Review of Curcoll et al. (2024), **„Estimation of seasonal methane fluxes over a Mediterranean rice paddy area using the Radon Tracer Method (RTM)”**

This paper aims to estimate methane flux and its seasonal variability using the "local scale" Radon Tracer Method (RTM) in a rice paddy in the Ebro Delta. The manuscript is well-written and logically structured, with a clear presentation of methodology and results. Given the importance of methane emissions from rice paddies as a significant contributor to agricultural greenhouse gases, as highlighted in the introduction, this study is highly relevant for meeting the goals of the Paris Agreement.

However, before recommending publication, I would suggest further clarification on some major points.

Methodology:

The "near flat variability of radon" raises significant concerns in this study. The authors are conducting measurements at a site that remains flooded for much of the year, which would lead to no radon flux. In contrast, methane (CH_4) emissions may still occur from this area. Given that this is a coastal site, radon flux from the ocean is minimal, but coastal regions can still generate CH_4.

The local accumulation version of the Radon Tracer Method (RTM) assumes that the fluxes of both radon and methane are similarly distributed in space and homogeneously spread across the measurement footprint. This homogeneity is crucial for neglecting advective effects. Additionally, the "accumulation" model is valid only under very stable conditions, specifically when wind speeds near the surface are less than or equal to 1.5 m/s. If the wind speeds exceed this threshold, the correlations observed between radon and CH_4 would likely result from fetch effects rather than local accumulation *(This limitation is also addressed in greater detail by another reviewer, Scott Chambers.).*

To achieve their objectives more effectively, the authors would need to focus exclusively on low wind speed nights, which, based on their wind rose data, are quite rare. This restriction would significantly reduce their dataset. Furthermore, they could only apply their methodology during times when the rice paddies were not flooded, ensuring that radon flux was indeed present from the relevant fetch region.

More detailed points in methodology:

- The dataset spans from 2013 to 2019, providing approximately seven years of data with a total of 61,320 hourly measurements. *I would appreciate some clarification on the rationale for specifically including data from 2019.*
- However, only about 30% of this data (approximately 18,396 observations) was usable due to instrument maintenance issues. Additionally, limiting the analysis to just six hours per night further restricts the dataset, raising concerns about the representativeness of the flux measurements. For example, Table 2 shows that data from February across seven years includes usable measurements from only two nights.
- The diurnal variability of radon appears very flat, which may be attributed to the site's proximity to the coast and the limited radon fluxes present.
- Finally, the data selection criterion employed a threshold of $R^2 \geq 0.5$, which is relatively low; a minimum of 0.7 is generally preferred. This suggests that the observed diurnal radon signal may arise from air advected to the study site or from minimal contributions from exposed ground. During other times of the year, wind speeds are often quite high. When comparing

this distance to the scale of the rice paddy fetch in various directions from the measurement site, the implications for flux representativeness become even more pronounced.

I understand that you are using the Weather Research and Forecasting (WRF) model, which operates on a mesoscale, to estimate the influenced footprints (for a local). However, in section 3.2.1, regarding the evaluation of the meteorological model, it's noted that the correlation between simulated wind speed and observed wind speed is 0.57. Additionally, it appears that the model tends to overestimate wind speeds for most of the assessment period.

Given these findings, it is important to critically evaluate the accuracy of the model in estimating the footprint dimensions. Overestimating wind speed can lead to significant discrepancies in how gases are predicted to disperse. If the model is consistently overpredicting wind speeds, the resultant footprints may be smaller or inaccurately positioned, which could misrepresent the true spatial extent of influence.

Furthermore, it is essential to clarify whether this correlation of 0.57 applies uniformly across the entire study period from 2013 to 2019, or if it is specific to certain times of day, such as nighttime or daytime. Variations in wind patterns between day and night can significantly affect dispersion characteristics. If the correlation is weaker during certain periods, this could further impact the model’s reliability in estimating footprint dimensions.

To improve the robustness of your analysis, consider the following:

1. **Temporal Analysis**: Investigate whether the correlation varies by time of day. This can provide insights into how well the model performs under different meteorological conditions.
2. **Sensitivity to Overestimation**: Assess how the overestimated wind speeds affect gases and footprint dimensions. This sensitivity analysis can help quantify the potential impacts of the model's inaccuracies.

More specific comments:

L328: I remain unconvinced that the average radon levels observed in December can be attributed solely to local radon fluxes. The high standard deviation associated with this mean suggests significant variability in the data, indicating that the observed radon levels may be more representative of a unique, one-time event rather than a consistent, ongoing trend driven by local sources. Could you please look in more detail for this?

L422-427: “In the present work, the correlation between observed wind and modelled wind (0.52) is higher than the correlation between observed and modelled radon concentrations (0.38 – 0.43). Moreover, differences between the three radon exhalation models are much lower than between observations and models. Therefore, although no observational data was available on BLH for DEC station, it can be deduced that most of the disagreement between models and observations may have come from the nocturnal boundary layer simulation rather than by radon exhalation maps uncertainties.” – Could you clarify how you reached this conclusion? I’d like to understand the reasoning behind attributing most discrepancies to boundary layer simulations instead of the radon exhalation models. What specific evidence or analysis supports this interpretation? I’m not entirely convinced that relying solely on correlation coefficients provides a complete picture.

L505-512: The authors compare the RTM results from 2013 to 2019, focusing on nocturnal accumulation, with the findings from Martinez-Eixarch et al. (2018), who conducted a campaign in

2015 using static chambers for approximately 1 to 3 days (10 am to 3 pm) each month across 15 spatially distributed locations.

It would be particularly insightful to include direct comparisons for the specific year of 2015. This year serves as a common reference point for both datasets, allowing for a more precise evaluation of how the RTM monthly averages align with the static chamber measurements from that same year. Highlighting these comparisons could clarify the performance of the RTM in accurately reflecting radon dynamics during 2015.