Response to Referee #2

We thank Referee #2 for an unusually thorough and technically acute review. The comments helped us sharpen a single overarching distinction — between effects directly demonstrated by the experiments and mechanisms inferred from them — that now runs through the entire manuscript.

In response, we have

(1) clarified the dependence and error treatment of the OLCI vs. multi-sensor products;

(2) separated consistency-with-assimilated-data from independent validation and added the robust MedAE metric;

(3) added a new multi-variable, multi-season station diagnostic analysis (Fig. S2, Table S1) to ground our physical-mediation interpretation in results; and

(4) moderated interpretations of MLD, pycnocline, nutrient entrainment, heat redistribution, and deep-water oxygen accordingly.

(5) clarified the weakly coupled assimilation framework and the role of bio-optical feedbacks;

(6) corrected figure, caption, terminology, and typographical inconsistencies

(7)revised the discussion, Table 2, and conclusions so that they remain consistent with the analyses presented.

All changes are marked; line numbers refer to the revised version.

General Comment 1 — 1. Use of OLCI data
The study assimilates both the multi-sensor and OLCI-only products, despite OLCI data already being included in the multi-sensor products, essentially meaning the OLCI data are double-counted. Was this accounted for in the assimilation methodology?
The manuscript justifies using both due to the different resolutions of the products:
“The combined assimilation of these datasets enhances spatial coverage and representation of variability across scales relative to a single-product approach. Differences in sensor characteristics and resolution result in complementary sampling from basin to coastal scales. Specifically, the 1 km dataset constrains basin-scale variability, while the 300 m dataset captures finer-scale features, particularly in coastal and heterogeneous waters. Their joint use enables consistent representation of large-scale patterns while improving the characterization of small-scale phytoplankton dynamics.”
This seems reasonable on the face of it, except that the model being assimilated into is 3.7 km resolution, so cannot resolve processes even on the scale of 1 km, let alone 300 m. Accordingly, “When multiple observations are available within the same model grid cell and depth level, they are averaged to provide a representative value.” This seems to negate the argument made above?
The very different behaviour seen in the OLCI data compared with the multi-sensor data in Fig. 5 is intriguing and worrying. Further assessment of the differing product characteristics, or at least references to any studies which have looked at that, would be very useful. I am not an expert in Baltic Sea phenology, but the OLCI data does not show the kind of seasonal cycle I would expect. The discussion that:
“None of the DA experiments are capable of replicating this early-season peak in Sentinel-3 product. This persistent mismatch originates from a hard-coded structural thresholding within the biogeochemical formulation rather than a failure of the assimilation scheme itself (Eilola et al, 2009; Ruvalcaba Baroni et al., 2024). Specifically, overly restrictive thresholds for minimum light availability and background phytoplankton biomass artificially suppress primary production under cold, low-light, early-season conditions. Therefore, this limits the model’s intrinsic ability to initiate the late-winter to early-spring bloom, regardless of the strength or configuration of the observational constraints applied.”
is potentially very useful and interesting to the community, but needs further interrogation. Firstly, it assumes the OLCI data are correct (implying the multi-sensor data are in error) which needs to be demonstrated. Secondly, to conclude that the thresholds are “overly restrictive” it would be useful to know what they are and how they compare to other models and lab studies. Suppression of primary production in low-light conditions is deliberately built into models based on empirical understanding, and so use of the word “artificial” could be disputed. Perhaps SCOBI does use more restrictive values than other models or observational studies would suggest, in which case this has revealed a model tuning issue. But if the understanding on which the model (and probably others) is built is flawed, then this is a valuable result for the wider community that merits greater discussion.

Response: We thank the referee for this incisive point and agree on both counts.

On double-counting: The referee is correct that our resolution-based justification was untenable at ~3.7 km, and we have removed it. We now state plainly that the two products are "not statistically independent" (OLCI feeds the multi-sensor product) and clarify the actual procedure: before assimilation, valid observations from both products are "pooled and averaged into super-observations on the model grid", so coincident raw observations are never assimilated as independent constraints. We also now acknowledge, as an explicit limitation, that cross-product observation-error covariance is not represented in the current system.

The revised text (lines 179-185) in Sect. 2.2.1 reads:

"Valid satellite Chl-a observations from the multi-sensor and OLCI products were pooled and averaged within each model grid cell and analysis window to form superobservations. All valid observations were equally weighted, so the relative contribution of each product depended on the number of available observations. This aggregation provides a Chl-a constraint at the ~3.7 km model-grid scale rather than at the native 300 m or 1 km product resolution. The observation-error specification is described in Sect. 3.4. Because OLCI observations also contribute to the multi-sensor product, the two products are not fully independent, and their shared information and correlated errors are not explicitly represented in the assimilation system."

We have consolidated the description of observation aggregation. The construction of satellite Chl-a super-observations is described in Sect. 2.2.1, while Sect. 3.4 now refers to this procedure and specifies the averaging rule for other observation types. This avoids duplication and clarifies that the multi-sensor and OLCI Chl-a products are pooled before assimilation.

Section 3.4 adds a corresponding statement (lines 323–325).

"The observation-error specification applied to the satellite superobservations does not explicitly include cross-product observation-error covariance between the multi-sensor and OLCI inputs."

We have removed the previous wording suggesting that the products provide statistically independent information or that their native 300-m and 1-km resolutions are explicitly represented by the approximately 3.7-km model. Their joint treatment is intended to form a spatially aggregated satellite Chl-a constraint, rather than to resolve the original sensor-specific spatial scales.

On the OLCI seasonal cycle: We agree our original "overly restrictive / artificial threshold" explanation was too strong, because it presupposed that OLCI is correct and the multi-sensor product is error — which our experiments cannot establish. We have removed the statements that the mismatch "originates from" hard-coded thresholds and that these are "overly restrictive" and "artificially suppress" primary production.

To provide additional observational context, we added a comparison against station-sampled CMEMS in situ surface Chl-a (new Fig. S1), analysed for both 0–2 m and 0–5 m depth ranges (similar patterns). The in situ record shows a pronounced spring maximum peaking in April; OLCI shows a spring maximum of similar timing but lower magnitude; the multi-sensor product and the model are lower still. The early-season signal is therefore not unique to OLCI, but its magnitude is strongly product-dependent. Because the in situ data are station-sampled while the satellite/model fields are spatially aggregated, we treat this comparison as context, not as validation identifying a "true" product.

The revised text now reads:

"None of the DA experiments are capable of replicating this early-season peak in Sentinel-3 product. However, the present experiments do not isolate the threshold-based treatment of low-light conditions in the biogeochemical formulation as the unique cause of this mismatch. The discrepancy may reflect a combination of limitations in the representation of low-light seasonal dynamics, satellite retrieval uncertainty, and differences in the spatial and temporal sampling of the available observations."

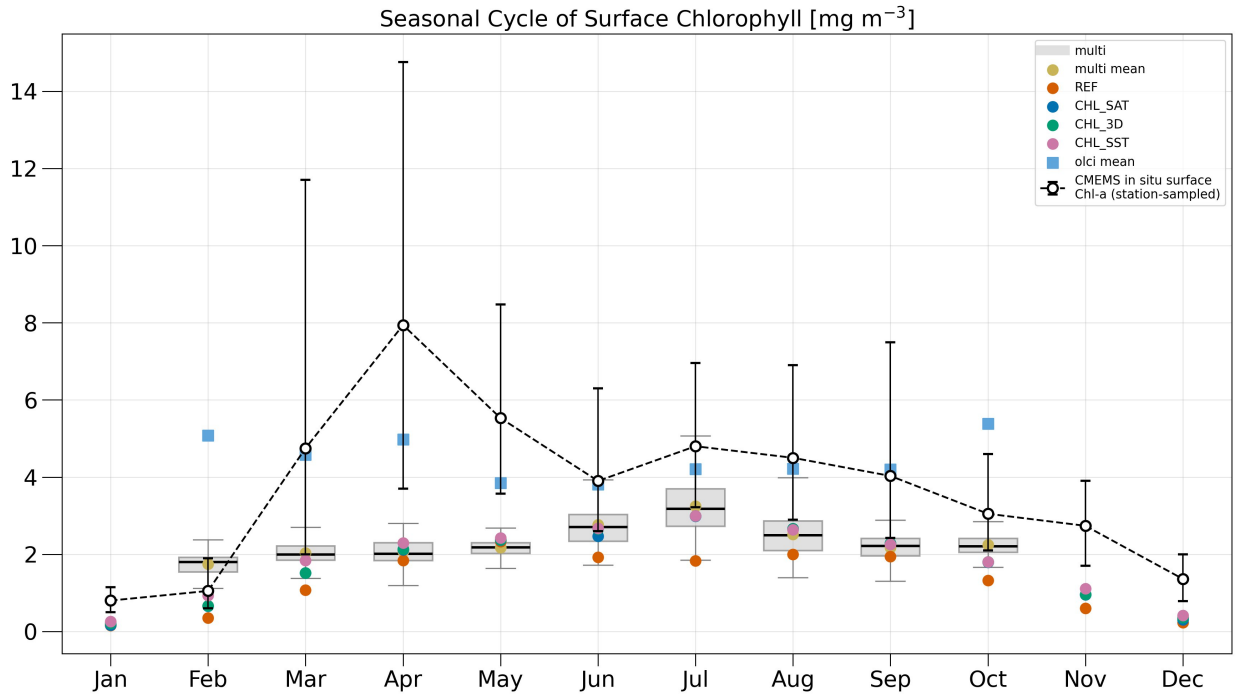


Figure S1. Seasonal cycle of surface chlorophyll-a (Chl-a) in the model experiments and observational products. Grey boxplots show the distribution of daily spatial means from the multi-sensor satellite product for each month, and gold circles indicate the corresponding monthly means. Blue squares show monthly OLCI means, while coloured symbols show monthly climatologies from the model experiments. Black open circles and vertical bars indicate the monthly median and 25th–75th percentile range, respectively, of station-sampled CMEMS in situ Chl-a observations from 0–2 m depth. Results based on the 0–5 m depth range are similar. The satellite and model fields are spatially aggregated, whereas the in situ observations are station sampled; the latter therefore provide observational context rather than a spatially equivalent basin-wide reference.

General Comment 2 — Validation

It's not clearly stated in the manuscript, but it looks like all the SST and chlorophyll data used for validation is the same data that's used for assimilation, meaning it's not independent. When data is sparse this can be unavoidable, but it does mean the validation results come with caveats. The purpose of a data assimilation scheme is to bring the model closer to the observations, so comparing to the assimilated data essentially just demonstrates that the assimilation scheme is coded correctly. That doesn't mean the validation presented isn't meaningful, but in the absence of independent data sets the validation is not as robust as it could be and that caveat should be clearly stated in the manuscript.

The oxygen data used is independent which is great. I'll discuss this more in my next comment, but I would like to see further analysis of unassimilated variables, whether compared to observations or not. Where this is the case, as with the oxygen data, please introduce these observations in Section 2 alongside the SST and chlorophyll data, making clear which data sets are used for assimilation, which for validation, and which for both.

Finally on this point, validation is performed using metrics such as bias and RMSE. These work best for variables with a Gaussian distribution, whereas chlorophyll is highly non-Gaussian. Robust equivalents such as median absolute error may be more appropriate.

Response:

We agree on all three points and have implemented them.

Independence caveat: We now state explicitly that comparisons against assimilated Chl-a and SST observations quantify consistency with the observational constraint rather than independent predictive skill, and we have revised the interpretation of the RMSE reductions accordingly. We also clarified the independence of the different observational comparisons. We also clarified the independence of the different observational comparisons in Sect. 3.6 (lines 361–367):

"… Chl-a and SST observations used for assimilation are not considered independent validation data for the corresponding experiments. In situ Chl-a profiles therefore provide an independent observational reference for REF and CHL_SAT, but comparisons in CHL_3D and CHL_SST assess consistency with the assimilated profiles. Oxygen observations remain independent for all experiments because oxygen is not assimilated. SCM depth is treated as a diagnostic derived from Chl-a profiles, and its independence therefore depends on whether the underlying profiles are assimilated. Similarly, comparisons with satellite products used in assimilation are interpreted as consistency with the assimilated observations rather than as independent validation."

Consistent with this clarification, the surface Chl-a result now reads (lines 408–411):

"…The corresponding reductions in RMSE in the DA experiments quantify the degree to which assimilation brings the modeled surface Chl-a toward the assimilated observational constraint; they should not be interpreted as independent predictive validation." The basin-wide RMSE values are 2.30, 1.62, 1.58, and 1.41 mg m^{-3} for REF, CHL_SAT, CHL_3D, and CHL_SST (reductions of ~30%, 31%, 39% relative to REF).

Dataset roles in Section 2: A new Table 1 lists, for each variable and experiment, the assimilation dataset, the validation dataset, and whether the validation is independent, directly addressing the request to consolidate dataset usage in the Methods. The source of the oxygen observations is also described in Sect. 2.2.3.

Table 1. Observational datasets used for assimilation and evaluation, and their experiment-specific assimilation status.

Observations	Assimilated in	Used for Validation	Independent validation
OSISAF SST	CHL_SST	Yes	CHL_SAT and CHL_3D: Yes CHL_SST: No
Multi-sensor and OLCI Satellite Chl-a	CHL_SAT, CHL_3D, and CHL_SST	Yes	No
In-Situ Chl-a	CHL_3D and CHL_SST	Yes	CHL_SAT:Yes CHL_3D and CHL_SST: No
Oxygen	No	Yes	Yes

Robust metric: We added the median absolute error (MedAE) alongside RMSE. The result reinforces our conclusion: MedAE values are 1.35, 0.65, 0.63, and 0.53 mg m^{-3} for REF, CHL_SAT, CHL_3D, and CHL_SST — the same ranking as RMSE — indicating the improvement is not driven by a few high-Chl-a outliers (lines 405–408). We agree a formal hold-out validation would be the gold standard; we now flag its absence as an explicit limitation but have not added one, as it would require a separate experimental design beyond this study's scope.

General Comment 3 — Non-assimilated variables and vertical structure

Further assessment of non-assimilated variables, particularly nutrients and 3D temperature and salinity and density, would bring interesting insight into the impact of the assimilation on the wider model state, and help support some of the statements and hypotheses made. Ideally this would include validation against observations, but simply presenting model results would be sufficient.

In particular, presentation of the vertical density structure. Much is discussed about the importance of the pycnocline depth for instance, and the impact of the assimilation on the physical environment that governs nutrient entrainment

and phytoplankton growth. But the only physics variable presented is SST, and nutrients are not presented. This means much of the discussion feels speculative, rather than grounded in results.

This feels especially important given that it's repeatedly pointed out that the Baltic is primarily salinity-stratified, which presumably limits the impact that could be achieved by just changing the thermal structure through SST assimilation and feedbacks from chlorophyll.

Response: We agree this is the single most important way to strengthen the paper, and we have added the requested evidence rather than merely softening the text. We have added two complementary sets of diagnostics. Table S1 summarizes differences in MLD, density stratification, DIN, DIP, and SCM depth at four stations across all four seasons. Figure S2 presents spring and autumn vertical profiles at BY5, BY15, and BY29. Supplementary Sect. S2 defines the diagnostics, including the model-output MLD criterion and the common SCM-depth definition used in Fig. 8 and Table S1. These show that the additional SST constraint in CHL_SST is associated with station- and season-dependent differences in the hydrographic background, nutrient distributions, and SCM depth relative to CHL_3D — for example, the largest ΔMLD (-3.35 m at BY29 in spring) and the largest ΔSCM depth (+4.2 m at BY29 in winter).

Because these diagnostics compare two assimilation configurations rather than evaluating an observed pycnocline, we present them as evidence for possible physical mediation of the biogeochemical correction, not as a demonstrated correction. Accordingly, the original statement that SST assimilation "corrects the seasonal pycnocline and mixed-layer depth" is replaced by (lines 639-646):

"The CHL_SST result further highlights the contrasting response to the additional physical constraint. Relative to CHL_3D, SST assimilation produces station- and season-dependent differences in temperature, salinity, density structure, MLD, nutrient distributions, and SCM depth (Fig.S2; Table S1). Seasonal mean MLD differences are generally small but reach approximately 3.3 m at BY29 station in spring, while the largest SCM-depth differences occur during autumn and winter (Table S1). The corresponding changes in DIN and DIP distributions are consistent with physical mediation of the vertical biogeochemical response. Thus, SST assimilation provides a broader physical context for the biogeochemical response, whereas Chl-a assimilation primarily constrains phytoplankton distributions and induces relatively small thermal adjustments. "

We did not add a basin-wide heat-content or deep-ventilation analysis, as those processes are not constrained by the surface-focused assimilation and would broaden the scope substantially; we state this boundary explicitly.

Specific comments

Line 53 — "Assimilation of biogeochemical variables, including nutrients, dissolved oxygen, and occasionally Chl-a, has recently emerged". Chl-a from ocean colour is by far the most commonly assimilated biogeochemical variable, so I'm puzzled by "occasionally". "Recently" is a relative term, but the first chl-a assimilation study I'm aware of was Ishizaka (1990) (https://doi.org/10.1029/JC095iC11p20183) so 36 years ago. Although it has only become established operationally much more recently.

Response: We agree that the original wording understated the established use of Chl-a assimilation. The sentence now reads:

"Assimilation of biogeochemical variables, including nutrients, dissolved oxygen, and Chl-a, has increasingly been applied in regional and coastal ocean appliations(Liu et al, 2014; Ciavatta et al, 2018; Teruzzi et al., 2021). However, the relative availability and use of these observations differ among regions, and the effectiveness of jointly constraining physical and biogeochemical states remains less systematically assessed in the Baltic Sea (Goodliff et al., 2019; Skákala et al., 2021). "

Section 2.1 — Section 2.1: Given its importance to the study, please add details of the two-way coupling and light penetration scheme.

Response: The coupling pathway has been clarified as follows:

"NEMO and SCOBI are coupled through the exchange of physical and biogeochemical variables. In the present configuration, phytoplankton and other optically active constituents influence the attenuation of shortwave radiation

through their contribution to the water-column light field. Changes in the simulated Chl-a concentration therefore modify the vertical distribution of available light and can subsequently alter the temperature tendency through the physical model. Conversely, temperature, mixing, and other physical variables influence the evolution and vertical distribution of the biogeochemical state through the coupled model dynamics."

Lines 99 — "A coupled ocean–biogeochemical modeling system was configured for the Baltic–North Sea" – if that was specifically done in this study, more details of the model need to be presented. If the modelling system was pre-existing, please provide references where more information can be found.

Response: We have clarified that the present configuration is based on the existing NEMO–SCOBI Baltic Sea modelling framework and adapted for the present assimilation experiments.

The revised text is(lines 99-104):

"A coupled ocean–biogeochemical modeling system was configured for the Baltic–North Sea based on the NEMO (Nucleus for European Modelling of the Ocean) framework , with a horizontal resolution of approximately 3.7 km (2 nautical miles) and 56 vertical levels with variable thickness (Ruvalcaba Baroni et al., 2024). The model domain is the same as that used by Ruvalcaba Baroni et al. (2024) and covers the the Baltic Sea and the wider North Sea. In the present study, however, data assimilation and subsequent analysis are restricted to the subdomain shown in Fig. 1, which encompasses the Baltic Sea and the adjacent part of the North Sea. "

Ruvalcaba Baroni, I., Almroth-Rosell, E., Axell, L., Fredriksson, S. T., Hieronymus, J., Hieronymus, M., Brunnabend, S.-E., Gröger, M., Kuznetsov, I., Fransner, F., Hordoir, R., Falahat, S., and Arneborg, L.: Validation of the coupled physical–biogeochemical ocean model NEMO–SCOBI for the North Sea–Baltic Sea system, Biogeosciences, 21, 2087–2132, https://doi.org/10.5194/bg-21-2087-2024, 2024.

L99: "Baltic–North Sea" – presumably this means observations were only assimilated in part of the model domain (the Baltic)? In the next section on use of observations, this would be worth clarifying.

Fig.1 does not seem to be referenced in the main text.

Response: we clarified that "Baltic–North Sea" is the full model domain, while assimilation and analysis are restricted to the subdomain in Fig. 1, and an explicit Fig. 1 reference has been added (lines 102–104):

"…In the present study, however, data assimilation and subsequent analysis are restricted to the subdomain shown in Fig. 1, which encompasses the Baltic Sea and the adjacent part of the North Sea."

L133: "statisticaled" – I'm not sure what is meant here.

Response: It has been revised to "The Baltic sub-basins and correponding relative errors between satellite-derived Chl-a and the in situ Chl-a observations are shown in (b) for ..."

L147: "all data were filtered using quality flags" – please provide more details.

Response: We agree that the description of the observational data processing warrants a more detailed explanation. The revised text now reads:

"The satellite Chl-a products are Level-3 products that have undergone quality control within the CMEMS processing chain. Only observations flagged as valid in the source products were used." As an additional screening step, "satellite observations of both SST and Chl-a within 5 km of the coastline are removed to reduce contamination from land adjacency effects." We have revised Sects. 2.2.1 and 3.4 to distinguish the product-level quality control from the additional coastal screening applied in this study.

L159-161: "(b) Basin-wide spatial sampling density mapping derived via smooth Gaussian kernel density estimation. The color scale highlights the geometric footprint of high-frequency undulating cross-basin transects overlaid on traditional fixed monitoring stations." – I think I understand the general message being conveyed, but I don't understand the details of the method or the plot; please rephrase/expand the explanation to be clearer.

Response: The caption now clarifies the method and interpretation :

"Basin-wide spatial sampling density of in-situ observations estimated using Gaussian kernel density estimation. The color scale represents the relative spatial density of observations, with higher values indicating areas where observations are more frequently concentrated. The resulting density pattern illustrates the sampling footprint of the high-frequency undulating cross-basin transects together with the fixed monitoring stations."

L174: "Propoer" – Proper

Response: Corrected to "Proper".

L197-8: "The LSEIK filter is implemented in an offline configuration, where ensemble statistics are derived from a long-term free running model simulation rather than from a dynamically evolving ensemble." Please signpost here to the later section where more details are given.

Response: We agree that a signpost should be provided here to guide the methodological details regarding the LSEIK configuration. We have added the methodological signposting so that the description reads:

"The LSEIK filter implementation and construction of the background ensemble are described in Sect. 3.3. The ensemble perturbations of the LSEIK filter are constructed from a long-term free-running simulation rather than from a dynamically evolving ensemble."

L201-2: "physical and biogeochemical variables are included in a common state vector" – please specify which variables are included. Just temperature and chlorophyll? Is chlorophyll a single model state variable?

Response: We agree that the state variables in the data assimilation configuration should be given. We have clarified:

"...physical variables (temperature and salinity) and biogeochemical variables (phytoplankton groups 1-3, oxygen, nitrate, ammonium, and phosphate) are combined into a single state vector. Chl-a is a diagnostic variable calculated as the sum of the three phytoplankton groups and is linked to observations through the observation operator. Only temperature and total phytoplankton biomass receive direct analysis increments. The total phytoplankton increment is distributed among groups 1-3 in proportion to their pre-analysis contributions to total biomass. The temperature and total phytoplankton analyses are performed independently, without explicit physical–biogeochemical cross-variable error covariances."

L229: "20 samples per year" – please be clearer what is meant by this.

Response: Thanks, we have revised this to:

"This results in a total ensemble size of 120 members, obtained by selecting 20 model snapshots for each of the six years within the temporal sampling window."

L245-6: "this window provides sufficient integration time for the model's radiative transfer scheme to respond to the updated chlorophyll fields." – I don't understand why this needs to be accounted for in the choice of assimilation window. Please expand on your rationale.

Response: We have clarified the rationale and no longer imply that the radiative scheme requires a specific integration time within the analysis step itself (lines 270–274):

"This window provides a compromise between resolving the rapid evolution of phytoplankton blooms and allowing sufficient time for the coupled model to respond to the assimilated Chl-a field within the assimilation cycle. In particular, changes in Chl-a can modify light attenuation and thus the vertical distribution of shortwave radiation, affecting the subsequent evolution of the upper-ocean physical and biogeochemical state."

L246-8: “chlorophyll concentration updates modify the light attenuation coefficient, which alters the vertical distribution of shortwave radiation and consequrently contribute to changes in the temperature tendency (Ciavatta et al., 2018, Skákala et al., 2021)” – I don’t think either of these studies address this, better references could be found.

Response:We agree that the previous references did not directly support every step of the stated causal chain.

We have revised the text with updated references:

“Specifically, updates to the Chl-a-related biogeochemical state modify the light attenuation coefficient, which alters the vertical distribution of shortwave radiation and consequentlycontribute to changes in the temperature tendency (Skákala et al., 2022; Cahill et al., 2023). ”

L247: “consequrently” – consequently

Response: corrected.

L249: “Kang e al,” – Kang et al.

Response: The reference was removed and the text revised (lines 276–277):

"In the present weakly coupled configuration, the physical response develops during the forecast integration between analysis steps."

L251-4: “This weakly coupled assimilation framework circumvents…” – please list some references which discuss these issues.

Response: We agree that it is necessary to provide reference here to support the statements made. We revised the text and added a reference:

This weakly coupled assimilation framework avoids direct cross-variable analysis increments based on physical–biogeochemical background-error covariances. This limits the risk that poorly constrained or non-stationary cross-variable covariances generate spurious correlations, unrealistic analysis increments, or initialization shocks (e.g., Yu et al., 2018).

L257-8: “the faster temporal variability of physical processes compared to biogeochemical dynamics” – surely this depends on which processes are being considered, I’m not convinced it’s true as a general statement. Plus, it seems to contradict L242-3: “The window is sufficiently short to capture the rapid onset and peak of phytoplankton blooms”.

Response: We agree that the original statement was overly general. We have revised the wording to clarify that the different update intervals are a pragmatic design choice for the offline background ensemble and computational efficiency, rather than a general assertion that biogeochemical processes are always slower than physical processes. The revised text reads:

“To reduce computational cost, ensemble perturbations are updated every 6 days for physical variables and every 30 days for biogeochemical variables. The different intervals were selected pragmatically to balance the temporal variability represented in the offline background ensemble with computational efficiency.”

L261-2: “excluded, along with duplicate records” – is this just for in situ data, or does it cover OLCI and multi-sensor chlorophyll?

Response: Thanks for point this out, the excluded processing is for in situ data here. The satellite observations are aggregated into super-observations on the model grid by averaging the available observations within each grid cell. We have clarified it by the revised text:

"In-situ Chl-a observations located over land, as defined by the model land–sea mask, were excluded, together with duplicate records within the corresponding observational datasets. Satellite observations of both SST and Chl-a within 5 km of the coastline were additionally removed to reduce land-adjacency contamination."

The processing of OLCI and multi-sensor chlorophyll observations are given in section 2.2.1.

L272: "corresponding to an approximate relative uncertainty of 10%, reflecting a stable measurement uncertainty." – this seems low, especially for fluorescence data (it's not stated if the in situ measurements are fluorescence, HPLC, or a mixture, this would be worth clarifying in the previous section). And representation error seems to be neglected completely. Please provide justification for this choice, it seems to me like it would lead to overfitting the observations in the assimilation (one reason why it's important to validate against independent data, but that's an aside).

Response: We thank the referee for this comment. We have clarified that CPHL identifies the reported Chl-a concentration variable, not the measurement technique. The prescribed 10% relative error is a pragmatic weighting assumption rather than a validated estimate of total observation uncertainty. Instrument- or method-dependent uncertainty and point-to-grid representativeness error are not explicitly included, which may give excessive weight to some profiles. We now state this consequence explicitly and do not interpret agreement with assimilated profiles as evidence that the error specification is optimal. Sensitivity to the prescribed in situ observation error has not been quantified in the present experiments.

L276: "εrel is a sub-basin–specific relative error coefficient" – please list these somewhere, either in a table or plotted on a map. Either in the main text or an appendix.

Response: We clarified statement as follows: "The sub-basin–specific relative error coefficient ($\varepsilon_{\mathrm{rel}}$), estimated from statistical comparisons between collocated satellite and in situ Chl-a observations, is shown spatially in Fig. 1b"

L290: "three numerical experiments" – four including REF.

Response: Corrected (lines 327–328): "three assimilation experiments and a reference run without DA were conducted using NEMO–SCOBI to assess the impact of DA on ecosystem dynamics."

L323-4: "This indicates that, in this weakly coupled framework, biological updates do not propagate directly into surface temperature fields." – is that not by design/definition? In which case, "indicates that" is not the best phrasing.

Response: We agree. The absence of a direct biological-to-physical analysis increment is a feature of the weakly coupled framework, rather than a result diagnosed from the experiments. We have therefore removed "This indicates that" and revised the text to distinguish the assimilation design from the simulated response (lines 377-380):

"Because the weakly coupled analysis does not apply direct biological-to-physical cross-variable increments, the SST response to Chl-a assimilation develops through the subsequent coupled-model integration. In the present experiments, this response is small in the basin mean and spatially heterogeneous relative to REF."

L327-8: "provides a physical consistent baseline for interpreting the biogeochemical state." – this relates to my third general comment. You conclude that CHL_SST provides a sound physical baseline for the biogeochemistry, but have only looked at SST, when parameters such as mixed layer depth and mixing have a more direct impact on biogeochemistry.

Response: We agree that our original wording extended beyond what the SST results demonstrate. The revised sentence is:

"The direct SST constraint reduces the regional SST biases present in REF and provides a better-constrained surface thermal state for interpreting the accompanying biogeochemical response."

L334: "The domain-averaged value for each experiment is indicated in the upper-left corner of each panel." – lower-right, not upper-left. Furthermore, for b-d this looks like the average of the difference in TSS rather than of the TSS

itself, which in my view would be more informative. Also, please express it as e.g. 0.977 rather than in scientific notation.

Response: We thank the referee for noting the incorrect annotation location and the ambiguity in the displayed quantities. We corrected "upper-left" to "lower-right." Panel (a) shows absolute TSS for REF, whereas panels (b–d) show the fractional relative TSS change with respect to REF, calculated as $100(TSS_{DA}-TSS_{REF})/TSS_{REF}$. The lower-right annotations in panels (b–d) are therefore domain-mean relative changes, not absolute TSS values.

We have revised the caption and annotations to make this distinction explicit. The relative changes are now reported in fixed decimal percentage notation rather than scientific notation. The absolute REF TSS is retained in panel (a), which provides the reference level for interpreting the relative-change maps.

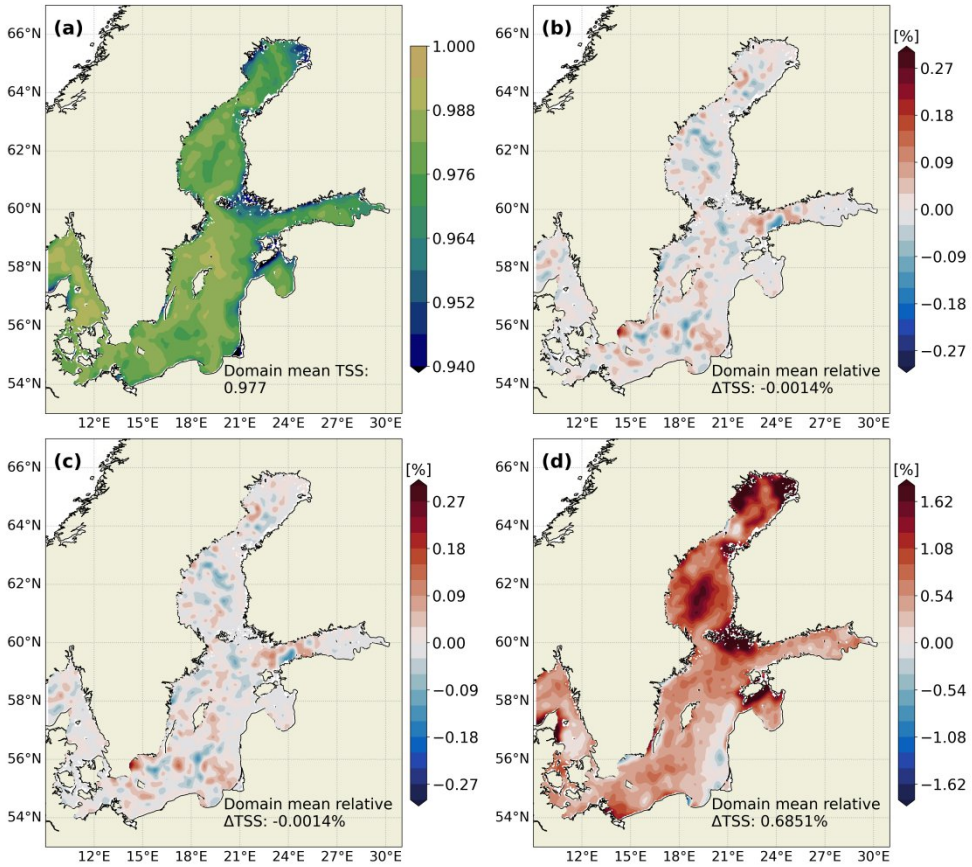


Figure 3. Spatial distribution of the Taylor Skill Score (TSS) for simulated sea surface temperature and its change relative to the reference experient. Panel (a) shows absolute TSS for the reference simulation (REF). Panels (b–d) show the fractional relative change in TSS with respect to REF for CHL_SAT, CHL_3D, and CHL_SST, respectively. The lower-right annotation in panel (a) gives the domain-mean absolute TSS for REF. The annotations in panels (b–d) give the domain-mean relative TSS change, expressed in percent. Note the different colour scale used in panel (d).

L336-7: "CHL_SST allows for the developement of a more accurate thermal environment" – again, a conclusion is drawn about the whole water column based only on the surface.

Response: We agree that, given the current assimilation system employs vertical localization, SST constraints should affect only the near-surface layer and should not impact the entire vertical water column. We revised the text to

"CHL_SST experiment provides an improved representation of the surface thermal state relative to REF, as indicated by the SST skill metrics"

L336: "developement" – development

Response: Corrected.

L337: “thermal environment, which is a prerequisite for correctly representing the stratification-dependent processes” – at some point in this manuscript, more discussion is needed about how true this statement is for primarily salinity-stratified waters.

Response: We agree the original wording was too broad. It has been replaced by:

“The CHL_SST experiment provides an improved representation of the surface thermal state relative to REF, as indicated by the SST skill metrics. This provides a better-constrained near-surface physical background for interpreting the associated biogeochemical response.”

L338: “SCM” – this hasn’t been defined yet.

Response: Thanks for noticing this. The “SCM” was defined at first use as the subsurface chlorophyll maximum (SCM) in the introduction section (Line 102).

L338-40: “The subsequent thermal response to biological corrections is therefore interpreted as a secondary, nonlinear adjustment within the model rather than a basin-scale physical reorganization.” – please explain further what you mean by this and why you conclude it.

Response: Thank you for pointing this out. We have revised the text to clarify how the thermal response is diagnosed and why it is interpreted as a secondary model-mediated response. The corresponding text has been revised as

“the thermal response associated with biological corrections is therefore diagnosed primarily from the differences between CHL_SAT or CHL_3D and REF, where SST is not directly assimilated. CHL_SST is used to assess the combined effect of physical and biological constraints rather than as an isolated diagnostic of the Chl-a-induced thermal response. The thermal response associated with biological corrections is therefore interpreted as a secondary model-mediated adjustment rather than a basin-scale physical reorganization.”

L343: “systematically underestimation” – systematic

Response: Thanks to noticed this grammatical error. We corrected to “systematic underestimation”

L344-5: “The spatial distribution of model errors shows a clear south–north gradient (Fig. 4).” – this isn’t clear to me.

Response: we rephrased the text as follows: "The spatial distribution of model errors is heterogeneous, with relatively smaller errors in the Arkona and Bornholm basins and larger errors in the northeastern Baltic, particularly in the Gulf of Finland and Gulf of Riga (Fig. 4)."

Section 4.2.1 – some of these RMSE values seem quite high, it would be useful to add some more discussion of what typical chlorophyll magnitudes are. Or at least say that this is looked at more in Section 4.2.2.

Response: We thank the reviewer for pointing this out. We have added a sentence to provide context for the magnitude of the reported RMSE, noting that it reflects the substantial spatial and seasonal variability of Chl-a concentrations across the Baltic Sea. The seasonal and spatial distribution of Chl-a concentrations is examined in more detail in Section 4.2.2:

“The magnitude of the basin-wide RMSE should be considered in the context of the substantial spatial and seasonal variability of Chl-a concentrations across the Baltic Sea, which is examined in detail in Section 4.2.2. ”.

L355: “efficienicy” – efficiency

Response: Corrected to “efficiency”

L356: “correponding a” – corresponding to a

Response: Corrected to “corresponding to a”

L359: “reduces” – reducing

Response: Corrected to “reducing”

L362: “complementary rather than redundant” – rephrase for clarity

Response: We have rephrased the text to clarify that the addition of SST observations influences the coupled response beyond the effect of Chl-a assimilation alone.

We revised the text to “This comparison also indicates that adding SST observations influences the coupled response beyond the effect of Chl-a assimilation alone.”

Fig. 5 – OLCI is spelt “ocli” in the key.

Response: Corrected to “OLCI”.

L372: “approxmiately” – approximately

Response: Corrected to “approximately”.

L382: “The impact of assimilation on the spatial distribution of simulated surface Chl-a shows a clear seasonal divergence” – this and subsequent text just seems to be describing the seasonal cycle of Chl-a in the model, rather than in the impact of assimilation.

Response: We have separated the description of the observational seasonal cycle from the assessment of the assimilation response. The revised text now makes clear that the observational seasonal cycle provides context, while the effects of assimilation are evaluated from the differences between each DA experiment and REF.

The revised text begins:

“The seasonal distribution of the observational Chl-a product provides the background against which the effects of assimilation can be evaluated. The DA experiments reproduce the broad seasonal increase in surface Chl-a, but their impacts differ among seasons and regions. We therefore focus below on the differences between each DA experiment and REF rather than interpreting the observational seasonal cycle itself as an assimilation effect.”

L387-8: “Notably, assimilating SST data yields an error reduction approximately 1.8 times greater than that achieved through the biological constraints alone” – assuming this statement is comparing Fig. 6m and 6q, then this error reduction seems to be limited to a small geographic area and is still rather low in absolute terms, which should be noted.

Response: Thanks for pointing this out. We have qualified this comparison to clarify its limited spatial extent and magnitude. The revised text reads:

“Notably, the additional error reduction associated with the SST constraint is approximately 1.8 times that achieved with the biological constraint alone in the affected region (Fig. 6m,q). This reduction occurs mainly over a relatively small area, including the Kattegat, Danish Straits, Arkona Basin, and Bornholm Basin, and corresponds to a comparatively small absolute difference. It should therefore not be interpreted as a basin-wide improvement.”

L389: “implying that correcting the thermodynamic state is more effective under extreme light-limited conditions” – following on from the above, this seems to be making a bolder conclusion than the evidence implies. I may be wrong about that, but in that case please provide more details about the likely mechanisms at play here.

Response: We agree the conclusion should be narrowed based on present anaysis. The original claim that correcting the thermodynamic state is “more effective under extreme light-limited conditions” has been removed. The revised text states only that the assimilation response varies seasonally and that the present experiments do not isolate light limitation from other concurrent seasonal processes. The text has been replaced with:

"The stronger assimilation response during this period indicates that the effect of the physical constraint varies seasonally, although the present experiments do not isolate the contribution of light limitation from other concurrent seasonal processes."

L390: "bloom" – blooms

Response: "bloom" has been corrected to "blooms" where grammatically appropriate.

L392: "CHL_SAT" – surely CHL_3D is the appropriate experiment to compare CHL_SST to here?

Response: We agree and revised the text : "Under these conditions, CHL_SST reduces the Chl-a bias by 64% (Fig. 6r), substantially outperforming the 39% reduction achieved by assimilating biological observations alone (CHL_3D)"

Fig. 6 – given the results in Fig. 5, it would be interesting to plot OLCI here too.

Response:We appreciate the suggestion. We retained Fig. 6 as a comparison of the seasonal assimilation response relative to REF, using the observational constraint adopted in the experiments. Adding OLCI would introduce a second observational reference whose differences from the multi-sensor product are themselves substantial and are discussed in Sect. 4.2.2 and Fig. S1. We therefore discuss the product dependence of the early-season Chl-a signal separately rather than adding OLCI to the assimilation-response maps.

L411-2: "Physical constraint is most important in winter and spring, while biological constraint becomes more important in summer and autumn" – true, but I think this is just reflecting the underlying oceanic processes, which is worth clarifying.

Response: We agree that the original statement did not sufficiently distinguish the seasonal assimilation response from the underlying seasonal variability of the coupled system. We have therefore revised the text to clarify that the relative contribution of the SST and Chl-a constraints varies with the background state and seasonal variability, rather than implying a strict separation between physical and biological control.

has been revised to:

"The relative magnitude of the assimilation response varies seasonally. During winter and spring, the SST constraint contributes more strongly to the modeled surface response, whereas during summer and autumn the direct Chl-a constraint has a larger influence on the biological fields. This seasonal contrast reflects differences in the background state and variability of the coupled system rather than a strict separation between physical and biological control."

Fig. 7: "Note the enhanced representation of the Subsurface Chlorophyll Maximum (SCM) in the CHL_3D and CHL_SST experiments following the inclusion of vertical profile data." – the results from CHL_SAT and CHL_3D look extremely similar to me. And I'm struggling to see much evidence of SCMs in the observations in the first place. What am I missing?

Response: We agree that the original caption overstated the evidence. In Fig 7, the profile-constrained experiments generally provide a closer representation of the observed vertical Chl-a structure than the surface-only assimilation experiment, particularly during stratified periods.

We have removed the blanket claim that the figure demonstrates an "enhanced SCM" in CHL_3D and CHL_SST.

Section 4.3.1 and Fig. 8. The caption for Fig. 8 describes a different figure than the one actually shown. I think much of the text in Section 4.3.1 is based on the figure described rather than the one shown, hence why it seems to be drawing conclusions not backed up by the plot.

Response: We thank the reviewer for pointing this out. We agree that the original caption and parts of Section 4.3.1 did not clearly distinguish the information provided by Fig. 8 from that provided by the other figures. We have corrected the caption and revised Section 4.3.1 to ensure that the discussion is consistent with the figure shown. In particular, Fig. 8 is now used specifically to discuss the temporal evolution of SCM depth, while the surface Chl-a

response is discussed based on Fig. 7 and the subsurface profile response is evaluated using Table 2 and Fig. 9. We have also removed or revised statements that were not directly supported by Fig. 8.

L441-2: "the basin-mean bias in the CHL_SAT run worsens" – but the RMSE improves, which seems worth noting.

Response: Thanks to point this out. We have revised the text: "The basin-wide profile validation (Table 2) further illustrates that bias and RMSE can respond differently to assimilation. Although the basin-mean RMSE decreases relative to REF, the basin-mean bias in the CHL_SAT run worsens relative to that in the REF run (increasing from 0.14 to 0.62 mg m-3 at 5–10 m and from 0.49 to 0.68 mg m-3 at 10–20 m). This indicates that the reduction in overall error does not imply a uniform improvement at every depth. "

L443-4: "However, the seasonal mean profiles at representative offshore stations show that the satellite assimilation better resolves the observed mixed layer depth relative to the REF run(Fig. 9)" – Fig. 9 shows chlorophyll, not mixed layer depth, so I would hesitate to draw conclusions about mixed layer depth purely based on that plot.

Response:We agree. Fig. 9 doesn't show the better mixed layer depth. We revised the text to "the seasonal mean profiles at representative offshore stations show that the satellite assimilation better resolves the observed Chl-a structure relative to the REF run(Fig. 9)"

L455: "The largest improvement" – since this is just based on comparison against the assimilated data, some caution is needed.

Response:We rephrased the text "The largest improvement relative to REF occurs in intermediate layers (10–20 m)"

L459: "reducing localized profile RMSE reduction" – rephrase

Response:We have substantially moderated this paragraph. In lines 532–542, we stated:

"Incorporation of in situ Chl-a profiles (CHL_3D and CHL_SST) enhances the representation of the vertical biomass distribution, reducing the systematic subsurface chlorophyll underprediction in the unconstrained simulation (Table 2). The largest improvement relative to REF occurs in intermediate layers (10–20 m), where both profile-based configurations reduce the basin-mean bias by 67%, from 0.49 mg m^{-3} in REF to 0.16 mg m^{-3}. Additionally, RMSE decreases from 0.23 mg m^{-3} in REF to 0.14 mg m^{-3} in CHL_3D and 0.15 mg m^{-3} in CHL_SST below 40 m. The seasonal evolution of the SCM depth (Fig. 8c) shows that both CHL_3D and CHL_SST represent the seasonal deepening of the SCM during summer stratification … During episodic mixing, such as the early-2018 event at station BY29, the profile-constrained experiments show a downward shift in SCM depth following the event (Fig. 8d)."

L461-2: "This adaptive capability is further evident in their precise responsiveness to transient hydrodynamic forcing. During episodic mixing, such as early 2018 at station BY29, these constraints promptly induce sharp downward shift in SCM following a mixing event (Fig. 8)." – this statement does appear to match the version of Fig. 8 shown. But I'm sceptical about the cause and effect described. The text is suggesting that the assimilation changed the wider model environment which led to the model reproducing the SCM as an emergent property. But is it just that the chlorophyll was assimilated and that's the simple reason for the change in chlorophyll? This is an example where analysis of the vertical physics and nutrients would greatly help elucidate.

Response: We agree and have reconstructed these statements (see response to L459 above). We no longer claim that the assimilation reorganized the wider physical environment to produce the SCM as an emergent property; we state only that the profile-constrained experiments adjust the SCM depth following mixing, consistent with the direct constraint from in situ Chl-a profiles. The new station diagnostics (Fig. S2, Table S1) provide the requested context (lines 643-646).

Fig. 9: "correct the timing" – how do these plots show that?

Response: The term has been removed from the caption. The Figure 9 caption now reads:

"Figure 9. Seasonal mean vertical profiles of Chl-a at monitoring stations BY1, BY5, BY15 and BY29. Shaded regions indicate the ±1 standard deviation of the observations at the corresponding depth, and lines show the corresponding simulation results for the four experiments."

L471: "Seasonal mean profiles (Fig. 9) show the greatest RMSE reductions occur in spring and summer" – perhaps, but this isn't obvious just from looking at them.

Response: We have softened the point and revised the text to "Seasonal mean profiles (Fig. 9) suggest that, relative to REF, the improvements are most pronounced occur in spring and summer, whereas autumn mixing reveals a clear difference between the two profile-constrained configurations."

L473: "successfully guide the simulated SCM" – this seems to be referring to Fig. 9d, which does not show any SCM?

Response: We agree the paragraph did not distinguish Figs. 8 and 9. The downward-displacement statement refers to Fig. 8, now explicitly cited, and "successfully guide the simulated SCM downward" is replaced by "better capture the downward displacement of the simulated SCM" (lines 549–551):

"As shown by the temporal evolution of the SCM in Fig. 8, autumnal convection deepens the mixed layer, both the CHL_3D and CHL_SST runs better capture the downward displacement of the simulated SCM into intermediate depths (approximately 15–30 m)."

L474: "However, the REF and CHL_SAT runs fail" – again, is this referring to Fig. 9d? If so, CHL_SAT and CHL_3D look extremely similar. Or is it referring to Fig. 9a (winter)? This does have a clear distinction between REF/CHL_SAT and CHL_3D/CHL_SST, but I still don't see any SCM or change in the depth of vertical gradients. All of this paragraph, it's not clear what I'm meant to be looking at.

Response: This statement refers to Fig. 8 (inability to reproduce the autumnal downward displacement of the SCM), while the subsequent discussion of vertical differences refers to the autumn mean profiles in Fig. 9d. We have added explicit figure references, replaced "fail to adapt to this vertical feature" with "do not reproduce this autumnal deepening of the SCM," and removed the stronger inference about direct modulation of the seasonal pycnocline (lines 552–557).

"However, the REF and CHL_SAT runs do not reproduce this autumnal deepening of the SCM (Fig. 8). Furthermore, a distinct vertical partitioning emerges between the two profile-constrained configurations (Fig. 9d): The CHL_3D run primarily improves the representation of the water column within the upper 30 m, whereas the CHL_SST run provides a more accurate structural representation of the deeper chlorophyll features in the 30–55 m depth range. This divergence suggests the two profile-constrained configurations influence the simulated subsurface chlorophyll structure differently during the autumn transition."

Fig. 10 only plots REF and CHL_SAT, but the corresponding text in Section 4.4 repeatedly describes features "across Chl-a–only assimilation experiments" as if all experiments were shown.

Response: We have revised the narrative in Sect. 4.4 to ensure consistency with the experiments actually displayed (lines 568, 571).

Fig. 10 caption references "deep-water warming" but the plot only shows near-surface warming?

Response:We corrected it to "subsurface" range.

L502: "TSS change < 0.01%" – is this actually %, or should it just be 0.01?

Response:Here it is relative TSS change, so we revised the text to "relative TSS change < 0.01%, see Fig. 3"

L502-3: "This near-neutral response suggests that the bio-optical feedback primarily redistributes heat vertically rather than driving large-scale surface temperature anomalies" – it's perhaps worth noting that since the ocean model is forced by atmospheric fluxes, this will limit the extent to which SST can vary.

Response: We agree and have made this limitation explicit (lines 581–585). The revised text now reads:

"The thermal changes are small and spatially heterogeneous, and they do not produce a basin-wide improvement in SST statistics (relative TSS change < 0.01%, See Fig. 3). The localized temperature differences are consistent with changes in the vertical distribution of shortwave heating associated with altered Chl-a, but the present analysis does not directly quantify the corresponding upper-ocean heat-content redistribution. The weak surface response is also consistent with the strong role of salinity stratification, wind mixing, and surface heat fluxes in controlling Baltic Sea upper-ocean temperature."

L503-5: "In salinity-stratified marginal seas like the Baltic Sea, buoyancy and horizontal gradients are primarily influenced by wind mixing and riverine inputs, which dominate the weaker thermodynamic signatures of biological light absorption." – as per my previous comments, how much does this limit what you can expect SST assimilation to achieve?

Response: We now state that salinity stratification, wind mixing, riverine inputs, and atmospheric forcing constrain the thermal impact of SST assimilation, so its direct effect is strongest near the surface, while it can still influence subsurface biogeochemistry through coupled processes (lines 584–588).

L517: "negative values represent the equivalent hydrogen sulfide (H2S) debt" – what would you expect observed values to be in these circumstances? Zero?

Response: We thank the referee for raising this point. We have added Sect. 2.2.3 to describe the dissolved-oxygen observations used for evaluation. Oxygen observations at BY1, BY5, BY15, and BY29 were obtained from the SHARK database and were not assimilated, providing an independent evaluation of all experiments. Negative oxygen values in SHARK represent oxygen equivalents associated with hydrogen sulfide under anoxic or sulfidic conditions and are compared directly with the model oxygen-equivalent field, in which SCOBI likewise represents hydrogen sulfide as negative oxygen equivalents (1 mL H_2S L^{-1} = -2 mL O_2 L^{-1}; Fonselius, 1962).

L522: "by current" – "by the current"

Response: Correct to "by the current DA framework."

L526: "observational uncertainty" – please provide a reference if possible.

Response: Thank you for pointing this out. We have clarified that the horizontal error bars represent ±1 standard deviation of the in situ observations at each station over the simulation period.

L527-8: "increasing model bias and observational uncertainty within the oxycline (50–100 m)" and L530: "The persistently wide observational uncertainty in the 50–100 m oxycline zone" – the increased observational uncertainty seems to be confined to the steepest gradients, whereas the model bias increases over a larger depth range. It's worth being clear on this distinction.

Response: Thank you for pointing this out. To avoid this ambiguity, we have removed this statement from the Figure 11 caption. The caption of Figure 11 now reads :

"Comparison of simulated vertical profiles of the dissolved oxygen (ml L^{-1}) against in situ observations at monitoring station (a) BY1 (Arkona Basin), (b) BY5 (Bornholm Basin), (c) BY15 (east Gotland Basin), and (d) BY29 (north Baltic proper). Horizontal error bars on the in situ observation represent ±1 standard deviation of the observations over the simulation period."

L531-2: "These results delineate the boundaries of what surface-focused, weakly coupled DA." – this sentence is incomplete.

Response:We revised the text to: “These results help delineate the limits of surface-focused, weakly coupled DA”

L542-4: “Here, we extend these findings by separating physical control on ecosystem dynamics from the thermal response to biological corrections, allowing the direction and strength of biophysical coupling to be evaluated explicitly in a single modeling framework.” – the results presented here are valuable, but only the impact of assimilation is assessed, not the strength of the coupling itself, so I think this sentence is over-selling things.

Response: We agree that the original wording overstated the inference. The experiments assess the response of the present coupled assimilation system to different physical and biogeochemical observational constraints; they do not provide a direct quantitative measure of the intrinsic direction or strength of biophysical coupling. We have therefore removed the claim that the coupling strength is evaluated explicitly. The revised text now states:

“This result does not quantify the intrinsic strength of biophysical coupling in the Baltic Sea, but it demonstrates the value of jointly evaluating hydrographic and biogeochemical diagnostics when interpreting vertical Chl-a responses and when testing coupled assimilation strategies in halocline-dominated coastal seas.”

L549-50: “indicates a redistribution of heat rather than a substantial change in upper-ocean heat content.” – it would be worthwhile and presumably straightforward to calculate upper-ocean heat content to confirm this.

Response: We agree that the original wording was too specific in the absence of an upper-ocean heat-content calculation. Although CHL_SAT and CHL_3D show weak surface warming together with subsurface cooling, this vertical temperature pattern alone does not quantify the net upper-ocean heat-content response. We have therefore removed the statement that the response represents heat redistribution rather than a substantial heat-content change.

L551-2: “they may still modify the local density gradients near the thermocline and influence the timing of seasonal mixed-layer deepening” – again, presumably it would be relatively straightforward to check this?

Response: We agree that this interpretation requires additional hydrographic diagnostics. In the revised manuscript, we have added station-based temperature, salinity, potential density, and MLD diagnostics across all four seasons (Fig. S2; Table S1). These show station- and season-dependent differences in density structure and MLD among the assimilation configurations. However, they do not establish that the relatively small Chl-a-related temperature anomalies modify local density gradients near the thermocline or alter the timing of seasonal mixed-layer deepening. We have therefore removed this specific causal statement and restrict the revised discussion to the diagnosed differences in hydrographic structure and MLD.

L552-3: “implications for nutrient entrainment” – again, a first-order check on how nutrients differ between experiments would be informative here.

Response: We agree that the nutrient response provides useful context for this interpretation. Nutrient entrainment is not directly diagnosed in the present analysis, and differences in nutrient distributions alone cannot distinguish the contributions of mixing, stratification, biological uptake, and assimilation increments. We therefore removed the causal statement and now identify nutrient entrainment among the processes not quantified here (lines 662–667, see the above response to general comments 3). The new Fig. S2 and Table S1 nevertheless provide station-level nutrient context by showing DIN and DIP distributions alongside the hydrographic and SCM diagnostics.

L554-5: “ecological sensitivity may exceed thermal sensitivity, with small thermal perturbations influencing nutrient supply, oxygen ventilation, and the timing of seasonal transitions” – the two parts of this sentence seem to contradict each other.

Response: We agree with the reviewer. The present experiments show different magnitudes of response in biological and thermal variables, but they do not provide a general measure of ecological sensitivity, thermal sensitivity, or the effects of small temperature anomalies on nutrient supply, oxygen ventilation, or seasonal transitions.We have removed this statement and now describe the contrasting responses of the tested configurations to SST and Chl-a constraints, while stating that the experiments do not quantify the intrinsic strength of biophysical coupling.

The revised text reads:

"Our experiments demonstrate a pronounced asymmetric response of the present coupled assimilation system to physical and biogeochemical observational constraints in the Baltic Sea. The SST constraint strongly regulate ecosystem evolution through their effects on the simulated thermal and density structure, mixed-layer depth (MLD), and nutrient distributions, whereas Chl-a corrections produce only a localized and vertically confined thermal response. "

L555: "These mechanistic insights" – without having actually checked if the model is responding as suggested above, this is just speculation rather than insights.

Response: We agree. The original wording overstated the mechanistic interpretation of the experiments. We have removed the phrase "These mechanistic insights" and revised the discussion to describe only the diagnosed response of the present weakly coupled assimilation system. The revised text does not claim that the experiments establish the mechanisms underlying density, mixing, nutrient, or oxygen responses. The relevant text reads:

"Our experiments demonstrate a pronounced asymmetric response of the present coupled assimilation system to physical and biogeochemical observational constraints in the Baltic Sea."

L558: "state(e.g.," – add space

Response: We corrected the spacing and punctuation in the revised text. The sentence now reads:

"This physical setting differs from more weakly stratified systems, in which bio-optical feedbacks can generate larger thermal signals (e.g., Löptien et al., 2009; Gnanadesikan and Anderson, 2009; Park et al., 2014). "

L565-6: "improves the physical environment governing phytoplankton dynamics" – again, this is purely based on SST, which is not the most important physical factor affecting phytoplankton dynamics. Even just looking at mixed layer depth would be helpful.

Response: We agree that SST alone does not provide a sufficient description of the physical environment relevant to phytoplankton dynamics. We have therefore revised this statement and added station-based diagnostics of temperature, salinity, potential density, MLD, DIN/DIP, and SCM depth at BY1, BY5, BY15, and BY29 across all four seasons (Fig. S2; Table S1). The revised discussion now refers to differences in the hydrographic and biogeochemical background rather than implying that SST assimilation alone improves the physical environment governing phytoplankton dynamics.

The revised text reads:

"Relative to CHL_3D, SST assimilation produces station- and season-dependent differences in temperature, salinity, density structure, MLD, nutrient distributions, and SCM depth (Fig. S2; Table S1). Seasonal mean MLD differences are generally small but reach approximately 3.3 m at BY29 in spring, while the largest SCM-depth differences occur during autumn and winter (Table S1). The corresponding changes in DIN and DIP distributions are consistent with physical mediation of the vertical biogeochemical response. Thus, SST assimilation provides a broader physical context for the biogeochemical response, whereas Chl-a assimilation primarily constrains phytoplankton distributions and induces relatively small thermal adjustments"

L570-2: "Biophysical coupling in the Baltic Sea is therefore not weak but strongly asymmetric, with physical forcing dominating biogeochemical variability and biological feedbacks exerting only a secondary influence on the thermal state" – maybe I'm misunderstanding, but this just seems to be stating well-known and basic oceanography that's true almost everywhere, rather than providing novel insight.

Response: We agree and have removed the broad claim, replacing it with a statement limited to the diagnosed response (lines 648–650):

"the coupled assimilation system responds asymmetrically to the two classes of observational constraint: the SST constraint produces a clearer response in the coupled biogeochemical state, whereas the Chl-a constraints produce only small and localized changes in the simulated thermal state. "

L572-3: “provides a mechanistic basis for designing observation-sensitive assimilation strategies in halocline-dominated coastal seas.” – please elaborate.

Response: We agree that “provides a mechanistic basis” was too strong because the present study does not directly diagnose all processes that may mediate the response, including density structure, mixing, nutrient availability, and ventilation. We have therefore removed this formulation.

The revised text now states:

“This result does not quantify the intrinsic strength of biophysical coupling in the Baltic Sea, but it demonstrates the value of jointly evaluating hydrographic and biogeochemical diagnostics when interpreting vertical Chl-a responses and when testing coupled assimilation strategies in halocline-dominated coastal seas.”

L590-2: “For example, at station BY15, CHL_SST brought the SCM depth closer to the observed pycnocline during the 2018 winter transition (Fig. 8), indicating a more realistic coupling between surface forcing and subsurface biomass structure.” Please reference other results too, otherwise this seems cherry-picked.

Response: We removed the BY15 example and the claims about alignment with the observed pycnocline/nutricline. Section 5.2 now refers to the station diagnostics (Fig. S2, Table S1) only as supporting context (lines 662–667):

"The hydrographic and biogeochemical diagnostics in Fig. S2 and Table S1 show that CHL_SST and CHL_3D differ in density structure, MLD, nutrient distributions, and SCM depth across the analyzed stations. These differences support a role of the physical background state in mediating the vertical expression of the assimilated Chl-a information, although the present analysis does not establish a unique causal sequence linking SST assimilation, mixing, nutrient distributions, and SCM depth."

Table S1. Station- and season-dependent differences in mixed-layer depth (MLD), vertical potential density stratification, nutrients, and subsurface chlorophyll maximum (SCM) depth between the CHL_SST and CHL_3D configurations. Positive values indicate larger values in CHL_SST than in CHL_3D. ΔMLD and ΔSCM depth are in metres; ΔStratification is in kg m^{-3}; and ΔDIN and ΔDIP are in mmol m^{-3}.

Station	Season	ΔMLD	ΔStratification	ΔDIN	ΔDIP	ΔSCM depth
BY1	Spring	+0.08	+0.054	+0.038	+0.017	+0.9
BY5	Spring	+0.16	+0.068	+0.092	+0.008	+1.8
BY15	Spring	−0.07	+0.022	−0.084	−0.003	+0.0
BY29	Spring	−3.35	+0.053	−0.083	−0.006	+0.9
BY1	Summer	−0.02	+0.037	+0.106	+0.003	−0.4
BY5	Summer	+0.36	−0.023	+0.144	+0.010	−0.2
BY15	Summer	+0.19	+0.034	+0.032	−0.002	−0.0
BY29	Summer	+0.01	+0.057	+0.141	−0.000	−0.1
BY1	Autumn	−0.41	−0.081	−0.183	−0.016	+3.5
BY5	Autumn	−0.66	−0.098	−0.030	−0.004	+0.5
BY15	Autumn	+0.12	−0.085	−0.040	−0.011	+0.0
BY29	Autumn	+0.70	−0.084	+0.057	−0.005	+2.1
BY1	Winter	+0.26	+0.047	−0.131	+0.003	+0.3
BY5	Winter	+0.35	−0.061	−0.180	−0.008	+0.7
BY15	Winter	+0.62	+0.003	−0.019	−0.001	+1.9
BY29	Winter	+2.34	−0.016	+0.025	−0.006	+4.2

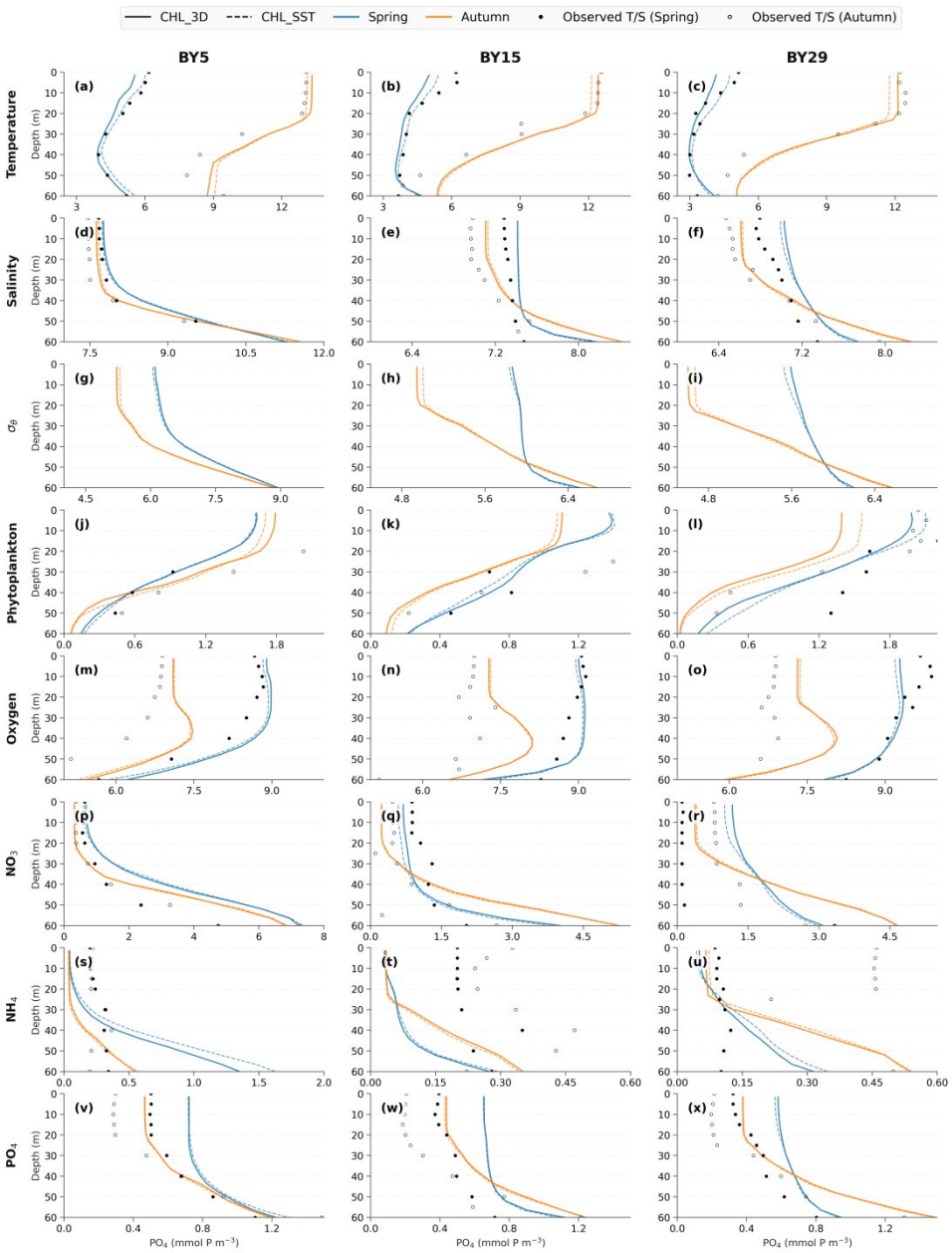


Figure S2. Seasonal mean vertical profiles of temperature, salinity, surface-referenced potential density (σθ), phytoplankton, oxygen, nitrate (NO3), ammonium (NH4), and phosphate (PO4) at BY5, BY15, and BY29. Solid and dashed curves denote CHL_3D and CHL_SST, respectively; blue and orange curves represent spring and autumn, respectively. Black symbols show the corresponding in situ observations where available. The comparison illustrates station- and season-dependent differences in the coupled physical–biogeochemical response to the additional SST constraint. The profiles provide contextual evidence for possible physical mediation of the vertical Chl-a response but do not directly diagnose mixed-layer depth, vertical nutrient fluxes, upper-ocean heat-content changes, or oxygen ventilation.

L592-3: “This spatial correction helps align the simulated biological productive center with the observed nutricline and pycnocline (Teruzzi, et al, 2021)” – without providing any assessment of the observed nutricline and pycnocline, I’m not sure you can confidently conclude this. Teruzzi et al. (2021) is referenced, but that study assimilated nitrate and chlorophyll and no physics. It’s certainly worth referencing here and discussing how your results compare to theirs, but that needs expanding on.

Response: We agree. The present study does not evaluate observed nutricline or pycnocline depths and therefore cannot demonstrate alignment of the simulated productive centre or SCM with these observed features. We have removed this statement and the citation to Teruzzi et al. (2021), rather than introducing an extended comparison with a study based on a different assimilation configuration. The revised text now refers only to differences in the simulated vertical Chl-a response and to the hydrographic and biogeochemical context provided by Supplementary Fig. S2 and Table S1.

L596-7: "SST constraints do not only modify the thermal field but also determine the effective vertical extent of biogeochemical corrections (Goodliff et al, 2019)." – similarly, does this statement apply to your study, to Goodliff et al. (2019), or to both? Given Goodliff et al. (2019) looked at a similar region, it would be interesting to discuss further.

Response: We agree that the original wording did not clearly distinguish the findings of Goodliff et al. (2019) from those of the present study. We have revised the text so that Goodliff et al. (2019) is cited only as external context for coupled shelf-sea physical–biogeochemical interactions. The revised discussion states that, in the present experiments, the additional SST constraint modifies the simulated thermal background and is associated with a different vertical expression of the Chl-a correction. We do not claim that the present analysis determines the effective vertical extent of the correction or isolates the mechanisms responsible for it.

The revised text reads:

"The additional SST constraint therefore modifies not only the thermal field but also the physical background within which the vertical expression of biogeochemical corrections develops. Goodliff et al. (2019) similarly showed that physical–biogeochemical interactions can influence the vertical propagation of biogeochemical information in coupled shelf-sea systems. "

L603-6: "The primary advantage of integrating SST lies precisely in its ability to dynamically re-adjust this hydrodynamic boundary. In this dynamic context, our results show that biogeochemical corrections affect the SCM only when the hydrographic environment retains sufficient vertical coherence to allow surface anomalies to penetrate below the mixed layer" – perhaps this is true, but it hasn't been demonstrated in the results section.

Response: We agree with this reviewer's comment that the original formulation implied a causal mechanism that was not directly demonstrated. we have removed these statements. The revised text now states:

"... These differences support a role of the physical background state in mediating the vertical expression of the assimilated Chl-a information, although the present analysis does not establish a unique causal sequence linking SST assimilation, mixing, nutrient distributions, and SCM depth. "

L606: "(Kahru et al., 2012; Quartly et al., 2023)." – again, references to other studies are attached to a statement about your own results rather than separately discussed.

Response: We agree that the citations could be interpreted as supporting a statement about our own results. We have removed the citations to Kahru et al. (2012) and Quartly et al. (2023) from this passage, together with the statement to which they were attached. We have revised the paragraph so that previous studies are discussed as contextual evidence, while our findings are described separately. The revised discussion now uses the following formulation:

"Goodliff et al. (2019) similarly showed that physical–biogeochemical interactions can influence the vertical propagation of biogeochemical information in coupled shelf-sea systems."

.**L608-9**: "SST therefore defines the stratified domain" – are you sure?

Reponse: We thank the referee for pointing this out. We agree that this formulation was inaccurate. SST assimilation does not define the stratified domain, particularly in the Baltic Sea where salinity makes a major contribution to density stratification. We have removed this statement and now explicitly recognize the role of salinity:

"This response asymmetry should be interpreted in the context of the strong stratification of the Baltic Sea, where salinity makes a major contribution to the density structure (Liblik and Lips, 2019)." and "In the Baltic Sea, SST

assimilation modifies the thermal component of the upper-ocean background within a density structure that is also strongly controlled by salinity."

L612-4: "The nearly invariant oxycline trajectories in our results (Fig. 11) reveal that the static benthic boundary layers in the model effectively seal the sediment-water interface, completely insulating deep-water redox renewals from any pelagic biological updates." – please elaborate on why you conclude it's due to the benthic boundary rather than lack of mixing across the pycnocline.

Response: We agree the attribution was too specific; the experiments do not isolate the relative roles of mixing, ventilation, exchange, benthic demand, or sediment exchange. We removed the statement and revised to (lines 687–691):

"The nearly invariant mean oxygen profiles in our results (Fig. 11) indicate that the tested surface-focused assimilation configurations have limited ability to modify the deep-water oxygen state. This result should not be interpreted as evidence that surface and deep-water processes are dynamically independent. Surface and subsurface biogeochemical states remain physically connected through mixing, sinking, nutrient availability, stratification and remineralization, while the information supplied by surface Chl-a observations is insufficient to constrain the processes controlling deep-water oxygen."

L627-8: "the assimilation system remains blind to the deep anoxic pool" – rephrase this, I think a different phrasing could get your intended meaning across better.

Response: We agree. We have replaced the expression with:

"The Chl-a and SST observations assimilated here provide only limited constraints on the deep oxygen state."

L631-4: "This suggests that surface biogeochemical corrections cannot be expected to reorganize the deep redox environment unless they are accompanied by independent changes in ventilation or benthic flux conditions. From this perspective, the remaining deep-oxygen bias should not be interpreted simply as a residual model truncation error of the DA algorithm, but rather as a physical manifestation of missing deep-water constraints." – I don't fully understand what's being said here, is it basically: "surface biogeochemistry is weakly- or un-correlated with deep-water oxygen, so it is not possible for assimilation of surface biogeochemistry to correct deep-water oxygen errors, observations of the deep ocean are required", similar to the conclusions of Fontana et al. (2013) (https://doi.org/10.5194/os-9-37-2013)?

Response: Yes — that is exactly our intended meaning, and we are grateful for both the clarification and the apposite reference. We rewrote the passage accordingly (lines 708–712):

"The weak deep-oxygen response does not imply that surface-focused biogeochemical assimilation is irrelevant to deep redox dynamics. The Chl-a and SST observations assimilated here provide only limited constraints on the deep oxygen state. Substantially modifying the simulated deep redox field in strongly stratified marginal seas thus requires direct observational constraints on hydrographic structure, deep-water ventilation events, and benthic biogeochemical processes (Fontana et al., 2013)."

We have added Fontana et al. (2013) as relevant context, because that study similarly highlights the value of deep observations and constraints for improving deep-water biogeochemical representation.

Fontana, C., Brasseur, P., and Brankart, J.-M.: Toward a multivariate reanalysis of the North Atlantic Ocean biogeochemistry during 1998–2006 based on the assimilation of SeaWiFS chlorophyll data, Ocean Sci., 9, 37–56, https://doi.org/10.5194/os-9-37-2013, 2013.

L634-7: "The Baltic Sea should therefore be viewed as a system of partially connected compartments. The photosynthetic zone forms the upper compartment, where biomass, light, and the local thermal structure interact

closely. Conversely, the deep basin operates as a separate domain. Its oxygen state is largely controlled by external renewal events and internal consumption processes." – I'm not familiar enough with the Baltic Sea to know, but is this a novel insight or just reaffirming something that's already known?

Response: We agree that the original wording could be interpreted as claiming a novel conceptual view of Baltic Sea functioning. The vertical separation of upper-ocean and deep-basin processes by strong haline stratification, as well as the importance of deep-water renewal and oxygen consumption for deep-water oxygen, are established features of the Baltic Sea rather than new findings of the present study. We have therefore removed the "partially connected compartments" formulation and instead focus on the result directly demonstrated by the experiments:

"The nearly invariant oxycline trajectories in our results (Fig. 11) indicate that the tested surface-focused assimilation configurations have limited ability to modify the deep-water oxygen state. This result should not be interpreted as evidence that surface and deep-water processes are dynamically independent."

"Tabel 2" (change to Table) is very interesting, but in reference to my previous comments it doesn't seem as well backed up by validation results as it could be.

Response: We agree the original title ("Hierarchy of physical–biogeochemical controls") overstated the inference. We retitled it "Summary of diagnosed responses to physical and biogeochemical observational constraints in the simulated Baltic Sea system," restructured it to report diagnosed responses rather than mechanistic controls, and now cite the new Fig. S2/Table S1 in the corresponding discussion. We do not present the table as independent validation of all mechanisms.

Table 3. Summary of diagnosed responses to physical and biogeochemical observational constraints in the simulated Baltic Sea system.

Layer / variable	Observational constraint or process context	DA configuration	Diagnosed response
Surface SST	Direct SST constraint	CHL_SST	Reduced SST bias and improved SST skill.
Surface Chl-a (spring)	Surface Chl-a and SST constraints	CHL_SAT / CHL_3D / CHL_SST	Different seasonal surface Chl-a responses among experiments.
Surface Chl-a (summer)	Surface and profile Chl-a constraints	CHL_SAT / CHL_3D / CHL_SST	Reduced mismatch with assimilated Chl-a observations.
Subsurface Chl-a / SCM	In situ Chl-a profiles and SST constraint	CHL_3D / CHL_SST	Different vertical Chl-a structures; station- and season-dependent SCM depth.
Hydrographic context	Temperature, salinity, density, MLD	CHL_3D / CHL_SST	Station- and season-dependent differences in hydrographic background.
Nutrient context	DIN and DIP distributions	CHL_3D / CHL_SST	Differences in nutrient distributions at selected stations and seasons.
Bio-optical response	Chl-a-dependent light attenuation	CHL_SAT / CHL_3D	Localized changes in vertical temperature structure.
Deep-water oxygen	Surface-focused Chl-a and SST constraints	All configurations	Limited change in deep oxygen among configurations.

The new station-based hydrographic and biogeochemical diagnostics in Supplementary Fig. S2 and Table S1 are now cited in the corresponding discussion text. We nevertheless avoid presenting the table as independent validation of all mechanisms; it summarizes diagnosed responses of the present assimilation configuration

L659: "correcting the seasonal pycnocline and mixed-layer depth" – this isn't shown.

Response: We agree that "correcting" is too strong. Although we added station-based diagnostics of density structure and MLD, these diagnostics compare the CHL_SST and CHL_3D simulations and do not establish correction against independent observations of pycnocline depth or MLD.

We have removed this wording and replaced it with the model-state comparison:

"Relative to CHL_3D, SST assimilation produces station- and season-dependent differences in temperature, salinity, density structure, MLD, nutrient distributions, and SCM depth (Fig. S2; Table S1)."

L674: "which may be adequate at the basin scale" – is this based on the results or just speculation?

Response: We thank the referee for pointing this out. We agree that the statement "which may be adequate at the basin scale" was not directly supported by our results. We have therefore removed this qualification to avoid implying that basin-scale nutrient budget adequacy was demonstrated.

L683: "The 30–39% reduction in Chl-a RMSE improves the reliability of bloom detection" – as a caveat, this is against the assimilated data. What you ideally want from a model used for monitoring is for it to be skilful in places you don't have observations – by only comparing against assimilated data it is hard to make conclusions about that.

Response: We agree. The original statement overstated the implication of the Chl-a RMSE reduction because the satellite Chl-a products are assimilated. The comparison therefore measures consistency with the applied observational constraint, rather than independent predictive skill in unobserved locations or periods.

We have removed the claim that the 30–39% RMSE reduction improves the reliability of bloom detection. The revised conclusion states:

"The 30–39% reduction in Chl-a RMSE demonstrates improved consistency with the assimilated Chl-a observations, but it should not be interpreted as an independent measure of bloom-detection reliability."

L684-5: "by providing a baseline that better separates anthropogenic nutrient forcing from natural variability" – this hasn't been shown.

Response: We agree that the statement about separating anthropogenic nutrient forcing from natural variability was not demonstrated by the present experiments. We have removed this claim and now restrict the conclusion to the demonstrated model–observation agreement:

"The assimilation framework provides an improved representation of the observed Chl-a distribution and vertical structure, but does not by itself separate anthropogenic nutrient forcing from natural variability."