

Linear Predictability: A Sea Surface Height Case Study

MAIKE SONNEWALD

Massachusetts Institute of Technology, Cambridge, Massachusetts

CARL WUNSCH

Massachusetts Institute of Technology, and Harvard University, Cambridge, Massachusetts

PATRICK HEIMBACH

The University of Texas at Austin, Austin, Texas, and Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 3 March 2017, in final form 14 November 2017)

ABSTRACT

A benchmark of linear predictability of sea surface height (SSH) globally is presented, complementing more complicated studies of SSH predictability. Twenty years of the Estimating the Circulation and Climate of the Ocean (ECCOV4) state estimate (1992–2011) are used, fitting autoregressive moving average [ARMA(n, m)] models where the order of the coefficients is chosen by the Akaike information criteria (AIC). Up to 50% of the ocean SSH variability is dominated by the seasonal signal. The variance accounted for by the nonseasonal SSH is particularly distinct in the Southern and Pacific Oceans, containing >95% of the total SSH variance, and the expected prediction error growth takes a few months to reach a threshold of 1 cm. Isolated regions take 12 months or more to cross an accuracy threshold of 1 cm. Including the trend significantly increases the time taken to reach the threshold, particularly in the South Pacific. Annual averaging has expected prediction error growth of a few years to reach a threshold of 1 cm. Including the trend mainly increases the time taken to reach the threshold, but the time series is short and noisy.

1. Motivation

The variability and change of future sea surface height (SSH, denoted η) is the center of much of the concern about the ongoing global warming. Understanding and predicting these key values, globally and regionally, involves projection and space–time integration of the numerous factors that influence SSH. These factors include the wind field, atmospheric pressure, tides, ice melt, river runoff, heat and freshwater exchange, and the shifting ocean circulation itself (Parker 1992; Church et al. 2013). The diverse physics spans a large range of time scales for oceanic response (e.g., Wunsch 2015).

Compared to the atmosphere, most relevant oceanic time scales are very long, ranging from months to thousands

of years. The presence of that long time scale (long memory) and the observed small perturbations in the oceanic state suggest that many of the major components determining future values of η can be predicted from a knowledge of the present and past states of the ocean. The expected prediction error (PE) growth of η is not well established. Attempting to estimate the PE via ensembles of climate model simulations reveals large ensemble spread [e.g., the Intergovernmental Panel on Climate Change (IPCC)] (Church et al. 2013; Stainforth et al. 2005; Palmer 2012).

The goal here is to assess quantitatively the extent to which η variability is predictable using linear methods, describing both the deterministic (seasonal) changes as well as the underlying continuum treated as a wide-sense stationary linear process.¹ As discussed, for example, by

 Denotes content that is immediately available upon publication as open access.

¹“Wide sense” stationarity is the terminology of electrical engineering; mathematicians call it “weakly” stationary, and in both cases, only the first two moments (mean and variance) are assumed time independent (Priestley 1981).

Corresponding author: Maike Sonnewald, maike_s@mit.edu

DOI: 10.1175/JCLI-D-17-0142.1

© 2018 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Wunsch (2013), such an approach provides a baseline against which predictions made with considerably more complex methods (nonlinear, nonstationary, extended to spatial structure) can be compared. The general case involves much more complex computations and raises the purely practical issue of whether the linear, univariate, stationary approach is adequate for SSH, and for how long.

An extensive body of literature explores the variability of η with varying degrees of complication, ranging from elementary statistics to the application of hierarchies of general circulation models (GCMs). These methods have varying degrees of regional success (Gille 1994; Chowdhury et al. 2007; Melillo et al. 2014) or global application (Rahmstorf et al. 2012; Church et al. 2013). The purely statistical approach is less concerned with capturing the underlying physics, while the GCM approach treats the η field as the deterministic integrated sum of ocean and atmospheric physics. This present study uses the simplest statistical approach to present a benchmark for more complex studies.

Treating oceanic change as linear may be counterintuitive. However, the modern observational record shows no major shifts in the large-scale baroclinic structure of the ocean (e.g., Roemmich et al. 2012). Apart from small regions of sea ice or convection, well-understood theory also supports the inference of only perturbation changes over periods from decades to centuries (Hirschi et al. 2013).

Interpretation of statistics from short records is difficult (see, e.g., Wunsch 1999; Percival et al. 2001; Ocaña et al. 2016). The methods that underlie much of what is presented here rely on the assumptions that η changes from the superposition of deterministic seasonal components and from a wide-sense stationary stochastic process. Of most relevance to the latter are general red noise processes and the extreme of white noise, which is, by definition, linearly unpredictable. Detection of true nonstationarity is not possible with the short records at hand. Similarly, an infinite number of generalizations to nonlinear representations are possible, but unless the linear assumption can be excluded, it remains an important reference point.

Local and global predictability are in many ways distinct; for example, regional variability in η has been attributed to shifts in wind features, tropical modes, and features such as the North Atlantic Oscillation (Yin and Goddard 2013; Roberts et al. 2016). Here, the approach is that of a univariate “black box,” with the underlying mechanisms (e.g., determining the changing global mean of η) having been discussed in many published papers (Parker 1992; Piecuch and Ponte 2011; Forget and Ponte 2015; Ocaña et al. 2016). The oceans store

large portions of the added heat from global warming, and land ice is retreating, along with other external forcings, but discussion of these specific physical contributions as functions of time and position is postponed.

The methods are detailed in section 2, and the results are presented in section 3, where the seasonal and nonseasonal contributions to the variance of η are presented. As defined in this paper, the seasonal component is perfectly predictable, and the nonseasonal portion involves stochastic forecasting. A set of four reestimates is presented: 1) using monthly or annual means of η , 2) with apparent linear trends included as part of the background red noise, and 3) and 4) with the trends removed in both cases. Section 4 presents the discussion and conclusion.

2. Numerical and ARMA models

Predictability of η is studied using the ECCOv4 global bidecadal state estimate, as described by Wunsch and Heimbach (2013), Forget et al. (2015), and others (see also ECCO Consortium 2017a,b). The state estimate is global, with latitudinal 1° resolution with tropical mesh refinement. A least squares with Lagrange multipliers approach is used to obtain the state estimate. The result is an adjusted, yet free-running, version of the MIT general circulation model (MITgcm; Adcroft et al. 2004). In contrast to most “reanalysis” products, the ECCO oceanic state satisfies basic conservation laws for enthalpy, salt, volume, and momentum, while remaining largely within error estimates of a diverse set of global data (Wunsch and Heimbach 2007, 2013; Stammer et al. 2016). Regions without data are filled in a dynamically consistent way, avoiding the use of untested statistical hypotheses (e.g., Reynolds et al. 2013).

At each point of latitude and longitude (θ, λ) of the state estimate, the temporal mean (1992–2011) is removed, and $\eta(\theta, \lambda)$ is defined as $\eta = \eta'(\theta, \lambda) + \bar{\eta}(\theta, \lambda)$, where $\bar{\eta}(\theta, \lambda)$ denotes the seasonal and $\eta'(\theta, \lambda)$ the nonseasonal $\eta(\theta, \lambda)$. Throughout this study, each θ, λ coordinate is used, and, for simplicity, the spatial indices are dropped hereafter. At this stage, any temporal trend is being included as part of η' .

The seasonal component here includes its first two harmonics, as illustrated in Fig. 1. Variances of the seasonal and nonseasonal components are additive:

$$\sigma_{\eta}^2 = \sigma_{\eta'}^2 + \sigma_{\bar{\eta}}^2. \quad (1)$$

To the extent that $\sigma_{\bar{\eta}}^2 \gg \sigma_{\eta'}^2$, useful prediction is purely deterministic. When the seasonal variability is not dominant, the predictability of the nonseasonal process

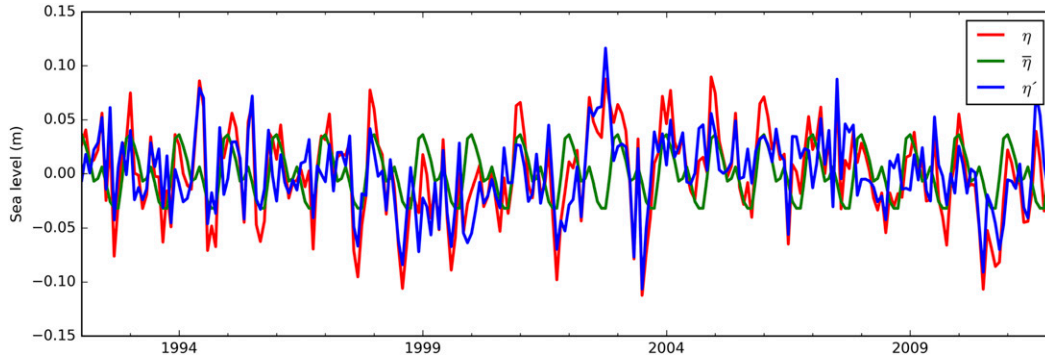


FIG. 1. Example of the process of removing the seasonal signal at 67°S, 149°W. The mean (1992–2011) is removed, and the green line illustrates the fitted seasonal model $\bar{\eta}$. The blue line is the nonseasoned data η' , while the red line shows the original estimate from ECCOV4 η .

has to be examined. Deterministic prediction of sinusoidal components is straightforward.

Linear predictability of wide-sense stationary stochastic processes, not distinguishable from Gaussian, is well understood, with a very large literature including standard textbooks (e.g., Box and Jenkins 1970). Here, the formalism is discussed only insofar as it develops the notation to be applied. Most linear methods are based on the autoregressive (AR) process of order n [AR(n)], the moving average (MA) process of order m [MA(m)], or the mixed autoregressive moving average (ARMA) of order n, m [ARMA(n, m)].

As the textbooks show (Box and Jenkins 1970), these representations are interchangeable, with a choice being mainly one of convenience, efficiency of representation, or a combination of those two. If $\zeta(t)$ is a zero-mean wide-sense stationary time series, two of the representations are

$$\begin{aligned}
 \text{AR}(n): \zeta(t) &= \overbrace{a_1}^{\text{coeff.}} \zeta(t-1) + a_2 \zeta(t-2) \\
 &+ \dots + a_n \zeta(t-n) + \overbrace{e(t)}^{\text{white noise}} \quad \text{and} \quad (2) \\
 \text{MA}(m): \zeta(t) &= e(t) + \underbrace{b_1}_{\text{coeff.}} e(t-1) + b_2 e(t-2) \\
 &+ \dots + b_m e(t-m). \quad (3)
 \end{aligned}$$

A combination gives the general ARMA(n, m) model:

$$\begin{aligned}
 \zeta(t) &= a_1 \zeta(t-1) + a_2 \zeta(t-2) + \dots + a_n \zeta(t-n) + e(t) \\
 &+ b_1 e(t-1) + b_2 e(t-2) + \dots + b_m e(t-m), \quad (4)
 \end{aligned}$$

where a_i and b_i are regression coefficients; $e(t)$ is near-Gaussian white noise with zero mean and variance σ_e^2 ; and t is any time, past, present, or future, measured in

units producing an implied time step $\Delta t = 1$. Conversion of one form to another or to the mixed representation is discussed in textbooks (e.g., Box and Jenkins 1970). In the MA form, the white noise increments $e(t)$ are determined for past values leading up to the present time t . Parameter $e(t)$ is known (estimated), but no future values $e(t + \Delta t)$, $\Delta t > 0$ are known. In the AR form, past values $\zeta(\tau)$ are assumed to have been estimated, as is $e(t)$, but again, no future values are available. In the presence of noise, these representations can become unstable, being indistinguishable from apparent nonstationarity. Tests for stability/nonstationarity are based upon the zeros and poles of complex polynomials formed from the various coefficients a_i, b_j (Box and Jenkins 1970).

The MA form gives the simplest representation of the growth of prediction error, from Eq. (5), although the final growth rate is the same for all consistent ARMA forms. (All results here are based on the conversion to the MA after determination of the more general ARMA.) Converting the ARMA to the MA form using the Wold representation, the τ -ahead PE is

$$\underbrace{\langle [\hat{\zeta}(t+\tau) - \zeta(t+\tau)]^2 \rangle}_{\text{prediction}} = \underbrace{\sigma_e^2}_{\text{noise variance}} \sum_{p=0}^{\tau} \underbrace{b_p^2}_{\text{coeff.}}, b_0 = 1. \quad (5)$$

This quadratic error growth depends upon the values of b_p , whose sum can never exceed the time series variance $\langle \zeta^2 \rangle$, for which the best prediction would be the time mean.

The performance of the ARMA(n, m) is assessed in terms of the PE growth over time. This criterion is expressed as the time it takes the error to grow beyond a given threshold, and good model performance refers to a relatively small PE at a specific time.

In practice, regression coefficients a_i, b_j are most often found using one of several versions of least squares

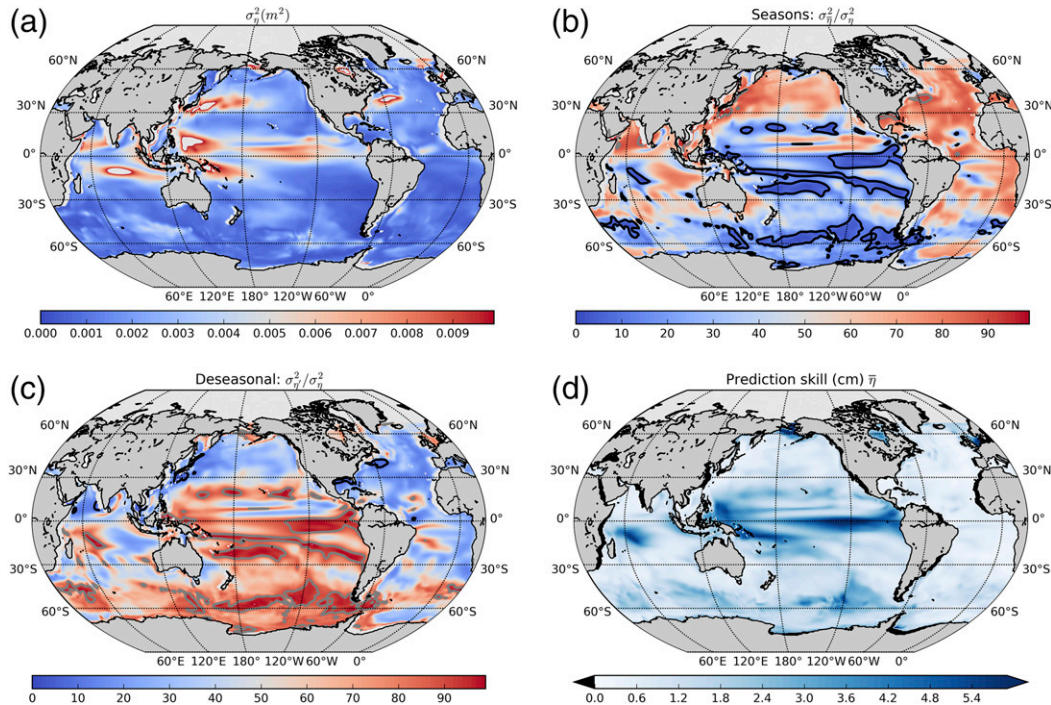


FIG. 2. From ECCOV4 of monthly averages from 1992 to 2011. (a) The variance of η (m^2)—note the high variance associated with equatorial, western boundary currents, and monsoonal regions. (b) The percentage of variance contained in $\bar{\eta}$ ($\sigma_{\bar{\eta}}^2/\sigma_{\eta}^2$). (c) The percentage of variance contained in η' ($\sigma_{\eta'}^2/\sigma_{\eta}^2$). Contours correspond to 5% (black) and 95% (gray) of values. (d) The amplitude of the variance (cm) the $\bar{\eta}$ contributes toward the total variance of η .

in which autocovariances are estimated along the way. The main difficulty is determining the orders n , m for the particular representation. Orders are increased incrementally until a stopping criterion is met (see, e.g., Akaike 1973; Hughes and Williams 2010; Aho et al. 2014). Adding regression parameters improves the fit to the data but risks overfitting. The Akaike information criteria (AIC) is used here—minimizing the expectation of the PE where k is the number of parameters:

$$AIC = 2k - 2 \ln(\mathcal{L}),$$

where \mathcal{L} is the likelihood:

$$\mathcal{L} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\sum_{i=1}^N \frac{(\zeta_i - \hat{\zeta}_i)^2}{2\sigma^2} \right].$$

Parameter ζ_i is the observed, and $\hat{\zeta}_i$ is the prediction, so $(\zeta_i - \hat{\zeta}_i)^2$ are the prediction residuals. In the estimate, the AIC value is minimized, which determines the smallest appropriate order to represent the time series. As discussed by Priestley (1981) and Yang (2005), the AIC can overestimate the order; see the appendix for more detail.

In the following discussion of the SSH time series, two cases are considered: one where only a time mean has

been removed (η), and one where a best-fitting linear trend has been subtracted as well (η^\dagger). Separate analyses of η^\dagger , using both monthly and annual mean time series, are considered. Further removing a seasonal cycle $\bar{\eta}$ leads to time series η' and $\eta^{\dagger\prime}$, respectively. A significant linear or quadratic trend can, itself, be used to make a prediction. By including trend structures in the stochastic process, the predictability of the time series will be enhanced.

3. Results

a. Seasonal variance

Figure 2a shows the total variance σ_{η}^2 of η from monthly means of 1992–2011. Western boundary currents and their extensions are associated with higher variance, particularly in the Northern Hemisphere. The tropics have a large SSH variance, particularly in the Pacific Ocean. The Indian Ocean is dominated by monsoonal effects, particularly in the Arabian Sea warm pool, and complicated interactions of jet dynamics and the Indonesian throughflow seen particularly around 15°S (Schott et al. 2001). Elevated standard deviations occur in the eastern Indian Ocean that are not seen in the Atlantic or Pacific Oceans. The Southern Ocean shows some excess variance, particularly in the Indian

and Pacific sectors. Bathymetric features, such as the Pacific–Antarctic Ridge, are associated with higher variance (Ponte and Piecuch 2014).

Figure 2b illustrates the percentage of the total variance included in the $\overline{\eta}$ component. A striking, but well-known, seasonal hemispheric difference appears across the equator (Pattullo et al. 1955), where much of the η variance in the Northern Hemisphere is dominated by the seasonal component. Exceptions include a large zonal band in the Pacific and areas at a similar latitude in the North Atlantic. The Irminger and Labrador Seas also have large areas where the seasonal signal is less dominant, as is also true of the Bering Sea. The South Pacific Ocean has large areas where the seasonal signal is weak, but a signal extends westward off the coast of Peru, likely associated with the upwelling there. A similar feature is seen in the East Australian Current. In the Indian Ocean, fluctuations in the Arabian Sea are almost entirely captured by the seasonal monsoonal component. South of the equator, the eastern half of the Indian Ocean in the Southern Hemisphere is largely dominated by $\overline{\eta}$, but less clearly so in the west. In the Southern Ocean, a clear, seasonal dominance appears in the Weddell Gyre region, as well as in the area of the standing meander off the Agulhas coast of South Africa.

b. Seasonal prediction

The seasonal component $\overline{\eta}$ is predictable, as defined over the bidecadal time interval covered by ECCOV4, and Fig. 2d illustrates the associated contribution to the prediction of the standard deviation of the seasonal component in cm over 1992–2011. As expected, where the seasonal variance is a large fraction of the total, good predictability is found (e.g., associated with the seasonal component in the Mascarene Basin area, as well as in the Southern Ocean in the Indian and Pacific Ocean sectors).

c. Nonseasonal variance

Figure 2c shows the percentage of the total variance accounted for by the nonseasonal background process. Here, a visually striking signal appears across the Pacific Ocean along the equator, probably associated with the El Niño–Southern Oscillation (ENSO) climate mode. Further zonal bands appear to the north of the equator. A strong signal occurs in the Bering Strait region. In the Atlantic Ocean, the subtropical regions show active areas, as well as along the paths of the Labrador and southern tip of the East Greenland Currents. In the South Atlantic, the nonseasonal component of variance accounts for a large portion of the variance in the Brazil Current and the Zapiola region of the Argentine Basin,

as well as in zonal bands. In the Indian Ocean, the nonseasonal component of variance also accounts for a large portion of the total variance to the west, in the Mascarene Basin region, as well as along the western Indonesian coast associated with the propagation of the throughflow. The Southern Ocean produces a very large signal associated with regions of deep mixed layers and possible mode water formation, as well as in areas where the Antarctic Circumpolar Current (ACC) is directed southward. Predictability associated with this nonseasonal variance is addressed in the remainder of this paper.

d. Predictability after trend removal

A linear trend is now removed from the SSH values, meaning that a possibly perfectly predictable component is eliminated. The η^{fi} is fit to an ARMA(n, m) process of the ECCOV4 state estimate from 1992 to 2011. A four-point smoother is applied to η^{fi} to reduce noise (two points in latitude and two in longitude), equivalent to moving from the tracer (t point) to the vorticity point (f point) in the Arakawa C grid. Using the smoother tends to make the data adhere more closely to a normal distribution, but it can exaggerate the spatial covariance of isolated outliers. The performance of the different choices is given by the rate of the PE growth over time. Figure 3a shows the order n of the ARMA(n, m) chosen with the AIC. The order chosen shows how many coefficients are used to optimally represent η^{fi} and the associated prediction error.

The simplest linear theory assumes that the underlying values are Gaussian, or close to it, an assumption tested in the ECCO estimated SSH in Fig. 4a using the Shapiro–Wilk test for normality (Shapiro and Wilk 1965). Large areas associated with features such as the ENSO signal appear to deviate from normality (i.e., p values close to 0). This result has implications for the predictability because these departures are important when interpreting the PE.

Figure 5 shows the error growth asymptoting to its upper bound: the full variance of the background residual time series. Large differences as a function of region appear in the PE, as well as in their asymptotic rate of growth. The expected error e -folding structure is shown to illustrate the rate of predictability decay, independent of magnitude.

The ARMA(n, m) expected PE growth associated with the η^{fi} is illustrated in Fig. 6a. The PE growth is expressed in terms of the time taken (months) before a target (1 cm) is exceeded. Large areas of the ocean show limited performance based on the PE growth over time, but certain (mostly isolated) areas have good performance, exceeding a year. In interpreting these figures, it

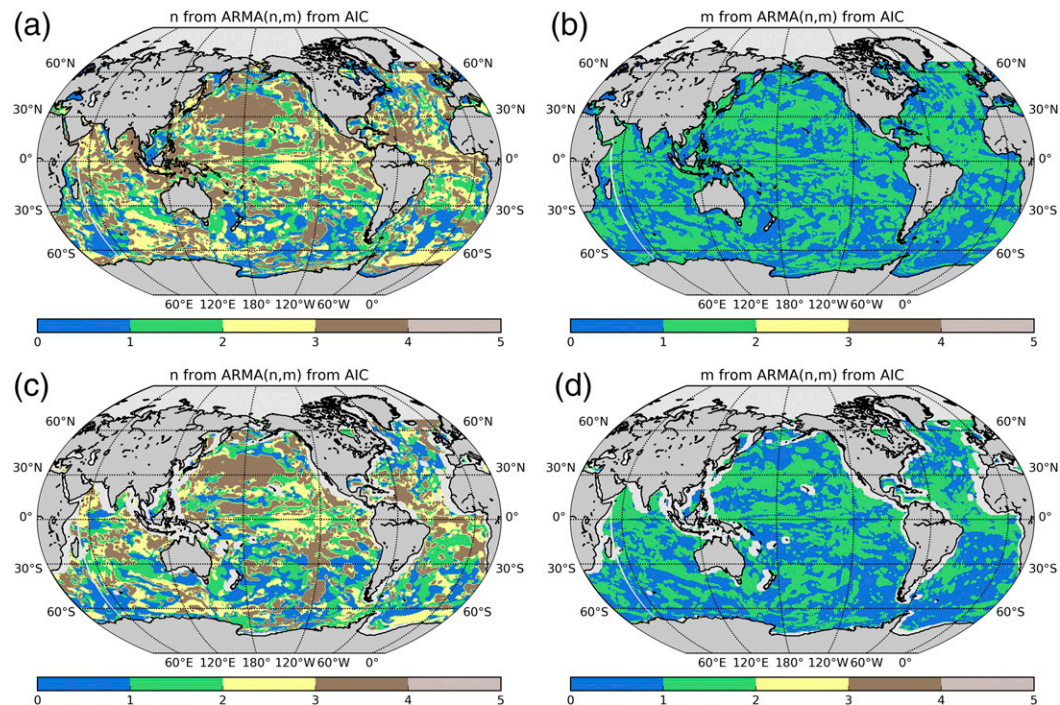


FIG. 3. The chosen order of $ARMA(n, m)$ using the AIC for the η^l over 1992–2011: (a) the n and (b) the m of the $ARMA(n, m)$ and the chosen order for nonseasoned η^l data with the linear trend not removed: (c) the n and (d) the m of the $ARMA(n, m)$.

is prudent to keep in mind the assumption of normality (Fig. 4a).

High predictability is seen in the North Pacific Ocean. Although the physics here are beyond the intended scope of this paper, these areas are associated with the Kuroshio crossing the Pacific and wave–eddy interactions stretching from Hawaii to the coast of California in a banana shape. Areas in the equatorial Pacific also show better performance of PE growth over time but are clearly nonnormal. Features associated with the Pacific–Antarctic Ridge in the Southern Ocean also produce

better performance in terms of the $ARMA(n, m)$ expected PE growth over time, along with some areas in the Irminger Sea.

e. Predictability with apparent trends

Tests of predictability are now made with the linear trend left in the time series. Including the trend treats it as an unresolved component of a red noise process. The inclusion of the trend is expected to increase the performance of the $ARMA(n, m)$ PE growth over time. One cannot distinguish this variability from a red noise

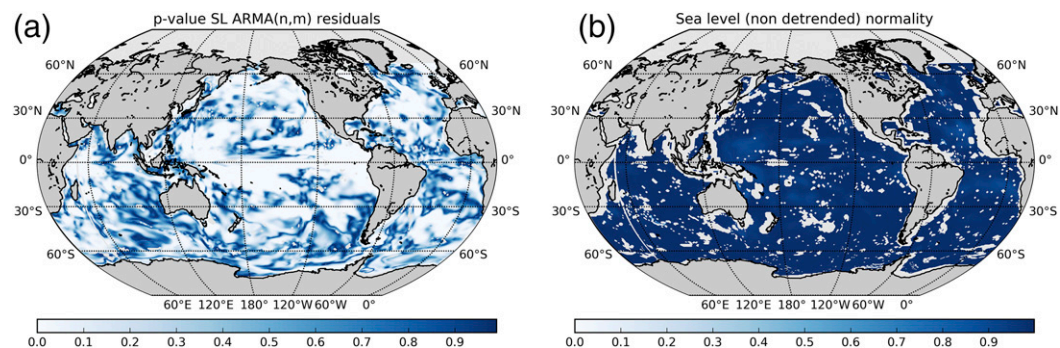


FIG. 4. The Shapiro–Wilk test for normality for (a) η^l and (b) η^l . Darker blue indicates increasing confidence in accepting the null hypothesis that the data are from a normally distributed population. White areas indicate areas where the AR coefficient roots were outside the unit circle or where the MA coefficients were noninvertible.

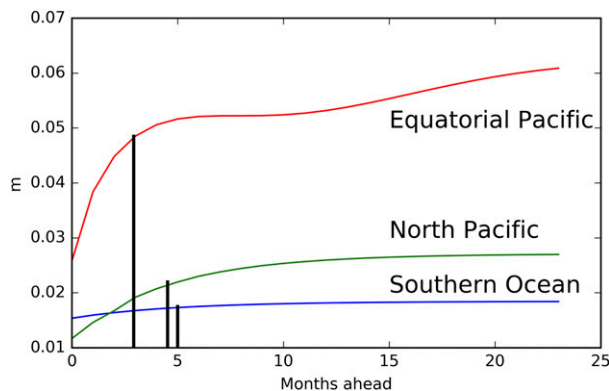


FIG. 5. The prediction error as defined in Eq. (5) for η^{\dagger} in three locations: the Southern Ocean (blue line; 50°S, 160°E), the North Pacific (green line; 44°N, 150°E), and the equatorial Pacific (red line; 2°S, 100°E). The rate of error growth varies regionally, showing the e -folding time scale (black bars).

process with existing data (Church et al. 2013; Lyu et al. 2014; Ocaña et al. 2016). To assess the impact of including the trend, analysis of the background process is repeated. Figures 3c and 3d illustrate the ARMA(n, m) order chosen by the AIC. This result is similar to Figs. 3a and 3b, but sometimes smaller values of n, m result, as is physically plausible with a trend.

The normality of the stochastic background process with the trend is retested in Fig. 4b, illustrating that most of the ocean remains indistinguishable from having a normal distribution in η^{\dagger} . Figure 6b shows the prediction performance based on the ARMA(n, m), phrasing the result in terms of the number of months it takes for the PE to cross the accuracy target of 1 cm. Retaining the trend adds predictability, with large areas taking over 12 months before exceeding the 1-cm threshold. Areas where the trend is important are in bands in the subtropics in both the Atlantic and Pacific, as well as large areas off the coast of Greenland, the Drake

Passage, the Kuroshio path across the Pacific, the southern Indian Ocean, and a remarkably large region in the South Pacific poleward of 30°S. Generally, Fig. 6b is seen to amplify Fig. 6a, but with notable exceptions. These exceptions could be associated with the spreading of a thermosteric signal, particularly the subtropical South Pacific associated with Pacific decadal oscillation dynamics. The subtropical Atlantic stands out as another region where the trend is key, along with the South Pacific. The mechanisms, particularly in terms of linear dynamics, are not clear. For example, regions associated with bathymetric features like the Pacific–Antarctic Ridge do not stand out as intuitive regions of heat storage.

f. Predictability with annual averages

Interannual and monthly physics are distinct. Assessing the annually averaged η^{\dagger} and η^{\dagger} from ECCOv4 separately, the assumptions of estimating the covariance and a Gaussian distribution are likely inaccurate, owing to the short, 20-value record.

Figure 7 shows the chosen ARMA(n, m) order. A four-point smoother is again used for reduced noise in the ECCOv4 η^{\dagger} for 1992–2011, where the trend is removed and the data are annually averaged. The n and m of the ARMA(n, m) now reflect the annually averaged data rather than the monthly. The n of the ARMA(n, m) is generally lower for the annually averaged η^{\dagger} than for monthly η^{\dagger} . However, large regions show very different patterns than in the monthly data. Examples include the Arabian Sea region and the North Pacific. However, the areas where one order dominates are generally larger than for monthly η^{\dagger} . Regions that have higher orders for yearly averaged data than for monthly averages are areas such as the Pacific sector of the Southern Ocean.

Figure 8a assesses the extent to which the 20-yr time series can be viewed as coming from a normally distributed

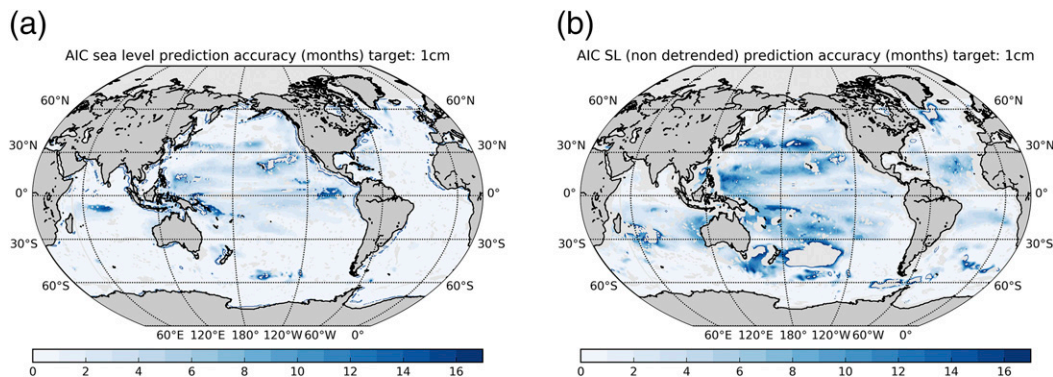


FIG. 6. The ARMA(n, m) expected prediction performance of (a) η^{\dagger} and (b) η^{\dagger} phrased in terms of the time it takes the expected prediction performance to be less than 1 cm (months). Areas saturating the color scale indicate predictability longer than 17 months.

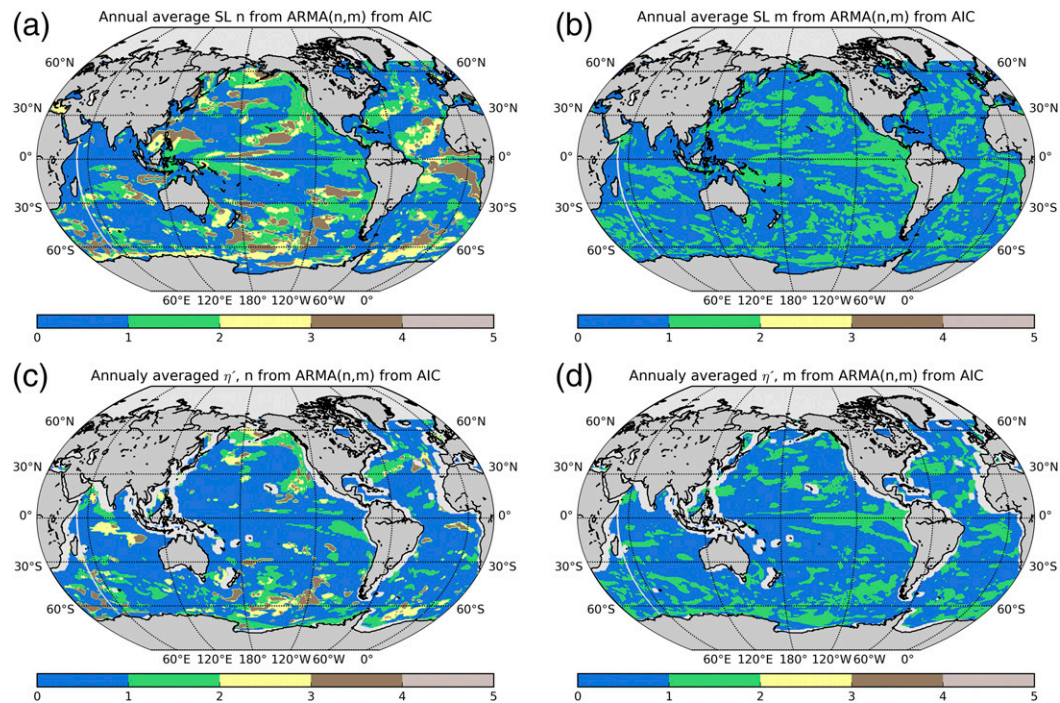


FIG. 7. Chosen order of $ARMA(n, m)$ using the AIC of nonseasoned η^t annually averaged with the linear trend removed: (a) the n and (b) the m of the $ARMA(n, m)$ and parameter η' annually averaged with the linear trend not removed: (c) the n and (d) the m of the $ARMA(n, m)$.

population when using annual averages. Most of the ocean passes this test for annually averaged η^t , but with large areas that appear noisy/nonstationary, presumably owing primarily to the presence of noise.

The associated prediction performance based on the $ARMA(n, m)$ expected PE growth over time is shown in Fig. 9a. In most areas, the time to reach the criterion of 1-cm expected error is 1 year, but areas exceeding 3 years are seen. These regions are clustered in the Pacific,

with a patch in the western equatorial Pacific, a band stretching from the central North Pacific to the coast of the United States (from Hawaii to California), and small, isolated patches elsewhere. A larger patch of good PE performance over time exists in the Pacific sector of the Southern Ocean.

As expected, the role of the linear trend is also important in the annually averaged η' . Figures 7a–c illustrate the associated orders of the $ARMA(n, m)$. As with

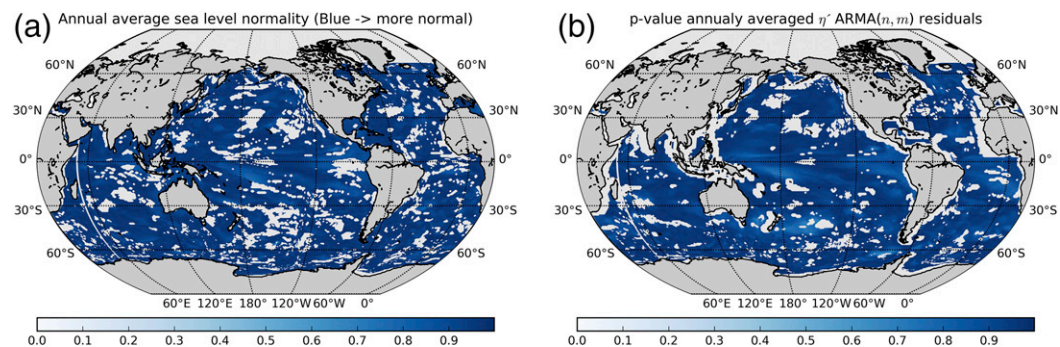


FIG. 8. The Shapiro–Wilk test for normality for the annually averaged (a) η^t and (b) η' data where the linear trend is not removed. Darker blue indicates increasing confidence in accepting the null hypothesis that the data are from a normally distributed population. White areas indicate areas where the AR coefficient roots were outside the unit circle or where the MA coefficients were noninvertible consistent with apparent nonstationarity, likely due to noise.

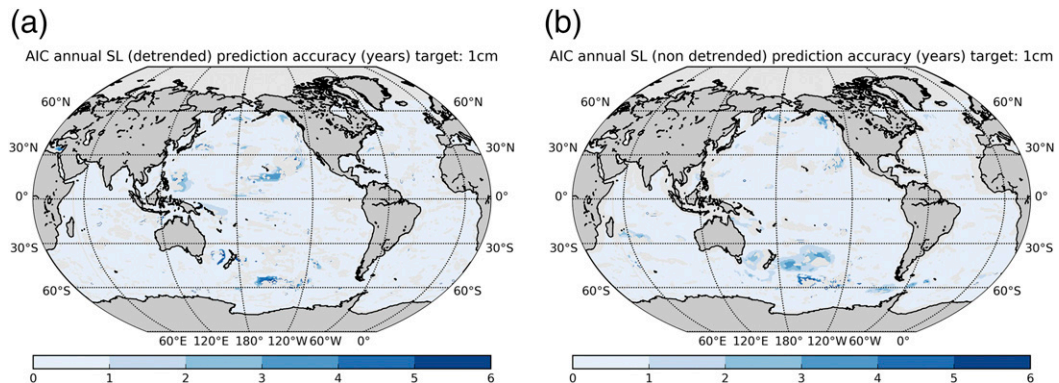


FIG. 9. The prediction performance based on the ARMA(n, m) expected prediction error for annually averaged (a) η^I and (b) η' . Performance is phrased in terms of the time it takes the prediction performance to be less than 1 cm (years). Areas saturating the color scale indicate nonstationary ARMA coefficients.

the monthly data, the orders are generally lower. Higher orders exist in a band stretching from the central Pacific northward to the United States (from Hawaii to California), as well as a feature associated with the Pacific–Antarctic Ridge.

Figure 8b, using annually averaged η' , suggests that most of the ocean passes the test of normality, but again with large areas of failure. The associated prediction performance based on the ARMA(n, m) expected PE growth over time is shown in Fig. 9b. This result is similar to that in Fig. 9a, but most of the longer-term PE performance over time is found in the South Pacific. A band stretching from the central Pacific northward to the United States still exists, but large areas do not pass the test of a stable or wide-sense stationary ARMA. Patches of longer-term PE performance over time are also seen in the Indian Ocean and Drake Passage, with isolated regions elsewhere.

4. Discussion and conclusions

In this paper, linear univariate predictability of SSH η is discussed, and benchmarks for more elaborate prediction methods are presented. In general, more complex models and prediction methods (e.g., GCM projections) would need to exceed this PE performance over time to be proven worthwhile. Prediction performance is presented here in terms of the time it takes the expected PE to grow beyond 1 cm. More complex models should necessarily do better, and their use may well be justified, particularly in specific, physically identifiable regions. This approach is supported by work such as Goddard et al. (2015), where certain events in η' have been attributed to factors such as changes in the Atlantic meridional overturning circulation and the North Atlantic Oscillation. Existing spreads in ensemble studies, such as the CMIP5 models,

would suggest that many difficulties remain (Church et al. 2013).

In the ECCOv4 state estimate, the seasonal cycle $\bar{\eta}$, seen in Fig. 2b, accounts for more than 80% of the variability over 20+ years in large parts of the Atlantic and Northern Pacific, as well as in the Weddell Gyre area in the Southern Ocean. In these regions, the seasonal component is likely sufficient for estimating the variation in η for at least a few decades.

The percentage of 20-yr variance in the stochastic, nonseasonal component η' , seen in Fig. 2c, is important over large areas of the ocean, particularly in the Southern Hemisphere. Parameter η' is treated as a weakly stationary univariate random process, with the assumption of being normally distributed. Areas where η' is particularly important for predictions, as seen in Fig. 6b, lie in the Southern Ocean and in extensive regions of the Pacific, as well as in the Indian Ocean. When a linear trend is removed, as in Fig. 6a, regions where the ARMA prediction still performs well are mainly clustered in the Pacific, particularly in a band extending northeast from Hawaii, as well as in small, isolated areas elsewhere. A patch of higher predictability exists in the Pacific sector of the Southern Ocean. However, the ARMA prediction, as expected, tends to perform less well when the trend is removed.

An important next step is to distinguish the physical mechanisms, whether atmospheric or oceanic, in regions where the ARMA prediction procedure does well and those where it works poorly. Stationary linear prediction methods can be of limited utility for a variety of reasons. These include dominance by unpredictable white noise (e.g., from atmospheric forcing), non-Gaussian forcing functions, strong nonlinearities in ocean physics, and nonstationary behavior from external forcing and lack of equilibrium in the ocean.

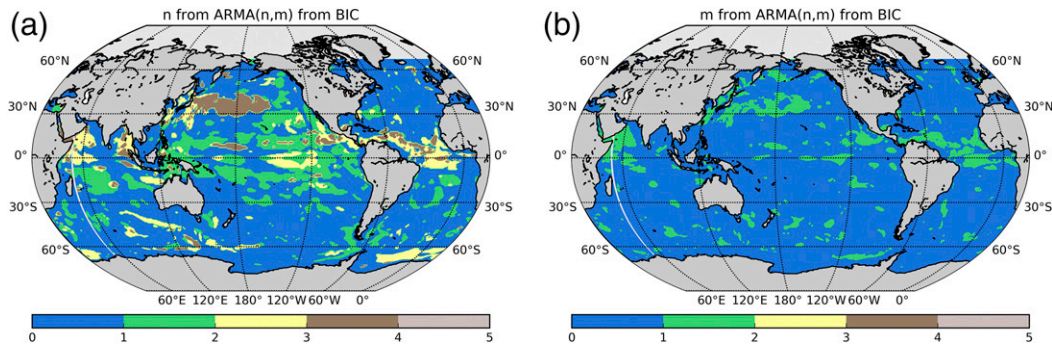


FIG. A1. Chosen order of ARMA(n, m) for annually averaged η^{\dagger} using the BIC: (a) the n and (b) the m of the ARMA (n, m). We use a four-point smoother for the η^{\dagger} for 1992–2011. Note the difference to Fig. 3, where higher orders are seen.

Investigating the contribution of specific mechanisms to the predictability structures of η is outside the scope of this study, but results presented suggest such analysis is merited. For example, work with linear models of η includes that of Hughes and Williams (2010). Using altimetry alone, and without a prediction focus, they fit AR(n) models, choosing the orders using the Bayesian information criterion (BIC). With their higher temporal resolution, the AR(n) fits are not easily compared to those from the use of the ARMA(n, m) on monthly and annual average values. They concluded that Rossby wave patterns are important within $\pm 30^{\circ}$ of the equator, while advective processes become more influential at higher latitudes, allowing features such as the Pacific–Antarctic Ridge to influence η' predictability. The ARMA fits to the ECCOv4 data also have distinct features associated with the Pacific–Antarctic Ridge and different behavior within 30° of the equator, suggesting the physics that give rise to these features are coherent across these time scales and lend themselves to linear modeling approaches.

Predictability from annually averaged data, as in Figs. 9a and 9b, proves generally different. With the trend included, a region in the western South Pacific Ocean has striking performance. Paradoxically, an increase in prediction performance is seen in a band extending northeast of Hawaii when the trend is removed. This is likely a stochastic artifact. Given the long time scales controlling oceanic physics, the records remain far too short to infer statistically stable results. In this context, the continuing difficulties, generally experienced in distinguishing the lowest frequencies present between a general red noise process and a true secular trend of multidecadal applicability, remain a major issue. Whether unconstrained models, such as the CMIP5 ones used by Lyu et al. (2014), have true prediction skill remains unknown. Note, too, that the univariate approach used here is readily extended to accommodate multivariate predictive models employing correlated spatial

structures of many different types, which may work much more effectively in some areas.

In brief summary, the present study produces a benchmark of univariate linear skill in predicting η . Figure 2b illustrates that up to 50% of the ocean η variability is accounted for $>80\%$ using only the seasonal signal over the 20 years of the ECCO state estimate. The remaining ocean η variability has a significant stochastic component, with expected prediction error growth largely taking over 2 months to exceed 1 cm. Figure 6b shows that treating the linear trend as part of the continuum enhances the predictive performance, as expected. In areas in the Southern and Pacific Oceans, the stochastic continuum η' contains more than 95% of the total variance of η , with expected prediction performance of 1 cm exceeding a year in significant portions of these regions. Moving forward, extending in time the global measurements of η and understanding the underlying physical processes remains the key to progress in regional sea surface height prediction.

Acknowledgments. This work was funded by the U.S. National Aeronautics and Space Administration Sea Level Change Team (contract NNX14AJ51G) and through the ECCO Consortium funding via the Jet Propulsion Laboratory.

APPENDIX

Influence of Chosen Information Criteria

The choice of information criteria to determine the ARMA(n, m) significantly influences the predictive performance. The AIC [Eq. (2)] was used, as it demonstrated better performance for applications where the “true” model is likely not available (Yang 2005). The BIC has been used in studies such as Hughes and Williams (2010). The BIC is defined as

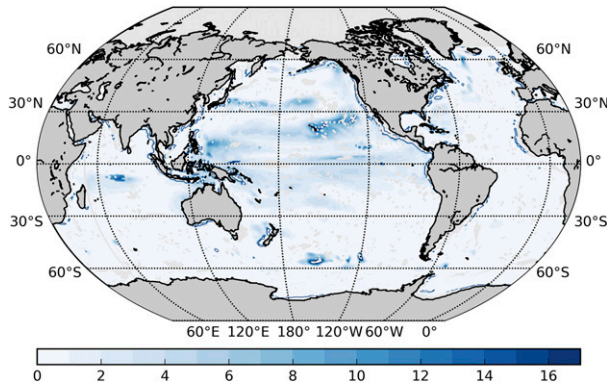


FIG. A2. The ARMA(n, m) expected prediction performance with BIC estimating the order of ARMA(n, m) models of η^{\dagger} phrased in terms of the time it takes the prediction performance to be less than 1 cm (months).

$$\text{BIC} = k \ln(n) - 2 \ln(\mathcal{L}),$$

where n is the number of data points, k is the number of parameters, and \mathcal{L} is the likelihood shown in Eq. (2).

As discussed by Priestley (1981) and Yang (2005), the AIC tends to overestimate the true order, and the BIC tends to underestimate it. The AIC results showed better predictive power. The two criteria give different weights to penalizing the number of regression coefficients, with the BIC having a larger penalty term.

Figure A1 illustrates the orders chosen by the BIC for η^{\dagger} . Note the difference from Fig. 3, where higher orders are chosen using the AIC. Overall, the AIC produces estimates at higher orders than the BIC, but results for large areas are similar (e.g., the North Pacific and equatorial Atlantic). The information criteria, particularly the AIC, tend to pick out differing dynamical regions.

The predictive potential associated with the η^{\dagger} using the BIC is illustrated in Fig. A2. This shows a very similar pattern to Fig. 6a, and we highlight the differences by showing the difference between the two (AIC – BIC) in Fig. A3a. Here, we see that the AIC largely gives higher predictive performance based on the ARMA(n, m) expected prediction error growth. The difference between the AIC and BIC prediction performance is highlighted in Fig. A3b. We see slight differences, with the AIC having higher predictive performance in bands $\pm 30^{\circ}$ of the equator, particularly in the Pacific and Indian Oceans, but also in the Atlantic. The BIC shows somewhat higher predictive performance in the higher latitudes, particularly in the Pacific.

For annually averaged η' , the difference between using the AIC and BIC is smaller. Figure A3c highlights the difference between the AIC and BIC criteria, showing that overall, the AIC has higher performance based on the ARMA(n, m) expected prediction error

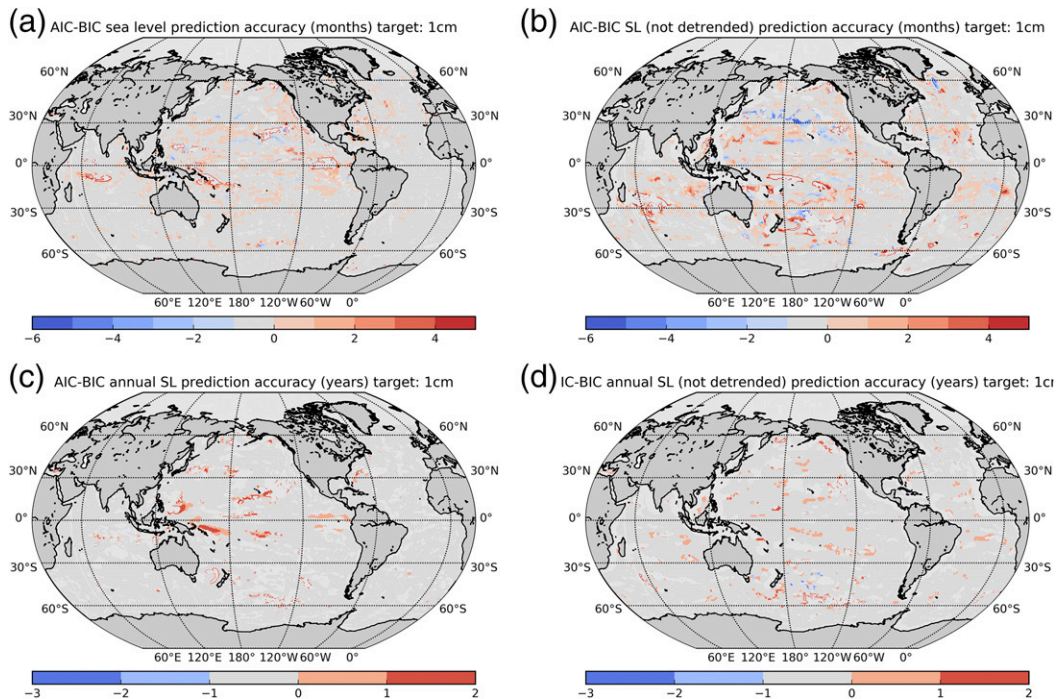


FIG. A3. The prediction performance difference with AIC – BIC using monthly averaged (a) η^{\dagger} (months), (b) monthly averaged η' (months), (c) annually averaged η^{\dagger} (years), and (d) annually averaged η' (years). Note that the AIC offers better predictability (red predominates) in most areas, but some prefer the BIC (blue).

growth. With annually averaged η^{\dagger} , the AIC and BIC also give similar prediction accuracies, with the differences highlighted in the difference plot shown in Fig. A3d. However, small areas show higher performance based on ARMA(n, m) expected prediction error growth with the BIC.

The difference in predictability with the AIC and BIC informing the choice of the order suggests that the AIC has the highest utility. The BIC underestimates the order, and the AIC is found to be more suitable. This is illustrated in detail, looking at the predictability with the apparent trend in η' , in Fig. A3d and is demonstrated throughout with the higher prediction performance using the AIC. The AIC is better particularly in the equatorial regions. The difference between the performance of the prediction is larger in the monthly data with the trend. For the case that is not detrended, the BIC occasionally does better than the AIC, but no obvious spatial pattern is apparent. The differences are larger for the annually averaged data, as the chosen ARMA(n, m) orders would suggest. The higher orders chosen by the AIC are not surprising, as the AIC penalizes adding parameters less strongly than the BIC. Burnham and Anderson (2002) show that the AIC can actually be derived from the BIC using a different prior in the Bayesian framework. They suggest the AIC has advantages over the BIC, first being based on information theory and, second, having a more sensible prior. Our results are in agreement with Burnham and Anderson (2002) and similar work by Yang (2005), suggesting better performance based on the ARMA(n, m) expected prediction error growth is achieved using the AIC. The AIC has been seen to have higher performance than the BIC, as is discussed further by Burnham and Anderson (2004) and Aho et al. (2014). The implications of how well the different models capture the different dynamical regimes is not discussed, as this relies on large generalizations of the prediction performance of the fitted ARMA(n, m) over vast areas.

REFERENCES

- Adcroft, A., C. Hill, J. M. Campin, J. Marshall, and P. Heimbach, 2004: Overview of the formulation and numerics of the MIT GCM. *Proc. ECMWF Seminar Series on Numerical Methods, Recent Developments in Numerical Methods for Atmosphere and Ocean Modelling*, Reading, United Kingdom, ECMWF, 139–150.
- Aho, K., D. Derryberry, and T. Peterson, 2014: Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, **95**, 631–636, <https://doi.org/10.1890/13-1452.1>.
- Akaike, H., 1973: Information theory and an extension of the maximum likelihood principle. *Proc. Second Int. Symp. on Information Theory*, Budapest, Hungary, Institute of Electrical and Electronics Engineers, 267–281.
- Box, G., and G. Jenkins, 1970: *Time Series Analysis: Forecasting and Control*. Holden-Day, 553 pp.
- Burnham, K. P., and D. R. Anderson, 2002: *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. 2nd ed. Springer-Verlag, 488 pp., <https://doi.org/10.1007/b97636>.
- , and —, 2004: Multimodel inference: Understanding AIC and BIC in model selection. *Sociol. Methods Res.*, **33**, 261–304, <https://doi.org/10.1177/0049124104268644>.
- Chowdhury, M. R., P.-S. Chu, T. Schroeder, and N. Colasacco, 2007: Seasonal sea-level forecasts by canonical correlation analysis—An operational scheme for the U.S.-affiliated Pacific Islands. *Int. J. Climatol.*, **27**, 1389–1402, <https://doi.org/10.1002/joc.1474>.
- Church, J. A., and Coauthors, 2013: Sea level change. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1137–1216.
- ECCO Consortium, 2017a: A twenty-year dynamical oceanic climatology: 1994–2013. Part 1: Active scalar fields: Temperature, salinity, dynamic topography, mixed-layer depth, bottom pressure. ECCO Consortium Rep., 54 pp.
- , 2017b: A twenty-year dynamical oceanic climatology: 1994–2013. Part 2: Velocities, property transports, meteorological variables, mixing coefficients. ECCO Consortium Rep., 45 pp.
- Forget, G., and R. M. Ponte, 2015: The partition of regional sea level variability. *Prog. Oceanogr.*, **137**, 173–195, <https://doi.org/10.1016/j.pocean.2015.06.002>.
- , J.-M. Campin, P. Heimbach, C. N. Hill, R. M. Ponte, and C. Wunsch, 2015: ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geosci. Model Dev.*, **8**, 3071–3104, <https://doi.org/10.5194/gmd-8-3071-2015>.
- Gille, S. T., 1994: Mean sea surface height of the Antarctic Circumpolar Current from Geosat data: Method and application. *J. Geophys. Res.*, **99**, 18 255–18 273, <https://doi.org/10.1029/94JC01172>.
- Goddard, P. B., J. Yin, S. M. Griffies, and S. Zhang, 2015: An extreme event of sea-level rise along the northeast coast of North America in 2009–2010. *Nat. Commun.*, **6**, 6346, <https://doi.org/10.1038/ncomms7346>.
- Hirschi, J. J.-M., A. T. Blaker, B. Sinha, A. Coward, B. de Cuevas, S. Alderson, and G. Madec, 2013: Chaotic variability of the meridional overturning circulation on subannual to interannual timescales. *Ocean Sci.*, **9**, 805–823, <https://doi.org/10.5194/os-9-805-2013>.
- Hughes, C. W., and S. D. P. Williams, 2010: The color of sea level: Importance of spatial variations in spectral shape for assessing the significance of trends. *J. Geophys. Res.*, **115**, C10048, <https://doi.org/10.1029/2010JC006102>.
- Lyu, K., X. Zhang, J. A. Church, A. B. A. Slangen, and J. Hu, 2014: Time of emergence for regional sea-level change. *Nat. Climate Change*, **4**, 1006–1010, <https://doi.org/10.1038/nclimate2397>.
- Melillo, J. M., T. C. Richmond, and G. W. Yohe, Eds., 2014: Climate change impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program Rep., 841 pp., <https://doi.org/10.7930/J0Z31WJ2>.
- Ocaña, V., E. Zorita, and P. Heimbach, 2016: Stochastic secular trends in sea level rise. *J. Geophys. Res. Oceans*, **121**, 2183–2202, <https://doi.org/10.1002/2015JC011301>.
- Palmer, T. N., 2012: Towards the probabilistic Earth-system simulator: A vision for the future of climate and weather prediction. *Quart. J. Roy. Meteor. Soc.*, **138**, 841–861, <https://doi.org/10.1002/qj.1923>.

- Parker, B., 1992: Sea level as an indicator of climate and global change. *Mar. Technol. Soc. J.*, **25**, 13–24.
- Pattullo, J., W. Munk, R. Revelle, and E. Strong, 1955: The seasonal oscillation in sea level. *J. Mar. Res.*, **14**, 88–155.
- Percival, D. B., J. E. Overland, and H. O. Mofjeld, 2001: Interpretation of North Pacific variability as a short- and long-memory process. *J. Climate*, **14**, 4545–4559, [https://doi.org/10.1175/1520-0442\(2001\)014<4545:IONPVA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<4545:IONPVA>2.0.CO;2).
- Piecuch, C. G., and R. M. Ponte, 2011: Mechanisms of interannual steric sea level variability. *Geophys. Res. Lett.*, **38**, L15605, <https://doi.org/10.1029/2011GL048440>.
- Ponte, R. M., and C. G. Piecuch, 2014: Interannual bottom pressure signals in the Australian–Antarctic and Bellingshausen Basins. *J. Phys. Oceanogr.*, **44**, 1456–1465, <https://doi.org/10.1175/JPO-D-13-0223.1>.
- Priestley, M. B., 1981: *Spectral Analysis and Time Series, Volumes 1–2*. Academic Press, 890 pp.
- Rahmstorf, S., M. Perrette, and M. Vermeer, 2012: Testing the robustness of semi-empirical sea level projections. *Climate Dyn.*, **39**, 861–875, <https://doi.org/10.1007/s00382-011-1226-7>.
- Reynolds, R. W., D. B. Chelton, J. Roberts-Jones, M. J. Martin, D. Menemenlis, and C. J. Merchant, 2013: Objective determination of feature resolution in two sea surface temperature analyses. *J. Climate*, **26**, 2514–2533, <https://doi.org/10.1175/JCLI-D-12-00787.1>.
- Roberts, C. D., D. Calvert, N. Dunstone, L. Hermanson, M. D. Palmer, and D. Smith, 2016: On the drivers and predictability of seasonal-to-interannual variations in regional sea level. *J. Climate*, **29**, 7565–7585, <https://doi.org/10.1175/JCLI-D-15-0886.1>.
- Roemmich, D., W. J. Gould, and J. Gilson, 2012: 135 years of global ocean warming between the *Challenger* expedition and the Argo Programme. *Nat. Climate Change*, **2**, 425–428, <https://doi.org/10.1038/nclimate1461>.
- Schott, A., P. Friedrich, and J. McCreary, 2001: The monsoon circulation of the Indian Ocean. *Prog. Oceanogr.*, **51**, 1–123, [https://doi.org/10.1016/S0079-6611\(01\)00083-0](https://doi.org/10.1016/S0079-6611(01)00083-0).
- Shapiro, S. S. and Wilk, M. B., 1965: An analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611, <https://doi.org/10.1093/biomet/52.3-4.591>.
- Stainforth, D., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406, <https://doi.org/10.1038/nature03301>.
- Stammer, D., M. Balmaseda, P. Heimbach, A. Köhl, and A. Weaver, 2016: Ocean data assimilation in support of climate applications: Status and perspectives. *Annu. Rev. Mar. Sci.*, **8**, 491–518, <https://doi.org/10.1146/annurev-marine-122414-034113>.
- Wunsch, C., 1999: The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull. Amer. Meteor. Soc.*, **80**, 245–255, [https://doi.org/10.1175/1520-0477\(1999\)080<0245:TIOSCR>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0245:TIOSCR>2.0.CO;2).
- , 2013: Covariances and linear predictability of the Atlantic Ocean. *Deep-Sea Res. II*, **85**, 228–243, <https://doi.org/10.1016/j.dsr2.2012.07.015>.
- , 2015: *Modern Observational Physical Oceanography: Understanding the Global Ocean*. Princeton University Press, 511 pp.
- , and P. Heimbach, 2007: Practical global oceanic state estimation. *Physica D*, **230**, 197–208, <https://doi.org/10.1016/j.physd.2006.09.040>.
- , and —, 2013: Dynamically and kinematically consistent global ocean circulation and ice state estimates. *Ocean Circulation and Climate: A 21 Century Perspective*, G. Siedler et al., Eds., International Geophysics Series, Vol. 103, Academic Press, 553–579.
- Yang, Y., 2005: Can the strengths of AIC and BIC be shared? *Biometrika*, **92**, 937–950, <https://doi.org/10.1093/biomet/92.4.937>.
- Yin, J., and P. B. Goddard, 2013: Oceanic control of sea level rise patterns along the East Coast of the United States. *Geophys. Res. Lett.*, **40**, 5514–5520, <https://doi.org/10.1002/2013GL057992>.