La Paz, Baja California Sur, 30 July 2026

To the Editor of Ocean Science

Re: First revised version (R1) of manuscript egusphere-2026-1040

Dear Editor,

On behalf of my co-authors, I am pleased to submit the first revised version (R1) of manuscript **egusphere-2026-1040** entitled "**Net primary productivity in the Gulf of California: spatio-temporal patterns and drivers in a warming climate (2003–2024)"** for further consideration in Ocean Science. We thank you and the two reviewers for their careful and constructive assessment. We have addressed every comment in a detailed point-by-point response and have revised the manuscript and supplementary material accordingly.

The principal methodological revision concerns the classification of net primary productivity (NPP). We now explain how the number of clusters was selected by comparing alternative solutions with the elbow and silhouette methods. These diagnostics supported a small number of clusters, particularly K = 2–4. We retained K = 3 as the most parsimonious and ecologically interpretable solution for low-, mid-, and high-productivity waters, and added the diagnostic analysis as Fig. A1 in the supplementary material. We also clarified the role of the discriminant analysis. The NPP-derived classes were used as the grouping variable, whereas hydrographic and biogeochemical variables were used to assess their environmental separability. NPP was excluded because it defined the original classes. The revised manuscript now states explicitly that this analysis evaluates environmental differentiation and does not validate either the clustering procedure or the selected number of clusters.

A second major revision concerns the analysis of environmental associations with NPP. The original Gaussian model was replaced by a Generalised Linear Mixed Model with a Gamma error distribution and log-link function, which is appropriate for positive, right-skewed NPP data. Continuous predictors were standardised before model fitting, month was included as a random intercept, and the model description and diagnostics were expanded. Table 1 and the corresponding Results and Discussion were revised to reflect the new analysis.

To avoid circularity, sea-surface temperature (SST) and euphotic-zone depth (EZD) were removed from the final GLMM because they are directly involved in, or closely linked to, the VGPM-derived NPP calculation. Silicate (SiO_2) was also excluded after the collinearity assessment. The final predictor set comprises mixed-layer depth, nitrate, phosphate, along-gulf wind and current components, stratification, and sea-surface salinity. Interpretations that treated SST or euphotic-zone depth as independent GLMM drivers were therefore removed or revised. Temporal analyses were also strengthened. Spearman rank correlations between PDO, NPGO, MEI.v2, and productivity-zone extent were reassessed using an effective-sample-size correction for temporal autocorrelation following Pyper and Peterman (1998). The revised interpretation recognises that the negative association between MEI.v2 and HPZ extent is weak, although statistically significant after correction, whereas PDO and NPGO are not significantly related to the extent of any productivity zone.

The trend analysis was replaced by the Mann–Kendall test for monotonic trends and the Theil–Sen estimator for trend magnitude. The supplementary material now includes the revised spatial trend map as Fig. A10, with statistically significant pixels identified. The Methods section now distinguishes clearly between satellite-derived VGPM NPP and the CMEMS reanalysis products used for subsurface and water-column variables. Data sources, product identifiers, regridding procedures, and the treatment of climate indices have been clarified. Statements presenting PDO or NPGO as demonstrated direct drivers have been removed or moderated. Figures, tables, and supplementary material were also revised. Figures, tables, and supplementary material were also revised. Figure 5 was enlarged to improve readability, and geographical locations requested by the reviewers were added to the study-area map. Productivity-zone thresholds and labels were checked for consistency. Supplementary Table A1 now reports the original and autocorrelation-corrected significance values, and the supplementary figures include the cluster diagnostics, seasonal coverage maps, anomaly time series, and revised trend analysis. Finally, the Discussion was shortened and reorganised to reduce repetition and improve the progression from regionalisation and seasonal variability to environmental associations, anomalous warming, and longer-term trends. Ambiguous terminology, including “tipping point”, was removed. Seasonal terminology, abbreviations, units, figure citations, references, and British English usage were reviewed for consistency.

The revised manuscript retains the central contribution of the study: a 22-year, basin-scale assessment of seasonal and interannual NPP variability in the Gulf of California, including the pronounced decline during 2014–2016, the extended below-average period during 2013–2019, and the recovery during 2020–2024. We think that the revisions have addressed the principal concerns of both reviewers and have substantially strengthened the methodological transparency, internal consistency, and interpretation of the study. We appreciate the opportunity to revise the manuscript and hope that the revised version is now suitable for publication in Ocean Science.

Yours sincerely,

Dr. Juan David Acevedo-Acosta
Corresponding author: acevedoacosta.juandavid@gmail.com
On behalf of the rest of coauthors

Reviewer 1

This is potentially an interesting paper, but it requires a major revision for the following reasons:

1. While K-means clustering is an accepted method of analysis the key question is how the number K of clusters is determined. There is no explanation of this in the paper. It is simply stated that NPP was grouped into 3 clusters using the k-means^{++} algorithm. How was the number of clusters (3) determined? Did the authors use, for example, the elbow method? Or some other approach? Why is 3 the optimum number of clusters for NPP?

Response: We thank the reviewer for raising this important methodological point. We agree that the selection of the number of clusters *(K*), required clearer justification in the manuscript.

To classify net primary productivity (NPP) in the Gulf of California, we first removed outliers following the criterion of Moore & McCabe (1999), using the same filtered dataset subsequently used for the clustering analysis. To support the selection of K, we evaluated alternative cluster solutions using both the elbow and silhouette methods. These diagnostics indicated that solutions with a small number of clusters, particularly K = 2–4, provided the most robust classifications, whereas higher values, $K \geq 5$, led to reduced clustering performance and increased fragmentation of primary productivity regimes, shown in Fig. R1A,B.

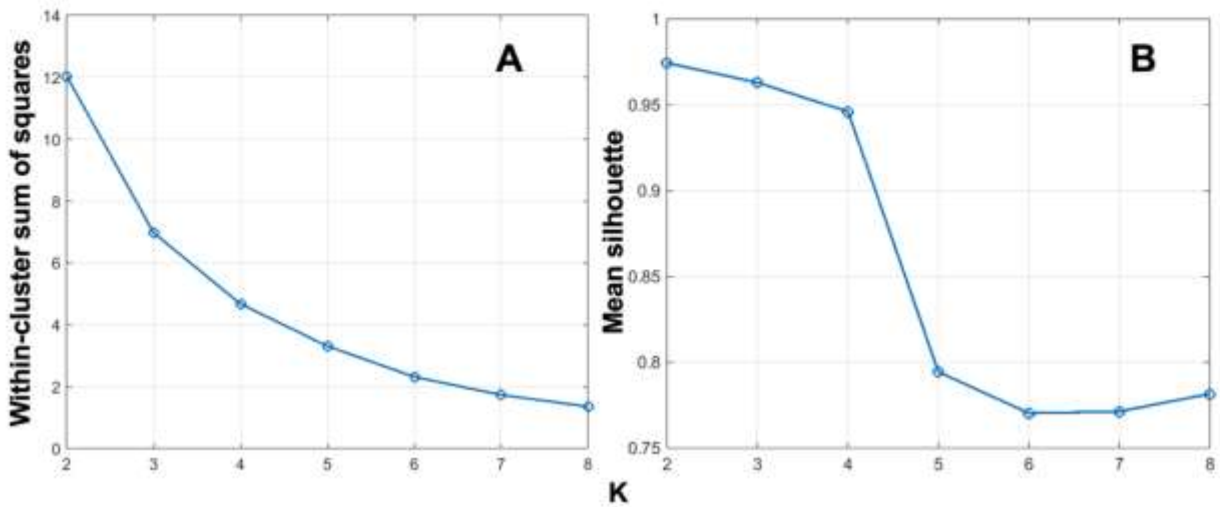

Figure R1. Evaluation of the optimal number of clusters for net primary productivity classification in the Gulf of California using a) the elbow method and b) silhouette analysis. Both diagnostics supported solutions with a small number of clusters, while higher values of K, particularly K ≥ 5, provided limited additional explanatory gain and showed reduced clustering performance.

Because NPP in the Gulf of California represents a continuous environmental gradient rather than a set of strictly discrete states, no single mathematically unique optimum was expected. We therefore selected K = 3 as the most parsimonious and ecologically interpretable solution. This classification consistently resolved three well-defined productivity regimes corresponding to low-, intermediate-, and high-productivity waters, in agreement with the known biogeographic and oceanographic structure of the Gulf of California. Although increasing the number of clusters produced finer subdivisions, these did not reveal additional coherent NPP regimes and instead reduced interpretability by increasing spatial fragmentation.

We have revised the Methods section to explicitly describe the procedure used to select the number of clusters and added the following reference:

Moore, D. and G. McCabe (1999). *Introduction to the Practice of Statistics. W. H. Freeman.*

In the text (page 7 Lines 186 to 195) now says: "To analyse spatio-temporal variability and identify productivity regimes, we classify the Gulf of California according to differences in satellite-derived NPP from January 2003 to December 2024. Prior to clustering, outliers were removed following the interquartile range (IQR) criterion (1.5 × IQR) of Moore and McCabe (1999), the resulting filtered dataset was used in all subsequent analyses. NPP was then classified using the k-means algorithm with k-means^{++} initialisation (Kapoor et al., 2017; Zhang et al., 2025). To evaluate the appropriate number of clusters, we compared alternative solutions using the elbow and silhouette methods. These diagnostics indicated that a small number of clusters, particularly K = 2–4, provided the most robust classifications, whereas higher values, K ≥ 5, resulted in lower clustering performance and increased spatial fragmentation. Because NPP varies along a continuous environmental gradient, K = 3 was selected as the most parsimonious and ecologically interpretable solution, resolving three coherent productivity regimes corresponding to low-, intermediate-, and high-productivity waters."

2. It is unclear how discriminant analysis was applied, as no details are given other than to state that it was applied to all variable except NPP – why? Again, the authors need to explain the details of their analysis for the results to be convincing.

Response: We thank the reviewer for pointing this out. We agree that the manuscript required a clearer explanation how the discriminant analysis was applied and why NPP was excluded from this step. The discriminant analysis was conducted after the k-means^{++} classification to evaluate whether the NPP-derived groups also corresponded to distinct environmental regimes. Specifically, the three NPP classes identified by the k-means^{++} algorithm with k-means^{++} initialisation, namely the LPZ, MPZ, and HPZ were used as the grouping variable. The environmental variables, including SST, EZD, MLD, NO_3, PO_4^{-3}, SiO_2, URT, UAR, and SI, were included as fixed effects within each model.

NPP was intentionally excluded from the discriminant analysis because it was the variable used to define the original k-means^{++} classification. Including NPP in the discriminant analysis would therefore have introduced circularity and artificially inflated classification performance. Instead, the purpose of this analysis was to test the NPP-based productivity groups could be independently distinguished using hydrographic and biogeochemical variables.

The discriminant analysis showed a low misclassification rate of 8.7% and high assignment accuracy among groups, with values of 92.4%, 96.1%, and 80.1% for Groups 1–3, respectively. These results indicate strong environmental separation among the productivity classes. We have clarified these methodological details in the revised Methods section and added the corresponding results to the Results section.

Method section, page 7, Lines 196 to 203: "A discriminant analysis was subsequently performed to evaluate whether the NPP-derived productivity classes corresponded to distinct environmental regimes. The productivity classes obtained from the k-means classification were used as the grouping variable, whereas the environmental variables, SST, EZD, MLD, NO_3, PO_4^{-3}, SiO_2, URT, UAR, SI, and SAL, were included as fixed effects within each model (Johnson and Wichern, 2018). NPP was deliberately excluded from this analysis because it was the variable used to define the original k-means classification. Its inclusion would therefore have introduced circularity and artificially inflated the classification performance. Thus, the purpose of the discriminant analysis was not to validate the clustering procedure or the selected number of clusters, but rather to determine whether the resulting NPP-based productivity classes could be independently distinguished by hydrographic and biogeochemical characteristics.".

In the section of result, page 10, Lines 263 to 265): "The discriminant analysis indicated clear environmental separation among the NPP-derived productivity classes. The overall misclassification rate was low (8.7 %), while assignment accuracy was high across groups, with values of 92.4 %, 96.1 %, and 80.1 % for Groups 1–3, respectively".

3. The clustering was done on NPP only. Why not on NPP together with other variables, such as SST and/or EZD? Was this attempted? If not, why not?

Response: We thank the reviewer for this important comment. The clustering was intentionally performed using NPP alone because the purpose of the regionalisation was to identify productivity regimes, rather than hydrographic or multivariate environmental provinces. Our objective was to classify the Gulf of California according to spatial differences in net primary productivity and then to evaluate the environmental factors associated with each productivity regime.

Including variables such as SST, EZD, MLD, salinity, or nutrients (NO_3, PO_4^{-3}, or SiO_2) directly in the clustering process would have fundamentally changed the aim of the analysis. Such an approach would have produced mixed hydrographic–biogeochemical clusters rather than NPP-based productivity classes. Moreover, because these environmental variables were subsequently analysed as potential drivers of NPP variability, including them in the clustering step could have introduced circularity and reduced interpretability of their explanatory role.

We therefore adopted a two-step analytical framework: First, the Gulf of California was classified solely on the basis of NPP; second, the environmental consistency and potential drivers of the resulting productivity classes were evaluated using independent hydrographic and biogeochemical variables. This approach allowed us to identify low-, intermediate-, and high-productivity regimes while preserving the ability to assess their environmental controls independently. We have clarified this rationale in the revised manuscript.

In the text page 7, Lines 186 to 187: "To analyse spatio-temporal variability and identify productivity regimes, we classify the Gulf of California according to differences in satellite-derived NPP from January 2003 to December 2024.

In the text page 7, Lines 196 to 197: A discriminant analysis was subsequently performed to evaluate whether the NPP-derived productivity classes corresponded to distinct environmental regimes.".

4. Was any comparison made between reanalysis products used and observations for this region? How confident can the authors be in their results based on reanalyses? Would the use of different reanalyses lead to different results (one might a priori think so)? Reanalyses are known to differ significantly.

Response: We thank the reviewer for this important comment. We clarify that the NPP analysed in this study was obtained from the satellite based Vertically Generalised Production Model (VGPM) product distributed through NOAA´s Environmental Research Division's Data Access Program (ERDDAP). This product was selected because it provided spatially coherent NPP fields for the Gulf of California and reproduces the well-established seasonal productivity cycle of the region, characterised by enhanced winter–spring production and reduced summer productivity under stronger stratification. We did not use the Copernicus-derived NPP product because preliminary comparisons showed spatial and seasonal inconsistencies in the Gulf of California relative to the expected regional productivity cycle.

Reanalysis products were used only for subsurface and water-column variables, including mixed-layer depth, nutrients, salinity, temperature profiles, and derived stratification index. These variables were required because long-term, basin-scale observational datasets with consistent spatial and temporal coverage are not currently available for the Gulf of California. We acknowledge that different reanalysis products may differ in their representation of oceanographic conditions, particularly in marginal seas and semi-enclosed basins. Therefore, the absolute magnitude of some environmental variables may vary depending on the product used.

However, our confidence in the main results does not rely on any single reanalysis field. Rather, it is based on the internal consistency of the relationships among NPP, temperature, mixed-layer depth, stratification, and nutrient availability, and on their agreement with the known oceanographic functioning of the Gulf of California. The identified mechanisms, enhanced productivity during cooler, more mixed conditions and reduced productivity during warmer, more stratified periods, are physically coherent and consistent with previous regional observational studies. Although the use of alternative reanalysis products could affect the magnitude of specific estimates, we do not expect the main spatial patterns or inferred mechanisms to change fundamentally. We have clarified this point and explicitly acknowledged this limitation in the revised manuscript.

In the Methods section, 2.2 Data collection pages 4–5 Lines 106–113: "Net primary productivity (NPP) was obtained from the satellite-based VGPM product distributed through NOAA ERDDAP. This product was selected because it provides spatially coherent NPP fields for the Gulf of California and reproduces the expected seasonal cycle of regional productivity, with enhanced winter–spring production and reduced summer productivity under stronger stratification. Reanalysis products were used only for subsurface and water-column variables, including reanalysis estimates of mixed-layer depth, nutrients, salinity, and temperature profiles, and the derived stratification index, for which long-term basin-scale observational datasets with consistent spatial and temporal coverage are not currently available in the Gulf of California. The following subsections summarize the datasets used for NPP, physical, biogeochemical, and climate variables."

5. What assumptions are implicit in the use of a generalised linear mixed model with Gaussian errors? Are these assumptions consistent with the behaviour of the variables being fitted? Also given that there is an effect due to the Colorado River (freshwater?) why wasn't salinity included in the analysis?

Response: We thank the reviewer for this observation about the assumptions of the Generalised Linear Mixed Model (GLMM) and the role of salinity in the regional analysis.

In the original version of the manuscript, NPP was modelled using a GLMM with Gaussian errors. We acknowledge that this formulation assumes normally distributed residuals, homoscedasticity, an approximately linear relationship between the predictors and the response variable under an identity link function, and independence of observations conditional on the specified random-effects structure. Following the reviewer´s suggestion, we re-evaluated these assumptions through residual diagnostics. These diagnostics indicated deviations from normality and evidence of heteroscedasticity, suggesting that the Gaussian formulation was not optimal for the distributional properties of NPP.

Regarding the potential influence of freshwater input from the Colorado River, we acknowledge that salinity is an important variable in such systems. To that end, we have now included salinity in the new analysis.

We therefore re-fitted the model using a Generalised Linear Mixed Model (GLMM) with a Gamma error distribution and a log link function, which is more appropriate for positive, continuous, and right-skewed ecological response variables such as NPP. Diagnostic plots for the revised model showed a clear improvement relative to the Gaussian formulation. The Q–Q plot showed a good agreement with the expected distribution across most of the range, with only minor deviations in the upper tail associated with a small number of high NPP values. Residuals plotted against fitted values showed no strong systematic pattern, indicating that heteroscedasticity was substantially reduced. The residual histogram was approximately symmetric, with a slight right tail consistent with the distribution of the response variable. Autocorrelation analysis indicated only weak short-term structure, which did not substantially affect model performance. Overall, these diagnostics suggest (Figure R2) that the revised GLMM provides a more suitable representation of the NPP data.

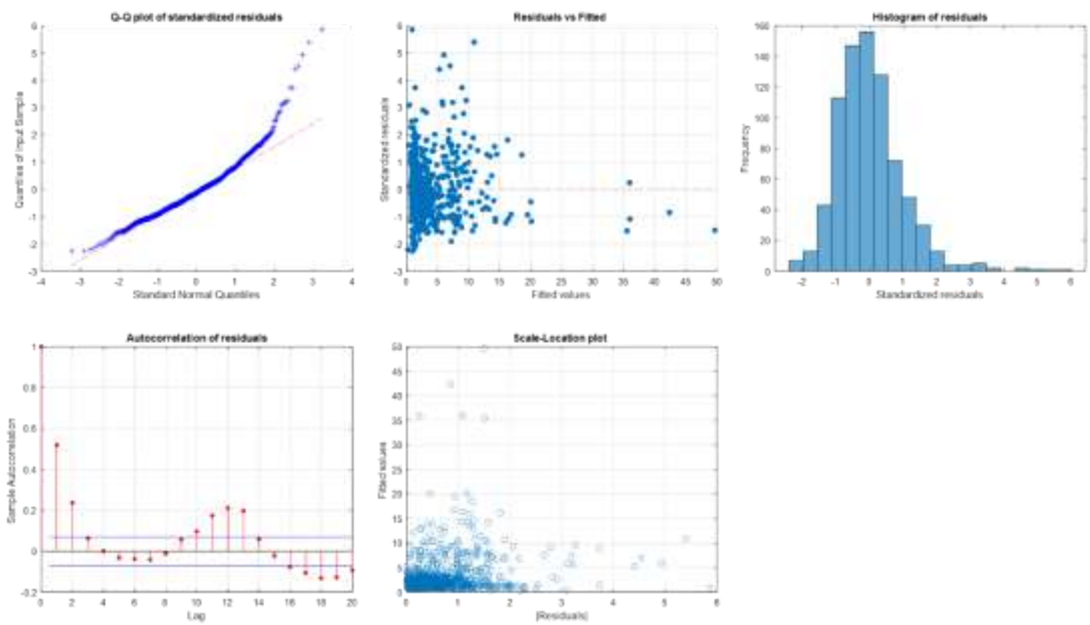

Figure R2. Q-Q plot shows standardised residuals; Residuals vs Fitted showed no strong systematic pattern; the residual histogram was approximately symmetric; autocorrelation analysis indicated only weak short-term structure.

The Colorado River no longer discharges regularly into the Gulf of California, as most of its freshwater flow is retained by dams or diverted for urban consumption and agricultural irrigation. This alteration resulted primarily from the construction and filling of the two major dams on the Colorado River: Hoover Dam (1935–1941) and Glen Canyon Dam (1963–1980). However, the influence of this river regulation has been largely restricted to the Upper Gulf of California, particularly the delta region, located approximately 70 km from the coast. Moreover, our study covers the period 2003–2024, during which freshwater release events into the Gulf of California were rare, and considers the Gulf of California as a whole. Consequently, the effect of the Colorado River on salinity is regional and negligible at the scale of the entire Gulf of California (Lavín & Sánchez, 1999).

We therefore do not consider it necessary to discuss the potential effects of the Colorado River in detail in the present manuscript neither quote Lavín & Sánche (1999) in the main document, as the study focuses on basin-wide patterns across the Gulf of California rather than specifically on the Upper Gulf. Furthermore, sea surface salinity (SSS) was included as a predictor variable in the Generalised Linear Model (GLM). Although SSS was inadvertently omitted from the textual description of the model, it was incorporated in all analyses and model calculations. This omission has now been corrected in the revised manuscript.

Reference

Lavín, M. F., & Sánchez, S. (1999). On how the Colorado River affected the hydrography of the upper Gulf of California. Continental Shelf Research, 19(12), 1545–1560.

6. The discussion section seems very long (and quite involved) compared to the overall length of the paper. This could do with being more succinct and easier to follow.

Response: We thank the reviewer for this helpful suggestion. Following the reviewer´s recommendation, we have carefully revised the Discussion section to make it more succinct and easier to follow. The section has been shortened by removing repetitive text, consolidating overlapping arguments, and emphasising only the findings most relevant to the objectives of the study. We have also improved the logical flow of the section so that the interpretation of the results is presented more clearly. These revisions have reduced the overall length of the Discussion while preserving the main ecological interpretations and implications of the work.

Minor comments:

- With regard to reanalysis variables (2.2.2) why weren't these obtained from CMEMS to be consistent with the use of CMEMS for currents? Why use different reanalyses?

Response: We thank the reviewer for this observation and apologise if the description in the Methods section was unclear. The reanalysis variables used in this study were in fact obtained consistently from Copernicus Marine Environment Monitoring Service (CMEMS) products. Specifically, ocean currents were derived from SEALEVEL_GLO_PHY_L4_NRT_008_047 (https://doi.org/10.48670/moi-00148) (satellite product); for the mixed-layer depth we used the Global Ocean Physics Reanalysis (GLOBAL_REANALYSIS_PHY_001_030), and nutrients variables NO_3, PO_4^{-3} and SiO_2 were obtained from the corresponding CMEMS biogeochemical reanalysis GLOBAL_REANALYSIS_BIO_001_029.

The apparent confusion may have arisen because NPP was not derived from CMEMS, but from the satellite-based VGPM product distributed through NOAA ERDDAP. This product was selected because it provided more spatially coherent estimates within the Gulf of California. We have clarified this distinction in the revised Methods section to avoid ambiguity and to make clear which variables were obtained from CMEMS and which were derived from external satellite-based products.

- In section 3.2. how did the authors determine which observations were misclassified? What is their definition of misclassification?

Response: We thank the reviewer for pointing out this ambiguity and agree that additional clarification is needed. In our analysis, misclassification refers to observations which the group assignment predicted by the discriminant analysis differed from the original class assigned by the k-means^{++} clustering.

Specifically, the three productivity classes obtained from the k-means^{++}algorithm with k-means^{++} initialisation, low-productivity zone (LPZ), mid-productivity zone (MPZ), and high-productivity zone (HPZ), were treated as the reference grouping variable. Discriminant analysis was then performed using the environmental variables, SST, EZD, MLD, NO_3, PO_4^{-3}, SiO_2, URT, UAR, and SI to evaluate whether the original NPP-derived groups could be independently distinguished based on hydrographic and biogeochemical conditions. Observations were considered misclassified when the discriminant analysis reassigned them to a different group than their original k-means^{++} classification. Using this criterion, 8.7% of observations were reassigned, whereas 92.4%, 96.1%, and 80.1% of observations remained correctly assigned to Groups LPZ, MPZ or HPZ, respectively. We have clarified this definition in the revised Methods and Results sections.

- In section 3.3 there is mention of the effect of the Colorado River outflow, but its position doesn't seem to be marked on any of the figures.

Response: We thank the reviewer for this helpful suggestion. Colorado River mouth is mentioned twice in the manuscript as specific location of HPZ. In response, we have added a marker indicating the location of the Colorado River delta/outflow region to the study-area map. We also reviewed the map scale to ensure that the distance scale is correctly represented. This revision improves the spatial context for the discussion of the Colorado River influence in the Upper Gulf of California.

- Figure 5 is quite small (even smaller when printed in the journal) so it is difficult to see the details. Also, in (b) there seems to be some anti-correlation between LPZ and HPZ – is this significant?

Response: We have revised Figure 5 by increasing its size and improving its readability, including the visibility of labels, axes, and line patterns.

Regarding the apparent anti-correlation between LPZ and HPZ in panel (b), we agree that the temporal patterns suggest an inverse relationship: periods with increased HPZ extent tend to coincide with reduced LPZ extent, and vice versa. This pattern is expected because the productivity zones represent spatial fractions of the same system; therefore, an expansion of one productivity class necessarily reduces the relative area available for one or both other classes. We have now evaluated this relationship statistically and clarified the result in the revised text. Where significant, we describe this as an inverse association between the spatial extent of LPZ and HPZ, while avoiding over interpretation because the relationship partly reflects the compositional nature of the classified area.

- Line 263 states that the regions determined by the authors align with previous classifications, so what is new / different here?

Response: We thank the reviewer for pointing out this ambiguity. We agree that the previous wording could have suggested that our classification simply reproduced earlier regional divisions. To avoid this misunderstanding, we have removed that statement from the revised manuscript.

Our classification differs from previous classifications because it was derived directly from spatial and temporal patterns of net primary productivity using an objective clustering approach, rather than being based on predefined hydrographic, geomorphological, or biogeographical regions. We have revised the text to emphasise that the novelty of our approach lies in identifying productivity-based zones across the Gulf of California and evaluating their environmental drivers and temporal variability.

- Line 337 mentions a “tipping point” but what do the authors mean by this term? There is confusion in the literature about use of the term “tipping point” and here the authors are referring to a period 2014-2016 rather than a point in time.

Response: We thank the reviewer for pointing out this ambiguity. We agree that the term “tipping point” may be inappropriate in this context, as it implies a threshold-type transition occurring at a specific point in time. In our manuscript, we intended to refer to a change in the temporal tendency of climate and productivity-related conditions during the 2014–2016 period, rather than to a formal ecological or climatic tipping point.

To avoid ambiguity, we have removed the term “tipping point” from the revised manuscript. We now describe this period more precisely as a marked shift or change in trend associated with regional climatic variability during 2014–2016. This revised wording better reflects the nature of the observed pattern and avoids implying the existence of a threshold response that was not explicitly tested in this study.