Our answers are provided in blue.

I find the manuscript highly informative, as it provides a comprehensive long-term dataset on benthic phosphorus fluxes in the Baltic Sea, offering valuable insights into nutrient cycling and contributing to more effective management of coastal eutrophication. I also appreciate the discussion of the study's limitations and the well-considered suggestions for future research directions.

We thank the reviewer for their feedback, which has helped us to clarify several methodological points.

General Comments:

1. In the first sentence of the abstract, I recommend avoiding exclusive emphasis on coastal eutrophication, as phosphorus recycling is a process of broader significance across marine ecosystems. Furthermore, the majority of the data presented in the study are not from coastal environments, so a more general framing would be more appropriate.

   We agree, and have removed the word “coastal”.

2. Please provide the source of the oxygen data used in Figure 1. If the data were collected as part of this study and extrapolated, a detailed description of the methodology should be included in the Methods section.

   We thank the reviewer for pointing out this omission. As described in the main text, the extent of hypoxic and anoxic conditions was taken from the SMHI oxygen survey from 2021 (Hansson and Viktorsson, 2021) – this information has been added to the figure caption.

3. According to many standard recommendations for phosphate sample collection and preservation, filtration (typically at 0.45 µm) is required to remove suspended particles and microorganisms. If unfiltered samples were used, a turbidity blank must be applied to correct for potential interference. Additionally, acidification is generally not advised for phosphate preservation. Given that various collection and preservation methods appear to have been used in this study, please discuss the uncertainties this introduces into the fluxes and extrapolations.

   All samples were filtered before analysis; the text has been corrected. Samples are routinely acidified to avoid the potential precipitation of iron oxides, which could scavenge phosphate (Bray et al., 1973, doi: 10.1126/science.180.4093.1362; Slomp et al., 1996, doi: 10.1357/0022240963213745).

The fluxes are not affected by potential differences in concentrations between measurements between methods, as they are calculated based on concentration changes. We have clarified this in the text.

L177-179: “All methods used to measure the DIP are routinely used for seawater, so no differences in concentrations are expected between methodologies. Furthermore, the resulting fluxes are not affected by methodological differences, as they are calculated from the change in DIP concentration in samples measured with the same methodology.”

4. I assume that samples were analyzed in different laboratories. If so, please provide detailed information on the accuracy and precision of the analytical methods used, as this can significantly affect the comparability of the results and calculations.

   While the samples were analysed in different laboratories, we have clarified in the text that the same method was used and similar analytical precisions were obtained:

   L175-177: “Concentrations of DIP were, in all cases, determined by the ammonium molybdate method using segmented flow colorimetric analysis (Koroleff, 1983) with an analytical precision better than 3%.”

5. The authors state that the dataset has limited use for interpreting seasonal trends. However, it would be interesting to deliberate a bit on the temporal variability at stations where monthly coverage was substantial, such as KH104.

   We have added a figure to the appendix showing the DIP fluxes as a function of month at sites that were sampled in at least three different months (Figure A3).

   Station KH104 is the only site in the dataset that has been sampled in more than two seasons. The DIP flux patterns at this station have been discussed extensively by Ekeroth et al. (2016, doi: 10.1016/j.jmarsys.2015.10.005), and changes to the O_2 conditions in the bottom water were concluded to have a larger impact than seasonality. We have clarified in the text why seasonal patterns in the data are not discussed.

   L248-250: “At 14 stations, sampling occurred during three or four different months (Fig. A3). However, except for at station KH104, samplings were conducted across different years, only during two seasons, and aimed at observing shifts in O_2 conditions. As a result, seasonal trends cannot be determined from this dataset.”