CHAPTER 3

Understanding and Predicting Subseasonal Variations

Given that the details of daily weather are unpredictable beyond about a week, the questions of what aspects of the circulation remain predictable and what useful information can be extracted from predicting them present interesting challenges. The forecast problem is particularly difficult for Week 2, because boundary conditions have begun to become important but initial conditions have not yet completely lost their influence; at the same time, the chaos from unpredictable nonlinear interactions has nearly saturated. This is mainly why prediction efforts have traditionally focused on shorter (synoptic) and longer (seasonal to interannual) time scales. And yet there is much to said for shifting some of the focus to the subseasonal scale, if only because variability on this scale accounts for a large fraction of the total atmospheric variability from synoptic to decadal scales. Also, episodes of springtime floods, summertime droughts, and prolonged wet or dry spells are phenomena with obvious societal consequences.

CDC scientists are addressing these issues by focusing on the variability and predictability of weekly averages, through both modeling and diagnosis of the observed statistics, and through detailed investigations of NCEP's operational forecast ensembles for Week 2. A significant recent accomplishment was the construction of a low-dimensional 37-component linear empirical-dynamical model that not only successfully represents the statistics of weekly anomalies but also has comparable forecast skill in Week 2 to that of NCEP's operational ensemble. There is evidence that much of this model's skill arises from processes not well represented in the NCEP or other numerical models, such as subseasonal variations of tropical convection. On the other hand, much of the skill of the numerical models is likely due to processes not well represented in the empirical model, such as nonlinear baroclinic cyclogenesis or blocking development in Week 1. It is therefore possible that an intelligent combination of the empirical and numerical model forecasts will yield a Week 2 forecast that is superior to either in isolation. Constructing such a combination is now one of our primary efforts. This effort will benefit from and build on our recent success in improving both statistical and numerical forecast products for this time scale.

3.1 Modeling and understanding the statistics of weekly averages

The principal mechanism of tropical-extratropical interaction is through diabatically forced Rossby waves. On seasonal and longer scales, tropical diabatic heating is strongly linked to tropical SST; hence one speaks of an "SST-forced" global response as in Chapter 2. On the subseasonal scales of interest here, the SST variability is relatively weak, and its coupling to the heating variability is much less rigid. The heating variability itself is considerable, however, and has a significant extratropical impact. Some of this variability (especially that associated with the MJO) is predictable, and raises the hope that at least some aspects of subseasonal extratropical variability may therefore also be predictable. Unfortunately, for various reasons the simulation and predictability of subseasonal tropical heating variations has thus far proved difficult in general circulation models. This has been a major stumbling block in capitalizing on this source of subseasonal extratropical predictability.

Inspired by the success in **Figs 2.1–2.4** of simple empirical predictions of seasonal tropical SST variations and their global impact, we have recently constructed a linear inverse model (LIM) suitable for studies of atmospheric variability and predictability on weekly time scales using global observations of the past 30 years. Notably, it includes tropical diabatic heating as an evolving model variable rather than as an externally specified forcing. It also includes, in effect, the feedback of the extratropical weather systems on the more slowly varying circulation. We have found both of these features to be important contributors to the model's realism.

The model is concerned with the behavior of 7-day running mean anomalies of extratropical streamfunction and column-averaged tropical diabatic heating. It assumes that atmospheric states separated by time lags τ are related as $\mathbf{x}(t+\tau) = \mathbf{G}(\tau)\,\mathbf{x}(t) + \varepsilon$, where $\mathbf{G}$ is a linear operator and ε is noise. This implies that the zero-lag and time-lag-covariance matrices of $\mathbf{x}$ are related as $\mathbf{C}(\tau) = \mathbf{G}(\tau)\mathbf{C}(0)$. We use this relationship at a particular lag, say $\tau = 5$ days, to obtain $\mathbf{G}(5)$ from observational estimates of $\mathbf{C}(5)$ and $\mathbf{C}(0)$. We then make another assumption that is at the heart of the LIM formalism, that distinguishes it from other empirical models, and that enables one to make dynamically meaningful diagnoses of direct relevance to modelers. This is that $\mathbf{G}(\tau)$ satisfies the relation $\mathbf{G}(\tau) = \exp(\mathbf{L}\tau)$, where $\mathbf{L}$ is a *constant* linear operator. We use this to obtain $\mathbf{L}$ from $\mathbf{G}(5)$, and having done so, use it again to obtain $\mathbf{G}$ for all other lags. We are finally in a position to make forecasts for all lags as $\mathbf{x}(t+\tau) = \mathbf{G}(\tau)\,\mathbf{x}(t)$. Crucially, having obtained $\mathbf{L}$, we can also diagnose the relative importance of its elements associated with tropical-extratropical and internal extratropical interactions. For example, we can use $\mathbf{L}$ to estimate what the statistics of extratropical variability would be without diabatic forcing from the tropics.

Figure 3.1 demonstrates the success of this model in reproducing the observed variance and 21-day lag covariance of 7-

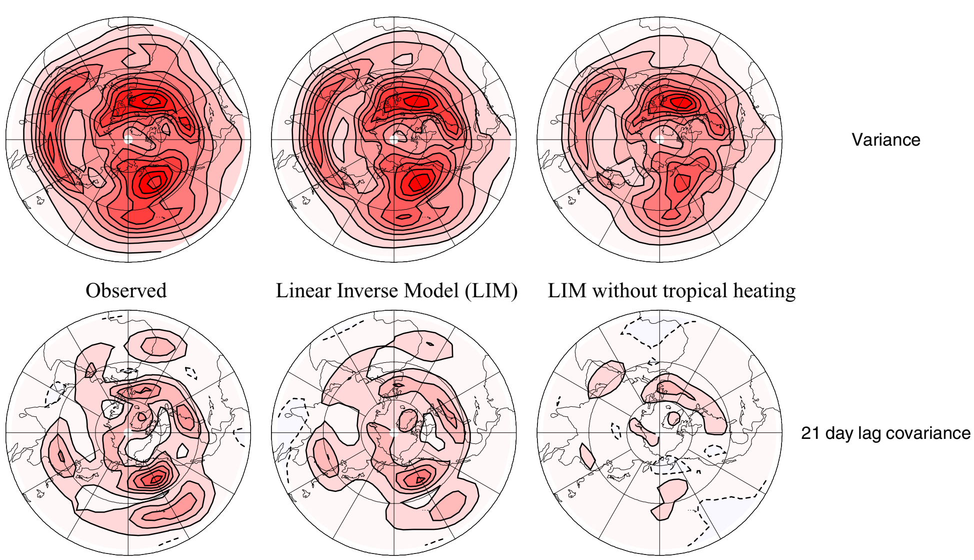

Fig. 3.1: Observed and modeled (using the full LIM and a version of the LIM in which the effects of tropical heating are removed) statistics of weekly 250 hPa streamfunction anomalies.

day running-mean anomalies of 250 mb streamfunction during northern winter. Note again that we are effectively using the observed 5-day lag covariances to predict the 21-day lag covariances here. The comparison of the observed and predicted covariances is clearly encouraging. The right column shows that the effect of tropical heating is relatively small on the variance but relatively large on the 21-day lag covariance. This is consistent with our finding that although tropical heating contributes a relatively small portion of the extratropical variability, it contributes a large portion of the *predictable* variability.

Forecast skill is an important test of any model. The LIM is better at forecasting Week 2 anomalies than a dynamical model based on the linearized baroclinic equations of motion (with many more than the LIM's 37 degrees of freedom) that is forced with *observed* tropical heating throughout the forecast. Indeed at Week 2 the LIM's skill is competitive with NCEP's MRF model with nominally O(10^6) degrees of freedom. The upper panel of **Fig. 3.2** shows such a comparison of Week 3 forecast skill during the winters of 1985/86–1988/89. Other experiments show that this encouraging forecast performance is not limited to years of El Niño or La Niña episodes.

The LIM assumes that the dynamics of extratropical low-frequency variability are linear, stable, and stochastically forced. The approximate validity of these assumptions has been demonstrated through several tests. A potentially limiting aspect of such a stable linear model with decaying eigenmodes concerns its ability to predict anomaly growth. We have nevertheless found, through a singular vector analysis of the model's propagator **G**, that predictable anomaly growth can and does occur in this dynamical system through constructive modal interference. Examination of the initial structures associated with optimal anomaly growth further confirms the importance of tropical heating anomalies associated with El Niño and La Niña as well as Madden-Julian oscillation episodes in the predictable dynamics of the extratropical circulation.

The LIM formalism also allows one to estimate predictability limits in a straightforward manner. Indeed it allows one to estimate the expected skill of any *individual* forecast from the strength of its predicted signal. Given that in many cases the predictable signal is associated with tropical forcing, one can quantify the effect of that forcing on extratropical predictability. Our general conclusion is that without tropical forcing, extratropical weekly averages may be predictable only about two weeks ahead, but with tropical forcing, they may be predictable as far as seven weeks ahead. This difference is highlighted in the lower panel of **Fig. 3.2**. This suggests that accurate prediction of tropical diabatic heating, rather than of tropical sea surface temperatures *per se*, is key to enhancing extratropical predictability on these time scales.

As mentioned earlier, most current GCMs have difficulty in representing and predicting heating variations on these scales. This is especially true of the

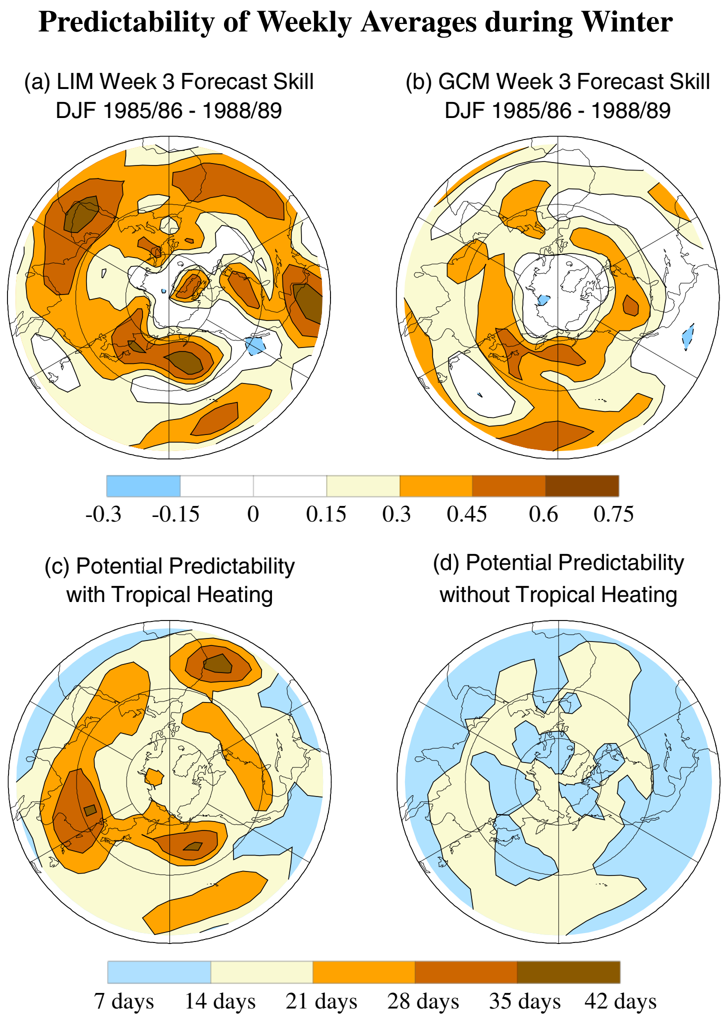

Fig. 3.2: Forecast skill and predictability of weekly averages during winter. Top: Correlation of observed and Week 3 forecasts of upper tropospheric streamfunction anomalies averaged over 52 forecast cases in the winters of 1985/86–1988/89 for (a) LIM and (b) the NCEP MRF. Bottom: Potential predictability limit: forecast lead at which skill (i.e., the correlation of observed and predicted anomalies) drops below 0.5. (c) Determined from the full LIM. (d) Determined from a version of the LIM in which the effects of tropical forcing are removed.

NCEP MRF model. We have documented significant deficiencies in the "reanalysis version" of that model in maintaining and propagating MJO-related heating and circulation anomalies. **Figure 3.3** shows that forecasts initialized when the MJO is active over the Indian ocean are unable to represent the subsequent eastward propagation of 850-mb zonal wind anomalies; indeed they do not predict propagation at all but a rapid decay. This has been demonstrated to have a negative impact on extratropical forecasts.

Figure 3.4 shows that the LIM's forecast skill over the PNA region is comparable to that of the operational MRF ensemble mean, especially in summer. The MRF can represent some phenomena that the LIM cannot, such as nonlinear baroclinic cyclogenesis and blocking. To the extent

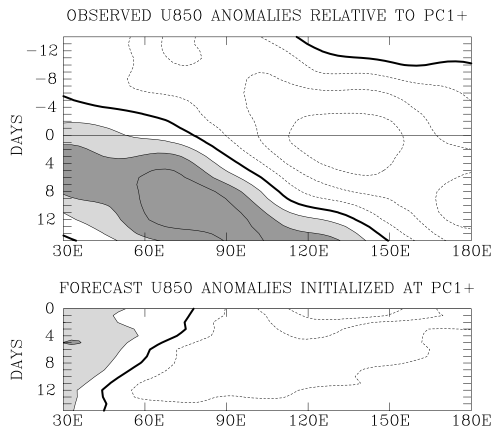

Fig. 3.3 Composite anomalies of 850 mb zonal wind averaged between 5°N–15°S relative to the maximum of the first EOF of subseasonal tropical OLR anomalies, when MJO activity is maximum over the east Indian ocean. The upper panel is for observed anomalies from Days –14 to +14, where Day 0 refers to the time of maximum EOF coefficient. The lower panel is for the anomalies predicted by the NCEP MRF model, with the mean model error removed.

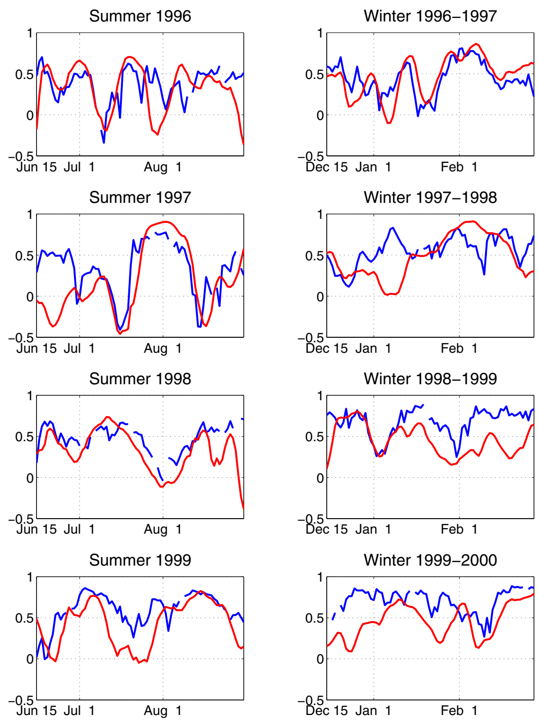

Fig. 3.4: Week two forecast skill (as measured by pattern anomaly correlation over the PNA regions) for the operational NCEP MRF ensemble mean (blue curve) and the LIM (red curve) for four winter and summer seasons.

that these phenomena are predictable, the MRF should have an advantage. This is indeed the case in Week 1. By Week 2, these phenomena become unpredictable; even so, their role in exciting larger scale, slowly evolving structures such as the PNA pattern in Week 1 can contribute to maintaining forecast skill in Week 2. On the other hand, the LIM is much better at predicting subseasonal variations of tropical convection than the MRF, and being an anomaly model, also does not suffer from climate drift by construction. Therefore, it seems likely that the comparable skill of the LIM and the operational MRF models is not arising entirely from the same sources. This is in contrast to the seasonal prediction problem discussed in Chapter 2, in which the comparable skill of GCMs and simple statistical models arises from essentially the same source. To the extent that the sources of Week 2 forecast skill in the statistical and dynamical models are distinct, combining the two forecasts should, in principle, yield forecasts that are superior to either in isolation. Constructing such a combination is currently one of our main priorities.

3.2 Subseasonal variations in tropical convection and predictability of California rainfall.

Figure 3.2 demonstrates that much of the LIM skill in the extratropics arises from its ability to predict tropical heating variations. Operational models are notoriously poor at this. This has implications for predictions of extratropical rainfall:

for example, rainfall along the west coast of North America is known to be influenced by subseasonal tropical heating. This suggests that operational precipitation forecasts over North America could be improved, particularly in Week 2–3 range, by using statistical methods to augment the numerical product. CDC scientists have obtained a conservative lower bound on the potential improvement through a statistical prediction model of weekly precipitation over western North America in winter.

The model is based on Canonical Correlation Analysis (CCA), with tropical Outgoing Longwave Radiation (OLR) anomalies as the predictor and NCEP Reanalysis precipitation over the eastern Pacific and western North America as the predictand. A single CCA mode accounts for most of the predictable signal. The rank correlation of this mode and observed rainfall anomalies over Southern California over a 25-winter period is 0.2 for a two-week lag, which is comparable to the correlation between a weekly ENSO index and weekly rainfall in this region. **Figure 3.5** shows that this corresponds to a 50% increase above the climatological risk (33%) of above-normal rainfall in California when the projection of tropical OLR on the leading CCA mode two weeks earlier is high, i.e. in the upper quintile of its distribution.

The leading CCA mode represents suppressed convection over the equatorial Indian Ocean and enhanced convection east of the dateline (**Fig 3.6**). Associated with this canonical tropical OLR anomaly pattern is the development of upper tropospheric westerly wind anomalies near 30°N in the eastern Pacific (not shown). Synoptic-scale weather systems are steered farther east toward California by these enhanced westerlies.

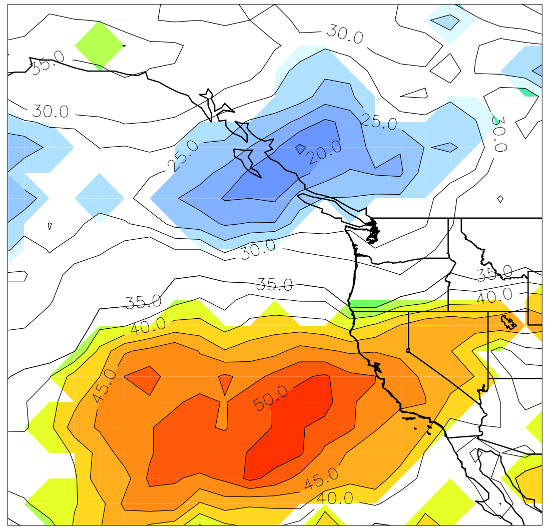

Fig. 3.5: Probability that rainfall will be in the upper tercile of its distribution when the projection of the tropical OLR on the leading CCA mode two weeks prior is in the upper quintile of its distribution.

An analysis of four years of operational Week 2 ensemble forecasts indicates that the skill of this statistical model is comparable to that of the operational ensemble mean, just like the LIM forecasts discussed earlier (see **Fig. 3.4**). Since by Week 2 the operational forecast model has lost its ability to represent subseasonal tropical heating variability, the statistical model provides essentially independent guidance to the forecaster. The fact that the skill of the two models is comparable suggests a significant potential for improvement of the operational Week 2 precipitation forecasts. We are investigating ways of optimally combining the numerical and statistical forecasts. This requires estimating the covariances of the ensemble mean fore-

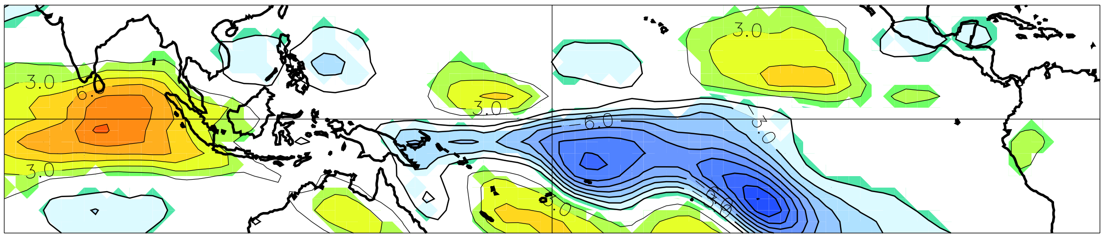

Fig. 3.6: OLR regressed on leading canonical predictor vector, scaled for one standard deviation of the canonical predictor variable. Contour interval is 1.5 W m^{-2}. Positive contours are thicker, the zero line is omitted and shading indicates statistical significance at the 95% level.

cast errors and observed tropical heating variability. Since the tropical convective events associated with predictability on this scale occur only about once or twice a season, a long (20+ year) record of numerical forecasts with a frozen model is needed to estimate the required forecast error statistics. Work is underway at CDC to create such a retrospective forecast database.

3.3 The role of ENSO-related tropical heating on operational weather forecasts

The crucial role of tropical heating in the evolution of at least some extratropical weather events is evident in a study of the effect of the 1997–98 El Niño on the operational 1–14 day ensemble forecasts. This study was motivated by the hypothetical question: What would happen to the medium-range (up to 14-day) forecast if the anomalous tropical SSTs were replaced by climatological values? The difference of such forecasts from those made with the actual SSTs— “the ENSO signal”—could then be used to diagnose the influence of El Niño or La Niña on evolving midlatitude storm systems. Because of the uncertainty inherent in weather forecasting, we used the operational MRF ensemble to look at the average of many forecasts of a given storm. The ensemble also provided us with a rigorous way of assessing the statistical significance of our results. It is perhaps worth mentioning that the ensemble with climatological SSTs was run in-house at CDC in real time. Results from the comparison with NCEP's operational ensemble with actual SSTs were made available to public on the web, also in real time.

This study provided the first demonstration of a direct impact of El Niño SST anomalies on individual extratropical weather systems. Perhaps the most interesting case was the devastating ice storm that hit Canada in early January 1998. **Figure 3.7** shows the predicted 500 mb ensemble-mean height anomaly patterns with and without the El Niño SST forcing, as well as the observed verification. The forecast was a lot closer to the observed with the El Niño forcing included, showing that it played an important role in the evolution of this storm. The area of unusually warm mid-level air associated with the production of freezing rain is indicated by the red arrow. The operational runs (with El Niño SSTs) show a wavetrain aligned

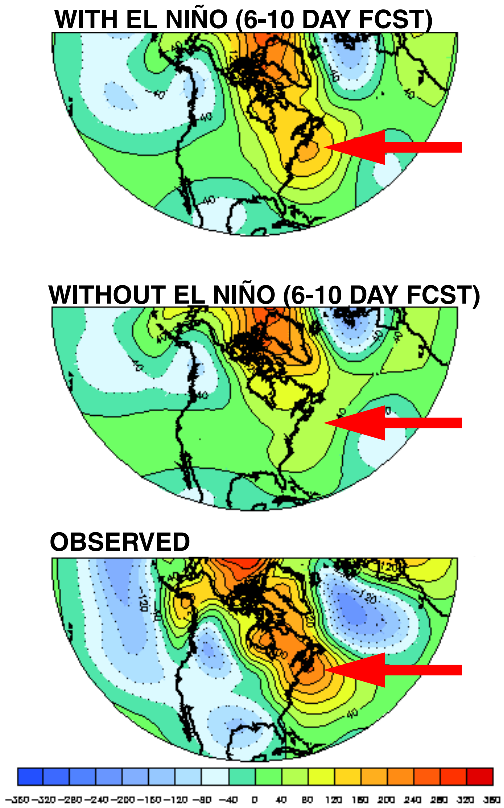

Fig. 3.7: Ensemble mean 6–10d average 500 hPa height anomalies for the ensemble with observed tropical SSTs (upper panel) and the ensemble with climatological tropical SSTs (middle panel) for forecasts verifying the first week of January 1997. The lower panel shows the verifying analysis, and the red arrows indicate where unusually warm air at mid-levels contributed to the development of freezing rain at the surface.

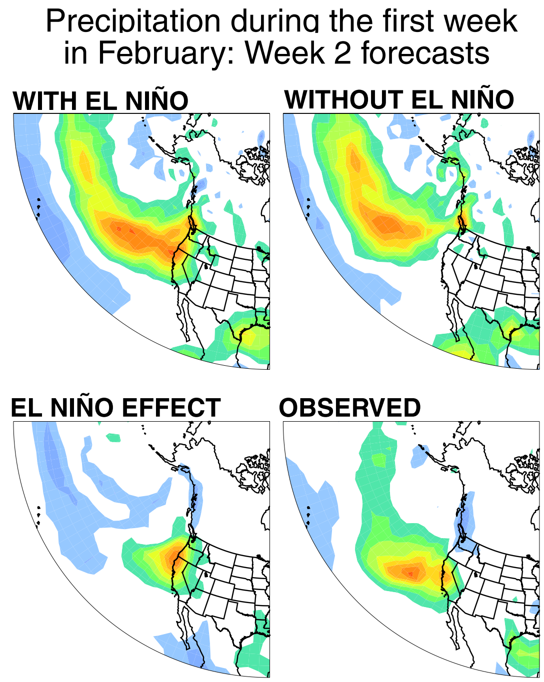

Fig. 3.8: Precipitation accumulated over the second week of the forecast for the ensemble with observed tropical SSTs (upper left), climatological SSTs (upper right), and the difference between the two (lower left) for forecasts verifying the first week of February 1998.

along the Atlantic Coast of the United States that is nearly absent in the runs without the El Niño forcing. This wavetrain appears to have been a product of convective forcing in the eastern tropical Pacific interacting with a deep midlatitude/subtropical trough over Mexico: a previously unnoticed mechanism for El Niño teleconnections.

The much-anticipated "El Niño rains" in California provide another example. California rainfall is episodic, even during El Niño years. Rain also falls during non-El Niño years, making attribution of an individual storm to El Niño nearly impossible using historical data alone. However, our use of a dynamical model allowed us to make the attribution directly. In **Fig. 3.8** the runs with and without the El Niño forcing differ substantially. The "El Niño effect" is a clear eastward extension of the rainfall into California.

We conducted a similar study for the following winter, during which there was a substantial La Niña event. However, the results were less conclusive, partly because of model changes and partly due to the weaker SST forcing. Nonetheless,

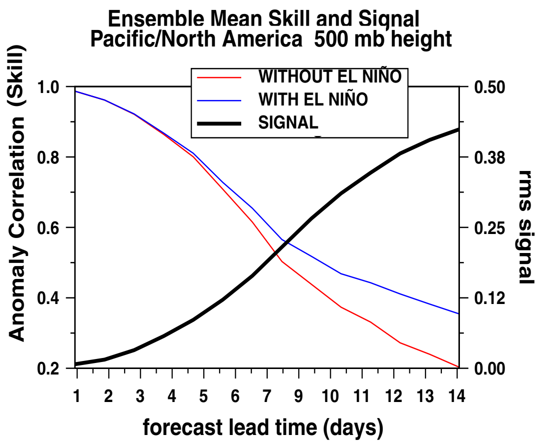

Fig. 3.9: Week 2 500 hPa height anomaly correlation skill over the PNA region for the ensemble mean with (blue) and without (red) El Niño SST anomalies in the tropics. The black curve shows the RMS 500 hPa height El Niño Week 2 forecast "signal", defined as the difference in the ensemble mean forecasts with and without tropical SST anomalies.

over the course of these two winters this experiment provided evidence of a large influence of tropical convection on medium and extended-range forecasts in midlatitudes, especially in Week 2. **Figure 3.9** shows that the tropical influence on forecast skill became significant after Week 1 in this experiment, and was responsible for almost all of the skill by day 14. This study also underscored the great difficulty, but also the great rewards, of using an operational forecast model in research mode. As an added benefit, we were able to assist NCEP during their fire-related computer outage by running the operational ensemble forecasts at CDC in real time.

3.4 The relationship between spread and skill in the operational NCEP ensemble forecasts

The studies discussed in sections 3.1 and 3.2 show that statistical models can have skill comparable to NWP models in the Week 2 to Week 3 range because subseasonal variations of tropical convection, which the NWP models do not simulate well, provide significant predictive information. However, the NWP models, unlike the statistical models, can provide information on day-to-day variations of both the signal (the amplitude of the predictable component of the forecast) and noise (the amplitude of the unpredictable component of the forecast). The statistical models assume the noise to be stationary, i.e. to not vary from forecast to forecast. An NWP ensemble can be used to estimate the noise in each forecast case, and hence a case-dependent estimate of the RMS error of the ensemble-mean forecast.

The simplest measure of forecast noise is the width, or spread, of the forecast probability distribution for any quantity of interest. CDC scientists have investigated the relationship between spread and skill in the operational NCEP forecast ensembles using an archive of operational forecasts maintained at CDC since 1995. Simple statistical considerations show that such a measure is most useful when the case-to-case variability of the ensemble spread is large. This was shown to be true in two winters of operational ensemble predictions. However, the short data record precluded a detailed analysis of the dynamical mechanisms of the spread variability.

To get around this limitation, a five-level linear quasi-geostrophic (QG) model, linearized about three-day segments of the observed flow for 21 years, was used to model the spread variability. The fundamental assumption was that day-to-day variations of spread are due primarily to day-to-day variations in the growth rate of small perturbations during the forecast period, and that day-to-day variations in the initial error, i.e. in the spread of the analysis-error distribution, are either unimportant or not well sampled. The five-level model was able to reproduce the main results (not shown) of the shorter 2-winter study mentioned above. When run for 21 years, the QG model showed the largest spread variability of 3-day forecasts over the eastern Pacific and eastern Atlantic oceans, which was associated with modulations of the local jets by the PNA and NAO modes of low-frequency variability (**Fig. 3.10**). To the extent that such modulations are predictable, the results from this study suggest that skill should also be most predicable in these regions.

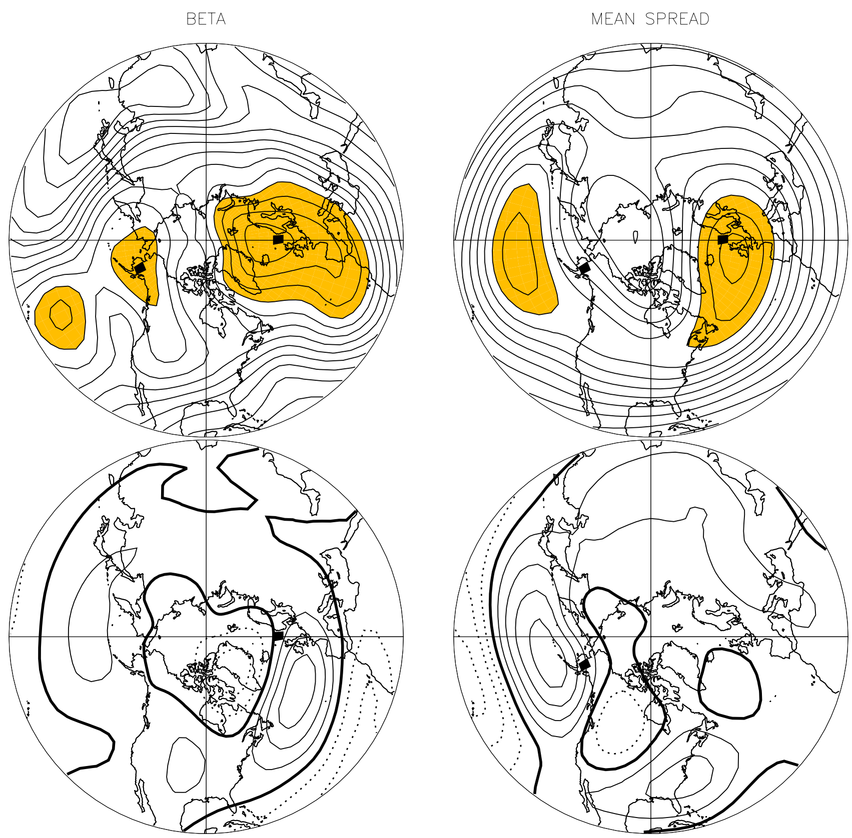

Fig. 3.10: Upper panels: 21 winter mean 300 hPa streamfunction spread (S) and standard deviation of lnS (=β), estimated from 3-day integrations of the five-level linear QG model. S is normalized by the mean amplitude of the initial perturbations used in the ensemble integrations. Contour interval for β is 0.01, with values greater than 0.28 shaded. Contour interval for normalized S is 0.25, with values greater than 4 shaded. Lower Panels: Map of correlations between time series of lnS at points indicated by the black rectangles and three-day averaged 300 hPa streamfunction. Contour interval is 0.1, negative values are dashed, and the zero line is thick solid.

3.5 Experimental week 2 forecasts of extreme events using the operational NCEP ensemble

The existence of a significant spread-skill relationship (at relatively short forecast ranges) means that changes of both the mean and width of the forecast probability distribution from their climatological values can be used to estimate the probability that the verification will lie in the tails of the climatological probability distribution (see **Fig. 2.6**). The statistical models discussed earlier assume that the spread is constant, and that only shifts of the mean are important in altering the probability that the verification will be an "extreme event".

Unfortunately, this advantage of ensemble forecasts, which is modest but significant in Week 1, is lost by the middle of Week 2. The main reason is that by Week 2 the forecast ensemble spread nearly saturates to its climatological mean value, so that there are no significant spread variations from case to case. In other words, most of the predictable variation of forecast skill in Week 2 is associated with predictable variations of the signal, not of noise. For several years CDC has exploited this fact in producing an experimental real-time Week 2 forecast product based on the NCEP ensemble (http://www.cdc.noaa.gov/~jsw/week2/). Tercile probability forecasts of 500 mb height, 850 mb temperature, 250 mb zonal wind, sea-level pressure and precipitation are provided. Only the signal, not noise, is used to construct these probability forecasts. The procedure involves converting maps of the predicted standardized anomalies into maps of extreme quantile (in this case, tercile) probabilities. This calibration is done empirically, using the available historical record of ensemble forecasts and verifying analyses. The procedure is as follows: 1) for a positive standardized forecast anomaly α, all instances in which a forecast *exceeded* this value in the data record are found, and the probability β that the verifying analysis fell in the upper tercile of the climatological distribution is computed, 2) the standardized anomaly contour α is relabeled as a probability of above-normal equal to β. If α is negative, the probability that the verifying analysis fell into the lower tercile is computed, and the contour is relabeled "probability of below-normal". If the model has systematic errors, these probabilities need not be symmetric, i.e. the probability of below-normal for a negative α need not be the same as the probability of above-normal for a positive α. Our calibration thus provides one simple way of accounting for model error in probabilistic predictions.

Figure 3.11 shows an example of such a probability forecast. Note that the interpretation of this map is slightly different from that for a conventional probability forecast. If all the points on the map inside the yellow contour (as opposed to those inside the yellow *band*) are counted over a large sample of forecasts, 50–60% of these points will verify in the upper tercile of the climatological distribution. Similarly, for points falling in the darkest red regions on the map, over 90% will verify in the upper tercile. The conventional interpretation would be that points

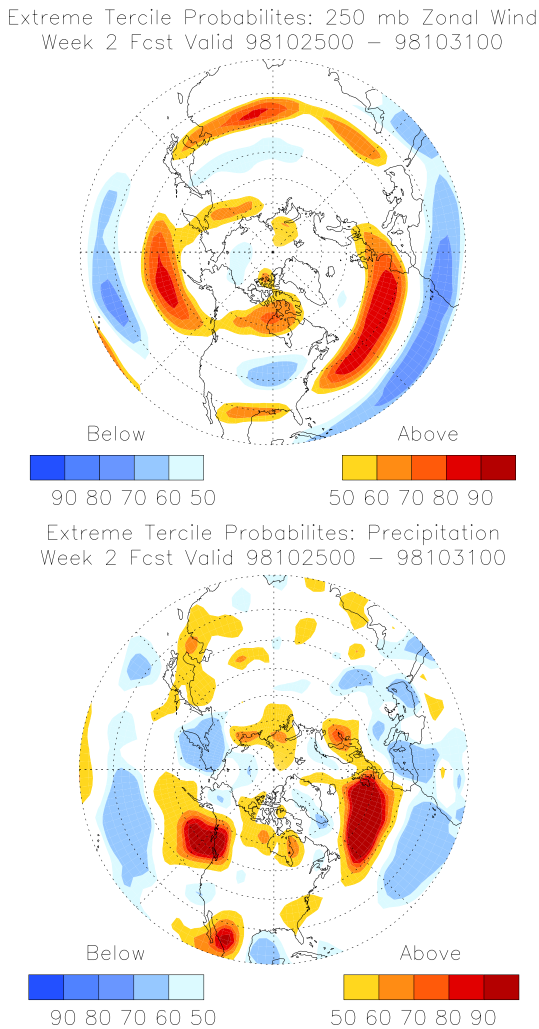

Fig. 3.11: Example of an experimental week 2 forecast verifying the last week of October 1998.

in the yellow *band* would have a 50–60% chance of verifying in the upper tercile. Such a calibration would require a lot more forecasts to compute reliably, since there are far fewer points inside the yellow band than there are inside the yellow contour.

Since we assume that the signal, not the noise, contains all of the useful predictive information, the useful subspace of the ensemble can be isolated through an EOF analysis of the correlation matrix of the ensemble-mean predictions. (The idea here is similar to that in **Fig 2.8**). The right panels of **Fig. 3.12** show the three leading EOFs thus obtained. For comparison, the three leading EOFs of the correlation matrix of observed 7-day averages is also shown, in the left panels. There are two notable aspects to **Fig. 3.12**: 1) the signal and observed EOF patterns are similar, and 2) the three leading EOFs explain considerably more variance of the ensemble-mean forecasts than they do of the observed variability (36% vs. 22%). To understand this better, note that the total forecast covariance can be decomposed into a part due to the predictable signal ($\mathbf{C}_{signal}$) and a part due to unpredictable noise ($\mathbf{C}_{noise}$). If the forecast model is unbiased and the noise is uncorrelated with the signal, the observed variance ($\mathbf{C}_{obs}$) is approximately the sum of the two. This relationship is exact for the LIM discussed in section 3.1. The fact that the signal variation occurs in a lower dimensional subspace than the observational 7-day averages then simply means that the variance contained in the noise is non-trivial. The similarity of the observed and signal EOF patterns has a subtler interpretation: it implies that the noise component of the covariance is nearly white, and that the ensemble-mean does indeed capture most of the extractable signal with coherent spatial structure.

The product shown in **Fig. 3.11** has been quite popular with operational forecasters. A similar method has been adopted in operations by NCEP/CPC. A detailed analysis of the performance of this

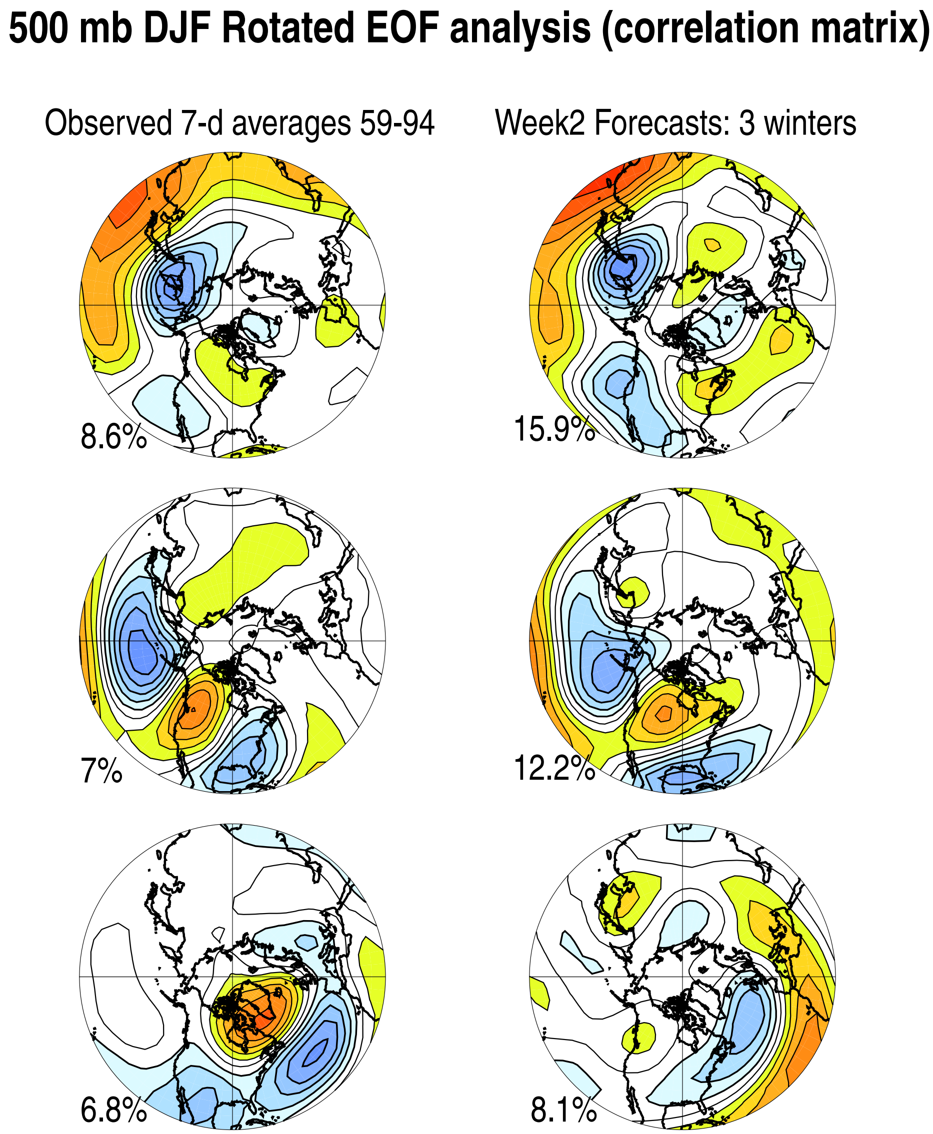

Fig. 3.12: Rotated EOFs of weekly average 500 mb height computed using the correlation matrix for DJF 1958-1994 (left panels) and the correlation matrix of week 2 operational ensemble mean forecasts for DJF 1995/96 to 1997/98.

scheme, and its implications for Week 2 predictability, is underway.

3.6 Unifying ensemble forecasting and data assimilation.

The fundamental goal in subseasonal prediction, just as in seasonal prediction, is to predict the forecast probability distribution function (PDF) accurately. In the previous sections, we have discussed research efforts at CDC toward this goal. It is hoped that statistical methods like the LIM or CCA, when combined with an NWP ensemble, will improve the mean of the forecast PDF. The spread-skill relationship discussed in section 3.4 shows that useful information can be extracted at short forecast ranges from the second moment of the NWP ensemble. One obvious way to improve the accuracy of the forecast PDF is to improve the accuracy of the initial PDF. Currently, all operational centers construct an ensemble of initial conditions by perturbing a single control analysis, obtained from a three-dimensional (as at NCEP) or a simplified four-dimensional (as at ECMWF) data assimilation system. The methods used to generate the perturbations to the control analysis, breeding vectors at NCEP and singular vectors at ECMWF, are fundamentally ad-hoc and not representative of analysis uncertainty. CDC scientists have been investigating new ways of coupling the ensemble forecast and data assimilation steps, in order to improve both the initial and forecast PDFs.

The coupling of ensemble forecasting and data assimilation is natural. The essence of data assimilation is statistical, in that it amounts to blending "first guess" forecasts with new observations using weights determined by their respective error statistics. Carefully constructed forecast ensembles can provide such statistics. Currently, operational methods make rather simplistic assumptions about the error statistics, assuming, for example, that the correlation of forecast errors at two locations depends only on the distance between them and not on the location or whether the atmosphere has recently been quiescent or stormy (**Fig 3.13a**). Results from simple model experiments using sophisticated "Ensemble Kalman Filter" techniques suggest

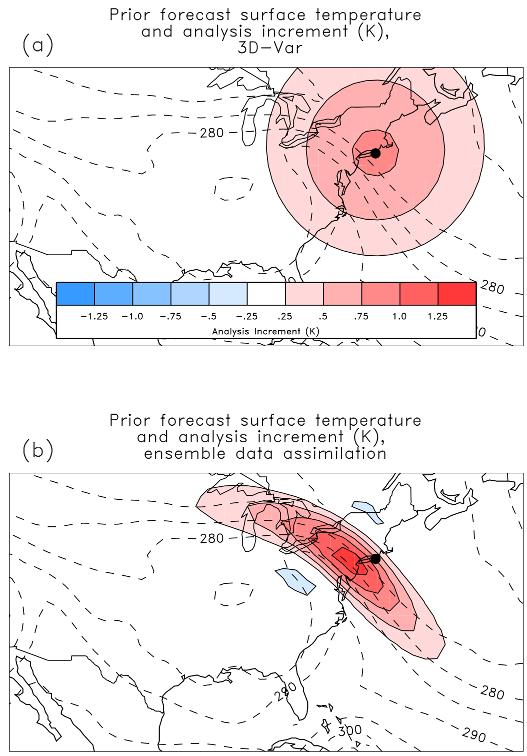

Fig. 3.13: Examination of the structure of the "analysis increment" (the initial condition minus the prior "first guess" forecast) for the traditional method of doing data assimilation, where error statistics do not change from location to location or day-to-day. In this experiment, an observation that is 1 K warmer than the prior forecast is found at the location denoted by the dot. (a) Analysis increments using the "3D-Var" data assimilation methodology. The "one size fits all" increments are a simple decreasing function of increasing distance from the observation location. (b) Analysis increments using the new ensemble data assimilation methodology. Changes to the prior forecast are now stretched out along the frontal zone, so that the entire position of the warm front is changed by the one observation.

that the quality of initial conditions can be dramatically improved by using forecast error statistics estimated from a specially constructed ensemble. For example, error statistics from the ensemble permits a single observation at a fixed location to make very different corrections to the first-guess depending on the flow of the day (**Fig 3.13b**). By estimating the analysis increment to the first guess in this flow-dependent manner, ensembles of initial conditions can be dramatically improved, perhaps even to the point that they are more accurate than analyses based on four-dimensional variational methods.

Recent CDC efforts in this area have focused on algorithmic details of ensemble-based data assimilation experiments. We have sought to understand how the statistics of forecast errors estimated from an ensemble depend on the size of the ensemble, and how one might extract useful information from smaller ensembles—an important issue, since larger ensembles make heavier demands on computational resources. This research has demonstrated that with an accurate specification of forecast error statistics, new problems can be tackled in a theoretically justifiable manner, including problems such as determining where supplementary observations would be most beneficial for reducing analysis or forecast error (the problem of "targeting" observations). In addition, since ensemble-based data assimilation techniques are particularly useful when observations are sparse, CDC scientists are planning to adapt such techniques to extend the NCEP reanalysis back into the pre-radiosonde era (pre-1948).

EPILOGUE

The problem of how to make useful forecasts at lead times between a week and a month is a challenging and often neglected one. Forecast information on these time scales is in great demand from users. This is an area that NOAA has traditionally not focused on in the past. CDC researchers have been addressing the problem on two fronts; 1) by trying to extract the maximum information from ensemble NWP model forecasts, and 2) by investigating statistical forecast methods that complement the NWP ensembles by exploiting predictable signals not well represented in current models. Our research thus far suggests that the NWP and statistical approaches are complementary, and provide information that is independent to some degree. The challenge is to combine the two in an optimal manner, yielding forecasts that are superior to either individually. Due to the low-frequency nature of the phenomena at these forecast ranges, determining the optimal combination would require generating a long (20+ year) dataset of ensemble forecasts with a fixed model to estimate the forecast error statistics with the necessary accuracy. Work is currently underway at CDC to create such a dataset, which will also be useful in several other applications not discussed here.

Contributed by: *J. Barsugli, T. Hamill, H. Hendon, B. Liebmann, M. Newman, P. Sardeshmukh, and J. Whitaker.*