Predicting the Real-time Multivariate Madden-Julian Oscillation Index through a Low-Order Nonlinear Stochastic Model

Nan Chen* and Andrew J. Majda

Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY.

*Corresponding author address: Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA.
E-mail: chennan@cims.nyu.edu
ABSTRACT

We develop a new low-order nonlinear stochastic model to improve the predictability of the real-time multivariate Madden-Julian Oscillation (MJO) index (RMM index), which is a combined measure of convection and circulation. A recent data driven physics-constrained low-order stochastic modeling procedure is applied to the RMM index. The result is a four dimensional nonlinear stochastic model for the two observed RMM variables and two hidden variables involving correlated multiplicative noise defined through energy conserving nonlinear interaction. The special structure of the low-order model allows efficient data assimilation for the initialization of the hidden variables that facilitates the ensemble prediction algorithm. An information-theoretic framework is applied to the calibration of model parameters in a short training phase of three years. This framework involves generalizations of the anomaly pattern correlation, the RMS error and the information deficiency in the model forecast. The nonlinear stochastic models show skillful prediction for 30 days on average in these metrics. More importantly, the predictions succeed in capturing the amplitudes of the RMM index and the useful skill of forecasting strong MJO events is around 40 days. Furthermore, information barriers in prediction for linear models imply the necessity of the nonlinear interactions between the observed and hidden variables as well as the multiplicative noise in these low-order stochastic models.
1. Introduction

The Madden-Julian Oscillation (MJO) is the dominant component of tropical intraseasonal variability. It is a slow moving planetary scale envelope of convection propagating eastward across the equatorial Indian and western/central Pacific oceans. As a naturally occurring component of the coupled ocean-atmosphere system, the MJO effects tropical precipitation, the frequency of tropical cyclones, and extratropical weather patterns (Lau and Waliser 2012). A central problem in contemporary meteorology with large societal impacts is to understand and predict the MJO (Zhang et al. 2013). Predicting the MJO is a major enterprise through either low-order statistical models (Jiang et al. 2008; Seo et al. 2009; Kang and Kim 2010; Oliver and Thompson 2011; Kondrashov et al. 2013; Cavanaugh et al. 2014) or operational dynamical models (Gottschalck et al. 2010; Vitart and Molteni 2010; Zhang et al. 2013; Kim et al. 2014; Mani et al. 2014).

The real-time multivariate MJO (RMM) index (Wheeler and Hendon 2004) is one of the most popular metrics for assessing the large-scale skill in MJO prediction. The RMM index involves both the winds at the top and bottom of the troposphere and the outgoing longwave radiation (OLR) which is a surrogate for convective activity. The statistical low-order models utilized for forecasting the RMM indices are mainly based on multivariate regression (Maharaj and Wheeler 2005; Jiang et al. 2008; Seo et al. 2009; Kang and Kim 2010), time series analysis (Seo et al. 2009; Love and Matthews 2009; Kang and Kim 2010) and analogs (Seo et al. 2009) with the model uncertainty typically represented by additive stochastic noise. The useful prediction skill for the MJO of these models is about 15 – 20 days, which is similar to that of the operational dynamical models. Although incorporating the past-noise forecasting (PNF) method into the empirical statistical models (Kondrashov et al. 2013) extends the empirical MJO prediction to 25 days regarding
the anomaly pattern correlation, the severe underestimation of the amplitudes in prediction by this model, especially for the strong MJO events, hinders potential useful forecasting in practice.

Here we improve the predictability of the RMM index in two aspects. First, a recent systematic strategy for data-driven physics-constrained low-order stochastic modeling procedure (Majda and Harlim 2013; Harlim et al. 2014) is applied to the RMM index, which results in a four-dimensional nonlinear stochastic model for the two variables representing the two RMM components and two hidden variables. This low-order model involves correlated multiplicative noise defined through energy conserving nonlinear interactions between the observed and hidden variables as well as additive stochastic noise. The special structure of the low-order model allows efficient data assimilation for the initialization of the hidden variables. This together with the initialization of the observed variables provided by the singular spectrum analysis (SSA) reconstruction (Vautard and Ghil 1989) of the RMM index facilitates the ensemble prediction algorithm. Second, due to the failure of measuring the disparity in the peaks between the observed and forecast RMM indices by the path-wise approaches utilizing anomaly pattern correlation, an information-theoretic framework (Roulston and Smith 2002; Weisheimer et al. 2014; Branicki and Majda 2014b) is applied to the model calibration in a short training period. This framework involves generalizations of the anomaly pattern correlation, the RMS error and the information deficiency in the model forecast which is an indicator for assessing the amplitudes in the forecast RMM index.

The remainder of the paper is organized as follows. Section 2 includes the preliminaries of the RMM index and SSA-based initialization. The nonlinear physics-constrained low-order stochastic models as well as the prediction algorithm and data assimilation algorithm for the hidden variables are presented in Section 3. The information-theoretic framework is introduced in Section 4, followed by the calibration of model parameters through information theory in a short training period. Section 5 illustrates the prediction skill of the nonlinear physics-constrained stochastic models as
well as that of the linear stochastic models. The relationship between the MJO prediction skill and El Niño Southern Oscillation (ENSO) is investigated in this section as well. The paper is concluded in Section 6.

2. Preliminaries

a. The real-time multivariate MJO (RMM) index

The real-time multivariate MJO (RMM) index is a combined measure of convection and circulation of tropical intraseasonal variability. It is based on the first two Empirical Orthogonal Functions (EOFs) of the combined fields of near-equatorially-averaged 850 hPa zonal wind, 200 hPa zonal wind, and satellite-observed outgoing longwave radiation (OLR) data, where the OLR data is that measured by the NOAA polar-orbiting satellites and the winds data are from the NCEP/NCAR Reanalysis and the NCEP Operational analysis. Projecting the daily observed data onto these multiple-variable EOFs and removing the annual cycle and components of interannual variability yield principal component (PC) time series that vary mostly on the intraseasonal time scale of the MJO. These two PC time series are defined as the RMM index (Wheeler and Hendon 2004). Since the publication of the work (Wheeler and Hendon 2004), the RMM index has become the leading method for identifying the state of the MJO in observations (Jiang et al. 2011; Riley et al. 2011; Wang et al. 2012; Straub 2013) and model analysis (Kim et al. 2009; Gottschalck et al. 2010).

As shown in Figure 1, the RMM indices (black lines) are dominated by intraseasonal oscillations around a 40 – 50 day band. Yet, both the phase and the amplitude of the RMM indices are stochastic and quite a few small-scale random fluctuations are observed in these indices. Despite some intermittently large bursts, the time-averaged equilibrium probability density functions (PDFs) for
both the components are nearly Gaussian with unit standard deviation. As widely accepted in literature (Lin et al. 2010; Rashid et al. 2011), the strong and weak MJO events are defined as

\[
\text{Strong MJO events : } RMM_1^2 + RMM_2^2 \geq 1, \\
\text{Weak MJO events : } RMM_1^2 + RMM_2^2 < 1,
\]

and the skillful prediction of the RMM index is measured by bivariate correlation exceeding 0.5 and RMS error below \(\sqrt{2}\) (which is the climatological forecast) (Lin et al. 2008; Gottschalck et al. 2010; Vitart et al. 2010).

**b. Initialization reconstruction through singular spectrum analysis (SSA)**

Initialization plays a significant role in the effective short and medium range forecasting. Yet, employing the raw noisy RMM indices for the initialization impedes the skillful prediction. Various statistical procedures are often utilized to improve the initialization.

We rely on the singular spectrum analysis (SSA) (Vautard and Ghil 1989) to reconstruct the RMM indices, serving as the initialization for prediction. SSA is a data adaptive, nonparametric method for spectral estimation that extends classic principal component analysis into the time lagged domain. In the following, we apply SSA for the entire dataset (years 1980 – 2013) containing both the training and prediction phases with a lagged embedding window of 50 days, which is consistent with the intraseasonal time scale. The SSA reconstruction of the RMM indices can be understood as extracting the dominant part of the signal from the noisy time-series. The SSA reconstruction for the entire dataset in predicting the RMM indices was also utilized in (Kang and Kim 2010) to reconstruct the predicted values from the autoregression models and in (Kondrashov et al. 2013) to identify the low-frequency mode. The leading two and four SSA reconstructed components (hereafter SSA(1-2) and SSA(1-4)), accounting for 56% and 84% of the total energy
of the RMM indices respectively, are shown in Figure 1. They are utilized as the reconstructed initialization for prediction.

Since SSA(1-4) reconstruction removes only the small-scale random fluctuations in the RMM indices, the prediction with SSA(1-4) initialization is expected to be skillful in short range but share the same problem in medium range as that with initialization based on the raw RMM indices. On the other hand, SSA(1-2) reconstruction extracts the large-scale principal components and brings about an obvious discrepancy compared with the RMM indices. Therefore, although the skill of forecasting at a very short lead time cannot be improved due to this intrinsic barrier, both short- and medium-range initialized predictions with the SSA reconstructions can be more skillful.

We point out that the direct application of SSA reconstruction for initialization is not practical in real-time prediction since it utilizes the “future” information of the indices. Yet, for the predictability study in this work, we aim at presenting the optimal prediction skill of the nonlinear physics-constrained low-order stochastic model with model calibration given by the information-theoretic framework as proof of concept and therefore SSA reconstruction based on the entire dataset is adopted here as has been done in previous work (Kang and Kim 2010; Kondrashov et al. 2013).

3. The nonlinear physics-constrained low-order stochastic model

Denote by \( u_1 \) and \( u_2 \) the two components, RMM1 and RMM2, depicted in Figure 1. We propose the following family of low-order stochastic models to describe the variability of the time series
Besides the two observed RMM variables $u_1$ and $u_2$, the other two variables $v$ and $\omega_u$ are hidden and unobserved, representing the stochastic damping and stochastic phase, respectively. In (2), $\dot{W}_{u_1}, \dot{W}_{u_2}, \dot{W}_v$ and $\dot{W}_\omega$ are independent white noise. The constant coefficients $d_u$, $d_v$, and $d_\omega$ represent damping for each stochastic process and have physical dimension $t^{-1}$; $a$ (also of dimension $t^{-1}$) is the background state of the phases of $u_1$ and $u_2$; $\sigma_u$, $\sigma_v$, and $\sigma_\omega$ are noise amplitudes with dimension $t^{-1/2}$; the non-dimensional constant $\gamma$ is the coefficient of the nonlinear interaction. All the model variables are real.

The hidden variables $v$, $\omega_u$ interact with the observed RMM variables $u_1, u_2$ through energy conserving nonlinear interactions following the systematic physics-constrained nonlinear regression strategies for time series developed recently (Majda and Harlim 2013; Harlim et al. 2014). The energy conserving nonlinear interactions are seen by first writing down the nonlinear parts of (2),

\[
\frac{du_1}{dt} = \gamma v u_1 - \omega_u u_2, \quad \frac{du_2}{dt} = \gamma v u_2 + \omega_u u_1, \quad \frac{dv}{dt} = -\gamma (u_1^2 + u_2^2), \quad \frac{d\omega_u}{dt} = 0. \tag{3}
\]

Then multiplying the four equations in (3) by $u_1, u_2, v$ and $\omega_u$ respectively, and summing them up yields

\[
\frac{d\tilde{E}}{dt} = 0, \tag{4}
\]

where $\tilde{E} = (u_1^2 + u_2^2 + v^2 + \omega_u^2)/2$ is the energy from nonlinear interactions. The vanishing of the right hand side in (4) is due to the opposite signs of the nonlinear terms involving $v$ multiplying $u_1$
and \( u_2 \) in the first two equations in (3) and those in the third equation multiplying by \( v \) as well as the trivial cancelation of skew-symmetric terms involving \( \omega_\nu \) in the first two equations in (3).

The low-order stochastic nonlinear models in (2) are fundamentally different from those utilized earlier (Kravtsov et al. 2005; Kondrashov et al. 2013) which allow for nonlinear interactions only between the observed variables \( u_1, u_2 \) and only special linear interactions with layers of hidden variables. The stochastic damping \( \nu \) and stochastic phase \( \omega_\nu \) as well as their energy conserving nonlinear interaction with \( u_1 \) and \( u_2 \) also distinguish the models in (2) from the classic damped harmonic oscillator with only constant damping \( d_u \) and phase \( a \). It is evident that a negative value of the stochastic damping \( \gamma \nu \) serves to strengthen the damping of the oscillator. On the other hand, when \( \gamma \nu \) becomes positive and overwhelms \( d_u \), an exponential growth of \( u_1 \) and \( u_2 \) will occur, which corresponds to the intermittent instability. Further motivation for the models in (2) is provided by the stochastic skeleton model which predicts key features of the MJO (Majda and Stechmann 2009, 2011; Thual et al. 2013; Stechmann and Majda 2014); these are coupled nonlinear oscillator models of the MJO where if we identify the observed variables with the envelope of synoptic scale convective activity and the zonal wind velocity, then the hidden variables \( \nu, \omega_\nu \) and their dynamics become phenomenological surrogates for the energy conserving interactions in the skeleton model involving the synoptic scale convective activity, the equatorial dynamic equations for the zonal wind velocity and those for the meridional wind velocity, temperature and moisture. Thus, the model in (2) can be regarded as a low-order nonlinear stochastic oscillator model for the MJO.

A more sophisticated version of (2) with additional time-periodic damping has been shown to have significant skill for determining the predictability limits of the large-scale cloud patterns of boreal winter MJO (Chen et al. 2014b). Note that these models are a special case of the models described in (Majda and Harlim 2013; Harlim et al. 2014).
a. Prediction algorithm and data assimilation for the hidden variables

The full time-series are divided into the training and prediction periods. The training period is utilized to calibrate the model parameters and the prediction is studied over a later phase.

The ensemble prediction algorithm is adopted by running the forecasting model (2) forward given the initial values. The initial data of the two state variables $\mathbf{u} = (u_1, u_2)$ are obtained directly from the observations. The more important and challenging issue is to determine the initial ensemble, i.e., initialization, of the two hidden variables $\Gamma = (v, \omega_u)$. To this end, data assimilation is incorporated into the prediction algorithm.

The estimates of the hidden parameters $\Gamma = (v, \omega_u)$ during the training period and initialization of these parameters during the prediction phase exploit the special structure of the low-order nonlinear stochastic model; the equations in (2) are a conditional Gaussian system with respect to the observation of $\mathbf{u} = (u_1, u_2)$, meaning that once $u_1$ and $u_2$ are given, there are closed analytic equations for the conditional Gaussian distributions of the hidden parameters $\Gamma = (v, \omega_u)$ (Liptser and Shiryaev 2001). These conditional Gaussian distributions are obtained from the posterior estimations in the Bayesian framework. Appendix B contains the details and explicit equations. We utilize this fact to construct an initial ensemble for forecasting at each time in the training and prediction phases for $t \in [t_0, t_1, \ldots, t_s]$. Starting from a “burn in” time $t_-$ earlier than $t_0$ with arbitrary initial conditions, solve the associated analytic formula (B2) until time $t_0$ to obtain the conditional Gaussian distribution $p_0(\Gamma|\mathbf{u}(t_0))$. The initial ensemble of the hidden variables $\Gamma = (v, \omega_u)$ for prediction starting from $t_0$ is drawn from this distribution. The initial ensemble for prediction starting from the next day $t_1$ is drawn from $p_1(\Gamma|\mathbf{u}(t_1))$, where $p_1(\Gamma|\mathbf{u}(t_1))$ is solved by running the analytic formula (B2) forward for one more step. Following the same procedure, the initial distributions of the hidden variables $v, \omega_u$ for prediction starting from each time $t_i$ are obtained “on
the fly” when the new observations \( \mathbf{u}(t_i) \) are available. In the prediction below with (2), we use \( N \) ensemble members with \( N = 50 \). This is a practical on-line data assimilation algorithm for the stochastic models in (2).

4. Calibration of model parameters through information theory

The commonly adopted forecasting scores for MJO in the literature (Lin et al. 2008; Gottschalck et al. 2010; Rashid et al. 2011) are the root-mean-squared error and the bivariate correlation, both of which are in the path-wise sense. However, as illustrated in the following motivating example, these path-wise measures are insufficient in assessing the prediction skill of the RMM indices in the sense that they are incapable of measuring the lack of information in the model forecast compared with the truth and consequently evaluating the forecasts with these metrics can fail to capture the peaks that correspond to the strong MJO events in the true signals.

To improve the prediction skill, an information-theoretic framework (Roulston and Smith 2002; Majda and Gershgorin 2010; Majda and Branicki 2012; Weisheimer et al. 2014; Branicki and Majda 2014b), which incorporates both the surrogates of path-wise errors and the measure of the lack of information, is utilized to calibrate the model parameters in (2) and only a short training phase of three years is needed. These information measures are also adopted to assess the forecasting skill during the prediction phase.

a. A motivating example

Denote by \( \mathbf{u}_t = (u_{t,1}, u_{t,2}) \) the RMM indices and \( \mathbf{u}_{t}^{\text{pred}} = (u_{t,1}^{\text{pred}}, u_{t,2}^{\text{pred}}) \) the predicted time series. The two path-wise measures, i.e., the root-mean-squared error (RMSE) and the bivariate correlation (Corr), are defined by

\[
\text{RMSE}(\mathbf{u}_t, \mathbf{u}_t^{\text{pred}}) = \sqrt{\frac{\sum_{t=1}^{n} (u_{t,1} - u_{t,1}^{\text{pred}})^2 + (u_{t,2} - u_{t,2}^{\text{pred}})^2}{n}},
\]
and

$$\text{Corr}(u_t, u_{t}^{\text{pred}}) = \frac{\sum_{t=1}^{n} (u_{t,1} u_{t,1}^{\text{pred}} + u_{t,2} u_{t,2}^{\text{pred}})}{\sqrt{\sum_{t=1}^{n} (u_{t,1}^2 + u_{t,2}^2)} \sqrt{\sum_{t=1}^{n} ((u_{t,1}^{\text{pred}})^2 + (u_{t,2}^{\text{pred}})^2)},}$$  \hspace{1cm} (6)$$

where $n$ is the number of the points in the time series. Since the RMM index has zero mean for both the components, the bivariate correlation is essentially the same as the anomaly pattern correlation.

In order to illustrate the insufficiency of the two path-wise measures (5) and (6) in assessing the skill of the RMM prediction, the ensemble forecasting of the RMM indices at a lead time of 25 days utilizing the low-order stochastic model (2) with the same SSA(1-2) initializations but two different sets of parameters for model in (2) are shown in Figure 2. The prediction time interval is from August 2005 to December 2008.

Looking at the RMM1 index (thin black) and comparing its ensemble mean predictions (thick blue) as shown in panel (a) and (c), the severe underestimation of the forecasting amplitudes in Prediction #2 leads to a much less skillful prediction than that of Prediction #1; despite the fact that they have nearly the same anomaly pattern correlation and RMS error, the time-averaged PDF associated with the predicted signal in Prediction #1 as shown in panel (b) is almost perfectly overlapped with that of the truth while the PDF corresponding to Prediction #2 shown in panel (d) is highly concentrated around the origin which indicates a large lack of information in the forecasting statistics. In fact, the phases with large amplitudes, corresponding to the strong MJO events, are of more practical concern and the failure of capturing the peaks as in Prediction #2 implies an almost useless forecasting. However, the comparable path-wise scores fail to distinguish the two predictions because neither of the traditional path-wise measures assesses the lack of information in forecasting.
An information-theoretic framework is utilized as a systematical procedure to calibrate the parameters \( \theta = (d_u, \sigma_u, d_v, \sigma_v, d_\omega, \sigma_\omega, a, \gamma) \) in the model (2) during a short training phase. The calibration is implemented by optimizing a certain prescribed information criterion \( \mathcal{M}(\theta) \), which combines different information measures, over the parameter space. The information measures and information criterion in this work are restricted to the Gaussian framework for both simplicity and computational efficiency. These information measures are also utilized to assess the forecasting skill in the prediction stage. See (Majda and Branicki 2012; Branicki et al. 2013) for applications of information theory in the non-Gaussian framework.

Consider the following three information-theoretic measures (Branicki and Majda 2014b) in a Gaussian framework.

- **The Shannon entropy** of the residual \( U_t = u_t - u^\text{pred}_t \),
  \[
  S(U_t) = \frac{1}{2} \ln \det \left[ C_{(11)t} + C_{(22)t} - 2 C_{(12)t} \right] + \frac{1}{2} q (1 + \ln 2\pi). \tag{7}
  \]

- **The relative entropy** of the PDF \( \pi^{\text{pred}} \) associated with \( u^\text{pred}_t \) compared with the truth \( \pi \),
  \[
  \mathcal{P}(\pi, \pi^{\text{pred}}) = \frac{1}{2} \mathbb{E}[u_t - u^{\text{pred}}_t]^* C_{(22)t}^{-1} \mathbb{E}[u_t - u^{\text{pred}}_t]
  + \frac{1}{2} \left[ \text{tr} \left( C_{(11)t} (C_{(22)t})^{-1} \right) - \ln \det \left( (C_{(11)t})^{-1} C_{(22)t} \right) - q \right]. \tag{8}
  \]

- **The mutual information** between the true signal \( u_t \) and the predicted one \( u^\text{pred}_t \),
  \[
  M(u_t, u^{\text{pred}}_t) = -\frac{1}{2} \ln \det \left( I - (C_{(22)t})^{-1} C_{(11)t}^* (C_{(11)t})^{-1} C_{(12)t} \right). \tag{9}
  \]

Here \( q \) is the dimension of the observed variables and \( \mathbb{E}(\cdot) \) represents the expectation value of \( \cdot \); the quantities \( C_{(11)t} \) and \( C_{(22)t} \) are the variance of \( u_t \) and \( u^\text{pred}_t \) respectively and \( C_{(12)t} \) is the
covariance between $u_t$ and $u_t^{pred}$. The general definitions of the three information measures and their properties are listed in Appendix A.

Each one of the three measures provides different information about the forecasting skill. The Shannon entropy of the residual $S(U_t)$ measures the uncertainty in the prediction $u_t^{pred}$ compared with the truth $u_t$. The mutual information $M(u_t, u_t^{pred})$ measures the dependence between $u_t$ and $u_t^{pred}$. These two information measures are the surrogates for the RMS error and the anomaly pattern correlation in the path-wise sense, respectively. For example, maximizing the mutual information in (9) is the same as maximizing the anomaly pattern correlation and the exponential of the Shannon entropy (7) for the forecast error generalizes the RMS error. The relative entropy $P(\pi, \pi_{pred})$ quantifies the lack of information in the statistics of the prediction $u_t^{pred}$ relative to that of the truth $u_t$ (Roulston and Smith 2002; Majda and Gershgorin 2010; Majda and Branicki 2012). Therefore, it is an indicator of assessing the disparity in the amplitudes and peaks between $u_t^{pred}$ and $u_t$.

Although the comparable path-wise skill scores of the two predictions in Figure 2 correspond to a small difference in both the Shannon entropy and the mutual information, the disparity in amplitude of the two predictions leads to a significant difference in the relative entropy $P(\pi, \pi_{pred})$. The small gap in the predicted PDFs compared with the truth for Prediction #1 implies a small lack of information with $P(\pi, \pi_{1}^{pred}) = 0.1014$ while the severe underestimation in amplitude of prediction #2 results in a huge amount information loss with $P(\pi, \pi_{2}^{pred}) = 4.5010$. Therefore, by incorporating the relative entropy $P(\pi, \pi_{pred})$ into the skill scores, Prediction #1 in Figure 2 is expected to be more skillful than Prediction #2.
c. Calibration of the model with information theory

To combine the Shannon entropy of the residual (7), the relative entropy (8) and the mutual information (9) into an information criterion, we follow the same idea proposed in (Branicki and Majda 2014b). We assess the prediction skill in the training phase through the functional

$$M(\theta) = \frac{\left( \exp(\tilde{S}(u_t - u_{t}^{\text{pred}}(\theta))) \right) + P(\pi(u_t), \pi_{\text{pred}}^{\text{pred}}(\theta))}{M(u_t, u_{t}^{\text{pred}}(\theta))},$$

(10)

where the constant prefactor in the Shannon entropy (7) is removed in (10)

$$\tilde{S}(u_t - u_{t}^{\text{pred}}(\theta)) = S(u_t - u_{t}^{\text{pred}}(\theta)) - \frac{1}{2}q(1 + \ln(2\pi)).$$

This guarantees the weights of the three information measures in the information criterion (10) are of the same order in the calibration of the model parameters for the RMM indices.

The information criterion $M(\theta)$ decreases monotonically with the decreasing of both the Shannon entropy and the relative entropy and the increasing of the mutual information. Hence, the parameters in the model (2) are calibrated via minimizing the information criterion $M(\theta)$ over the parameter space.

What remains is to select an appropriate expression of $u_{t}^{\text{pred}}(\theta)$ in the training phase $T_{\text{Training}}$. In this work, the same $u_{t}^{\text{pred}}(\theta)$ in the training phase as the target in the prediction phase is adopted to optimize the parameters $\theta$. Depending on the prediction of interest, two approaches are utilized:

**Approach (1).** Minimizing the information criterion $M$ for a specific $S$-day lead prediction skill, where in (10) $u_{t}^{\text{pred}}$ is the $S$-day lead prediction of $u_t$ in $T_{\text{Training}}$.

**Approach (2).** Minimizing the averaged information criterion $\bar{M} = (M_1 + \ldots + M_m)/m$, where $u_{t}^{\text{pred}}$ in $M_i$ is the $S_i$-day lead prediction of $u_t$ in $T_{\text{Training}}$. 
The first approach aims at optimizing the prediction skill for a specific lead time while the second one is regarded as an approximation of optimizing the prediction skill within a certain range of forecast leads.

The MJO prediction at lead times of 15 and 25 days and the overall medium-range forecasting are of particular concern. According to the properties of SSA reconstructed initialization as described in Section 2, we adopt the following strategies in the calibration stage.

**Strategy-1a.**

\[ S = 15 \text{ with SSA(1-4) initialization of the observed variables in Approach (1).} \]  

**Strategy-1b.**

\[ S = 25 \text{ with SSA(1-2) initialization of the observed variables in Approach (1).} \]  

**Strategy-2.**

\[ S_1 = 25, \ S_2 = 35 \text{ and } S_3 = 45 \ (m = 3) \]  

with SSA(1-2) initialization of the observed variables in Approach (2).

Two forecast experiments are studied below. In the first forecast a 3-year short training period \( \mathcal{T}_{training_1} \) from January 2002 to December 2004 is utilized for model calibration with different strategies and the forecasts are validated in an independent time interval from August 2005 to December 2008 in order to compare with other work in the literature (Kondrashov et al. 2013); in the second forecast we adopt another 3-year training period \( \mathcal{T}_{training_2} \) from January 1981 to December 1983 and compare the forecast skill over a long range of different years from 1985 to 2013. In the prediction stage, ensemble prediction with SSA(1-4) initialization of the observed variables is utilized when the parameters are calibrated by Strategy-1a (11) while that with SSA(1-2) initialization is adopted when the parameters are calibrated by the other two strategies (12) and (13).
Trained in the period $T_{Training}$ from January 2002 to December 2004, the optimized parameters utilizing the three strategies are listed in Table 1. The posterior estimates of the hidden processes in the training period via data assimilation utilizing Strategy-1b are shown in Figure 3. These estimates vary as a function of time, depending on the observations from the RMM indices. Particularly, the posterior mean estimates of stochastic damping $\nu$ at the phases corresponding to large bursts in the observed variables, e.g., May 2002, January 2003 and April 2004, are strongly positive with small uncertainty. Starting from such phases, the stochastic anti-damping $\gamma\nu$ overwhelms the deterministic damping $-d_u$ and therefore running the model (2) forward leads to the solutions with local exponential growth in the observed variables $u_1$ and $u_2$, corresponding to the intermittent instability. The three strategies provide similar qualitative results for the posterior estimations. Yet, the posterior mean estimation utilizing Strategy-1a (11) has a more significant variation and the posterior covariance is smaller because more information is included in SSA(1-4) initialization. On the other hand, Strategy-2 (13) leads to a slightly larger posterior variance in the stochastic damping $\nu$ than Strategy-1b (12) due to the fact that the feedback from the observations utilizing Strategy-2 is weaker and thus the estimation contains more uncertainty.

Before systematically studying the prediction skill of the low-order stochastic model (2), we provide the evidence to support that the calibrated parameters in the 3-year short training phase are also the nearly optimal ones in the prediction phase. To this end, the prediction skill for the RMM indices in an independent time interval August 2005 to December 2008 at a lead time of 25 days as a function of parameter variations around the optimal values are shown in Figure 4, where the optimized parameters are given by Strategy-1b (12), i.e., the calibration of the forecasting skill at a lead time of $S = 25$ days in the training phase $T_{Training}$. In row (a) of Figure 4, the optimized parameters in the training phase almost coincide with the minimizer of the information criterion $M(\theta)$, indicating the sufficiency of such a 3-year short training period. The minimizer
of the information criterion is the combination of the three information measures given in (10) and therefore it is not necessarily the minimizer of the RMS error or the maximizer of the bivariate correlation. Referring back to the two predictions from Figure 2 in Section 4, it is also evident from the last column that the prediction with \( \gamma = 1.5 \) results in a comparable Shannon entropy (RMS error) and only slightly lower mutual information (bivariate correlation) than that with \( \gamma = 0.5 \) but a much larger relative entropy is observed with \( \gamma = 1.5 \), which enhances the total information criterion \( \mathcal{M} \) significantly. This indicates the capability of information-theoretic framework in distinguishing the skill of the two predictions in the motivating example.

5. Prediction results

a. Prediction skill of the RMM indices utilizing the nonlinear physics-constrained low-order stochastic model with different calibration strategies

In the first experiment, the nonlinear physics-constrained low-order stochastic model (2) is trained in the time interval \( \mathcal{T}_{\text{Training}} \) from January 2002 to December 2004. We validate the forecasts with different strategies proposed in Section 4 in an independent time interval from August 2005 to December 2008, which involves sufficient strong and weak MJO events as well as regular and irregular phases. Note that the forecasting skill utilizing PNF method was studied extensively in the same time interval (Kondrashov et al. 2013) as well. The comparison of the forecasting skill of the prediction algorithm proposed in this work with that of the PNF method will be included in this subsection.

In the following, we report the prediction skill of the nonlinear physics-constrained low-order stochastic model (2) with the optimized parameters from Table 1 and the ensemble initialization scheme for the hidden variables described in Section 3. The path-wise skill scores and the infor-
mation measures for prediction as a function of lead days $S$ utilizing different calibration strategies are shown in Figure 5. Recall that calibration Strategy-1a (11) aims at optimizing the parameters for prediction at a lead time of 15 days with SSA(1-4) initialization of the observed variables while Strategy-1b (12) and Strategy-2 (13) are designed for calibrating the model parameters at a lead time of 25 days and at lead times of 25, 35 and 45 days respectively with SSA(1-2) initialization.

Since SSA(1-4) removes the noise-like fluctuations in the smallest scales but nevertheless contains most of the information in the RMM indices, the skillful short-term prediction with SSA(1-4) initialization is extended to 20 days compared with the predictability limit of 15 days utilizing the raw RMM indices for initialization (not shown here) in both the path-wise sense (panel (a)-(b)) and the measure of the lack of information in prediction (panel (d)). The skillful predictions in the time domain at lead times of $S = 5, 10$ and 15 days are shown in Figure 6 and the small difference in time-averaged PDFs between the truth and the prediction verifies the insignificant lack of information in prediction. Yet, the medium range forecasting with SSA(1-4) initialization is unskillful due to the remaining small scale fluctuations in the initial values. Besides, as seen in panel (d) of Figure 5, since the calibration is designed to optimize the 15-day lead prediction, the lack of information in the forecast RMM indices shoots up at lead times that are more than 30 days. We have also tested the calibration strategy aiming at minimizing the averaged information criterion at lead times of 25, 35 and 45 days with SSA(1-4) initialization (not shown here) and find that skillful prediction remains only up to 20 days as well.

In order to extend the useful prediction to medium range, SSA(1-2) initialization is adopted. Although SSA(1-2) initialization leads to some intrinsic barriers for a very short range forecasting as shown in the zoomed-in panel (d) of Figure 5, the skillful prediction lasts up to 30 days for both Strategy-1b (12) and Strategy-2 (13). The difference in prediction between these two strategies is that Strategy-2 (13) leads to a smaller lack of information in the forecast RMM indices within
the 30 ∼ 60 days range. This is because Strategy-2 (13) focuses on the medium-range forecasting within the interval that covers 25-, 35- and 45-day lead predictions and therefore the relative entropy remains low for the medium range forecasting as shown in panel (d) of Figure 5. On the other hand, Strategy-1b (12) optimizes the prediction skill only at a lead time of $S = 25$ days and therefore it has little effect on the prediction when $S$ is far from 25. It is worthwhile noticing that although both the information criterion $\mathcal{M}$ and relative entropy $P$ of Strategy-1b (12) are larger than those of Strategy-2 (13) in the medium range forecasting for $S > 30$ as expected, the RMS error (Shannon entropy) of Strategy-1b (12) is smaller instead. This again implies the potentially misleading assessment of the prediction skill utilizing only the path-wise measures.

Figure 7 shows prediction in the time domain at lead times of 25, 35 and 45 days utilizing Strategy-2 (13) and SSA(1-2) initialization of the observed variables. The prediction at a 25-day lead time is quite accurate in most phases regarding both the path-wise error and the anomaly pattern correlation. Particularly, the peaks and strong MJO events are well predicted. The prediction is unskillful only in the strongly irregular phases, e.g., April 2006 to July 2006, November 2006, and November 2008 to December 2008, which can be attributed to both the deficiency of SSA initialization in these phases and the model error in (2) compared to the perfect physics. The predictions at lead times of 35 and 45 days are nevertheless capable of capturing the main trend of the truth and the bivariate correlation of the strong MJO events even in a 45-day lead prediction is still close to 0.5, indicating the skillful prediction. In addition, the insignificant difference in the time-averaged PDFs for all the 25-, 35- and 45-day lead predictions implies the lack of information in the forecast RMM indices compared with the truth is small.

The skill scores for prediction utilizing the nonlinear physics-constrained low-order stochastic model (2) as a function of lead time are illustrated in Figure 8. Information-theoretic Strategy-2 (13) with SSA(1-2) initialization of the observed variables is utilized for model calibration. The
overall prediction is skillful up to 30 days and the useful prediction of the strong MJO events is about 40 days, both of which are much improved compared with those utilizing empirical model reduction (EMR) and past-noise forecasting (PNF) as shown in Figure 2 of Kondrashov et al. (2013). The only unskillful prediction via information theory is the bivariate correlation of the weak MJO events, which is, however, of less concern in practice. Furthermore, the forecast RMM indices in the time domain at a lead time of 25 days utilizing both EMR and PNF methods as shown in Figure 1 of Kondrashov et al. (2013) look quite similar to those of Prediction #2 in Figure 2 of the motivating example, demonstrating the severe underestimation of the forecasting amplitudes by these methods. This contrasts with the prediction here incorporating information theory in the calibration stage, where the forecast RMM indices as shown in panel (a) of Figure 7 capture all the peaks and extreme events and minimize the information deficiency in prediction.

Finally, the long range forecasting of the nonlinear physics-constrained low-order stochastic model (2) with SSA(1-2) initialization calibrated by Strategy-2 (13) is shown in Figure 9. Different panels show the prediction starting from the first day of different months in year 2007 and each prediction lasts for 6 months. The ensemble mean predictions are skillful up to about 1-2 months with the ensemble spread evolving with the same trend as the truth. The ensemble mean predictions do not have any long range skill but the ensemble spread automatically predicts this lack of skill and the envelope of the ensemble predictions is able to capture the true signal. This is a significant attractive feature of the methods developed here.

b. Prediction skill of the RMM indices utilizing the nonlinear physics-constrained low-order stochastic model in different years

We now explore the prediction skill of the RMM indices in different years utilizing the nonlinear physics-constrained low-order stochastic model (2) with case studies. To check the prediction skill
for a longer period, we modify the training phase to an early period $\mathcal{S}_2^{\text{Training}}$ from 1981 to 1983 and predict RMM indices in each year from 1985 to 2013. We calibrate the model parameters via Strategy-2 (13) and the initialization of the observed variables in the prediction stage is given by SSA(1-2). Although this training period contains one strong ENSO (ENSO: El Niño Southern Oscillation) event, it is remarkable that essentially the same parameters for (2) occur in this 3-year calibration as those in $\mathcal{S}_1^{\text{Training}}$. Thus, we omit these values here and refer the readers to Table 1.

Below, the RMS error and bivariate correlation at a lead time of 25 days are utilized as the measures of skillful prediction. The case studies cover the periods of two field experiments: Tropical Ocean Global Atmosphere Coupled Ocean Atmosphere Response Experiment (TOGA-COARE; November 1992 to February 1993) (Webster and Lukas 1992; Yanai et al. 2000) and Dynamics of MJO initiated and organized internationally under the Cooperative Indian Ocean Experiment on Intraseasonal Variability (CINDY/DYNAMO; October 2011 to March 2012) (Yoneyama et al. 2013; Zhang et al. 2013). In addition, motivated by the well-known MJO-ENSO interactions, we also study the prediction skill of the RMM indices accompanying strong ENSO phases.

The RMS error and bivariate correlation for prediction in different years are shown in panel (b) and (c) in Figure 10. Note that the yearly averaged skill scores of the $i$-th year are put in the middle of the $i$-th and the $(i+1)$-th years. The Multivariate ENSO Index (MEI) (Wolter and Timlin 1993, 1998) is shown in panel (a) for comparison. MEI is a monthly averaged ENSO index, based on the six main observed variables over the tropical Pacific: sea-level pressure, zonal and meridional components of the surface wind, sea surface temperature, surface air temperature, and total cloudiness fraction of the sky. The predictions of the RMM indices in time domain in 8 different years to be discussed below are shown in Figure 11.

TOGA-COARE was conducted in November 1992 to February 1993, where two pronounced Madden-Julian oscillation (MJO) events associated with super cloud clusters and westerly wind
bursts were observed and are well reflected in the RMM indices. Actually, besides the TOGA-COARE period, the RMM indices illustrate the strong MJO events with regular intraseasonal variability frequencies around 50 days throughout the whole year 1992 (panel (c) in Figure 11) and therefore a skillful overall prediction (RMSE = 1.08 and Corr = 0.587) and an excellent prediction of strong events (Corr(S) = 0.672) are obtained for the low-order model in (2) at a lead time of 25 days. The MJO events throughout the year 1993 (panel (d) in Figure 11) are also moderately strong and the prediction skill of the strong MJO events is skillful (Corr(S) = 0.592). Yet, the termination phase of the pronounced MJO event in March and the high-frequency variability in December are both hard to predict, which deteriorate the overall prediction skill in this year.

CINDY/DYNAMO collected unprecedented observations during October 2011 to March 2012. Coincidentally, a La Niña event took place at nearly the same time with about −1°C anomaly in the sea surface temperature (SST) (Zhang et al. 2013). Three MJO events, occurred over the tropical Indian Ocean in late October, late November and late December of year 2011, were observed. However, the first two of them barely reached the Pacific Ocean, partially because of the La Niña condition while the late December MJO event is not recognized by the RMM indices as an independent MJO event. All these three MJO events have high-frequency variability with only 30 days period, which leads to a tough prediction. Besides, similar fast oscillation of MJO phases are observed in the RMM indices during May and June in year 2011 (Panel (h) in Figure 11). All these factors lead to the unskillful prediction (Corr = 0.416) of this year.

Looking at the skill scores in Figure 10, the most poorly predicted years are 1987, 1998 and 2011 regarding both the overall and strong MJO forecasting skill. The unskillful prediction in year 2011 due to a large amount of high-frequency variability was explained above. The other two poorly predicted years 1987 and 1998 are accompanied with the strong ENSO phases. The oscillation frequencies of the RMM indices in late 1986 and the whole year 1987 are irregular,
containing quite a few fast oscillations, and therefore the correlations for both the overall events and the strong events are below the skillful level (Corr = 0.384 and Corr($S$) = 0.411). Similar high-frequency and irregular oscillated phases are observed from the late 1997 to September 1998 and the small amplitudes of the RMM indices during this strong ENSO phase suggest that ENSO sometimes serves to damp MJO activity (Szekely et al. 2014). This contrasts with the relatively strong MJO activity with nearly 50-day periodic bursts and breaks and their skillful predictions during the pre-ENSO transition phases, year 1985 (Corr = 0.588 and Corr($S$) = 0.677) and 1996 (Corr = 0.621 and Corr($S$) = 0.696), and the post-ENSO transition phases, year 1988 (Corr = 0.707 and Corr($S$) = 0.759) and 1999 (Corr = 0.606 and Corr($S$) = 0.661). Actually, the fact that a dramatic increased amplitude in MJO events is observed before ENSO peaks at a lead time of 8 months (Hendon et al. 2007) is evidence to support the skillful predictions during the pre-ENSO phases. Thus, the transition to ENSO is a barrier for skillful medium-range prediction in the low-order stochastic models from (2).

c. Prediction skill of the RMM indices utilizing the nonlinear physics-constrained low-order stochastic model in the phase space

To explore the progression of the MJO in prediction through different phases, we study the phase diagrams of the RMM indices utilizing the nonlinear physics-constrained low-order stochastic model (2). Strategy-2 (13) is utilized in the calibration period $\mathcal{F}_2^{Training}$ from year 1981 to 1983 and SSA(1-2) initialization is adopted in the prediction period from year 1985 to 2013.

Four phase diagrams of RMM1 and RMM2 prediction up to 45 days are included in Figure 12. Panel (a) shows a weak MJO period while panel (c) shows a moderate period and panel (b) and (d) show two strong MJO periods. The ensemble mean prediction in panel (a)-(c) succeeds in capturing the trend of the truth which is covered by the ensemble spread. Due to the large error
in the initialization, the short-range prediction in panel (d) is not skillful, but after a few days’
adjustment the forecasting in medium-range describes the truth accurately.

Next, we study the forecasting skill of the nonlinear low-order stochastic model (2) in different
phases. Table 2 demonstrates the skill scores with respect to the RMS error and bivariate correla-
tion in the prediction period at a lead time of 25 days starting from and ending at different phases.
In the same Table, the percentages of the strong MJO events within each phase are included. It is
evident that the prediction utilizing the nonlinear low-order stochastic model (2) is skillful in all
the phases. Particularly, starting from phase 2 and phase 3, the bivariate correlations at a lead time
of 25 days are almost 0.6 and the RMS errors (1.1767 and 1.2145) are far below the climatological
forecast (\(\sqrt{2}\)). In addition, the bivariate correlations at a lead time of 25 days are all above 0.55
when MJO ends at phase 1, 2, 3, 5, 6 and 8. These facts contrast with the predictability of most
current models (Maharaj and Wheeler 2005; Woolnough et al. 2007; Vitart et al. 2007; Vitart and
Molteni 2010), which are unskillful in the MJO initialization phase (phase 2) or when MJO is over
the Maritime continent (phase 5). Actually, the improvement in prediction utilizing the nonlinear
low-order stochastic model (2) is associated with the relatively large amount of the strong MJO
events at these phases and the fact that the model is particularly skillful in forecasting the strong
MJO events.

\textit{d. Prediction skill utilizing the linear models}

To understand the role of nonlinearity in the nonlinear physics-constrained low-order stochastic
model (2) in predicting the RMM indices, the prediction skill utilizing the following two linear
models are studied.

The first linear model is the reduced version of (2) by dropping the nonlinear interaction with
the hidden variables. The result is a two-dimensional (2D) stochastic oscillator system containing
only the observed variables,

\[ du_1 = (-d_u u_1 - a u_2) dt + \sigma_u dW_{u_1}, \] (14a)

\[ du_2 = (-d_u u_2 + a u_1) dt + \sigma_u dW_{u_2}. \] (14b)

The second linear model is a four-dimensional (4D) system, in which the two observed variables \( u_1, u_2 \) interact with the two hidden variables \( b_1, b_2 \) only in an additive way.

\[ du_1 = (-d_u u_1 - a u_2 + \alpha b_1) dt + \sigma_u dW_{u_1}, \] (15a)

\[ du_2 = (-d_u u_2 + a u_1 + \alpha b_2) dt + \sigma_u dW_{u_2}, \] (15b)

\[ db_1 = (-d_b b_1 + \beta u_1) dt + \sigma_b dW_{b_1}, \] (15c)

\[ db_2 = (-d_b b_2 + \beta u_2) dt + \sigma_b dW_{b_2}. \] (15d)

Note that the 4D model (15) is a linear stochastic oscillator driven by the stochastic forcing with memory effect (Majda and Harlim 2012) and its simplified version with \( \sigma_u = \beta = 0 \) is the damped harmonic oscillator model in (Oliver and Thompson 2011).

To compare the forecasting skill of the nonlinear stochastic model (2) with the two linear models (14) and (15), we calibrate the two linear models in the same training period \( T_{Training} \) as the nonlinear model (2), i.e., from January 2002 to December 2004, utilizing calibration Strategy-2 (13). Note that the 2D model (14) contains only the observed variables and therefore the ensemble prediction algorithm is straightforward while the same data assimilation scheme as in Section 3 is utilized to estimate the hidden variables \( b_1 \) and \( b_2 \) in the 4D model (15). The resulting optimized parameters are listed in Table 3.

Figure 13 illustrates the comparison of the information measures and the information criterion in prediction as a function of lead days utilizing the nonlinear physics-constrained stochastic model (2) and the two linear stochastic models (14) and (15) equipped with the optimized parameters.
The validation interval for prediction is again from August 2005 to December 2008 and SSA(1-2) initialization is adopted in the prediction stage. The prediction results utilizing different models lead to indistinguishable mutual information and thereby almost identical anomaly pattern correlation. The Shannon entropy of the residual in prediction utilizing the two linear models (14) and (15) is even slightly smaller than that utilizing the nonlinear physics-constrained stochastic model (2), indicating the smaller RMS error in prediction with the linear models. However, conspicuous information barriers are revealed in the relative entropy utilizing the two linear models for prediction; the information barrier in the 2D linear model (14) is more significant than that in the 4D linear model (15) due to its more simplified form. These information barriers (Majda and Gershgorin 2010; Majda and Branicki 2012) imply the failure of capturing the extreme events of the two linear models in prediction as shown in panel (a)-(c) of Figure 14, which illustrates the predictions of the three models in the time domain. It is obvious that many extreme events, such as those around 10 Sep 2005, 15 Oct 2006, 15 Jul 2007 and 15 Oct 2007, are captured well by the nonlinear physics-constrained stochastic model (2) but are missed by the two linear models (14) and (15) and therefore the time-averaged PDFs of prediction utilizing the two linear models, shown in panel (e)-(f), have a significantly smaller variance compared with that associated with the RMM1 index.

The information barriers in prediction by the linear models even with additive correlated stochastic forcing indicate the necessity of including the multiplicative noise and the nonlinear interaction between the observed and hidden variables in the low-order stochastic model. The stochastic damping, which is the mechanism to stimulate the local exponential growth of the observed variables in prediction, plays a significant role in capturing the extreme events. See (Chen et al. 2014b) for another example. Therefore, the results from the linear models emphasize once again that the traditional path-wise measures alone are insufficient in assessing the prediction skill.
and assessing the lack of information is essential for useful prediction. We also find that the ensemble spread utilizing the linear models (14) and (15) is not as skillful as that of the nonlinear physics-constrained low-order stochastic model (2) in representing the lack of information in long range forecasting, but we omit detailed discussion here.

6. Conclusions

In this paper, we predict the RMM indices utilizing suitable low-order stochastic models. The systematic physics-constrained nonlinear regression strategies for time series developed recently (Majda and Harlim 2013; Harlim et al. 2014) as well as the stochastic skeleton model (Thual et al. 2013) suggest a four-dimensional model (2) with two observed variables $u_1, u_2$ for the RMM indices and two hidden variables $v, \omega_u$ representing stochastic damping and stochastic phasing. This low-order model involves correlated multiplicative noise defined through energy conserving nonlinear interactions between the observed and hidden variables as well as additive stochastic noise.

The special structure of the low-order model allows efficient data assimilation for the initialization of the hidden variables that facilities the ensemble prediction algorithm. The initialization of the observed variables is given by the SSA reconstruction for the sake of removing the small-scale noise-like fluctuations.

The failure of measuring the disparity in the peaks between the observed and forecast RMM indices in Figure 2 indicates the insufficiency of the standard path-wise measures, i.e., anomaly pattern correlation and RMS error. Therefore, an information-theoretic framework (Branicki and Majda 2014b) is applied to the calibration of model parameters in a short training phase of three years. This framework involves generalizations of the anomaly pattern correlation, the RMS error and the information deficiency in the model forecast. The nonlinear stochastic models (2) show skillful prediction for 30 days on average in these metrics. More importantly, the predictions
succeed in capturing the amplitudes of the RMM index and the useful skill of forecasting strong
MJO events is around 40 days. In addition, the prediction at a lead time of 25 days is skillful in all
the eight phases. Regarding the prediction in different years, the prediction is quite skillful in the
pre-ENSO and post-ENSO transition years while the transition to ENSO is a barrier for skillful
medium-range forecasting in the low-order stochastic model (2). The very long range forecasts
by the nonlinear stochastic model also succeed in capturing the truth within the ensemble spread,
another attractive feature (see Figure 9). Furthermore, the information barriers in prediction by the
linear models imply the necessity of the nonlinear interactions between the observed and hidden
variables as well as the multiplicative noise in these low-order stochastic models.

Many General Circulation Models (GCMs) and multi-model ensemble systems (Maloney and
Kiehl 2002; Liess and Bengtsson 2004; Zhang et al. 2006; Vitart et al. 2007; Vitart and Molteni
2010; Fu et al. 2013; Ling et al. 2014) are suffering from underestimating the MJO amplitudes in
prediction. In these operational models, additive noise is the main source of the noise and path-
wise skill scores are the typical measures of the predictability. Motivated by the improvement
of the nonlinear stochastic models with multiplicative noise (2) compared with the linear models
with only additive noise (14) and (15) in predicting RMM indices (see Figure 14), involving the
multiplicative noise in GCMs is able to enhance the amplitudes of prediction, especially for captur-
ing the local extreme events. Stochastic parameterization based on multiplicative noise improves
the capability of coarse resolution models to represent the MJO (Thual et al. 2013; Deng et al.
2014). In addition, the insufficiency of measuring the disparity in amplitudes by the path-wise
measures in predicting the RMM indices (see Figure 2) also implies the necessity of incorporating
the information-theoretic framework in GCMs for predicting MJO and other climate variabilities.

Two issues will be taken into consideration in the future. First, as mentioned in Section 2, SSA
initialization of the observed variables is impractical in real-time prediction because it utilizes
“future” information. Exploring effective and practical data smoothing approaches, which utilize only the information in the “historic” time-series, is of importance. In addition, as noted in Section 4, the three information measures and the information criterion in this work are restricted to the Gaussian framework. Although the climatological PDFs associated with the RMM indices are nearly Gaussian, those of the predictions are not necessarily Gaussian, e.g., panel (e) in Figure 7. It is interesting to see in the future if there is improved predictability utilizing information theory in the non-Gaussian framework.

Finally, we point out that information theory can be applied to many other problems. Imperfect predictions via Multi Model Ensemble forecasts are improved with the information-theoretic framework (Branicki and Majda 2014a). Applying information theory to the noisy Lagrangian tracers reveals the practical information barrier as a function of the number of tracers (Chen et al. 2014c,d). Information theory also has many desirable properties for characterizing model error (Majda et al. 2002; Majda and Gershgorin 2010; Majda and Branicki 2012; Branicki et al. 2013) and predictive skill (Kleeman 2002; Branicki and Majda 2012; Giannakis and Majda 2012; Giannakis et al. 2012; Chen et al. 2014a).

Acknowledgments. This research of A.J.M is partially supported by the Office of Naval Research grant ONR MURI N00014-12-1-0912. N.C. is supported as a graduate research assistant on this grant.

APPENDIX A

General definitions of the three information measures

This Appendix includes the general definitions and properties of the three information measures as stated in Section 4.
The Shannon entropy. The Shannon entropy $S(U_t)$ of the residual $U_t = u_t - u_t^{\text{pred}}$ is given by

$$S(U_t) := - \int p(U_t) \ln p(U_t) dU_t, \quad (A1)$$

which measures the uncertainty in the prediction $u_t^{\text{pred}}$ compared with the truth $u_t$.

The relative entropy. The relative entropy $P(\pi, \pi^{\text{pred}})$ of the PDF $\pi^{\text{pred}}$ associated with $u_t^{\text{pred}}$ compared with the truth $\pi$ is given by

$$P(\pi, \pi^{\text{pred}}) := \int \pi(u) \ln \frac{\pi(u)}{\pi^{\text{pred}}(u)} du, \quad (A2)$$

which quantifies the lack of information in the statistics of the prediction $u_t^{\text{pred}}$ relative to that of the truth $u_t$. The relative entropy is often interpreted as a 'distance' between the two probability densities but it is not a true metric. It is non-negative with $P = 0$ only when $\pi = \pi^{\text{pred}}$ and it is invariant under nonlinear changes of variables.

The mutual information. The mutual information $M(u_t, u_t^{\text{pred}})$ between the true signal $u_t$ and the predicted one $u_t^{\text{pred}}$ is given by the symmetric formula

$$M(u_t, u_t^{\text{pred}}) := \int \int p(u_t, u_t^{\text{pred}}) \ln \frac{p(u_t, u_t^{\text{pred}})}{\pi(u_t) \pi^{\text{pred}}(u_t^{\text{pred}})} du_t du_t^{\text{pred}}, \quad (A3)$$

which measures the dependence between these two processes. One useful interpretation of the mutual information is as a measure of the lack information in the factorized density $\pi(u_t) \pi^{\text{pred}}(u_t^{\text{pred}})$ relative to the joint density $p(u_t, u_t^{\text{pred}})$ which follows from the identity

$$M(u_t, u_t^{\text{pred}}) = P(p(u_t, u_t^{\text{pred}}), \pi(u_t) \pi^{\text{pred}}(u_t^{\text{pred}})).$$

Similar to the relative entropy (A2), the mutual information is nonnegative and it is invariant under nonlinear change of variables.

The explicit forms of these information measures in Gaussian framework are given by (7), (8) and (9), respectively.
Mathematical details of data assimilation and prediction algorithm

This Appendix involves the mathematical details of utilizing data assimilation algorithm to estimate the hidden variables \((v, \omega_u)\) and initialization of them during the prediction phase.

Denote by \(U = (u_1, u_2)^T\) and \(\Gamma = (v, \omega_u)^T\). The abstract form of the low-order stochastic model (2) are given as follows:

\[
dU_t = \left[ A_0(t, U) + A_1(t, U)\Gamma_t \right] dt + \Sigma_U(t, U) dW_U(t), \tag{B1a}
\]

\[
d\Gamma_t = \left[ a_0(t, U) + a_1(t, U)\Gamma_t \right] dt + \Sigma_\Gamma(t, U) dW_\Gamma(t), \tag{B1b}
\]

where

\[
A_0 = \begin{pmatrix}
-d_u u_1 - a u_2 \\
-d_u u_2 + a u_1
\end{pmatrix}, \quad
A_1 = \begin{pmatrix}
\gamma u_1 & -u_2 \\
\gamma u_2 & u_1
\end{pmatrix}, \quad
\Sigma_U = \begin{pmatrix}
\sigma_u \\
\sigma_u
\end{pmatrix},
\]

\[
a_0 = \begin{pmatrix}
-\gamma(u_1^2 + u_2^2) \\
0
\end{pmatrix}, \quad
a_1 = \begin{pmatrix}
-d_v \\
-d_\omega
\end{pmatrix}, \quad
\Sigma_\Gamma = \begin{pmatrix}
\sigma_v \\
\sigma_\omega
\end{pmatrix}.
\]

The model (B1) is a conditional Gaussian system conditioned on the observation \(U\). The closed analytic equations for the conditional Gaussian distributions of the hidden parameters \(v\) and \(\omega_u\) are given by (Liptser and Shiryaev 2001):

\[
d\mu_t = [a_0(t, U) + a_1(t, U)\mu_t] dt + (R_t A_1^*(t, U)) (\Sigma_U \Sigma_U^*)^{-1}(t, U) \times
\]

\[
[dU_t - (A_0(t, U) + A_1(t, U)\mu_t)] dt,
\]

\[
dR_t = \{a_1(t, U) R_t + R_t a_1^*(t, U) + (\Sigma_\Gamma \Sigma_\Gamma^*)((t, U)
\]

\[
- (R_t A_1^*(t, U)) (\Sigma_U \Sigma_U^*)^{-1}(t, U) (R_t A_1^*(t, U))^* \} dt,
\]

where \(\mu_t\) and \(R_t\) are the posterior mean and posterior covariance of the conditional distribution, respectively. The asterisk represents the complex conjugate.
As a remark, the formulae (B2) are optimal if and only if the signal is generated from system (B1). Since our observed signal RMM indices are not from the low-order nonlinear stochastic model (B1), the evolutions of the conditional Gaussian distributions (B2) are suboptimal.

References


Oscillation: A new perspective from Cloudsat. *Journal of the Atmospheric Sciences*, 68 (12),
3032–3051.

Roulston, M. S., and L. A. Smith, 2002: Evaluating probabilistic forecasts using information

Seo, K.-H., W. Wang, J. Gottschalck, Q. Zhang, J.-K. E. Schemm, W. R. Higgins, and A. Kumar,
2009: Evaluation of MJO forecast skill from several statistical and dynamical forecast models.
*Journal of Climate*, 22 (9), 2372–2388.

Stechmann, S. N., and A. J. Majda, 2014: Identifying the skeleton of the Madden-Julian Oscilla-

Straub, K. H., 2013: MJO initiation in the real-time multivariate MJO index. *Journal of Climate*,
26 (4), 1130–1151.

Szekely, E., D. Giannakis, and A. J. Majda, 2014: Extraction and prediction of coherent intrasea-
sonal signals in infrared brightness temperature data. *Climate Dynamics*, submitted.

Thual, S., A. J. Majda, and S. N. Stechmann, 2013: A stochastic skeleton model for the MJO.
*Journal of the Atmospheric Sciences*, 71, 697–715.

Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics, with applica-

Vitart, F., A. Leroy, and M. C. Wheeler, 2010: A comparison of dynamical and statistical predic-
138 (9), 3671–3682.


## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table 1.</th>
<th>Optimized parameters in the nonlinear low-order stochastic model (2) calibrated by the information criterion (10) in the training period $T_{1}^{Train}$ from January 2002 to December 2004.</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.</td>
<td>Skill scores of 25 days lead prediction from year 1985 to 2013 utilizing the nonlinear low-order stochastic model (2) starting and ending in different phases. The model parameters in predicting RMM index are calibrated via Strategy-2 (13) in the training period $T_{2}^{Train}$ and SSA(1-2) initialization of the observed variables is adopted in prediction.</td>
<td>43</td>
</tr>
<tr>
<td>Table 3.</td>
<td>Optimized parameters in the 2-D and 4-D linear models (14) and (15) calibrated by the information criterion (10) via Strategy-2 (13) in the training period $T_{1}^{Train}$ from January 2002 to December 2004.</td>
<td>44</td>
</tr>
</tbody>
</table>
TABLE 1. Optimized parameters in the nonlinear low-order stochastic model (2) calibrated by the information criterion (10) in the training period $\mathcal{H}_1^{\text{Training}}$ from January 2002 to December 2004.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Description</th>
<th>$\gamma$</th>
<th>$\alpha$</th>
<th>$d_u$</th>
<th>$\sigma_u$</th>
<th>$d_v$</th>
<th>$\sigma_v$</th>
<th>$d_\omega$</th>
<th>$\sigma_\omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy-1a (11)</td>
<td>SSA(1-4) initialization and $S = 15$</td>
<td>0.6</td>
<td>4.1</td>
<td>0.1</td>
<td>0.4</td>
<td>1.2</td>
<td>1.2</td>
<td>1.6</td>
<td>0.6</td>
</tr>
<tr>
<td>Strategy-1b (12)</td>
<td>SSA(1-2) initialization and $S = 25$</td>
<td>0.5</td>
<td>4.0</td>
<td>0.1</td>
<td>0.4</td>
<td>0.2</td>
<td>1.0</td>
<td>1.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Strategy-2 (13)</td>
<td>SSA(1-2) initialization and $S_1 = 25, S_2 = 35, S_3 = 45$</td>
<td>0.3</td>
<td>4.0</td>
<td>0.1</td>
<td>0.4</td>
<td>0.2</td>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>
TABLE 2. Skill scores of 25 days lead prediction from year 1985 to 2013 utilizing the nonlinear low-order stochastic model (2) starting and ending in different phases. The model parameters in predicting RMM index are calibrated via Strategy-2 (13) in the training period $\mathcal{F}_2^{Training}$ and SSA(1-2) initialization of the observed variables is adopted in prediction.

<table>
<thead>
<tr>
<th>I. Starting phase</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) RMSE</td>
<td>1.2280</td>
<td>1.1767</td>
<td>1.2145</td>
<td>1.2335</td>
<td>1.2256</td>
<td>1.2170</td>
<td>1.1825</td>
<td>1.2598</td>
</tr>
<tr>
<td>(b) Corr</td>
<td>0.5575</td>
<td>0.5957</td>
<td>0.5937</td>
<td>0.5216</td>
<td>0.5290</td>
<td>0.5427</td>
<td>0.5769</td>
<td>0.5618</td>
</tr>
<tr>
<td>(c) Strong events</td>
<td>0.62%</td>
<td>0.62%</td>
<td>0.63%</td>
<td>0.59%</td>
<td>0.58%</td>
<td>0.58%</td>
<td>0.64%</td>
<td>0.63%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Ending phase</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) RMSE</td>
<td>1.1878</td>
<td>1.1471</td>
<td>1.2360</td>
<td>1.2534</td>
<td>1.2199</td>
<td>1.1997</td>
<td>1.2572</td>
<td>1.2363</td>
</tr>
<tr>
<td>(b) Corr</td>
<td>0.5653</td>
<td>0.5710</td>
<td>0.5642</td>
<td>0.5192</td>
<td>0.5755</td>
<td>0.5791</td>
<td>0.5486</td>
<td>0.5627</td>
</tr>
<tr>
<td>(c) Strong events</td>
<td>0.62%</td>
<td>0.60%</td>
<td>0.61%</td>
<td>0.58%</td>
<td>0.64%</td>
<td>0.63%</td>
<td>0.61%</td>
<td>0.60%</td>
</tr>
</tbody>
</table>
TABLE 3. Optimized parameters in the 2-D and 4-D linear models (14) and (15) calibrated by the information criterion (10) via Strategy-2 (13) in the training period $\mathcal{T}_{training}$ from January 2002 to December 2004.

<table>
<thead>
<tr>
<th>Model</th>
<th>$a$</th>
<th>$d_a$</th>
<th>$\sigma_a$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$d_b$</th>
<th>$\sigma_b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-D linear model (14)</td>
<td>3.8</td>
<td>0.1</td>
<td>0.3</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>4-D linear model (15)</td>
<td>4.0</td>
<td>0.1</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td>1.75</td>
<td>0.5</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Fig. 1. The real-time Multivariate MJO (RMM) indices. Panels (a) and (c) show the two components of RMM indices (solid black) from January 2000 to May 2002 as well as the leading two (solid cyan) and four (dashed magenta) SSA reconstructed components (RCs). The time-averaged PDFs of the two RMM components over years 1980 to 2013 compared with their Gaussian fits are shown in panel (b) and (d). The PDFs are computed utilizing a histogram approach.

Fig. 2. Comparison of two predictions (thick blue) of the RMM indices (thin black) utilizing the real-time Multivariate MJO (RMM) indices. Panels (a) and (c) show the two components of the prediction #1 and #2 is γ = 0.5 and γ = 1.5 respectively. Other parameters adopted in the two predictions are the same: a = 4, σu = 0.4, σv = 1, d_u = 0.1, d_v = 0.2, σ_ω = 0.5 and d_ω = 1.2. The relative entropy is defined in (8), which measures the lack of information in the statistics of the prediction relative to that of the truth.

Fig. 3. Data assimilation utilizing Strategy-1b for the hidden variables in the training period $T_{1}^{\text{Training}}$ from January 2002 to December 2004. Panel (a) shows the SSA(1-2) initialization (thick blue) compared with the RMM1 index (thin black). Panel (b) and (c) show the posterior mean and posterior variance of the stochastic damping $\nu$ while panel (d) and (e) show those of the stochastic phase $\omega$. The dotted lines in panel (b) and (d) represent $\nu = d_u/\gamma = 0.2$ and $\omega_0 = 0$ while those in panel (c) and (e) indicate the variance associated with the prior distribution without filtering.

Fig. 4. Prediction skill of nonlinear physics-constrained low-order stochastic model (2) utilizing different measures at a lead day of 25 days in the time interval August 2005 to December 2008 as a function of the parameter variations around the optimal values (the red circles) which are calibrated by Strategy-1b (12) in the training phase $T_{1}^{\text{Training}}$. Row (a): Information criterion $\mathcal{H}(\theta)$; Row (b)-(d): three information measures, i.e., exponential of the rescaled Shannon entropy $\exp(S(\theta))$, the relative entropy $S(\theta)$ and the mutual information $M(\theta)$; Row (e)-(f): two path-wise measures, i.e. the RMS error $\text{RMSE}(\theta)$ and the bivariate correlation $\text{Corr}(\theta)$.

Fig. 5. Path-wise skill scores (panel (a)-(b)), information measures (panel (c)-(e)) and information criterion (panel (f)) of predicting the RMM indices utilizing the nonlinear physics-constrained low-order stochastic model (2) in the time interval August 2005 to December 2008 at lead times of 1 to 60 days with the optimal parameters calibrated by the three strategies (11), (12) and (13) in the training period $T_{1}^{\text{Training}}$, respectively. SSA(1-4) initialization of the observed variables is adopted in the prediction stage when model parameters are calibrated by Strategy-1a (11) while SSA(1-2) initialization is utilized with calibration Strategy-1b (12) and Strategy-2 (13). The sub-panels in (d) and (f) show the zoomed-in curves for the relative entropy and the information criterion up to 30 days.

Fig. 6. Prediction of nonlinear physics-constrained low-order stochastic model (2) (thick blue) in the time domain of RMM indices (thin black) with SSA(1-4) initialization (thick cyan) at lead times of 5, 10 and 15 days (panel (a)-(c)) and the time-averaged PDFs of prediction (panel (d)-(f)). In each caption, $\cdot(S)$ and $\cdot(W)$ represent the skill scores of strong and weak MJO events defined by (1) respectively. Only RMM1 prediction is shown here. The parameters are calibrated by Strategy-1a (11) in the training period $T_{1}^{\text{Training}}$. x axis is time in calendar months within the validation interval of August 2005 to December 2008.
Fig. 7. Prediction of nonlinear physics-constrained low-order stochastic model (2) (thick blue) in the time domain of RMM indices (thin black) with SSA(1-2) initialization (thick cyan) at lead times of 25, 35 and 45 days (panel (a)-(c)) and the time-averaged PDFs of prediction (panel (d)-(f)). In each caption, \((S)\) and \((W)\) represent the skill scores of strong and weak MJO events defined by (1) respectively. Only RMM1 prediction is shown here. The parameters are calibrated by Strategy-2 (13) in the training period. x axis is time in calendar months within the validation interval of August 2005 to December 2008.

Fig. 8. The path-wise skill scores for prediction of the nonlinear physics-constrained low-order stochastic model (2) as a function of lead time within the validation interval August 2005 to December 2008. The parameters are calibrated by information-theoretic Strategy-2 (13) in the training period. The skill scores for predicting the overall, the strong and the weak MJO events are shown in the solid, dashed and dash-dotted blue lines and the statistics between the prediction and SSA(1-2) reconstruction of the RMM indices is shown in the solid cyan lines. Panel (a) and (b) show the RMS error and bivariate correlation at lead times of 1 to 30 days while panel (c) and (d) show those at lead times of 1 to 60 days.

Fig. 9. Long-range forecasting of the nonlinear physics-constrained low-order stochastic model (2). The parameters are calibrated by Strategy-2 (13) in the training period. Different panel shows the prediction starting from the first day of each month in year 2007 and predicting with SSA(1-2) initialization up to 6 months. The label in x axis indicates the calendar month. In each panel, the solid black line is the RMM1 index; the solid thick blue line is the ensemble mean prediction which is averaged over \(N = 50\) ensemble members shown in the thin green lines; the SSA(1-2) reconstruction is given by the thick red dashed-dotted line. The ensemble spread with 95% confidence interval is indicated by the gray area.

Fig. 10. Comparison of Multivariate ENSO Index (MEI) with the prediction of RMM index at a lead time of 25 days. The model parameters in predicting RMM index are calibrated via Strategy-2 (13) in the training period from January 1981 to December 1983 and SSA(1-2) initialization of the observed variables is adopted in prediction. x axis is time in year with the label representing the first day of every fifth year. In panel (a), the two dotted lines indicate \(\pm 0.5\), the events outside which are regarded as strong El Niño or La Niña events. In panel (b) and (c), the yearly averaged skill scores of the \(i\)-th year are put in the middle of the \(i\)-th and the \((i + 1)\)-th years for the sake of the comparison with the monthly averaged MEI. The skill scores of overall, strong and weak MJO events are shown by blue solid, red dashed and green dash-dotted lines, respectively. The black dotted lines in panel (b) and (c) indicate one standard deviation of RMM indices and Corr = 0.5, respectively.

Fig. 11. Predictions (thick blue) of the RMM indices (thin black) at a lead time of 25 days in different years. The model parameters in predicting RMM index are calibrated via Strategy-2 (13) in the training period and SSA(1-2) initialization of the observed variables (thick cyan) is adopted in prediction. The label in x axis indicates the calendar month.

Fig. 12. Phase Diagrams of RMM1 and RMM2 prediction utilizing nonlinear physics-constrained low-order stochastic model (2) with SSA(1-2) initialization, where the parameters are calibrated by Strategy-2 (13) in the training period. The ensemble mean prediction (thick blue) averaged over \(N = 50\) ensemble members (thin green) is compared with the RMM observation (thick black). The black and blue circles represent the initial state of the RMM observation and the initialization of the model.

Fig. 13. Information measures and information criterion for prediction within the validation interval August 2005 to December 2008 utilizing nonlinear physics-constrained low-order stochastic...
model (2) and the two linear stochastic models (14) and (15) as a function of lead days. The
parameters in all the three models are calibrated via Strategy-2 (13) in the training period

Fig. 14. Predictions (thick blue) in time domain of RMM indices (thin black) with SSA(1-2) ini-
tialization of the observed variables (thick cyan) at a lead time of 25 days and the time-
averaged PDFs of prediction. The comparison of the prediction utilizing nonlinear physics-
constrained low-order stochastic model (2) (panel (a) and (c)) and the two linear stochastic
models (14) (panel (c) and (d)) and (15) (panel (e) and (f)) is shown. The parameters in all
the three models are calibrated via Strategy-2 (13) in the training period $\mathcal{T}_{\text{Training}}$. 
FIG. 1. The real-time Multivariate MJO (RMM) indices. Panels (a) and (c) show the two components of RMM indices (solid black) from January 2000 to May 2002 as well as the leading two (solid cyan) and four (dashed magenta) SSA reconstructed components (RCs). The time-averaged PDFs of the two RMM components over years 1980 to 2013 compared with their Gaussian fits are shown in panel (b) and (d). The PDFs are computed utilizing a histogram approach.
25–day lead prediction

(a) Prediction #1: RMSE = 1.1217 Corr = 0.6212 Relative entropy = 0.1014

(b) Time–averaged PDF

(c) Prediction #2: RMSE = 1.1582 Corr = 0.5829 Relative entropy = 4.5010

(d) Time–averaged PDF

Fig. 2. Comparison of two predictions (thick blue) of the RMM indices (thin black) utilizing the low-order stochastic model (2) with SSA(1-2) initialization (thick cyan) at 25-day lead. The prediction time interval is from August 2005 to December 2008. The parameter $\gamma$ in Prediction #1 and #2 is $\gamma = 0.5$ and $\gamma = 1.5$ respectively. Other parameters adopted in the two predictions are the same: $a = 4, \sigma_u = 0.4, \sigma_v = 1, d_u = 0.1, d_v = 0.2, \sigma_\omega = 0.5$ and $d_\omega = 1.2$. The relative entropy is defined in (8), which measures the lack of information in the statistics of the prediction relative to that of the truth.
FIG. 3. Data assimilation utilizing Strategy-1b for the hidden variables in the training period $\mathcal{S}_1^{\text{Training}}$ from January 2002 to December 2004. Panel (a) shows the SSA(1-2) initialization (thick blue) compared with the RMM1 index (thin black). Panel (b) and (c) show the posterior mean and posterior variance of the stochastic damping $v$ while panel (d) and (e) show those of the stochastic phase $\omega_u$. The dotted lines in panel (b) and (d) represent $v = d_u/\gamma = 0.2$ and $\omega_u = 0$ while those in panel (c) and (e) indicate the variance associated with the prior distribution without filtering.
Fig. 4. Prediction skill of nonlinear physics-constrained low-order stochastic model (2) utilizing different measures at a lead day of 25 days in the time interval August 2005 to December 2008 as a function of the parameter variations around the optimal values (the red circles) which are calibrated by Strategy-1b (12) in the training phase $\mathcal{P}_{\text{Training}}$. Row (a): Information criterion $M(\theta)$; Row (b)-(d): three information measures, i.e., exponential of the rescaled Shannon entropy $\exp(\tilde{S}(\theta))$, the relative entropy $S(\theta)$ and the mutual information $M(\theta)$; Row (e)-(f): two path-wise measures, i.e. the RMS error $\text{RMSE}(\theta)$ and the bivariate correlation $\text{Corr}(\theta)$. 

Information criterion $M$

Exponential of the rescaled Shannon entropy $\exp(\tilde{S})$

Relative entropy $S$

Mutual information $M$

RMS error $\text{RMSE}$

Bivariate correlation $\text{Corr}$
Fig. 5. Path-wise skill scores (panel (a)-(b)), information measures (panel (c)-(e)) and information criterion (panel (f)) of predicting the RMM indices utilizing the nonlinear physics-constrained low-order stochastic model (2) in the time interval August 2005 to December 2008 at lead times of 1 to 60 days with the optimal parameters calibrated by the three strategies (11), (12) and (13) in the training period \( T_{\text{Training}} \), respectively. SSA(1-4) initialization of the observed variables is adopted in the prediction stage when model parameters are calibrated by Strategy-1a (11) while SSA(1-2) initialization is utilized with calibration Strategy-1b (12) and Strategy-2 (13). The sub-panels in (d) and (f) show the zoomed-in curves for the relative entropy and the information criterion up to 30 days.
FIG. 6. Prediction of nonlinear physics-constrained low-order stochastic model (2) (thick blue) in the time domain of RMM indices (thin black) with SSA(1-4) initialization (thick cyan) at lead times of 5, 10 and 15 days (panel (a)-(c)) and the time-averaged PDFs of prediction (panel (d)-(f)). In each caption, $\cdot(S)$ and $\cdot(W)$ represent the skill scores of strong and weak MJO events defined by (1) respectively. Only RMM1 prediction is shown here. The parameters are calibrated by Strategy-1a (11) in the training period $T_{Training}$. x axis is time in calendar months within the validation interval of August 2005 to December 2008.
Fig. 7. Prediction of nonlinear physics-constrained low-order stochastic model (2) (thick blue) in the time domain of RMM indices (thin black) with SSA(1-2) initialization (thick cyan) at lead times of 25, 35 and 45 days (panel (a)-(c)) and the time-averaged PDFs of prediction (panel (d)-(f)). In each caption, $\cdot(S)$ and $\cdot(W)$ represent the skill scores of strong and weak MJO events defined by (1) respectively. Only RMM1 prediction is shown here. The parameters are calibrated by Strategy-2 (13) in the training period $\mathcal{T}$. x axis is time in calendar months within the validation interval of August 2005 to December 2008.
Fig. 8. The path-wise skill scores for prediction of the nonlinear physics-constrained low-order stochastic model (2) as a function of lead time within the validation interval August 2005 to December 2008. The parameters are calibrated by information-theoretic Strategy-2 (13) in the training period $\mathcal{T}_1^{\text{Training}}$. The skill scores for predicting the overall, the strong and the weak MJO events are shown in the solid, dashed and dash-dotted blue lines and the statistics between the prediction and SSA(1-2) reconstruction of the RMM indices is shown in the solid cyan lines. Panel (a) and (b) show the RMS error and bivariate correlation at lead times of 1 to 30 days while panel (c) and (d) show those at lead times of 1 to 60 days.
Fig. 9. Long-range forecasting of the nonlinear physics-constrained low-order stochastic model (2). The parameters are calibrated by Strategy-2 (13) in the training period $T_{\text{Training}}$. Different panel shows the prediction starting from the first day of each month in year 2007 and predicting with SSA(1-2) initialization up to 6 months. The label in x axis indicates the calendar month. In each panel, the solid black line is the RMM1 index; the solid thick blue line is the ensemble mean prediction which is averaged over $N = 50$ ensemble members shown in the thin green lines; the SSA(1-2) reconstruction is given by the thick red dashed-dotted line. The ensemble spread with 95% confidence interval is indicated by the gray area.
Fig. 10. Comparison of Multivariate ENSO Index (MEI) with the prediction of RMM index at a lead time of 25 days. The model parameters in predicting RMM index are calibrated via Strategy-2 (13) in the training period $\mathcal{F}_{\text{Training}}$ from January 1981 to December 1983 and SSA(1-2) initialization of the observed variables is adopted in prediction. X axis is time in year with the label representing the first day of every fifth year. In panel (a), the two dotted lines indicate $\pm 0.5$, the events outside which are regarded as strong El Niño or La Niña events. In panel (b) and (c), the yearly averaged skill scores of the $i$-th year are put in the middle of the $i$-th and the $(i+1)$-th years for the sake of the comparison with the monthly averaged MEI. The skill scores of overall, strong and weak MJO events are shown by blue solid, red dashed and green dash-dotted lines, respectively. The black dotted lines in panel (b) and (c) indicate one standard deviation of RMM indices and $\text{Corr} = 0.5$, respectively.
Fig. 11. Predictions (thick blue) of the RMM indices (thin black) at a lead time of 25 days in different years. The model parameters in predicting RMM index are calibrated via Strategy-2 (13) in the training period $\mathcal{T}^{Training}$ and SSA(1-2) initialization of the observed variables (thick cyan) is adopted in prediction. The label in x axis indicates the calendar month.
Fig. 12. Phase Diagrams of RMM1 and RMM2 prediction utilizing nonlinear physics-constrained low-order stochastic model (2) with SSA(1-2) initialization, where the parameters are calibrated by Strategy-2 (13) in the training period $\mathcal{T}_{\text{Training}}$. The ensemble mean prediction (thick blue) averaged over $N = 50$ ensemble members (thin green) is compared with the RMM observation (thick black). The black and blue circles represent the initial state of the RMM observation and the initialization of the model.
Fig. 13. Information measures and information criterion for prediction within the validation interval August 2005 to December 2008 utilizing nonlinear physics-constrained low-order stochastic model (2) and the two linear stochastic models (14) and (15) as a function of lead days. The parameters in all the three models are calibrated via Strategy-2 (13) in the training period $\mathcal{D}^{Training}$. 

F13.png

(a) Exponential of Shannon entropy $\exp(\tilde{S})$

(b) Relative entropy $P$

(c) Mutual information $M$

(d) Information criterion $M$
25–day lead prediction utilizing SSA(1−2) initialization with parameters calibrated via Strategy–2

(a) Nonlinear model \[ \text{RMSE}=1.1534 \text{ Corr}=0.6170 \text{ RMSE}(S)=1.2197 \text{ Corr}(S)=0.6882 \text{ RMSE}(W)=1.0336 \text{ Corr}(W)=0.3955 \]

(b) 2–D Linear model \[ \text{RMSE}=1.1080 \text{ Corr}=0.5997 \text{ RMSE}(S)=1.2356 \text{ Corr}(S)=0.6775 \text{ RMSE}(W)=0.8556 \text{ Corr}(W)=0.3721 \]

(c) 4–D Linear model \[ \text{RMSE}=1.1189 \text{ Corr}=0.5969 \text{ RMSE}(S)=1.2321 \text{ Corr}(S)=0.6740 \text{ RMSE}(W)=0.9004 \text{ Corr}(W)=0.3700 \]

(d) Time–averaged PDF utilizing nonlinear model

(e) Time–averaged PDF utilizing 2–D linear model

(f) Time–averaged PDF utilizing 2–D linear model

FIG. 14. Predictions (thick blue) in time domain of RMM indices (thin black) with SSA(1-2) initialization of the observed variables (thick cyan) at a lead time of 25 days and the time-averaged PDFs of prediction. The comparison of the prediction utilizing nonlinear physics-constrained low-order stochastic model (2) (panel (a) and (c)) and the two linear stochastic models (14) (panel (c) and (d)) and (15) (panel (e) and (f)) is shown. The parameters in all the three models are calibrated via Strategy-2 (13) in the training period $Training$. 

61