A COMPONENT-RESAMPLING APPROACH FOR ESTIMATING PROBABILITY DISTRIBUTIONS FROM SMALL FORECAST ENSEMBLES

MICHAEL DETTINGER

U.S. Geological Survey, Scripps Institution of Oceanography, Dept. 0224, 9500 Gilman Drive, La Jolla, CA 92093-0224, USA
E-mail: mddettin@usgs.gov

Abstract. In many meteorological and climatological modeling applications, the availability of ensembles of predictions containing very large numbers of members would substantially ease statistical analyses and validations. This study describes and demonstrates an objective approach for generating large ensembles of "additional" realizations from smaller ensembles, where the additional ensemble members share important first-and second-order statistical characteristics and some dynamic relations within the original ensemble. By decomposing the original ensemble members into assuredly independent time-series components (using a form of principal component decomposition) that can then be resampled randomly and recombined, the component-resampling procedure generates additional time series that follow the large and small scale structures in the original ensemble members, without requiring any tuning by the user. The method is demonstrated by applications to operational medium-range weather forecast ensembles from a single NCEP weather model and application to a multi-model, multi-emission-scenarios ensemble of 21st Century climate-change projections.

1. Problem

Ensembles of weather and climate forecasts from a single model often are used to describe prediction uncertainties associated with uncertain initial conditions. Super-ensembles of forecasts from several models or with several external climate forcings can be used to describe prediction uncertainties associated with model differences and indeterminate future forcings. In weather- and climate-prediction applications, several studies have argued that ensemble means are typically better predictors than are any individual members of the contributing ensembles (e.g., Krishnamurti et al., 2000; Richardson, 2001; Zhu et al., 2002). This finding eventually may be found to extend also to climate-change projections. However, ensemble forecasts also presumably contain information about higher order statistics of prediction uncertainties. Typically, members of an ensemble of weather or climate predictions diverge as time from the initial conditions passes. Under some dynamical conditions, the members can also converge temporarily. Both of these behaviors are much as one might expect from the actual (unknown) forecast uncertainties, with general increases attributable to sensitive dependencies of dynamical

Climatic Change (2006)
DOI: 10.1007/s10584-005-9001-6

systems to initial uncertainties but also with the occasional, temporary confluences of trajectories associated with visits to ghost limit points and cycles, and homoclinic orbits, shared by ensemble members (Ghil and Childress, 1987; Ghil et al., 2002).

If the ensemble includes many members (e.g., Allen and Ingram, 2002), then characterizing the evolving ensemble spread can be as simple as ranking the forecasts for each time and binning the results to directly form histograms or crude probability-distribution functions (pdfs). Even if the pdfs so estimated were crude, they could provide useful measures for comparing forecasts to subsequent observations and would provide a useful basis for comparing different ensembles (e.g., Toth et al., 2003). These pdfs, like all the pdfs discussed herein, are measures of the probabilities of obtaining various outcomes with the models applied and thus are only estimates of the probabilities of various real-world outcomes. Thus the pdfs might more accurately be considered to be *prediction* distribution functions.

When the number of ensemble members is smaller, however, developing even a rough estimate of the forecast pdfs involves assumptions about the character of the forecast uncertainties sampled by the ensemble. One approach is to sort and rank the ensemble forecasts, use them as mileposts of the forecast pdf (say the median forecast value at a given lead time marks the median in the pdf), and then smooth algebraically to fill in interpolated values. Alternatively, one can attribute error bars of some weight and shape to each ensemble member and then essentially sum the error bars from all the ensemble members to arrive at the overall ensemble pdf, but then important assumptions need to be made regarding the growth rate of the error bars for the individual ensemble members. Both of these approaches have the advantage that they are simple, but have the disadvantage that they require critical subjective choices or assumptions by the analyst.

In the present study, a third alternative is presented that, in its simplest form, has no subjectively tunable parameters. The method requires no tunable parameters because it characterizes the ensemble spread by a data-adaptive principal components analysis and then resamples the independent components obtained from that analysis as often as necessary to provide a smooth pdf estimate. Using the orthogonality properties that are designed into principal components analysis (PCA), the resampling method provides an almost unlimited number of other realizations that are distinct from each other while retaining the essential characteristics of the ensemble members, including evolving ensemble means and standard deviations and all the lag and intervariable correlations. Because the method is based on PCA, the components of the resampled ensemble are based on its first and second statistical moments, so that the resulting smoothed pdfs tend toward Gaussian shapes. However, that tendency is relatively weak and, as will be shown with examples here, the method readily handles ensembles that diverge into two or more clusters of trajectories and handles non-Gaussian characteristics like heavy-tails, as a matter of course.

2. Method

Consider an ensemble of n forecasts of, say, temperature at a given model grid cell, each m time steps long and each containing elements $\{x_i^j, i = 1, m\}$ where j indicates which forecast sequence from the ensemble is being referenced. A given forecast sequence can be compiled into a forecast (column) vector x^j, where the superscript j in absence of a subscript indicates that the j-th column vector is being referenced. The resulting n forecast vectors, in turn, can be compiled into an $m \times n$ matrix X. A PCA, with all expectations calculated across the ensemble members – in the n direction, will decompose the original ensemble X into m temporal empirical-orthogonal functions (EOFs; von Storch and Zwiers, 2002, ch. 13), with the k-th temporal EOF vector, $\{e^k, i = 1, m\}$, accompanied by a corresponding vector of EOF coefficients, $\{p^k, j = 1, n\}$, wherein the i-th element, p_i^k, of the k-th coefficients vector is the projection (strength) of the k-th EOF pattern in the i-th ensemble member. Together the EOFs form an $m \times m$ EOF matrix E and the coefficient sets form an $n \times m$ coefficients matrix P, such that

$$X = \mathrm{EP}^T. \tag{1}$$

In the unusual form of PCA decomposition used here – wherein expectations are taken across ensemble members rather than, say, time – the EOFs are the eigenvectors of the correlation matrix $X'X'^T$, where X' is a zero-mean version of X and the superscript T indicates a matrix transpose, so that $X'X'^T e^k = \lambda_k e^k$, where λ_k is the variance in X captured by the k-th EOF (Ghil et al., 2002). Because $X'X'^T$ is a symmetric matrix, its eigenvectors form an orthogonal basis for decomposing and reconstructing the original time series (Blyth and Robertson, 2002). That is, when each e^k vector is scaled to unit length, as in standard EOF analysis (von Storch and Zwiers, 2002), $E^T E$ is an $m \times m$ identity matrix. The coefficients P are the projections of X' onto the EOFs; that is, $p^k = X'^T e^k$. Because the EOFs are eigenvectors of $X'X'^T$ and are orthogonal, $p^{kT} p^m = (X'^T e^k)^T \; X'^T e^m = e^{kT} X'X'^T e^m = \lambda_m e^{kT} e^m = \lambda_m \delta_{mk}$, where δ_{km} equals 1 if $m = k$ and 0 if $m \neq k$. Thus the coefficients for the different EOFs are orthogonal (uncorrelated) with each other, with expectations calculated across the n ensemble members.

The orthogonality of the coefficients means that one EOF's coefficient for a particular ensemble member is uncorrelated (linearly independent) with any other EOF's coefficient for that ensemble member. In standard PCA (or EOF analysis) with expectations collected across time, the dimension spanned by each EOF is grid location (von Storch and Zwiers, 2002) and the value of the k-th coefficient series at time j is uncorrelated with, or tells nothing – on average over time – about, the value of any other coefficient series at the same times. In the PCA applied herein, expectations are instead collected across the ensemble (rather than time) and the dimension spanned by the EOFs is time (rather than grid location). Thus, by substituting "ensemble member" for "time" in the preceding

statement, we see that, in the present analysis, the value of the k-th coefficient series at ensemble member j is uncorrelated with, or tells nothing – on average across the ensemble members – about, the value of any other coefficient series for the same ensemble member. The resampling method developed here uses the independence of the coefficients associated with the various ensemble members to scramble the orthogonal components of the original ensemble without changing its statistics.

Having completed this nonstandard PCA-based decomposition, the i-th time step of the j-th original (properly standardized) ensemble member can be recovered completely by

$$x_i'^j = \sum_{k=1}^{m} e_i^k p_j^k \tag{1'}$$

which is just an expansion of Equation (1) into the column vector (superscripts) and vector elements (subscripts) notation used here. Another forecast vector that is indistinguishable from the original ensemble elements, to second order, can be obtained by picking the j indices in equation (1′) at random, with replacement, from the set $\{j = 1 \text{ to } n\}$, at each step k in the summation. Because the coefficients associated with a given (new or original) ensemble member are uncorrelated from EOF to EOF, by the orthogonality relations described earlier, the second-order statistics of the resulting sum do not depend on which ensemble member's coefficient is chosen at each step as long as – in the limit of many resamplings – all coefficients are eventually included. As before, this point may be made more intuitive by considering its analogy in standard PCA (EOF) analysis wherein new maps that have the same first and second order statistics as the original collection of maps can be constructed as the sums of the products of each spatial EOF, in turn, with one of that EOF's temporal coefficient values at a randomly selected observation time. In the present reconstruction, new ensemble members are constructed as the sums of the products of each temporal EOF with one of its coefficients for a randomly selected ensemble member.)

The procedure is as follows:

1. Calculate the ensemble-mean values μ_i of x_i^j for each time i, with expectation taken across the ensemble:

$$\mu_i = \sum_{j=1}^{j=n} x_i^j \Big/ n \tag{2}$$

and subtract these time-varying means from all of the forecast vectors to obtain an ensemble of centered (zero-mean at each lead time) forecast vectors. Any mean or time variation shared by all the ensemble members is removed by this step. After the ensemble is resampled, this time-varying ensemble mean can readily be added back.

2. Calculate the ensemble standard deviations σ_i of the centered forecasts at each time i, with expectation taken across the ensemble,

$$\sigma_i = \left[\frac{1}{n} \sum_{j=1}^{j=n} \left(x_i^j - \mu_i \right)^2 \right]^{1/2}, \tag{3}$$

and divide the centered forecast vectors at each lead time by the corresponding standard deviation to obtain a standardized forecast ensemble (zero mean and unit variance at each lead time) X', with

$$x_i'^j = \frac{\left(x_i^j - \mu_i \right)}{\sigma_i} \quad \text{for } j = 1, n; \quad \text{and} \quad i = 1, m. \tag{4}$$

This standardization ensures that any temporal evolution of the spread of the ensemble is captured and can be reintroduced (multiplying by σ_i) after resampling. Removing the temporally varying standard deviation at this point in the analysis ensures that inter-ensemble variations in the early part of the forecasts (when the ensemble typically has not spread much) are treated in the same detail as those later in the forecasts.

3. Compute the cross correlations of the standardized forecasts at all different times i and k, with expectations taken across the ensemble,

$$c_{ik} = \frac{1}{n} \sum_{j=1}^{n} x_i'^j x_k'^j \quad \text{for } i = 1, m; \quad \text{and} \quad k = 1, m. \tag{5}$$

The resulting cross correlation matrix is $m \times m$, and summarizes the covariation (averaged across all ensemble members) of each forecast with itself and each of the forecasts at each of the other time steps.

4. This cross correlation matrix is decomposed into EOFs and their coefficients by the PCA described previously. The resulting EOFs describe the temporal evolution of the ensemble members in most economical form (von Storch and Zwiers, 2002). For a given EOF, the coefficients measure the weight (similarity) of each ensemble member in turn to the EOF's temporal pattern. For example, perhaps most ensemble members trend throughout from warmer towards cooler, while a few might warm for a while and then cool like the others. These two behaviors would tend to be captured by two distinct EOFs, and those ensemble members in the former category would have larger coefficients for the former EOF whereas the latter ensemble members would have larger coefficients for the latter EOF.

5. Construct additional "forecast" realizations, x'^l, by recombining the PCA components to generate new combinations of EOF and their coefficients as

often as necessary, for each time step i,

$$x_i'^{l} = \sum_{k=1}^{m} e_i^k \, p^k_{\text{random}(j)} \tag{6}$$

where the index j (in Equation (1′)) is now selected at random at each step k of the summation. Because the coefficients for the various EOFs are linearly independent of each other, because of their origins as projections onto orthogonal eigenvectors, the first and second order statistics of these additional realizations do not depend on which ensemble member's coefficient for a given EOF is mixed with which other ensemble member's coefficient for the next EOF.

With m EOFs, each of which can take on any of the n coefficient values, the number of distinct realizations that can be constructed ideally would be m^n; e.g., in a 10-member ensemble of 14-day forecasts, 14^{10} distinct resamples could be generated. However, typically only a fraction of the coefficient series from a PCA contributes much variance to the reconstructions (which is why PCA is so often used to reduce the dimensionality of complex collections of data). In a typical temporal EOF analysis, perhaps 20% of the EOFs capture, or contribute, much variance (e.g., Ghil annd Vautard, 1991), so that in the resampling method developed here, only a similar fraction of the components would contribute to significant differences between the new realizations. Alternatively, if $n < m$, the number of non-zero PCA components may be roughly equal to n, rather than m, and only these n components would contribute to the new realizations. In either of these cases, the practical number of new realizations that can be generated is reduced. If only 20% of the EOFs contributed significant variance to the reconstructions in the case of the 10-member ensemble of 14-day forecasts, the number of EOFs that would contribute most of the variance in the new realizations might be only 3, and the number of distinguishable realizations would drop to about 3^{10} or 59,000, a considerable reduction albeit still be a useful expansion of the size (10) of the original ensemble. In a typical application of PCA, the choice of how many components to retain is important, because the goal generally is to reduce the dimensionality of the data set, to strip away noisy components. In the present use of the PCA decomposition method, the aim is to retain noise and signal in realistic new realizations. Thus all components need to be retained. Assuming that many components may not contribute much to the new realizations, one can choose to evaluate the results of the PCA to limit the resampling to eliminate very small, or zero, contributors; however, because these "extra" components add little or nothing to the new realizations, one can as easily simply include all the PCA components in the resampling procedure with no significant change in the results. The latter option makes the method strictly reproducible, and eliminates a subjective judgment (what subset of the components should be retained). Thus, in the present application, the latter

option – retaining all the components, whether they contribute much or not – is employed.

6. Having reconstructed a "new" member of the standardized forecast ensemble by resampling, in step 5, rescale the result by the time-varying ensemble standard deviations and then add the time-varying ensemble means, essentially reversing Equation (4). By this rescaling, the time-varying ensemble means and scatter are restored, and the large numbers of resulting forecast realizations can be ranked and summarized in detailed histograms to estimate pdfs in as much detail as desired.

The method ensures that features shared by the original ensemble members are shared by the component-resampled ensemble members, that variations shared by subsets of the ensemble members are reproduced realistically and in proportion to their occurrence in the original ensemble, and – because no components of the original ensemble, however small or erratic, are left out of the reconstructions–even that the noisy (unshared) variations are faithfully captured and reproduced in the component-resampled ensemble.

One way to picture the method is to imagine that the original ensemble X has been filtered into a large number of narrow and nonoverlapping frequency bands. PCA can have this effect (e.g., Dettinger et al., 2001), although it is not constrained to take that form. The result of such filtering is that an ensemble member would be recognized to have power A in the first frequency bin, have power B in the second frequency bin, and so on. These powers would correspond (conceptually) to the EOF coefficients in the PCA described above. Another ensemble member would have a different set of powers in each frequency bin. Now, assuming that an ensemble's power in the first frequency bin had little bearing on its power in the second, and so on, one can imagine generating new ensemble surrogates with the same statistical properties as the original ensemble, by taking the filtrate from one ensemble member (at random) from the first frequency bin, adding to it the filtrate from an ensemble member (at random) the second bin, and so on, until samples from all the frequency bins had been incorporated. The sum of the frequency components would constitute a new time series with statistical and spectral properties that are derived strictly from the ensemble's overall power spectrum. For example, if the 10-day periodicities in the ensemble members were most powerful and 8-day periodicities notably lacking, the resampling would still yield members with powerful 10-day periodicities and weak 8-day periodicities, because the resampling only used observed values from each frequency bin. The present method improves on this hypothetical frequency-binned resampling (i) by guaranteeing – by the construction of the PCA decomposition of the ensemble members – that the various elements resampled (the loading patterns, which would correspond to sine waves of given frequency in the hypothetical) are always independent of each other, and (ii) by allowing more flexibility of loading-pattern shape than is offered by a simple frequency-domain approach.

3. Example Application: Medium-Range Temperature Forecasts

An example of the results of the PCA-based component resampling of a forecast ensemble is provided by application to a 12-member ensemble of 14-day forecasts of air temperatures at the 850-mbar pressure level over Northern California, from the National Centers for Environmental Prediction's (NCEP) operational medium-range forecast (MRF) model (Toth and Kalnay, 1993; Traction and Kalnay, 1993). The MRF ensembles are generated by simulating the next 14 days of weather (globally) from 12 different initial conditions that are intended to represent the general levels of uncertainty about the current conditions at the beginning of the forecasts. Special methods have been developed to ensure that the differences in the initial conditions include perturbations that will diverge most quickly in the model's multidimensional state space (Toth and Kalnay, 1993). This choice of initial-condition perturbations is intended to span the most virulent uncertainties while using the smallest perturbations and smallest number of ensemble members.

The operational MRF forecasts of 850-mbar temperatures at 40°N 120°W, initiated on 11 November 1998 and spanning the 14 days from 12 November to 25 November, are shown as black dots in Figure 1. Initially, the MRF ensemble forecasts are tightly clustered within a few degrees of $+5\,^{\circ}$C, but they gradually diverge until, by the final day of the forecast, they span a temperature range of about $12\,^{\circ}$C. The forecasts warm for a day or two initially, then cool until the midpoint in the forecast cycle and then generally warm yet again. At various times (e.g., day 10), the ensemble diverges into several, apparently distinct clusters of forecasts.

The resampling method described above was applied to this ensemble forecast (an 14×12 forecast matrix) to obtain 10,000 component-resampled realizations, which then were histogrammed and contoured to map the pdfs shown by contours in Figure 1. Gaussian distributions reflecting the daily means and standard deviations of the (original) ensemble are shown as bell curves in the figure as well. The resampled pdfs capture much of the same gradual spread of the ensemble indicated by the daily Gaussian fits, and closely enclose the clustered MRF ensemble members throughout the forecast period. Initially the pdfs are very narrowly focused on the small range of forecasts, and gradually the pdfs spread even more than do the MRF members. This additional spread reflects a tendency for the initial days of the MRF ensemble to be more structured, more parallel, more completely captured by leading EOFs from the PCA, than are the later forecasts, which include variations that are increasingly uncorrelated (noisy) with respect to the other ensemble members. Notice also that the component-resampling method yields pdf estimates that readily follow the general (shared) trends in the MRF ensemble. The resampled pdfs also are able to diverge into several clusters of trajectories, most notably in the example on days 9 to 11 when three clusters of MRF forecasts (dots) develop temporarily and then dissolve into each other again by day 12. Such temporary divergences in the forecast ensemble can correspond to true dynamical bifurcations and divergences in the atmospheric dynamics (e.g., Ghil and Childress, 1987) and

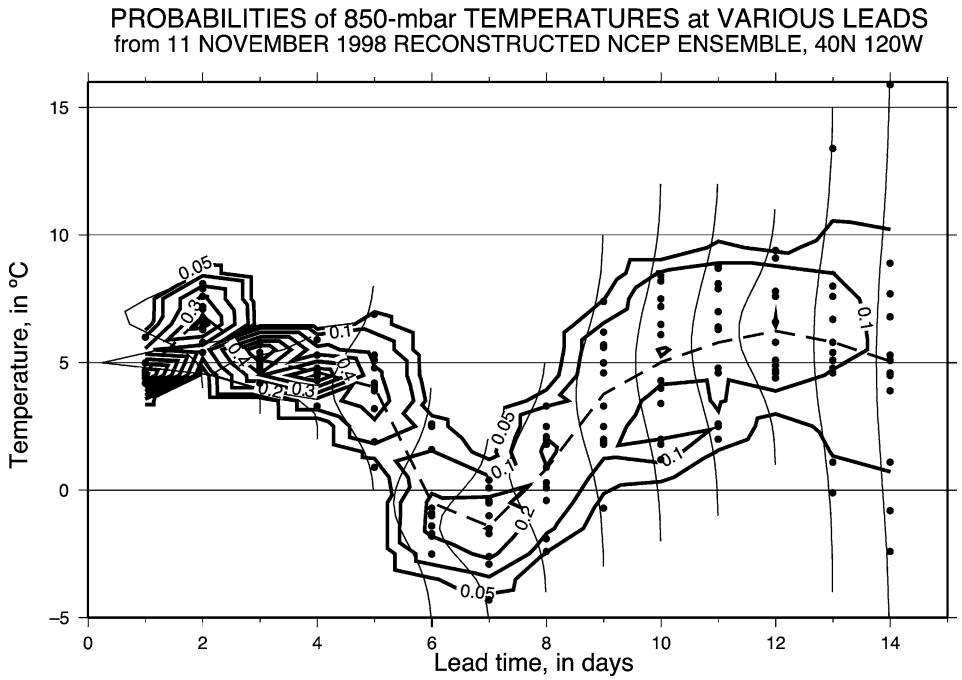

Figure 1. Probability density functions and original ensemble-forecast members during an NCEP medium-range forecast (MRF) cycle beginning 11 November 1998; black dots are individual members of the 12-member operational MRF ensemble; dashed curve traces the original-ensemble means and bell-shaped curves (rotated to vertical) are normal distributions with same standard deviations and means as original ensemble at each time. Contours represent the distributions in an ensemble of 10,000 component-resampled realizations.

thus are important to capture when estimating the forecast pdfs. The simple Gaussian distributions shown in the figure cannot replicate this ensemble behavior. In the present methodology, such forecast divergences are captured in the EOFs and weighted appropriately (in both amplitude and numbers of participating ensemble members) by the corresponding coefficient series. Then, when the coefficient series are resampled randomly, both the shapes and relative frequencies of the divergences are naturally reproduced in a satisfying way.

Simply by extending the MRF forecast ensemble considered, e.g., by appending another 14 days of forecast from another grid location (in the manner of Weare and Nasstrom, 1982), the same resampling method can be used to estimate the joint probability distributions for more than one forecast location or variable. An example application of the methodology to the MRF forecasts of temperatures at two California grid cells is shown in Figure 2. The joint distributions are constructed from loading patterns that capture concurrent variations of forecasts at the two points, so that if, say, among MRF ensemble members, those forecasts that warmed in the south tended to cool in the north, and vice versa, the EOFs would reflect that

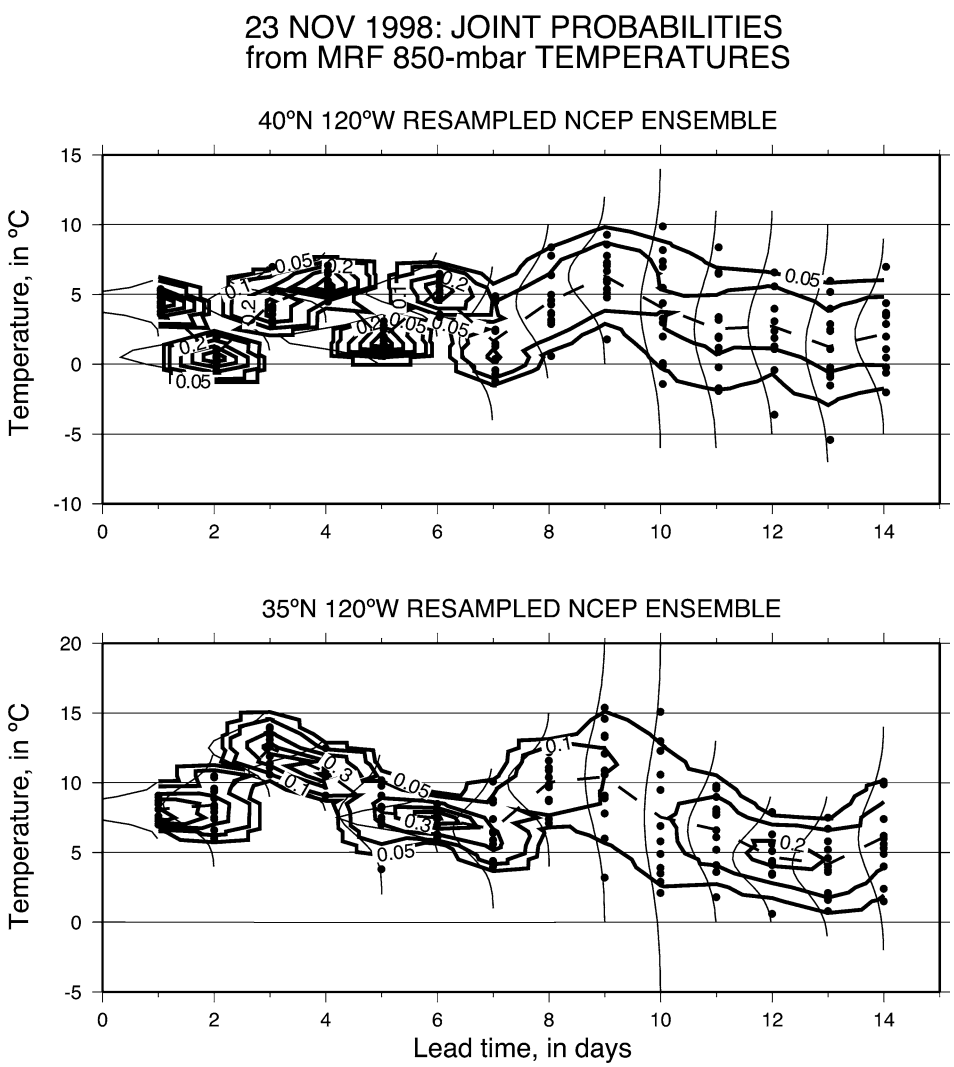

Figure 2. Same as Fig. 1, but showing the joint distributions of temperatures for two California grid cells, for forecasts initiated on 23 November 1998.

joint behavior and the component-resampled realizations would maintain this joint behavior. Thus the component-resampling approach can provide pdfs that are more informed than would be a simpler histogram or smoothing of the raw ensemble members. The component-resampled realizations themselves also provide examples of the kinds of shared variations among forecast points or variables, for use in other calculations.

The component-resampled method does not directly add information to the ensemble of forecasts, but it does simplify manipulation of the probabilities implied

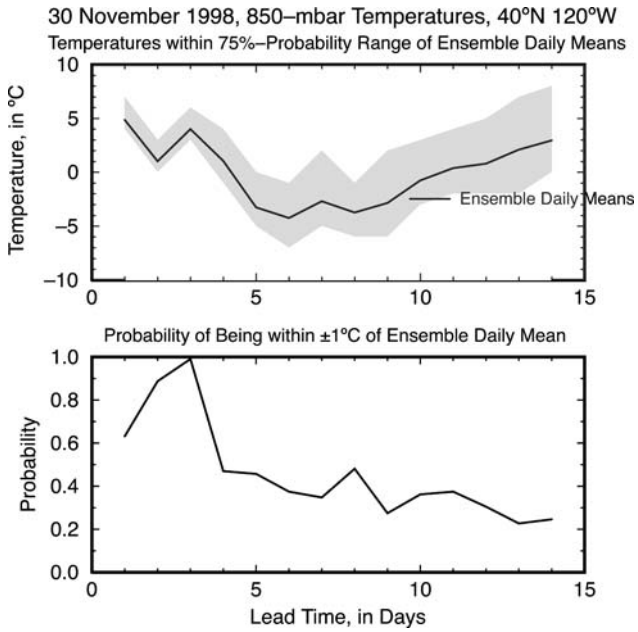

Figure 3. Range of temperatures that fall within the 75%-probability range of the daily ensemble-mean temperature in the NCEP operational forecast ensemble (upper panel) and the daily probability of being within 1 °C of the ensemble mean (lower panel); probabilities were estimated from 10,000 resampled forecast realizations, based on the 30 November 1998 operational forecast ensemble of air temperatures at 850 mbar pressure level above 40°N 120°W.

by a forecast ensemble. The smoothed pdfs obtained from the component-resampled forecast realizations can be used as objective (nonparametric) estimates of relatively fine gradations in the ensemble spread at each forecast lead time, gradations that are much finer than could be directly estimated from the original small ensembles. Two examples of uses of these additional gradations are illustrated in Figure 3. The upper panel shows the daily ranges of temperatures in a resampled forecast ensemble falling within a 75%-probability range of the daily ensemble means. This measure could be used to estimate the risks of various temperature outcomes associated with deviations around the ensemble-mean forecasts or, given an archive of past forecast ensembles, could be used to determine the historical reliability of spread-skill forecast relations linking the ensemble spread at any given lead time to the (historical) skill of the ensemble-mean forecast on the same day.

The lower panel in Figure 3 essentially inverts the probabilities suggested by the upper panel to estimate the probability that a given (resampled) ensemble member falls within 1 °C of the ensemble mean each day. This provides as fine a measure of the ensemble spread as the analyst desires for use in a variety of forecast applications. As with the upper panel, the probability estimates in the lower panel can also be used to hone assessments of the historical forecast skill of the ensemble means.

4. Example Application: Climate-Change Projections

As an illustration of the value of the more detailed pdf descriptions that can be obtained by component resampling, the procedure is applied here to an ensemble of climate-change projections of 21st Century (2001–2099) climate over Northern California. The ensemble considered here was compiled from six climate models, each simulating responses to each of three specified greenhouse-gas-plus-sulfate-aerosols emissions scenarios (Figure 4). The ensemble includes three projections each by the US PCM, Canadian CCCM, German ECHAM4, British HadCM3, Japanese NIES, and Australian CSIRO coupled ocean-atmosphere global climate models; the emissions scenarios are the A2, IS92a, and B2 scenarios (Houghton et al., 2001), which represent projections of relatively rapid, moderate, and intermediate, rates of 21st Century emissions increases, respectively. By considering

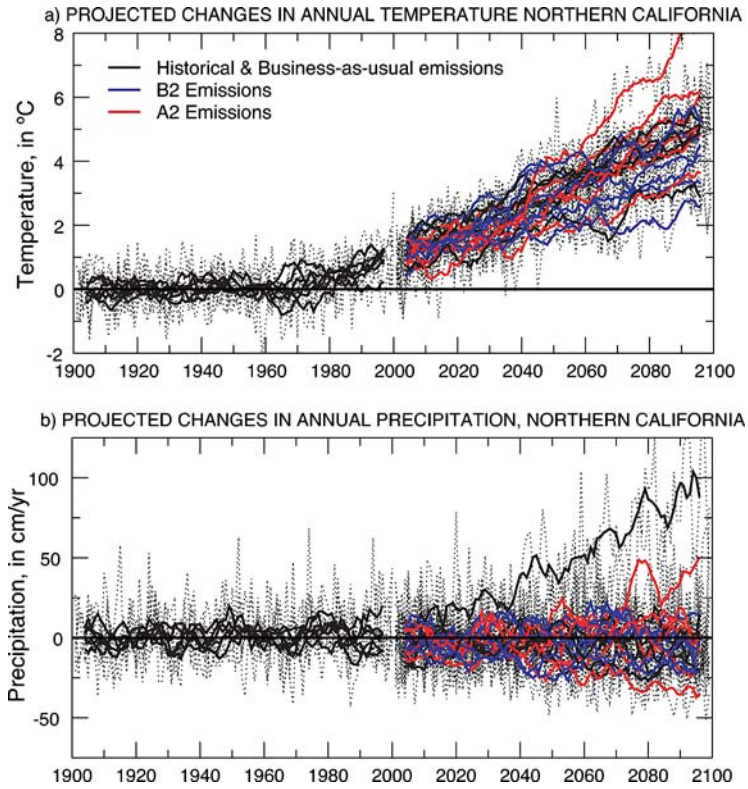

Figure 4. Ensembles of historical and future temperature and precipitation projections from 6 coupled ocean-atmosphere general-circulation models, each forced by historical and then – in the 21st Century – the A2, B2, and IS92a SRES emissions scenarios (Houghton et al., 2001); background of dashed curves shows annual deviations from the 1951–1980 simulated means whereas heavy curves show 7-yr moving averages. Projections are for a single model grid cell from each model centered over northern California.

this even balancing of models and scenarios, no model or scenario is emphasized over the others. This even-handed treatment of the models and scenarios was valued here sufficiently so that other models that did not have all the scenarios available, and scenarios that have not been run in all the models, were excluded from the present analysis. Ideally, in the absence of known deficiencies in one or another of the ensemble members, the pdfs to be interpreted should reflect, in an even handed way, the combination of uncertainties associated with models and the uncertainties associated with future emissions. However, not all models (or scenarios) are equally skillful at reproducing or projecting climate variations. A simple extension of the resampling procedure to allow an uneven treatment of the models, to weight the most skillful models most and least skillful models least, is outlined at the end of this section.

The 18 99-yr long projections of Northern Californian climate changes compiled here all share rapid warming tendencies after about 1970 and, by about 2020, temperatures have all warmed beyond most of the background of historical temperature variability. The general spread of temperature changes by 2100 is from $+2.5\,^{\circ}$C to $+9\,^{\circ}$C. Notably, the scatter among scenarios is not substantially larger or different than the scatter among models considered here; however, emission scenarios that diverge even more than the scenarios analyzed here (e.g., the A1 and B1 scenarios of Houghton et al., 2001) might be different enough to spread the projections considerably more (Hayhoe et al, 2004). Projections of precipitation in the 21st Century are less unanimous, with some projections becoming much wetter (the wettest projections are both from the Canadian model) and some drier. Plotted in this way, the eye naturally focuses on the outliers in the ensemble and many studies have been constructed to address the bounds of such projection ensembles, rather than exploring the more common results.

To improve visualization, interpretation, and – for some applications – the usefulness of this ensemble, the 18 projections of temperature and precipitation (Figure 4) were resampled according to the component-resampling procedure. In this application of the procedure, mixing of the ensemble loading patterns was restricted to only allow projections by a single model to be intermixed. This restriction avoids the possibly inappropriate mixing of incompatible components from the projections by very different climate models. The restriction is easily accomplished by beginning each resampling cycle with the choice of one of the 6 models at random, followed by random sampling among only the 3 amplitude series for that model, to obtain the new realization. With this restriction, even if only 10% of the components (about 20) contribute significantly to any one realization, the number of independent combinations is about 6×3^{20} or about 10^{10}, certainly a sufficient number for applications.

Results of a 20,000 member resampling of the 18-member 99-year climate-change projection ensemble – not the result of the 11 separate sub-applications described in the next paragraph–are shown in Figure 5. The PCA applied in step 4 of the procedure was extended (in the manner of Weare and Nasstrom, 1982) so that

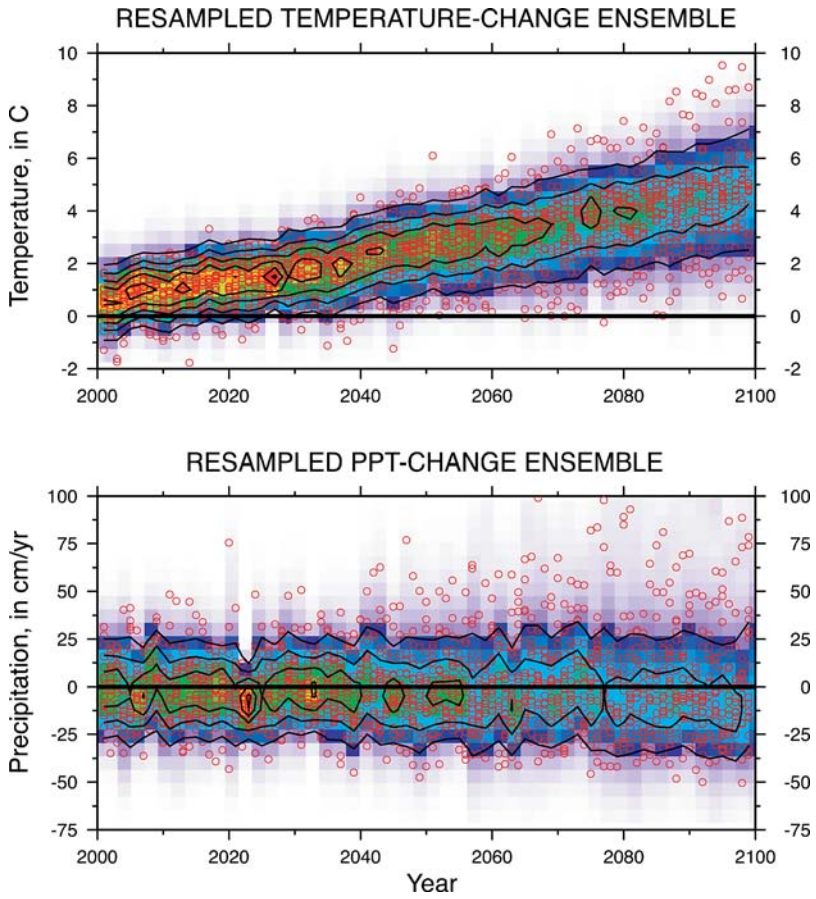

Figure 5. Distributions of original and component-resampled projections of annual 21st Century surface-air temperatures and precipitation changes for a grid cell over Northern California (40°N 120°W), from the ensemble of projections shown in Fig. 4. Red circles show the raw ensemble projections; contours and shading show resampled joint temperature-precipitation probabilities, with contour interval of 0.025.

temperature and precipitation changes were analyzed as a single projection matrix composed of the 99-year time series projections of temperature change appended to bottom of the 99-year projections of precipitation change, to form a 198 × 18 projection-ensemble matrix. This extended projection matrix is decomposed by PCA and resampled, as described in Section 2. The pdfs shown are thus joint pdfs of temperature and precipitation. Consequently, for example, if a particular model has a tendency for excursions of temperature and precipitation to occur simultaneously, the component-resampled realizations emulate those linkages.

PCA of the present 18-member ensemble of the paired 99-year projections of temperature and precipitation change raises the concern that a relatively few examples (ensemble members) are being decomposed into a much larger number of

EOFs for resampling. In a more typical PCA application, where only a few components are retained from further analysis, such a dearth of examples of variability can lead to misinterpretations of the significance or character of the modes isolated. In the present application, however, the problem is much less debilitating because no components are left out of the resampling. That is, all the "parts" are saved and find their ways into the resampled realizations so that the particular form that the PCA decomposition takes is less crucial. Recall that the primary role of PCA in the present analysis is to provide a ready supply of uncorrelated components for resampling with components that together still comprise all, and only, the variance in the original ensemble. To illustrate the strength of the simple form of component resampling outlined previously, the climate-change ensemble was resampled by two different methods: first, precisely as outlined in Section 2 (with results illustrated in Figure 5), and then a second time, in 11 separate applications of the method to subsets of the ensemble containing the projections of precipitation and temperature change for 9 years each. The 9 years in these subsets were selected to span the 99 years of the original projections. The 9 years in each subset could be staggered randomly through the projection period (results shown here) or spaced uniformly through it. The pdfs from the 11 applications were then averaged by decade and compared (in later figures) to the results from the former, simpler application of the method. Considered at this decadal scale, the pdfs obtained by the two methods will be seen to be all but distinguishable.

Early in the 21st Century, the projections are closely clustered, somewhat warmer and somewhat drier on average than the 1951–1980 climatology (because, even by 2000, greenhouse forcings were larger than during that climatology period; Dai et al., 2001). The ensembles spread over the course of the 21st Century, until by the final decade of the projections, temperature-change projections range (mostly) from about $+2$ to $+7\,^{\circ}$C, and precipitation-change projections range from about -30 to $+25$ cm/yr with two outlying exceptions. The smoothing that is provided by the component-resampling procedure is illustrated in Figure 6 by the averages of 11-yr time slices through the time varying projection pdf in Figure 5. Heavy curves correspond to the pdfs from the direct application of the method as described in Section 2, ignoring the mismatch between number of ensemble members (18) and number of time steps and thus number of components resampled (twice 99 years). Light curves correspond to the 11-subset revision of the method, outlined earlier in this section. No substantial differences between the results were found, and even the year-by-year pdfs in Figure 5 are in excellent agreement with both sets of decadal pdfs in Figure 6.

The resampled temperature-change pdfs spread and trend toward warmer conditions as the 21st Century climate evolves. The pdfs measure the combination of uncertainties from simulated natural variability in the projections, from the differences between models, and from the differences between emissions scenarios. The gradual widening of the pdfs is mostly a result of divergence between the models and divergence of the emissions scenarios, with relatively little contribution by

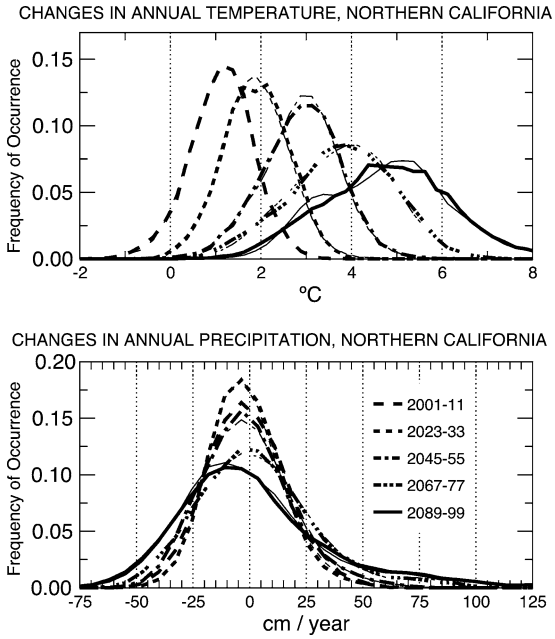

Figure 6. Time slices of the distributions of resampled ensemble realizations from Figure 5; heavy curves are eleven-year averages of distributions directly from Figure 5, thin lines are results of distributions estimated from separately resampling the ensemble in 11 9-yr subsets of the original 99-yr ensemble (described in text).

increasing interannual variability within any given model's projections (a conclusion obtained by resampling subsets of the ensemble containing only one model, one emissions scenario, and so forth). Notice that, by the 11-year period centered on 2028, realizations that are cooler than the 1951–1980 "normal" are already exceedingly rare.

The precipitation-change pdfs translate and spread much less than do the temperature pdfs (Figure 6). Overall, the component-resampled realizations (as in the raw projections) most commonly exhibit only modest 21st Century precipitation changes over California. The modes of the smoothed (resampled) pdfs in Figure 6 trend toward drier conditions, which is much more difficult to perceive in the scattered red dots of Figure 5 or in Figure 4. Thus, although no new information is introduced by the component-resampling procedure, the smoothing it allows can be very informative nonetheless. The general rate of expansion of the ensemble spread around this mean precipitation-change behavior is small, except for a distinct heavy tail spread towards substantially wetter conditions. That heavy tail spread reflects the contributions to the ensemble from the Canadian model's projections, the two outlying much wetter projections in the original 18-member ensemble. That model, under each of the emissions scenarios, evolves towards a much wetter California.

The component-resampled realizations of the projections provide a ready supply of examples of coordinated temperature and precipitation changes for use in evaluating the risk of various climate-change impacts. As a simple example, the 20,000 temperature-and-precipitation-change realizations generated for Figures 5 and 6 were introduced to the streamflow amount and timing response surfaces mapped by Jeton et al. (1996) for the North Fork American River in the central Sierra Nevada. Those response surfaces (Figures 16b and 17c in Jeton et al., 1996) describe the mean simulated changes in annual streamflow amounts and in the median-flow dates (days of year by which half the year's flow is past), in response to 100-year long synthetic climate series with arbitrarily specified mean-climate changes ranging from cooler to warmer, and from drier to wetter. The mean streamflow changes, as mapped by Jeton et al. (1996), that would be expected from the temperature and precipitation changes in each of the 20,000 resampled ensemble realizations (from each of the time slices in Figure 6) were accumulated, and the resulting pdfs of changes in streamflow amount and timing are shown in Figure 7. The realizations from both resampling approaches used in Figure 6 (light and heavy curves) were used to estimate streamflow amount and timing pdfs with no substantial difference in the results.

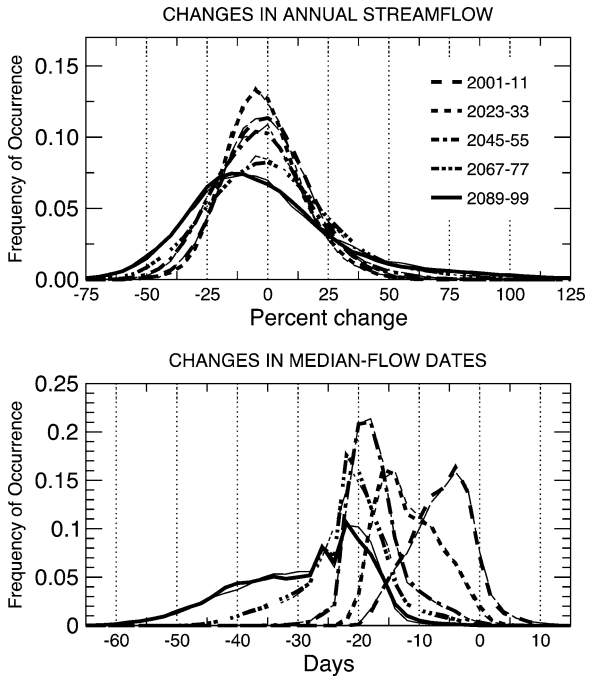

Figure 7. Distributions of annual streamflow amounts and median-flow dates (date by which half of a year's flow is past) in response to 20,000 resampled climate-change projections (illustrated in Fig. 6), streamflow responses were estimated from response surfaces mapped in Jeton et al. (1996); curves weighted as in Figure 6.

The pdfs of annual streamflow changes in Figure 7 are similar to the pdfs of precipitation change in Figure 3, reflecting the strong control that precipitation change exerts on total streamflow amount, as well as the nearly complete buffering of streamflow amounts against responses to temperature changes, discussed at length by Jeton et al. (1996). By the end of the 21st Century, streamflow amounts are significantly biased towards a drier mean and mode, although the much wetter Canadian climate models ensures a heavy tail of significantly wetter streamflow-amount realizations. Streamflow timing mostly reflects the warmer temperatures projected by all the models, although concurrent precipitation changes in the realizations couple nonlinearly with the temperature effects in the Jeton et al. (1996) response surfaces to yield much broader and more multimodal timing distributions. By the 11-year period centered on 2028, years with earlier than normal (1951–1980) median-flow dates are all but eliminated among the resampling-driven realizations. By the end of the 21st Century, the most common median-flow date projections are nearly a month earlier than the 1951–1980 norms; see Stewart et al. (2004) for a more comprehensive and geographically far-reaching discussion of this phenomenon.

Now, consider the differences between the messages and information contents of Figure 4 and Figure 6 (or 7). From the tangle of projections in Figure 4, we conclude mostly that projected temperature changes are more unanimous than are the projections of precipitation change, and that very wet futures are a significant threat (or opportunity). The pdfs, in contrast, suggest that the envelope of (most likely) temperature projections spreads more through time than does the envelope of precipitation changes. The less-than-obvious tendency for the mode of precipitation changes to drift towards drier conditions is also much clearer in the pdfs. In fact, no new information has been added to the ensemble by the component-resampling procedure, but our understanding of the potentialities that the ensemble represents is arguably much clearer. In addition, the users of such an ensemble has much more freedom to select their own levels of risk aversion when ensemble results are quantified by pdfs rather than by a tangle of intertwining projections. The pdfs amount to (crude) estimates of likelihood of various future climates. Risk is essentially a product of likelihood and cost of a given outcome, so that each application needs the opportunity to identify its own levels and climates. A more pdf-centric approach to projection ensembles makes this opportunity available.

Although the resampling procedure as applied in this section added no real information to the ensembles, the procedure can be used to add information in simple and helpful ways. For example, the resampling procedure shown above treated each model's projections as equally likely and each emissions scenario as equally likely. However, the procedure can be modified to reflect any assumed weighting of the various models and scenarios. For example, if the accuracies of each model could be quantitatively indexed by a measure of the likelihood that its projections were most accurate (among all the models considered), then those indices could be used to weight the number of times that its components would contribute to the resampling procedure. This would mean that the most accurate models

would contribute most to the resampled distributions and the least accurate models would contribute least. Similarly, if the likelihood of emissions scenarios could likewise be ranked quantitatively, the resampling probabilities could be adjusted to reflect those outcomes as well. Thus measures of the "skill" of various models and scenarios can be objectively and simply included in the estimation of risks and probabilities.

5. Summary

In many meteorological and climatological modeling applications, the availability of ensembles of predictions containing very large numbers of members would substantially ease statistical analyses and validations, and could be used to improve prediction skill. This study describes and demonstrates an objective approach for generating large ensembles of "additional" ensemble members from smaller ensembles, where the additional ensemble members shared important first- and second-order statistical characteristics and some dynamic relations within the original ensemble. By decomposing the original ensemble members into assuredly independent time-series components (using PCA) that can then be resampled randomly and recombined, the component-resampling procedure generates additional time series that follow the large and small scale structures in the original ensemble members, without requiring any tuning by the user. The method is demonstrated by applications to operational medium-range weather forecast ensembles from a single NCEP weather model and to a multi-model, multi-emission-scenario ensemble of 21st Century climate-change projections. Although little or no new information is introduced by the component-resampling procedure, it provides an objective method for developing reproducible estimates of detailed probability distributions from small ensembles. The procedure also offers a natural way to incorporate information about model or ensemble-member skills, if that additional information is available. Thus the procedure helps to bridge the gap between subjectively interpreted "spaghetti" plots of small ensembles and the kinds of visualizations and calculations that could be accomplished with much larger ensembles.

Acknowledgments

The idea for this procedure was spawned by an offhand remark from Cecile Penland, Climate Diagnostics Center, over lunch several years ago. This study was funded by the California Energy Commission's California Climate Change Center at Scripps Institution of Oceanography, and by the Geological Survey's Hydroclimatology Program.

References

Allen, M. R. and Ingram, W. J.: 2002, 'Constraints on future changes in climate and the hydrologic cycle', *Nature* **419**, 224–232.

Blyth, T. S. and Robertson, E. F.: 2002, *Basic Linear Algebra*, Springer, New York, 232 p.

Dai, A., Wigley, T. M. L., Boville, B. A., Kiehl, J. T., and Buja, L. E.: 2001, 'Climates of the twentieth and twenty-first centuries simulated by the ncar climate system model', *J. Clim.* **14**, 485–519.

Dettinger, M. D., Battisti, D. S., Garreaud, R. D., McCabe, G. J., and Bitz, C. M.: 2001, 'Interhemispheric Effects of Interannual and Decadal ENSO-like Climate Variations on the Americas', in V. Markgraf (ed.), *Interhemispheric Climate Linkages*, Academic Press, San Diego, pp. 1–16.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: 2002, 'Advanced spectral methods for climatic time series', *Reviews of Geophysics* **40**, doi:10.1029/2000RG000092, pp. 1–41.

Ghil, M. and Childress, S.: 1987, *Topics In Geophysical Fluid Dynamics – Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics*, Springer-Verlag, New York, 485 p.

Ghil, M. and Vautard, R.: 1991, 'Interdecadal oscillations and the warming trend in global temperature time series', *Nature* **350**, 324–327.

Hayhoe, K., Cayan, D., Field, C., Frumhoff, P., Maurer, E., Miller, N., Moser, S., Schneider, S., Cahill, K., Cleland, E., Dale, L., Drapek, R., Hanneman, R. M., Kalkstein, L., Lenihan, L., Lunch, C., Neilson, R., Sheridan, S., and Verville, J.: 2004, 'Emissions pathways, climate change, and impacts on California', *Proc., National Academy of Sciences* **101**, 12422–12427.

Houghton and others (eds.): 2001, *Climate Change 2001 – The Scientific Basis*, Cambridge University Press, 525–582.

Jeton, A. E., Dettinger, M. D., and Smith, J. L.: 1996, *Potential Effects of Climate Change on Streamflow, Eastern and Western Slopes of the Sierra Nevada, California and Nevada*, U.S. Geological Survey Water Resources Investigations Report 95 – 4260, 44 p.

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T., Bachiochi, D., Zhang, Z., Williford, C. E., Gadgil, S., and Surendran, S.: 2000, 'Multimodel superensemble forecasts for weather and seasonal climate', *J. Clim.* **13**, 4196–4216.

Richardson, D. S.: 2001, 'Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size', *Q. J. Royal Met. Soc*. **127**, 2473–2489.

Stewart, I., Cayan, D. R., and Dettinger, M. D.: 2004, 'Changes in snowmelt runoff timing in western north america under a 'Business As Usual' climate change scenario', *Climatic Change*, 15 p.

Toth, Z., and Kalnay, E.: 1993, 'Ensemble forecasting at NMC: The generation of perturbations', *Bull. Amer. Meteor. Soc.* **74**, 2317–2330.

Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: 2003, 'Probability and ensemble forecasts', in I. T. Jolliffe and D. B. Stephenson (eds.), *Forecast Verification – A Practitioner's Guide in Atmospheric Science* John Wiley & Sons, San Francisco, pp. 137–164.

Traction, M. S. and Kalnay, E.: 1993, 'Ensemble forecasting at NMC: Operation implementation', *Weather and Forecasting* **8**, 379–398.

Von Storch, H. and Zwiers, F. W.: 2002, *Statistical Analysis in Climate Research*, Cambridge University Press, New York, 494 p.

Weare, B. C., and Nasstrom, J. S.: 1982, 'Examples of extended empirical orthogonal function analysis', *Mon. Wea. Rev.* **110**, 481–485.

Zhu, Y., Toth, Z., Wobus, R., Richardson, D., and Mylne, K.: 2002, 'The economic value of ensemble-based weather forecasts', *Bull. Amer. Met. Soc*. **83**, 73–83.

(Received 2 February 2004; in revised form 18 July 2005)