Dear Editor,

The authors have made extensive improvements to the manuscript, making it far easier to review on this occasion. There are some use useful results in this study making it worthy of eventual publication. Yet, this manuscript is not in an acceptable state for publication at this stage with many improvements required. I therefore recommend further major revisions.

General comments

I struggle to understand the point of the horizontal flux component of this work. The introduction states that "strong spatial heterogeneity" and "non-negligible horizontal transport processes" are responsible for introducing "substantial uncertainties into quantitative estimates of ecosystem-atmosphere" carbon dioxide exchanges. It then goes on to state that resolving horizontal transport is therefore required. Why? And how does this specific work help to achieve this. As far as I can see, this work shows variability in fluxes between different sized horizontal boxes above the tree canopy. What is the use of this analysis? This work simply confirms that forests are spatially heterogeneous. How does this horizontal analysis help in the overall aim of quantifying carbon dioxide fluxes between forest ecosystems and the atmosphere? This really must be emphasised and made totally clear in this manuscript.

In a formal scientific publication, it is important to clearly define any terms when they are first introduced in the manuscript. Please use consistent terminology throughout the manuscript and define any terms at their first point of use. This is the responsibility of the authors and not the responsibility of the reviewers, the editors or the copy-editing/typesetting staff. Even if a term may seem obvious (for example, t), this could mean temperature, time, thickness, torque or tonnes, as some examples. Or it could mean anything else, depending on how it is defined, especially in environmental science journals where conventions are less entrenched.

How is $E_{C,VT}$ calculated for the different tests? The UAV conducts both vertical transects as well as box-type transects. Is the box-type data used for vertical fluxes when deriving horizontal components for different sized boxes or is the vertical transect data used? Please can this be clear in the manuscript. The methods section it states that "z_1 and z_2 correspond to 5 m below and 5 m above the target height, respectively, during the UAV ascent" when calculating $E_{C,VT}$. From this, I assumed that the single vertical UAV transect was used to derive a single $E_{C,VT}$ value from each flight whilst separate corresponding $E_{C,H}$ values were derived from box sampling. But it looks like different $E_{C,VT}$ values are used, as the median $E_{C,VT}$ values given in Table 5 do not scale linearly with box area. What is the chosen "target height" (defined in section 2.3) used to derive the $E_{C,VT}$ values given in section 4.3 and why? I really think this could be made simpler to follow for a reader. I am expending a lot of effort in trying to ascertain which data is used in which way, with it still not being clear to me.

Serious work in required on the uncertainty explanation. There seems to be no analysis of methodological errors, associated with the flux quantification method itself, when deriving a so-called "total uncertainty". The authors themselves partially acknowledge this at the end of subsection 4.4.4. Instrumental, wind and thermodynamic variability are treated as the only sources of error, assuming the method itself to be free from error. Then a description of the way in which these errors are quantified is totally lacking in the manuscript. It simply states that "propagating these errors through the flux equations" allows a fractional error to be obtained. This is not adequate. In addition, the instrumental error includes errors in wind measurements, despite there being a separate term for wind error. This seems illogical and lacks rationale in the manuscript. I raised this issue in the previous review and the authors claim to incorporate all wind measurement uncertainty into the wind term. However, it appears that they have failed to remove the wind component from the instrumental term, resulting in it being counted twice.

Specific comments

Line 40: "evaluated against a collocated eddy-covariance (EC) flux tower"

- Consider replacing this with "evaluated against fluxes derived using a collocated eddy-covariance (EC) flux tower". It is not the EC tower itself that is compared but fluxes from the EC tower.
- It is not clear how a tower can be collocated with a UAV (*i.e.* in the same place). Maybe "nearby" is a better word than "collocated".

Line 41: “Validation against a long-term flux tower”

- Again, it is not validation against a tower but validation against fluxes from a tower.
- Also it is not clear what a long-term flux tower is. I think the authors mean a permanently installed tower or a tower with a long-term installation.

Line 42: “gradient method”

- What is a gradient method? There is no information about any gradient method in the abstract up until this point.

Line 43: “correlation coefficient r of 0.9”

- How many individual flux estimates does this correspond to and over which time period (*i.e.* how many months and years).

Line 44: “Horizontal measurements”

- What are horizontal measurements? All measurements have a horizontal and a vertical component. Does this mean measurements using horizontal profiles?

Line 47: “multi-scale grid boxes”

- Perhaps state that these are grid boxes in the horizontal plane (if I understand correctly). The vertical extent of each grid box remains the same.

Line 53: “complex”

- What is a complex forest?

Line 60: “however”

- There is no need for this word. This statement is not in contrast to the opening statement.

Line 72: “satellite-based remote sensing”

- I am a bit confused with what it trying to be said here. A satellite is a measurement platform. Satellite-based remove sensing is a type of measurement method. I would recommend simply replacing this with the word “satellites” to be consistent with the start of this list.

Line 72: “measurements”

- Remove this word. A measurement is not a platform.

Line 73: “most existing methods”

- Most methods of what? Perhaps replace this with “most existing flux quantification methods”.

Line 74: “inherently confined to either vertical or horizontal dimensions”

- I am totally lost here. Is this referring to measurement methods or sampling platforms? If this is referring to sampling platforms, I disagree. An aircraft can fly in three dimensions. Conversely, a stationary tower can be zero-dimensional (with no vertical or horizontal component), as they often measure at a single point (although some can also sample at multiple heights providing a vertical dimension).
- I hope there can be consistency on this point: a sampling platform is different to a measurement method. A sampling platform can be a satellite, an aircraft or a tower, for example. A measurement method can be Eddy covariance, Gaussian plume inversion or a Bayesian inversion (for example). Gaussian plume modelling is a method that can be applied to sampling from multiple platforms (for example, car sampling or UAV sampling).

Line 80: “approaches”

- Please specify which approaches. Is this referring to flux quantification approaches?

Line 80: “both horizontal and vertical CO2 transport”

- I do not understand how resolving horizontal transport helps to understand canopy-atmosphere carbon dioxide interactions. I understand that there is spatial heterogeneity, but why is it important to quantify the flux from one part of the forest to another part of the forest. The important thing here is the vertical flux between the forest and the overall atmosphere.
- I agree that horizontal fluxes are important if fluxes between the entire forest and outside the forest and considered at the forest perimeter wall. But this study appears to look at the horizontal flux between one box area above the forest and another area above the forest. How is this useful?

Line 82: “the most widely”

- Consider changing this to “a widely”.

Line 99: “In addition, stationary towers (e.g. EC flux towers) provide flux information at a fixed height”

- I disagree. Stationary towers can provide sampling at multiple heights.

Line 109: “vertical fluxes are difficult to determine”

- Why?

Line 120: “over short time”

- Stationary towers can also measure over short time scales.

Line 123: “these UAV applications”

- Which UAV applications? I do not follow this.

Line 125: “both of which”

- I do not understand what this is referring to. Both of what?

Line 255: “a lower elevation (~15 m lower)”

- 15 m is a large change in elevation. Please can the authors comment on the topography of the area. Was the take-off location in a ditch? If so, it is very important that the UAV flew at a fixed height above ground level, rather than above take-off level.

Line 256: “which facilitates safe UAV operations”

- Why does being 15 m below the tower help to facilitate UAV operations?

Line 265: “results from 168 flights were retained for analysis”

- So what is the cumulative flight time of these good flights? The time of the discarded flights (above) is a bit irrelevant.

Line 267: “seasons”

- Are these meteorological or astronomical seasons?

Line 328: “becomes sufficiently mixed”

- Why does a species need to be sufficiently mixed? A mass balance approach should be able to account for high (or low) concentration emission plumes, so long as all hotspots (or coldspots) are sufficiently measured.

Line 351: “u, v”

- Please define these vectors here (see comment below).

Line 352: “30-minute period”

- Each flight was on average 29 minutes long (100 minutes of flight time by 209 flights). So where did this 30-minute averaging period come from?
- What was done for any flights of less than 30 minutes?

Line 355: “mean horizontal wind speed”

- This does not make sense as the mean horizontal wind speed is by definition the average of a scalar, with u(bar) instead defined above as a vector components.
- If u(bar) is the mean horizontal wind speed, what is v(bar)? u(bar) can be the zonal wind component and v(bar) can be the meridional wind component (following the convention). Or perhaps the authors set u(bar) to be in the direction of mean wind direction, in which case v(bar) (which is orthogonal to u(bar)) is automatically zero in magnitude.
- In any case, u(bar) should be clearly defined above (where it is first introduced) rather than being defined here.

Line 356: “main contributor”

- What does it mean to be a main contributor to a term? Both concentration and wind speed effect fluxes multiplicatively. So there is no main contributor, as far as I can see.

Line 357 “horizontal turbulent velocity”

- Which direction does this refer to? The horizontal plane has two dimensions.

Line 367: “$E_{C,M}$”

- Unfortunately, this equation is useless at this stage as none of the terms in the equation have been defined.

Line 367: “M_R”

- What is this?

Line 367: “χ_C”

- What is this?

Line 367: “ρ_a”

- What is this?

Line 367: “t”

- What is this?

Line 367: “x”

- What is this?

Line 367: “y”

- What is this?

Line 367: “z”

- What is this?

Line 382: “mole fraction”

- “Mole fraction” and “mixing ratio” are used interchangeably throughout the manuscript. Please use one term consistently. I think “mole fraction” is better as it is a more precise definition of the concentration.

Line 384: “flight trajectory”

- It is important to state that this represents a closed loop (*i.e.* the start and end of each s coordinate should be in the same location) to avoid any hole in the flux field.

Line 392: “flight distance”

- Perhaps specify that this is the distance between consecutive sampling points.

Line 394: “To suppress small-scale variability”

- Please specify that this refers to variability in the vertical direction.
- What is the purpose of suppressing this variability? Why not include it in the overall flux calculation?

Line 394: “moving average”

- I am totally lost. The consecutive values are in which dimension? Is this an average over time or over altitude? How many transects were performed at each measurement height?
- As far as I can see, a single line flux value is obtained at each height for each flight. So what is there to average over?

Line 408: “The resulting vertical fluxes were assumed to be representative of the broader domain, an assumption evaluated through direct comparison with eddy covariance (EC) flux tower measurements (see Section 4.2.1).”

- If the vertical sampling from a UAV is deemed to be equally good as sampling from a flux tower, what is the point of the additional horizontal fluxes? It seems like the vertical work is sufficient according to what is stated here.

Line 420: “concentration”

- What sort of concentration (in which units)? It this a mass concentration? Is this a volumetric concentration? A mole fraction is a type of concentration too.

Line 424: “friction velocity”

- In which direction is this fraction velocity? The horizontal plane has two dimensions.

Line 443: “wind speed”

- Is this a scalar (absolute) wind speed or does this term refer to a vector wind component with a certain direction in the horizontal plane?

Line 455: “surface-layer mean wind profile”

- Is this a scalar wind profile or a wind vector profile with an associated direction?

Line 507: “The air inlets of the two analyzers were positioned at the same height to ensure sampling of the same air mass”

- According to my understanding, the UAV conducted an extended vertical profile from which sampling at the same height as the EC instrumentation was used. It was not that the UAV was necessarily positioned at a certain height.

Line 522: “The MIRA measurements tend to report slightly lower CO2 mixing ratios than the LI-7500.”

- I see the opposite from Figure 4. Maybe provide the mole fraction averages for both datasets for comparison.

Line 551: “mole fraction”

- “Mole fraction” and “mixing ratio” are used interchangeably throughout the manuscript. Please use one term consistently. I think “mole fraction” is better as it is a more precise definition of the concentration.

Line 569: “z_2=26 m and z_1=36 m”

- Please clarify that this is from the vertical profile sampling data during UAV flights, rather than the box sampling data.

Line 585: “sample size”

- So does each sample refer to a single UAV flight? Or were multiple fluxes obtained from a single UAV flight? This point could be made clearer in this manuscript.

Line 707: “physiological”

- Could this be “photosynthetic”?

Line 717: “UAV measurement results”

- Why is this subsection entitled “UAV measurement results”? The previous subsection also contains UAV measurement results. I think this subsection is dedicated to net fluxes including the box-type sampling data.

Line 718: “mole fraction”

- “Mole fraction” and “mixing ratio” are used interchangeably throughout the manuscript. Please use one term consistently. I think “mole fraction” is better as it is a more precise definition of the concentration.

Line 732: “mole fraction”

- “Mole fraction” and “mixing ratio” are used interchangeably throughout the manuscript. Please use one term consistently. I think “mole fraction” is better as it is a more precise definition of the concentration.

Line 749: “Therefore, the positive E_c primarily reflects net CO_2 inflow driven by horizontal transport (i.e., lateral convergence) rather than local vertical emission from the canopy.”

- But does this help us to understand net carbon dioxide exchange with the overall atmosphere? This result shows that there is a next flux between the inside and the outside of a box over the tree canopy. Some boxes will have a net negative flux and some boxes will have a net positive flux. As the entire forest is not placed inside a horizontal box, this does not show the net flux from the entire forest to outside the forest (to the overall atmosphere). So what is the point?

Line 754: “This underscores the importance of resolving spatial gradients and lateral boundary fluxes-features that can be captured by UAV sampling but are not directly observable from a single-point flux tower measurement.”

- Why? To achieve what?

Line 759: “To verify whether net emissions (E_C) calculated using the mass-balance approach differ across spatial scales”

- Why? What is the point of proving this? We know that forest emissions are heterogeneous.

Line 763: “The vertical contribution term $E_{C,VT}$ and the horizontal transport contribution term $E_{C,H}$ within each control volume were separately computed as per the detailed outlines above.”

- The outlines above do not explain how $E_{C,VT}$ was calculated from box-type sampling. It does not say which “target height” was chosen and why. It does not say whether the vertical transect from each UAV flight was used or if the box sampling was somehow used to obtain a vertical flux.

Line 770: “To ensure the validity of the comparison, measurements from the three virtual control volumes must be obtained within 20 minutes of each other”

- I am confused. Each flight (of on average 29 minutes) included sampling of each of the three control volumes. Why not just compare three fluxes for different control volumes from each flight? I do not understand why there is this 20-minute requirement. Were fluxes from different UAV flights compared? This seems like a strange approach.

Line 815: “Table 5”

- According to my understanding, $E_{C,VT}$ is calculated from the single vertical transect conducted during each UAV flight. If this is the case, why do the $E_{C,VT}$ values not scale linearly with box size, for each box size? For example, the median $E_{C,VT}$ is -2.77 for the 50x50x50 box. I would expect it to be -11.08 for the100x100x50 box as it is four times bigger, in terms of surface area. And the fact that the 150x150x50 box is larger than -13.6 (-12.5) is even stranger. I am not sure what is happening here. The vertical flux should be independent of the box size, unless the box sampling is used to derive $E_{C,VT}$ using some unexplained method.

Line 877: “independent sources”

- These are not independent sources. The wind measurement error is treated as one error component, yet the anemometer error is incorporated into the instrumental error.

Line 897: “Propagating these errors through the flux equations”

- How are wind direction measurement uncertainties propagated through the equations? The equations do not include wind direction.
- Please can some details be provided about how exactly these errors are propagated in the flux equations. It seems a bit vague.

Line 901: “methodological effects”

- What are the methodological effects and have they been quantified? I cannot see any analysis of methodological uncertainties in this work.

Line 904: “includes uncertainties arising from sensor performance”

- But I thought wind measurement errors were dealt with within the instrumental error above. That is what the text says. It is being included twice.

Line 916: A conservative $\delta_V \approx 8\text{-}22\%$ is assigned to this uncertainty term.

- Is this uncertainty range arbitrarily assigned or computed. The word “assigned” implies that this range is made up.
- Assuming these uncertainty ranges are somehow computed, how is their computation achieved. There is no detail here. It is impossible to know whether this uncertainty has any value without these details.
- Assuming the wind speed has a certain percentage error, how is this converted into an error in the flux? Is it by propagating the error in the equations, as for the instrumental error? Please describe this.

Line 919: “Temperature and pressure effects”

- The authors can include this if they insist, but air density variations over a flight of approximately 29 minutes are likely to be negligible. This could be summarised in one sentence and ignored from the uncertainty analysis.

Line 931: “independent components”

- They are not independent. Wind measurement errors appear in both instrumental and wind uncertainties, according to the description given in the text.

Line 933: “total relative uncertainty”

- It is not a total uncertainty, as it assumes that there are no methodological errors. There is a sentence at the end of the paragraph partially confirming this. So I do not really understand the point of this uncertainty if it is an incomplete uncertainty. What is the value of this incomplete uncertainty analysis?

Line 943: “quantify three-dimensional CO_2 transport”

- Is this the correct terminology? Is it the transport that was quantified or the flux? Perhaps the transport was characterised rather than being quantified.

Line 952 “Pearson correlation coefficient”

- From how many samples is this value derived?

Line 965: “local CO_2 budget”

- What does this mean? Does a “local” scale correspond to a box above the tree canopy?