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 2

This study uses two decades of satellite and reanalysis data to map and track three productivity zones, high, medium, and low, in the Gulf of California and to examine their seasonal and interannual variability. The descriptive contribution is a genuine strength. The manuscript presents a long-term, basin-scale account of where and when productivity is high or low, including a well-documented 2014 to 2016 decline and subsequent recovery. The use of machine learning methods is appropriate for this kind of question, and the manuscript has the potential to be an excellent contribution.

However, I have serious concerns about the methods as applied, which currently prevent the central conclusions from being supported. The most important is a lack of independence among the variables. Causal control over NPP is attributed to predictors, notably SST and euphotic zone depth, that are themselves inputs to the VGPM algorithm used to compute NPP. As a result, the driver analysis is in part recovering the structure of the production model rather than independent environmental control. This is not acknowledged or accounted for, and it must be addressed before the conclusions can stand. A second concern is that the number of clusters is not justified by any cluster validity analysis, and there are a few indications that the clustering method is not fully understood, which I detail below. I recommend major revisions. With the predictor set properly decorrelated, the classification justified or reframed, and model performance cross validated, this could become a sound paper. My specific comments follow.

Response: As explained earlier, 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.

Major Revisions

1. The manuscript uses a Generalized Linear Mixed Model to identify significant drivers of NPP (Section 2.3.4). A mixed model requires a random effect; which models structure in the data such as the grouping of repeated observations within zones or within months. The random effect is not defined anywhere in the methods, and the stated model structure (line 160) contains only fixed effects. If a random effect was included, please specify it. If one was not, the model is a fixed effects GLM rather than a GLMM and should be described as such. This distinction matters because a properly specified random structure is one of the standard tools for handling non-independence in the data, which is itself a central concern with the current analysis (see comments below). Additionally, the model mixes standardized predictors (SST, EZD, MLD, per line 153) with apparently unstandardized ones (nutrients, currents, winds, SI), evident in the order-of-magnitude differences in standard errors in Table 1 (e.g. UAR ~1.8–2.2 vs SST ~0.08). Because coefficient magnitudes are only comparable across predictors on a common scale, the conclusion that "SST was the dominant predictor" cannot be read from coefficient size as currently presented. Consider standardizing (z-scored) all continuous predictors before fitting.

Response: We thank the reviewer for this valuable observation. We agree that the original description of the model did not explicitly specify the random-effects structure, which could lead to ambiguity regarding the use of a Generalised Linear Mixed Model-effects framework. Following the reviewer's suggestion, we revised the model specification and incorporated month as a random intercept to account for temporal dependence among observations collected within the same month. The final model was fitted as:

$$NPP_{zone} \sim MLD + NO_3 + PO_4 + UWR + UAR + SI + SSS + (1|Month)$$

where NPP_{zone} indicates that the same model structure was fitted independently for each productivity zone (LPZ, MPZ, and HPZ). Thus, productivity zone was used only to stratify the dataset prior to model fitting and was not included as a fixed effect. In this formulation, Month represents the random intercept, whereas all environmental variables (MDL = mixed-layer depth, NO_3 = nitrate concentration, PO_4^{-3} = phosphate concentration, UAR = along-gulf (u) sea current component; URT = along-gulf (u) wind component; SI = stratification index, SSS = sea surface salinity) were included as fixed effects.

The Methods section has been revised accordingly to explicitly describe the mixed-effects structure and the rationale for including Month as a random effect. To verify that the mixed-effects formulation was justified, we compared the GLMM with an equivalent Generalised Linear Mixed Model lacking the random term using a likelihood ratio test. The inclusion of the random effect significantly improved model performance (LR = 152.51, $\Delta df = 1$, $p < 0.001$) and reduced both AIC (1062.3 vs. 1212.8) and BIC (1169.8 vs. 1315.6), supporting the use of the GLMM in the final analysis.

We also appreciate the reviewer's observation regarding the scaling of predictor variables. We agree that comparing regression coefficients across predictors requires a common measurement scale. Accordingly, all continuous predictor variables were standardised using z-score transformation prior to model fitting. The models were subsequently refitted using the standardised predictors, allowing direct comparison of effect sizes while also improving numerical stability during model estimation. Consequently, Table 1 and the corresponding Results section have been updated, and statements regarding the relative importance of environmental drivers are now based on standardised coefficients rather than raw coefficient magnitudes.

In the text page 8-9, Lines 212 to 234: "To explain spatial and temporal variability in NPP, the following biophysical variables were analysed for each NPP zone: SST, EZD, MLD, NO3, PO4-3, SiO2, URT, UAR, SI, and SSS. Monthly time series were computed by zone for each variable. Nutrients (NO_3, PO_4^{-3}, SiO_2) were vertically integrated from the surface to the EZD prior to analysis. Potential collinearity among predictors was evaluated with correlation matrices considering Pearson's correlation > 0.7; highly correlated variables were excluded from the final predictor set. To avoid potential circularity between the response variable and predictor variables, SST, euphotic zone depth (EZD), and SiO2 were excluded from the final GLMM because these variables are directly involved in, or derived from quantities used in, the VGPM algorithm employed to estimate NPP. To identify the environmental drivers of net primary productivity (NPP) and evaluate whether their effects differed among productivity zones, we fitted a Generalised Linear Mixed Model (GLMM) using a Gamma error distribution and a log-link function, which is appropriate for continuous, positive, and right-skewed response variables such as NPP (Corrales and Cepeda-Cuervo, 2019; Miyamoto et al., 2019). Prior to modelling, all continuous predictor variables were standardised (z-scores) to facilitate comparison of effect sizes and improve model convergence. Separate models were fitted for each productivity zone; MLD, NO_3, PO_4^{3-}, URT, UAR, SI, and SSS were included as fixed effects within each model. To account for temporal dependence among observations and repeated measurements through time, month was included as a random intercept term (1|Month). The model structure was:

$$NPP_{zone} \sim MLD + NO_3 + PO_4 + UWR + UAR + SI + SSS + (1|Month)$$

where NPP_{zone} indicate that the same GLMMs structure was fitted independently to each net productivity zone class (low-productivity zone (LPZ), mid-productivity zone (MPZ), and high productivity zone (HPZ) as defined in result section); therefore, zone was used only to stratify the dataset prior to model fitting and was not included as a predictor in the models. MDL = mixed-layer depth, NO_3 = nitrate concentration, PO_4^{-3} = phosphate concentration, UAR = along-gulf (u) sea current component; URT = along-gulf (u) wind component; SI = stratification index, SSS = sea surface salinity, and 1|Month represents the random effect, and the environmental variables correspond to fixed effects."

In the Methods section, Section 3.5, pages 15–16 Lines 331–351: "Best-fitting Generalised Linear Mixed Models (GLMMs) for each NPP class identified the significant environmental drivers (Table 1). The model provided a good fit to the data (AIC = 1057.4, BIC = 1178.9, log-likelihood = −502.69). Month was included as a random intercept to account for temporal dependence among observations, yielding an estimated random-effect standard deviation of 0.504.

Table 1. GLMM results for main effects of biophysical variables on NPP by productivity zone classification (LPZ, MPZ and HPZ). Coefficients $p < 0.05$ are in bold. Abbreviations: LPZ = low-productivity zone; MPZ = mid-productivity zone; HPZ = high-productivity zone; MLD = mixed layer depth; NO_3 = nitrate concentration; PO_4^{-3} = phosphate concentration; UAR = along-gulf (*u*) sea current component; and URT = along-gulf (*u*) wind component; SI = stratification index, and SSS = sea surface salinity.

Productivity zone	Environmental drivers	Estimate	Std. Error	*t*-ratio	*p*-value
LPZ	Intercept	0.124	0.168	0.737	0.461
	MLD	0.079	0.086	0.921	0.357
	NO_3	-1.841	0.261	-7.063	**3.7×10^{-12}**
	PO_4^{-3}	-0.091	0.051	-1.772	0.077
	URT	0.326	0.072	4.504	**7.7×10^{-6}**
	UAR	-0.061	0.050	-1.206	0.228
	SI	1.050	0.105	10.026	**2.6×10^{-22}**
	SSS	-0.176	0.048	-3.682	**2.5×10^{-4}**
MPZ	Intercept	0.484	0.089	5.465	**6.3×10^{-8}**
	MLD	0.081	0.094	0.862	0.389
	NO_3	1.926	0.266	7.247	**1.0×10^{-12}**
	PO_4^{-3}	0.097	0.060	1.620	0.106
	URT	-0.122	0.090	-1.357	0.175
	UAR	0.033	0.063	0.524	0.600
	SI	-1.116	0.110	-10.109	**1.2×10^{-22}**
	SSS	0.196	0.055	3.555	**4.0×10^{-4}**
HPZ	Intercept	1.307	0.090	14.502	**2.7×10^{-42}**
	MLD	-0.059	0.106	-0.558	0.577
	NO_3	2.008	0.260	7.720	**3.7×10^{-14}**
	PO_4^{-3}	0.294	0.057	5.109	**4.1×10^{-07}**
	URT	-0.336	0.094	-3.572	**3.8×10^{-4}**
	UAR	-0.048	0.056	-0.855	0.393
	SI	-1.117	0.105	-10.645	**9.0×10^{-25}**
	SSS	0.182	0.054	3.370	**7.9×10^{-4}**

The stratification index (SI), sea surface salinity (SSS), and nitrate (NO_3) were significant predictors of NPP across the three productivity zones, although their effects differed in both magnitude and direction (Table 1). In the LPZ, SI was positively associated with NPP ($\beta = 1.050$, $p < 0.001$), whereas salinity ($\beta = -0.176$, $p < 0.001$) and nitrate ($\beta = -1.841$, $p < 0.001$) showed significant negative relationships. In contrast, both the MPZ and HPZ exhibited the opposite pattern. SI was negatively associated with NPP (MPZ: $\beta = -1.116$; HPZ: $\beta = -1.117$; $p < 0.001$), whereas salinity (MPZ: $\beta = 0.196$; HPZ: $\beta = 0.182$; $p < 0.001$) and nitrate (MPZ: $\beta = 1.926$; HPZ: $\beta = 2.008$; $p < 0.001$) were positively associated with productivity. The effect of URT also varied among productivity zones. URT was positively associated with NPP in the LPZ ($\beta = 0.326$, $p < 0.001$), whereas a significant negative relationship was observed in the HPZ ($\beta = -0.336$, $p < 0.001$). In contrast, URT was not a significant predictor in the MPZ. Phosphate (PO_4^{-3}) was significant only in the HPZ, where it exhibited a positive association with NPP ($\beta = 0.294$, $p < 0.001$)."

2. On lines 155 to 156 the authors describe the model as a GLMM with Gaussian errors. Choosing Gaussian errors assumes the data are symmetric around the average, with roughly even spread above and below it, like a normal bell curve. NPP does not behave this way. Because it is a productivity measurement it cannot go below zero, and values like these are usually bunched up at the low end with a long tail of high values, rather than spread evenly. This is a common and well known feature of productivity and concentration data, and it is the reason researchers often work with the logarithm of these values instead. The manuscript does not show any check that the Gaussian assumption is reasonable here, such as a plot of the model residuals or a test for skew. This matters because the flexibility to handle data that are not bell shaped is the main advantage of this family of models. Choosing the Gaussian option sets that flexibility aside and treats the skewed NPP data as if it were symmetric, which can distort the model and make the significance values for each driver unreliable. I recommend the authors either fit the model using an error type suited to positive, skewed data, such as a Gamma or log normal, or take the logarithm of NPP before fitting. If the

authors prefer to keep the Gaussian model, they should show that the residuals are well behaved and justify that choice.

Response: We thank the reviewer for this insightful comment and agree that the original Gaussian specification was not the most appropriate choice for modelling NPP. As noted by the reviewer, NPP values are strictly positive and exhibited a positively skewed distribution, characteristics that are more consistent with a Gamma-distributed response than with a Gaussian error structure. Following this recommendation, we refitted the model using a Gamma error distribution with a log-link function, which is widely used for positive, continuous, and right-skewed ecological response variables. This formulation preserves the original scale of the response variable while explicitly accounting for its non-normal distribution. In addition, all continuous predictor variables were standardised prior to model fitting, and month was incorporated as a random intercept to account for temporal dependence among observations. The revised model was specified as:

$$NPP_{zone} \sim MLD + NO_3 + PO_4 + UWR + UAR + SI + SSS + (1|Month)$$

with a Gamma distribution and log-link function.

To evaluate model adequacy, we examined residual diagnostics, including quantile–quantile plots, residual-versus-fitted value plots, and Pearson and deviance residual distributions. These diagnostics indicated no major departures from model assumptions and supported the suitability of the Gamma GLMM framework for the analysed data. The Methods section has been revised to describe the new model structure and diagnostic procedures, and the Results section has been updated accordingly.

Please see our detailed response in Reviewer 2's Comment 1 and the corresponding revisions in the Methods and Results sections.

3. A central problem with the feature importance analysis here is the circularity of the features going into the GLM. NPP in this study is not measured, it is calculated by a model (VGPM) based on measured satellite products. Two of the input variables are SST and euphotic zone

depth. When the analysis then asks what controls NPP, SST and euphotic-zone depth emerge as the strongest predictors because they are the same quantities used to compute NPP in the first place. The coupling is especially tight for euphotic-zone depth, which the authors derive directly from chlorophyll-a (lines 91 to 92), and chlorophyll is itself the primary driver of the VGPM calculation. EZD and NPP are therefore linked mathematically before any ocean process is considered. A further consequence is that the analysis dismisses genuinely independent variables that likely exert real control, such as the wind components, nitrate, and stratification, all of which were non-significant in every zone (Table 1). This circularity is visible in the euphotic-depth result itself, which carries the wrong sign. The physically expected relationship is negative, because more productive water is greener and attenuates light more strongly, so high NPP should correspond to a shallower euphotic zone. The authors describe exactly this negative relationship in their discussion (lines 310 to 311), stating that enhanced mixing "reduces euphotic zone depth anomalies, reflecting increased phytoplankton biomass and reduced light penetration associated with productive conditions." Yet the GLMM reports a significant positive coefficient for EZD in the HPZ ($\beta = 0.775$, $p < 0.01$; Table 1, lines 234 to 235). The model thus contradicts both the established physical relationship and the authors' own text, which is most consistent with the coefficient reflecting collinearity among the entangled predictors rather than a physical effect.

Response: We thank the reviewer for this thoughtful and important observation. We agree that including variables that are directly involved in, or closely linked to, the computation of VGPM-derived NPP may introduce circularity into the analysis and potentially confound the interpretation of environmental controls on productivity. Following the reviewer's comment, we re-evaluated the predictor set. SST and EZD were removed because they are directly involved in, or closely linked to, the computation of VGPM-derived NPP. In addition, SiO_2 was excluded because of its strong collinearity with PO_4. The revised model focuses exclusively on environmental variables that are not directly used in the calculation of NPP, including nitrate, phosphate, salinity, wind-related variables, and stratification indices. This modification reduces the potential for circularity and allows a more robust assessment of the physical and biogeochemical factors associated with NPP variability across productivity zones. Consequently, all interpretations related to SST and EZD have been removed from the Results and Discussion sections, and the corresponding analyses have been updated throughout the manuscript. The Methods, Results, Table 1, and Discussion sections have been revised accordingly.

Please see our detailed response in Reviewer 2's Comment 1 and the corresponding revisions in the Methods and Results sections.

4. The GLM(M) treats each observation as independent through time. However, measured oceanographic data are inherently autocorrelated. The temperature and chlorophyll of a given month are closely related to those of the previous month, and the pattern repeats every 12 months with the seasonal cycle. The consequence is that significance values can be artificially low, because the test assumes far more independent observations than the data actually contain. The *p*-value for the HPZ SST coefficient in Table 1 may be an indicator of this. It is on the order of 10^-48, and it would be reasonable to expect a 22-year monthly series to produce a much larger p-value than that, even if the relationship ultimately remains statistically significant. This issue would be present anywhere significance is claimed, for example Table 1, the Spearman correlations (Methods, line 122), the TREND least-squares regression (Methods, lines 123 to 124), and the ENSO correlation $r_s(264) = -0.21$, $p < 0.05$ (lines 318 to 319). The latter is also a weak relationship in absolute terms, explaining only about 4% of the variance, so even setting the significance aside, ENSO accounts for very little of the variability in HPZ coverage.

   The autocorrelation issue affects the ability to determine whether trends or slopes are significant. It does not necessarily mean the relationships are wrong. Autocorrelation is straightforward to account for, and I propose the following changes, though other equally valid methods exist. For the GLMM, include an autoregressive error term or a properly specified random effect, and report the effective degrees of freedom. When analyzing trends, use a statistical test designed for serially dependent data, such as Mann-Kendall with Theil-Sen, which is also available in the same Climate Data Toolbox. For the ENSO correlation, correct the effective sample size for autocorrelation before claiming significance, and report the effect size.

Response: Below is a revised version of the response, organised clearly into the three components identified by the reviewer: (1) GLMM, (2) trend analysis, and (3) correlations.

We thank the reviewer for this important observation. We agree that temporal autocorrelation is a relevant issue when analysing monthly environmental time series, as it may inflate the effective degrees of freedom and lead to overestimation of statistical significance. In response, we revised the statistical analyses and clarified the manuscript in three specific ways.

First, regarding the GLMM analysis, we acknowledge the reviewer's concern about temporal dependence among monthly observations. The GLMMs now included Month as a random intercept, which accounts for repeated observations through time in the data. We have now clarified this aspect of the model specification in the revised manuscript and improved the model formulation to make its interpretation more explicit. In addition, the model structure was revised to avoid over parameterisation and to remove variables that could introduce bias or redundancy in the analysis. The final model was specified as:

$$NPP_{zone} \sim MLD + NO_3 + PO_4 + URT + UAR + SI + SSS + (1 \mid Month)$$

with a Gamma distribution and a log-link function. This revised formulation provides a clearer and more appropriate framework for evaluating the environmental drivers of primary productivity while accounting for repeated monthly observations. Please see our detailed response in Reviewer 2's Comment 1 and the corresponding revisions in the Methods and Results sections.

Second, regarding the trend analysis, we replaced the original least-squares regression approach with a non-parametric method that is more appropriate for serially dependent and non-normally distributed environmental time series. Specifically, we used the Mann–Kendall test to evaluate the statistical significance of monotonic trends and the Theil–Sen slope estimator to quantify their magnitude, using the Mann–Kendall and Theil–Sen functions available in the Climate Data Toolbox. This approach is more robust to non-normality and less sensitive to outliers, making it better suited for long-term oceanographic data. Accordingly, the trend analysis was recalculated for each grid cell of the monthly primary productivity dataset. The corresponding figure was also updated to show the spatial distribution of Theil–Sen slopes, with statistically significant trends identified by the Mann–Kendall test indicated by black dots. This revised representation simultaneously shows the direction, magnitude, and statistical significance of long-term changes in primary productivity.

In the text page 7, Lines 171 to 174: "Monotonic trends were assessed using the Mann–Kendall test to determine their statistical significance and the Theil–Sen slope estimator to quantify their magnitude. Both analyses were performed using the Mann–Kendall and Theil–Sen functions implemented in the Climate Data Toolbox for MATLAB (Greene et al., 2019)."

In the Supplementary material:

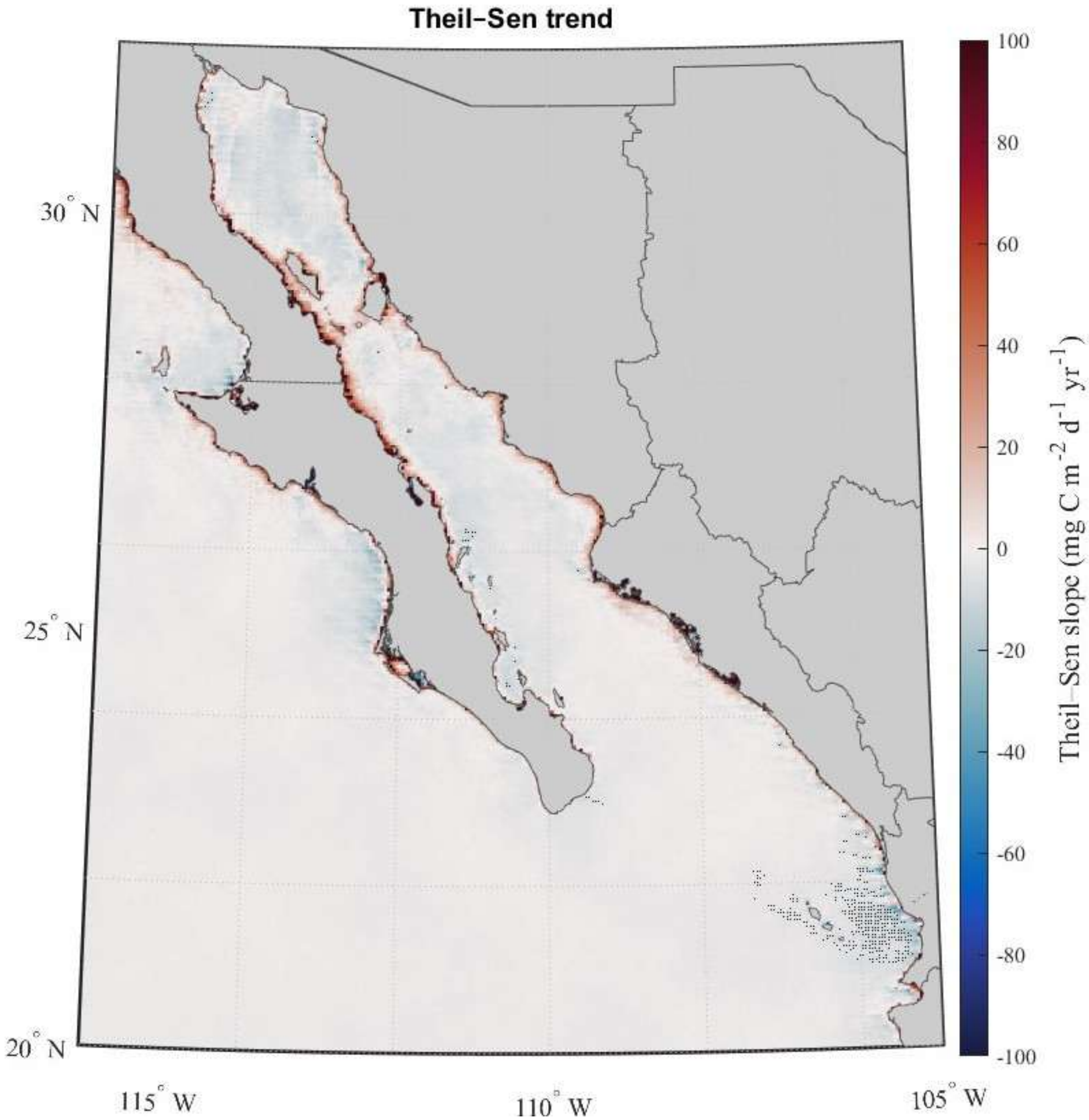

Figure A10. Spatial trends in the NPP (mg C m^{-2} d^{-1} $year^{-1}$) across the Gulf of California during 2003–2024. Cool colours indicate negative trends, and warm colours indicate positive trends. Dots identify pixels with statistically significant trends ($p<0.05$).

Third, regarding the correlation analyses, including the relationship between ENSO and the extent of productivity zones, we re-evaluated the Spearman correlations by accounting for temporal autocorrelation through an effective sample size correction based on the autocorrelation structure of each time series, following Pyper and Peterman (1998). We now report Spearman's correlation coefficient (r_s), which provides a measure of the strength of the association, together with corrected significance values. After applying this correction, most correlations remained weak and statistically non-significant once temporal autocorrelation was considered. The only relationship that remained statistically significant was the weak negative association between MEI v2 and the high-productivity zone, although the effect size was small ($r_s = -0.136, corrected p = 0.027$). We have therefore revised the manuscript to avoid overstating the influence of ENSO and now describe this result as a weak but statistically significant association.

Reference: Pyper, B. J., & Peterman, R. M. (1998). Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(9), 2127–2140.

In the text page 7, Lines 169 to 171: "Spearman's rank correlations coefficients were calculated to evaluate the relationships among hydrographic variables, climate indices, and NPP across gulf regions. To account for temporal autocorrelation, statistical significance was assessed using an effective sample size correction following the method of Pyper and Peterman (1998)."

In the Supplementary material: Table A1. Spearman rank correlations (r_s) between the climate indices (PDO, NPGO, and MEI.v2) and productivity zone extent (LPZ, MPZ, and HPZ). Significance levels were corrected for temporal autocorrelation using the effective sample size method of Pyper and Peterman (1998). N_{eff} denotes effective sample size after accounting for temporal autocorrelation; p_{Orig} = the original significance level obtained assuming independent observations; p_{Corr} = the autocorrelation-correction significance level. Significant correlations (α=0.05) are shown in bold.

Climate index	Zone	r_s	N_{eff}	p_{Orig}	p_{Corr}
PDO	HPZ	-0.118	91.33	0.055	0.263
	MPZ	0.055	139.78	0.373	0.518
	LPZ	0.058	106.98	0.344	0.550
NPGO	HPZ	-0.097	141.99	0.118	0.253
	MPZ	-0.059	140.09	0.337	0.486
	LPZ	0.065	189.32	0.293	0.374
MEI .V2	HPZ	-0.136	264	**0.027**	**0.027**
	MPZ	-0.043	264	0.486	0.486
	LPZ	0.116	264	0.061	0.061

5. For the k-means clustering, it is not clear where k=3 comes from. The authors do not provide any elbow, silhouette, gap statistic, or stability analysis to support this choice. The authors further suggest that k=3 is consistent with other analyses that divide the region but those values range from 12-22 and general cardinal directions (lines 253-254). An easy fix is to report a cluster-validity criterion or reframe the three zones as threshold-based opposed to data-determined clusters.

Response: We thank the reviewer for this observation. This issue was also raised by Reviewer 1, and we have addressed it comprehensively in the revised manuscript. Specifically, we now describe the rationale for selecting K = 3, including the results of the elbow and silhouette analyses, and clarify that the final choice was based on the most parsimonious and ecologically interpretable solution rather than on arbitrary thresholds.

Please see our detailed response to Reviewer 1 (Comment 1) and the corresponding revisions in the Methods section.

6. K-means does not produce misclassified observations; it assigns each point to its nearest centroid. However, the authors report "8.7% misclassified" and per-group "accuracies" of 92.4/96.1/80.1% (lines 173–174) which can only come from the discriminant analysis (how well the non-NPP variables reproduce the cluster labels), not from the clustering itself. Presenting these as if they validate the clustering conflates two different things and suggests the method was not fully understood. Additionally, discriminant reclassification accuracy does not validate the number of clusters. Any number of k would yield high accuracy because separating bands of a gradient by correlated variables is trivial. As evidence, the weakest group is the middle one (MPZ, 80.1%), exactly the signature of a boundary-straddling band in a continuum rather than a discrete regime. Reclassification accuracy is a valid separability check when the classes are defined independently and assessed with cross-validation, but here the classes were produced by k-means on NPP itself, so high accuracy is largely circular and cannot demonstrate that three distinct regimes exist.

Response: We agree with the reviewer and have clarified in the Methods that the discriminant analysis was not intended to validate the clustering or the selected number of clusters, but only to evaluate the environmental separability of the NPP-derived productivity classes.

Please see our detailed response to Reviewer 1 (Comment 1) and the corresponding revisions in the Methods section.

7. The GLMM can remain a valid tool if the issues above are addressed. Another machine learning algorithm worth considering is Random Forest with feature permutation and cross-validation. Random Forest makes no assumptions about the data distribution, so right-skewed data are not a problem. The GLMM also imposes a linear relationship for each predictor, but phytoplankton growth is well established as a non-linear process, and the temperature response in particular rises to an optimum and then declines. Random Forest accounts for non-linearity automatically, whereas the GLMM may dismiss a real driver

simply because its effect is not linear. As discussed above, the GLMM p-values are also unreliable due to autocorrelation and the non-Gaussian response. Feature permutation in Random Forest is a more direct measure of how much information a predictor carries about NPP, because it measures how much the model's accuracy degrades when that variable is shuffled, rather than relying on a coefficient and standard error that depend on assumptions the GLMM violates.

It is important to reiterate that Random Forest does not resolve every issue. Autocorrelation would still need to be handled through blocked (temporal) cross-validation so that adjacent, correlated months are not split across training and testing. The circularity among SST, euphotic depth, and NPP would also persist with the current predictor set, regardless of model, and must be fixed by removing the VGPM-input variables. Finally, because the predictors here are collinear, standard permutation importance can distribute importance arbitrarily among correlated variables, so a correlation-aware variant would be needed for the rankings to be trustworthy.

Response: We thank the reviewer for this thoughtful suggestion. We agree that Random Forest is a powerful non-parametric approach for modelling complex, nonlinear relationships and assessing predictor importance. However, the primary objective of this study was not predictive modelling but to evaluate the direction and relative magnitude of associations between environmental variables and NPP within each productivity regime. For this reason, we chose Generalised Linear Mixed Models because they provide interpretable parameter estimates while accounting for the hierarchical structure of the data. We agree that Random Forest represents a valuable complementary approach and will consider it in future work specifically aimed at predictive modelling and nonlinear ecological responses.

8. The discussion and conclusions repeatedly attribute NPP variability to the PDO, and in one case the NPGO. The recovery after 2019 is tied to "the negative phase of the PDO that began in late 2019" (line 377), the 2014 to 2016 event to "a positive PDO and the transition to a negative NPGO" (line 339), and the conclusions state broadly that positive phases of MEI v2 and PDO suppress NPP (lines 379 to 380). However, the only climate index actually included in the methods and results is MEI v2. There is no PDO or NPGO time series,

correlation, or any other analysis anywhere in the manuscript. These are causal claims that the analysis never tested. The authors should either add PDO and NPGO to the methods and results, analyzing them against the zone metrics the same way MEI v2 was handled, or demote them to context and present them as plausible co-occurring modes rather than demonstrated drivers. If they are added, two cautions apply. PDO is partly an integration of ENSO forcing, so it is correlated with MEI v2 and should not be treated as an independent driver without addressing that collinearity. The ENSO and PDO teleconnections to this region are also concentrated in the cold season, so any test should be season-resolved rather than run on annual aggregates.

Response: We thank the reviewer for this thoughtful comment. We apologize that the original manuscript did not clearly describe the analyses performed for the PDO and NPGO indices, which may have led to the impression that only MEI.v2 was evaluated.

In fact, Spearman's rank correlations were calculated for all three climate indices (PDO, NPGO, and MEI.v2) against the extent of each productivity zone (LPZ, MPZ, and HPZ). To account for temporal autocorrelation, statistical significance was evaluated using the effective sample size correction of Pyper and Peterman (1998). These analyses are now explicitly described in the Methods section, and the complete results have been included in Supplementary Table A1. The revised analyses indicate that neither PDO nor NPGO exhibited statistically significant correlations with productivity zone extent after correcting for temporal autocorrelation ($p > 0.05$ for all comparisons). In contrast, MEI.v2 showed a weak but statistically significant negative correlation with the HPZ extent ($r_s = -0.136$, corrected $p = 0.027$), whereas correlations with MPZ and LPZ were not significant. Accordingly, we have revised the Discussion and Conclusions to ensure that the interpretation reflects these results. Statements referring to PDO and NPGO as direct drivers of NPP variability have been removed or rephrased. PDO and NPGO are now discussed only as large-scale climate modes that provide regional climatic context, whereas conclusions regarding statistically supported climate–NPP relationships are restricted to MEI.v2. We also appreciate the reviewer's comments regarding the potential collinearity between PDO and ENSO and the seasonality of their teleconnections. Because PDO and NPGO did not show significant relationships with productivity zone extent in our analyses, we did not pursue additional multivariate or season-resolved analyses. We agree that these represent valuable avenues for future research.

In the text page 7, Lines 169 to 171: "Spearman's rank correlations coefficients were calculated to evaluate the relationships among hydrographic variables, climate indices, and NPP across gulf regions. To account for temporal autocorrelation, statistical significance was assessed using an effective sample size correction following the method of Pyper and Peterman (1998)."

In the Supplementary material: Table A1. Spearman rank correlations (r_s) between the climate indices (PDO, NPGO, and MEI.v2) and productivity zone extent (LPZ, MPZ, and HPZ). Significance levels were corrected for temporal autocorrelation using the effective sample size method of Pyper and Peterman (1998). N_{eff} denotes effective sample size after accounting for temporal autocorrelation; p_{Orig} = the original significance level obtained assuming independent observations; p_{Corr} = the autocorrelation-correction significance level. Significant correlations (α=0.05) are shown in bold.

Climate index	Zone	r_s	N_{eff}	p_{Orig}	p_{Corr}
	HPZ	-0.118	91.33	0.055	0.263
PDO	MPZ	0.055	139.78	0.373	0.518
	LPZ	0.058	106.98	0.344	0.550
	HPZ	-0.097	141.99	0.118	0.253
NPGO	MPZ	-0.059	140.09	0.337	0.486
	LPZ	0.065	189.32	0.293	0.374
	HPZ	-0.136	264	**0.027**	**0.027**
MEI .V2	MPZ	-0.043	264	0.486	0.486
	LPZ	0.116	264	0.061	0.061

In the text page 18-19, Lines 412 to 416: "The 2003–2024 NPP time series exhibited significant interannual variability. Spearman's rank correlation analyses indicated that neither the PDO nor the NPGO was significantly associated with the extent of any productivity zone (Table A1). In contrast, MEI.v2 showed a weak but statistically significant negative correlation with the extent of the HPZ ($rs_{[264]:}$ −0.136, p_{corr} = 0.027), suggesting that ENSO-related variability exerts a measurable, albeit limited, influence on regional productivity patterns."

Minor Revisions:

- Overall: Please fix the labels on the colour bars of the NPP zones. They should read <900, 900-1700, >1700.

Response: In response to the reviewer's observation, we changed the information in the colour bars of every figures.

- Line 9: "mid (MPZ; 900-1700…)" is repeated

Response: We thank the reviewer for this observation. The repeated definition of "mid productivity zone (MPZ; 900–1700 …)" has been removed.

- Line 44: consider changing "mid and upper trophic levels" to "mid- and upper-trophic levels"

Response: We thank the reviewer for this suggestion. The text has been revised accordingly, and "mid and upper trophic levels" has been changed to "mid- and upper-trophic levels."

- Line 49: When introducing the Gulf of California, consider including longitude coordinates as well.

Response: We thank the reviewer for this suggestion. Longitude coordinates have been added to the geographic description of the Gulf of California.

- Line 50: It would be helpful to reference Figure 1 in the first sentence.

Response: In attended your observation, we added a reference Fig. 1 in the first sentence.

- Line 73: Bahía de Banderas is not labeled on the map and unfamiliar readers may not know where this is. Consider rewording or adding the location to the map.

Response: In response to your suggestion, we have attached the location of Banderas Bay on the map.

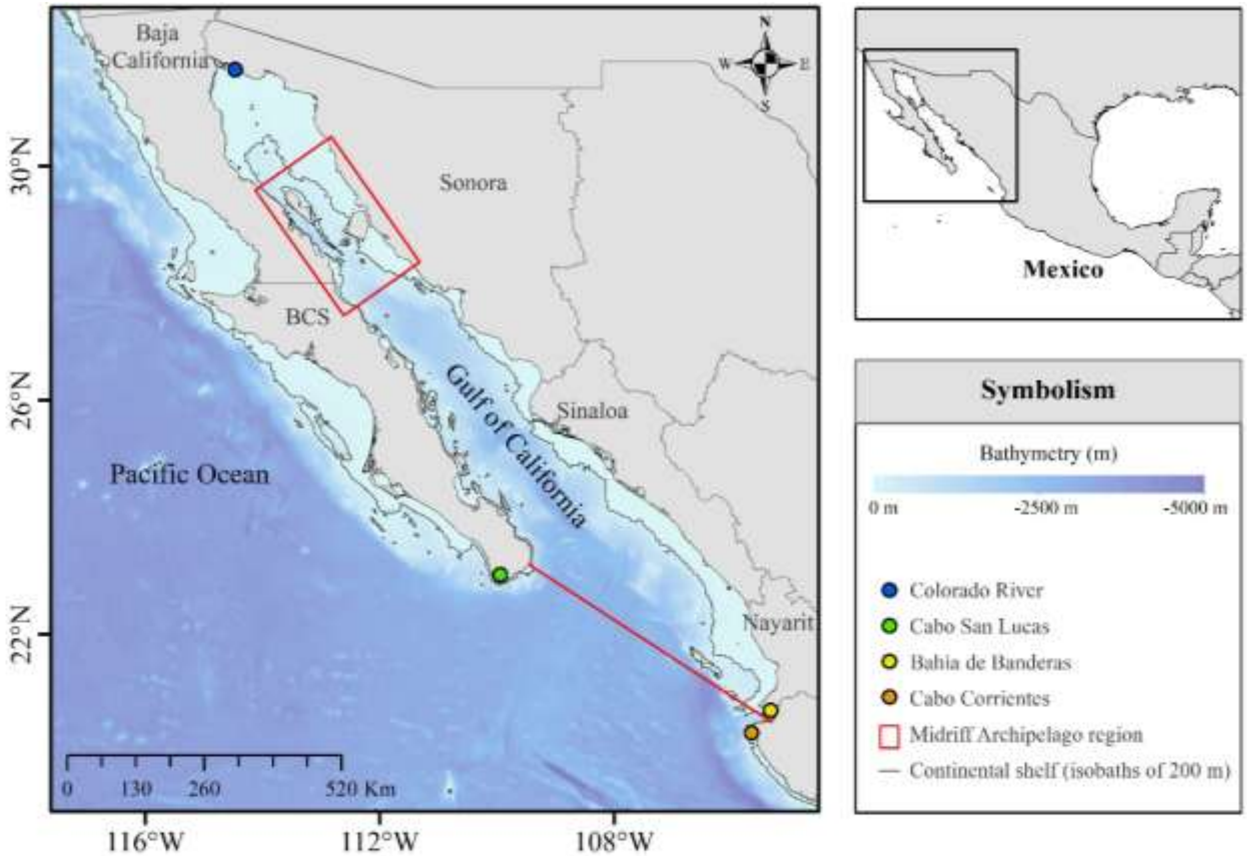

Figure 1. Bathymetric map of the Gulf of California, conceptually delimited from the Pacific Ocean by a red line. The gulf is a large and deep pull-apart basin formed by right-lateral strike-slip motion along faults associated with the San Andreas system. Tectonic opening on the order of 3–6 cm y^{-1} has shaped the oceanic crust underlying most of the basin. The resulting geological configuration supports oceanographic processes that sustain high phytoplankton productivity, with contrasting dynamics along the opposing coasts.

- Line 85: Consider adding the longitudes to Cabo San Lucas and Cabo Corrientes

Response: In response to your suggestion, we have attached the location of Cabo San Lucas and Cabo Corrientes on the map.

- Line 86: Consider adding a quick overall summary of the data collection methods above the first sub-section.

Response: We thank the reviewer for this suggestion. To improve the organisation of the Methods section, we have added a brief introductory paragraph summarizing the datasets used in the study before the description of each individual data source.

In the Methods section, 2.2 Data collection; pages 4–5 Lines 105–112: "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. NPP reanalysis products were used only for subsurface and water-column variables, including reanalysis estimates of mixed-layer depth, nutrients, salinity, 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."

- Line 94: "Vertially" should be corrected to "Vertically"

Response: We thank the reviewer for this observation. We modified this typo as Vertically.

- Line 94: If applicable, this website would be more informative than the current ERDDAP listed one. https://orca.science.oregonstate.edu/1080.by.2160.monthly.hdf.vgpm.m.chl.m.sst.php

Response: Thanks for this suggestion. We would like to clarify that the ERDDAP server was used as the data access and extraction platform for the VGPM-based NPP product employed in this study, ensuring consistent retrieval and processing. Therefore, we have retained the ERDDAP reference in the manuscript, as it reflects the actual data access pathway used in our analysis.

- Line 107: Broken link. Please update. Is it possible you meant this one? https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_PHY_001_030/description

Response: We thank the reviewer for identifying this issue. The broken link has been corrected and updated accordingly.

- Line 108: Consider changing the subsection title to “Data Regridding” or similar. Homogenization is not wrong, but it is a bit vague.

Response: We thank the reviewer for this suggestion. The subsection title has been revised to “Data Regridding” for improved clarity and specificity.

- Line 115: “(Hobday et al. 2016)” should be Hobday et al. (2016)

Response: We made this editorial change as suggested by the reviewer.

- Line 117: “Bimonthly” can be ambiguous and mean either twice a month or every two months. Please consider clarifying or using a different term.

Response: We thank the reviewer for pointing out this ambiguity. The term “bimonthly” has been replaced with a more precise expression throughout the manuscript to avoid confusion. Specifically, we now use “overlapping two-month periods” to clearly indicate the intended sampling frequency.

In the text page 6, Lines 149 to 151: “The Multivariate ENSO Index version 2 (MEI.v2) was used to represent ENSO variability during the study period. MEI.v2 is an index for overlapping two-month periods (Dec-Jan, Jan-Feb, etc.) based on the leading combined empirical orthogonal function of five ocean–atmosphere variables over the tropical Pacific…”

- Line 138: “(Moore and McCabe, 1999)” should be “Moore and McCabe (1999)”. Please also consider reporting exactly how outliers were defined (1.5*IQR?) and how many/what percent of data points this excluded.

Response: We thank the reviewer for this helpful comment. We add the information suggested for more clarification that outliers were identified and removed using the interquartile range criterion (1.5 × IQR).

In the text page 7, Lines 191 to 192: “Prior to clustering, outliers were removed following the interquartile range (IQR) criterion (1.5 × IQR) of Moore and McCabe (1999) …”

- Line 139: k-means^{++} is an initialization method for the k-means algorithm. Consider changing to “...grouped into three clusters using the k-means algorithm with k-means^{++} initialization.”

Response: We thank the reviewer for this suggestion. The text has been revised accordingly to improve clarity and explicitly state that the data were grouped into three clusters using the k-means^{++} algorithm with k-means^{++} initialisation.

In the text page 7, Lines 193 to 198: "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 (Fig. A1)."

- Line 141: "Spatial distribution and seasonal patterns (cold season: December–May; warm season: June–November) were visualized" would be helpful to point to a figure here.

Response: The reference to Figure 3, corresponding to said information, is attached.

- Line 151: URT and UAR seem to be undefined above.

Response: These variables are defined in the revised Methods section where they are first introduced.

- Line 154: Please provide thresholds used. For example, Pearsons correlation >0.7.

Response: Thanks for the observation, we have already added the correlation type and the cutoff threshold of 0.7.

In the text page 8, Lines 218 to 220: "Potential collinearity among predictors was evaluated with correlation matrices considering Pearsons's correlation > 0.7; highly correlated variables were excluded from the final predictor set…"

- Line 160: Is Zone a predictor? If so, that also becomes circular because NPP defines the zone. So the model would say "this pixel predicts high NPP because it is in the High NPP zone"

Response: We thank the reviewer for pointing out this potential source of confusion. Zone was not included as an explanatory predictor in the GLMM, and therefore no circularity exists between the response variable and the model predictors. Instead, the productivity zones (LPZ, MPZ, and HPZ) were used solely to stratify the dataset prior to model fitting. Separate Generalised Linear Mixed Models were then fitted independently for each productivity zone to identify the environmental drivers associated with NPP variability within each productivity regime. In each model, the environmental variables were included as fixed effects, whereas Month was included as a random intercept to account for temporal dependence among observations collected within the same month. Consequently, the models evaluate the relationships between NPP and the environmental predictors within each productivity regime, rather than using zone membership to explain NPP.

In the text page 8, Lines 222 to 229: "To identify the environmental drivers of net primary productivity (NPP) and evaluate whether their effects differed among productivity zones, we fitted a Generalised Linear Mixed Model (GLMM) using a Gamma error distribution and a log-link function, which is appropriate for continuous, positive, and right-skewed response variables such as NPP (Corrales and Cepeda-Cuervo, 2019; Miyamoto et al., 2019). Prior to modelling, all continuous predictor variables were standardised (*z*-scores) to facilitate comparison of effect sizes and improve model convergence. Separate models were fitted for each productivity zone; MLD, NO_3, PO_4^{-3}, URT, UAR, SI, and SSS were included as fixed effects within each model. To account for temporal dependence among observations and repeated measurements through time, month was included as a random intercept term (1|Month)."

- Line 176: The last sentence is a little ambiguous. Consider rewriting to something like "By dividing the Gulf into three productivity zones, we can evaluate how abnormal warming alters their physical size, their total carbon contribution, and the underlying physical forces that drive them."

Response: We agree that the original sentence was ambiguous and could be improved for clarity. Following the reviewer's recommendation, we have revised the sentence to more clearly articulate the objective of the zonal division and its link to anomalous warming.

In the text page 10, Lines 268 to 270: "By dividing the Gulf into three productivity zones, we evaluate how anomalous warming affects their spatial extent, their total carbon contribution, and the physical drivers underlying these changes".

- Line 179: This would be helpful to put a figure reference.

Response: We include the reference to the corresponding figure.

- Line 179: Does “climatology” refer to the average annual seasonal cycle max and min? Or just the max and min across this specific time range?

Response: In response to your observation, we changed the word to "mean seasonal cycle" for greater accuracy and clarity in the text.

- Line 180: I believe “longitudinal” should be “latitudinal”.

Response: You're right, the change has already been made.

- Line 180: Do these numbers represent the climatological mean? or do these also include the cold and warm seasons? could be clarified by including the figure number/panel associated with the paragraph.

Response: We thank the reviewer for this helpful comment. The text has been revised to clarify that the reported values correspond to mean seasonal cycle. In addition, we have explicitly indicated the corresponding figure and panel in the text to improve clarity and facilitate interpretation of the results.

- Line 182: Is the latitude value only for the archipelago? If so, why are the latitudes of the coasts previously listed left out?

Response: To avoid potential confusion and ensure consistency in the geographic description, the latitude values have been removed from this section of the manuscript, as they were not consistently provided for all regions described.

- Line 190: Is this the per-month average over 22 years? Why is this important? It might make more sense to compare high values to high-only climatology, etc. Also including a figure reference here would be helpful.

Response: We thank the reviewer for this insightful comment. The manuscript has been revised to clarify that the reported values correspond to a 22-year climatological monthly mean. We now explicitly state this in the text and clarify that seasonal values (cold and warm periods) are derived from this climatological framework. In addition, the corresponding figure reference has been added to improve clarity and facilitate interpretation of the results.

In the text page 11, Lines 289 to 293: “Gulf-wide climatological monthly mean NPP over the 22-years period was 8.6 Tg C per month. During the cold season, NPP peaked in March at 14.9 Tg C per month, whereas during the warm season it reached a minimum in August (4.4 Tg C per month) (Fig. 4a)”.

- Line 195: at first read “zone-integrated” is a little confusing but upon subsequent reads it is clear what you are writing here. Consider rewriting to “NPP per year per productivity zone” or similar and refer to Figure 4, panel b.

Response: In attended your observation, the term “zone-integrated” has been revised to improve clarity and is now expressed as “NPP per year per productivity zone” throughout the manuscript.

- Line 205: In Section 3.4, please ensure that spatial percentages (e.g., '% of gulf area') are clearly delineated from the productivity contribution percentages (e.g., '% of total NPP') used in Section 3.3 to avoid reader confusion.

Response: In attended your observation, we now explicitly state that all percentages refer to the spatial extent of each productivity zone. We also clarify that these differ from the relative contribution to total NPP described in Section 3.3, to avoid potential confusion for the reader.

- Line 206: I believe it should be Fig. 5, not 4.

Response: Thank you for your observation, the corresponding change has been made.

- Line 207: Much of the paper discusses latitude, it would be helpful to include latitude on at least one of the maps in Figure 5.

Response: Thank you for your observation, we add the latitude in one map in Figure 5.

- Line 307: This is one of the situations where “winter” and “cold season” get a little unclear. Intuitively, 6 months in a near-tropical region is hard to classify as winter. Additionally, in 3.3 "winter HPZ was approx. 54%” and the "cold period" in 3.4 gave PHZ 44.2%. Please explicitly define winter versus cold period and remain consistent.

Response: We thank the reviewer for this comment. We agree that the use of "winter" and "cold season" may lead to confusion. To ensure clarity and consistency, we have revised the manuscript to use the term cold period (December–May) and warm season (June–November).

- Line 311: Consider starting a new paragraph at "During winter." and combining with the sentence below starting "Conversely..."

Response: We thank the reviewer for this helpful suggestion. The paragraph has been reorganised to improve clarity and flow. Specifically, we have separated the description of cold season conditions from the warm season dynamics into distinct paragraphs. This restructuring improves readability and better highlights the seasonal contrast in NPP and productivity zone distribution.

In the text page 18-19, Lines 410 to 416: "…During this period, HPZ and MPZ together typically cover ~77 % of the Gulf of California. Hakspiel-Segura et al. (2022) reported a negative correlation of SST and meridional wind stress with surface Chl-a in the central gulf, reinforcing the role of wind-driven mixing.

Conversely, during the warm season (June–November), oligotrophic conditions dominate: NPP < 900 mg C m^{-2} d^{-1} covers ~68 % of the basin due to weaker southeasterly winds, the intrusion of warm tropical waters, and strong stratification that limits nutrient supply (Lavín and Marinone, 2003; Escalante et al., 2013; Mercado-Santana et al., 2017; Trasviña-Castro et al., 2024)."

- Line 326: "this physical mechanism" is a bit ambiguous.

Response: We thank the reviewer for this comment. The ambiguous phrase "this physical mechanism" has been removed from the manuscript to improve clarity and avoid potential confusion.

In the text page 19, Lines 427 to 429: "Stratification index (SI) analysis indicated varying productivity regimes (Fig. A9): HPZ had low SI (0–0.5 kg m–3) due to mixing processes (upwelling, tidal forcing, and north-westerly winds), while LPZ exhibited high SI (up to ~2.0 kg m–3), indicating strongly stratified waters."

- Line 337: "tipping point" implies a point of no return. Consider changing to regime change or similar.

Response: We thank the reviewer for pointing out this ambiguity, we eliminate this affirmation and we add a description more details about this change of the trend of the climate in the region.

In the text page 19, Lines 435 to 438: “The 2014–2016 period marked a significant decline in Net Primary Production (NPP) in the Gulf of California, characterised by severe negative anomalies. This period was associated with strong atmospheric forcing related to ENSO (MEI.v2) and coincided with a positive phase of the PDO, a transition toward a negative NPGO, and widespread marine heatwaves.”

- Line 354: The listed values are wind speeds, not wind stress.

Response: The variable has been corrected from “wind stress” to “wind speed” to accurately reflect the units and data presented in this section of the manuscript.

In the text page 19, Lines 445 to 446: “A key mechanism affecting the marine ecosystem between 2014 and 2016 was the weakening of north-westerly winds (~8 $m\ s^{-1}$ versus climatological values >20 $m\ s^{-1}$), which reduced wind speed and coastal upwelling efficiency.”

- Line 367: Does “Long-term trends” refer to your data set? Or others?

Response: We have clarified that the “long-term trends” refer to results derived from our dataset, as shown in Fig. A10. The text has been revised to explicitly distinguish between our findings and previously reported trends, improving clarity and avoiding potential ambiguity.

In the text page 20, Lines 454 to 455: “Long-term trends derived from this study indicate declining NPP of 50–100 mg C m–2 d–1 per year in the upper gulf, in La Paz Bay, and in front of Nayarit coasts (Fig. A10), while values at the mouth and near ~28° N remained relatively stable.”

- Line 367: Are the 50-100 mg C $m^{-2}\ d^{-1}$ also per decade?

Response: The reported trend is per year, not per decade; the unit has been clarified in the text and in Fig. A10.

- Line 371: (Maishal, 2024a) should not be in parenthesis.

Response: Now this was corrected as reviewer suggested, now says “These patterns are consistent with global trends reported by Maishal (2024a,b),