GENERAL COMMENTS

This manuscript presents a novel and well-documented autonomous platform for high-resolution monitoring of aquatic carbon fluxes and environmental variables and it fits well within the scope of biogeoscience/environmental instrumentation journals. The study represents a valuable technological contribution that fits well within the scope of AMT. The platform design, field demonstrations, sensor intercomparison, and open-access documentation are major strengths. However, before publication, the manuscript would benefit from a more thorough discussion of flux uncertainty, stronger justification for the use of linear chamber flux calculations, a more cautious interpretation of comparisons with eddy covariance observations, and additional clarification regarding the robustness of spatial interpolation procedures. These concerns do not undermine the overall quality and novelty of the work but should be addressed prior to acceptance.

Overall assessment

The BlueMinerva platform fills an important gap between point-scale chamber measurements and ecosystem-scale eddy covariance observations. The combination of autonomous navigation, chamber flux measurements, water-quality monitoring, meteorological sensing, and bathymetry in a single platform is innovative The authors provide unusually detailed information on hardware, software, communication architecture, power management, payload, and operational constraints. This greatly enhances reproducibility and community uptake. The evaluation in two contrasting environments (a temperate German lake and a sub-Arctic Swedish lake) strengthens the manuscript and the dataset is reasonably extensive (485 flux measurements over 72 h).

Main concerns (Major revisions)

1. Insufficient validation of flux accuracy

The paper validates agreement between two gas analyzers and compares measurements with eddy covariance, but it does not fully validate the accuracy of flux estimates themselves.

The authors essentially compare:

- sensor A vs sensor B,
- chamber measurements vs EC measurements.

However, neither comparison establishes whether the measured fluxes are correct.

The manuscript provides a convincing evaluation of instrument consistency but less evidence regarding the absolute accuracy of the chamber-derived fluxes. I encourage the authors to discuss in more detail possible biases associated with chamber deployment, chamber-induced turbulence, pressure perturbations, and non-linearity of concentration accumulation. A quantitative uncertainty analysis of flux estimates would substantially strengthen the study.

2. Exclusive use of linear flux estimation

The authors calculate all fluxes through linear fits (Eq. 5). This is potentially problematic because chamber accumulation often deviates from linearity. Usually also exponential fit is used. The manuscript would benefit from a justification for the exclusive use of linear regressions for flux estimation. Why were non-linear chamber models (e.g., Hutchinson-Mosier type approaches or alternative accumulation models) not considered? Please discuss potential biases resulting from assuming linear accumulation during the 5-minute enclosure period.

3. Eddy covariance comparison is based on a small sample size

The EC comparison uses only:

- 17 chamber observations,
- 29-38 EC observations after footprint filtering

The conclusions are somewhat stronger than warranted. The comparison with eddy covariance measurements is useful but based on a limited number of overlapping observations (TABLE 3). The authors should moderate statements regarding agreement between methods and explicitly acknowledge the limited statistical power of this comparison because a non-significance difference does not demonstrate the agreement. In my opinion a scatter plot between the two measurements, with RMSE and determination coefficient is a better way to compare two data sets.

4. Potential temporal confounding at Inre Harrsjön

The authors correctly note that measurements were acquired in two different time period (one week apart).Because measurements at Inre Harrsjön were conducted on different dates with substantially different environmental conditions, some of the observed spatial patterns may partly reflect temporal variability. Please discuss this limitation more prominently when interpreting hotspot distributions and smoothed flux maps.

5. Interpolated flux maps may overstate spatial patterns

The generated maps (Fig. 8) are visually attractive but could give the impression of greater certainty than justified.

No sensitivity analysis is presented for:

- grid size,
- kernel bandwidth (σ),
- measurement density.

The interpolation approach requires further justification. Please provide a sensitivity analysis demonstrating whether the identified hotspots remain robust under different kernel bandwidths and interpolation parameters.

Minor comments

Methods

- Explain why k = 5 and k = 8 were selected for Moran's I.
- Clarify how measurement duration of 5 min was chosen initially.
- Report uncertainty on bathymetric interpolation in the main text, not only as RMSE values.

Results

- Include confidence intervals around median fluxes.
- Add effect sizes consistently when discussing all statistical comparisons.

Figures

- Figure 8 is attractive but uncertainty is not shown.
- Consider adding observation density layers beneath interpolated maps.
- Several figures are information-dense and fonts appear small.