INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **19**: 821–834 (1999)

PRINCIPAL PREDICTORS†

W.C. THACKER*
Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, FL 33149, USA

Received 16 April 1998
Revised 5 November 1998
Accepted 6 November 1998

ABSTRACT

Principal predictors are linear combinations of variables from one set that efficiently describe the collective variability of those from a second set. Their defining eigenproblem is similar to that of canonical correlation analysis, and when the two sets are taken to be the same, principal predictors reduce to principal components. Within the context of a forecast model for the circulation in the Gulf of St. Lawrence, they are shown to be capable of providing a low-dimensional characterization of high-resolution model dynamics.

KEY WORDS: principal predictor analysis; Gulf of St. Lawrence; water circulation

1. MOTIVATION

High-resolution numerical models for predicting near-term climatic conditions can be run only on the largest, most expensive computers. Empirical models, on the other hand, are quite inexpensive, but their skill is limited by the few decades of observations from which they can be inferred. Why not combine these two approaches: use a high-resolution model to simulate seasonal-to-interannual variability over many decades and base an empirical model on the simulated data? The resulting model might capture all of the low-frequency, large-scale predictive skill of the numerical model.

Identifying such a model from simulated data would not be an easy task. First, there is the problem of *too many* data. Which fields should be sampled, at what resolution, how often, and for how long? What sort of filtering or averaging is needed? How will they be stored and manipulated? Can existing software handle the high data volume? Then there are the important issues of how to formulate a low-dimensional empirical model. In order to avoid most of the problems associated with too many data, the context for this study has been shifted from climate to wind-driven shallow-water circulation. The question of how to make a simple model that captures low-frequency behaviour remains essentially the same.

The approach taken here is to replace the many variables of the high-resolution model by a relatively few linear combinations that account for most of the variability at a specified time in the future. To distinguish these linear combinations from principal components, which account for contemporaneous variability, they are referred to as *principal predictors*. This approach is explored for modeling the wind-driven circulation in the Gulf of St. Lawrence. The goal is to determine how many principal predictors capture how much of the low-frequency variability.

Principal predictors have much in common with principal components (Jolliffe, 1986). The essential difference is that, while principal components are intrinsic to a single set of variables, principal predictors link two sets. In this regard, principal predictor analysis resembles canonical correlation analysis (Johnson and Wichern, 1992). But while that seeks the most predictable features without regard for whether those features are energetically significant, principal predictors are designed to account for as much predictand

* Correspondence to: Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, FL 33149, USA. E-mail: thacker@aoml.noaa.gov

Published in 1999 by John Wiley & Sons, Ltd.

variance as possible. This issue has also been addressed by the singular value decomposition of the cross-correlation matrix for the two sets of variables (Bretherton *et al*., 1992); actually, both canonical correlation analysis and principal predictor analysis can be regarded as a metric-based singular value decomposition of the cross-correlation matrix.

After the shallow water model and the simulated data are described in the next section, the theory of principal predictors is developed in the following one. The first step of principal predictor analysis is to remove the linear dependencies of the original variables, which is done via a principal component transformation as described in the section entitled 'Principal components of the simulated data'. The issue of which principal components should be identified as characterizing only the peculiarities of the training sample and thus discarded is resolved by comparisons with independent data. Similarly, in the following section, independent data are used to judge how well the principal predictors are able to account for future low-frequency variability.

2. MODEL AND SIMULATED DATA

An objective of this study is to determine how well a relatively small number of principal predictors can capture the low-frequency dynamics of a high-resolution numerical model. Although this study was motivated by the desire for more economical climatic predictions, a model[1] for the wind-driven circulation in the Gulf of St. Lawrence offers a computationally easier context in which to work. Because the model is based on the linear shallow-water equations, it is clear that the state of the future circulation depends linearly on the present state and on the subsequent forcing. Complications due to non-linear dynamics[2] are thus avoided, allowing attention to be focused on how few variables capture how much of the low-frequency variability. Because the seasonality of the circulation enters only through the wind, and the dynamics are time-independent, there is the hope of capturing the dynamics with fairly short simulations.

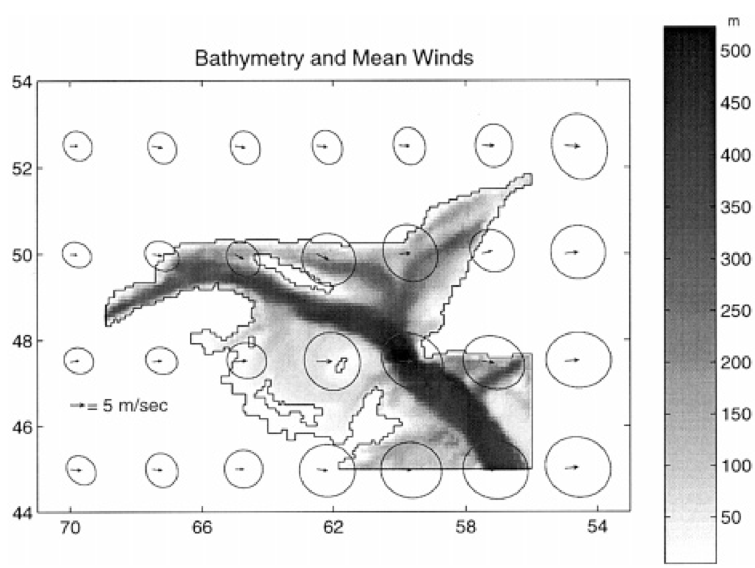

Figure 1. Shades of grey indicate the depth in meters within the Gulf of St. Lawrence, arrows indicate the annual mean wind speed in m/s, and ellipses indicate wind variability. Map co-ordinates are degrees west longitude and degrees north latitude

[1] The model was provided by the Department of Oceanography, Dalhousie University.

[2] If a non-linear model was used, it would be necessary to argue that its low-frequency behaviour can be captured by a linear model. A familiar example is the linear parameterization of non-linear Reynolds stresses.

Figure 1 shows the boundaries of the numerical model's computational grid, an Arakawa C grid with one-twelfth degree resolution. The straight segments in the southeast corner are open boundaries separating the Gulf from the Atlantic Ocean, which are modeled with radiative boundary conditions (Chapman, 1985). The Strait of Belle Isle in the northeast is closed off and modeled with the same no-flow boundary conditions used for land boundaries. Similarly, the St. Lawrence River is closed off at the western boundary. Flow is modeled in the Jacques Cartier Passage and the Honguedo Strait on either side of the Ile d'Anticosti and in the Northumberland Strait by Prince Edward Island. The Iles de la Madeleine are lumped into a single island and Ile Lameque is replaced by an artificial strait between Miscon Island and the peninsula south of Chaleur Bay, but these discrepancies are unimportant for present purposes. Shading indicates that the greatest depths lie along a direct path passing through the Cabot Strait at the Gulf's entrance and proceeding northwestward to the St. Lawrence River.

The model is forced by wind stress computed from 10 m wind analyses from the European Centre for Medium Range Weather Forecasts, which are available twice daily with 2.5° resolution. Surface stress components were computed at this resolution using bulk formulae (Large and Pond, 1982), interpolated bilinearly to the higher-resolution model grid, and then linearly interpolated in time to provide values for every 2 min time step. Three years of wind data were at hand, starting at 00:00 January 1, 1985. The first 2 years were used to generate training data from which principal predictors could be identified, and the third year was used to generate data for verifying their predictive skill. Figure 1 shows the annual mean wind vectors and their standard deviation ellipses based on all 3 year's data. There is a strong seasonal component to the wind's variability, as well as an interannual component, which suggests that a 2 year training record is really inadequate. However, as this is an exploratory study focusing on the principal predictor methodology, this limited wind record was deemed sufficient.

The model's initial state was generated by integrating forward from rest for 9 days using the annual mean wind stress and then for a tenth day using wind stress interpolated linearly between the annual mean and the stress for January 1, 1985. The model was then integrated for 3 years and surface elevations and velocity components were saved twice daily. Present values of the state variables, together with the present and next wind stress values, completely determine the state of the numerical model for the next 12 h, so they should provide a sufficient basis for an empirical 12 h forecast.

For simplicity, the two velocity components were interpolated to cell centers from their positions on the edges of the C-grid cells. Thus, the state of the circulation is described by three numbers for each of the 5977 cells, and the forcing is described by two stress components at 26 locations[3]. The objective of the empirical model is to capture the dynamical evolution of as much of the spatial details as possible using far fewer variables. The training record is limited to 2 years of twice daily values[4], so it can distinguish no more than 1460 independent modes of variability. Clearly, there is the possibility that significant variability that the model might generate will not have been sampled in this limited interval, and much of the variability that is captured reflects the idiosyncrasies of the sample and should be regarded as 'noise'. At most, a few hundred variables can be expected to prove useful. While principal components are known to be most efficient for capturing the variability, principal predictors might be better suited to the task of capturing the dynamics.

3. THEORY OF PRINCIPAL PREDICTORS

Principal predictors relate two sets of variables: a pool of potential predictors p_i, $i = 1, \ldots, i_{\max}$, and a collection of variables q_j, $j = 1, \ldots, j_{\max}$, that are to be predicted. Within the present context, the

[3] The wind data at the southwest and northwest corners of the domain do not influence the model simulations.

[4] No attempt was made to remove the mean seasonal cycle from the data for two reasons. First, the structure of the model generating the data was independent of season with information about seasonality entering only via the imposed forcing, so the same sort of model should be reasonable for the low-frequency approximation. Second, the 3 year record was insufficient for defining a meaningful seasonal cycle to remove.

predictors are the current values of the model state and *forcing variables*, so $i_{\max} = (3 \times 5977) + (2 \times 26) = 17983$; the predictands are the values of the state variables 12 h in the future, so $j_{\max} = 17931$.

The defining idea of a principal predictor π is that it should be well-suited for predicting *each* of the many predictands, i.e. it should be as highly correlated as possible with *all* predictands. Thus, the sum of the squared correlation coefficients for all predictands should be maximized. Furthermore, a principal predictor should be a linear combination of the variables in the predictor pool. If the predictors are represented collectively by the vector p and the coefficients defining the linear combination by α, then $\pi = \alpha^T p$. Finally, the principal predictors should be mutually uncorrelated. The first principal predictor π_1 accounts for the largest possible fraction of the collective variability of the predictands, the second π_2 accounts for the largest fraction of the remaining variability, etc. The problem of defining a set of principal predictors π_k, $k = 1, \ldots, k_{\max}$, becomes that of specifying the corresponding coefficient vectors α_k.

Maximizing the sum of squared correlation coefficients leads[5] to an eigenvalue problem for the coefficient vectors:

$$X^T[\operatorname{diag} Q]^{-1} X\alpha_k = r_k^2 P\alpha_k, \tag{1}$$

which involves three matrices P, $\operatorname{diag} Q$ and X. The first is the predictor covariance matrix, $P_{i,i'} = \operatorname{cov}(p_i, p_{i'})$; it accounts for the fact that the predictors are correlated. The second is the diagonal of the predictand covariance matrix, $[\operatorname{diag} Q]_{j,j} = \operatorname{var}(q_j)$; it appears because the correlations with all predictands are treated equally and serves to standardize each to unit variance. The third is the cross-covariance matrix between predictors and predictands, $X_{j,i} = \operatorname{cov}(q_j, p_i)$; it provides the predictive link. The eigenvalue r_k^2 is the sum of the squared correlation coefficients for π_k, and the eigenvalue α_k is the coefficient vector defining π_k.

The eigenproblem (1) involves two symmetric matrices $X^T[\operatorname{diag} Q]^{-1}X$ and P (Wilkinson, 1965). Qualitatively, it is very like the usual one matrix eigenvalue problem for principal components[6] (Kendall *et al.*, 1983). From the symmetry of these two matrices it follows that

$$\alpha_k^T P \alpha_{k'} = 0, \tag{2}$$

if $r_k^2 \neq r_{k'}^2$. The left-hand-side of this orthogonality condition for the coefficient vectors can be recognized to be $\operatorname{cov}(\pi_k, \pi_{k'})$, which indicates that the principal predictors are indeed uncorrelated.

If all the variables in the predictor pool were statistically independent[7], P would be positive definite and thus serve as a proper metric, the normalization $\alpha^T P \alpha = 1$ for the coefficient vector being equivalent to choosing π to have unit variance. This could be the case theoretically, but as P must be estimated from data, it cannot be the case in practice unless the number of samples used for estimating P is greater than the number of predictor variables. The 17983 predictors considered here would require a training record of at least 25 years. Fortunately, the singularity of P poses no essential problem[8]; it only means that some of the coefficient vectors defining the principal predictors are not unique, due to the fact that some members of the predictor pool can be estimated from the others. All that is needed to correct the situation is to replace the predictor pool by a smaller set of variables, e.g. all principal components with positive variance, removing only redundant information. Thus, P should be interpreted as the covariance matrix of the new, smaller pool of predictors, and p should be interpreted as the vector of these variables.

---

[5] See Appendix for the derivation.

[6] If the predictors and predictands are the same set of variables, $X = P = Q$ and (1) reduces to $P\alpha_k = r_k^2 \operatorname{diag} P\alpha_k$, which in turn reduces to the eigenvalue equation for correlation-matrix principal components (Thacker and Lewandowicz, 1996). In that context, $\operatorname{diag} P$ serves as a metric; variables are measured in units of their own standard deviation.

[7] Here, statistical independence means that no variable in the pool can be fully explained by a linear combination of other variables from the pool.

[8] This same situation is encountered in canonical correlation analysis (Kendall *et al.*, 1983), which involves the similar eigenvalue equation, $X^T Q^{-1} X a_k = p_k^2 P a_k$, and the singularity there can be handled in the same way.

The fact that the matrix $X^T[\text{diag}\, Q]^{-1}X$ may be singular is not important. For example, if the predictand set contains only a single variable, the rank of this matrix is 1; the single non-zero eigenvalue corresponds to the only useful principal predictor[9]. Just as the principal predictors with no sample correlations to any of the predictands should be ignored, so should those with small eigenvalues, as they reflect the idiosyncrasies of the sample from which the covariance matrices have been estimated. In this regard, choosing principal predictors to discard is similar to the problem of deciding which principal components constitute ‘noise’. In both cases, the best way to proceed is to see how well they account for independent data (Thacker and Lewandowicz, 1997).

The set of principal predictors can be regarded as an orthogonal rotation of the principal components of the variables in the predictor pool. The principal components are ordered according to their ability to explain the observed variability of the predictors themselves, while the principal predictors are aligned to explain variability of the predictands. In general, each predictor p_i is related to the uncorrelated principal predictors π_k via multiple regression:

$$p_i = \sum_k \text{cov}(p_i, \pi_k)\pi_k + e_i, \quad i = 1, \ldots, i_{\text{max}}, \tag{3}$$

where the principal predictors are normalized to have unit variance. When the sum is over all principal predictors, even those corresponding to zero eigenvalues, all residuals ϵ_i for the training data vanish. On the other hand, if some principal predictors are discarded as noise, some fraction of the variability of the data is ignored and the training residuals are increased. Note that the EOF-like vector $\beta_k = P\alpha_k = \text{cov}(p, \pi_k)$ represents how well the kth principal predictor accounts for the variability of the individual predictors.

Similarly, there is a multiple regression model for each of the predictands:

$$q_j = \sum_k \text{cov}(q_j, \pi_k)\pi_k + \varepsilon_j, \quad j = 1, \ldots, j_{\text{max}}. \tag{4}$$

The vector of Gauss–Markov weights $\xi_k = X\alpha_k = \text{cov}(q, \pi_k)$ is like an out-of-set EOF in that it indicates the influence of the principal predictor π_k on the various predictands[10]. When there are more predictors than samples, all of the sample variability of the predictands can be explained and the residuals ε_j for the training data can be zero. Limiting the predictor pool to a subset of the principal components of the original variables and limiting the number of principal predictors to those with large eigenvalues r_k^2 will increase the size of the residuals for the training sample, but might decrease those for independent data.

Equation (4) constitutes a low-dimensional model. Within the present context, 1460 principal components are sufficient to capture all of the variability exhibited by the 17983 predictor variables and the 17931 predictand variables over the 2 year training interval, while most of this variability is captured by a few hundred. Note that these models might collectively be regarded as being a vector autoregressive model with the wind forcing entering as in a transfer function model (Wei, 1990). If the wind stress variables were removed from the predictor pool, it would be entirely vector autoregressive.

The idea here is to include in the predictor pool all variables that are expected to have a substantial impact on what is to be predicted. In the example considered here, the winds play a major role in the evolution of the circulation, but the present circulation cannot be argued to have any influence on the future winds. In fact, the present winds might be poor predictors of their own evolution, which is better explained by a synoptic scale model, so including the wind stress variables in the set of predictands was thought to be counterproductive. Instead, the empirical model should be regarded exactly like the original numerical model, where some additional model is needed to forecast future winds needed for several step ahead forecasts[11].

[9] In this example principal predictor analysis reduces to canonical correlation analysis.

[10] Davis’ (1977) ‘principal estimator patterns’ are essentially the same as the principal predictor patterns β_k and ξ_k.

[11] To make such extended forecasts, the predicted model state is combined with independently predicted wind stress to evaluate the future principal predictors, which are then used to take the next step.

It is useful to note the similarities to canonical-correlation forecasting (Graham *et al.*, 1987). Seeking canonically correlated predictor–predictand pairs is fruitless because the large numbers of variables involved guarantees many very highly correlated pairs, which need not account for any appreciable fraction of the total variability and which are not likely to be reflected in independent data. Graham *et al.* (1987) not only use principal components to encapsulate the robust behaviour of the predictors, they also characterize the predictands by their principal components, and then they express the dynamical relationships between these two reduced sets as canonical correlations. In principal predictor analysis, the explicit compact representation of the predictands is generally not necessary.

Principal predictor analysis can also be compared with principal-oscillation-pattern analysis (von Storch *et al.*, 1988). That method typically focuses on a two-dimensional slice through a single three-dimensional field, which is encapsulated as a handful of principal components. A first-order vector autoregressive model is fitted to the components, and the patterns (POPs) are computed as eigenvectors of the autoregressive model matrix, the eigenvalues specifying the frequency of oscillation and rate of decay. Just as each principal predictor pattern evolves into its future correlate, each POP evolves into its 'complex' correlate, *viz.*, a rotated and dissipated version of itself. However, as principal predictors are not limited to autoregression, this suggests generalizations of POPs that include a transfer function component of the model to accommodate causes of change external to the field itself.

4. PRINCIPAL COMPONENTS OF THE SIMULATED DATA

As explained above, because the 2 year interval provides fewer samples than the number of model variables, it is necessary to define the principal predictors from a compressed representation of the original predictors. This is accomplished here using principal components. Because different types of variables serve as predictors, it is important to decide how to measure them on a common scale. One possibility is to use an energy metric[12] (Thacker, 1996). Another is to re-scale each variable type so that the total variance of each field is the same (Xu, 1992). The approach adopted here is to measure each variable in units of its sample standard deviation, which results in the use of correlation matrix principal components. A possible drawback is that shallow regions, where there is little kinetic energy, have as much influence as deep regions. Another is that wind stress might be under-represented, as only the values at 26 wind-grid points are used. However, if all principal components are retained when defining the principal predictors, this is not an issue, because collectively they describe all of the observed variability. The specific method used in defining the principal components becomes important only when some of the components are discarded as noise.

Which of the principal components should be regarded as being mostly noise? For correlation-matrix principal components, Kaiser's criterion (Mardia *et al.*, 1979) is that correlation-matrix principal components corresponding to eigenvalues less than unity are suspect. Here, there are 99 with eigenvalues greater than unity, and the remaining 1360 could be discarded by this rule. The only really good way to decide how many components carry predictive information is to examine their ability to explain independent data. Principal components for the verification data are defined as linear combinations using the coefficients that were determined from the training data; likewise, the estimates of the individual variables are linear combinations of the components defined by coefficients determined from the training data (Thacker and Lewandowicz, 1996).

Figure 2 summarizes the ability of correlation-matrix principal components to account for the behaviour of the individual variables in the predictor pool. The top panel indicates how well the first few components reproduce the data comprising the 2 year record from which they were derived, while the lower panel shows their ability to approximate the independent behaviour encountered during the third

[12] Each surface elevation value could be scaled so that its square would be proportional to the local potential energy and each velocity could be multiplied by a function of the local depth to reflect its contribution to the kinetic energy. Stress values are more problematic, at best representing the potential for transferring energy into the flow over the 12 h period.

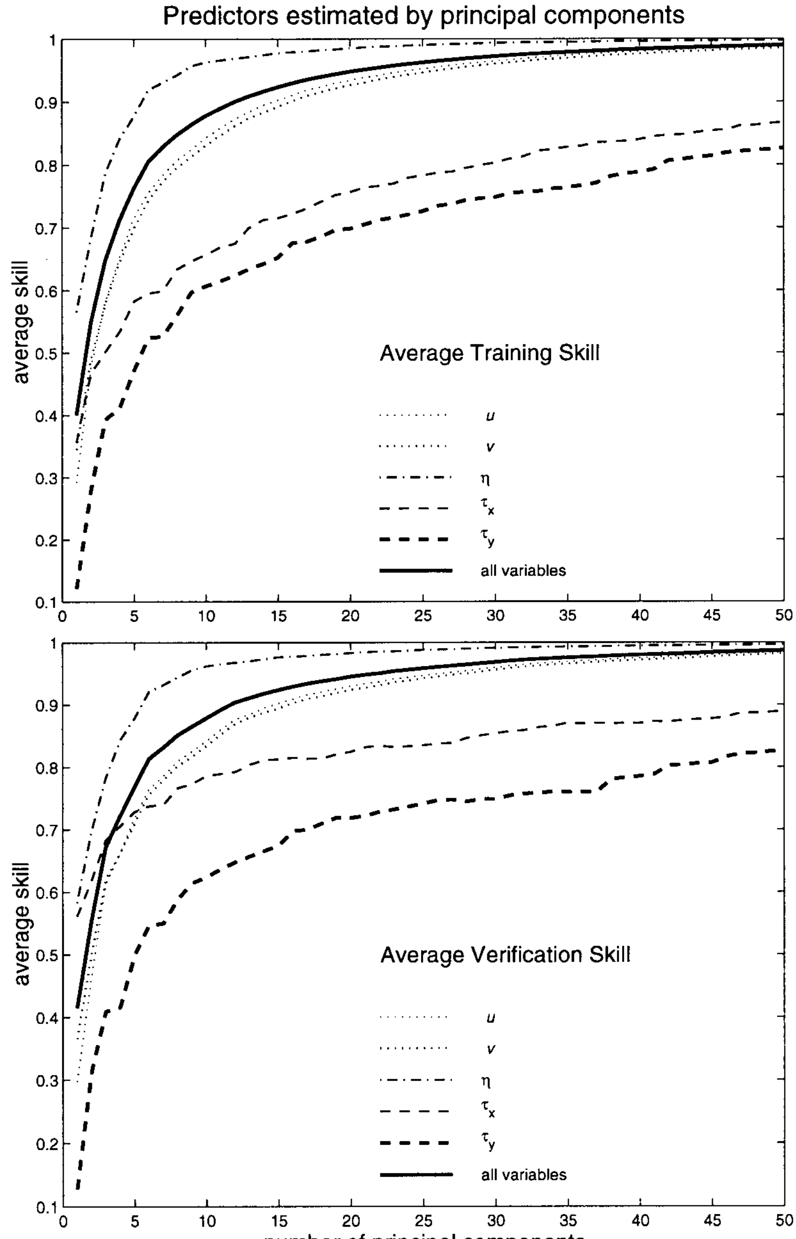

Figure 2. Average skill with which each of the 17983 state and forcing variables are estimated by their correlation-matrix principal components. The different curves correspond to averages over the indicated fields

year. Note that the verification skills are remarkably similar to the training skills. All of the first 50 principal components appear to be useful for characterizing the behaviour of the predictor variables in the independent data to be used for verification.

The skill with which any individual variable is estimated is defined in terms of the average of the squared error of the estimate: $s = 1 - \langle \epsilon^2 \rangle / \sigma^2$, where the angular brackets indicated an average over the record and where σ is the sample standard deviation determined from the training set. For training skill, this reduces to the ratio of the variance of the estimate to σ^2; for a single principal component, this is the square of its correlation coefficient with that variable; and for several components, because they are uncorrelated, the skill becomes the sum of the squared correlation coefficients. However, this is not the case for independent data, so the definition in terms of errors must be used.

The various curves correspond to averages of the individual skills over various subsets of the variables in the predictor pool. The eigenvalues of the correlation matrix represent the sum of the training skills with which the corresponding principal components explain the individual variables, and because the components are uncorrelated, the sum of the first few eigenvalues give the total combined training skill for the corresponding set of components (Thacker and Lewandowicz, 1996). Thus, the solid curve in the top panel of Figure 2 could be computed either by averaging squared sample correlation coefficients or by normalized partial sums of eigenvalues.

The first few principal components characterize the 5977 surface-elevation variables η better, on average, than they do the velocity components (n, v). These principal components carry relatively less information about the 26 wind stress vectors (τ_x, τ_y). However, components with indices greater than 20 carry more information about wind stress, as can be seen from the fact that the dashed curves are increasing faster than the other curves.

Figure 3 shows how well the principal components of the present state and forcing variables could account for the variability of the state variables 12 h later. Conceptually, this involved fitting a large number of linear models, like (4) with principal predictors replaced by principal components, to the training data: one for each variable and for each combination of principal components. However, because the principal components are uncorrelated and have unit variance over the training interval, the coefficients that enter these models are simply the sample covariances of the individual principal components with the individual variables.

The top panel of Figure 3 shows how well the principal components of the predictors were able to be fitted to the first 2 years data for the predictands. The overall average skill (solid curve) levels off at less than 0.8, indicating that the first 50 principal components do not carry as much information about what will be happening 12 h in the future as they do about what is happening now. If the curves were extended to the right, they would be seen to converge to unity as more components are used; the solid curve reaches a skill of 0.8 for 70 principal components, 0.9 for 491, 0.99 for 1129, and the complete set of 1460 principal components explains all of the behaviour of the predictands in the training interval, even though they were computed from the predictors. This can be most easily understood in terms of time series: the fit is perfect because the 1460 principal component time series are orthogonal functions of time that are capable of representing any zero-mean function over 2 year intervals. However, this will not be the case for independent data, which are not used to determine the parameters used in the estimates.

Note that, while the first few principal components carry the greatest amount of information about the contemporaneous surface-elevation field (Figure 2), they carry substantially less information about the state of that field 12 h in the future. In fact, they provide better 12 h ahead estimates of the velocity than of elevation. Such results might be different if some other metric were used to define the principal components. However, as the focus here is on principal predictors, no effort has been made to track down the cause of this behaviour. Note, however, that similar results are to be expected for the principal predictors, as they are simply linear combinations of the principal components. They derive their predictive ability from the principal components, which obtained theirs from the entire predictor pool, so they cannot be expected to provide additional information about the future state of the elevation field. They can only be asked to provide a more compact representation of the limited information that is available. The interesting question of what limits the information that can be exploited statistically still remains.

The bottom panel of Figure 3 shows how well the principal components estimate the predictors during the verification interval. Again, the similarity with the training-skill curves of the top panel is remarkable. Fifty principal components of the variables in the predictor pool clearly carry useful information about the future state. However, principal components after the 10th individually carry little predictive information. All curves have leveled off and there is no indication that basing principal predictors on more than 50 will improve 12 h forecasts. The objective here is to determine whether considerably fewer principal predictors might do as well as the 50 principal components.

5. PRINCIPAL PREDICTORS OF THE SIMULATED DATA

The above results suggest that, for predictions involving independent data, there is no advantage to using more than 50 principal components for defining the principal predictors for 12 h ahead forecasts of the circulation in the Gulf. Before examining whether this is indeed the case, it is useful first to demonstrate that the theory of principal predictors is fully supported by the training data.

Figure 4 shows that principal predictors provide a more efficient explanation of the predictands in the training interval than do principal components. The lower curve is the average training skill with which

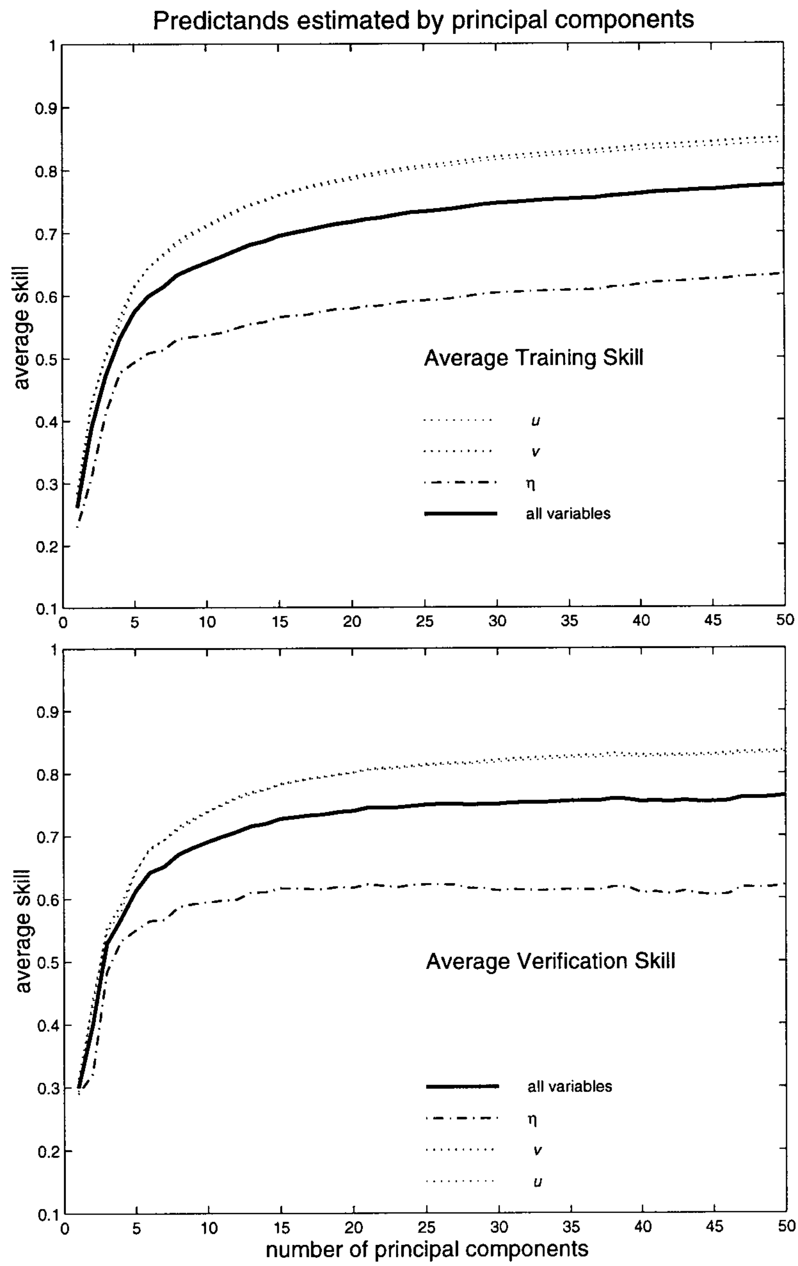

Figure 3. Average skill with which each of the 17931 state variables 12 h in the future are estimated by principal components of the current state and forcing variables

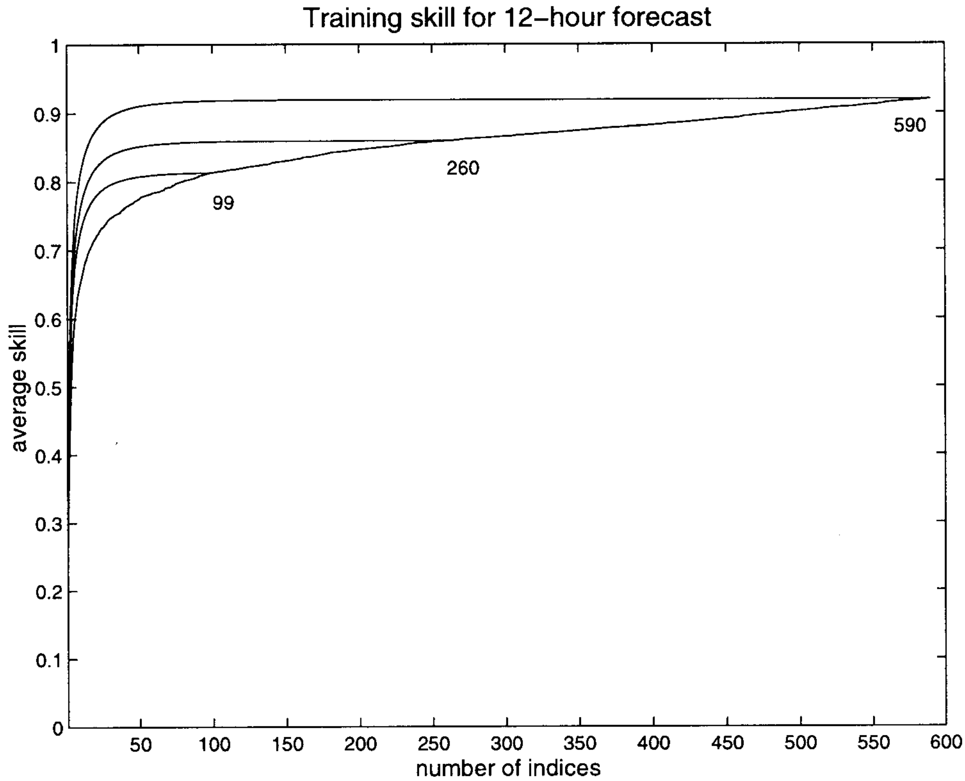

Figure 4. Lower curve is average training skill, based on fitting to a 2 year record, with which the predictands are estimated by principal components. Upper curves are training skills for estimation of principal predictors based on 99, 260 and 590 principal components

principal components account for the collective behaviour of the 17931 predictands. The curve labeled 99 is the average skill as a function of the number of principal predictors, when the first 99 principal components[13] comprise the representation of the predictor pool; note that there are only 99 principal predictors, the first few of which account for the bulk of the skill, and all 99 affording exactly the same skill as the first 99 principal components. When the reduced representation is expanded to 260 principal components[14], there are 260 principal predictors, and the first few of these explain an even greater fraction of the 12 h ahead variability. And the first few principal predictors based on 590 principal components[15] carry even more information about the future[16]. If all 1460 principal components were used to define the principal predictors, and all 1460 principal predictors were used to determine the future model state, the training skill would be perfect.

Thus, the principal predictors do what they were designed to do. The eigenproblem is seen to be formulated correctly. Furthermore, it is clear that, however many principal components are used to characterize the variables of the predictor pool, no more 50 principal predictors are needed to characterize most of the details of the circulation in the Gulf 12 h in the future. However, this conclusion is based on the 2 year training record. The real test is how well they characterize future variability for independent data, and the answer is expected to depend on the number of principal components used to characterize the robust variability of the initial state and forcing variables.

[13] These correspond to eigenvalues greater than unity, a standard selection criterion.
[14] These correspond to eigenvalues greater than 10^{-1}.
[15] These correspond to eigenvalues greater than 10^{-3}.
[16] If the principal predictors were based on all 1460 principal components, the training skill curve would reach unity.

Figure 5 shows the training and verification skills for the 12 h forecast using principal predictors based on the first 50 principal components. The first 20 principal predictors do about as well as all 50, both for the training data and for the verification data. As expected, the training skill was significantly higher for the first few principal predictors than for the same number of principal components (Figure 3), the differences diminishing as the numbers increase. Comparing the lower panels of Figures 5 and 3 shows the differences in verification skills to be much less dramatic. However, the hard-to-predict elevation field was much better predicted with fewer than 15 principal predictors than with the same number of principal components. So, if the Gulf were to be modeled with fewer than 15 variables, principal predictors would

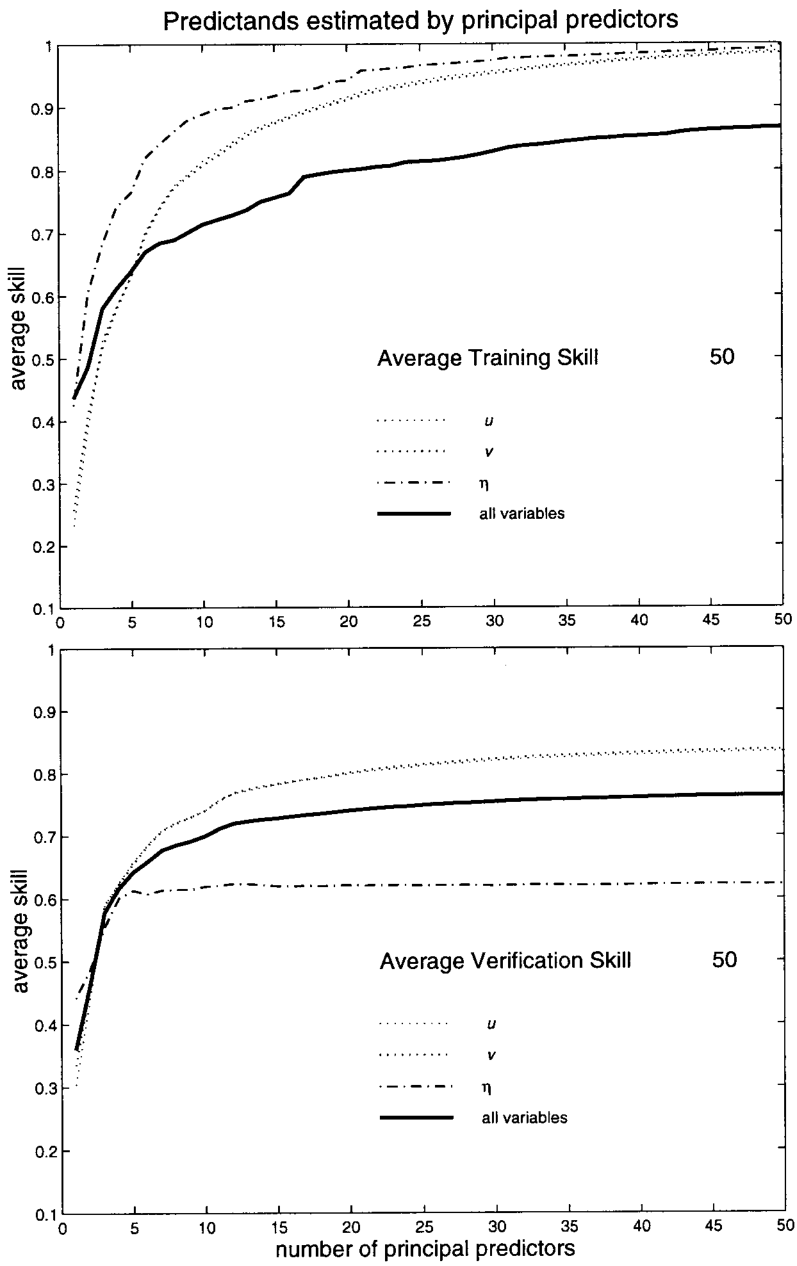

Figure 5. Average skill with which the predictands are estimated by principal predictors

offer an advantage over principal components. For example, ten principal predictors can account for about two-thirds of the variance, on average, of each of the 17931 variables characterizing the high-resolution surface-elevation and velocity fields.

Twelve-hour-ahead predictions made using principal predictors based on the first 99 principal components gave essentially the same results, reinforcing the conclusion that little or no additional predictive skill was carried by the higher principal components. When the principal predictors were based on the first 260 principal components, results were worse. Verification skills were noticeably less when using more than five of these principal predictors, with average verification skill for surface elevation decreasing as the number of principal predictors increased beyond five. This fact suggests that the principal components beyond 99 actually carry some misinformation, which contaminates the principal predictors. This conclusion is verified by the even worse results for principal predictors based on 431 principal components: the decrease for elevation skill with more than five predictors was quite dramatic, and more than ten doing worse than only one. The first 50 principal components seem to carry essentially all of the predictive information, so there is no advantage in basing the principal predictors on more.

6. DISCUSSION

The primary purpose of this paper has been the presentation of the theory of principal predictors. While it was motivated by the desire for compact representations of numerical forecast models, its scope is much more general. For example, it might be used to estimate the surface pressure field above a sea surface temperature field. Its applicability, as well as its mathematics, is similar to that of canonical-correlation analysis. It can offer the advantage of associating aspects of the two groups of variables that account for a greater part of the total variability. Within the context of climate, principal predictors have much in common both with POPs and with canonical-correlation forecasting and might offer an alternative worth considering. For example, very recently in a POP-like application, Selten *et al.* (1997) have exploited what are essentially autoregressive principal predictors, i.e. predictors differ from predictands only by a lag in time, to study propagating patterns in the upper ocean temperature field from their climate model.

One aspect of the theory that has not been illustrated here is the utility of the EOF-like spatial patterns associated with the principal predictors. For each principal predictor, there is a pattern of predictors and an associated pattern of predictands that can be displayed graphically, which can be useful for diagnosing patterns of evolution. As the primary emphasis of the paper was not on the dynamics of the Gulf of St. Lawrence nor on the actual behaviour of the empirical model, but on the theory of principal predictors and their ability to capture the variability of the future state, examples of such patterns were not included.

The theory has been verified within the context of simulated data from a numerical model for the Gulf of St. Lawrence. The computed principal predictors behaved exactly as expected. When attention is confined entirely to the training data used in determining their coefficient vectors, principal predictors are seen to provide a significantly more efficient representation of the dynamical relationships than do principal components. This advantage was less dramatic for independent data but still noticeable: fewer than 15 principal predictors exhibited greater verification skill than the same number of principal components.

One point that should be emphasized is the large number of principal components that were found to carry useful information about the set of predictors. While it is commonly accepted that less than a dozen principal components are useful, the others being considered to be noise, here 50–100 seemed to account for robust variability. This should not be rationalized as being due to the fact that the data were generated by a numerical model, as a similar result was found for seasonal mean sea surface temperature data in another study (Thacker and Lewandowicz, 1997). This should be contrasted with the smaller number of variables, roughly 15, that carry information about the state of the circulation of the Gulf 12 h in the future. Still, 15 would be considered a fairly large number when compared with current canonical-correlation forecasting practice.

An unresolved question is why the surface elevation field is so difficult to predict. Clearly, the 2 year training record used to identify expected variability is insufficient, but that does not account for the fact that the difficulty already shows up in the training data. If this problem stems from the use of correlation-matrix principal components to remove the statistical dependencies caused by the large predictor pool and the limited training record, then there is the chance that it can be overcome by using principal components defined using an energy metric, so that the energetically important aspects are given more emphasis. It could also be a consequence of the metric used here for defining principal predictors, which treats the skill for each predictand equally, without regard to its energetic significance. If so, the use of an energy weighted average when deriving the eigenproblem for principal predictors might result in comparable skills for the velocity and elevation fields.

APPENDIX. DERIVATION OF EIGENVALUE PROBLEM

To derive Equation (1), the eigenproblem defining the principal predictors, start by expressing the correlation coefficients in terms of variances and covariances in the sum of squared correlations that are to be maximized:

$$J = \sum_j \mathrm{cor}^2(q_j, \pi) = \sum_j \frac{\mathrm{cov}^2(q_j, \pi)}{\mathrm{var}(q_j)\mathrm{var}(\pi)}. \tag{A1}$$

In the denominator, recognize that $\mathrm{var}(\pi) = \alpha^T P\alpha$, where P is the covariance matrix of the variables from which π is built. Similarly, $\mathrm{var}(q_j) = Q_{j,j}$. In the numerator, $\mathrm{cov}(q_j, \pi) = \mathrm{cov}(q_j, p^T)\alpha$, where $\mathrm{cov}(q_j, p^T)$ is the jth row of the cross-covariance matrix X. So,

$$J = \frac{\alpha^T X^T (\mathrm{diag}\, Q)^{-1} X\alpha}{\alpha^T P\alpha}. \tag{A2}$$

Requiring derivatives of J to vanish gives an eigenproblem for the coefficient vector α:

$$X^T (\mathrm{diag}\, Q)^{-1} X\alpha = JP\alpha. \tag{A3}$$

The eigenvalue r^2 in (1) can be recognized to be the sum of the squared correlation coefficients J.

REFERENCES

Bretherton, C.S., Smith, C. and Wallace, J.M. 1992. 'An intercomparison of methods for finding coupled patterns in climate data', *J. Clim.*, **5**, 541–560.

Chapman, D.D. 1985. 'Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model', *J. Phys. Oceanogr.*, **15**, 1060–1075.

Davis, R.E. 1977. 'Techniques for statistical analysis and prediction of geophysical fluid systems', *Geophys. Astrophys. Fluid Dyn.*, **8**, 245–277.

Graham, N.E., Michaelson, J. and Barnett, T.P. 1987. 'An investigation of the El Niño–Southern Oscillation cycle with statistical models. 1. Predictor field characteristics', *J. Geophys. Res.*, **92**, 14251–14270.

Johnson, R.A. and Wichern, D.W. 1992. *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs NJ.

Jolliffe, I.T. 1986. *Principal Component Analysis*, Springer, New York.

Kendall, M.G., Stuart, A. and Ord, J.K. 1983. *The Advanced Theory of Statistics*, vol. 3, Charles Griffin, London.

Large, W.G. and Pond, S. 1982. 'Sensible and latent heat flux measurements over the ocean', *J. Phys. Oceanogr.*, **12**, 464–482.

Mardia, K.V., Kent, J.T. and Bibby, J.M. 1979. *Multivariate Analysis*, Academic Press, Reading MA.

Selten, F.M., Haarsma, R.J. and Opsteegh, J.D. 1997. *On the mechanism of North Atlantic decadal variability*, KNMI Preprint 97-31, Royal Netherlands Meteorological Institute.

Thacker, W.C. and Lewandowicz, R. 1996. 'Climatic indices, principal components, and the Gauss–Markov theorem', *J. Clim.*, **9**, 1942–1958.

Thacker, W.C. and Lewandowicz, R. 1997. 'A comparison of low-dimensional representations of sea surface temperature anomalies in the North Atlantic', *Int. J. Climatol.*, **17**, 953–967.

Thacker, W.C. 1996. 'Metric-based principal components: Data uncertainties', *Tellus*, **48A**(4), 584–592.

von Storch, H., Bruns, T., Fischer-Bruns, I. and Hasselmann, K. 1988. 'Principal oscillation analysis of the 30- to 60-day oscillation in a GCM', *J. Geophys. Res.*, **93**, 11022–11036.

Wei, W.W.S. 1990. *Time Series Analysis*, Addison-Wesley, Redwood City CA.
Wilkinson, J.H. 1965. *The Algebraic Eigenvalue Problem*, Oxford University Press, Oxford.
Xu, J.-S. 1992. 'On the relationship between the stratospheric quasi-biennial oscillation and the tropical southern oscillation', *J. Atmos. Sci.*, **49**, 725–734.