

1  
2 **Seasonal Predictability of Global and North American Coastal Sea Surface**  
3 **Temperature and Height Anomalies**

4  
5 **Sang-Ik Shin<sup>1</sup>, and Matthew Newman<sup>1</sup>**

6 <sup>1</sup>CIRES, University of Colorado Boulder & NOAA Physical Sciences Laboratory, Boulder,  
7 Colorado.

8  
9 Corresponding author: Sang-Ik Shin ([Sangik.Shin@noaa.gov](mailto:Sangik.Shin@noaa.gov))

10  
11 **Key Points:**

- 12 • A Linear Inverse Model (LIM) is constructed that can predict monthly mean global sea  
13 surface temperature (SST) and height (SSH) anomalies.
- 14 • The LIM outperforms operational numerical models for SSH skill in the Atlantic and US  
15 East Coast tide gauge stations.
- 16 • LIM coastal skill is largely due to capturing ENSO teleconnections impacting the West  
17 Coast, and Gulf Stream modulation for the East Coast.  
18

## 19 **Abstract**

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

## 31 **Plain Language Summary**

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

## 40 **1 Introduction**

41 Coastal marine ecosystems are strongly impacted by changes in ocean circulation and  
42 stratification, which lead to the geographic redistribution of available nutrients and oxygen and,  
43 in turn, marine species (Doney et al., 2012; Capotondi et al., 2019a). Therefore, skillful  
44 prediction of coastal ocean conditions is increasingly needed for marine resource management  
45 (Jacox et al., 2020). In fact, within the Large Marine Ecosystems (LMEs), relatively large coastal  
46 zones where primary productivity is higher than in the open ocean, sea surface temperature (SST)  
47 hindcast skill from the numerical models of the North American Multi-Model Ensemble (NMME;  
48 Kirtman et al., 2014) is large enough to be potentially useful (Stock et al., 2015; Hervieux et al.,  
49 2019). Recently, in a comprehensive review of seasonal-to-interannual prediction methods and  
50 their skill within North American coastal marine ecosystems, Jacox et al. (2020) suggested not  
51 only that prediction skill may be maximized via a hybrid statistical-numerical prediction  
52 approach, but also that statistical prediction alone could be competitive with numerical  
53 prediction.

54 This paper explores North American coastal forecast skill using a global Linear Inverse  
55 Model (LIM; Penland and Sardeshmukh, 1995; hereafter PS95). The LIM is an empirical  
56 dynamical model that may diagnose potential spatiotemporal variations of skill, since its  
57 forecasts are often competitive with numerical model forecasts. For example, Newman and  
58 Sardeshmukh (2017; hereafter NS17) showed that a tropical LIM had similar tropical Indo-  
59 Pacific SST skill to the grand ensemble mean of the NMME operational models and that the LIM  
60 itself largely predicted the spatial and temporal skill variations of both models. LIMs have been  
61 constructed for other regions, including the Pacific and the Atlantic Ocean basins (Hawkins and

62 Sutton, 2009; Hawkins et al., 2011; Zanna, 2012; Newman et al., 2016; Huddart et al., 2017;  
 63 Dias et al., 2018; and many others). However, the LIM has rarely been applied globally, nor has  
 64 its coastal skill explicitly been evaluated. Here we do both, extending the approach of NS17 to  
 65 cover most of the ice-free global ocean (60°S-65°N), yielding a global skill benchmark of the  
 66 current generation of coupled climate models that allows for estimating potential source(s) of  
 67 predictability. We then specifically evaluate prediction skill along the North American coastline,  
 68 focusing on monthly SST anomalies within LMEs and sea surface height (SSH) anomalies  
 69 observed at numerous tide gauge stations, and diagnose how large-scale global dynamics impact  
 70 the potential for such skill.

## 71 **2 Methods**

### 72 **2.1 Data**

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

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

### 86 **2.2 Linear Inverse Model**

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

$$89 \quad \frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \mathbf{S}\boldsymbol{\eta}, \quad (1)$$

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

$$97 \quad \hat{\mathbf{x}}(t + \tau) = \mathbf{G}(\tau) \cdot \mathbf{x}(t), \quad (2)$$

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

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

### 105 2.3 Hindcasts and skill evaluation

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

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

### 120 2.4 LIM anomaly growth and predictability

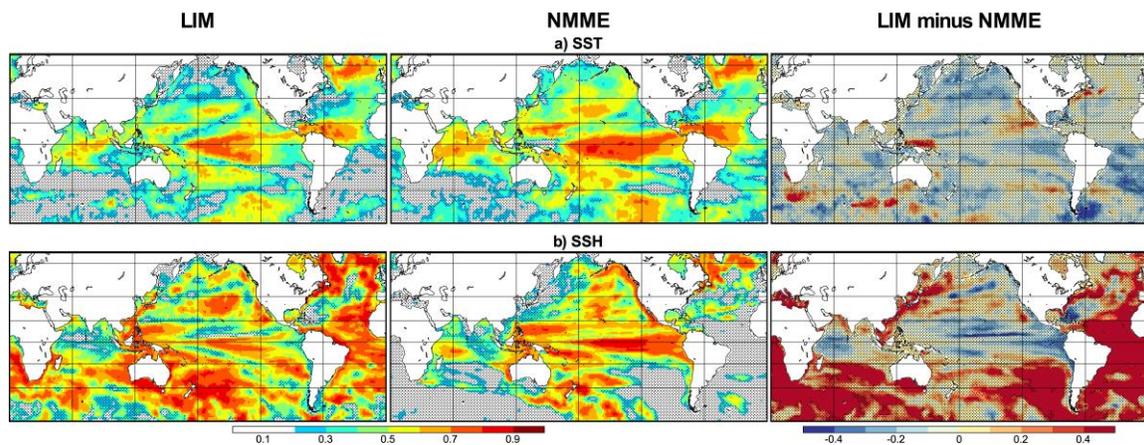
121 Assessing LIM predictability is straightforward since (1) distinguishes between  
122 predictable signal and unpredictable noise. For any infinite-member ensemble-mean perfect  
123 model forecasts, the average AC between forecasts at the lead time  $\tau$  and the corresponding  
124 verification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

### 137 3 Hindcast Skill Evaluation

#### 138 3.1 Global skill

139 Figure 1 compares the Month 6 LIM and NMME-mean AC skill of predicted SST (top  
 140 row) and SSH (bottom row) anomalies. Results at other leads (not shown) present similar  
 141 patterns and lead to similar conclusions, as does the RMSSS (Fig. S3). Overall, the LIM and  
 142 NMME-mean have comparable SST skill patterns and amplitudes, although there are some  
 143 notable regions of lesser LIM skill. For example, skill is quite similar in the central equatorial  
 144 Pacific (see also Fig. S4a) but not in the eastern equatorial Pacific. [In fact, a Tropics-only LIM  
 145 is also more skillful than our global LIM in the eastern equatorial Pacific (not shown, but see  
 146 NS17), which might be a drawback of the global approach.] On the other hand, the global LIM  
 147 still has higher skill than the NMME-mean in the far western equatorial Pacific, where the  
 148 NMME-mean forecasts cannot produce the correct anomalous SST horseshoe ENSO pattern  
 149 (NS17), and in the subtropical eastern Pacific, extending up towards the US West Coast.

150



151

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

156

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

167 The LIM also provides skillful SSH hindcasts. For example, Month 6 AC skill is higher  
 168 than 0.6 in most regions. This contrasts with the NMME-mean, whose hindcast skill is largely

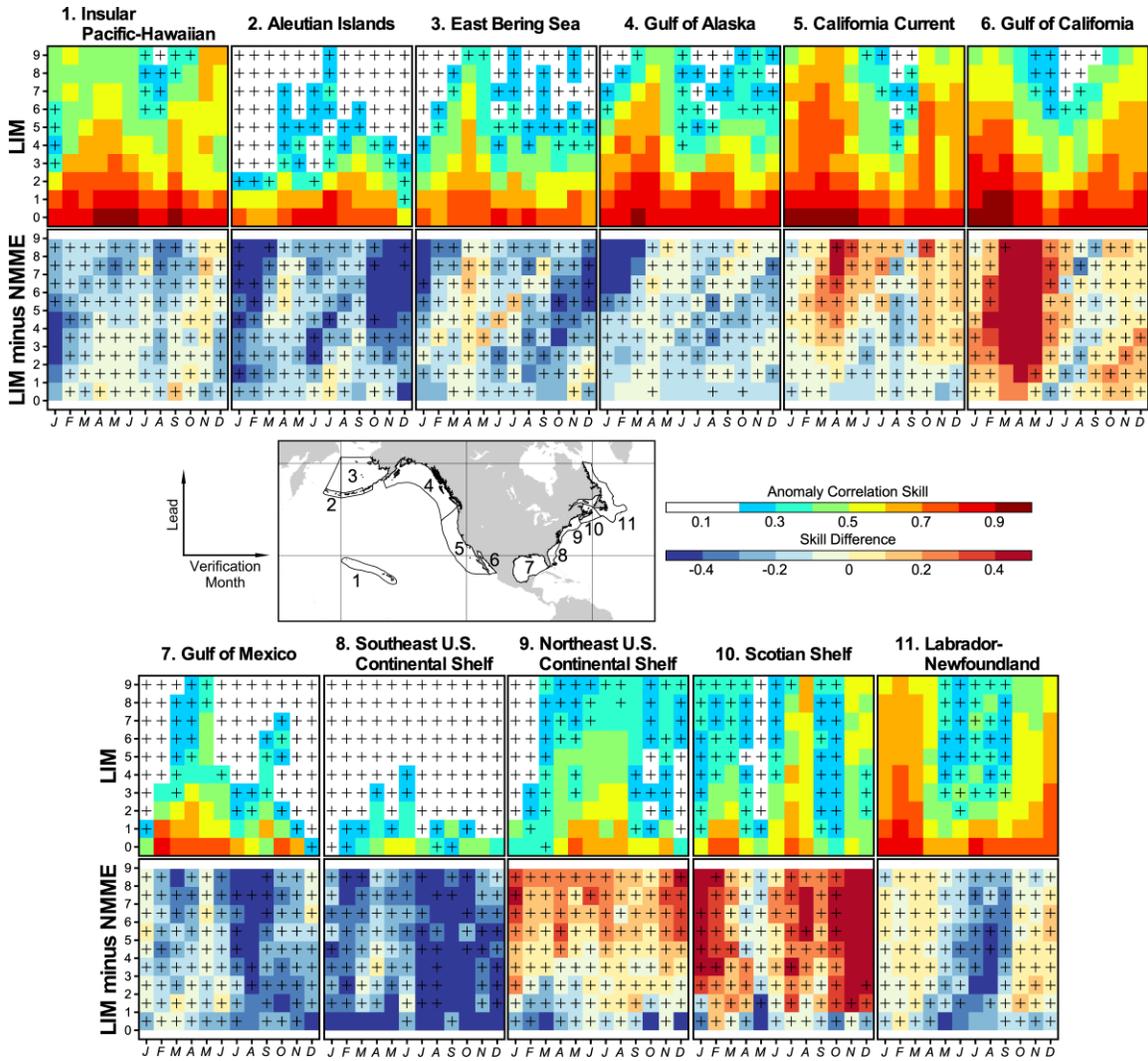
169 limited to the ENSO-impacted Indo-Pacific and along the west coast of North America, where  
170 skill somewhat exceeds the LIM. The absence of skill elsewhere for the NMME-mean,  
171 especially in comparison to the LIM (Fig. 1b), is particularly stark throughout much of the  
172 Atlantic and Southern Oceans and within the North Pacific western boundary region.

### 173 3.2 North American coastal skill

174 Both models generally have comparable skill along much of the world's coastlines (Fig.  
175 1), but here we will focus upon North America. We repeated the Hervieux et al. (2019) analysis,  
176 comparing LIM and NMME-mean SST skill within the eleven North American LMEs (see map  
177 in Fig. 2). Figure 2 shows LIM skill for each LME as a function of verification month and  
178 forecast lead, alongside the LIM and NMME-mean skill difference (see Fig. S5 for NMME-  
179 mean skill). In general, skill differences are small, statistically significant at the 95% confidence  
180 level for only a few LMEs and only specific seasons. For example, the NMME-mean is more  
181 skillful in the Aleutian Islands, East Bering Sea, and Gulf of Alaska LMEs, but not uniformly  
182 and significantly during winter. In contrast, the LIM is more skillful within regions strongly  
183 influenced by ENSO, such as the California Current and Gulf of California LMEs, although only  
184 significantly during spring. Interestingly, the LIM is also more skillful along the northeastern US  
185 coast, especially in the Scotian Shelf, perhaps because the Gulf Stream in some CGCMs tends to  
186 separate from the coast further north than observed (Griffies et al., 2015; Schoonover et al.,  
187 2016).

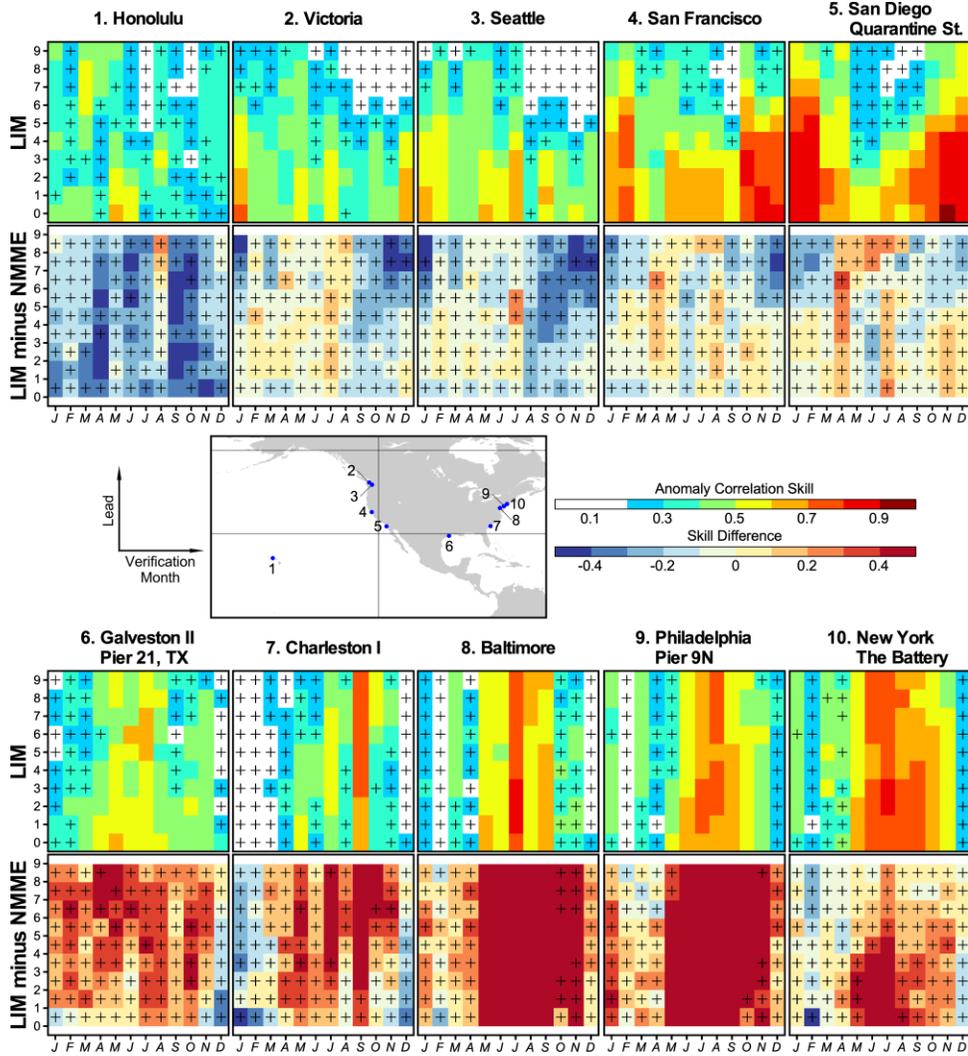
188 Next, we verified SSH skill against monthly sea-level observations at ten tide gauge  
189 stations along the North American coast plus Honolulu (Fig. 3), drawn from the Revised Local  
190 Reference (RLR; <https://www.psmsl.org>) dataset during 1982-2010. Since the hindcast grid is  
191 too coarse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

197



198

199 **Figure 2.** AC hindcast skill of SST anomalies from (top rows) the LIM, and (bottom rows) skill  
 200 differences (LIM minus NMME-mean), for the eleven LMEs along the North American coast  
 201 during 1982-2010. Plus signs indicate skill and differences that are insignificant at the 95%  
 202 confidence level.  
 203



204  
205  
206  
207  
208  
209

**Figure 3.** AC hindcast skill of SSH anomalies from (top rows) the LIM, and (bottom rows) skill differences (LIM minus NMME-mean), for the ten tide gauge stations along the US coast during 1982-2010. Plus signs indicate skill and differences that are insignificant at the 95% confidence level.

### 210 3.3 Assessing the impact of the secular trend on skill

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

216 In some regions and for some variables, secular trends may significantly enhance  
217 seasonal forecast skill (Huang et al., 1996; Ding et al., 2019). This is not necessarily dependent  
218 upon predicting the trend, since standard skill scores defined relative to a long-term fixed  
219 climatology are inflated simply by including the trend within the climate anomaly definition.  
220 SSH hindcast skill could be particularly impacted, since LIM initialization includes the  
221 pronounced post-1992 SSH trend (Balmaseda et al., 2013), while NMME initializations largely  
222 do not, although they do include historically evolving external forcings including increasing CO<sub>2</sub>  
223 levels (Kirtman et al. 2014 and references therein), as well as the trend that is present within their  
224 atmospheric initializations<sup>1</sup>. Removing the externally-forced trend from the hindcasts also could  
225 be problematic since 30 years may be too short to distinguish between this trend and natural  
226 decadal variability (Solomon et al., 2011; Frankignoul et al., 2017).

227 Instead, following the approach in Tippett et al. (2020), we evaluated the skill of the  
228 *tendency* of both SST and SSH anomalies over the six months after initialization [i.e., the skill of  
229  $\mathbf{x}(6) - \mathbf{x}(0)$  rather than  $\mathbf{x}(6)$ ]. This skill metric removes the impact of the trend component, to  
230 the extent that it is relatively constant over any six-month increment, as is largely the case for the  
231 LIM. The results (Fig. S7) suggest that the relative skill differences between the LIM and  
232 NMME-mean in Figure 1 are not solely due to the trend. We also repeated the skill assessment  
233 shown in Figure 3 with this new metric, determining the LIM and NMME-mean AC skill of SSH  
234 tendencies [i.e., the skill of  $\mathbf{x}(\tau) - \mathbf{x}(0)$ ] for all the tide gauge stations (Fig. S8). The LIM still  
235 has higher skill for all the Atlantic coastal stations with this new metric, though the differences  
236 are generally not significant at the 95% confidence level. Collectively, these results suggest that  
237 seasonal ocean SSH variations are better captured by the LIM than the NMME-mean in many  
238 regions outside of the tropical Pacific, notably in much of the Atlantic basin.

## 239 4 Evaluating Sources of Skill

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

244 Figure 4 shows the four growing initial SVs (left column) and their corresponding  
245 evolved structures (middle column) after six months. The leading SV shows canonical ENSO-  
246 like evolution, with SST and SSH anomalies consistent with a recharge-discharge mechanism  
247 (Jin, 1997) and similar to the leading SV of the tropics-only LIM (Newman et al., 2011a), along  
248 with a subsequent “atmospheric bridge” extratropical response (Alexander et al., 2002). Overall,

---

<sup>1</sup> 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 Fig. S9).

249 this SV essentially evolves into the dominant pattern of Pacific SST variability (e.g., PDO;  
250 Newman et al., 2016).

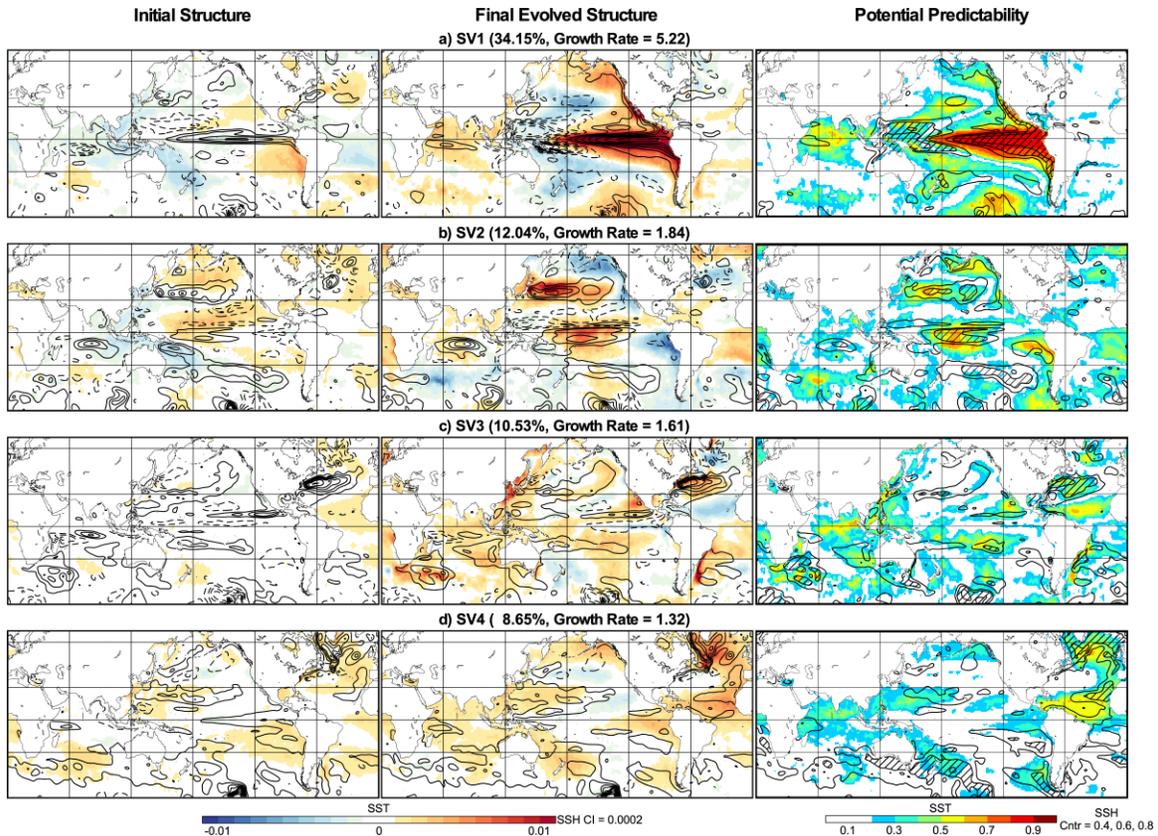
251 SV2 represents dynamics driving an east-west equatorial SST dipole and North Pacific  
252 Gyre Oscillation-like (NPGO; Di Lorenzo et al., 2008) SSH anomalies. Different initial  
253 combinations of SV1 and SV2 can therefore evolve into a wide diversity of ENSO events  
254 (Newman et al., 2011b;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

261 Predictable anomaly growth in the North Atlantic is primarily due to the other two SVs,  
262 which maximize their growth at six months. The evolved SST anomaly structure for SV3  
263 includes the tripole pattern associated with the North Atlantic Oscillation (NAO; Czaja and  
264 Frankignoul, 2002). SV4 is more similar to the horseshoe pattern associated with the AMO  
265 (Trenberth and Shea, 2006). The SSH portions also support this NAO/AMO distinction: SV3 is  
266 somewhat similar to the second Atlantic SSH EOF, which has been suggested to represent Gulf  
267 Stream variability arising from NAO-related wind stress curl (Häkkinen et al., 2011), while SV4  
268 contains more of the Atlantic SSH trend component (Fig. 2 of Häkkinen et al., 2013). Both SVs  
269 also show pronounced signals along the North American East Coast, with SV3 maximizing along  
270 the US coast and SV4 maximizing further north.

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

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

286



287  
 288 **Figure 4.** (Left) SST (shading) and SSH (contours) optimal initial structures of the (a) leading, (b) 2<sup>nd</sup>,  
 289 (c) 3<sup>rd</sup>, and (d) 4<sup>th</sup> SVs of  $\mathbf{G}(6)$  under the L2 norm of total  $\mathbf{L}$ . (Middle) Corresponding final  
 290 evolved structures after a 6-month interval. Contour intervals of both SST and SSH  
 291 anomalies are arbitrary but are the same for the initial and final structures. For SSH  
 292 anomalies, negative values are dashed, and zero contours are suppressed. The fractional  
 293 variance explained by each SV, and its growth rate (singular value squared) are shown.  
 294 (Right) SST and SSH predictability, assessed as local potential AC skill, at 6-month lead  
 295 determined for each SV. For SSH, 0.6 and 0.8 skill contours are shown, with regions where  
 296 potential skill is higher than 0.8 indicated by hatching.

297

## 298 5 Conclusions

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

310 In the LIM, combinations of the two leading SVs drive ENSO diversity and diversity of  
311 North Pacific anomalies, consistent with diagnosis from many previous studies (Di Lorenzo et al.,  
312 2015; Newman et al., 2016; Chen and Wallace, 2016), with related impacts on North American  
313 West Coast predictability. That is, US West Coast forecast skill depends on various aspects of  
314 ENSO evolution, not merely the canonical ENSO pattern itself (Capotondi et al., 2019b). LIM  
315 SV evolution and related predictability may also prove useful for diagnosing coupled model  
316 deficiencies responsible for locally poorer skill relative to the LIM, especially within the coastal  
317 region.

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

327 Our LIM could be used to support coastal ocean seasonal prediction, providing skillful  
328 SST and SSH forecasts that complement climate model ensembles. Given that the coarse-grained  
329 LIM has hindcast skill even when verifying against local tide gauge observations, it seems  
330 promising to consider downscaling LIM forecasts, either statistically or dynamically, to finer  
331 coastal ocean scales to better predict the physical environment there.

332

### 333 **Acknowledgments and Data**

334 This work was supported by NOAA/CPO and DOE (grant #0000238382). SS was also supported  
335 in part by NOAA/CPO COCA program (grant #NA-15OAR4310133). NMME hindcasts are  
336 available at <https://www.earthsystemgrid.org/search.html?Project=NMME>; HadISST data are  
337 available at <https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html>; ECMWF ORAS4  
338 data were downloaded from [http://apdrc.soest.hawaii.edu/datadoc/ecmwf\\_oras4.php](http://apdrc.soest.hawaii.edu/datadoc/ecmwf_oras4.php); and  
339 Revised Local Reference (RLR) sea-level data are available at <https://www.psmsl.org>.

340

### 341 **References**

- 342 Alexander, M. A., Bladé, I., Newman, M., Lanzante, J. R., Lau, N.-C., & Scott, J. D. (2002), The  
343 atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over  
344 the global oceans. *J. Climate*, 15, 2205-2231.
- 345 Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013), Evaluation of the ECMWF ocean  
346 reanalysis system ORAS4. *Q. J. R. Meteorol. Soc.*, 139, 1132-1161.
- 347 Barnston, A. G., Tippett, M. K., van den Dool, H. M., & Unger, D. A. (2015), Toward an  
348 improved multimodel ENSO prediction. *J. Climate*, 54, 1579-1595.
- 349 Capotondi, A., Wittenburg, A. T., Newman, M., Di Lorenzo, E., Yu, J.-Y., Braconnot, P., et al.  
350 (2015), Understanding ENSO diversity. *Bull. Amer. Meteor. Soc.*, 96, 921-938.

- 351 Capotondi, A., Jacox, M., Bowler, C., Kavanaugh, M., Lehodey, P., Barrie, D., et al. (2019a).  
352 Observational Needs Supporting Marine Ecosystems Modeling and Forecasting: From  
353 the Global Ocean to Regional and Coastal Systems. *Front. Mar. Sci.*, 6, 623.  
354 <https://doi.org/10.3389/fmars.2019.00623>
- 355 Capotondi, A., Sardeshmukh, P. D., Di Lorenzo, E., Subramanian, A. C., & Miller, A. J. (2019b).  
356 Predictability of US West Coast Ocean Temperatures is not solely due to ENSO. *Sci.*  
357 *Rep.*, 9, 10993. <https://doi.org/10.1038/s41598-019-47400-4>
- 358 Chang, P., Zhang, L., Saravanan, R., Vimont, D. J., Chiang, J. C. H., Ji, L., Seidel, H., & Tippett,  
359 M. K. (2007), Pacific meridional mode and El Niño-Southern Oscillation. *Geophys. Res.*  
360 *Lett.*, 34, L16608, doi:10.1029/2007GL030302.
- 361 Chen, X., & Wallace, J. M. (2016), Orthogonal PDO and ENSO indices. *J. Climate*, 29, 3883–  
362 3892.
- 363 Chiang, J. C. H., & Vimont, D. J. (2004), Analogous Pacific and Atlantic meridional modes of  
364 tropical atmosphere–ocean variability. *J. Climate*, 17, 4143–4158.
- 365 Czaja, A., & Frankignoul, C. (2002), Observed impact of Atlantic SST anomalies on the North  
366 Atlantic Oscillation. *J. Climate*, 15, 606–623.
- 367 Di Lorenzo, E., Schneider, N., Cobb, K. M., Franks, P. J. S., Chhak, K., Miller, A. J., et al.  
368 (2008), North Pacific Gyre Oscillation links ocean climate and ecosystem change.  
369 *Geophys. Res. Lett.*, 35, L08607, doi:10.1029/2007GL032838.
- 370 Di Lorenzo, E., Liguori, G., Schneider, N., Furtado, J. C., Anderson, B. T., & Alexander, M.  
371 (2015), ENSO and meridional modes: A null hypothesis for Pacific climate variability.  
372 *Geophys. Res. Lett.*, 42, 9440–9448.
- 373 Dias, D. F., Subramanian, A., Zanna, L., & Miller, A. J. (2018), Remote and local influences in  
374 forecasting Pacific SST: a linear inverse model and a multimodel ensemble study. *Clim.*  
375 *Dyn.*, 52, 3183–3201.
- 376 Ding, H., Newman, M., Alexander, M., & Wittenberg, A. T. (2019), Diagnosing secular  
377 variations in retrospective ENSO seasonal forecast skill using CMIP5 model-  
378 analogs. *Geophys. Res. Lett.*, 4, 1147.
- 379 Doney, S. C., Ruckelshaus, M., Duffy, J. E., Barry, J. P., Chan, F., English, C. A., et al. (2012),  
380 Climate change impacts on marine ecosystems. *Ann. Rev. Mar. Sci.*, 4, 11–37.
- 381 Farrell, B. (1988), Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.*, 45, 163–172.
- 382 Frankignoul, C., Gastineau, G., & Kwon, Y.-O. (2017), Estimation of the SST response to  
383 anthropogenic and external forcing and its impact on the Atlantic Multidecadal  
384 Oscillation and the Pacific Decadal Oscillation. *J. Climate*, 30, 9871–9894.
- 385 Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., et al.  
386 (2015), Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate  
387 models. *J. Climate*, 28, 952–977.
- 388 Häkkinen, S., Rhines, P. B., & Worthen, D. L. (2011), Warm and saline events embedded in the  
389 meridional circulation of the northern North Atlantic. *J. Geophys. Res. Atmos.*, 116,  
390 C03006, doi.org/10.1029/2010JC006275.

- 391 Häkkinen, S., Rhines, P. B., & Worthen, D. L. (2013), Northern North Atlantic sea surface  
392 height and ocean heat content variability. *J. Geophys. Res. Oceans*, 118, 3670–3678.
- 393 Hawkins, E., & Sutton, R. (2009), Decadal predictability of the Atlantic Ocean in a coupled  
394 GCM: Forecast skill and optimal perturbations using linear inverse modeling. *J. Climate*,  
395 22, 3960–3978.
- 396 Hawkins, E., Robson, J., Sutton, R., Smith, D., & Keenlyside, N. (2011), Evaluating the potential  
397 for statistical decadal predictions of sea surface temperatures with a perfect model  
398 approach. *Clim. Dyn.*, 37, 2495–2509.
- 399 Hervieux, G., Alexander, M. A., Stock, C. A., Jacox, M. G., Pegion, K., Becker, E., et al. et al.  
400 (2019), More reliable coastal SST forecasts from the North American multimodel  
401 ensemble. *Clim. Dyn.*, 53, 7153–7168.
- 402 Huang, J., Dool, H. M. van den, & Barnston, A. G. (1996), Long-lead seasonal temperature  
403 prediction using optimal climate normals. *J. Climate*, 9, 809–817.
- 404 Huddart, B., Subramanian, A., Zanna, L., & Palmer, T. (2017), Seasonal and decadal forecasts of  
405 Atlantic sea surface temperatures using a linear inverse model. *Clim. Dyn.*, 49, 1833–  
406 1845.
- 407 Hudson, D., Alves, O., Hendon, H. H., Lim, E.-P., Liu, G., Luo, J.-J., et al. (2017), ACCESS-S1:  
408 The new Bureau of Meteorology multi-week to seasonal prediction system. *J. So.  
409 Hemisph. Earth*, 67(3), 132–159. doi: 10.22499/3.6703.001
- 410 Jacox, M., Alexander, M. A., Siedlecki, S., Chen, K., Kwon, Y.-O., Brodie, S., et al. (2020),  
411 Seasonal-to-interannual prediction of North American coastal marine ecosystems:  
412 Forecast methods, mechanisms of predictability, and priority developments. *Prog.  
413 Oceanogr.*, 183, 102307, <https://doi.org/10.1016/j.pocean.2020.102307>.
- 414 Jin, F.-F. (1997), An equatorial recharge paradigm for ENSO: I. Conceptual model. *J. Atmos.  
415 Sci.*, 54, 811–829.
- 416 Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., et al.  
417 (2019), SEAS5: the new ECMWF seasonal forecast system. *Geosci. Model Dev.*, 12,  
418 1087–1117.
- 419 Kirtman, B. P., Min, D., Infanti, J. M., Kinter III, J. L., Paolino, D. A., Zhang, Q., et al. (2014),  
420 The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction;  
421 Phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, 95, 585–  
422 601.
- 423 Lacarra, J.-F., & Talagrand, O. (1988), Short-range evolution of small perturbation in a  
424 barotropic model. *Tellus*, 40A, 81–95.
- 425 Newman, M. (2013). An Empirical Benchmark for Decadal Forecasts of Global Surface  
426 Temperature Anomalies. *J. Climate*, 26(14), 5260–5269. [https://doi.org/10.1175/jcli-d-](https://doi.org/10.1175/jcli-d-12-00590.1)  
427 12-00590.1
- 428 Newman, M., Sardeshmukh, P. D., Winkler, C. R., & Whitaker, J. S. (2003), A study of  
429 subseasonal predictability. *Mon. Wea. Rev.*, 131, 1715–1732.

- 430 Newman, M., Alexander, M. A., & Scott, J. D. (2011a), An empirical model of tropical ocean  
431 dynamics. *Clim. Dyn.*, 37, 1823-1841.
- 432 Newman, M., Shin, S. I., & Alexander, M. (2011b), Natural variation in ENSO flavors. *Geophys.*  
433 *Res. Lett.*, 38, L14705, doi:10.1029/2011GL047658.
- 434 Newman, M., Alexander, M. A., Ault, T. R., Cobb, K. M., Deser, C., Di Lorenzo, E., et al.  
435 (2016), The Pacific Decadal Oscillation, revisited. *J. Climate*, 29, 4399-4427.
- 436 Newman, M., & Sardeshmukh, P. D. (2017), Are we near predictability limit of tropical sea  
437 surface temperatures? *Geophys. Res. Lett.*, 44, 8520-8529.
- 438 Penland, C., & Matrosova, L. (2006), Studies of El Niño and interdecadal variability in tropical  
439 sea surface temperatures using a nonnormal filter. *J. Climate*, 19, 5796-5815.
- 440 Penland, C., & Sardeshmukh, P. D. (1995), The optimal growth of tropical sea surface  
441 temperature anomalies. *J. Climate*, 8, 1999-2014.
- 442 Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., et al.  
443 (2003), Global analyses of sea surface temperature, sea ice, and night marine air  
444 temperature since the late nineteenth century. *J. Geophys. Res. Atmos.*, 108, 4407,  
445 doi:10.1029/2002JD002670.
- 446 Sardeshmukh, P. D., Compo, G. P., & Penland, C. (2000), Changes of probability associated  
447 with El Niño. *J. Climate*, 13, 4268-4286.
- 448 Schoonover, J., Dewar, W., Wienders, N., Gula, J., McWilliams, J., Jeroen, M., et al.  
449 (2016), North Atlantic barotropic vorticity balances in numerical models. *J. Phys.*  
450 *Oceanogr.*, 46, 289–303.
- 451 Solomon, A., Goddard, L., Kumar, A., Carton, J., Deser, C., Fukumori, I., et al. (2011),  
452 Distinguishing the roles of natural and anthropogenically forced decadal climate  
453 variability. *Bull. Amer. Meteor. Soc.*, 92, 141–156.
- 454 Stock, C. A., Pegion, K., Vecchi, G. A., Alexander, M. A., Tommasi, D., Bond, N. A., et al.  
455 (2015) Seasonal sea surface temperature anomaly prediction for coastal ecosystems. *Prog.*  
456 *Oceanogr.*, 137, 219-236.
- 457 Thomas, E. E., Vimont, D. J., Newman, M., Penland, C., & Martinez-Villalobos, C., 2018: The  
458 role of stochastic forcing in generating ENSO diversity. *J. Climate*, 31, 9125-9149, doi:  
459 10.1175/JCLI-D-17-0582.1.
- 460 Tippett, M. K., L'Heureux, M. L., Becker, E. J. & Kumar, A. (2020), Excessive momentum and  
461 false alarms in late-spring ENSO forecasts. *Geophys. Res. Lett.*, 47, doi:  
462 10.1029/2020GL087008.
- 463 Trenberth, K. E., & Shea, D. J. (2006), Atlantic hurricanes and natural variability in 2005.  
464 *Geophys. Res. Lett.*, 33, L12704, doi:10.1029/2006GL026894.
- 465 Vimont, D. J., Alexander, M., & Newman, M. (2014), Optimal growth of central and east Pacific  
466 ENSO events. *Geophys. Res. Lett.*, 41, 4027–4034.
- 467 Zanna, L. (2012), Forecast skill and predictability of observed Atlantic sea surface temperatures.  
468 *J. Climate*, 25, 5047–5056.

469 Zwiers, F. W., & von Storch, H. (1995), Taking serial correlation into account in tests of the  
470 mean. *J. Climate*, 8, 336-351.  
471