Empirical mode reduction in an atmospheric GCM

Sergey Kravtsov
Department of Mathematics, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201

Motivation and methodology

General circulation models (GCMs) used for climate prediction are extremely difficult to analyze. One way of simplifying such an analysis is by constructing a reduced model (the one with fewer degrees of freedom), which will nevertheless capture essential dynamics of a GCM.

$$
\begin{aligned}
dx_i &= \tau_{ijkl}\; x_j x_k x_l + \alpha_{ijk}\; x_j x_k + \beta_{ij}^{(0)}\; x_j + \gamma_i + r_i^{(0)} \\
dr_i^{(0)} &= \beta_{ij}^{(1)}\; [x;\; r^{(0)}]_j + r_i^{(1)} \qquad (1) \\
dr_i^{(1)} &= \beta_{ij}^{(2)}\; [x;\; r^{(0)};\; r^{(1)}]_j + dw_i
\end{aligned}
$$

In this study, we apply an empirical methodology for such a reduced-model construction (Kravtsov et al. 2006) to analyze a realistic GCM.

The reduced model (1) operates in a phase space $\boldsymbol{x}$ of model's EOFs (see below), has two hidden levels $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$, and is driven by a spatially correlated white noise $\boldsymbol{w}$. All coefficients ($\boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$) in the model are found directly, using GCM output, by least squares. The increments d$\boldsymbol{x}$, d$\boldsymbol{r}_1$, and d$\boldsymbol{r}_2$ are approximated as daily differences ($\boldsymbol{x}^{(n+1)}-\boldsymbol{x}^{(n)}$ and so on).

GCM data

We decompose daily 300-mb streamfunction fields from the output of a 10^6-day-long perpetual-winter simulation of an early version of the Community Climate System model CCM0 (Branstator and Berner 2005) into empirical orthogonal functions (EOFs; **Fig. 1**). Cubic and quadratic ($\boldsymbol{\tau}$=0) three-level models (1) are constructed in a phase space spanned by *N* leading EOFs, where the optimal number *N* and the degree of nonlinearity in the model are determined by cross-validation: we aim to choose the model that will fit best both linear and nonlinear properties of a few leading EOFs (see below).

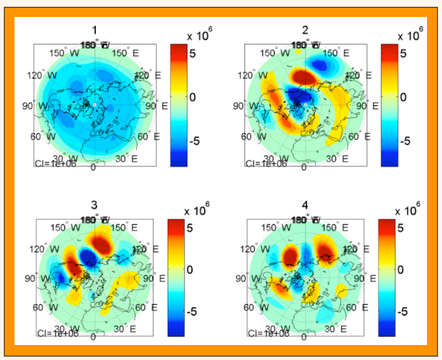

Fig. 1. Leading EOFs of 300-mb streamfunction (Northern Hemisphere parts of global patterns are shown). EOFs 2 and 3 resemble the Arctic Oscillation (AO) and Pacific–North American pattern (PNA), respectively. The four EOFs account for 8%, 3.5%, 3%, and 2.5% of variance, respectively.

Choice of the best model and results

The model (1) has many more nonlinear terms than the linear ones, but the linear tendency in fact dominates (see **Fig. 2**): inclusion, in a given 20-variable regression model, of a linear dependence on all 20 variables results in the most dramatic reduction of unexplained variance (compare left and right panels of **Fig. 2**), while turning up nonlinearity does not lead to any such appreciable increase (right panel). The main nonlinear effects manifest in the evolution of EOFs 2 and 3 (not shown). Both linear and nonlinear properties of the full-model time series are captured best by the 9-variable cubic model (**Fig. 3**).

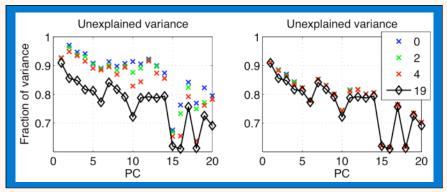

Fig. 2. Fraction of unexplained variance [i.e., Var{$r_i^{(0)}$}/Var{dx_i}] in the main level of model (1). The models, whose results are shown on the left, all predict the evolution of leading 20 EOFs, but the right-hand side of the equation for each variable only contains this variable, and its *M* nearest neighbors [e.g., for *M=2,* the equations for dx_1 and dx_2 contain x_1, x_2, and x_3; the equation for dx_3 contains x_2, x_3, and x_4, and so on]. The models on the right use *M* nearest neighbors in computing nonlinear terms, but all 20 variables to estimate the linear part of the operator. Results for *M* = 0, 2, 4, and 19 are shown.

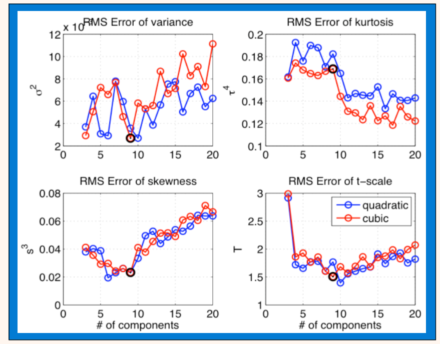

Fig. 3. The root-mean-square (rms) error of four different statistical characteristics of a reduced-model data: variance, skewness, kurtosis, and integral correlation time scale. The rms is computed by averaging the errors (i.e., difference in a given statistical characteristic between full-model and reduced-model data sets) among *first three* model variables: nonlinear effects are most pronounced in the time dependence of EOFs 2 and 3 (not shown). The results from quadratic (blue) and cubic (red) models are shown as a function of the number of variables *N* in these models.

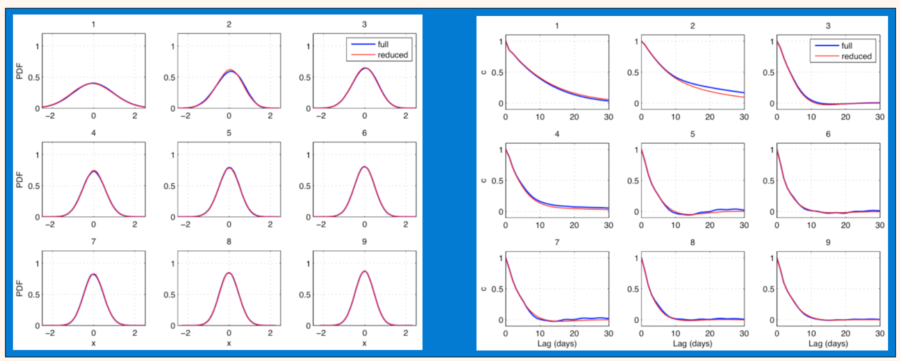

Fig. 4. Performance of 9-variable cubic model. Left frame: Probability density function (PDF); right frame: Autocorrelation function. Each panel's title refers to the number of variable being displayed. The results from the reduced model are shown in red, GCM data results are plotted in blue.

An excellent performance of the 9-variable cubic model, in terms of capturing both the probability density function (PDF), and the autocorrelation function of the full-model time series, is summarized in **Fig. 4**. Two- and three-dimensional PDFs of the synthetic data (not shown) also describe fairly accurately the "zero-order" Gaussian structure of full-model PDFs, as well as deviations from Gaussianity in the latter.

Discussion

While the reduced model produces synthetic multi-channel time series whose statistical properties resemble closely those of the full-model time series, the discrepancies between the two are also notable: the reduced model fails to capture the fourth-order moment of the distribution of EOF-2, and underestimates the time scale of the same EOF. Both deficiencies may be related to a restrictive form of a regression model (1), which only has additive noise terms. Majda et al. (1999) have developed a systematic mode-reduction strategy, which predicts a slightly different formulation of reduced-model equations, namely that including not only additive, but also multiplicative noise terms. The latter terms may represent, dynamically, the so-called *synoptic-eddy feedback*, and may result, in principle, in both "flatter" PDF distributions and increased persistence of the patterns of low-frequency variability.

The reduced model (1), which mimics the GCM data pretty well, can be further used to infer dynamics behind the "observed," full-model statistics (cf. Kondrashov et al. 2006), or be combined with the full model into a medium-range prediction scheme; here, the reduced model (which is thousands times less computationally expensive than full GCM) produces "most probable" long-term trajectory for its variables, for further assimilation into the GCM to forecast other variables.

References

Branstator, G., and J. Berner 2005: Linear and nonlinear signatures of a planetary wave dynamics of an AGCM: Phase space tendencies. *J. Atmos. Sci.*, **62**, 1792–1811.
Kravtsov, S., D. Kondrashov, and M. Ghil, 2006: Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *J. Climate*, **18**, 4404–4424.
Kondrashov, D., S. Kravtsov, and M. Ghil, 2006: Empirical mode reduction in a model of extratropical low-frequency variability. *J. Atmos. Sci.*, in press.
Majda, A. J., I. Timofeyev, and E. Vanden-Eijnden, 1999: Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. USA*, **96**, 14687–14691.

Acknowledgments

The author thanks G. Branstator for providing the CCM0 data set and for helpful discussions; the author is also grateful to Andrew Mai for technical assistance. This research was supported by the Department of Energy grant DE-FG-03-01ER63260.

For further information

Please contact *kravtsov@uwm.edu*. A PDF-version of this poster can be found at *www.uwm.edu/~kravtsov/presentations/DOE2006a_poster.pdf*.

UNIVERSITY of WISCONSIN
UWMILWAUKEE
College of Letters and Science
ATMOSPHERIC SCIENCE
DEPARTMENT OF MATHEMATICAL SCIENCES