Research Articles

Nan Chen* and Andrew J Majda

Predicting the Cloud Patterns for the Boreal Summer Intraseasonal Oscillation Through a Low-Order Stochastic Model

Abstract: We assess the predictability limits of the large-scale cloud patterns in the boreal summer intraseasonal variability (BSISO), which are measured by the infrared brightness temperature, a proxy for convective activity. A recent developed nonlinear data analysis technique, nonlinear Laplacian spectrum analysis (NLSA), is applied to the brightness temperature data, defining two spatial modes with high intermittency associated with the BSISO time series. Then a recent developed data-driven physics-constrained low-order modeling strategy is applied to these time series. The result is a four dimensional system with two observed BSISO variables and two hidden variables involving correlated multiplicative noise through the nonlinear energy-conserving interaction. With the optimal parameters calibrated by information theory, the non-Gaussian fat tailed probability distribution functions (PDFs), the autocorrelations and the power spectrum of the model signals almost perfectly match those of the observed data. An ensemble prediction scheme incorporating an effective on-line data assimilation algorithm for determining the initial ensemble of the hidden variables shows the useful prediction skill in the non-El Niño years is at least 30 days and even reaches 55 days in those years with regular oscillations and the skillful prediction lasts for 18 days in the strong El Niño year (year 1998). Furthermore, the ensemble spread succeeds in indicating the forecast uncertainty. Although the reduced linear model with time-periodic stable-unstable damping is able to capture the non-Gaussian fat tailed PDFs, it is less skillful in forecasting the BSISO in the years with irregular oscillations. The failure of the ensemble spread to include the truth also indicates failure in quantification of the uncertainty. In addition, without the energy-conserving nonlinear interactions, the linear model is sensitive with parameter variations.

***Corresponding author: Nan Chen:** Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, USA. chennan@cims.nyu.edu

Andrew J Majda: Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, USA and Center for Prototype Climate Modeling, NYU Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE. jonjon@cims.nyu.edu

Finally, the twin experiment with nonlinear stochastic model has comparable skill as the observed data, suggesting the nonlinear stochastic model has significant skill for determining the predictability limits of the large-scale cloud patterns of the BSISO.

Keywords: Nonlinear Laplacian spectrum analysis (NLSA), data-driven physics-constrained low-order modeling strategy, information theory, data assimilation, predictability limits

PACS: 60G25, 60H10, 76E20, 94A15

DOI: ..., Received ...; accepted ...

Communicated by: ...
Für ...

1 Introduction

The boreal summer intraseasonal oscillation (BSISO) is one of the prominent modes of tropical intraseasonal variability. As a slow moving planetary scale envelope of convection propagating northward [1, 2], the BSISO distinguishes itself from the Madden-Julian oscillation (MJO), which prevails during boreal winter and propagates eastward. The BSISO is known to affect summer monsoon onset and active/break phases [3, 4, 5], and the seasonal means of summer monsoons [6, 7]. It also has fundamental impacts on the tropical precipitation, the frequency of tropical cyclones and extra-tropical climate variations [8, 9]. The studies of the prediction of the BSISO are mostly through operational dynamical models [1, 10, 11, 12, 13, 14, 15, 16] with only a few works focusing on low-order statistical models [17, 18, 19, 20]. One reason for the lack of utilizing low-order models for prediction is that different from the MJO (e.g., real-time multivariate MJO (RMM) index [21]) low-dimensional real-time monitoring and forecast verification metrics for the BSISO were not available until the recent time (since 2013) [22, 23]. The BSISO index in [23] is based on an extended empirical orthogonal function (EEOF) analysis on daily unfiltered rainfall anomalies. The derivation of the BSISO index in [22] mimics that for the RMM index and is based on the multivariate EOF analysis of daily anomalies of the zonal wind at 850 hPa and outgoing long-wave radiation (OLR), a proxy for convective activity. While the use these indices improves the quantification of skill of extended range forecasts of the BSISO, these linear

data analysis methods have limitations in capturing the highly nonlinear and intermittent characters of the BSISO [24].

Here we assess the predictability limits of the BSISO of the large-scale cloud patterns, which are measured by the infrared brightness temperature, a proxy for convective activity, alone. This is achieved in two steps. In the first step, a recent developed advanced nonlinear data analysis technique, nonlinear Laplacian spectrum analysis (NLSA), is applied to the brightness temperature data to define two spatial patterns associated with the BSISO. A key advantage of NLSA is that it requires no preprocessing such as bandpass filtering or seasonal partitioning of the input data, enabling simultaneous recovery of the dominant BSISO modes [25, 26, 27, 28, 29]. NLSA provides two time series representing the BSISO. These two BSISO time series are highly intermittent with non-Gaussian fat tailed probability distribution function (PDF), which differs from those derived by straightforward linear methods [22, 23]. The second step is to apply a recent systematic strategy for data-driven physics-constrained low-order stochastic modeling of time series to the two BSISO time series [30, 31]. The result is a four-dimensional nonlinear stochastic model with two observed state variables representing the two BSISO indices and two hidden variables. This low-order model involves correlated multiplicative noise through the energy-conserving nonlinear interaction between the observed and hidden variables as well as the additive noise. Note that this nonlinear low-order stochastic model has been shown to have significant skill for determining the predictability limits of the large-scale cloud patterns of the boreal winter MJO [32]. In addition, incorporating a new information-theoretic strategy in the training phase, a slight simplified version of (1) has been adopted to significantly improve the predictability of the real-time multivariate MJO indices [33].

The remainder of the paper is organized as follows. Section 2 describes the source of the brightness temperature dataset and the application of NLSA to the brightness temperature data to form the two BSISO time series. Section 3 involves the nonlinear low-order stochastic model as well as the information-theoretic calibration strategy and data assimilation algorithm for determining the initial ensemble of the hidden variables in the ensemble prediction scheme. This is followed by the prediction results, shown in Section 4. To understand the role of the nonlinearity in the low-order model, the linear model as well as its forecasting skill is studied in Section 5. Section 6 includes the perfect model twin experiment in checking the model error and the predictability limits. The paper is concluded in Section 7.

2 The Boreal Summer Intraseasonal Oscillation Through NLSA

Here tropical convection is analyzed in satellite observations of infrared brightness temperature (T_b) from the Cloud Archive User Service (CLAUS) Version 4.7 [34], which is sampled every 3 hours from 1 July 1983 to 30 June 2006 with a spatial resolution of 0.5^o. Brightness temperature is a measure of the earth's infrared emission in terms of the temperature of a hypothesized blackbody emitting the same amount of radiation at the same wavelength ($\sim$ 10-11 μm in CLAUS). It is a highly correlated variable with the total terrestrial longwave emission. In the tropics, positive (negative) T_b anomalies are associated with reduced (increased) cloudiness, hence suppressed (enhanced) deep convection.

The NLSA algorithm is applied to the full CLAUS dataset restricted to the tropical belt 15°N–15°S, with a lagged embedding window of 60 days. A variety of extended spatial cloud patterns emerge from the analysis, including annual, interannual, intraseasonal and diurnal modes, but the focus here is on the two spatial cloud patterns with time series depicted in Figure 1. It is clear that these time series are active from May to September of each year corresponding to boreal summer. Hereafter, we utilize the terminology, BSISO indices, for the two time series in Figure 1. The details of the NLSA method is available in [25, 26, 27, 28, 29]. Applying NLSA to the CLAUS dataset to obtain the BSISO indices are well documented in [24], where the spatial patterns associated with the BSISO indices in different phases are shown in Figure 9b in [24] and the spatiotemporal evolution of BSISO is shown in Movie 1h there. The figure and movie indicate the character of BSISO that a cluster of positive T_b develops in the central Indian Ocean and then moves northward towards the Bay of Bengal and India and branches off towards the western Pacific and the Monsoon Trough, bypassing the Maritime Continent from the north. Then following the dry phase of BSISO, a cluster of anomalously high convection develops in the central Indian Ocean, and propagates towards India and the western Pacific, completing the BSISO cycle. In addition, Figure 11 and Movie 2 in [24] compares the BSISO indices derived from NLSA with those from the EOF-singular spectrum analysis which is a linear data analysis method. The reconstructed BSISO and MJO patterns from the EOF method have coarser structures than their NLSA-based counterparts and appear to mix BSISO and MJO propagation. The highly non-Gaussian fat-tailed distributions in the BSISO indices from NLSA, which corresponding to intermittency and strong seasonality, also contrast with the nearly Gaussian distributions in the EOF related indices. We point out that the BSISO indices studied in this work is slightly cleaner than those in [24] by

reconstructing the eigenfunctions associated with NLSA through the shift map [35] while the corresponding spatiotemporal patterns have almost no changes. See Appendix A for details.

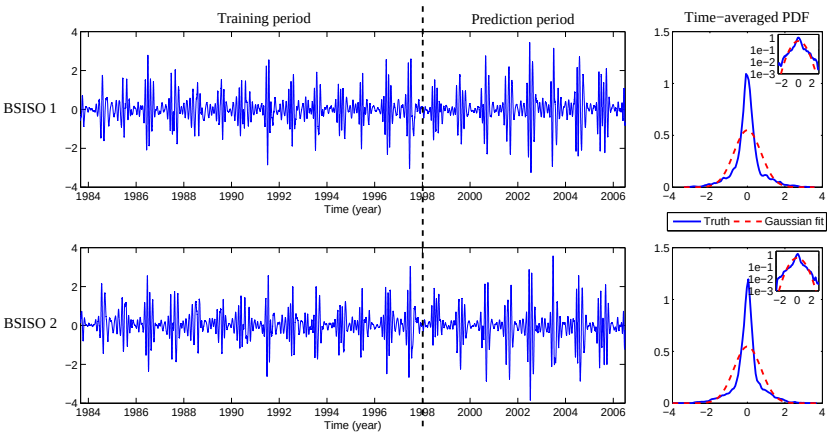

Fig. 1. (Left) BSISO indices from NLSA ranging from 3 September 1983 to 30 June 2006. The time series from September 1983 to December 1997 is utilized as training period to get the statistics and that from January 1998 to December 2005 represents the predicting period which will be predicted by the nonlinear low-order stochastic model (1). (Right) The associated time-averaged probability distribution function (PDF) of each index and the its Gaussian fit. Here the time-averaged PDF means we take all the points in each one-dimensional time series and compute the PDF based on these points. The small panel inside each subplot shows the PDF in the logarithm scale.

Note that the derivation of the BSISO indices from NLSA shown in Figure 1 is based on the dataset within the band 15°N–15°S. Yet, the BSISO activities typically propagate northward beyond 15°N [1]. Actually, Movie (1h) in [24] also shows that the cluster of convection moves outside the north boundary. To check the cluster propagation behavior in a farther northward region, we have looked at the projected spatiotemporal patterns in the extended band 30°N–15°S with the given BSISO indices in Figure 1. The northward propagation of the cluster extending up to 20°N-30°N in the spatiotemporal patterns provides the evidence that the time series in Figure 1 are suitable BSISO indices. The extension of the NLSA and of this work to a wider band, e.g., 30°N–15°S, for a better coverage of the Northward propagating signals is more involved and it is thus left for future investigations.

3 The Nonlinear Physics-Constrained Low-Order Stochastic Model

Denote by u_1 and u_2 the two components, BSISO 1 and BSISO 2, described in Figure 1. The PDFs for u_1 and u_2 are highly non-Gaussian with fat tails indicative of the temporal intermittency in the large-scale cloud patterns associated with the BSISO. To describe the variability of the time series u_1 and u_2, we propose the following family of low-order stochastic models:

$$\frac{du_1}{dt} = (-d_u\, u_1 + \gamma\,(v + v_f(t))\, u_1 - (a + \omega_u)\, u_2) + \sigma_u\, \dot{W}_{u_1}, \tag{1a}$$

$$\frac{du_2}{dt} = (-d_u\, u_2 + \gamma\,(v + v_f(t))\, u_2 + (a + \omega_u)\, u_1) + \sigma_u\, \dot{W}_{u_2}, \tag{1b}$$

$$\frac{dv}{dt} = (-d_v\, v - \gamma\,(u_1^2 + u_2^2)) + \sigma_v\, \dot{W}_v, \tag{1c}$$

$$\frac{d\omega_u}{dt} = (-d_\omega \omega_u) + \sigma_\omega\, \dot{W}_\omega, \tag{1d}$$

where

$$v_f(t) = f_0 + f_t \sin(\omega_f t + \phi). \tag{2}$$

Besides the two observed BSISO variables u_1 and u_2, the other two variables v and ω_u are hidden and unobserved, representing the stochastic damping and stochastic phase, respectively. In (1), $\dot{W}_{u_1}, \dot{W}_{u_2}, \dot{W}_v$ and $\dot{W}_\omega$ are independent white noise. The constant coefficients d_u, d_v, and d_ω 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; σ_u, σ_v, and σ_ω are noise amplitudes with dimension $t^{-1/2}$; the non-dimensional constant γ is the coefficient of the nonlinear interaction. The time periodic damping $v_f(t)$ in the equations in (1a) and (1b) is utilized to crudely model the active phase of the BSISO and the quiescent winter season in the seasonal cycle. The constant coefficients ω_f and ϕ in (2) are the frequency and phase of the damping, respectively. All of the model variables are real.

The hidden variables v, ω_u interact with the observed BSISO variables u_1, u_2 through energy-conserving nonlinear interactions following the systematic physics-constrained nonlinear regression strategies for time series developed recently [30, 31]. The energy conserving nonlinear interactions between u_1, u_2 and v, ω_u are seen in the following way. First, by dropping the linear and external forcing terms in (1), the remaining equations involving only the nonlinear parts

of (1) read,

$$\frac{du_1}{dt} = \gamma\, v\, u_1 - \omega_u\, u_2, \tag{3a}$$

$$\frac{du_2}{dt} = \gamma\, v\, u_2 + \omega_u\, u_1, \tag{3b}$$

$$\frac{dv}{dt} = -\gamma\,(u_1^2 + u_2^2), \tag{3c}$$

$$\frac{d\omega_u}{dt} = 0. \tag{3d}$$

To form the evolution equation of the energy from nonlinear interactions $\tilde{E} = (u_1^2 + u_2^2 + v^2 + \omega_u^2)/2$, we multiply the four equations in (3) by u_1, u_2, v and ω_u respectively and then sum them up. The resulting equation yields

$$\frac{d\tilde{E}}{dt} = 0. \tag{4}$$

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 (3a) and (3b) and those in (3c) multiplying by v as well as the trivial cancelation of skew-symmetric terms involving ω_u in (3a) and (3b).

The nonlinear low-order stochastic models in (1) are fundamentally different from those utilized earlier [36, 37] in predicting various kinds of climate phenomena 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 v and stochastic phase ω_u as well as their energy conserving nonlinear interaction with u_1 and u_2 also distinguish the models in (1) from the classic damped harmonic oscillator with only constant damping d_u and phase a. It is evident that a negative value of $\gamma(v + v_f)$ serves to strengthen the total damping of the oscillator. On the other hand, when $\gamma(v + v_f)$ becomes positive and overwhelms d_u, an exponential growth of u_1 and u_2 will occur, which corresponds to the intermittent instability.

The nonlinear low-order stochastic model (1) has been shown to have significant skill for determining the predictability limits of the large-scale cloud patterns of the boreal winter MJO [32]. In addition, incorporating a new information-theoretic strategy in the training phase, a simplified version of (1) without the time-period damping $v_f(t)$ has been adopted to improve the predictability of the real-time multivariate MJO indices [33]. Note that these models are a special case of the models described in [30, 31].

3.1 Calibration of the nonlinear low-order stochastic models

As shown in Figure 1, the full time series are divided into the training (Year 1983-1997) and prediction (Year 1998-2005) periods.

The parameters of the stochastic model in (1)–(2) are calibrated by systematically minimizing the information distance of the highly non-Gaussian PDFs of the stochastic model compared with that of the actual data [38, 39] and taking into consideration of the autocorrelations of the two BSISO variables u_1, u_2. Details are presented in Appendix B which also demonstrates the robustness of these optimal parameters to their variation. Table 1 records the optimal parameter values while Figure 2 displays the skill of the stochastic model with these parameters in recovering the statistics of the two BSISO indices. Panels (a) and (b) show that the stochastic model (1) succeeds in capturing the autocorrelations almost perfectly for a three-month duration and even the wiggles that appears with lags around one year. Panel (c) shows that the stochastic model captures the fat tailed highly non-Gaussian PDFs of the two BSISO indices due to intermittency. Panel (d) shows that the power spectrums of the two BSISO indices from the data and those from the stochastic model match very well.

	d_u	d_v	d_ω	σ_u	σ_v	σ_ω	γ	a	f_0	f_t	ω_f	ϕ
Nonlinear model (1)	0.9	0.9	0.5	0.3	0.8	1.0	0.3	−4.25	1.0	4.0	$2\pi/12$	−3.4
Linear model (5)	0.9			0.35			0.3	−4.25	0.0	4.5	$2\pi/12$	−3.4

Table 1. Parameters for the nonlinear low-order stochastic model (1) and linear stochastic model (5). The parameters $d_u, a, \gamma, f_0, f_t, \omega_f, d_v$ and d_ω have units m^{-1}; σ_u, σ_v and σ_ω have units $\mathrm{m}^{-1/2}$; ϕ is dimensionless. Here m stands for month.

3.2 Prediction algorithm and data assimilation for the hidden variables

The ensemble prediction algorithm is adopted to study the predictability of the nonlinear low-order stochastic model (1), which involves running the forecast model (1) forward in time given the initial values. The initial data of the two state variables $\mathbf{U} = (u_1, u_2)$ are obtained directly from the observations, i.e., BSISO 1 and BSISO 2 indices, and all the ensembles have the same initial values of $\mathbf{U} = (u_1, u_2)$. The more important and challenging issue is to determine

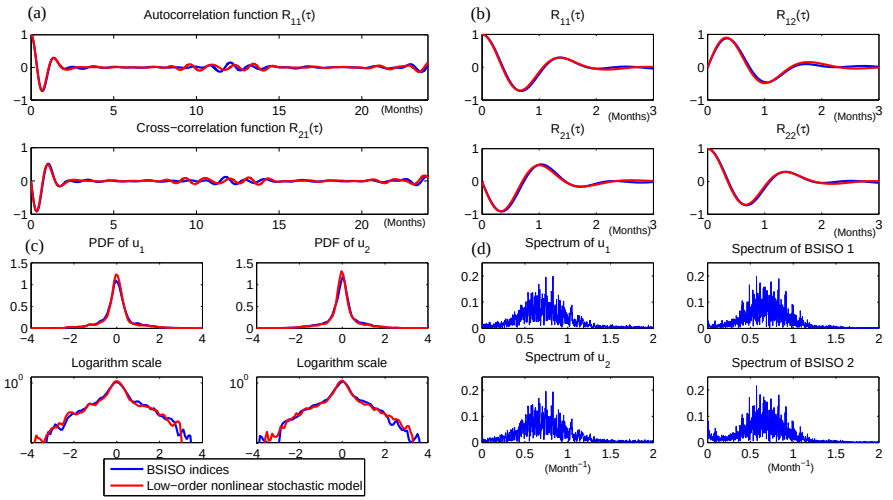

Fig. 2. Statistics of the nonlinear low-order stochastic model (1) with the optimal parameters in Table 1. (a) Long-term autocorrelation function $R_{11}(\tau)$ and cross-correlation function $R_{21}(\tau)$ from 0 to 24 months. (b) Short-term autocorrelation functions $R_{11}(\tau)$, $R_{22}(\tau)$ and cross-correlation functions $R_{21}(\tau)$, $R_{12}(\tau)$ from 0 to 3 months. (c) Equilibrium PDFs of the signal u_1, u_2 from stochastic model compared with that of the BSISO indices. (d) Spectrum of u_1, u_2 compared with that of the BSISO indices.

the initial ensemble of the two hidden variables $\Gamma = (v, \omega_u)$. To this end, an active data assimilation algorithm is incorporated into the ensemble forecasting scheme.

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 nonlinear low-order stochastic model (1). The equations in (1) are a conditional Gaussian system with respect to the observations $\mathbf{U} = (u_1, u_2)$, meaning that once u_1 and u_2 are given the time evolution of the distributions of $\Gamma = (v, \omega_u)$ is Gaussian. Such special feature of (1) allows the closed analytic equations for the conditional Gaussian distributions of the hidden parameters $\Gamma = (v, \omega_u)$ obtained from the posterior estimations in the Bayesian framework [40]. Appendix C 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 $t \in [t_0, t_1, \ldots, t_s]$ in the following way.

1. Starting from a "burn in" time t_- earlier than t_0 with arbitrary initial conditions for Γ, solve the associated analytic formula (12) 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.

2. The initial ensemble for prediction starting from the next time 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 (12) forward from time t_0 to t_1 with the initial value $p_0(\Gamma|\mathbf{u}(t_0))$.
3. Following the same procedure, the initial distributions of the hidden variables $\Gamma = (v, \omega_u)$ for prediction starting from each time t_i are obtained "on the fly" by running the analytic formula (12) forward from time t_{i-1} to t_i with the initial value $p_{i-1}(\Gamma|\mathbf{u}(t_{i-1}))$ when the new observations up to $\mathbf{u}(t_i)$ are available.

This is an effective and practical on-line data assimilation algorithm for the stochastic models in (1). Note that this algorithm is an improved version of the one in [32] for predicting the cloud patterns of the boreal winter MJO, in which the initial ensemble at the current time is the combination of all the ensembles corresponding to the analogous observations from the historic data and thus the algorithm there is more expensive. In the prediction below with (1), we use N ensemble members with $N = 50$.

4 Prediction Results

With the optimal parameters shown in Table 1 and the effective data assimilation algorithm for the ensemble initialization described in Section 3.2, we now study the prediction skill of the nonlinear low-order stochastic model (1) in the prediction period from year 1998 to year 2005. The BSISO 1 and BSISO 2 indices from year 1998 to 2005 are shown in panel (a) and (b) of Figure 3. The BSISO activities in year 1998 and 1999 are weaker compared with those from year 2000 to 2005. Note that year 1998 accompanies strong El Niño events.

Panel (c) and (d) in Figure 3 illustrate the root-mean-squared (RMS) error and bivariate correlation in prediction as a function of lead time in different years. The threshold level for skillful prediction is indicated by the horizontal dotted line in both panels. It is clear that the skillful prediction lasts for more than 30 days from year 1999 to 2005. Particularly, the prediction skill in year 2001 and 2002 with relatively regular intraseasonal oscillations reaches about 55 days. However, the useful prediction skill in year 1998, 18 days, is significantly lower than those in other years. Yet, it is still much improved compared with

the result from persistence prediction, which is at most 8 days for all the years as shown in the sub-panels in panel (c) and (d).

To understand the disparity in the prediction skill in different years, the predictions of BSISO 2 index regarding the ensemble mean at lead times of 15 and 25 days and 35 and 45 days are shown in Figure 4 and 5, respectively. The 15-day predictions in all years and 25-day predictions in year 1999 to 2005 are all very skillful (Corr > 0.6), capturing not only the patterns and amplitudes of the intraseasonal oscillations but some small-scale fluctuations as well. The prediction skill decreases as the lead time increases as expected. Yet, the predictions in year 2001 and 2002 at a lead time of 35 days are still quite accurate (Corr > 0.7) and even the 45-day predictions illustrate the similar patterns (Corr > 0.6) as the truth. The unskillful 25-day prediction (Corr $= 0.32803$) in year 1998 ascribes to the anomalous feature in BSISO 2 index, that is, the mean state of the index is significantly above zero (See also Figure 1 for a clearer visual). Yet, the mean state of the nonlinear oscillator (1), in consistency with the indices in other years, is around zero. This indicates an intrinsic model error in describing the BSISO events in year 1998 and in turn results in a low prediction skill. Actually, strong El Niño events occur at the same the period as the anomaly in BSISO 2 index. As shown in [41], many examples reveal that strong El Niño events are coincident with severely weakened BSISO events. The decrease in the amplitude of BSISO signal implies the increase of the noise to signal ratio, which prevents the skillful prediction and is consistent with the findings in [16]. The prediction skill of BSISO 1 is almost the same as that of BSISO 2, except that BSISO 2 involves a clear demonstration for the intrinsic model error for predicting the strong El Niñ year 1998. In addition, as indicated in [33], at the strong El Niño phases another intraseasonal oscillation – MJO – also shows the irregular behaviors and are hard to predict.

Figure 6 shows the predictions including the ensemble members starting from three different dates. The date, April 1, is a time at the transition between the quiescent phase and the active phase of the BSISO; June 1 is a starting date in the active mature phase while September 1 is a starting date in the decaying phase of BSISO activity. Although the ensemble mean forecasts starting from the April 1 have no long-term skill, which is consistent with the results in Figure 5, the ensemble spread automatically indicates such lack of skill. The envelope of the ensemble predictions includes the truth for all the years and is a good indicator of the active and quiescent phases of the BSISO. The ensemble mean predictions for the June 1 starting date are skillful in short and medium ranges and the ensemble spreads are the successful uncertainty indicator at long lead times for all the years. The forecasts starting from September 1 have both an accurate mean and a small ensemble spread for very long lead times. Note that

April 1, June 1 and September 1 are the three typical starting days that reflect different prediction behaviors. In general, starting from the quiescent phases, the ensemble spread accurately tells the active and quiescent phases of BSISO although the ensemble mean has no long-term prediction skill for the active phase. On the other hand, starting from the active phases, the short and medium range forecasting skill is obtained by the ensemble mean and the uncertainty at the long lead time is well indicated by the ensemble spread. Therefore, these results implies that the important target of predicting the onset and demise time of BSISO is well predicted by the ensemble spread.

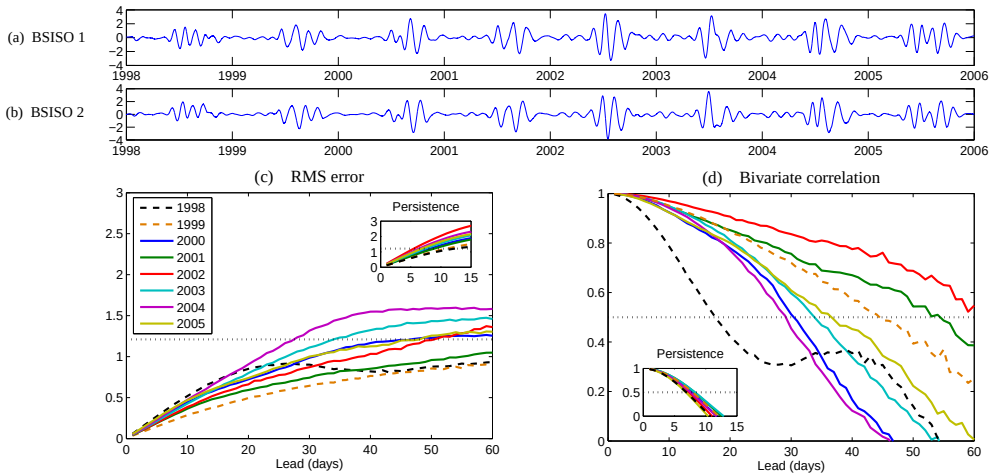

Fig. 3. (Panel (a) and (b)) BSISO indices in the prediction period. (Panel (c) and (d)) Skill scores with RMS error and bivariate correlation for ensemble mean prediction as a function of lead days in different years. The sub-panels inside panel (c) and (d) show the persistence prediction skill up to 15 lead days. The black dotted horizontal line in panel (c) shows the standard deviation of the BSISO indices, i.e., the climatological forecast skill, and that in panel (d) indicates Corr $= 0.5$, the typical threshold for skillful prediction.

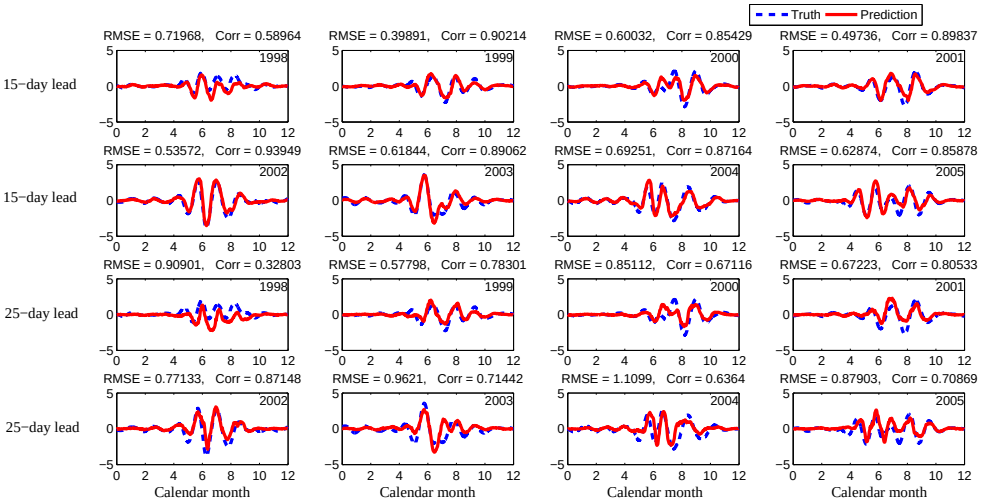

Fig. 4. Prediction of BSISO 2 index regarding the ensemble mean at lead times of 15 and 25 days in different years. The skill scores are computed from both BSISO 1 and 2 indices. The blue curve shows the truth and the red curve shows the ensemble average of the predicted signal. The number of ensemble utilized for prediction is 50.

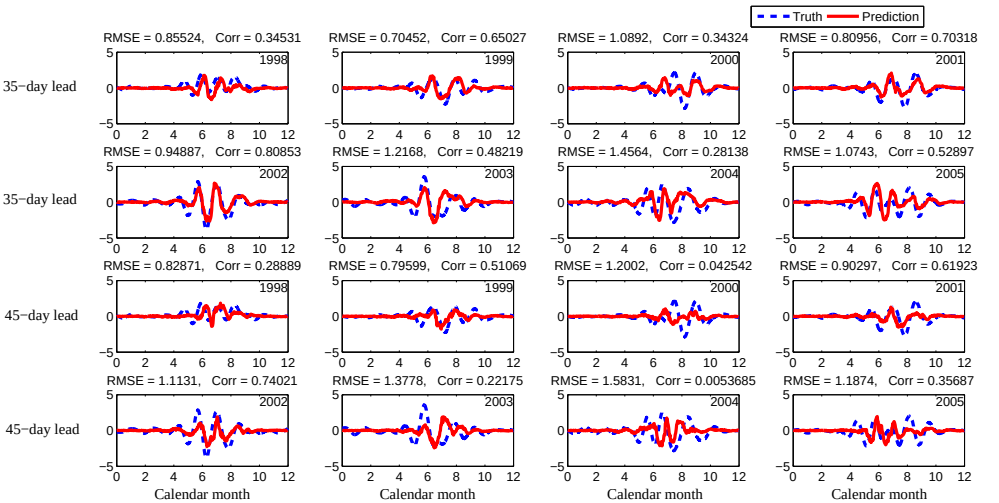

Fig. 5. Same as 4 but at lead times of 35 and 45 days.

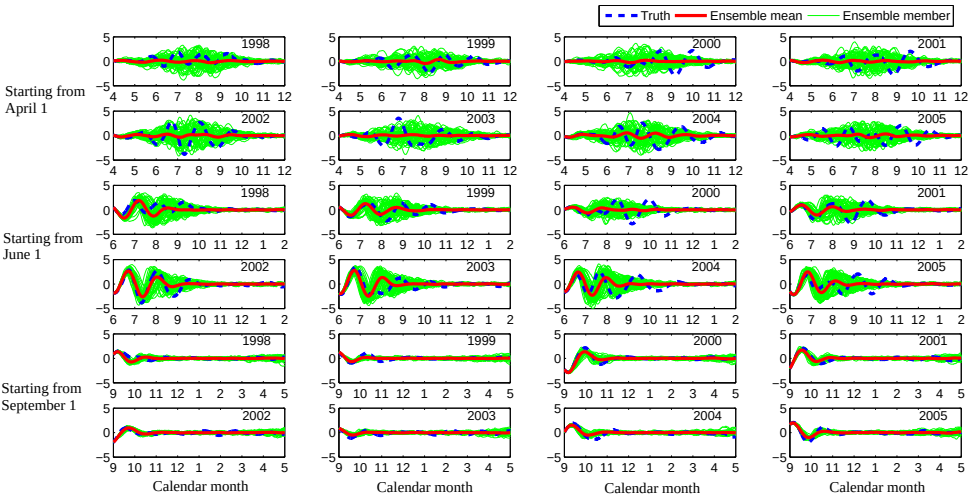

Fig. 6. (First and second rows) Prediction of BSISO 2 starting from April 1 for different years. Each panel shows the prediction skill of 8 months with the label in x axis indicating the calender month. (Third and fourth rows) Same as first and second rows but starting from June 1. (Fifth and sixth rows) Same as first and second rows but starting from September 1. The thick blue dashed curve is the BSISO 2 index. The thick red solid curve is the ensemble mean with 50 members, which are shown by the thin solid green curves.

5 The Role of the Nonlinearity in the Low-Order Models

To understand the role of the nonlinearity in the nonlinear low-order stochastic model (1), we study the prediction skill of the reduced linear version of (1) by removing the stochastic damping v and stochastic phase ω_u,

$$\frac{du_1}{dt} = (-d_u\, u_1 + \gamma\, v_f(t)\, u_1 - a\, u_2) + \sigma_u\, \dot{W}_{u_1}, \tag{5a}$$

$$\frac{du_2}{dt} = (-d_u\, u_2 + \gamma\, v_f(t)\, u_2 + a\, u_1) + \sigma_u\, \dot{W}_{u_2}. \tag{5b}$$

Albeit linear, the model in (5) is still able to generate intermittent instability due to the time-periodic damping $v_f(t) = f_0 + f_t \sin(\omega_f t + \phi)$ because the coefficient $-d_u + \gamma v_f(t)$ switches between negative (stable) and positive (unstable) intervals during a time period. Therefore, as shown in panel (c) of Figure 7, the time-averaged PDF associated with the linear model (5) succeeds in capturing the non-Gaussian fat tailed PDF of the observed data given the optimal parameters in Table 1. However, obvious discrepancies in the autocorrelation functions and the power spectrums are observed in panel (a), (b) and (d) of Figure 7. Although the oscillations in the autocorrelation functions of the linear model are of the same frequency as that of the observed data, the decaying rate of the autocorrelation functions of the linear model is linear and much slower than the nonlinear decay rate of those associated with the BSISO indices, indicating a barrier in describing the nature of the BSISO indices by the linear model. In addition, the power spectrums of the linear model with only the additive noise are more concentrated compared with that of the observed data, implies the insufficient with only additive noise. Therefore, nonlinearity and multiplicative noise are necessary in order to match the statistics of the BSISO indices.

Column (c) in Figure 8 shows the skill scores for ensemble mean prediction utilizing the linear model (5) with the optimal parameters. The linear model has a comparably high prediction skill as the nonlinear low-order stochastic model (1) in the short and medium range up to 20-25 days (column (a) in Figure 8). Above this range, the linear model becomes less skillful than the nonlinear model in years 1998, 2003, 2004 and 2005. On the other hand, the linear model shows the same skill with respect to the RMS error as the nonlinear model and even slightly more skillful prediction with respect to the bivariate correlation at lead times of 50-60 days in year 2001 and 2002 with regular oscillations. This possibly comes from the linear nature of the intraseasonal oscillation in these two years.

To further understand the difference in prediction utilizing the linear and nonlinear models with optimal parameters, the predictions in year 2005 are shown in column (a) and (c) of Figure 9. The forecasts at lead times of 25 and 35 days are shown in row I and II. The prediction at a lead time of 25 days in June and July utilizing the nonlinear model has almost the same pattern as the truth while that utilizing the linear model is slightly shifted forward in time and therefore results in a lower correlation. The superiority of the nonlinear model in this medium range forecasting is due to the phase correction from the stochastic part ω_u shown in the fourth row of Figure 13 in Appendix C. Similarly, the predicted amplitude around June 1 at lead times of both 25 and 35 days utilizing the nonlinear model being more close to that utilizing the linear model attributes to the initial correction of the overall damping by the stochastic part v, as is shown in the third row of Figure 13. The prediction including the ensemble members starting from June 1, the beginning of the active mature phase, is shown in row IV. Except slightly shifted forward in time in the prediction utilizing the linear model, the two models have the comparable good ensemble mean prediction skill up to 20 days. However, the medium- and long-range forecast up to 5 months utilizing the linear models shows not only a large bias in the ensemble mean forecasting but a small ensemble spread as well, which fails to include the truth and thus indicates a false uncertainty quantification in prediction. The misleading forecast of the linear model is associated with the long memory (slow decaying autocorrelation) and the single frequency dominated nature (concentrated power spectrum) of the system. Actually, the opposite patterns in the truth and prediction of the linear model during the period from the middle of July to the end of October are good indicators of the nonlinear nature of the oscillations in year 2005.

Another important thing to check is the model sensitivity. The model error dependence on the parameters is shown in Figure 12 in Appendix B. Clearly, the nonlinear low-order stochastic model (1) is more robust with respect to the parameter variations around the optimal values compared with the linear stochastic model (5). Note that when $\gamma v_f(t) > d_u$, a slight increase in $\gamma v_f(t)$ leads to an exponential increase in u_1 and u_2 in the linear model (5) and therefore the linear model is sensitive to the parameter variations. On the other hand, the energy-conserving nonlinear interaction plays a significant role in increasing the robustness in the nonlinear low-order stochastic model (1). Large values in u_1 and u_2 in the nonlinear low-order stochastic model (1) strongly reduce the stochastic damping v due to the nonlinear feedback $-\gamma(u_1^2 + u_2^2)$ in (1c), which then prevents the unbounded exponential increase in u_1 and u_2 and guarantees the robustness of the model.

To check the sensitivity of the prediction skill with respect to the parameter variations. We pick up the following random suboptimal parameters in each ensemble for both the nonlinear and linear models,

$$\begin{aligned} \sigma_u &= \sigma_u^* + \mathcal{U}(0, 0.2), & d_u &= d_u^* + \mathcal{U}(-0.4, 0), \\ \sigma_v &= \sigma_v^* + \mathcal{U}(-0.2, 0.2), & d_v &= d_v^* + \mathcal{U}(-0.2, 0.2), \\ \sigma_\omega &= \sigma_\omega^* + \mathcal{U}(-0.2, 0.2), & d_\omega &= d_\omega^* + \mathcal{U}(-0.2, 0.2), \\ f_t &= f_t^* + \mathcal{U}(0, 1), & \gamma &= \gamma^* + \mathcal{U}(0, 0.3), \end{aligned} \tag{6}$$

where the variables with asterisk are the optimal parameters and $\mathcal{U}(a, b)$ is the uniform distribution within the interval $[a, b]$. The forecasting skill scores are shown in column (b) and (d) of Figure 8. As expected, due to the energy-conserving nonlinear interactions, the nonlinear low-order stochastic model (1) with the random suboptimal parameters has the comparable prediction skill as that with the optimal parameters. On the other hand, the RMS error in prediction utilizing the linear model (5) with the suboptimal parameters increases dramatically as a function of lead time, contrasting the skillful prediction with the optimal parameters. The overestimation in amplitude utilizing the linear model (5) with the random suboptimal parameters at lead times of 25 and 35 days shown in Figure 9 is a good evidence. In addition, with the random suboptimal parameters, the ensemble spread for predicting starting from April 1 overestimates the uncertainty at long ranges and that starting from June 1 fails to include the truth, which together with the huge bias in the ensemble mean estimation indicates the useless prediction.

As a remark, the nonlinear model (1) without the nonlinear feedback term $-\gamma(u_1^2 + u_2^2)$ is the stochastic parameterized extended Kalman filter (SPEKF) model [42, 43], which is able to capture the intermittent nature of the BSISO indices and is skillful as a short-term forecast model. Yet, without the energy-conserving nonlinear interactions, the SPEKF model is not robust for long-range predictions and has sensitive dependence on parameters.

In summary, the linear model (5) with time-periodic damping $v_f(t)$ is able to capture the non-Gaussian fat tailed PDF of the truth and has comparably high prediction skill as the nonlinear low-order stochastic model (1) in the years with regular oscillations given the optimal parameters. The failure of the linear model in capturing the autocorrelation functions and power spectrums leads to a significant bias in medium- and long-range prediction with respect to ensemble mean in the years with nonlinear oscillations and the ensemble spread fails to include the truth in those years. Without the energy-conserving nonlinear interaction, the linear model is also sensitive to the parameter variations around

the optimal values. The error in prediction increases dramatically utilizing the linear model with the random suboptimal parameters.

Note that the bivariate correlation in prediction utilizing the linear model with the random suboptimal parameters remains almost the same as that with the optimal parameters due to the fact that the linear oscillator model is nevertheless able to capture the averaged oscillation frequency. Actually, this is a representative example to demonstrate that the bivariate correlation should not be overemphasized in measuring the prediction skill. In [33, 44], an information-theoretic framework is proposed to evaluate the forecast skill, which contains the information surrogates of the RMS error and the bivariate correlation as well as the information deficiency in comparing the amplitudes of the prediction and the truth. As shown in [33], two different predictions can have nearly the same bivariate correlation with the truth but have quite different information deficiency. Note that although only the RMS error and bivariate correlation are utilized here in measuring the prediction skill of the BSISO indices, the short- and medium-range ensemble mean prediction and the long-term ensemble spread of the nonlinear low-order stochastic model (1) have nearly the same amplitude as the truth, implying the small information deficiency in prediction.

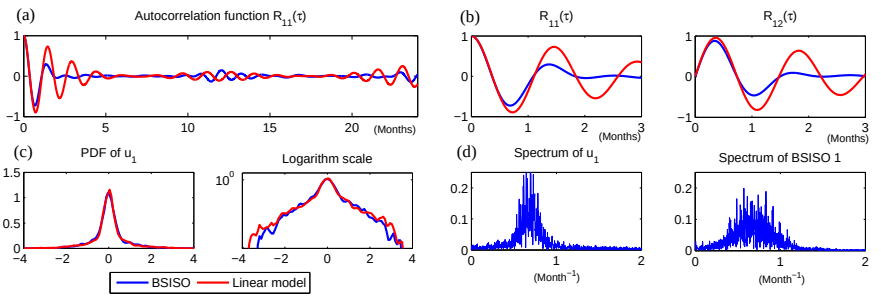

Fig. 7. Statistics of the linear stochastic model (5) with the optimal parameters in Table 1. (a) Long-term autocorrelation function $R_{11}(\tau)$ from 0 to 24 months. (b) Short-term autocorrelation function $R_{11}(\tau)$ and cross-correlation function $R_{12}(\tau)$ from 0 to 3 months. (c) Equilibrium PDFs of the signal u_1 from stochastic model compared with that of the BSISO 1 index. (d) Spectrum of u_1 compared with that of the BSISO 1 index.

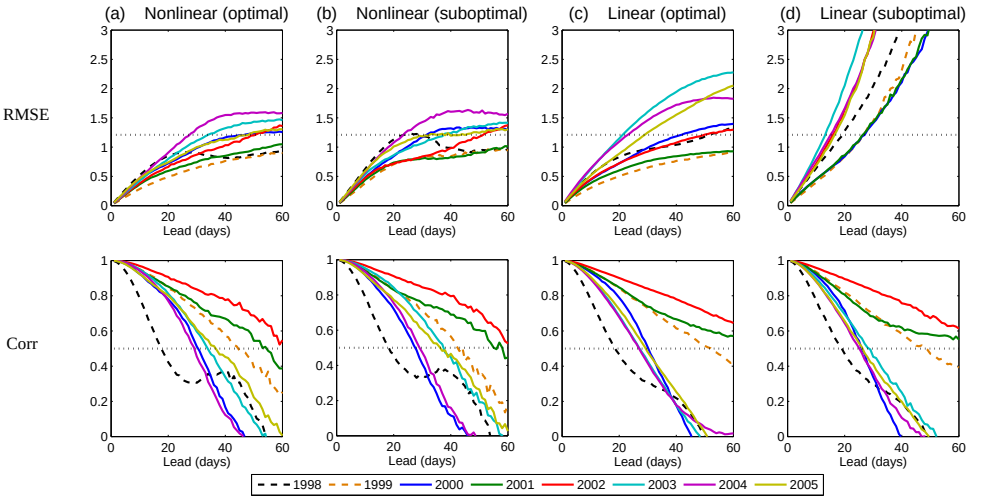

Fig. 8. Comparison of the skill scores with RMS error and bivariate correlation for prediction utilizing different models and parameters as a function of lead days in different years. Column (a): Nonlinear stochastic model (1) with optimal parameters; Column (b): Nonlinear stochastic model (1) with suboptimal parameters; Column (c):Linear stochastic model (5) with optimal parameters; Column (d): Linear stochastic model (1) with suboptimal parameters.

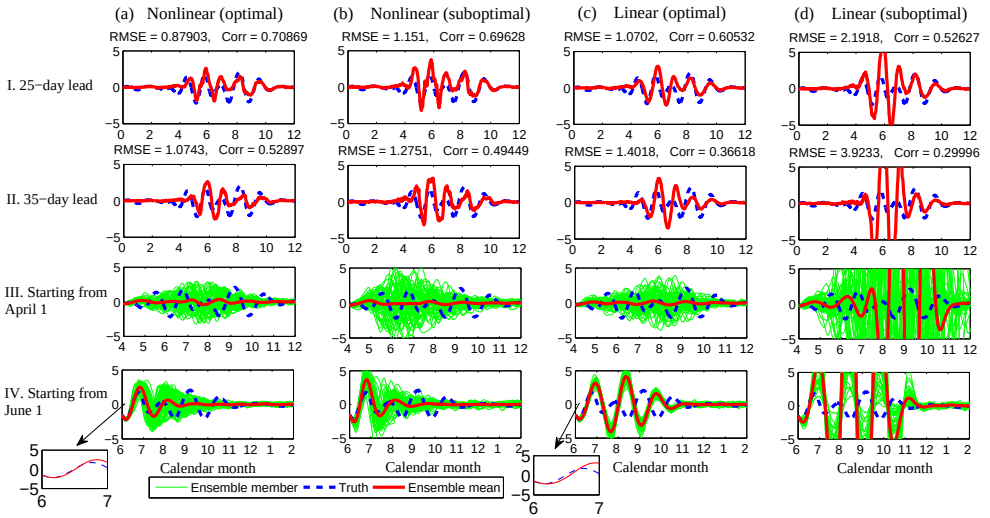

Fig. 9. Comparison of the prediction skill in year 2005 utilizing different models and parameters at lead times of 25 (Row I) and 35 (Row II) days and the long-term prediction up to 8 months starting from April 1 (Row III) and June 1 (Row IV). The skill scores are computed from BSISO 1 and 2 indices but only the curves of BSISO 2 index are shown. Column (a): Nonlinear stochastic model (1) with optimal parameters; Column (b): Nonlinear stochastic model (1) with suboptimal parameters; Column (c):Linear stochastic model (5) with optimal parameters; Column (d): Linear stochastic model (1) with suboptimal parameters.

6 Twin Experiment

To explore the model error and the predictability limits of the nonlinear low-order stochastic model (1) in predicting the BSISO indices, we include the results from the perfect model twin experiment. In the twin experiment, the truth signal is generated from the nonlinear low-order stochastic model (1) and therefore the data assimilation algorithm in Section 3.2 is based on a perfect filter. The length of the two signals representing u_1 and u_2 in the twin experiment is the same as the two BSISO indices in Figure 1 and as in predicting the BSISO indices we predict the signals in the last eight years of these two time series. See panel (a) and (b) in Figure 10.

The prediction skill scores as a function of lead time are shown in panel (c) and (d) in Figure 10 and the ensemble mean predictions at lead times of 25 and 35 days are shown in Figure 11. Actually, the internal prediction skill in this twin experiment indicates the predictability limit of the model. Comparing panel (c) and (d) in Figure 10 and panel (c) and (d) in Figure 3, it is clear that the BSISO prediction skill is comparable to this internal prediction skill, suggesting that the nonlinear low-order stochastic model (1) has a significant skill for determining the predicability limits of the large scale cloud patterns of the BSISO.

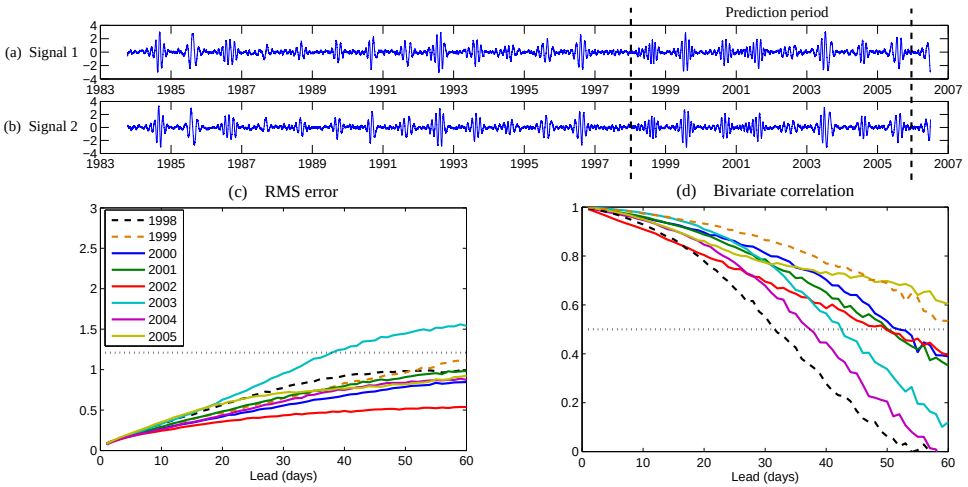

Fig. 10. Twin experiment. (Panel (a) and (b)) The two components of the signal generated from the nonlinear low-order stochastic model (1), where the eight year period from 1998 2006 is utilized as the prediction period. (Panel (c) and (d)) Skill scores with RMS error and bivariate correlation for prediction as a function of lead days in different years.

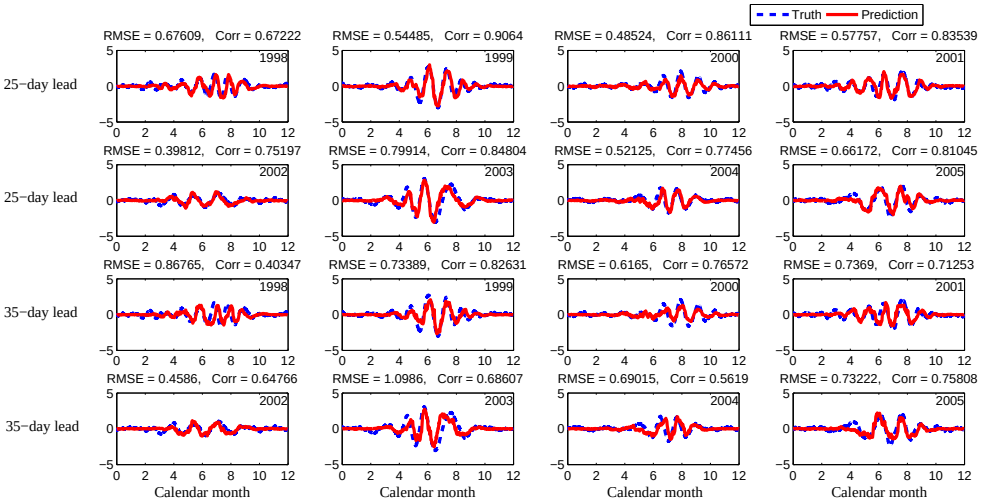

Fig. 11. Twin experiment. Prediction of the twin experiment signal 2 (panel (b) in Figure 10) generated from the nonlinear low-order stochastic model (1) regarding the ensemble mean at lead times of 25 and 35 days in different years. The blue curve shows the truth in panel (b) of Figure 10 and the red curve shows the ensemble average of the predicted signal. The number of ensemble utilized for prediction is 50.

7 Conclusions and Discussions

A recent developed nonlinear data analysis technique NLSA has been applied to the CLAUS data to define two spatial patterns associated with the BSISO without detrending or spatial-temporal filtering. This provides two BSISO indices with strong intermittency and non-Gaussian fat-tailed PDF (Figure 1). Then a recent systematic strategy for data driven physics-constrained low-order stochastic modeling is applied to the BSISO indices. The result is a four dimensional nonlinear stochastic model (1) with two state variables denoting the observed BSISO indices and two hidden variables representing stochastic damping and stochastic phase. The model contains correlated multiplicative noise through the energy-conserving nonlinear interactions between the observed and hidden variables as well as the additive noise.

The parameters in the nonlinear low-order stochastic model (1) are calibrated in minimizing the model error of the time-averaged PDF compared with that of the truth in the information framework. The model with the optimal parameters succeeds in recovery the highly intermittent non-Gaussian PDF; the autocorrelations and the power spectrums of the model signals almost perfectly match those of the observed BSISO indices (Figure 2). The special structure of model (1) allows an effective data assimilation algorithm for determining the initial ensemble of the hidden variables in the ensemble forecasting scheme. This on-line prediction algorithm shows that the ensemble mean forecasting skill of the BSISO in the non-El Niño year is at least 30 days and even reaches 55 days in the years with regular intraseasonal oscillation. In the strong El Niño year (year 1998) the useful prediction is around 18 days due to both the increase of noise to signal ratio and the intrinsic model error (Figure 3–5). Furthermore, the ensemble spread is a good indicator of the forecasting uncertainty at long range (Figure 6).

The twin experiment (Figure 10 and 11) shows the skill of forecasts in the perfect modeling setting is comparable with that of predicting the BSISO indices, implying that the nonlinear low-order stochastic model (1) has a significant skill for determining the predicability limits of the large scale cloud patterns of the BSISO.

To check the role of the nonlinearity in the low-order models, the prediction skill of the linear model (5) is studied. With time-periodic damping $v_f(t)$, the linear model is able to capture the non-Gaussian fat tailed PDF of the truth and has comparably high prediction skill as the nonlinear low-order stochastic model (1) in the years with regular oscillations given the optimal parameters. The failure of the linear model in capturing the autocorrelation functions and

power spectrums (Figure 7) leads to a significant bias in medium- and long-range prediction with respect to ensemble mean in the years with nonlinear oscillations and the ensemble spread fails to include the truth in those years (Figure 8 and 9). In addition, the linear model is sensitive to the parameter variations around the optimal values (Figure 12). The error in prediction increases dramatically utilizing the linear model with random suboptimal parameters (Figure 8 and 9). On the other hand, the SPEKF model, which is a nonlinear model but has no energy-conserving nonlinear interaction, is also quite sensitive with respect to the parameter variations for long-range forecasting.

Although in this paper, we focus only on the skill of predicting the two BSISO time series, it is straightforward to translate to the prediction of the location and evolution of BSISO convection itself. Recall that the original spatiotemporal patterns of the MJO cloud clusters are illustrated in the video in [24]. The predicted spatial patterns are a rank-2 reconstruction constructed from the predicted temporal patterns (u_1 and u_2) multiplied by the dataset projected onto the original temporal patterns. Because u_1 and u_2 evolve in near-quadrature (as do the original temporal patterns), we do not expect major qualitative differences between the structure of the predicted cloud clusters in the reconstruction and the original clusters when the time series can be predicted reasonably well.

Most of the current studies of BSISO prediction is up to 30 days. In [45], a regression scheme is designed to study the forecasts of central India precipitation. The prediction skill lasts for 30 days with respect to the pattern correlation. Yet, their prediction underestimates the amplitudes and fails to predict some monsoon onsets. Similar results are found in [46] that the skill utilizing the coupled ocean-atmosphere forecast models for monsoon prediction is up to 30 days as well. Thus, our nonlinear low-order stochastic model combined with NLSA data analysis tool are potentially able to extend the prediction limit of BSISO and boreal summer monsoon.

Note that what we have predicted utilizing the nonlinear low-order stochastic model (1) is the anomalies of the BSISO. Actually, when we apply NLSA to the CLAUS data, besides the BSISO modes, we also obtain those large-scale modes such as the annual and seasonal modes, which reflect the background of rainfall. These modes are quite regular and easy to predict. Therefore, to recover the total rainfall, we simply need to combine the predicted BSISO anomaly with the background.

We have also looked at the prediction skill utilizing the nonlinear low-order stochastic model (1) in predicting the boreal winter MJO derived from NLSA with shift map and with the improved prediction scheme incorporating the effective data assimilation algorithm in Section 3.2. The boreal winter MJO time

series also have non-Gaussian fat tailed time-averaged PDF and the oscillation frequency is slightly lower than that of the BSISO. Yet, unlike BSISO with moderate or strong amplitudes in almost all the years, the amplitudes of the boreal winter MJO range from weak to strong in different years. The forecasting skill for the BSISO and the boreal winter MJO is comparable in the moderate and strong years while the small signal to noise ratio in the weak MJO years, including year 1998, deteriorates the prediction skill. The results are similar to those reported in [32]

We point out that the information-theoretic calibration procedure utilized in this work is widely adopted as training strategy for improving the predictive skill in many different issues. Imperfect predictions via Multi Model Ensemble forecasts are improved with the information-theoretic framework [47]. The prediction skill of imperfect large-dimensional turbulent models are enhanced through statistical response and information theory [48]. The forecasting skill of the RMM index is also greatly enhanced by combining three different information measures compared with adopting only path-wise measures [33, 44].

We also note that the conditional Gaussian models as in (11) and the effective data assimilation algorithm in Section 3.2 are innovative tools for studying the hidden processes from the observations in the turbulent flows. One example is the understanding of the practical information barrier in recovering the fluid flows with noisy Lagrangian tracers [49, 50], which contains the application of the information theory as well.

Acknowledgment: 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. A.J.M also gratefully acknowledges the financial support given by the Earch System Science Organization, Ministry of Earth Sciences, Government of India (Grant no./Project no MM/SERP/CNRS/2013/INT-10/002) to conduct this research under Monsoon Mission. The authors thank Dimitrios Giannakis and Eniko Szekely for discussion on NLSA applied to CLAUS data.

8 Appendix

8.1 A. Reconstruction of the eigenfunctions from the shift map in NLSA

Following Section 2 in [24], the Laplace-Beltrami operator L is given by formula (3) there. In formula (4) there, the eigenfunction ϕ_i with $i \in \{0, 1, 2, \ldots\}$ is solved via

$$L\phi_i = \lambda_i \phi_i, \tag{7}$$

where λ_i is the associated eigenvalue. Now we utilize the shift map S_τ [35], where $S_\tau f(x(t_i)) = f(x(t_{i+\tau}))$, to formulate a new operator A,

$$A_{ij} = \langle \phi_i, S_\tau \phi_j \rangle.$$

Then we solve the following eigenvalue problem

$$Av_i = \tilde{\lambda}_i v_i, \tag{8}$$

where the eigenvectors v_i are utilized as the time series of different modes. Similar to the formula (5) in [24] but replacing ϕ_i by v_i that the spatiotemporal patterns are recovered through the operation

$$\tilde{X}_i = XDv_i v_i^T, \tag{9}$$

where X is the lagged embedding of the raw data and D is the weight matrix.

To derive the BSISO indices in Figure 1, τ is set to be 1.

8.2 B. Calibration of the nonlinear and linear low-order stochastic models and sensitivity analysis

The optimal parameters in the nonlinear low-order stochastic model (1) and the linear model (5) are calibrated by systematically minimizing the information distance, i.e., the model error, of the time-averaged PDF of the model π^M compared with that of the BSISO index π [38, 39, 51, 52, 53],

$$\mathcal{P}(\pi, \pi^M) = \int \pi \log\left(\frac{\pi}{\pi^M}\right). \tag{10}$$

The model error dependence on the variation of different parameters is shown in Figure 12, which indicates that the nonlinear low-order stochastic model (1)

is robust with respect to the parameters around their optimal values. The huge model error with the underestimation of σ_u, f_t and γ and the overestimation of d_u is due to the failure of capturing the intermittency. Note that the model error has only a weak dependence on the background phase a since the contribution of the oscillation in the signal has been averaged out in the time-averaged PDF. However, the parameter a is crucial in describing the frequency of intraseasonal oscillation, and it is calibrated by matching the autocorrelation functions associated with the model and the truth. The other parameters d_v, σ_v, d_ω and σ_ω in describing the stochastic processes affect not only on the model error but more on the autocorrelations and power spectrums as well. A large discrepancy appears in the statistics if these parameters are outside the optimal range. The parameter f_0 is not an independent parameter given d_u and γ and therefore we fix its value. The frequency ω_f in the time-periodic damping $v_f(t)$ is prescribed to be $2\pi/12$ such that one time unit of the model corresponds to one month in reality. The phase ϕ in $v_f(t)$ is tuned to make the strong intermittency occur in the boreal summer in accordance with the BSISO indices. Note that none of the parameters is redundant in the nonlinear stochastic model (1). In fact, without the hidden variables v and ω_u, even if the time-period damping $v_f(t)$ is able to crudely describe the active phase of BSISO in the reduced linear model, a distinguished disparity is observed in the model statistics compared with the truth, indicating the intrinsic barrier [39, 52]. In addition, as seen in Figure 12, the linear model (5) is more sensitive with respect to the parameter variations around the optimal values.

8.3 C. Mathematical details of effective data assimilation and prediction algorithm

Recall the nonlinear low-order stochastic model (1). Denote by $\mathbf{U} = (u_1, u_2)^T$ and $\mathbf{\Gamma} = (v, \omega_u)^T$. The abstract form of the low-order stochastic model (1) is given as follows:

$$d\mathbf{U}_t = [\mathbf{A}_0(t, \mathbf{U}) + \mathbf{A}_1(t, \mathbf{U})\mathbf{\Gamma}_t]dt + \mathbf{\Sigma}_U(t, \mathbf{U})d\mathbf{W}_U(t), \tag{11a}$$

$$d\mathbf{\Gamma}_t = [\mathbf{a}_0(t, \mathbf{U}) + \mathbf{a}_1(t, \mathbf{U})\mathbf{\Gamma}_t]dt + \mathbf{\Sigma}_\Gamma(t, \mathbf{U})d\mathbf{W}_\Gamma(t), \tag{11b}$$

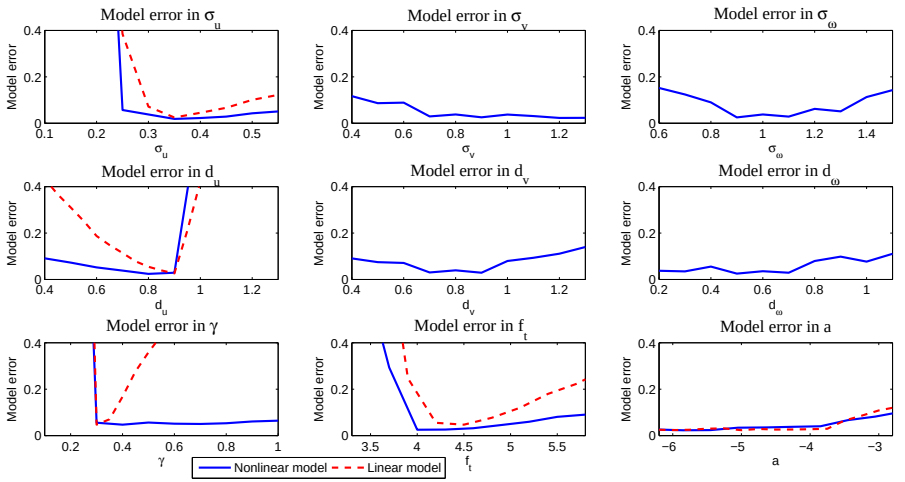

Fig. 12. Sensitivity analysis of the parameters in the nonlinear low-order stochastic model(1) (solid blue curves) and the linear stochastic model (5) (dashed red curves).

where

$$\mathbf{A}_0 = \begin{pmatrix} -d_u\, u_1 + \gamma\, v_f(t)\, u_1 - a\, u_2 \\ -d_u\, u_2 + \gamma\, v_f(t)\, u_2 + a\, u_1 \end{pmatrix}, \qquad \mathbf{A}_1 = \begin{pmatrix} \gamma\, u_1 & -u_2 \\ \gamma\, u_2 & u_1 \end{pmatrix},$$

$$\mathbf{a}_0 = \begin{pmatrix} -\gamma\,(u_1^2 + u_2^2) \\ 0 \end{pmatrix}, \qquad \mathbf{a}_1 = \begin{pmatrix} -d_v & \\ & -d_\omega \end{pmatrix},$$

$$\mathbf{\Sigma}_U = \begin{pmatrix} \sigma_u & \\ & \sigma_u \end{pmatrix}, \qquad \mathbf{\Sigma}_\Gamma = \begin{pmatrix} \sigma_v & \\ & \sigma_\omega \end{pmatrix}.$$

The model (11) is a conditional Gaussian system conditioned on the observations $\mathbf{U}$, meaning that once the observations $\mathbf{U}$ are given the dynamics of $\mathbf{\Gamma}$ in (11) becomes a Gaussian system. The special structure of system (11) allows the closed analytic formulae for the evolution of the conditional Gaussian distributions of the hidden parameters v and ω_u [40] obtained in the Bayesian framework:

$$\begin{aligned} d\mu_t =& [\mathbf{a}_0(t,\mathbf{U}) + \mathbf{a}_1(t,\mathbf{U})\mu_t]dt + (R_t\mathbf{A}_1^*(t,\mathbf{U}))(\mathbf{\Sigma}_U\mathbf{\Sigma}_U^*)^{-1}(t,\mathbf{U})\times \\ & [d\mathbf{U}_t - (\mathbf{A}_0(t,\mathbf{U}) + \mathbf{A}_1(t,\mathbf{U})\mu_t)dt], \\ dR_t =& \{\mathbf{a}_1(t,\mathbf{U})R_t + R_t\mathbf{a}_1^*(t,\mathbf{U}) + (\mathbf{\Sigma}_\Gamma\mathbf{\Sigma}_\Gamma^*)(t,\mathbf{U}) \\ & -(R_t\mathbf{A}_1^*(t,\mathbf{U}))(\mathbf{\Sigma}_U\mathbf{\Sigma}_U^*)^{-1}(t,\mathbf{U})(R_t\mathbf{A}_1^*(t,\mathbf{U}))^*\}\, dt, \end{aligned} \tag{12}$$

where μ_t and R_t are the posterior mean and posterior covariance of the conditional distributions, respectively. The asterisk represents the complex conjugate.

As a remark, the formulae (12) are optimal if and only if the signal is generated from system (11). Since our observed signal, i.e., the BSISO indices, are not from the nonlinear low-order stochastic model (11), the evolutions of the conditional Gaussian distributions (12) are suboptimal.

In Figure 13 we show the posterior mean and variance of stochastic damping v and stochastic phase ω_u in (1) as a function of time compared with the observations of the BSISO indices from January 1994 to December 2005. Note that the corrections of the stochastic variables v and ω_u from the data assimilation algorithm are significant and important in the intermittent phase. In addition, the posterior covariance at the intermittent phase is smaller than that at the quiescent phase, indicating the small uncertainty in the recovered stochastic variables.

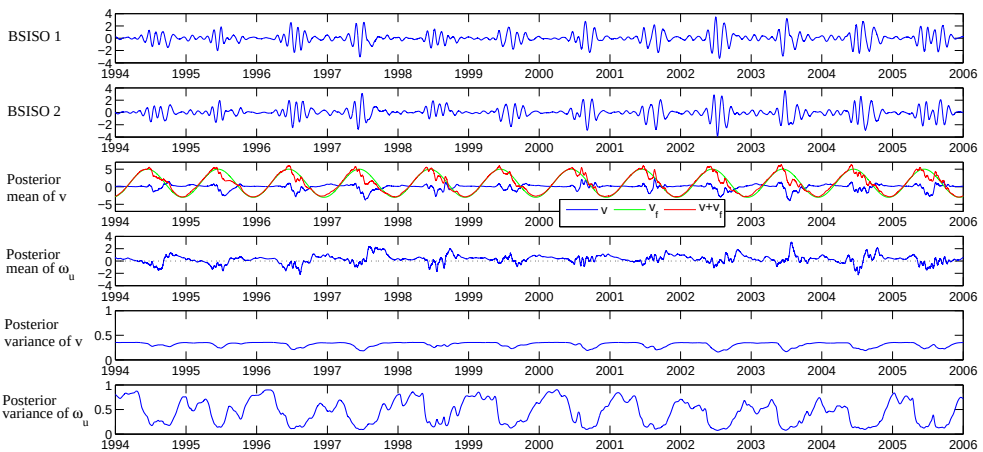

Fig. 13. Recovery of posterior mean and variance of stochastic damping v and stochastic phase ω_u in (1) as a function of time compared with the observations of the BSISO indices from January 1994 to December 2005. The cross-covariance of v and ω_u is negligible of order $O(10^{-18})$ and is not shown here.

References

[1] William KM Lau and Duane E Waliser. *Intraseasonal variability in the atmosphere-ocean climate system*. Springer, 2012.

[2] Peter J Webster, Vo Oo Magana, TN Palmer, J Shukla, RA Tomas, M u Yanai, and T Yasunari. Monsoons: Processes, predictability, and the prospects for prediction. *Journal of Geophysical Research: Oceans (1978–2012)*, 103(C7):14451–14510, 1998.

[3] Tiruvalam Natarajan Krishnamurti and D Subrahmanyam. The 30-50 day mode at 850 mb during MONEX. *Journal of the Atmospheric Sciences*, 39(9):2088–2095, 1982.

[4] In-Sik Kang, Chang-Hoi Ho, Young-Kwon Lim, and KM Lau. Principal modes of climatological seasonal and intraseasonal variations of the Asian summer monsoon. *Monthly weather review*, 127(3):322–340, 1999.

[5] Qinghua Ding and Bin Wang. Predicting extreme phases of the Indian summer monsoon. *Journal of Climate*, 22(2):346–363, 2009.

[6] V Krishnamurthy and J Shukla. Intraseasonal and seasonally persisting patterns of Indian monsoon rainfall. *Journal of climate*, 20(1):3–20, 2007.

[7] V Krishnamurthy and J Shukla. Seasonal persistence and propagation of intraseasonal patterns over the Indian monsoon region. *Climate Dynamics*, 30(4):353–369, 2008.

[8] Bin Wang, June-Yi Lee, In-Sik Kang, J Shukla, C-K Park, A Kumar, J Schemm, S Cocke, J-S Kug, J-J Luo, et al. Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dynamics*, 33(1):93–117, 2009.

[9] June-Yi Lee, Bin Wang, I-S Kang, J Shukla, A Kumar, J-S Kug, JKE Schemm, J-J Luo, T Yamagata, X Fu, et al. How are seasonal prediction skills related to models' performance on mean state and annual cycle? *Climate dynamics*, 35(2-3):267–283, 2010.

[10] In-Sik Kang, June-Yi Lee, and Chung-Kyu Park. Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. *Journal of climate*, 17(4):834–844, 2004.

[11] Bin Wang, Qinghua Ding, Xiouhua Fu, In-Sik Kang, Kyung Jin, J Shukla, and Francisco Doblas-Reyes. Fundamental challenge in simulation and prediction of summer monsoon rainfall. *Geophysical Research Letters*, 32(15), 2005.

[12] Hye-Mi Kim and In-Sik Kang. The impact of ocean–atmosphere coupling on the predictability of boreal summer intraseasonal oscillation. *Climate dynamics*, 31(7-8):859–870, 2008.

[13] DR Pattanaik and Arun Kumar. Prediction of summer monsoon rainfall over India using the NCEP climate forecast system. *Climate dynamics*, 34(4):557–572, 2010.

[14] Nachiketa Acharya, Sarat C Kar, UC Mohanty, Makarand A Kulkarni, and SK Dash. Performance of GCMs for seasonal prediction over india – a case study for 2009 monsoon. *Theoretical and applied climatology*, 105(3-4):505–520, 2011.

[15] Archana Nair, UC Mohanty, Andrew W Robertson, TC Panda, Jing-Jia Luo, and Toshio Yamagata. An analytical study of hindcasts from general circulation models for Indian summer monsoon rainfall. *Meteorological Applications*, 21(3):695–707, 2014.

[16] Sun-Seon Lee, Bin Wang, Duane E Waliser, Joseph Mani Neena, and June-Yi Lee. Predictability and prediction skill of the boreal summer intraseasonal oscillation in the intraseasonal variability hindcast experiment. *Climate Dynamics*, pages 1–13, 2015.

[17] Timothy DelSole and J Shukla. Linear prediction of indian monsoon rainfall. *Journal of Climate*, 15(24):3645–3658, 2002.

[18] Charles Jones, Leila MV Carvalho, R Wayne Higgins, Duane E Waliser, and JK E Schemm. A statistical forecast model of tropical intraseasonal convective anomalies. *Journal of climate*, 17(11):2078–2095, 2004.

[19] M Rajeevan, DS Pai, R Anil Kumar, and B Lal. New statistical models for long-range forecasting of southwest monsoon rainfall over india. *Climate Dynamics*, 28(7-8):813–828, 2007.

[20] Sun-Seon Lee and Bin Wang. Regional boreal summer intraseasonal oscillation over indian ocean and western pacific: comparison and predictability study. *Climate Dynamics*, pages 1–17, 2015.

[21] Matthew C Wheeler and Harry H Hendon. An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, 132(8):1917–1932, 2004.

[22] June-Yi Lee, Bin Wang, Matthew C Wheeler, Xiouhua Fu, Duane E Waliser, and In-Sik Kang. Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. *Climate Dynamics*, 40(1-2):493–509, 2013.

[23] E Suhas, JM Neena, and BN Goswami. An indian monsoon intraseasonal oscillations (miso) index for real time monitoring and forecast verification. *Climate dynamics*, 40(11-12):2605–2616, 2013.

[24] Eniko Székely, Dimitrios Giannakis, and Andrew J Majda. Extraction and predictability of coherent intraseasonal signals in infrared brightness temperature data. *Climate Dyn*, 2014.

[25] Dimitrios Giannakis, Wen-wen Tung, and Andrew J Majda. Hierarchical structure of the Madden-Julian oscillation in infrared brightness temperature revealed through nonlinear Laplacian spectral analysis. In *Intelligent Data Understanding (CIDU), 2012 Conference on*, pages 55–62. IEEE, 2012.

[26] Dimitrios Giannakis and Andrew J Majda. Comparing low-frequency and intermittent variability in comprehensive climate models through nonlinear Laplacian spectral analysis. *Geophysical Research Letters*, 39(10), 2012. doi: 10.1029/2012GL051575.

[27] Dimitrios Giannakis and Andrew J Majda. Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proceedings of the National Academy of Sciences*, 109(7):2222–2227, 2012.

[28] Dimitrios Giannakis and Andrew J Majda. Nonlinear Laplacian spectral analysis: capturing intermittent and low-frequency spatiotemporal patterns in high-dimensional data. *Statistical Analysis and Data Mining*, 6(3):180–194, 2013.

[29] Wen-wen Tung, Dimitrios Giannakis, and Andrew J Majda. Symmetric and antisymmetric convection signals in the Madden-Julian oscillation. Part I: basic modes in infrared brightness temperature. *Journal of the Atmospheric Sciences*, 71(9):3302–3326, 2014.

[30] Andrew J Majda and John Harlim. Physics constrained nonlinear regression models for time series. *Nonlinearity*, 26(1):201–217, 2013.

[31] John Harlim, Adam Mahdi, and Andrew J Majda. An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models. *Journal of Computational Physics*, 257:782–812, 2014.

[32] Nan Chen, Andrew J Majda, and Dimitrios Giannakis. Predicting the cloud patterns of the Madden-Julian oscillation through a low-order nonlinear stochastic model. *Geophysical Research Letters*, 41(15):5612–5619, 2014.

[33] Nan Chen and Andrew J Majda. Predicting the real-time multivariate Madden-Julian oscillation index through a low-order nonlinear stochastic model. *Monthly Weather Review*, 2015. in press.

[34] KI Hodges, DW Chappell, GJ Robinson, and G Yang. An improved algorithm for generating global window brightness temperatures from multiple satellite infrared imagery. *Journal of Atmospheric & Oceanic Technology*, 17(10):1296–1312, 2000.

[35] Tyrus Berry, Dimitrios Giannakis, and John Harlim. Nonparametric forecasting of low-dimensional dynamical systems. *arXiv preprint arXiv:1411.5069*, 2014.

[36] S Kravtsov, D Kondrashov, and M Ghil. Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *Journal of Climate*, 18(21):4404–4424, 2005.

[37] D Kondrashov, MD Chekroun, AW Robertson, and M Ghil. Low-order stochastic model and "past-noise forecasting" of the Madden-Julian Oscillation. *Geophysical Research Letters*, 40(19):5305–5310, 2013.

[38] Andrew J Majda and Boris Gershgorin. Quantifying uncertainty in climate change science through empirical information theory. *Proceedings of the National Academy of Sciences*, 107(34):14958–14963, 2010.

[39] Andrew J Majda and Boris Gershgorin. Improving model fidelity and sensitivity for complex systems through empirical information theory. *Proceedings of the National Academy of Sciences*, 108(25):10044–10049, 2011.

[40] Robert S Liptser and Albert N Shiryaev. *Statistics of Random Processes II: II. Applications*, volume 2. Springer, 2001.

[41] M Berkelhammer, A Sinha, M Mudelsee, H Cheng, K Yoshimura, and J Biswas. On the low-frequency component of the enso–indian monsoon relationship: a paired proxy perspective. *Climate of the Past*, 10(2):733–744, 2014.

[42] Boris Gershgorin, John Harlim, and Andrew J Majda. Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *Journal of Computational Physics*, 229(1):32–57, 2010.

[43] Boris Gershgorin, John Harlim, and Andrew J Majda. Test models for improving filtering with model errors through stochastic parameter estimation. *Journal of Computational Physics*, 229(1):1–31, 2010.

[44] M Branicki and AJ Majda. Quantifying bayesian filter performance for turbulent dynamical systems through information theory. *Communications in Mathematical Sciences*, 12(5), 2014.

[45] Peter J Webster and Carlos Hoyos. Prediction of monsoon rainfall and river discharge on 15-30-day time scales. *Bulletin of the American Meteorological Society*, 85(11):1745–1765, 2004.

[46] MJ MCPHADEN. Rama: The research moored array for african-asian-australian monsoon analysis and prediction. *Bull. Amer. Meteor. Soc.*, 90:459–480, 2009.

[47] Michal Branicki and Andrew J Majda. An information-theoretic framework for improving imperfect predictions via Multi Model Ensemble forecasts. *J. Nonlinear Science*, 2014. in press.

[48] Andrew J Majda and Qi Di. Improving prediction skill of imperfect turbulent models through statistical response and information theory. *J. Nonlinear Science*, 2015. submitted.

[49] Nan Chen, Andrew J Majda, and Xin T Tong. Information barriers for noisy Lagrangian tracers in filtering random incompressible flows. *Nonlinearity*, 27(9):2133–2163, 2014.

[50] Nan Chen, Andrew J Majda, and Xin T Tong. Noisy Lagrangian tracers for filtering random rotating compressible flows. *Journal of Nonlinear Science*, 2015. in press.

[51] Richard Kleeman. Measuring dynamical prediction utility using relative entropy. *Journal of the atmospheric sciences*, 59(13):2057–2072, 2002.

[52] Andrew J Majda and Michal Branicki. Lessons in uncertainty quantification for turbulent dynamical systems. *Discrete Contin. Dyn. Syst*, 32(9):3133–3231, 2012.

[53] Michal Branicki, Nan Chen, and Andrew J Majda. Non-Gaussian test models for prediction and state estimation with model errors. *Chinese Annals of Mathematics, Series B*, 34(1):29–64, 2013.