Empirical modeling and stochastic simulation of vector wind for the use in regional downscaling and risk assessment schemes

Sergey Kravtsov, Boyko Dodov, Peter Sousounis and Yucheng Song

Summary: The scope of this project is stochastic simulation of surrogate vector-wind fields for the use in regional downscaling with the purpose of risk assessment. Proof-of-concept results for the 850-mb NCEP/NCAR reanalysis wind indicate that it is possible to construct a stand-alone empirical stochastic vector-wind model that produces variable wind fields statistically indistinguishable from the observed wind over a wide range of statistical measures; these measures include those involving the statistics of mid-latitude as well as tropical cyclone tracks. We propose to extend these encouraging results to simulations of the vector-wind variability and co-variability at three vertical levels, and also as a function of external forcing associated with the observed variability of selected climate indices, using the 1° resolution CFSR reanalysis wind and SST products. The project's timeline places tentative completion date in late July 2015.

1. Introduction

A wide class of atmospheric and climatic phenomena is governed by prognostic equations that relate tendencies (time derivatives) of the geophysical fields (velocity, temperature etc., which we will hereafter refer to as to the state vector) to the quadratic function of the state vector itself. In the data-modeling world, we are given observations of this state vector's time series, and we would like to fit, empirically, a model for the evolution of this state vector that would capture, as accurately as possible, the observed statistics.

$$\frac{dx_i}{dt} = a_{ijk}x_jx_k + b_{ij}x_j + c_i \tag{1}$$

To do so, we specify the form of the model, for example as in (1), and seek to find the "correct" values of *a*, *b*, *c* parameters so that the model, when integrated in time, would produce behavior similar to the actual observations of the state vector. Since we are looking at only a subset of the important dynamical variables, we have to parameterize interaction with the other, unresolved variables via introducing additional random forcing to our equation. The time series of *x*'s on the right hand side and *dx*/*dt*'s on the left-hand side of (1) are considered given, and the problem of finding the coefficients of our empirical model, as well as the noise characteristics, is formally equivalent to that of multiple linear regression. The theoretical and methodological aspects of this regression-problem solution have been well developed for a long time. Recently Kravtsov and collaborators (see the References section) have shown, in a series of papers, that this methodology — which they termed Empirical Model Reduction (EMR) — can be applied with a great success to the modeling of multidimensional geophysical fields. In the applications of the EMR methodology thus far, the emphasis was on modeling, in some sense, the “slow manifold” of the observed atmospheric dynamics. Here we demonstrate that using the EMR technique, it is feasible to simulate the entire spatiotemporal range of the atmospheric 850-mb wind variability. We propose to extend these results to the simulations of multi-level wind variations conditioned on the variable observed sea-surface temperature states of the World Ocean.

2. Preliminary results

The particulars of the 850-mb U, V model based on the daily NCEP/NCAR 2.5º vector wind data (Kalnay et al. 1996) over the Northern Hemisphere are as follows. The model has 1100 effective variables (U, V combined principal components (PCs) that account for > 99% of the total 850-mb U, V’s variability), arranged, along with hidden variables (previous level residuals) in 3 *linear* levels (see Kravtsov et al. references for more info). Seasonal dependence is accounted for by fitting separate models to seasonal data (however, PCs were computed for the year-round data, thus providing seamless year-round integration). Stepwise regression/PLS regularization was used for each model level to provide robust estimates of model coefficients. A novel aspect of the model is in the application of the state-dependent stochastic forcing at the third (hidden) model level, based on libraries of the observed level residual and conditioning on the main-level observable state. One 1948–2008 simulation of the stochastic model takes about 10 minutes of wall-clock time on the standard 2.7 GHz Intel Core i7 Apple MacBook Pro.

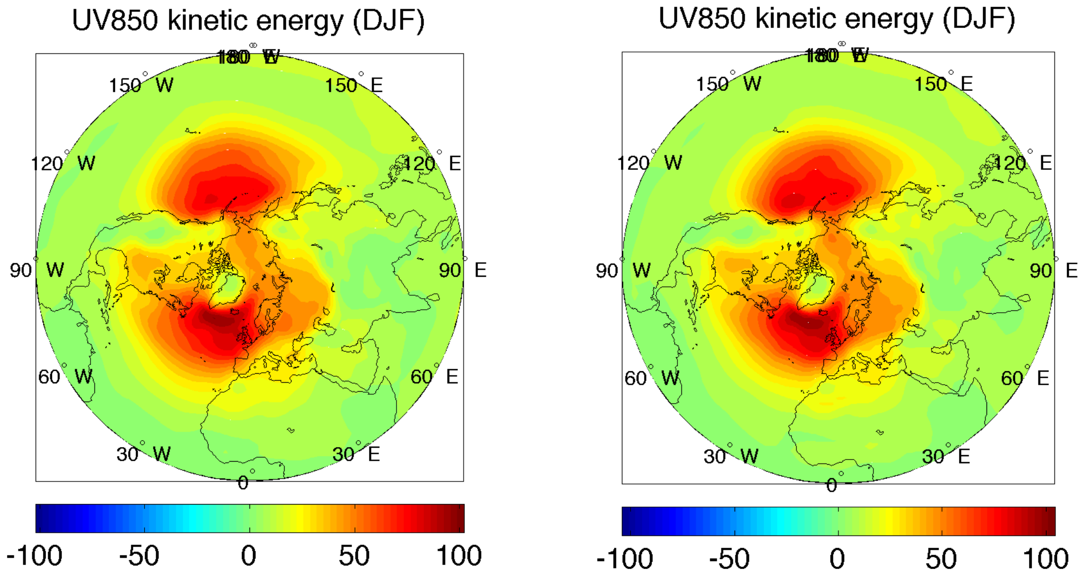

Figure 1: Climatology of kinetic energy for 850-mb level based on NCEP/NCAR (u,v) field (1948–2008), for DJF season. Left panel: observed data; right panel: simulated data. Seasonal cycle was filtered out prior to computing the energy.

The model generates synthetic wind patterns whose statistics is nearly indistinguishable from that of the observed wind (see Fig. 1 as an example), including statistics associated with mid-latitude and tropical storm tracks (not shown).

3. Proposed work

We propose to use the above methodology to construct an empirical wind model based on the combined 850-mb, 500-mb, and 250-mb winds; this models should also include the dependence on a set of climate indices to match the variability of simulated fields observed in real climate; the most efficient way of doing so would be to make the model coefficients linearly dependent on the representative climate indices: global-mean SST, AMO, PDO, ENSO, etc. (see Fig. 2). The proposed data sets used for the model construction will be taken from the 1° resolution CFSR reanalysis project (Saha et al. 2014).

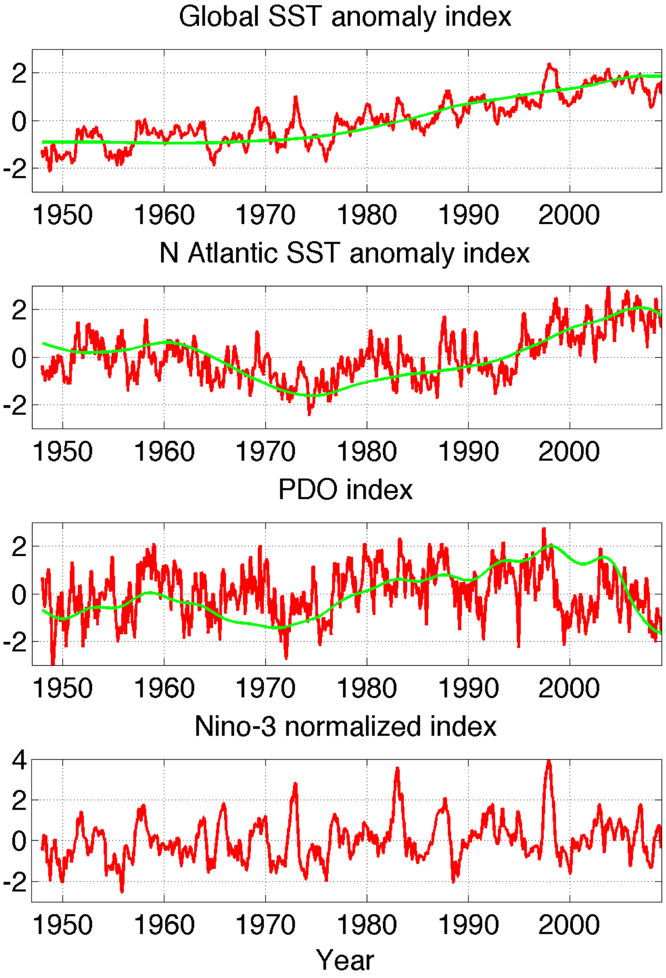

Figure 2: Sea-surface temperature (SST) climate indices based on the ERSST data set (Smith et al. 2008).

4. Project timeline, goals and deliverables

4.1. Stage I: Analysis and construction of 3-level NCEP/NCAR reanalysis based data model

- **Timeline:** Jul 1 – Sep 30, 2014
- **Scope:** Construction of a 3-level NCEP/NCAR reanalysis based model of instantaneous U and V fields at 2.5°x2.5° resolution
- **Goals:** Simulate multilevel (e.g. 850mb, 500mb and 250mb) U and V fields with consistent baroclinicity. The test criterion will be developed by AIR scientists in consultation with SK early in the course of Stage I.
- **Deliverables:**
    - Input data files and surrogate realizations of the multi-level wind fields with reasonable baroclinicity according to the criteria developed by AIR in consultation with SK
    - MATLAB © codes used to train the empirical model, and to produce these realizations

4.2. Stage II: Conditioning of the data model from Stage I on a set of seasonal and climate indices

- **Timeline:** Oct 1 – Dec 31, 2014
- **Scope:** Construction of a 3-level NCEP/NCAR reanalysis based model of instantaneous U and V fields at 2.5°x2.5° resolution and 6 hourly time step with consistent seasonal variability and conditioned on a set of climate indices.
- **Goals:** Simulate multilevel (e.g. 850mb, 500mb and 250mb) U and V fields with consistent baroclinicity, seasonality and dependence on a set of climate indices. The test criteria will be developed by AIR in consultation with SK early in the course of Stage II.
- **Deliverables:**
    - Surrogate realizations of the multi-level wind fields with reasonable baroclinicity, seasonal variability and dependence on the selected set of climate indices, according to the criteria developed by AIR in consultation with SK
    - MATLAB © codes and input data files used to train the empirical model, and to produce these realizations

4.3. Stage III: Analysis and construction of a 3-level higher-resolution CFSR based data model

- **Timeline:** Jan 1 – May 31, 2015
- **Scope:** Based on the results from stages I and II, a higher resolution 3-level U and V data model will be constructed based on CFSR data at 1°x1° resolution and 6 hourly time step with consistent seasonal variability and conditioned on a set of climate indices.
- **Goals:** Increase the resolution of simulated U and V fields from Stage II to 1°x1°. The consistency of baroclinicity, seasonality and dependence on a set of climate indices has to be preserved compared to CFSR data and measured according to the criteria developed at stages I and II.
- **Deliverables:**

    - Surrogate realizations of the multi-level wind fields at 1°x1° resolution with reasonable baroclinicity, seasonal variability and dependence on the selected set of , according to the criteria developed by AIR in consultation with SK at previous stages
      MATLAB © codes and input data files used to train the empirical model, and produce these realizations

4.4. Stage IV: Final report and submission of a methodological paper to a leading atmospheric science journal and preparation of industry/academia (GOALI) collaborative NSF proposal on the application of the proposed model in high-resolution downscaling schemes

- **Timeline:** Jun 1 – Jul 31, 2015
- **Scope:** The results from the collaborative research, if successful, would be a significant scientific achievement that, when published, would bring recognition to both parties and provide a significant marketing value to AIR. Upon success, this work should be further extended in a GOALI supported contract to develop a framework for high-resolution regional downscaling using a multilevel version of the proposed empirical model.
- **Deliverables:**
    - Final report with detailed description of the modeling approach
    - Joint paper submitted to a leading peer review journal
    - A GOALI proposal submitted to NSF

5. Personnel

The proposed project is a collaboration between UWM (SK) and AIR scientists (Boyko Dodov, Peter Sousounis, Gerhard Zuba and Yucheng Song). The AIR group will be responsible for initial CFSR data setup and for the detailed data/empirical model output comparison for both NCEP/NCAR based and CFSR based results. Sergey Kravtsov will lead the empirical model development and will be responsible for model codes and running basic model diagnostics with regards to matching the observed wind-data statistics. All participants will take part in writing the manuscripts and proposal and interpreting the results of this work.

References

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.

Kravtsov, S., I. Kamenkovich, D. Kondrashov, and M. Ghil, 2011: Empirical stochastic model of sea-surface temperatures and surface winds over the Southern Ocean. *Ocean Sciences*, **7**, 755–770.

Kondrashov, D., S. Kravtsov, and M. Ghil, 2010: Signatures of nonlinear dynamics in an idealized atmospheric model. *J. Atmos. Sci.*, **68**, 3–12.

Strounine, K., S. Kravtsov, D. Kondrashov, and M. Ghil, 2010: Reduced models of atmospheric low-frequency variability: Parameter estimation and comparative performance. *Physica D*, **239**, 145–166, doi:10.1016/j.physd.2009.10.013.

Kravtsov, S., M. Ghil, and D. Kondrashov, 2010: *Empirical Model Reduction and the Modeling Hierarchy in Climate Dynamics and the Geosciences*. *Stochastic Physics and Climate Modeling*, T. Palmer and P. Williams, Eds., Cambridge University Press, pp. 35–72.

Kravtsov, S., Hoeve, J. E. T., S. B. Feldstein, S. Lee, and S.-W. Sun, 2009: The relationship between statistically linear and nonlinear feedbacks and zonal-mean flow variability in an idealized climate model. *J. Atmos. Sci.*, **66**, 353–372.

Kondrashov, D., S. Kravtsov, and M. Ghil, 2006: Empirical mode reduction in a model of extratropical low-frequency variability. *J. Atmos. Sci.*, **63**,1859–1877.

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

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.

Saha, Suranjana, and Coauthors, 2014: The NCEP Climate Forecast System Version 2. *J. Climate*, **27**, 2185–2208.

Smith, Thomas M., Richard W. Reynolds, Thomas C. Peterson, Jay Lawrimore, 2008: Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006). *J. Climate*, **21**, 2283–2296.