Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL091886

Seasonal Predictability of Global and North American Coastal Sea Surface Temperature and Height Anomalies

Sang-Ik Shin[1] and **Matthew Newman**[1]

[1]CIRES, University of Colorado Boulder & NOAA Physical Sciences Laboratory, Boulder, CO, USA

Key Points:

- A Linear Inverse Model (LIM) is constructed that can predict monthly mean global sea surface temperature (SST) and sea surface height (SSH) anomalies
- The LIM outperforms operational numerical models for SSH skill in the Atlantic and US East Coast tide gauge stations
- LIM coastal skill is largely due to capturing ENSO teleconnections impacting the West Coast, and Gulf Stream modulation for the East Coast

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S.-I. Shin,
Sangik.Shin@noaa.gov

Citation:

Shin, S.-I., & Newman, M. (2021). Seasonal predictability of global and North American coastal sea surface temperature and height anomalies. *Geophysical Research Letters*, *48*, e2020GL091886. https://doi.org/10.1029/2020GL091886

Received 28 NOV 2020
Accepted 7 APR 2021

Abstract A Linear Inverse Model (LIM) is constructed to evaluate predictability of seasonal sea surface temperature (SST) and sea surface height (SSH) anomalies over the ice-free global ocean. Its ensemble-mean hindcast skill is also compared to that of the North American Multi-Model Ensemble (NMME) for 1982–2010. Both have similar skill for dominant modes of SST variability, but regional NMME SST skill is somewhat higher in many locations. However, the LIM has considerably more Atlantic and Southern Ocean SSH skill. Skill is generally comparable along the North American coastline, but LIM skill is greater for several highly productive coastal zones and East Coast tide gauge stations. Diverse, often predictable ENSO events drive teleconnections providing predictability in the North Pacific and along the US West Coast. Predictability in the Atlantic and along the US East Coast is associated with Gulf Stream strength modulation. Overall, the LIM shows potential for seasonal prediction of coastal ocean conditions.

Plain Language Summary Marine resource management requires a skillful forecast of coastal ocean conditions. Here, we developed and used an empirical prediction system to benchmark the operational system's seasonal prediction skill and understand the source(s) of predictability over the North American Coast. Both systems' skills are generally comparable, but the empirical system's skill is greater for several highly productive coastal zones and East Coast tide gauge stations. The results indicate that the remote impacts of ENSO provide predictability along the US West Coast. Predictability along the US East Coast is associated with Gulf Stream modulation. Overall, our empirical system shows potential for seasonal prediction of coastal ocean conditions.

1. Introduction

Coastal marine ecosystems are strongly impacted by changes in ocean circulation and stratification, which lead to the geographic redistribution of available nutrients and oxygen and, in turn, marine species (Capotondi, Jacox, et al., 2019; Doney et al., 2012). Therefore, skillful prediction of coastal ocean conditions is increasingly needed for marine resource management (Jacox et al., 2020). In fact, within the Large Marine Ecosystems (LMEs), relatively large coastal zones where primary productivity is higher than in the open ocean, sea surface temperature (SST) hindcast skill from the numerical models of the North American Multi-Model Ensemble (NMME; Kirtman et al., 2014) is large enough to be potentially useful (Hervieux et al., 2019; Stock et al., 2015). Recently, in a comprehensive review of seasonal-to-interannual prediction methods and their skill within North American coastal marine ecosystems, Jacox et al. (2020) suggested not only that prediction skill may be maximized via a hybrid statistical-numerical prediction approach, but also that statistical prediction alone could be competitive with numerical prediction.

This study explores North American coastal forecast skill using a global Linear Inverse Model (LIM; Penland & Sardeshmukh, 1995; hereafter PS95). The LIM is an empirical dynamical model that may diagnose potential spatiotemporal variations of skill, since its forecasts are often competitive with numerical model forecasts. For example, Newman and Sardeshmukh (2017; hereafter NS17) showed that a tropical LIM had similar tropical Indo-Pacific SST skill to the grand ensemble mean of the NMME operational models and that the LIM itself largely predicted the spatial and temporal skill variations of both models. LIMs have been constructed for other regions, including the Pacific and the Atlantic Ocean basins (Dias et al., 2018; Hawkins & Sutton, 2009; Hawkins et al., 2011; Huddart et al., 2017; Newman, Alexander, Ault, et al., 2016; Zanna, 2012; and many others). However, the LIM has rarely been applied globally, nor has its coastal skill explicitly been evaluated. Here, we do both, extending the approach of NS17 to cover most of the ice-free global ocean (60ºS–65ºN), yielding a global skill benchmark of the current generation of coupled climate

models that allows for estimating potential source(s) of predictability. We then specifically evaluate prediction skill along the North American coastline, focusing on monthly SST anomalies within LMEs and sea surface height (SSH) anomalies observed at numerous tide gauge stations, and diagnose how large-scale global dynamics impact the potential for such skill.

2. Methods

2.1. Data

We analyzed seasonal hindcasts for 1982–2010 made by nine NMME models (note that only five models provide SSH hindcasts; detailed in Table S1). Ocean observations, used to evaluate NMME hindcasts and construct the LIM, are monthly SSTs from the HadISST data set (Rayner et al., 2003) and SSHs from the ECMWF ocean reanalysis system ORAS4 (Balmaseda et al., 2013), during the period 1961–2010. All observations and NMME hindcasts were interpolated to a common 1° latitude × 1° longitude grid before analysis.

To determine the observed anomalies, we removed the climatological annual cycle at each grid point from the monthly SST and SSH observations. For the NMME hindcasts, each model's mean bias was corrected at each grid point by subtracting the mean difference between the model hindcasts and the observations for each target month and each forecast lead (Barnston et al., 2015). The multi-model ensemble mean (hereafter, NMME-mean) was then constructed by averaging all individual bias-corrected ensemble members of SST (SSH) within the available nine (five) models.

2.2. Linear Inverse Model

The evolution of an anomalous climate state vector $\mathbf{x}(t)$ may be approximated with the stochastically forced, stable linear dynamical system:

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \mathbf{S}\boldsymbol{\eta}, \tag{1}$$

where $\mathbf{L}\mathbf{x}$ represents predictable linear dynamics and $\mathbf{S}\boldsymbol{\eta}$ represents white noise forcing by the unpredictable rapidly decorrelating remainder (PS95). Determining (1) empirically from observed covariances yields a LIM, as described in PS95. LIMs are low-order models, typically constructed from the leading Principal Component (PC) time series of a reduced empirical orthogonal function (EOF) space. Given limited data availability, the matrices $\mathbf{L}$ and $\mathbf{S}$ are assumed state-independent; then, the infinite-member ensemble-mean forecast at lead τ becomes

$$\hat{\mathbf{x}}(t+\tau) = \mathbf{G}(\tau) \cdot \mathbf{x}(t), \tag{2}$$

where $\mathbf{G}(\tau) = \exp(\mathbf{L}\tau)$ is the linear system propagator.

Our LIM is a global extension of NS17 with these key differences: (i) The latitude range was extended from the Tropics to cover 60°S to 65°N; (ii) We used only SST and SSH anomalies within the state vector, whereas NS17 also used winds; (iii) Our EOF truncation captured less domain-integrated SST variance (62% here vs. 85% in NS17); and (iv) We used only HadISST for the SST data set. Other details of our LIM construction, including the "tau-test" of the linear approximation (PS95; Figure S1), are in the supporting information S1 and S2; see also NS17.

2.3. Hindcasts and Skill Evaluation

As in NS17, we used "ten-fold" cross-validation to evaluate the independent skill of ensemble-mean LIM SST and SSH hindcasts (see supporting information S1 for additional details). To match the start dates of the NMME-mean hindcasts, which were initialized on (or near) the first day of each month, each LIM hindcast was initialized with the monthly mean observations centered on the previous month. Therefore, the 1-month lead LIM forecasts were compared to the 0.5-months lead NMME forecasts, where both are named the "Month 1" forecast, and so on for increasing forecast leads.

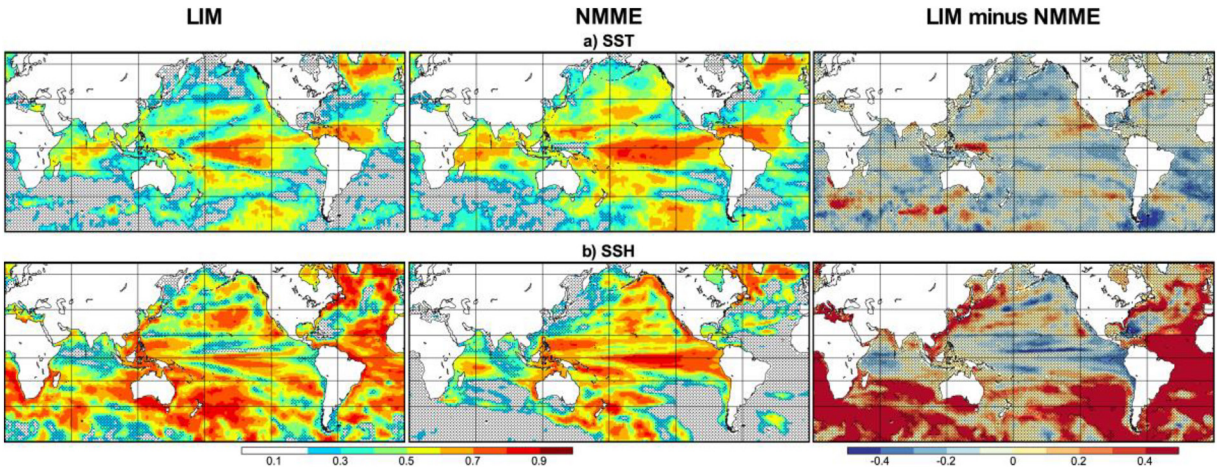

Figure 1. Maps of (a) SST and (b) SSH AC hindcast skill from the (left) LIM and (middle) NMME-mean at Month 6 during the period 1982–2010 (right) Difference maps of Month 6 AC skill during the same period. Dotting indicates skill and difference values that are insignificant at the 95% confidence interval.

Prediction skill was measured by local anomaly correlation (AC) and root-mean-square error based skill score (RMSSS; Barnston et al., 2015). We estimated 95% confidence intervals for AC and its differences using the two-tailed student's *t*-test. Since the sampling distribution of AC is generally not normal, the ACs were first converted to Fisher's *z*-statistic, and then the confidence interval was computed (Zwiers & von Storch, 1995). All significance tests were performed with domain-averaged effective degrees of freedom estimated at each grid point from observations during the hindcast evaluation period 1982–2010.

2.4. LIM Anomaly Growth and Predictability

Assessing LIM predictability is straightforward since (1) distinguishes between predictable signal and unpredictable noise. For any infinite-member ensemble-mean perfect model forecasts, the average AC between forecasts at the lead time τ and the corresponding verification is a function of the forecast signal-to-noise ratio (Sardeshmukh et al., 2000), which for the LIM is related to the predictable forecast signal covariance (Newman, Sardeshmukh, et al., 2003; NS17; see supporting information S3). Though LIM dynamics are stable, some anomalies undergo substantial growth over finite time intervals, yielding larger forecast signals and greater predictability. This growth is diagnosed through the Singular Value Decomposition of $\mathbf{G}(\tau)$, generating singular vector (SV) pairs where each right SV evolves into its corresponding left SV over the interval τ, with amplitude change determined by the corresponding singular value (Farrell, 1988; Lacarra & Talagrand, 1988; PS95; Vimont et al., 2014). The maximum increase in domain-integrated variance occurs when the initial state is proportional to the leading right singular vector, with maximum amplification factor equal to the square of the leading singular value. More generally, a subspace of growing SVs, associated with singular values greater than 1, together contribute to anomaly growth and related skill over a given time interval (Figure S2). Additional details are in the supporting information S4.

3. Hindcast Skill Evaluation

3.1. Global Skill

Figure 1 compares the Month 6 LIM and NMME-mean AC skill of predicted SST (top row) and SSH (bottom row) anomalies. Results at other leads (not shown) present similar patterns and lead to similar conclusions, as does the RMSSS (Figure S3). Overall, the LIM and NMME-mean have comparable SST skill patterns and amplitudes, although there are some notable regions of lesser LIM skill. For example, skill is quite similar in the central equatorial Pacific (see also Figure S4a) but not in the eastern equatorial Pacific [In fact, a Tropics-only LIM is also more skillful than our global LIM in the eastern equatorial Pacific (not shown, but see NS17), which might be a drawback of the global approach.] On the other hand, the global LIM still has higher skill than the NMME-mean in the far western equatorial Pacific, where the NMME-mean forecasts

cannot produce the correct anomalous SST horseshoe ENSO pattern (NS17), and in the subtropical eastern Pacific, extending up toward the US West Coast.

The global LIM provides substantial, but not uniform, extratropical SST skill. Both models are generally comparable in the North Atlantic, with the LIM even a little more skillful within the Gulf Stream region; both models also have similar skill for the Atlantic Multidecadal Oscillation (AMO) index (Figure S4b). In most of the South Atlantic, the NMME-mean skill is not high, but the LIM is worse. In the Pacific, both LIM and NMME-mean have similar skill patterns, especially in low and mid-latitudes, but the NMME-mean has somewhat higher values. Farther north, this skill difference increases, except along the west coast of North America where both models' skill is similar. Consequently, the difference in LIM and NMME-mean Pacific Decadal Oscillation (PDO) index hindcast skill is small (Figure S4c), with the LIM a little more (less) skillful than the NMME-mean for forecasts made for the spring (winter) months.

The LIM also provides skillful SSH hindcasts. For example, Month 6 AC skill is higher than 0.6 in most regions. This contrasts with the NMME-mean, whose hindcast skill is largely limited to the ENSO-impacted Indo-Pacific and along the west coast of North America, where skill somewhat exceeds the LIM. The absence of skill elsewhere for the NMME-mean, especially in comparison to the LIM (Figure 1b), is particularly stark throughout much of the Atlantic and Southern Oceans and within the North Pacific western boundary region.

3.2. North American Coastal Skill

Both models generally have comparable skill along much of the world's coastlines (Figure 1), but here we will focus upon North America. We repeated the Hervieux et al. (2019) analysis, comparing LIM and NMME-mean SST skill within the 11 North American LMEs (see map in Figure 2). Figure 2 shows LIM skill for each LME as a function of verification month and forecast lead, alongside the LIM and NMME-mean skill difference (see Figure S5 for NMME-mean skill). In general, skill differences are small, statistically significant at the 95% confidence level for only a few LMEs and only specific seasons. For example, the NMME-mean is more skillful in the Aleutian Islands, East Bering Sea, and Gulf of Alaska LMEs, but not uniformly, and significantly only for winter verifications. In contrast, the LIM is more skillful within regions strongly influenced by ENSO, such as the California Current and Gulf of California LMEs, although only significantly during spring. Interestingly, the LIM is also more skillful along the northeastern US coast, especially in the Scotian Shelf, perhaps because the Gulf Stream in some CGCMs tends to separate from the coast further north than observed (Griffies et al., 2015).

Next, we verified SSH skill against monthly sea-level observations at 10 tide gauge stations along the North American coast plus Honolulu (Figure 3), drawn from the Revised Local Reference (RLR; https://www.psmsl.org) data set during 1982–2010. Since the hindcast grid is too coarse to resolve the coastline, we used the nearest hindcast grid point for those stations outside the grid. The LIM is notably more skillful than the NMME-mean for all the Atlantic coastal stations, significantly during summer and fall for Charleston, Baltimore and Philadelphia (Pier 9N). The LIM has better skill at shorter leads for the Pacific coastal stations, especially for the southern stations (San Francisco and San Diego), but for increasing lead NMME-mean skill eventually surpasses the LIM's; these differences are not significant, however.

3.3. Assessing the Impact of the Secular Trend on Skill

In the LIM, the externally forced trend is captured by the least damped eigenmode of $\mathbf{L}$ (Frankignoul et al., 2017; Newman, 2013; Penland & Matrosova, 2006; see Figure S6). For seasonal leads, each LIM initialization includes a trend component comprised of this eigenmode's projection on each initial state. This trend component is then predicted to persist over the forecast period, apart from very weak damping with an e-folding time scale of 9 years.

In some regions and for some variables, secular trends may significantly enhance seasonal forecast skill (Ding et al., 2019; Huang et al., 1996). This is not necessarily dependent upon predicting the trend, since standard skill scores defined relative to a long-term fixed climatology are inflated simply by including the trend within the climate anomaly definition. SSH hindcast skill could be particularly impacted, since LIM

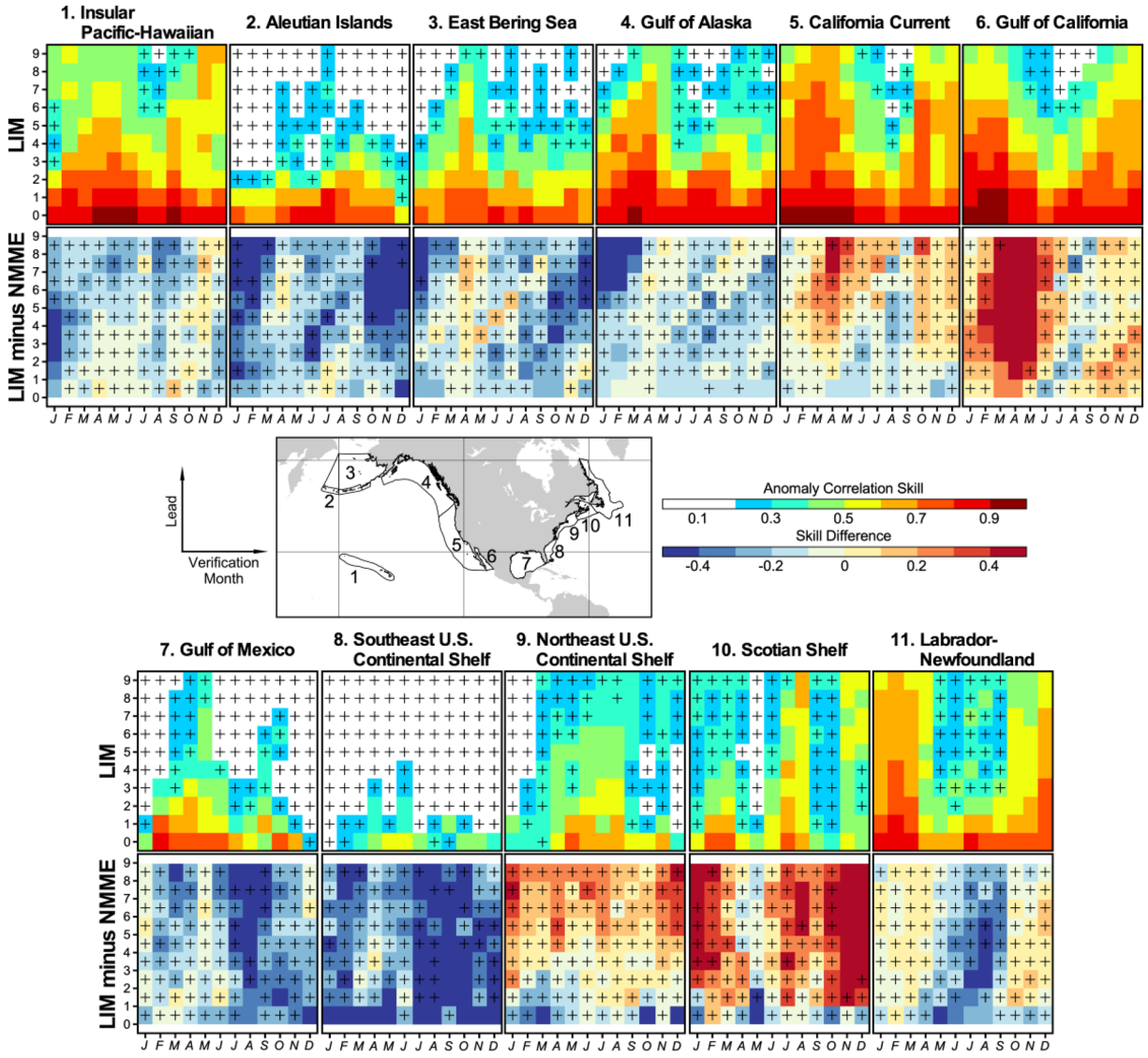

Figure 2. AC hindcast skill of SST anomalies from (top rows) the LIM, and (bottom rows) skill differences (LIM minus NMME-mean), for the 11 LMEs along the North American coast during 1982–2010. Plus signs indicate skill and differences that are not significant at the 95% confidence level.

initialization includes the pronounced post-1992 SSH trend (Balmaseda et al., 2013), while NMME initializations largely do not, although they do include historically evolving external forcings including increasing CO_2 levels (Kirtman et al., 2014 and references therein), as well as the trend that is present within their atmospheric initializations [However, we did not find a substantially different CGCM-LIM comparison when we used one set of model hindcasts (ACCESS-S1, Hudson et al., 2017) whose initialization did include the SSH trend (see Figure S9)]. Removing the externally forced trend from the hindcasts also could be problematic since 30 years may be too short to distinguish between this trend and natural decadal variability (Frankignoul et al., 2017; Solomon et al., 2011).

Instead, following the approach in Tippett et al. (2020), we evaluated the skill of the *tendency* of both SST and SSH anomalies over the 6 months after initialization [i.e., the skill of $\mathbf{x}(6) - \mathbf{x}(0)$ rather than $\mathbf{x}(6)$]. This skill metric removes the impact of the trend component, to the extent that it is relatively constant over any 6-month increment, as is largely the case for the LIM. The results (Figure S7) suggest that the relative skill differences between the LIM and NMME-mean in Figure 1 are not solely due to the trend. We also repeated the skill assessment shown in Figure 3 with this new metric, determining the LIM and NMME-mean AC

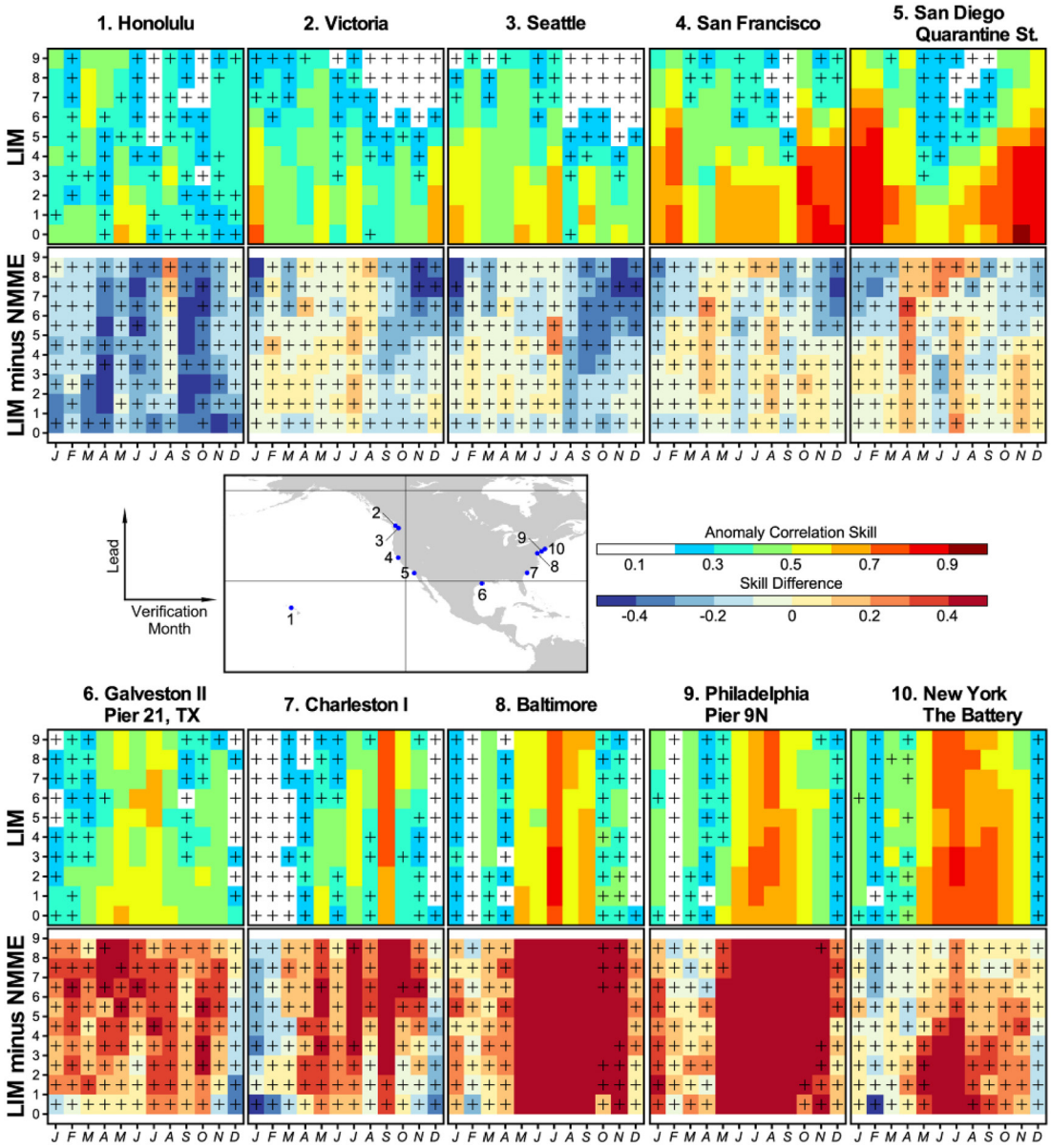

Figure 3. AC hindcast skill of SSH anomalies from (top rows) the LIM, and (bottom rows) skill differences (LIM minus NMME-mean), for the 10 tide gauge stations along the US coast during 1982–2010. Plus signs indicate skill and differences that are not significant at the 95% confidence level.

skill of SSH tendencies [i.e., the skill of $\mathbf{x}(\tau) - \mathbf{x}(0)$] for all the tide gauge stations (Figure S8). The LIM still has higher skill for all the Atlantic coastal stations with this new metric, though the differences are generally not significant at the 95% confidence level. Collectively, these results suggest that seasonal ocean SSH variations are better captured by the LIM than the NMME-mean in many regions outside of the tropical Pacific, notably in much of the Atlantic basin.

4. Evaluating Sources of Skill

Since the LIM captures basic details of coupled GCM hindcast skill, we next diagnose its regional and coastal skill variations. As discussed in Section 2.4, LIM potential predictability depends on forecast signal amplitude, so we discuss initial states that could undergo strong anomaly growth (i.e., the SVs) and relate their potential predictability to the hindcast skill.

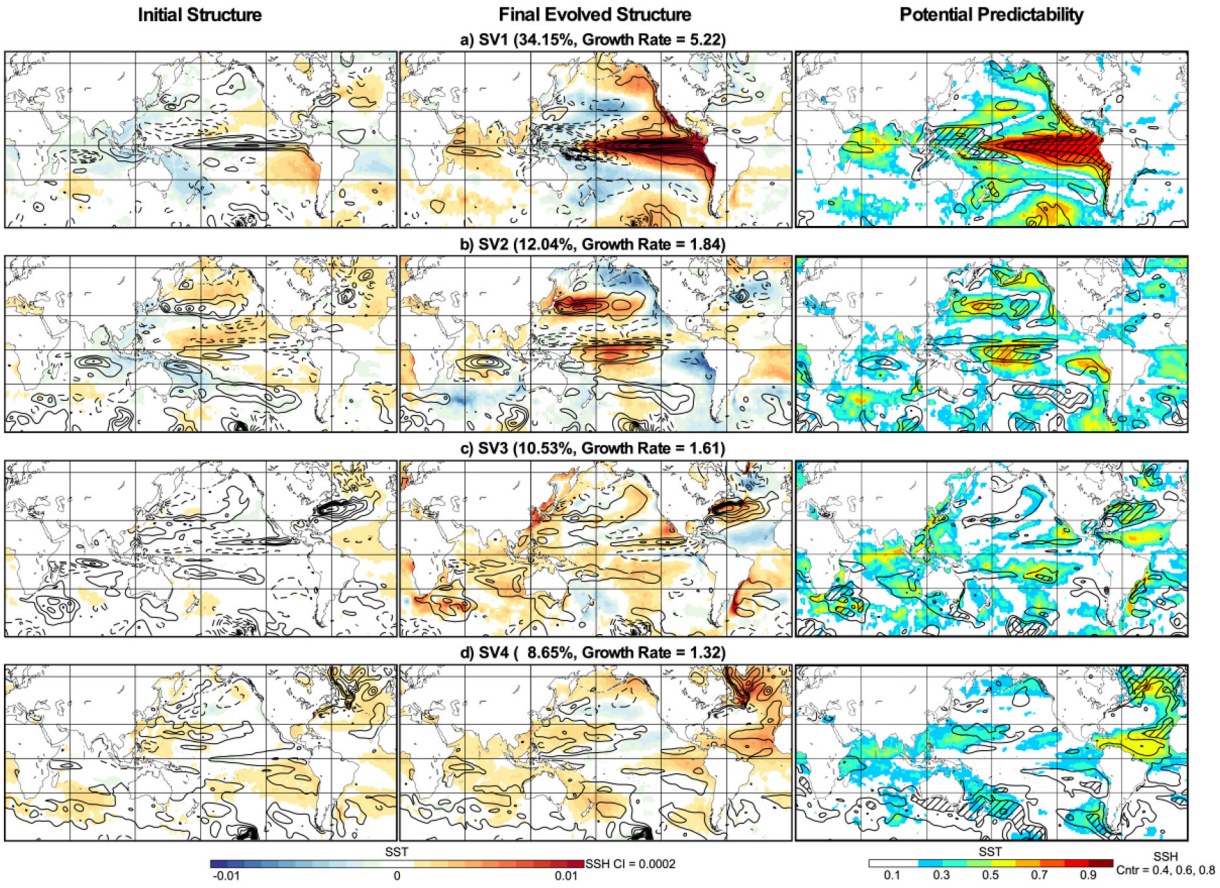

Figure 4. (Left) SST (shading) and SSH (contours) optimal initial structures of the (a) leading, (b) second, (c) third, and (d) fourth SVs of **G**(6) under the L2 norm of total **L** (Middle) Corresponding final evolved structures after a 6-months interval. Contour intervals of both SST and SSH anomalies are arbitrary but are the same for the initial and final structures. For SSH anomalies, negative values are dashed, and zero contours are suppressed. The fractional variance explained by each SV, and its growth rate (singular value squared) are shown (Right) SST and SSH predictability, assessed as local potential AC skill, at 6-months lead determined for each SV. For SSH, 0.6 and 0.8 skill contours are shown, with regions where potential skill is higher than 0.8 indicated by hatching.

Figure 4 shows the four growing initial SVs (left column) and their corresponding evolved structures (middle column) after 6 months. The leading SV shows canonical ENSO-like evolution, with SST and SSH anomalies consistent with a recharge-discharge mechanism (Jin, 1997) and similar to the leading SV of the tropics-only LIM (Newman, Alexander, & Scott, 2011), along with a subsequent "atmospheric bridge" extratropical response (Alexander et al., 2002). Overall, this SV essentially evolves into the dominant pattern of Pacific SST variability (e.g., PDO; Newman, Alexander, Ault, et al., 2016).

SV2 represents dynamics driving an east-west equatorial SST dipole and North Pacific Gyre Oscillation-like (NPGO; Di Lorenzo, Schneider, et al., 2008) SSH anomalies. Different initial combinations of SV1 and SV2 can therefore evolve into a wide diversity of ENSO events (Capotondi, Wittenburg, et al., 2015; Newman, Shin, & Alexander, 2011; Thomas et al., 2018; Vimont et al., 2014) and correspondingly diverse responses in the North Pacific and along the North American coast. For example, relative to SV1 alone, an initial condition with equal contributions from SVs 1 and 2 has an enhanced North Pacific meridional mode (Chang et al., 2007; Chiang & Vimont, 2004) and westward displacement of the initial SSH anomaly. It subsequently evolves into a central Pacific ENSO event (Vimont et al., 2014) with a delayed PDO SST signature and weaker SST/SSH anomalies along the US West Coast (see Figure S10).

Predictable anomaly growth in the North Atlantic is primarily due to the other two SVs, which maximize their growth at 6 months. The evolved SST anomaly structure for SV3 includes the tripole pattern associated with the North Atlantic Oscillation (NAO; Czaja & Frankignoul, 2002). SV4 is more similar to the horseshoe pattern associated with the AMO (Trenberth & Shea, 2006). The SSH portions also support this NAO/AMO

distinction: SV3 is somewhat similar to the second Atlantic SSH EOF, which has been suggested to represent Gulf Stream variability arising from NAO-related wind stress curl (Häkkinen et al., 2011), while SV4 contains more of the Atlantic SSH trend component (Figure 2 of Häkkinen et al., 2013). Both SVs also show pronounced signals along the North American East Coast, with SV3 maximizing along the US coast and SV4 maximizing further north.

Each SV's potential predictability (see Section 2.4 and supporting information S4) is shown in the right column of Figure 4. Note that the potential predictability of the sum of SVs is different than the sum of potential predictability of those SVs. The leading SV alone could drive substantial skill throughout the Pacific basin. However, the second SV also contributes to the central tropical Pacific and could be equally or more important to the North Pacific. LIM skill along the Americas' West Coast appears to reflect ENSO-related coastal Kelvin wave propagation, seen in the leading SV's evolved pattern. Likewise, LIM skill in the North Pacific likely results from ENSO-related atmospheric bridges, as does some of the tropical Indian Ocean skill, although the third and fourth SVs also contribute.

In the North Atlantic, SV3 and SV4 have largely complementary (overlapping) impacts on potential SSH (SST) skill. Along the northeast US coast, the LIM has somewhat higher skill than the NMME-mean, which SV3 (Figure 4c) suggests arises primarily through modulating the Gulf Stream's strength, inferred from the SV's SSH gradient. Additionally, over the Gulf of Mexico and along the southeast coast of the US, where the strong Gulf of Mexico Loop Current and Gulf Stream impact local SST variability, potential LIM skill is limited.

5. Conclusions

A near-global LIM and the NMME multi-model ensemble mean have comparable SST hindcast skill for leads up to 9 months, including for large-scale patterns of climate variability such as ENSO, the PDO, and the AMO. While regional skill is sometimes higher for one model or the other, skill differences are mostly not statistically significant given the relatively short hindcast period. Additionally, while the LIM and NMME-mean (based on a subset of the models) have comparable SSH skill over the Indian and Pacific Oceans, the LIM has significantly more SSH skill over much of the Atlantic and Southern Oceans. Overall, along the North American coastline, both models have comparable skill for SST anomalies within the highly productive coastal zones (LMEs) and SSH anomalies at selected coastal stations. However, greater LIM skill for Atlantic SSH anomalies translates into generally more LIM skill for predicting monthly mean tide gauge observations for several stations along the US East Coast.

In the LIM, combinations of the two leading SVs drive ENSO diversity and diversity of North Pacific anomalies, consistent with diagnosis from many previous studies (Chen & Wallace, 2016; Di Lorenzo, Liguori, et al., 2015; Newman, Alexander, Ault, et al., 2016), with related impacts on North American West Coast predictability. That is, US West Coast forecast skill depends on various aspects of ENSO evolution, not merely the canonical ENSO pattern itself (Capotondi, Sardeshmukh, et al., 2019). LIM SV evolution and related predictability may also prove useful for diagnosing coupled model deficiencies responsible for locally poorer skill relative to the LIM, especially within the coastal region.

One aim here was to use the LIM to benchmark skill of participating NMME coupled models. Note, however, that the five NMME GCMs used for SSH hindcasts do not assimilate satellite altimetry information in their initializations, which might be one reason for their relatively poor seasonal SSH prediction skill. On the other hand, the tendency skill metric (Section 3.3) suggests that higher skill for the LIM is not simply due to its initialization with a more realistic SSH trend. More recent seasonal forecasting systems such as ECMWF's SEAS5 (Johnson et al., 2019), which have finer spatial resolution and assimilate altimetry measurements, might be more skillful than the NMME models and therefore closer to (or even exceed) the LIM benchmark.

Our LIM could be used to support coastal ocean seasonal prediction, providing skillful SST and SSH forecasts that complement climate model ensembles. Given that the coarse-grained LIM has hindcast skill even when verifying against local tide gauge observations, it seems promising to consider downscaling

LIM forecasts, either statistically or dynamically, to finer coastal ocean scales to better predict the physical environment there.

Data Availability Statement

NMME hindcasts are available at https://www.earthsystemgrid.org/search.html?Project=NMME; HadISST data are available at https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html; ECMWF ORAS4 data were downloaded from http://apdrc.soest.hawaii.edu/datadoc/ecmwf_oras4.php.; and Revised Local Reference (RLR) sea-level data are available at https://www.psmsl.org.

Acknowledgments
This work was supported by NOAA/CPO and DOE (grant #0000238382). SS was also supported in part by NOAA/CPO COCA program (grant #NA-15OAR4310133).

References

Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N.-C., & Scott, J. D. (2002). The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans. *Journal of Climate*, *15*, 2205–2231. https://doi.org/10.1175/1520-0442(2002)015<2205:tabtio>2.0.co;2

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*, 1132–1161. https://doi.org/10.1002/qj.2063

Barnston, A. G., Tippett, M. K., van den Dool, H. M., & Unger, D. A. (2015). Toward an improved multimodel ENSO prediction. *Journal of Climate*, *54*, 1579–1595. https://doi.org/10.1175/jamc-d-14-0188.1

Capotondi, A., Jacox, M., Bowler, C., Kavanaugh, M., Lehodey, P., Barrie, D., et al. (2019). Observational needs supporting marine ecosystems modeling and forecasting: From the global ocean to regional and coastal systems. *Frontiers Marine Science*, *6*, 623. https://doi.org/10.3389/fmars.2019.00623

Capotondi, A., Sardeshmukh, P. D., Di Lorenzo, E., Subramanian, A. C., & Miller, A. J. (2019). Predictability of US West Coast Ocean Temperatures is not solely due to ENSO. *Scientific Reports*, *9* .10993. https://doi.org/10.1038/s41598-019-47400-4

Capotondi, A., Wittenburg, A. T., Newman, M., Di Lorenzo, E., Yu, J.-Y., Braconnot, P., et al. (2015). Understanding ENSO diversity. *Bulletin of the American Meteorological Society*, *96*, 921–938. https://doi.org/10.1175/bams-d-13-00117.1

Chang, P., Zhang, L., Saravanan, R., Vimont, D. J., Chiang, J. C. H., Ji, L., et al. (2007). Pacific meridional mode and El Niño-Southern Oscillation. *Geophysical Research Letters*, *34*, L16608. https://doi.org/10.1029/2007GL030302

Chen, X., & Wallace, J. M. (2016). Orthogonal PDO and ENSO indices. *Journal of Climate*, *29*, 3883–3892. https://doi.org/10.1175/jcli-d-15-0684.1

Chiang, J. C. H., & Vimont, D. J. (2004). Analogous Pacific and Atlantic meridional modes of tropical atmosphere–ocean variability. *Journal of Climate*, *17*, 4143–4158. https://doi.org/10.1175/jcli4953.1

Czaja, A., & Frankignoul, C. (2002). Observed impact of Atlantic SST anomalies on the North Atlantic Oscillation. *Journal of Climate*, *15*, 606–623. https://doi.org/10.1175/1520-0442(2002)015<0606:oioasa>2.0.co;2

Dias, D. F., Subramanian, A., Zanna, L., & Miller, A. J. (2018). Remote and local influences in forecasting Pacific SST: A linear inverse model and a multimodel ensemble study. *Climate Dynamics*, *52*, 3183–3201. https://doi.org/10.1007/s00382-018-4323-z

Di Lorenzo, E., Liguori, G., Schneider, N., Furtado, J. C., Anderson, B. T., & Alexander, M. (2015). ENSO and meridional modes: A null hypothesis for Pacific climate variability. *Geophysical Research Letters*, *42*, 9440–9448. https://doi.org/10.1002/2015gl066281

Di Lorenzo, E., Schneider, N., Cobb, K. M., Franks, P. J. S., Chhak, K., Miller, A. J., et al. (2008). North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophysical Research Letters*, *35*, L08607. https://doi.org/10.1029/2007GL03283810.1029/2007gl032838

Ding, H., Newman, M., Alexander, M., & Wittenberg, A. T. (2019). Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs. *Geophysical Research Letters*, *4*, 1147.

Doney, S. C., Ruckelshaus, M., Duffy, J. E., Barry, J. P., Chan, F., English, C. A., et al. (2012). Climate change impacts on marine ecosystems. *Annual Review Marine Science*, *4*, 11–37. https://doi.org/10.1146/annurev-marine-041911-111611

Farrell, B. (1988). Optimal excitation of neutral Rossby waves. *Journal of the Atmospheric Sciences*, *45*, 163–172. https://doi.org/10.1175/1520-0469(1988)045<0163:oeonrw>2.0.co;2

Frankignoul, C., Gastineau, G., & Kwon, Y.-O. (2017). Estimation of the SST response to anthropogenic and external forcing and its impact on the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation. *Journal of Climate*, *30*, 9871–9894. https://doi.org/10.1175/jcli-d-17-0009.1

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., et al. (2015). Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *Journal of Climate*, *28*, 952–977. https://doi.org/10.1175/jcli-d-14-00353.1

Häkkinen, S., Rhines, P. B., & Worthen, D. L. (2011). Warm and saline events embedded in the meridional circulation of the northern North Atlantic. *Journal of Geophysical Research*, *116*, C03006. https://doi.org/10.1029/2010JC006275

Häkkinen, S., Rhines, P. B., & Worthen, D. L. (2013). Northern North Atlantic sea surface height and ocean heat content variability. *Journal of Geophysics Research: Oceans*, *118*, 3670–3678. https://doi.org/10.1002/jgrc.20268

Hawkins, E., Robson, J., Sutton, R., Smith, D., & Keenlyside, N. (2011). Evaluating the potential for statistical decadal predictions of sea surface temperatures with a perfect model approach. *Climate Dynamics*, *37*, 2495–2509. https://doi.org/10.1007/s00382-011-1023-3

Hawkins, E., & Sutton, R. (2009). Decadal predictability of the Atlantic Ocean in a coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling. *Journal of Climate*, *22*, 3960–3978. https://doi.org/10.1175/2009jcli2720.1

Hervieux, G., Alexander, M. A., Stock, C. A., Jacox, M. G., Pegion, K., Becker, E., et al. et al. (2019). More reliable coastal SST forecasts from the North American multimodel ensemble. *Climate Dynamics*, *53*, 7153–7168. https://doi.org/10.1007/s00382-017-3652-7

Huang, J., Dool, H. M., & Barnston, A. G. (1996). Long-lead seasonal temperature prediction using optimal climate normals. *Journal of Climate*, *9*, 809–817. https://doi.org/10.1175/1520-0442(1996)009<0809:llstpu>2.0.co;2

Huddart, B., Subramanian, A., Zanna, L., & Palmer, T. (2017). Seasonal and decadal forecasts of Atlantic sea surface temperatures using a linear inverse model. *Climate Dynamics*, *49*, 1833–1845. https://doi.org/10.1007/s00382-016-3375-1

Hudson, D., Alves, O., Hendon, H. H., Lim, E.-P., Liu, G., Luo, J.-J., et al. (2017). ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system. *Journal South Hemisphere Earth*, *67*(3), 132–159. https://doi.org/10.22499/3.6703.001

Jacox, M., Alexander, M. A., Siedlecki, S., Chen, K., Kwon, Y.-O., Brodie, S., et al. (2020). Seasonal-to-interannual prediction of North American coastal marine ecosystems: Forecast methods, mechanisms of predictability, and priority developments. *Progress in Oceanography*, *183*, 102307. https://doi.org/10.1016/j.pocean.2020.102307

Jin, F.-F. (1997). An equatorial recharge paradigm for ENSO: I. Conceptual model. *Journal of the Atmospheric Sciences*, *54*, 811–829. https://doi.org/10.1175/1520-0469(1997)054<0811:aeorpf>2.0.co;2

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al. (2019). SEAS5: The new ECMWF seasonal forecast system. *Geoscientific Model Development*, *12*, 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., III, Paolino, D. A., Zhang, Q., et al. (2014). The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, *95*, 585–601. https://doi.org/10.1175/bams-d-12-00050.1

Lacarra, J.-F., & Talagrand, O. (1988). Short-range evolution of small perturbation in a barotropic model. *Tellus*, *40A*, 81–95. https://doi.org/10.1111/j.1600-0870.1988.tb00408.x

Newman, M. (2013). An empirical benchmark for decadal forecasts of global surface temperature anomalies. *Journal of Climate*, *26*(14), 5260–5269. https://doi.org/10.1175/jcli-d-12-00590.1

Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., et al. (2016). The Pacific Decadal Oscillation, revisited. *Journal of Climate*, *29*, 4399–4427. https://doi.org/10.1175/jcli-d-15-0508.1

Newman, M., Alexander, M. A., & Scott, J. D. (2011). An empirical model of tropical ocean dynamics. *Climate Dynamics*, *37*, 1823–1841. https://doi.org/10.1007/s00382-011-1034-0

Newman, M., & Sardeshmukh, P. D. (2017). Are we near predictability limit of tropical sea surface temperatures? *Geophysical Research Letters*, *44*, 8520–8529. https://doi.org/10.1002/2017gl074088

Newman, M., Sardeshmukh, P. D., Winkler, C. R., & Whitaker, J. S. (2003). A study of subseasonal predictability. *Monthly Weather Review*, *131*, 1715–1732. https://doi.org/10.1175//2558.1

Newman, M., Shin, S. I., & Alexander, M. (2011). Natural variation in ENSO flavors. *Geophysical Research Letters*, *38*, L14705. https://doi.org/10.1029/2011GL047658

Penland, C., & Matrosova, L. (2006). Studies of El Niño and interdecadal variability in tropical sea surface temperatures using a nonnormal filter. *Journal of Climate*, *19*, 5796–5815. https://doi.org/10.1175/jcli3951.1

Penland, C., & Sardeshmukh, P. D. (1995). The optimal growth of tropical sea surface temperature anomalies. *Journal of Climate*, *8*, 1999–2014. https://doi.org/10.1175/1520-0442(1995)008<1999:togots>2.0.co;2

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, *108*, 4407. https://doi.org/10.1029/2002JD002670

Sardeshmukh, P. D., Compo, G. P., & Penland, C. (2000). Changes of probability associated with El Niño. *Journal of Climate*, *13*, 4268–4286. https://doi.org/10.1175/1520-0442(2000)013<4268:copawe>2.0.co;2

Solomon, A., Goddard, L., Kumar, A., Carton, J., Deser, C., Fukumori, I., et al. (2011). Distinguishing the roles of natural and anthropogenically forced decadal climate variability. *Bulletin of the American Meteorological Society*, *92*, 141–156. https://doi.org/10.1175/2010bams2962.1

Stock, C. A., Pegion, K., Vecchi, G. A., Alexander, M. A., Tommasi, D., Bond, N. A., et al. (2015). Seasonal sea surface temperature anomaly prediction for coastal ecosystems. *Progress in Oceanography*, *137*, 219–236. https://doi.org/10.1016/j.pocean.2015.06.007

Thomas, E. E., Vimont, D. J., Newman, M., Penland, C., & Martinez-Villalobos, C. (2018). The role of stochastic forcing in generating ENSO diversity. *Journal of Climate*, *31*, 9125–9149. https://doi.org/10.1175/JCLI-D-17-0582.1

Tippett, M. K., L'Heureux, M. L., Becker, E. J., & Kumar, A. (2020). Excessive momentum and false alarms in late-spring ENSO forecasts. *Geophysical Research Letters*, *47*. https://doi.org/10.1029/2020GL087008

Trenberth, K. E., & Shea, D. J. (2006). Atlantic hurricanes and natural variability in 2005. *Geophysical Research Letters*, *33*, L12704. https://doi.org/10.1029/2006GL026894

Vimont, D. J., Alexander, M., & Newman, M. (2014). Optimal growth of central and east Pacific ENSO events. *Geophysical Research Letters*, *41*, 4027–4034. https://doi.org/10.1002/2014gl059997

Zanna, L. (2012). Forecast skill and predictability of observed Atlantic sea surface temperatures. *Journal of Climate*, *25*, 5047–5056. https://doi.org/10.1175/jcli-d-11-00539.1

Zwiers, F. W., & von Storch, H. (1995). Taking serial correlation into account in tests of the mean. *Journal of Climate*, *8*, 336–351. https://doi.org/10.1175/1520-0442(1995)008<0336:tsciai>2.0.co;2