Final Technical Report for DOE Award SC0006616

Title: Collaborative Research, Type 1: Decadal Prediction and Stochastic Simulation of Hydroclimate Over Monsoonal Asia

Principal Investigators:
Drs. Andrew W. Robertson (IRI, lead PI), Edward R. Cook and Rosanne D'Arrigo (LDEO Tree Ring Laboratory), Michael Ghil (UCLA), Upmanu Lall (Columbia Water Center, CWC), and Padhraic J. Smyth (UCI)

Award Number: DOE-DE-FOA-000041 1
Award period: September 2011-August 2015

This collaborative project brought together climate dynamicists (UCLA, IRI), dendroclimatologists (LDEO Tree Ring Laboratory), computer scientists (UCI), and hydrologists (Columbia Water Center, CWC), together with applied scientists in climate risk management (IRI) to create new scientific approaches to quantify and exploit the role of climate variability and change in the growing water crisis across southern and eastern Asia.

The project objectives were to:
- Develop new tree-ring based streamflow reconstructions for rivers in monsoonal Asia;
- Improve understanding of hydrologic spatio-temporal modes of variability over monsoonal Asia on interannual-to-centennial time scales;
- Assess decadal predictability of hydrologic spatio-temporal modes;
- Develop stochastic simulation tools for creating downscaled future climate scenarios to 2050 and estimates of uncertainties, based on historical/proxy data and GCM climate change
- Develop stochastic reservoir simulation and optimization for scheduling hydropower, irrigation and navigation releases.

The final technical results of the project over the award period are summarized below by sub-topic. This Final Technical Report for this collaborative research project includes material from those submitted separately by PIs M. Ghil (UCLA, SC0006694) and P. Smyth (UC Irvine, SC0006619).

1. Tree ring-based streamflow reconstructions

We have published a paper on the reconstruction of Upper Indus River discharge from long tree-ring records located in the Karakoram region of northern Pakistan (Cook et al., 2013). While support for the tree-ring sampling came from USAID, the work conducted in producing the reconstruction and its novel method of estimating uncertainties utilized support from this project. The method of uncertainty estimation is based on the maximum entropy bootstrap (Vinod, 2006). It is noteworthy for being totally data-adaptive and much more natural in the way it expresses uncertainty in the reconstructed values back in time. It also automatically produces ensembles of pseudo-reconstructions that can be used for further analyses relevant to the goals of this project.

Previous studies indicate that there may be some modulating effect from the Pacific Decadal Oscillation or PDO on monsoon climate across Myanmar. To further investigate relationships between Myanmar hydroclimate and larger-scale climate dynamics, we analyzed the instrumental record and 20th century reanalysis data, along with the Myanmar teak series, to examine the large-scale climate patterns (precipitation, winds, moisture fluxes, monsoon indices) over Myanmar and adjacent southeast Asia during cold and warm PDO phases. We found significant decadal to multidecadal modes of variability linked to monsoon rainfall in both instrumental/reanalysis and the much longer teak tree-ring data from Myanmar, which appear to reflect the remote influence of the PDO and related climatic regime shifts on South Asian summer monsoon rainfall (D'Arrigo and Ummenhofer, 2014). The boreal winter PDO may precondition tree growth at the onset of the wet monsoon season over this region of south Asia. Our spectral analyses of the Maingtha teak identified ~48 and ~18.9yr modes that are similar to those identified for the PDO and other long instrumental time series of North Pacific climate variability. A ~48-year mode is also evident in the MHS Thailand teak, and is also believed to reflect decadal Pacific variability. The ~27yr mode seen in the teak record in Figure 4 is similar to that observed in a tree-ring reconstruction of Upper Indus, Pakistan streamflow over the past millennium (Cook et al., 2013). A ~18yr mode has been associated with lunar-tidal forcing in some North Pacific tree-ring records (Wilson et al. 2007). The ~14-15yr mode corresponds to bandwidths associated with lower-frequency ENSO variability (Allan 2000), and the PDO (Krishnan and Sugi 2003). The amplitudes of the latter two modes vary considerably over time, and all may interact with each other and be modulated by various climatic forcings and synoptic phenomena.

2. Identification of hydroclimate modes of variability

Correlation maps with the Upper Indus streamflow reconstruction and a variety of datasets suggests the importance of the Arabian Sea moisture source, and ENSO & NAO teleconnections. We have explored more systematically the three-way interactions between the Indian monsoon and the two ocean basins in the light of the synchronization theory of chaotic oscillators (Feliks et al. 2013). Four climate records were analyzed: the monsoon rainfall in the core region and the Peninsula region for India, the Southern Oscillation Index (SOI) for the Tropical Pacific, and the NAO index for the North Atlantic. The individual records exhibit highly significant oscillatory modes with spectral peaks at 7--8 yr and in the well-known quasi-biennial and quasi-quadrennial bands. A key result is that the 7--8-yr and 2.7-yr oscillatory modes in all three regions are completely synchronized, and the energy ratio analysis suggests that the NAO induces these modes in the other two regions. Both these modes in the NAO appear to be connected to intrinsic modes of variability of the Gulf Stream front.

An assessment was made of the ability of general circulation models in the CMIP5 ensemble to reproduce observed modes of low-frequency winter/spring (DJFMAM) precipitation variability in the region of the Upper Indus basin (UIB) in south-central Asia. This season accounts for about two thirds of annual precipitation totals in the UIB and is characterized by "western disturbances" propagating along the eastward extension of the Mediterranean storm track (Pal et al. 2014). Observational data are used, first for spatiotemporal characterization of the precipitation seasonal cycle. Seasonal spectra and teleconnections are then examined and links to global patterns of sea-surface temperature (SST) and atmospheric circulation explored. Annual and lowpassed winter variability are found to be associated primarily with large-scale SST modes in the tropical and extratropical Pacific. A more obscure link to North Atlantic SST, possibly related to variability of the North Atlantic Oscillation (NAO), is also noted (Greene and Robertson 2016).

An ensemble of CMIP5 models was then assessed with regard to these characteristics.

Eight of the 31 models are found to reproduce well the two leading modes (summer and winter/spring) of the observed seasonal cycle. Low-frequency (decadal-band) spectral peaks for winter/spring precipitation similar in frequency to those observed are found in four of the eight models, deriving from just two model families. Key features of the observed SST teleconnection patterns are best identified in the two models having higher resolution, but with varying degrees of resemblance to observations. Future precipitation and snowpack trends for the region of the UIB seem not to depend on model resolution, although the evidence for this is not definitive. Whether low-frequency variability represents a useful source of predictive information remains to be determined. This work is reported in Greene and Robertson (2016).

3. Empirical prediction using data-based closure models via multilayered stochastic models

a) Empirical model reduction

Kondrashov et al. (2014) undertook a rigorous mathematical analysis of empirical model reduction (EMR: Kondrashov et al. 2005; Kravtsov et al. 2005; Kravtsov et al. 2009) in their continuous-time limit; the authors called this limit multilayered stochastic models (MSMs). An MSM is a system of stochastic differential equations (SDEs) that models the dynamics of the macroscopic, observed variables along with their interactions with the microscopic, hidden variables.

The hidden variables are modeled through a “matrioshka” of layers, in which each additional layer contains an extra hidden variable that is less correlated with the observed variables than those in the previous layer, until some de-correlation criterion is reached. In practice, MSMs are learned in a polynomial basis by multilevel regression techniques that lead to an EMR model.

An MSM model can be written as a system of stochastic integro-differential equations that yields in practice a good approximation of the generalized Langevin equation of the Mori-Zwanzig formalism of statistical mechanics (Zwanzig, 2001). Furthermore, conditions were identified that guarantee the existence of a global random attractor for MSMs that generalize the EMR models used so far and allow for non-polynomial predictors. These generalized stochastic-dynamic, empirical models do not require energy-preserving nonlinearities — such as those present in fluid-dynamic models but not in ENSO or other climate models — while the global attractor still prevents numerical blow-up.

b) “Past-Noise Forecasting” (PNF) of Madden-Julian Oscillation

In the presence of both low-frequency variability (LFV) and noise, Chekroun et al. (2011) had shown that linear pathwise response of a nonlinear stochastic model to perturbations allows one to develop a novel Past-Noise Forecasting (PNF) method. The PNF method is based on the knowledge of “past noise” at times in the system’s history that resemble in LFV phase the present from which one wishes to forecast. These authors successfully applied PNF to an EMR model for ENSO; they constructed an ensemble of forecasts that accounts for interactions between high-frequency variability (“noise”), estimated by EMR, and the LFV mode of ENSO, as captured by singular-spectrum analysis (SSA).

Following up on the development of PNF and its application to ENSO, Kondrashov et al. (2013) have shown that PNF improves the prediction of the Madden-Julian Oscillation (MJO), an important intraseasonal phenomenon that affects hydroclimate in Asia. Applying the EMR-PNF method to the two leading indices of the MJO, RMM1 and RMM2, yields a bivariate correlation skill that exceeds 0.4 at 30 days, a skill that is comparable to that of state-of-the art dynamical models. A key result is that — compared to an EMR ensemble driven by generic white noise — PNF is able to considerably improve prediction of the MJO's phase. When forecasts are initiated from weak MJO conditions, the useful skill is of up to 30 days. PNF also significantly improves MJO prediction skill for forecasts that start over the Indian Ocean.

c) Data-driven Stochastic Modeling and Prediction for Asian Hydroclimate

The UCLA team's multivariate EMR stochastic methodology has been applied to monthly gridded (2.5 x 2.5 degrees) Palmer Drought Severity Index (PDSI) for a 700-yr–long time interval (1300–2005). The data were based on the Monsoonal Asia Drought Atlas (MADA) and were projected onto the 8 leading PCs in the 10N–56N latitude band. The optimal EMR multi-level model was 2-level quadratic and energy conserving; it was compared with data in terms of robust spectral peaks identified by Multichannel Singular Spectrum Analysis (MSSA). We found robust 5-yr and 8-yr low-frequency modes (LFMs) over the Indus river basin, in the actual dataset, as well as in the EMR simulations.

The tree-lab team at Lamont produced a 300-member ensemble of pseudo-reconstructions of Indus River discharge, based on maximum entropy bootstrapping of tree-ring proxy records. Our UCLA team analyzed this ensemble by MSSA and identified a robust 27-yr LFM. We assessed decadal predictability of hydrologic spatio-temporal modes by applying the SSA-MEM prediction methodology to the ensemble mean of the Indus River discharge reconstructions, and carried out retroactive forecasts with no look-ahead over the 1702–2005 time interval.

Validation shows some predictability of up to 15 yr, although this predictability appears to have been somewhat smaller in the 20th century, and it is largely due to the 27-yr LFM. We also generated a very long, 5000-yr stochastic simulation of a univariate EMR model of Indus streamflow reconstruction with realistic spectral features. This simulation will serve as a synthetic dataset for further analysis and testing of newly developed strategies for stochastic reservoir simulation and optimization.

4. Stochastic Simulation of Daily Precipitation Data for Downscaling using Hidden Markov Models

a) Methodological development

We developed a framework for a new type of hidden Markov model (HMM), including the statistical and mathematical foundations of the model as well as implementing in software in the R software package and making it publicly available. In this new modeling framework, an HMM is coupled with a generalized linear model (GLM) within a Bayesian framework to downscale exogenous input variables to spatio-temporal daily rainfall. This model (which we refer to as a GLM-HMM) is an extension of the non-homogeneous hidden Markov model (NHMM) to include uncertainty quantification in a Bayesian framework. The GLM-HMM also allows the exogenous inputs to directly modulate the emission distribution (governing the occurrence and amounts at each station) in a more flexible manner than a traditional NHMM, allowing for modeling of non-stationarity (for both seasonal and inter-annual variation).

The GLM-HMM includes hidden weather states for each day which have Markov transition dynamics that capture the time dependence and spatial relationship of the rainfall much like a traditional NHMM. Each day is modeled by one of the K hidden weather states, where each state has a distinct spatial pattern across stations in terms of distribution on precipitation occurrence and amount. The Markov state transitions are modeled through a set of transition probability matrices that vary over time, allowing the capture of seasonal variation in the relative frequencies of different states.

The distributions of daily rainfall depend on both the daily state variable as well as the exogenous input variables (which can for example be a large-scale variable reflecting inter-annual variability). The precipitation distribution for each station is a mixture of a delta function at zero and two gamma distributions (one for lower rainfall amounts and one for higher amounts). The exogenous variables influence which of the three component distributions is chosen given a particular daily weather state. As well as large scale annual exogenous input variables, seasonal or daily inputs can also be included in the model. The Bayesian framework allows for uncertainty estimation of the dependence of daily rainfall on particular exogenous input variables.

b) Algorithm and Software Development

In our work we have found that Bayesian implementations of NHMM models are computationally expensive and can be difficult to tune. By incorporating the exogenous variables into a different portion of the model (via the new GLM-HMM approach described above) we can retain the general form and functionality of the NHMM while removing the need for extensive tuning of algorithm parameters. Latent variables are added to ensure the mathematical conjugate properties are available to use Gibbs steps throughout the Markov chain Monte Carlo (MCMC) algorithm. The model is implemented in R with the aid of the Rcpp package to construct compilable modules in C++ for faster run time. This reduced the run-time 100-fold due to the compiled nature of the Rcpp modules. N=2000 iterations can be run in 4 hours for a field of 23 stations and 31 monsoon seasons (June, July, Aug, Sept) without the need for specialized user knowledge of MCMC tuning methods. In addition, the R code has been parallelized and, using the snowfall package, is run on a multi-core Linux machine to reduce computational time.

In more recent work we have developed a further extension of our Bayesian inference algorithms that allow us to fit these types of models to much larger data set, e.g., to almost 700,000 daily measurements of precipitation over 30 years across India. This new extension is based on a technique known as Polya-Gamma data augmentation, which has been proposed and applied successfully recently to multivariate data – our work extends this approach to handle temporal data using our GLM-HMM approach.

We implemented or developed metrics to assess the fit and predictive performance of the GLM-HMM model. Given particular settings of the exogenous variables, we collect N=500 simulated daily downscale runs from the model for each station. Assessment of the spatial features of the model is done through comparison of pair-wise station correlations of the simulated time series. The skill of the model to capture the temporal aspects of the data is assessed by run lengths of wet and dry spells of these 500 simulated runs. The main difference in the GLM-HMM and a traditional NHMM is in how the exogenous variables handle station level seasonality and downscale at a daily level; seasonal plots with uncertainty bands were compared to the raw data. The NHMM can have difficulty because it extends one seasonal input to the entire field of stations, whereas the GLM-HMM can have a seasonal trend for each station. The GLM-HMM allows the emission distribution of a state to change throughout the season, instead of being stationary. These trends were apparent in our model assessment plots. Additionally, larger model fit assessment was done through the computation of the recursive log-likelihood and annual metrics to measure the influence of different exogenous variables on both training and test sets of data. 6-fold cross validation is used to split the data into six portions for training and test (holding out 5 year portions of data at a time). This work is reported in Holsclaw et al. (2016b).

The resulting software has been publicly released as an R software package, NHMM: Bayesian NHMM Modeling, https://cran.r-project.org/web/packages/NHMM/index.html

c) Application of NHMMs to downscaling of rainfall projections in India and China

Downscaling rainfall using the GLM-HMM has been applied to a region in India covering the Upper Indus basin that feeds into the reservoir system, and in China for the Upper Yangtze basin. The impact of exogenous variables such as the El Nino-Southern Oscillation (ENSO), standardized anomaly index (SAI), and wind shear (WSI) have been assessed by this model for their skill in out-of-sample prediction of downscaling the monsoon season for these regions. Thirty years of daily rainfall data is available for training and testing the model for these regions. Test data is held out to assess the ability of the model to forecast or hindcast given the exogenous variables. This work is reported in Holsclaw et al. (2016a).

5. Multi-Timescale Climate Informed Stochastic Hybrid Simulation – Optimization Model (McISH) for a Single Multi-Purpose Reservoir

Streamflow forecasts at multiple time scales (e.g., season and year ahead) provide a new opportunity for reservoir management to address competing objectives. Market instruments such as forward contracts with specified reliability are considered as a tool that may help address the perceived risk associated with the use of such instruments in lieu of a traditional operation and allocation. A water allocation process that enables multiple contracts with different durations, to facilitate participatory management of the reservoir by users and system operators, is presented here. Since these contracts are based on a verifiable reliability they may in turn be insurable. A Multi-timescale climate informed Stochastic Hybrid Simulation – Optimization Model (McISH) is developed, featuring (1) dynamic flood control storage allocation at a specified risk level; (2) multiple duration energy/water contracts with user specified reliability and prices; and (3) contract sizing and updating to reflect changes in both demands and supplies. The model incorporates multi-timescale (annual & seasonal) streamflow forecasts, and addresses uncertainties across both timescales. The intended use is as part of an interaction between users and water operators to arrive at a set of short-term and long term contracts through disclosure of demand or needs and the value placed on reliability and contract duration. An application is considered using data for the Bhakra Dam, India. The issues of forecast skill and contract performance given a set of parameters are examined to illustrate the approach. Prospects for the application in a general setting are discussed in Lu (2014).

This section formed Part IV of Mengqian Lu's PhD thesis (Lu 2014).

Publications resulting from this grant

Arnesen, P., T. Holsclaw, P. Smyth, 2016, `Bayesian detection of changepoints in finite-state Markov chains for multiple sequences, in press.

Cook, E.R., Palmer, J.G., Ahmed, M., Woodhouse, C.A., Fenwick, P., Zafar, M.U., Wahab, M., and Khan, N. 2013. Five centuries of upper Indus River flow from tree rings. Journal of Hydrology 486:365–375.

D'Arrigo, R. and C. Ummenhofer. 2014. The climate of Myanmar: evidence for effects of the Pacific Decadal Oscillation. Int. J. Climatol. doi:10.1002/joc.3995.

D'Arrigo, R., J. Palmer, C. Ummenhofer, Nyi Nyi Kyaw and P. Krusic. 2013. Myanmar monsoon drought variability inferred by tree rings over the past 300 years: linkages to ENSO. Special issue on ENSO, PAGES newsletter 21: 2, P. Braconnot, C. Brierley, S. Harrison, L. von Gunten and T. Kiefer, editors.

D'Arrigo, R., J. Palmer, C. Ummenhofer, Nyi Nyi Kyaw and P. Krusic. 2012. Three centuries of Myanmar monsoon climate variability inferred from teak tree rings. Geophys. Res. Lett. 38: L24705, doi:10.1029/2011GL049927.

Feliks, Y., A. Groth, M. Ghil, and A. W. Robertson, 2013: Oscillatory climate modes in the Indian monsoon, North Atlantic and Tropical Pacific. J. Climate, 26, 9528–9544.

Greene, A., T. Holsclaw, A. Robertson, P. Smyth, 2015: A Bayesian multivariate nonhomogeneous Markov model, in Machine Learning and Data Mining Approaches to Climate Science, Springer, pp.61–69.

Greene A., A. W. Robertson, 2016: Interannual and low-frequency variability of Upper Indus Basin winter/spring precipitation in observations and CMIP5 models. In preparation.

Holsclaw, T., A.M. Greene, A.W. Robertson, and P.J. Smyth, 2016a: A Bayesian hidden Markov model of daily precipitation over South and East Asia, Journal of Hydrometereorology, doi: 10.1175/JHM-D-14-0142.1, 17(1):3–25.

Holsclaw, T., A. W. Robertson, A. M. Greene, and P. Smyth, 2016b: Bayesian non-homogeneous Markov models via Polya-Gamma data augmentation with applications to precipitation modeling. In preparation.

Kondrashov, D., M.D. Chekroun, and M. Ghil, 2014: Data-driven non-Markovian closure models, Physica D, doi:10.1016/j.physd.2014.12.005.

Kondrashov, D., M.D. Chekroun, A.W. Robertson, M. Ghil, 2013: Low-order stochastic model and "Past Noise Forecasting" of Madden-Julian Oscillation, Geophys. Res. Lett., 40, 5303–5310.

Lu, M., 2014: “From Diagnosis to Water Management: The role of Atmospheric Dynamics and Climate Variability on Hydrological Extremes”, PhD Thesis, Columbia University, New York, pp270.

Moron, V., and A. W. Robertson, 2013: Interannual variability of Indian summer monsoon rainfall onset date at local scale. Int. J. Climatol., DOI: 10.1002/joc.3745.

Pal, I., A. W. Robertson, U. Lall, and M. A. Cane, 2014: Modeling Winter Rainfall in Northwest India using a Hidden Markov Model: Understanding Occurrence of Different States and their Dynamical Connections. Climate Dynamics, doi: 10.1007/s00382-014-2178-5.

Ummenhofer, C., R. D'Arrigo, K. Anchukaitis and E. R. Cook. 2012. Links between Indo-Pacific Variability and Drought in Monsoon Asia in the MADA.
Climate Dynamics: DOI 10.1007/s00382-012-1458-1.

Additional references cited

Allan, R. J. 2000. ENSO and climatic variability in the past 150 years, in ENSO: Multiscale Variability and Global and Regional Impacts, edited by H. F. Diaz and V. Markgraf, pp. 3– 55, Cambridge Univ. Press, New York.

Chekroun, M. D., D. Kondrashov and M. Ghil, 2011: Predicting stochastic systems by noise sampling, and application to the El Niño-Southern Oscillation, Proc. Natl. Acad. Sciences USA, 108 (29), 11766–11771.

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

Krishnan R, and M. Sugi. 2003. Pacific Decadal oscillation and variability of the Indian summer monsoon rainfall. Climate Dynamics 21: 233–242.

Kravtsov, S., M. Ghil, and D. Kondrashov, 2009: Empirical model reduction and the modeling hierarchy in climate dynamics and the geosciences. Stochastic Physics and Climate Modelling, T. Palmer and P. Williams, Eds., Cambridge University Press, pp. 35–72.

Kravtsov S., D. Kondrashov, and M. Ghil, 2005: Multi-level regression modeling of nonlinear processes: Derivation and applications to climatic variability. J. Climate, 18, 4404–4424.

Vinod, H.D., 2006. Maximum entropy ensembles for time series inference in economics. Journal of Asian Economics. 17(6):955–978.

Wilson, R., G. Wiles, R. D’Arrigo and C. Zweck. 2007. Cycles and shifts: 1300 years of multi-decadal temperature variability in the Gulf of Alaska. Clim. Dyn. DOI 10.1007/s00382-006-0194-9.

Zwanzig, R., Nonequilibrium Statistical Mechanics, Oxford University Press, 2001.