Reply to reviewer #2

This paper presents the GRAMM/GRAL model setup to simulate CO_2 concentrations in the city of Zurich. The manuscript is generally well-written and presents robust results, forming a solid basis for future studies. However, additional analysis and more in-depth discussion are necessary before the manuscript is ready for publication in *Atmospheric Chemistry and Physics*.

Major Comments

- **Line 330: Selection of sites for matching**

  Your claim is that the model can resolve flow within street canyons. Therefore, including stations located within street canyons for model validation should be appropriate. If including these stations deteriorates the model performance, this may indicate that GRAMM/GRAL does not adequately capture the street-level flow, raising questions about the justification for using it in such environments.

The idea of the matching procedure is to capture the mesoscale flow, which then forces the microscale flow including the flow around buildings. Even at the high resolution of our model, it is necessary to select sites measuring winds with sufficient spatial representativeness. It is true that we did not evaluate the quality of the microscale flow in this study. Unfortunately, the wind measurements in street canyons in Zurich (Stampfenbachstrasse, Rosengartenstrasse, Schimmelstrasse) are not suitable for this purpose. These measurements are taken at air quality stations a few meters away from busy roads and trees (Rosengartenstrasse, Schimmelstrasse) or very close to a building facade (Stampfenbachstrasse), situations that can not be represented by the model.

Although we didn't explicitly claim that we can resolve the flow in street canyons, this may have been concluded from our statement that "*the model operates at a high horizontal, building-resolving resolution 10 m x 10 m, which is required to capture the situation at the low-cost sensor sites at street level*". Since most street canyons in Zurich are no wider than a few tens of meters, a resolution of 10 m poses clear limitations to representing the flow in these canyons. A higher resolution (5 m or better) would be desirable, but so far has been prohibitive due to memory limitations (the model has only OpenMP but no MPI parallelization). We will add the following sentences in the introduction clarifying the capability and limitations of a model at this resolution:

The model operates at a high horizontal, building-resolving resolution, which is required to capture the situation at the low-cost sensor sites at street level, but it may also be an advantage to represent the CO_2 concentrations at rooftop level. Due to computational constraints, the horizontal resolution was limited to 10 m, which is sufficient to capture the flow in wide streets and open spaces such as parks, squares and rivers but it poses limitations for resolving the flow in most street canyons in Zurich, which are no wider than a few tens of meters. Despite these limitations, the resolution is orders of magnitude better compared to atmospheric transport models used in previous urban CO_2 studies, which typically operated at resolutions of one kilometer or coarser (Staufer et al., 2016; Feng et al., 2016; Turner et al., 2020; Lauvaux et al., 2020; Nalini et al., 2022). Computational fluid dynamics models such as OpenFoam (Kubilay et al., 2018) and PALM (Maronga et al., 2020) have also been run over cities at building-resolving scale down to sub-meter resolution, but these simulations were limited to individual districts or short time periods due to the excessive computational costs.

Furthermore, we added a figure (new Fig. 15) comparing simulated CO_2 with one of the traffic sites. The site Schimmelstrasse was selected for this, as the model's performance is in between the other two traffic sites Stampfenbachstrasse and Rosengartenstrasse. We added the following text:

Figure 15 shows the time series of daily afternoon CO_2 at Schimmelstrasse, one of three street-level traffic sites. While the observed seasonal and day-to-day variability is generally well captured, there is a tendency to overestimate CO_2 at this site. An even greater tendency to overestimate CO_2 is evident at the Rosengartenstrasse traffic site, whereas at the third traffic site (Stampfenbachstrasse), the model's performance is comparable to that at rooftop sites (not shown). One likely reason for the overestimation at Schimmelstrasse and Rosengartenstrasse is the model's limited capability to resolve flow in street canyons. Visual inspection of the model simulations suggests that wind speeds are underestimated in many street canyons due to limited resolution. As a result, CO_2 accumulates too strongly. Another possible factor is that the model does not represent turbulence induced by traffic. This is consistent with several modelling studies (e.g. Wang and Zhang, 2009) that have demonstrated the importance of vehicle-induced turbulence for predicting the spatial gradients of air pollutants near roadways.

- **Line 330: Omission of the mountain station**

  Excluding the mountain station is a limitation, especially since the manuscript claims GRAMM/GRAL performs particularly well at such sites. While it is true that models often struggle to represent flow at mountain tops, this issue warrants discussion. Additionally, excluding higher elevation sites may point to problems in representing the vertical concentration profile. Please address these points in the revised version.

We agree that this deserves further discussion. We've added the following explanation:

The simple approach of driving a GRAMM simulation with a single vertical profile seems not to be able to capture the actual contrast between the flow near the surface and the free troposphere. Winds are measured at Uetliberg at 1043 m above mean sea level (amsl) on top of a 187 m tall television tower on the highest mountain in Zurich, often above the planetary boundary layer. Hilltop sites at lower elevation including ETH Hönggerberg (540 m amsl), Fluntern (556 m amsl), or Gubrist Gipfel (615 m amsl), on the other hand, were included in the matching procedure. All mid-cost CO_2 measurements presented in this study were collected below these hilltop sites. To better capture actual wind profiles throughout the lower troposphere, it would likely be beneficial to use the newly developed GRAMM-SCI model (Oettl and Verati, 2021; Oettl 2021), which allows nesting the model into ERA5 reanalysis fields of the European Centre for Medium-Range Weather Forecasts. However, driving with ERA5 winds would require a largely different approach to generating the catalogue.

- **Section 2.8: Background concentration**

  Please discuss how the choice of background concentration influences the results. Can you provide an uncertainty estimate, perhaps based on the spread among background stations? This is particularly important given your observation that a substantial portion of the concentration signal and its variability originates from the background. Moreover, this discussion is essential groundwork for any future inverse studies.

Uncertainties in the background will indeed be a critical aspect for the inverse modeling. We have revised and extended the discussion of background concentrations in two ways: First, instead of showing hourly values, we now present daily afternoon means but for the full 2 years instead of only a few months. Second, we added a panel b to Figure 10 showing pairwise differences between the background stations as a function of wind direction. To the first point, the paragraph was changed to

Figure 10a shows daily afternoon mean (12-16 UTC, 13-17 local time in winter) CO_2 concentrations at the three sites together with the weighted mean background for the two years of simulation. Here and in the following we mostly limit the analysis to afternoon concentrations as the model struggles reproducing nighttime concentrations (see Sect. 3.2). Overall, the concentrations followed each other rather closely. The site Breite-Birchwil to the northeast of the city occasionally showed large deviations from the other sites, especially during winter and most prominently on a few days in December 2023 and late January 2024.

To the second point, the following paragraph was added. The last sentence summarizes the expected uncertainty in background concentrations.

Uncertainties in background levels will be crucial for the inverse estimation of CO_2 emissions. Figure 10b shows the differences between individual station pairs as a function of wind direction. Wind directions in which a given station is located perfectly upstream of Zurich are labelled. The frequency of occurrence of individual wind directions (grey line) shows two distinct peaks for southwesterly to westerly and for northwesterly (Bise) directions. During southwesterly to westerly winds, Beromünster and Lägern are located upstream, with their concentrations agreeing within about 2.5 to 4.5 ppm. During Bise situations, Lägern and Beromünster showed average differences of about 4 ppm from the upstream Breite-Birchwil site. The largest differences were observed during southeasterly to southerly winds, but these situations were infrequent. As previously mentioned, Breite-Birchwil occasionally had much higher concentrations when winds blew from the southwest and the station was in the city's outflow. The lowest station-to-station differences of around 2–3 ppm occurred during northwesterly to northerly winds, but these situations were rare. These station-to-station differences provide an indication of the uncertainty of the background levels. Assuming that the true uncertainty is somewhat lower because the background is a weighted average of the three stations, uncertainties in the background are probably in the order of 1 to 3 ppm.

- **Line 384: Limited presentation of data**

  Since this paper aims to demonstrate the model’s capabilities (and is not an inverse study relying only on afternoon values), presenting only the mean diurnal cycle and two stations during the afternoon is insufficient. I recommend showing the full time series for the two selected stations and then arguing why a focus on afternoon values reduces mismatches. Otherwise, the comparison may be misleading and overly optimistic.

We added the full time series to the appendix. Note that the mean diurnal cycles also show the range in terms of +/- 1 standard deviation of the hourly values. We think this a clearer way of presenting the agreement (or disagreement) between simulations and observations than showing a busy time series of hourly values, where it becomes difficult to see the details.

- **Figure 15: Large discrepancies in diurnal cycles**

  The differences between modeled and observed diurnal cycles are substantial. While you mention possible causes such as incorrect VPRM input, flawed scaling factors, or PBLH errors, the discussion remains superficial. Given Zurich's extensive observational infrastructure, these discrepancies should be examined in more detail—for example, through comparison with other top-down estimates or vertical mixing data from tall towers. The discussion of catalog probabilities is weak and inconclusive; consider moving this part to the appendix and replacing it with a more thorough analysis of the discrepancies.

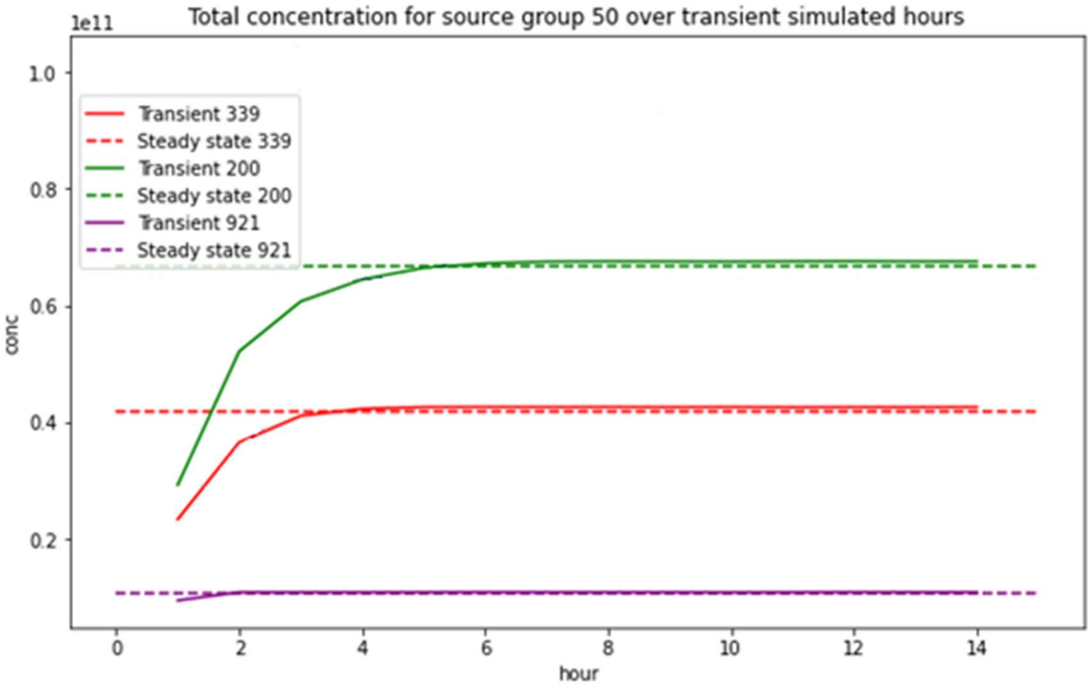
The discussion of the potential reasons for the substantial differences in the diurnal cycles is indeed somewhat unsatisfactory. We tested many hypotheses but, unfortunately, were unable to identify one major driving factor. One of the hypotheses was that the static hourly solutions of the catalogue do not properly account for the accumulation of CO_2 over multiple hours, which is expected to be particularly important under stable low-wind conditions. See our reply to the corresponding point below. Another possibility that we mentioned is too strong mixing at night, possibly due to the selection of not sufficiently stable situations. In response to reviewer #1, we added a discussion of the impact of selecting a wrong situation. Finally, background concentrations deduced from stations dozens of kilometers away may not be appropriate for low wind speed situations. Another approach for stable night-time conditions with low winds could be to use the concentrations in the city in the preceding afternoon as background. However, applying this idea would require substantial additional work as it will have to be tested under which situations this is useful and how the method could be blended with the existing background method, for example depending on wind speed and stability.

The discussion of the discrepancies in the mean diurnal cycles was thus expanded with a discussion of the dynamic simulations and with the potential impact of selecting a wrong stability class. We feel that the paper is already very long and that additional discussions would dilute the overall message. We will certainly continue investigating this aspect, for example following the above-mentioned idea of using afternoon concentrations as background for nights with stable low-wind conditions.

- **Line 475: Dynamic CO_2 simulation**

You refer here to simulating CO_2 in a dynamic manner, but this concept is introduced rather abruptly in the discussion section without being presented in the results. Please clarify and provide context. This aspect deserves better integration into the manuscript.

We agree that this was not explained in sufficient detail. GRAMM/GRAL allows simulating in dynamic mode, i.e. Lagrangian particles are transported in a dynamically changing flow field. Because of the strong underestimation of CO_2 in stable situations at night, we suspected that the static approach may not properly represent the accumulation of CO_2 over many hours. To check this, we performed simulations for a few selected situations in dynamic mode (but keeping the same situation over the simulation). The concentrations in the dynamic simulations increased with time until they reached a steady state, which almost exactly matched the concentration levels of the static simulation for the same situations. This convinced us that the catalogue of static situations was built in a way that fully accounts for the buildup of CO_2 over multiple hours.
As an example, the figure below compares the results between transient (solid lines) and static simulations from the catalogue (dashed lines) for three situations with different stability and wind speed. The figure shows domain-integrated mass of CO_2 emitted by one arbitrarily selected source group. It shows that the transient simulations reach a constant level very closely matching the results from the static simulations. As expected, it takes longer to reach the steady state for an extremely stable situation with low winds of 0.25 m/s (situation 200, green) than for a neutral situation with stronger winds of 4 m/s (situation 921).



We moved the discussion of these dynamic simulations from the conclusions to the discussion of the comparison of the seasonal mean diurnal cycles in Section 3.2 and expanded it as follows:

We also investigated whether the static CO_2 distributions of the catalogue account for the potential accumulation of CO_2 over multiple hours. Such accumulation is expected to be particularly important under stable low-wind situations. To test this, we performed a series of dynamic simulations over 14 hours for one arbitrarily selected source group and for different stability classes and wind speeds. These simulations are identical to those performed when building the catalogue, but the Lagrangian particles carrying the emitted CO_2 are followed over the full time of the simulation or until they leave the domain. In these dynamic simulations, CO_2 integrated over the whole domain was initially zero and then increased with time until it reached a steady-state equilibrium between emissions and loss through transport out of the domain. The equilibrium was reached after about 7 hours in case of an extremely stable situation (stability class G) with low winds (0.25 ,m s^{-1}) but already after less than 2 hours in case of a neutral situation (stability class D) with stronger winds (4 m s^{-1}). These equilibrium levels were almost identical to the corresponding catalogue entries, confirming that accumulation over multiple hours is fully accounted for by the static solutions of the catalogue.

- **Line 479: Potential for further investigation**

  As mentioned earlier, there are additional data and modeling resources that could help investigate the discrepancies observed in this study. It would strengthen the manuscript to make use of these tools or at least outline how they could be used in follow-up work.

There is indeed additional data that could help investigate the model's performance. As mentioned in the manuscript, we made use of ecosystem observations collected in Zurich to evaluate the performance of the VPRM model (see Stagakis et al., 2025; https://bg.copernicus.org/articles/22/2133/2025/). This analysis showed comparable performance to other vegetation models, but the analysis was limited to urban parks and gross photosynthetic production (GPP) could only be evaluated qualitatively. Additional work is in progress: Li et al. have just submitted a publication comparing different VPRM versions and evaluating them against ecosystem observations in both Zurich and Munich and including a quantitative evaluation of GPP. Still, one major drawback is that so far no observations are available for street trees. Such observations are currently being collected in the framework of a Swiss research project.

Another potentially useful data set is vertical wind profiles from two Doppler lidars deployed in Zurich during about 7 months. The first lidar (Metek Windranger 200) probed the lowest 100 m above ground, the second (Leosphere Wind Cube 100) higher altitudes starting at 200 m. Unfortunately, the profiles of the Windranger were strongly influenced by nearby high-rise buildings, but the measurements at the highest level (100 m) agreed very closely with the in-situ measurements at Hardau II. Since the latter were included in the catalogue matching procedure, the measurements from the Windranger do not add much information. Profiles from the second lidar measuring above 200 m, on the other hand, could be useful to evaluate the GRAMM/GRAL wind profiles. Since the lidar was deployed for only a few months, its measurements could not be integrated into the catalogue selection procedure.

A third potentially useful data set is the Eddy covariance measurements at the Hardau tower. Some of these measurements have been described by Hilland et al. (2025, under review), a publication now cited in our manuscript. These measurements could potentially be used to evaluate the diurnal cycles of the probability distribution of stability classes presented in Figure 16 (now Fig. 17). However, flux measurements performed on top of a tall building at 112 m above ground in a heterogenous urban area need to be treated with great care, e.g., to account for issues such as vertical decoupling during stable situations, storage fluxes, distortion of the flow by the building, etc. There is not enough space in this manuscript to describe such a complex data set and how it could be applied to determine stability classes. Further studies using the Eddy covariance observations are currently in preparation, which will serve as a much better basis and reference for future analyses.

Instead of providing additional analyses, we added a paragraph at the end of the results section describing different potential sources of error and how additional observations could help address them in follow-up studies. Please see our response to reviewer #1 asking for "additional quantification of the uncertainties in each aspect of the model".

Minor Comments

- **Line 58:** Please clarify what "As an alternative" refers to, or consider removing the phrase for clarity.

Thank you, this was not clear. The sentence was reformulated as follows:

To enable an alternative method for quantifying the CO_2 emissions of the city, an Eddy covariance system for direct CO_2 flux measurements was installed on a 16.5 m mast on top of one of the tallest (95.3 m) buildings in the city. These measurements are presented in Hilland et al. (2025) and were not used in this study.

- **Line 208ff:** Remove the brackets for Glauch et al. 2025.

The reference was changed as suggested.

- **Table 1, last line:** Please delete this line.

This was a mistake also pointed out by reviewer #1. The line was deleted.

- **Line 251:** The sentence "Possible differences … cannot be accounted for in this way" is not entirely accurate. In principle, these differences could have been addressed by distinguishing between inner-city and outer emissions through separate source groups.

This is true. We added
.. unless they are simulated as separate categories.

- **Line 306:** The statement "requires a uniform scaling of the fluxes from a given vegetation type" suggests a limitation, but this could be addressed with a more refined approach.

This is indeed a limitation and we don't think it can easily be overcome, unless vegetation in different parts of the city (or different altitudes) is treated as separate tracers.

- **Line 358:** "Rather small gradients" is vague. Please provide a quantitative value or range.

They were of the order of 4 ppm. We will add this information. Note that the section on the background has been revised significantly.

- **Line 461:** It would be appropriate here to acknowledge that GRAMM/GRAL struggles to reproduce dynamics outside of afternoon hours.

These difficulties are mentioned a few lines later.

- **Literature:** Please refer To Vardag and Maiwald (2024), who have already applied GRAMM/GRAL for inversion studies.

Thank you, this reference was badly missing. We already added it in response to a request of reviewer #1 to include more recent relevant literature in the introduction (see corresponding response to reviewer #1).