Ensemble Regression: Using Ensemble Model
Output for Atmospheric Dynamics and Prediction

by

Daniel Gombos

B.S. Atmospheric Science, Cornell University, 2004

Submitted to the Department of Earth, Atmospheric, and Planetary Science in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Atmospheric Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2009

© Massachusetts Institute of Technology 2009. All rights reserved

Signature of Author ............................................................... Daniel Gombos

Department of Earth, Atmospheric, and Planetary Sciences

January 30, 2009

Certified by ................................................................. Kerry Emanuel

Kerry Emanuel
Breene M. Kerr Professor of Meteorology
Thesis Supervisor

Accepted by ................................................................. Maria Zuber

Maria Zuber
E.A. Griswold Professor of Geophysics
Head, Department of Earth, Atmospheric, and Planetary Sciences
Abstract

Ensemble regression (ER) is a linear inversion technique that uses ensemble statistics from atmospheric model output to make dynamical inferences and forecasts. ER defines a multivariate regression operator using ensemble forecasts and analyses to determine the most probable predictand perturbation associated with the prescribed predictor perturbation resolved by linear combinations of the predictor ensemble anomalies.

Because it employs flow-dependent ensemble data, as opposed to the stationary time series data typically used to make statistical forecasts, ER is capable of modeling synoptic scale processes with rapidly evolving covariances. This characteristic is applied in several ways. Firstly, it is shown that the classical dynamical piecewise potential vorticity (PV) inversion of the PV perturbation effectively resolved by the ER operator yields nearly identical geopotential heights to those deduced from an ER performed in the subspace of the leading PV singular vectors. Secondly, using the example of the lagged sensitivity of tropical cyclone tracks to preexisting midtropospheric heights, ER is used to infer dynamical relationships from statistical sensitivities, to identify, in real-time, the dynamical processes that are particularly relevant to specific forecast decisions, and to make preemptive forecasts. Thirdly, it is shown that singular vectors deduced from the ER operator approximate those from the analysis error covariance normed tangent linear model operator, suggesting a simple alternative method for computing singular vectors.

Given that ER results are a function of forecast ensemble reliability, theory and applications of a multivariate ensemble reliability verification technique called the minimum spanning tree rank histogram are presented. Experiments using Euclidean, variance, and Mahalanobis norms for defining minimum spanning tree distances imply that, unless the number of ensemble members is less than or equal to the number of dimensions being verified, the Mahalanobis norm transforms a spanning tree into a space where model imperfections are most readily identified.
Acknowledgements

First and foremost, I would like to thank Jim Hansen (Naval Research Laboratory) with the utmost sincerity. Jim welcomed me into his research group during my first year at MIT, upon my realization that my academic interests had gravitated from climate dynamics to ensemble methods. Through his organization of frequent group meetings and social outings, Jim created a friendly environment that helped strengthen relationships between students and faculty with similar academic interest. Most importantly, his unrestricting style of advising facilitated my development as a researcher by giving me substantial freedom to explore ideas that appealed to me, while still providing invaluable suggestions to help me regain focus when necessary. Despite being across the country, Jim continued to act as a mentor after leaving MIT by collaborating with me on papers and enabling two trips to Monterey that helped progress my research. I thank Jim for all of his effort, time, ideas, and support that contributed to this thesis and my development as a scientist.

I would also like to thank several MIT faculty members. Firstly, upon Jim’s departure from MIT, Kerry Emanuel graciously welcomed me into his research group and provided me with helpful guidance and support. In addition to him offering valuable insights for my research, Kerry’s undeniable scientific prowess also provided academic inspiration. Prior to joining Jim’s group, John Marshall welcomed me into research group and was instrumental in facilitating my transition to MIT. I would also like to thank Alan Plumb, who contributed significant research insights and support as a member of my thesis committee and chair of PAOC. I am also grateful for MIT Department of Aeronautics and Astronomics faculty members Jon How and Nicholas Roy for including me in their research group and providing financial assistance during the latter part of my MIT education.

I gratefully acknowledge Gregory Hakim (University of Washington), Ryan Torn (University of Albany), Christopher Davis (National Center for Atmospheric Research) and Kerry Emanuel, whose previous works on PV inversion inspired my interest in statistical PV regression, and Leonard Smith (University of Oxford) and Daniel Wilks
(Cornell University) whose previous works inspired my interest in minimum spanning tree rank histograms, and Cecile Penland (NOAA Earth Systems Research Laboratory), whose previous works on Linear Inverse Modeling fostered my interest in statistical inversion techniques. I also thank Cecile for being a member of my thesis committee and for inspiring me to explore new research angles. I am appreciative for the valuable comments about potential vorticity inversion from John Nielsen-Gammon (Texas A&M University), Lodovica Illari (MIT), Gregory Hakim, Craig Bishop (NRL), and an anonymous reviewer and acknowledge Lodovica Illari, Gregory Hakim, Ryan Torn, and Christopher Davis for providing useful data and code.

I am also indebted to my high school science teacher Robert Mallin, whose excellent chemistry and physics teaching developed my interest in science, and Daniel Wilks, whose advising and teaching during my undergraduate education at Cornell University fostered my interest in statistical atmospheric science and ensemble methods.

I would also like to thank my friends at MIT for their academic support, camaraderie, and encouragement. I would particularly like to thank Brian Tang, John Williams, Tim Whitcomb, John Moskaitis, Vikram Khade, Greg Lawson, Angela Zalucha, Kelly Klima, Kevin Albert, Amit Gupta, and Andreas Kellas, who, along with being partners for stimulating academic discussions, helped make my time at MIT outside of the lab enjoyable.

Away from MIT, I am extremely grateful for my dearest friends Lucy Mansdorf, Scott Mishara, Rich Palese, Dan Story, Tim Matlack, Teddy Salgado, and Tim Sikorski, who provided the support and necessary distractions for the success of my academic endeavors.

None of this would have been possible without the love and support from my parents and Carrie, who consoled me through my struggles, heightened my elation during the victorious moments, and undyingly provided guidance and encouragement through my time at MIT, and undoubtedly for all times in the future.
Contents

1. Introduction: Ensembles, Data Assimilation, and Field Covariability .......... 9
   1.1 Motivation and background: Atmospheric field coupling ........................ 9
   1.2 Time series field covariability techniques ........................................ 10
   1.3 Probabilistic forecasting ..................................................................... 16
   1.4 Data assimilation techniques .............................................................. 19
   1.5 Choosing initial ensemble members ...................................................... 23
   1.6 Introduction overview ....................................................................... 27

2. Theory and Applications of the Minimum Spanning Tree Rank Histogram . . . . 31
   2.1 Introduction ....................................................................................... 31
   2.2 \(L_2\), variance, and Mahalanobis MST distance norms. ......................... 37
   2.3 Averaged univariate rank histograms and MST rank histogram comparison 47
   2.4 Description of data ........................................................................... 49
   2.5 Analysis of the multidimensional reliability of weather components .......... 50
   2.6 Conclusion ......................................................................................... 60

3. Ensemble Regression .................................................................................. 63
   3.1 Ensemble Synoptic Analysis (ESA) ....................................................... 63
   3.2 Defining Ensemble Regression (ER) ..................................................... 65
   3.3 Truncated SVD and principal component subspaces ............................. 69
   3.4 Perturbation dynamical representativity and resolvability ...................... 73
3.5 Chapter summary ........................................... 75

4. Ensemble Regression Prediction Error ........................................... 77

4.1 Introduction .............................................. 77
4.2 ER Lorenz ’63 forecast skill and tangent linear model comparison .... 78
4.3 ER Lorenz ’95 ensemble size and sampling error sensitivity ........... 84
4.4 Operational data ER forecasting and singular vector truncation sensitivity . 89
4.5 Chapter summary ........................................... 94

5. Piecewise Potential Vorticity Ensemble Regression ........................................... 96

5.1 Introduction .............................................. 96
5.2 PV thinking and Davis and Emanuel (1991) piecewise PV inversion .... 97
5.3 Defining PV ER ........................................... 100
5.4 Defining PV ER and DE91 inversion differences ............... 103
5.5 Implementation and results of the PV ER and DE91 inversion comparison 109
5.6 PV ER and DE inversion results analysis .......................... 116
5.7 The utility of PV ER ........................................... 119
5.8 PV ER and covariance localization .................................. 121
5.9 Chapter summary ........................................... 123

6. Tropical Cyclone Track ER Sensitivity Analysis and Forecasting ............... 124

6.1 Introduction: Motivating non-contemporaneous ER .................... 124
6.2 A contrived example of dynamical inference from ER .............. 126
6.3 Geostrophy ER ............................................... 130
6.4 Experimental motivation and design ......................... 132
6.5 Sepat track ER Results and analysis .......................... 140
6.6 ER preemptive forecasting .................................... 151
6.7 Comparing univariate ESA sensitivities to multivariate ER sensitivities . 157
6.8 Summary and discussion ...................................... 163

7. Comparing ER and Tangent Linear Model Singular Vectors .......... 164
  7.1 Introduction: Defining analysis error covariance normed singular vectors .164
  7.2 ER and singular vectors ....................................... 168
  7.3 Experimental setup ........................................... 169
  7.4 Results ............................................................. 170
  7.5 Conclusions ......................................................... 173

8. Conclusions .......................................................... 176
  8.1 Summary and conclusion ......................................... 176
  8.2 Future work ......................................................... 180
Chapter 1

Introduction: Ensembles, Data Assimilation, and Field Covariability

1.1 Motivation and background: Atmospheric field coupling

The ultimate goal of atmospheric research is the complete understanding and perfect prediction of the atmosphere. A natural prerequisite for and implication of the success of this goal is the knowledge of the relationships between components of the atmosphere. This thesis seeks to develop and apply a new technique for computing and understanding these field couplings.

Much of the current understanding of the relationships between atmospheric fields comes from analytics derived from atmospheric physics. Modern numerical weather prediction models, for example, relate an initial condition field with a forecast field via approximations to several physical equations governing the time evolution of momentum (Navier-Stokes equations), temperature (first law of thermodynamics), density (mass conservation) and water (water vapor mixing ratio conservation). Similarly, sophisticated high-order approximations to atmospheric balance equations enable cyclogenesis diagnostics by relating the potential vorticity field to the geopotential field (e.g. Davis and Emanuel 1991).
However, despite the atmospheric science community’s remarkable success in forecasting and diagnostics achieved by employing physical models, several limitations motivate the use of development and use of alternative methods. First and foremost, analytical equations relating most fields are unknown, and so quantitative estimations of their couplings cannot be derived from known physics. Secondly, although they are still often highly effective diagnostic tools, some analytical models employ approximations and conditions that are unrepresentative of reality, making any derived inferences conditional on the assumptions of these toy atmospheres. Thirdly, computational costs, speed, and coding difficulties motivate faster and more easily implementable alternatives to physics-based field coupling diagnostics and forecasts.

1.2 Time series field covariability techniques

One alternative is to empirically determine field couplings from observational data. In a class of methods often referred to as field covariability techniques, relationships between multivariate fields are inferred from the covariance between two fields of interest, as determined by treating each observation drawn from the historical observational time series as an independent statistical sample.

Several field covariability techniques have been developed and used extensively to make forecasts and inferences about the dynamics of coupled fields. Although differences exist between each of the methods detailed below, at the heart of each is the use of covariance information between one field and another to make predictions about how one field changes given a change in the coupled field. This idea will be further
discussed in section 3.2. The following describes several leading time series based field covariability techniques.

Linear Inverse Modelling (LIM; e.g. Penland 1989) is a field covariability technique that makes the assumption that the state anomaly field is modeled by a linear Markov process,

$$\frac{dx}{dt} = Bx + \xi .$$

(1.1)

such that the rate of change of a state anomaly, $x$, is equal to the product of $x$ and a linear constant feedback matrix, $B$, plus Gaussian white noise, $\xi$. In other words, LIM assumes that system evolution has both a linear deterministic forcing and a forcing that is assumed to be white-noise that can be correlated in space, but not over time. Solving 1.1 gives

$$x(t + \tau) = G(\tau)x(t) + \sigma(t + \tau),$$

(1.2)

where $\tau$ is some lag after initial time $t$, $\sigma(t + \tau)$ is the forecast error, and

$$B = \tau^{-1} \ln[G(\tau)],$$

(1.3)

where

$$G(\tau) = \exp(B \tau) = \langle x(t + \tau)x^T(t) \rangle \langle x(t)x^T(t) \rangle^{-1},$$

(1.4)

and the angle brackets denote an ensemble average. That is, the linear deterministic forcing is approximated using multiple least regression from contemporaneous and time-lagged covariance estimates computed using observational time series data. The covariance of the stochastic forcing can be estimated by a fluctuation-dissipation relation as

$$\langle \xi \xi^T \rangle dt = -B \langle x(t)x(t)^T \rangle - \langle x(t)x(t)^T \rangle B^T ,$$

(1.5)
which can be used to analyze the nature of the noise. Given that the forecast error is independent of the initial state, the best estimate of the state at time $t + \tau$, in a least squares sense, is given by

$$\hat{x}(t + \tau) = G(\tau)x(t).$$

(1.6)

This forecasting technique has been used to make seasonal and interannual predictions of tropical Atlantic sea surface temperatures (Penland and Matrosova 1998), and northern hemisphere 700 hPa geopotential height anomalies (Penland and Ghil 1993), among others.

One significant virtue of LIM over other field covariability models is that it relates the observed lagged statistics to the system dynamics. Once it is shown that the model given by 1.1 satisfactorily describes the observations, it is possible to extract meaningful information from the structures of $B$ that can further the understanding of the system dynamics. Such analysis has been used to study the influences of El Niño and the Madden-Julian oscillation on extratropical low-frequency variability (Winkler et. al 2001) and the Pacific decadal oscillation (Alexander et. al 2008).

Another standard field covariability technique is canonical correlation analysis (CCA; Barnett and Preisendorfer 1987; Nicholls 1987). CCA operates by separately projecting the two fields onto vector pairs with successively decreasing maximum intrapair correlation and zero cross-pair correlation. That is, the predictor, $X$, and predictand, $Y$, are projected onto the canonical variates, $V$ and $W$, respectively, such that

$$\text{corr}(v_{m}, w_{m}) = \begin{cases} r_{m}, & \alpha = m \\ 0, & \alpha \neq m \end{cases}$$

and
\[ \text{corr}(v_1, w_1) \geq \text{corr}(v_2, w_2) \geq \ldots \geq \text{corr}(v_m, w_m) \geq 0 \] (Wilks 2006). \hfill (1.8)

Here, \textit{corr} is shorthand notation for the Pearson product-moment correlation of the indicated arguments, \( v \) and \( w \) are columns of \( V \) and \( W \), respectively, the \( r_c \) are the canonical correlations, \( \alpha = 1, 2, \ldots, M \), \( m = 1, 2, \ldots, M \), and \( M \) is the number of pairs of canonical variates. The \( v_m \) and \( w_m \) are computed as

\[
V = AX \hfill (1.9)
\]

and

\[
W = BY , \hfill (1.10)
\]

where

\[
A^T = S_{xx}^{-1/2} E \hfill (1.11)
\]

and

\[
B^T = S_{yy}^{-1/2} F , \hfill (1.12)
\]

Here, \( E \) and \( F \) are the eigenvectors of \( S_{xx} \) and \( S_{yy} \), respectively, where \( S_{xx} \) and \( S_{yy} \) are the covariances of the subscripted arguments. \( E \), \( F \), and \( R_c \) can be computed as

\[
\text{SVD}\left( S_{xx}^{-1/2} S_{xy} S_{yy}^{-1/2} \right) = ER_c F^T , \hfill (1.13)
\]

where SVD is shorthand notation for the singular value decomposition of the argument and \( R_c \) is a diagonal matrix of canonical correlations of decreasing magnitude.

The primary virtue of CCA is that the projection onto the canonical variates ensures that the analysis is performed in the subspace of maximal predictor and predictand correlations, thereby facilitating the identification of the \textit{coupled} variability between the predictor and predictand field. This characteristic contrasts other methods, such as principal component analysis, that identify patterns of variability within a single
dataset. Typically, CCA proceeds by analyzing canonical vector and canonical
correlation maps to make inferences about the nature of the relationship between
variations of the predictor and predictand data. Such analysis has suggested a coupling of
the variability of the North Atlantic and North Pacific sea surface temperatures due to a
mutual relationship with planetary wave patterns (Wallace et al. 1992). Additionally, the
United States Climate Prediction Center has employed CCA for seasonal weather
forecasting via
\[ \hat{W} = R_c V, \]
where the $\hat{\cdot}$ denotes the predicted value of the predictand as determined by the canonical
predictor (Barnston et al. 1999).

The fluctuation-dissipation theorem (FDT), a cornerstone of modern statistical
mechanics, is another field covariability technique used by climate researchers. The FDT
states that the average mean response of a statistical system to small external
perturbations can be calculated through the knowledge of suitable correlation functions of
the unperturbed system (Majda et al. 2005). Following Leith (1975), the FDT implies
that the regression matrix for linear regression prediction of a time series, $G(\tau)$, can be
determined as
\[ G(\tau) = U(\tau)U^{-1}(0), \]
where $U$ denotes the lagged covariance matrix of two arbitrary fields with the lag denoted
by the parenthetical argument. This form of the FDT regression matrix is very similar to
that of LIM (viz. 1.4), however, unlike LIM, more general forms of the FDT are not
strictly limited to linear dynamics (Gritsun and Branstator 2007; Gritsun et al. 2008).
LIM, CCA, FDT and other field covariability techniques all assume weak stationarity (e.g. Wilks 2006), which asserts that the mean and autocovariance function of the data time series are constant in time. This assumption implies that that correlations between variables are only functions of lag and not state, thereby admitting the use of statistics describing the past to make inferences about the present and future.

However, as exemplified by the strong diurnal temperature cycle, few geophysical processes are truly stationary; even many climate-scale phenomena, such as El Niño, have distinct annual and seasonal periods (e.g. Penland and Sardeshmukh 1995). Accordingly, a cyclostationarity transformation is often applied to periodic data so that methods that assume stationarity can be applied. This transformation attempts to remove cycles that illegitimize the stationarity assumption by normalizing the mean and variance of subsections of the time series with common statistical moments, so that the mean and variance of the resulting time series is homogenized (e.g. Wilks 2006).

However, even cyclostationarity is a poor assumption for the modeling of some atmospheric processes. From the comparison of the statistics of temporally neighboring ensemble analysis probability distribution functions (to be described below), random data samples drawn from the same time series subsection are seldom characterized by the same mean and covariance (Kalnay 2003; Toth and Kalnay 1997; Pu et al. 1997). Synoptic-scale statistics have “errors of the day” that are typically driven by rapidly evolving baroclinic instabilities in the background flow (e.g. Toth and Kalnay 1997); such instabilities can alter the atmospheric statistics so frequently that often only instantaneous statistics, not even time series statistics from the most recent past, can capture the current atmospheric uncertainty.
Therefore, despite the proven utility of time series based field covariability techniques for the modeling of quasi-cyclostationary atmospheric processes, quickly evolving “errors of the day” motivate an alternative approach for the statistical modeling of specific markedly flow-dependent atmospheric phenomena. This alternative is the use of *instantaneous flow-dependent ensemble statistics* for synoptic-scale field covariability studies, which will be the primary focus of this thesis. Before discussing how ensembles can be used for field covariability, the next sections review the motivation for and formulation of ensembles in the atmospheric sciences.

### 1.3 Probabilistic forecasting

Although only recently have ensembles become fundamental to atmospheric prediction, the motivation for ensemble forecasting dates back to Edward Lorenz’s 1963 experiment that laid the groundwork for the study of atmospheric predictability. Attempting to illustrate the futility of statistical prediction for atmosphere-like systems characterized by nonperiodicity (Lorenz 1993), Lorenz integrated a low-order system of equations

\[
\begin{align*}
\dot{x} &= -a(x - y) \quad (1.16) \\
\dot{y} &= rx - xz - y \quad (1.17) \\
\dot{z} &= xy - bz \quad (1.18)
\end{align*}
\]

(where \( a = 10 \), \( b = 8/3 \), and \( r = 28 \) to ensure chaotic dynamics that mimic those of the real atmosphere (Lorenz 1963)) twice using initial conditions that differed only by a round-off error. Although highly similar at the start of the integrations, Lorenz observed
that the two solutions diverged after a few model days and eventually showed little resemblance. He concluded that the atmosphere is chaotically unstable with respect to perturbations of small amplitude and postulated that influences as seemingly trivial as a butterfly flapping its wings could potentially determine whether or not a tornado forms in another part of the world (Lorenz 1993).

Given the sensitive dependence of atmospheric flows to small perturbations, it became clear that the quality of atmospheric forecasts is a function of the proper specification of the conditions used to initialize the model forecast equations. However, as was speculated by Lorenz and his contemporaries and later proven by Judd and Smith (2001), even given the exact system dynamics and a highly accurate time series of past observations, it is impossible to know the exact true state of a chaotic system when observations are imperfect, as is always the case; many past trajectories through state space that are indistinguishable from the true trajectory may have “shadowed” the true state (Judd and Smith 2001). Accordingly, atmospheric scientists slowly began to abandon the notion of strictly deterministic weather forecasting, and sought to develop probabilistic approaches that recognized the unavoidable uncertainty of the true atmospheric state. The following further describes the development of probabilistic weather forecasting.

Probabilistic weather prediction involves both a forecasting step and an assimilation step. The forecasting step evolves a probability distribution function (PDF), \( \phi \), from an initial time to a final forecast time. In a method called stochastic dynamic prediction, this evolution can be solved using the Fokker-Planck equation, which, for systems with deterministic dynamics, can be simplified to the Liouville equation.
\[ \frac{\partial \phi}{\partial t} + \nabla \cdot (\dot{x} \phi) = 0 \] (Epstein, 1969),

(1.19)

where \( \nabla \) is a gradient operator, and \( \dot{x} \) denotes the time rate of change at a point, \( x \), in state space. However, although the Liouville equation, which models the continuity of probability through phase space, is strictly the correct method to evolve a PDF, typical numerical prediction weather model state dimensions exceed \( 10^7 \) numbers, making solving 1.19 prohibitively expensive. Epstein (1969) showed that this computational limitation can be overcome by using the Liouville equation to integrate individual moments of the PDF, rather than the full PDF. By invoking Gaussian moment closure, Epstein (1969) approximated the full PDF evolution by evolving only the mean and covariance estimates.

However, inaccuracies attributable to neglecting higher order moments of the PDF motivated alternative approaches. Leith (1974) showed that PDF evolution can be approximated using Monte Carlo methods. He suggested that the evolved PDF can be approximated by integrating the system equations multiple times, each time using a different initial state drawn from the initial PDF; the atmosphere’s sensitive dependence to the initial conditions could simply be addressed by launching a set of forecasts from different initial states, rather than a single deterministic forecast from a single state. Leith found that, as the number of ensemble members approached infinity, the PDF approximated from the ensemble of integrations approached that deduced from stochastic dynamic prediction. Moreover, he showed that even small ensemble sizes yield adequate approximations, thereby providing a sufficiently accurate and computationally feasible approximation to the Liouville equation. This Monte Carlo technique laid the
groundwork for the forecasting step of ensemble weather prediction, the gold standard of modern operational probabilistic prediction.

Naturally, the accuracy of ensemble forecasts hinges on whether initial ensemble members are drawn from the PDF that best approximates the uncertainty of the initial time state. Accordingly, the implementation of Monte Carlo ensemble forecasting requires an estimate of the current atmospheric state and its associated uncertainty. Section 1.4 describes this assimilation procedure and outlines several data assimilation techniques. Section 1.5 discusses methods by which ensemble members used as initial conditions for the following forecast step can be chosen based on the state estimation.

1.4 Data assimilation techniques

The majority of state estimation techniques employ a variational or Kalman filter approach to approximate the state of the atmosphere and its associated uncertainty. Both approaches seek the most probable estimate of the current state by combining uncertainties in the observations with uncertainties in prior estimates of the current state. This can be cast as an application of Bayes rule attempting to find the probability of the state at time $t$, $x(t)$, given observations up to and including the present, $Y(t)$, such that

$$
\phi(x(t) \mid Y(t)) \propto \phi(y^o(t) \mid x(t)) \phi(x(t) \mid Y(t-1)).
$$

(1.20)

That is, $\phi(x(t) \mid Y(t))$ is proportional to the product of the probability of the current observations, $y^o(t)$, given the current state and the probability of the state given prior observations, $Y(t-1)$, which can be estimated from a short-term forecast (D’Agostini 2003; Lawson, 2005).
Variational assimilation approaches seek to find the optimal analysis field that minimizes a cost function, $J(x)$, defined by the distances of this analysis to a background state (typically a short-term forecast estimate), $x_b$, and to observations, $y_o$, with a typical form given by

$$J(x) = \frac{1}{2} (x - x_b)^T B^{-1} (x - x_b) + [y - H(x)]^T R^{-1} [y - H(x)] \quad (e.g. \text{Lorenc 1986}).$$

(1.21)

Here, $B$ and $R$ respectively denote the uncertainties of the background and observations, and $H$ represents the measurement matrix, a linear operator that maps the state from the model space to the observation space.

Although variational schemes are widely used operationally in the forms of a three-dimensional variational schemes (3D-Var; e.g. Lorenc 1986; Parrish and Derber 1992), which typically employs a cost function similar to that given by 1.21, and four-dimensional variational scheme (4D-Var; e.g. Rabier et al. 2000), which adds a constraint that the minimizing trajectory satisfy the model dynamics, they make two assumptions about the background covariance matrix that deem them suboptimal. Firstly, such approaches typically assume a spectrally separable background covariance matrix that implies that the matrix is diagonal when represented by spectral elements. Given that nonlinear flows are marked by the transfer of information between scales, assuming the independence of elements of different spatial scales is inappropriate (e.g. Kalnay 2003). Secondly, variational approaches yield only an analysis estimate and not an estimate of its associated uncertainty (e.g. Kalnay 2003), which is a particularly unattractive quality given that initial ensemble members should ideally be chosen from this uncertainty distribution. Thirdly, the 3D-Var approach assumes covariance stationarity, rendering it incapable of accounting for forecast errors of the day, such as those attributable to
quickly evolving synoptic scale baroclinic instabilities (e.g. Kalnay 2003). This
assumptions makes 3D-Var particularly suboptimal for producing the flow-dependent
ensemble data desirable for synoptic-scale field covariability studies. Note, however, that
the 4D-Var approach is implicitly able to evolve the forecast error covariance (Thepaut et
al., 1993) and can therefore produce flow-dependent statistics.

Kalman filter techniques overcome these shortcomings by employing explicitly
evolving covariance estimates. Denoting $F$ as a linear model, superscripts $a$ and $f$ as
abbreviations for analysis and forecast, respectively, and the covariances of current and
forecast estimates of the state as $P^a$ and $P^f$, respectively, the Kalman filter equations
are given by

$$x^f(t) = F[x^a(t - 1)]$$  \hspace{1cm} (1.22)

$$P^f(t) = FP^a(t - 1)F^T$$  \hspace{1cm} (1.23)

$$K(t) = P^f(t)H(t)^T[H(t)P^f(t)H(t)^T + R(t)]^{-1}$$  \hspace{1cm} (1.24)

$$x^a(t) = x^f(t) + K(t)[y^o(t) - H(t)x^f(t)]$$  \hspace{1cm} (1.25)

$$P^a(t) = [I - K(t)H(t)]P^f(t)$$  \hspace{1cm} (1.26)

(Kalman 1960; Houtekamer and Mitchell, 2005). Using the Kalman gain matrix, $K$,
which weights the influences of the observations and the prior estimate of the state based
on their associated uncertainties (viz. 1.24), this set of equations yields both the analysis,$x^a(t)$, the best estimate of the current state (viz. 1.25), and its associated error
covariance, $P^a$ (viz. 1.26). Note that the evolving forecast error covariance matrix is
explicitly computed by integrating the model at each forecast step (viz. 1.23), thereby
ensuring that it is a function of the most current state estimate.
Several variations have been made to the Kalman filter equations to improve the estimation of the state, the most popular of which is called the extended Kalman filter (EKF; Jazwinski 1970; Ghil and Malanotte-Rizzoli 1991). Accounting for the nonlinearity of the atmospheric model equations, the EKF allows for the use of a nonlinear model rather than linear equations to integrate the state (viz. 1.22), a linearized version of the nonlinear model to integrate the analysis covariance (viz. 1.23), and a linearized version of the nonlinear measurement matrix (viz. 1.24-1.26). However, although it more appropriately accounts for the nonlinear system physics, EKF state estimations can be inaccurate because of their neglect of higher order moments in the integration of the error covariance, as was the case with Epstein’s (1969) approximate stochastic dynamic prediction; propagating only the first two moments can potentially lead to unbounded error variance growth (Evensen 1992) and is inappropriate for data assimilations of the model state variables such of moisture that are sensitive to motions at small scales where errors grow and saturate rapidly (Evensen 1992).

Analogous to how inaccuracies attributable to stochastic dynamic prediction closure can be mitigated by implementing a Monte Carlo forecasting technique (e.g. Leith 1974), Kalman filter closure errors can be mitigated by using a Monte Carlo approximation to the Kalman filter forecast step. The Ensemble Kalman Filter (EnKF; Evensen 1994; Burgers et al. 1998) is a Kalman filter variation that integrates an ensemble of states to estimate the forecast error covariance. The EnKF equations are given by

\[ x^f(t) = F[x^f(t-1)] \]  
\[ P^f(t) = \frac{1}{n_{ens} - 1} X^f X^{f'} \]  

\(1.27\)  
\(1.28\)
\[ K(t) = P^f(t)H(t)^T[H(t)P^f(t)H(t)^T + R(t)]^{-1} \]  
(1.29)

\[ x_i^a(t) = x_i^f(t) + K(t)[y_i^o(t) - H(t)x_i^f(t)] \]  
(1.30)

\[ P^a(t) = \frac{1}{n_{ens} - 1}X^a^T X^a, \]  
(1.31)

where \( i = 1, \ldots, n_{ens} \), \( n_{ens} \) is the number of ensemble members, the prime superscript denotes that the ensemble mean has been removed, and \( X \) is the matrix comprised of \( n_{ens} \) columns \( x_i \). The analysis and forecast error covariance matrices are simply estimated using the ensemble of nonlinearly estimated states at analysis and forecast time, respectively, as statistical samples; no moments are truncated in the uncertainty propagation, thereby avoiding closure errors.

Although its treatment of closure errors is a primary benefit of the EnKF over the EKF, arguably the most important benefit of using the EnKF is its natural provision of the initial ensemble members appropriate for ensemble forecasting, as will be discussed in the following section.

### 1.5 Choosing initial ensemble members

The previous section outlined several techniques that estimate the state of the atmosphere and its associated uncertainty. This section discusses how these estimates are used to choose the initial conditions for the ensemble forecasts that yield the flow-dependent output statistics required for ensemble-based field covariability techniques. These ensemble formulation techniques can generally be distinguished between those that define ensemble members by adding perturbations to the analysis mean state (such as the
random perturbation, bred vector, and singular vector techniques), and those that define members as draws from the analysis uncertainty distribution (such as Kalman filter-based techniques). Note too that there exists yet another class of ensemble formulation techniques that yields probabilistic forecasts by running model integrations with varying model physics, rather than perturbed initial conditions (e.g. multisystem or poor-man’s ensembles; Krishnamurti et al. 2000).

Among the first class of techniques, Leith’s (1974) method proposed that ensemble members can be simply defined as random perturbations from the analysis control state with magnitudes characteristic of the instrumental uncertainty. These members, however, lacked appropriate dispersion, underestimated forecast uncertainties due to dynamical inconsistencies, and led to model energy dissipation (Palmer et al. 1990), which motivated a dynamical basis for ensemble selection.

One dynamics-based approach currently used operationally by the National Center for Environmental Prediction is the bred vector technique (e.g. Toth and Kalnay 1993), which defines the directions of the perturbations as the directions of the vector differences of a nonlinear control forecasts and random nonlinear perturbation forecasts. These vectors differences are scaled down to the amplitude of the initial perturbation, added to the analysis mean, and then used as the perturbation for the subsequent breeding cycle. Bred vector ensemble members capture flow-dependent statistics because they sample the errors in recent phase space trajectories, which are functions of the current underlying flow instabilities.

The European Center for Medium Range Forecasting (ECMWF) employs the singular vector (SV) ensemble prediction system. As will be discussed in chapter 7, SV
perturbations are defined as the perturbations at initial time that evolve into the leading eigenvectors, under a specified norm, of the covariance of the forward propagator (typically approximated by the tangent linear model). Because these eigenvectors represent the most highly varying perturbations at the end of the optimization interval, the initial ensemble members are optimized to sample the most dynamically sensitive regions of the analysis distribution (e.g. Molteni et al. 1996).

Although ensemble members derived from bred and singular vectors are dynamically sensitive, they are not inherently conditioned to be realistic draws from the analysis uncertainty distribution; perturbations that potentially evolve into the most dynamically meaningful perturbations may consistently be associated with low probability portions of the analysis uncertainty distribution. Since only ensemble members that appropriately sample the full initial uncertainty distribution can be expected to correctly estimate the forecast distribution after being integrated, choosing members without regard to their probability of occurrence may yield skewed forecast uncertainty information.

Addressing this shortcoming, a second class of ensemble formulation techniques ensures that members are potentially realizable by defining them as draws from the analysis uncertainty distribution derived from Kalman filtering techniques (e.g. 1.26 and 1.31). This class defines the likelihood of choosing an ensemble member with a given direction in phase space by the value of the PDF that corresponds with that direction. Moreover, in addition to being realizable draws, members chosen from Kalman filter-based techniques are ensured to be flow-dependent because their directions are defined by the most recent trajectories through model phase space.
When using the EKF to assimilate the data, ensemble members can simply be chosen as random draws from the analysis EKF PDF. Alternatively, when using the EnKF, ensemble members can be defined simply as the analysis states from the previous assimilation step (viz. 1.30); in other words, the output from EnKF data assimilation defines the EnKF initial ensemble. This highly attractive property of the EnKF highlights the natural coupled relationship between data assimilation and ensemble forecasting for these sophisticated ensemble prediction systems. Unfortunately, however, no operational centers currently employ this EnKF technique as of the time of this writing.

Once an ensemble is initialized by one of the methods previously outlined, each member is used as the initial conditions for a numerical weather prediction model run. Given that the initial ensemble correctly samples the initial uncertainty distribution, each individual ensemble forecast state defines an equally likely realization of the future atmospheric state and the joint distribution of ensemble forecasts defines the forecast uncertainty distribution (e.g. Kalnay 2003). Ensemble forecasts are typically displayed as a spaghetti diagram, in which each spaghetti strand depicts a potential forecast state and the ensemble of strands portrays the range of potential realizations. Figure 1.1, for example, displays the five-day ensemble forecast spaghetti diagram illustrating the uncertainty of the position of the trough over the eastern half of the United States on 05 November 1995 (from Sivillo 1997).

Because ensembles yield random draws from density functions of potentially realizable model states, it is natural and effectual to use them to quantify weather forecast uncertainty. It is regrettably squandering, however, that uncertainty estimation is virtually the exclusive use of ensemble forecast output; this thesis asserts that the utility
Figure 1.1: This figure displays the five-day ensemble forecast spaghetti diagram that illustrates the uncertainty of the position of the trough over the eastern half of the United States on 05 November 1995 (from Sivillo 1997).

... of ensembles transcends being a measure of forecast spread and develops techniques that extract grossly underutilized information supplied by ensembles.

1.6 Introduction overview

Recent progress in climate science has benefited greatly from field covariability studies. The physical inferences and regression equations deduced from the analysis of statistical relationships between atmospheric and oceanic fields have furthered the understanding, modeling, and forecasting of the El Niño Southern Oscillation (Nicholls
1987), the coupling of 500-mb height anomalies and wintertime sea surface temperatures (Wallace et al. 1992), and other climate mechanisms.

Synoptic meteorology, however, has not seen a significant corresponding advancement. The assumption of covariance stationarity underlying climatological field coupling studies (e.g. Penland 1989; Bretherton et al. 1992) is inappropriate for the statistical modeling of specific synoptic events, whose covariances are strong functions of position in state-space (e.g. Lorenz 1963). In contrast to climatological studies, in which the stationarity assumption allows one to obtain statistical samples over significant time intervals, the requirement of flow-dependent cross-covariances for synoptic field covariability studies implies the necessity of instantaneous flow-dependent samples. Obtaining these samples by taking simultaneous observations violates the requirement of sample independence, necessitating an alternative approach.

Using ensemble analyses and/or forecasts as samples is this alternative. Ensemble analyses eliminate the climatological constraint on field covariability studies by providing instantaneous and independent state estimations that are consistent with the uncertainty associated with the propagation of imperfect observations under nonlinear chaotic dynamics (Lorenz 1963). Ensembles that come from data assimilation systems that use evolving error covariance estimates, such as methods based on the Kalman filter and 4D-Var, are particularly attractive because these ensembles capture the "errors of the day" typically associated with the current atmosphere's underlying baroclinic instabilities (e.g. Toth and Kalnay 1997).

One overlooked application of ensembles is Ensemble Synoptic Analysis (ESA; Hakim and Torn 2008), the use of ensemble analysis and forecast covariances to make
inferences about the atmosphere. By treating individual ensemble members as independent samples, ESA employs standard statistical techniques to detect sensitivities, infer dynamical couplings, and aid forecasters in identifying dynamical processes that are particularly relevant for specific weather predictions. The focus of this thesis is to develop and apply an ESA method called Ensemble Regression (ER), which facilitates inference about the relationship between two multidimensional atmospheric fields via the regression of a perturbation using an operator defined by the covariances of the fields’ ensemble forecasts and/or analyses.

Noting that the significance of the results of ensemble-based field covariability techniques such as ER is a function of the reliability of the ensemble distributions, it is advisable to assess the multidimensional reliability of the ensemble being used for ER. Therefore, before defining and applying ER in chapters 3-7, chapter 2 develops the theory and presents applications for a multivariate ensemble verification tool called the minimum spanning tree rank histogram.

Chapter 3 returns to the focus of this thesis by defining Ensemble Regression and discussing key properties of the technique. Chapter 4 employs low order Lorenz models and high-dimensional operational data to apply ER as a tool for predicting future states and discusses potential sources of forecast error. Chapter 5 develops the use of field covariances to invert potential vorticity and compares geopotential height perturbations deduced from a potential vorticity ensemble regression to those deduced from the classical piecewise inversion technique of Davis and Emanuel (1991). Chapter 6 applies ER as a non-contemporaneous multidimensional sensitivity tool to analyze the sensitivities of tropical cyclone tracks to prior mid-tropospheric geopotential height
perturbations and motivates ER's use for preemptive forecasting and sensitivity forecast guidance. Chapter 7 presents how ER can be used for singular vector analysis and uses a Lorenz model to compare singular vectors deduced from ER to those from a tangent linear model. Chapter 8 presents conclusions and ideas for future work.
Chapter 2

Theory and Applications of the Minimum Spanning Tree Rank Histogram

2.1 Introduction

This thesis focuses on a multivariate field covariability technique that uses the covariances from ensemble model output to make predictions and dynamical inferences about the atmosphere. The significance of the results of ensemble regression is intimately related to the validity of the ensemble covariances that describe the uncertainty of analysis and forecast model output. Therefore, before defining and applying ER, this thesis develops the theory and presents applications for a verification technique called the minimum spanning tree rank histogram that assesses the extent to which multivariate forecast ensembles used for ER and other applications are potential realizations of the true future state of the atmosphere.

The identification of ideal verification techniques requires an understanding of the nature of goodness in weather forecasting. Weather forecast goodness is typically defined in terms of a forecast’s consistency, value, and quality (Murphy 1993), which is further subdivided into components that include sharpness, resolution, and reliability (Murphy 1993). Since no known verification measure satisfactorily addresses all aspects of goodness, it is necessary for a verification tool to address an individual aspect. This chapter focuses on the assessment of ensemble reliability, which is defined as the
correspondence between the mean of the observations associated with a particular forecast and that forecast, averaged over all forecasts. A perfectly reliable 30% chance precipitation forecast, for example, verifies exactly 30% of the time (Murphy 1993). It is important to reiterate that, although they are extremely important measures of forecast goodness, this chapter is not concerned with the assessment of forecast sharpness and resolution.

Reliability can be measured by the degree to which the ensemble forecast members and truth are random samples from the same PDF. For scalar forecasts, this degree can be assessed by the shape of a rank histogram (RH), or Talagrand diagram (Anderson 1996; Talagrand et. al. 1997). The scalar RH is simply a histogram of the $N$ verification ranks over $N$ independent forecast occasions. Each verification rank is defined as the rank of the verification entry in a forecast’s $n_{ens} + 1$ member vector comprised of an individual forecast’s $n_{ens}$ ensemble entries and the corresponding verification entry, sorted in ascending order. Therefore, the histogram’s shape depends on the population of the $n_{ens} + 1$ bins, as determined by the $N$ ranks of the verification entries in the $N$ vectors.

An equal representation of ranks, as indicated by a flat histogram, implies that the members of the ensemble forecast and the verification are random draws from the same PDF: they are statistically indistinguishable. This is easily conceptualized by thinking of the forecast cumulative distribution function (CDF) as a transfer function between the forecast PDF and a uniform distribution. Figure 2.1 shows a continuous schematic of this idea. The forecast PDF is shown in figure 2.1c (upside down for convenience), the associated CDF in figure 2.1b, and the uniform distribution that results from using the
CDF to transform random draws from the PDF in figure 2.1a. The circles in figure 2.1c represent the boundaries between areas of equal probability; the integral of the PDF between each circle is the same. Note that the functional form of the PDF (which is arbitrary) results in unequal spacing between the circles. When these points are transformed by the CDF in figure 2.1b (dashed lines guide the eye), they result in the uniform distribution in figure 2.1a. In the construction of an RH, the circles in figure 2.1 are defined by the forecast ensemble, their rank ordering approximates the forecast CDF, and verification populates the bins of the transformed distribution (Gombos and Hansen 2007).

Figure 2.1: An illustration of the CDF as a transfer function from a PDF to a uniform distribution. See text for details. From Gombos and Hansen (2007).
Traditional RHs are used to assess one-dimensional forecasts. The atmosphere, however, is far from one-dimensional. Because of the covariance between dimensions, averaging univariate RHs to assess the multidimensional reliability can give misleading information (Smith and Hansen 2004). Therefore, in order to accurately assess the reliability of multidimensional fields, it is desirable to formulate a multidimensional extension of the RH that accounts for this covariance. A contrived comparison of univariate and multivariate MST rank histograms is presented in section 2.3.

One such extension is the minimum spanning tree (MST) RH. Consider a $K$ dimensional space, where each dimension could correspond to one of $K$ individual weather components, such as temperature or pressure, to the same component in one of $K$ different locations, or to a combination of components and locations. Let each point, $x_{i,j,k}$, in this $K$ dimensional space correspond to the value of the $k$th element of the $j$th ensemble member on the $i$th forecast occasion, where $i = 1,...,N$, $j = 1,...,\text{ens}$, and $k = 1,...,K$. The MST of this set of points is defined by the sum of the lengths (under a chosen norm) of the $\text{ens} - 1$ line segments that connect these points, subject to the restrictions that the resulting network has no closed loops and that the distance is minimized (Smith and Hansen 2004; Wilks 2004). Figure 2.2a shows an example of a two-dimensional MST with $\text{ens} = 15$ and 24 hour lead time, where one dimension corresponds to the forecast temperature in Bangor, ME and the other to the forecast temperature in Portland, ME on 21 August 2004. Each circle represents $x_{i,j,k}$ and the sum of the $\text{ens} - 1$ line segments represents the MST. Figure 2.2b shows a three-dimensional
example of an MST, where the third dimension corresponds to the forecast temperature in Albany, NY (Gombos and Hansen 2007).

The calculation of each increment of an MST RH requires the computation of $n_{\text{ens}} + 1$ MST lengths. The first of these lengths is the MST distance of the $n_{\text{ens}}$ ensemble points alone. The other $n_{\text{ens}}$ lengths are the MST distances of the $n_{\text{ens}}$ points consisting of the union of $n_{\text{ens}} - 1$ ensemble points and the verification. The verification replaces a different ensemble member for each of these $n_{\text{ens}}$ lengths. If the ensemble members and

\[
\begin{align*}
\text{Bangor Temperature (K)} & \\
\text{Portland Temperature (K)} & \\
\text{Albany Temperature (K)} & \\
\end{align*}
\]

Figure 2.2: Illustrations of two (a) and three-dimensional (b) MSTs for a 24 hour forecast. The dimensions are represented by the cities and the norm is 2-m temperature on 21 August 2004. Circles represent the $n_{\text{ens}} = 15$ points that could be comprised of either the ensemble only, or the union of $n_{\text{ens}} - 1$ ensemble members and the verification. The sum of the line segments represents the MST distance. From Gombos and Hansen (2007).
the verification are random draws from the same PDF, the MST length of the ensemble-only points should be statistically indistinguishable from the \( n_{\text{ens}} \) MST lengths that include the verification. Analogous to a traditional scalar RH being a plot of the rank of the verification within the \( n_{\text{ens}} + 1 \) member vector over \( N \) one-dimensional forecasts, the MST RH is a plot of the rank of the ensemble-only MST length within the \( n_{\text{ens}} + 1 \) member MST length vector.

The degree to which the ensemble and verification points are statistically indistinguishable can be quantified using the Cramér-von-Mises (CvM) goodness-of-fit test for a uniform distribution. The CvM test statistic, \( W^2 \), is given by

\[
W^2 = N^{-1} \sum_{q=1}^{n_{\text{ens}}+1} Z_q^2 m_q,
\]

(2.1)

where \( m_q \) is the probability of an observation landing in the \( q \)th bin, \( O_q \) and \( E_q \) are the observed and expected number of counts in the \( q \)th bin, respectively, and

\[
Z_q = \sum_{r=1}^{q} (O_r - E_r).
\]

(2.2)

Given the independence of each of the \( N \) forecast occasions, a histogram will be considered flat if this test statistic is less than the CvM critical value with \( n_{\text{ens}} \) degrees of freedom. Note that the CvM statistic was chosen to assess flatness because, unlike the \( \chi^2 \) statistic, it is sensitive to rank ordering and gives a more powerful goodness-of-fit assessment for small sample sizes (Elmore 2005). The CvM statistic is particularly sensitive to skewed histograms (Elmore 2005) and is therefore appropriate for the assessment of de-biased MST RHs, which are characteristically right skewed for underdispersed ensembles and left skewed for overdispersed ensembles (Wilks 2004;
Gombos and Hansen 2007). CvM critical values can be found in Table 1 of Elmore (2005).

This chapter addresses both theory and applications of the MST RH. Section 2.2 details MST distance norms and how the improper use of such norms causes misleading MST RH shapes. A contrived comparison of univariate and multivariate MST rank histograms is presented in section 2.3. Section 2.4 describes the data used to construct the MSTs used in the application section of this chapter. Section 2.5 is an analysis of separate MST RHs that were constructed by using a common city cluster, but different weather component norms. This section also compares the MST RHs from a southwestern United States city cluster and a northeastern United States city cluster. Section 2.6 presents conclusions.

2.2. $L_2$, variance, and Mahalanobis MST distance norms

A multidimensional ensemble reliability assessment determines the statistical similarities of the ensemble forecast distribution and the verification distribution. Because the MST RH determines this likeness using the ranks of MST distances, it is crucial to choose a norm for these distances that most accurately measures this statistical similarity.

The three choices of the norm considered in this section are the Euclidean $L_2$, variance, and Mahalanobis norms. Each will be described below. Other than the circumstance using the Mahalanobis norm when $n_{ens} \leq K$ described below, in the limit of large numbers of realizations, the use of each of these norms in the construction of MST
RHSs will qualitatively yield the same determination of whether or not the two distributions are alike. However, as the number of realizations decreases, the choice of norm can potentially influence the CvM statistic’s evaluation of population histogram flatness, motivating the use of the most sensitive and justifiable norm. The following describes how each of these norms can give misleading measurements about the degree of reliability of an ensemble forecast and outlines circumstances when certain norms should not be used.

The familiar Euclidean $L_2$ norm is the most intuitive and straightforward norm to use when constructing MST RHSs. However, because it does not homogenize the variances of the data, the $L_2$ norm yields a misleading MST RH when the standard deviation of the data in each of the $K$ dimensions is not the same. Consider the $K = 8$, $n_{ens} = 15$, $N = 140$ $L_2$ MST RH depicted in figure 2.3a. For this contrived example, each of the $K = 8$ dimensions represents the 2-m temperature in an individual city, and suppose that the true distribution is known. Assume that the true standard deviations of the temperatures in the first four cities are 5 K and that the forecasts for these cities are perfectly reliable, with standard deviations also equal to 5 K. Also assume that the true standard deviations in the other four cities are 1 K but that the forecasts for these cities are underdispersed, with standard deviations of only 0.1 K. Despite the underdispersion of half of the forecasts, the $L_2$ MST is relatively flat. Because it does not homogenize variances, the $L_2$ MST distances are dominated by the distances associated with the high-standard deviation dimensions; the incorrect, but small, distances associated with the low variance cities are “lost in the noise”. Of course, with a large enough number of
samples, the $L_2$ MST RHs would correctly indicate that the ensembles are drawn from the incorrect $K = 8$ distribution (Gombos and Hansen 2007).

The variance norm transforms each entry, $x_{i,j,k}^*$, of $X_i^*$ into $x_{i,j,k}^{\text{var}}$ such that

$$x_{i,j,k}^{\text{var}} = x_{i,j,k}^* / \sigma_{i,k}, \tag{2.3}$$

where $\sigma_{i,k}$ is the standard deviation of the data in the $k$th dimension and $X_i^*$ is an $(n_{\text{ens}} + 1) \times K$ matrix formed by the union of the verification vector, $o_i$, and $n_{\text{ens}}$ ensemble row vectors of length $K$, $x_{i,j}^*$. (The star superscript indicates that the ensemble has been de-biased. The bias transformation procedure will be explained in a following section.) The MST distance is then formed using the transformed $x_{i,j,k}^{\text{var}}$ entries.

A variance norm MST RH will equally weight the ensemble and verification dispersion differences in the unit directions of the cities. After each data point is divided by the standard deviation of its respective dimension, the data in each transformed dimension has unit variance. Therefore, from the previous example, a distance of 1 K in the low standard deviation temperature axis and a distance of 5 K in the high standard deviation temperature axis will be weighted equally under the variance norm, thereby enabling the MST distance to equitably account for each dimension in the reliability assessment (Gombos and Hansen 2007).

Figure 2.3b shows the variance norm MST of the same $K = 8$, $n_{\text{ens}} = 15$, $N = 140$ temperature data as used in the $L_2$ example above. As portrayed by its right skewed shape, the variance-norm MST RH properly shows the underdispersed relationship between the ensemble and verification. By homogenizing the variances of all
Figure 2.3: The $L_2$ MST RH (a) is nearly flat, even though four of the eight ensembles are highly underdispersed. However, since it homogenizes variances and thereby equally weights all dimensions, the variance norm MST RH (b) is underdispersed. The solid line represents the expected number of counts in each bin, $p$, given a perfectly flat histogram. Dotted lines represent a one standard deviation bound of this expectation, $(1/\sqrt{N})\sqrt{p(1-p)}$ (Smith and Hansen 2004). From Gombos and Hansen (2007).

dimensions, the variance norm equally weights all dimensions when computing MST distances and therefore is able to capture the extreme underdispersion of the low standard deviation elements (Gombos and Hansen 2007).

Although it averts the problems presented in the previous example, the variance norm is not ideal when the ensemble dimensions covary. Consider a two-dimensional linear cluster of highly correlated ensemble points and two hypothetical verification points
portrayed in figure 2.4a. The x-dimension has a standard deviation of 0.1 and the y-dimension has a standard deviation of 10. Verification point $B$ has a short Euclidean distance but a large statistical distance from the mean of the cluster of points measured in terms of standard deviation units of the associated two-dimensional PDF. Verification point $A$ has a relatively large Euclidean distance but a short statistical distance from the mean of the cluster of points compared to point $B$. Because it is farther than $A$ from the mean in terms of standard deviation units, point $B$ is significantly less statistically similar to the ensemble than is point $A$ (Gombos and Hansen 2007).

Figure 2.4: A comparison of the behavior of ensemble and verification points under the $L_2$, variance, and Mahalanobis norms. See text for details. From Gombos and Hansen (2007).
Figures 2.4a and 2.4b show how these ensemble and verification points behave under the $L_2$ and variance norms, respectively. Since it has undergone no transformations of variance or covariance, the $L_2$ norm simply reflects the case described above. The collection of points in the variance norm space portrays a similar structure as the $L_2$ norm, with the notable difference that the $x$ and $y$ dimensions have been homogenized. However, since it has not accounted for the covariance of the data, the variance norm improperly implies that point A is less similar to the ensemble than is point B, as seen by its greater Euclidean distance (in variance-normed space) from the ensemble PDF. Therefore, a variance-normed MST RH systematically using points similar to verification point A for all forecast occasions will be significantly less flat than one using points similar to point B, even though each B point is less likely to be a random draw from the same distribution that forms the respective ensemble (Gombos and Hansen 2007).

The Mahalanobis transformation is a conversion of the verification vector and ensemble vectors to the multivariate counterparts of the z-score,

$$z_{i,o} = \mathbf{C}_i^{-1/2} (\mathbf{o}_i - \bar{x}_i^*) \quad \text{and}$$

$$z_{i,j} = \mathbf{C}_i^{-1/2} (\mathbf{x}_{i,j}^* - \bar{x}_i^*) \quad \text{(Wilks 2004).}$$

Here $\bar{x}_i^*$ is the length $K$ vector whose entries are the averages of the columns of $\mathbf{X}_i^*$ (as defined above), $\mathbf{C}_i$ is the covariance matrix of $\mathbf{X}_i^*$, and

$$\mathbf{C}_i^{-1/2} = \mathbf{E}_i \mathbf{D}_i^{-1/2} \mathbf{E}_i^T,$$
where the columns of $E_i$ are the eigenvectors of $C_i$ and the entries of the diagonal matrix $D_i$ are the corresponding eigenvalues of $C_i$. Mahalanobis-normed MSTs are computed in the same way as are $L_2$-normed MSTs, except that $z_i$ is substituted, in turn, for one of the $z_{i,j}$, instead of the $L_2$ verification vector being substituted, in turn, for one of the $L_2$ ensemble vectors (Wilks 2004; Mardia et. al. 1979). Note that this forecast error covariance norm, $C^{-1/2}$, performs the same function as the analysis error covariance norm used to transform non-isotropic initial uncertainty into isotropic initial uncertainty in singular vector computations. In each case, this operation simply defines the mean of the multivariate distribution to be zero and the covariance to be the identity matrix.

The Mahalanobis transformation homogenizes the variances and decorrelates the points that form the MST by operating on the ensemble and verification points with the covariance matrix, thereby eliminating the problems associated with the $L_2$ and variance norms. This operation effectively alters the Euclidean distances (in Mahalanobis-normed space) so that they properly reflect the statistical “closeness” of points from the mean (Wilks 2004; Mardia et. al. 1979). Therefore, as depicted in figure 2.4c, the Mahalanobis transformation decreases the Euclidean distance (in Mahalanobis-normed space) of point $A$ from the mean of the cluster and increases the Euclidean distance (in Mahalanobis-normed space) of point $B$ from the mean of the cluster (Gombos and Hansen 2007).

Although it effectively accounts for covariance information, the Mahalanobis norm gives misleading results when $R-1 \leq K$, where $R$ is the number of samples used to compute the covariance. (In the case of the Mahalanobis-normed MST RH, $R$ was
previously defined to be $n_{\text{ens}} + 1$, the number of rows of $X_i^*$. This circumstance results in a symmetric configuration of the Mahalanobis-normed points in which every pair of points in the transformed space is separated by a distance of exactly $\sqrt{2(R-1)}$. These points form a perfect $R$-hedron, analogous to a two dimensional equilateral triangle or three dimensional tetrahedron. Figure 2.5 shows nine examples of a $K = 2$ and $R = 3$ set of random points with zero mean and unit variance. The x-marks and circles represent these same points under the $L_2$ and Mahalanobis norm, respectively. The line segments connecting the circles are shown to indicate the shapes of the triangles formed by the three points; they are not the MST, but they do indicate the problem encountered by the MST RH. For $R-1 \leq K$, all MST distances are exactly the same, rendering the Mahalanobis-normed MST RH useless. Note that the implication of $R-1 \leq K$ is that the ensemble is spanning a rank deficient space, and the Mahalanobis norm always chooses $R$-hedrons as the most efficient way to isotropically span that space (Gombos and Hansen 2007).

Figure 2.6 presents six Mahalanobis-normed MST RHs in which the verification and the ensembles are random draws from the same distribution. The only difference between the panels is the number of dimensions used to calculate the MST distances. Given that the ensembles and the verification are random draws from the same distribution, the MST RHs should be flat. However, when $n_{\text{ens}} \leq K$ (panels d, e, and f), the Mahalanobis-normed MST RHs appear to indicate an underdispersed ensemble. In reality, all MST distances are exactly the same (to within machine precision) and the verifying MST distance satisfies the condition for populating the leading bin. The fact that bins other than the leading bin are populated is an artifact of round-off error.
Figure 2.5: The three x-marks represent three randomly chosen points with zero mean and unit variance. Line segments connect the x-marks in order to show the shape of the triangle formed by the three points. These lines do not represent the MST. The three circles indicate these same points under the Mahalanobis norm. From Gombos and Hansen (2007).

Creating differences in MST distances at the level of machine precision. Note that when \( n_{ens} \leq K \), it is necessary to calculate \( C_i^{-1/2} \) using truncations of \( E_i \) and \( D_i \). The \( E_i \) is comprised of the \( n_{ens} \) columns corresponding to the nonzero eigenvalues of \( C_i \) and the entries of the diagonal \( n_{ens} \times n_{ens} \) matrix \( D_i \) are the nonzero eigenvalues of \( C_i \) (Wilks 2004).
Figure 2.6: Mahalanobis-normed MST RHs with $n_{\text{ens}} = 10$ and varying dimensions. Since both the ensemble and the verification are random draws from a random Gaussian distribution with zero mean and unit variance, all RHs should be flat. However, when $n_{\text{ens}} \leq K$, the Mahalanobis-normed MST RHs spuriously indicate an underdispersed ensemble. The solid line represents the expected number of counts in each bin, $p$, given a perfectly flat histogram. Dotted lines represent a one standard deviation bound of this expectation, $(1/\sqrt{N})\sqrt{p(1-p)}$ (Smith and Hansen 2004). From Gombos and Hansen (2007).

Researchers familiar with the ensemble-based data assimilation literature will likely be concerned about spurious correlations due to sampling errors. While sampling errors will certainly exist for small ensemble sizes, since each increment to a
Mahalanobis-normed MST RH uses a common covariance matrix, each of the $n_{\text{ens}} + 1$ MST distances are subject to the same errors. The sampling errors may increase the number of forecast occasions needed to discern that a Mahalanobis-normed MST RH is non-flat, but they will not make flat histograms appear non-flat (Gombos and Hansen 2007).

2.3 Comparing averaged univariate rank histograms to an MST rank histogram

Before the advent of the MST rank histogram, the reliability of multidimensional ensemble forecasts was measured by averaging univariate rank histograms. The ranks of the histograms of a Boston temperature forecast, for example, would be averaged with the ranks of the individual New York City and Philadelphia temperature rank histograms to holistically assess temperature forecasts for this entire area.

Although each independent univariate rank histogram (URH) properly measures the likeness of the ensemble and verification distributions, the averaged URH’s neglect of the covariance lends itself to misinterpretation. Consider the pathological example portrayed in figure 2.7, where $n_{\text{ens}} = 30$ and $K = 2$. The dots, solid line PDF, $x$s, and dashed line PDF in panel a respectively depict the ensemble members, Gaussian ensemble PDF, verification members, and Gaussian verification PDF. The mean and variance of the ensemble and verification distributions are identical and equal to 0 and 1, respectively, but the two distributions negatively covary. Since each dimension’s URH assesses the likeness of the mean and variance of the dimension’s ensemble and verification distributions, both the $x$ and $y$ dimension’s URH is appropriately flat, as seen
Comparing univariate RHs to MST RHs

a. Verification and ensemble distributions

b. URH for $x$ dimension

$$S_{0.04} \quad S_{0.03} \quad 0 \quad -0.01$$

$$-2 \quad 0 \quad 2$$

$$x \, \text{dimension (Standard deviation units)}$$

Ensemble Bin

$$0 \quad 10 \quad 20 \quad 30$$

c. URH for $y$ dimension

d. Averaged URH for $x$ and $y$ dimensions

e. Typical MST

$$m_{\text{ens}} = 30, K = 2, N = 1000$$

f. Mahalanobis MST

Ensemble Bin

$$0 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \quad 0.5$$

$$0 \quad 0.01 \quad 0.02 \quad 0.03 \quad 0.04$$

$$-2 \quad -1 \quad 0 \quad 1 \quad 2$$

$$x \, \text{dimension (Standard deviation units)}$$

Ensemble Bin

$$0 \quad 10 \quad 20 \quad 30$$

Figure 2.7: Panel a shows dots, a solid line PDF, $x$s, and a dashed line PDF that respectively depict the ensemble members, Gaussian ensemble PDF, verification members, and Gaussian verification PDF. The mean and variance of both of these distributions are 0 and 1, respectively. Panels b and c show the univariate rank histogram for the $x$ and $y$ dimensions, respectively. Panel d shows the rank histogram that averages the histogram in panels b and c. Panel e shows a randomly chosen ensemble-verification union MST from this distribution. The $x$ represents the random verification member and the dots represent the $n_{\text{ens}} - 1$ ensemble members. Panel f shows the Mahalanobis MST rank histogram.

in panels b and c. However, as depicted in panel d, despite the obvious dissimilarities between the two multidimensional distributions, the averaged URH is also flat, since it is simply the average of two flat URHs and is independent of the covariance. Clearly, a multidimensional reliability assessment tool requires the incorporation of the covariance of the dimensions to properly assess the true likeness.
The MST rank histogram fulfills this requirement. By incorporating all
dimensions into the same geometric space, the MST distance accounts for the
covariances. Panel e of figure 2.7 shows a typical ensemble-verification union MST for
the given distributions. Because, due to the difference in covariance, the verification
PDF is not aligned with the ensemble PDF, a typical verification point will be
geometrically distant from the ensemble cluster. The characteristically distant
verification points result in characteristically small ensemble-only MST distances that
cause MST rank histograms to properly depict differences in the distributions. Panel f
depicts an underdispersed MST rank histogram, affirming that the ensemble and
verification distributions are different despite their identical mean and variance.

2.4 Description of data

The following section applies the MST rank histogram to assess multidimensional
ensemble reliability using the National Centers for Environmental Prediction’s (NCEP)
Short-Range Ensemble Forecast (SREF) data sets. SREF consists of fifteen ensemble
members, five of which come from the ETA model with a BMJ convective scheme, five
of which comes from the same ETA model but with a Kain-Fritch convective scheme,
and five of which come from the Regional Spectral Model (RSM) with a SAS convective
scheme. All fifteen SREF ensembles are perturbed in their initial conditions. For a
description of this system, the reader is referred to Du et al. (2003).

Two separate SREF data sets are used to construct the MST rank histograms in
this work. Aiming to improve ensemble diversity and forecast spread, SREF physics
diversity was modified by increasing the number of convective schemes (from three to
six) and cloud microphysics parameterizations (Du et al. 2004; McQueen et al. 2005).

This upgrade, which occurred on 17 August 2004, also included an increase in the model resolution; the ten ETA members have 60 levels and a 32 km horizontal resolution and the five RSM members have 28 levels and a 40 km horizontal resolution. The first data set, which will be referred to as SREF1, is comprised of these upgraded ensemble forecasts from 18 August 2004 to 13 May 2005. The second set, which will be referred to as SREF2, is an older ensemble prediction system with slightly less ensemble diversity and decreased model resolution; the ten ETA members have 45 levels and a 48 km resolution and the five RSM members have 28 levels and a 48 km horizontal resolution. This set uses data from 3 June 2004 to 17 August 2004.

The SREF forecasts were verified using two separate verification data sets. The first verification was obtained by randomly selecting between the two ETA analysis controls and the one RSM analysis control. These controls were not averaged because averaging significantly reduced the variance of the verification. The second verification consisted of station observations obtained from the National Climate Data Center that undergo extensive automated quality control. Ensemble forecast values were linearly interpolated to the station locations.

2.5 Analysis of the multidimensional reliability of weather components

This section uses MST RHs to compare the multivariate reliabilities of forecast components for various cities clusters, forecast components, and lead times. For all of the following examples $K = 7$ (seven different cities) and $n_{ens} = 15$ (the ten ETA
forecasts and the five RSM forecasts). Because of unlike variances of the data in the different dimensions, significant covariance between dimensions, and $n_{ens} > K$, the Mahalanobis norm has been used to calculate MST distances. Note that for the cases considered, $L_2$-normed MST RHs give qualitatively similar results to the Mahalanobis-normed MST RHs, but the Mahalanobis-normed MST RHs are significantly less flat.

Mahalanobis-normed MST RHs were separately computed for two clusters of $K = 7$ cities. The first cluster is comprised of the northeastern United States cities shown in figure 2.8a: Boston, New York City, Philadelphia, Washington D.C., Albany, Portland (Maine), and Bangor; the second cluster is comprised of the southwest United States cities shown in figure 2.8b: San Diego, Los Angeles, San Francisco, Sacramento, Fresno, Reno, and Las Vegas. The northeastern and southwestern city clusters were chosen because they are the most populous regions in the United States. The reader should be aware that the SREF reliabilities depicted by the MST RHs in this chapter may be significantly different than the SREF reliabilities for other regions. Each MST RH uses all seven cities to separately assess the ensemble forecast reliability of one of four different weather components: mean sea level pressure ($P_{MSL}$), 2-m temperature ($T_{2m}$), 10-m wind speed ($u_{10m}$), and the temperature-humidity index ($THI$). SREF1 was used to compute the $P_{MSL}$, $T_{2m}$, and $u_{10m}$ MST RHs and SREF2 was used to compute the $THI$ MST RHs. SREF1 was not used to compute the $THI$ MST RHs because necessary dew point temperature information was not available as part of the SREF1 data set. Also note that the analysis $THI$ MST RHs were verified using the SREF2 analyses.
The choice of $P_{MSL}$, $T_{2m}$, and $u_{10m}$ was motivated by their obvious importance in typical weather forecasts. The temperature-humidity index, as defined by

$$THI(^\circ F) = 0.55 \times T_{2m} (^\circ F) + 0.2 \times T_{d_2} (^\circ F) + 17.5 \quad \text{(Glickman 2000)}, \quad (2.7)$$

where $T_{d_2}$ is the 2-m dew point temperature, was chosen because of its importance in energy markets. This index is an indicator of the sultriness due to the combined effects of temperature and humidity. Therefore, the accurate prediction of the $THI$ is crucial for markets that are sensitive to supply and demand fluctuations induced by air conditioner energy usage. Because energy companies are particularly interested in regional forecasts, a multivariate reliability assessment of the $THI$ is especially important. As it is an indicator of sultriness, the $THI$ RHs were computed using only summer data.

Following Stensrud (1996), seven day running mean biases have been removed from all MST RHs in this section. Define $x^*_{i,j,k}$ to be
\[ x_{i,j,k}^* = x_{i,j,k} - \frac{1}{7n_{ens}} \sum_{m=1}^{7} \sum_{j=1}^{n_{ens}} (x_{i-m,j,k} - o_{i-m,k}), \tag{2.8} \]

where \( o_{i,k} \) is an individual verification data point in the \( k \)th dimension. This transformation simply subtracts the average bias of each dimension, for the seven day period prior to the \( i \)th day, from each ensemble data point of the corresponding dimension on the \( i \)th day. All MST RHs in the application sections of this chapter have been computed using the de-biased \( x_{i,j,k}^* \) points. The biases reported in Tables 1-4 are the averages of these seven day running mean biases for each city and weather component.

Figures 2.9-2.12 show Mahalanobis-normed MST histograms for 24 hour forecasts valid at 09 UTC, the associated CvM test statistic, and a histogram flatness assessment at the 1% significance level. The verification for each increment in figures 2.9 and 2.11 is a random selection of one of the three SREF control analyses for the corresponding day; the verification for each increment in figures 2.10 and 2.12 is the actual observation for the corresponding day, with forecast values interpolated to the location of the observing station. Figures 2.9 and 2.10 are for the northeast cluster of cities and figures 2.11 and 2.12 are for the southwest cluster. Note that the number of counts in each bin has been divided by \( N = 81 \) for the \( P_{MSL} \), \( T_{2m} \), and \( u_{10m} \) histograms and by \( N = 21 \) for the THI histogram to yield relative frequency histograms. The solid line represents the expected number of counts in each bin, \( p \), given a perfectly flat histogram. In order to give an indication of the effects of the small sample size on the flatness, dotted lines representing a one standard deviation bound of this expectation, \( (1/\sqrt{N})\sqrt{p(1-p)} \), have also been included (Smith and Hansen 2004). Also note that, because the proper interpretation of an MST RH requires that each increment be
statistically independent of others, the MST RHs are constructed using data from every third day, the lag at which bin population autocorrelations were found to be relatively negligible (Gombos and Hansen 2007).

Regardless of the verification type, city cluster location, or weather component, multidimensional SREF forecasts are underdispersed, as indicated by the right skewed MST RHs of figures 2.9-2.12. Despite recent attempts by NCEP to increase ensemble diversity, short range ensembles members lack sufficient differences to capture the PDF of the verification. Further initial condition, physics, and/or parameterization diversifications are needed. Although all RHs are right skewed, the degree of underdispersion depends on the choice of the verification, city cluster location, and weather component. Figures 2.9 and 2.10 indicate that forecasts for the THI are the most reliable (or more accurately, the least unreliable) for 24 hour lead times in the northeast cluster, followed by $P_{MST}$, $T_{2m}$, and $u_{10m}$. This bodes well for those that rely on THI forecasts in the energy markets. Note, however, that the small sample sizes for these and all MST RHs in this section limits the significance of the differences between the CvM statistics. Although it is clear that the forecasts for these weather components are underdispersed, the relative reliability may change with increased sample sizes (Gombos and Hansen 2007).

Differences between figures 2.9 and 2.10 can be attributed to differences between and limitations of the two choices of verification. Using observations as the verification introduces representativeness errors that reflect the fact that the observations resolve scales that the model does not; a simple interpolation of low resolution forecast fields to an observation station location is a particularly crude form of downscaling. Geographic
Figure 2.9: Mahalanobis-normed and de-biased MST RHs for the northeast cluster for 24 hour lead time valid at 09 UTC. The verification for each increment is taken as a random selection of one of the three SREF control analyses. The dotted lines represent a one standard deviation bound on this expectation. CvM statistics are also included, as well as an assessment of flatness at the 1% significance level. A rejection implies that the histogram is not flat, whereas an acceptance indicates that the histogram is flat. From Gombos and Hansen (2007).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{MSS}(mb)$</td>
<td>-0.63</td>
<td>-0.39</td>
<td>-0.51</td>
<td>-0.27</td>
<td>-0.50</td>
<td>-0.63</td>
<td>-0.76</td>
</tr>
<tr>
<td>$T_{2m}(C)$</td>
<td>-0.08</td>
<td>-0.41</td>
<td>-0.37</td>
<td>-0.81</td>
<td>-0.77</td>
<td>-0.26</td>
<td>-0.19</td>
</tr>
<tr>
<td>$u_{10m}(ms^{-1})$</td>
<td>2.46</td>
<td>3.30</td>
<td>2.46</td>
<td>3.32</td>
<td>2.53</td>
<td>2.49</td>
<td>2.80</td>
</tr>
<tr>
<td>$THI(°F)$</td>
<td>0.64</td>
<td>-0.59</td>
<td>-0.50</td>
<td>-1.03</td>
<td>-0.62</td>
<td>-0.30</td>
<td>-0.33</td>
</tr>
</tbody>
</table>

Table 2.1: Averaged seven day running biases for northeast cluster cities from figure 2.9. From Gombos and Hansen (2007).
Figure 2.10: Same as figure 2.9, except observations are used as the verification.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{MSL}(mb)$</td>
<td>-0.59</td>
<td>-0.52</td>
<td>-0.81</td>
<td>-0.33</td>
<td>-0.17</td>
<td>-0.60</td>
<td>-0.64</td>
</tr>
<tr>
<td>$T_{2m}(C)$</td>
<td>-0.46</td>
<td>0.54</td>
<td>-1.30</td>
<td>-2.13</td>
<td>-2.93</td>
<td>-1.86</td>
<td>-2.20</td>
</tr>
<tr>
<td>$u_{10m}(ms^{-1})$</td>
<td>-1.95</td>
<td>-0.92</td>
<td>-1.67</td>
<td>-4.52</td>
<td>-4.64</td>
<td>-3.83</td>
<td>-2.19</td>
</tr>
<tr>
<td>$THI(F)$</td>
<td>-0.82</td>
<td>-0.33</td>
<td>-1.95</td>
<td>-2.18</td>
<td>-3.38</td>
<td>-1.61</td>
<td>-2.61</td>
</tr>
</tbody>
</table>

Table 2.2: Same as Table 2.1, except observations are used as the verification.

areas that tend to generate steep gradients in the forecast component are particularly prone to such representativeness errors. A more fair comparison would be to compare station observations with forecast values that have been mapped from model space into
observation space via model output statistics (MOS) or some other form of calibration (Gombos and Hansen 2007). Such a comparison is beyond the scope of this work.

Although they mitigate representativeness errors, RHs using the analysis as the verification are subject to forecast dependence errors. Because the control analysis is a weighted combination of a short term forecast and observations, the analysis and forecast are incestuously dependent, especially at short lead times. For example, by construction, the ensemble control analysis is not an outlier of the ensemble forecast at analysis time; all other ensemble members are perturbations around this control analysis. Therefore, a sufficient lead time is required to ensure that the ensemble can evolve such that the verifying analysis is different from the median of the forecast ensemble (Saetra, et. al. 2004). Additionally, because analyses are in model space, not observation space, one expects model forecasts to be in some sense “closer” to analyses than to observations, which lie in a completely different space (Gombos and Hansen 2007).

Because of the incestuous relationship between ensemble forecasts and the analysis at short lead times, the 24 hour northeast cluster analysis RHs of figure 2.9 are susceptible to forecast dependence errors. However, the observation RHs of figure 2.10 of weather components that can support relatively steep spatial gradients, such as $T_{2m}$, $THI$, and particularly $u_{10m}$, are highly prone to representativeness errors. Therefore, it is difficult to determine which figures’ histograms measure the ensemble reliability most accurately. It is the view of the author that verification in observation space is preferred. Note, however, that the relative similarities of the two $P_{MSL}$ histograms (figures 2.9a and 2.10a), which are not prone to high representativeness errors, may indicate that
representativeness errors have a larger impact than dependence errors, even at 24 hour lead times. Again, the author reiterates that the preferred method of forecast assessment is to project forecast ensemble members into observation space using some form of
Figure 2.12: Same as figure 2.10, except for the southwest cluster. From Gombos and Hansen (2007).

Table 2.4: Same as Table 2.2, except for the southwest cluster. From Gombos and Hansen (2007).

As can be seen from figures 2.11 and 2.12, the reliabilities of weather components in the southwest cluster at 24 hour lead times also depend on the type of verification.
THI reliability is consistently poor, whereas the poorness of the $T_{2m}$, $u_{10m}$, and $P_{MSL}$ reliabilities differs. Because of the greater topographic changes in the southwest than in the northeast, we speculate that representativeness errors are more influential in the southwest cluster RHs than in the northeast cluster RHs. Topographic channeling effects and valley inversion layers induce high mesoscale wind speed and temperature variability. Because mesoscale $P_{MSL}$ gradients are primarily thermally driven in this region (Zhong et al. 2004), even $P_{MSL}$ histograms are prone to representativeness errors. Therefore, to a greater extent than for the northeast cluster RHs, the southwest analysis RHs are likely to be more accurate than the southwest observation RHs (Gombos and Hansen 2007).

Comparing figure 2.11 with figures 2.9 and 2.10, forecast reliability is generally worse in the southwest than in the northeast. This is especially true for the THI, which is particularly foreboding considering the heavy air conditioner usage in this region. Note, however, that the observation RHs of these two clusters are extremely similar, other than that of the THI (Gombos and Hansen 2007).

2.6 Conclusions

The MST RH is an effective multidimensional ensemble reliability assessment tool. After eliminating biases, spatial and temporal correlations, and variance inconsistencies among the $K$ dimensions, the shape of an MST RH can be used to diagnose the relationship between the distribution of the ensemble and of the verification. This information can ultimately help improve forecast reliability through the modification
of the ensemble prediction system and can be used to assess whether forecast ensembles used for ensemble regression accurately sample the forecast uncertainty space.

The Mahalanobis norm transforms the forecast data in the most meaningful and interpretable way when the number of ensemble members is greater than the number of forecast locations and/or weather components; this chapter advocates the use of the Mahalanobis norm under this circumstance. However, given the misleading results when $n_{ens} \leq K$, it is suggested that the variance norm be used when $n_{ens} \leq K$ and the variances in all dimensions are not identical. The $L_2$ norm should only be used when the covariance matrix is a scalar multiple of the identity matrix.

Although results are somewhat obscured by verification errors, the analysis of Mahalanobis-normed MST RHs has revealed several important characteristics of the SREF ensemble forecast system. For the components and city clusters analyzed, the right skewed RHs imply that SREF ensembles are underdispersed at a 24 hour lead time. For the northeast cluster, $THI$ forecasts are the least underdispersed, followed by $P_{MSL}$, $T_{2m}$, and $u_{10m}$ forecasts. Depending on the type of verification used, the most reliable weather component forecasts in the southwest cluster are for $T_{2m}$, followed by $P_{MSL}$, $u_{10m}$, and the $THI$. Reliability in the northeast cluster is generally greater than southwest cluster reliability, especially when the analysis is used as the verification.

It is important to note that absolute uniformity of a reliable RH requires that initial ensemble distributions are correct, and that ensembles be evolved under a perfect forecast model. Since no models of the atmosphere are perfect, RH interpreters must realize that both model error and initial distribution error will impact the histograms (Smith and Hansen 2004), and that it is not clear how to disentangle these two types of inadequacies.
This chapter has presented some preliminary applications of the MST RH. Subsequent studies of the detailed effects of imperfect model scenarios, variable sample sizes, ensemble sizes, dimension sizes, and norm definitions, among others, are needed. Given the multidimensionality of the atmosphere and the need to jointly assess the reliability of these dimensions, the MST RH will hopefully evolve into a standard ensemble reliability assessment tool that is available to all ensemble forecasting practitioners and researchers trying to assess the goodness of forecast ensembles used for ensemble regression.
Chapter 3

Ensemble Regression

3.1 Ensemble Synoptic Analysis (ESA)

As discussed in section 1.2, statistics deduced from climatological time series are typically assumed to be stationary, making field covariability techniques that employ these statistics (e.g. LIM, CCA, and the FDT) inappropriate for the modeling of specific flow-dependent synoptic-scale atmospheric events. On the other hand, as discussed in section 1.4, statistics deduced from ensembles derived from state estimations that employ evolving covariance estimates (e.g. EnKF; EKF; 4D-Var) capture these “errors of the day” typically attributable to baroclinic instabilities. Therefore, the use of ensemble statistics rather than time series statistics enables the progress reaped by climatologists from field covariability techniques to be shared by synopticians.

Hakim and Torn (2008, hereafter HT) coined Ensemble Synoptic Analysis (ESA) to describe the method of statistically inferring synoptic dynamical relationships from the covariability of fields’ ensemble analyses and forecasts. By treating individual ensemble members as independent samples, ESA employs standard statistical techniques to detect sensitivities, infer dynamical couplings, and aid forecasters in identifying dynamical processes that are particularly relevant for specific weather predictions.
Ensemble sensitivity is the tool most commonly used in the pioneering papers on 
ESA. Hakim and Torn (2008) defined ensemble sensitivity, which measures the 
univariate sensitivity of a member of a state vector to any univariate variable, as 
\[
\frac{\partial J_e}{\partial x} = \frac{\text{cov}(J_e, x)}{\text{var}(x)},
\]
(3.1)

where \(J\) and \(x\) are \(1 \times n_{\text{ens}}\) ensemble anomaly (e.g. anomaly with respect to the ensemble 
mean) estimates of a forecast metric and state variable, respectively, and \(\text{var}\) denotes the 
variance of the parenthesized argument. Ensemble sensitivity has been used to unveil 
otherwise obscured linkages between a midlatitude cyclone and the subtropical jetstream 
(Hakim and Torn 2008), to predict the impact of observations on sea level pressure and 
precipitation forecast metrics (Torn and Hakim 2008), and to define optimal 
climatological (Torn and Hakim 2008) and adaptive (Ancell and Hakim 2007) observing 
sites. It has also been shown to be proportional to the projection of the analysis-error 
covariance onto the adjoint-sensitivity fields, such that 
\[
\frac{\partial J_e}{\partial x_0} = D^{-1} A \frac{\partial J}{\partial x_0}, \quad \text{(Ancell and Hakim 2007)}
\]
(3.2)

where \(A\) is the ensemble estimated analysis error covariance, \(D\) is a diagonal matrix with 
initial-time error variance, and \(\frac{\partial J}{\partial x_0}\) is the adjoint sensitivity (Ancell and Hakim 2007).

Hakim and Torn (2008) also presented a multivariate extension to ensemble 
sensitivity in a general proof-of-concept manner in order to explore the feasibility of 
statistical potential vorticity inversion. They introduced a multivariate operator 
computed from matrices of state and potential vorticity ensemble anomalies that can be 
used to find the state perturbation estimate statistically associated with any potential
vorticity perturbation. Statistical potential vorticity inversion has since been further
developed by Gombos and Hansen (2008) and Hakim (2008) and will be explored in
section 5 of this thesis. This operator laid the groundwork for Ensemble Regression
(ER), which is the focus of this thesis.

3.2 Defining Ensemble Regression (ER)

This thesis focuses on a particular ESA technique called Ensemble Regression
(ER; Gombos and Hansen 2008, hereafter GH08). ER facilitates inference about the
relationship between two multidimensional atmospheric fields $P$ and $Y$ via the regression
of a perturbation using an operator defined by the fields’ ensemble forecasts and/or
analyses. Let $Y_e$ and $P_e$ be $K \times n_{ens}$ and $k \times n_{ens}$ ensemble anomaly matrices,
respectively, where each of the $K$ or $k$ rows corresponds to a point on a vectorized grid,
each of the $n_{ens}$ columns corresponds to an individual ensemble member, and the prime
denotes an anomaly with respect to the ensemble mean. Let $Y_e$ and $P_e$ be related as

$$ Y_e' = LP_e' \quad (3.3) $$

and let $L$ be a linear operator that maps the ensemble anomalies of $P$ into ensemble
anomalies of $Y$ such that

$$ L = Y_e^{-1} P_e'. \quad (3.4) $$

(GH08; Hakim and Torn 2008). Assuming a sufficiently large ensemble and a linear
relationship between $Y$ and $P$, given any perturbation, $p'$, $L$ is a Green’s function that
approximates the $K \times 1$ ensemble anomaly, $\hat{y}'$, with which $\bar{p}'$ is statistically associated
via
\[
\hat{y}' = Lp'.
\]  
(3.5)

Here,

\[
\hat{p}' = P\left(P^TP\right)P^Tp'.
\]  
(3.6)

is the effective perturbation that is resolved by \( L \) and effectively regressed via 3.5 (GH08). The notion of the effective perturbation will be discussed in section 3.4.

Symbols used to define ER are listed in Table 3.1; the meaning of some symbols will be made more clear in chapter 5.

ER uses the \( n_{en} \) known estimates of the predictor and predictand fields in the form of ensemble model output analyses and forecasts to train an operator \( L \) that maps linear combinations of the predictor ensemble anomalies to linear combinations of the predictand ensemble anomalies. That is, based on ensemble anomaly examples of how the predictand anomaly field \( Y_e \) is typically configured when the configuration of the predictor anomaly field \( P_e \) is known, \( L \) is trained to predict the most probable state of the predictand field, \( \hat{y}' \), given any resolved perturbation of the predictor field, \( \hat{p}' \), in cases when the actual predictand state is not explicitly known.

The operator \( L \) can be alternatively interpreted as a perturbation mapping function based on the covariances between the predictor and predictand fields. Right multiplying 3.3 by \( P_e^T \) and rearranging yields

\[
L = \text{cov}(P_e', Y_e')\text{cov}(P_e', P_e')^{-1} \quad \text{(Hakim and Torn 2008),}
\]  
(3.7)

which shows that ER operator is simply a function of the cross-covariance of the
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Ensemble predictand matrix</td>
</tr>
<tr>
<td>P</td>
<td>Ensemble predictor matrix</td>
</tr>
<tr>
<td>Y_e'</td>
<td>Ensemble predictand perturbation matrix, with respect to the ensemble mean</td>
</tr>
<tr>
<td>P_e'</td>
<td>Ensemble predictor perturbation matrix, with respect to the ensemble mean</td>
</tr>
<tr>
<td>p'</td>
<td>Prescribed predictor perturbation</td>
</tr>
<tr>
<td>L</td>
<td>Linear ensemble regression operator</td>
</tr>
<tr>
<td>n_ens</td>
<td>Number of ensemble members</td>
</tr>
<tr>
<td>K</td>
<td>Number of grid points for each ensemble member of Y</td>
</tr>
<tr>
<td>k</td>
<td>Number of grid points for each ensemble member of P</td>
</tr>
<tr>
<td>w_t</td>
<td>Number of retained singular values and singular vectors of P_e'</td>
</tr>
<tr>
<td>n_and_p</td>
<td>Number of retained predictand principal components</td>
</tr>
<tr>
<td>n_or_p</td>
<td>Number of retained predictor principal components</td>
</tr>
<tr>
<td>p̃</td>
<td>p', except all entries not being inverted (regressed) are set to zero</td>
</tr>
<tr>
<td>p̂</td>
<td>Effective predictor perturbation comprised of perturbations associated with p̃</td>
</tr>
<tr>
<td>ŷ</td>
<td>Estimated predictand perturbation associated with p̂</td>
</tr>
<tr>
<td>ŷ̃</td>
<td>Ensemble predictand anomaly estimate associated with p̃</td>
</tr>
<tr>
<td>U_Y</td>
<td>Leading principal components of Y_e'</td>
</tr>
<tr>
<td>U_P</td>
<td>Leading principal components of P_e'</td>
</tr>
</tbody>
</table>

Table 3.1: A symbol definitions table for symbols used in ensemble regression.
ensemble predictor and predictand anomalies, normalized by the inverse covariance of the predictor anomalies. This concept is schematically portrayed using contrived data in figure 3.1. The small dots in the panel in the lower right (upper left) of the figure depict $n_{\text{ens}} = 50$ ensemble anomaly estimates of an arbitrary two-dimensional predictor (predictand) field. The ellipses respectively represent the 0.5, 1, and 2 standard deviations of the covariance matrix of the ensemble predictor (predictand) distribution. The large center panel depicts 0.5, 1, and 2 standard deviations ellipses defined by the $2 \times 2$ covariance operator $L$. The large black dot in the lower-right panel represents an arbitrary predictor perturbation, $p'$, which was not one of the ensemble perturbations used to define $L$. The operator $L$ can be employed to estimate the most probable predictand perturbation, $\hat{y}'$, given $p'$ and can be interpreted as a transfer function between the two fields. Schematically, the covariance transfer function interpretation of $L$ is depicted by the dashed lines that guide the eye between the predictor perturbation’s location in the predictor distribution to the covariance function $L$ and then, based on the ensemble training samples of the two fields that define $L$, to the most probable position in the predictand distribution, $\hat{y}'$, given $p'$. 

ER can also be considered a multivariate extension to the univariate point-correlation ESA ensemble sensitivity techniques (viz. 3.1) used in Hakim and Torn (2008), Torn and Hakim (2008), and Ancell and Hakim (2007). Whereas point-correlations compute field sensitivities by independently and iteratively computing ensemble correlations between individual elements of the state vector, ER computes field sensitivities by computing covariance-based regression operators between significant portions of the (or the entire) state vector. The use of a multidimensional operator
enables inferences of how entire fields jointly relate, rather than just of how scalars individually relate. This comparison will be further explored in section 6.7.

### 3.3 Truncated SVD and principal component subspaces

In realistic applications, $K$ is typically much greater than $n_{ens}$, making $P_e^{-1}$ rank deficient. Therefore, the inverse of $P_e$ in 3.4 is technically a pseudoinverse and can be calculated using the singular value decomposition (SVD). Taking the SVD of $P_e$ and retaining the $n_{ens}$ nonzero singular values, $P_e^{-1}$ can be expressed as
\[ P_e^{-1} = VW^{-1}U^T, \] (3.8)

where \( V \) denotes the matrix composed of the right singular vectors, \( U \) the matrix composed of the left singular vectors, and \( W \) a diagonal matrix composed of the singular values, \( w \) (e.g. Hakim and Torn 2008).

However, despite the computational stability of the SVD, multicollinearities due to geographic proximity are likely to render the estimate of \( P_e^{-1} \) via 3.8 ill-conditioned. Singular value spectra of the data used in the following sections of this thesis (not shown) reveal that the retained singular values of \( P_e^T P_e \) (and undoubtedly those of other realistic \( P_e^T P_e \) matrices) decay to zero, leading to high condition numbers and correspondingly inflated regression parameter variances. This ill-conditionedness implies that the regression of predictor perturbations other than those that define \( P_e \) may yield inaccurate results, necessitating the regularization of \( P_e^{-1} \).

One way to potentially alleviate this ill-conditionedness is to employ the truncated SVD (TSVD; Golub and van Loan 1996) in lieu of 3.8. The TSVD, which is simply an SVD performed with all but the largest \( w \) values of \( w \) set to zero, stabilizes the inversion by eliminating the small singular values that inflate the regression parameter variances (Golub and van Loan 1996).

Another way to alleviate problems relating to multicollinearity and inflated regression parameter variances is to first project the predictor and predictand ensembles onto their first \( n_{pc}^o \) and \( n_{pc}^a \) principal components, \( U_p \) and \( U_y \), respectively, before computing the ER operator (viz. 3.4). Here,
\[ U_p = EP'_e, \] (3.9)

and

\[ U_y = FY'_e, \] (3.10)

where \( E \) and \( F \) are the eigenvectors of the covariances of \( P'_e \) and \( Y'_e \), respectively (e.g. Barnett and Preisendorfer 1987; Wilks 2006). After also projecting the predictor perturbation, \( p' \), onto this subspace via

\[ u_p = E p'. \] (3.11)

the ER proceeds as usual (viz. 3.4-3.5), except \( U_p, U_y, \) and \( u_p \) are substituted for \( P'_e, Y'_e, \) and \( p' \), respectively. At the conclusion of the ER, the estimated predictand perturbation, \( u_y \), must be projected back into physical space from the principal component subspace via

\[ \hat{y}' = Fu_y \] (3.12)

for plotting and analysis purposes.

Because multicollinearities are typically associated with principal components with the smallest eigenvalues (Jolliffe 2002), performing the regression in the subspace of the leading principal component reduces multicollinearity problems that can lead to ill-conditioned singular covariance estimates, inflated regression parameter estimates, and poor out-of-sample regression results. In physical space, it is difficult to isolate predictors plagued by multicollinearities, but, because eigenvectors are defined to be orthogonal and thus independent, the eigenvector subspace enables the removal of problematic principal components without affecting others (e.g. Wilks 2006). This
practice, however, comes at the expense of potentially discarding principal components with valuable predictive information.

In addition to regularizing the regression, another extremely important advantage of performing ER in the subspace of principal components is it makes computations more tractable; the computing power required to multiply and invert large matrices exceeds the limitations of many computing systems at the time of this writing. The projections reduces the dimensionality of the predictor ensemble, predictand ensemble, predictor perturbation, and regression operator to $n_{or} \times n_{ens}$, $n_{pc} \times n_{ens}$, $n_{or} \times 1$, and $n_{or} \times n_{pc}$, respectively, dramatically increasing the speed and the computations.

The determination of the optimal values of $w_r$, $n_{or}^{pc}$, and, $n_{pc}^{and}$ is a subjective tradeoff between numerical stability and the conservation of the degrees of freedom of $P_e'$ and $Y_e'$; more severe truncations result in lower regression parameter variance, but also in estimates of $L$ that lack information about the truncated low variance singular vectors and principal components of the predictors and predictands. The literature suggests that the optimal value of $w_r$ or $n_{or}^{pc}$ is $w_r = r$, where $r$ is the “numerical rank” of $P_e'$ (Hansen 1998). As a general rule-of-thumb, $w_l$ and $n_{pc}^{or}$ ($n_{pc}^{and}$) can be estimated as the number of singular values that corresponds to a marked flattening of the singular value spectrum of $P_e^T P_e$ ($Y_e^T Y_e$) (Hansen 1998; Wilks 2006). The singular vectors that correspond with singular values smaller than $w_l$ or $n_{pc}^{or}$ ($n_{pc}^{and}$) contribute to the destabilization of $P_e^{-1}$ ($Y_e^{-1}$), but insignificantly to the variance of $P_e'$ ($Y_e'$). The optimal values can also be estimated via a leave-one-out cross-validation technique, as explained in section 4.4.
3.4 Perturbation dynamical representativity and resolvability

Although ER can be used to determine predictand state anomalies associated with any predictor perturbation (as will be illustrated in the sections 6.2 and 6.5), significant inference can be gained only via the regression of meaningful perturbations of $P$.

There are two fundamental characteristics that deem a perturbation meaningful in the context of ER. The first characteristic, dynamical representativity, refers to how well the regressed perturbation describes an underlying dynamical state or mechanism; dynamical inferences from ER are possible only if the regressed perturbation represents a dynamically meaningful state. Examples of perturbations with potential dynamical representativity are a time-mean normed perturbation, which could be used to represent and model synoptic-scale anomaly storm systems (e.g. Davis and Emanuel 1991), and an outlying ensemble anomaly (with respect to the ensemble mean), which could facilitate understanding of the causes of anomalous forecasts. Predictor principal components may also have significant dynamical representativity because highly variable directions in state space often correspond to uncertainties in state dynamics that require understanding, as will be further explained later. Note that when using ER for prediction (viz. chapter 4), rather than dynamical inference, dynamical representativity is unnecessary.

Resolvability, the second fundamental characteristic of a meaningful ER perturbation, describes the correspondence between the perturbation, $p'$, one attempts to regress to the perturbation, $\tilde{p}'$, that is effectively regressed. Because the regression operator, $L$, is defined in terms of $Y_e$ and $P_e'$, the ER machinery is only capable of
yielding the predictand states associated with perturbations that are linear combinations of the ensemble members of $\mathbf{P}_e'$; predictor perturbations not spanned by $\mathbf{P}_e'$ are imperfectly resolved as the least squares estimate of $\mathbf{P}_e'$. Therefore, although any perturbation can be prescribed to ER, the researcher must analyze how similar this perturbation is to the effectively regressed least squares estimate of the perturbation before interpreting the associated predictand state (Gombos and Hansen 2008).

Resolvability can be better understood by taking a closer look at 3.5. Because 3.5 states that $\hat{y}'$ is a linear combination of the columns of $\mathbf{L}$, with weights given by $\mathbf{p}'$, $\hat{y}'$ is inexact if the true solution is linearly independent of the column-space of $\mathbf{L}$. Therefore, $\hat{y}'$ is not the state perturbation statistically associated with $\mathbf{p}'$, but is instead the perturbation associated with the effective PV perturbation, $\tilde{\mathbf{p}}'$, that is defined in terms of $\mathbf{L}$, or equivalently, is a linear combination of $\mathbf{P}_e'$. That is, because $\mathbf{L}$ is defined in terms of $\mathbf{P}_e'$, any $\mathbf{p}'$ is re-expressed in terms of $\mathbf{P}_e$ when it is regressed via 3.5. The resolved perturbation, $\tilde{\mathbf{p}}'$, that is effectively regressed via 3.5 is the least squares projection of $\mathbf{p}'$ onto $\mathbf{P}_e$ (viz. 3.6).

Note that, for realistic ensembles with imperfect correlations, $\tilde{\mathbf{p}}' = \mathbf{p}'$ if there exists a linear combination of $\mathbf{P}_e'$ that can perfectly resolve the desired perturbation. This will occur if $\mathbf{p}'$ is not dynamically coupled with other predictor perturbations, or, equivalently, if enough ensemble members of $\mathbf{P}_e'$ reflect that the ensemble perturbations at the location of $\mathbf{p}'$ are uncorrelated with the ensemble perturbations elsewhere. Therefore, given uncoupled model dynamics and a sufficiently large ensemble that spans...
the subspace of the applied perturbation, the least squares estimate of the effective perturbation will equal the directly applied desired perturbation. However, if \( p' \) is dynamically coupled with other perturbations or if the ensemble is too small to infer that it is not, ER will regress the effective perturbation, not the directly applied desired perturbation (Gombos and Hansen 2008). The significances of perturbation dynamical representativity and resolvability will be made clearer in chapters 4-6.

3.5 Chapter summary

Ensemble Synoptic Analysis (ESA) is a method of statistically inferring synoptic dynamical relationships from the covariability of fields’ ensemble analyses and forecasts. This section defines and discusses fundamental characteristics of Ensemble Regression (ER), the particular type of ESA that is the focus of this thesis.

ER facilitates inference about the relationship between two multidimensional atmospheric fields \( P \) and \( Y \) via the regression of a perturbation using an operator defined by the fields’ ensemble forecasts and/or analyses. Matrices of ensemble anomalies of the predictor field \( P \) and the predictand field \( Y \) are used to train an ER regression operator, \( L \), which is simply a function of the covariances of these fields (viz. 3.7). Note that, in order to maximize numerical stability, the leading singular vectors of \( P \) can be used for the regression or both \( P \) and \( Y \) can be projected onto the subspace of their leading principal components. \( L \) can be used to predict the most probable state of the predictand field, \( \hat{y}' \), given a perturbation of the predictor field, \( p' \). The perturbation effectively regressed,
however, is not the prescribed perturbation, $\mathbf{p}'$, but is instead $\mathbf{\tilde{p}}'$, the perturbation resolved by $\mathbf{L}$ as the least squares estimate of $\mathbf{p}'$.

By applying ER to low-order Lorenz models and sophisticated operational atmospheric model data, the following section further develops the notion of the effective perturbation and illustrates fundamental properties of ER such as sensitivities to ensemble size, sampling error, and principal component truncations.
Chapter 4

Ensemble Regression Prediction Error

4.1 Introduction

ER is fundamentally a prediction technique; at its core, ER simply estimates the most probable predictand state given a predictor perturbation. Not only can ER be explicitly applied for forecasting (e.g. preemptive forecasting, as discussed in section 6.6), but its dynamical applications also implicitly require skillful predictand estimations. For example, because ER’s use as a synoptic dynamics and field sensitivity analysis technique hinges on the nature of the anomaly patterns of the regressed and resultant perturbations (see sections 6.2 and 6.5), making meaningful dynamical inferences from ER requires that the resultant predictand perturbation accurately corresponds to the true predictand state with which the effectively regressed predictor perturbation is associated. Therefore, it is necessary to prove ER a proficient prediction method in order to justify its use for forecasting and synoptic analysis.

The goodness of ER forecasts depend primarily on four factors: 1) the strength of the direct (i.e. independent of other mechanisms or variables) or indirect (i.e. mutually dependent on a mechanism not explicitly included in the ER model) physical relationship between the predictor and predictand, 2) the model’s ability to capture this physical relationship, 3) the sufficiency of the ensemble size and ability of the ensemble to model
the linear dynamics to a desired statistical confidence, and 4) the degree of non-linearity in the system dynamics that illegitimize the use of linear modeling techniques. The Lorenz (1963) model (hereafter L63) and Lorenz ’95 model (Lorenz and Emanuel 1998) are employed here because they are deterministic (removing the affect of the first factor on ER forecast goodness, given that the entire state is integrated), perfect (eliminating the affect of the second factor), and simple (enabling computational tractability and facilitating interpretation), and therefore disentangle the effects of non-linearity and ensemble size from model imperfections and predictor-predictand selection on ER forecast goodness. L63 and L95 are used in sections 4.2 and 4.3, respectively, to illustrate properties of ER forecasts such as error magnitudes, error comparisons with standard linear forecasting techniques, and ensemble size and sampling error sensitivities. In section 4.4, ER is applied to high-dimensional operational data to display an example of ER prediction skill for real atmospheric applications and to analyze singular vector truncation sensitivities.

4.2 Lorenz ’63 ER forecast skill and tangent linear model comparison

ER skill in the L63 system is explored in the following manner. The system equations (viz. 1.16-1.18, where \( a = 10 \), \( b = 8/3 \), and \( r = 28 \) to ensure chaotic dynamics), are integrated for 5000 iterations to force the initial condition onto the system attractor. Then, a \( n_{ens} = 50 \) ensemble is created by randomly perturbing the final spin-up state (the control state) and integrating the entire ensemble for 500 steps, while ensemble square-root Kalman filtering (Whitaker and Hamill 2002) the ensemble at each step using
the control state as the observations. The timestep used for the integration is 0.01. Given that the L63 error doubling time (i.e. the time after the size of at least 50% of all initial perturbations doubles) is approximately 0.8 Lorenz time units (e.g. Harlim 2007) and that the doubling time in the real atmosphere is approximately 2.5 days (Lorenz 1969), about 32 L63 timesteps approximates one day.

The L63 experiment proceeds as a variation of leave-one-out cross validation (e.g. Wilks 2006), a tool that enables ER forecast skill evaluation by comparing the ER-computed predictand perturbation to the actual predictand perturbation. Figure 4.1 schematically describes the steps of this leave-one-out cross validation variation, starting from the end of the aforementioned ensemble initialization process. The ellipse on the left side of figure 4.1 depicts the initialized ensemble distribution and the small circles represent initial ensemble members. First, the control state around which the ensemble was perturbed (the striped member) is nonlinearly integrated $n_{\tau au} = 70$ steps. Note the control state is defined to be the mean of the initial ensemble distribution. Then, a randomly chosen member (the black member) is removed from the resulting ensemble and is defined as the perturbation whose future state is forecast using ER. This member is nonlinearly integrated (viz. 1.16-1.18) for $n_{\tau au}$ steps (depicted by the dotted line) to determine its nonlinear state estimate at each of the $n_{\tau au}$ steps (the connected black circle). The remaining ensemble members (the gray members) are separately integrated (viz. 1.16-1.18) for $n_{\tau au}$ steps and a separate regression operator is computed at each of the $n_{\tau au}$ steps (viz. 3.4) using the initial ensemble as the predictor and the $\tau au$-step ensemble as the predictand. The ellipse on the right side of figure 4.1 represents the ensemble distribution at step $n_{\tau au}$ and the large gray circles depict the state of the
ensemble at step $n_{\text{tau}}$. The state of the black left-out ensemble perturbation at the $tauth$ step is then ER forecast (viz. 3.5) using the $tauth$-step operator formed using the (gray) ensemble members to determine its ER estimated state at step $n_{\text{tau}}$ (the connected black dot on the right side). The entire experiment, including the ensemble initialization, is repeated 100 times.

Let $E_{NL}$ denote the average, over the 100 repetitions of the experiment, of the magnitude of the vector difference between the nonlinearly evolved state of the perturbation of interest at the $tauth$ step and the nonlinearly evolved state of the control member. This error measures the nonlinear growth of the perturbation with respect to the control state. Let $E_{ER}$ denote the average, over the 100 repetitions of the experiment, of the magnitude of the vector difference between the nonlinearly evolved state of the perturbation of interest at the $tauth$ step and the ER evolved state of the perturbation of interest. This error measures the degree to which the linear ER forecast state differs from the nonlinear forecast state using the system equations.

The L63 ER forecast error, $E_{ER}$, is compared to two quantities to gauge ER’s usefulness as a linear prediction tool in the nonlinear L63 system. The first quantity is forecasts from the L63 tangent linear model (TLM; e.g. Lorenz 1965; Kalnay 2003), the transpose of which is the adjoint model. The TLM linear operator is defined as

$$\mathbf{M} = \frac{\partial F}{\partial \mathbf{x}}$$ (Kalnay 2003),

where $F$ is an operator that represents the nonlinear integration of the system equations; the TLM operator is the product, over many short integration steps, of the linearized
version of the nonlinear system equations. Let $E_{TLM}$ denote the average, over the 100 repetitions of the experiment, of the magnitude of the vector difference between the non-linearly evolved state of the perturbation of interest at the $tauth$ step and the TLM evolved state of the perturbation of interest. This error measures the degree to which the linear TLM forecast state differs from the non-linearly forecast state using the system equations and serves as a comparison to the forecast performance of a well-established technique. The second quantity is simply the magnitude of the ensemble average non-linear forecast error. Let $E_i$ denote the average, over the 100 repetitions of the experiment, of the average, over the all ensemble members, of the magnitude of the vector difference between the non-linearly evolved state of the perturbation of interest at step the $tauth$ step and the non-linearly evolved state of each ensemble member; ER
forecasts are skillful only if their error magnitudes are small compared to the typical error magnitudes of similarly sized initial perturbations.

Figure 4.2 depicts the results of the L63 experiment. The dashed-dotted, solid, dotted, and dashed lines respectively depict $E_{NL}$, $E_{ER}$, $E_{TLM}$, and $E_t$. To give a sense of the relative size of these errors, all magnitudes are defined as a percentage of the size of the L63 attractor. One notable characteristic of figure 4.2 is that the goodness of the linear ER and TLM forecasts varies inversely with size of the nonlinear error. During the period of small nonlinear error (i.e. insensitive dependence of the integrations to small perturbations in the initial conditions) through approximately 30 integration steps (roughly one day), both $E_{ER}$ and $E_{TLM}$ are approximately zero. That is, during a window of linearity, both ER and TLM forecasts closely approximate the true nonlinearly integrated state, indicating that short-range linear forecasts of nonlinear systems can potentially be highly skillful. Moreover, within this linear window and even up to 70 integration steps, $E_{ER}$ is small compared to $E_t$, implying that ER forecasts much more closely resemble the actual nonlinear state than other “climatological” nonlinear states associated with similarly sized initial perturbations.

$E_{ER}$ closely mirrors $E_{TLM}$ throughout the integration, suggesting that ER forecasts can potentially have similar skill as those from the TLM. In fact, although not particularly applicable for this L63 experiment because the entire state is used to form the regression operator, it is expected that ER forecasts using sufficiently sized ensembles
Figure 4.2: ER forecast error in the L63 system. The dashed-dotted, solid, dotted, and dashed lines respectively depict $E_{NL}$, $E_{ER}$, $E_{TLM}$, and $E_{L}$. To give a sense of the relative size of these errors, all magnitudes are defined as a percentage of the size of the L63 attractor. See text for details.

and resolvable perturbations will generally be more skillful than those from the TLM for many systems. Unlike the TLM, ER linearly parameterizes all physical processes and state perturbations correlated to those used to compute the regression operator, thereby implicitly propagating the perturbation using a relatively more complete representation of
the system dynamics\textsuperscript{1}. Moreover, ER implicitly includes those dynamics that are omitted from TLMs due to a lack of differentiability.

Given that L63 is a three-dimensional system, only three ensemble members that span the initial perturbation are necessary to fully resolve the perturbation. Therefore, ensembles of reasonable size are capable of resolving most perturbations, making L63 integrations relatively insensitive to ensemble size once a minimum ensemble size threshold is reached. However, as will be shown in the following section using the 40-dimensional Lorenz '95 model (Lorenz and Emanuel 1998), because the resolvability of the prescribed perturbation is a strong function of ensemble size in higher-dimensional systems, the $E_{ER}$ will correspondingly be inversely related to the number of ensemble members.

4.3 Lorenz '95 ER ensemble size and sampling error sensitivity

This section employs the 40-dimensional Lorenz '95 model (hereafter L95; Lorenz and Emanuel 1998) to explore characteristics of ER forecast skill in higher-dimensional idealized atmosphere-like systems. The L95 system is defined as

\textsuperscript{1}This idea can be understood using the analogy that an ER with a predictor and predictand defined as shoe size and reading ability among youths, respectively, might show some skill because variables correlated to shoe size, such as age, are implicitly included in the regression; the regression operator may have predictive power attributable to the two fields that define operator being correlated to an extraneous field or mechanism.
\[
\frac{dx_k}{dt} = (x_{K+1} - x_{K-2})x_{K-1} - x_K + F \quad \text{(Lorenz and Emanuel 1998)}
\] 

(4.2)

where \( k = 1, \ldots, K \), \( K = 40 \), \( F = 8 \), \( x_0 = x_K \), \( x_{-1} = x_{K-1} \), and \( x_{K+1} = x_1 \). The model is run using a timestep of 0.025, which corresponds to approximately 3 hours (Lorenz and Emanuel 1998).

A similar variation of leave-one-out cross validation to the one described in the previous section is applied in this section using the L95 model. Here, however, the experiment is performed using only one randomly chosen initial perturbation (rather than 100, as for the L63 experiment), but the initial ensemble members and the size of the ensemble are varied to assess ER sensitivities to ensemble sampling and size. First, a grand ensemble is created with 5000 members and an initial ensemble perturbation of interest is defined from this grand ensemble. Then, 50 separate random combinations of ensembles of size \( n_{ens} = 100 \), \( n_{ens} = 40 \), \( n_{ens} = 30 \), or \( n_{ens} = 20 \) are randomly chosen from the grand ensemble and ER errors are computed for each of these 200 different ensembles. Based on the observed flattening of the respective singular vector spectra (see section 3.3), the \( n_{ens} = 100 \), \( n_{ens} = 40 \), \( n_{ens} = 30 \), and \( n_{ens} = 20 \) ERs respectively use the leading 40, 36, 22, and 12 singular vectors in order to regularize the regressions. Also note that the ensemble mean is recomputed for each ER and the control state is defined as this ensemble mean, so that ER errors are always defined with respect to the appropriate ensemble mean.

Figure 4.3 illustrates fundamental properties of ER ensemble size and sampling error sensitivities. The solid grays lines in panels a-d of figure 4.3 respectively depict the mean value of \( E_{ER} \) over the 50 ERs using different random combinations of ensembles.
from the grand ensemble with the indicated ensemble size. The dotted gray lines bound
1.5 standard deviations from the mean. The black dashed line represents $E_{TLM}$. It is
crucial to note that these results depend substantially on the randomly chosen initial
perturbation of interest; the results here are used to generally illustrate ER properties.

The most obvious characteristic of figure 4.3 is that $E_{ER}$ is a function of $n_{ens}$; L95
$E_{ER}$ increases as the number of ensemble members used to define the operator decreases.
Most noteworthy is the significant difference in $E_{ER}$ between $n_{ens} = 40$ and $n_{ens} = 30$
(panels b and c). When $n_{ens} < K$, it is impossible to fully resolve the initial perturbation
of interest; in this case, the regressed perturbation is the effectively resolved least squares
perturbation, and predictions are necessarily imperfect. Given that L95 is $K = 40$
dimensional, it comes as no surprise, therefore, that the $E_{ER}$ values of the rank deficient
$n_{ens} = 30$ and $n_{ens} = 20$ ERs are substantially greater than those of the $n_{ens} = 100$ and
$n_{ens} = 40$ ERs. In fact, because the ER is forward propagating the effective perturbation
while the nonlinear system equations are integrating a different perturbation (the actual
prescribed perturbation of interest), the significance of $E_{ER}$ for the rank deficient ERs is
actualized after only a single integration step, as portrayed by the large $E_{ER}$ values at the
second step in panels c and d.

As depicted in panels a and b of figure 4.3, on the other hand, the $n_{ens} = 100$ and
$n_{ens} = 40$ $E_{ER}$ are significantly smaller than those of the rank deficient ensembles (panels
c and d). Moreover, the $n_{ens} = 100$ and $n_{ens} = 40$ $E_{ER}$ values do not increase drastically
over the first time step because these ensembles are capable of regressing an effective
perturbation that corresponds well with the prescribed perturbation. However, even the
$E_{ER}$ values of the $n_{ens} = 100$ and $n_{ens} = 40$ ERs are non-negligible. Although only ensemble with $n_{ens} \geq K$ are capable of fully resolving the initial perturbation and therefore (under perfectly linear dynamics) yielding perfect ER forecasts, the $n_{ens} \geq K$ condition does not guarantee skillful forecasts, even within the window of linearity. In addition to ensemble size, resolvability is a function of the degree to which ensemble members span the subspace of the initial perturbation; a larger ensemble increases the likelihood that the ensemble spans this subspace, but the condition of $n_{ens} \geq K$ does not guarantee perfect resolvability.

Another noteworthy characteristic of figure 4.3 is the significant variation of the $E_{ER}$ among ERs using the same numbers of ensemble members to define the ER operator, as illustrated by the dotted lines bounding the 1.5 standard deviation $E_{ER}$ values for the 50 separate ER ensembles for each ensemble size. Because different ensembles span the subspace of the perturbation and resolve the prescribed perturbation with varying degrees, ER results depend on the choice of ensemble members and are subject to significant sampling errors, even when ensemble members are drawn from the same distribution. The portion of $E_{ER}$ attributable to sampling errors becomes increasingly significant at longer lead times, after small changes between initial errors (i.e. between different effective perturbations) have sufficiently amplified. Note that sampling errors are inversely related to ensemble size, as the effective perturbations for larger ensembles have less variations and more uniformly resemble the prescribed perturbation. See section 5.8 for more discussion on ER sampling errors.
Figure 4.3: ER error analysis, ensemble size sensitivity, and sampling error sensitivity in the 40-dimensional L95 system. The solid gray lines in panels a-d of figure 4.3 respectively depict the mean value of $E_{ER}$ over the 50 ERs using different random combinations of ensembles from the grand ensemble with the indicated ensemble size. The dotted gray lines bound 1.5 standard deviations from the mean. The black dashed line represents $E_{TLM}$. See text for more details.

The solid gray lines and dashed black lines of figure 4.3 compare the mean $E_{ER}$ to $E_{TLM}$. For the rank deficient $n_{ens} = 30$ and $n_{ens} = 20$ ensembles (panels c and d), $E_{ER}$ is significantly greater than $E_{TLM}$. However, for the $n_{ens} = 100$ ensemble (panel a), $E_{ER}$ approximately equals $E_{TLM}$ through ten integration steps (approximately 30 hours in the
real atmosphere) and the magnitude of the 1.5 standard deviation of $E_{ER}$ is smaller than $E_{TLM}$ through 20 integration steps (approximately 60 hours). In fact, other chosen initial conditions yielded significantly more skillful ER forecasts than TLM forecasts (not shown). This illustrates that ER using a sufficiently sized and optimized ensemble is capable of outperforming the TLM.

4.4 Operational data ER forecasting and singular vector truncation sensitivity

The previous two sections illustrated important characteristics of ER predictions using low-order Lorenz models. Although a great deal can be learned from these simple atmosphere-like models, the illustrated ER L63 and L95 skill is certainly not guaranteed to translate to skill for high-dimensional sophisticated atmospheric model ERs.

In order to illustrate ER forecast skill for real atmospheric data, this section employs leave-one-out cross validation to the $n_{ens} = 50$ ER forecast of the Japanese Meteorological Association (JMA) 1000 hPa geopotential heights for the $K = 2562$ region of grid points between 10N-40N latitude and 105E-140E longitude on 12UTC August 14 2007 and after. The $K = 15372$ predictor is defined as the 12UTC August 14 2007 analysis JMA geopotential heights at 1000, 850, 700, 500, 300, and 200 hPa for the same region. It is crucial to note that, although only geopotential heights are included as predictors, ER implicitly includes all predictors that are coupled with the explicitly included predictors (see section 4.2). Also note that the choice of predictor is not optimized to yield the greatest ER forecast skill; the results might certainly improve by
adding carefully chosen predictors. Future work might include using the method of screening predictors for this optimization process (e.g. Wilks 2006).

The variation of leave-one-out cross validation employed here begins by removing the first ensemble member from the ensemble and defining it as the control perturbation. Then, a randomly chosen member is removed from the ensemble and is defined as the perturbation of interest. The remaining 48 members are used to form the ER operator and to forecast the future state of the ensemble member of interest. This procedure is repeated for each ensemble member, each time leaving out a different member. This method produces \( n_{\text{ens}} - 2 \) sample ER forecasts that indicate the likely correspondence of the predictand perturbation yielded from the regression of any arbitrary perturbation with the actual state statistically associated with that perturbation.

Figure 4.4 illustrates the results of the experiment. Using the definitions of the errors from section 4.2, the solid line depicts \( E_{\text{ER}} \), the dashed line depicts \( E_{\text{NL}} \), and the dotted line depicts \( E_{\text{f}} \). Keep in mind that errors are defined with respect to the nonlinear model integration, not to truth, and so they indicate ER and operational forecast correspondence rather than ER forecast goodness. To give a sense of the magnitudes of the errors, errors are expressed as mean absolute errors, with units of meters. The numbers of predictand and predictor principal components used for this ER are the optimal numbers determined by the data in figure 4.5, as described below.

As expected, \( E_{\text{ER}} \) increases with increasing \( E_{\text{NL}} \), which increases with lead time. Average absolute differences between the ER forecast and the JMA model predictand state are approximately 2 meters at analysis time and 10 meters at 4 days lead time. Comparing \( E_{\text{ER}} \) to \( E_{\text{f}} \), ER forecasts deviate from the nonlinear state much less than do
Figure 4.4: Cross validated mean absolute error as a function of lead time. The solid line depicts $E_{ER}$, the dashed line depicts $E_{NL}$, and the dotted line depicts $E_i$. See text for details.

other “climatological” nonlinear states associated with initial ensemble perturbations drawn from the same distribution, indicating significant forecast skill past 4 days.

Using the same cross validation procedure, the goodness of high-dimensional operational data ER predictions is also assessed via the ensemble median anomaly correlation coefficient (ACC; e.g. Wilks 2006) of the ER-computed predictand field and the actual ensemble predictand field. Along with being an important aspect of forecast skill, the ability of ER to correctly forecast the locations of perturbation anomaly features
is particularly important because dynamical inferences from ER derive primarily from these perturbation anomaly patterns, as will be shown in section 6.2. Also, in order to assess the ER prediction skill sensitivity to the principal component truncations (see section 3.3), this ACC cross validation experiment is repeated for all combinations of \( n_{pc}^{and} \) and \( n_{pc}^{or} \) ranging from 5 through 50 principal components, in intervals of 5.

Figure 4.5 portrays the ensemble median leave-one-out cross validated anomaly correlation coefficients as a function of lead time. The black line depicts the maximum value, among the combinations of the numbers of predictand and predictor principal component from 5 to 50 in intervals of 5, of the ensemble median ACC for each combination as a function of lead time. The most significant finding portrayed by figure 4.5 is that the ACCs using the denoted \( n_{pc}^{and} \) and \( n_{pc}^{or} \) combinations range between 0.86 and 0.58. These statistically significant ACCs suggest that, although some subtle features of the predictand fields may be inaccurately forecast, ER is highly capable of capturing the majority of the gross features. However, although the chosen predictor and predictand ensembles are certainly viable for synoptic-scale analysis, confidence in the exact nature of the predicted smaller-scale features should be somewhat low especially at the later predictand lead times. As expected, figure 4.5 also indicates that the median ACC decreases as the difference between the predictor and predictand lead times increases; that is, ER predictions worsen as the predictand and predictor ensembles become less contemporaneous.

Figure 4.5 also illustrates the sensitivity of these ACCs to the number of predictor and predictand principal components used to regularize the ER. The values of \( n_{pc}^{and} \) and \( n_{pc}^{or} \) corresponding to this maximum ACC at each lead time are displayed in parentheses.
Black dots at each lead time depict the ensemble median ACC computed using other combinations of $n_{pc}^{and}$ and $n_{pc}^{or}$ for that lead time. Parenthetical values of $n_{pc}^{and}$ and $n_{pc}^{or}$ range between 10 and 30, suggesting that one-quarter to three-quarters of the principal components should be truncated to yield forecasts that best resolve significant anomaly features.

Figure 4.5: Ensemble median leave-one-out cross validated anomaly correlation coefficients as a function of lead time. The black line depicts the maximum value, among the combinations of the numbers of predictand and predictor principal component from 5 to 50 in intervals of 5, of the ensemble median ACC for each combination as a function of lead time. The optimal values of $n_{pc}^{and}$ and $n_{pc}^{or}$ at each lead time are displayed in parentheses. Black dots at each lead time depict the ensemble median ACC computed using other combinations of $n_{pc}^{and}$ and $n_{pc}^{or}$ for that lead time.
ACCs for individual lead times vary by as much as approximately 0.3 (6 hour lead time) and as little as 0.1 (60 hour lead time) depending on the number of truncated principal components. Although this suggests very high truncation sensitivity, note that, the dots for most lead times cluster at an ACC range close to the maximum ACC value for the respective lead time; in general, the median ACC for each lead time increases with increasing numbers of principal components until a threshold number of principal components, beyond which the median ACC remains nearly constant or decreases (not shown) with increasing numbers of principal components. This suggests that $n_{pc}^{nd}$ and $n_{pc}^{er}$ values close to the parenthetical ones in figure 4.5 yield ACCs similar to the cross validated maximum values, and therefore a relatively large range of $n_{pc}^{nd}$ and $n_{pc}^{er}$ values yields similar results. Additionally, because overfit regressions yield high variance results (e.g. Wilks 2006), it is advantageous to define regression operators using the fewest possible predictand and predictor principal components, without sacrificing regression accuracy. Therefore, the optimal values of $n_{pc}^{nd}$ and $n_{pc}^{er}$ at each lead time may in fact be closer to those towards the bottom of the cluster, rather than the top; using values of $n_{pc}^{nd}$ and $n_{pc}^{er}$ corresponding to the maximum ER ACCs yields ERs with only slightly greater expected regression accuracy, but potentially significantly greater ER variance compared to ERs that use $n_{pc}^{nd}$ and $n_{pc}^{er}$ values towards the bottom of the cluster.

4.5 Chapter summary

This section assessed the skill of ER forecasts using leave-one-out cross validation. Low-order atmosphere-like Lorenz 63 and Lorenz 95 models are used to
compare differences between ER and tangent linear model (TLM) predictions of the future state of an ensemble perturbation, with respect to the actual state determined by the integration of the deterministic nonlinear system equations.

L63 ER yields highly accurate forecasts comparable to those of the TLM within a window of linearity, suggesting that linear ER forecasts can potentially be skillful even for highly nonlinear systems. L95 is used to assess ER forecast sensitivity to ensemble size and sampling error. It is found that ensembles with sizes greater than the dimensionality of the system equations yield forecasts significantly more skillful than those with smaller sizes. This difference is attributable to larger ensembles being significantly more likely to span the subspace of the ensemble perturbation of interest; L95 forecasts of well resolved ensemble perturbations have skill comparable to that of TLM forecasts. Also, L95 ER forecasts showed significant sensitivities, particularly for long lead times, to the choice of ensemble members used to defined the ER operator, suggesting that ER forecasts are potentially subject to significant sampling errors.

Ensemble data from the Japanese Meteorological Association is used to illustrate ER skill for high-dimensional sophisticated operational data. ER mean absolute errors, with respect to JMA forecasts, of 1000 hPa geopotential height perturbations ranged from approximately 2 meters at analysis time to 10 meters at 4 days lead time and median anomaly correlation coefficients ranged from 0.86 at analysis time to 0.58 at 72 hours lead time. Principal component sensitivity analysis suggested that, although highly sensitive when considering the full spectrum of potential degrees of truncation, ER ACC values are significantly less sensitivity within a relatively wide range of principal component numbers close to the optimal numbers.
Chapter 5

Piecewise Potential Vorticity Ensemble Regression

5.1 Introduction

Hakim and Torn (2008; hereafter HT08) presented ESA in a general “proof-of-concept” manner to study extratropical cyclones and piecewise potential vorticity (PV) inversion. This chapter reinterprets HT08’s statistical piecewise PV “inversion” technique as being a PV ensemble regression and compares PV ER results to those from the established dynamical inversion technique of Davis and Emanuel (1991, hereafter DE) in an effort to explore the applicability of the approach. Section 5.2 introduces PV, presents the theory of piecewise PV inversion, and describes the DE piecewise PV inversion method. Section 5.3 describes the PV ER technique (Gombos and Hansen 2008, hereafter GH08). Section 5.4 defines the fundamental differences between DE inversion and PV ER and illustrates these differences in a contrived two-layer atmosphere. Section 5.5 presents the procedure and results of a DE-ER comparison using a 100 member Weather Research and Forecasting (WRF) ensemble. Section 5.6 analyzes these results and accounts for discrepancies. Section 5.7 generalizes the utility of PV ER. Section 5.8 discusses the potential application of covariance localization to PV ER and section 5.9 provides conclusions. Note that much of the following comes from GH08.
5.2 PV thinking and Davis and Emanuel (1991) piecewise PV inversion

The dual properties of conservation and invertibility have long secured “PV thinking” as a cornerstone of atmospheric dynamics. The conservation of Ertel’s PV\(^2\) (EPV) (Ertel 1942), defined as

\[
q = \frac{1}{\rho} \eta \cdot \nabla \theta , \quad (5.1)
\]

and the conservation of quasi-geostrophic (QG) PV\(^3\) (Charney and Stern 1962), defined as

\[
q_p = \nabla^2 \psi + f_0 + \beta(y - y_0) + \frac{f_0^2}{\bar{\rho}} \left( \frac{\bar{\rho}}{N^2} \psi_z \right)_z , \quad (5.2)
\]

renders PV a dynamically active Lagrangian tracer useful for identifying the origin and trajectories of air parcels. Here, \(\eta\) is the absolute vorticity vector, \(\theta\) the potential temperature, \(\bar{\rho}\) the density (a function of the vertical coordinate, only), \(\psi\) the geostrophic streamfunction, \(N\) the Brunt Vaisalla frequency, \(f\) the Coriolis parameter, and \(\beta = \partial f / \partial y\). The second fundamental property of PV is the invertibility principle: one can diagnostically deduce the distribution of velocity and mass from the spatial distribution of PV given suitable boundary conditions (e.g. Hoskins et al. 1985). PV inversion is the primary focus of this chapter.

---

\(^2\) EPV is only conserved following three-dimensional, adiabatic, inviscid motion.

\(^3\) QGPV is only conserved following geostrophic, adiabatic, and inviscid motion for \(R_o \equiv U / f L \ll 1\). Here, \(R_o\) is the Rossby number, \(U\) the horizontal wind speed, \(L\) the horizontal length scale, and \(f\) the Coriolis parameter.
Employing the simplifications of the Eady (1949), Bretherton (1966), and Hoskins et al. (1985) models, the invertibility principle can be used to depict baroclinic instability as resulting from the mutual amplification of phase-locked Rossby waves. It is clear from these models that the interactions of PV anomalies contribute significantly to cyclogenesis. In order to disentangle the dynamics of cyclogenesis, it is necessary to quantify the relative contributions of subsets of the PV anomaly field to a cyclone’s velocity and mass fields. This can be accomplished via piecewise PV inversion (DE; Davis 1992a). Whereas full-field PV inversion recovers the entire three dimensional velocity and mass fields (thereby obscuring the relative PV contributions to these fields), piecewise PV inversion attempts to recover the subset of the fields attributable to a subset of the PV field (thereby enabling diagnosis of the relative PV contributions to the fields). This method has been used to study aspects of cyclogenesis including the interaction of PV anomalies (Huo et al. 1999), the importance of initial structure and condensational heating (Davis 1992b), and the effects of jet-streak-induced circulations (Huo et al. 1995). An analysis, development, and comparison of existing piecewise PV inversion techniques constitutes the remainder of this chapter.

The most notable of the existing PV inversion techniques is that presented in the seminal paper of Davis and Emanuel (1991). The importance of the DE formulation is rooted in its ability to invert high order PV balance approximations. Because the QGPV invertibility relation given by 5.2 is limited by geostrophic flow and QG scaling constraints, DE recognized that piecewise inverting the relatively unconstrained EPV (viz. 5.1) would allow for more generalized and accurate cyclogenesis diagnosis. However, a strict invertibility relation for EPV does not exist, necessitating the use of
Charney’s (1962) balance equations to approximate one. By defining two Rossby numbers, $R_\varphi = V_\varphi / f_0 L$ and $R_\chi = V_\chi / f_0 L$, retaining $O(R_\varphi)$ terms, and neglecting $O(R_\chi)$ terms, the horizontal divergence of the horizontal momentum equation can be expressed as

$$\nabla^2 \Phi = \nabla \cdot (f \nabla \Psi) + \frac{2}{a^4 \cos^2 \phi} \frac{\partial (\partial \Psi / \partial \lambda, \partial \Psi / \partial \phi)}{\partial (\lambda, \phi)}$$

(5.3) (DE). Replacing the horizontal velocity of EPV by the nondiagonal wind, 5.1 can be approximated as

$$q = \frac{g \kappa \pi}{p} \left[ (f + \nabla^2 \Psi) \frac{\partial^2 \Phi}{\partial \pi^2} - \frac{1}{a^2 \cos^2 \phi} \frac{\partial^2 \Psi}{\partial \lambda \partial \pi} \frac{\partial^2 \Phi}{\partial \lambda \partial \pi} - \frac{1}{a^2 \partial \pi} \frac{\partial^2 \Psi}{\partial \phi \partial \pi} \frac{\partial^2 \Phi}{\partial \phi \partial \pi} \right]$$

(5.4) (DE). Here, $\Phi$ is the geopotential, $\Psi$ the nondiagonal streamfunction, $\lambda$ the longitude, $\phi$ the latitude, $a$ the Earth’s radius, $\kappa = R_d / C_p$, and $\pi = C_p (p / p_0)^{\kappa}$ is the vertical coordinate Exner function. The coupled system 5.3 and 5.4 are used to recover the balanced mass and nondiagonal wind fields from the approximated EPV (DE; Davis 1992a). $\Psi$ and $\Phi$ are used for any lateral boundary conditions and, noting that the boundary potential temperature is a surrogate for PV (Bretherton 1966), Neumann conditions on the horizontal boundaries are defined as

$$\frac{\partial \Phi}{\partial \pi} = -\theta, \quad (\pi = \pi_o; \pi = \pi_r)$$

(5.5) (DE).

By defining a PV perturbation as a deviation from a synoptic-scale time-mean and characterizing the invertibility relation as $q(t) = A(t)B(t)$, the PV perturbation field can be decomposed into $N$ parts such that
\[ q_n = \left( A + \frac{1}{2} \sum_{n=1}^{N} A_n \right) B_n + \left( \hat{B} + \frac{1}{2} \sum_{n=1}^{N} B_n \right) A_n \]  

(5.6)

(DE). Using 5.6, one can derive the perturbation forms of 5.3,

\[ \nabla^2 \Phi_n = \nabla \cdot (f \nabla \Psi_n) + \frac{2}{a^4 \cos^2 \phi} \left( \frac{\partial^2 \Psi_n}{\partial \lambda^2} \frac{\partial^2 \Psi_n}{\partial \phi^2} + \frac{\partial^2 \Psi_n}{\partial \phi^2} \frac{\partial^2 \Psi_n}{\partial \lambda^2} - 2 \frac{\partial^2 \Psi_n}{\partial \lambda \partial \phi} \frac{\partial^2 \Psi_n}{\partial \lambda \partial \phi} \right) \]  

(5.7)

(DE) and 5.4

\[ q_n = \frac{8 \kappa \pi}{p} \left[ (f + \nabla^2 \Psi_n) \frac{\partial^2 \Phi_n}{\partial \pi^2} - \frac{\partial^2 \Phi_n}{\partial \pi^2} \nabla^2 \Psi_n - \frac{1}{a^4 \cos^2 \phi} \left( \frac{\partial^2 \Psi_n}{\partial \lambda \partial \pi} \frac{\partial^2 \Phi_n}{\partial \lambda \partial \pi} + \frac{\partial^2 \Phi_n}{\partial \lambda \partial \pi} \frac{\partial^2 \Psi_n}{\partial \lambda \partial \pi} \right) \right. \]

\[ \left. - \frac{1}{a^2} \left( \frac{\partial^2 \Psi_n}{\partial \phi \partial \pi} \frac{\partial^2 \Phi_n}{\partial \phi \partial \pi} + \frac{\partial^2 \Phi_n}{\partial \phi \partial \pi} \frac{\partial^2 \Psi_n}{\partial \phi \partial \pi} \right) \right], \]  

(5.8)

(DE) where \[ \left[ \begin{array}{c} \Psi_n \\ \Phi_n \end{array} \right] = \left[ \begin{array}{c} \Psi_n^* \\ \Phi_n^* \end{array} \right] \frac{1}{2} \sum_{n=1}^{N} \left[ \begin{array}{c} -1 \\ 1 \end{array} \right], \] which can then be solved through successive overrelaxation to yield the streamfunction and geopotential perturbation fields attributable to a particular subset of the PV perturbation field (DE). This arbitrary choice of partitioning linearizes the PV around the time-mean state, enabling superposition of solutions, but retains the nonlinear terms in the coefficients of the linear differential operator (DE). Homogeneous boundary conditions for \( \Psi_n \) and \( \Phi_n \) are used on any lateral boundaries. Neumann conditions on the horizontal boundaries are defined as

\[ \left( \frac{\partial \Phi_n}{\partial \pi}, \frac{\partial \Psi_n}{\partial \pi} \right) = -\Theta_n, \quad (\pi = \pi_0; \pi = \pi_f) \]  

(5.9)

(DE).

### 5.3 Defining PV ER

Another PV “inversion” technique is that presented in HT08 and developed in GH08. Unlike traditional PV inversion techniques that employ deterministic physics to
deduce the state attributable to a PV perturbation, the GH08 technique determines the state associated with a PV perturbation via the covariances between the PV, potential temperature, and state fields. Therefore, the author has chosen to distinguish the GH08 technique from classical inversions by labeling it a PV "regression" algorithm. However, unlike most regressions, which employ educated guesswork to determine the optimal predictor set, atmospheric dynamical theory has supplied the exact, necessary, and sufficient set for PV ER; because classical PV inversion theory states that the balanced height field can be fully determined from PV and potential temperature boundary conditions, a regression inversion operator defined by covariance estimates of predictor PV and potential temperature and predictand geopotential height is sufficient to determine the height perturbations statistically associated with the regressed PV perturbation.

GH08 PV ER is defined as follows. Refer to Table 3.1 to reference definitions of the symbols used to describe ER. Let $Y_e'$ and $P_e'$ be $K \times n_{ens}$ ensemble state anomaly and $k \times n_{ens}$ ensemble predictor anomaly matrices, respectively, where each of the $K$ or $k$ rows, respectively, corresponds to a point on a vectorized grid, each of the $n_{ens}$ columns corresponds to an individual analysis ensemble member, the prime denotes a perturbation, and the subscript $e$ denotes that this perturbation is an anomaly with respect to the ensemble mean. $P_e'$ is comprised of $k_b$ entries of upper and lower boundary potential temperature ensemble anomalies and $k - k_b$ entries of interior PV ensemble anomalies. Let $L$ be a $K \times k$ linear inversion operator defined by 3.4 such that the PV ER invertibility relation is given by 3.5, where $\hat{y}'$ is the $K \times 1$ estimated height
perturbation associated with the $k \times 1$ PV perturbation $\mathbf{p}'$. Note that the boundary potential temperature was not included in the predictor matrix in HT08; it has been included here because, according to atmospheric dynamical theory, it is one of the necessary and sufficient predictors required to recover the height perturbation field.

Considering each entry in the rows of $\mathbf{p}'$ to be a point source of PV, the Green’s function piecewise inversion technique (Pedlosky 1979) can be mimicked by zeroing out all entries of $\mathbf{p}'$ except those whose influence is of interest for the piecewise regression (HT08; GH08). PV can then be piecewise regressed via

$$\mathbf{\tilde{y}}' = L\mathbf{\tilde{p}}',$$  \hspace{1cm} (5.10)

where $\mathbf{\tilde{p}}'$ is this sparse version of $\mathbf{p}'$ and $\mathbf{\tilde{y}}'$ is the ensemble state anomaly estimate attributable to the point sources of PV in $\mathbf{\tilde{p}}'$ (HT08; GH08).

As discussed in section 3.3, multicollinearities due to geographic proximity are likely to render the estimate of $\mathbf{P}_{e}^{-1}$ ill-conditioned. This ill-conditionedness implies that the regression of PV perturbations other than those that define $\mathbf{P}_{e}'$ may yield inaccurate results, necessitating the regularization of $\mathbf{P}_{e}^{-1}$. PV ER corrects the ill-conditionedness of $\mathbf{P}_{e}^{-1}$ by employing the truncated SVD (TSVD; Golub and van Loan 1996) in lieu of 3.8. See section 3.3 for more details.

Whereas the DE approach piecewise inverts PV using a diagnostic dynamical inversion relation, the GH08 PV ER approach defines the relation in terms of the covariance between the PV, potential temperature, and state fields (viz. 3.7). That is, the regression operator is simply the covariance between ensemble PV and potential
temperature anomalies and the state estimate, normalized by the inverse of the covariance of the ensemble PV anomalies. Alternatively, PV ER can be interpreted as a multidimensional application of Bayes’ Theorem, in which the conditional probability of the state variables given the predictor variables is “inverted” (Hakim 2008). Note that PV ER requires no finite differencing, solution convergence, map factors, or modifications to account for grid shapes, thereby avoiding associated ambiguities, limitations, and errors (HT08; GH08).

5.4 Defining PV ER and DE inversion differences

Both the DE and PV ER algorithms are designed to deduce the state given a PV perturbation and boundary conditions. However, the preceding sections highlighted the fundamental difference between the two techniques: DE inversion yields the state physically attributable to a PV perturbation, whereas PV ER yields the state statistically associated with a PV perturbation. Because, as explained below, all height perturbations statistically associated with a PV perturbation will not necessarily be physically attributable to that perturbation, height perturbations resulting from the GH08 and DE techniques should not necessarily be identical.

However, neglecting the aforementioned physical approximations in the DE formulation and the sampling and nullspace errors described in section 5.6, the only reason why the two techniques yield different states is because a different PV perturbation is effectively applied by each algorithm; DE inversions yield the state attributable to a given PV perturbation, \( \mathbf{\tilde{p}} \), whereas PV ERs determine the state
associated with an "effective" PV perturbation, $\tilde{\mathbf{p}}^\prime$, that is comprised of PV perturbations that are statistically associated with $\tilde{\mathbf{p}}^\prime$. GH08 posit that, neglecting the aforementioned errors, PV ER and DE inversion yield the same state if the same "effective" PV perturbation is applied.

Before demonstrating this principle using real atmospheric data, this concept is illustrated by performing DE PV inversion and PV ER in a contrived two layer balanced atmosphere (see figure 5.1). Suppose that this atmosphere contains two mutually amplifying Rossby waves, such that the $\mathbf{P}_e^\prime$ at the northeast quadrant of the upper level (denoted with an A) is perfectly correlated with the $\mathbf{P}_e^\prime$ at the southwest quadrant of the lower level (denoted with a B), as expressed by a $n_{ens} = 500$ member ensemble. Because of the requirement for atmospheric balance, the $\mathbf{P}_e^\prime$ at each level is perfectly negatively correlated with the collocated height perturbations, $\mathbf{Y}_e$, and because of the PV "action at a distance" principle (e.g. Bishop and Thorpe 1994), the $\mathbf{P}_e^\prime$ at each level induces a $\mathbf{Y}_e^\prime$ at the opposing level. Assume that the $\mathbf{P}_e^\prime$ associated with the induced $\mathbf{Y}_e^\prime$ at the opposing level is negligible, so that the correlation between the $\mathbf{P}_e^\prime$ at areas A and B and the $\mathbf{P}_e^\prime$ at the opposite quadrant of the respective level is also negligible. That is, the $\mathbf{P}_e^\prime$ at each level is perfectly negatively correlated with the collocated $\mathbf{Y}_e^\prime$ and the $\mathbf{Y}_e^\prime$ at the opposite quadrant of the respective level, and perfectly positively correlated with the $\mathbf{P}_e^\prime$ at the opposite quadrant of the opposing level. Figure 5.1, which depicts the correlation of $\mathbf{P}_e^\prime$ at point A with $\mathbf{P}_e^\prime$ at all other points, illustrates this scenario (GH08).
Figure 5.2 shows results from a classical piecewise inversion and an ensemble regression of the upper level PV perturbation, $\tilde{\mathbf{p}}'$, in figure 5.2a, given the aforementioned ensemble covariances. Note that figure 5.2a depicts only the upper level PV perturbation because the lower level PV perturbation is set to zero (as explained in section 5.3, piecewise PV ER requires the zeroing out of all PV perturbations other than the perturbation of interest, which in this case is the perturbation at the upper level).

Figure 5.2b shows the inverted height perturbation using the Laplacian as the inversion method.

---

**Correlation Map of the PV Ensemble Anomalies at Area A**

- a. Upper Level
- b. Lower Level

Figure 5.1: Correlation map of the PV ensemble anomalies at area A. Each point represents the correlation, as defined by the contrived $n_{ens} = 500$ member ensemble described in section 5.4, of the point A with all other points. Black areas indicate a zero correlation and white areas indicate a unity correlation. From GH08.
Comparison of HT Regression and Classical Inversion Results

a. Inverted PV Field

b. Upper Level Inverted Height Field using the Laplacian Operator

c. Upper Level Inverted Height Field using HT

Figure 5.2: A comparison of ER and classical inversion results, using the contrived ensemble described in section 5.4. (a) The nonzero entries of the inverted PV field, $\tilde{p}'$. Note that black areas represent zero values and white areas represent unity values. (b) The inverted height field attributable to $\tilde{p}'$, using the Laplacian operator. Note that white areas represent zero values and black areas represent negative values. (c) The regressed height field statistically associated with $\tilde{p}'$, using the ER method with $w_r = 100$ singular vectors. Note that white areas represent zero values and black areas represent negative values. From GH08.

operator and figure 5.2c shows the regressed height perturbations using the PV ER technique with $w_r = 100$ singular vectors. Because the Laplacian of a function tends to

4 In a QG barotropic system, the PV and height fields are exactly related by the Laplacian and so the Laplacian is used here to illustrate the behavior of a classical inversion.
be a maximum where the function itself is a minimum, the location of the negative height perturbation in the classical inversion results are collocated with the location of the inverted positive PV perturbation. The statistically associated heights of figure 5.2c, however, appear at both the northeast and southwest quadrants of the domain. At these locations, $\tilde{p}'$ strongly covaries with the height field, producing high magnitudes for the corresponding entries in L and therefore high magnitudes for the corresponding entries of $\tilde{y}'$ from 5.10 (GH08).

Although the height perturbation at the northeast quadrant of figure 5.2b is both statistically correlated with and dynamically attributable to $\tilde{p}'$, the height perturbation in the southwest quadrant of figure 5.2c appears only because it is statistically correlated with $\tilde{p}'$; it is dynamically attributable to the zeroed out PV perturbation at area B (not shown; recall that there is no perturbation at the lower level of the regressed PV perturbation) (GH08).

This point leads to the following alternative interpretation of the PV ER technique: because $p'$ (the nonzero elements of $\tilde{p}'$) is perfectly correlated with $P_e$ at area B, it is statistically impossible for $p'$ to exist without an equivalent PV perturbation at area B. Therefore, piecewise regressing $\tilde{p}'$ is statistically equivalent to regressing an “effective” perturbation, $\tilde{p}'$, defined by both $p'$ and an additional PV perturbation at area B, thereby accounting for the height perturbation in the southwest quadrant of figure 5.2c. That is, even though the lower level PV perturbation is not explicitly included as part of the regressed PV perturbation, ensemble correlations require that it be effectively included
because the linear combinations of $P_e$ that resolve the upper level perturbation necessarily also resolve a PV perturbation at the lower level (at area B). See section 3.4 for more discussion of the effective perturbation. Note that the violation of ensemble probabilities is irrelevant in the deterministic DE framework, since each DE inversion operator is defined with respect to only the PV perturbation being inverted and is independent of other perturbations.

At the beginning of this section, the author posited that, neglecting the mentioned sources of error, classical inversions and PV ERs yield the same inverted heights if the same PV perturbation is applied. This statement is supported by the inversion of the effective perturbation, $\tilde{p}'$, illustrated in figure 5.3. Figures 5.3a and 5.3b depict $\tilde{p}'$ at the upper and lower levels, respectively. Note that, even though the PV perturbation at area B is not directly included as part of the perturbation applied to ER, $\tilde{p}'$ (figure 5.2a), it is effectively included in the regression (see figure 5.3b) because linear combinations of $P_e$ are unable to resolve the upper level perturbation at area A without also resolving the perfectly correlated lower level perturbation at area B. Figure 5.3c shows the inversion of $\tilde{p}'$ using the Laplacian operator. Clearly, the same pattern as that from ER illustrated in figure 5.2c emerges when this effective perturbation $\tilde{p}'$, not the applied perturbation $\tilde{p}'$, is inverted (GH08). Section 5.5 will illustrate this principle in the real atmosphere.
Figure 5.3: The effects of the effective PV perturbation, \( \tilde{\mathbf{p}}' \). (a) The upper level effective PV perturbation. Note that black areas represent zero values and white areas represent unity values. (b) The upper level effective PV perturbation. Note that black areas represent zero values and white areas represent unity values. (c) The inverted heights attributable to the effective PV perturbation using the Laplacian operator. Compare with 5.2c. Note that white areas represent zero values and black areas represent negative values. From GH08.

### 5.5 Implementation and results of the PV ER and DE91 inversion comparison

The contrived example presented in the previous section illustrates that PV ERs and DE inversions yield the same results if the same PV perturbation is applied. The following sections demonstrate this assertion by comparing the results from a DE
piecewise inversion and ER of the same effective PV perturbation using output from a sophisticated atmospheric model. This section elaborates on the implementation of the two techniques, describes modifications performed to ensure a fair comparison, and presents results.

The DE and ER techniques are implemented as follows. The full “balanced” PV, $P$, and the full balanced heights, $Y$, are calculated via 5.3 and 5.1 separately for $n_{ens} = 100$ ensemble members from the temperature, wind, and geopotential height Weather Research and Forecasting model (WRF; Michalakes et al. 2001; Hakim and Torn 2006) output data of 06 UTC 29 March 2003. Observations consisting of 250 randomly spaced surface pressure readings sampled from a truth run are assimilated using the ensemble square-root filter (Whitaker and Hamill 2002). As discussed in section 1.4, the ensemble square-root filter, which is a variation of the EnKF, uses evolving non-stationary error covariances, so that the flow-dependent errors of the day associated with the underlying baroclinic instabilities are captured by the ensemble statistics (Whitaker and Hamill 2002; Kalnay 2003). All data is calculated for the 500, 400, 300, 250, 200, 150, and 100 hPa pressure levels. This model has ~100 km horizontal grid spacing on a 90×90 grid with 28 vertical levels that have been interpolated to pressure coordinates at a pressure interval of 50 hPa. Model parameterizations include warm-rain microphysics, the MRF planetary boundary layer scheme (Hong and Pan 1996), and the Janjic (1994) convective parameterization scheme. The data is the same as that used in HT08 and GH08 and the author refers the reader to HT08 for a more detailed description of the model and data generation algorithm.
Although any PV perturbation can be inverted (regressed) using the DE (ER) technique, the PV perturbation chosen to be inverted (regressed) here is the ensemble mean of the PV anomaly defined with respect to the synoptic-scale time mean, scaled down by a factor $\alpha = 0.2$ (as explained below). That is, the PV perturbation is simply the scaled ensemble mean of the PV associated with the UTC 29 March 2003 cyclone, or $p' = \alpha(\overline{p}_e - \overline{p}_t)$, where the overbar denotes an average of the subscripted norm and $\overline{P}_t$ is taken as the average PV, as calculated (viz. 5.1) using the Global Forecasting System (GFS) model’s analyses for the 00, 06, 12, and 18 UTC for the seven consecutive days centered on 29 March 2003. The PV perturbation associated with a cyclone was chosen to be inverted because most PV inversion applications are attempts to anatomize cyclones and because doing so maintains consistency with the inversion procedure used in DE (GH08).

It is necessary to scale down the perturbation by a factor $\alpha = 0.2$ so that the DE inversion solutions are relatively independent of the choice of mean state around which the DE perturbation equations 5.7 and 5.8 are linearized. Note that inversion solutions are functions of the mean state chosen for the linearization because the linearization of 5.7 and 5.8 introduces mean-dependent coefficients in the inversion operator. However, if the perturbation is sufficiently small, the differences between inversion solutions resulting from linearizations around the ensemble mean and time mean (the two natural choices for this mean state, given that $p' = \alpha(\overline{p}_e - \overline{p}_t)$) are negligible; Taylor expansions around the time and ensemble means are approximately equal for sufficiently small perturbations. Therefore, the perturbation is scaled by a factor $\alpha = 0.2$, the greatest factor resulting in negligible inversion differences (GH08).
For larger perturbations, the ensemble-mean linearization may be more appropriate than the time-mean linearization for capturing the state dependent dynamics. The flow-dependent ensemble-mean linearization captures the current synoptic situation, causing the inversion operator to account for the current flow-dependent relationship between the PV and height fields. The time-mean linearization, on the other hand, accounts for the relationship over the entire week, thereby diluting the current dynamics.

GH08 arbitrarily chose to compute the 400, 300, 250, 200, and 150 hPa geopotential height perturbations associated with the PV perturbation at 300 hPa for this DE-ER comparison. \( \tilde{p} \) is calculated from \( p \) by zeroing out all entries in \( p \) other than those that correspond with the 300 hPa PV and the 500 hPa and 100 hPa boundary potential temperatures. Figure 5.4 depicts all non-zero entries in \( \tilde{p} \); Figures 5.4a and 5.4b respectively illustrate the ensemble mean of the 500 hPa and 100 hPa boundary potential temperature time-mean perturbation and figure 5.4c shows the ensemble mean of the 300 hPa time-mean PV perturbation.

Figure 5.5 shows the results of the PV ER and DE inversion comparison. Figures 5.5a,c,e,g,i depict \( y^{ER} \), the 400, 300, 250, 200, and 150 hPa ER height perturbation, respectively, statistically associated with the 300 hPa PV perturbation and boundary potential temperature, \( \tilde{p} \) (figure 4), computed by solving 5.10. Note that \( P_{e}^{-1} \) is calculated via 3.8 with \( w_{i} = 20 \) singular vectors and \( L \) is calculated via 3.4. Recall that \( \tilde{p} \) (figure 5.6), not \( p \) (figure 4), is the PV perturbation effectively ensemble regressed. The effective PV perturbations depicted in figure 5.6, \( \tilde{p} \), are the linear combinations of \( P_{e} \) at 500, 400, 300, 250, 200, and 150, and 100 hPa (the seven pressure
Figure 5.4: The nonzero entries of the regressed perturbation, $\tilde{p}'$, the ensemble-mean of the time-mean anomaly. (a) The ensemble-mean of the upper boundary 500 hPa time-mean potential temperature perturbation. (b) The ensemble-mean of the lower boundary 100 hPa time-mean potential temperature perturbation. (c) The ensemble-mean of the 300 hPa time-mean PV perturbation in PVU, where $1 \text{PVU} = 10^6 \text{m}^2 \text{Kkg}^{-1} \text{s}^{-1}$ (GH08).

levels considered in the calculation of the ER operator) that yield the best least squares estimate of the applied perturbation, $\tilde{p}'$. Therefore, because $\tilde{p}'$ is the PV perturbation effectively regressed, the ER results (figures 5.5a,c,e,g,i) are compared to $y^{DE}$ (figures 5.5b,d,f,h,j; viz. 5.7 and 5.8), the 400, 300, 250, 200, and 150 hPa DE height
Figure 5.5: (a,c,e,g,i) Regressed height perturbation in meters at 400, 300, 250, 200, and 150 hPa, respectively, statistically associated with $\mathbf{p}'$, the PV perturbation in figure 5.4, using the ER technique (viz. 5.10) with $w_s = 20$ singular vectors. Note that, because $\mathbf{p}'$ is unable to fully resolve $\mathbf{p}'$, the PV perturbation effectively regressed is the perturbation in figure 5.6. (b,d,f,h,j) Inverted height perturbation in meters at 400, 300, 250, 200, 150 hPa, respectively, attributable to $\mathbf{p}'$, the effective PV perturbation in figure 5.6, using the DE technique. From GH08.
Figure 5.6: \( \mathbf{\tilde{p}} \), the effective PV perturbation resolved by linear combinations of the ensemble anomalies, \( \mathbf{P}_e \), at the indicated pressure level, in PVU, where

\[ 1 \text{ PVU} = 10^6 \text{ m}^2 \text{ K kg}^{-1} \text{ s}^{-1} \]. From GH08.

perturbations, respectively, attributable to \( \mathbf{\tilde{p}} \). For comparison, \( \mathbf{y}^{DE^*} \), the DE height perturbation attributable to the directly applied perturbation, \( \mathbf{p}' \), is illustrated in figure 5.7. Note that, in order to mitigate the effects of the lateral boundaries on the results, the plotted field is smaller than the computed field by five degrees longitude and latitude for all figures (GH08).
5.6 PV ER and DE inversion results analysis

This section discusses how the ER and DE inverted heights compare to each other and to theoretical expectations and how ensemble deficiencies, null spaces, and linearity assumptions contribute to and account for the discrepancies.

The ER and DE inversion results are consistent with their respective statistical and physical theoretical expectations. Considering the negatively covarying height and PV fields (not shown), the negative (positive) height perturbation in the central (northeastern and southwestern) portion of the ER results (figures 5.5a,c,e,g,i) is accordant with the collocated positive (negative) regressed PV perturbation (figure 5.6). Likewise, considering that the analytical inversion relation is essentially a Laplacian operator (viz. 5.7 and 5.8) and that the Laplacian of a function tends to be a maximum where the function itself is a minimum, the locations of the negative (positive) height perturbations in the DE results are consistent with the locations of the positive (negative) PV perturbations (GH08).

Most importantly, the ER and DE results are unequivocally alike, supporting the author’s prior posit that the two algorithms yield the same state if the same effective PV perturbation is applied, neglecting the aforementioned and following sources of discrepancies. The primary features in each of the panels are collocated, as supported by the 98.8, 98.6, 98.5, 98.7, and 99.1% correlations, respectively, between the five pairs of corresponding ER and DE fields depicted in figure 5.5. The magnitudes of the heights are also similar, with the respective root mean squared error between the pairs of fields being only 0.8, 0.8, 0.9, and 0.9 meters. The primary differences between the two
fields are the slightly higher magnitude ER heights (figures 5.5a,c,e,g,i) in the northeastern section compared to the collocated DE heights (figures 5.5b,d,f,h,j) (GH08).

Figure 5.7 displays $y^{DE^*}$, the 400, 300, 250, 200, and 150 hPa heights attributable to $\tilde{p}'$, the 300 hPa time-mean PV perturbation and the boundary potential temperature in figure 5.4, via the DE inversion method. That is, figure 5.7 displays the DE inverted heights attributable to the perturbation directly applied to the ensemble regression. As expected, although they are consistent with theoretically expected results, the inversion results in figure 5.7 do not closely resemble the ER of $\tilde{p}'$ presented in figure 5.5; the positive height perturbations in the southwestern sections of the ER are not present in the DE inversions of $\tilde{p}'$ and the magnitudes are non-negligibly different. This is simply because $\tilde{p}'$ (figure 5.4), not $\tilde{p}$ (figure 5.6), is the effective PV perturbation that resulted in the inverted heights displayed in figure 5.7. Note that the discrepancies between the height perturbations in figure 5.5 and 5.7 are consistent with the discrepancies between the PV perturbations in figure 5.6 and 5.4. This is exemplified by the relative strength (weakness) of the negative PV perturbations in the southwestern sections of the panels of figure 5.6 (5.4) and the corresponding presence (absence) of positive height perturbations in the southwestern sections of the panels of figure 5.5 (5.7) (GH08).

Despite the obvious likeness of the PV ER and DE inversion results depicted in figure 5.5, null spaces and sampling and model errors result in solution discrepancies. The first of these null spaces is due to both $P_e$ and $Y_e$ being underdetermined. This null space can be geometrically conceptualized by noting that, when $n_{ens} < K$, the $K$-
dimensional $\mathbf{P}_e$ and $\mathbf{Y}_e$ data cannot be fully resolved by their respective $K n_{ens}$-dimensional ensemble vectors (HT08; GH08).

The second null space is attributable to the truncation of singular vectors of $\mathbf{P}_e$ with nonzero singular values necessary to ensure defined inverses and ER stability. The significance of this null space is related to the proportion of the variance of $\mathbf{P}_e$ that cannot be explained by the retained singular vectors. Given that $w_i = 20$ singular vectors are retained here, this proportion equals 35%, which may have significantly contributed to the discrepancies (GH08).

The significance of this null space is also related to the proportion of the cross-covariance that cannot be explained by the truncated predictor. A determination of singular vectors associated with small singular values that explain a significant percentage of cross-covariability is beyond the scope of this analysis, but should be addressed in further work.

Sampling and model errors are additional likely contributors to the differences between the DE and ER results. Because the ensemble is finite and susceptible to WRF model errors, ER covariances are necessarily misrepresentative of the true underlying covariances. Such misrepresentations may allow for spurious correlations between points whose true population covariances are negligible, resulting in spurious inverted height signals. Moreover, DE inversion dynamics are different than WRF dynamics (DE dynamics do not account for diabatic effects, for example) and therefore inversion results should not be expected to be identical (GH08).
5.7 The utility of PV ER

Although the previous sections have demonstrated that PV ER is a viable method for determining the heights associated with a PV perturbation, the primary role of the ER algorithm will likely not be as a replacement for the well documented and established DE inversion method. Instead, the utility of PV ER stems from its potential to regress fields...
with unknown analytical relationships and from the implications of treating PV perturbations as part of a correlated system of perturbations, not as independent entities, as in the DE technique. This section outlines several potential applications of PV ER and comments on the use of covariance localization in improving the PV ER technique.

The fact that the ER technique regresses an effective PV perturbation defined by the PV ensemble covariances, not the directly specified perturbation, may be undesirable if ER is used to implement “classical” PV inversion; ER may yield results that are inconsistent with classical PV thinking, as exemplified by the presence of a positive height perturbation in the southwest quadrant of the ER height perturbation results (figure 5.5a,c,e,g,i) and the corresponding absence of a collocated 300 hPa negative PV perturbation of significant strength in figure 5.4.

However, this characteristic is desirable if ER is used for PV predictability research. Past PV “surgery” (e.g. Roebber et. al 2002) and bogussing research (e.g. Leslie and Holland 1995) has attempted to deduce analyzed and forecast states corresponding to modified and supplemented vortex perturbations. Such studies have failed to recognize that the PV dynamics, as expressed by the covariance of the PV ensemble, dictate that altering the PV perturbation at one location implies corresponding alterations of the PV at other locations. Unlike classical PV inversion methods that treat PV perturbations in isolation, ER addresses this shortcoming by identifying a set of PV perturbations (the “effective” perturbation) that is statistically associated with the PV perturbations of interest. That is, the effective PV perturbation, not the directly specified perturbation, defines the most statistically likely PV configuration (with respect to the employed ensemble prediction system) that will occur if a PV perturbation of interest is
modified. Therefore, if one seeks the state associated with a modified perturbation, it is more appropriate to regress the effective PV perturbation via ER than invert the specified perturbation via DE inversion. Future research may find that the state attributable to the altered PV is drastically different than the state attributable to an effective PV perturbation that accounts for the covariance between the PV perturbation undergoing surgery and other correlated perturbations (HT08).

5.8 PV ER and covariance localization

It was mentioned in section 5.6 that sampling and model errors are likely contributors to inversion and regression errors. The problem of spurious signals due to sampling and model errors in the background covariance, $P^b$, is a well studied problem in the data assimilation community (Houtekamer and Mitchell 1998; Hamill et al. 2001). Assuming that the true covariances are distance-dependent (Hamill et al. 2001), the problem can be addressed by localizing $P^b$ by weighting the entries of $P^b$ by a function of the Euclidean distance between the two points that contribute to that entry of $P^b$ (Hamill et al. 2001). This can be accomplished by an element-by-element multiplication (Schur product) of $P^b$ with a separate correlation matrix, $S_k$, for each grid point, given by Equation 4.10 of Gaspari and Cohn (1999). Equation 4.10 is a quasi-Gaussian correlation function whose magnitude decreases monotonically from unity at the $k$th grid point to zero at a tunable distance, $l_e$. 
By performing a Schur product of the $k$th row of the analysis covariance with $S_k^5$, this same localization technique can be applied to the ensemble *analysis* covariance in order to mitigate the effects of spurious distance-dependent signals caused by sampling and model errors in ER. However, caution should be taken when estimating the population covariance structure because setting inappropriate localization parameters can unjustifiably preserve or eliminate covariance signals, resulting in spurious regressions. One justifiable method for estimating the true population correlation structure and the corresponding localization parameters is to calculate covariances multiple times with an “ensemble of ensembles” (e.g. Anderson 2004).

In addition to using it to mitigate the effects of sampling and model errors, covariance localization may also be used to impose the same effective perturbation on both the ER and DE techniques. It has been stated throughout this chapter that the fundamental difference between PV ER and DE PV inversion is that, because of covariances between the regressed PV and other PV perturbations, the PV perturbation effectively regressed via the ER method is not the regressed PV, but is instead the projection of this perturbation onto $\mathbf{P}_e'$. Future work could potentially localize $\mathbf{P}_e'$ so that the regressed PV perturbation would be forced to converge to the effective perturbation, thereby forcing ER to effectively regress $\tilde{\mathbf{p}}'$, not $\tilde{\mathbf{p}}^+$, as what occurs in the DE technique. However, it is imperative to reiterate that caution must be taken when estimating appropriate localization scales, since using inappropriate scales will unjustifiably set the extent of the regressed PV perturbation, resulting in spurious regressions.

---

$^5 S_k$ must be reshaped column-wise such that it is a $K \times 1$ vector.
5.9 Chapter summary

Following GH08, this chapter has presented the first rigorous application of statistical potential vorticity regression. Using flow-dependent covariances of WRF model output assimilated with an ensemble square-root filter, ER has enabled the statistical inference of the synoptic dynamical relationships between the PV and geopotential height fields from the covariability of fields’ ensemble analyses. It is shown that, if performed in the subspace of the leading PV singular vectors, the statistical ER technique yields height perturbations nearly identical to those resulting from the piecewise PV dynamical inversion technique of Davis and Emanuel (1991), if the same PV perturbation is regressed (inverted). Discrepancies between ER and DE result are attributed to null spaces and sampling and model errors.

PV ER is an unexplored method of studying synoptic meteorology with a potentially fruitful future. It is the view of the author that the demonstrated accuracy and versatility of ER PV regression will improve the understanding, modeling, and forecasting of cyclogenesis and other atmospheric processes in the years to come.
6.1 Introduction: Motivating non-contemporaneous ER

Following Gombos and Hansen (2008), the last chapter verified the validity of using ensemble covariances to form a regression operator by comparing the statistically predicted state using ER to the physically determined state using a dynamical model. By defining an operator with the covariances of potential vorticity (PV), potential temperature (PT), and geopotential height (Z) ensemble analyses, GH08 showed that, if the same effective perturbation is regressed (inverted, in the case of the dynamical model), piecewise PV regression and Davis and Emanuel (1991) dynamical piecewise PV inversion yield an almost identical Z perturbation.

In the case of contemporaneous (i.e. predictors and predictands are ensemble analyses and/or forecasts with the same initialization and lead times) piecewise PV regression, atmospheric dynamical theory supplies the necessary and sufficient set of predictors to nowcast collocated Z anomalies. That is, because inversion theory (e.g. Ertel 1942; Hoskins 1985; Davis and Emanuel 1991) asserts that only PV and PT boundary conditions are required to compute Z, PV and PT analyses account for all of the variance of height analyses and are perfect predictors of Z analyses, assuming the
dynamical relationship is balanced, adiabatic, inviscid, and linear and the ensemble is sufficiently large; as evidenced by the near perfect correspondence of the $Z$ yielded by the non-linear dynamical model of Davis and Emanuel (1991) and a one-hundred ensemble member ER (GH08), these can be valid assumptions for the purposes of ER.

However, most regressions involve imperfect predictor-predictand relationships and/or nonlinearities that potentially illegitimize the use of covariance-defined operators. The current chapter extends the work of GH08 by analyzing the utility of ER when the predictor-predictand relationship is imperfect. More specifically, via the non-contemporaneous (i.e. predictors and predictands are ensemble analyses and/or forecasts with the same initialization time, but different lead times) regression of a PV perturbation associated with a tropical cyclone, this chapter analyzes the utility of ER as a method for improving the understanding of dynamical mechanisms and as a means for forecasters to identify relevant forecast-specific dynamical processes. Via a contrived example, section 6.2 develops an ER dynamical analysis technique, which is then applied to diagnose geostrophic relationships in section 6.3 and then to the ER of a tropical cyclone PV perturbation using sophisticated operational Japanese Meteorological Association (JMA) ensemble data in section 6.5. Section 6.4 presents the experimental setup for the regression performed in section 6.5. Section 6.6 discusses how ER can be used for preemptive forecasting and how forecasters might make use of ER sensitivity information. Section 6.7 compares the multivariate ER sensitivity analysis discussed in this chapter to the univariate ESA sensitivity technique discussed in section 3.1. Section 6.8 provides chapter conclusions.
6.2 A contrived example of ER inference

There are a myriad of ways to make dynamical inferences from ER. One such method, which will be employed to analyze tropical cyclone track sensitivities to geopotential heights in section 6.5, is to analyze predictor and predictand anomaly patterns. This section will illustrate the use of ER anomaly patterns to infer physical relationships via a simple contrived system.

Consider the synthetically generated ensemble $P$ of size $[K \times n_{ens}]$. $P$ is comprised of two independently varying sinusoidal functions $B$ and $C$, such that

$$P = B + C,$$

where

$$B = -b \sin \left( \frac{\pi x}{n_x} \right) \cos \left( \frac{3\pi y}{2n_y} \right),$$

$$C = c \sin \left( \frac{\pi x}{n_x} \right) \sin \left( \frac{3\pi y}{2n_y} \right),$$

$b = N(10,1)$, $c = N(10,16)$, $x = 1,2,\ldots,n_x$, $y = 1,2,\ldots,n_y$, $n_x = n_y = 40$, $n_{ens} = 500$, and $K = n_xn_y$. $P$ can be thought of as any atmospheric field and $B$ and $C$ as the two independent modes of variability of $P$. The black dots in figures 6.1a and 6.1b mark the location of the maximum value of the ensemble mean of $P$.

Let $Y$ be representative of a field that covaries with $P$ whose functional form is given by

$$Y(P) = \exp \left( -f(x - d_x)^2 - g(y - d_y)^2 \right),$$

$$6.4$$
where \( d_x \) and \( d_y \) are the respective x and y coordinates of the maximum value of \( A \) for each ensemble member and \( f = g = 0.05 \). \( \mathbf{Y} \) is simply an ensemble of Gaussian functions in which each member is centered on the maximum value of the respective member of \( \mathbf{P} \). The black dots in figures 6.1c and 6.1d mark the location of the maximum value of the ensemble mean of \( \mathbf{Y} \).

Because they are orthogonal (ensuring the independence and uniqueness of the ERs), defined by linear combinations of \( \mathbf{P}_e \) (guaranteeing perfect resolvability (see section 3.4)) and capture the underlying dynamics of \( \mathbf{P}_e \) (indicating dynamical representativity (see section 3.4)), the perturbations chosen to be regressed in order to understand how fields \( \mathbf{P} \) and \( \mathbf{Y} \) relate via ER are the two leading principal components of \( \mathbf{P} \). These principal components, \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \), are defined as the columns of \( \mathbf{U} \), where

\[
\mathbf{U} = \mathbf{S} \mathbf{V}^T, 
\]

and \( \mathbf{S} \) and \( \mathbf{V} \) denote the diagonal matrix of singular values and the right singular matrix, respectively, of \( \mathbf{P}_e \), and are depicted in figures 6.1a and 6.1b, respectively. Note that, as expected given the underlying functions of \( \mathbf{P} \), the patterns of the ensemble means of \( \mathbf{B} \) and \( \mathbf{C} \) (not shown) are nearly identical to those of \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) (figures 6.1a and 6.1b).

Note that eigenvectors are defined up to a sign and therefore the signs of \( \mathbf{P} \) need not be the same as the signs of the patterns in \( \mathbf{B} \) and \( \mathbf{C} \). Because \( \mathbf{u}_1 \) (\( \mathbf{u}_2 \)) depicts a dipole in which the maxima is located to the north (south) of the maxima of the mean value of \( \mathbf{P} \) (marked by the dots in figures 6.1c and 6.1d) and the minima is located to the south (north) of the maxima of the mean value of \( \mathbf{P} \), \( \mathbf{u}_1 \) (\( \mathbf{u}_2 \)) represents a perturbation of \( \mathbf{P}_e \) located anomalously north (south) of the ensemble mean. Knowing the functional form
of \( P \), it is straightforward to predict the state of \( Y \) associated with perturbations \( u_1 \) and \( u_2 \). Because the maxima of \( P \) gives the location of the maxima of \( Y \) (viz. 6.4), the state of \( Y \) dynamically associated with \( u_1 \) (\( u_2 \)) is a Gaussian function centered anomalously north (south) of the ensemble mean location marked by the dot in figures 6.1c and 6.1d. Therefore, the ensemble anomaly pattern of \( Y \) associated with \( u_1 \) (\( u_2 \)) is a dipole with positive values to the north (south) and negative value to the south (north) of the location of the mean.

However, if the analytical equation 6.4 is unknown, understanding the relationship between \( P \) and \( Y \) may not be so straightforward, but can be inferred using ER. Let \( L \) compute the changes from the mean state of \( Y \), \( y_{u_1} \) and \( y_{u_2} \) (figures 6.1c and 6.1d, respectively), with which \( u_1 \) and \( u_2 \) are respectively statistically associated, such that

\[
y_{u_1} = Lu_1 \tag{6.6}
\]

and

\[
y_{u_2} = Lu_2 \tag{6.7}
\]

As expected from the governing equations of the dynamics, the dipole perturbation depicted in Fig 6.1c (6.1d) indicates that \( u_1 \) (\( u_2 \)) is statistically associated with a perturbation of \( Y \) located to the north (south) of the ensemble mean of \( Y \). That is, without use of the governing equation, ensemble statistics indicate that a perturbation of \( P \) located north (south) of average is associated with a perturbation of \( Y \) located north (south) of average. Although it is strictly inappropriate to conclude from statistical
A Contrived Example of ER Inference from Anomaly Patterns

Figure 6.1: (a). $u_1$, the first principal component of $P$ (viz. 6.1 and 6.5). (b). $u_2$, the second principal component of $P$ (viz. 6.1 and 6.5). (c). The predictand perturbation, $y_{u_1}$, resulting from the ER of the perturbation from panel a. (d). The predictand perturbation, $y_{u_2}$, resulting from the ER of the perturbation from panel b. The black dot in all panels marks the location of the maximum value of the ensemble mean of $Y$. The tip of the arrowhead is located at the location of maximum value of $u_1$ (panel a) and $u_2$ (panel b). The direction of the arrow indicates the direction of the shift of the perturbation of $Y$ relative to the location of the mean value of $Y$ (denoted by the black dot).

evidence that the location of $Y$ is a function of the location of $P$, ER has certainly supplied evidence supporting this claim.
6.3 Geostrophy ER

Before applying the notion of ER anomaly patterns for the sensitivity analysis of tropical cyclone tracks to geopotential heights in section 6.5, this section uses the simple geostrophic relationship to show how ER anomaly patterns can verify atmospheric theorems through observational data, or even potentially suggest new dynamical relationships that can subsequently be analytically verified.

In isobaric coordinates the zonal geostrophic wind is given by

$$u_g = \frac{g}{f} \frac{\partial Z}{\partial y}$$  \hspace{1cm} (e.g. Holton 1992), \hspace{1cm} (6.8)

where $y$ denotes the meridional distance. Therefore, under geostrophic conditions of an approximately equal Coriolis and pressure gradient force, a westerly wind blows at the location of a negative $Z$ gradient.

This property can be shown using ER using only ensemble model output and with no prior knowledge of the dynamical relationship. Let $L$ be the ER operator formed using the leading ten principal components of the ensemble anomalies of the JMA 12UTC 14 August 2007 6 hour forecast 300 hPa geopotential heights as the predictor and the leading 10 principal components of the ensemble anomalies of the JMA 14 August 2007 12UTC 6 hour forecast 300 hPa zonal wind as the predictand. Let the dynamically representative perturbation of interest be the negative meridional gradient of $Z$, depicted by the filled contours of figure 6.2. Given this negative meridional gradient of $Z$, the most probable configuration of the 300 hPa zonal wind is determined by applying this perturbation to the ER operator. The line contours of figure 6.2 depict the associated zonal winds as westerlies whose maximum magnitude is collocated with the maximum...
negative Z gradient. Note that lighter shades represent positive values and darker shades represent negative values.

With prior knowledge of geostrophy, figure 6.2 simply confirms that the ensemble model output corroborates the physical expectations of the relationship between the zonal winds and the gradient of the geopotential heights. Through ER model...
intercomparison and by comparing model ER relationships to those from physical theory, this analysis can potentially be useful for identifying model errors and differences for more complicated atmospheric phenomena. Moreover, without prior knowledge of geostrophy, the relationship depicted by figure 6.2 can be used to identify physical relationships between the zonal wind and Z fields that can potentially be used to uncover the geostrophic relationship, or other currently unrealized relationships. That is, because ER uncovered the geostrophic relationship expressed by the covariances of the wind and geopotential height fields, ER can certainly be used to statistically reveal clues about other physical field relationships unknown a priori.

6.4 Experimental motivation and design

Although atmospheric forecasts are inherently probabilistic, public demand, political pressures, and evacuation preparations often call for the issuance of deterministic forecasts. Operational forecasters use many tools to aid them in choosing a single best forecast, including ensemble means, multi-model consensuses, and operational “experience”, but forecasts are sometimes ultimately subjective. Figure 6.3 depicts the 72-hour JMA ensemble track forecasts (initialized on 12 UTC 14 August 2007) of the 1000 hPa PV associated with Supertyphoon Sepat, a category five tropical cyclone that devastated parts of China; a forecaster given this highly uncertain forecast guidance would undoubtedly struggle to definitively decide on a single best forecast track. This chapter illustrates how ER can help objectify this decision process by identifying elements of the state on which Sepat’s track is highly dependent and how ER can be used by researchers a posteriori to understand Sepat’s track dynamics.
Although Sepat’s track is undoubtedly sensitive to many atmospheric variables, this chapter focuses on track sensitivities to mid-tropospheric Z. This section discusses the selection of the predictor and predictand ensembles and perturbations that a forecaster or researcher could potentially use to analyze this sensitivity via ER. The section concludes by comparing ER to singular vector analysis.

Figure 6.3: The $n_{ens} = 50$ ensemble forecast tracks of Supertyphoon Sepat, initialized on 12 UTC August 14 2007. Grey lines depict the forecast track of each ensemble member out to 72 hours. The black dots represent the actual best track of Sepat every three hours. The thick black line represents the coastline for purposes of determining the forecast average landfall time.
The ER predictor and predictands used here to analyze the sensitivities and
dynamics of Sepat’s track to mid-tropospheric Z are 1000 hPa PV and 500 hPa Z,
respectively. Despite the likelihood that additional inferences about this sensitivity can
be made by using additional state elements to define the regression operator, the choice
of using only 1000 hPa PV and 500 hPa Z is made for several reasons. Firstly, PV
inversion theory (e.g. Ertel 1942; Hoskins 1985; Davis and Emanuel 1991) states that PV
and Z are dynamically linked, thereby ensuring the statistically significant covariances
required for ER. Secondly, exploring non-contemporaneous PV and Z ER is a natural
extension to the contemporaneous ER performed in GH08 and chapter 5; seen in that
light, this chapter is an exploration of the time scales over which the PV invertibility
relation holds. Thirdly, choosing two specific fields simplifies the presentation and
analysis of ER results. It is crucial to note that, although only two variables are directly
involved in the regression, all variables correlated to the 1000 hPa PV are effectively
involved (see section 4.2).

The sensitivity of Sepat’s track to the 500 hPa Z is explored using a backward-in-
time ER. An ER operator is defined using predictor 72-hour 1000 hPa PV ensemble
forecasts and predictand 500 hPa Z ensemble forecasts, both initialized at 12 UTC 14
August 2007. Note that the lead-time of the Z forecasts varies and will be specified for
each ER in section 6.5. Also note that 72 hours is the ensemble average forecast time of
potential landfall at Taiwan (for those ensemble members that are forecast to strike
Taiwan) and therefore the predictor ensemble represents the 1000 hPa signature of the
cyclone at landfall. The coastline for the purpose of defining the time of landfall is
defined as the thick black line in figure 6.3. Thus, by defining the ER operator with a
predictor ensemble of forecast states further into the future than that of the predictand ensemble, it is possible to determine landfall location sensitivities to pre-landfall Z.

Before using them to make meaningful dynamical inferences from ER, it is necessary to assess whether this set of predictors and predictands provides sufficiently good forecasts, so that ER predictand anomaly patterns can be considered probable perturbation estimates. Using the predictor and predictand ensembles defined in the previous paragraph, leave-one-out cross validation (e.g. Wilks 2006) is employed here to assess the likely goodness of the rank-deficient \( n_{\text{ens}} = 50 \) and \( K = 2562 \) Sepat track ERs in section 6.5 and to choose the optimal numbers of predictor and predictand principal components for the regression. Refer to section 4.4 on details of this procedure.

Figure 6.4 shows the ensemble median ACC as a function of predictand lead time, for the optimal values of \( n_{\text{pc}}^{\text{and}} \) and \( n_{\text{pc}}^{\text{or}} \) at each lead time (in parentheses). Figure 6.4 indicates that the median ACC increases as the difference between the predictor and predictand lead times decreases; recalling that the predictor is the 72-hour forecast PV, ER predictions improve as the predictand and predictor ensembles become more contemporaneous. Given that the predictand ensemble becomes less Gaussian with lead time (not shown), this result implies that the increase in the strength of the fields’ dynamical relationship as the difference in lead times decreases influences the goodness of the ER prediction more significantly than does the decrease in Gaussianity of the predictand. Figure 6.4 also illustrates that, in general, fewer predictor principal components than predictand principal components optimally regress the 1000 hPa PV perturbations; typical optimal predictand principal component numbers range between

135
and \( n_{pc}^{and} = 15 \) and \( n_{pc}^{or} = 20 \) whereas typical optimal predictor principal component numbers range between \( n_{pc}^{or} = 10 \) to \( n_{pc}^{or} = 15 \).

The most significant finding portrayed by figure 6.4 is that the ACCs using the optimal \( n_{pc}^{and} \) and \( n_{pc}^{or} \) combinations range between 0.38 and 0.64. These statistically significant ACCs suggest that, although some subtle features of the predictand fields may be inaccurately forecast, ER is highly capable of capturing the majority of the gross features. It is crucial to note that, since some of the error is attributable to poor resolvability of the predictor perturbation, and given that the actual perturbations chosen to be regressed in the Sepat track sensitivity ER in section 6.5 are nearly perfectly resolvable (as explained below), the ER predictions for the Sepat track sensitivity are likely to be good relative to the cross validated ER predictions. Also note that ER prediction accuracy would likely improve with the addition of qualified predictors, but this analysis has been confined to just 1000 hPa PV and 500 hPa Z for the aforementioned reasons.

The PV perturbations chosen to be regressed for the Sepat track sensitivity ER are the second and third principal components of the filtered 72-hour 1000 hPa PV ensemble forecast field. Note that the field was pre-filtered to isolate Sepat before the computation of the principal components by zeroing out all points outside of a five degree radius of the cyclone PV maximum. These prescribed perturbations (not shown) are both highly resolvable, as they nearly identically match the perturbations effectively resolved by the predictor ensemble, and dynamically representative (see section 3.4), as can be understood by noting the form of their associated eigenvectors (figures 6.5a and 6.5b, respectively). Fig 6.5a depicts a dipole with the positive (negative) anomaly lobe.
Figure 6.4: This plot shows the ensemble median ACC as a function of predictand lead time, for the optimal values of $n_{pc}^{and}$ and $n_{pc}^{or}$ at each lead time (in parentheses, ordered as $n_{pc}^{and}$ and then $n_{pc}^{or}$). That is, the plot presents the expected ACC at each lead time when the ER is performed using the optimal numbers of principal components. See text for more details.

positioned north (south) of the ensemble mean position of the 72-hour forecast 1000 hPa PV maxima (denoted by the black dot) and figure 6.5b depicts a dipole with the positive (negative) anomaly lobe positioned west (east) of this ensemble mean position. For the purposes of an eigen-analysis, these dipoles respectively imply that the second and third most variable direction in the ensemble state space corresponds with the uncertainty in
the meridional and zonal positions of the cyclone. But, for the purposes of ER, given that
the perturbations are comprised of linear combinations of ensemble anomalies, the

Figure 6.5: (a). The second leading eigenvector of the filtered JMA 72-hour forecast
1000 hPa PV initialized on 12 UTC 14 August 2007. The regressed perturbation for the
Sepat track sensitivity ER is the second principal component of the predictor ensemble
(the projection of the predictor ensemble onto this eigenvector). (b) The third leading
eigenvector of the filtered JMA 72-hour forecast 1000 hPa PV initialized on 12 UTC 14
August 2007. The regressed perturbation for the Sepat track sensitivity ER is the third
principal component of the predictor ensemble (the projection of the predictor ensemble
onto this eigenvector).
dipoles simply represent, respectively, a cyclone located anomalously north and west of
the ensemble mean location. The fields yielded from ERs of these perturbations thus
represent the most probable states of 500 hPa Z anomalies when the eventual cyclone is
located north and west, respectively, of the ensemble mean location at 72 hours, making
the dipoles ideally dynamically representative for track sensitivity ERs.

It is worth noting that, given that the predictor perturbation is a principal
component, the ER described here can be considered a variation of a singular vector (SV)
analysis (e.g. Kalnay 2003). Traditional singular vector analysis involves an initial-time
SV (i.e. the right SV of a tangent linear model formed from the entire state vector) that
evolves into the final-time SV (i.e. the left SV of a tangent linear model formed from the
entire state vector), the most highly varying direction of the final state (e.g. Kalnay 2003;
Palmer et al. 1998). For this ER experiment, on the other hand, the predictands are the
states of the 500 hPa Z, only, that are statistically associated with the second and third
most highly varying final-time states of the 1000 hPa, only; unlike traditional initial and
final SVs, the predictands and predictors in this experiment are not comprised of the
entire state vector and, more importantly, there is no implication that the initial-time
predictands evolves into the final-time predictors. Instead, it can only be asserted that the
initial predictand perturbations are the most statistically probable 500 hPa Z
perturbations, given that the states of the final-time 1000 hPa PV field are equal to the
that of the resolved regressed perturbations.
6.5 Sepat track ensemble regression results and analysis

The method of using anomaly patterns from dynamically meaningful predictor perturbations and statistically significant predictand perturbations used in section 6.2 is applied here to a $n_{ens} = 50$ operational JMA ensemble to study Typhoon Sepat track sensitivities. Note that JMA ensemble data was retrieved from the Thorpex Interactive Global Grand Ensemble data archive (e.g. Bougeault et. al, 2008). The ensembles used for this regression are the same as those described and cross validated in section 6.4 (e.g. predictor 1000 hPa PV and predictand 500 hPa Z) and therefore the ER predictand patterns are expected to be statistically significant. In fact, because of the high resolvability of the regressed perturbations, the ER results are expected to exceed the 0.35 and 0.65 ACC range presented in figure 6.4.

The following discussion is intended to outline dynamical inferences and sensitivity assessments that can be made using ER. These ideas can potentially be used by researches a posteriori to understand what may have contributed to changes in forecast tracks and can also be used a priori by operational forecasters to identify the most relevant features that influence the track, thereby determining the “forecast problem of the day”. Note that the following analysis uses only forecast, not verification information, and so all necessary data is available in real-time.

Figure 6.6a shows the results of the regression of the perturbation associated with the eigenvector in figure 6.5a using the 12 UTC 14 August 2007 6-hour forecast 500 hPa Z predictand ensemble and $n_{pc} = 15$ and $n_{pc}^{or} = 15$ (viz. figure 6.4). That is, figure 6.6a depicts the most probable state of the 500 hPa Z 6-hr forecast given that the 1000 hPa PV
signature of the 72-hour (e.g. ensemble average forecast landfall time) forecast cyclone is
located anomalously north of the ensemble mean location. Note that the 6-hour Z
forecasts are used rather than the analysis because analysis perturbations need not be
balanced, are not necessarily random draws from the analysis PDF, and because
forecasters would typically be interested in future, rather than current, sensitivities in
order to adjust forecasts. Filled contours show ER anomaly patterns; lighter (darker)
colors depict ER anomaly values greater (less) than zero. The ER anomaly zero line is
labeled as a thick black dashed line. Labeled line contours depict ensemble mean 12
UTC 14 August 2007 6-hour forecast 500 hPa Z values in decameters. Irregular black
tables represent coastlines and bold black boxes, labeled with numbers in the lower right
corners, isolate specific patterns of interest.

Box 1 of figure 6.6a illustrates a straightforward statistical representation of
dynamical persistence. The line contour of Box 1 of figure 6.6a represents the ensemble
mean position of the 500 hPa Z signature of Sepat. The negative anomaly to the
northeast of this position illustrates that anomalously northern Sepat tracks are likely
preceded by anomalously northeastern 500 hPa low Z. This implies that ensemble
members whose initial condition mid-tropospheric height signatures of Sepat are
northeast of average are likely to have associated surface PV features that remain north of
average as the cyclone is integrated by the model. The fact that the Z lobe is northeast,
not north, of the mean location may imply that 1) a westward force steers anomalously
northeastern cyclones to the west, or 2) the 1000 hPa PV signatures are west of their 500
hPa Z counterparts, or 3) there is an ER error due to an insufficient ensemble size or
retained number of singular vectors. Note, also, that the lack of a significant positive
anomaly lobe to the southwest of the ensemble mean location indicates that forecast
southern Sepat tracks have no statistical correspondence with southward 500 hPa Z
analyses.

Box 2 of figure 6.6a depicts a strong trough to the northwest of Sepat and three
significant ER anomaly signals. Firstly, the negative ER anomaly along the southeastern
edge of the trough implies that deepening the trough (by further decreasing the heights
along the southeastern edge via a southeastward shift of the trough) is associated with a
northward typhoon track. This statistical result is very consistent with physical
expectations; the increased and southward shifted counterclockwise circulation associated
with a strengthened and southeastward displaced trough would increase the southerly
steering component controlling Sepat’s motion. Secondly, the negative ER anomaly
along the western and southern fringes of the trough reflects the sensitivity of the landfall
location to the overall size of the trough; a broader and larger trough is associated with an
increased southerly steering current. Thirdly, the positive ER anomaly in the northeast
corner of box 2 indicates that an increase in the heights along the northeastern section of
the trough, possibly via the negative tilting of the heights, is associated with a northward
storm track. Note that decreasing the heights on the western and southeastern edges of
the trough and increasing the heights on the northeastern edge of the trough would result
in a negative tilting of the trough axis. Therefore, together, the three ER anomalies of
box 2 jointly imply that Sepat’s meridional landfall location is sensitive to the tilt and
extent of the trough.

The line contours of box 3 of figure 6.6a show the strong subtropical ridge high Z
center to the northeast of Sepat. To the east of this high are strong positive ER contours
a. 6 hour height field associated with the second leading 1000 hPa PV eigenvector

b. 6 hour height field associated with the third leading 1000 hPa PV eigenvector
Figure 6.6: (a). Results of the ensemble regression of the perturbation associated with the eigenvector in figure 6.5a using the 12 UTC 14 August 2007 6-hour forecast 500 hPa Z predictand ensemble and $n_{pc}^{and} = 15$ and $n_{pc}^{or} = 15$ (viz. figure 6.4). Filled contours show ER anomaly patterns; lighter (darker) colors depict ER anomaly values greater (less) than zero. The ER anomaly zero line is labeled as a thick black dashed line. Labeled line contours depict ensemble mean 12 UTC 14 August 2007 6-hour forecast 500 hPa Z values in decameters. Bold black boxes, labeled with numbers in the lower right corners, isolate specific patterns of interest. (b). Same as (a), except the regressed perturbation is the eigenvector in figure 6.5b.

and to the southwest are weak negative ER contours, indicating that increased (decreased) Z to the east (southwest) of the high Z center are statistically associated with anomalously northern cyclone track forecasts. Therefore, it is probable that the 72-hr forecast surface cyclone makes landfall anomalously north of average if the location of the subtropical high in the initial conditions is shifted anomalously east of the ensemble mean. This statistical result is also very consistent with physical expectations; strong blocking high pressure centers located directly north of cyclones are likely to inhibit northward cyclone movement, whereas eastward shifted blocking highs are less likely to impede the natural northward trajectory.

Figure 6.6b shows the results of the regression of the perturbation associated with the eigenvector in figure 6.5b using the 12 UTC 14 August 2007 6-hour forecast 500 hPa Z predictand ensemble and $n_{pc}^{and} = 15$ and $n_{pc}^{or} = 15$ (viz. figure 6.4). That is, figure 6.6b depicts the most probable state of the 500 hPa Z 6-hour forecast given that the 1000 hPa PV signature of the 72-hour (e.g. ensemble average forecast landfall time) forecast cyclone is located anomalously west of the ensemble mean location.
Box 1 of figure 6.6b is the zonally-oriented counterpart to the meridionally-oriented result from box 1 of figure 6.6a; figure 6.6b indicates that a 6-hour forecast of the 500 hPa Z signature of Sepat located anomalously west of the ensemble average is, as expected, associated with an anomalously western landfall location at 1000 hPa.

Box 2 of figure 6.6b depicts a strong positive ER contour on the southeastern edge of the high Z center to the northeast of Sepat. This area represents one of the most uncertain features of the 12 UTC 14 August 2007 6-hour forecast; several ensemble members showed the emergence of a tongue of high pressure jutting southeast of the high Z center (as resolved in box 2), whereas others maintained a more circular structure. The strong positive ER contour in box 2 indicates that those ensemble members with a southeasterly movement or extension of the high Z center are strongly associated with anomalously western 1000 hPa landfall locations. Box 2 certainly represents a highly dynamically important location for Sepat’s easterly steering component at the 6-hour forecast time.

The high pressure collocated with weak negative ER contours in box 3 of figure 6.6b indicates that a weakening of the high Z to the west of Sepat is statistically associated with an anomalously western track. This too is consistent with physical expectations; the weakening of the westerly steering current to the north of the high and to the west of Sepat would result in an increase in Sepat’s easterly steering current.

The positive ER contours to the north and northwest of Sepat in box 4 of figure 6.6b generally imply that increasing the geopotential heights in this area results in an anomalously western Sepat track. More specifically, this swath of positive ER contours may be a combination of two separate signals. The first signal signifies that a weakened
trough to the northwest of Sepat may reduce the westerly steering flow to the northwest of Sepat, resulting in an anomalously western track. Also consider, however, that this signal may be a consequence of the coupling of the trough tilt and extent with the *meridional* landfall position. Because, as aforementioned, a negatively tilted and anomalously strong trough increases the southerly steering current, it also effectively reduces the westward displacement of the cyclone (and therefore a positively tilted and anomalously weak trough may increase Sepat’s westward displacement, as is depicted in box 4 of figure 6.6b). Therefore, this feature, along with others, may not directly impact the cyclone’s steering flow; instead, it may covary with the landfall location only because it is coupled with a feature that does directly impact the steering flow.

The second signal depicted in box 4 of figure 6.6b is the positive ER anomaly swath to the southwest of the high Z center located to the northwest of Sepat. This signal likely implies that the southwestward extension or displacement of the subtropical high corresponds to an anomalously western cyclone track. Because a southwestward displaced or extended high pressure center would shift the easterlies to the south of the high closer to the north of Sepat, thereby strengthening Sepat’s easterly steering current, these ER contours are consistent with physical expectations.

Figure 6.7a shows the results of the regression of the perturbation associated with the eigenvector in figure 6.5a using the 12 UTC 14 August 2007 60-hour forecast 500 hPa Z predictand ensemble and $n_{pc}^{md} = 20$ and $n_{pc}^{or} = 15$ (*viz.* figure 6.4). That is, figure 6.7a depicts the most probable state of the 60-hour forecast 500 hPa Z given that the 1000 hPa PV signature of the 72-hour (*e.g.* the ensemble average forecast landfall time) forecast cyclone is located anomalously *north* of the ensemble mean location.
Box 1 of figure 6.7a shows the forecast 500 hPa Z signature of Sepat 12 hours before ensemble mean landfall time. The strong north to south oriented dipole on the meridional fridges of the 500 hPa Z typhoon signature simply reflects forecast track persistence; model integrations that take Sepat anomalously north at forecast hour 60 are likely to be the same as those that take Sepat anomalously north at forecast hour 72.

Box 2 of figure 6.7a illustrates an inverted ridge, which evolved from the high Z tongue in box 2 of figure 6.6b, to the west of an inverted trough. The positive ER contours within this inverted ridge are consistent with expectations; the increased counterclockwise flow associated with a strengthening of the southeastern ridge extension acts to significantly influence Sepat’s southerly steering current. By inspecting the individual ensemble 500 hPa Z (not shown), it is evident that only some ensembles forecast the development of this ridge extension and the associated inverted trough. Judging from the strength of the ER signal associated with the extension, it can be inferred that this feature is (or is correlated with a factor that is) significantly and directly impacts Sepat’s ultimate meridional position.

Box 3 of figure 6.7a isolates the northwestern section of the subtropical high Z center to the north of Sepat. The strong positive ER contours to the northwest of this high Z center imply that a northwestward shift of this high is associated with an anomalously northward landfall location. Noting that the strong high Z center acts to block the natural northward trajectory of Sepat, this result matches expectation considering that the northwestward movement of the high might delay the blocking until the typhoon reaches a more northern location.
a. 60 hour height field associated with the second leading 1000 hPa PV eigenvector

b. 60 hour height field associated with the third leading 1000 hPa PV eigenvector
Figure 6.7: (a). Same as 6.6a, except the predictand ensemble is the 12 UTC 14 August 2007 60-hour forecast 500 hPa Z and $n_{pc}^{end} = 15$ and $n_{pc}^{or} = 15$ (viz. figure 6.4). (b). Same as 6.6b, except the predictand ensemble is the 12 UTC 14 August 2007 60-hour forecast 500 hPa Z.

Figure 6.7b shows the results of the regression of the perturbation associated with the eigenvector in figure 6.5b using the 12 UTC 14 August 2007 60-hour forecast 500 hPa Z predictand ensemble and $n_{pc}^{end} = 20$ and $n_{pc}^{or} = 15$ (viz. figure 6.4). That is, figure 6.7b depicts the most probable state of the 60-hour forecast 500 hPa Z given that the 1000 hPa PV signature of the 72-hour (i.e. ensemble average forecast landfall time) forecast cyclone is located anomalously west of the ensemble mean location.

Box 1 of figure 6.7b depicts a weak southwest to northeast dipole collocated with the weak low to the west of Sepat. The dipole implies that ensembles with lows positioned farther southwest of average also tend to have Sepat landfall locations farther west of average. This statistical result matches physical expectations; a low centered anomalously southwest and farther from Sepat would have a weaker westerly steering influence on its southern edge than one located anomalously northeast and closer to Sepat.

Box 2 of figure 6.7b shows the forecast 500 hPa Z signature of Sepat 12 hours before ensemble mean landfall time. The strong zonally oriented dipole on the north and south fridges of the 500 hPa Z signature reflect that model integrations that take Sepat anomalously west at forecast hour 60 are likely to be the same as those that take Sepat west of average at forecast hour 72.
Box 3 of figure 6.7b indicates that the westward motion of Sepat is statistically associated with the position of the inverted trough to the east of the southeastern subtropical ridge extension; as shown by the southwest to northeast oriented dipole, the flattening of this trough is associated with a westward movement of Sepat. Because of the intervening subtropical ridge extension and the relatively far distance of this trough from Sepat, it is unlikely that the trough directly influences the steering flow. Instead, this dipole likely indicates that the position of the ridge to the north, which certainly does directly influence Sepat’s zonal steering flow, is coupled with the position of the trough.

Box 4 of figure 6.7b isolates the northern and western fringes of the subtropical ridge to the north of Sepat. Firstly, the negative ER pattern to the north of the ridge simply indicates that if the ridge were to shift southward, the ridge would have a stronger influence on the steering current, thereby forcing the track westward. Secondly, the meridionally oriented ER dipole on the western fringe of the subtropical ridge indicates that anomalously strong (weak) southern (northern) Z at the western fringe of the subtropical high corresponds with an anomalously westward Sepat track. Because the southern section of the dipole is at approximately the same latitude as the center of the subtropical ridge, this dipole implies that a zonally oriented, rather than slightly negatively tilted (as would be the case if the sign of the dipole was reversed), Z is correlated with a western track. This statistical pattern matches physical expectations, considering that a more zonally oriented high would have an increased easterly steering current to the north of Sepat.

The preceding discussion has illustrated that qualitative statistical conclusions computed from ER are consistent with physical expectations and, more importantly,
without the need for an a priori understanding of the system dynamics, ER has identified and ranked the features most strongly coupled with Sepat’s track at different lead times. Forecasters could potentially use these ER maps in a real-time forecasting situation to infer, for example, that the extent of the southeastern extension of the subtropical ridge at the 6-hour forecast time (box 2 of figure 6.6b) has a very significant relationship with Sepat’s zonal landfall position. If, for example, other models’ ensembles uniformly extend this ridge well to the southeast, a forecaster attempting to issue a single best track forecast could accordingly reduce the likelihood of the eastern JMA landfall solutions. Refer to section 6.6 for more discussion on the use of ER sensitivity information for forecasting.

6.6 ER preemptive forecasting

ER sensitivity information is not only valuable to researchers aiming to understand model dynamics and differences, it can also be extremely valuable to forecasters through its use for preemptive forecasting. Preemptive forecasts (e.g. Etherton 2007), use computationally inexpensive flow-dependent lagged ensemble covariances, in conjunction with recently available analyses, to issue forecasts in advance of the completion of the relatively time-consuming nonlinear operational model run. Given that the error propagation via ensemble statistics takes seconds to perform, whereas the nonlinear ensemble integration of the operation model takes hours, preemptive forecasts are available hours before the operational forecasts are available.
In EnKF preemptive forecasting (Etherton 2007), ensemble covariance statistics from a previous ensemble forecast (launched at a past time \( t = \eta \) and first available at the current time \( t = 0 \)) are used to propagate forward the analysis increment from the most recent observations (assimilated at \( t = 0 \)) to produce a statistical forecast valid at some future verification time \( t = \nu \). Similarly to the method proposed by Etherton (2007), preemptive forecasts can be performed using ER. Let

\[
L_{\eta, \nu} = Y_{\eta, \nu}^{-1},
\]

where the first subscripted argument denotes the time of initialization and the second subscripted argument denotes the time of validity. That is, \( L_{\eta, \nu} \) represents an ER operator from a past time that associates the predictor, which is initialized in the past and valid at the present, to a predictand, which is initialized in the past and valid in the future. Let \( P_{0,0} \) represent a new ensemble of analyses and let \( P_{0,0}^{\eta} \) represent analysis ensemble anomalies with respect to the ensemble mean of \( P_{\eta,0} \). That is, \( P_{0,0}^{\eta} \) represents updated perturbations that incorporate new assimilated observation data, but are perturbations defined with respect to the ensemble mean used to define \( L_{\eta, \nu} \). These and other preemptive forecasts have value if, over a statistically significant sample of cases, they are more similar to the future currently unavailable operational forecast than the most recently available operational forecasts are to the future currently unavailable operational forecasts. That is, if the mean and distribution of

\[
\hat{Y}_{0, \nu} = L_{\eta, \nu} P_{0,0}^{\eta}
\]

are more similar to that of \( Y_{0, \nu} \), than that of \( Y_{\eta, \nu} \) is to that of \( Y_{0, \nu} \), then preemptive ER forecasting has value, since the preemptive ER forecast therefore provides a better
Figure 6.8: A timeline for ER preemptive forecasting. Note that any times can be used, not just the ones labeled here.

estimate of the most probable future state than does any forecast currently available.

Figure 6.8 shows an example timeline for preemptive ER forecasting, in which \( \eta = -6 \) hrs and \( v = 60 \) hrs with respect to the current time \( t = 0 \).

Preemptive ER forecasts are performed on the National Center for Environmental Prediction’s (NCEP) Supertyphoon Sepat tracks. Let \( t = \eta \) be 12 UTC 14 August 2007 and \( t = 0 \) be 18 UTC 14 August 2007. Let the predictor field be the 1000 hPa PV and the predictand field be the 1000 hPa PV. Figure 6.9 displays three ellipses for each of two verification times, 06 UTC 16 August 2007 and 06 UTC 17 August 2007. The dashed ellipse depicts the one standard deviation PDF of the locations of the forecast 1000 hPa PV from \( Y_\eta,v \) and the x depicts the ensemble mean. The solid (dotted) and
solid (circle) dots respectively represent the one standard deviation PDF and means for the forecast 1000 hPa PV from \( \hat{Y}_{0,v} \) and \( Y_{0,v} \), respectively.

It is clear that, in the case of typhoon Sepat and this particularly set of predictors and predictands, that ER preemptive forecasting has significant value to forecasters. For both verification times explored, the preemptive ER forecasts (solid ellipses and black dots) are more similar to the future operational forecasts (dotted ellipses and circle) than the old operational forecasts (dashed ellipses and xs) are to the future operational forecasts. Assuming that forecasts always improve as lead times decrease, the relative similarity of the ER forecast to the future forecast indicates that the ER forecast offers the best forecasts available at the current time. Given that operational ensemble forecasts can often take several hours to run, whereas ER forecasts take only seconds to complete, ER forecasts supply forecasters and emergency preparation decision makers with forecasts that utilize the most recently available observation data several hours in advance of the operational models that make use of this information.

It is crucial to note, however, that given the probabilistic nature of this technique, one case study certainly does not validate the use of ER preemptive forecasting; the Sepat results presented here merely represent a promising preliminary result and a proof-of-concept of ER preemptive forecasting that should be validated with a significantly greater sample size of cases. Given a statistically significant sample of preemptive forecasts, the reliability of the ensemble distributions can be assessed using the Mahalanobis-normed minimum spanning tree rank histogram (see chapter 2; Gombos and Hansen 2008; Wilks...
Figure 6.9: Preemptive ER forecasting results for cyclone Sepat. Predictors and predictands are the 1000 hPa PV. $t = \eta$ is 12 UTC 14 August 2007 and $t = 0$ is 18 UTC 14 August 2007 and two separate verification times, 06 UTC 16 August 2007 and 06 UTC 17 August 2007, are displayed. The dashed ellipse depicts the one standard deviation PDF of the locations of the forecast 1000 hPa PV from $\hat{Y}_{\eta,v}$ and the $x$ depicts the ensemble mean. The solid (dotted) ellipses and solid (circle) dots respectively represent the one standard deviation PDF and means for the forecast 1000 hPa PV from $\hat{Y}_{0,v}$ and $Y_{0,v}$, respectively.
One could define the ensemble mean of the future forecast as the verification data point and assess, over a large sample, whether this verification point is more likely to be a member of the ER forecast distribution or the old operational forecast distribution via the MST RH CvM goodness of fit statistics. Alternatively, an ensemble of rank histograms (or a PDF rank histogram) can be computed that uses each member of the future forecast distribution as a verification point for the MST RH analysis.

Although quantitative ER preemptive forecasting, as validated by an MST RH analysis, can provide valuable information to a forecaster, it is likely that the most practical application of ER preemptive forecasting comes as qualitative sensitivity guidance. Given old ER sensitivity information and newly available observations or analyses, a forecaster can perform a qualitative estimate of the impact of new observations on forecasts. For example, knowing from figure 6.6b that the development of a tongue of anomalously high heights along the southern edge of the subtropical high for the 12 UTC 14 August 2007 6-hour JMA forecast is highly correlated with an anomalously western Sepat track forecast, a forecaster might qualitatively put increased confidence in the western Sepat track forecasts if the newly available observation or analysis data is consistently greater than the ensemble mean used to define the old ER sensitivity fields. However, it is crucial that such analysis be quantitatively validated using MST RH verified ER preemptive forecasts before forecasters should feel confident using this qualitative ER guidance.
6.7 Comparing univariate ESA sensitivities to multivariate ER sensitivities

An alternative ensemble-based sensitivity tool to ER is the univariate point correlation ESA technique discussed in section 3.1 (Hakim and Torn 2008; Torn and Hakim 2008; Ancell and Hakim 2007). Whereas ER computes field sensitivities by computing covariance-based regression operators between significant portions of the (or the entire) state vector, the point-correlation technique estimates field sensitivities by independently and iteratively computing ensemble correlations between individual elements of the state vector, and then normalizing by the predictand variance (viz. 3.1). That is, for example, ESA point correlation techniques approximate the sensitivity of the 500 hPa temperature field to the 500 hPa geopotential height at one specific location by individually computing the correlation of the 500 hPa heights at the point of interest to each of the points of the 500 hPa temperature field, and then multiplying the respective correlation by the variance of the temperature field at that respective point.

Because ER approximates sensitivities using ensemble covariances of vector-value functions, whereas the point correlation approach employs ensemble correlation estimates of scalar-valued function, ER can be considered a multivariate extension to the ESA point correlation technique. Expectedly, ER and point correlation sensitivities are very similar when the ensemble regressed perturbation is a multivariate representation of the scalar-valued function, as will be shown below.

The sensitivity of the forecast landfall zonal location of typhoon Sepat to pre-existing mid-tropospheric geopotential heights is used here to compare ER and point correlation sensitivities. The (ensemble-estimated) scalar-valued function employed for
this comparison is the JMA 72-hour forecast (at the approximate forecast landfall time and location) longitudes of the 1000 hPa PV of typhoon Sepat; the (ensemble-estimated) predictand field is the JMA 6-hour forecast 500 hPa geopotential height field. To ensure that the two methods are approximately measuring the sensitivity of the same quantity, the ER vector perturbation is a multivariate representation of the scalar-valued function, a dipole oriented approximately along the axis of the zonal landfall locations. That is, both the scalar-valued landfall longitudes (used as the independent variable in the univariate ensemble-sensitivity analysis) and the multidimensional forecast 1000 hPa PV zonally oriented dipole (used as the regressed perturbation for the ER) represent measures of the location of the 1000 hPa PV along approximately the same zonally oriented axis. Also note that, to ensure a numerically stable regression, the predictand and predictor ensembles are projected onto the leading four singular vectors and the point correlations are performed in this truncated space.

The filled contours of figure 6.10a show the prescribed 1000 hPa PV zonally oriented dipole perturbation (the ER predictor perturbation) and the black dots depict the ensemble estimates of the 72-hour forecast longitudes of the 1000 hPa PV (the ESA point-correlation independent variable). Note that the axis of the dipole is approximately equal to the line formed by the ensemble estimates of the longitudes. Figure 6.10b shows the effectively regressed perturbation, the ensemble resolved least squares estimate of the prescribed perturbation. By comparing figures 6.10a and 6.10b, it is clear that dynamical couplings of the JMA model require that the existence of the 72-hour 1000 hPa PV zonally oriented dipole anomaly in figure 6.10a requires the simultaneous presence of the meridional dipole and other features to the north and northeast of this zonally oriented
dipole, as depicted in figure 6.10b; the fact that the ensemble cannot resolve just the
dipole of figure 6.10a implies that no ensemble linear combinations exist that construct
the dipole of interest independently of the other features of figure 6.10b. Therefore, the
regression of the prescribed dipole associated with the zonal forecast landfall location
also includes the effects of this meridionally oriented northeastern dipole, as well as
several other features to its west.

The filled contours of figure 6.11 illustrate the 6-hour forecast 500 hPa
geopotential height anomalies associated with the effective perturbation (figure 6.10b), as
deduced via ER. Figure 6.12 depicts the univariate ensemble sensitivities (viz. 3.1) of the
6-hour forecast 500 hPa geopotential heights to the longitude of the 72-hour forecast
location of the 1000 hPa PV signature of Sepat. The gray line contours in figures 6.11
and 6.12 depict the ensemble mean 6-hour forecast 500 hPa geopotential heights, which
are included to inform the reader of the locations of the fields’ primary features.

As evidenced by comparing figures 6.11 and 6.12 the ER and ensemble
correlation sensitivities are unequivocally similar, with a 0.90 correlation coefficient
between the two fields. This similarity supports the notion that ER is a multivariate
extension to the univariate point correlation technique; when the ER perturbation is a
multivariate representation of the point-correlation independent variable of interest, the
two approaches yield similar sensitivity fields, as both techniques use the same ensemble
statistics to gauge the relationship between the perturbation function of interest and the
predictand field. Note, however, that in general multidimensional ER sensitivity analysis
has the distinct advantages over ESA point correlation analysis in that the ER operator
enables inferences of how entire fields jointly relate, rather than just of how scalars
Figure 6.10: a) Filled contours show the prescribed 1000 hPa PV zonally oriented dipole perturbation (the ER predictor perturbation) and the black dots depict the ensemble estimates of the 72-hour forecast longitudes of the 1000 hPa PV (the ESA point-correlation independent variable). Note that the axis of the dipole is approximately equal to the line formed by the ensemble estimates of the longitudes. b) Filled contours depict the effectively regressed perturbation.

individually relate. This allows the entire state vector to be regressed, enabling ER to diagnose the coupled variability of fields and to statistically forecast future states using lagged covariances; forecasting using univariate sensitivities is inappropriate given the couplings inherent to atmospheric dynamical evolution.
It is important to point out the significance of the positive perturbation and negative perturbation in the north-central and northeastern sections, respectively, of both figures 6.11 and 6.12. Although these features are certainly statistically related to the Sepat landfall longitudes, their collocation with the perturbations in the northern section of figure 6.10b suggests that their presence is most directly attributable to (or statistically associated with) these northern effective (yet not directly prescribed) perturbation of figure 6.10b. In the case of ER (figure 6.11), this result is unsurprising and consistent with the notion that an effective perturbation, rather than the prescribed perturbation, is regressed during ER. However, the evident effects of similar implied perturbations in the case of the univariate point-correlation technique (figure 6.12) may be unexpected, since the northern perturbations are not explicitly included in the sensitivity analysis; this may lead researchers to inappropriately consider these northern perturbations of 6.12 to be directly linked to the point perturbation of interest, when really they are likely linked to northern predictor perturbations correlated to the point perturbation.

This result highlights the importance of considering the prescribed scalar-valued independent variable in point correlation sensitivity studies as only one element of a multidimensional effectively resolved variable; for both univariate and multivariate sensitivity analysis, just because a perturbation is not explicitly incorporated into the regression does not mean that it is not implicitly included. A perturbation will be implicitly included if it is correlated to the prescribed perturbation or independent variable. It is recommended that users of univariate point correlation techniques identify perturbations correlated to the point perturbation of interest to better interpret and analyze sensitivity fields.
Figure 6.11: Filled contours illustrate the 6-hour forecast 500 hPa geopotential height anomalies associated with the effective perturbation (figure 6.10b), as deduced via ER. The gray line contours depict the ensemble mean 6-hour forecast 500 hPa geopotential heights.

Figure 6.12: Filled contours depict the ensemble sensitivities (viz. 3.1) of the 6-hour forecast 500 hPa geopotential heights to the longitude of the 72-hour forecast location of the 1000 hPa PV signature of Sepat. The gray line contours depict the ensemble mean 6-hour forecast 500 hPa geopotential heights.
6.8 Chapter summary

This chapter has presented ER as a means to understand the sensitivities of the track of Supertyphoon Sepat to mid-tropospheric geopotential heights. Using forecast ensemble information available in real-time, ER has quantified the relative sensitivities to track of dynamically meaningful atmospheric features including the position and extent of the subtropical ridge to Sepat’s north, the ridge extension to its east, and the retreating trough to its northwest. This information can potentially be invaluable to forecasters who wish to modify their predictions according to these changing sensitivities and to researchers attempting to understand the dynamics contributing to Sepat’s ultimate landfall location.

Despite its unconventionality, ER is not a drastic departure from traditional means of estimating sensitivities. Analogous to the countless numerical experiments that draw physical conclusions from model sensitivities to parameter tunings, ER draws inferences from the sensitivities of models to changes in initial (or forecast) conditions. The primary discrepancy is that traditional sensitivity experiments typically tune a single parameter or variable, whereas ensemble techniques necessitate the joint tunings of all initial condition fields in accordance with physical requirements. Although one may consider the inability to isolate the cause of model changes as a drawback of ER, its accordance with physical (i.e. balance considerations) and probabilistic requirements (i.e. probability laws imply that tuning one parameter or variable implies tuning all correlated parameters) dictated by atmospheric laws results in ER offering more realistic sensitivity information than that yielded by contrived single parameter tuning techniques. These two types of sensitivity experiments should be considered complementary.
Chapter 7

Comparing ER and Tangent Linear Model Singular Vectors

7.1 Introduction: Defining analysis error covariance normed singular vectors

Singular vectors are fundamental to several atmospheric science disciplines including dynamical sensitivity analysis and ensemble prediction systems. However, the calculation of relevant singular vectors via the analysis error covariance-normed tangent linear model, as explained below, often requires difficult coding, computational expenses, and frequent coding upkeep that can impede progress and provide difficulties in research related to singular vectors. This chapter uses a low-order model to show that singular vectors and tangent linear models can relatively simply be approximated using ensemble regression.

Singular vectors (SVs) represent perturbations that grow the most (under a specified norm) during the linearized integration of nonlinear dynamical equations. More specifically, initial-time SVs are the state-dependent directions in phase space at the beginning of the optimization interval that linearly evolve into the final-time SVs, the most highly varying directions (under a specified norm) at the end of the optimization interval (e.g., Kalnay 2003; Khade and Hansen 2004). SVs define the perturbations associated with the greatest forecast uncertainty and potentially with the most extreme
weather conditions, making them fundamental to research on the precursors to extreme atmospheric phenomenology (e.g. Farrell 1989), the identification of optimal targeted observation sites (Palmer et al. 1998), and the determination of the most relevant initial conditions for ensemble forecasts (Molteni et al. 1996).

Traditionally, initial- and final-time SVs, \( V \) and \( U \), respectively, are computed as the leading left and right SVs computed from the singular value decomposition (SVD) of the tangent linear model (TLM) operator, \( M \) (viz. 4.1), or equivalently as the eigenvectors of \( MM^T \) and \( M^T M \). The TLM is the product, over many short integration steps, of the linearized version of the nonlinear system equations, and can be considered an operator that maps an initial error to a final error, assuming linear dynamical evolution. The left SVs from an SVD of \( M \), \( U \), are eigenvectors of the state at the end of the optimization interval; these vectors represent the most highly varying directions at that time. The right SVs from an SVD of \( M \), \( V \), represent the unit magnitude vectors at the beginning of the optimization interval that map into \( U \) via \( M \) with growth factors given by the singular values. That is,

\[
MV = US, \quad (7.1)
\]

and

\[
v_i^T v_i = 1, \quad (7.2)
\]

where \( S \) is a diagonal matrix with singular values along its diagonal and \( v_i \) is the \( i \)th initial time singular value. Note that the unit magnitude constraint implies that the initial distribution from which the initial SVs are drawn is assumed to be isotropic and, given that the initial vectors all have the same magnitude, ensures that the final-time SVs are those vectors that have grown the most over the optimization interval.
However, the assumption of the initial distribution being isotropic implicit in traditional SV computations is poor. Given that the most appropriate method for estimating the initial ensemble distribution for SV analysis is by optimally combining a nonlinearly propagated short-term forecast and an observational uncertainty distribution (e.g. EnKF data assimilation; Evensen 1994; Houtekamer et al. 1998), there is an insignificant and merely coincidental chance that the resulting analysis is isotropic (e.g. Khade and Hansen 2004). Falsely assuming an isotropic analysis can result in SV directions that are highly improbable given the actual initial uncertainty distribution. Moreover, these directions will not be those that evolve into the eigenvectors of the final-time uncertainty distribution, and are therefore useless if one is interested in using SVs to determine the sensitive directions that lead to the greatest forecast error (e.g. Khade and Hansen 2004).

Because perturbations drawn from the analysis error uncertainty distribution (from an EnKF) are, by definition, statistically feasible and evolve into the final-time distribution from which the final-time SVs must come, relevant SVs assume that the initial uncertainty distribution is that given by the analysis error uncertainty distribution (Ehrendorfer and Tribbia 1997; Barkmeijer et al. 1998; Palmer et al. 1998; Bishop and Toth 1999; Hamill et al. 2003; Khade and Hansen 2004). Noting that perturbations, $\mathbf{x}'$, drawn from a distribution with zero mean and an identity covariance matrix can be transformed into perturbations, $\hat{\mathbf{x}}'$, drawn from a distribution with zero mean and covariance $\mathbf{P}^a$ via

$$
\hat{\mathbf{x}}' = \mathbf{P}^{a^{-1/2}} \mathbf{x}' \quad (e.g. \text{DelSole and Tippet 2008}),
$$

(7.3)
it is possible to transform an isotropic initial distribution into one having the same
distribution as the analysis by simply applying $P^{a^{-1/2}}$ to the initially isotropic
distribution. Moreover, noting that applying any norm $W$ to the TLM propagator $M$ is
equivalent to transforming the covariance of the initial distribution implicit in $M$ to $W^{-1}$,
it is straightforward to show that applying the norm $P^{a^{1/2}}$ to $M$ is equivalent to imposing
a covariance of $P^{a^{-1/2}}$ onto the initial distribution (e.g. DelSole and Tippett 2008). That
is, by applying the Analysis Error Covariance (AEC; e.g. Hamill et al. 2003) norm (also
known as the Mahalanobis norm), $P^{a^{1/2}}$, to $M$, the initial distribution is effectively
transformed into the appropriate analysis error uncertainty distribution, $P^a$. Therefore,

$$ SVD(MP^{a^{1/2}}) \quad (\text{Khade and Hansen 2004}) $$(7.4)

yields SVs relevant for predictability studies (Ehrendorfer and Tribbia 1997; Khade and
Hansen 2004).

Hamill et al. (2003) showed that AEC SVs can be computed without a TLM;
AEC initial SVs can be computed as the projections onto the analysis ensemble (formed
via an EnKF) of the linear combinations of ensemble perturbations that define the leading
final-time eigenvectors (Hamill et. al 2003). That is, instead of estimating the
eigenvectors of the final-time distribution from the SVD of the TLM operator, the
eigenvectors of the final-time distribution can be estimated from the distribution of the
nonlinearly integrated final-time states of an ensemble initially drawn from the analysis
distribution. Similarly, instead of using $M$ to determine the initial-time perturbations that
evolve into the eigenvectors of the final-time distribution, the initial-time SVs are
computed by applying to the analysis ensemble the same linear combination of ensemble
members that compose the eigenvectors of the final-time distributions \(i.e.\) the final-time SVs).

**7.2 ER and singular vectors**

An alternative method for producing AEC SVs comes from ER. By letting 
\[ P = X(t = 0) \] \(i.e.\) the analysis state vector and 
\[ Y = X(t = \tau) \] \(i.e.\) the forecast state vector at time \(\tau\), the end of the optimization interval, the ER operator \(L\) can be used to approximate AEC SVs. Analogous to the procedure for computing SVs from \(M\) \(viz.\) 7.4), the initial- and final-time SVs can be estimated from \(L\) as simply the right and left SVs, respectively, from

\[ SVD(L). \tag{7.5} \]

Because the initial distribution implicit in \(L\) is \(P^a\), the SVs computed \(viz.\) 7.5) are AEC SVs; that is, no further normalization is required, as one can consider the initial distribution as being an isotropic distribution with the AEC norm already applied.

ER-based SVs are equivalent to AEC SVs. These singular vectors, however, are presented somewhat differently; whereas ER-based initial and final SVs are cast as the right and left eigenvectors, respectively, of an ensemble-based linear regression operator, AEC initial SVs are defined as the projection onto the analysis ensemble of the linear combination of ensemble perturbations that defines the leading final-time eigenvector \(Hamill\ et.\ al\ 2003\). This equivalence introduces alternative interpretations of both ER and AEC SVs: the ER operator propagates perturbations by conserving ensemble linear combinations through phase space and the computation of AEC SVs can be cast as a linear regression of state perturbations.
This chapter uses the L95 model (Lorenz and Emanuel 1998) to show that, neglecting the null spaces discussed in section 7.5, 1) ER SVs are equivalent to those from an AEC-normed TLM operator and 2) ER can be used to approximate the TLM. Section 7.3 discusses the L95 model and the experimental setup. Section 7.4 compares SVs computed from the ER operator and the AEC-normed L95 TLM operator and also compares TLM SVs with ensemble-based inverse-AEC-norm approximations to TLM SVs. Section 7.5 summarizes and discusses the results.

### 7.3 Experimental Setup

The goal of this chapter is to show that ensembles can very closely approximate adjoint and TLM model operators and the singular vectors estimated from them. This aim will be addressed by computing the TLM and ER operators for the L95 model, and showing that SVs computed from the SVD of the AEC-normed TLM operator are approximately equivalent to those computed from the SVD of the ER operator. The inverse-AEC normed ER operator will be shown to approximate the TLM operator.

The L95 ER experiment is performed in the following manner. The system equations (viz. 4.2) are integrated for 5000 iterations to force the initial condition onto the system attractor. Then, a $n_{env} = 20$ ensemble is created by randomly perturbing the final spin-up control state and integrating the entire ensemble for 500 steps, while ensemble Kalman filtering the ensemble at each step using the control state as the observations. This ensemble is integrated (viz. 4.2) for $t = 1, 2, ..., \tau$ steps, where $\tau = 10$ with a time-step of $\Delta t = 0.025$; note that, because $\Delta t = 0.025$ corresponds to a time-step of approximately 3 hours (Lorenz and Emanuel 1998), an integration of $\tau = 10$ steps
corresponds to an approximately 30 hour model run. An ER operator is computed at the \( \tau = 10 \) step (viz. 3.4) using the initial ensemble as the predictor and the \( \tau \) th-step ensemble as the predictand. Separately, the TLM is evaluated as the product of the TLM at the previous time-step with the Jacobian of the time-varying basic state (viz. 4.1).

It is crucial to note that multicollinearities are likely to render the estimate of \( P_e^{-1} \) (the inverse of the analysis state perturbations) for the computation of \( L \) ill-conditioned. Therefore, \( P_e^{-1} \) is computed using only the first few leading singular vectors (see section 3.3). For the same reason, the \( P^{a1/2} \) and its inverse \( P^{a-1/2} \) are computed using only the leading singular vectors from the SVD of the analysis state perturbations.

### 7.4 Results

This section presents results of the comparison between SVs of the AEC-normed TLM operator and SVs of the ER operator. The solid line in figure 7.1a (7.1b) shows the leading initial-time (final-time) SV from \( SVD(MP^{a1/2}) \) and the dashed line depicts the leading initial-time (final-time) SV from \( SVD(L) \) for a randomly chosen case from the 100 independent trials of the experiment described in section 7.3. It is clear that the AEC-normed TLM SVs are approximately equivalent to ER SVs, neglecting the null spaces discussed in section 7.5. In fact, the average, over the 100 independent trials, correlation coefficient between the leading initial-time (final-time) SVs was 0.99 (0.95).

It is sometimes preferential to apply to \( M \) a norm other than the AEC. For example, if one is interested in the directions that have grown the most over the
Figure 7.1: The leading initial-time (a) and final-time (b) singular vectors from a randomly chosen run of the L95 model with $n_{ens} = 20$. The solid line depicts the AEC-normed TLM SV approximation (viz. 7.4). The dashed line depicts the ER SV approximation (viz. 7.5). Note that the dashed line is overlapped by the solid line in panel a.

optimization interval, rather than the directions that have the greatest variance at the end of the optimization interval (as what is derived from $SVD(MP^{a1/2})$), then the appropriate norm to apply to $M$ is the Euclidean $L_2$ norm (e.g. Khade and Hansen 2004). Other norms, such as total energy, kinetic energy, enstrophy, and streamfunction variance are also commonly applied to $M$ for SV computations (Palmer et al. 1998).

In such cases, it is possible to approximate $M$ via ER, thereby eliminating the
Figure 7.2: The leading initial-time (a) and final-time (b) singular vectors from a randomly chosen run of the L95 model with \( n_{\text{ens}} = 20 \). The solid line depicts the TLM SVs computed from \( \text{SVD}(M) \). The dashed line depicts the ER approximation of the TLM SVs, computed from \( \text{SVD}(LP^{a=1/2}) \).

need to code and update the TLM. Given that

\[
L = MP^{a1/2}
\]  

(7.6)

it is straightforward to show that

\[
M = LP^{a^{-1/2}}.
\]  

(7.7)

The solid line in figure 7.2a (7.2b) shows the leading initial-time (final-time) SV from \( \text{SVD}(M) \) and the dashed line depicts the leading initial-time (final-time) SV from \( \text{SVD}(LP^{a^{-1/2}}) \) for a randomly chosen case from the 100 independent trials of the...
experiment described in section 7.3. It is clear that the TLM can be approximated using an inverse-AEC-normed ER operator, neglecting the null spaces discussed in section 7.5. In fact, the average, over the 100 independent trials, correlation coefficient between the leading initial-time (final-time) SVs was 0.50 (0.80).

Note that the average correlations between the leading SV from $SVD(M)$ and $SVD(LP^{-1/2})$ are significantly lower than those between $SVD(L)$ and $SVD(MP^{a^{1/2}})$. This is attributable to $MP^{a^{1/2}}$ being dependent on the ensemble statistics intrinsic to $L$, causing an inbreeding of the compared quantities; contrarily, $M$ is computed independently of $LP^{-a^{1/2}}$.

### 7.5 Discussion and conclusions

Hamill et al. (2003) showed that the eigenvectors of a forecast ensemble and the projection onto the analysis ensemble of the linear combination of ensemble perturbations that defines these eigenvector respectively represent final- and initial-time AEC-normed singular vectors, the relevant SVs for studies of observation targeting and initial conditions for ensemble prediction systems. This chapter reinterprets the work of Hamill et al. (2003) by casting ensemble-based SVs as the singular vectors computed from an SVD of a multivariate regression operator relating the analysis ensemble to the forecast ensemble. More importantly, this chapter extends the work of Hamill et al. by showing that, for the L95 model, intrinsically AEC-normed ER SVs are approximately equivalent to explicitly AEC-normed tangent linear model SVs and that the TLM can be estimated by applying the inverse AEC-norm to the ER operator.
The approximate equivalence of ER SVs and AEC-normed TLM SVs implies that the existence of the TLM is not necessary for the computation of SVs. This result is highly attractive given the difficulties and expenses of coding and updating the TLM of sophisticated atmospheric models. Moreover, ER SVs enable the computation of SVs for models with TLMs that do exist because of a lack of differentiability of the system dynamics.

Although highly similar, ER SVs and AEC-normed TLM SVs had non-negligible differences that are predominantly attributable to ER null spaces. One such null space is due to the ensemble being rank-deficient (see chapter 4). Increasing the ensemble size to $n_{ens} = 500$, such that $n_{ens} > K$, reduced this null space and correspondingly increased the average, over the 100 independent trials, correlation coefficient between the leading initial-time (final-time) ER and AEC-normed TLM SVs to 0.99 (0.95) and between the leading initial-time (final-time) TLM and inverse-AEC-normed ER SVs to 0.61 (0.90).

The second null space is attributable to the truncation of singular vectors of the analysis ensemble with nonzero singular values necessary to ensure defined inverses and ER regression stability. The significance of this null space is related to the proportion of the variance of the analysis ensemble that cannot be explained by the retained singular vectors and also to the proportion of the cross-covariance between the analysis and forecast ensembles that cannot be explained by the truncated predictor. In addition to the inbreeding effect mentioned in section 7.4, this truncation null space is likely another factor causing the correlations between $SVD(M)$ and $SVD(LP^{-1/2})$ to be significantly lower than that between $SVD(L)$ and $SVD(MP^{-1/2})$; the error is double counted in the $SVD(M)$ and $SVD(LP^{-1/2})$ comparison, since both $L$ and $P^{-1/2}$ are subjected,
whereas the error is somewhat offset in the case of the \( SVD(L) \) and \( SVD(MP^{a^{1/2}}) \) comparison, since both \( L \) and \( P^{a^{1/2}} \) contain this null space.

It is expected that ER SVs using sufficiently sized ensembles, resolvable perturbations, and negligible truncation null spaces will be more accurate than those from the TLM. Unlike the TLM, ER linearly parameterizes all physical processes and state perturbations correlated to those used to compute the regression operator, thereby implicitly including a relatively more complete representation of the system dynamics into its estimate of the system propagator. Moreover, ER implicitly includes those dynamics that are omitted from TLMs due to a lack of differentiability.

It is the hope of the author that the relative simplicity of using ER compared to using the TLM and the proven approximate equivalence of ER and AEC-normed TLM SVs will encourage and facilitate future SV sensitivity research.
Chapter 8

Conclusions

8.1 Summary and conclusions

A potentially fruitful application of ensemble model output is Ensemble Synoptic Analysis (ESA; Hakim and Torn 2008), the use of ensemble analysis and forecast covariances to make inferences about the atmosphere. By treating individual ensemble members as independent samples, ESA employs standard statistical techniques to compute sensitivities, infer dynamical couplings, and aid forecasters in identifying dynamical processes that are particularly relevant for specific flow-dependent weather predictions that stationary time series data are ill-equipped to model.

The focus of this thesis is to develop and apply an ESA method called Ensemble Regression (ER), which facilitates inference about the relationship between two multidimensional atmospheric fields \( P \) and \( Y \) via the regression of a perturbation using an operator \( L \) defined by the covariances of the fields’ ensemble forecasts and/or analyses. Matrices of ensemble anomalies of the predictor field \( P \) and the predictand field \( Y \) are used to train \( L \), which can be used to predict the most probable state of the predictand field, \( \hat{y}' \), given a perturbation of the predictor field, \( p' \).

Noting that the significance of the results of non-contemporaneous ERs is a function of the reliability of the ensemble forecast distribution, chapter 2 presents theory
and applications for a multivariate ensemble verification tool called the minimum spanning tree rank histogram (MST RH). After eliminating biases, spatial and temporal correlations, and variance inconsistencies among the $K$ dimensions, the shape of an MST RH can be used to diagnose the relationship between the distribution of the ensemble and of the verification. It is shown that the Mahalanobis norm transforms the forecast data in the most meaningful and interpretable way when the number of ensemble members is greater than the number of forecast locations and/or weather components. However, given the misleading results when $n_{ens} \leq K$, it is suggested that the variance norm be used when $n_{ens} \leq K$ and the variances in all dimensions are not identical. Reliability information ascertained from the MST RH can ultimately help improve forecast reliability through the modification of the ensemble prediction system and can be used to assess whether forecast ensembles used for ER accurately sample the forecast uncertainty space.

Chapter 3 returns to the focus of this thesis by defining ER and discussing several of its key properties. This chapter discusses methods to maximize the numerical stability of ER, such as performing ER in the subspace of the leading predictor singular vectors or the subspace the leading predictor and predictand principal components. Also, chapter 3 introduces the notion of the effective perturbation; the perturbation effectively regressed via ER is not the prescribed perturbation, $\mathbf{p}'$, but is instead $\hat{\mathbf{p}}'$, the perturbation resolved by $\mathbf{L}$ as the least squares estimate of $\mathbf{p}'$.

Chapter 4 employs low order Lorenz models and high-dimensional operational data to assess ER prediction skill and potential error sources using leave-one-out cross validation. L63 ER yields highly accurate forecasts comparable to those of the TLM
within a window of linearity, suggesting that linear ER forecasts can potentially be skillful even in highly nonlinear systems. The 40-dimensional L95 model is used to show that, because larger ensembles are significantly more likely to span the subspace of the predictor perturbation of interest, ensembles with sizes greater than the dimensionality of the system equations yield ER forecasts significantly more skillful than those with smaller sizes; L95 forecasts of well resolved ensemble perturbations have skill comparable to that of TLM forecasts. Also, L95 ER forecasts are significantly sensitive, particularly for long lead times, to the choice of ensemble members used to defined the ER operator (even from the same ensemble distribution), suggesting that ER forecasts are potentially subject to non-negligible sampling errors.

Ensemble data from the Japanese Meteorological Association is used to illustrate ER skill for high-dimensional sophisticated operational data. ER mean absolute errors, with respect to JMA forecasts, of 1000 hPa geopotential height perturbations ranged from approximately 2 meters at analysis time to 10 meters at 4 days lead time and median anomaly correlation coefficients ranged from 0.86 at analysis time to 0.58 at 72 hours lead time. Principal component sensitivity analysis suggested that, although highly sensitive when considering the full spectrum of potential degrees of truncation, ER ACC values are relatively insensitive within a moderately wide range of principal component numbers close to the optimal numbers.

Following Gombos and Hansen (2008), chapter 5 presents the first rigorous application of statistical potential vorticity (PV) regression, which piecewise inverts a PV perturbation using an ER operator defined by PV, potential temperature, and geopotential height ensemble analyses. It is shown that, if performed in the subspace of the leading
PV singular vectors, the statistical ER technique yields height perturbations nearly identical to those resulting from the piecewise PV dynamical inversion technique of Davis and Emanuel (1991), if the same PV perturbation is effectively regressed (inverted).

Chapter 6 applies ER as a non-contemporaneous multidimensional sensitivity tool to analyze the sensitivities of tropical cyclone tracks to prior mid-tropospheric geopotential height perturbations. Using forecast ensemble information available in real-time, ER has quantified the relative sensitivities to track of dynamically meaningful atmospheric features including the position and extent of the subtropical ridge to Sepat’s north, the ridge extension to its east, and the retreating trough to its northwest. This information can potentially be invaluable to forecasters who wish to modify their predictions according to these changing sensitivities and to researchers attempting to understand the dynamics contributing to Sepat’s ultimate landfall location. This chapter also illustrated promising proof-of-concept results for ER preemptive forecasting, a technique that can potentially supply forecasters and emergency preparation decision makers with skillful forecasts several hours in advance of operational models.

Chapter 7 discusses ER applications for singular vector analysis and uses a Lorenz model to compare singular vectors deduced from ER to those from a tangent linear model. This chapter extends the work of Hamill et al. (2003) by showing that, for the L95 model, intrinsically AEC-normed ER SVs are approximately equivalent to explicitly AEC-normed tangent linear model SVs and that the TLM can be estimated by applying the inverse AEC-norm to the ER operator. The approximate equivalence of ER SVs and AEC-normed TLM SVs implies that the existence of the TLM is not necessary
for the computation of SVs. This result is highly attractive given the difficulties and expenses of coding and updating the TLM of sophisticated atmospheric models.

Moreover, ER SVs enable the computation of SVs for models with TLMs that do exist because of a lack of differentiability of the system dynamics.

8.2 Future work

The generality of ER presents innumerable potential applications, several of which are outlined in this section. Note that several ideas for future work, such as verifying large samples of ER preemptive forecasts using the MST RH, have already been discussed in other sections.

PV and geopotential height are only one of many pairs of fields that be studied via ER. One can piecewise regress subsections of the PV field to estimate statistically associated precipitation, for example, rather than heights, to understand the coupled dynamical relationship between PV and precipitation (e.g. Hakim and Torn 2008).

Bearing in mind the EPV conservation principle and using finite time differencing of the PV field to isolate PV of nonconservative origin, PV-precipitation ER using nonconservative PV can be used to study the poorly understood feedbacks of latent heating and precipitation. Note, however, that unlike the PV-height regressions presented in chapter 5, the exact dynamical relationship between PV and precipitation is unknown. Therefore, using PV as a predictor for precipitation may prove futile if only an inadequate percentage of the variance of precipitation can be explained by the PV predictor ensemble (Gombos and Hansen 2008).
Another potential area of research is to further explore the implications of coupled PV perturbations in the context of PV inversion. For example, because stratospheric and tropospheric PV perturbations are typically dynamically linked, performing classical piecewise PV inversion on isolated stratospheric PV perturbations can potentially be dynamically inconsistent; future research might account for dependent perturbations by employing the notion of the effective perturbation when using PV inversion to study the stratosphere.

Future work can apply ER to better understand the dynamics of important currently poorly forecast atmospheric phenomena, such as tropical cyclone intensity, through the use of ER anomaly patterns and other methods. Note, however, that since current models poorly forecast tropical cyclone intensity, ER might be better suited to study intensity-related model errors rather than the dynamics of intensity.

ECMWF ER sensitivities of Sepat tracks to 500 hPa geopotential heights (not shown) are non-negligibly different from those from the JMA ER (viz. chapter 6). Such model sensitivity comparisons can potentially be used to identify fundamental model differences and model errors, to suggest model-specific rankings of the importance of dynamical features for specific forecast decisions, and to suggest that optimal targeting observing sites are highly model-dependent, necessitating model-specific targeting tracks and fleets.

ER also presents the potential to combine aspects of different models. An ER operator formed using ensemble analyses and/or forecasts from multiple models, each with different physical specifications and parameterizations, samples various regions of the model parameter PDF and effectually combines characteristics of multiple model into
a single model. Future work might assess the value of multi-model ER and identify how its forecast errors compare to those from the constituent model ERs.

It is the hope of the author that the demonstrated accuracy and versatility of ensemble regression will improve the understanding, modeling, and forecasting of the atmosphere in the years to come.
Bibliography


temperatures using Linear Inverse Modeling, *J. Climate*, 11, 483-496.

Penland, C. and P. D. Sardeshmukh, 1995: The optimal growth of sea surface

NCEP global 3D vibrational analysis system, *Wea. Forecasting.*, 12, 689-695.

Rabier, E., H. Jarvinen, E. Klinker, J. Mahfouf, and A. Simmons, 2000: The ECMWF
operational implementation of four-dimensional variational assimilation. I:
Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, 126,
1143-1170.


Saetra, O., H. Hersbach, J.R. Bidlot, D.S. Richardson, 2004: Effects of observation errors
1501.

Forecasting.*, 12, 809-818.


Stensrud, D.J. and J.A. Skindlov, 1996: Gridpoint predictions of high temperature from a

Talagrand, O., B. Strauss and R. Vautard, 1997: Evaluation of probabilistic prediction
systems. *Proc. of the ECMWF workshop on predictability*, Reading, United
Kingdom, 1-25.

Thepaut, J.N., R.N. Hoffman, and P. Courtier, 1993: Interactions of dynamics and
observations in a 4-dimensional variational assimilation. *Mon. Wea. Rev.*, 121,
3393-3414.


Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of

136, 663-677.


