Closeout of Award No. DE-FG03-01ER63260, fund #22414

Final Report

Title: Predicting Coupled Ocean-Atmosphere Modes with a Climate Modeling Hierarchy
Project ID: 0008263
PI: M. Ghil (UCLA) and Andrew W. Robertson (IRI)
Award Register#: ER63413 0008263

Developments of a family of Hidden Markov Models (HMMs)

Nonhomogeneous Hidden Markov Models (NHMMs) have been developed to model local daily observed rainfall on a regional network of raingauge-stations, and to study the ``downscalability'' of climate variability to local daily scales. As part of this work we have developed statistical estimation techniques, parameter fitting algorithms, and software in C code, for a general class of NHMMs. Our algorithms and software generalize earlier work by Hughes and Guttorp (1994) and Hughes, Guttorp, and Charles (1999), who originally proposed the application of NHMMs to downscaling. In general, an NHMM is characterized by three modeling components: (1) an unobserved finite-state discrete-time Markov process S(t), (2) a vector of observations R(t) at time t that has a stochastic dependence on the Markov state S(t), and (3) an exogenous set of measurements X(t) that modulate the state transition probabilities.

We have developed and implemented a general Expectation-Maximization (EM) procedure that can efficiently estimate the parameters of a family of NHMMs, that includes the following variations (in addition to more-basic mixture models with no explicit time dependence): (1) Binary outputs that can be modeled as conditionally independent given the state; (2) Binary outputs that can be modeled as having a joint tree-dependence multivariate structure (to capture spatial dependence), conditioned on each state; (3) Real-valued outputs that can be modeled as being multivariate Gaussian conditioned on the state, and independent of all other variables in the model; (4) Outputs that can be modeled as having autoregressive (or Markov) dependence on the output vector from the previous day in addition to dependence on the state; (5) State transition probabilities that are parameterized via a logistic function, such that the Markov chain transition probabilities are now inhomogenous in time by virtue of the dependence on X(t).

The EM algorithms developed allow these variations to be combined in multiple different ways within one general estimation procedure. Special cases of the models include both ``stateless" models with Markov dependence between rain stations and the NHMM of Hughes and Guttorp with conditionally independent outputs. More generally this allows models that allow hidden states, inhomogeneous Markov transitions on the states, and a variety of spatial and dependence structures on the outputs (the rain stations). Software in C code has been developed and tested. The software has a cross-validation mode that allows for repeatedly partitioning the historical data into data for training and data for testing, where the model is fit to the training data and then prediction metrics are calculated on the test data; the average prediction metrics over all such test data sets are then reported as the ``cross-validated" estimates, allowing one (for example) to fairly compare models with different complexities.

The C code used to implement this algorithm has been tested and validated against the original Fortran code of Hughes and Guttorp for NHMMs. The new C code is more general and flexible than the original Fortran implementation, and has been found to be significantly faster at run time, providing the opportunity to explore more modeling options and fit larger data sets.

The estimation and prediction parts of the algorithm can also handle missing data, which are quite common in historical daily rainfall records, particularly in regions such as Africa. The optimal approach is to infer a probability distribution over the missing values given the rest of the observed data and the model, and the EM algorithm provides the mechanism for making these inferences. The software we have developed can handle arbitrary patterns of missing rainfall observations using EM and, to our knowledge, is the first full-scale implementation of an optimal EM algorithm for this purpose.

Application of HMMs to GCM downscaling

The HMM has been used to investigate daily rainfall occurrence at ten gauge stations in northeast Brazil during the February-April wet season 1975-2002. The model assumes that rainfall occurrence is governed by a few discrete states, with Markovian daily transitions between them. Four hidden rainfall states are identified. One pair of the states represents wet vs. dry conditions at all stations, while a second pair of states represents north-south gradients in rainfall occurrence. The estimated daily state-sequence is characterized by a systematic seasonal evolution, together with considerable variability on intraseasonal, interannual and longer time scales. The first pair of states are shown to be associated with large-scale displacements of the tropical convergence zones, and with teleconnections typical of the El Nino-Southern Oscillation and the North Atlantic Oscillation. A trend toward greater rainfall occurrence in the north of Ceara compared to the south since the 1980s is identified with the second pair of states.

A non-homogeneous HMM (NHMM) was then used to downscale daily precipitation occurrence at the ten stations, using GCM simulations of seasonal-mean large-scale precipitation, obtained with historical sea surface temperatures prescribed globally. Interannual variability of the GCM's large-scale precipitation simulation is well correlated with seasonal- and spatial-averaged station rainfall-occurrence data. Simulations from the NHMM are found to be able to reproduce this relationship. The GCM-NHMM simulations are also able to capture quite well interannual changes in daily rainfall occurrence and 10-day dry spell frequencies at some individual stations. This work has demonstrated that the NHMM provides a useful tool (a) to understand the statistics of daily rainfall occurrence at the station level in terms of large-scale atmospheric patterns, and (b) to produce station-scale daily rainfall sequence scenarios for input into crop models etc.

Intermediate coupled model results and their relation to observed climate variability

We have considered two groups of intermediate coupled models: dynamical models that are stripped-down versions of GCMs, and statistical predictive (inverse) models, whose propagator is derived directly from the observed data.

Dynamical models

Our research has concentrated on exploring two possible mechanisms of decadal climate variability: (i) interaction between mesoscale oceanic eddies associated with nonlinear dynamics of wind-driven circulation, and midlatitude atmospheric jet stream; (ii) interaction between the oceanic thermohaline circulation (THC) and sea ice through tropically forced changes of the middle- and high-latitude atmospheric circulation.

A. Midlatitude coupled modes. We have coupled a high horizontal resolution ocean and atmosphere quasi-geostrophic models through a diagnostic oceanic mixed layer. The atmospheric behavior in the uncoupled runs, forced by spatially varying, but constant in time sea-surface temperature (SST), is characterized by the existence of bimodality in the modeled zonal-mean flow for realistic values of surface friction. The dominant atmospheric low-frequency variability (LFV) in this regime consists of irregular transitions between the high-latitude and low-latitude jet states. In the coupled run, where SST is allowed to respond to and, in turn, cause changes in the oceanic and atmospheric flow, the transitions between the two states are regularized and lead to a coupled decadal cycle, where the atmospheric sensitivity to SST is due to the bimodal nature of the modeled midlatitude jet, while the time scale of the oscillation is determined by nonlinear adjustment of the oceanic inertial recirculations.

B. Coupled THC--sea-ice--atmosphere behavior. The intrinsic THC variability is explored in a coarse resolution primitive-equation ocean model coupled to a thermodynamic sea-ice model and an energy-balance model of atmosphere; stochastic forcing is also included to represent the effect of intrinsic atmospheric variability. Two types of interdecadal oscillations are identified. The first type is characterized by a stationary dipolar SST pattern in the northwestern corner of the ocean basin and westward-propagating temperature anomalies at depth. The second type involves larger-scale westward-propagating temperature anomalies in the upper and deep ocean. Both oscillations exist in a model where the sea ice is artificially suppressed. Sea ice tends to damp both modes due to a phase-transition effect, by which SST in the sea-ice region is bound to be close to the freezing temperature. Despite that, both modes re-appear subject to stochastic atmospheric forcing. We have shown that they are associated with the least-damped eigenmodes of the linearized model.

C. Relation to observed climate variability. The nonlinear nature of observed atmospheric LFV was addressed by considering the NCEP/NCAR Northern Hemisphere zonal wind data sets. Zonal averaging is essential to concentrate on eddy-driven multiple states of the midlatitude jet stream, rather than on topographically driven barotropic modes often considered previously. We have found multiple regimes of atmospheric circulation characterized by the midlatitude jet shifted to the north and to the south of its time-mean position, as well as ultra-low-frequency intraseasonal oscillations associated with this regime, similar to those found in the Southern Hemisphere observations. The existence of both multiple regimes and low-frequency oscillations are predicted by our idealized atmospheric model (A) in the relevant range of the surface-stress values. In addition, the statistics of the observed regimes are correlated with SST patterns in the North Atlantic ocean. The associated SST indices are correlated with the regimes' occurrence frequency time series at multi-year lags (not shown), suggesting coupled behavior. One type of interdecadal variability is characterized by a spatial pattern resembling the one obtained in our coupled model with wind-driven ocean (A). The other pattern resembles the one obtained in our hybrid coupled model (B).

Inverse stochastic models

Isolating coupled ocean-atmosphere modes from ICM and GCM simulations or observed data is not trivial. Inverse models are derived directly from observed or simulated (ICM or GCM) data and are assessed by their ability to forecast climate anomalies. We have used the concept of linear inverse models (LIMs) as a basis of our analysis. LIMs are constructed in the phase space of the data's leading empirical orthogonal functions (EOFs). They assume a linear dependence of climate tendency on climate-state variables and parameterize unresolved processes as a spatially correlated white in time noise. The model's propagator and the noise covariance matrix are estimated from the data's statistics by multiple linear regression (MLR). We have developed generalizations of such models that relax the assumptions of linearity and white noise, since the two assumptions are generally not satisfied for the atmospheric data. Our novel procedure relies on a regularized multiple regression and results in a construction of a nonlinear model that faithfully represents the data's statistics and generally outperforms linear model in terms of forecast skill.

We have built a nonlinear inverse stochastic model based on Northern Hemisphere geopotential height data. It captures well the daily midlatitude atmosphere's time dependence and non-Gaussian distribution in the phase space. The behavior is characterized by the existence of multiple planetary flow regimes like those identified in an intermediate dynamical model. We have also applied the inverse model to global SST data, and found it to perform similarly or better than all alternative statistical and dynamical models in predicting interannual ENSO signal in terms of forecast skill, rms error, extreme event prediction, and seasonal dependence of skill. A useful skill is not confined to tropical Pacific and Indian oceans, but extends to tropical Atlantic and to middle and high latitudes of the Northern Hemisphere. There is, therefore, a potential for a significant atmospheric tropical-midlatitude teleconnection associated with tropical SST forcing. Such a forcing in the Atlantic may be due to the THC variability.

Publications supported by the grant

Ghil, M., and A. W. Robertson, 2002: ``Waves'' vs. ``particles'' in the atmosphere's phase space: A pathway to long-range forecasting?, Proc. Natl. Acad. Sci., 99 (Suppl. 1), 2493--2500.

Kirshner, S., P. Smyth, and A. W. Robertson, 2003: Use of input-driven hidden Markov models with application to multi-site precipitation modeling, 2003 Neural Information Processing Conference, Vancouver, Canada, December 2003.

Kirshner, S., P. Smyth, and A. W. Robertson, 2004a: Semi-parametric Modeling of Time Series of Discrete-Valued Vectors using Hidden Markov Models. 20th Annual Uncertainty in Artificial Intelligence Conference, submitted, March 2004.

Kirshner, S., P. Smyth, and A. W. Robertson, 2004b: Conditional Chow-Liu Tree Structures for Modeling Discrete-Valued Vector Time Series. Technical Report, ICS-TR 04-04. Dept. Information and Computer Science, University of California, Irvine, CA.

Kondrashov, D., S. Kravtsov, A. W. Robertson and M. Ghil, 2005: A hierarchy of data-based ENSO models. J. Climate, 18, 4425–4444.

Koo, S., A. W. Robertson, and M. Ghil, 2003: Multiple regimes and low-frequency oscillations in the Southern
Hemisphere's zonal-mean flow. J. Geophys. Res., doi: 10.1029/2001JD001353.

Kravtsov, S., and A. W. Robertson, 2002: Midlatitude ocean-atmosphere interaction in an idealized coupled model. Climate Dynamics, 19, 693-711.

Kravtsov, S. V., A. W. Robertson, and M. Ghil, 2003: Low-frequency variability in a baroclinic beta-channel with land--sea contrast. J. Atmos. Sci., 60, 2267--2293.

Kravtsov, S., D. Kondrashov, and M. Ghil, 2004: Multi-level regression modeling of nonlinear processes: Derivation and applications to climatic variability. J. Climate, 18, 4404–4424.

Kravtsov, S., and M. Ghil, 2004: Interdecadal variability in a hybrid coupled ocean--atmosphere--sea-ice model. J. Phys. Oceanogr., 34, 1756–1775.

Kravtsov, S., A. W. Robertson and M. Ghil, 2005: Bimodal Behavior in the Zonal Mean Flow of a Baroclinic Beta-Channel Model. J. Atmos. Sci., 62, 1746-1769

Kravtsov, S., D. Kondrashov and M. Ghil. 2005: Multilevel Regression Modeling of Nonlinear Processes: Derivation and Applications to Climatic Variability. J. Climate, 18, 4404–4424.

Kravtsov, S., P. Berloff, W. K. Dewar, M. Ghil, and J. C. McWilliams, 2006: Dynamical origin of low-frequency variability in a highly nonlinear mid-latitude coupled model. J. Climate, accepted.

Lott, F., A. W. Robertson, and M. Ghil, 2004a: Mountain torques and Northern

Hemisphere low-frequency variability. Part I: Hemispheric and oscillatory aspects. J. Atmos. Sci., 61, 1259-1271.

Lott, F., A. W. Robertson, and M. Ghil, 2004b: Mountain torques and Northern Hemisphere low-frequency variability. Part II: Regional and episodic aspects. J. Atmos. Sci., 61, 1272-1283.

Robertson, A. W., S. Kirshner, and P. Smyth, 2006: Subseasonal-to-Interdecadal Variability of the Australian Monsoon Over North Queensland. Quart. J. Royal Meteor. Soc., 132, 519-542.