EGUSPHERE-2026-3057
“A dynamical downscaling approach for the improved representation of sub-annual coastal sea level variability: the importance of shelf and slope ocean dynamics”
Steinberg, J.M. and Krasting, J.P. and Ross, A.C.

Response to Reviewers - RC1 - Referee #1

This is a high-quality piece of work that advances both process understanding and practical modeling approaches for coastal sea level. The suggested revisions are mostly clarifications and modest quantitative additions rather than new analyses. Once the IB handling is clearly explained and a few supporting quantifications are added, the manuscript will be a valuable contribution to the coastal ocean modeling and sea-level forecasting literature. I am happy to review a revised version.

1. In Section 2.2 and the results, the authors state that “both SPEAR and NWA12 oceans do not feel atmospheric pressure variations” and therefore add the IB effect offline for comparison to GESLA observations. This needs careful clarification.

- In the coupled SPEAR system, does the saved sea level field already include the IB response to the atmospheric pressure field, or was it excluded/post-processed? If SPEAR sea level already contains the IB signal, adding it again would double-count the contribution.

  In the coupled SPEAR system, the ocean does not feel atmospheric pressure variations and thus sea level diagnostic output does not include IB effects. The IB response is calculated offline using sea level pressure output and then added to sea level time series. This is now clarified (L123-124).

  “In both NWA12 and SPEAR configurations, the ocean does not feel atmospheric pressure variations as computational limitations prohibit model timesteps from being short enough to resolve surface gravity waves.”

- Please explicitly state how the IB term was calculated (which pressure field, reference pressure, etc.) and whether high-frequency (sub-daily) pressure variability was neglected. This matters most at higher latitudes where IB can explain ~33% of variance.

  The IB contribution is calculated as the sea level pressure anomaly with respect to the global ocean mean and local time mean divided by (rho_0*g) as in Ponte 2006 (10.1175/JTECH1864.1). Sub-daily pressure variations are neglected as we use daily means. These daily means were also used to assess NWA12 IB contributions as no higher frequency pressure output from SPEAR was saved (note, we do use the 6hrly wind field to force NWA12). We have added a new figure that compares ERA5 and SPEAR daily mean sea level pressure variability, expressed as an IB contribution, at each tide gauge (Figure 7 - included here). Text has been added for clarification (L147-149).

  "In using daily mean surface pressures, referenced to global ocean mean, higher frequency variability associated with synoptic scale storms may result in an underestimation of extremes and distribution tails.”

- A short additional sentence or two explaining why pressure forcing was omitted (or not saved) in the NWA12 configuration would help readers understand the experimental design.

Pressure forcing was omitted as the model timestep is not short enough to resolve surface gravity waves and instabilities immediately arise resulting in a simulation crash. Text has been added (L123-124).

"In both NWA12 and SPEAR configurations, the ocean does not feel atmospheric pressure variations as computational limitations prohibit model timesteps from being short enough to resolve surface gravity waves."

1. The claim that downscaling "enhances variability by nearly an order of magnitude across all resolved frequencies" is central to the paper and visually supported by the spectra. However, it would be much stronger if supported by quantitative ratios.

- Please add a supplementary table or a small panel showing NWA12/SPEAR variance ratios at the tide-gauge stations for the key frequency bands (or at least for the 5-day–2-week and >10-week bands).

We have created a new figure (Figure 7) that compares model bathymetry, observed coastal sea level variance, and variance ratios at all tide gauge stations. Variance ratios are computed in each frequency band subsequently analyzed and to characterize ensemble spread. New text is added discussing the information provided in this figure (L189-198).

At the fifteen tide gauge stations considered here, SPEAR and NWA12 variance differences are explored more closely and alongside station depth differences and observed variance (Fig. \ref{f7}). At nearly all stations, coastline-adjacent NWA12 grid cells are shallower that those in SPEAR (Fig. \ref{f7}, left). These differences, greatest north of Cape Hatteras and along the Florida coastline, contribute to a decrease in SPEAR coastal sea level variance arising from less coastal amplification or shoaling. Diagnosed IB contributions compare favorably to reanalyses and when added to model coastal sea level variance, NWA12 values approach observed values (Fig. \ref{f7}, middle). Ratios of NWA12 to SPEAR variance (Fig. \ref{f7}, right) reveal a five-fold or greater increase in coastal sea level variance in this region that is largely constrained to shorter timescales. Further north, SPEAR and NWA12 stations show comparable variance suggesting more limited downscaling improvements in this region. South of Cape Hatteras, NWA12 increases appear linked to coastal station depth difference. These comparisons reveal the extent and pattern of NWA12 coastal variance increases and show that they approach and better match observed values.

- In Figure 6, the difference maps are shown after coarsening NWA12 to the SPEAR grid. It would be helpful to also show (or describe) the raw high-resolution NWA12 variance field so readers can appreciate the full magnitude of the enhancement before any spatial averaging.

The full NWA12 variance is shown in the left-most column of Figure 6. We have also added variance comparisons and ratios in Figure (7). See above response.

1. NWA12 and SPEAR differ in horizontal resolution, bathymetry, and likely in some aspects of vertical mixing, bottom friction, or tide implementation. Although the use of identical surface forcing helps isolate the ocean response, the paper should briefly acknowledge that not every difference is purely attributable to resolved shelf/slope dynamics. A short sentence in Section 2.1 or 4 noting that the dominant differences are still expected to arise from bathymetric resolution and the ability to represent coastal trapped waves and cross-shelf transport would be sufficient and intellectually honest.

We have added text to this effect L(110-113, 286-288).

“We also acknowledge that additional factors, for example, non-linear tide-current interactions, may contribute to enhanced NWA12 near-coast variance, but we expect these impacts to be smaller as NWA12 does not include a dynamic land/sea interface where these effects may become large.”

“While we acknowledge that model configuration differences contribute to differences in coastal process representation, we focus on and contrast NWA12 and SPEAR coastal sea level variability linked to continental shelf and slope representations.”

2. The 5th and 95th percentiles in Figures 10–12 are based on roughly 150 daily values per lead month (5 members × ~30 days). This is acceptable, but the tails are sensitive to sample size.

- Please add a brief comment on the robustness of these percentiles (e.g., via bootstrap resampling across ensemble members or years, or by noting that the patterns are consistent across individual ensemble members).

We now show the 5th and 9th percentile ensemble spread (range of calculated values) in Figure 13 (included here). This reveals a consistent pattern across individual ensemble members. A comment is added (L270-271).

"Ensemble member spread, shaded bands in Figure \ref{f13}, reveal consistency in these NWA12-SPEAR differences.”

- Consider also showing the inter-ensemble spread (shading) on the percentile curves in Figure 12 or in a supplementary figure.

This is now added.

1. The introduction and conclusions position the work as a step toward a “practical time-varying statistics” framework for forecasting. This is a fair aspiration, but the current experiments are 1-year retrospective hindcasts initialized on 1 December.

- A short paragraph (or expanded sentence in the conclusions) discussing computational feasibility for true operational use, the need for real-time bias correction or anomaly nesting, and the missing ocean–atmosphere feedback would strengthen the applied relevance without over-selling the current setup.

Done (L318-321).

“Next steps focused on generating statistics-oriented forecasts to be used in applied settings are feasible as SPEAR seasonal forecasts are generated seasonally, bias correction is not needed as ocean anomalies are applied to climatology, and single ensemble member NWA12 run times are only a few hours.”

- The ocean-only configuration means that SST differences can alter surface heat fluxes even though the wind field is identical. While this is probably secondary for shelf sea level, it is worth a one-sentence acknowledgment.

This is acknowledged (L103-105).

"We acknowledge a missing feedback in the NWA12 ocean-only configuration in which the ocean currents and temperature in the regional model do not feed back onto the atmosphere in the forcing model. As a result, heat fluxes in NWA12 and SPEAR differ due to different ocean temperatures, but configurations and bulk formulae for the two are the same."

Figures

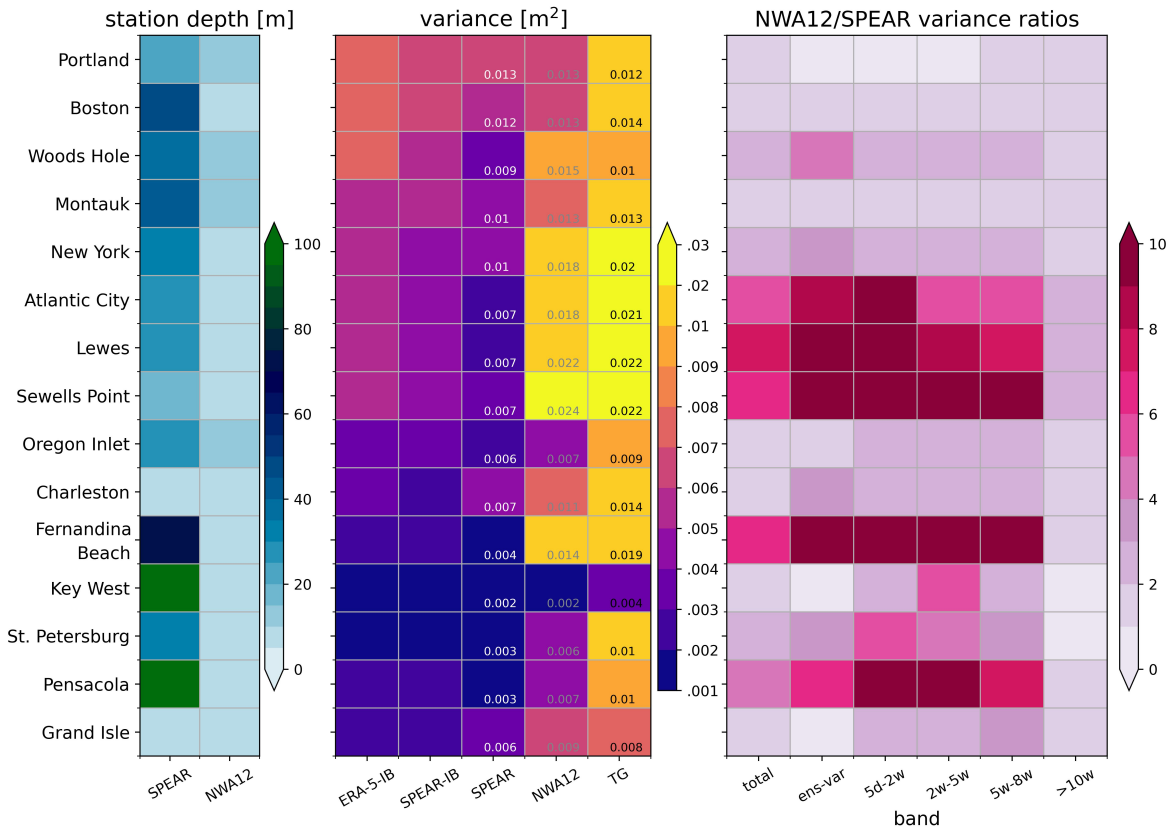


Figure 7 - left) SPEAR and NWA12 ocean depth at 15 tide gauge stations. middle) Coastal sea level variance: ERA-5 inverted barometer contribution (ERA-5-IB), SPEAR inverted barometer contribution (SPEAR-IB), SPEAR, NWA12, tide gauge observations (TG). Text reports model sea level variance with the IB contribution included and observed sea level variance. Note colormap and colorbar are on a log scale. right) NWA12/SPEAR variance ratios: total, ensemble spread (ens-var), and for each frequency bands.

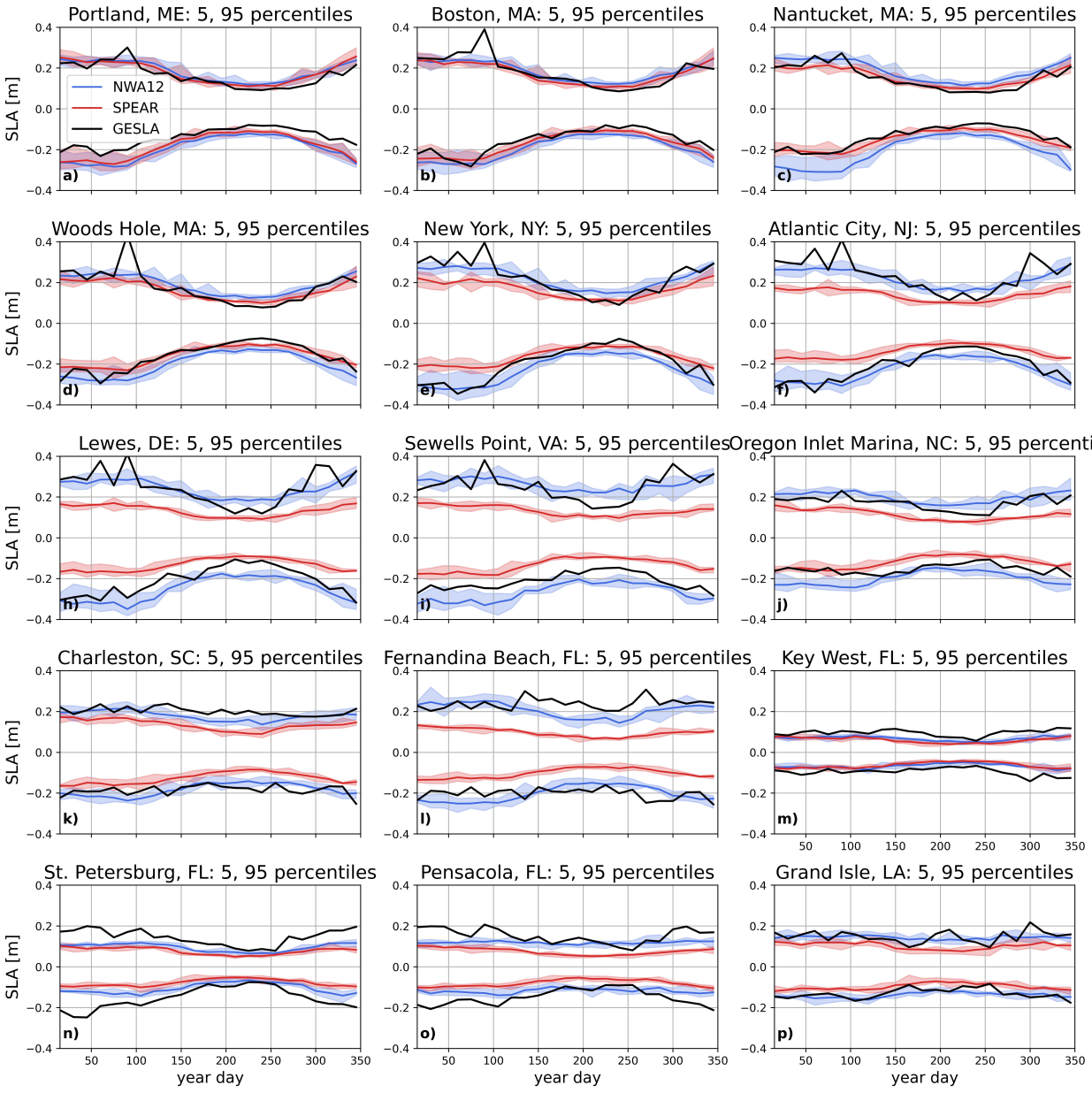


Figure 13 - 5th and 9th percentiles of daily sea level anomaly distributions as a function of simulation lead time (days since model initialization). SPEAR is in red, NWA12 in blue, and observations (GESLA) in black. Red and blue shading shows the mean ensemble spread.

EGUSPHERE-2026-3057
“A dynamical downscaling approach for the improved representation of sub-annual coastal sea level variability: the importance of shelf and slope ocean dynamics”
Steinberg, J.M. and Krasting, J.P. and Ross, A.C.

Response to Reviewers - RC2 - Referee #2

1. Summary and Overall Assessment

This manuscript evaluates a 100-member ensemble of one-year dynamical downscaling hindcasts for the period 1999–2018 along the U.S. East and Gulf Coasts. The hindcasts are produced using the high-resolution, 1/12° Northwest Atlantic regional ocean model (NWA12), driven by retrospective seasonal forecasts from the global coupled SPEAR model, which has an atmospheric resolution of 0.5° and an ocean resolution of 1°.

The authors investigate how increased ocean-model horizontal resolution and improved representation of continental-shelf and slope bathymetry affect subannual coastal sea-level variability and short-term daily extremes.

Overall, this is a well-designed, timely, and insightful study that addresses an important gap in coastal sea-level forecasting: the need to connect global climate and seasonal prediction systems with fine-scale ocean dynamics over the continental shelf and slope. The methodology is robust and makes effective use of a comprehensive 100-member ensemble dataset. The manuscript is logically organized, and the figures are generally clear and of high quality.

The physical interpretations linking model resolution, shelf hypsometry and bathymetry, coastal-trapped waves, and western boundary-current behavior are physically sound and well supported by the results. The study also provides useful and actionable insights for regional coastal-risk assessment and the prediction of extreme water levels.

Recommendation: Minor Revision. The manuscript is suitable for publication after the authors address several methodological clarifications, expand the discussion of certain physical mechanisms, and correct a small number of technical and typographical issues.

2. Major Comments

2.1 Uncoupled Air–Sea Interaction and Surface-Stress Formulation
The authors acknowledge in Section 2.1, Lines 97–104, that NWA12 is an ocean-only model forced by atmospheric fields from SPEAR. Consequently, regional ocean sea-surface temperatures and currents do not feed back onto the atmospheric circulation or surface winds. Please clarify whether the wind-stress calculation in NWA12 accounts for relative air–sea velocity, that is, whether the stress is calculated using surface wind velocity minus ocean surface-current velocity. In regions of strong horizontal shear, particularly near the Gulf Stream, current–wind stress feedback can substantially influence upper-ocean energetics and shelf–slope exchange.

A brief statement indicating whether absolute winds or relative winds are used in the NWA12 bulk-flux formulation would strengthen the methodological description in Section 2.1.

The NWA12 regional model configuration used here calculates the relative wind to determine the stress to apply to the ocean; in other words, the NWA12 ocean surface current was subtracted from the SPEAR winds being applied. We acknowledge that the use of relative wind forcing in a high resolution uncoupled ocean model can often have an overly strong eddy killing effect (e.g., Renault et al. 2020). In recent experiments with a different atmospheric forcing

product, we found that using a correction similar to Renault et al. 2016 to make the wind forcing closer to absolute rather than relative to the current increased the eddy kinetic energy in the model. Therefore, we expect that the increased sea level variance in NWA12 would be boosted even further if this correction was included. To address this comment, we have inserted a description of how the wind forcing was applied into the methods (L91-93).

“While surface ocean conditions do not feed back onto the atmosphere, wind stress calculations in NWA12 account for relative air-sea velocity, with NWA12 surface currents subtracted from the SPEAR surface winds.”

2.2 Treatment of Tides and Nonlinear Tide–Surge Interactions
NWA12 includes tidal constituents, which are removed offline using UTide before the model output is averaged to daily resolution. In contrast, SPEAR does not explicitly resolve tides. Nonlinear tide–surge interactions, including the modulation of bottom friction by tidal currents, can influence non-tidal residual sea level in shallow coastal regions. The authors should briefly discuss whether some portion of the enhanced coastal variance in NWA12 relative to SPEAR could arise from nonlinear tidal interactions represented in NWA12, rather than solely from improved bathymetric resolution, shelf dynamics, and wind-driven Ekman transport.
The authors do not necessarily need to quantify this contribution but acknowledging it as a possible source of difference between the two models would provide a more complete interpretation of the results.

Thank you for this comment, we now acknowledge non-tidal interactions as a possible source of SPEAR-NWA12 differences (L110-113).

“We also acknowledge that additional factors, for example, non-linear tide-current interactions, may contribute to enhanced NWA12 near-coast variance, but we expect these impacts to be smaller as NWA12 does not include a dynamic land/sea interface where these effects may become large.”

2.3 Temporal Resolution of the Inverted Barometer Correction
The inverted barometer effect is added offline using daily averaged atmospheric surface pressure from SPEAR.

Because synoptic storms and coastal-flooding events often involve substantial pressure changes on hourly to six-hourly timescales, the use of daily averaged pressure may smooth short-lived pressure minima. Could this procedure lead to an underestimation of the upper tails of the daily sea-level distributions, particularly in Figures 10 and 12 and at higher-latitude stations such as Boston and Portland?

Please clarify whether higher-frequency pressure fields were available and, if so, why daily averaged pressure was selected. At minimum, the manuscript should acknowledge the potential effect of temporal averaging on the representation of extreme sea-level events.

Yes, upper tails of the distributions in both SPEAR and NWA12 may be underestimated. This will not impact SPEAR-NWA12 comparisons in Figures 11 and 12 (left columns), but will impact comparison to observations in Figures 11-13. Despite this, comparison between models and observations at northern stations are favorable. Higher-frequency pressure fields were not saved. See added text (L147-150, 262-263).

“In using daily mean surface pressures, referenced to global ocean mean, higher frequency variability associated with synoptic scale storms may result in an underestimation of extremes and distribution tails.”

“At all stations, model extremes, and distribution tails, are expected to be under-represented as daily mean surface pressures miss a higher-frequency component of coastal flooding and surge associated with synoptic storms.”

2.4 Coastal-Waveguide “Choke Point” Near Cape Hatteras
The manuscript identifies Cape Hatteras as a critical geographic transition where the continental shelf narrows, and the Gulf Stream separates from the coast. The discussion in Section 3.3 would benefit from a more explicit physical explanation of why SPEAR performs poorly south of Cape Hatteras.

Please discuss whether the 1° SPEAR ocean grid is too coarse to resolve either the relevant coastal Rossby radius of deformation or the physical width and bathymetric structure of the narrow continental-shelf waveguide. Such a discussion would help clarify the spatial-resolution threshold required to represent coastal-trapped-wave propagation and related shelf dynamics in this region.

It would also strengthen the broader interpretation of why dynamical downscaling provides substantially greater benefits in some coastal regions than in others.

SPEAR is unable to resolve the physical width and bathymetric structure of the narrow continental shelf in this region (shelf widths decrease to less than 1 degree latitude/longitude). The relevant coastal Rossby radius of deformation is too small for even NWA12 to properly resolve (O[2-10km]). Discussion is added (L245-250).

“While SPEAR horizontal resolution is too coarse to resolve the first baroclinic Rossby radius of deformation over the continental shelf ($\mathcal{O}$ 2-10 km), and NWA12 may at best resolve this radius with one grid cell, model resolutions relative to the shelf width are quite different. SPEAR may resolve the average US east coast shelf width of 100-150 km with one grid cell while NWA12 will have over twelve. Where the shelf narrows, coastal grid cells in SPEAR will generally be too deep as they are generated by averaging over shelf and slope depths. This difference impacts shelf process and coastal trapped wave representation significantly and reveals where dynamical downscaling will improve model coastal sea level. “

See new Figure 7 (included here), left column.

3. Minor Comments and Technical Corrections
Line 248: “...after removal a monthly seasonal climatology…” to: “...after removal of a monthly seasonal climatology….”

Corrected.

Line 411, References: In the citation for Wise et al. (2020), “oddshore-forced” should be corrected to “offshore-forced.”

Corrected.

Please ensure that the terms “non-tidal residual” and “sea-level anomaly,” as well as the abbreviations NTR and SLA, are used consistently throughout the text, figure legends, and captions. In particular, the terminology appears to differ between Figure 3 and Figures 10–11. If the terms refer to distinct quantities, the distinction should be clearly defined; otherwise, consistent terminology should be adopted throughout the manuscript.

We have reviewed and revised usage of NTR and SLA for improved consistency. As all comparisons are between daily means we remove NTR time series in Figure 3.

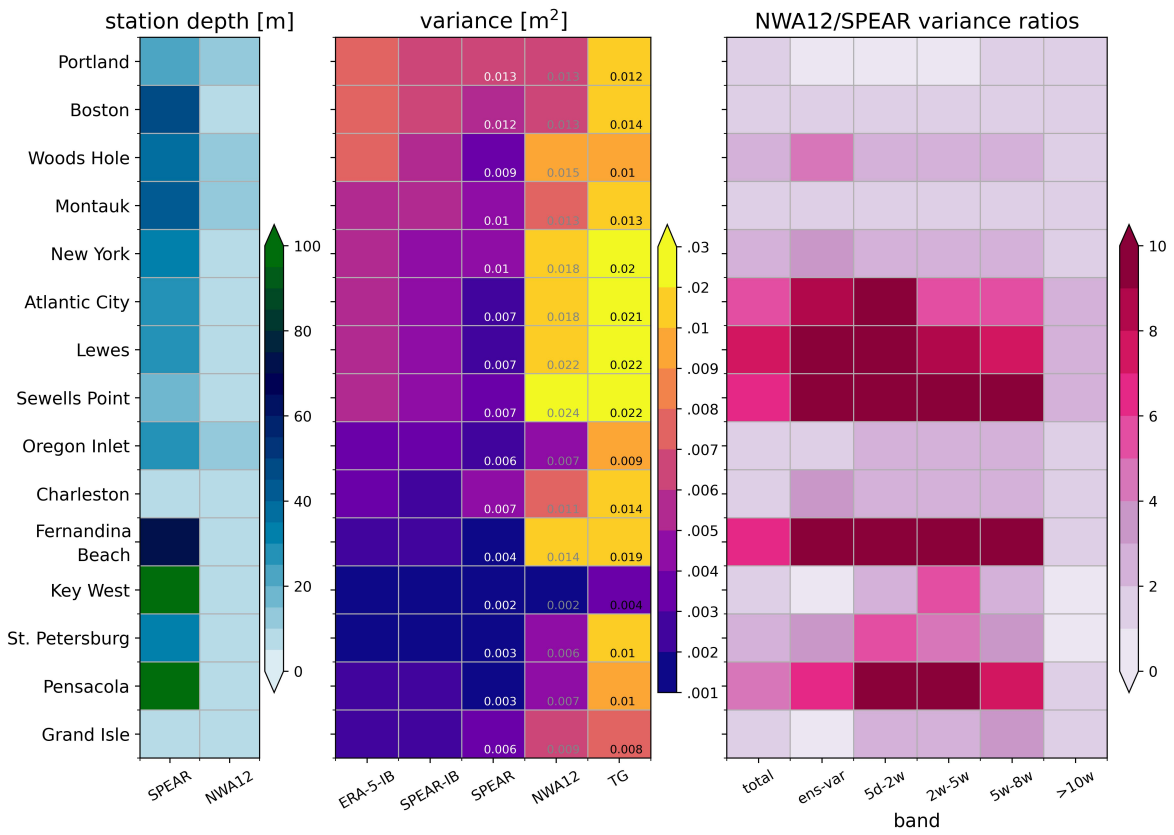


Figure 7 - left) SPEAR and NWA12 ocean depth at 15 tide gauge stations. middle) Coastal sea level variance: ERA-5 inverted barometer contribution (ERA-5-IB), SPEAR inverted barometer contribution (SPEAR-IB), SPEAR, NWA12, tide gauge observations (TG). Text reports model sea level variance with the IB contribution included and observed sea level variance. Note colormap and colorbar are on a log scale. right) NWA12/SPEAR variance ratios: total, ensemble spread (ens-var), and for each frequency bands.

EGUSPHERE-2026-3057
“A dynamical downscaling approach for the improved representation of sub-annual coastal sea level variability: the importance of shelf and slope ocean dynamics”
Steinberg, J.M. and Krasting, J.P. and Ross, A.C.

Response to Reviewers - RC3 - Referee #3

This manuscript presents a useful comparison of the representation of coastal sea-level variability along the U.S. East and Gulf coasts in a high-resolution regional ocean model (NWA12) and its coarser parent model (SPEAR). The 20-year, five-member ensemble is a clear strength, and the paper addresses an important problem for coastal sea-level modeling and potential forecasting applications. I found the study interesting and generally well motivated, and I hope that the framework is used more extensively going forward.

The manuscript occasionally makes stronger interpretive and forecasting claims than the analysis fully supports. In particular, the paper emphasizes improved predictability or forecast relevance, but it does not formally evaluate forecast skill (or even correlations/RMSE for hindcasts, which I think is an easily achievable improvement). I also think the mechanistic interpretation relies too heavily on alongshore winds without fully isolating causality (or its relationship to the spectral analyses), and the manuscript would benefit from a more quantitative treatment of uncertainty and ensemble spread.

Thank you for this review and summary. We have attempted to address your concerns below and recast language referencing predictability to assessment of hindcast importance.

Major comments

1. Forecast-skill language should be more carefully qualified.

The paper does show improved hindcast fidelity and more realistic variability in several respects, but it does not demonstrate event-by-event forecast skill. I recommend that the authors soften the language throughout the abstract, discussion, and conclusions so that the distinction between hindcast fidelity and forecast skill is explicit.

Suggestions for the authors: replace broad forecasting claims with phrasing such as “forecasting-relevant hindcasts” or “frameworks that may inform forecasting". Also, see later suggestion about correlation analyses that can enhance the paper's relevance to forcasting.

We now reference retrospective forecasts as hindcasts. Done L(59-61, 62-66)

“With the aim of improving simulations and, in the future, forecasts of sub-annual coastal sea levels, we leverage a recently developed dynamical downscaling framework to explore ocean and atmospheric influences along the U.S. East Coast as they depend on model horizontal resolution.”

“Our central focus is realized through comparison of retrospective forecasts, or hindcasts, of coastal sea level in a parent global atmosphere-ocean forecast model, the Seamless System for Prediction and Earth System Research (SPEAR) model \citep{delworth2020}, and a higher horizontal resolution regional ocean model of the Northwest Atlantic (NWA12, \cite{ross2023}) forced at the boundaries by SPEAR.”

2. The mechanistic interpretation is too strongly centered on alongshore winds without fully demonstrating causality (or its relationship to the spectral analyses).

The Methods emphasize alongshore wind stress as a key diagnostic. In the Results, the discussion of Atlantic City and Charleston again leans on wind-stress regressions and correlations (p. 8-9, lines 175-195). The Conclusions then state that the observed differences are attributed to the “differing resolution-dependent ability of along-shore winds to excite cross-shore transport and along-shore wave propagation” (p. 16, lines 257-260).

This interpretation is plausible, but I do not think the analysis as presented isolates winds as the dominant mechanism. The title and abstract emphasize shelf and slope ocean dynamics, and the paper itself also discusses boundary currents and shelf-slope structure, yet the quantitative attribution remains limited. In my view, the current wording risks sounding more definitive than the evidence warrants.

Suggestions for the authors: as much as possible, add quantitative evidence such as variance-explained diagnostics, coherence/cross-spectral analysis, or a partial attribution framework that separates wind forcing from shelf/slope and boundary-current effects.

We have revised our language throughout the manuscript to soften the claim that NWA12-SPEAR differences can be solely explained and understood as a different response to along-shore winds (L107-109). This was not our intent.

“Even so, and despite the decreased frequency at which the ocean-only model is forced by the atmosphere, we find enhanced NWA12 variability that we hypothesize arises partly as a result of winds acting on a higher-resolution ocean.”

We have added a fraction of variance explained diagnostic to Figure 8 and Figure 9 that shows, at zero lag, how much coastal variability at Atlantic City and Charleston can be explained by local and remote along-shore winds. We now discuss this addition in Section 3.3 (L213-216, 223-230).

“Where these NWA12 regression coefficients are elevated, the fraction of coastal sea level variance explained increases relative to SPEAR by 10-20\% (size of colored circles in Fig. \ref{f8}). This increase is largest in the 5-day to 2-week band and at longer timescales SPEAR and NWA12 fractions are comparable.”

“South of Cape Hatteras, at Charleston, wind stress anomaly regression coefficients and fraction of variance explained are also elevated in NWA12 compared to SPEAR (Fig. \ref{f9}). The most significant increases in fraction of variance explained, over 25\%, occur in the SAB in the 5-day to 2-week band and in the MAB in the 5-week to 8-week band. In NWA12, this shift to the MAB with increasing timescale suggests, as in \cite{little2026}, connectivity across Cape Hatteras that is limited or absent in SPEAR due to model horizontal resolution.”

3. The paper would be stronger with a more quantitative uncertainty analysis.

The ensemble design is an important strength, but most of the analysis is presented through ensemble means and qualitative comparisons. This is evident in Figures 5, 10, 11, and 12, where the ensemble-mean behavior is emphasized, but uncertainty bounds, confidence intervals, or ensemble spread are not clearly summarized in the text. The Results do note that the spectral and variance differences appear across all ensemble members or multiple initialization years (e.g., p. 7, lines 150-155), but the manuscript does not show the degree of consistency in a compact quantitative way.

Suggestions for the authors: add a table reporting station-level metrics such as mean spectral ratio, variance ratio, correlation/coherence, and/or ensemble spread, together with confidence intervals or another uncertainty measure for the key comparisons.

For tide gauge station spectra (Figure 4) we now show ensemble spread as shaded bands. We also include a new figure (Figure 7) that reports SPEAR-NWA12 variance ratios, ensemble spread, and variance comparisons to observation. Text is added to discuss these new additions L(166-167, 189-198).

“Across these frequencies SPEAR and NWA12 ensemble spreads minimally overlap suggesting distinct improvement.”

“At the fifteen tide gauge stations considered here, SPEAR and NWA12 variance differences are explored more closely and alongside station depth differences and observed variance (Fig. \ref{f7}). At nearly all stations, coastline-adjacent NWA12 grid cells are shallower that those in SPEAR (Fig. \ref{f7}, left). These differences, greatest north of Cape Hatteras and along the Florida coastline, contribute to a decrease in SPEAR coastal sea level variance arising from less coastal amplification or shoaling. Diagnosed IB contributions compare favorably to reanalyses and when added to model coastal sea level variance, NWA12 values approach observed values (Fig. \ref{f7}, middle). Ratios of NWA12 to SPEAR variance (Fig. \ref{f7}, right) reveal a five-fold or greater increase in coastal sea level variance in this region that is largely constrained to shorter timescales. Further north, SPEAR and NWA12 stations show comparable variance suggesting more limited downscaling improvements in this region. South of Cape Hatteras, NWA12 increases appear linked to coastal station depth difference. These comparisons reveal the extent and pattern of NWA12 coastal variance increases and show that they approach and better match observed values.”

Specific comments
Abstract (p. 1, lines 7-17)

- “forecasting framework” should be softened. Perhaps "framework suitable for forecasting”.

In the abstract we reference development of the framework used in Ross et al. 2024 for forecast generation and assessment. We clearly state that we use this framework to generate and assess hindcasts (L1-2)

“Using a recently developed dynamical downscaling framework, we build and assess an ensemble of year-long sea level hindcasts for the western North Atlantic.”

- “particularly from an extremes perspective” would be clearer if the relevant timescales are stated directly.

Clarified is added to reference daily extremes (L2-4). More detailed definitions follow in the main text.

“Downscaled high-resolution regional ocean model (NWA12) output of U.S. East Coast sea level at long established tide gauge stations is compared to parent coarse-resolution global ocean model (SPEAR) output to quantify varied impacts of downscaling in a forecasting framework, particularly from a daily extremes perspective.”

- “nearly an order of magnitude” would be stronger if quantified more precisely.

Detail added and provided in the main body L(5-6).

“Comparisons of modeled coastal sea level distributions and observations at tide gauge stations reveal downscaling enhances variability by a factor of two to ten across all resolved frequencies.”

- “practical” is vague and should be replaced with a more specific description of the intended utility.

We rephrase language and use the word “applied” (L9-12).

“Together, these results demonstrate how dynamical downscaling can offer application-relevant time varying statistics of higher frequency coastal sea level variability developed, like planning based on inundation return periods, and an improved understanding of the links between model resolution and the processes driving coastal sea level changes.”

- add "that the" before “added"

Done.

- "developed with" -> and”

Corrected.

Introduction (p. 1-2, lines 19-30 and 32-35)

- The introduction, before line 48, could use a fair bit of refocusing. In particular, I think the second half of the first paragraph is misdirected, or at least confusing. The cited approaches (CORA/statistical) actually do account for coast-to-open ocean processes. It's the large-scale models that don't represent them. Furthermore, I don't agree with the necessity of resolving the coast-to-open ocean transition to gain meaningful skill. Maybe to understand the processes underlying such skill.

Thank you for this comment, we have now added clarification regarding cited approaches. While a product like CORA does resolve variability due to some coast-to-open ocean processes, it is a two-dimensional barotropic ocean circulation model, and as such it misses process representation and impacts of baroclinic forcing and variability. To address this, CORA assimilates coastal water level observations, but not any offshore. Compared to NWA12 it misses coastal variability caused by cross-shore changes in vertical density structure linked to forcing and boundary currents. While this maybe secondary to barotropic processes, it is missing. Your second point is well received and comments to this effect are added in the manuscript (L18-26).

“Different approaches have sought predictive skill through the use of: simplified barotropic models and data assimilation \citep{rose2024}, statistical approaches \citep{long2023}, and damped persistence frameworks \citep{dusek2022} that reduce system complexity or isolate specific processes. These efforts all reflect a desire and need to characterize, understand, and predict coastal variability, especially extreme event occurrences associated with flooding \citep{thompson2021}. What is often absent in these approaches, however, as a result reductions in system or model complexity, is full consideration or representation of three dimensional continental shelf and slope ocean dynamics that shape connectivity between the coast and larger-scale or remote ocean and atmosphere dynamics. Resolving this coast-to-open-ocean transition is necessary to gain a process-level understanding of the drivers of coastal sea level change which can bolster interpretation of forecast skill.”

Methods (p. 4-5, lines 85-125)

- Please state more explicitly how winds were interpolated to the NWA12 grid.

Detail is added (L88-91) and the reader is referred to Ross et al. 2024 for configuration details.

“As in \cite{ross2024}, daily-averaged SPEAR atmospheric fields of pressure, temperature, and humidity are interpolated from the 0.5° SPEAR atmospheric grid to the 1/12° NWA12 grid and used as surface boundary conditions, with the exception of surface winds, which are updated at 6-hourly resolution.”

Figure 3 and related discussion (p. 6, lines 130-137)

- Figure 3 appears important, but the discussion moves quickly from the time series to the spectra. I recommend including correlations, and possibly coherence or cross-spectral measures, to bridge the gap between visual time-series agreement and the later frequency-domain analysis.

We now enhance the discussion on Figure 3 and include a fourth tide gauge time series. We avoid correlation between SPEAR and NWA12 as well as with observations as phasings diverge from model initialization states. We do now add a variance comparison figure (Figure 7, comparing NWA12, SPEAR, and observations). Figure 3 is included to qualitatively show the time series used, methods described in Section 2.2, and to highlight generally the decrease in variance moving north to south. This figure is discussed alongside Figure 4 in Section 3.1.

- Is the light blue line really showing non-tidal residuals? What's causing the diurnal variability?

Yes, one source of residual diurnal variability may be wind and heat forcing near the coast. At northern stations considered here, coastal waves may carry this signal equator-ward locally reflect remote origins. At northern stations, general agreement between daily means computed from NWA12 non-tidal residuals and those in SPEAR suggest these higher frequency signals to be effectively smoothed out for subsequent analysis. We have chosen to remove these component of the plot as it may distract from the main focus of the rest of the manuscript, daily mean comparisons. Hourly spectra are still included in Figure 4.

Frequency spectra and band-passed variance (p. 7, lines 140-171)

- The frequency bands are useful, but they should be defined consistently and linked more explicitly to the physical interpretation. (Note lines 154 and 164 state two different frequency bands that we "expect" to resolve).

The comment regarding expected variability in SPEAR is generalized (L166-169). In section 3.1 we highlight two timescale ranges, the first are those timescales associated with variability resolved by both SPEAR and NWA12 and the second are timescales associated with variability only resolved by NWA12. The distinction is made to highlight that not all enhanced NWA12 variance derives from variability at newly resolved short timescales. We clearly state the frequency bands of interest at the start of Section 3.2 and now add physical interpretation (L176-180).

"Across these frequencies SPEAR and NWA12 ensemble spreads minimally overlap suggesting distinct improvement. These differences between SPEAR and NWA12 span a range of frequencies including those which SPEAR is expected to resolve, as well as those associated with processes unexpected to be resolved by SPEAR due to its relatively coarse shelf resolution.”

“These bands were selected to partition variance across weather to seasonal timescales, the full span of resolved frequencies from these 1-year simulations. The 5-day to 2-week band was selected to contain variability linked to synoptic scale weather, while the others capture a range of sub-seasonal atmosphere and ocean variability. Importantly, this separation includes

frequencies expected to be well resolved by both SPEAR and NWA12 and frequencies that only NWA12 can resolve associated with tides and dynamics permitted due to higher horizontal resolution."

- The comment that NWA12 adds grid-scale variance through more realistic mesoscale eddy turbulence is plausible, but the impact of those eddies on coastal sea level is not clear.

This comment (L185-188) is referencing SPEAR-NWA12 differences away from the coast. Mesoscale turbulence representation in 1 degree and 1/12 degree ocean models is significantly different in and around the Gulf Stream (Chassignet et al. 2008). We do not explore the relationship between these eddies and coastal sea level.

- Strongly recommend adding a small table with summary statistics by station, and some interpretation.

Figure 7 is added along with interpretation (L189-198).

"At the fifteen tide gauge stations considered here, SPEAR and NWA12 variance differences are explored more closely and alongside station depth differences and observed variance (Fig. \ref{f7}). At nearly all stations, coastline-adjacent NWA12 grid cells are shallower that those in SPEAR (Fig. \ref{f7}, left). These differences, greatest north of Cape Hatteras and along the Florida coastline, contribute to a decrease in SPEAR coastal sea level variance arising from less coastal amplification or shoaling. Diagnosed IB contributions compare favorably to reanalyses and when added to model coastal sea level variance, NWA12 values approach observed values (Fig. \ref{f7}, middle). Ratios of NWA12 to SPEAR variance (Fig. \ref{f7}, right) reveal a five-fold or greater increase in coastal sea level variance in this region that is largely constrained to shorter timescales. Further north, SPEAR and NWA12 stations show comparable variance suggesting more limited downscaling improvements in this region. South of Cape Hatteras, NWA12 increases appear linked to coastal station depth difference. These comparisons reveal the extent and pattern of NWA12 coastal variance increases and show that they approach and better match observed values."

- Add uncertainty ranges, ensemble spread, or confidence intervals for key comparisons: spectral differences, uncertainty across ensemble members, significance of latitude-dependent changes, station-to-station variability

Ensemble spread has been added to Figures 4 and 13. Figure 7 is added that shows NWA12/ SPEAR variance fractions with a column that includes ensemble spread.

- Highlight site by site differences: Why is SPEAR so much better at Charleston (and maybe st petersburg) than elsewhere south of Hatteras?

Figure 7 now highlights site-by-site differences. We speculate that SPEAR performance at Charleston is linked to the model station's depth. This region of the South Atlantic Bight is wide enough that a SPEAR grid cell fits over the shelf and reports a very shallow depth (5m), similar to NWA12. As the shelf is narrower at Fernandina Beach, the adjacent SPEAR bathymetry is deeper than that in NWA12, due to averaging across one degree of longitude that spans shelf and slope depths (not just shelf depths). Variability at this deeper coastal grid cell is thus reduced due to decreased coastal amplification (L191-192).

"These differences, greatest north of Cape Hatteras and along the Florida coastline, contribute to a decrease in SPEAR coastal sea level variance arising from less coastal amplification or shoaling."

- shouldn't the observational spectra vary depending on start date? why not include that range?

As all time series considered span twelve months, the variability reflected in observational spectra should not depend on the start date of an extracted time series.

Wind-stress regression analysis (p. 8-9, lines 175-195)

- The alongshore-wind focus is interesting, but I suggest stating more clearly why this is the key diagnostic and what it can and cannot show.

This focus derives from previous results that demonstrate much coastal variability across the timescales considered here is wind driven (e.g. Wang et al. 2022 and Lentz 2024 - cited in Section 3.3). We do not claim it is the only relevant process or key diagnostic, as process relevance/importance varies with latitude. We suggest winds play a first order role, especially north of Cape Hatteras. Motivating text is added (L200-209).

"Sea level variability along continental shelves can be significantly influenced by local and remote along-shore winds driving across-shore Ekman transport and exciting coastal trapped waves \citep{wang2022}. Along the U.S. East Coast, winds are predominantly northwesterly and typically strongest in the winter, becoming light and variable in the summer. After removing a seasonal climatology to highlight a coastal sea level response to wind anomalies across daily to weekly timescales, daily SPEAR along-shore wind stress anomalies are regressed onto coastal sea level anomalies at Atlantic City, NJ, and Charleston, SC. This approach was adopted to reveal the spatial pattern and strength of a main driver of coastal sea level variability. As both SPEAR and NWA12 oceans see the same atmosphere, before calculation of relative wind stress, comparisons reveal differing coastal sea level response patterns to wind stress anomalies over the shelf that, while dependent on timescale, suggest the importance of continental shelf resolution."

- Interpret the regression coefficient in terms of added noise or weaker signal.

We add a variance-explained metric in the regression plots (Figures 8 and 9). If variance explained is positive, the size of the scattered point scales with fraction of variance explained (reference sizes are in the legend). Comparison of time series shows that in regions where regression coefficient is weak, what would be considered noise is this case is elevated.

-- Strongly recommend at least a variance-explained metric

Figures 8 and 9 now include a fraction of variance explained metric shown as the size of the scattered points. Considering the role of winds at Atlantic City, NWA12 shows a modest increase in fraction of variance explained relative to SPEAR, along the Mid Atlantic Bight and further north, across all frequencies. The jump also appears when considering Charleston, with short timescale increases localized to the South Atlantic Bight and longer timescale increases north of Cape Hatteras L213-216, 223-230).

"Where these NWA12 regression coefficients are elevated, the fraction of coastal sea level variance explained increases relative to SPEAR by 10-20\% (size of colored circles in Fig. \ref{f8}). This increase is largest in the 5-day to 2-week band and at longer timescales SPEAR and NWA12 fractions are comparable."

"South of Cape Hatteras, at Charleston, wind stress anomaly regression coefficients and fraction of variance explained are also elevated in NWA12 compared to SPEAR (Fig. \ref{f9}). The most significant increases in fraction of variance explained, over 25\%, occur in the SAB in

the 5-day to 2-week band and in the MAB in the 5-week to 8-week band. In NWA12, this shift to the MAB with increasing timescale suggests, as in \cite{little2026}, connectivity across Cape Hatteras that is limited or absent in SPEAR due to model horizontal resolution."

-- line 195: this seems like a key point that remains ambiguous. With the addition of some analysis of the variance explained by wind I think the figure will be sufficient to look at differences between spear and NWA. However, the increasing and more uniform regression coefficients as a function of time scale is going to be really tough to disentangle from the changing (increasing spatial coherence) of winds. I wonder if there is a better way to do the analysis, or to have a supplementary analysis of the fingerprint of wind/spatial decorrelation scale at lower frequencies.

Thank you for this comment. We have revised the analysis to include a variance explained metric. As you note, we agree that as timescale increases, so too does the spatial coherence of the winds (for example, at 8 weeks, the winds offshore of Atlantic City and Charleston correlate at ~0.38, at 2 weeks, this correlation is ~0.17). This is information we wish to convey.

Seasonal distributions and extremes (p. 13, lines 217-240; p. 16-17, lines 251-273)

- The section would benefit from a clearer purpose statement at the start, and some connection to the spectral (and if added, other time-series analyses).

Done (L253-255).

"Metrics such as frequency of occurrence, extreme event likelihood, or rate of return all derive from distribution characterizations that vary in space and time, but if properly assessed offer sources of predictability."

Discussion and conclusions (p. 17, lines 266-281)

- The ending would be stronger if it concluded with one or two concrete implications for coastal forecasting and flood-risk applications. For example, what are the implications for a statistical transfer function? Minimum resolution? Are results generalizable to other coastlines? Etc.

We do not have a Discussion section in this paper and the Conclusions section begins in the submitted manuscript on L238. We now add some next steps discussion regarding application in a true forecast framework (L318-321)

"Next steps focused on generating statistics-oriented forecasts to be used in applied settings are feasible as SPEAR seasonal forecasts are generated seasonally, bias correction is not needed as ocean anomalies are applied to climatology, and single ensemble member NWA12 run times are only a few hours."

- Operational relevance should be stated carefully and conservatively.

We do not directly address operational relevance in our initial submission.

- I would replace "dynamically downscaled forecasts" with "dynamically downscaled simulations" unless the paper directly evaluates forecasts.

This language has been corrected. We now reference "dynamically downscaled hindcasts"

There are quite a few typos that should be cleaned up -- I have noted them in some cases.

- “underpredction” should be corrected to “underprediction.”

- “frequencieis” should be corrected to “frequencies.”

- “conclusiosn” should be corrected to “conclusions.”

- Please check for other minor grammar issues such as missing articles, extra “that,” and inconsistent capitalization of place names.

We cannot find any of the typos referenced above. “underpredction”, “frequencieis”, “conclusiosn”, and extra “that”s do not appear in the submitted text.

Figure captions are relatively terse. if possible, please make captions more self-contained and define abbreviations.

We have added detail to figure captions.

Final assessment
I think the paper addresses an important and timely question, and the ensemble hindcast experiment is a real strength. My concerns are justified, but they can be addressed with modest revisions: more careful language about forecast skill, stronger mechanistic support for the wind-based interpretation, and a more quantitative uncertainty analysis. With those changes, the paper would read as more rigorous, more balanced, and more convincing.

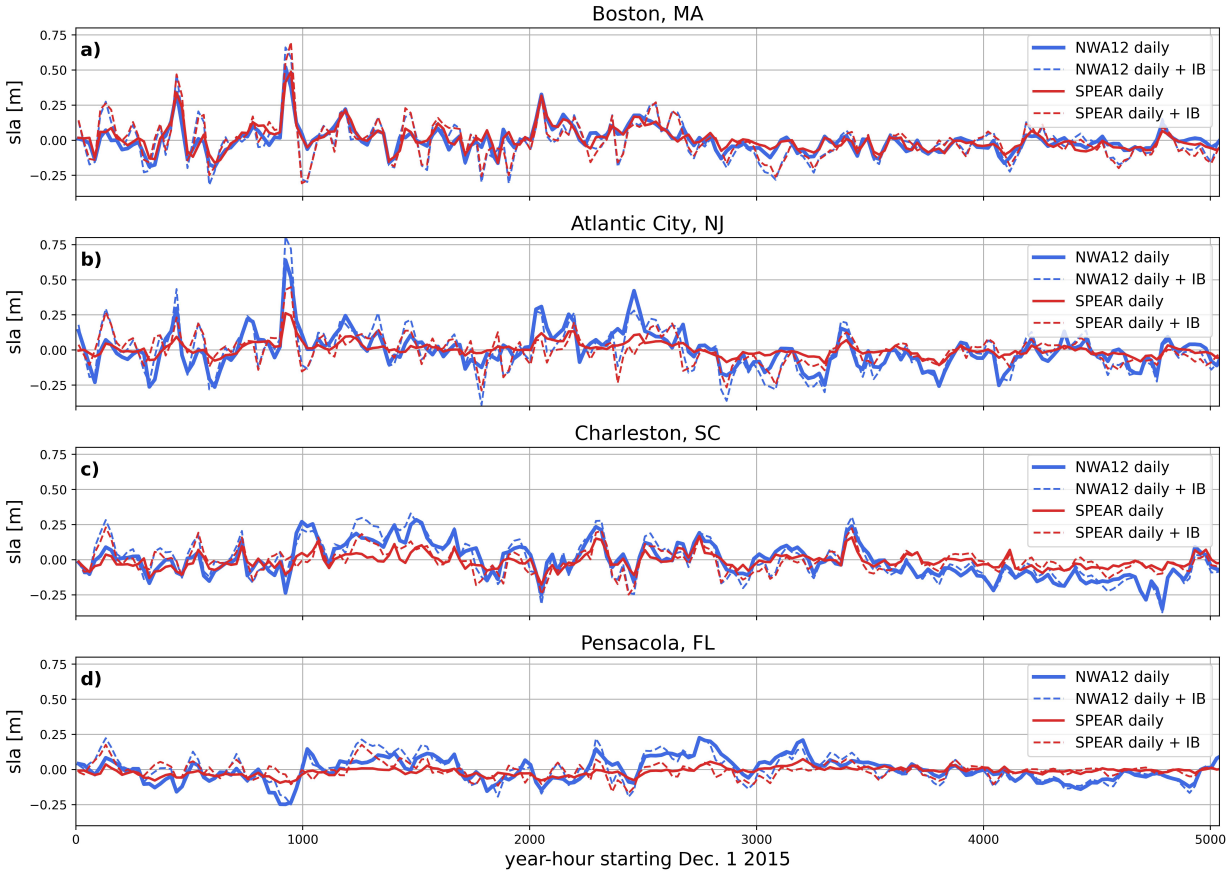


Figure 3 - SPEAR (red) and NWA12 (blue) coastal sea level at Boston, MA (a), Atlantic City, NJ (b), Charleston, SC (c), and Pensacola, FL (d) tide gauge stations for the 2015-12-01 -- 2016-06-30 period. Tides and a twenty-year mean season cycle have been removed. Hourly non-tidal residuals in NWA12 are averaged in 24-hour windows to obtain daily means. Solid and dashed lines are daily averages without and with the diagnosed IB contribution.

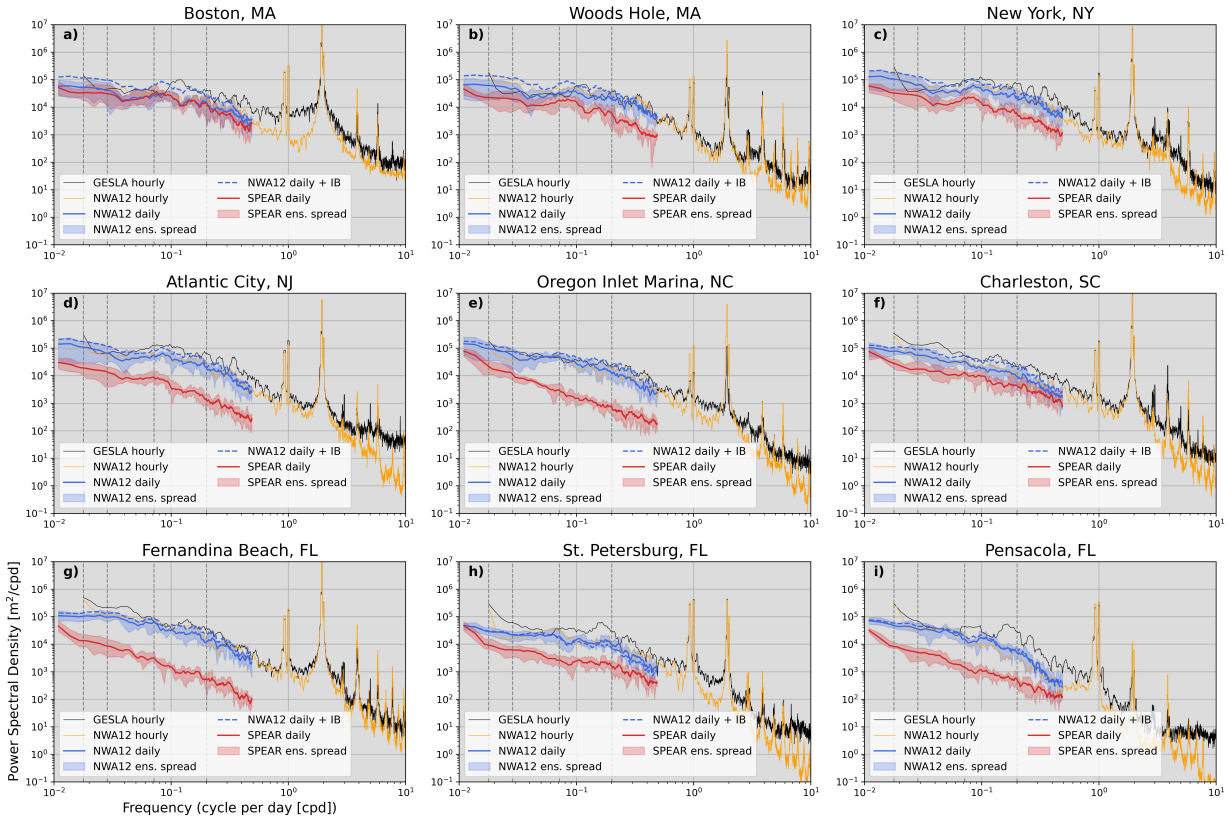


Figure 4 - Power spectral density at nine tide gauge stations. Black denotes observed and yellow NWA12 hourly spectra for the 2018-12-01 simulation start time. Solid red (SPEAR) and blue (NWA12) are ensemble-mean daily spectra with the blue dashed line adding the IB contribution. Red and blue shading encompass the ensemble spread. Moving left to right, light gray vertical dashed lines correspond to 2-month, 1-month, 2-week, and 5-day periods.

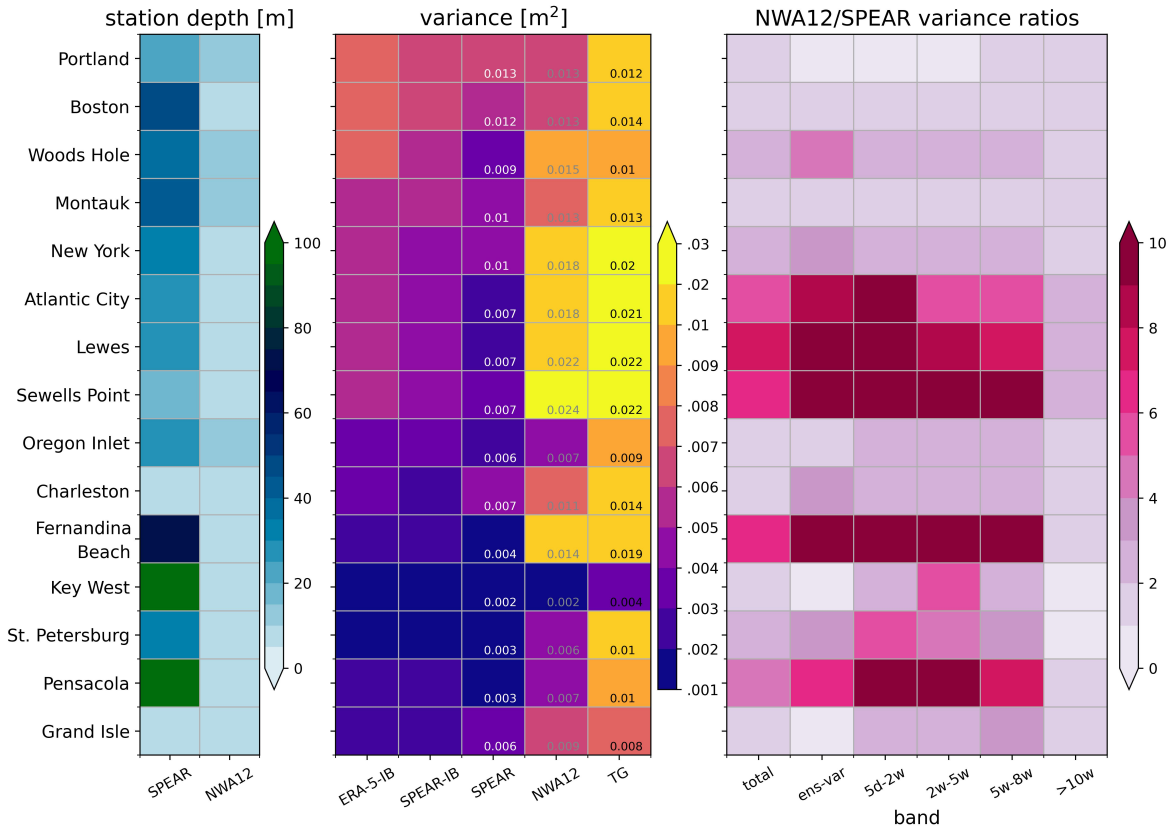


Figure 7 - left) SPEAR and NWA12 ocean depth at 15 tide gauge stations. middle) Coastal sea level variance: ERA-5 inverted barometer contribution (ERA-5-IB), SPEAR inverted barometer contribution (SPEAR-IB), SPEAR, NWA12, tide gauge observations (TG). Text reports model sea level variance with the IB contribution included and observed sea level variance. Note colormap and colorbar are on a log scale. right) NWA12/SPEAR variance ratios: total, ensemble spread (ens-var), and for each frequency bands.

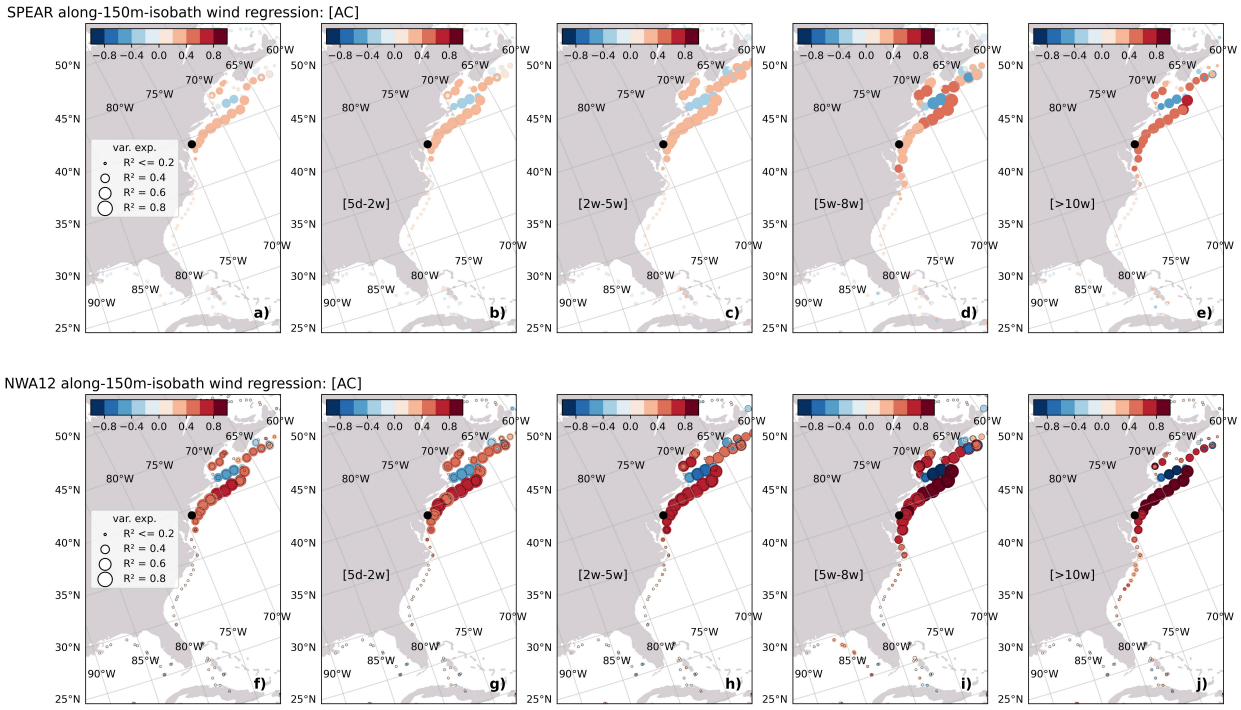


Figure 8 - Regression of along-150 m isobath winds with SPEAR total SLA at Atlantic City, NJ (black-circle) (a) and as a function of frequency band (b-e). f-j) Same as a-e but for NWA12. Colored circle size corresponds to the fraction of variance explained by along-isobath winds. In f-j, the open black circles are the fraction of variance explained in a-e.

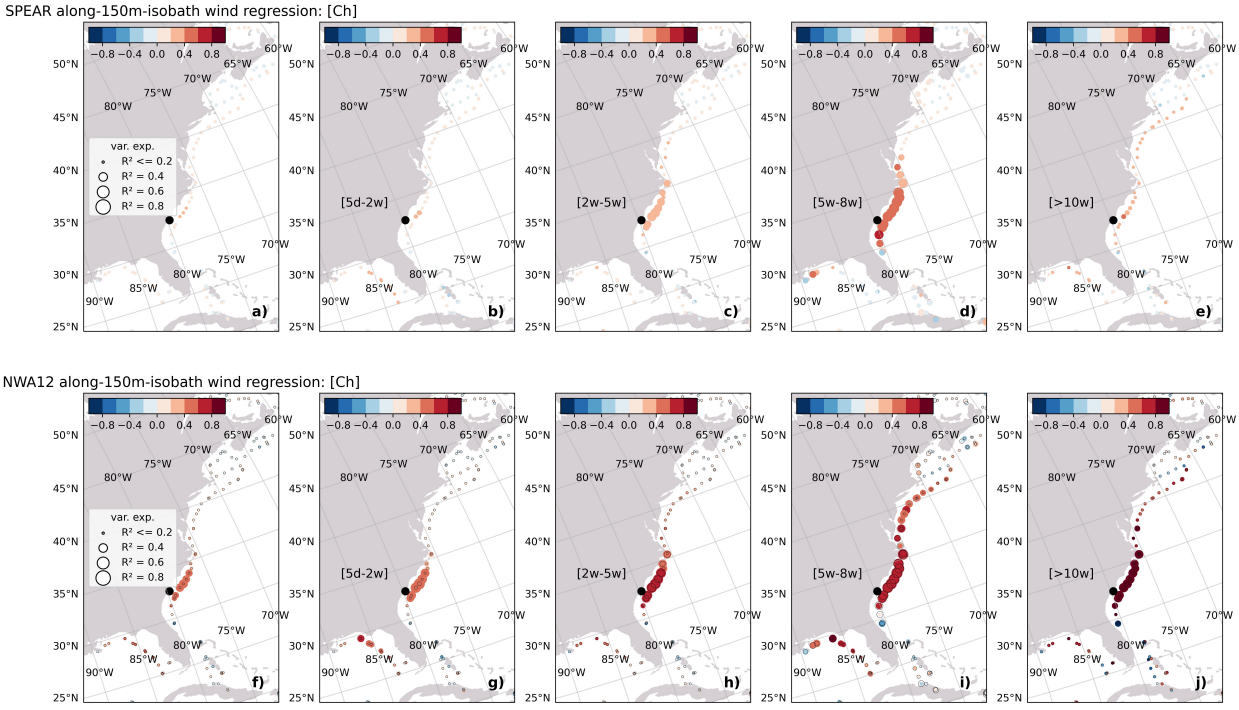


Figure 9 - Same as in Figure \ref{f8}, but for Charleston, SC.

EGUSPHERE-2026-3057
“A dynamical downscaling approach for the improved representation of sub-annual coastal sea level variability: the importance of shelf and slope ocean dynamics”
Steinberg, J.M. and Krasting, J.P. and Ross, A.C.

Response to Reviewers - RC4 - Referee #4

An analysis of dynamically downscaled seasonal predictions from SPEAR global seasonal predictions using regional MOM6 is presented. The downscaled predictions are compared with the native global predictions to show the benefit of downscaling. Improved spatial resolution in MOM6 allows realistic bathymetry and an enhanced shelf response to overlying winds. These improvements then lead to improved coastal sea level characteristics in the downscaled predictions. The presented analysis shows mechanistic underpinnings of the open-ocean to coastal dynamical transition which I believe adds to our understanding of the added value of dynamical downscaling. The manuscript is well composed and focused. I have only minor comments. It would help the readers if the authors consider the following comments and suggestions regarding their manuscript:

Methodology:
How were the SPEAR seasonal retrospective predictions generated? The authors, at multiple instances (Lines 119-120, 233-234), state that the winds in both SPEAR and the regional downscaled model are the same and hence any improvement in the quality of hindcasts has to be due to other factors or to the differing response to the same wind field. But, I believe the authors are aware that SPEAR is a coupled model and the SPEAR ocean would feed back onto the SPEAR atmosphere. Such feedback is, by design, absent in the regional MOM6 retrospective predictions. The authors point this out in Lines 97-100 but then still assert that both SPEAR ocean and MOM6 ocean see the same winds. I would suggest that the authors be careful about their statement that the atmosphere in SPEAR and the sampled (6-hourly or daily) atmospheric forcing from SPEAR, with which MOM6 was forced, are identical and hence the consequences on their interpretation of results.

Thank you for this comment. We have revised language in the manuscript. As you note, NWA12 is missing atmosphere-ocean feedbacks. We have added detail to further describe this configuration. Our response to Referee 2 is repeated below.

The NWA12 regional model configuration use here calculates the relative wind to determine the stress to apply to the ocean; in other words, the NWA12 ocean surface current was subtracted from the SPEAR winds being applied. We acknowledge that the use of relative wind forcing in a high resolution uncoupled ocean model can often have an overly strong eddy killing effect (e.g., Renault et al. 2020). In recent experiments with a different atmospheric forcing product, we found that using a correction similar to Renault et al. 2016 to make the wind forcing closer to absolute rather than relative to the current increased the eddy kinetic energy in the model. Therefore, we expect that the increased sea level variance in NWA12 would be boosted even further if this correction was included. To address this comment, we have inserted a description of how the wind forcing was applied into the methods (L91-93).

“While surface ocean conditions do not feed back onto the atmosphere, wind stress calculations in NWA12 account for relative air-sea velocity, with NWA12 surface currents subtracted from the SPEAR surface winds.”

Can the authors add an objective analysis to support that five ensemble members are sufficient for their seasonal prediction framework in the context of Figures 10, 11 and 12?

We now show ensemble spread in Figures 4 and 13. Excluding northern stations, the lack of overlap between SPEAR and NWA12 shaded regions suggest distinct drivers of variability. Plans for running forecasts include increasing the number of ensemble members to gain statistical power. Here, we determined five ensemble members was sufficient to identify SPEAR and NWA12 differences relative to internal variability. Comment added (L320-321).

“For single year forecasts, we plan to generate a larger ensemble to enhance statistical robustness. Five ensemble members were chosen here to focus on parent and downscaled model differences relative to internal variability.”

Results:

In Figures 6, 7, 8 and 9, observations-based maps and differences are not shown. The regression pattern in Figure 7 and 8 seems to be same in SPEAR and MOM6 with only the magnitude of the regression coefficients changing. It would help to see how much underestimation does SPEAR have and how much under- or overestimation, if any, does MOM6 have. Is the enhanced response of MOM6 to SPEAR-derived winds realistic?

We have added a new figure (Figure 7) which compares SPEAR, NWA12, and observed variance, including NWA12/SPEAR variance ratios, for each tide gauge station. Figures 4 and 5 highlight SPEAR and NWA12 representation of coastal sea level variability across timescales and relative to observations (comparisons made at nine tide gauge stations). These frequency spectra show SPEAR to have less energy across all but the northern-most stations and at lowest frequencies, indicating the pattern and extent of SPEAR under-estimation. The pattern of these differences is extended to all 15 the tide gauge stations in Figure 7 and assessment of over-/under-estimation relative to observations is included (L189-198).

“At the fifteen tide gauge stations considered here, SPEAR and NWA12 variance differences are explored more closely and alongside station depth differences and observed variance (Fig. \ref{f7}). At nearly all stations, coastline-adjacent NWA12 grid cells are shallower that those in SPEAR (Fig. \ref{f7}, left). These differences, greatest north of Cape Hatteras and along the Florida coastline, contribute to a decrease in SPEAR coastal sea level variance arising from less coastal amplification or shoaling. Diagnosed IB contributions compare favorably to reanalyses and when added to model coastal sea level variance, NWA12 values approach observed values (Fig. \ref{f7}, middle). Ratios of NWA12 to SPEAR variance (Fig. \ref{f7}, right) reveal a five-fold or greater increase in coastal sea level variance in this region that is largely constrained to shorter timescales. Further north, SPEAR and NWA12 stations show comparable variance suggesting more limited downscaling improvements in this region. South of Cape Hatteras, NWA12 increases appear linked to coastal station depth difference. These comparisons reveal the extent and pattern of NWA12 coastal variance increases and show that they approach and better match observed values.”

I am not fully convinced with the reasoning to completely exclude the established deterministic and probabilistic skill metric to evaluate both SPEAR and the downscaled predictions against observations. What information do Figures 10, 11 and 12 convey that can not be summarized into a skill metric at each lead time? Is it the size of your ensemble that would prevent any meaningful estimate of the the skill, for example, of short-term extremes in coastal sea level?

After forecast initialization, synoptic atmospheric variability responsible for much of coastal sea level variance at the timescales considered here decouples from observed variability. As we remove a seasonal cycle, the SPEAR and NWA12 simulated coastal sea levels considered here are shaped by ocean and atmospheric processes that evolve away from observations-informed initial conditions. Model and observed behaviors diverge from initial conditions, with unique phasing of variability, resulting in distinct coastal sea level variability. We feel that focus on a

deterministic skill metric would take away from the manuscript's main consideration of coastal sea level variability as a function frequency, model resolution, and in a distributional assessment sense. Similarly, we do not seek to determine skill in extreme event occurrence, but rather provide a characterization of distribution shape evolution that can be associated with event likelihood (this is information we wish to convey in Figures 11-13).

Figures:
Figure 1: Are the NWA12 maps from the same 30-year ERA-forced simulation from which the initial conditions were derived for the downscaled predictions? Please add this detail in the figure caption.

Clarification added. In this figure, we show annual means of daily averages spanning the year 1999. This comparison is made to give general sense of regional mean SSH structure and variance along the coast and associated with the Gulf Stream.

Mention the frequency band in the caption for Figure 9

Done.

Mention that IB effect is included in SPEAR and NWA12 in Figure 12 caption

Added. Figure 13 has been edited to include ensemble mean 5th and 9th percentile values of SPEAR and NWA12 as well as ensemble spread. In our initial submission we mistakenly did not include IB contributions (original Figure 12).

All figure legends, axis labels are too small to read on an A4 page, which some of us still read from.

We have increased the font size of all figure titles, axis labels, and legends wherever possible.

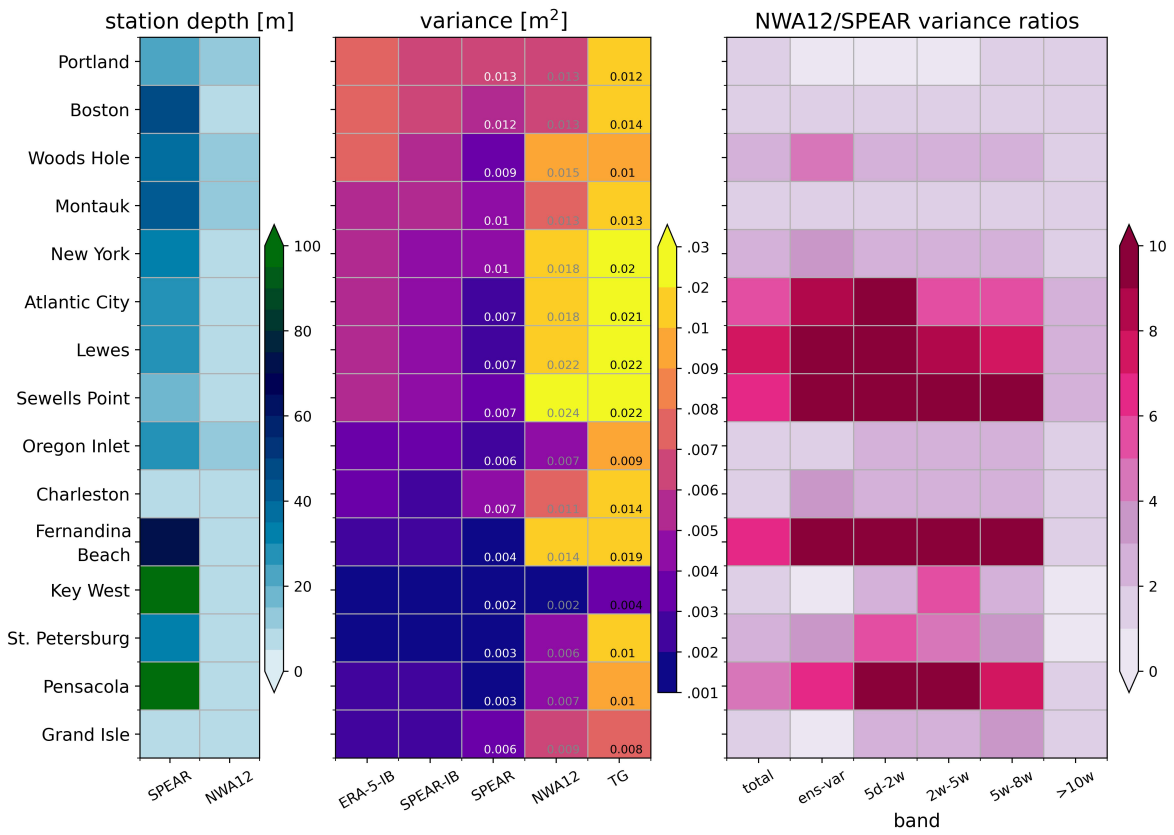


Figure 7 - left) SPEAR and NWA12 ocean depth at 15 tide gauge stations. middle) Coastal sea level variance: ERA-5 inverted barometer contribution (ERA-5-IB), SPEAR inverted barometer contribution (SPEAR-IB), SPEAR, NWA12, tide gauge observations (TG). Text reports model sea level variance with the IB contribution included and observed sea level variance. Note colormap and colorbar are on a log scale. right) NWA12/SPEAR variance ratios: total, ensemble spread (ens-var), and for each frequency bands.

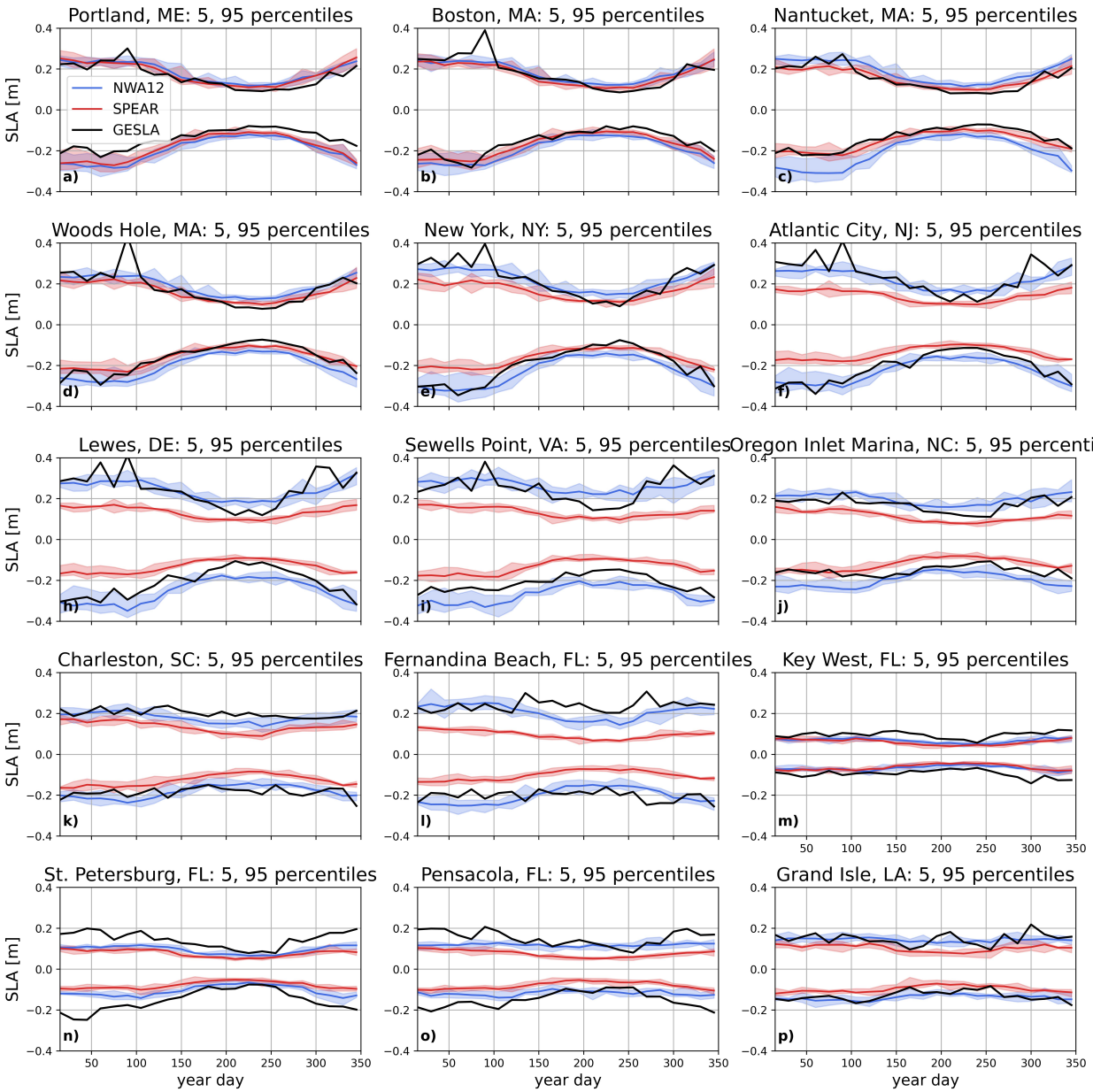


Figure 13 - 5th and 9th percentiles of daily sea level anomaly distributions as a function of simulation lead time (days since model initialization). SPEAR is in red, NWA12 in blue, and observations (GESLA) in black. Red and blue shading shows the mean ensemble spread.