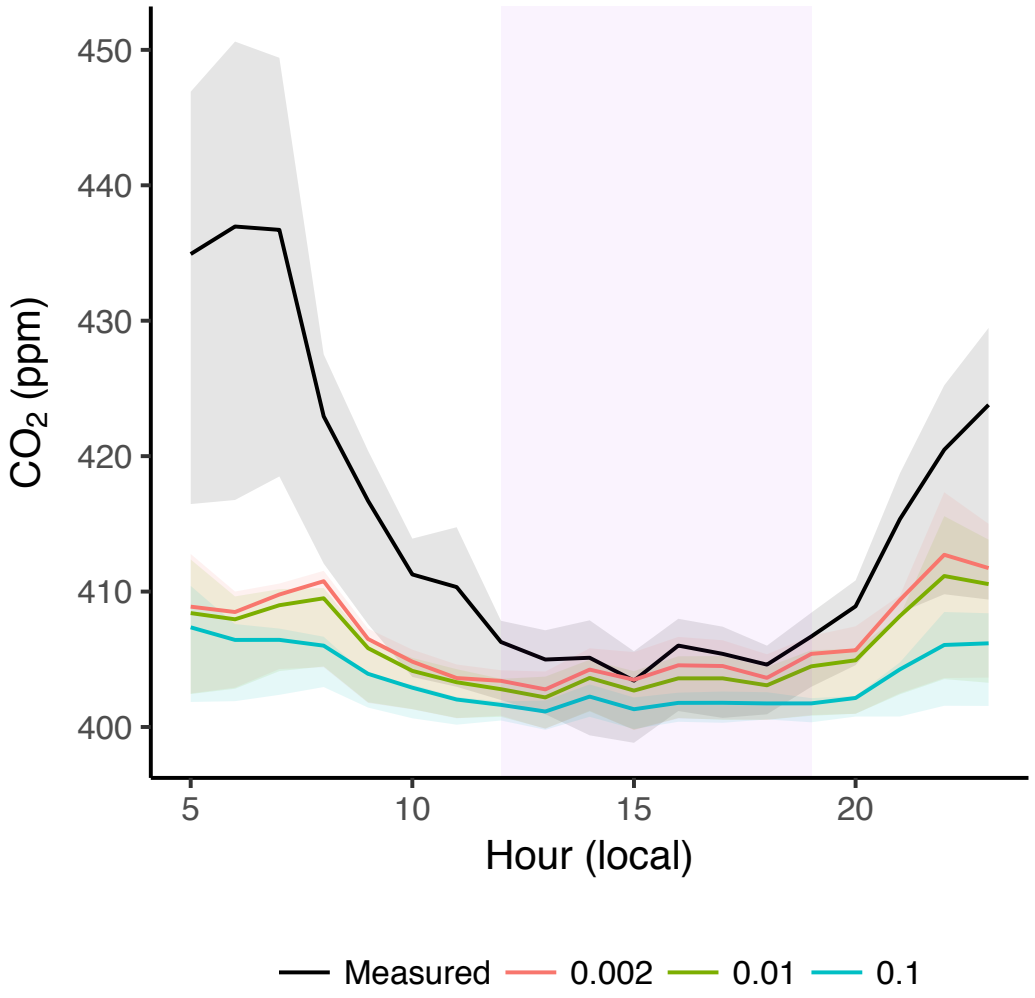

Figure 9. Modeled mole fraction at 0.002° as well as spatially degraded 0.01° and 0.1° grid resolutions by hour of day during light-rail operating hours. Purple shading indicates afternoon hours used for analyses. Solid lines represent the mean and shading represents the interquartile range. Mole fractions using finer grid resolutions (0.002°, 0.01°) agree more closely with observations than the coarser 0.1° due to the close proximity of the light-rail train to emissions sources.