The manuscript by Lena Feld et al. presents results from a field campaign conducted in Thessaloniki, Greece, aimed at estimating CO_2 emissions using two ground-based remote sensing FTIR spectrometers (EM27/SUN). Simulated XCO_2 was generated using the ICON-ART numerical weather prediction model, incorporating anthropogenic emissions from the high-resolution ODIAC inventory and biogenic fluxes from a combination of three datasets. CO_2 emissions were then inferred by scaling the modeled XCO_2 to match the observations.

This study demonstrates a valuable application of the ICON-ART model supported by ground-based total column measurements. The use of such measurements helps to reduce uncertainties in current emission estimates and shows potential for extension to other trace gases. However, some aspects, such as the uncertainty and sensitivity of the model need further investigation. This study can be considered for publication after the authors carefully address the concerns outlined below.

Specific comments:

Line4&5: Please spell out the full names of 'EDGAR' and 'ODIAC' upon first mention in the abstract for clarity, especially for readers who may not be familiar with these acronyms.

Line 118: The reported XCO_2 enhancements reached up to 2.03 ppm. I agree with the other reviewer that this signal appears relatively modest. Could the authors clarify the distance between the two instruments, as well as the wind conditions on that particular day? It is possible that the relatively small enhancement was influenced by the instrument spacing, mild wind speeds, or variable wind directions.

It would be helpful to compare this result with findings from similar urban campaigns. For example, in the early Berlin study (doi:10.5194/amt-8-3059-2015), a moderate XCO_2 enhancement of approximately 4 ppm was observed at the downwind site.

Line 124: The authors mentioned that simple models, such as the box model, are not suitable for this study and instead proposed a more precise simulation using ICON-ART. Could the authors clarify whether ICON-ART has previously been applied to greenhouse gas or trace gas emission studies? If so, referencing relevant prior work would help support its use in this context.

Line 141: the authors state that they “set the surrounding background concentration to 0 ppmv”. Could the authors provide more justification for this assumption? What are the characteristics of the surrounding areas, and is there evidence to suggest they have no significant influence on the target region? Including a spatial distribution of XCO_2 derived from satellite observations (or from CAMS XCO_2 predictions?) could help assess whether notable emission sources are present in the surrounding areas and validate this assumption.

Line 199: The purpose of the spatial regridding is not entirely clear for me. Specifically, it's unclear why four 1 km × 1 km pixels (from the original ODIAC inventory, which already matches the simulation resolution) are merged into larger 4 km² pixels for the city with the highest emissions. Since the ODIAC inventory has a comparable resolution to the model grid, why not retain the original 1 km × 1 km pixel resolution for emission allocation? Please clarify the rationale behind this regridding step and explain how it supports the goal of rescaling the inventory after the simulation runtime.

Line 214: The sentences of "Outside the subdivided area the emissions of the ODIAC inventory remain unchanged. The scaling factors, were restricted so that the prior could only be scaled in the range (1/6, 6), that was empirically chosen." are unclear to me. Could the authors clarify why emissions outside the subdivided area were kept constant? Additionally, why restricting the scaling factors to the (1/6, 6) range, and how was this empirical range determined?

Line 218-219: I do not think the cited paper by Frey et al., 2019 is relevant to the discussion of trajectories or the Tokyo campaign.

Line 224: the authors use two days of closely spaced measurements (~500 m apart) to assess the impact of spatial heterogeneity. Based on Figure 2, I assume the first instrument was located at Campus and the second at Seych Sou, which appears to be a more mountainous region. It would be helpful to explicitly state the locations of these measurements in the manuscript to provide better context for the reader.

Additionally, I'm curious about the wind conditions during these two days. Was there any prevailing wind direction that could have transported emissions from one site to the other? If so, this might have influenced the observed enhancements and should be discussed as a potential factor in the interpretation. A discussion of wind-related differences between the sites could further strengthen the assessment of spatial heterogeneity and its impact on the measurements.

Furthermore, during the calibration days, there appear to be significant biases of up to 1 ppm. Any reasons for this (e.g., could they be related to higher solar zenith angles which should be filtered?). This level of bias is comparable to the ΔXCO2 signal in this study used to estimate emissions.

Line 277-279: "This indicates that the variability in the simulated data is dominated by the biogenic source and sink for these days……" This statement would benefit from additional supporting evidence. Are there any specific indicators, such as back trajectories?

On several days (as 2022-07-07), the simulation aligns well with the observations. Does this suggest that ICON-ART can reliably model biogenic contributions and no significant anthropogenic emissions were generated on these days? Additionally, is it possible to isolate and

evaluate the biogenic and anthropogenic components separately within the simulation to better support this interpretation?

The no scaling XCO_2' values appear to be significantly lower in October compared to June and July, which does not align with the observations. Could this higher discrepancy in October relate to an underestimation of biogenic sources in the ICON-ART model? Or could this discrepancy be related to the background removal method? Specifically, the use of a constant daily background based on the 5th percentile may suppress meaningful variability and potentially remove part of the signal of interest. Given that two EM27/SUN instruments were deployed, was it considered to use the upwind site as a dynamic background reference? Alternatively, if satellite data are available, would using upwind satellite observations provide a more representative background than the current approach?

Technical comments:

Figure4: please specify the legend for the two different lines in the figure.

Line 135: 0,808 km >>> dot

Line 215: "This implies also that a change of sign was not possible in the optimization, and all emissions were required to be positive." >>> "This also implies that the optimization did not allow sign changes…"

Line 248: observation was >>> observations were.

Line 249: "expressively" >>> This can be clearly seen by…?

Line 252: observation >>> observations, difference >>> differences