Atmos. Chem. Phys. Discuss., 12, C687–C690, 2012
www.atmos-chem-phys-discuss.net/12/C687/2012/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Understanding and forecasting polar stratospheric variability with statistical models" *by* C. Blume and K. Matthes

Anonymous Referee #2

Received and published: 17 March 2012

This paper studies the performance of four statistical techniques with external factors in explaining and forecasting the variability of the north-polar stratospheric vortex. All of the techniques studied [viz., linear discriminant analysis (LDA), finite element clustering (FEM-VARX), multi-layer perceptrons (MLP), and support vector regression (SVR)] feature some aspects of nonlinearity and/or nonstationarity. Here, variability of the north-polar stratospheric vortex is measured through the leading principal component (PC) of area-weighted averages of either geopotential or temperature anomalies at the 10, 20, and 30 hPa levels. A total of 12 external factors are incorporated in the models, consisting of nine physical external factors (representing ENSO, the pacific decadal oscillation, the quasi-biennial oscillation, tropospheric blockings, solar variability, and aerosol optical depth), plus three baseline factors representing the seasonal cycle and a linear global warming trend. The models are trained for the period 07/01/1980–

06/31/2005 using ERA-Interim reanalysis data. MERRA reanalysis data for the period 07/01/2005–04/30/2011 are used for validation.

The authors select the optimal parameters for their models using well-established statistical methods, such as the Akaike information criterion (AIC) and cross-validation. All models apart from the LDA model are able to fit the training data with high accuracy ($R \simeq 0.9$), but there is a striking reduction of skill when the models are tested against the validation data. Among the studied models, only the FEM-VARX and MLP models are able to reproduce the test data with acceptable skill $R \simeq 0.4$. However, the latter skill scores are evaluated using a priori known values of external factors, and therefore do not represent genuine skill in an operational forecast environment where the external factors must be predicted. Relative response indices are computed to assess the importance of each external factor, indicating that the least important factors in this study are tropospheric blockings and aerosol optical depth, whereas all other factors are significant. The authors conclude with an operational prediction (including predictive estimates of external factors) that the stratospheric winter of 2011/12 would be anomalously warm (with a weak stratospheric vortex), and that a sudden warming would take place in late January or early February 2012.

Overall, this is an interesting and well-written article, suitable for publication in Atmospheric Chemistry and Physics, provided that the comments below are addressed.

1. The authors should comment on the rationale for projecting the three dimensional time series (measured at 10, 20, and 30 hPa) onto the leading PC instead of using all of the available data. Even before the data were projected onto the first PC, significant coarse-training took place in the area averaging step, which makes one wonder why additional coarse-graining (and the associated loss of information) is needed through projection. Was projection performed in order to reduce the number of parameters estimated in the training stage? If so, to what extent were prevention of overfitting and/or reduction of computational cost

significant considerations in the modeling process?

Assuming that the authors seek to train their models on scalar time series, I would think that more information of the original system would be retained by performing EOF analysis on the composite $\{10, 20, 30\}$ hPA fields without area-averaging, and projecting the data onto the leading EOF. Moreover, the justification in p. 5662 for retaining only PC1 on the basis of its high explained variance is not appropriate in the context of nonlinear predictive models, where it is well known that states carrying low variance may be important to reproduce the right dynamics (e.g., Aubry et al., 1993; Crommelin and Majda, 2004). In general, the authors should report or comment on how the results of the analysis would change if information from the 2nd or all three PCs were included.

2. A particularly revealing aspect of the paper (which is frequently glossed over in the climate literature), is the contrast in Fig. 1 of the performance of the models in fitting the training and validation data. However, the results of that figure cannot be used to infer that the models have actual forecasting skill, as the authors claim on p. 5668. This is because the true (hindcast) values of the external factors were used to produce Fig. 1, but those values would not be available in an operational forecast. Thus, labeling the right-hand panel of Fig. 1 as "forecast" is somewhat misleading for readers who have not read the paper in full, or are not familiar with external-factor-based models. For the same reason, the statement in the conclusions (p. 5671) that the models are able to satisfactorily forecast variability is also misleading. Perhaps a more accurate statement would be that the models are able to explain the variability of hindcast data with a priori known external factors. Also, the statement that the sudden stratospheric warming of January 2009 cannot be predicted because of internal chaotic variability is not necessarily true; the poor skill in predicting that event may well be due to model error of the statistical models, rather than intrinsic unpredictability of the atmosphere. On p. 5669 the authors correctly point out that operational forecasts with external factors require

a priori knowledge or predictability of the future values of the factors. This is a potentially major difficulty in climate applications, which should be made clear early on in the paper.

3. That a total of 12 external factors, plus a number of model regression parameters, are used to fit scalar time series raises questions related to model selection and overfitting. In particular, if some of the external factors with small relative impact are removed from the analysis, do the optimal models determined through the AIC and cross-validation change significantly? All of the external factors used in this analysis are natural from a physical point of view, but the authors should comment on whether including all nine of the external factors is warranted from the point of view of model complexity and risk of overfitting. Presumably, an AIC-type analysis could be performed to estimate an "optimal" number of external factors. Have the authors considered pursuing such an analysis?

4. Since winter 2011/12 is over at the time of writing of this review, it would be very interesting if the authors can comment on the actual skill of the forecast described in Section 6.

References

N. Aubry, W.-Y. Lian, and E. S. Titi. Preserving symmetries in the proper orthogonal decomposition. *SIAM J. Sci. Comput.*, 14:483–505, 1993. doi: 10.1137/0914030.

D. T. Crommelin and A. J. Majda. Strategies for model reduction: Comparing different optimal bases. *J. Atmos. Sci.*, 61:2206–2217, 2004. doi: 10.1175/1520-0469(2004)061<2206:sfmrcd>2.0.cO;2.

Interactive comment on Atmos. Chem. Phys. Discuss., 12, 5659, 2012.