University of Reading

School of Mathematics, Meteorology and Physics

Error growth in medium-range forecasting models

Using the Mean-Variance of Logarithms (MVL) diagram to characterise the spatiotemporal behaviour of ensembles

Zak Kipling

August 21, 2009

A dissertation submitted in partial fulfilment of the requirement for the degree of Master of Science in Mathematical and Numerical Modelling of the Atmosphere and Oceans.

Declaration

I confirm that this is my own work, and the use of all material from other sources has been properly and fully acknowledged.

Zak Kipling

Acknowledgements

I would like to thank my supervisors, Andrew Charlton-Perez[1] and Cristina Primo[2], for their support, encouragement and ideas throughout the project, as well as all those whose teaching has given me the necessary background knowledge.

During the course of this project, many open-source software tools have been a considerable help, in particular Python[3] (with the NumPy[4], netCDF4[5] and matplotlib[6] modules) for data processing and plotting and LaTeX[7] for typesetting the dissertation itself.

I would also like to thank the Natural Environment Research Council (NERC) for the financial support to undertake the MSc programme.

Finally, this work would not have been possible without the TIGGE[8] archive of medium-range ensemble forecasts, from which all the data for this project were obtained.

Abstract

Due to the chaotic nature of Numerical Weather Prediction (NWP) models, medium-range forecasting is usually based on ensemble techniques to quantify the uncertainty in the forecast products. There are a variety of techniques used operationally to produce initial perturbations for these ensembles: bred vectors, singular vectors, ensemble Kalman filters, etc. These different techniques, as well as differences in model formulation, lead to ensembles with different spatiotemporal dynamics.

Using the recently-developed mean-variance of logarithms (MVL) diagram, together with the TIGGE archive of medium-range ensemble forecasts from nine different centres, we present an analysis of the dynamics of their perturbations, and show how the differences between models and perturbation techniques can explain the shape of their characteristic MVL curves.

We also consider the use of the MVL diagram to compare the growth of perturbations within the ensemble with the growth of the forecast error, showing that there is a much closer correspondence for some models than others.

We conclude by looking at how the MVL technique might assist in selecting models for inclusion in a multi-model ensemble, and suggest an experiment to test its potential in this context.

iv

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Predictability and chaos

Beyond very short lead times, deterministic weather forecasts become un-
reliable as the chaotic nature of the atmospheric dynamics limits the pre-
dictability. However, as explained by Palmer (2006), it is not just that the
future state of the system depends on the initial conditions; the degree of
predictability itself also varies. Figure 1.1 illustrates this in the context of a
simple three-variable model (Lorenz 1963). On the left is a high-predictability
initial state, with a neighbourhood in state space that evolves in a consis-
tent way; on the right is a low-predictability state where the neighbourhood
rapidly diverges into very different states.

Thus, as we look to forecast at longer lead times, we must start to consider
probabilistic forecasts. In this framework, instead of predicting a single future
state based on the evolution of a single current state, we aim to predict the
evolution of the probability distribution function (PDF) of the atmospheric
state. There are three main problems to overcome:

- Although the time evolution of a PDF under a chaotic dynamical sys-
  tem can be computed explicitly via the Liouville equation (Ehrendorfer
  2006) for low-dimensional systems, this approach is not computation-
  ally feasible for the high-dimensional models used in numerical weather
  prediction (NWP).
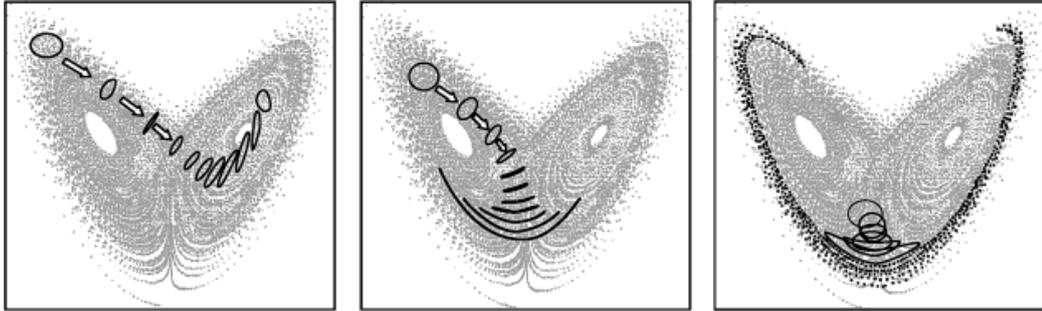
Figure 1.2 Finite time ensembles of the Lorenz (1963) system illustrating the fact
that in a non-linear system, the evolution of the forecast probability density $\rho(X, t)$
is dependent on initial state.

Figure 1.1: An illustration of how the predictability of a chaotic system
depends on the initial conditions (Palmer 2006, fig. 1.2).

- We need an estimate of the full PDF at the initialisation time, rather
  than just a single "best estimate" analysis of the current state.

- We also need to account for errors in the forecast model itself; otherwise
  the forecast PDF is likely to be overconfident.

The usual approach is to take an ensemble of forecasts each starting from
a state which samples the initial (analysis) PDF; the forecast PDF can then
be estimated from the density of ensemble members in the state space (see
figure 1.2). Modern data assimilation techniques provide an estimate of the
analysis error covariance; this allows us to sample the initial PDF provided we
make some assumptions about its distribution (e.g. that the analysis errors
are Gaussian). Model error is typically accounted for either by varying certain
parameters in the model equations between ensemble members, or by using
stochastic models for some of the physical processes.

This approach gives an estimate of the uncertainty in the forecast at
longer lead times, allowing meaningful medium-range forecasts to be pro-
duced. If the ensemble spread is small (i.e. the ensemble members are in
close agreement) at a given lead time, then we can have a high confidence in
the forecast; if the spread is large (i.e. there is little agreement between the
members) then the forecast should not be relied upon too heavily.
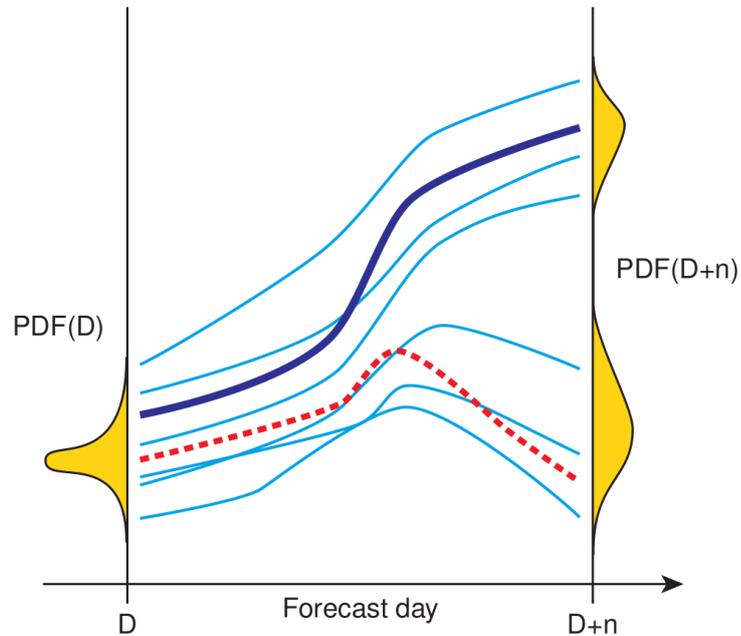
Figure 1 Schematic of ensemble prediction. The initial proba-
bility PDF(D) represents the initial uncertainties. From the best
estimate of the initial state a single deterministic forecast (bold
blue solid curve) is performed. This single deterministic forecast
fails to predict correctly the future state (red dotted curve). An
ensemble of perturbed forecasts (thin blue solid curves) start-
ing from perturbed initial conditions designed to sample the
initial uncertainties can be used to estimate the probability
PDF(D+n) at the future time D+n.

Figure 1.2: An illustration of how ensemble prediction estimates the evolution
of the PDF of an uncertain state (Buizza et al. 2001, fig. 1).

In a low-dimensional model it is easy to sample a region of state space effectively, e.g. by taking points randomly or uniformly within an ellipsoid around the "best-guess" initial conditions. Actual NWP models, however, may well have $\sim 10^6$ variables; if we are to have a manageable ensemble size (typically no more than 50 members in operational systems), we can only span a small subspace of the total state space with our perturbations. Early experiments in ensemble forecasting (Hollingsworth 1980) showed that in such cases an ensemble based on random perturbations does not produce a realistic distribution of forecast states.

Given this restriction, it is important that we choose this subspace to include those perturbations which will grow the fastest and thus dominate the uncertainty in the forecast. For the long-term growth of infinitesimal perturbations, these are the leading Lyapunov vectors; however for shorter-term growth and/or non-infinitesimal perturbations, this does not necessarily hold (Boffetta et al. 2002), and there are a number of different techniques used to select the perturbations: singular vectors, bred vectors, empirical orthogonal functions, and ensemble Kalman filter-based techniques.

We aim to explore the spatiotemporal dynamics of these perturbations, and in particular whether there are significant differences in the dynamics of the various methods, as used in operational medium-range weather forecasting models.

1.1.1   Measuring perturbations

We are interested in the growth over time of perturbations in a spatial field $\phi(\boldsymbol{x}, t)$. The model variables are typically in three spatial dimensions (latitude, longitude and a generalised vertical coordinate). In this investigation, however, we restrict ourselves to two-dimensional "slices" (e.g. $500\,\mathrm{hPa}$ geopotential height, $2\,\mathrm{m}$ temperature).

We define the perturbation field $\delta\phi^{(n)}$ as the difference between the nth perturbed ensemble member $\phi^{(n)}$ and an unperturbed control $\phi^{(0)}$:

$$\delta\phi^{(n)}(\boldsymbol{x}, t) = \phi^{(n)}(\boldsymbol{x}, t) - \phi^{(0)}(\boldsymbol{x}, t). \tag{1.1}$$

It is well known that, if the system is chaotic, infinitesimal perturbations will exhibit a long-term average exponential growth with a characteristic rate:

$$\delta\phi \sim \mathrm{e}^{\lambda_1 t}, \tag{1.2}$$

where λ_1 is referred to as the *leading Lyapunov exponent.* Growth rates for non-infinitesimal perturbations, and short time scales, however, may be larger or smaller. In particular, nonlinear effects will prevent the perturbations growing larger than the natural amplitude of the system (i.e. the range of possible weather is restricted by climatology).

In addition, the perturbations at different points in space typically do not grow independently, but rather in a spatially-correlated way on a variety of scales from mesoscale systems to large-scale planetary waves. It is this spatial correlation with which we are primarily concerned here, seeking a way to visualise its growth and decay as the perturbations increase in magnitude.

1.1.2 The significance of spatial correlation

What is the significance of the spatial correlation in the perturbation fields? The answer lies in what it can tell us about predictability. When the forecast no longer has any skill over climatology, the perturbations are akin to the difference between arbitrary states drawn from climatology, and we do not expect to see large-scale correlations. Thus, when large-scale correlation does exist in the perturbations of the ensemble members, they still contain some information about the large-scale evolution even if the perturbations are large in magnitude.

Lopez et al. (2004) show, for a simple model, how small perturbations initially grow according to the linearised model equations, with growing spatial correlation; as the perturbations become large and enter the nonlinear regime, the spatial correlation is destroyed and the perturbations tend to "spatiotemporal white noise".

In order to understand this growth of spatial correlation in the perturbations in more depth, we will make use of some results from another area of mathematics – the growth of rough interfaces. In the next section we give

a brief introduction to the theory, and show how it can be applied to the spatiotemporal growth of perturbations.

1.2   Interface roughening

There is, perhaps surprisingly, a strong theoretical connection between the spatiotemporal growth of perturbations in a chaotic system and the fractal growth of rough interfaces, of which there are many examples in materials science, biology and elsewhere. We will begin by reviewing some of the basic ideas involved; a more thorough introduction to interface-growth theory is given in Barabási & Stanley (1995).

1.2.1   Ballistic deposition

In this model, we consider randomly-distributed particles on a square lattice falling from the far field onto an initially-flat surface, or "interface" between the accumulated particles and the free space above. A particle comes to rest when it neighbours another, permitting both vertical and lateral growth as shown in figure 1.3.
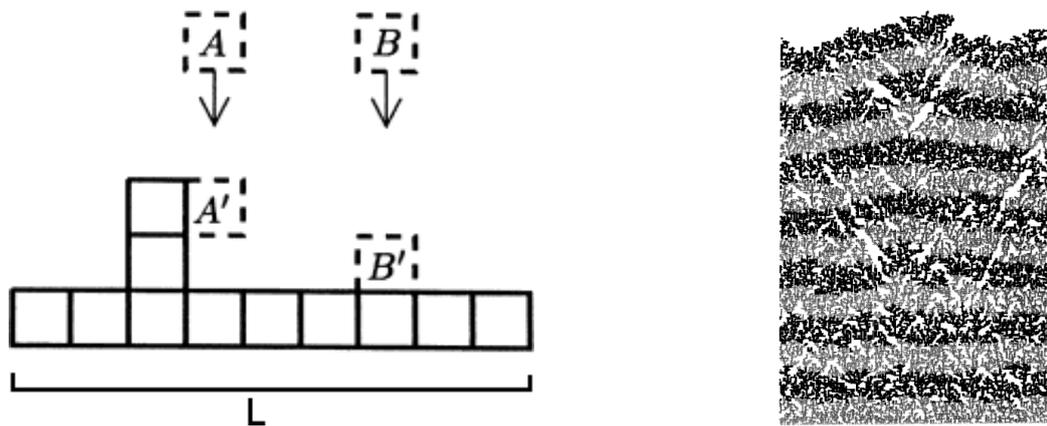
Figure 1.3: The ballistic deposition (BD) process (Barabási & Stanley 1995, figs 2.1, 2.2)

We can characterise the interface at any time t by two functions:

1. its *mean height,*

$$\overline{h}(t) = \frac{1}{L} \sum_{i=1}^{L} h_i(t), \tag{1.3}$$

   where $h_i(t)$ is the height of the highest particle in column i and L is the size of the system (i.e. the total number of columns); and

2. its *width* (the standard deviation of the column heights),

$$w(t) = \sqrt{\frac{1}{L} \sum_{i=1}^{L} \left(h_i(t) - \overline{h}(t)\right)^2}, \tag{1.4}$$

   which is a measure of the "roughness" of the interface.

1.2.2 Scaling

If we assume that particles fall in each column at a constant average rate, then the mean height increases linearly with time:

$$\overline{h}(t) \sim t. \tag{1.5}$$

The interface width grows in a more complex way (typical results from a numerical simulation are shown in figure 1.4) with initial power-law growth,

$$w(t) \sim t^{\beta}, \tag{1.6}$$

until the system becomes *saturated* at a width

$$w_{sat} \sim L^{\alpha}. \tag{1.7}$$

This occurs at the *crossover* or *saturation* time

$$t_\times \sim L^z. \tag{1.8}$$

The exponents in these three power laws are known as:

β – the *growth exponent*;

α – the *roughness exponent*;

z – the *dynamic expnonent*.

These exponents are not independent; for continuity at the crossover time (i.e. for the three straight lines on the previous figure to be concurrent), we must have

$$z = \frac{\alpha}{\beta}. \tag{1.9}$$

They are not only fixed for a given model, but are often found to be common to a number of models of very different phenomena which are then said to be in the same "universality class".

1.2.3   Saturation

The possibility of lateral growth introduces correlations between the height of nearby columns. Continuing with scaling arguments, we can introduce a typical *correlation length* $\zeta_\parallel$ over which such correlation occurs.

Initially, the interface is flat and all growth is vertical (and thus uncorrelated). Once lateral growth occurs, $\zeta_\parallel$ begins to grow. However it cannot become larger than the total length L of the system.

We make the plausible assumption that saturation corresponds to the entire interface becoming correlated, i.e.

$$\zeta_\parallel \sim L \sim t_\times^{1/z} \quad (t \gg t_\times). \tag{1.10}$$
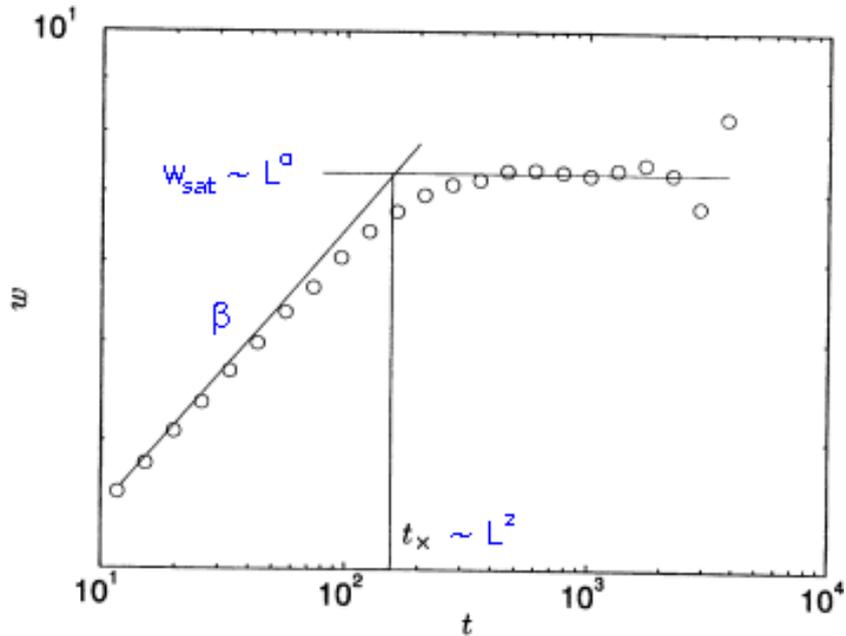
Figure 1.4: The growth of the interface width in a numerical simulation of the Ballistic Deposition process, on log-log axes (Barabási & Stanley 1995, fig. 2.3).

We can extrapolate to pre-saturation times to estimate the growth of these correlations:

$$\zeta_{\parallel} \sim t^{1/z} \quad (t \ll t_{\times}) \tag{1.11}$$

More sophisticated continuum models of the process – which we will not discuss here in detail, but see Barabási & Stanley (1995, ch. 6) – can validate these assumptions.

1.2.4 Self-affinity and fractal structure

Given the relation (1.9) between the three exponents ($z = \alpha/\beta$), we obtain a scaling relation between the correlation length and the interface width:

$$\zeta_{\parallel} \sim w^{1/\alpha}. \tag{1.12}$$

We can interpret this by viewing the interface as a *self-affine fractal*, whose statistical properties are invariant under a scaling with aspect ratio $1 : \alpha$.

This property is important because it allows us to characterise the growth of the correlation length in terms of the easily-measured interface width.

1.2.5   The Kardar-Parisi-Zhang (KPZ) equation

Based on physical and symmetry principles, the KPZ equation (Kardar et al. 1986) gives a continuum model of the BD process:

$$\frac{\partial h(\boldsymbol{x}, t)}{\partial t} = \nu \nabla^2 h + \frac{\lambda}{2} (\nabla h)^2 + \eta(\boldsymbol{x}, t), \tag{1.13}$$

where $h_i(t)$ is considered as the discretisation of $h(\boldsymbol{x}_i, t)$, $\eta(\boldsymbol{x}, t)$ represents uncorrelated random noise with zero mean, and ν and λ are parameters of the system.

In fact, this same model can be used to describe a wide range of growth processes – including, as we shall see shortly, perturbations in chaotic dynamical systems.

The exponents α, β and z depend on the dimensionality of $\boldsymbol{x}$, and are only easily found analytically for the 1D case, when

$$\alpha = \frac{1}{2}, \quad \beta = \frac{1}{3}, \quad z = \frac{3}{2}. \tag{1.14}$$

This theory of interface growth finds applications in many other fields:

- Deposition processes – BD is an idealised example, but many real-world processes behave similarly both in nature (e.g. snowfall) and in industry (e.g. in semiconductor manufacturing).

- Fluid flow in porous media, from the progressive "wetting" of cloth or paper to the flow of oil through rock.

- The spread of bacterial colonies in a nutrient medium.

Although these systems span a wide range of scales, they exhibit many common characteristics in terms of the scaling laws and self-affinity discussed

above. Many examples fall into one of a small number of "universality classes" with specific values for the exponents α, β and z, of which KPZ is one particular example.

1.2.6 The link to perturbation growth

By taking the (natural) logarithm of the perturbations (the *Hopf-Cole transformation*),

$$h(\boldsymbol{x},t) = \ln |\delta\phi(\boldsymbol{x},t)|, \tag{1.15}$$

at each point in space $\boldsymbol{x}$ and time t, it can be shown for many dynamical systems that h evolves according to the KPZ equation, and thus the associated results from interface-growth theory apply.

In particular, this means we can use the interface width as a measure of the spatial correlation of the perturbation field. Pikovsky & Kurths (1994) demonstrate the applicability of the KPZ equation to the special case of coupled-map lattices; this is extended to a range of other simple dynamical systems by Pikovsky & Politi (1998). More recently, Primo et al. (2007) show that perturbations in an operational weather forecasting model follow similar scaling laws. Specifically, they find that in the mid-latitudes the growth corresponds to the 1D KPZ equation; while at the equator the behaviour is akin to the 2D Edwards-Wilkinson equation – which models a different class of growing interface, with different values for α, β and z (Barabási & Stanley 1995, ch. 5).

In chapter 2, we will see how this theory can be used to construct a concise visualisation of the spatiotemporal dynamics of a given ensemble, and how this can be used to understand aspects of how different models behave.

Chapter 2

The Mean-Variance of Logarithms (MVL) diagram

By analogy with the characterisation of a rough interface by its mean height and width, discussed in section 1.2, we can characterise the perturbation field by the spatial mean and variance of $h_i(t) = \ln |\delta\phi_i(t)|$:

$$M(t) = \frac{1}{L} \sum_{i=1}^{L} h_i(t) \tag{2.1}$$

$$V(t) = \frac{1}{L} \sum_{i=1}^{L} (h_i(t) - M(t))^2, \tag{2.2}$$

where we have discretised the field $\phi_i(t) = \phi(\boldsymbol{x}_i, t)$ $(i = 1, 2, \ldots, L)$. To visualise the spatiotemporal growth of the perturbations, we plot a graph of $V(t)$ against $M(t)$. While such a curve can be plotted for an individual member of a forecast ensemble, we typically plot the mean of M and V over all ensemble members and a number of separate forecast initialisations to obtain a characteristic MVL curve for the model in question. This is the "Mean-Variance of Logarithms" diagram, introduced by Gutiérrez et al. (2008). Their results for a simplified model (Lorenz 1996) and its linearisation, using both random initial perturbations and bred vectors (see section 3.2.1), are shown in figure 2.1.
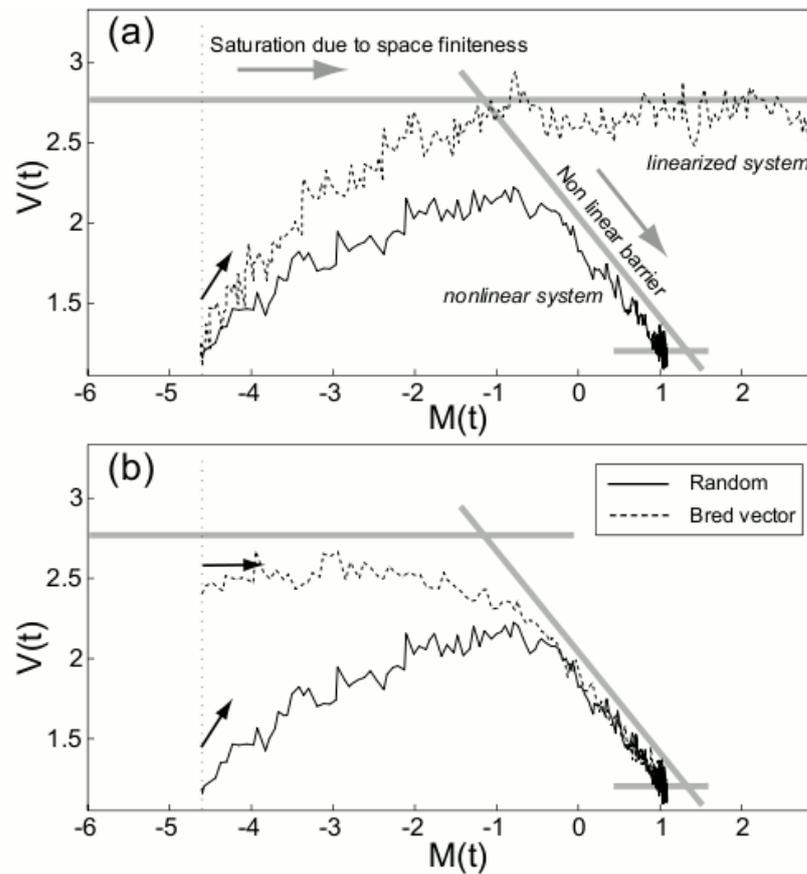
Fig. 2. (a) MVL diagram (see text) for the nonlinear Lorenz96 (solid) and the linearized system (dashed). (b) MVL diagram for two different initial perturbations: random (solid) and bred vectors (dashed).

Figure 2.1: MVL diagram for the Lorenz (1996) model and its linearisation, with random and bred-vector initial perturbations
(Gutiérrez et al. 2008, fig. 2).

There are a few points to note here:

- $M(t)$ increases with time during perturbation growth, so the progression is from left to right.

- From initially-random (i.e. spatially uncorrelated) perturbations, the spatial correlation grows until saturation occurs (just as for the interface width considered in section 1.2)

- In the linearised system, $M(t)$ continues to increase while the spatial correlation remains saturated; in the nonlinear system, the growth is curtailed by nonlinear effects which also destroy the spatial correlation as described by Lopez et al. (2004).

- The bred vectors, on the other hand, have a strong initial spatial correlation and are already saturated. They grow in amplitude, maintaining this correlation, until the nonlinear effects take over.

- For physical systems, where ϕ is not dimensionless, M is defined up to an arbitrary additive constant depending on the constant used to nondimensionalise $\delta\phi$ before taking the logarithm. V is unaffected, as this constant will cancel out.

2.1 Computational issues

There are two issues which slightly complicate the implementation of this method: grid points at which the perturbation is apparently zero, and the non-uniform distribution of grid points.

The first problem arises due to the use of logarithms and the limited numerical precision of real-world data sets. If the perturbation at a grid point $\boldsymbol{x}_i$ is too small to be represented in the data set, then it will appear that $\delta\phi_i = 0$, and therefore h_i is undefined. For data stored as 32-bit IEEE 754 floating point data, this is likely to occur when $|\delta\phi/\phi| < 10^{-7}$, e.g. a $0.1\,\mathrm{mm}$ perturbation in a geopotential height of $1\,\mathrm{km}$. To solve this problem, we

calculate h_i as:

$$h_i(t) = \ln \left| \max \left\{ \delta\phi_i(t), \epsilon \right\} \right|, \tag{2.3}$$

where ϵ is chosen to be an order of magnitude smaller than any perturbation which could be represented in the given field (e.g. for the above geopotential height field, $\epsilon = 0.01 \, \text{mm}$ would be appropriate). The exact choice of ϵ to represent the range of very small perturbations is somewhat arbitrary, however in practice has no significant impact on the resulting plots while removing the undefined values that would otherwise occur.

The second problem is that, when processing data from atmospheric models, we are generally dealing with a latitude/longitude grid, where the longitudinal spacing becomes closer toward the poles. The upshot of this is that the grid points near the equator represent a greater physical area than those near the poles. To correct for this, we modify (2.1) and (2.2) to use an area-weighted mean and variance:

$$M(t) = \frac{\sum_{i=1}^{L} h_i(t) \cos\theta_i}{\sum_{i=1}^{L} \cos\theta_i}, \qquad V(t) = \frac{\sum_{i=1}^{L} (h_i(t) - M(t))^2 \cos\theta_i}{\sum_{i=1}^{L} \cos\theta_i}, \tag{2.4}$$

where θ_i is the latitude of the grid point $\boldsymbol{x}_i$. In practice the correction only has a minor impact on the resulting MVL curves, and is barely noticeable if the polar regions are not included in the calculation.

2.2 Seasonal model intercomparison: DEMETER

The DEMETER project (Palmer et al. 2004) is a multi-model seasonal ensemble with hindcasts for seven models covering a common 22-year period (1980–2001).
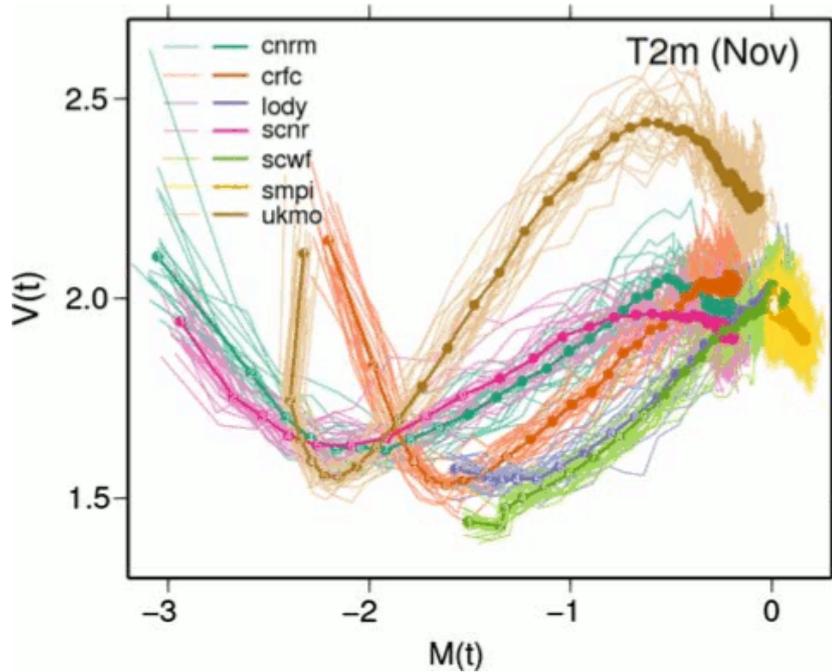
Figure 2.2: MVL diagram for 2 m temperature in the DEMETER models
(Fernández et al. 2009, fig. 2).

Fernández et al. (2009) applied the MVL approach to these seven models,
looking at both the 2 m temperature (figure 2.2) and 500 hPa geopotential
height (figure 2.3) fields.

For the 2 m temperature (figure 2.2), many of the models show a strong
initial spatial correlation (large V), which decreases rapidly before the main
growth stage. This is interpreted as an indication that the initial perturba-
tions have a large-scale spatial structure, but in a manner inconsistent with
the model dynamics (i.e. taking the system away from its attractor); thus this
structure is destroyed in an initial transient phase as the perturbed system
returns to its attractor. This is path (B) in the conceptual picture shown in
figure 2.4.

For the 500 hPa geopotential height (figure 2.3), all the models start with
very little spatial correlation (low V); this is because in seasonal models the
perturbations are generally introduced in the sea-surface temperature (SST)
fields and so couple strongly to near-surface variables but only weakly to
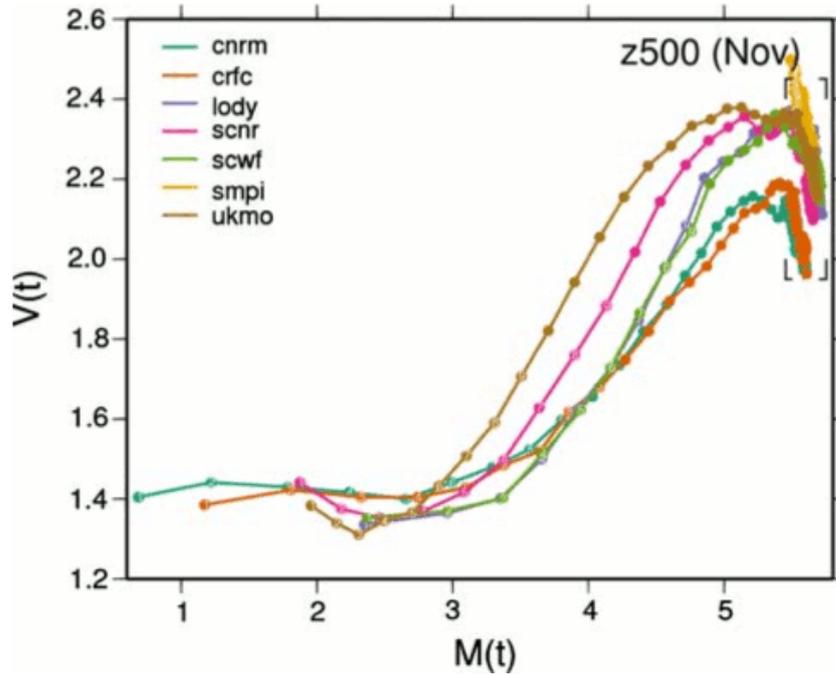
Figure 2.3: MVL diagram for 500 hPa geopotential height in the DEMETER models (Fernández et al. 2009, fig. 3).

higher levels where we see effectively random perturbations – path (A) in figure 2.4.

The third path (C) in figure 2.4 is not seen in any of the DEMETER models, and is expected for initial perturbations which are already consistent with the model dynamics. For seasonal forecasts, the presence of an initial transient period of hours or days is unlikely to be a problem; however some of the techniques used in medium-range and short-range ensemble forecasts are designed to produce such perturbations, as we shall see in chapter 3.

There has also been some work, as yet unpublished, using the MVL diagram to compare different techniques for generating initial perturbations, both in a simplified model (Pazó et al. 2009) and in an operational model (Primo & Magnusson 2009). We shall return to these in section 3.2.8 after reviewing the different perturbation techniques used in the medium-range models.
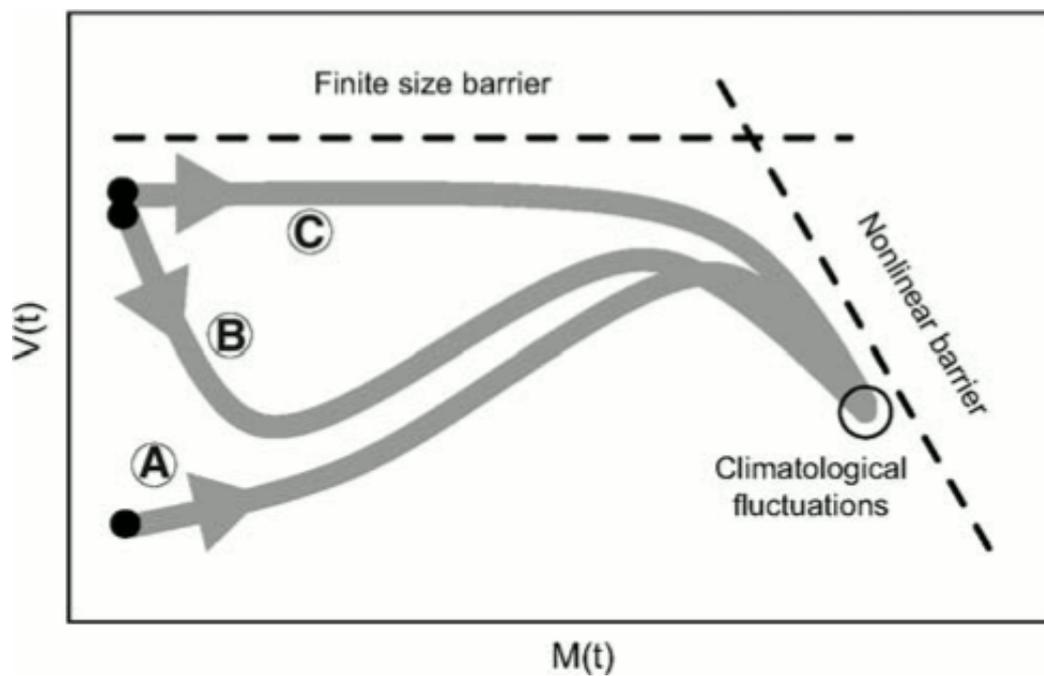
Figure 2.4: Conceptual picture of MVL curves for different perturbation types: (A) random, (B) spatially-correlated but dynamically inconsistent, and (C) spatially-correlated and dynamically consistent (Fernández et al. 2009, fig. 1).

Chapter 3

Medium-range models: TIGGE

3.1 TIGGE: the THORPEX Interactive Grand Global Ensemble

While the DEMETER project provides a useful archive of ensemble forecasts to analyse, these are *seasonal* forecast models designed for forecasts on a much longer timescale than the 15–20 days in which the initial atmospheric perturbations typically saturate.

Medium-range forecast models, on the other hand, are designed for precisely this timescale – where there is too much uncertainty for deterministic forecasts to be useful but before saturation has destroyed all trace of the initial atmospheric state.

A new project, TIGGE[1], provides an archive of medium-range ensemble forecasts from a range of different operational models from 2007 onwards (exact start times vary between the different models). A standardised set of model products is available across the range of models, eliminating the need to handle each model's native output separately.

The models and their relevant properties are listed in table 3.1; the geographical spread of the centres is shown in figure 3.1. Note that Météo-France has been omitted as their ensemble forecast only runs to a lead time of 60 hours which is too short for this analysis.

---

[1] http://tigge.ecmwf.int/

	Centre	Location	Horiz. res.	Levels	Lead time	Members	Perturbations	Region
	BOM	Australia	T119	19	10 days	1 + 32	SV (initial)	< 20°S, > 20°N
	CMA	China	T213	31	10 days	1 + 14	BV	global
	CMC	Canada	0.9°	58	16 days	20	EnKF (since 2005)	global
	CPTEC	Brazil	T126	28	15 days	1 + 14	EOF	45°S − 30°N
	ECMWF	Europe	T399	62	15 days	1 + 50	SV (initial+final) + stoch.	< 30°S, > 30°N, trop. cyc.
	JMA	Japan	T319	60	9 days	1 + 50	SV (since Nov 2007)	> 20°S
	KMA	South Korea	T213	40	10 days	1 + 16	BV	> 20°N
	NCEP	USA	T126	28	15 days	1 + 20	ETR (since May 2006) + stoch.	global
	UKMO	UK	1.25° × 0.833°	38	15 days	1 + 23	ETKF + IAU + stoch.	global

Centre Abbreviated name of the centre.

Location Geographical location of the centre.

Horiz. res. Horizontal model resolution (Tnnn refers to a triangular truncation in a spectral model.)

Levels Vertical levels in the model.

Lead time The maximum lead time to which the model runs (or to which its forecasts are archived in TIGGE).

Members Number of perturbed ensemble members (1+ indicates an additional unperturbed control run)

Perturbations Type of perturbations:

BV Bred Vectors (section 3.2.1)

EOF Empirical Orthogonal Functions (section 3.2.3)

EnKF Ensemble Kalman Filter (section 3.2.4)

ETKF Ensemble Transform Kalman Filter (section 3.2.6)

ETR Ensemble Transform with Rescaling (section 3.2.5)

IAU Incremental Analysis Update (section 3.2.7)

SV Singular Vectors (initial- and/or final-time) (section 3.2.2)

stoch. Stochastic parameterisation/physics

Region Region(s) of the globe over which perturbations are optimised.

Table 3.1: Characteristics of the TIGGE models, sourced from the TIGGE model documentation (http://tigge.ecmwf.int/models.html) with missing information taken from the latest available WMO technical progress reports from each centre (http://www.wmo.int/pages/prog/www/DPFS/ProgressReports/ – Champion (2008) was of great help in finding these) or Park et al. (2008).
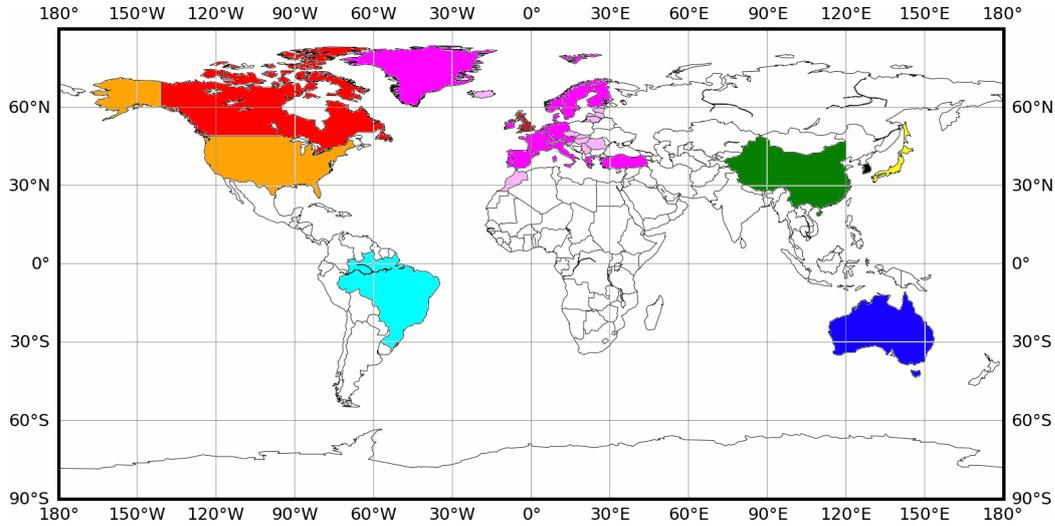
Figure 3.1: Geographical spread of the modelling centres in the TIGGE archive. Colours correspond to those used in the graphs which follow.

Thus it is now possible to apply the techniques described to analyse the spatiotemporal behaviour of these models, where the period of growing spatial correlation is within the operational forecast range.

3.2   Perturbation techniques

As is clear from table 3.1, the TIGGE models use a range of different techniques to generate their initial perturbations. In this section we present a brief overview of these techniques.

3.2.1   Bred Vectors (BV)

In the "bred vector" approach pioneered at NCEP (Toth & Kalnay 1993, 1997) and used at several other centres, random perturbations are grown using the model equations, allowing the fastest-growing vectors to dominate, in a manner that aims to mimic the growth of analysis errors in the data assimilation cycle. This is done by running a second short-range forecast from a randomly-perturbed initial condition alongside the short-range forecast used for the data assimilation to construct the analysis. After each cycle, the

forecast perturbation is rescaled to its original magnitude and added to the new analysis to produce the perturbed initial condition for the next cycle, as shown in figure 3.2. After a number of such cycles, the resulting perturbation is added to (or subtracted from) the operational analysis to create the initial condition for an ensemble member of the actual medium-range forecast run.

Multiple breeding cycles may be run in parallel, starting with different perturbations, to produce the required number of ensemble members. At least in the NCEP implementation, a regionally-varying rescaling is also applied to the perturbations to ensure that their distribution matches the time-average analysis error.

In this way, the method attempts to find the fastest-growing *finite*-size perturbations according to the *nonlinear* model dynamics, as opposed to the Lyapunov vectors, which are the fastest-growing infinitesimal ones under the linearised model dynamics. In addition, because the initial perturbations are generated as genuine perturbations of the full dynamical model, the perturbed initial state is expected to remain close to the model attractor.

NCEP have now moved to a more sophisticated variant of the breeding technique based on the Ensemble Transform (section 3.2.5); however the standard bred vector technique remains in use at CMA and KMA.

3.2.2 Singular Vectors (SV)

In the singular-vector technique pioneered at ECMWF (Buizza & Palmer 1995, Molteni et al. 1996), and also used at BOM and (since November 2007) JMA, we find the perturbation vectors with the largest growth over a finite time interval (typically 48 hours) under the linearised dynamics – in contrast, the Lyapunov vectors represent the fastest *long-term* growth.

We take the propagator $M(T, T_0)$ of the tangent linear model at the optimisation time T_0, and its adjoint M^* with respect to a suitable inner product $(\cdot\,;\cdot)$ so that

$$(Mx; y) = (x; M^*y) \tag{3.1}$$

for any pair of states x and y.

BREEDING OF GROWING MODES
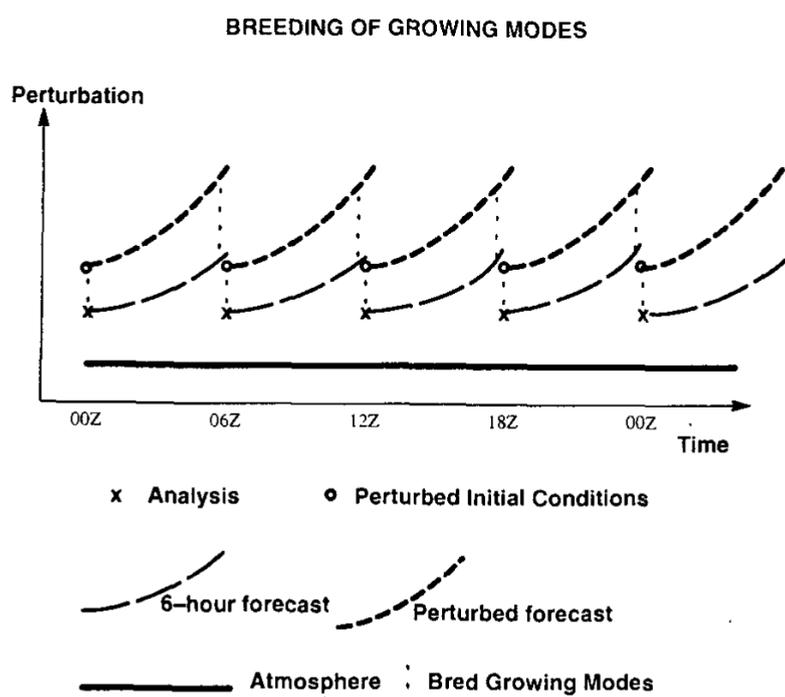
FIG. 4. Schematic of the 6-hour breeding cycle. Note that the breeding cycle depends on the analysis cycle but does not affect it. *A small arbitrary perturbation is introduced on the control analysis* initially. After a 6-h integration, the difference between the control and perturbed forecasts is scaled back to the size of the initial perturbation and this difference field is now added onto the new analysis. After 3–4 days of cycling, the perturbation is dominated by growing modes due to the "natural selection" of fast-growing perturbations.

Figure 3.2: Illustration of the bred vector technique (Toth & Kalnay 1993, fig. 4).

The operator M^*M is Hermitian, and thus its eigenvectors form an orthonormal basis for the state space. These eigenvectors are referred to as the (right) singular vectors of M; the square roots of the corresponding eigenvalues of M^*M are the singular values of M. The vectors associated with the largest singular values represent the perturbations which grow largest, as measured by the norm induced by the chosen inner product. The "dry total perturbation energy" inner product and norm are frequently chosen, however other choices have been used experimentally, for instance "moist total perturbation energy" (Ehrendorfer et al. 1999) and "Hessian singular vectors" (Barkmeijer et al. 1998).

The vectors described above are "initial-time" SVs, which have the fastest growth over the period $[T_0, T]$ following the analysis time T_0 and tend to be highly localised in space; there are also "final-time" or "evolved" SVs, which have the fastest growth over the period $[-T, T_0]$ preceding the analysis time and tend to have a larger spatial scale. Many SV implementations (e.g. BOM) use only initial-time vectors, although ECMWF uses a combination of initial- and final-time vectors (Bourke et al. 2004). Because the singular vectors tend to be highly localised, the actual perturbations are generated by taking linear combinations of the SVs.

Although not shown above for simplicity, a projection operator is often included in the formulation to maximise the perturbation energy over a specific portion of the globe (e.g. in the ECMWF implementation separate SVs are generated for the northern and southern extratropics, and in the vicinity of tropical cyclones).

3.2.3   Empirical Orthogonal Functions (EOF)

In this technique (Zhang & Krishnamurti 1999), unique to CPTEC amongst the TIGGE models, random perturbations are grown over a short time period (typically 36 hours) to produce time series of perturbation fields, which are

formed into a single matrix W for a model field w

$$
W = \begin{pmatrix}
\delta w_{11} & \delta w_{12} & \cdots & \delta w_{1S} \\
\delta w_{21} & \delta w_{22} & \cdots & \delta w_{2S} \\
\vdots & \vdots & \ddots & \vdots \\
\delta w_{T1} & \delta w_{T2} & \cdots & \delta w_{TS}
\end{pmatrix},
\tag{3.2}
$$

where the S columns are the model grid-points, and the T rows are the time steps over the calculation period. In the case of the horizontal wind field, a complex representation $\delta w = \delta u + \mathrm{i}\,\delta v$ is used to represent the zonal and meridional wind perturbations in a single matrix. These are subjected to an EOF (or Principal Component) analysis: we define the ($N \times N$, Hermitian) covariance matrix

$$
H = \frac{1}{T} W^* W
\tag{3.3}
$$

(where * represents the complex conjugate transpose) and express each row of W (which represents the model output at a given time) in terms of the basis of eigenvectors of H. The eigenvectors whose coefficients grow most rapidly with time are then selected for use as initial perturbations in the ensemble.

The current operational model at CPTEC perturbs the horizontal wind field (with the two components treated as a single complex field, as described above) and the temperature field; Mendonça & Bonatti (2009) show that improvements can be obtained by including additional perturbations in the pressure and humidity fields.

3.2.4 Ensemble Kalman Filter (EnKF)

In this approach, unique to CMC amongst the TIGGE models (Houtekamer & Mitchell 2005), an ensemble Kalman filter (Evensen 1994, Houtekamer & Mitchell 1998) is used for operational data assimilation; in this formulation, the analysis error covariance is estimated from the covariance of the ensemble members used in the data-assimilation cycle. By using a sample of these

ensemble members directly to provide the perturbed initial conditions, it is assured that the initial perturbations will have the same statistics as the estimated analysis error, and also that they are consistent with the model dynamics.

CMC switched to this method in 2005 from an earlier technique which used multiple runs of an Optimal Interpolation data assimilation scheme with perturbed observations (Houtekamer et al. 1996).

3.2.5   Ensemble Transform with Rescaling (ETR)

This is a modern variant on the bred-vector technique, used at NCEP (Wei et al. 2008) since May 2006. Instead of independently rescaling each bred perturbation after each cycle, the ensemble transform (ET) applies a matrix transformation to the whole set of n perturbations to produce $n-1$ orthogonal perturbations scaled according to the analysis error covariances from the operational (variational) data assimilation scheme. A simplex transformation (Wang et al. 2004) is then used to produce an ensemble of n members centred around the operational analysis while preserving the analysis error covariance. The differences between this method and the original bred-vector technique are illustrated in figure 3.3.

The "with rescaling" refers to a final step in which the perturbations are regionally rescaled to fit the expected geographical distribution, in a similar manner to the regional rescaling step mentioned in section 3.2.1 for bred vectors.

3.2.6   Ensemble Transform Kalman Filter (ETKF)

The ETKF (Bishop et al. 2001) provides a computationally-efficient way of calculating analysis perturbations which are approximately consistent with the EnKF (albeit constrained to be in the subspace spanned by the forecast perturbations) without running the whole EnKF process. This is particularly useful where a non-EnKF data assimilation scheme (e.g. 4D-var) is being used operationally – the ETKF-generated perturbations can be added to the operational analysis (using a simplex transformation for centring as in the
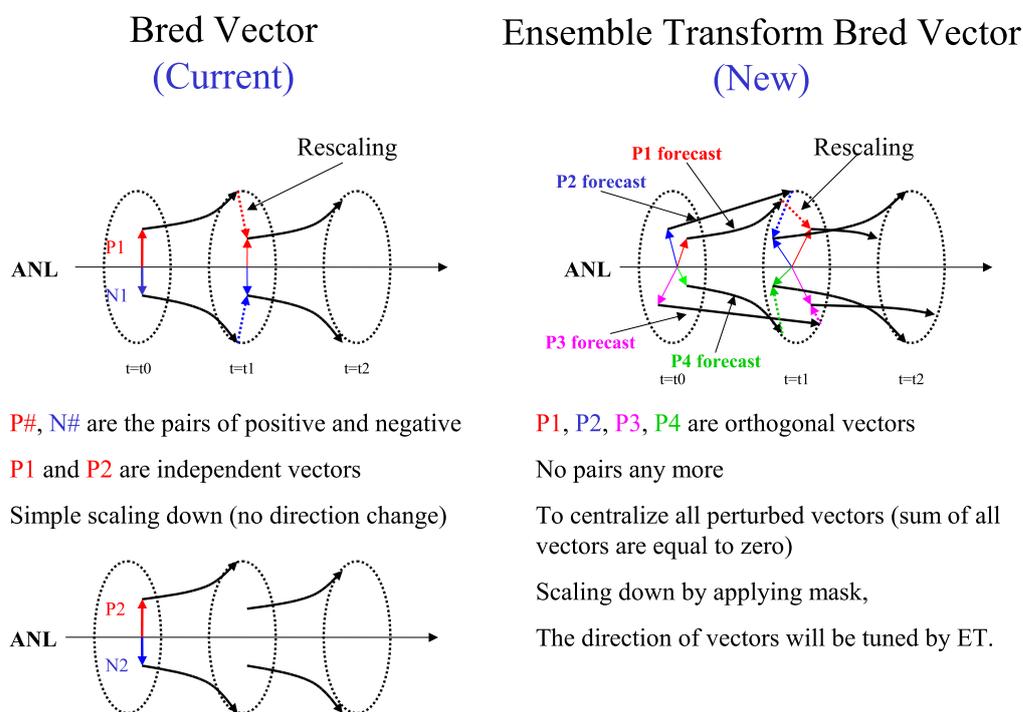
Figure 3.3: Illustration of the Ensemble Transform (ET) technique (Zhu & Toth 2006).

ET case), harnessing much of the ensemble-generating power of the EnKF without replacing the data assimilation scheme.

It is currently only used operationally at the UK Met. Office (Bowler et al. 2008), although it has been used experimentally at NCEP (Wei et al. 2006).

3.2.7 Incremental Analysis Update (IAU)

This is an ancillary technique (Bloom et al. 1996) used by the UK Met. Office to introduce the initial perturbations gradually over the first 6 hours of the forecast period. This results in an apparent very rapid growth from almost zero perturbation amplitude over that initial time period. The reasons for this are not entirely clear, but may be related to the model's origin in short-range ensemble forecasting (Bowler et al. 2008) where transient effects caused by the instantaneous application of perturbations may degrade fore-

cast performance over the first few hours if the model is pushed away from
its attractor.

3.2.8   Comparing different perturbation techniques

Although a variety of perturbation techniques are in operational use, it has
proved difficult to separate the differences in performance due to the ini-
tial perturbations from those due to other aspects of the model formulation
(Buizza et al. 2005). To address this, a number of studies have been carried
out comparing different techniques in one model, in terms of ensemble-mean
and probabilistic skill measures:

- Bowler (2006) and Descamps & Talagrand (2007) both compare the
  BV, SV, EnKF and ETKF techniques in the context of a simplified
  model (Lorenz 1996). Both find that the EnKF and ETKF techniques
  have significant advantages due to their initial perturbations better
  sampling the analysis error PDF. However on many performance mea-
  sures, random perturbations perform equally well; this cannot be ex-
  pected to hold for larger models, and shows the limitations of using a
  low-dimensional model for such experiments.

- Wei et al. (2008) compare the BV, ET(R) and ETKF techniques using
  the operational NCEP model, showing improved probabilistic skill from
  the ETR and ETKF methods, with the best results from ETR.

- Magnusson et al. (2008) compare the BV and SV techniques using the
  operational ECMWF IFS model, showing conflicting results in different
  regions: SV performs better in the NH, but BV performs better in the
  tropics, particularly for the early part of the forecast; the two methods
  perform similarly in the SH.

- In forthcoming papers, Pazó et al. (2009) and Primo & Magnusson
  (2009) apply the MVL technique described in chapter 2 to evaluate the
  effect of different initial perturbations on the spatiotemporal dynamics
  of the Lorenz (1996) model and the operational ECMWF EPS model
  respectively.

3.3 Previous studies

As TIGGE is a relatively new project, there have been only a small number of published studies using this data:

- Champion (2008) gives an overview of the different models, and looks at the differences in (global root-mean-square) perturbation amplitude and growth rate.

- Park et al. (2008) is focused more on evaluating the performance of the different models, giving comparative results for root-mean-square error (RMSE) and ranked probabilistic skill score (RPSS), and also looks at the issues involved in producing a multi-model ensemble – something we will consider further in chapter 5.

- Johnson & Swinbank (2009) consider further the calibration of such a multi-model ensemble (specifically using the ECMWF, NCEP and UKMO models).

- Pappenberger et al. (2008) and He et al. (2009) look at the application of the combined ensemble data from TIGGE to flood warning systems.

In the present study, however, we wish to concentrate instead on the spatiotemporal dynamics of the perturbations in these models, and in particular to ask what the MVL diagram introduced in chapter 2 can tell us about their behaviour. In particular, can we relate similarities and differences between the MVL curves of the different ensembles to known similarities and differences between either the forecast models or the techniques used to generate the initial perturbations?

Chapter 4

Results

Results from applying the MVL technique to these models over the whole globe, for a period of 21 days in February 2009, are shown in figure 4.1. The large number of forecast realisations (models $\times$ initialisations $\times$ ensemble members) makes this figure somewhat crowded; figure 4.2 shows the plot for each model separately. These "spaghetti" plots indicate that the means over realisations (the bold lines in figures 4.1 and 4.2, and shown on their own in figure 4.3) are representative of a typical realisation or ensemble member, although the spread amongst realisations does vary between models. We can clearly see distinct characteristics of the models in these curves; in particular:

- There is a large range in the initial amplitude $M(0)$ of the perturbations between different models. This may be partly due to the ensembles being calibrated to the analysis error estimates produced by different data assimilation procedures used at each centre. The UKMO curve is deceptive here, as their perturbations are introduced gradually over the first 6 hours of the forecast using the Incremental Analysis Update technique (Bloom et al. 1996); this shows as an artificially-small initial perturbation with very rapid growth in the first 6-hour time period.

- The BOM, CPTEC, ECMWF and JMA curves show a marked decrease in variance in the first stage of perturbation growth; this is typically observed while the model adjusts to initial perturbations which are not fully consistent with the model dynamics (Fernández et al. 2009). It is
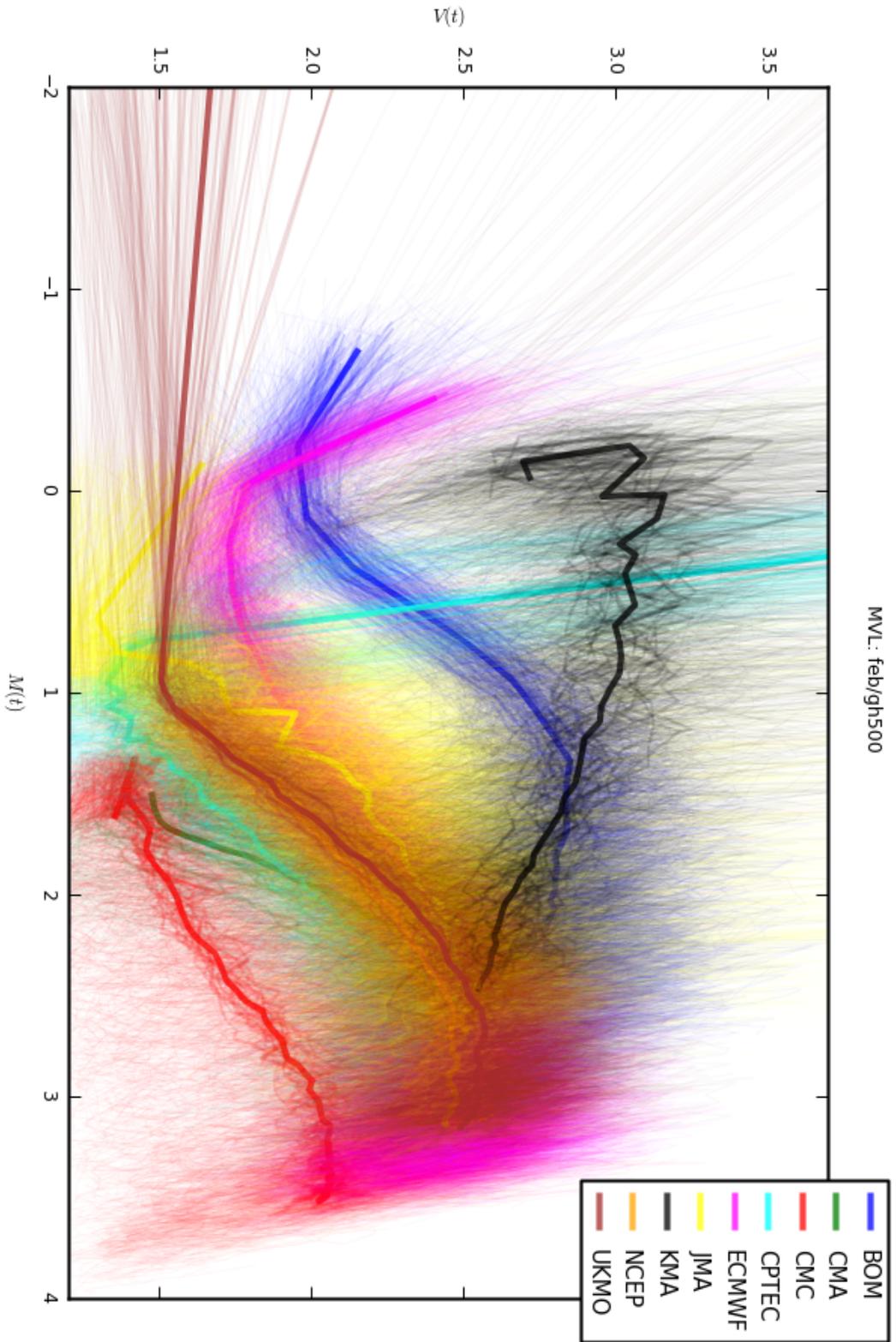
Figure 4.1: Combined MVL diagram for 500 hPa geopotential height in the TIGGE models, across the whole globe, for February 2009. The bold lines plot the mean of $M(t)$ and $V(t)$ across all ensemble members and all forecast initialisations for each lead time t; the medium-weight lines plot the mean over all ensemble members for each initialisation; the finest lines are the individual ensemble members for each initialisation.
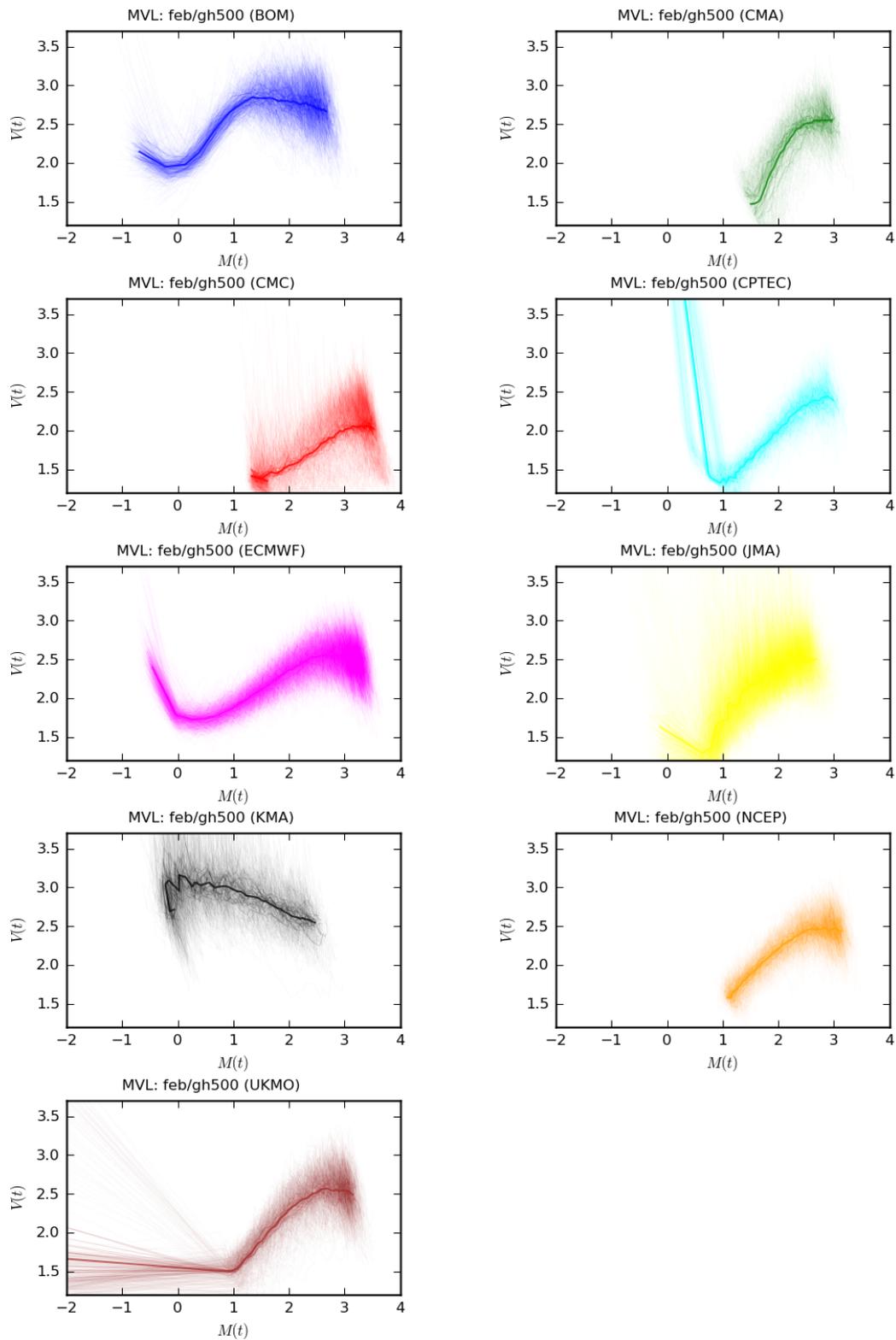
Figure 4.2: Separate thumbnail MVL diagrams for 500 hPa geopotential height in each of the TIGGE models, across the whole globe, for February 2009.
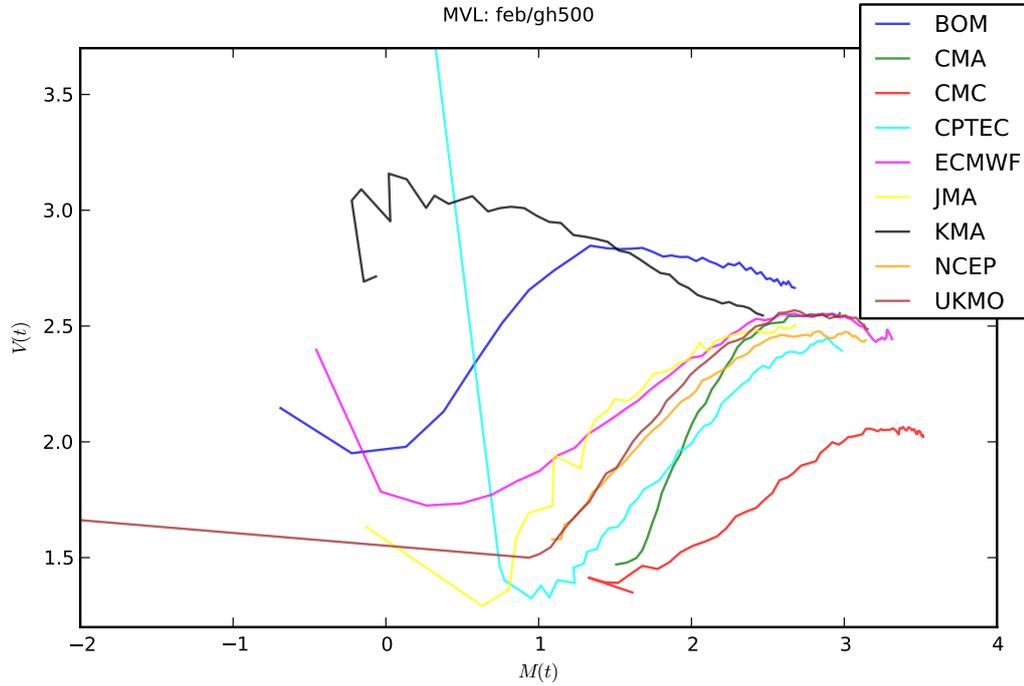
Figure 4.3: Mean-over-realisations MVL diagram for $500\,\mathrm{hPa}$ geopotential height in the TIGGE models, across the whole globe, for February 2009.

worth noting that these are precisely those models using the SV and EOF techniques (see table 3.1), although this is not necessarily the cause.

- The CMC and KMA curves appear to show a *decrease* in perturbation amplitude (leftward motion on the plot) over at least one of the early time periods, contrary to the expected linear average growth in M; it is not clear why this should be the case.

- Although there is some variation in the perturbation amplitude at which the variance saturates (i.e. the horizontal position of the main peak on the plot), most of the models behave similarly in this respect, suggesting that they transition from the linear to the nonlinear regime at a similar amplitude. The notable exceptions are BOM, CMC and (especially) KMA, which saturate at a lower or higher amplitude than the rest.

There are several questions which arise as to the generality of the above results. We might reasonably ask whether this picture is valid across the whole globe, given that (amongst other reasons) the dynamics of tropical and mid-latitude meteorology are known to be different. We might also consider seasonal variability: does an equivalent plot for a different month look similar? (We could also consider interannual variability; however the TIGGE dataset, unlike DEMETER, doesn't span a long historical period – less than two years for most models.) Finally, we are looking here at one model variable (geopotential height) on one level surface (500 hPa) – are the dynamics the same at other levels, or for other model variables (e.g. temperature or wind speed).

Fernández et al. (2009) looked at two of these questions – seasonal variability and differences between model variables (500 hPa geopotential height and 2 m temperature) – in the case of the DEMETER seasonal forecast models. There they found little difference between seasons, but a significant qualitative difference in the dynamics between variables, as noted in section 2.2.

4.1 Geographical variability

To investigate variation between different latitude bands, the MVL diagrams can be calculated separately for each band, as shown in figure 4.4. We have chosen $0° - 30°$ (tropics) and $30° - 75°$ (extratropics) in each hemisphere; this captures the relevant features without introducing an unnecessary number of bands.

It is clear from figure 4.4 that the differences between northern and southern extratropics are slight for most models. The notable exception of KMA is likely explained by the fact that their initial perturbations are optimised over only one of these regions (north of $20°$ N – see table 3.1); thus the perturbations remain small over the rest of the globe as the fast-growing perturbations in those regions are not effectively sampled by the ensemble.

On the other hand, there are quite strong differences between the tropics and extratropics. In all the models, the perturbation amplitude grows larger in the extratropics than in the tropics by the end of the forecast period. Also,
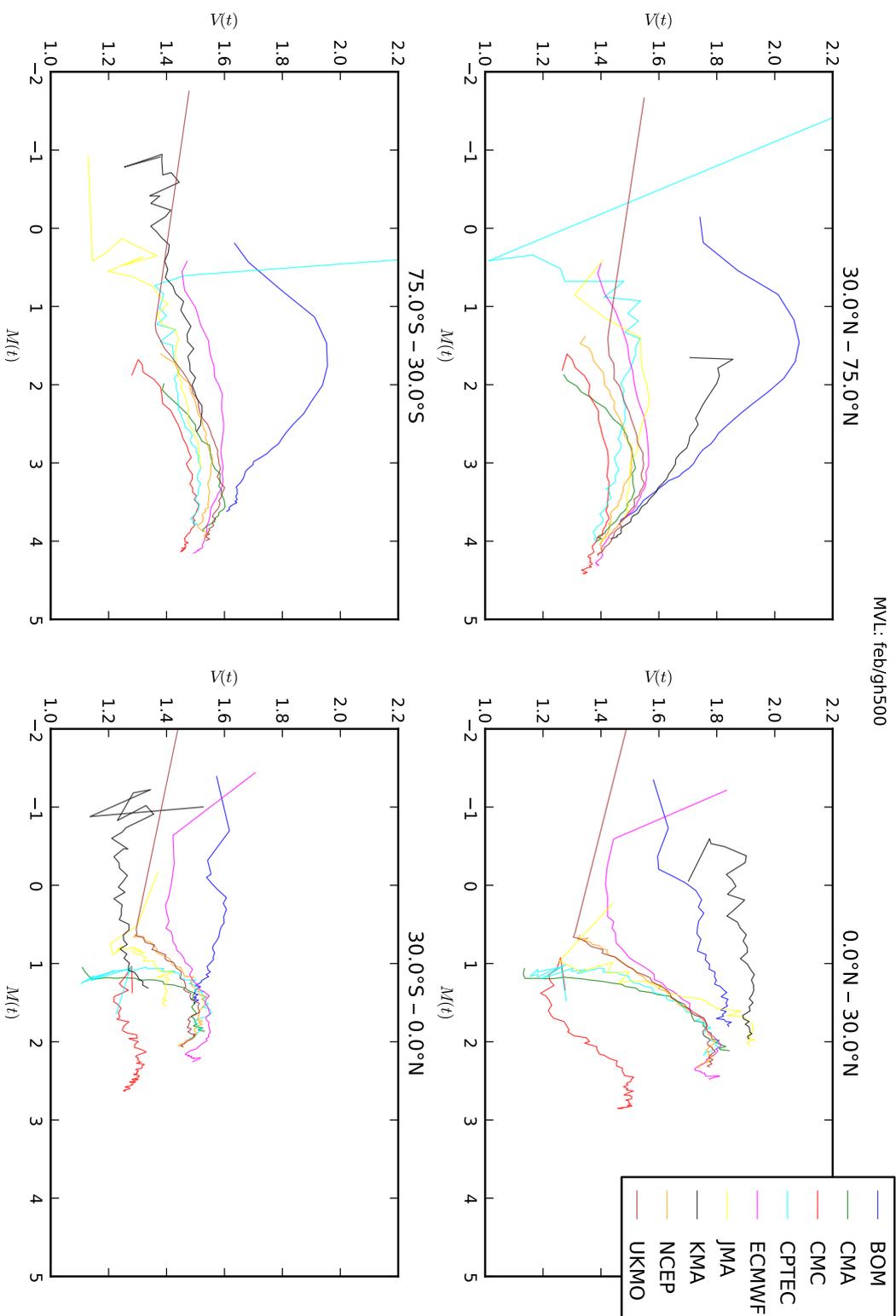
Figure 4.4: MVL diagrams for 500 hPa geopotential height in the TIGGE models, NH (top) and SH (bottom) extratropics (left) and tropics (right), for February 2009.

while in the extratropics the models all reach a similar "climatological" region of the diagram by the end of the forecast period, this is less true in the tropics; this may be due to differences in tropical climatology between the models, or it may be that the tropical dynamics have a longer predictability time and so the ensembles have not dispersed to the full range of climatology by the end of the forecast period.

There are also noticeable differences between the northern and southern tropics, with a much greater growth in spatial correlation in the north for all the models (except KMA whose optimisation region covers very little of the tropics). A seasonal explanation seems likely, perhaps related to the movement of the inter-tropical convergence zone (ITCZ) – we shall return to this matter in section 4.2.

We previously noted the initial decrease in variance seen in some of the models (BOM, CPTEC, ECMWF and JMA); this behaviour appears to have a geographical component. In the case of ECMWF, it is very pronounced in the tropics but barely discernible in the extratropics. As ECMWF has a separate perturbation strategy (also SV-based) for the tropics[1], it appears that this, rather than the perturbations to the extratropics, is the source of spatial structures which are inconsistent with the model dynamics. CPTEC is rather different, with a very large transient correlation primarily in the extratropics. This appears to be an artefact of the perturbation scheme, which produces perturbations concentrated in an equatorial band (the optimisation region is $45°\,\mathrm{S} - 30°\,\mathrm{N}$; see figure 4.5). The gradual decrease in perturbation amplitude outside this region leads to the very high variance in the extratropics. Mendonça & Bonatti (2009) discuss some of the issues surrounding this particular model in more detail. In particular, they show that a combination of additional EOF-based perturbations optimised over the extratropics and perturbing a more complete set of model variables improves both the geographical distribution of the initial perturbation field and the skill of the ensemble forecast.

---

[1]`http://www.ecmwf.int/research/predictability/projects/IC_pert/`
`tropical_SV/index.html`
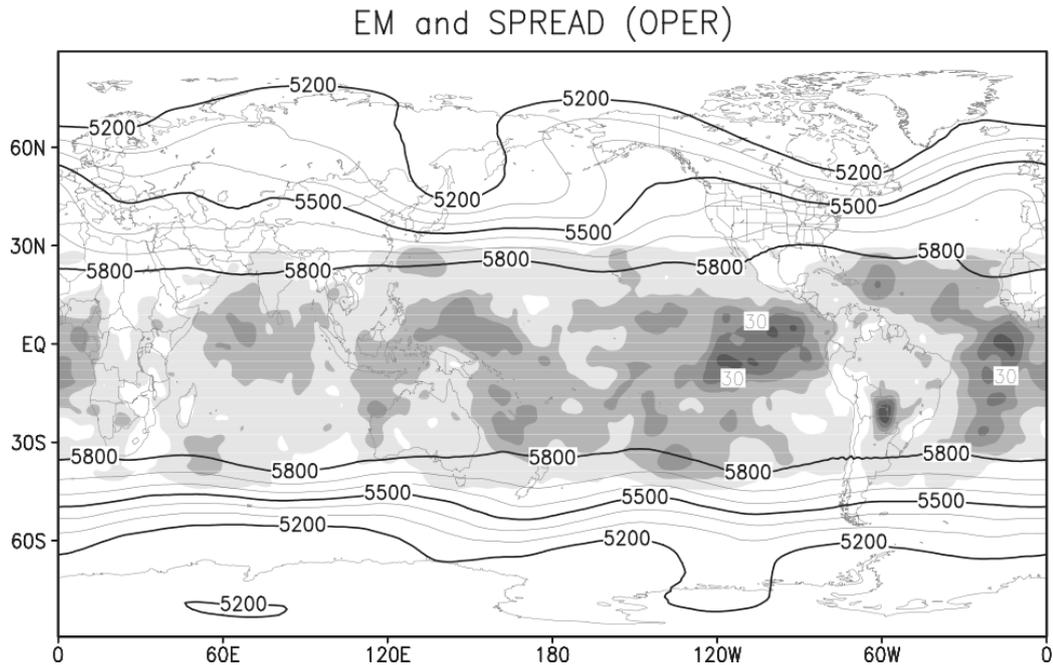
Figure 4.5: Example of initial ensemble state for the CPTEC model, showing the concentration of perturbations in the tropics (Mendonça & Bonatti 2009, fig. 1(a)). The shading represents the initial ensemble standard deviation, with darker shades indicating a larger spread, i.e. larger initial perturbations.

4.2   Seasonal variability

To check for inter-seasonal variations, we compare the previous results for February 2009 (figures 4.3 and 4.4) with their equivalents for August 2008, shown in figures 4.6 and 4.7.

The global plots (figures 4.3 and 4.6) show only minor changes between February and August, as in the DEMETER analysis by Fernández et al. (2009).

Looking at the plots for individual latitude bands however (figures 4.4 and 4.7) it is clear that, while the seasonal changes in extratropical regions are small, there is a "reversal" of the north/south split in the tropics observed in section 4.1: in February the spatial growth is stronger in the northern tropics, while in August it is stronger in the southern tropics. This suggests that the difference is not a geographical one between the two tropical regions,
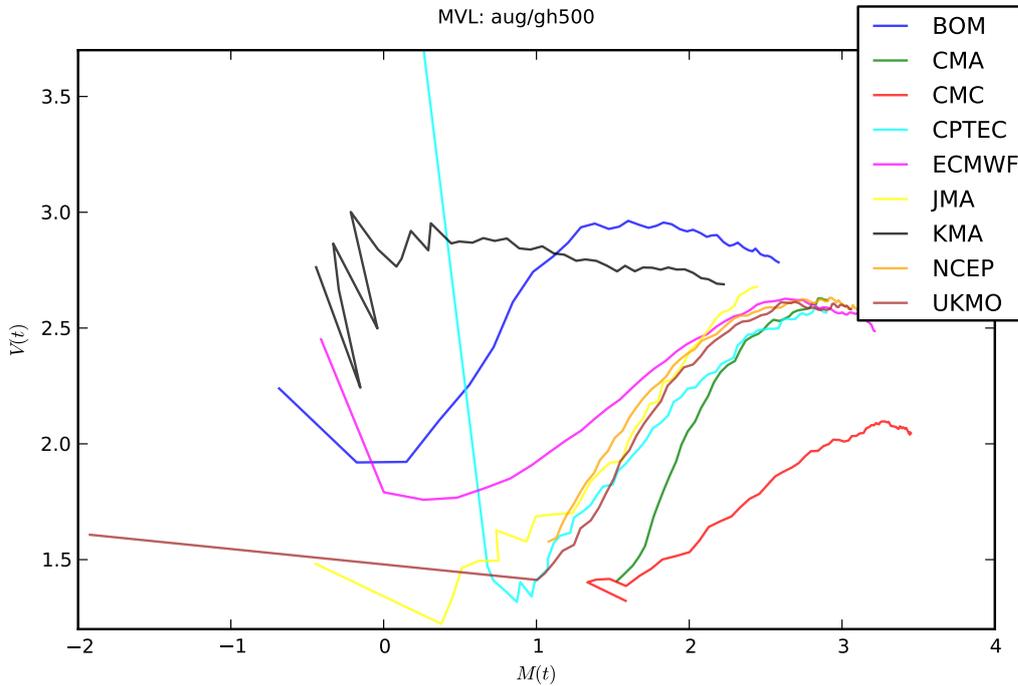
Figure 4.6: MVL diagrams for $500\,\mathrm{hPa}$ geopotential height in the TIGGE models, across the whole globe, for August 2008.

but rather a seasonal one with more spatial growth in the winter hemisphere. There are several factors which may be relevant here: the larger and stronger Hadley cell in the winter hemisphere (Vallis 2006, §11.4), the position of the ITCZ (associated with widespread convective-scale activity) in the summer hemisphere and the timing of monsoons. It may be that perturbations in the summer hemisphere are thus dominated by more more localised convective and monsoonal effects, compared to the winter hemisphere where large-scale structures can grow relatively undisturbed.

4.3 Differences between model variables

4.3.1 Near surface: $2\,\mathrm{m}$ temperature

We can also contrast the results above for $500\,\mathrm{hPa}$ geopotential height with those for a near-surface variable; for this purpose we use the $2\,\mathrm{m}$ temperature
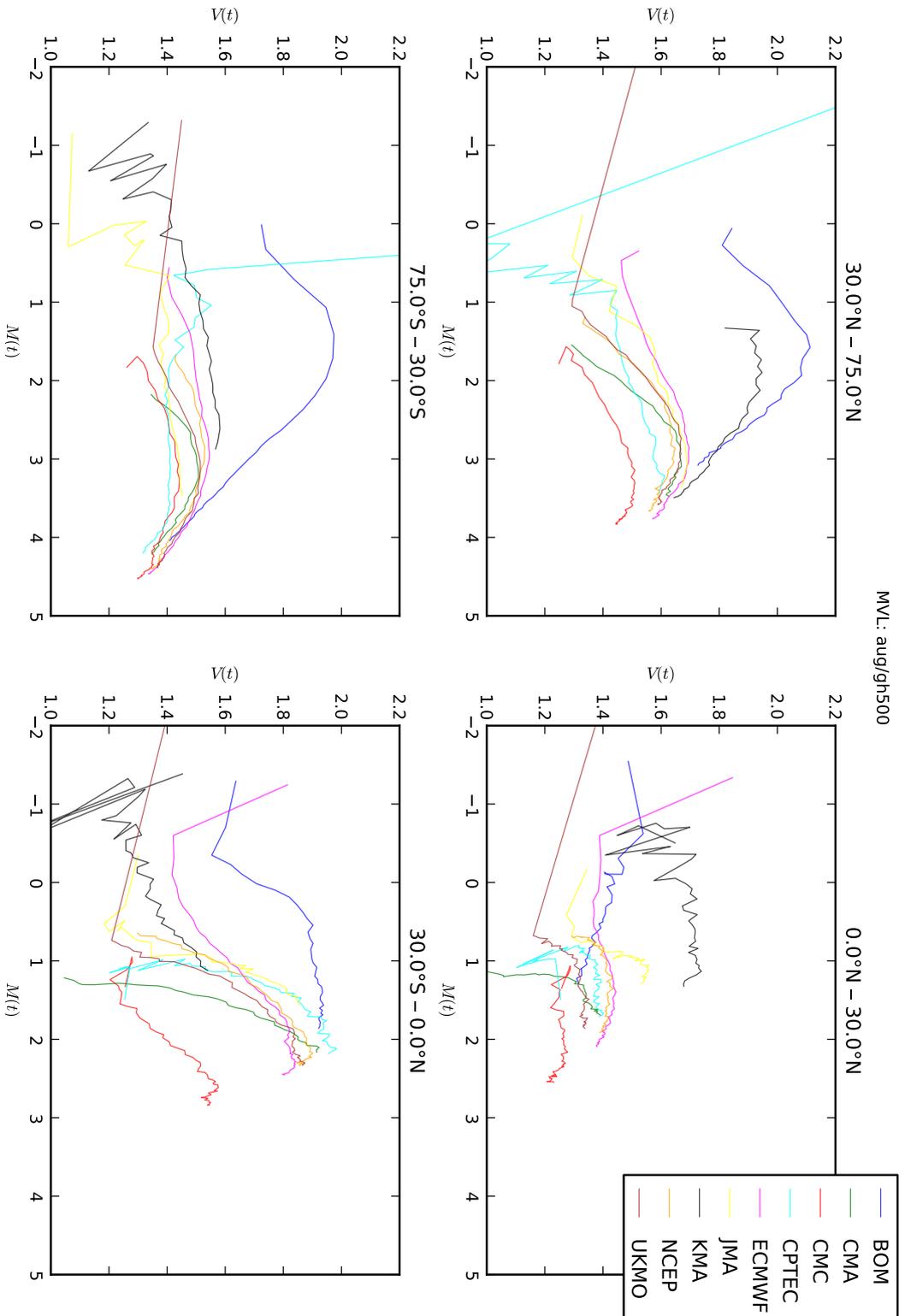
Figure 4.7: MVL diagrams for 500 hPa geopotential height in the TIGGE models, NH (top) and SH (bottom) extratropics (left) and tropics (right), for August 2008.
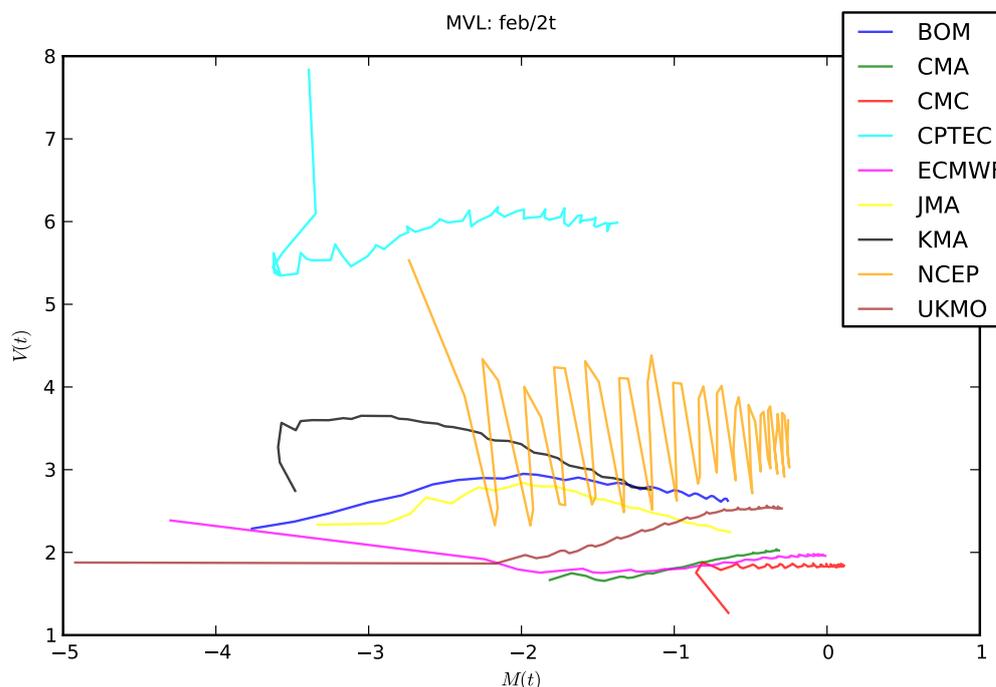
Figure 4.8: MVL diagrams for $2\,\mathrm{m}$ temperature in the TIGGE models, across the whole globe, for February 2009.

field (or $1.5\,\mathrm{m}$ temperature in the case of UKMO, as that is what is provided in the TIGGE archive). These results (for February 2009) are shown in figure 4.8. Two anomalous features are immediately apparent: the CPTEC curve shows a very high variance over the whole forecast period, and NCEP shows very large oscillations in variance with a period of four data points, i.e. 24 hours. We have already seen anomalous behaviour from CPTEC when looking at the results for $500\,\mathrm{hPa}$ geopotential height, and will return to the NCEP oscillations later, but these two features distort the vertical scale to the point where the dynamics of the remaining models are obscured; hence figure 4.9 shows a rescaled version.

Two things are apparent here. First, there is much less agreement between the models than for $500\,\mathrm{hPa}$ geopotential height, even towards the end of the forecast period (especially in the tropics, as can be seen in figure 4.10). Second, the curves are noisier in general and the 24-hour oscillations observed
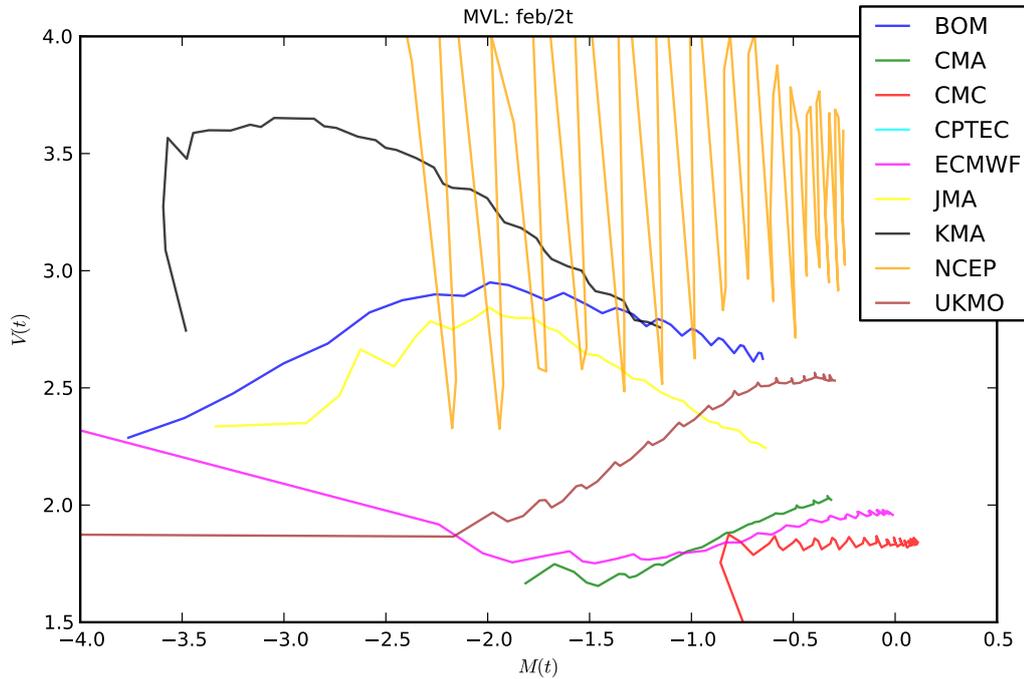
Figure 4.9: MVL diagrams for 2 m temperature in the TIGGE models, across the whole globe, for February 2009 (enlarged).

in the NCEP curve appear to be present in most of the other models, although with a much smaller amplitude.

There are two clear differences from the Fernández et al. (2009) DEME-TER analysis (figures 2.2 and 2.3). Because the medium-range models introduce their perturbations in the full initial model state, they do not show the difference in initial correlation between lower and upper levels that was seen in the seasonal models. Also, the daily oscillations in the 2 m temperature plot appear to be unique to medium-range models; however, this could be misleading since the lower time resolution used for the seasonal models (one point each day) would be insufficient to resolve such an effect even if it were present in those models.

Returning to the daily oscillations we saw particularly in the NCEP curve, we might reasonably ask what causes these. Since all the data points are zonal averages, we should expect simple diurnal cycles to average out. It is conceivable however that they do not average out perfectly, perhaps due to
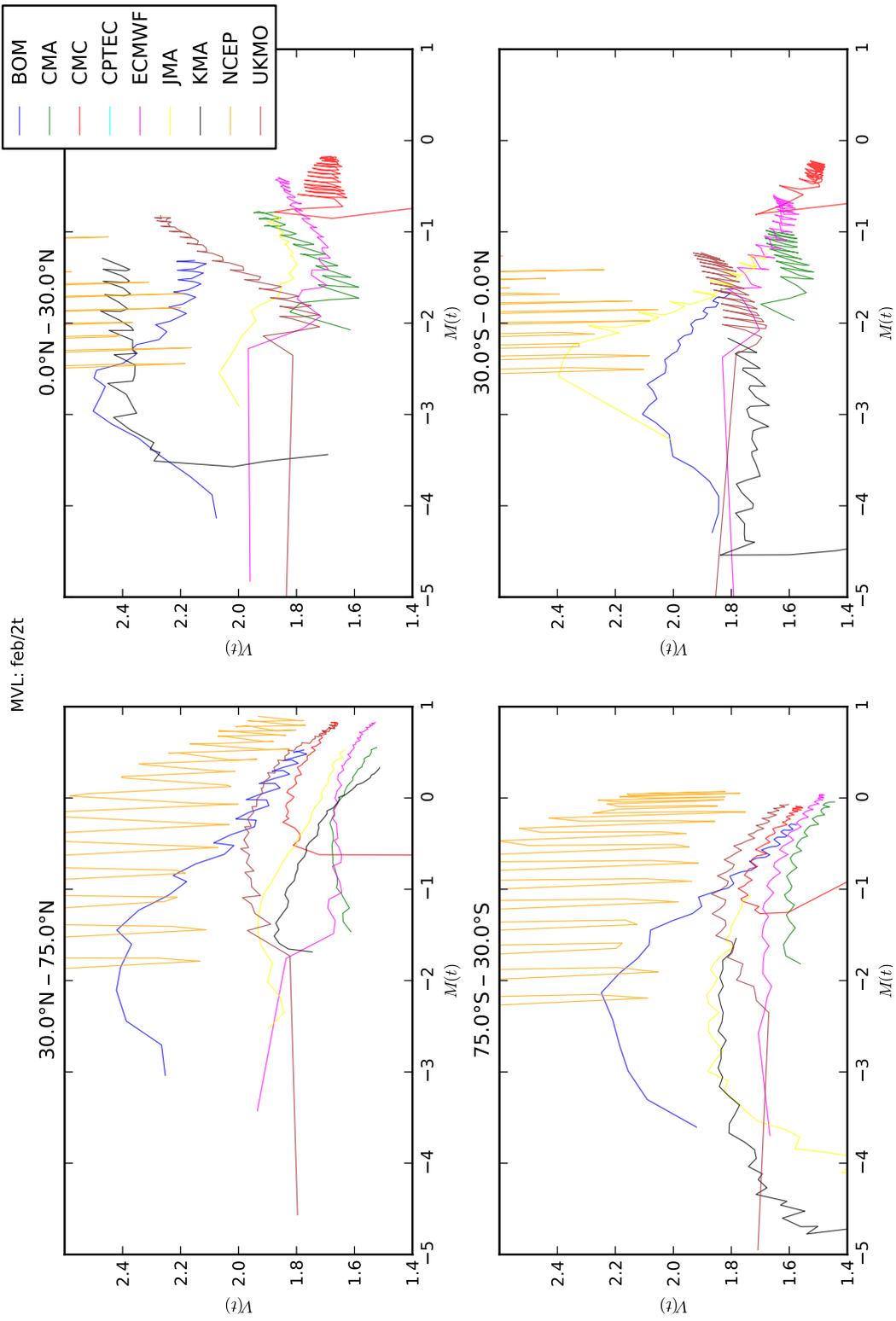
Figure 4.10: MVL diagrams for 2 m temperature in the TIGGE models, NH (top) and SH (bottom) extratropics (left) and tropics (right), for February 2009 (enlarged).

different effects when the sun is over land and ocean. If this were the case, we would expect the peaks to occur at approximately the same *local* time at all longitudes. However, if we plot time series of V against t for separate longitude slices (figure 4.11), we can see that there is little phase difference between the slices – the peaks are at approximately the same *universal* time. This suggests that we may be seeing a model artefact with a 24-hour period (for example introduced by the boundary-layer model required to calculate $2\,\mathrm{m}$ temperature from the raw model variables) rather than a representation of a genuine diurnal effect.

If we look at time series for individual ensemble runs (i.e. the mean over ensemble members for each separate initialisation time) for the first few days of the time period (the bold lines in figure 4.12), we see that this oscillation is not a continuous effect, but occurs in most runs for some portion of the forecast period. There is some variability in these oscillations between individual ensemble members (the faint lines), but they appear largely in agreement for each run.

4.3.2   Upper troposphere: $300\,\mathrm{hPa}$ meridional wind

For completeness, having looked at a mid-level field ($500\,\mathrm{hPa}$ geopotential height) and a near-surface one ($2\,\mathrm{m}$ temperature), we shall look briefly at the upper troposphere and consider $300\,\mathrm{hPa}$ meridional wind. (Ideally we would consider the full vector wind field; however the results for the zonal component are similar and this avoids the extra complexity of implementing vector-based analysis.)

Figure 4.13 shows the results over the whole globe, and once again these are broken down into latitude bands in figure 4.14. The plots are relatively smooth, much like those for $500\,\mathrm{hPa}$ geopotential height (figures 4.3 and 4.4). However the difference between hemispheres in the tropics is less apparent here, and some models show noticeably different behaviour:

- ECMWF shows very little growth in spatial correlation here after the initial transient decrease; for $500\,\mathrm{hPa}$ geopotential height there is sub-
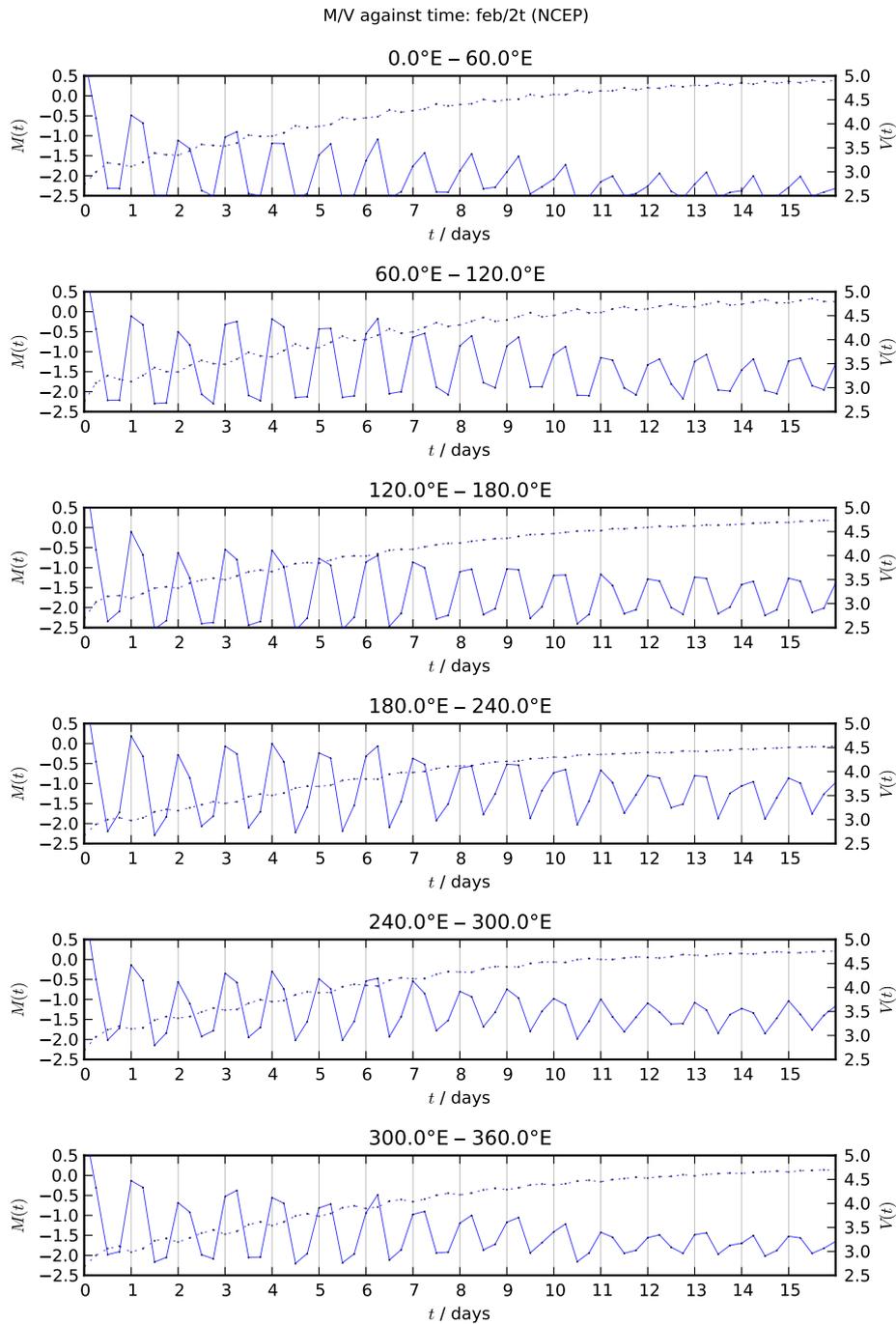
Figure 4.11: Mean-over-realisations time series of M (dotted line) and V (solid line) for 2 m temperature in the NCEP model, for 60° longitude slices, for February 2009.
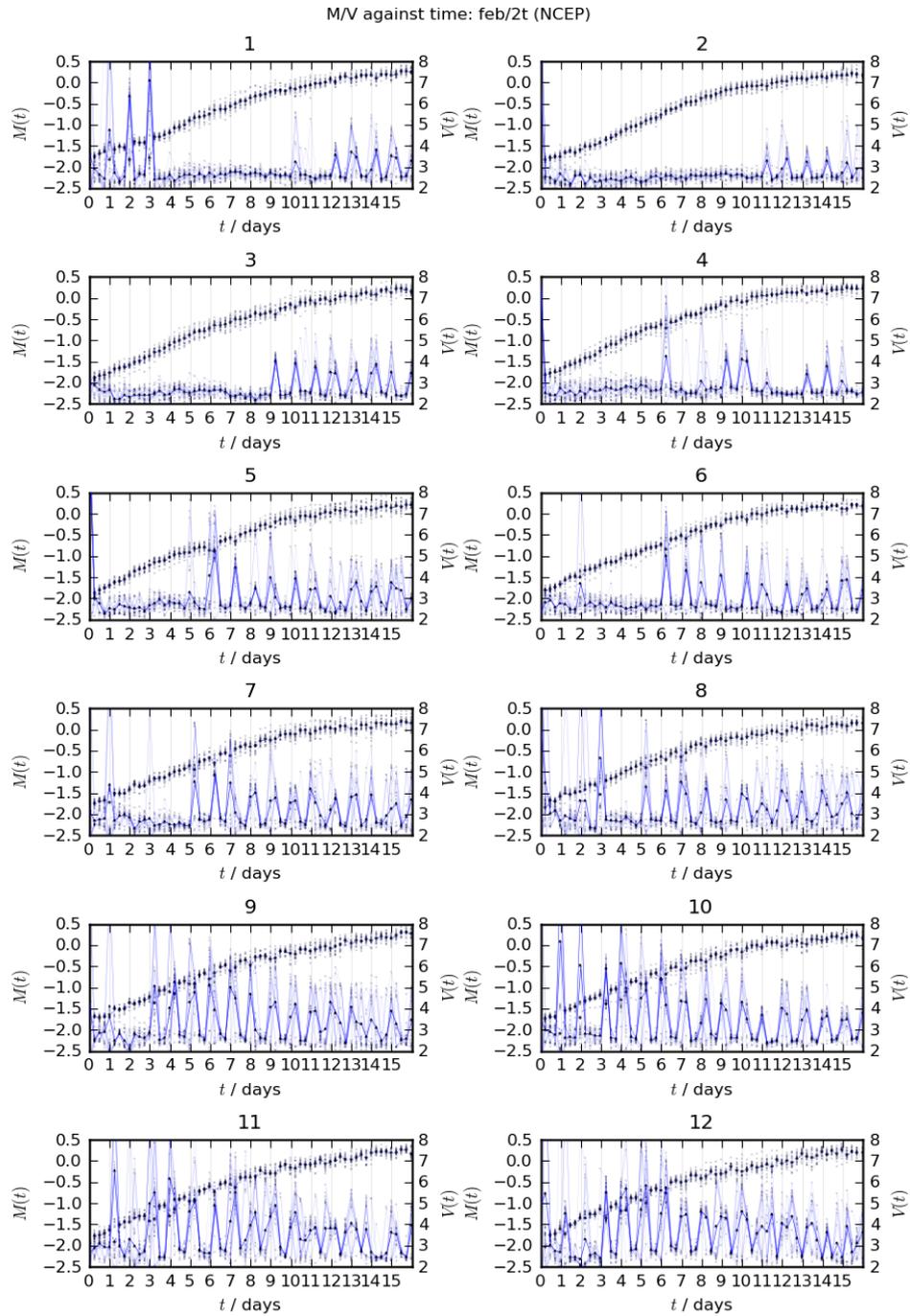
Figure 4.12: Separate time series of M (dotted line) and V (solid line) for $2\,\mathrm{m}$ temperature in the NCEP model, for the forecasts initialised on 1–12 February 2009 The bold lines show the mean of $M(t)$ and $V(t)$ over ensemble members; the faint lines show the individual members.
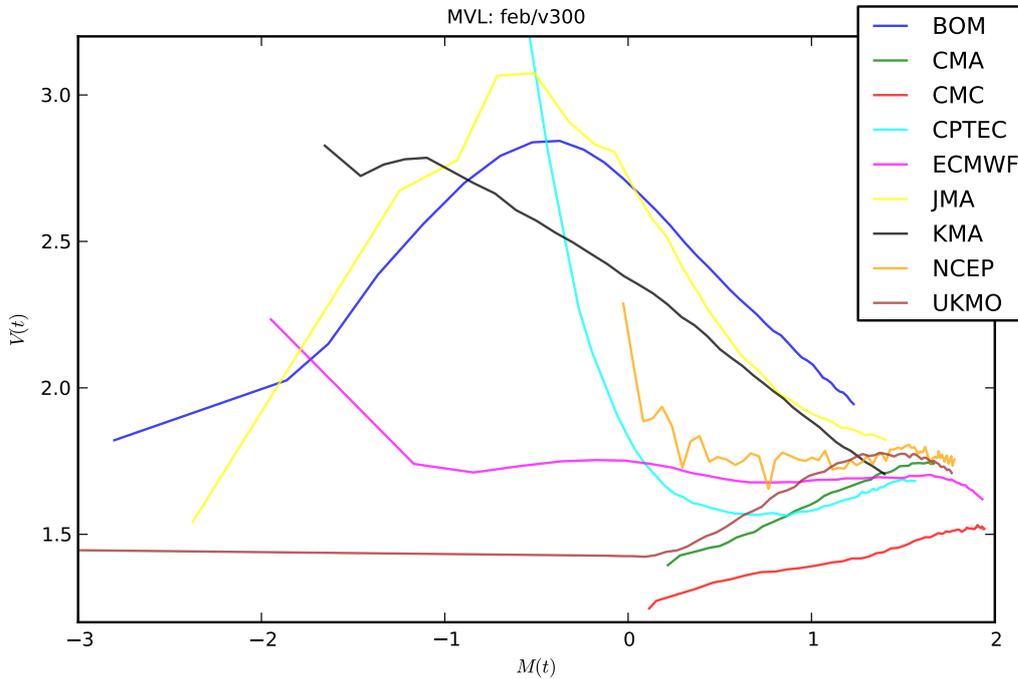
Figure 4.13: MVL diagrams for 300 hPa meridional wind in the TIGGE models, over the whole globe, for February 2009.

stantial regrowth. This is particularly true in tropical regions; some growth can still be seen in the extratropics.

- NCEP also exhibits initial decrease and little regrowth in this case, at all latitudes; this is in contrast to 500 hPa geopotential height, where there is strong growth in spatial correlation from the start.

- JMA shows a very high peak spatial correlation for 300 hPa meridional wind, especially in the tropics, although that for 500 hPa geopotential height is close to the majority of models.

4.4 Perturbations vs. errors

One question which arises when we look at the spatiotemporal dynamics of the perturbations in these ensembles is to what extent they mimic the
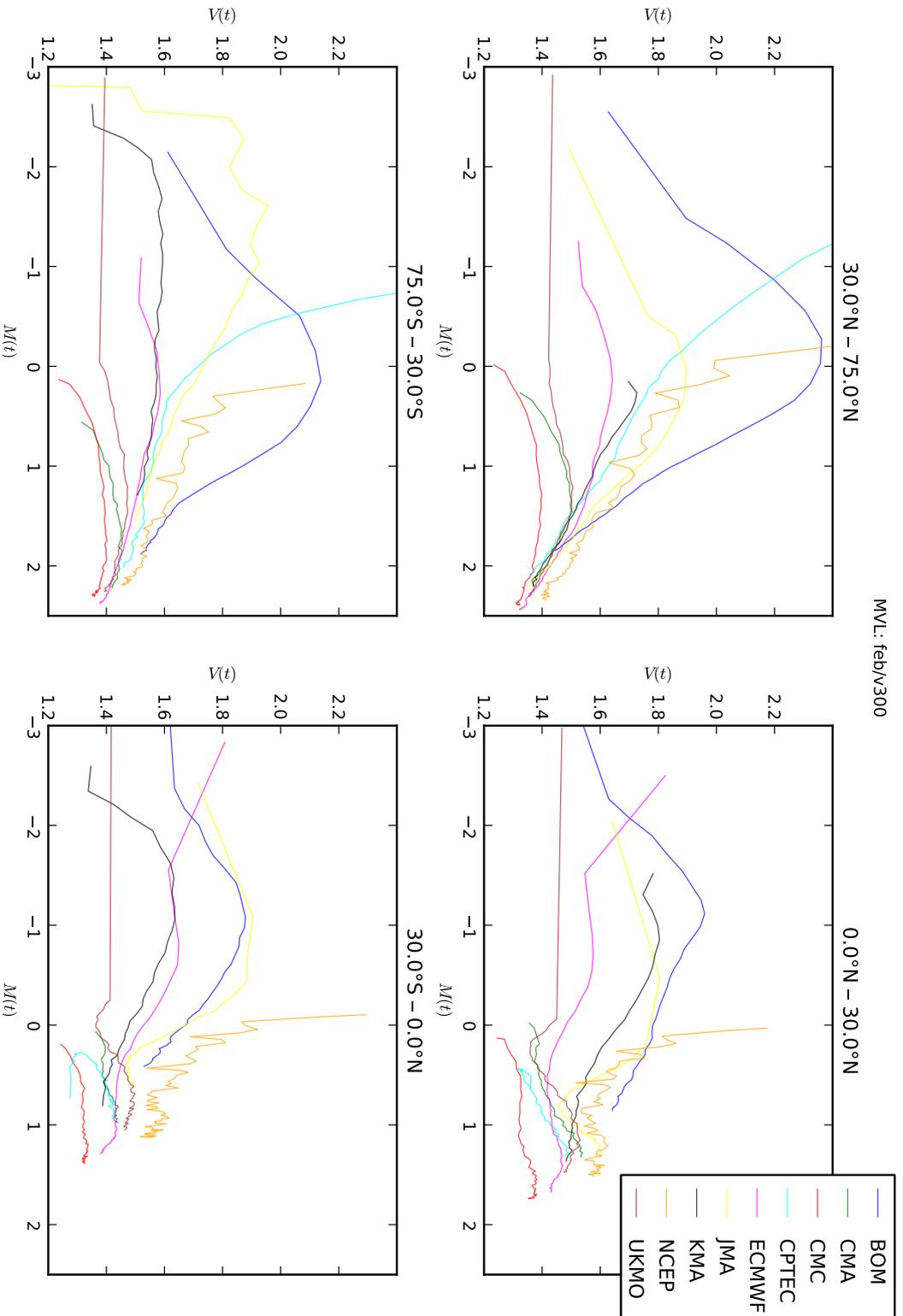
Figure 4.14: MVL diagrams for 300 hPa meridional wind in the TIGGE models, NH (top) and SH (bottom) extratropics (left) and tropics (right), for February 2009.

dynamics of the actual forecast error, since the main purpose of the ensemble is to give an estimate of the range of error in our forecast.

A word on definitions is in order at this point. By "perturbation", we mean the difference between an ensemble member and the corresponding control forecast at a particular time; when we talk about "error" we mean the difference between an ensemble member and the "true state" at that time. Of course we never know the true state, as we have a limited set of observational data; thus we use the subsequent analysis for that time as a "best estimate" of the true state.

With this in mind, we can draw a second MVL plot of the errors, taking differences from the subsequent analysis valid at each forecast time, rather than from the corresponding control forecast. For example, when calculating $M(24\,\mathrm{h})$ and $V(24\,\mathrm{h})$ for the forecast initialised at 00Z on 1 February, we consider differences from the analysis (control forecast at $t = 0$) that was subsequently generated at 00Z on 2 February as a best estimate of the error in each ensemble member (we include the control forecast here as a member of the ensemble). Figure 4.15 compares the evolution of these errors to the perturbations from the control member which we have been considering previously. Note that the first point ($t = 0$) is omitted as for the error curve it would involve comparing the control forecast to itself, resulting in a zero difference field with an undefined logarithm; for consistency it is omitted here from the perturbation curve as well.

For most models, the perturbations and errors appear to evolve with a similar shape. However, the actual errors never achieve as much spatial correlation as the perturbations, and the final "climatological" amplitude of the errors is greater than that of the perturbations. The latter suggests an incomplete representation of the sources of variability in these models, and is true even of those using stochastic physics to represent model error (e.g. NCEP and UKMO, less so ECMWF).

A notable exception is the CMC model, where the two plots are remarkably similar. It is unclear however whether this represents a genuine advantage in this model's perturbation dynamics, or is simply an artefact of using
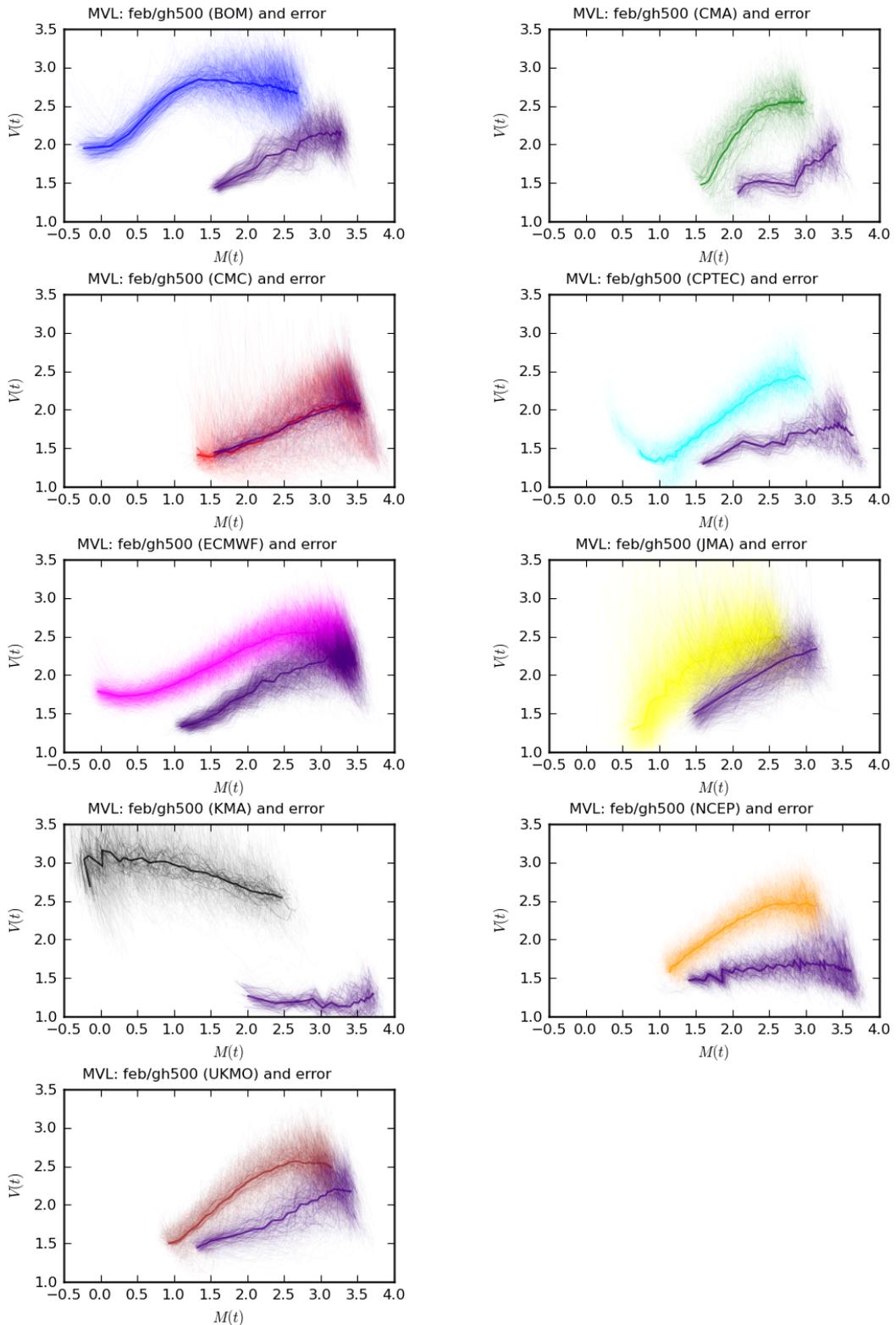
Figure 4.15: MVL comparison of perturbations from the control forecast (usual colour) against the error with respect to a subsequent analysis at the forecast time (indigo), for 500 hPa geopotential height in each of the TIGGE models, across the whole globe, for February 2009.

the same ensemble Kalman filter for perturbation generation and operational data assimilation.

Note that we have calculated the errors by comparing forecasts against analyses from the same centre. These analyses are produced by combining the model dynamics with a set of observations via a data assimilation scheme; however all three of these components differ between models. In particular, the contribution of the model dynamics may constrain the sequence of analyses to develop in a more similar way to the ensemble members than the "true" state (especially if a sparse set of observations are used). It might be enlightening to instead compare all the models against a single high-quality re-analysis data set to obtain a "best estimate" of the true errors.

Figure 4.16 shows how the relationship between perturbation and error evolution varies across the latitude bands.

4.5 Is the control forecast a good reference?

Another question to consider is the validity of using the control forecast as our reference. Although it is initialised using the best estimate of the true state, and starts off in a central position in the ensemble spread, it does not necessarily remain so; this is why the ensemble mean is generally used as the best estimate of the forecast state. However, the ensemble mean is a statistical construct rather than a genuine state of the dynamical system, and thus we cannot necessarily expect perturbations from the ensemble mean to have the same spatial structure, or for the results linking their spatiotemporal growth to the KPZ equation (Primo et al. 2007) to remain valid.

In order to eliminate any bias due to giving special status to the control forecast, while still considering real states of the dynamical system, we can instead calculate perturbations from the differences between all pairs of ensemble members (including the control as just another ensemble member). The disadvantage of this technique however is the computational complexity, which becomes quadratic in the number of ensemble members (as opposed to linear when considering only perturbations from the control forecast). Nevertheless, they can still be calculated in a feasible time, and figure 4.17 shows
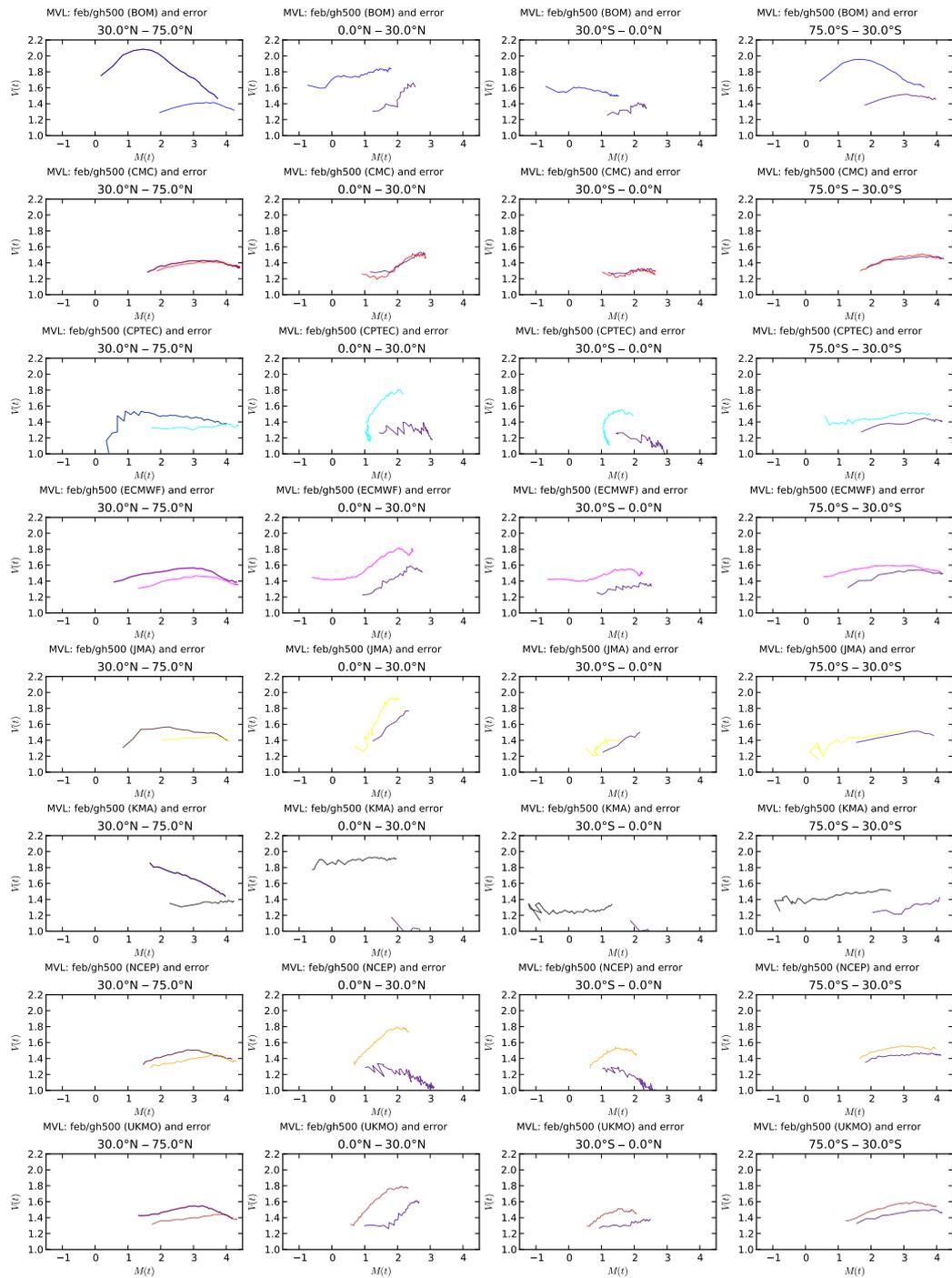
Figure 4.16: MVL comparison of perturbations from the control forecast (usual colour) against the error with respect to a subsequent analysis at the forecast time, for 500 hPa geopotential height in each of the TIGGE models, over (from left to right) northern extratropics, northern tropics, southern tropics and southern extratropics, for February 2009.
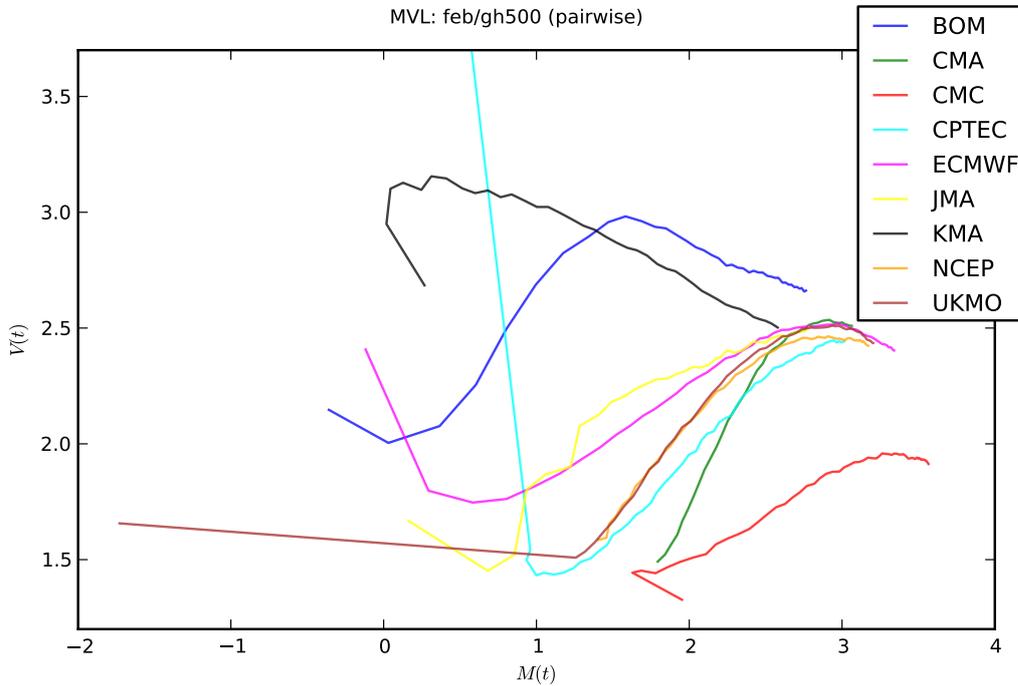
Figure 4.17: MVL diagrams for pairwise perturbations in $500\,\mathrm{hPa}$ geopotential height in the TIGGE models, over the whole globe, for February 2009. The centres are sorted by M_{peak}.

the results for $500\,\mathrm{hPa}$ geopotential height. Comparing this to figure 4.3, there is a slight shifting and smoothing of the curves, but the main features are unchanged. This suggests that for most purposes there is little to be gained from the extra computational effort.

4.6 Summary measures

In this section, we present an approach to calculating summary measures which characterise certain aspects of the MVL diagram, with the aim that models can be grouped or ranked according to particular characteristics. In particular, we look at the measures listed in table 4.1 and illustrated in figure 4.18.

The results for $500\,\mathrm{hPa}$ geopotential height in the TIGGE models are summarised in table 4.2 (these correspond to measurements of the curves

Symbol	Description	Equivalent to
t_{peak}	The lead time at which V achieves its maximum (excluding $t = 0$ for cases like CPTEC with an abnormally high initial value).	
t_{trgh}	The lead time at which V achieves its minimum prior to t_{peak}.	
M_{peak}	The value of M at which V achieves its maximum (excluding $t = 0$).	$M(t_{\text{peak}})$
M_{trgh}	The value of M at which V achieves its minimum prior to t_{peak}.	$M(t_{\text{trgh}})$
Δt_{drop}	The time taken to reach minimum V (in fact this is equal to t_{peak} since we always start at $t = 0$).	t_{peak}
Δt_{grow}	The time over which the main growth of spatial correlation occurs.	$t_{\text{peak}} - t_{\text{trgh}}$
ΔM_{drop}	The change in M over the initial drop in spatial correlation.	$M_{\text{trgh}} - M(0)$
ΔM_{grow}	The change in M over the main growth of spatial correlation.	$M_{\text{peak}} - M_{\text{trgh}}$
ΔV_{drop}	The change in V over the initial drop in spatial correlation.	$V(t_{\text{trgh}}) - V(0)$
ΔV_{grow}	The change in V over the main growth of spatial correlation.	$V(t_{\text{peak}}) - V(t_{\text{trgh}})$

Table 4.1: Summary measures of the MVL diagram.
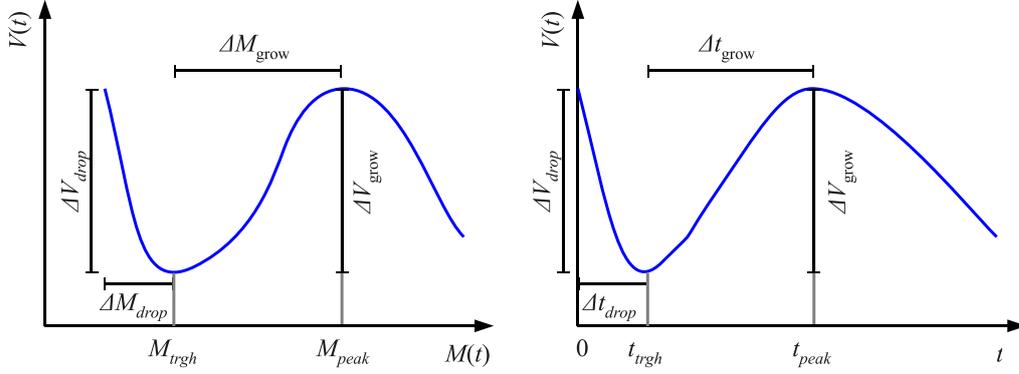
Figure 4.18: Schematic showing summary measures of the MVL diagram.

in figure 4.4). It is not clear if all of these measures are of practical value; however there are certain respects in which models can be differentiated:

- The models with an initial decrease in spatial correlation are picked out by large (negative) values of ΔV_{drop} – e.g. CPTEC in the extratropics and ECMWF in the tropics. Since this indicates an initial transient period where the model adjusts to the perturbations, such ensembles should probably not be considered reliable during that initial period, up to time t_{trgh}.

- As discussed earlier, most of the models have their peak variance at a similar value of M in a given latitude band; this is reflected in the M_{peak} values (which e.g. cluster around 3.0 in the extratropics), along with the exceptions noted previously (e.g. BOM and CMC which have unusually low and high values respectively in all latitude bands). This value gives an indication of the perturbation amplitude at which non-linear effects begin to dominate.

- t_{peak} gives an indication of the lead time beyond which predictability begins to be lost as the spatial structure in the perturbations is destroyed. However, high values do not necessarily imply that the forecast remains skilful at such lead times – merely that the ensemble retains some memory of its initial state.

```
75.0°S - 30.0°S
 centre  t_trgh  M_trgh  t_peak  M_peak dV_drop dt_drop dM_drop dV_grow dt_grow dM_grow
BOM          0  0.1925      30  1.7600  0.0000       0  0.0000  0.3206      30  1.5675
KMA          6 -0.7832     222  2.3520 -0.1234       6  0.1285  0.2680     216  3.1352
ECMWF        6  0.5554      84  2.5089 -0.0179       6  0.1305  0.1442      78  1.9534
NCEP         0  1.6046     132  2.9531  0.0000       0  0.0000  0.1800     132  1.3485
CPTEC       12  0.7314     204  3.0519 -6.0300      12  1.7482  0.1603     192  2.3204
JMA          0 -0.9205     216  3.0689  0.0000       0  0.0000  0.3906     216  3.9895
UKMO         6  1.2324     174  3.3139 -0.1166       6  2.9859  0.2398     168  2.0815
CMC          0  1.8764     156  3.5522  0.0000       0  0.0000  0.2322     156  1.6758
CMA          6  2.0645     162  3.5547 -0.0060       6  0.0749  0.2177     156  1.4901

30.0°S - 0.0°N
 centre  t_trgh  M_trgh  t_peak  M_peak dV_drop dt_drop dM_drop dV_grow dt_grow dM_grow
KMA          6 -0.8786      36 -0.9089 -0.3901       6  0.1235  0.2195      30 -0.0303
BOM          0 -1.3908       6 -0.6962  0.0000       0  0.0000  0.0441       6  0.6946
JMA         12  0.7463     126  1.1105 -0.1644      12  0.9131  0.2188     114  0.3643
NCEP         6  0.6591     174  1.4209 -0.0108       6  0.0399  0.2538     168  0.7618
UKMO        18  0.6417     216  1.6255 -0.1710      18  3.0919  0.2207     198  0.9838
CPTEC       30  1.2532     270  1.6371 -0.1217      30 -0.3876  0.4503     240  0.3839
CMA          0  1.0498     240  1.8630  0.0000       0  0.0000  0.4219     240  0.8132
ECMWF       30  0.2594     222  1.8769 -0.3103      30  1.6986  0.1566     192  1.6175
CMC         42  1.4886     156  2.1306 -0.0640      42  0.1178  0.1120     114  0.6420

0.0°N - 30.0°N
 centre  t_trgh  M_trgh  t_peak  M_peak dV_drop dt_drop dM_drop dV_grow dt_grow dM_grow
KMA          0 -0.0543     132  1.1421  0.0000       0  0.0000  0.2276     132  1.1964
BOM          0 -1.3463     228  1.7485  0.0000       0  0.0000  0.2695     228  3.0948
JMA         12  0.9692     198  1.9701 -0.1845      12  0.7397  0.6729     186  1.0009
NCEP         6  0.6931     252  1.9991 -0.0071       6  0.0565  0.4750     246  1.3060
ECMWF       18 -0.0161     222  2.0523 -0.4182      18  1.1990  0.3993     204  2.0685
CPTEC       30  1.1619     336  2.0726 -0.1242      30 -0.3130  0.6534     306  0.9107
CMA          6  1.1339     240  2.1137 -0.0041       6  0.0695  0.7112     234  0.9799
UKMO        18  0.6627     288  2.1291 -0.2152      18  3.1246  0.4901     270  1.4664
CMC         36  1.3796     198  2.4536 -0.0850      36  0.0393  0.3224     162  1.0741

30.0°N - 75.0°N
 centre  t_trgh  M_trgh  t_peak  M_peak dV_drop dt_drop dM_drop dV_grow dt_grow dM_grow
CPTEC        6  0.4147      42  0.9287 -2.3464       6  3.6156  0.5229      36  0.5140
BOM          0 -0.1409      30  1.4576  0.0000       0  0.0000  0.3426      30  1.5985
KMA          0  1.6539       6  1.6752  0.0000       0  0.0000  0.1483       6  0.0212
JMA          6  0.8544      36  2.2396 -0.0966       6  0.4410  0.2590      30  1.3853
NCEP         6  1.4753     120  2.8182 -0.0171       6  0.0924  0.1797     114  1.3430
ECMWF        6  0.5776     114  2.9690 -0.0110       6  0.1259  0.1760     108  2.3914
CMA          0  1.8736     102  3.0685  0.0000       0  0.0000  0.2472     102  1.1949
UKMO        12  1.4164     150  3.2588 -0.1255      12  3.0812  0.1260     138  1.8424
CMC          0  1.8156     156  3.7095  0.0000       0  0.0000  0.1624     156  1.8939
```

Table 4.2: Summary measures for 500 hPa geopotential height in the TIGGE models over each latitude band, for February 2009

There may well be scope for developing more robust measures, perhaps by filtering or pre-smoothing the curves; these relatively simple measures are easily distorted by "noisy" MVL curves that don't follow a smooth theoretical path (e.g. KMA outside the northern extratropics).

4.7 Summary

In this chapter we have analysed the spatiotemporal dynamics of the medium-range ensembles in the TIGGE archive, showing how each model has a distinct pattern. Concentrating mostly on $500\,\mathrm{hPa}$ geopotential height, we saw a large difference between the perturbation dynamics in the tropics and extratropics, with a further seasonal variation in the tropics. Looking closer to the surface, the perturbations in $2\,\mathrm{m}$ temperature produced less smooth MVL curves, with unexplained daily oscillations especially in NCEP. We also considered how the MVL diagram might be used to compare the dynamics of perturbations and forecast error. In chapter 5, we turn to the problem of constructing multi-model ensembles, and examine how the MVL diagram might be useful in this context.

Chapter 5

Multi-model ensembles

One of the reasons for looking at the dynamics of the various TIGGE models, apart from comparing their behaviour as independent ensembles, is to consider the potential for combining (all or a subset of) them to produce a multi-model ensemble with better performance (more skilful in either its ensemble mean or probabilistic forecasts).

5.1   Background

Experiments have been performed using multi-model ensembles for various purposes. For instance, Krishnamurti et al. (2006) discuss the application of one technique (the "superensemble"), to seasonal, medium-range and hurricane-track forecasts. The DEMETER project, already mentioned in section 2.2, has provided a framework for work on multi-model seasonal forecasting.

In the medium-range, the North American Ensemble Forecasting System (NAEFS; Zhu & Toth 2006) has been operational since May 2006, combining the CMC and NCEP models into a single 40-member ensemble. Candille (2009) shows that the combined ensemble offers improvements over each individual model and that this is true even for a random 20-member subset of the combined ensemble, i.e. that the improvement is not just due to the increased ensemble size.

Johnson & Swinbank (2009) present an experimental multi-model ensemble based on three of the TIGGE models: ECMWF, NCEP and UKMO, "chosen because they are all accessible in real time and because they have similar overall levels of skill". They also discuss the various calibration steps which may be desirable when combining ensembles with different statistical properties: bias correction, weighting and variance adjustment (their effects are shown in figure 5.1). The combined ensemble outperforms each of the individual ensembles, although the improvement is much smaller for mean-sea-level pressure and $500\,\mathrm{hPa}$ geopotential height (where the models exhibit similar forecast errors) than for $2\,\mathrm{m}$ temperature (where the forecast errors differ more between models).

5.2   Using MVL to choose models

Because the MVL diagram gives a concise picture of how the perturbations evolve in each model, it provides one approach to identifying models which will produce a dynamically diverse ensemble (Fernández et al. 2009). For example, referring back to figure 4.3, the NCEP and UKMO curves appear to have very similar spatiotemporal dynamics (after the first 6 hours, during which the Incremental Analysis Update takes effect in the UKMO model). This suggests that combining these ensembles will not add significant dynamical diversity, while adding e.g. CMC (as in NAEFS) or ECMWF (as in Johnson & Swinbank 2009) would increase the diversity.

It would therefore be illuminating to construct a combined NCEP-UKMO ensemble and compare its performance with the existing NAEFS (CMC-NCEP) ensemble, to see if this argument is supported by a smaller improvement over the individual ensembles for the former. Alternatively, one could examine the effect of removing either NCEP or UKMO from the ECMWF-NCEP-UKMO ensemble of Johnson & Swinbank (2009); the above argument would suggest that this would retain most of the dynamical diversity of the full three-model ensemble. The construction of such a new multi-model ensemble is beyond the scope of this work however.
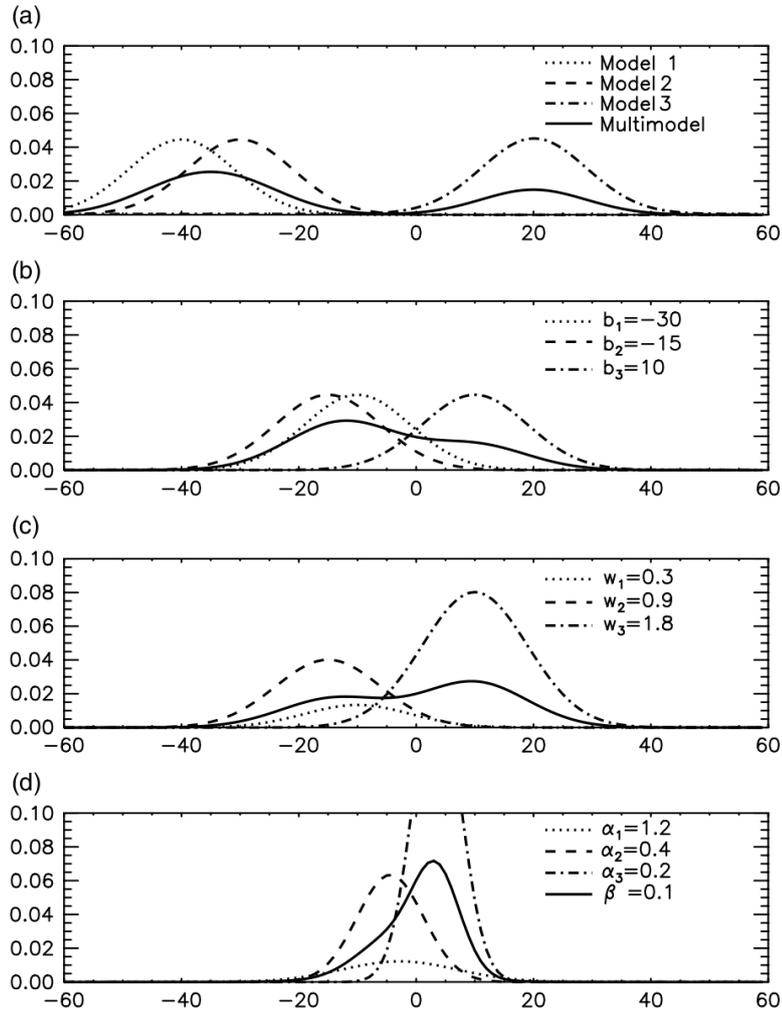
Figure 1. Illustration of the steps in calibrating and combining the multimodel ensemble. Each panel shows the pdfs for each ensemble, with the x-axis representing the value of the forecast variable, e.g. temperature. The panels show the pdfs (a) for raw data, (b) after bias correction (c) after model-dependent weights and (d) after variance adjustment. b_1, b_2 and b_3 are the biases of the three ensembles, w_1, w_2 and w_3 are the weights for the three ensembles, α_1, α_2 and α_3 are the multiplicative inflation factors for the within-model variance and β is the multiplicative inflation factor for the between-model variance.

Figure 5.1: Calibration of a multi-model ensemble (Johnson & Swinbank 2009, fig. 1).

5.3   MVL diagrams of multi-model ensembles

Having considered the use of MVL diagrams in choosing a multi-model ensemble, we now turn to the question of what the MVL diagram of that ensemble would look like. It is unclear, however, how we should measure the perturbations in such an ensemble: should we take each ensemble member relative to the control forecast from its own model, a (weighted) average of the different control forecasts, or the multi-model ensemble mean? The latter two options would introduce the problems discussed in section 4.5 of using a reference which is not a genuine state of the dynamical system; we thus take the first option (each ensemble's members relative to its own control). This approach has the additional property of being insensitive to differences in bias between the models (which will affect the control and perturbation equally and cancel out).

Taking this approach, and considering a multi-model ensemble without further calibration beyond bias correction and global weighting, each ensemble member produces exactly the same MVL curve as in its own ensemble, and we simply take the weighted mean of M and V across the members of all the constituent ensembles. For the case of equal weighting of all ensemble members, the result for several hypothetical combinations (along with their individual constituents) is shown in figure 5.2 (only models with a forecast range of at least two weeks have been included). However, where we combine models with very different curves (e.g. NCEP and CMC) the mean values are no longer representative as the ensemble becomes bimodal, as shown in figure 5.3.
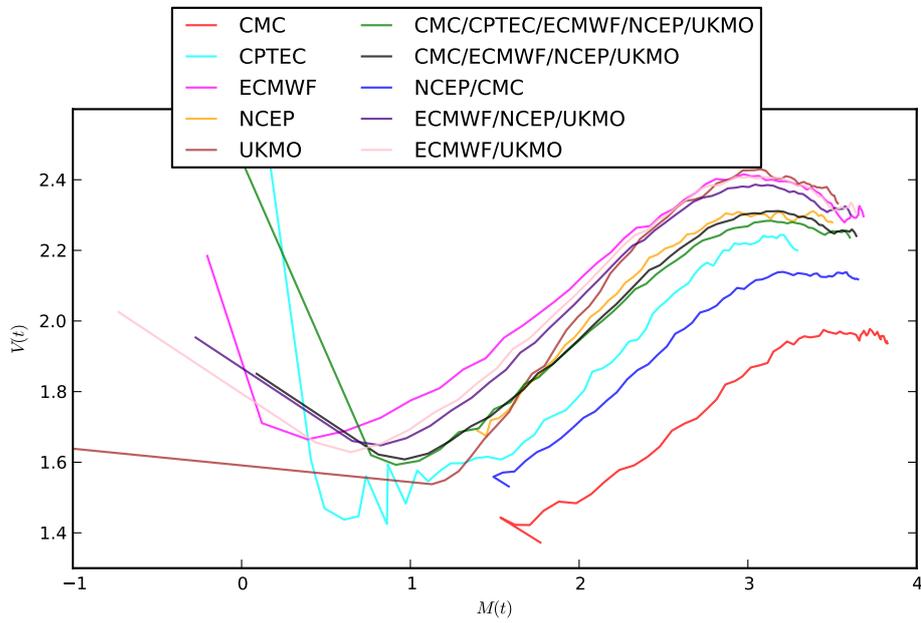
Figure 5.2: Mean-over-realisations MVL curves for hypothetical multi-model ensembles
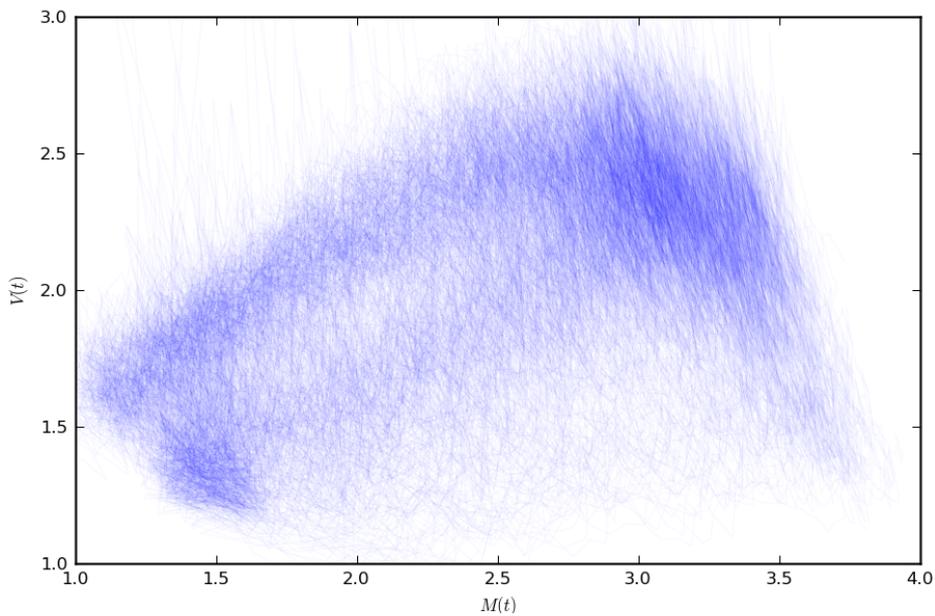
Figure 5.3: "Spaghetti" MVL plot for a combined NCEP/CMC ensemble

Chapter 6

Summary and conclusions

We have introduced the concept of spatiotemporal growth of perturbations, and shown how this can be related to the theory of interface growth if we take the logarithm of the perturbations. Using this theoretical framework, the MVL diagram provides a concise way to visualise the spatiotemporal dynamics of a complex model.

Following previous work applying the technique to the DEMETER seasonal weather forecasting models, we have presented an analysis of the TIGGE medium-range models. In both cases, each model's MVL curve has a characteristic shape for a given variable, but generally fitting into one of the expected patterns in figure 2.4. Many features of these curves can be related to known characteristics of the model (e.g. the poor growth of spatial correlation in the unperturbed regions of some models), although other features remain unexplained (e.g. the large oscillations in spatial correlation for the NCEP 2 m temperature field). Looking at separate MVL diagrams for the tropics and extratropics, we saw that the different dynamics are clearly reflected in very different curves for most of the models. There was also much more seasonal variation in the diagrams for the tropics than those for the extratropics.

There is a clear distinction between the models using singular-vector and EOF initial perturbations (which show a large initial drop in spatial correlation before the growth phase at some or all latitudes) and those using

bred-vector and ET/EnKF-based perturbations (which do not). This suggests that the latter produce perturbed states closer to the model attractor; while this does not necessarily imply greater skill in the medium range, it is probably important if the ensemble is also to be used for short-range probabilistic forecasts.

We also considered the relationship of the perturbation dynamics to the evolution of forecast error, showing a wide variation in how closely the MVL curves for the two agree. There are two major caveats, however, to using this as an indicator of how well the ensemble captures the true forecast uncertainty: firstly, while the perturbation (ensemble member minus control) is the difference between two states of the same dynamical system, the error (forecast minus analysis) is not, and thus the applicability of the MVL theory is uncertain; and secondly, the "almost too good to be true" fit for the EnKF-based CMC suggests that this analysis may be over-sensitive to artefacts of the interaction between the data assimilation and perturbation generation schemes.

We briefly presented a number of simple measurements which can be taken from the MVL diagram to characterise certain aspects of the perturbation growth; it is unclear whether this approach offers any insight beyond that apparent in the diagram itself, although with further work to make the measures more robust it could lead to a more systematic approach to categorising the model dynamics.

In the last chapter, we considered how these results might be applied to the construction of multi-model ensembles from the TIGGE models by choosing models with a range of different MVL curves, given the previous studies which highlight the importance of dynamical diversity between the constituent ensembles. We proposed an experiment to test the usefulness of such an approach; this would require the construction of at least one new multi-model ensemble, but could be an enlightening avenue for further work.

Bibliography

Barabási, A.-L. & Stanley, H. E. (1995), *Fractal Concepts in Surface Growth*, Cambridge University Press. 366 pp.

Barkmeijer, J., Van Gijzen, M. & Bouttier, F. (1998), 'Singular vectors and estimates of the analysis-error covariance metric', *Quarterly Journal of the Royal Meteorological Society* **124**(549, Part A), 1695–1713.

Bishop, C. H., Etherton, B. J. & Majumdar, S. J. (2001), 'Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects', *Monthly Weather Review* **129**(3), 420–436.

Bloom, S. C., Takacs, L. L., DaSilva, A. M. & Ledvina, D. (1996), 'Data assimilation using incremental analysis updates', *Monthly Weather Review* **124**(6), 1256–1271.

Boffetta, G., Cencini, M., Falcioni, M. & Vulpiani, A. (2002), 'Predictability: a way to characterize complexity', *Physics Reports* **356**(6), 367–474.

Bourke, W., Buizza, R. & Naughton, M. (2004), 'Performance of the ECMWF and the BoM ensemble prediction systems in the southern hemisphere', *Monthly Weather Review* **132**(10), 2338–2357.

Bowler, N. E. (2006), 'Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model', *Tellus A* **58**(5), 538–548.

Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B. & Beare, S. E. (2008), 'The MOGREPS short-range ensemble prediction system', *Quarterly Journal of the Royal Meteorological Society* **134**(632), 703–722.

Buizza, R., Houtekamer, P. L., Toth, Z., Pellerin, G., Wei, M. Z. & Zhu, Y. J. (2005), 'A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems', *Monthly Weather Review* **133**(5), 1076–1097.

Buizza, R. & Palmer, T. N. (1995), 'The singular-vector structure of the atmospheric global circulation', *Journal of the Atmospheric Sciences* **52**(9), 1434–1456.

Buizza, R., Richardson, D. S. & Palmer, T. N. (2001), 'The new 80km high-resolution ECMWF EPS', *ECMWF Newsletter* (90), 2–9.

Candille, G. (2009), 'The multiensemble approach: The NAEFS example', *Monthly Weather Review* **137**(5), 1655–1665.

Champion, A. (2008), Properties of ensemble forecasts from different weather centres, M.Sc. dissertation, Department of Meteorology, University of Reading.

Descamps, L. & Talagrand, O. (2007), 'On some aspects of the definition of initial conditions for ensemble prediction', *Monthly Weather Review* **135**(9), 3260–3272.

Ehrendorfer, M. (2006), The Liouville equation and atmospheric predictability, *in* Palmer & Hagedorn (2006), chapter 4, pp. 59–98.

Ehrendorfer, M., Errico, R. M. & Raeder, K. D. (1999), 'Singular-vector perturbation growth in a primitive equation model with moist physics', *Journal of the Atmospheric Sciences* **56**(11), 1627–1648.

Evensen, G. (1994), 'Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics', *Journal of Geophysical Research – Oceans* **99**(C5), 10143–10162.

Fernández, J., Primo, C., Cofiño, A., Gutiérrez, J. & Rodríguez, M. (2009), 'MVL spatiotemporal analysis for model intercomparison in EPS: application to the DEMETER multi-model ensemble', *Climate Dynamics* **33**(2), 233–243.

Gutiérrez, J. M., Primo, C., Rodríguez, M. A. & Fernández, J. (2008), 'Spatiotemporal characterization of ensemble prediction systems – the mean-variance of logarithms (MVL) diagram', *Nonlinear Processes in Geophysics* **15**(1), 109–114.

He, Y., Wetterhall, F., Cloke, H. L., Pappenberger, F., Wilson, M., Freer, J. & McGregor, G. (2009), 'Tracking the uncertainty in flood alerts driven by grand ensemble weather predictions', *Meteorological Applications* **16**(1, Sp. Iss. SI), 91–101.

Hollingsworth, A. (1980), An experiment in Monte Carlo forecasting, *in* 'Workshop on Stochastic Dynamic Forecasting', ECMWF, Shinfield Park, Reading, UK, pp. 65–85.

Houtekamer, P. L. H. & Mitchell, H. L. (2005), 'Ensemble Kalman filtering', *Quarterly Journal of the Royal Meteorological Society* **131**(613), 3269–3289.

Houtekamer, P. L., Lefaivre, L., Derome, J., Ritchie, H. & Mitchell, H. L. (1996), 'A system simulation approach to ensemble prediction', *Monthly Weather Review* **124**(6), 1225–1242.

Houtekamer, P. L. & Mitchell, H. L. (1998), 'Data assimilation using an ensemble Kalman filter technique', *Monthly Weather Review* **126**(3), 796–811.

Johnson, C. & Swinbank, R. (2009), 'Medium-range multimodel ensemble combination and calibration', *Quarterly Journal of the Royal Meteorological Society* **135**(640, Part A), 777–794.

Kardar, M., Parisi, G. & Zhang, Y. C. (1986), 'Dynamic scaling of growing interfaces', *Physical Review Letters* **56**(9), 889–892.

Krishnamurti, T., Kumar, T. S. V. V., Yun, W.-T., Chakraborty, A. & Stefanova, L. (2006), Weather and seasonal climate forecasts using the superensemble approach, *in* Palmer & Hagedorn (2006), chapter 20, pp. 532–560.

Lopez, J. M., Primo, C., Rodríguez, M. A. & Szendro, I. G. (2004), 'Scaling properties of growing noninfinitesimal perturbations in space-time chaos', *Physical Review E* **70**(5, Part 2).

Lorenz, E. N. (1963), 'Deterministic nonperiodic flow', *Journal of the Atmospheric Sciences* **20**, 130–141.

Lorenz, E. N. (1996), Predictability: a problem partly solved, *in* 'Seminar on Predictability', Vol. 1, ECMWF, Shinfield Park, Reading, UK, pp. 1–18.

Magnusson, L., Leutbecher, M. & Kallen, E. (2008), 'Comparison between singular vectors and breeding vectors as initial perturbations for the ECMWF ensemble prediction system', *Monthly Weather Review* **136**(11), 4092–4104.

Mendonça, A. M. & Bonatti, J. P. (2009), 'Experiments with EOF-based perturbation methods and their impact on the CPTEC/INPE ensemble prediction system', *Monthly Weather Review* **137**(4), 1438–1459.

Molteni, F., Buizza, R., Palmer, T. N. & Petroliagis, T. (1996), 'The ECMWF ensemble prediction system: Methodology and validation', *Quarterly Journal of the Royal Meteorological Society* **122**(529, Part A), 73–119.

Palmer, T. & Hagedorn, R., eds (2006), *Predictability of weather and climate*, Cambridge University Press.

Palmer, T. N. (2006), Predictability of weather and climate: from theory to practice, *in* Palmer & Hagedorn (2006), chapter 1, pp. 1–29.

Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M., Delecluse, P., Deque, M., Diez, E., Doblas-Reyes, F. J., Feddersen, H., Graham, R., Gualdi, S., Gueremy, J. F., Hagedorn, R., Hoshen, M., Keenlyside, N., Latif, M., Lazar, A., Maisonnave, E., Marletto, V., Morse, A. P., Orfila, B., Rogel, P., Terres, J. M. & Thomson, M. C. (2004), 'Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER)', *Bulletin of the American Meteorological Society* **85**(6), 853.

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H. L., Buizza, R. & de Roo, A. (2008), 'New dimensions in early flood warning across the globe using grand-ensemble weather predictions', *Geophysical Research Letters* **35**(10).

Park, Y.-Y., Buizza, R. & Leutbecher, M. (2008), 'TIGGE: Preliminary results on comparing and combining ensembles', *Quarterly Journal of the Royal Meteorological Society* **134**(637, Part B), 2029–2050.

Pazó, D., Rodríguez, M. & López, J. M. (2009), Spatio-temporal evolution of perturbations in ensembles initialized by bred, Lyapunov, and singular vectors. Unpublished.

Pikovsky, A. & Politi, A. (1998), 'Dynamic localization of Lyapunov vectors in spacetime chaos', *Nonlinearity* **11**(4), 1049–1062.

Pikovsky, A. S. & Kurths, J. (1994), 'Roughening interfaces in the dynamics of perturbations of spatiotemporal chaos', *Physical Review E* **49**(1), 898–901.

Primo, C. & Magnusson, L. (2009), Spatiotemporal comparison of initial perturbation methods. Unpublished.

Primo, C., Szendro, I. G., Rodríguez, M. A. & Gutiérrez, J. M. (2007), 'Error growth patterns in systems with spatial chaos: From coupled map lattices to global weather models', *Physical Review Letters* **98**(10).

Toth, Z. & Kalnay, E. (1993), 'Ensemble forecasting at NMC – the generation of perturbations', *Bulletin of the American Meteorological Society* **74**(12), 2317–2330.

Toth, Z. & Kalnay, E. (1997), 'Ensemble forecasting at NCEP and the breeding method', *Monthly Weather Review* **125**(12), 3297–3319.

Vallis, G. K. (2006), *Atmospheric and Oceanic Fluid Dynamics*, Cambridge University Press. 745 pp.

Wang, X. G., Bishop, C. H. & Julier, S. J. (2004), 'Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble?', *Monthly Weather Review* **132**(7), 1590–1605.

Wei, M., Toth, Z., Wobus, R. & Zhu, Y. (2008), 'Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system', *Tellus A* **60**(1), 62–79.

Wei, M. Z., Toth, Z., Wobus, R., Zhu, Y. J., Bishop, C. H. & Wang, X. G. (2006), 'Ensemble transform Kalman filter-based ensemble perturbations in an operational global prediction system at NCEP', *Tellus A* **58**(1), 28–44.

Zhang, Z. & Krishnamurti, T. N. (1999), 'A perturbation method for hurricane ensemble predictions', *Monthly Weather Review* **127**(4), 447–469.

Zhu, Y. & Toth, Z. (2006), 'Global ensemble and NAEFS', `http://www.nws.noaa.gov/oh/hrl/hsmb/docs/hep/events_announce/Zhu_GEFS_NAEFS_Final.pdf`. Presentation for RFC Short-Term Ensemble Workshop.