Looking for Nonlinearities in the Large Scale Dynamics of the Atmosphere*

C. Tebaldi†*, D. Nychka*‡*, and G. Branstator*§

1 Introduction

Linear models can accurately account for the fundamental circulation patterns observed in the large scale atmospheric flow. It is as if single realizations of a multivariate random variable were observed – what we call the weather maps – whose deviations from a mean equilibrium state have no preferred direction or sign. A multivariate Gaussian distribution would be a good model for these observations. Traditionally, the mean state is believed to be unique.

By now, though, the atmospheric sciences literature has well documented the existence of important low-frequency anomalies, their persistence in time and space, and their recurrence. These phenomena cannot be explained by simple linear models, hinting to more complex, nonlinear generator mechanisms.

Methods of addressing the issue – mainly work of the last two decades – range from strictly mathematical developments to data analysis of observations from nature. Scientists have written down the equations and analysed the behavior of simple and complex nonlinear dynamical systems, in order to confirm that the large scale dynamics are thus better represented – and the concepts of multiple

*This work was done for the main part when the first author was visiting scientist in the Geophysical Statistics Project at the National Center for Atmospheric Research (NCAR), Boulder, CO.

†Claudia Tebaldi is data modeler, Advanced Technology Group, Athene Software inc., Boulder, CO.

‡Doug Nychka is senior scientist in the Climate and Global Dynamics division, and leader of the Geophysical Statistics Project at NCAR.

§Grant Branstator is senior scientist in the Global Dynamics Section of the Climate and Global Dynamics division of NCAR.

equilibria, basins of attraction, stable and unstable points, limit cycles are fertile ground of investigation. (see [2] and [8])

From a different perspective, data exploration has tried to document the existence of multiple regimes in observations from nature, identified as multiple modes and skewness in those observations' distribution. However, the short span of the time series available makes difficult to obtain statistically significant results, especially when the analysis is carried out in high dimensional domains, and looks for subtle deviations from a dominant, well behaved, linear component. (see [5],[6], [3], [7], [9])

In an intermediate position between the two approaches, but more similar to the investigation of nature's dynamics than to the abstract mathematical formulations of the problem, is a third kind of analysis. General Circulation Models – like NCAR Community Climate Models (CCMs) – can simulate the dynamics of the atmosphere and can provide the investigator with potentially infinite records of statistics. These wealth of records can be used to search for multimodality and non-gaussian behaviors. Also, with this abundance of data points one can conduct more extensive and sophisticated statistical investigations. Results already published (for example [4]), in accordance to the observational and mathematical developments above mentioned, confirm the potential for interesting findings.

This study takes up this opportunity, and analyzes a long time series, available from a 1-million-day run of CCM0 in perpetual January mode (see [10]). At each time step (every 12 hour) the high dimensional quantity observed (the stream function at 300mb as produced by the model at a grid of points covering the entire globe) is summarized into a lower – 5 – dimensional state vector, as explained in detail in Section 2. Using such long record (2 million observations at 12 hour intervals), we first aim at estimating the functional form which governs the evolution of the state vector from one time-step to the next, in the lower dimensional phase space. We do so by means of a flexible nonparametric modeling tool: neural networks. We estimate the map, and assess the robustness of the resulting functional form, by using independent segments of the time series for the estimation process. Once stable estimates are obtained, we possess an accurate representation of the dynamics governing the evolution of the state vector through its phase space. This first stage is described in Sections 3 through 6. We then proceed to separate the dominant linear component from the nonlinear part, and concentrate our attention on the latter. We summarize the local behavior of the dynamics by computing the Jacobian of the nonlinear component of the map at a regularly spaced grid of points in the phase space. The Jacobians, strang out as vectors, are then clustered, as explained in Section 7.

We can now identify subregions of the phase space according to cluster membership. We then use the clusters obtained from the regular grid to classify a much larger set of observations, and observe how the two main clusters separate neatly, particularly for chosen 3-dimensional projections of the 5-dimensional domain, in Section 8. As the last step, reported in Section 9, the two subregions are analysed as separate in order to construct linear approximations of their internal dynamics. Once the separate blocks are identified and fitted by separate linear models, we can evaluate the power in approximating the nonlinear system by piecewise linear

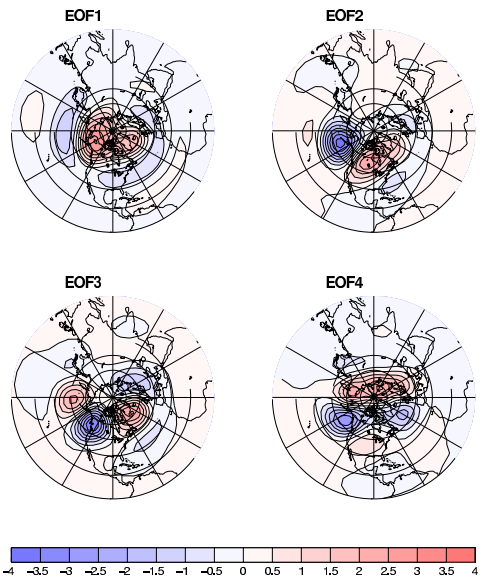

Figure 1. *The first* 4 *empirical orthogonal functions.*

subdynamics. Conclusions follow.

Hopefully this work will excite interest in the subject and thoughts of alternative models. After all, how many 1-million-day-long time series does a statistician see in a life time?

2 Data Structure

The basic data is global 300mb stream function twice per day on 7.5° longitude by approximately 4.4° latitude grids, for a 1-million-day run of CCM0 in perpetual January conditions. At each time step the global stream function field is projected onto the linear space of the first 5 empirical orthogonal functions (EOFs), and the corresponding 5 time coefficients constitute the coordinates of the state vector, whose time series we analyze. EOFs are the principal component basis vectors of the Variance-Covariance matrix of the gridded stream function (actually of its spectral coefficients). The matrix is estimated from the time series, and the first 5 EOFs explain approximately 50% of the variance of the eddy fields. When recombined, they would reproduce the statistically relevant modes of variability of the large-scale planetary waves, predominantly active in the Northern hemisphere. We show in Figure 1 the reconstruction of 4 of the 5 EOFs. The large part of the activity is concentrated over the Northern hemisphere, which makes our analysis – although based on global measures – qualitatively consistent with previous studies (especially [4]), that traditionally analyzed the large scale circulation over the Northern hemisphere and the Southern hemisphere separately. The time coefficients have been centered, thus eliminating a possible source of spurious low-frequency variabil-

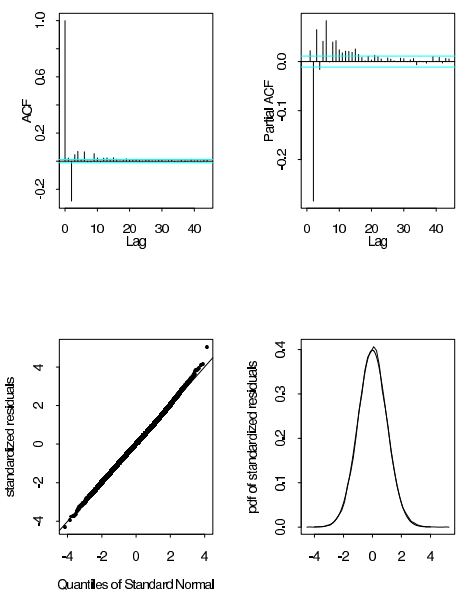

Figure 2. *Standardized residuals from the linear regression of EOF 1 at time t on EOF 1 through EOF 5 at time* $t-1$.

ity. Also, because the numerical experiment was done under permanent January conditions, we are not concerned with low-frequency seasonal cycles, which would introduce masking features.

3 Linear regression fit

The regression function estimated by least squares under the traditional assumption of a linear model as in

$$x_t = Ax_{t-1} + \epsilon_t, \tag{1}$$

where x_t represents the vector of the 5 time coefficients (from now on simply defined as the 5 EOFs), A is a 5×5 transfer matrix, and $\epsilon_t \sim MN(0,\sigma^2 I)$, accounts for over 88% of the variation of each of the components of the state vector. The transfer matrix so estimated, $\hat{A}$, is strongly diagonal, suggesting that each element of the state vector could be in first approximation modeled as an autoregressive process of low order.

The analysis of the residuals from the multivariate linear model fit (1), however, indicates that organized patterns – subtle but nevertheless significant – are left unexplained. Figure 2 refers to the standardized residuals from the linear fit of EOF 1 over the entire – lagged – state vector. The autocorrelation function indicates some correlation – of magnitude 0.2 or less – at the first few lags, and the quantile-quantile plot together with the estimated probability density distribution (superimposed to a standard normal distribution), give a slight indication of skew-

ness to the right, still significant when considering that these estimates are derived from 1 million observations.

The wealth of observations at our disposal permits us to pursue even such subtle hint (less than the 12% of the total variability is left unexplained by the linear regression). We train a neural network by a large amount of data, so that any weak signal of organized nonlinearity will be translated into a functional relation by the flexible nonparametric regression forms that neural networks are capable of fitting.

An alternative approach to modeling the dynamics by higher order AR processes is possible, but we choose not to pursue it. Our preferences go towards modeling the single lag structure, for a number of reasons. The CCM0 output at time $t+1$ only uses the state at time t as input, and we are trying to describe this process by our statistical models. In addition, previous work in this area, referred to in Section 1, has focused on the single lag dynamics and we want to compare our findings to it.

4 Neural Networks

Single-layer neural networks fit the following functional form:

$$\tilde{A}(x) = \gamma_0 + \sum_{l=1}^{h} \gamma_l \Phi \left(\alpha_l + \sum_{j=1}^{k} \beta_{lj} x_j \right). \tag{2}$$

The sigmoid function is usually taken as

$$\Phi(x) = \frac{1}{1+e^{-x}}$$

or a rational polynomial approximation to it. We fit 5 independent neural net regressions, one for each of the components of the state vector separately.

The design of the neural network requires an arbitrary choice for the number h of parallel transformations to operate on the regressors. The larger this number, the more flexible the fitted function will be, but this also translates into a larger cost in terms of computing time, degrees of freedom lost, and, possibly, overfitting. We train and test different architectures (i.e. different h values in (2)), and we see a steady increase in the relative goodness of fit up to $h = 4$ or 5, whence the increments become negligible, and the training and test errors level off. We choose $h = 4$ for the final functional form to be estimated. Hence, (2) becomes more specifically:

$$\hat{x}_{t+1} = \gamma_0 + \sum_{l=1}^{4} \gamma_l \Phi \left(\alpha_l + \sum_{j=1}^{5} \beta_{lj} x_{tj} \right), \tag{3}$$

and is fitted to each of the 5 components of the state vector separately.

Note that the optimal value for h being greater than one is an implicit confirmation that besides a linear component, nonlinear structure is present.

We train the 5 neural nets by a set of $6 * 10^5$ observations, found to be a large enough sample size to stabilize the result of the fit: independent, non-overlapping subsequences of the whole time series give results in accordance (we specify later what we mean by this). The 5 maps thus fitted constitute the cornerstones of our analysis.

5 "Slices" of the 5-Dimensional Phase Space

Consistent with the findings in Section 3, the joint distribution of the original time series is well described – at least in first approximation – by multivariate gaussian distributions centered at zero, and so are all the lower dimensional joint distributions for subsets of components of the state vector. We fix the value of three EOFs in turn and analyse the instantaneous direction of the trajectories in the residual bidimensional phase space. A justifiable choice for the fixed values is their mean, zero in all cases. Figures 3 and 4 give a suggestive depiction of the nonlinear dynamics that surface from this simplified perspective. For each couple of EOFs the first panel shows the vector field from the full nonlinear map as fitted by the neural net. Each point in the bidimensional phase space corresponds to a configuration of the state vector for which the two coordinates that label the axes assume the corresponding values and the residual three are set to zero. Each arrow originates from one such point, and aims in the direction of the subsequent step as predicted by the neural net maps. For the same origin points, the second panel just shows the directions contributed by the linear component of the maps. The linear component has been obtained by a first order Taylor expansion of the functional form fitted by the neural nets around the origin, as detailed in Section 6. Notice that the two pictures are almost indistinguishable, since the linear part accounts for the dominant process, and "hides" the nonlinear component. The third panel isolates the nonlinear component, obtained by subtracting the linear component from the complete map. The lengths of the vectors have been rescaled so that the longest vectors in the three panels have the same length, hence are not indicative of the relative weight of the two components. It is stunning how the differences in the first two maps, so similar at first sight, turn out to draw a coherent and organized pattern of flows. Coherence and organization are highlighted in the fourth panel, obtained by "connecting the arrows" of the third. This fourth panel is an artificial construction, obtained by recursively applying the nonlinear maps to just two coordinates of the state vector at a time, with the remaining three components fixed at their mean level of zero. In addition, the origin of each line in the graph is a point of a regular grid, that uniformly spans the ranges of the two EOFs under examination. Therefore, new features may appear in this last panel that were not in the third, as the origin of the arrows in the first three graphs correspond to a random sample of just 1000 observations from the whole time series. (We limit to 1000 points the representation to avoid cluttering up the figures.)

The whole picture of the time-dependent flow can be obtained by projecting the 5-dimensional trajectories, as predicted by the full set of 5 nonlinear maps onto the bidimensional planes. As can be expected, the projected trajectories appear

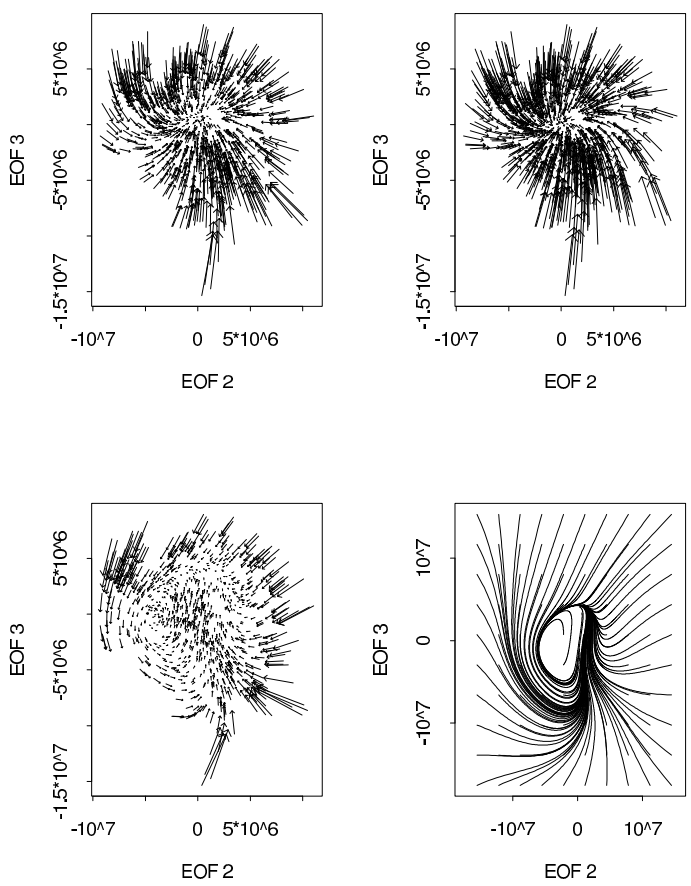

Figure 3. *EOF 2 – EOF 3.*

much noisier, with regard to the directions of the flow and the location and number of fixed points and limit cycles. As an example, Figure 5 shows such projection for the plane of EOF 2 – EOF 5. Compare it to the fourth panel in Figure 4.

6 Trajectories in the Full 5-Dimensional Phase Space

The bidimensional flows presented in Section 5 all converge to either a limit cycle (EOF 2 – EOF 3 plane, in Figure 3) or a number of fixed points (all the other planes, of which we only show EOF 2 – EOF 5). Each single step of the flow has been derived by applying the neural network map, $\tilde{A}(\cdot)$ of Equation (2), to the state-vector at time t, and subtracting the linear component, computed as the first term of the Taylor expansion, centered at the origin. The result is:

$$\tilde{x}_{t+1} = \tilde{A}(x_t) - J(\tilde{A})|_{\mathbf{0}} x_t, \tag{4}$$

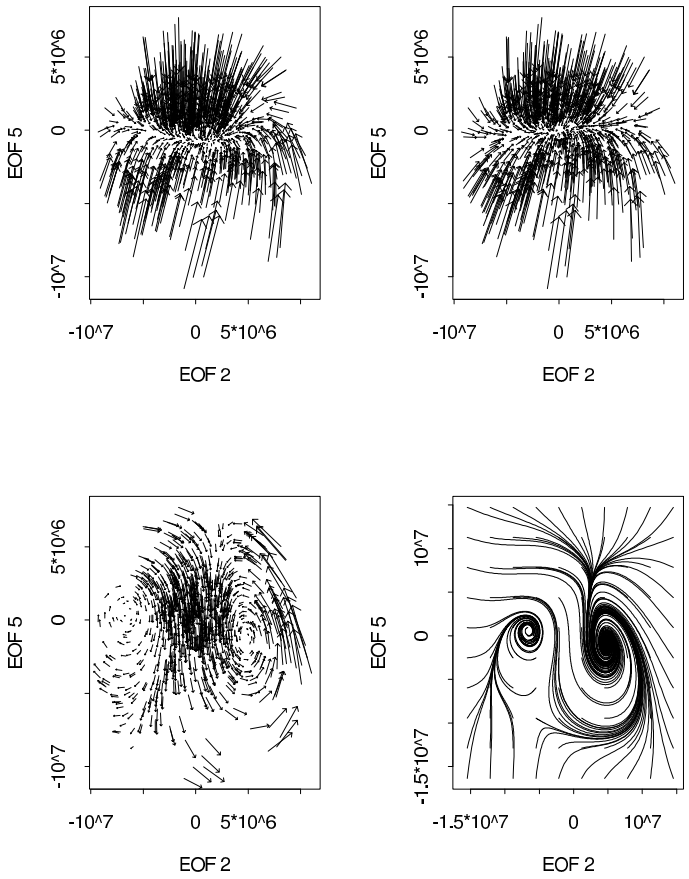

Figure 4. *EOF 2 – EOF 5.*

where $J(\tilde{A})|_{\mathbf{0}}$ indicates the Jacobian matrix of $\tilde{A}$ computed at 0. $\tilde{x}_{t+1}$ represents the increment from time t to time $t+1$ as due to the nonlinear component of the map. The complete step is computed as

$$x_{t+1} = x_t + \tilde{x}_{t+1}$$

which amounts to setting the linear part to the identity transformation and letting just the nonlinear map drive the motion.

Separating the phase space into subdomains with respect to the limit of the nonlinear trajectories, although a very natural proposition, turns out to be unpractical. Here are some reasons:

- the form of (2) and (4) makes exact mathematical derivations impossible;
- conducting a numerical search for the boundaries of the basins of attraction (hyperplanes in the 5 dimensional space) is computationally intractable;

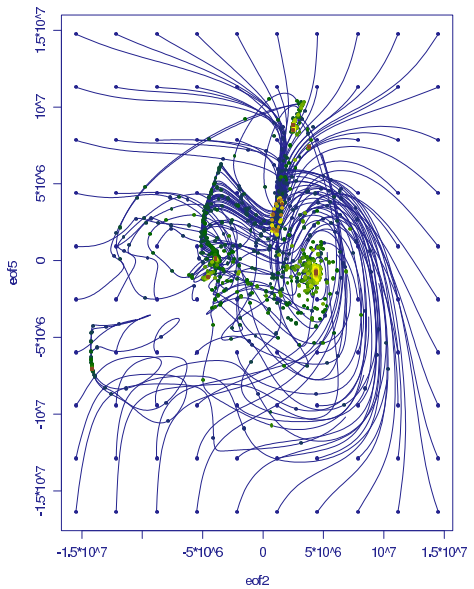

Figure 5. *EOF 2 – EOF 5.*

- slight perturbations of the coefficients of the fitted maps, either due to round-off errors or statistically unsignificant variations in the estimated values, produce diverging long-term trajectories and qualitatively different limit behaviors, especially when the full 5-dimensional dynamical system is observed through time, rather than the simplified bidimensional summaries of Section 5. Such perturbations, however, have no impact on the local, instantaneous quality of the dynamics.

This last remark suggests the use of local – instantaneous – dynamical features, rather than global limit behavior, as a reliable criterion for discriminating among subregions of the phase space.

7 Jacobians as a measure of local behavior

Specifically, we use the information contained in the Jacobians of the nonlinear map to characterize the local dynamics in different regions of the phase space.

For a map $F : X^n \to X^n$ – in our case the recursive relation defined in Equation 4, which maps the 5-dimensional domain onto itself – we can compute the Jacobian at any given point. The set of Jacobians, for a number of points in the domain, can be then clustered with respect to a distance measure, so to define homogeneous subregions and distinguish them from one another.

Hierarchical clustering is an expensive method, because of the need of computing distances between each pair of elements. We get around the problem by coarsely sampling the 5-dimensional domain, over a regular grid that spans evenly

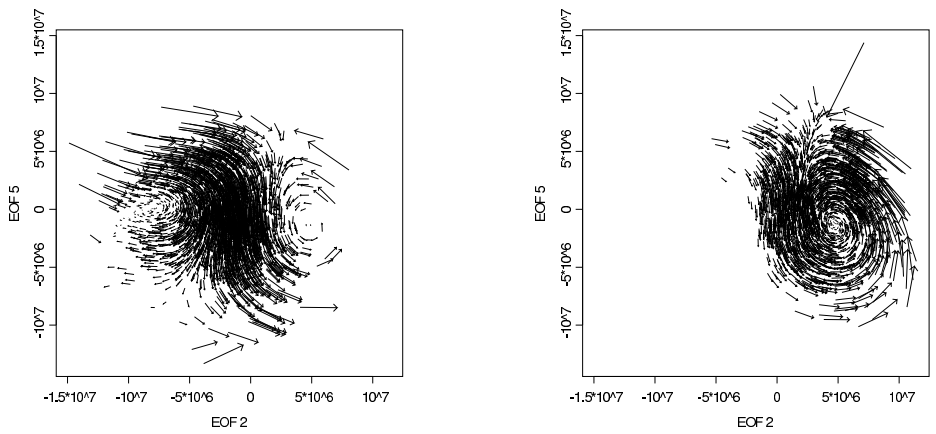

Figure 6. *Trajectories for subset of observations corresponding to the two clusters. EOF 2 – EOF 5 plane.*

each direction out to two standard deviations from its mean value. Hierarchical clustering is then applied to the Jacobians computed at the evenly spaced points of the 5-dimensional grid. The centers of the main clusters obtained (i.e. the mean of the Jacobians in each clusters) become the starting values for k-means clustering applied to a large subset of the observations.

We repeated the procedure starting from neural network maps estimated from a non-overlapping segment of the time series, to ensure that the cluster assignments remain the same, across independent fits. Satisfactorily, 85% of the observations are consistently allocated.

8 Boundaries between subdomains of the phase space

The clustering renders an implicit partition of the 5-dimensional phase space: each of the points is now labeled with respect to two kinds of local behavior as identified by the clustering of the Jacobians.

To illustrate the coherence of such classification with the intuitive separation of the phase space that surfaces from Figure 4 we look separately at local trajectories for points assigned to each cluster. Figure 6 shows how the swirls of Figure 4 separate rather neatly, the inexact boundaries being due to the oversimplified dynamics of the swirls in two dimensions, which hid a more complicated behavior in five dimensions. The clustering, in turn, is based on the full 5-dimensional dynamics, and the two representations cannot therefore adhere perfectly.

The spatial nature of the clusters is further characterized in Figure 7. The entire observational set has been clustered around the means obtained in the first

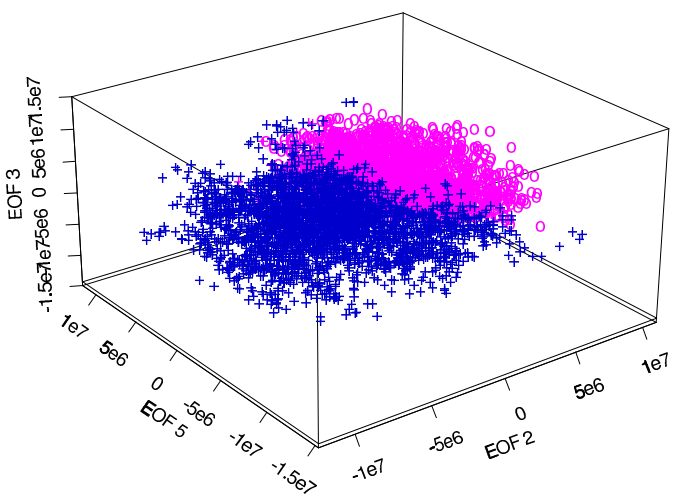

Figure 7. *Three-dimensional projection of the observations used for hierarchical clustering. The two clusters are coded by crosses and circles.*

hierarchical stage. A 3-dimensional projection of the observations (originally 5-dimensional vectors) is plotted, and each point is coded with respect to its cluster membership. A linear plane roughly perpendicular to the EOF 2 – EOF 5 plane and parallel to the EOF 5 – EOF 3 plane seems to divide neatly the two subsets of points.

If one is interested in identifying the position of the hyperplane separating the two clusters, tree models (see for instance chapter 9 of [1]) may help. Structure the dataset in predictors (EOF values) and predictand (cluster label) and apply binary recursive partitioning that successively split the full set into more and more homogeneous subsets by using values of one predictor at a time as splitting rule.

The collection of all splitting rules applied until the stopping criterion is met (either the "leaves" of the tree are perfectly homogeneous or it is unfeasible to proceed further) defines a piecewise linear boundary in the EOF space that optimally divides the clusters.

Besides an analytical form for the approximated boundaries, the tree delivers another important piece of information: the relative importance of the 5 coordinates in the discrimination process, indicated by the order in which the splits take place along the tree. Such order, in our case, highlights the importance of EOF 2 in combination with EOF 3, EOF 4 and EOF 5, whereas the first coordinate (EOF 1) is never used as a splitting rule.

Figure 8 shows the boundaries obtained by tree models in the coordinate system EOF 2 – EOF 3 and EOF 2 – EOF 5.

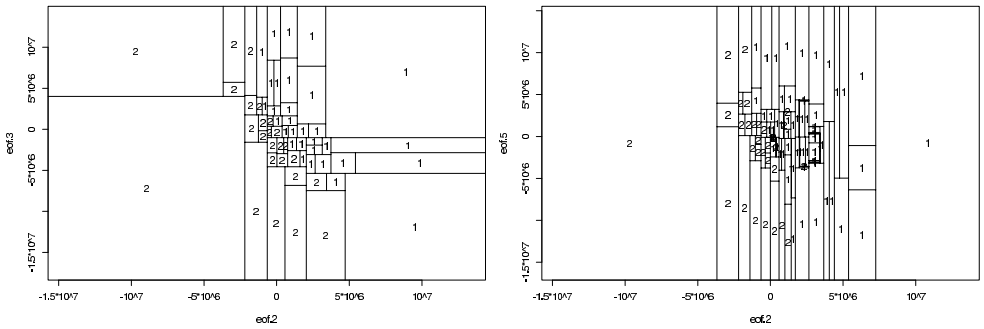

Figure 8. *The piecewise linear boundaries detected through recursive partitioning. Same bidimensional subspaces as in Figure 3 and 4.*

9 Subregional Dynamics and their Linear Approximation

How well can we approximate the nonlinear process underlying the time series of observations by piecewise linear components? Working separately on each of the two clusters, we match each value of the state vector to the observation following it in the original time series. The first state vector constitutes the regressor, the state vector that followed it in the observed time series (disregarding its cluster membership) constitutes the dependent variable in the linear regression.

The separate estimation shows that it is the pattern of interactions as expressed by the off-diagonal elements which represents the distinguishing factor. The differences in the coefficient values are all statistically significant, based on the large number of observations used to estimate them (on the order of 1.2 millions for one submodel, 0.8 millions for the other). Some of them are actually inverting the sign of the interaction between couples of EOFs.

Such differences become larger when we repeat the piecewise estimation after isolating the 'boundary' region. For this purpose we define as 'central' in the clusters those observations that belonged to one of the clusters in the 2-million time series and were separated from observations in the other cluster by at least 8 time steps in both temporal directions. The features are the same as the ones obtained in the two-piece estimation, but the magnitude of the differences are accentuated.

The next question is: how much of the total nonlinear dynamics can be approximated by pieceing together these different linear models? We show in table 1 an answer to the question.

In the table, we compare the percentage of the total mean variation in the quantity $t_t = x_t - x_{t-1}$ (tendency) explained by 5 alternative models: a simple linear model fitted to the entire domain, the linear component of the neural network map as defined in Section 6, a piecewise linear model made of two components fitted separately to the two clusters of observations as defined in Section 7, another made of three components, fitted to the *centers of the clusters* and the *boundary region* as previously defined, and the complete neural network map. The numbers in the

R2 %	LM	DELIN	2PLM	3PLM	NN
GLOB	3.45	3.42	3.52	3.67	3.63
CC1	3.71	3.75	3.98	4.23	4.09
CC2	3.25	3.18	3.42	3.50	3.54
BOUND	3.45	3.40	3.39	3.52	3.48
C1	3.55	3.52	3.63	3.83	3.72
C2	3.36	3.32	3.43	3.52	3.54

Table 1. Percentage of the total mean variation in the quantity $t_t = x_t - x_{t-1}$ (tendency) explained by the alternative models (by column: linear model (LM), linear component of the neural network map as defined in Section 6 (DELIN), piecewise linear model by two components (2PLM) – corresponding to the two clusters fitted separately, piecewise linear model by three components (3PLM) – the two central parts of the clusters and the boundary region fitted separately, neural network map (NN). Rows correspond to measures applied to the global phase space (GLOB) and, separately, the two centers of the clusters (CC1 and CC2), the boundary region (BOUND) or the two full clusters (C1 and C2).

increase in R2 %	LM	DELIN	2PLM	3PLM	NN
GLOB	0.00	-0.87	2.03	6.38	5.22
CC1	0.00	1.08	7.28	14.02	10.24
CC2	0.00	-2.15	5.23	7.69	8.92
BOUND	0.00	-1.45	-1.74	2.03	0.87
C1	0.00	-0.85	2.25	7.89	4.79
C2	0.00	-1.19	2.08	4.76	5.36

Table 2. As in Table 1, but the numbers here represent percentage increase in the explanatory power of the models, using the simple linear model as baseline

table may strike for their small order of magnitude. Their importance changes though when we rescale them by taking the value of the simplest linear model as the baseline and compare the contribution of the alternative more complex ones against it. Table 2 reproduces these results.

10 Conclusions

The weather tomorrow will be very similar to today's. Actually it will be even closer to the 'mean state' than today's is. This in layman's terms expresses the view that the larger part of the evolutionary dynamics in the atmosphere is a simple linear process, where the state vector x_{t+1} will be well approximated by x_t.

Our analysis was aimed at showing that something besides this linear behavior can be modeled in a significant and robust fashion.

We have applied neural network estimation and clustering techniques to de-

scribe the map of an atmospheric dynamic process, and then discriminated among different subregions of its phase space. Our results show the presence of a statistically significant – even though subtle – nonlinear component, robustly estimated in terms of instantaneous directions of the trajectory in its phase space.

Two neatly divided subregions have been characterized, and the results of applying piecewise linear models to them have been compared to two alternative representations – the single linear model and the full nonlinear mapping of the neural network fit.

The gains in the nonlinear representations – even if mostly unsignificant when compared to the full process variance, where the linear part is the dominant component swamping all the rest – become interesting and significant when compared to the part of the process left unexplained by the simple linear model.

The availability of large amounts of data has been fundamental to our study. The features surfacing could not have been detected with smaller data sets. The methods we use scale well to large problems, thanks to the numerical efficiency of neural network training, and the two stage clustering, that first detects a separation by hierarchical methods applied to a smaller set of observations, than uses the clusters' means so derived for k-means aggregation. We find that the combination of neural networks for estimating the autoregressive map followed by clustering on Jacobians is a useful way to understand a dynamical system. Also it allows to compute local Lyapunov (singular) values of the dynamics, and future work will address this, among other aspects of the characterization of the local dynamics.

Bibliography

[1] J. M. Chambers, and T. J. Hastie, *Statistical Models in S*, Wadsworth & Brooks, Pacific Grove, CA, (1992), pp. 377–419.

[2] J. G. Charney, and J. G. DeVore, *Multiple Flow Equilibria in the Atmosphere and Blocking*, Journal of the Atmospheric Sciences, 36 (1979), pp. 1205–1216.

[3] A. R. Hansen, and A. Sutera, *On the Probability Density Distribution of Planetary-Scale Atmospheric Wave Amplitude*, Journal of the Atmospheric Sciences, 43, 24, (1986), pp. 3250–3265.

[4] A. R. Hansen, and A. Sutera, *Weather Regimes in a General Circulation Model*, Journal of the Atmospheric Sciences, 47, 3, (1990), pp. 380–391.

[5] M. Kimoto, and M. Ghil, *Multiple Flow Regimes in the Northern Hemisphere Winter. Part I: Methodology and Hemispheric Regimes*, Journal of the Atmospheric Sciences, 50, 16, (1993), pp. 2625–2643.

[6] M. Kimoto, and M. Ghil, *Multiple Flow Regimes in the Northern Hemisphere Winter. Part II: Sectorial Regimes and Preferred Transitions*, Journal of the Atmospheric Sciences, 50, 16, (1993), pp. 2645–2673.

[7] F. Molteni, S. Tibaldi, and T. N. Palmer, *Regimes in the Wintertime Circulation over Northern Extratropics. Part I: Observational Evidence*, Quarterly Journal of the Royal Meteorological Society, 116, (1990), pp. 31–67.

[8] B. B. Reinhold , and R. T. Pierrehumbert, *Dynamics of Weather Regimes: Quasi-Stationary Waves and Blocking*, Journal of the Atmospheric Sciences, 110, 9, (1982), pp. 1105–1145.

[9] P. Smyth, K. Ide, and M. Ghil,*Multiple Regimes in Northern Hemisphere Height Fields via Mixture Model Clustering*, Journal of the Atmospheric Sciences, 56,(1999) pp. 3704–3723.

[10] D. L. Williamson, D.L., *Description of NCAR Community Climate Model (CCM0B)*, NCAR Technical Note, NCAR/TN-210+STR (1983).