

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 $(\rho_0 \cdot g)$ 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 than 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 $\times$ ~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

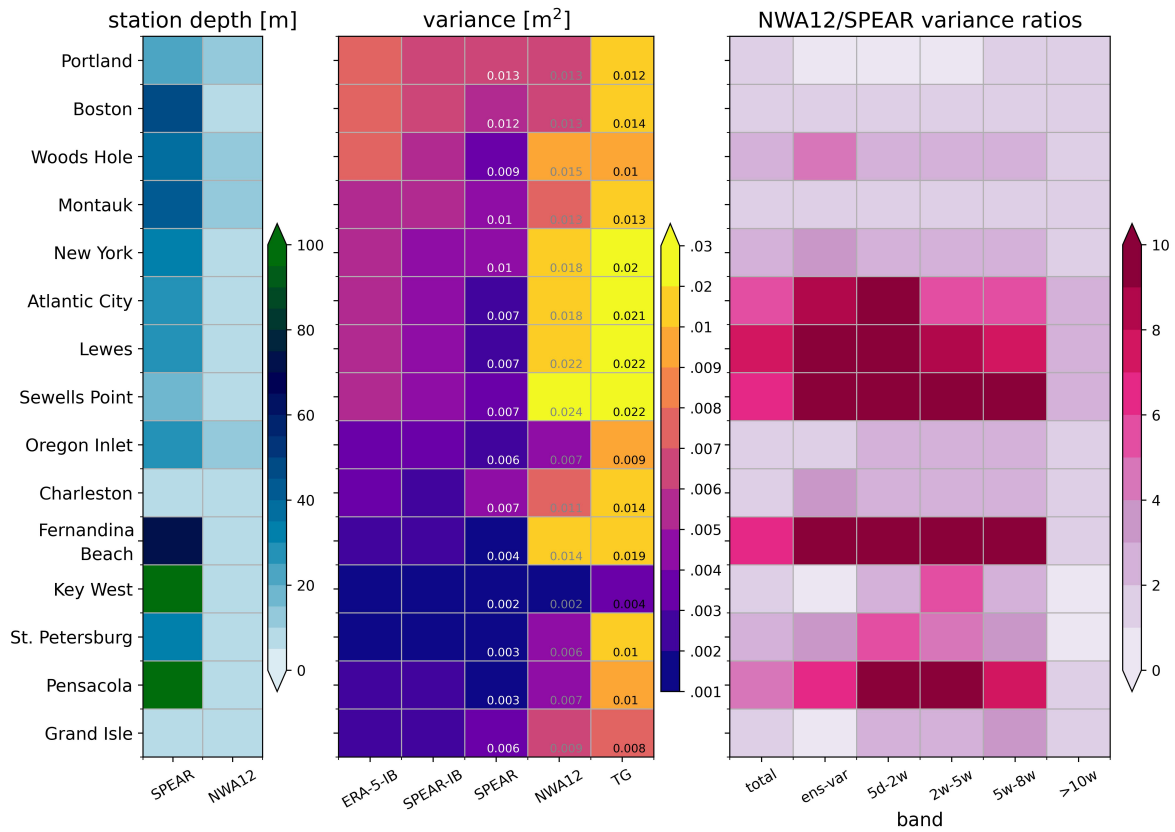


Figure 7 - left) SPEAR and NWA12 ocean depth at 15 tide gauge stations. middle) Coastal sea level variance: ERA-5 inverted barometer contribution (ERA-5-IB), SPEAR inverted barometer contribution (SPEAR-IB), SPEAR, NWA12, tide gauge observations (TG). Text reports model sea level variance with the IB 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.

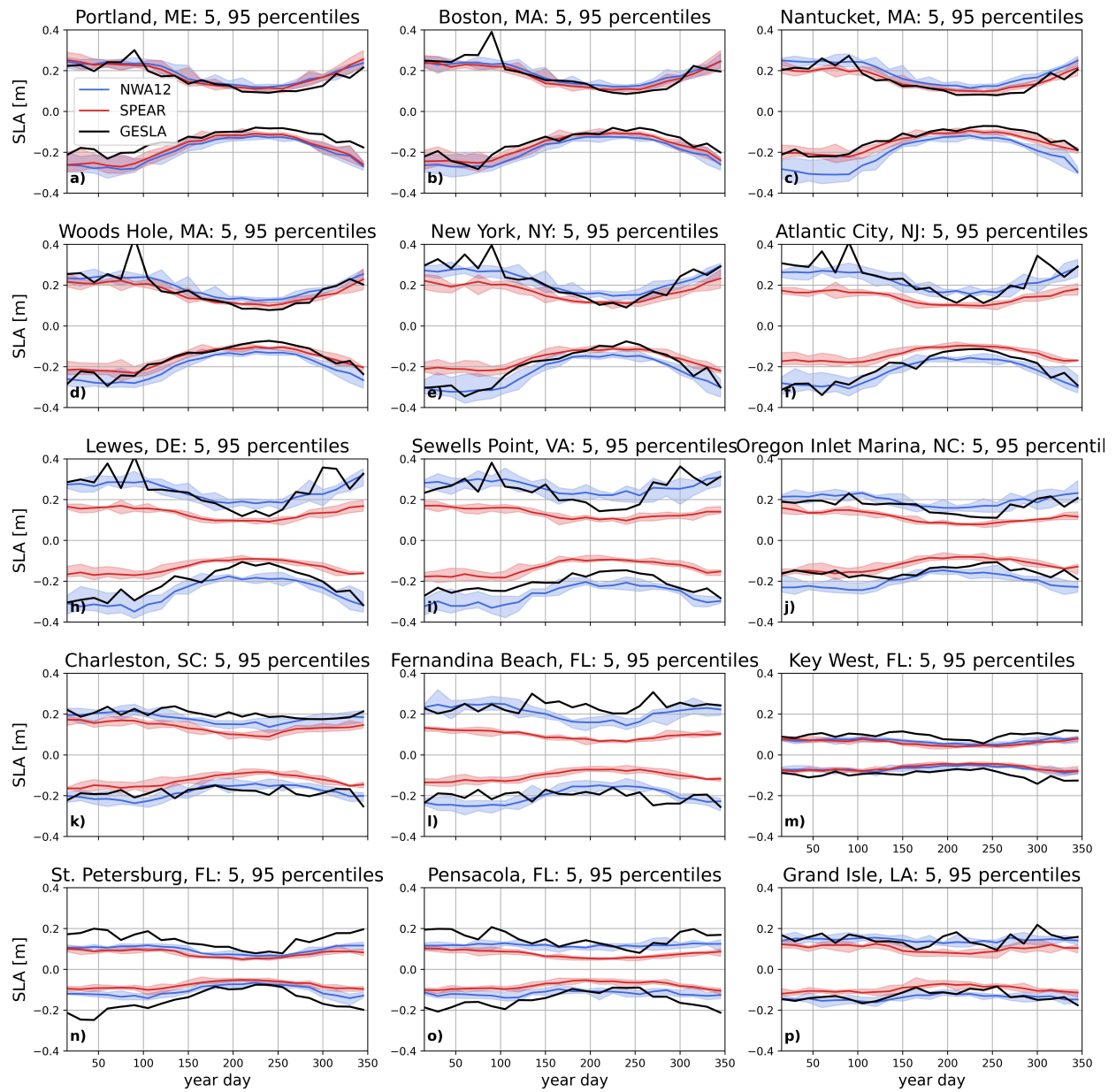


Figure 13 - 5th and 9th percentiles of daily sea level anomaly distributions as a function of simulation lead time (days since model initialization). SPEAR is in red, NWA12 in blue, and observations (GESLA) in black. Red and blue shading shows the mean ensemble spread.