DISTRIBUTION STATEMENT A. Approved for public release; distribution is unlimited.

2014 Annual Report: Physics Constrained Stochastic-Statistical Models for Extended Range Environmental Prediction

Award Number: N00014-12-1-0912

P.I. Andrew J. Majda
New York University
Courant Institute of Mathematical Sciences
251 Mercer Street, New York, NY 10012-2331
phone: (212) 998-3323 fax: (212) 995-4121 email: jonjon@cims.nyu.edu

CO-P. I. John Harlim
Penn State University
Department of Mathematics
214 McAllister Building, University Park, PA, 16802
phone: (814) 863-9024 fax: (814) 865-3735 email: jharlim@psu.edu

CO-P. I. Samuel Stechmann
University of Wisconsin—Madison
Department of Mathematics
480 Lincoln Drive, Madison, WI 53706
phone: (608) 263-4351 fax: (608) 263-8891 email: stechmann@wisc.edu

CO-P. I. Duane E. Waliser
UCLA - JIFRESSE
Young Hall, Room 4242
Los Angeles, CA 90095-7228
phone: (818) 393-4094 fax: (818) 393-3379 email: waliser@ucla.edu

Informal CO-P. I. Dimitrios Giannakis
Department of Mathematics and Center for Atmosphere Ocean Science
Courant Institute of Mathematical Sciences
New York University
251 Mercer St, New York, NY 10012
phone: (212) 998-3316 fax: (212) 998-415-9921 email: dimitris@cims.nuy.edu

LONG-TERM GOALS

Using a combination of theoretical, observational, and high-order model resources, we aim to develop and improve the modelling and forecasting capabilities for crucial problems such as long range weather forecasting of planetary scale convection patterns in the tropics and short term climate change in the polar regions such as sea ice reemergence.

OBJECTIVES

To develop simplified physics constrained stochastic statistical models and techniques for long range environmental forecasting by blending novel ideas from mathematics, statistics and physics and validating the skill of these new models on a suite of tests ranging from observational data to data output from high order models like GCM's to synthetic data from instructive toy models. These objectives include 1) the systematic development of low order (few dimensional) stochastic statistical models, 2) intermediate models with hundreds of variables, as well as 3) novel strategies for improvement of higher order models like GCM's. These objectives include the development and application of new techniques for 1) finding and assessing the intrinsic prediction skill in crucial variables associated with massive data outputs from observation or high order models, 2) the development of multi-scale data assimilation and parameter estimation algorithms, and 3) the crucial understanding of the role of model errors in data assimilation and prediction both to reveal intrinsic information barriers in model classes and to develop strategies to mitigate such errors.

APPROACH

The approach relies on synergy and serendipity among four different research pathways:

I. The use of a combination of mathematical, numerical, amd observational analysis to develop and evaluate intermediate models for the statistical prediction of the Madden Julian Oscilation (MJO), the dominant component of tropical intraseasonal variability with well known significant subseasonal forecasting impact.

II. The development of new kernel methods such as NLSA for data analysis of high dimensional time series and empirical forcasting using training data from models or observations. These methods use techniques from harmonic analysis, machine learning, dynamical systems, and information theory.

III. The development of physics constrained semi-empirical reduced stochastic models, to improve long range forecasting using stochastic and statistical analysis, scientific computing, and physics reasoning.

IV. The development of new strategies for the role of model errors, their mitigation, and intrinsic barriers in model classes using information theory, multi-scale modelling and stochastic-statistical techniques.

<u>MURI Post docs</u>

I. Courant NYU (Majda and Giannakis)

S. Thual

Y. Lee

E. Szekely

J. Zhao

D. Comeau

X. Tong

II. Penn State

T. Berry

III. U. Wisconsin (S. Stechmann)

H.R. Ogrosky

S. Chen

IV. Cal Tech - UCLA (D. Waliser)

J.Stachnik

MURI Graduate Students (Courant Institute)
Nan Chen (3rd year)
DiQi (2nd year)
Noah Brenowitz (2nd year)
Qiu Yang (2nd year)

Unsupported Participants
T. Sapsis (MIT)
A. Mahdi (North Carolina State U.)
M. Branicki (U. Edinburg)
I. Grooms (Courant Institute)
R. Herbei (Ohio State U)
W.W. Tung (Purdue U.)
M. Bushuk (Courant Grad Student)

WORK COMPLETED

The work completed and results are organized along the four topics I) – IV) described in the approach

I) Observations, MJO Prediction, and the Skeleton Model

1. Understanding MJO Termination Events from Observational Data (Stachnik, Waliser, Majda)
2. Development of a stochastic version of the MJO skeleton model (Thual, Majda, Stechmann).
3. Creation of computer codes for analyzing the MJO skeleton in observational data (Stechmann, Majda).
4. Development of an MJO Skeleton Index (MJOS) for identifying the MJO in observational data and climate model data (Stechmann, Majda).
5. Identification of outgoing longwave radiation (OLR) satellite data as a measure of diabatic heating (Stechmann, Ogrosky).
6. Further understanding of MJO skeleton dynamics through numerical simulations and mathematical analysis (Stechmann and S. Chen)
7. Theory for Planetary scale impact of Diurnal Cycle on Instraseasonal time scales (Yang and Majda)

II) Kernel Methods For Data Analysis of Large Dimensional Time Series and Empirical Forecasting

1. Reduced representation of arctic sea ice reemergence in GCMs and observations (Bushuk, Giannakis, Majda)
2. Extraction and predictability of interannual and intraseasonal modes of tropical organized convection from observations (Szekely, Giannakis, Majda, Tung)
3. Kernel-based methods for analog forecasting (Zhao, Comeau, Giannakis, Majda)
4. Kernels for dimension reduction and feature extraction of data generated by dynamical systems (Giannakis)
5. Markov chain Monte Carlo (MCMC) algorithms for parameter estimation of stochastic models with hidden intermittent instabilities (Chen, Giannakis, Majda, Herbei)
6. Extending the applicability of diffusion maps for sparsely sampled data (Berry and Harlim)
7. Develop a new uncertainty quantification method via diffusion map based non-parametric models (Berry and Harlim)

III) <u>Physics Constrained Low Order Nonlinear Stochastic Models for Long Range Prediction and Data Assimilation</u>

1. Predicting the cloud patterns of the MJO through a low-order nonlinear stochastic model
   Nan Chen (PhD. Student, Courant), Majda Giannakis (Courant)
2. An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models J. Harlim (Penn State), A. Mahdi (external collaborator, NC State), A. Majda (Courant)
3. Conceptual Dynamical Models for Turbulence
   A. Majda, Y. Lee (Post Doc, Courant)

IV) <u>Model Error, Information Barriers, and Multi-Scale Data Assimilation</u>

1. Information barriers for noisy Lagrangian traers in filering random incompressible flows
   Nan Chen (Ph.D. Student, Courant), A.J. Majda, X.T. Tong (post doc, Courant)
2. Noisy Lagrangian tracers for filtering random rotating compressible flows
   Nan Chen (Ph.D. Student, Courant), A.J. Majda, X.T. Tong (post doc, Courant)
3. Quantifying Bayesian Filter Performance for Turbulent Dynamical Systems through Information Theory
   M. Branicki (Unsupported, U. Edinburgh), A.J. Majda (Courant)
4. Linear theory for filtering nonlinear multiscale systems with model error
   T. Berry (Post Doc, Penn State), J. Harlim (Penn State)
5. Ensemble Kalman Filters for Dynamical Systems with Unresolved Turbulence
   Grooms (Courant unsupported), Y. Lee (Post doc, Courant), A.J. Majda
6. Blended Particle Filters for Large Dimensional Chaotic Dynamical Systems
   A.J. Majda, DiQi (Ph. D. Courant) T. Sapsis (MIT, unsupported)
7. Multi-scale methods for data assimilation in turbulent systems
   Y. Lee (Post doc, Courant), A.J. Majda, Courant

RESULTS

I) Observations, MJO Prediction, and the Skeleton Model

1. Precursor Environmental Conditions Associated with the Termination of Madden-Julian events
   Stachnik (UCLA, JPL)
   Waliser (UCLA-JPL)
   Majda (Courant)

   The study submitted to J.Atmos. Sci presents an analysis of the precursor environmental conditions related to the termination of Madden-Julian oscillation (MJO) events. A simple climatology is created using a real-time MJO monitoring index, documenting the locations and frequencies of MJO decay. Lead-lag composites of several atmospheric variables including temperature, moisture, and intraseasonal wind anomalies are generated from three reanalyses. There is remarkable agreement among the reanalyses with long-term, lower tropospheric moisture deficits over the local domain best identifying termination events over the Indian Ocean. MJO termination over the Indian Ocean appears to be linked to a northward shift of the Intertropical Convergence Zone (ITCZ) with possible lead times as much as 20 days prior to MJO decay.
   Statistically significant differences are also identified more than 10 days in advance of MJO termination events in the west Pacific, though the vertical velocity and moisture anomalies are more symmetric about the equator. Unlike the Indian Ocean and west Pacific, MJOs that terminate over the maritime continent appear to be related to its own intensity, rather than the downstream environmental conditions. As such, only the strongest MJOs tend to propagate into the warm pool region.
   Finally, a budget analysis is performed on the three-dimensional moisture advection equation in order to better elucidate what time-scales and physical mechanisms are most important for MJO termination. The combination of intraseasonal vertical circulation anomalies coupled with the mean state specific humidity best explain the anomalous moisture patterns associated with MJO termination, suggesting that the downstream influence of the MJO circulation can eventually lead to its upstream demise. (See Figure 1)

   A nearly completed project of Stachnik, Waliser, Majda, Stechmann, and Thual addresses the skill of the stochastic MJO skeleton model in predicting these termination events. (See Figure 2 for such a comparison)

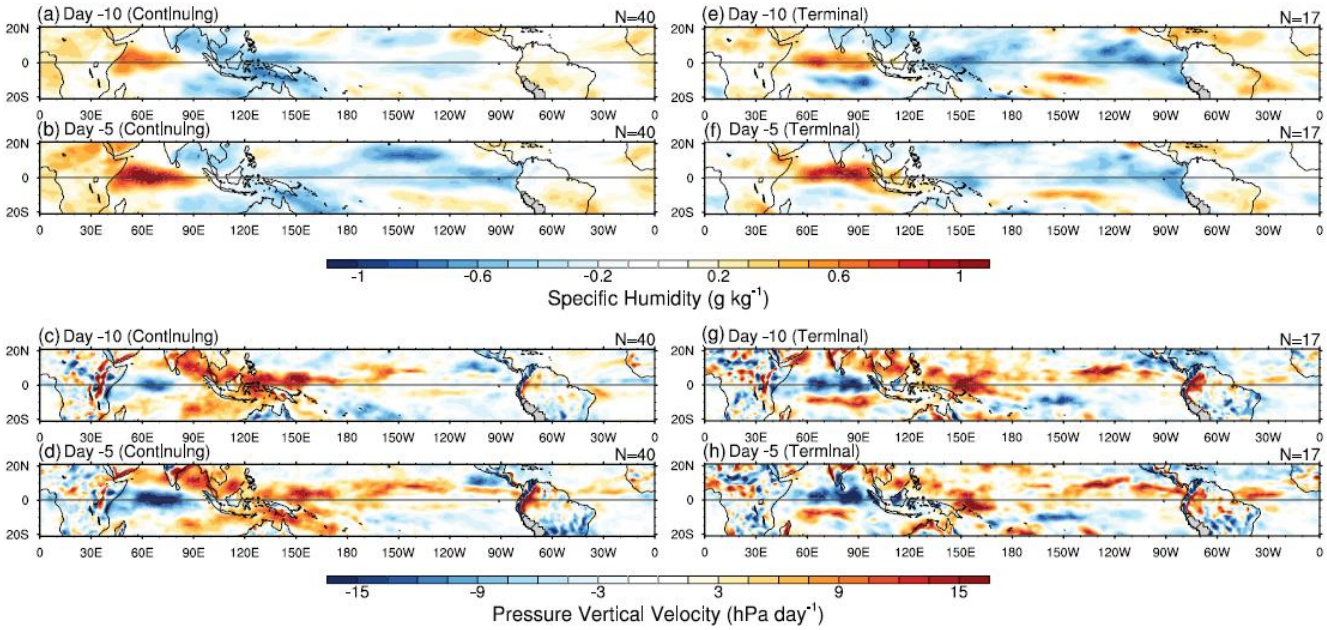


Figure 1. Spatial patterns of the 20-100 day filtered, (a), (b) low-level (850-700 hPa) moisture and (c), (d) low-level vertical velocity anomalies for continuing RMM Phase 2 MJO events from MERRA data. Composite snapshots are provided for lags at (a), (c) Day -10 and (b), (d) Day -5 prior to the start of a continuing event. The equator is plotted as a thin, gray line in all panels for reference. Panels (e)-(h) are the same as in (a)-(d), but for terminating MJO events.

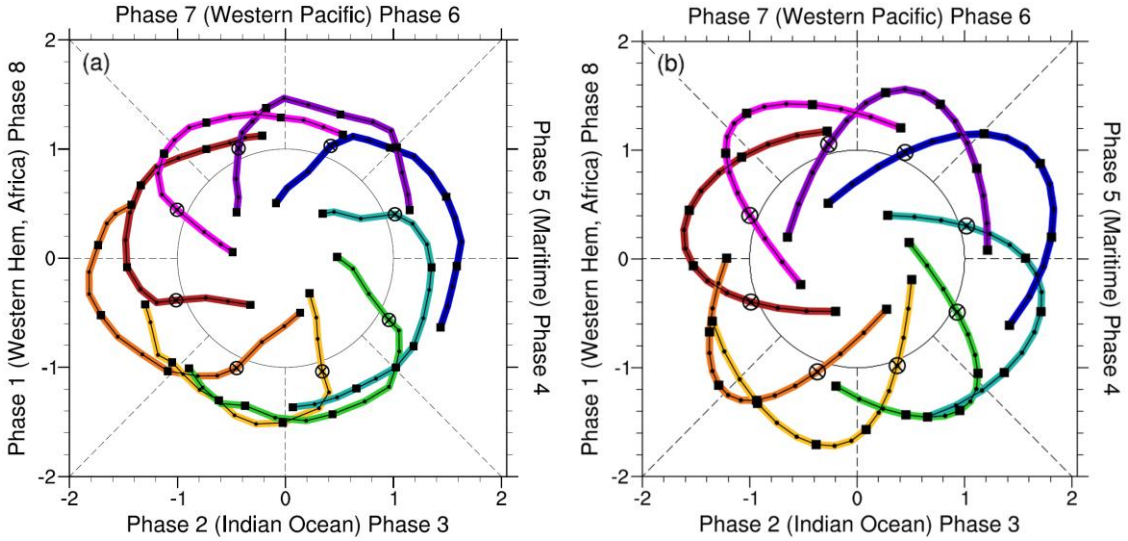


Figure 2. Phase space composites for MJO termination events using the (a) observed RMM and (b) model-derived RMM-like index for the warm pool case using a multivariate EOF analysis. The observational composites use 34 years of data (1979-2012) and the model composites use a similar period of ~34 years. The composite RMM values are shown for the start of a termination event (i.e., the last day before the combined amplitude decays below unity) as a circumscribed X, with different colors indicating the composite history for each phase before (12 days) and after (3 days) the

termination event. Individual daily average RMM values are shown with small dots, with every three days indicated with closed squares for reference. The geographic locations of the index phase (e.g., "Indian Ocean") are relative for the model warm pool case (right) as the PCs are currently calculated from different EOFs than the observed RMM index.

2. A stochastic skeleton model for the MJO.

S. Thual (Post Doc, Courant Institute)
A. J. Majda (PI, Courant Institute)
S. N. Stechmann (co-PI, Wisconsin)

(Background information: The Madden-Julian oscillation (MJO) is the dominant mode of variability in the tropical atmosphere on intraseasonal time scales and planetary spatial scales. Despite the primary importance of the MJO and the decades of research progress since its original discovery, a generally accepted theory for its essential mechanisms has remained elusive. In recent work by two of the authors, a minimal dynamical model has been proposed that recovers robustly the most fundamental MJO features of (i) a slow eastward speed of roughly 5 m/s, (ii) a peculiar dispersion relation with dw/dk=0, and (iii) a horizontal quadrupole vortex structure. This model, the skeleton model, depicts the MJO as a neutrally stable atmospheric wave that involves a simple multiscale interaction between planetary dry dynamics, planetary lower-tropospheric moisture, and the planetary envelope of synoptic-scale activity.)

In this article, it is shown that the skeleton model can further account for (iv) the intermittent generation of MJO events and (v) the organization of MJO events into wave trains with growth and demise, as seen in nature. The goal is achieved by developing a simple stochastic parameterization for the unresolved details of synoptic-scale activity, which is coupled to otherwise deterministic processes in the skeleton model. In particular, the intermittent initiation, propagation, and shut down of MJO wave trains in the skeleton model occur through these stochastic effects. This includes examples with a background warm pool where some initial MJO-like disturbances propagate through the western region but stall at the peak of background convection/heating corresponding to the Maritime Continent in nature.

3. Asymmetric intraseasonal events in the stochastic skeleton MJO model with seasonal cycle.

S. Thual (Post Doc, Courant Institute)
A. J. Majda (PI, Courant Institute)
S. N. Stechmann (co-PI, Wisconsin)

We analyze here the solutions of a stochastic skeleton model with an idealized seasonal cycle, namely a background warm pool state of heating/moistening displacing meridionally during the year. The present model considers both equatorial and off-equatorial convective heating structure, as well as tilted events with convective heating structure oriented north-westward and associated northward propagation that is reminiscent of the summer monsoon intraseasonal oscillation. The model also reproduces qualitatively the meridional migration of intraseasonal variability during the year, that approximatively follows the meridional migration of the background warm pool.

4. Identifying the Madden-Julian Oscillation skeleton in observational data.

S. N. Stechmann (co-PI, Wisconsin)
A. J. Majda (PI, Courant Institute)

The Madden-Julian oscillation (MJO) skeleton model offers a theoretical prediction of the MJO's structure. Here, a method is described for identifying this structure in observational data. The method utilizes projections onto equatorial wave structures, and a main question is: Can this method isolate the MJO without using temporal filtering or empirical orthogonal functions? For the data projection, a wide range of data is incorporated: multiple variables (wind, geopotential height, water vapor, and, as a proxy for convective activity, outgoing longwave radiation); multiple pressure levels (850 and 200 hPa); and multiple latitudes (both equatorial and off-equatorial). Such a data variety is combined using a systematic method, and it allows a distinction between the Kelvin and Rossby components of the MJO's structure. Results are illustrated for some well-known cases, and statistical measures are presented to quantify the variability of the MJO skeleton signal, MJOS(x,t), and its amplitude, MJOSA(t). The robustness of the methods is demonstrated through a suite of sensitivity studies, including tests with an alternative projection method, which also indicate a possible limitation in the water vapor formulation that will require further attention in the future. Finally, a simple interpretation is given for the MJO skeleton structure: it is related to the wave response to a moving heat source. From either perspective, the methods here identify signals that project onto coupled convection-circulation patterns; and the results suggest that a large portion of the MJO's structure is consistent with such a coupled pattern.

5. The Walker circulation, diabatic heating, and outgoing longwave radiation.

S. N. Stechmann (co-PI, Wisconsin)
H. R. Ogrosky (Post Doc, Wisconsin)

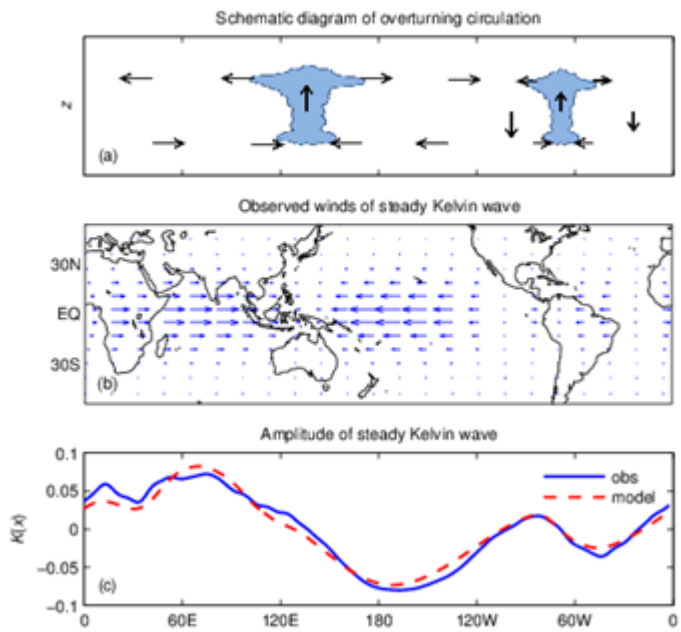
For the tropical atmosphere on planetary scales, it is common to model the circulation using strong damping. Here, with new data analysis techniques, evidence suggests that damping can actually be neglected. Specifically, near the equator, the east-west overturning circulation is in agreement with the undamped wave response to atmospheric heating. To estimate the heating, satellite observations of outgoing longwave radiation (OLR) are used. Frequently OLR is used as a heuristic indicator of cloudiness. Here, the results further suggest that OLR variations are actually proportional to diabatic heating variations, with a proportionality constant of 18 W/m^2 / (K/day). While the agreement holds best over long time averages of years or decades, it also holds over shorter periods of one season or one month. Consequently, it is suggested that the strength of the Walker circulation--and its evolution in time--could be estimated using satellite data. (See I) Figs. 3,4)

6. Nonlinear traveling waves for the skeleton of the Madden-Julian Oscillation.

S. Chen (Post Doc, Wisconsin)
S. N. Stechmann (co-PI, Wisconsin)

Traveling wave solutions are presented for the MJO skeleton model of Majda and Stechmann. The nonlinear traveling waves come in four types, corresponding to the four types of linear wave solutions, one of which has the properties of the MJO. In the MJO traveling wave, the convective activity has a pulse-like shape, with a narrow region of enhanced convection and a wide region of suppressed convection. Furthermore, an amplitude-dependent dispersion relation is derived, and it shows that the

nonlinear MJO has a lower frequency and slower propagation speed than the linear MJO. By taking the small-amplitude limit, an analytic formula is also derived for the dispersion relation of linear waves. To derive all of these results, a key aspect is the model's conservation of energy, which holds even in the presence of forcing. In the limit of weak forcing, it is shown that the nonlinear traveling waves have a simple sech-squared waveform.



I) Figure 3

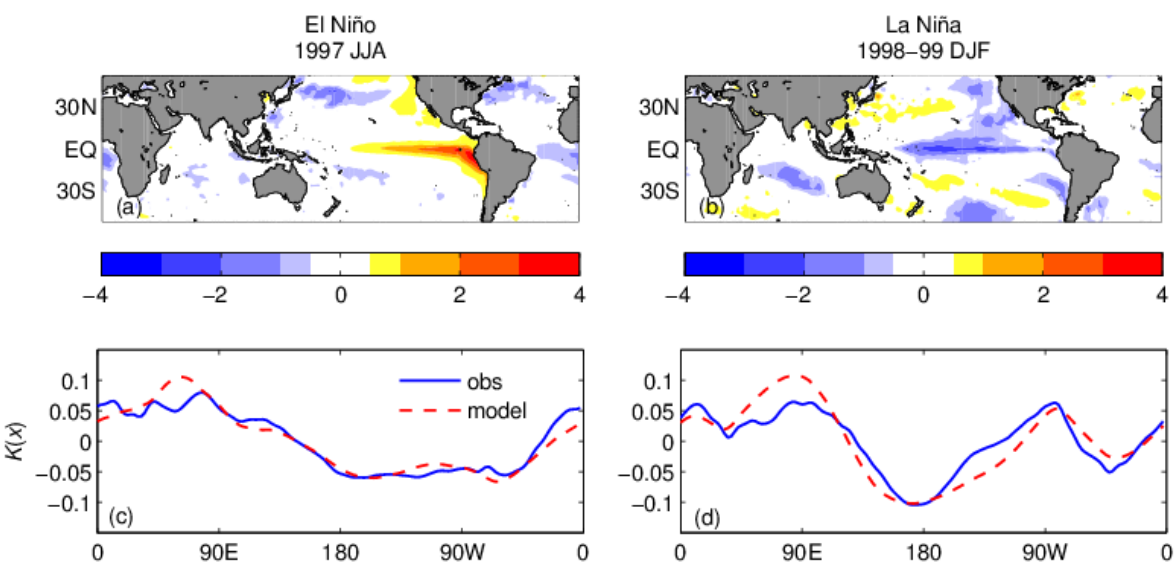


Figure 4

7. A Multi-Scale Model for the Intraseasonal Impact of the Diurnal Cycle of Tropical Convection

Qiu Yang (Phd. Student, Courant)
Andrew Majda

One of the crucial features of tropical convection is the observed variability on multiple spatiotemporal scales, ranging from cumulus clouds on the daily time scale over a few kilometers to intraseasonal oscillations over planetary scales. The diurnal cycle of tropical convection is a signifcant process but its large-scale impact is not well understood. Here we develop a multi-scale analytic model to assess the intraseasonal impact of planetary-scale inertial oscillations in the diurnal cycle. A self-contained derivation of a multi-scale model governing planetary-scale tropical ows on the daily and intraseasonal time scale is provided below, by following the derivation of systematic multi-scale models for tropical convection. This derivation demonstrates the analytic tractability of the model. The appeal of the multi-scale model developed here is that it provides assessment of eddy flux divergences of momentum and temperature and their intraseasonal impact on the planetary-scale circulation in a transparent fashion. Here we use it to study the intraseasonal impact of a model for the diurnal cycle heating with two local phase-lagged baroclinic modes with the congestus, deep, stratiform life cycle. The results show that during boreal summer the eddy flux divergence of temperature dominates in the northern hemisphere, providing significant heating in the middle troposphere of the northern hemisphere with large-scale ascent and cooling with subsidence surrounding this heating center. Due to the analytic tractability of the model, such significant eddy flux divergence of temperature is traced to meridional asymmetry of the diurnal cycle heating. In an ideal zonally symmetric case, the resulting planetary-scale circulation on the intraseasonal time scale during boreal summer is characterized by ascent in the northern hemisphere, southward motion in the upper troposphere, descent around the equator and northward motion in the lower troposphere. The intraseasonal impact of the diurnal cycle on the planetary scale also includes negative potential temperature anomalies in the lower troposphere, which suggests convective triggering in the tropics. Furthermore, a fully coupled model for the intraseasonal impact of the diurnal cycle on the Hadley cell shows that the overturning motion induced by the eddy flux divergences of momentum and temperature from the diurnal cycle can strengthen the upper branch of the winter cell of the Hadley circulation, but weaken the lower branch of the winter cell. The corresponding eddy fluxes from the diurnal cycle are very weak for the equinox case with symmetric meridional profiles and eddy momentum fluxes are small for all scenarios considered here.

II) <u>Kernel Methods For Data Analysis of Large Dimensional Time Series and Empirical Forecasting</u>

1. Reemergence Mechanisms for Arctic Sea Ice

M. Bushuk (PhD student, Courant Institute)
D. Giannakis (Courant Institute)
A. J. Majda (PI, Courant Institute)

Giannakis, PI Majda, and PhD student Mitchell Bushuk have studied mechanisms for spring to fall reemergence of arctic sea ice concentration in model output and satellite observations using an advanced data analysis technique called nonlinear Laplacian Spectral analysis (NLSA) [10]. Sea ice reemergence is a process in which sea ice anomalies originating in the melt season (spring) are positively correlated with anomalies in the growth season (fall) despite a loss of correlation in the intervening summer months. Because sea ice is a key dynamical component of the climate system, elucidating the role of the ocean, the atmosphere, and sea ice itself in this phenomenon may lead to significant advances in predictive capabilities for the state of the arctic on annual to interannual

timescales. In a paper published in the Journal of Climate [1], Bushuk, Giannakis, and Majda have constructed a reduced representation of sea ice reemergence in the North Pacific Ocean through a low-dimensional family of spatiotemporal modes extracted from global circulation model (GCM) output and satellite observations using NLSA. The low-dimensional modes extracted objectively via this algorithm provide a skeleton for a sea ice reemergence mechanism proposed earlier [11] using lagged correlation analysis where persistence in ocean temperature acts as a "memory" for sea ice anomalies during the summer months when no significant sea ice is present in the North Pacific. Moreover, it was found that the dominant pattern of sea ice reemergence in the Bering Sea of the North Pacific depends strongly on the phase on the North Pacific Gyre Oscillation (NPGO), a significant pattern of oceanic variability on interannual time scales. Thus, the state of the NPGO was identified as an important variable for extended-range prediction for ice concentration in the North Pacific through sea ice reemergence.

In a paper recently submitted to the Journal of Climate [2], Bushuk, Giannakis, and Majda have extended the analysis in [1] to cover the whole of the Arctic, and to include both ocean and atmosphere variables [sea surface temperature (SST) and sea level pressure (SLP), respectively]. A major finding of this work, illustrated in Figure 1, is that the North Pacific patterns identified in [1] are part of larger-scale patterns characterized by strong anti-correlation in sea ice concentration between the Bering Sea and spatially separated sectors in the North Atlantic (the Labrador and Barents Seas). The teleconnection between these domains is consistent with anomalous winds associated with large-scale atmospheric patterns such as the annular mode. Thus, the analysis in [2] has revealed a reemergence process where the ocean stores anomalies of sea ice in the summer and the atmosphere organizes these anomalies over spatially separated ocean basins. Currently, Giannakis and Bushuk are studying the causal structure of this process by analyzing via NLSA model output from GCMs with non-dynamical atmospheres or oceans.

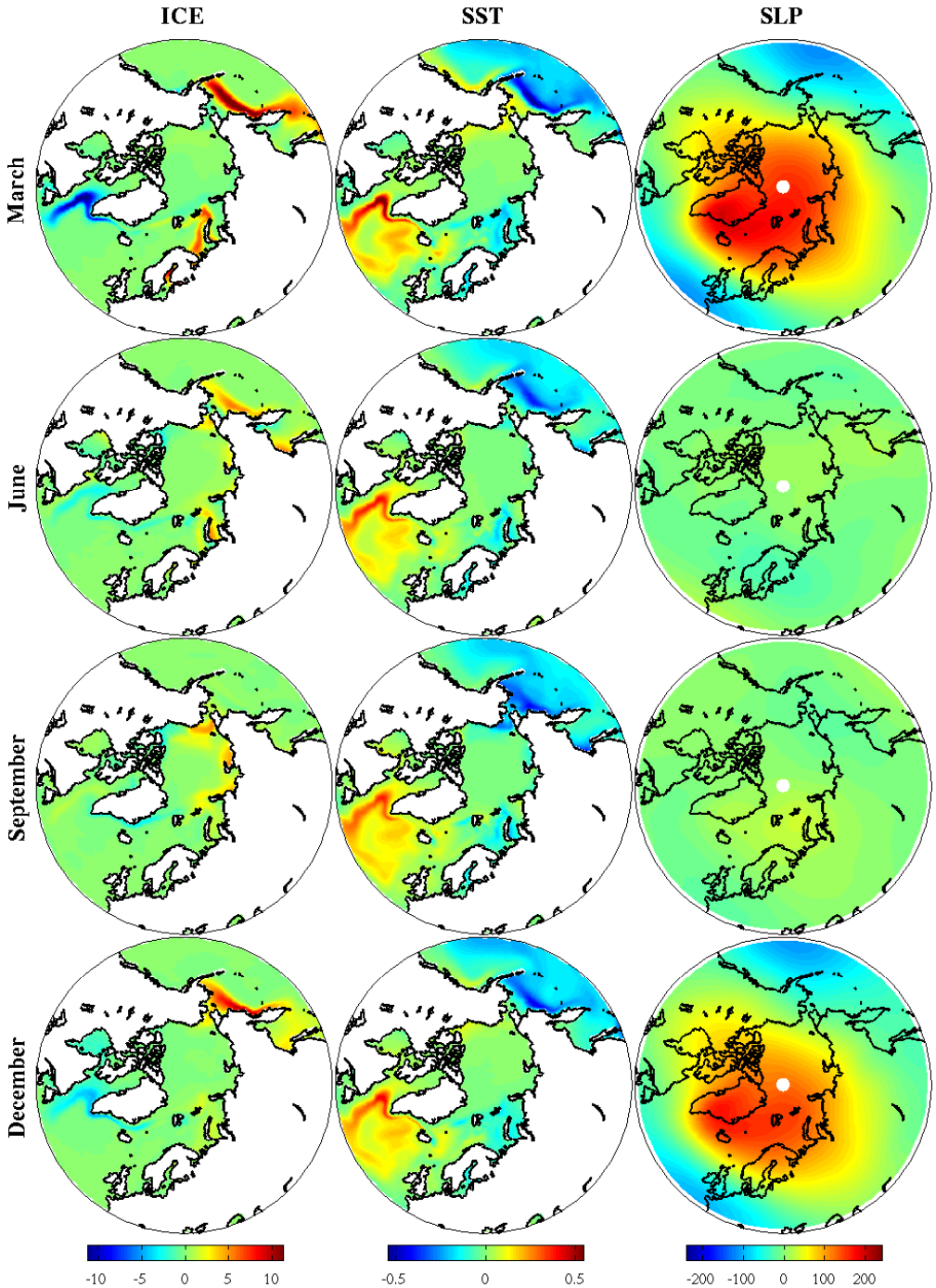


Figure II 1. Reconstruction of sea ice concentration, SST, and SLP anomalies in the arctic using NLSA reemergence modes during an active phase of the NPGO. In all panels red (blue) colors indicate positive (negative) anomalies. In March, regions of sea ice concentration anomalies of one sign are spatially collocated with SST anomalies of the opposite sign. In June, the sea ice concentration anomalies virtually disappear, but leave an imprint in SST anomalies. These positive (negative) anomalies inhibit (allow) sea ice formation, thus allowing the ice anomalies of the preceding winter to reemerge. The geostrophic winds associated with the annular SLP pattern in the right-hand column are cold Northerlies (warm Southerlies) in the North Pacific (North Atlantic), giving rise to anticorrelated ice anomalies between the two spatially separated basins.

2. Extraction and Predictability of Coherent Intraseasonal Signals in Infrared Brightness Temperature Data

D. Giannakis (Courant Institute)
A. J. Majda (PI, Courant Institute)
E. Szekely (Postdoc, Courant Institute)
W.-w. Tung (external collaborator, Purdue University)

In a series of papers in the *Journal of Atmospheric Sciences* [3, 4], *Climate Dynamics* [5], and the *4th International Workshop on Climate Informatics* [6], Giannakis, PI Majda, MURI postdoc Eniko Szekely, and collaborator Wen-wen Tung of Purdue University have investigated the extraction and predictability of coherent spatiotemporal patterns of organized tropical convection in satellite observations of infrared brightness temperature (T_b); an observational proxy for convective activity. The principal objective of this work is to recover from the high-dimensional T_b data the dominant instraseasonal oscillation (ISO) modes including the eastward-propagating Madden Julian Oscillation (MJO) and the poleward-propagating Boreal Summer Intraseasonal Oscillation (BSISO), applying minimal preprocessing to the data. These phenomena are of critical importance in explaining and predicting large-scale convective organization at subseasonal timescales (and also influence the global climate through extratropical interactions), yet their accurate simulation by large-scale numerical models remains elusive. A significant obstacle in advancing the scientific understanding and simulation capability of ISOs is that the phenomena themselves are defined subjectively through data analysis techniques such as empirical orthogonal functions (EOFs) applied to data which have been preprocessed to isolate the intraseasonal signals of interest. This has led to significant discrepancies in the inferred spatiotemporal structure of the ISOS, as well as inconsistencies in the identification of significant events [12]. In an effort to extract ISO modes objectively from observational data, the research in [3-6] applies NLSA to T_b data from the CLAUS multi-satellite archive. This work advances the state-of-art in ISO extraction through the use of a nonlinear analysis technique (NLSA) which requires no preprocessing of the data, and has high skill in capturing dynamically-significant patterns with temporal intermittency which may not accessible via conventional linear approaches such as EOFs and singular spectrum analysis (SSA).

In [3-6], the MJO and BSISO were found to emerge naturally from NLSA applied to equatorially averaged and two-dimensional CLAUS T_b data with significantly higher fidelity than SSA-type approaches. In particular, NLSA yields two distinct mode families representing MJO and BSISO, which are characterized by strong temporal intermittency, and are mainly active in boreal winter (MJO) or boreal summer (BSISO). In contrast, SSA applied to the same dataset yields modes which are active year-round and exhibit mixed eastward and poleward propagation. By avoiding the data preprocessing which eliminates non-intraseasonal modes, NLSA also produces a hierarchy of modes that influence and are influenced by MJO and BSISO, including the annual cycle and its harmonics, the El Nino Southern Oscillation (ENSO), and the diurnal cycle. In [5], these modes were employed to construct a low-dimensional space of predictor variables for the MJO, and the potential predictability of MJO in T_b data was assessed using the information-theoretic framework developed in [13, 14]. It was found that the predictability of the extracted MJO modes in fall (i.e., prior to the MJO active season in boreal winter) depends strongly on the state of ENSO, with MJO remaining predictable at 80-day leads during strong El Nino years. During the active season (boreal winter), the current state of MJO as well as ENSO were both found to be significant predictors. Ongoing and future work in this area will be to perform coupled NLSA using both T_b and atmospheric circulation variables, as well as

to carry out analyses over the West Pacific Warm Pool region to identify smaller-scale convectively coupled waves. (See II, Fig 2)

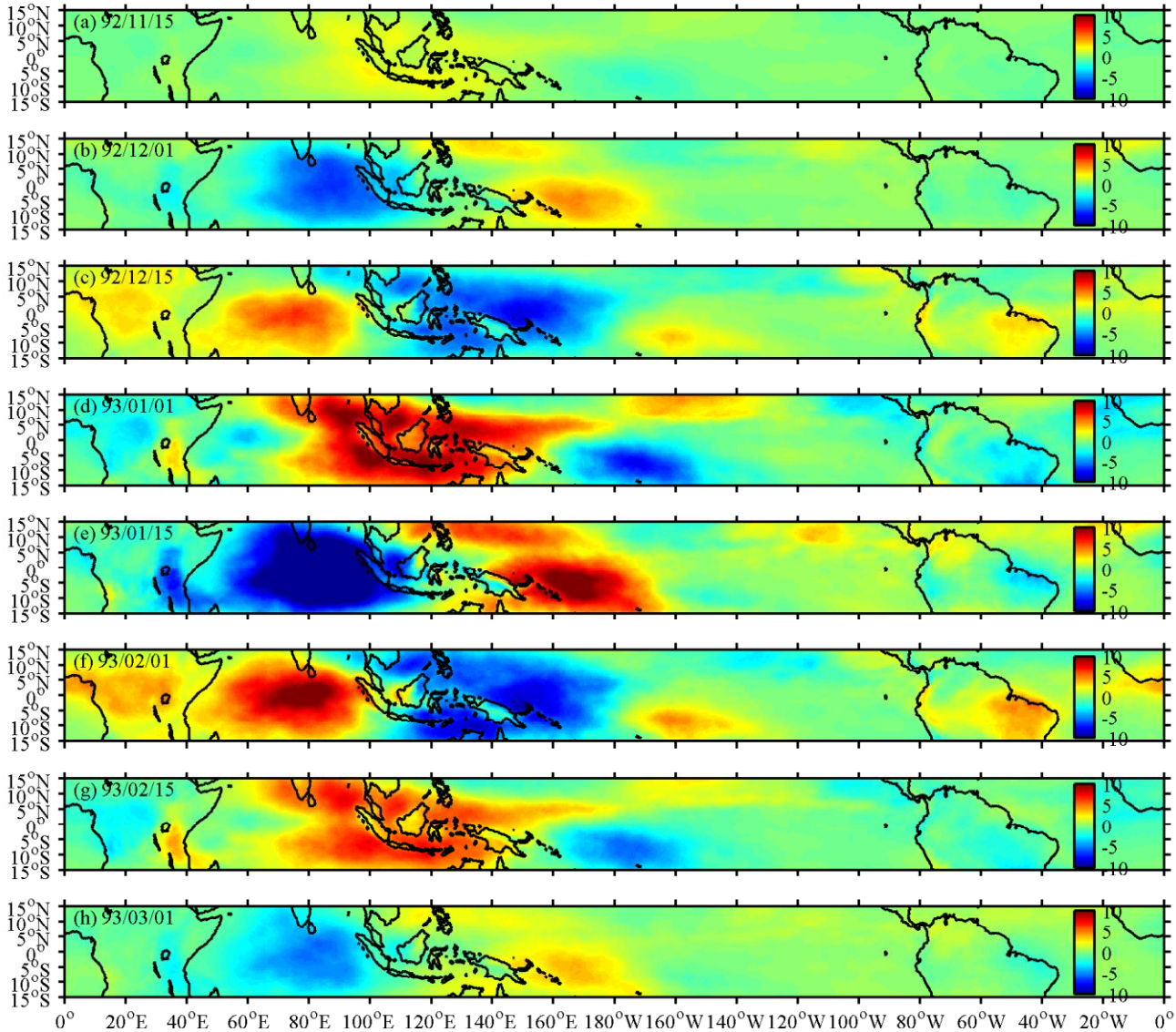


Figure 2. Reconstruction of the MJO wavetrain observed during the intensive observing period (IOP) November 1992–March 1993 of the TOGA COARE field campaign using spatiotemporal modes recovered via NLSA. The color maps show temperature $\mathbf{T_b}$ *anomalies in K. Blue (red) colors correspond to increased convection (decreased cloudiness). (a) No MJO activity is present; (b, c, d) the first MJO initiates over the Indian Ocean, propagates eastward over the Indonesian Maritime Continent, and decays after reaching the* $\mathbf{180^o}$ *dateline; (e,f,g) a second, stronger, MJO event with an initiation signal over East Africa; (h) a weak third event starting at the end of the TOGA COARE IOP.*

3. Kernel Methods for Analog Forescasting

D. Comeau (Postdoc, Courant Institute)
D. Giannakis (Courant Institute)
A. J. Majda (Courant Institute)
J. Zhao (Postdoc, Courant Institute)

Giannakis and MURI postdoc Jane Zhao have recently submitted a paper [7] to Nonlinearity on analog forecasting techniques using kernel methods. Analog forecasting is a technique pioneered by Lorenz [15] whereby one makes a prediction about the future value of an observable f(y) of a dynamical system using a reference record y(t) of past observations of the system state y. In conventional analog forecasting one identifies the state y(t^*) in the historical record which most closely resembles the state of the system x at initialization time, and makes a forecast based on the evolution of that state, i.e., f_forecast (t;x) = f(y(t^*+t)). In the technique developed in [6], this approach is combined with kernel methods and out-of-sample extension techniques originally developed in machine learning and harmonic analysis [16] to leverage the low-dimensional nonlinear manifold structure of the data. More specifically, the forecast based on a single analog y(t^*) is replaced by a locally-averaged forecast on the data manifold of historical observations using the similarity kernel K(x,y(t)) originally used in NLSA algorithms [9] to produce a forecast weighing all of the states in the historical record, viz. f_forecast (t;x)=∑_s⁒[K(x,y(s))f(y(s+t))]. A key feature of the NLSA kernel is that it is evaluated in Takes delay-coordinate space, i.e., it depends on the similarity of the state of the system at initialization and the historical data measured over a time window (as opposed to similarity between "snapshots" at a single time instance). As a result, the NLSA kernel is able to identify analogs with strong dynamical similarity, thus yielding more skillful forecasts than what is possible with conventional analog approaches. In [6], the advantages of kernel-based analog forecasting over conventional analog techniques were demonstrated in a challenging application involving North Pacific SST variability, where traditional linear regression models perform more poorly than trivial persistence forecasts; see Figure II 3.

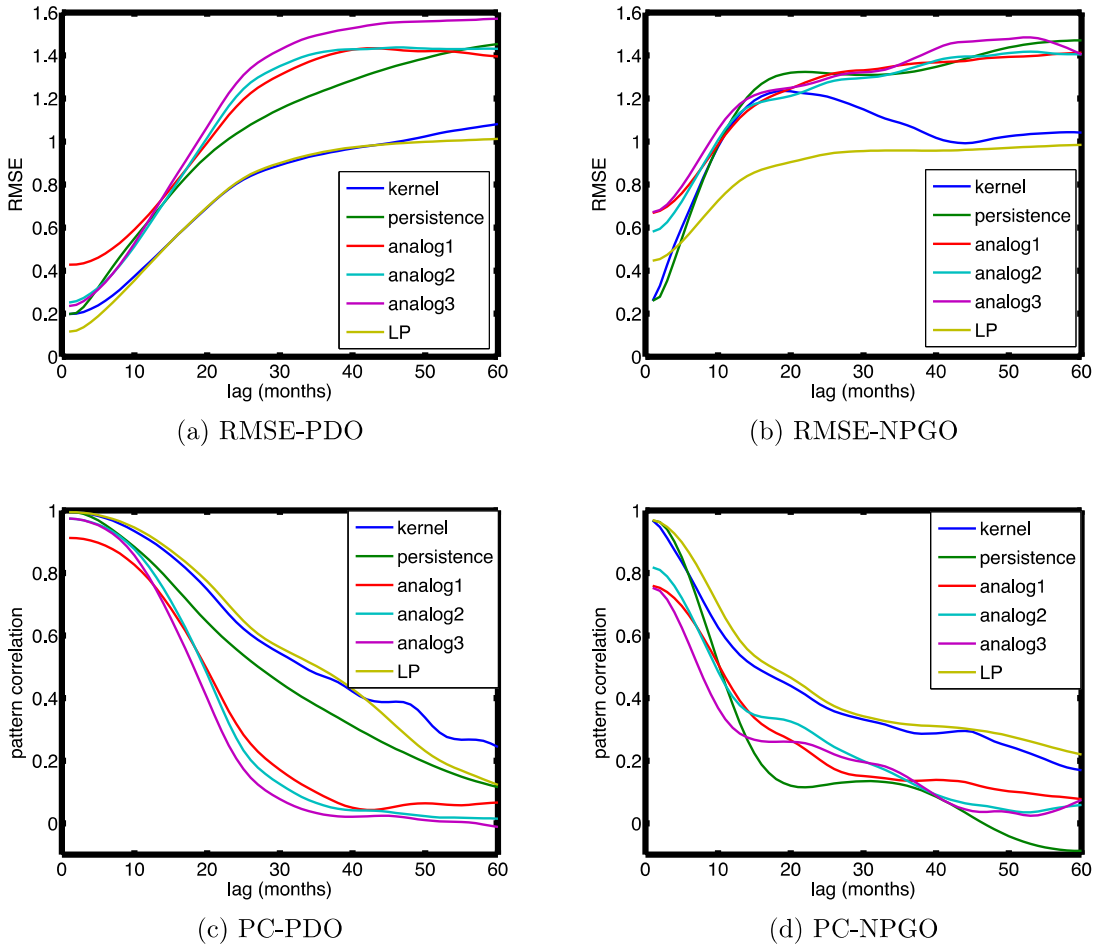
These topics are being further pursued by Giannakis, PI Majda, Zhao, and MURI postdoc Darin Comeau in a paper in preparation on long-range (interannual) forecasting of coupled SST and sea ice variability in the North Pacific using GCM output as the historical reference dataset and satellite observations as hindcast data. Compared to [7], an additional challenge in this work is that the reference dataset has model error. Nevertheless, hindcasts with the kernel analog approach for low-frequency modes such as the NPGO are found to significantly outperform conventional analog as well as regression-based forecasts.

4. Kernels for Data Analysis in Dynamical Systems

D. Giannakis (Courant Institute)

Giannakis has recently developed a family of kernels for dimension reduction and feature extraction of data generated by dynamical systems. This work, which has been submitted to the SIAM Journal on Applied Dynamical Systems [8], builds on the existing NLSA framework [10] for empirical extraction temporal and spatial modes of variability from high-dimensional observed data. The newly developed family of kernels (so-called cone kernels) include the NLSA kernels as a special case, and feature an angular dependence on the time-tendency of the data, designed to enhance the capability of the kernel to extract modes of variability which are slowly-varying in the course of dynamical evolution (i.e., are potentially more predictable) and have strong invariance properties under nonlinear transformations of

the data (i.e., the modes are intrinsic to the dynamical system generating the data). The paper [8] provides a theoretical interpretation of the cone and NLSA kernels, and demonstrates their timescale separation and invariance properties in toy models and real-world applications on low-frequency and intermittent SST variability in GCMs. More generally, these kernels should be applicable in a variety of data-driven MURI activities, including the predictability applications in B and C.



(a) RMSE-PDO (b) RMSE-NPGO

(c) PC-PDO (d) PC-NPGO

Figure II 3. Predictive skill for the Pacific Decadal Oscillation (PDO) and the North Pacific Gyre Oscillation (NPGO) in the North Pacific Sector of the CCSM3 climate model, showing root mean square error (RMSE) and pattern correlation (PC) with respect to a hindcast dataset independent of the training dataset. "analog1-analog3" are results based on different implementations of the traditional analog forecasting technique based on a single analog. "kernel" and "LP" are results based on two implementations of the forecasting technique. "persistence" is a trivial forecast based on the state of the PDO and NPGO at initialization. In all cases, the kernel methods significantly outperform the conventional analog forecasting methods as well as the persistence forecast.

5. Markov Chain Monte Carlo Algorithms for Parameter Estimation in Systems with Hiden Intermittent Instability

N. Chen (PhD student, Courant Institute)
D. Giannakis (Courant Institute)
R. Herbei (external collaborator, Ohio State University)
A. J. Majda (PI, Courant Institute)

Giannakis, PI Majda, PhD student Nan Chen, and collaborator Radu Herbei of Ohio State University have developed a Markov Chain Monte Carlo (MCMC) algorithm for estimating from training data the parameters of reduced stochastic models with intermittent instability. This class of models (called SPEKF models in a filtering context where they were first introduced [17]) is able to describe a broad class of signals with intermittency and non-Gaussian statistics generated by turbulent dynamical systems (such as the atmosphere and the ocean), including the MJO and BSISO discussed in B above. Despite their attractiveness in reduced modeling applications, parameter estimation in these models is challenging due to the fact that the instability is driven by an augmented variable (used to model the dynamical effect of the unobserved turbulent degrees of freedom on the observed signal), which is not observed in the training data. In an MCMC context, this means that conventional samplers for the model parameters have low acceptance rates and poor mixing properties, leading to reduced models with poor predictive skill. In a paper in press at the SIAM/ASA Journal on Uncertainty Quantification [9], Chen, Giannakis, Herbei, and Majda have developed a novel MCMC algorithm which overcomes these limitations, and have demonstrated the skill of this algorithm in forecasting intermittent signals in perfect-model scenarios as well as cases with model error. A key element of the newly developed MCMC algorithm is a preconditioning step which places an informative prior on the unobserved process sampled at the observation times using a mean stochastic model. In this manner, the new sampler is able to explore efficiently and with sufficient accuracy the joint posterior distribution for the parameters and unobserved path, despite that the unobserved path is infinite-dimensional. Figure 4 illustrates the parameter estimation and prediction skill for a reduced model constructed with this algorithm trained on a signal where the intermittent instability is caused by an infrequently occurring negative damping. Besides MURI extended-range forecasting applications, this tool should have broad applicability in the MCMC and stochastic modeling communities.

6. Variable Bandwidth Diffusion Kernels

T. Berry (post doc, Penn State)
J. Harlim (Co-PI, Penn State)

In this paper, we show that the restriction of compact domain data in utilizing diffusion maps theory as an operator estimation algorithm can be overcome by varying the bandwidth of the kernel spatially. This is particularly important for accurate estimation in regions where the data is sparsely sampled (e.g., on the tail of a distribution).

We present an asymptotic expansion of these variable bandwidth kernels for arbitrary bandwidth functions; generalizing the theory of diffusion maps. Subsequently, we present error estimates for the corresponding discrete operators, which reveal how the small sampling density leads to large errors; particularly for fixed bandwidth kernels. By choosing a bandwidth function inversely proportional to the sampling density (which can be estimated from data) we are able to control these error estimates uniformly over a non-compact manifold, assuming only fast decay of the density at infinity in the ambient space.

We numerically verify these results by constructing the generator of the Ornstein-Uhlenbeck process on the real line using data sampled independently from the invariant measure. In this example, we find that the fixed bandwidth kernels yield reasonable approximations for small data sets when the bandwidth is carefully tuned, however these approximations actually degrade as the amount of data is increased. On the other hand, an operator approximation based on a variable bandwidth kernel does converge in the limit of large data; and for small data sets exhibits reduced sensitivity to bandwidth selection. Moreover, even for compact manifolds, variable bandwidth kernels give better approximations with reduced dependence on bandwidth selection. These results extend the classical statistical theory of variable bandwidth density estimation to operator approximation.

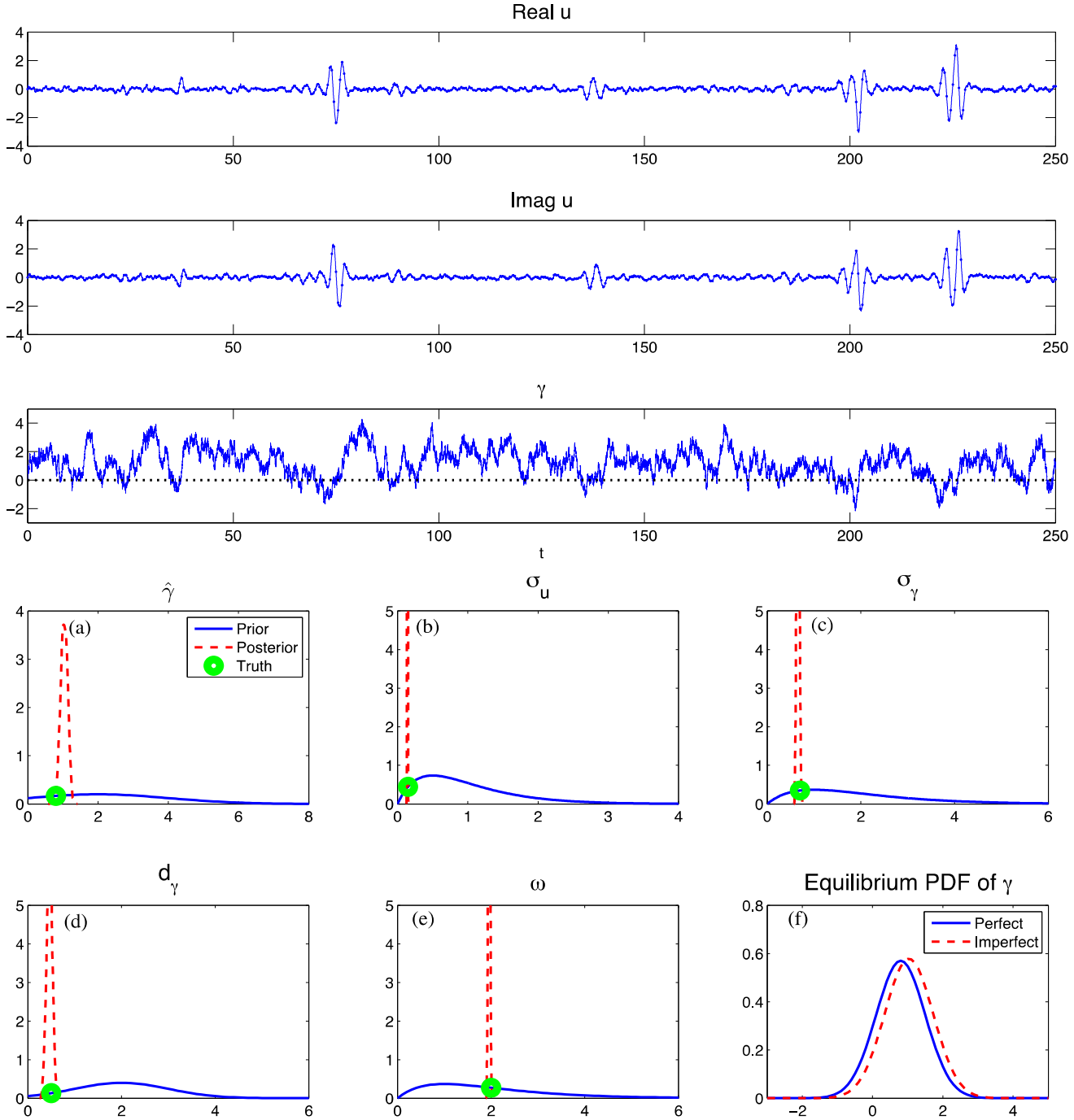


Figure II 4. Top panels: Training time series for parameter estimation of SPEKF models via the MCMC algorithm developed in [9]. $\boldsymbol{u(t)}$ *and* $\boldsymbol{\gamma(t)}$ *are the resolved and unresolved (augmented) variables, respectively. In this case,* $\boldsymbol{u(t)}$ *is complex, and* $\boldsymbol{\gamma(t)}$ *is a real stochastic damping so that intermittent instability in* $\boldsymbol{u(t)}$ *occurs whenever* $\boldsymbol{\gamma(t)}$ *is positive. Only observations of* $\boldsymbol{u(t)}$ *are used for parameter estimation via MCMC. The* $\boldsymbol{\gamma(t)}$ *process is hidden from the algorithm. Bottom panels: Parameter estimation performed via the new MCMC algorithm. Panels (a)–(e) show the prior distribution (solid lines), posterior distribution from MCMC (dashed lines) and the true value of the parameters (circles) respectively. Panel (f) compares the equilibrium probability density (PDF) of the unresolved variable of the true model and the model equipped with the maximum a posteriori estimates of the parameters. Note the accuracy in the posterior distribution of the parameters, which significantly exceeds the accuracy obtained via conventional MCMC algorithms with data augmentation.*

7. Nonparametric Uncertainty Quantification for Stochastic Gradient Flows

T. Berry (post doc, Penn State)
J. Harlim (Co-PI, Penn State)

This paper presents a nonparametric statistical modeling for quantifying uncertainties of stochastic gradient systems with isotropic diffusion. The central idea is to use the sampled data to construct a stochastic matrix whose generator is a discrete approximation to the backward Kolmogorov operator of the underlying dynamics with diffusion maps algorithm. In particular, we use the variable bandwidth kernels from previous paper.

The resulting eigenvectors of this stochastic matrix, which we will refer to as the diffusion coordinates, form an orthonormal basis of the given data set and represents a discrete approximation to the eigenfunctions of the Kolmogorov operator of the underlying dynamics. Given this basis, we consider a projection of three uncertainty quantification (UQ) problems (prediction, filtering, and response) on the diffusion coordinates. In these coordinates, the nonlinear prediction and response problems reduce to solving systems of infinite-dimensional linear ordinary differential equations whereas the continuous-time nonlinear filtering problem reduces to solving a system of infinite-dimensional linear stochastic differential equations. The resulting nonparametric UQ methods consist of solving the corresponding truncated linear systems of ordinary or stochastic differential equations in the finitely many diffusion coordinates. We numerically verify these new formulations on low-dimensional linear and nonlinear gradient flow systems.

III)

1. Predicting the cloud patterns of the Madden Julian Oscillaion through a low-order nonlinear stochastic model

Nan Chen (Ph.D. Courant)
A.J. Majda (Courant)
D. Giannakis (Courant)

A recently developed technique for nonlinear time series analysis NLSA [Giannakis and Majda, 2012a, b, 2013] has been utilized to de_ne two MJO indices of the boreal winter MJO for the large scale cloud patterns based only on OLR from the CLAUS data set without detrending or spatial-temporal filtering (Figure 1). The observed time series have non-Gaussian fat-tailed PDF's as a consequence of intermittency. Both systematic strategies for physics constrained regression models [Majda and Harlim,

2013; Harlim et al., 2014] and the dynamic stochastic skeleton model for the MJO [Majda and Stechmann, 2009, 2011; Thual et al., 2013] suggest a four dimensional stochastic model with two hidden variables representing stochastic damping and random phasing with energy conserving nonlinear feedback interaction. In a calibration phase, these models can successfully capture the observed non-Gaussian PDFs and autocorrelations (Figure 2). The models have a special structure which leads to efficient data assimilation and ensemble initialization algorithms for the hidden variables. The low-order nonlinear stochastic model has been applied to prediction of the OLR-based indices for boreal winter MJO's with forecasting skill up to 40 days in strong MJO years, 25 days in moderate MJO years and roughly 18 or 19 days in weak MJO years (Figure 3, 4 and 5); furthermore, the ensemble spread in the stochastic model has been shown to be an accurate predictive indicator of forecast uncertainty at long range (Figure 5). Perfect twin experiments with the stochastic model have comparable skill as with the observed data suggesting that the low-order nonlinear stochastic model has significant skill for determining the predicability limits of the large scale cloud patterns of the boreal winter MJO. All the techniques used here have been developed in this MURI proposal. Work by Nan Chen and Majda on applying these methods to the Wheeler-Hendon Index is underway.

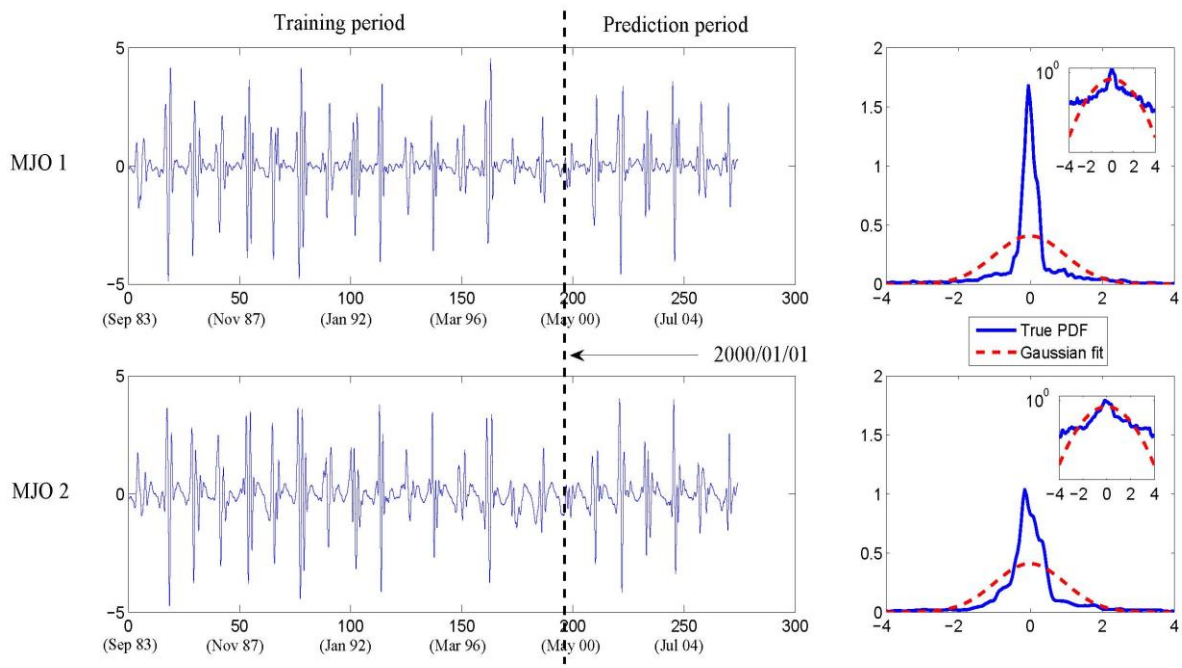


Figure 1. Left: MJO indices from NLSA (modes 8 and 9) ranging from 1983/09/03 to 2006/06/30. The time-series before 2000/01/01 is utilized as training period to get the statistics and that after 2000/01/01 represents the prediction period using the low-order stochastic model.
Right: The associated PDF of each index and the Gaussian _t. The small panel inside each subplot shows the PDF in the logarithm scale.

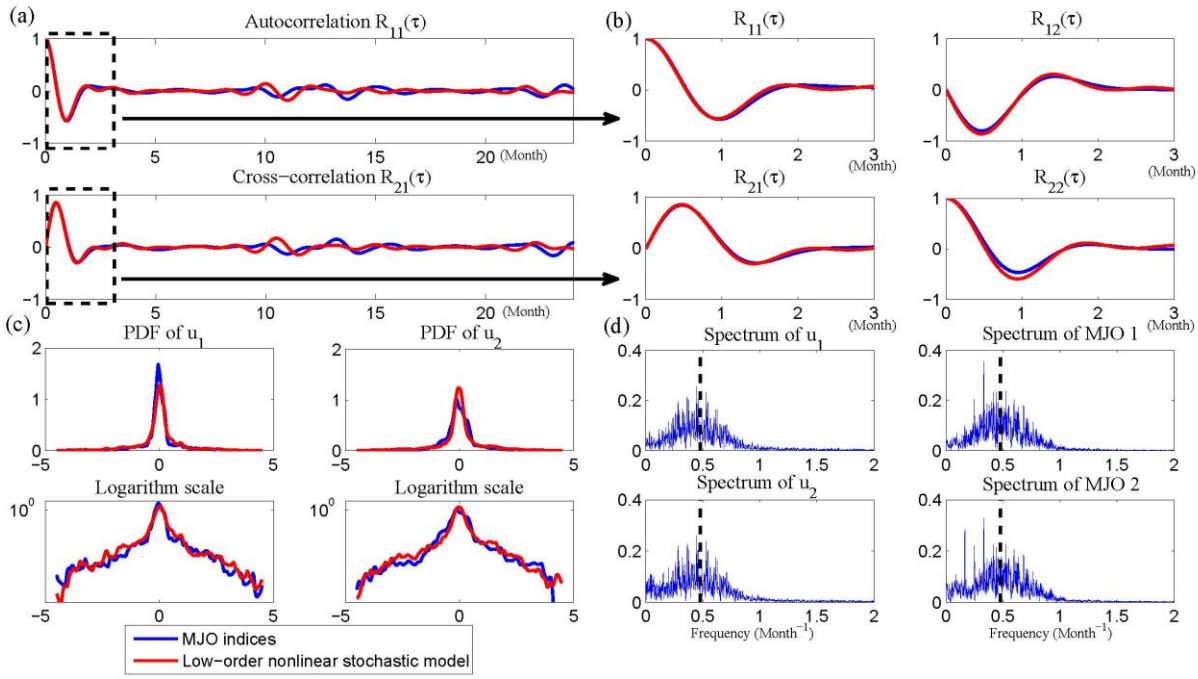


Figure 2. Statistics of the stochastic model with the optimal parameters in Table 1. (a) Long-term autocorrelation function R11(_) and cross-correlation function R21(_) from 0 to 24
months. (b) Short-term autocorrelation functions R11(_), R22(_) and cross-correlation functions R21(_), R12(_) from 0 to 3 months. (c) Equilibrium PDFs of the signal u1; u2 from stochastic model compared with that of the MJO indices. (d) Spectrum of u1; u2 compared with that of MJO indices. Here, the black dashed line indicates the frequency a=(2_).

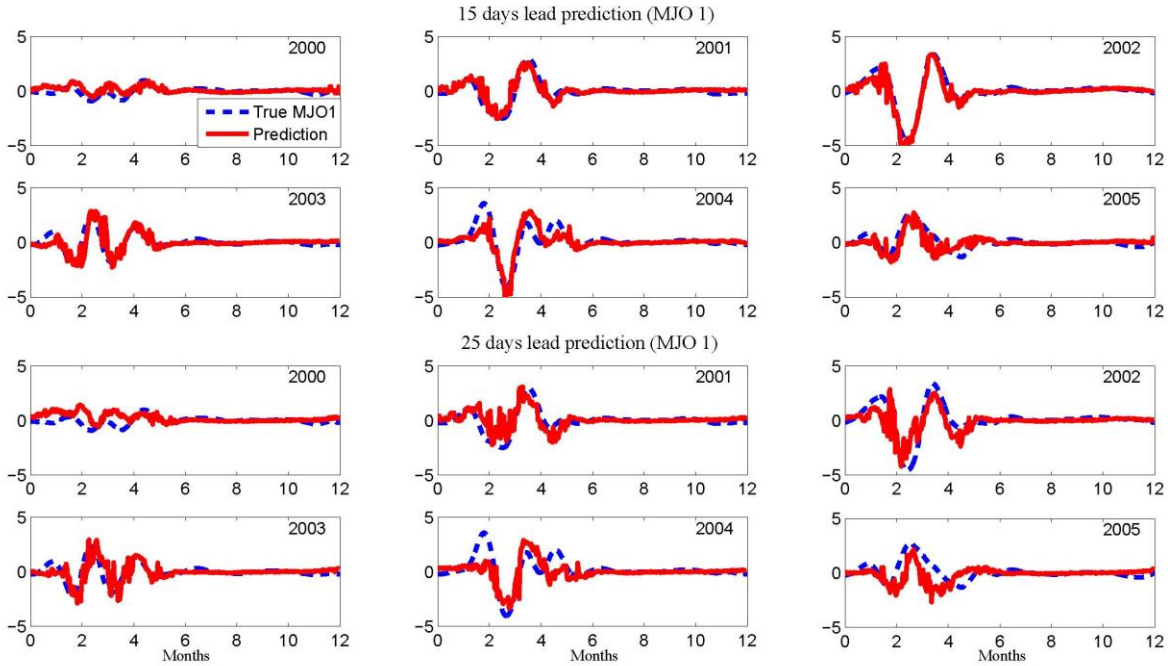


Figure 3. Prediction of MJO 1 at a 15 (top) and 25 (bottom) days lead. The blue line

shows the true signal and the red line shows the ensemble average of the predicted signal with 50 ensemble members.

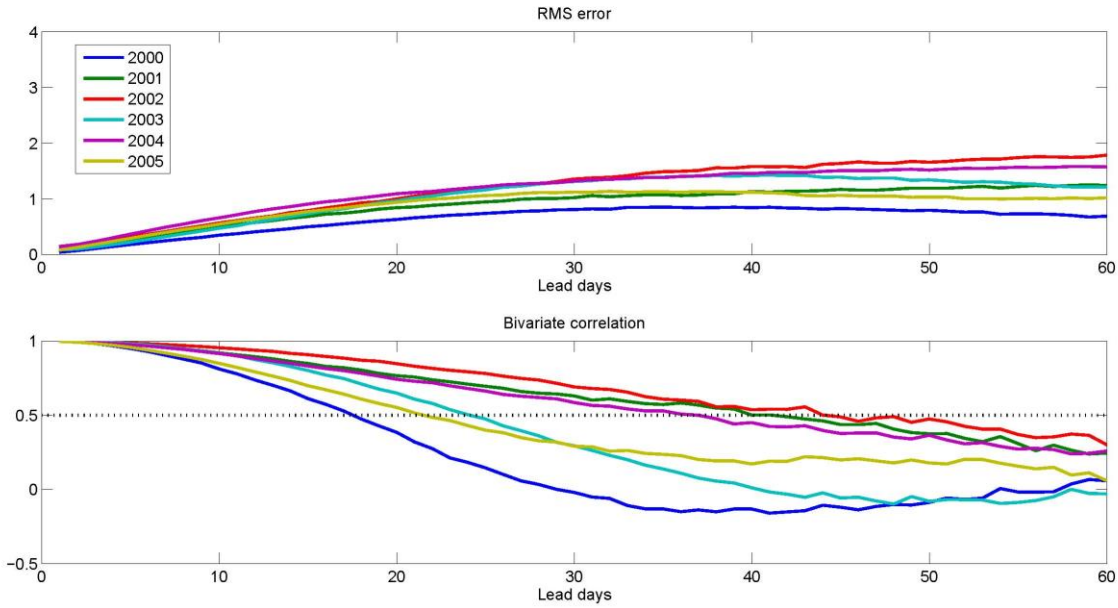


Figure 4. Skill scores with RMS error (top) and bivariate correlation (bottom) for prediction in different years.

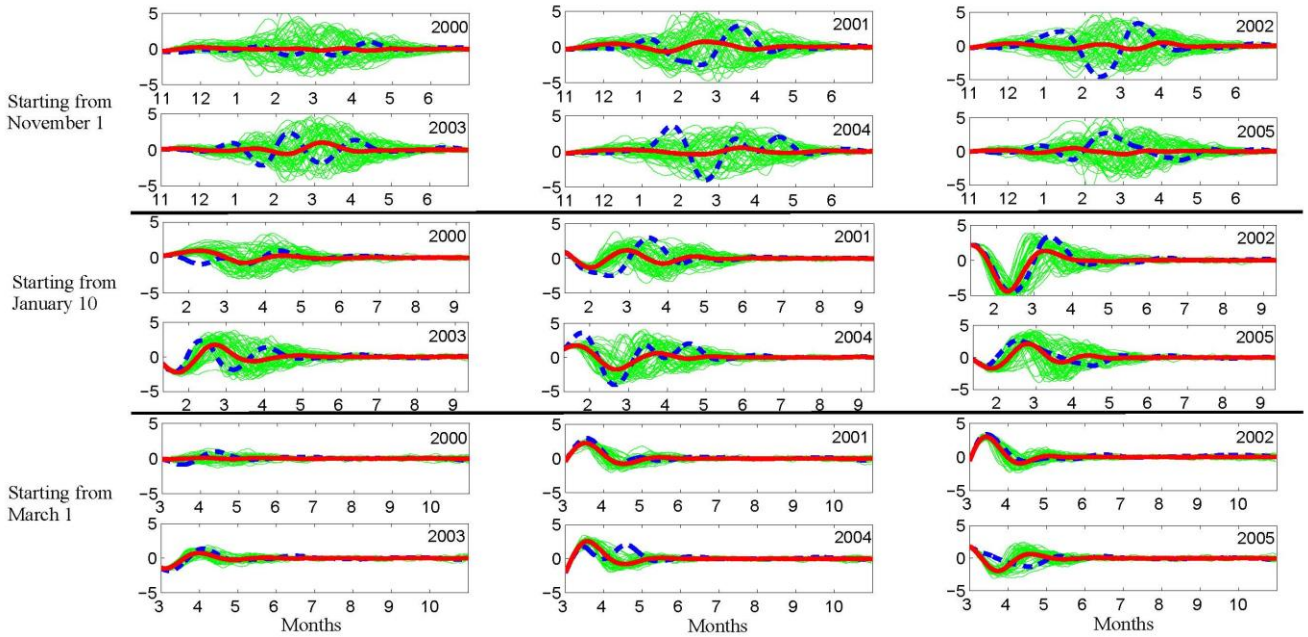


Figure 5. First and second rows: Prediction of MJO 1 starting from November 1 for different years. Each panel show the prediction skill of 8 months with the label in x-axis indicating the month. Third and forth rows: Same but starting from January 10. Fifth and Sixth rows: Same but starting from March 1. The thick blue dashed line is the MJO 1 index. The thick red solid line is the ensemble mean with 50 members, which are shown by the thin solid lines.

2. An ensemble kalman filter for statistical estimation of physics constrained nonlinear regression models

J. Harlim (Co-PI, Penn State)
A. Mahdi (unsupported collaborator, NC State)
A.J. Majda (PI, Courant Institute)

In this article, a new finite ensemble Kalman filtering algorithm is developed for estimating the state, the linear and nonlinear model coefficients, the model and the observation noise covariances from available partial noisy observations of the state.

Several stringent tests and applications of the method are developed here. In the most complex application, the perfect model has 57 degrees of freedom involving a zonal (east-west) jet, two topographic Rossby waves, and 54 nonlinearly interacting Rossby waves; the perfect model has significant non-Gaussian statistics in the zonal jet with blocked and unblocked regimes and a non-Gaussian skewed distribution due to interaction with the other 56 modes. We only observe the zonal jet contaminated by noise and apply the ensemble filter algorithm for estimation. Numerically, we find that a three dimensional nonlinear stochastic model with one level of memory mimics the statistical effect of the other 56 modes on the zonal jet in an accurate fashion, including the skew non-Gaussian distribution and autocorrelation decay. On the other hand, a similar stochastic model with zero memory levels fails to capture the crucial non-Gaussian behavior of the zonal jet from the perfect 57-mode model.

3. Conceptual dynamical models for turbulence

A.J. Majda (PI, Courant Institute)
Yoonsang Lee (Post doc, Courant Institute)

Understanding the complexity of anisotropic turbulent processes in engineering and environmental fluid flows is a formidable challenge with practical significance since energy often flows intermittently from the smaller scales to impact the largest scales in these flows. Conceptual dynamical models for anisotropic turbulence are introduced and developed here which, despite their simplicity, capture key features of vastly more complicated turbulent systems. These conceptual models involve a large scale mean flow and turbulent fluctuations on a variety of spatial scales with energy conserving wave-mean flow interactions as well as stochastic forcing of the fluctuations. Numerical experiments with a six dimensional conceptual dynamical model conform that these models capture key statistical features of vastly more complex anisotropic turbulent systems in a qualitative fashion. These features include chaotic statistical behavior of the mean flow with a sub-Gaussian probability distribution function (pdf) for its fluctuations while the turbulent fluctuations have decreasing energy and correlation times at smaller scales with nearly Gaussian pdfs for the large scale fluctuations and fat-tailed non-Gaussian pdfs for the smaller scale fluctuations. This last feature is a manifestation of intermittency of the small scale fluctuations where turbulent modes with small variance have relatively frequent extreme events which directly impact the mean flow. The dynamical modes introduced here potentially provide a useful test bed for algorithms for prediction, uncertainty quantification, and data assimilation for anisotropic turbulent systems.

IV)
1. Information Barriers for Noisy Lagrangian Tracers in Filtering Random Incompressible Flows

N. Chen (Ph. D student)
A. Majda
X. T. Tong (Post Doc, Courant Institute)

An important practical problem is the recovery of a turbulent velocity field from Lagrangian tracers that move with the fluid flow. Here the filtering skill of L moving Lagrangian tracers in recovering random incompressible flow fields defined through a finite number of random Fourier modes is studied with full mathematical rigor. Despite the inherent nonlinearity in measuring noisy Lagrangian tracers, it is shown below that there are exact closed analytic formulas for the optimal filter for the velocity field involving Riccati equations with random coefficients for the covariance matrix. This mathematical structure allows a detailed asymptotic analysis of filter performance both as time goes to infinity and as the number of noisy Lagrangian tracers, L, increases. In particular, the asymptotic gain of information from L-tracers grows only like lnL in a precise fashion, i.e., an exponential increase in the number of tracers is needed to reduce the uncertainty by a fixed amount; in other words, there is a practical information barrier. The proofs proceed through a rigorous mean field approximation of the random Ricatti equation. Also, as an intermediate step, geometric ergodicity with respect to the uniform measure on the period domain is proved for any fixed number L of noisy Lagrangian tracers. All of the above claims are confirmed by detailed numerical experiments presented here.

2. Noisy Lagrangian Tracers for Filtering Random Rotating Compressible Flows
N. Chen (Ph. D student)
A. Majda
X. T. Tong (Post Doc, Courant Institute)

The recovery of a random turbulent velocity field using Lagrangian tracers that move with the fluid flow is a practically important problem of intense interest in oceanography and has wide interest for the Navy. This paper studies the filtering skill of L noisy Lagrangian tracers in recovering random rotating compressible flows that are a linear combination of random incompressible geostrophically balanced (GB) flow and random rotating compressible gravity waves. The idealized random fields are defined through forced damped random amplitudes of Fourier eigenmodes of the rotating shallow water equations with the rotation rate measured by the Rossby number ε ". In many realistic geophysical flows, there is fast rotation so ε satisfies $\varepsilon \ll 1$ and the random rotating shallow water equations become a slow-fast system where often the primary practical objective is the recovery of the GB component from the Lagrangian tracer observations. Unfortunately, the L-noisy Lagrangian tracer observations are highly nonlinear and mix the slow GB modes and the fast gravity modes. Despite this inherent nonlinearity, it is shown here that there are closed analytical formulas for the optimal filter for recovering these random rotating compressible flows for any ε involving Ricatti equations with random coefficients. The performance of the optimal filter is compared and contrasted through mathematical theorems and concise numerical experiments with the performance of the optimal filter for the incompressible GB random flow with L noisy Lagrangian tracers involving only the GB part of the flow. In addition, a sub-optimal filter is defined for recovering the GB flow alone through observing the L noisy random compressible Lagrangian trajectories, so the effect of the gravity wave dynamics is unresolved but effects the tracer observations. Rigorous theorems proved below through suitable stochastic fast-wave averaging techniques and explicit formulas rigorously demonstrate that all these filters have comparable skill in recovering the slow GB flow in the limit $\varepsilon \to 0$ for any bounded time interval. Concise numerical experiments confirm the mathematical theory and elucidate

various new features of filter performance as the Rossby number ε, the number of tracers L and the tracer noise variance change.

3. Quantifying Bayesian Filter Performance for Turbulent Dynamical Systems through Information Theory

M. Branicki (U. Edinburgh, unsupported)
A.J. Majda (Courant)

This work develops families of information measures involving the entropy of the forecast error (a generalization of RMS error), the mutual information (a generalization of anomaly pattern correlation), and the relative entropy (a nonlinear statistical measure of forecast fidelity) to cope with imperfect model errors in filtering and prediction. All three measures are optimized by the Kalman filter for perfect models but emphasize different properties of the filter or forecast distribution for imperfect models. This paper introduces a new optimization principle over parameters for imperfect models which combines all three probabilistic measures and incorporates both means and covariances of the imperfect model in assessing skill. This tool should have wide use for the training phase of low order imperfect models for data assimilation and prediction as well as for targeted statistics in more complex models.

4. Linear Theory for Filtering Nonlinear Multiscale Systems with Model Error

T. Berry (post doc, Penn State)
J. Harlim (Co-PI, Penn State)

We studied filtering with model error on simple low-order stochastic dynamical systems. Here, we rigorously showed that for a two-scale linear systems of SDE's, given continuous-time direct observations of the slow components, there exists a unique reduced stochastic model that simultaneously produces optimal filtering and climatological statistical prediction. The optimality here is in the sense that the mean and covariance of the reduced model are close to the corresponding solutions with perfect model. We also found that it is possible to extend this linear theory to nonlinear problems as long as the parametric form for the model error is known. When the appropriate parametric form is not available, parameters chosen for good filtering may give poor climatological statistical estimates and vice versa. When the appropriate parametric form is known, it is more natural to estimate the parameters simultaneously and to account for the nonlinear feedback of the stochastic parameters into the reduced filter estimates. This study confirmed that the unsatisfactory climatological statistical prediction in the presence of model error (as numerically found in the first project above) is completely due to the difficulty of choosing the appropriate parametric form for nonlinear regression modeling.

5. Ensemble Kalman Filters for Dynamical Systems with Unresolved Turbulence

I. Grooms (unsupported, Courant),
Yoonsang Lee (Post doc, Courant Institute)
A.J. Majda (PI, Courant Institute)

Ensemble Kalman filters are developed for turbulent dynamical systems where the forecast model does not resolve all the active scales of motion. Coarse-resolution models are intended to predict the large-scale part of the true dynamics, but observations invariably include contributions from both the resolved large scales and the unresolved small scales. The error due to the contribution of unresolved

scales to the observations, called ‘representation’ or ‘representativeness’ error, is often included as part of the observation error, in addition to the raw measurement error, when estimating the large-scale part of the system. It is here shown how stochastic superparameterization (a multiscale method for subgridscale parameterization) can be used to provide estimates of the statistics of the unresolved scales. In addition, a new framework is developed wherein small-scale statistics can be used to estimate both the resolved and unresolved components of the solution.
The one-dimensional test problem from dispersive wave turbulence used here is computationally tractable yet is particularly difficult for filtering because of the non-Gaussian extreme event statistics and substantial small scale turbulence: a shallow energy spectrum proportional to $k^{-5/6}$ (where k is the wavenumber) results in two-thirds of the climatological variance being carried by the unresolved small scales. Because the unresolved scales contain so much energy, filters that ignore the representation error fail utterly to provide meaningful estimates of the system state. Inclusion of a time-independent climatological estimate of the representation error in a standard framework leads to inaccurate estimates of the large-scale part of the signal; accurate estimates of the large scales are only achieved by using stochastic superparameterization to provide evolving, large-scale dependent predictions of the small-scale statistics. Again, because the unresolved scales contain so much energy, even an accurate estimate of the large-scale part of the system does not provide an accurate estimate of the true state. By providing simultaneous estimates of both the large- and small-scale parts of the solution, the new framework is able to provide accurate estimates of the true system state.

6. Blended Particle Filters for Large Dimensional Chaotic Dynamical Systems

A.J. Majda (PI, Courant Institute)
D. Qi (Post doc, Courant Institute)
T. Sapsis (MIT, unsupported)

A major challenge in contemporary data science is the development of statistically accurate particle filters to capture non-Gaussian features in large-dimensional chaotic dynamical systems. Blended particle filters that capture non-Gaussian features in an adaptively evolving low-dimensional subspace through particles interacting with evolving Gaussian statistics on the remaining portion of phase space are introduced here. These blended particle filters are constructed in this paper through a mathematical formalism involving conditional Gaussian mixtures combined with statistically nonlinear forecast models compatible with this structure developed recently with high skill for uncertainty quantification. Stringent test cases for filtering involving the 40-dimensional Lorenz 96 model with a 5-dimensional adaptive subspace for nonlinear blended filtering in various turbulent regimes with at least nine positive Lyapunov exponents are used here. These cases demonstrate the high skill of the blended particle filter algorithms in capturing both highly non-Gaussian dynamical features as well as crucial nonlinear statistics for accurate filtering in extreme filtering regimes with sparse infrequent high-quality observations. The formalism developed here is also useful for multiscale filtering of turbulent systems and a simple application is sketched below.

7. Multi-scale Methods for Data Assimilation in Turbulent Systems

Yoonsang Lee (Post doc, Courant Institute)
A.J. Majda (PI, Courant Institute)

Data assimilation of turbulent signals is an important challenging problem because of the extremely complicated large dimension of the signals and incomplete partial noisy observations which usually mix the large scale mean flow and small scale fluctuations. Due to the limited computing power in the foreseeable future, it is desirable to use multi-scale forecast models which

are cheap and fast to mitigate the curse of dimensionality in turbulent systems; thus model errors from imperfect forecast models are unavoidable in the development of a data assimilation method in turbulence. Here we propose a suite of multi-scale data assimilation methods which use stochastic superparameterization as the forecast model. Superparameterization is a seamless multi-scale method for parameterizing the effect of small scales by cheap local problems embedded in a coarse grid. The key ingredient of the multi-scale data assimilation methods is the systematic use of conditional Gaussian mixtures which make the methods efficient by filtering a subspace whose dimension is smaller than the full state. The multi-scale data assimilation methods proposed here are tested on a six dimensional conceptual dynamical model for turbulence which mimics interesting features of anisotropic turbulence including two way coupling between the large and small scale parts, intermittencies, and extreme events in the smaller scale fluctuations. Numerical results show that suitable multi-scale data assimilation methods have high skill in estimating the most energetic modes of turbulent signals even with infrequent observation times.

IMPACT/APPLICATIONS

1. The development of the stochastic skeleton model, the observational skeleton index, and the observatorial study of the MJO termination events pave the way for the practical development of the stochastic MJO skeleton model as a predictive tool for the MJO in nature. This next stage of the MURI project is currently underway with the collaboration of the UCLA group (Stachnik, Waliser), Courant (Thual, Majda), and Wisconsin (Stechmann).
2. The development of data driven methods for analog forecasting combined with advanced nonlinear time series techniques (NLSA) for large dimensional data sets has great promise as a new predictive tool for central short term climate problems such as the prediction of sea ice reemergence.
3. The class of physics constrained low order stochastic models for predicting the cloud patterns of the MJO hold great promise as tools to assess the predictability and its limits for the Wheeler-Hendon index, monsoon intraseasonal oscillations, etc. This is a priority of the Courant group (Chen, Majda, Giannakis)
4. The existence of transparent and analyzable information barriers in data assimilation of noisy Lagrangian tracers is a highly significant result. The two papers on noisy Lagrangian tracers comprise the first rigorous mathematical analysis on this important practical topic. The development of genuinely multi-scale methods for data assimilation of unresolved turbulence could have major impact pn pratical data assimilation of mesocale and submesoscale ocean dynamics as well as tropical weather prediction.

RELATED PROJECTS

The JIFRESSE MJO research group has worked on three related lines of research that have benefited from the MURI support and research agenda. The first, with primary funding from ONR, is an examination of the relationship between convectively-coupled Kelvin waves and the MJO, which has important implications of the large-scale modulation of Kelvin waves, particularly over the South America-Atlantic-Africa sector, with possible implications for MJO initation in the Indian Ocean (Guo et al. 2014). The second, with primary funding from NOAA, is the first comprehensive estimate of MJO predictability and prediction skill based a multi-model collection of hindcasts, in this case from the Intraseasonal Variability Hindcast Experiment (ISVHE) (Mani et al. 2014). The third, with primary

funding from NSF, is a comprehensive assessment of MJO simulation fidelity from a set of 28 contemporary weather/climate models participating in the WCRP-WWRP/THORPEX MJO Task Force and GEWEX GASS mulit-model experiment on physical processes and the MJO (Jiang et al. 2014).

REFERENCES

I)

Majda, A. J., and S. N. Stechmann, 2009: The skeleton of tropical intraseasonal oscillations. Proc. Natl. Acad. Sci. 106, 8417-8422.

II)

D. Giannakis and A. J. Majda (2012), "Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability", Proc. Natl. Acad. Sci., 109(7), 2222-2227

E. Blanchard-Wrigglesworth, K. C. Armour, C. M. Bitz, and E. DeWeaver, "Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations", J. Climate, 24, 231–250.

K. H. Straub (2013). "MJO initiation in the real-time multivariate MJO index", J. Climate, 26, 1130–1151.

D. Giannakis and A. J. Majda (2012), "Quantifying the predictive skill in long-range forecasting. Part I: Coarse-grained predictions in a simple ocean model", J. Climate, 25, 1793-1813.

D. Giannakis, A. J. Majda, and I. Horenko (2014), "Information theory, model error, and predictive skill of stochastic models for complex nonlinear systems", Phys. D, 24, 1735-1752.

E. N. Lorenz (1969), "Atmospheric predictability as revealed by naturally occurring analogues", J. Atmos. Sci., 23, 636-646.

R. R. Coifman and S. Lafon (2006), "Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions", J. Appl. Comput. Harmon. Anal., 21, 31-52.

B. Gershgorin, J. Harlim, and A.J. Majda (2010), "Test models for improving filtering with model errors through stochastic parameter estimation", J. Comput. Phys., 229, 1-31.

PUBLICATIONS

I)

Stachnik, J.P., D.E. Waliser, and A.J. Majda, 2014Presursor Environmental Conditions Associated with Termination of Madden-Julian Oscillation Events. J. Atmos Sci. [submitted, refereed]

Thual, S., A. J. Majda, and S. N. Stechmann, 2014: A stochastic skeleton model for the MJO. *J. Atmos. Sci.* **71**, 697-715. [published, refereed]

Thual, S., A. J. Majda, and S. N. Stechmann, 2014: Asymmetric intraseasonal events in the stochastic skeleton MJO model with seasonal cycle. *Climate Dynamics.* [in press, refereed]

Stechmann, S. N., and A. J. Majda, 2014: Identifying the Madden-Julian Oscillation skeleton in observational data. *Monthly Weather Review.* [submitted, refereed]

Stechmann, S. N., and H. R. Ogrosky, 2014: The Walker circulation, diabatic heating, and outgoing longwave radiation. *Geophysical Research Letters.* [submitted, refereed}

Chen, S., and S. N. Stechmann, 2014: Nonlinear traveling waves for the skeleton of the Madden-Julian Oscillation. *Communications in Mathematical Sciences.* [submitted, refereed]

Yang, Q. and A.J. Majda, 2014: A multi-scale model for the intraseasonal impact of the diurnal cycle of tropical convection (Theoret. Comp. Fluid Dyn. [submitted, refereed]

II)

M. Bushuk, D. Giannakis, and A. J. Majda (2014), "Reemergence mechanisms for North Pacific sea ice revealed through nonlinear Laplacian spectral analysis", *J. Climate*, 27, 6265-6287 [published, refereed].

M. Bushuk, D. Giannakis, and A. J. Majda (2014), "Arctic sea ice reemergence: The role of large-scale oceanic and atmospheric variability", *J. Climate* [submitted, refereed].

W.-w. Tung, D. Giannakis, and A. J. Majda (2014), "Symmetric and antisymmetric signals in MJO deep convection. Part I: Basic modes in infrared brightness temperature", *J. Atmos. Sci.*, 71, 3302-3326 [published, refereed].

W.-w. Tung, D. Giannakis, and A. J. Majda (2014), "Symmetric and antisymmetric signals in MJO deep convection. Part II: Kinematics and thermodynamics", *J. Atmos. Sci.* [submitted, refereed].

E. Szekely, D. Giannakis, and A. J. Majda (2014), "Extraction and predictability of coherent intraseasonal signals in infrared brightness temperature data", *Climate Dyn.* [submitted, refereed].

E. Szekely, D. Giannakis, and A. J. Majda (2014), "Extraction and predictability of coherent intraseasonal signals in infrared brightness temperature data", *14th International Workshop on Climate Informatics,* National Center for Atmospheric Research, Boulder, Colorado, September 25-26, 2014 [in press, refereed].

J. Zhao and D. Giannakis (2014), "Kernel Methods for Analog Forecasting", *Nonlinearity* [submitted, refereed].

D. Giannakis (2014), "Dynamics-adapted cone kernels", *SIAM J. Appl. Dyn. Syst.* [submitted, refereed].
N. Chen, D. Giannakis, R. Herbei, and A. J. Majda (2014), "An MCMC algorithm for parameter estimation in signals with hidden intermittent instability", *SIAM/ASA J. Uncertainty Quantification* [in press, refereed].

T. Berry and J. Harlim, 2014, Variable Bandwidth Diffusion Kernels, Applied and Computational Harmonics Analysis (2014), [submitted].

T. Berry and J. Harlim, 2014, Nonparametric Uncertainty Quantification for Stochastic Gradient Flows, SIAM/ASA J. Uncertainty Quantification [submitted]. [Refereed]

III)

(N. Chen, A. Majda, D. Giannakis) "Predicting the cloud patterns of the Madden-Julian Oscillation through a low-order nonlinear stochastic model", GRL, August 2014, DOI: 10.1002/2014GL060876 [published, refereed]

(J. Harlim, A. Mahdi, A Majda) "An Ensemble Kalman Filter for Statistical Estimation of Physics Constrained Nonlinear Regression Models", J. Comp. Phys., January 2014 doi:10.1016/j.jcp.2013.10.025 [published, refereed]

(A. Majda, Y. Lee) "Conceptual dynamical model for turbulence", PNAS, vol. 111 no. 18, March 2014 6548–6553, doi: 10.1073/pnas.1404914111 [published, refereed]

IV)

(N. Chen, A. Majda, X. Tong) "Information Barriers for Noisy Lagrangian Tracers in Filtering Random Incompressible Flows", Nonlinearity 27 pg 2133-2163, August 2014, DOI 10.1088/0951-7715/27/9/2133 [published, refereed]

(N. Chen, A. Majda, X. Tong) "Noisy Lagrangian Tracers for Filtering Random Rotating Compressible Flows" submitted, J. Nonlinear Sci., April 2014 [refereed]

(Majda, M. Branicki) "Quantifying Bayesian Filter Performance for Turbulent Dynamical Systems through Information Theory", Comm. Math. Sci., 2014 DOI: 10.4310/CMS.2014.v12.n5.a6 [published, refereed]

(T. Berry, J. Harlim) "Linear Theory for Filtering Nonlinear Multiscale Systemes with Model Error", Proc. R. Soc. A (2014) 470, 20140168 [arXiv] [published, refereed]

(I. Grooms, Y. Lee, A. Majda) "Ensemble Kalman Filters for Dynamical Systems with Unresolved Turbulence", J. Com. Phys, October 2013, DOI: 10.1016/j.jcp.2014.05.037 [published, refereed]

(Majda, D. Qi, T. Sapsis) “Blended Particle Filters for Large Dimensional Chaotic Dynamical Systems”, pgs 7511–7516, PNAS, May 2014 doi: 10.1073/pnas.1405675111 [published, refereed]

(Majda, Y. Lee) "Multi-scale Methods for Data Assimilation in Turbulent Systems”, SIAM J. MMS, July 2014 [submitted, refereed]

Guo, Y., X. Jiang, D.E. Waliser, 2014: Modulation of the Convectively Coupled Kelvin Waves over South America and the Tropical Atlantic Ocean in Association with the Madden–Julian Oscillation. *J. Atmos. Sci.*, 71, 1371–1388. doi: http://dx.doi.org/10.1175/JAS-D-13-0215.1. [pubished, referreed]

Jiang, X., D. E. Waliser, P. K. Xavier, J. Petch, N. P. Klingaman, S. J. Woolnough, B. Guan, G. Bellon, T. Crueger, Charlotte DeMott, C. Hannay, H. Lin, W. Hu, D. Kim, C.-L. Lappen, M.-M. Lu, H.-Y. Ma, T. Miyakawa, J. A. Ridout, S. D. Schubert, J. Scinocca, K.-H. Seo, E. Shindo, X. Song, C. Stan, W.-L. Tseng, W. Wang, T. Wu, K. Wyser, X. Wu, G. J. Zhang, and H. Zhu (2014), Vertical Structure and Physical Processes of the Madden-Julian Oscillation: Exploring Key Model Physics in Climate Simulations, *J. Geophys. Res. Atmos.* [submitted, in review]

Mani, N. J., J. Y. Lee, D.E. Waliser, B. Wang, and X. Jiang, 2014: Predictability of the Madden-Julian Oscillation in the Intraseasonal Variability Hindcast Experiment (ISVHE). *J. Climate*, 10.1175/JCLI-D-13-00624.1. [pubished, referreed]

HONORS/AWARDS/PRIZES

1. PI, Andrew J. Majda will receive the 2015 Lagrangian Prize awarded every four years by the International Congress of Industrial and Applied Mathematics, for significant research contributions throughout his career. The citation even includes reference to topics of his research in the MURI award.
2. CO-PI, Sam Stechmann has been awarded a Sloan Foundation Fellowship honoring his research in applied mathematics
3. Co-PI Duane Waliser has been elected as a 2014 Fellow of the American Meteorology Society
4. MURI Postdoc Eniko Szekely received a travel award and recognition for outstanding paper submission by an early-career scientist (see Publication [6]) at the 14th International Workshop on Climate Informatics, held at the National Center for Atmospheric Research, NCAR, at Boulder, Colorado on September 25-26, 2014.
5. PhD student Mitchell Bushuk received the Moses A. Greenfield Research Award at the Courant Institute for the research published in [1].

STATISTICS

Graduate students 4
Total Postdocs 10
Total Women post docs 3
Total Minority postdocs 0

BEST ACCOMPLISHMENTS

I) Development of Stochastic skelton model andcomparison with observations
II) Comprehensive Nonlinear Data Analysis for Teleconnections among basins in Sea Ice Reemergence
III) Skillfully Predicting the Cloud Patterns of the MJO through a nonlinear low order stochastic model
IV) First rigorous mathematical theory for filering random incompressible and compressible flows by noisy Langrangian tracers.