

# A Vector Autoregressive ENSO Prediction Model

DAVID CHAPMAN,\* MARK A. CANE, NAOMI HENDERSON, DONG EUN LEE, AND CHEN CHEN

*Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York*

(Manuscript received 28 April 2015, in final form 7 August 2015)

## ABSTRACT

The authors investigate a sea surface temperature anomaly (SSTA)-only vector autoregressive (VAR) model for prediction of El Niño–Southern Oscillation (ENSO). VAR generalizes the linear inverse method (LIM) framework to incorporate an extended state vector including many months of recent prior SSTA in addition to the present state. An SSTA-only VAR model implicitly captures subsurface forcing observable in the LIM residual as red noise. Optimal skill is achieved using a state vector of order 14–17 months in an exhaustive 120-yr cross-validated hindcast assessment. It is found that VAR outperforms LIM, increasing forecast skill by 3 months, in a 30-yr retrospective forecast experiment.

## 1. Introduction

Linear El Niño–Southern Oscillation (ENSO) forecasting models of low dimensionality have proven to be useful tools for prediction, analysis, and improved physical understanding. Because of their simplicity and forecast skill, the study and improvement of linear ENSO forecasting models remains an important area of research (Blumenthal 1991; Penland and Magorian 1993; Barnston et al. 1994; Penland and Sardeshmukh 1995; Jin 1997a,b; Xue et al. 2000; Newman et al. 2011; Barnston et al. 2012). A popular framework in ocean and climate sciences is the linear inverse method (LIM), a first-order approximation to a dynamical system in which the evolution operator is reduced to a matrix product plus white Gaussian noise (Penland and Magorian 1993; Penland and Sardeshmukh 1995). As the linear term dominates the evolution of tropical SSTA at interseasonal time scales, the LIM framework exhibits useful skill (Barnston et al. 1994; Penland and Magorian 1993; Penland and Sardeshmukh 1995).

Although capable of predicting El Niño, the SSTA-only LIM is not ideal, because it ignores the vital role of

the subsurface ocean (Wyrтки 1975; Cane et al. 1986). More sophisticated methods and/or additional variables are necessary to capture the subsurface interaction, because western Pacific heat content is not solely related to the present SSTA. To resolve this limitation, the LIM must be extended, either explicitly or implicitly, to include additional variables that take the subsurface ocean into account.

The most straightforward way to improve upon LIM is to include subsurface information explicitly into the model (Xue et al. 2000; Newman et al. 2011). Xue et al. (2000) report that the forecast skill of linear Markov (i.e., LIM) models can be dramatically improved if sea level information and wind stress is incorporated into the state vector. Similarly, Newman et al. (2011) showed that forecast skill of LIM greatly improves if it is extended to include additional terms of thermocline depth and zonal wind stress. By including subsurface information explicitly (or sea level as a proxy), the approaches of Xue et al. (2000) and Newman et al. (2011) provide a much more realistic linear approximation to the true dynamical system.

It has also been demonstrated that the surface–subsurface interaction can be represented without explicitly resolving the subsurface. Rather, much of the subsurface forcing in the tropical Pacific can be modeled using only SSTA as a predictor. As SST is one of the earliest oceanographic variables to be recorded, SST-only models are valuable for long-term and historical studies of ENSO variability. The empirical model reduction (EMR) algorithm of Kondrashov et al. (2005),

---

\* Current affiliation: Oceaneering International Inc., Hanover, Maryland.

---

*Corresponding author address:* David Chapman, Lamont-Doherty Earth Observatory, Columbia University, 61 Route 9W, Palisades, NY 10964.  
E-mail: dchapman@oceaneering.com

further refined by Kravtsov et al. (2010), was capable of extracting the signal of the subsurface ocean implicitly from the residual of LIM forecasts, which is only possible because the residual of LIM, while uncorrelated with SSTA in the same time step, is not entirely white. Rather, the persistence of subsurface forcing leaves a discernable red noise autocorrelation in the LIM residual. Thus, EMR constructs a hierarchy of LIM models to this red-tinted noise signal and extends the state vector of LIM to include these additional linear models. By whitening the residual, EMR extracts meaningful information about the surface–subsurface interaction, which markedly improves the forecast skill.

The vector autoregressive (VAR) model described by this paper is an extension of LIM to include many months of recent history as a predictor. We will show that doing so captures the linear interaction between surface and subsurface using only SSTA. The extension of ENSO forecasting models to include recent historical SSTA in initial conditions is not a new idea. Barnston and Ropelewski (1992) was the first to do so for a canonical correlation analysis (CCA) model, by using four seasons of past and present SSTA in addition to present sea level pressure (SLP) conditions. It was found that the inclusion of one year of recent SSTA greatly improves forecast skill, although the rationale was not fully understood at the time. Van den Dool (2006) in the context of construct analog (CA) models, justifies such an approach as follows: “We are forecasting a single variable from a single variable in a world where many variables are interrelated” (p. 106) and that “In effect we use the initial conditions [to substitute] for the first, second, third derivative at recent times” (p. 106). Although this is true in general, it is of particular utility to the ENSO problem due to the integrated response of western Pacific heat content to tropical SSTA history. We show in the discussion section that (in the univariate case) such a relationship arises naturally from a recharge oscillator framework.

A further justification for the use of a VAR model comes from its equivalence to EMR as proven in the appendix. We show that the VAR and EMR (in its linear version with the quadratic term omitted) are algebraically equivalent, differing only in their algebraic form. We may conclude that like EMR, VAR can be interpreted as extracting subsurface information from the redness (autocorrelation) of the LIM residual. Moreover, the redness of the LIM residual extracted by EMR must be a linear combination of the prior months of SSTA included in VAR. Although the VAR and EMR are theoretically equivalent, there are practical reasons why one might prefer to implement a VAR

model. The VAR is a classical time series analysis technique that carries with it a substantial body of literature, particularly for financial and economic forecasting (Yule 1927; Zellner 1962; Stock and Watson 2001; Box et al. 2008). Furthermore, being less general than EMR, it is a conceptually simpler model with a more compact definition.

## 2. VAR model

The VAR( $L$ ) model is a linear–stochastic time series model for causal stationary processes. The model is “vector” in that it assumes a multivariate state vector  $\mathbf{x}$ , and it is “autoregressive” in that the next state is determined by regression against  $L$  levels of the present and recent past state. The VAR model is denoted as

$$\mathbf{x}_t = \sum_{i=1}^L \mathbf{A}_i \mathbf{x}_{t-i} + \epsilon. \quad (1)$$

As we are working with monthly average time series anomalies, we assume that the next month’s forecast is a linear function of the present anomaly as well as  $L - 1$  months of prior anomaly. In our case, the multivariate state vector  $\mathbf{x}$  represents an empirical orthogonal function (EOF) decomposition of the monthly mean anomaly of SSTA. The EOF decomposition is employed to reduce dimensionality. Each  $\mathbf{A}_i$  is an  $M \times M$  square matrix of autoregression coefficients, and  $\epsilon$  is stochastic random forcing, modeled by spatially correlated but temporally white Gaussian noise.

The VAR model approximates the class of causal linear–stochastic processes that are stationary (i.e., the covariance between any two time steps does not depend on time but is only a function of lead  $l$ ). Thus, the covariance matrix separating states by a lead of  $l$  is equal to the autocovariance matrix as follows:

$$\Gamma_l = \mathbf{X}_t \mathbf{X}_{t-l}^T, \quad (2)$$

where  $\mathbf{X}_t$  denotes the matrix time series of state vector  $\mathbf{x}$  centered around time step  $t$ . Model parameters are estimated using the Yule–Walker equations (Yule 1927; Box et al. 2008) that relate autoregression coefficients  $\mathbf{A}_l$  to the autocovariance matrices  $\Gamma_l$ :

$$\Gamma_l = \sum_{i=1}^L \Gamma_{l-i} \mathbf{A}_i^T \quad \text{for } l = 1, 2, \dots \quad (3)$$

The LIM, also known as the linear Markov model is a special case of the VAR model in which the change in state as a function of time is considered to be linearly proportional to its present state anomaly as follows:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x} + \epsilon. \quad (4)$$

LIM is typically implemented as a discrete time series model in which the temporal derivative is approximated by the change in state from one time step to the next. The LIM is equivalent to the following VAR(1) process:

$$\mathbf{x}_t = (\mathbf{I} + \Delta t \mathbf{A})\mathbf{x}_{t-1} + \epsilon \Delta t. \quad (5)$$

### 3. Data and parameter selection

We use the Kaplan et al. (1998) historical monthly mean SST data gridded at  $5^\circ \times 5^\circ$  resolution. (Kaplan SST data is available for access and download from the IRI data library <http://iridl.ldeo.columbia.edu/>.)

We restricted the data to the tropical latitudes of  $20^\circ\text{S}$ – $20^\circ\text{N}$  and separated it temporally into two pieces: a 120-yr calibration period of 1861–1980 and a 30-yr verification period of 1981–2010. A relatively long 120-yr calibration period was selected to increase the chance that the fitting of the VAR and LIM models are not biased toward the regime behavior of any particular decade. The skill of the VAR model is sensitive to the dimensionality of its state vector  $\mathbf{x}_t$ , which represents the retained part of monthly mean SSTA of present conditions and  $L - 1$  months prior. The total number of coefficients for a VAR model is the product of the square of the number of principal components (PCs) of SSTA  $M^2$  and the number of monthly temporal levels  $L$ . The purpose of this calibration procedure is to determine the most suitable VAR and LIM models for further analysis.

We need to determine values that are in some sense optimal for the number of retained EOF modes  $M$  and the number of time levels  $L$ . We choose as our skill benchmark the correlation coefficient of the predicted Niño-3.4 index with the observed one at a 9-month lead (Fig. 1b). We compute the correlation for the 120 years (1861–1980) with a leave 10-yr-out cross validation. That is, each nonoverlapping 10-yr period (1861–70, ..., 1971–80) is left out of the model calibration step, and the resultant model is verified on the left-out period. The average of these 12 correlation coefficients are displayed in Fig. 1. As there are 12 independent decades, we also refer to this procedure as a “12-fold cross validation.” Cross validation is used here to ensure that artificial skill is not a factor in choosing optimal values for  $M$  and  $L$  since an in-sample measure is expected to show artificially increased skill as the number of predictors is increased.

Figure 1b exhibits two bull’s-eye regions of high forecast skill greater than 0.54 correlation. The first is a region of lower complexity with parameters  $10 \leq M \leq 12$  and

$14 \leq L \leq 17$ . The second is a higher-complexity model, with  $M = 19$  and  $L = 14$ . Interestingly, the lower-complexity bull’s eye overlaps a region of greatest skill at lead 6 months (Fig. 1a) above 0.71 correlation, whereas the higher-complexity bull’s eye overlaps a region of greatest skill at lead 12 months (Fig. 1c) above 0.36 correlation. Although the models in both regions demonstrate comparable forecast skill, the model with  $M = 19$  and  $L = 14$  requires the fitting of 5054 coefficients as opposed to 1400–2448 for the  $10 \leq M \leq 12$  bull’s eye. Thus, we selected for further analysis the model with  $M = 11$  and  $L = 15$  [i.e., VAR(15)–11] at the center of the lower-complexity bull’s eye.

Since the LIM is the special case of VAR for  $L = 1$ , the left edge of Fig. 1b shows the best LIM models according to this cross-validated criteria. Figure 1b shows that the best LIM models achieve Niño-3.4 correlation skill between 0.46 and 0.48 at a 9-month lead in the yellow shaded region of  $19 \leq M \leq 27$ . This is lower than the 0.54–0.56 correlation skill achieved by the best VAR models. While the complexity of the best LIM models (361–729 coefficients) is considerably lower than the complexity of the best VAR models (1400–2448) coefficients, these LIM models require many more EOF spatial patterns. Although LIM does not have the temporal memory to explain western Pacific heat content as a delayed response, the teleconnections of ENSO contain some indirect information of SSTA several months prior.

Irrespective of the number of levels, we see in Figs. 1a–c, skill increases rapidly from 4 to 6 modes whereas skill improvements with greater than 6 modes are more gradual. Recall that the EOFs span the global tropical oceans and that the modes are ordered by variance explained over this global domain. A visual inspection of the EOF structures shows that EOF 4 contains information about tropical North Atlantic (TNA) variability, whereas EOFs 5 and 6 contain information about the northern and southern subtropical Pacific variability.

For most choices of  $L$ , we also see a gradual increase in skill as we approach 11 EOF modes. For the models with less than a year of SSTA history, the best skill is achieved when the number of levels  $L$  is equal to 5 months; adding months up to 12 adds only artificial skill that is removed via cross validation. However, much greater skill is achieved when the number of levels  $L$  is greater than 12 months and even more so between 14 and 17 months. As the recharge (La Niña) phase can last up to or in excess of a year, the VAR model must observe this recharge phase for many months in order to estimate the western Pacific heat content as a linear combination of past and present SSTA.

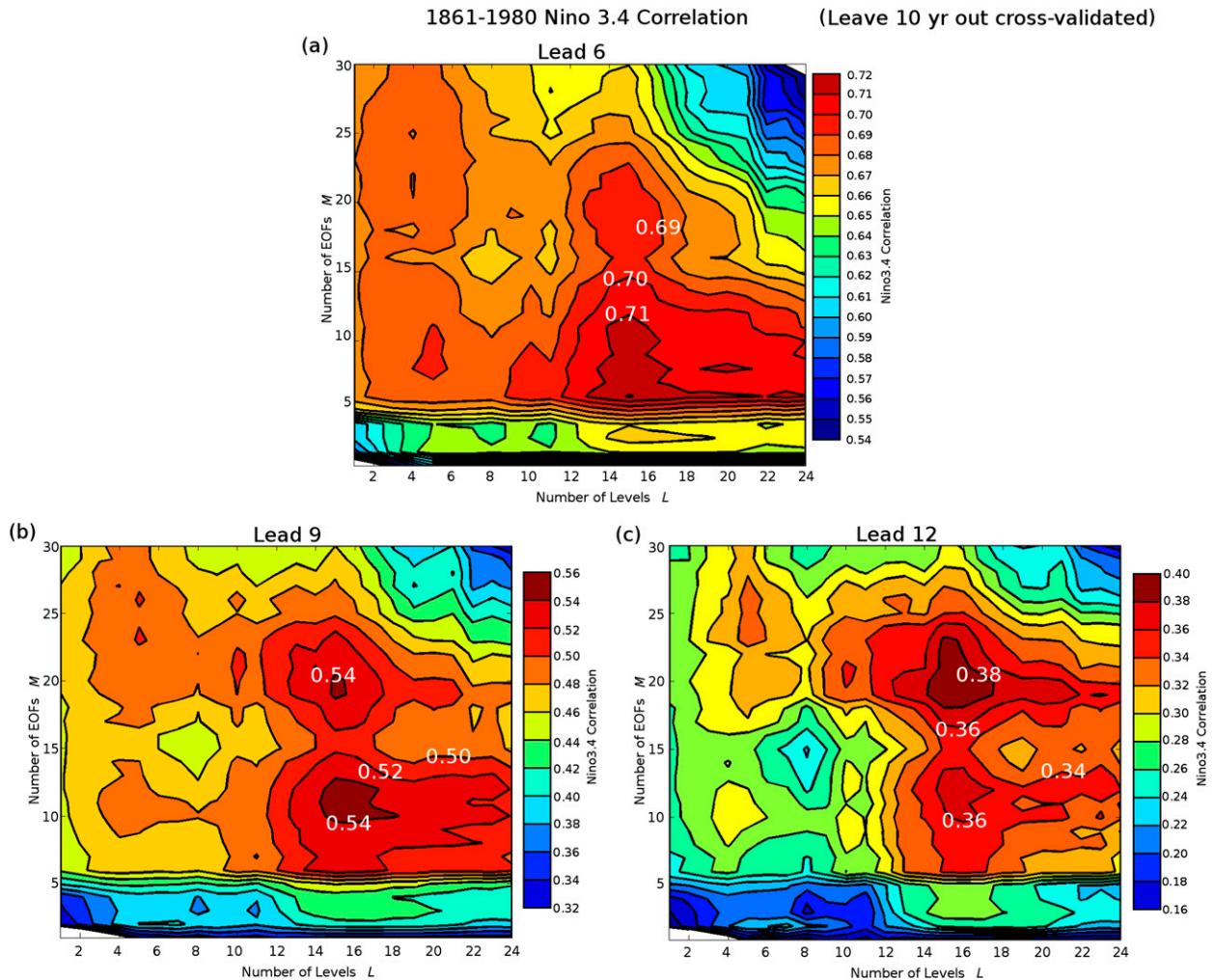


FIG. 1. Model selection by means of cross-validated hindcast procedure of Niño-3.4 correlation skill on the 120-yr calibration interval 1861–1980. Correlation skill is shown for (a) lead 6, (b) lead 9, and (c) lead 12 months.

#### 4. Goodness of fit

We wish to consider and to compare the fitness of the 15-level 11-mode VAR model [VAR(15)–11] against the 23-mode LIM model (LIM–23) because of their strong performance over the 120-yr cross-validated assessment. We begin by regressing both models onto the 120-yr calibration over the 1861–1980 period in order to evaluate their goodness of fit.

A major consideration of the goodness of fit of a linear regression model is the model's ability to capture all causal linear relationships of the observed variables. For a linear stationary process, the noise forcing is assumed to be spatially correlated but temporally white Gaussian noise. We find that the residuals after regression are well approximated by a Gaussian distribution. Histograms of the residual for the first principal

component are shown in Fig. 2 for the VAR(15)–11 model and the LIM–23 model. We visually inspected the histograms of the residuals for all other principal components (not shown) and found that none appeared to deviate substantially from a Gaussian distribution for either the VAR(15)–11 model or the LIM–23 model.

Ideally, the residual between model fit and observation would be temporally uncorrelated. However, the mathematics of linear regression guarantee only that the residual will be uncorrelated with the state vector. It is possible for the residual to contain a small autocorrelation if there exists a physical process not linearly related to the state vector. In this case, the process would not be adequately captured by the model and would persist in the residual as correlated (red) noise.

In Fig. 3 we compare the autocorrelation of the leading principal component of the VAR(15)–11 model

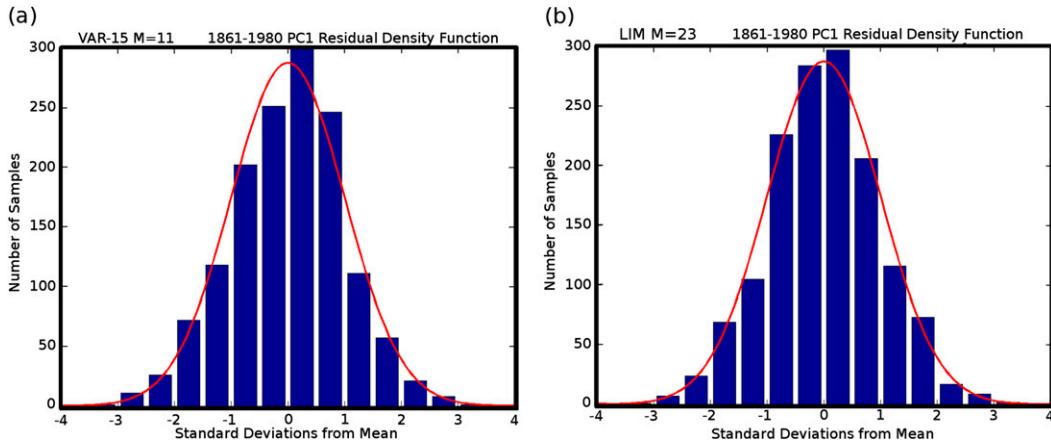


FIG. 2. Probability density function of PC1 residual (bars) in comparison with a Gaussian distribution (bell curve) for (a) the VAR(15)–11 model and the (b) LIM–23 model.

versus the LIM–23 model over the years 1861–1980 with 12-fold cross validation. A true white noise time series would exhibit no significant autocorrelation at any lead (except trivially at lead 0). We find that the PC1 residual autocorrelation for the VAR(15)–11 model is statistically indistinguishable from white noise, but the PC1 residual for the LIM–23 model contains a significant autocorrelation detectable at a variety of leads, most notably lead 1 month. For a true white noise process, the confidence intervals of autocorrelation are proportional to the inverse square of sample size. For a sample size of 1440 months, the size of a  $2\sigma$  interval is a  $\pm 0.05$  correlation for any temporal displacement, whereas the PC1 residual autocorrelation for LIM–23 is 0.10 for a 1-month displacement. We conclude that the VAR(15)–11 model is a better statistical fit to the leading principal component of global SST than the LIM–23 model due to its significantly whiter residual.

### 5. Retrospective forecast (1981–2010)

We evaluate the retrospective forecast skill of VAR (15)–11 and LIM–23 on the 30-yr verification period of 1981–2010. No data from this period was used to construct any of the models. Correlation and root-mean-square error (RMSE) are shown in Figs. 4a,b. VAR (15)–11 shows a marked improvement relative to LIM–23 in correlation skill (Fig. 4a). A horizontal line is drawn at a correlation coefficient of 0.5, which we consider to be a lower bound for useful prediction skill. The yellow dashed line represents the skill of a persistence forecast from initial conditions. VAR(15)–11 achieves skill above 0.5 correlation up to a 9-month lead, as opposed to a 6-month lead for LIM. Beyond 6 months, the forecast skill of LIM–23 continues to decline at a nearly linear rate. This improvement in lead time is also

apparent when the VAR and LIM models are compared in terms of RMSE (Fig. 4b). VAR forecasts up to lead 9 months achieve RMSE less than  $0.8^{\circ}\text{C}$ , whereas all LIM forecasts greater than lead 6 months exhibit RMSE greater than  $0.8^{\circ}\text{C}$ . It is clear that the 15-level VAR delivers improved forecast skill over LIM on the 30-yr retrospective forecast period considered.

Figures 4c,d shows similar correlation and RMSE skill graphs over the cross-validated 120-yr historical period of 1861–1980. We see that VAR(15)–11 achieves a much more modest improvement of a 1-month lead over LIM–23 over this longer period. It is also worth noting that RMSE errors of the historical period (Fig. 4d) are across the board lower than that of 1981–2010 (Fig. 4b), suggesting that the historical period of 1861–1980 was more predictable than the recent period of 1981–2010. However, despite cross validation, it is not clear to what extent this increased predictability might be an artifact of

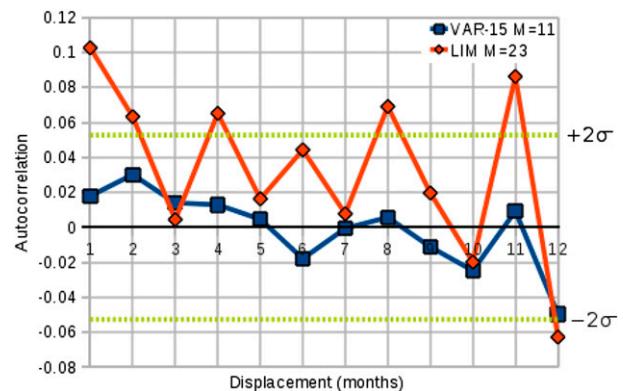


FIG. 3. Autocorrelation of the PC1 residual for VAR(15)–11 model (blue) and LIM–23 model (red) over the cross-validated 1861–1980 calibration period. Green dotted lines show the  $2\sigma$  confidence interval for which autocorrelation is indistinguishable from white noise.

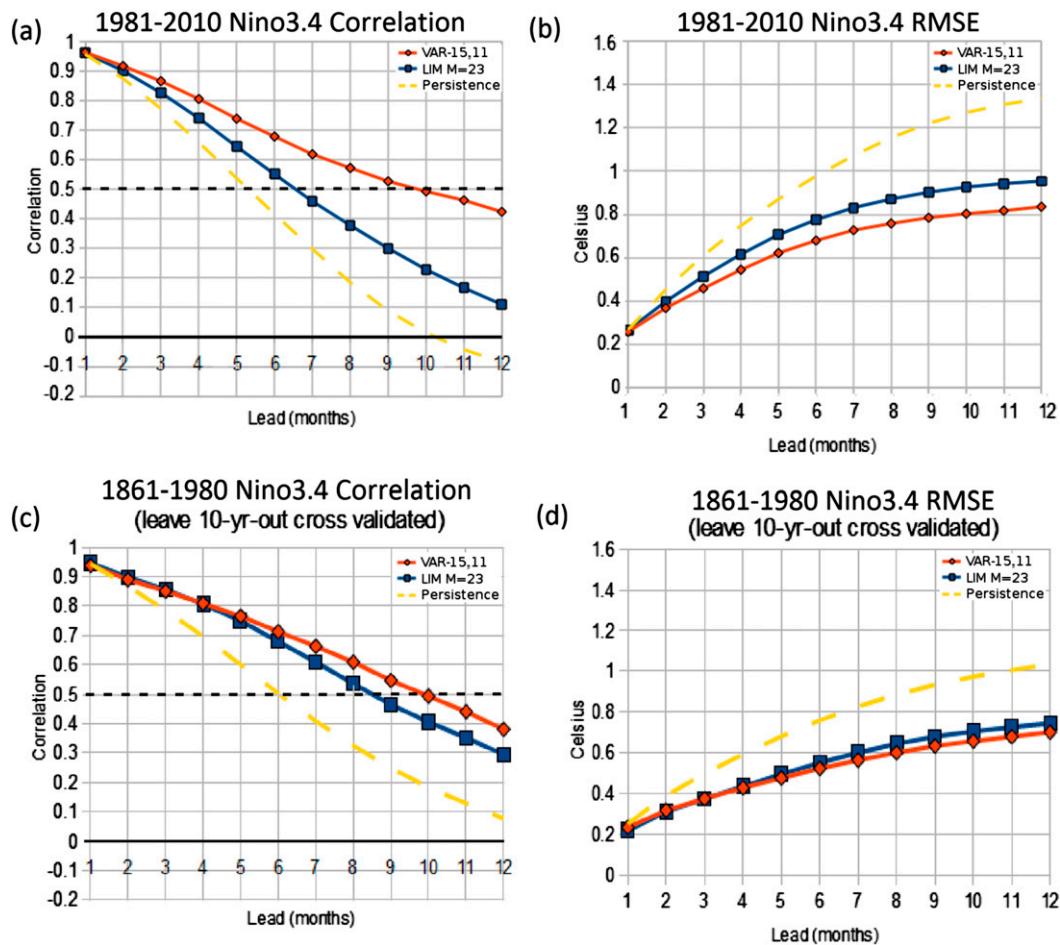


FIG. 4. (a),(c) Correlation and (b),(d) root-mean-square error forecast skill over (a),(b) 1981–2010 and (c),(d) 1861–1980 for the 11-mode VAR(15) model (red diamond) compared against the 23-mode LIM model (blue square) and persistence (yellow dashed).

our testing procedures, as our models were tuned for optimal skill over this historical period.

The seasonal variation in forecast skill as a function of lead time is shown in Fig. 5 for the VAR(15)–11 (right) and LIM–23 (left) over the verification period 1981–2010. We divide the calendar year into 12 overlapping 3-month target periods. We estimate monthly mean values for Niño-3.4 from between 1 and 12 months prior, and plot the correlation skill of all monthly forecasts within each respective target period from a given lead prior.

We see in Fig. 5 that the forecast skill of the VAR(15)–11 and LIM–23 models is highly seasonally dependent, with the strongest predictability occurring in the late boreal winter to early spring months, and the weakest predictability occurring during the summer months. This behavior is a classical spring predictability barrier, and is typical for most dynamical and statistical ENSO forecasting models to demonstrate such a curve (Barnston et al. 2012). The autocorrelation of Niño-3.4

variability is similar (Barnston et al. 2012), indicating that much of the spring predictability barrier may be due to seasonal variation in persistence skill.

The October–December (OND) forecast is of operational importance, because it is the time period in which El Niño events typically begin to reach their peak. We see in Fig. 5 that the VAR(15)–11 model achieves a skill of 0.5 OND correlation up to a 10-month lead, whereas the LIM–23 achieves 0.5 OND correlation between 6 and 7 months.

The greatest skill for both models in Fig. 5 is achieved during the late winter–early spring months. During these months, both El Niño and La Niña events are often en route to decay into more neutral conditions. This season is more predictable because both El Niño and La Niña events often persist, albeit damped, into these months. Thus, the OND forecast is a major clue as to the conditions in late winter–early spring. The February–April (FMA) season achieves 0.5 correlation skill up to the full

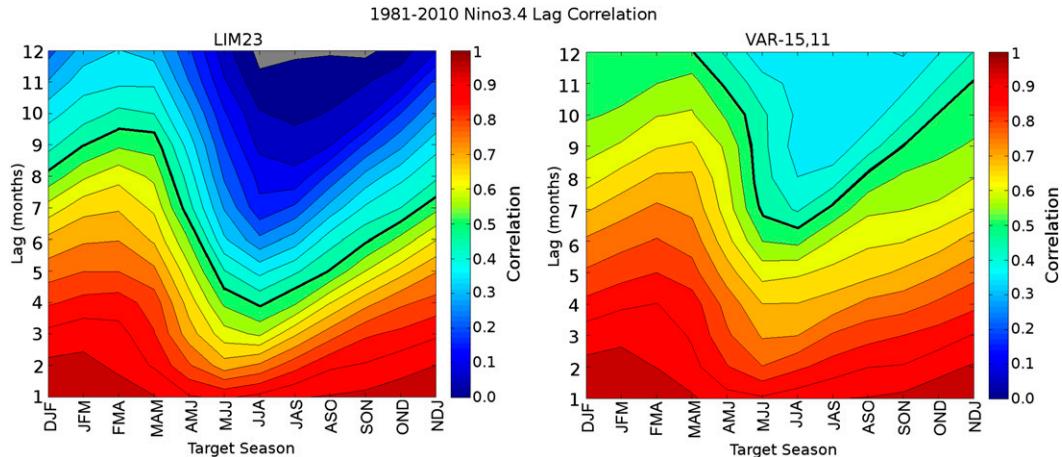


FIG. 5. Temporal correlation skill of (right) VAR(15)–11 and (left) LIM–23 models for running 3-month seasons, as predicted from conditions “lag” months prior. A black line is drawn at 0.5 correlation skill.

12-month lead using the VAR(15)–11 model as opposed to the 9-month lead using the LIM–23 model.

The most difficult period to predict is the summer months of June–August (JJA). This is a period of relatively low amplitude in SSTA, so the correlation is relatively more sensitive to high-frequency atmospheric noise in May. The strongest El Niño events are usually triggered by westerly wind bursts in May instigating Kelvin waves advecting warm surface water to the east. It is likely that this triggering mechanism would be underestimated by models that assume linear–stochastic forcing including VAR (An 2009). Nevertheless, most of the predictability of JJA at longer leads still comes from improving estimates of the precursor ocean conditions. As shown in Fig. 5, the JJA season achieves 0.5 correlation skill between 6 and 7 months with the VAR(15)–11 as compared to 4 months with the LIM.

Figure 6 shows the forecast time series of the LIM–23 and VAR(15)–11 models over the 1981–2010 verification period. For each month over this 30-yr period a red and blue streamline is drawn showing the 9-month forecast trajectory for the VAR(15)–11 and LIM–23 models, respectively. Under most circumstances the VAR(15)–11 model is more skillful than the LIM–23 model, although there is a great deal of variation in prediction accuracy on an event-to-event basis.

We observe a general tendency for the VAR and LIM models to err on the side of caution. The models do not appear prone to overpredicting the magnitude of events. Rather, they are more likely to underpredict growth, and in certain circumstances overpredict persistence. This behavior is not unusual for linear statistical ENSO forecast models, which are known to produce more conservative forecasts than their dynamical counterparts. In a study of the nonlinear processes that influence

ENSO, An (2009) points out that the westerly wind burst signal plays a role in the sudden onset of ENSO events, and that nonlinear convective heating plays a role in their sudden decay. A linear framework such as VAR or LIM cannot capture these nonlinearities, and can at best approximate them by the first-order linear term. This may lead to the linear model predicting slower growth and decay than a dynamical model or a nonlinear statistical model.

## 6. Discussion

Use of SSTA history as a linear proxy for western Pacific thermocline depth can be derived from the recharge oscillator framework of Jin (1997a,b). In this section we discuss the theoretical relation between western Pacific thermocline depth and the history of SSTA PC1. We include a regression experiment between thermocline depth and SSTA history to help support our argument.

The western Pacific thermocline depth is forced by zonal wind stress. The recharge oscillator model also includes an ocean-adjustment term to slowly draw western Pacific thermocline depth back toward its climatological mean:

$$\Delta h_w \approx -\alpha\tau - rh_w, \quad (6)$$

where  $h_w$  is the thermocline depth at the west,  $\tau$  is wind stress,  $\Delta h_w$  is the change of  $h_w$  in one month, and  $\alpha$  and  $r$  are constant coefficients.

Zonal wind stress includes a stochastic high-frequency term that we assume to be white Gaussian noise, as well as a predictable low-frequency term that is part of the Walker circulation. The Walker circulation is driven by

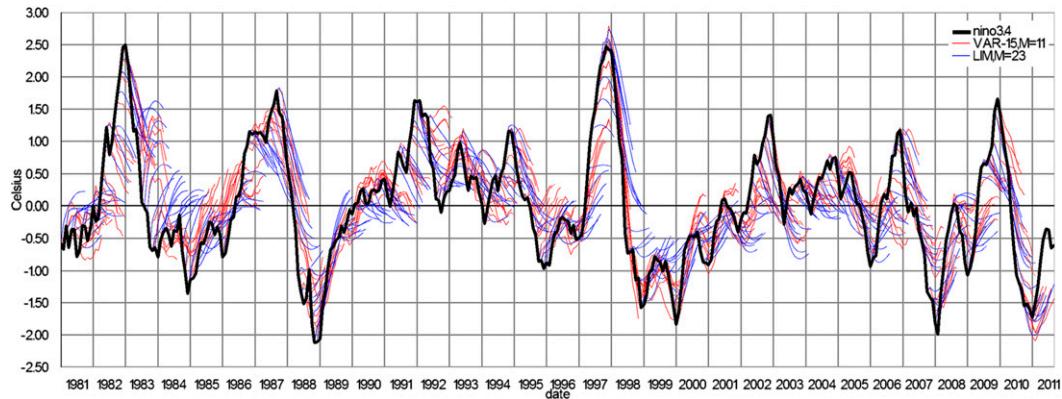


FIG. 6. Time series of ENSO forecast 1981–2010 (black) with streamlines (red, blue) showing the 9-month projected forecast time series initialized each month over this 30-yr period. The VAR(15)–11 (red) forecasts are plotted against the LIM–23 (blue).

pressure gradients in the western and eastern Pacific and proportional to the east minus west temperature gradient. This temperature gradient resembles the shape pattern of SSTA EOF1. Zonal wind stress thus becomes

$$\tau \approx b(T_e - T_w) + \epsilon = bPC_1 + \epsilon. \quad (7)$$

Substituting Eq. (7) for  $\tau$  into Eq. (6) we arrive at western Pacific thermocline depth written as an infinite sum of prior time steps of the dominant mode of monthly SSTA:

$$h_{w,t+1} \approx -\alpha b \lim_{L \rightarrow \infty} \sum_{i=1}^L (1-r)^{i-1} PC_{1,t-i} + \epsilon. \quad (8)$$

As this infinite sum decays exponentially as a function time, it is sufficient to use a finite number of monthly time steps of SSTA history  $L$  as employed by VAR. Thermocline depth is thus approximated as a function of the recent history of SSTA PC1.

The derivation of Eq. (8) suggests that it is reasonable to employ a linear combination of the history of SSTA PC1 as a proxy for  $h_w$  in the event that subsurface information is not available. On the other hand, it is well known that the shallower eastern Pacific thermocline depth  $h_e$  exhibits a much larger signature in the present SSTA patterns. As such, one might expect a linear model unaware of SST history to have a much greater ability to estimate  $h_e$  than  $h_w$ .

In support of this reasoning we have included an experiment in which we attempt to explain the present thermocline depth anomaly using linear regression from SSTA principal components. In this experiment we approximate thermocline depth data using the 20°C isotherm of the Simple Ocean Data Assimilation (SODA), version 2.4.2, from 1951 to 2010. In Fig. 7, we

calculate grid cell by grid cell the percent of variance of thermocline depth explained by using  $M$  principal components and  $L$  time steps of SSTA history for selected values of  $M$  and  $L$ . The 10-fold cross validation is employed, and the percent variance explained is defined to be 100 times the coefficient of determination  $r^2$ .

It can be seen in Fig. 7a that both the eastern and western thermocline depths share a trivial zeroth-order phase relationship to the SSTA PC1. Linear regression from 23 PCs of SSTA, as in Fig. 7b, greatly improves our ability to explain the eastern Pacific thermocline depth, but does relatively little to improve the regression of western Pacific thermocline depth. However, using

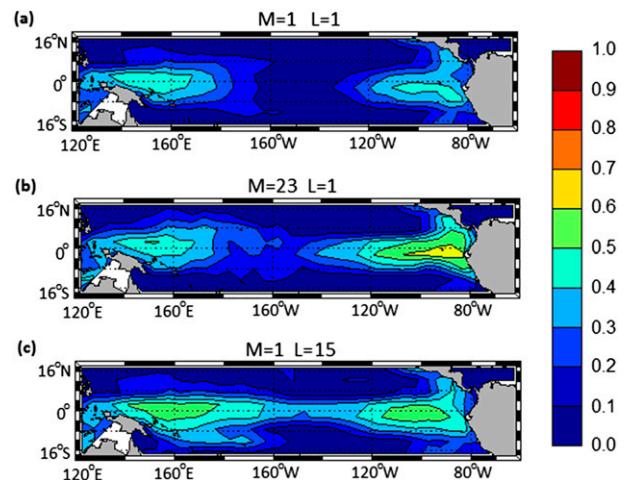


FIG. 7. Percent of explained variance of 20°C isotherm explained by linear regression from SSTA principal components over the 1951–2010 period with 10-fold cross validation. (a) The 20°C isotherm explained by a single PC of SSTA with no SST history. (b) The 20°C isotherm explained using 23 PCs of SSTA with no history. (c) The 20°C isotherm explained using a single PC of SSTA with 15 months of SST history.

only a single principal component of SSTA but including 15 months of SSTA history as in Fig. 7c, improves the ability to reconstruct both the eastern and western thermocline depths relative to Fig. 7a, but the improvement in the east is less dramatic than in Fig. 7b. We also see that the linear SST models have relatively little ability to reconstruct thermocline depth in the central Pacific where the depth changes are smaller.

## 7. Conclusions

Both LIM and VAR use only tropical SSTA to predict future SSTA. The LIM uses only the current (or most recent) SSTA, but the VAR includes recent historical SSTA as well. The inclusion of recent historical SSTA is found to extract meaningful information of subsurface observation and greatly improve forecasting skill. This is an alternative approach to extending the LIM to include subsurface observation directly as investigated by Xue et al. (2000) and Newman et al. (2011), and allows the VAR to train using very long observational records prior to the completion of the TOGA–TAO array.

We find that an SSTA VAR model improves retrospective forecast skill by a 3-month lead relative to SSTA LIM over the 1981–2010 period. VAR also improves cross-validated hindcast skill of the 1861–1980 period by a 1-month lead. Furthermore, we find that the hindcast residuals of the VAR model are statistically white containing no discernable autocorrelation. The use of many months of SSTA history as a proxy for western Pacific thermocline depth can be justified from a recharge oscillator point of view, and we find experimentally that the history of SSTA PC1 contains proxy information about the western Pacific thermocline depth.

*Acknowledgments.* This research was supported by the Office of Naval Research under Grant N00014-12-1-0911, as well as the Provost's Postdoctoral Research Scientist and Scholar Program of Columbia University. We would also like to thank two anonymous reviewers for their comments.

## APPENDIX

### Proof of Equivalence of VAR and Linear EMR

The empirical model reduction (EMR) algorithm of Kondrashov et al. (2005) is a linear and/or quadratic ENSO prediction scheme that, using only SSTA, is capable of capturing the subsurface forcing that remains in the forecast residual of LIM. EMR is a hierarchical method designed to reduce this residual autocorrelation by fitting additional LIM models to an extended state

vector that includes the autocorrelated residual as a predictor. Although the first term in EMR is quadratic, a linear version was also analyzed by Kondrashov et al. (2005) and shown to exhibit comparable forecast skill to the quadratic version.

Here we present an inductive proof that the linear version of EMR is equivalent to the VAR method, in the sense that any linear EMR model can be algebraically restructured to a VAR model of equal order. The linear EMR is an extension of LIM in which the residual term  $\epsilon$  is replaced by additional LIM models using an extended state vector that includes the autocorrelated residual terms:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}_1 \mathbf{x}_t + \mathbf{r}_{t+1}^{(2)} \\ \mathbf{r}_{t+1}^{(2)} &= \mathbf{A}_2 [\mathbf{x}_t \mathbf{r}_t^{(2)}] + \mathbf{r}_{t+1}^{(3)} \\ \mathbf{r}_{t+1}^{(3)} &= \mathbf{A}_3 [\mathbf{x}_t \mathbf{r}_t^{(2)} \mathbf{r}_t^{(3)}] + \mathbf{r}_{t+1}^{(4)} \\ &\vdots \\ \mathbf{r}_{t+1}^{(L)} &= \mathbf{A}_L [\mathbf{x}_t \mathbf{r}_t^{(2)} \mathbf{r}_t^{(3)} \dots \mathbf{r}_t^{(L)}] + \epsilon \end{aligned} \quad (\text{A1})$$

Henceforth, we refer to linear EMR simply as EMR. The square bracket notation refers to the vector concatenation of terms and is used to construct an extended state vector. As the additive white noise term  $\epsilon$  is identical in either the EMR or VAR models, for simplicity we omit this term. It is sufficient to prove equivalence of the expected value of the forecasts, which for a linear model is equivalent to the absence of zero-mean noise. For conciseness, we define an additional term  $\mathbf{r}^{(l)}$  as an alias of  $\mathbf{x}$  [such that  $\mathbf{r}^{(l)} = \mathbf{x}$ ]. EMR can be written as follows for all  $1 \leq l \leq L$ :

$$\mathbf{r}_{t+1}^{(l)} = \sum_{i=1}^l \mathbf{A}_{l,i} \mathbf{r}_t^{(i)} + \mathbf{r}_t^{(l+1)}. \quad (\text{A2})$$

It is clear that EMR and VAR are equivalent in the case  $L = 1$  [i.e.,  $\text{EMR}(1) \equiv \text{VAR}(1)$ ]. This is the special case for which both methods are identical to LIM. It is less obvious that EMR and VAR are equivalent in the nontrivial case of  $L > 1$ . We will prove this by induction with the hypothesis: If  $\text{EMR}(L - 1) \equiv \text{VAR}(L - 1)$ , then  $\text{EMR}(L) \equiv \text{VAR}(L)$ .

An  $\text{EMR}(L)$  model is an  $\text{EMR}(L - 1)$  model extended to include one extra level. This extra level is implemented by adding the term  $\mathbf{r}_{t+1}^{(L)}$  to the  $\text{EMR}(L - 1)$  model. For induction we have assumed that  $\text{EMR}(L - 1)$  is equivalent to  $\text{VAR}(L - 1)$ , yielding the following:

$$\text{EMR}(L) \equiv \text{VAR}(L - 1) + \mathbf{r}_{t+1}^{(L)}. \quad (\text{A3})$$

As any linear combination of VAR models of orders  $p$  and  $q$  yields a VAR model of order  $\max(p, q)$ , it is sufficient to show that  $\mathbf{r}_{t+1}^{(L)}$  is a  $\text{VAR}(L)$  model. This can be

proven by strong induction assuming  $\mathbf{r}_{t+1}^{(l)}$  is a VAR( $l$ ) model for any  $l \in [1, L - 1]$ . We work with Eq. (A2) for the term of order  $L - 1$  as it draws a relationship to the term of order  $L$ :

$$\mathbf{r}_{t+1}^{(L-1)} = \sum_{i=1}^{L-1} \mathbf{A}_{t,i} \mathbf{r}_t^{(i)} + \mathbf{r}_{t+1}^{(L)}. \quad (\text{A4})$$

According to our inductive hypothesis, the first term  $\mathbf{r}_{t+1}^{(L-1)}$  is a VAR( $L - 1$ ) model. The term  $\mathbf{r}_t^{(i)}$  is in fact a VAR( $i + 1$ ). This is because  $\mathbf{r}_{t+1}^{(i)}$  is a VAR( $i$ ) model, thus by backtracking one step in time  $\mathbf{r}_t^{(i)}$  is a linear combination of the  $i + 1$  terms prior to time step  $t + 1$ . Thus, Eq. (A4) is rearranged as follows:

$$\text{VAR}(L - 1) - \sum_{i=1}^{L-1} \mathbf{A}_{t,i} \text{VAR}(i + 1) = \mathbf{r}_{t+1}^{(L)}. \quad (\text{A5})$$

Thus,  $\mathbf{r}_{t+1}^{(L)}$  is a VAR( $L$ ) model. Therefore, by substitution back into Eq. (A3), we get

$$\text{EMR}(L) \equiv \text{VAR}(L). \quad (\text{A6})$$

#### REFERENCES

- An, S.-I., 2009: A review of interdecadal changes in the non-linearity of the El Niño–Southern Oscillation. *Theor. Appl. Climatol.*, **97**, 29–40, doi:10.1007/s00704-008-0071-z.
- Barnston, A. G., and C. F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, **5**, 1316–1345, doi:10.1175/1520-0442(1992)005<1316:POEEUC>2.0.CO;2.
- , and Coauthors, 1994: Long-lead seasonal forecasts—Where do we stand? *Bull. Amer. Meteor. Soc.*, **75**, 2097–2114, doi:10.1175/1520-0477(1994)075<2097:LLSFDW>2.0.CO;2.
- , M. K. Tippett, M. L. L’Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, doi:10.1175/BAMS-D-11-00111.1.
- Blumenthal, M. B., 1991: Predictability of a coupled ocean–atmosphere model. *J. Climate*, **4**, 766–784, doi:10.1175/1520-0442(1991)004<0766:POACOM>2.0.CO;2.
- Box, G. E., G. M. Jenkins, and G. C. Reinsel, 2008: *Time Series Analysis: Forecasting and Control*. Wiley Series in Probability and Statistics, Vol. 734, John Wiley & Sons, 784 pp.
- Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature*, **321**, 827–832, doi:10.1038/321827a0.
- Jin, F.-F., 1997a: An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *J. Atmos. Sci.*, **54**, 811–829, doi:10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2.
- , 1997b: An equatorial ocean recharge paradigm for ENSO. Part II: A stripped-down coupled model. *J. Atmos. Sci.*, **54**, 830–847, doi:10.1175/1520-0469(1997)054<0830:AEORPF>2.0.CO;2.
- Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.*, **103**, 18 567–18 589, doi:10.1029/97JC01736.
- Kondrashov, D., S. Kravtsov, A. W. Robertson, and M. Ghil, 2005: A hierarchy of data-based ENSO models. *J. Climate*, **18**, 4425–4444, doi:10.1175/JCLI3567.1.
- Kravtsov, S., D. Kondrashov, and M. Ghil, 2010: Empirical model reduction and the modelling hierarchy in climate dynamics and the geosciences. *Stochastic Physics and Climate Modelling*, T. Palmer and P. Williams, Eds., Cambridge University Press, 35–72.
- Newman, M., M. A. Alexander, and J. D. Scott, 2011: An empirical model of tropical ocean dynamics. *Climate Dyn.*, **37**, 1823–1841, doi:10.1007/s00382-011-1034-0.
- Penland, C., and T. Magorian, 1993: Prediction of Niño-3 sea surface temperatures using linear inverse modeling. *J. Climate*, **6**, 1067–1076, doi:10.1175/1520-0442(1993)006<1067:PONSST>2.0.CO;2.
- , and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999–2024, doi:10.1175/1520-0442(1995)008<1999:TOGOTS>2.0.CO;2.
- Stock, J. H., and M. W. Watson, 2001: Vector autoregressions. *J. Econ. Perspect.*, **15**, 101–115, doi:10.1257/jep.15.4.101.
- Van den Dool, H., 2006: *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, 240 pp.
- Wyrtki, K., 1975: El Niño—The dynamic response of the equatorial Pacific Ocean to atmospheric forcing. *J. Phys. Oceanogr.*, **5**, 572–584, doi:10.1175/1520-0485(1975)005<0572:ENTDRO>2.0.CO;2.
- Xue, Y., A. Leetmaa, and M. Ji, 2000: ENSO prediction with Markov models: The impact of sea level. *J. Climate*, **13**, 849–871, doi:10.1175/1520-0442(2000)013<0849:EPWMMT>2.0.CO;2.
- Yule, G. U., 1927: On a method of investigating periodicities in disturbed series, with special reference to Wolfer’s sunspot numbers. *Philos. Trans. Roy. Soc. London*, **A226**, 267–298.
- Zellner, A., 1962: An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Amer. Stat. Assoc.*, **57**, 348–368, doi:10.1080/01621459.1962.10480664.