Dear Editor,

In this work by Sun *et al.*, a UAV was used to conduct 209 UAV flights above a forest canopy in China resulting in 100 hours of flight data. Flights were conducted at different times of day and throughout the year. The UAV carried an on-board closed-path carbon dioxide sensor and a high-frequency anemometer. Each UAV flight consisted of a single vertical profile. In addition the UAV flew in multiple concentric squares at three fixed distances from a central point and at seven different heights. These data were used to derive net carbon dioxide fluxes composed of a horizontal advective flux term (derived from concentric square sampling) and a vertical turbulent flux term (derived from vertical profile sampling). A stationary sensor tower was placed near the UAV sampling location, from which Eddy covariance fluxes and friction velocities were derived for direct comparison against those from UAV sampling, showing apparently good vertical correlation, although correlation coefficients are lacking from the manuscript. Horizontal advective fluxes were derived using sampling with different box sizes, for comparison against the vertical turbulent flux component. However, the value of the horizontal flux component to this work is concerning, especially with the lowest UAV sampling taking place 17 m above the top of the tree canopy. This horizontal flux component has been presented as an important flux component at certain times of day, but the actual magnitude of such a flux component may be small, which should be discussed. I found that the text is often muddled, using inconsistent terminology. The flux methodology section is poorly written. I suggest a major revision to this manuscript before publication.

General comments

Unfortunately, the abstract lacks clarity what sort of data is collected and what is happening to it. Please can it be explicitly stated that this work is dedicated to flux calculation? What sort of sensor is used? Perhaps this is not an important detail but then what is? At least some details need to be provided in this abstract to appreciate the novelty of this work. Otherwise, the abstract simply tells us that a UAV was used to measure forest emissions in this work, which is vague.

The introduction lacks citations to more recent work. Many citations are old. This is fine, but they should be used in addition to more recent work.

The manuscript conflates measurement platforms and flux quantification methods. This occurs throughout the manuscript and particularly in the introduction. The authors should make a clear distinction between platforms (such as stationary towers, aircraft, satellites and UAV) and flux quantification methods (such as the Eddy covariance method and mass balance box modelling). This is particularly problematic when referring to Eddy covariance towers. It is true that many of these towers are sold to process data using the Eddy covariance technique. However, any tower with high frequency concentration and vertical wind sampling can be used for Eddy covariance fluxes. Similarly, measurements from a system "sold" as an Eddy covariance tower do not have to be used for Eddy covariance flux calculation; the data can be used in other ways.

The flights are conducted at a minimum height of 35 m above ground level. Yet the canopy height is 18 m above ground level. Therefore, any mass-balance box modelling in the horizontal direction is fundamentally flawed as the source is not included within the box and nowhere near the top of the tree canopy. Any horizontal component to the overall flux will miss most of the emission or absorption. If the vertical wind speed is negligible (as stated by the authors), then it will take a long time for activity below the forest canopy to influence air at a higher elevation. Perhaps the horizontal box sampling is not a very useful component of this work.

The equations are unfortunately incredibly difficult to follow. Many terms are not defined and they are written incorrectly, using inconsistent terminology. The authors must revise all of the equations and ensure that each term is correctly defined in the main text. The style of identical terms must remain consistent. For example, if a lower-case "u" represents wind speed, it cannot be changed into an upper-case "U" in a different equation, unless it has been defined as a different term in the main text. There are countless examples like this. It is really important that the authors correct this and make everything consistent as the typesetters will not necessarily know that "U" and "u" mean the same thing.

Section 3.2.2 provides extensive details on how the vertical turbulent flux component is derived with many complex steps. However, there are absolutely no citations in this subsection to justify why such steps are taken. For example, m is defined as the stratification correction parameter which can be calculated using equation 3.11. But where does this equation come from? Did the authors derive these equations themselves from first principles?

Section 4.1 provides a useful inter-comparison between measurements from the two co-located gas analysers. It would be good if some correlation statistics could be provided, with a linear regression fit. The 10 Hz data can be averaged to every second for this comparison.

Equation 3.16 proposes that a vertical flux can be derived from mean concentrations at two heights. However, the heights are not provided in the manuscript. Section 4 presents vertical fluxes without describing which two specific heights these fluxes are derived from.

Many of the conclusions of this work are based on the coefficient of determination of the linear fit between Eddy covariance and UAV fluxes, with each of these two types of fluxes calculated independently of each other. The authors should instead present the Pearson correlation coefficient where the coefficient of determination is currently used throughout this manuscript. A coefficient of determination is used to evaluate the ability of a model to fit a data input to a data output. Yet, in this work, the purpose of the linear fit is not to use one flux to derive the other flux but rather to compare two flux values expected to have the same value. A coefficient of determination is therefore better is it evaluates how two sets of results, expected to have the same value, are correlated together.

Section 4.3 compares different total flux estimates, including different boxes covering different sized areas. However, different sets of UAV flights are used for each of the different area sizes (this is clear because the number in Table 5 is not the same for each square size). These different datasets are compared directly in Figure 9 and Figure 10. I do not think direct comparisons can be made unless the same flights are used for all three distances. Averages from the same set of flights should be compared. If one flight is rejected at 50 m, it should also be rejected at all distances.

The authors assert that this horizontal flux component can be as important as the vertical turbulent component. This was evaluated by taking the ratio between the two flux components. While this may be the case, it is not clear whether this is due to an especially large horizontal flux component or if the vertical turbulent flux component is naturally small at certain periods. Figure 6 suggests that the vertical flux component is naturally small in the morning, which might explain why the horizontal flux component appears to be so dominant in the autumn morning. In reality, if the flux magnitudes are compared, perhaps it is not so important.

Specific comments

Line 32: "evaluated"
- How can a platform be evaluated? I think this refers to a UAV flux quantification methodology, rather than the UAV platform itself.

Line 32: "applied to a forest ecosystem"
- How can a UAV-based platform be "applied" to a forest ecosystem? This doesn't really make sense (applying a hardware system to an environment). Again, perhaps this refers to the overall flux quantification methodology.

Line 34: "box-pattern and profile-pattern"
- It isn't clear to me what this means at this stage. What is a box-pattern and profile-pattern flight? Perhaps a few details can be given here.

Line 36: "resolve horizontal and vertical exchange"
- It is really not clear what this refers to in the abstract. Is this referring to the exchange of carbon dioxide (*i.e.* carbon dioxide flux) or some sort of study on atmospheric dynamics? I think the authors can be clearer here that this manuscript quantifies horizontal advective and vertical turbulent carbon dioxide flux components.
- Based on the manuscript, the horizontal turbulent component is ignored and the vertical non-turbulent component is ignored. This is not reflected here in the abstract, where it seems like all vertical and horizontal components are considered.

Line 38: "$R^2 \approx 0.76$–0.77"
- Why is the coefficient of determination used here? When comparing like-for-like flux estimates it would be better to provide the correlation coefficient. The coefficient of determination is more apt when evaluating the goodness of a model fit, which is not taking place here.
- Why is an approximately equal sign used for a range? If a range is provided, it is simpler to just precisely give the upper and lower bounds of the range.

Line 38: "Box-pattern flights"
- The UAV flights themselves cannot be used to infer diurnal variation but rather the measurements from flights are used to makes such evaluations.

Line 41: "Sensitivity analyses across scales of 50 m, 100 m, 150 m indicate that CO_2 emission intensity is sensitive to control-volume dimensions and shows spatial heterogeneity above the forest."
- This sentence is currently unhelpful in the abstract, without any details on the nature of the sensitivity analysis.
- What does it mean to say that emission intensity is sensitive to control-volume dimensions? What are control-volume dimensions? I do not follow this.
- A forest ecosystem being heterogeneous is not a new discovery. Again, without any details on the nature of the sensitivity analysis, I don't know the novelty what is being revealed here and its significance.
- Having read the manuscript, I now understand what the three different distances mean. However, at this point in the abstract, this is vague. It is not clear that these refer to different sized boxes.

Line 62: "estimate CO_2 exchange"
- The list provided in this sentence is not a list of methods to estimate carbon dioxide exchange rates but rather a list of sampling platforms. Please can the authors clarify this? Measurements from different sampling platforms can be treated in different ways to evaluate gas exchange rates and fluxes.

Line 62: "eddy covariance (EC) towers"
- I would recommend replacing this with "stationary towers" because that is what all systems from which Eddy covariance fluxes are derived actually are, as a sampling platform. Fluxes from towers are often evaluated using the Eddy covariance methodology, but not necessarily.
- The Eddy covariance is a method of flux calculation whereas a tower is a platform (like a satellite).

Line 73: "most widely used method"
- Please provide some citations here to support this assertion.

Line 75: "(Aubinet et al., 2012; Baldocchi et al., 2001; Baldocchi, 2003; Wilson et al., 2002)"
- Are there any recent references? The youngest citation in this list is from 14 years ago. If there is no recent work with FLUXNET, please can the authors state that this is a historical initiative that is no longer operational?

Line 104: "complex"
- In what way are these environments complex?

Line 107: "fixed EC towers"
- I would recommend replacing with this with "stationary towers" as not all measurements from towers are used for Eddy covariance fluxes.

Line 114: "flight attitude"
- What does flight attitude mean? Maybe this is a technical UAV term, but it must be defined in an atmospheric science journal.
- Why does this effect carbon dioxide flux estimates?

Line 118: "disturbances"
- What does this mean? What sort of disturbances does this refer to?

Line 126: "GHG"
- This term has not been defined and is not used elsewhere in the manuscript.

Line 127: "3-D"
- This term has not been defined.

Line 130: "UAV platform"
- It is not clear what "evaluating" a "UAV platform" means. Does this mean testing the sensor on the platform? Does it mean evaluating a flux quantification algorithm using data acquired from on-board the UAV? Perhaps it means something else.

Line 133: "emission"
- Maybe "emission" is not the best word here and throughout the manuscript as "emission" implies that carbon dioxide is generally released into the atmosphere. Yet, during photosynthesis, it is absorbed resulting in a negative flux. I would recommend evaluating the use of this word throughout the manuscript.

Line 137: "mass flux measurements"
- The flux is not itself measured, but rather computed from other measurements. So I don't think "measurements" is the right word here.

Line 141: "tower-based EC flux"
- This is the correct way to describe EC fluxes derived from a stationary tower. This phrase makes it clear that measurements from a tower (the platform) are used within the Eddy covariance method (the flux quantification technique). This statement correctly does not imply that all towers are used for Eddy covariance fluxes. Please can the authors apply this correct approach to the rest of the manuscript to avoid ambiguity?

Line 156: "the height from the body stands to the body top is 565 mm"
- Does this include the height of the air inlet? Please clarify this.

Line 168: "global positioning system (GPS)"
- The UAV probably contains a "GPS sensor" or "GPS tracker". The GPS itself is in space.

Lin 186: "< 50 g"
- Is it less than 50 g or equal to 50 g? In Table 2, it is given as equal to 50 g, not less than. In any case, it is better to give the actual mass, rather than saying it is less than something.

Line 188: "propeller blade downwash"
- This is a very good idea and often lacking in other studies. I am glad that the authors have taken this into account.

Line 206: "Table 2"
- Where are these accuracy values from? Are these values provided by the manufacturer or were the sensors laboratory tested for the authors to determine these values independently?
- Why is the carbon dioxide accuracy given in ppb per second? Accuracy is generally independent of time (at short time-scales).
- The terms "T" and "P" have not been defined.

Lin 209: "Figure 1"

- Which pump does this diagram refer to? Did the carbon dioxide sensor have an in-built pump or was an external plump used? Please provide details on this pump in the main text. If an external pump was used, was it tested to ensure a constant flow rate?

Line 242: "40 flights in spring, 55 in summer, 53 in autumn, and 20 in winter"

- Do these refer to astronomical or meteorological seasons? It is also worth stating that these refer to Northern Hemispheric seasons.

Line 246: "Figure 2"

- Are the flight tracks from a single flight (if so, please provide flight details)? Are all UAV flights identical?

Line 257: "Using flight-path fitting and screen interpolation (Han et al., 2024), the three-dimensional spatial distributions of CO2 volume mixing ratio, wind vectors, and air density were reconstructed within the box domain (Fig. 3b)."

- Was this interpolated data used during flux calculation or is this purely for visualising. Please make this very clear here and in Section 3. It is incredibly important to know if fluxes are derived from raw measurements or from interpolated measurements.

Line 278: "Figure 3"

- It is difficult to tell that there are two flight strategies here. After reading the text, I now understand that there is a vertical profile in the middle of the box. Perhaps the two flight strategies can be labelled in this figure. Otherwise it is easy to miss.

Line 286: "characterize the flow-field distortion by the UAV"

- Was this characterisation conducted for the specific UAV used in this work, with the finalised UAV configuration including the mounted carbon dioxide sensor and air inlet?

Line 309: "equals the net outward flux"

- It may be worth stating here that this integration includes negative flux components. If the wind is blowing into the box horizontally, the flux contribution is negative and if the wind blowing out of the box horizontally, the flux contribution is positive.

Line 324: "corresponding wind fluctuations"

- What do these wind fluctuations represent and how were they derived? Do they represent the variance in wind speed over a certain integrating time? Please explain this.

Line 348: "χ"

- Where is this term defined in the manuscript? I can't find it.

Line 348 "s"

- What is "s"? I can see that "Δs" has been defined, but not "s".

Line 348: "$U_\perp$"

- Where is this term defined in the manuscript? I can't find it.

Line 352: "$u_\perp$"

- Is this lower-case u different to the upper-case U in equation 3.3? In any case, neither has been defined.

Line 359: "Σ"

- What is the index of this summation? I can see that it spans from 0 to N, but what is summing over? Is it z? This should be placed underneath the sigma symbol.
- Assuming the index of this summation is z, this is apparently summed from a height of 0 m. However, the lowest UAV sampling height is 35 m. So it is actually a summation from 35 m. It is not possible to sum data from a data point that does not exist. Please clarify this.

Line 359: "u_n"

- What is u_n? Where is it defined? Is it related to the upper-case U in equation 3.3 and the lower-case u in equation 3.4?

Line 360: "Δz denotes altitude intervals"

- I guess that the lowest interval spans from 0 m to 35 m (the lowest sampling height). Therefore, a single measurement at 35 m represents the entire forest plane (with a canopy height of 18 m), with no measurement at ground level. So the UAV sampling misses the entire forest. Please discuss this.

Line 379 "T and P"

- These terms have not been defined.

Line 382: "K"

- What is "K"? Has it been defined?

Line 399: "surface-layer wind profile"

- How was this wind profile devised? What is this equation based on? A citation could be useful here.

Line 409: "z_1 and z_2"

- According to this equation, any two heights can be chosen. So which heights were chosen in this work and why? This important detail is missing from the manuscript. Was this an arbitrary choice?

Line 413: "Eq. (3.8) yields"

- Why not substitute equation 3.12 into 3.15, to eliminate R_i?

Line 414: "K"

- Are "K" and "K_C" the same thing? "K" has not been defined.

Line 417: "z_1 and z_2"

- So which heights were chosen in this work and why? I can't find this vital information in the manuscript.

Line 436: "The friction velocity (u_*) and scalar fluxes (F_v) were computed"

- Please provide citations for the source of these equations.

Line 445: "open-path eddy covariance analyzer"

- It is not an Eddy covariance analyser. It is a carbon dioxide gas analyser used by the Eddy covariance system.

Line 457: "Figure 4"

- What is the horizontal axis of this plot? What is "No."? It cannot be the number of measurements as the two gas analysers operate at sampling at different frequencies. I would recommend plotting this data against time.

Line 468: "interactions with the sampling line material"

- Which material was used for the sampling line?
- Was the effect of this sampling line on carbon dioxide measurements tested in the laboratory?

Line 485: "coefficient of determination"

- Please provide the Pearson correlation coefficient, which is a better metric to evaluate the linear correlation between two datasets. The coefficient of determination is usually better-suited to evaluate the goodness of a model fit.

Line 489: "Figure 5"
- What does SMA mean? This has not been defined.
- It is impossible to tell which dataset is plotted on which axis. Please make this clear and make the axis labels bigger.
- Why is a linear equation given with "y" as the subject of the equation? "y" has not been defined. It is better to give this equation as one vertical flux in terms of the other vertical flux.

Line 498: "coefficient of determination"
- As stated previously, the Pearson correlation coefficient is really required here and not the coefficient of determination. This is not an evaluation of a model fit but rather how one set of data compares against another.
- Here the coefficient of determination is a lower-case r^2 whereas in the main text it is an upper-case R^2. There are many inconsistencies in this manuscript which are unacceptable in a scientific publication.

Line 500: "$y_{UAV} =$ slope $x_{TOWER} +$ intercept
- As stated above, the equation should not be given in terms of undefined "y" and "x" but rather, in terms of vertical flux values.

Line 506: "coefficient of determination"
- The correlation coefficient is really required here to draw such conclusions.

Line 533: "UAV platform reproduces"
- The UAV platform does not reproduce anything, but rather the flux derived from the measurements made on-board the UAV platform.

Line 551: "Figure 6"
- As for Figure 5, the axis labels here are very difficult to read.

Line 580: "low coefficient of determination"
- I would be interested to know what the correlation coefficient is and if the vertical turbulent flux correlation is really that bad.

Line 589: "as radiative forcing weakens and a stable stratification develops, turbulence above the canopy diminishes"
- Please provide some citations to support this assertion.

Line 597: "seasonal distributions of nighttime CO_2 flux differences"
- How many data points are there for each season during the night?
- What is the definition of night in the manuscript? What hours of the day does this correspond to?
- What is the motivation for looking at night-time fluxes independently? Please provide some justification here.

Line 600: "0.1"
- What are the units for this value?

Line 604: "F=2.50"
- What is F? Has it been defined? Maybe it is related to ANOVA; if so, some background should be provided here.

Line 606: "p"
- What is p?

Line 632: "reflecting reduced turbulent mixing"
- Why does a positive flux indicate reduced turbulent mixing?

Line 659: "height=50m"
- According to equation 3.5, the boxes are integrated across all heights. So I don't understand why a fixed height of 50 m is being used here. This is a very important point and perhaps represents some missing information in the methodology.

Line 676: "This result may suggest relatively weak spatial heterogeneity during the summer-noon period."
- This is probably because sampling at a fixed height of 50 m is used, when the tree canopy only reaches up to 18 m. The spatial heterogeneity of the forest may be lost at this height. So this entire analysis seems a bit pointless to me.

Line 678: "F"
- What is F? Has it been defined?

Line 701: "$E_{C,V}$"
- Should this be $E_{C,VT}$? I thought the vertical advective flux component is negligible and only the vertical turbulent flux component is considered.

Line 705: "The mean values in Fig.10 indicate that the emission intensities (e_c) derived for different box sizes (0.02, 0.21, and 0.11, respectively) differ to some extent."
- This is probably because the exact same set of UAV flights are not used for the different box sizes. It is impossible to compare them directly, if they do not represent identical sampling.

Line 714: "3.214 for 50×50, 0.847 for 100×100, and 1.213 for 150×150"
- The $E_{C,H}$ and $E_{C,VT}$ mean (or median) values and standard deviation values also need to be provided separately in Table 5. Perhaps autumn morning $F_{C,VT}$ fluxes are very close to zero, amplifying this horizontal flux dominance.
- As mentioned above, it is difficult to make direct comparisons, if different sets of UAV flights were used for each box size.

Line 715: "These results suggest that there is a significant spatial heterogeneity in the CO2 emissions/uptake in the range covered by the different boxes"
- It is difficult to draw these sorts of conclusions with only seven or eight data points.

Line 726: "$E_{C,V}$"
- According to section 3, $E_{C,V}$ is negligible and only $E_{C,VT}$ needs to be considered. So I am a little lost with the terminology here.

Line 739: "Figure 10"
- The closing bracket is missing for the units on the vertical axis.
- What are F and p?

Line 753: "Table 6"
- I am not sure that this is the best way to represent uncertainties. It seems like twelve emission estimates have been cherry-picked out of the full dataset. Why not just give the average percentage uncertainty for each season, rather than picking a single day?

Line 764: "better than 0.2 ppm"
- According to Table 2, the accuracy (not the precision) is better than 0.2 ppm. Please clarify whether this refers to accuracy or precision.

Line 768: "uncertainty on the order of $\delta_M \approx 5\%$"
- How can the uncertainty be approximately equal to the order of 5%? This doesn't really make sense.
- Table 6 shows that the uncertainty is only 5% for one of the twelve fluxes. Most of the time it is much lower than 5%.

Line 771: "4.4.4 Wind-calibration uncertainties"
- It seems strange that wind calibration uncertainties are treated separately to wind measurement uncertainties. It would make more sense to treat wind uncertainties as one single term and carbon dioxide uncertainties as a separate term. Please discuss the rationale for the approach used here.

Line 792: "can be considered negligible"
- But they are not treated as negligible as they are included in the error propagation equation. So I don't understand this point.

Line 803: "CO_2 enhancement field"
- What are the enhancements above? Are they enhanced above some sort of background? How is this background determined?
- Enhancements above a background are only useful assuming a positive emission flux. A negative emission flux characteristic of dominant photosynthetic activity would be better represented as mole fraction diminutions.

Line 815: "Figure 11"
- Please remind the reader that this interpolated data is derived following the procedure described in Section 2.3 and that this figure does not show raw mole fraction sampling. This data has been processed.

Line 848: "excellent agreement"
- This is a strong phrase Consider rewording this.
- Pearson correlation coefficient is required to evaluate such agreement.

Line 854: "are ecologically consistent with long-term site behavior"
- How is this shown from UAV fluxes? Where is the long-term site behaviour presented in this manuscript? Perhaps a time-series should be shown to make this point.

Line 859: "dominant term under certain conditions"
- Horizontal transport may be dominant when vertical turbulent fluxes are naturally small. It is important to evaluate if the horizontal flux magnitude is actually that large in absolute terms, instead of only taking the ratio between the two flux components.

Line 865: "reflecting the spatial heterogeneity of sources/sinks and the changing relative contribution of horizontal transport with control-volume size"
- This spatial heterogeneity may be amplified if the vertical turbulent flux component is naturally small at certain times of day. It may be totally insignificant if the vertical turbulent flux component is larger. This should be discussed and clarified.