Towards high resolution data assimilation and ensemble forecasting

Roel Stappers

ISBN 978-90-393-9000-9
Copyright ©2013 Roel Stappers, The Netherlands

Towards high resolution data assimilation and ensemble forecasting

Naar hoge resolutie data assimilatie
en ensemble verwachtingen

(met een samenvatting in het Nederlands)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

vrijdag 14 juni 2013 des middags te 4.15 uur

door

Roeland Johannes Jozef Stappers

geboren op 31 mei 1976 te Horst

Promotor:	**Prof.dr.ir. H.A. Dijkstra**
Co-promotor:	**Dr. J. Barkmeijer**

CONTENTS

1 Introduction **1**
1.1 Motivation 1
1.2 Tangent linear and adjoint models in NWP 2
1.3 4D-VAR and the Kalman filter 3
1.4 This thesis 5

2 Properties of CAPE singular vectors **9**
2.1 Introduction 10
2.2 Theory 11
2.3 Case study 15
2.4 Linearity test twin experiments 21
2.5 Conclusions and discussion 22

3 Optimal linearization trajectories for tangent linear models **29**
3.1 Introduction 30
3.2 Bilinear differential equations 31
3.3 Optimal linearization trajectories 34
3.4 Iterative relinearization 35
3.5 Estimation using the inverse TL-model 37
3.6 Applications 37
3.7 Discussion 46
3.8 Conclusions 50
3.A Quasi-geostrophic model 51
3.B Lorenz 96 model 51
3.C Bilinearity preserving discretizations 56
3.D Multilinear systems 57

4 Gaussian quadrature 4D-VAR **61**
4.1 Introduction 62
4.2 Notation 63
4.3 Fundamental theorem for line integrals over vector fields 65
4.4 The Gaussian quadrature tangent linear model 66
4.5 Contractions and fixed points 72
4.6 Incremental 4D-VAR (Gauss-Newton estimation) 72
4.7 Incremental 4D-VAR as a fixed point iteration 73
4.8 Merging of inner and outer loops 76
4.9 Adjoint based observation impact 79
4.10 Conclusions 81
4.A Modified Lorenz 96 model 82

5 Conclusions and further research **85**
5.1 Conclusions 85
5.2 Further research 87

Bibliography **89**

Samenvatting **95**

Dankwoord **101**

Curriculum Vitae **103**

Acronyms **105**

CHAPTER 1

INTRODUCTION

1.1 Motivation

Due to the chaotic nature of the atmosphere, small errors in the initial conditions of numerical weather prediction (NWP) models may amplify rapidly (the so called butterfly effect (Gleick, 1987)). To produce skillful weather forecasts it is therefore essential to have accurate estimates of the current state of the atmosphere. For this reason the assimilation of observations into the model (data assimilation) is one of the most important aspects of numerical weather prediction. The aim of data assimilation is to combine the model and observations to give an optimal estimate (the analysis) of the current state of the atmosphere.

As a result of improvements in data assimilation algorithms and NWP models and an increase in the number of observations the errors in the analyses have decreased during the past decades. However, the analyses will always be uncertain and because of the amplification of small initial condition errors the forecasts will be uncertain. To account for this uncertainty several NWP centres use ensemble prediction methods. These ensemble prediction methods aim to characterize the forecast probability density function (pdf) by running the model several times with slightly different initial conditions (and different physics). One method to construct these initial conditions makes use of singular vectors and has been succesful in ensemble prediction systems for the medium-range.

Both variational data assimilation and singular vector calculations make use of tangent linear and adjoint models. In sections 1.2 and 1.3 we discuss some of the problems related to the use of these linear models in high resolution NWP models which will lead to the research questions in section 1.4.

1.2 Tangent linear and adjoint models in NWP

Tangent linear models and, in particular, adjoint models have been very useful in several applications in numerical weather prediction (see Errico, 1997, 2003; Errico and Ehrendorfer, 2007; Barkmeijer, 2010, for an overview). For example, at the European Centre for Medium-Range Weather Forecasts (ECMWF) these linear models play a crucial role in the computation of initial condition perturbations used in the ensemble prediction system (Leutbecher and Palmer, 2008) and in their 4D-VAR data assimilation system (Courtier et al., 1994). One of the major limitations of the application of linear models is that the results are useful only when the linear approximation is valid (Errico, 1997). By this we mean that the difference between two integrations of the nonlinear model can be described by the associated linearized version of the nonlinear model. In adiabatic NWP models at synoptic scales the lead time during which the linear model can capture the nonlinear behaviour is usually assumed to be on the order of 2 days.

Due the increase in computational power of supercomputers the grid resolution of high resolution numerical weather prediction models is now reaching the 1 km scale. As a result, mesoscale processes related to high impact weather (such as deep convection) can now explicitly be resolved by the models. Compared to synoptic scale systems (~1000 km) nonlinearity plays a much more prominent role at the mesoscale (Hohenegger and Schär, 2007). Furthermore the relation between observations of these mesoscale processes and the model state is strongly nonlinear. The most successful synoptical scale data assimilation method is known as incremental 4D-VAR and makes use of linearised versions of both the model dynamics and the observation operator that maps the model state to the observed values. Due to the use of these linear approximations there are serious concerns that the incremental 4D-VAR approach will be defeated by these nonlinearities when the resolution of the model is increased to the km scale (Andersson et al., 2005).

Due to the success of ensemble prediction systems (EPSs) for the medium-range several numerical weather prediction consortia have developed, or are in the process of developing, EPSs for the short-range (SREPS) (see e.g. Iversen et al. (2007); Hágel and Horányi (2007); Bowler et al. (2008); Du et al. (2009)). The main goal of these short-range ensemble prediction systems is to aid forecasters in situations of severe weather. In medium-range EPSs several methods are being used to create initial condition perturbations. At the European Centre for Medium-Range Weather Forecasts (ECMWF) part of the initial condition perturbations are created by taking linear combination of singular vectors. These singular vectors are perturbations of the initial condition that show maximum linear growth measured in a certain norm over a fixed time period. The structure of the singular vectors is dependent on the norms used, the time interval considered and on the background flow around which the linear operators are evaluated. In medium-range forecasting the common choice is to maximize the two day perturbation in the total energy (TE) norm. As noted by Badger and Hoskins (2001) and

Hoskins et al. (2000) these TE-SVs with short optimization time yield structures that achieve growth mainly through upward propagating Rossby waves. These initial structures tend to be in the mid troposphere while at final time the energy is at the jet stream level. As such, TE-SVs with short optimization times are not optimally suited for SREPS where we are mainly interested in creating sufficient ensemble spread at the surface in the short-range. Furthermore, because the tangent linear and adjoint model play a key role in the singular vector calculation, there are concerns that singular vector based techniques are not suitable for high resolution numerical weather prediction models (Hohenegger and Schär, 2007).

1.3 4D-VAR and the Kalman filter

Data assimilation is an estimation problem that aims at producing the best estimate of the atmosphere given the model and observations. In the field of estimation theory it is well known that for imperfect linear models with Gaussian errors the Kalman filter (KF) provides the best linear unbiased estimate. Due to the high dimensionality of the state vector ($\sim 10^9$) in NWP models the propagation of the covariance matrices required in the KF equations is not possible. In NWP the covariance matrices are either approximated by low rank approximations (e.g. the ensemble Kalman filter) or variational methods (3D-VAR, 4D-VAR) are used. For linear models the variational methods produce the same estimate as the unapproximated full rank KF at the end of the assimilation window but without requiring the explicit propagation of the covariance matrix. For state vector dimensions typical for current operational NWP systems, 4D-VAR is far more efficient than the full rank Kalman filter (Fisher et al., 2005).

In this section we highlight some characteristics of variational data assimilation problem in NWP by taken the ECMWF model as an example. In the data assimilation system at ECMWF a new optimal initial condition is estimated every 12 hours using approximately 10 million observations distributed (irregularly in space and time) over the globe. For a given 12 hour time window the optimal initial condition (the analysis) is defined as the minimum of a cost function. This cost function measures the distance between the observations in this time window and the value of these observations predicted by the model (see figure 1.1). The 4D-VAR problem is challenging for a number of reasons (Fisher et al., 2009)

1. There are 7 billion unknown parameters (temperature, humidity, pressure, wind speed and direction at each grid point in the model with 30 minute intervals in the 12 hour window). The number of observations is almost 3 orders of magnitude less than the number of unknowns, making this an ill-posed problem. Additional information in the form of a short-range forecast (the background) is included in the cost function to make the problem well-posed.

2. There is a strict time budget available for solving the problem. Forecasts

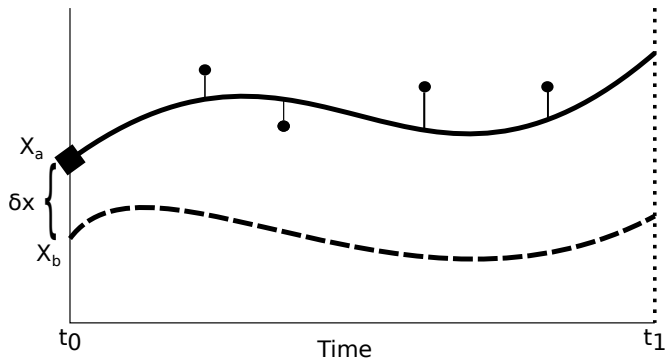

Figure 1.1: For a given background X_b incremental strong constraint 4D-VAR finds a (small) increment δx such that the forecast (solid line) from the analysis $X_a = X_b + \delta x$ (square) is close to the observations (dots). The tangent linear and adjoint models are linearized around the nonlinear forecast (dashed line) from the background. In the outer loops of incremental 4D-VAR the linearisation trajectory for the tangent linear and adjoint model is updated.

must be produced in a 12-hour cycle, less than two hours of which are available for solving the data assimilation problem.

3. The cost function is non-convex and the constraints are nonlinear.

4. Evaluations of the constraints via forward integration of the model is expensive and can only be performed a few times in each forecast cycle.

To efficiently solve the data assimilation problem Courtier et al. (1994) proposed an incremental formulation of 4D-VAR that makes use of low resolution linearised versions of the high resolution nonlinear atmospheric model. The minimization of the incremental 4D-VAR cost function is an iterative method structurally similar to Gauss-Newton methods (Lorenc, 1988; Lawless et al., 2005a,b; Fisher et al., 2009). It is known that if too many (outer loop) iterations are performed in 4D-VAR the estimates can diverge. Trémolet (2007b) attributed this to differences in phase speed of gravity waves between the low resolution tangent linear model and the high resolution nonlinear model. Eliminating these phase speed differences requires much higher resolutions and smaller time steps in the linear integration than can be afforded. Currently the most promising method to account for such deficiencies are the so called weak constraint 4D-VAR methods (Lindskog et al., 2009; Trémolet, 2006). Here the model is no longer assumed to be an exact representation of the true atmosphere. Besides estimating the initial condition, possible biases in the model are also estimated. The main drawback of this approach is the increase in the, already large, number of unknown parameters. As models reach higher resolutions and highly nonlinear mesoscale processes

can be resolved by the model, the linearity assumption is becoming increasingly problematic (Tanguay et al., 1995). Furthermore, the relation between measurements of small scale atmospheric features and the state of the model is becoming highly nonlinear which makes the linear assumption even more problematic.

1.4 This thesis

As indicated in sections 1.2 and 1.3 there are problems with linear approximations in high resolution models both for data assimilation and for ensemble prediction. This leads us to consider the following research questions:

1. Can the singular vector technique, which has been successful for the medium-range at the synoptic scale, be used in short-range high resolution ensemble prediction systems? In particular can the singular vectors inform about the predictability of high impact weather?

2. Can the tangent linear model and adjoint model be improved, e.g. can the tangent linear model at the synoptic scale be used beyond 2 days?

3. Are there alternative formulations for the incremental 4D-VAR data assimilation method that are more suitable for high resolution models?

Chapter 2 addresses research question 1 by examining the feasibility of using the singular vector technique to create initial condition perturbations for short-range ensemble prediction systems (SREPS) focussing on predictability of severe local storms and in particular deep convection. For this a new final time semi-norm based on the convective available potential energy (CAPE) is introduced. We compare singular vectors using the CAPE-norm with SVs using the more common total energy (TE) norm for a two week summer period in 2007, which includes a case of mesoscale extreme rainfall in the south west of Finland. The CAPE singular vectors perturb the CAPE field by increasing the specific humidity and temperature of air parcels and increase the lapse rate above these parcels in the lower troposphere, consistent with physical considerations. The CAPE-SVs are situated in the lower troposphere. This in contrast to TE-SVs with short optimization times which predominantly remain in the high troposphere. By examining the time evolution of the CAPE singular values we observe that the convective event in the south west of Finland is clearly associated with high CAPE singular values. With respect to the first research question we therefore conclude that it is possible to compute flow dependent initial condition perturbations that have maximum impact on the CAPE. Because the CAPE-SVs are situated in the lower troposphere ensemble based prediction systems based on the CAPE-SVs will have largest spread in the lower troposphere. We conclude that EPSs based on CAPE-SVs seem more suitable for the short-range than EPSs based on TE-SVs.

Chapter 3 addresses research question 2 by examining differential equations where nonlinearity is a result of the advection part of the total derivative or the use

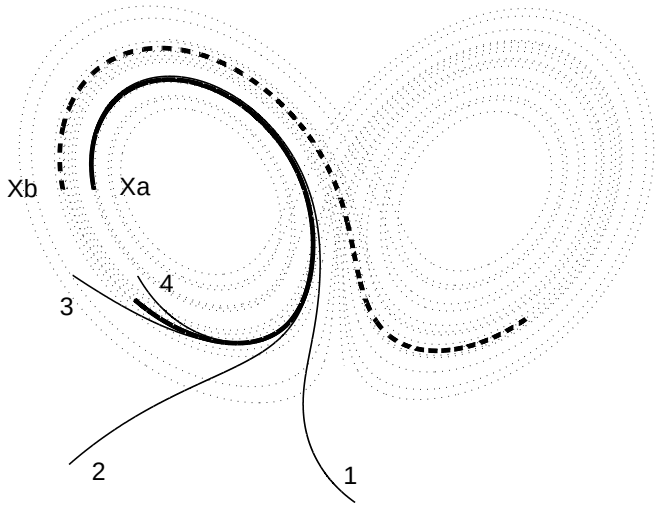

Figure 1.2: Illustration of the relinearisation method of chapter 3 using the Lorenz 63 model. The first linearisation trajectory for the tangent linear model is the dashed thick line with initial condition X_b. The tangent linear model prediction for the time evolution of the increment $X_a - X_b$ is given by the thin solid line (1). The method uses this tangent linear prediction to update the linearisation trajectory for the tangent linear model and produce a new estimate for the nonlinear time evolution (2), etc. After 4 iterations we obtain a solution that is close to the true nonlinear time evolution (solid thick line).

of quadratic algebraic constraints between state variables (such as the ideal gas law). We show that these types of nonlinearity can be accounted for in the tangent linear model by a suitable choice of the linearisation trajectory. In this chapter we show that the tangent linear model can be used to reproduce the exact nonlinear error growth of perturbations for more than 200 days in a quasi geostrophic model and more than (the equivalent of) 150 days in the Lorenz 96 model. We introduce an iterative method, purely based on tangent linear integrations, that converges to this optimal linearisation trajectory (see figure 1.2).

With respect to the second research question we therefore conclude that the tangent linear model can be used for much longer than 2 days. This eliminates one of the major limitations for using linear models.

Chapter 4 addresses research question 3 by generalizing the result of chapter 3 to nonlinear models with higher order nonlinearities and by introducing a new incremental 4D-VAR algorithm based on the relinearisation method of chapter 3. It is shown that this introduces the possibility to update the linearisation trajectories at negligible computational cost and that the distinction between

inner and outer loops is no longer necessary. Furthermore it is shown that it is straightforward to account for multiple outer loops in adjoint based observation impact calculations when the proposed 4D-VAR method is used. The method is illustrated within the context of a quasi geostrophic model and the Lorenz 96 model. With respect to the third research question we therefore conclude that the proposed data assimilation method seems more suitable for high resolution models but more research is needed to determine for which constraint nonlinear least squares problems the proposed method is preferable over the Gauss-Newton method.

In chapter 5 more detailed conclusions and possible directions for further research are given.

CHAPTER 2

PROPERTIES OF CAPE SINGULAR VECTORS

We study the feasibility of using the singular vector technique to create initial condition perturbations for short range ensemble prediction systems (SREPS) focussing on predictability of severe local storms and in particular deep convection. For this a new final time semi-norm based on the convective available potential energy (CAPE) is introduced. We compare singular vectors using the CAPE-norm with SVs using the more common total energy (TE) norm for a two week summer period in 2007, which includes a case of mesoscale extreme rainfall in the south west of Finland.

The CAPE singular vectors perturb the CAPE field by increasing the specific humidity and temperature of the parcel and increase the lapse rate above the parcel in the lower troposphere consistent with physical considerations. The CAPE-SVs are situated in the lower troposphere. This in contrast to TE-SVs with short optimization times which predominantly remain in the high troposphere. By examining the time evolution of the CAPE singular values we observe that the convective event in the south west of Finland is clearly associated with high CAPE singular values.

This text is published as: Properties of singular vectors using convective available potential energy as final time norm, Tellus A, (Stappers and Barkmeijer, 2011)

2.1 Introduction

Due to the success of ensemble prediction systems (EPSs) for the medium range several numerical weather prediction consortia have developed, or are in the process of developing, EPSs for the short-range (SREPS) (see e.g. (Iversen et al., 2007; Hágel and Horányi, 2007; Du et al., 2009)). The main goal of these short-range ensemble prediction systems is to aid forecasters especially in situations of severe weather. Almost all severe local storms are associated with deep convection and require three necessary ingredients (Johns and Doswell III, 1992; Doswell III, 1987; Schultz and Schumacher, 1999),

1. A moist layer of sufficient depth
2. A steep enough lapse rate
3. Sufficient lifting of a parcel from the moist layer to allow it to reach its level of free convection (LFC).

The requirement for moisture (i) and steep enough lapse rate (ii) can be understood in terms of convective available potential energy (CAPE). In general the large scale flow creates the favorable thermodynamic environment while the mesoscale provides the lift needed to initiate convection (Doswell III, 1987).

In medium range EPS several methods are being used to create initial condition perturbations. At the European Centre for Medium-Range Weather Forecasts (ECMWF) the initial condition perturbations are created by taking linear combination of singular vectors. Singular vectors are norm dependent structures and can therefore be designed to focus on particular forecast aspects. Given the requirement that the SREPS should inform about predictability of severe local storms in general and deep convection in particular, we investigate the properties of singular vectors that have maximum impact on the 12 hour CAPE forecasts. We will refer to these singular vectors as CAPE-SVs.

Based on the fact that forecasters are able to improve numerical weather predictions of mesoscale convective systems, Xu et al. (2001) propose to let forecasters identify the area of forecast concern and use the adjoint model to determine key atmospheric parameters influencing the forecast. The output from these sensitivity experiments is then used to create initial condition perturbations for a short-range ensemble prediction system, which is specifically designed to investigate the forecast concern of the day. They introduced nine cost functions that are believed to be important for deep convection. In their conclusions they state that with the perturbation sizes they used their method is superior in terms of the equitable threat scores and ranked probability scores between 6 and 24 h to ensembles based on physics perturbations or a certain Monte Carlo approach both described in Stensrud et al. (2000). Homar et al. (2006) compared an EPS using the method of Xu et al. (2001) to the National Centers for Environmental Prediction (NCEP) operational Short-Range Ensemble Forecasting (SREF) system

and the Eta model for a three week period. They show that the adjoint based system improved the numerical forecasts of severe weather and produces better probabilistics scores of heavy precipitation than the SREF. We show that by a suitable projection the CAPE-SVs are equal to optimal perturbations and could therefore be used as a cost function in the method of Xu et al. (2001).

One of the assumptions that has to be made in the SV calculation is that forecast errors are linearly related to initial condition errors. When going to higher resolutions and at the same time focusing on mesoscale aspects of the forecast, such as instantaneous convective precipitation, the linearity assumption becomes increasingly problematic, see Mahfouf and Bilodeau (2007), Hohenegger and Schär (2007) and Errico et al. (2003). As CAPE is associated with the large scale flow and the CAPE-field is a relatively smooth field compared to convective precipitation it is plausible that the tangent linear assumption holds for longer lead times compared to convective precipitation.

In section 2.2 we introduce singular vectors and derive a norm based on CAPE. We show how the CAPE-norm can be used in adjoint sensitivity studies and for the creation of optimal perturbations. In section 2.3 we investigate the properties of the CAPE-SVs for a two week summer period in 2007 and compare them with TE-SVs. In section 2.4 we dicuss the validity of the tangent linear assumption using twin experiments. Finally in section 2.5 we end with conclusions.

2.2 Theory

2.2.1 Singular vectors

In the singular vector calculation it is assumed that forecast errors $\delta x(T)$ are linearly related to errors $\delta x(0)$ at analysis time

$$\delta x(T) = \mathbf{M}(T, 0)\delta x(0) \tag{2.1}$$

where $\mathbf{M}(T, 0)$ is called the propagator or resolvent. The action of $\mathbf{M}$ on a vector $\delta x(0)$ is obtained by integrating the tangent linear model

$$\dot{\delta x}(t) = J(x(t))\delta x(t) \tag{2.2}$$

starting from the initial condition δx_0 and where $x(t)$ is the solution of the nonlinear forecast model

$$\dot{x} = F(x) \tag{2.3}$$

starting from the initialized analysis x_0. There are practical problems with obtaining the tangent linear model by linearizing the nonlinear atmospheric model such that $J(x(t)) = \frac{\partial F}{\partial x}\big|_{x(t)}$ due to the strong nonlinearities in the physical parameterizations of the models (Janisková, 2003; Vukićević and Errico, 1993; Fillion and Bélair, 2004).

Singular vectors are vectors $\delta x(0)$ that maximize the ratio

$$\frac{(\mathbf{P}\delta x(T), \mathbf{DP}\delta x(T))}{(\delta x(0), \mathbf{E}\delta x(0))} \tag{2.4}$$

where $\mathbf{D}$ is a positive semi-definite symmetric operator, $\mathbf{E}$ is a positive definite symmetric operator w.r.t. the Euclidean inner product $(\cdot,\cdot)$ and $\mathbf{P}$ is a projection operator (Barkmeijer, 1993). By using Lagrange multipliers it is easy to show that the solution to equation 2.4 is equal to the leading eigenvector of the following generalized eigenvector problem.

$$\mathbf{M}^*\mathbf{P}^*\mathbf{DPM}\delta x(0) = \lambda \mathbf{E}\delta x(0) \tag{2.5}$$

Here $\mathbf{M}^*$ is the adjoint of the propagator $\mathbf{M}$ w.r.t. the inner product $(\cdot,\cdot)$. This is a symmetric generalized eigenvalue problem and therefore the eigenvalues are real and can be ordered $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N \geq 0$. The n-th singular vector maximizes the ratio in equation 2.4 in a direction $\mathbf{E}$-orthogonal to the space spanned by the leading $(n-1)$- singular vectors. For each singular vector the singular value is given by $\sqrt{\lambda}$.

2.2.2 CAPE norm

The convective available potential energy based on the virtual temperature is defined as (Doswell III and Rasmussen, 1994)

$$\mathrm{CAPE} = g\int_{\mathrm{LFC}}^{\mathrm{EL}} \frac{T_v(z') - \bar{T}_v(z')}{\bar{T}_v(z')}\mathrm{d}z' \tag{2.6}$$

Where T_v is the virtual temperature of the parcel rising adiabatically from its source layer and $\bar{T}_v$ is the virtual temperature of the environment. The integral is only evaluated between heights where the integrand is positive. These heights are called the level of free convection (LFC) and the equilibrium level (EL). For computational simplicity we derived the final time CAPE-norm from a simplified CAPE calculation used in the post processing of an earlier version of the operational ECMWF-model (Bechtold, personal communication)

$$\mathrm{CAPE} = g\int_{\mathrm{LFC}}^{\mathrm{EL}} \frac{\theta_{e_{\mathrm{up}}} - \theta_{e_{\mathrm{sat}}}(z)}{\theta_{e_{\mathrm{sat}}}(z)}\mathrm{d}z' \tag{2.7}$$

Where $\theta_{e_{\mathrm{up}}}$ is the equivalent potential temperature of a parcel based on the average conditions in the lowest 50 hPa in the model and $\theta_{e_{\mathrm{sat}}}$ is the saturation equivalent potential temperature. The value of $\theta_{e_{\mathrm{up}}}$ is kept constant during the parcel ascent. The CAPE values from equation 2.7 are approximately 25 % higher than those computed using equation 2.6 (Bechtold, personal communication). Note that multiplying the final time norm in the singular vector calculation by a scalar has no impact on the structure of the singular vector. From the Taylor expansion of the CAPE calculation around a model state x we obtain

$$\mathrm{CAPE}(x+\delta x) - \mathrm{CAPE}(x) = \frac{\partial \mathrm{CAPE}}{\partial x}\delta x + O(\delta x^2) \tag{2.8}$$

Therefore the impact of an infinitesimal perturbation δx of the model state x on CAPE, measured in the Euclidean norm, is

$$||\mathrm{CAPE}(x+\delta x) - \mathrm{CAPE}(x)||^2 = ||\mathbf{C}\delta x||^2 = (\delta x, \mathbf{C}^*\mathbf{C}\delta x) \tag{2.9}$$

We will refer to $\mathbf{C} = \frac{\partial \mathrm{CAPE}}{\partial x}$ as the CAPE-operator. This operator maps 3 dimensional perturbations of the model state to 2 dimensional CAPE perturbations. The operator $\mathbf{C}^*\mathbf{C}$ is positive semidefinite and therefore defines a semi-norm on the model states.

The eigenvalue problem 2.5 can now be written as

$$\mathbf{E}^{-1/2}\mathbf{M}^*\mathbf{P}^*\mathbf{C}^*\mathbf{CPME}^{-1/2}v = \lambda v \tag{2.10}$$

where we have assumed that $\mathbf{E}$ is a simple metric in the sense that $\mathbf{E}^{-1/2}$ can easily be obtained, e.g. the metric used in the TE-norm (Errico et al., 2004). The solutions to the original maximization problem are now given by $\delta x = \mathbf{E}^{-1/2}v$

2.2.3 Squared mean versus mean squared SVs

Both positive and negative CAPE perturbations contribute to the CAPE norm. This typically leads to nearby areas of negative and positive CAPE perturbations, similar to the wave train structures when the TE-norm is used. In particular there is the possibility that perturbations which are large in the CAPE-norm have small impact on the mean CAPE value in a certain area. This may prohibit the analysis of how CAPE perturbations are associated with, for example, severe convective events. For this reason we also computed perturbations that have maximum impact on the square of the mean CAPE values as opposed to the mean of squared CAPE values. We will refer to these SVs as SCAPE-SVs. If we define a projection operator $\mathbf{\Sigma}$ on the CAPE perturbations as

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \dots & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{bmatrix} \tag{2.11}$$

Then, maximizing the squared mean CAPE perturbation using the total energy metric $\mathbf{E}$ at initial time is equivalent to solving the eigenvector problem

$$\mathbf{E}^{-1/2}\mathbf{M}^*\mathbf{P}^*\mathbf{C}^*\mathbf{\Sigma}^*\mathbf{\Sigma}\mathbf{CPME}^{-1/2}v = \lambda v \tag{2.12}$$

The projection operator $\mathbf{\Sigma}$ is a rank 1 mapping. Therefore there is only one singular vector with $\lambda \neq 0$. This singular vector is obtained up to a normalization by

$$\delta x = \mathbf{E}^{-1}\mathbf{M}^*\mathbf{P}^*\mathbf{C}^*1 \tag{2.13}$$

Here 1 denotes a vector of ones. In particular there is no need for a Lanczos scheme to determine this "leading" SCAPE-SV.

Similar expressions are obtained in Mahfouf and Bilodeau (2007) and Errico et al. (2003), where a cost function is maximized subject to constraints on the initial conditions. The resulting structures are referred to as optimal perturbations. The equivalence between their formulation and the squared mean SVs given here can be demonstrated by defining a cost-function as the sum of CAPE-values inside a projection domain

$$J = \left(\tilde{\mathbf{P}}\mathrm{CAPE}(x), 1\right) \tag{2.14}$$

Here $\tilde{\mathbf{P}}$ is a projection operator that sets CAPE-perturbations outside the projection domain to zero. Small perturbations δx in the model state lead to perturbations in CAPE

$$\delta\mathrm{CAPE} = \mathbf{C}\delta x \tag{2.15}$$

where $\mathbf{C}$ is the CAPE operator. We get

$$\delta J = \left(\tilde{\mathbf{P}}\delta\mathrm{CAPE}, 1\right) = \left(\mathbf{C}\delta x_T, \tilde{\mathbf{P}}^*1\right) = \left(\delta x_0, \mathbf{M}^*\mathbf{C}^*\tilde{\mathbf{P}}^*1\right) \tag{2.16}$$

and by definition

$$\delta J = (\nabla_{x_0} J, \delta x_0) \tag{2.17}$$

So we conclude that the adjoint sensitivity of the sum of CAPE is given by

$$\nabla_{x_0} J = \mathbf{M}^*\mathbf{C}^*\tilde{\mathbf{P}}^*1 \tag{2.18}$$

Maximization of the cost function under constraints on the initial conditions and introducing a projection $\mathbf{P}$ such that $\tilde{\mathbf{P}}\mathbf{C} = \mathbf{C}\mathbf{P}$ leads to expression 2.13. The optimal pertubations maximize the mean CAPE values while SCAPE-SVs maximize the square of the mean CAPE values. Therefore the remark in Mahfouf and Bilodeau (2007) that SVs are not equal to optimal perturbations is true but only in the sense that the SVs do not necessarily increase the mean CAPE values. The structures obtained using both methods are equal. The CAPE cost function can now be used in the method of Xu et al. (2001) to create initial condition perturbations.

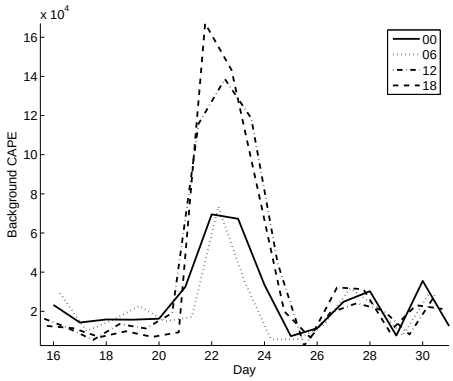

Figure 2.1: Sum of CAPE values inside the projection domain for the 12 hour forecast as a function of the verification time.

2.3 Case study

For all experiments we used HIRLAM 7.3beta1 with incremental 4DVAR using two outer loops and digital filtering initialization (DFI) as a weak constraint. The domain is approximately equal to the Euro-Atlantic region ($\approx 40^o$W–40^oE 35^oN–80^oN) with a resolution of 0.5^o. The singular vectors are calculated every 6 hours with an optimization time of 12 hours for a two week summer period starting from August 15, 2008 at 00 UTC till August 30, 12 UTC. This period is also used in the grand limited area model ensemble prediction system (GLAMEPS) project (Iversen et al., 2007) to test different short-range ensemble strategies. During this two week period a severe mesoscale convective system developed in the south west of Finland, which the operational Finnish model failed to predict in any cycle verifying at the same time (Iversen et al., 2007). We use analysis blending (Yang, 2005) such that the large scales are mixed with the ECMWF analysis. For a detailed description of the HIRLAM system we refer to Undén et al. (2002). We use the default settings for the tangent linear model which only includes vertical diffusion of the Météo France simplified physics package (Janisková et al., 1999). The nonlinear trajectory was updated every hour in the tangent linear model. At final time we target the fields to the area defined by 55^oN–65^oN and 10^oE–30^oE. At initial time we use a dry total energy norm similar to the computation of singular vectors used in the EPS at ECMWF (Leutbecher and Palmer, 2008), i.e. we do not perturb the specific humidity field.

Figure 2.1 shows the sum of CAPE values inside the projection domain for the 12 hour forecast as a function of verification time using the approximate CAPE calculation from which the CAPE norm was derived. Around August 22, during the mesoscale convective event, the CAPE values are relatively high and during

this period there is a clear diurnal cycle with lower CAPE values for forecasts that verify at 00 and 06 UTC, likely due to the cooler, stable boundary layer that develops at night.

2.3.1 Singular values

In all singular value plots we plot the average singular values of the leading ten SVs for the TE-SVs and CAPE-SVs and the singular value for leading SCAPE-SVs. We will refer to these as the average singular value. Figure 2.2 shows the average singular value using the final time TE-norm, CAPE-norm and SCAPE-norm. Both the CAPE singular values and the SCAPE singular values clearly indicate that the leading singular vectors have a large impact on the CAPE in the background flow around August 22. In the period around August 22 there is a daily cycle in the CAPE singular values with lower singular values for forecasts starting from 00 and 18 UTC.

By comparing the CAPE and SCAPE singular values with figure 2.1 it is clear that the higher singular values around August 22 correspond to high CAPE values in the 12 hour forecasts inside the projection domain during this period. Note however that the low CAPE values verifying at 00 UTC correspond to relatively high singular values according to the SCAPE and CAPE singular values starting from the 12 UTC analysis of the previous day. These results should be contrasted with the singular values for the TE-SVs. Here there is no increase in the singular values around August 22 and no daily cycle. In Figure 2.2 (lower right) we show the impact of the TE-SVs on CAPE, i.e. we measure the size of the evolved TE-SVs in the CAPE-norm. In this case we do see a clear signal in the "singular values". By comparing with the singular values of the CAPE-SVs we see that the CAPE-SVs are approximately 10 times more effective at perturbing CAPE compared to TE-SVs around August 22.

2.3.2 SV structures for August 21 at 18 UTC

The mesoscale convective system in the south west of Finland developed around August 22 at 6 UTC. Here we discuss the properties of the SVs using the TE, CAPE and SCAPE norm starting from the August 21 18 UTC analysis and verifying at August 22 at 6 UTC. Figure 2.3 shows the structure at initial and final time of the leading TE-SVs. At initial time the temperature perturbations show the typical tilted wave-train structure against the wind-shear. After 12 hours the structures are untilted and have propagated upward towards the jet-stream level. These characteristics are well known features of TE-SVs (Palmer et al., 1998). Due to the short optimization time used here the temperature perturbations at initial time are high in the troposphere. At initial time the energy is almost completely in the temperature field which is converted to kinetic energy at final time. These are consistent features of all the TE-SVs in the two week period (see figure 2.6).

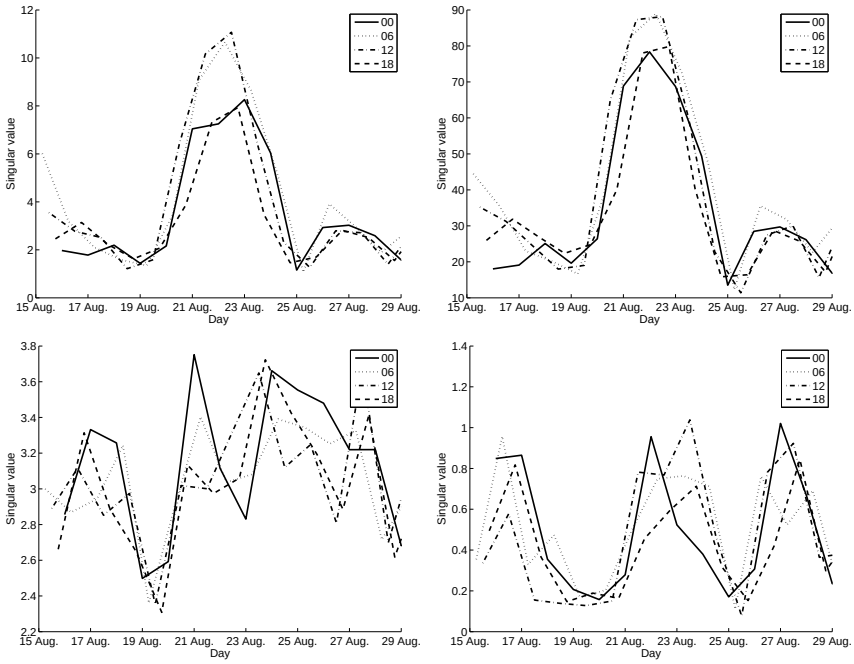

Figure 2.2: Average singular value of the CAPE-SV (top left), SCAPE-SV (top right) and TE-SVs (bottom left) as a function of the analysis time. In the lower right we show the average "singular value" of the TE-SVs in the CAPE-norm.

Figure 2.4 and 2.5 show the structure at initial and final time of the leading CAPE-SV and the leading SCAPE-SV. Compared with the leading TE-SV both types of CAPE-SVs are situated much lower in the troposphere and the CAPE-SV is more localized in the horizontal. At final time the temperature field is increased in the low troposphere and decreased in the mid troposphere for both the CAPE and SCAPE-SVs. The impact of the evolved TE-SVs on CAPE is small compared to the CAPE-SVs.

For this particular day the leading CAPE-SVs increases CAPE for almost all grid points inside the projection domain and therefore also increases the mean CAPE values, i.e. we do not have wave train structures in the evolved CAPE-field. As a result of this the SCAPE-SV, which maximizes the mean CAPE, projects strongly on the CAPE-SVs in the south west of the projection domain. More generally, using the sign ambiguity in the CAPE-SVs, we can multiply all CAPE-SVs with the sign of the mean evolved CAPE field inside the projection domain such that the CAPE-SVs all have a nonnegative impact on the mean CAPE-field. When the SCAPE-SV is written as a linear combination of these CAPE-SVs it

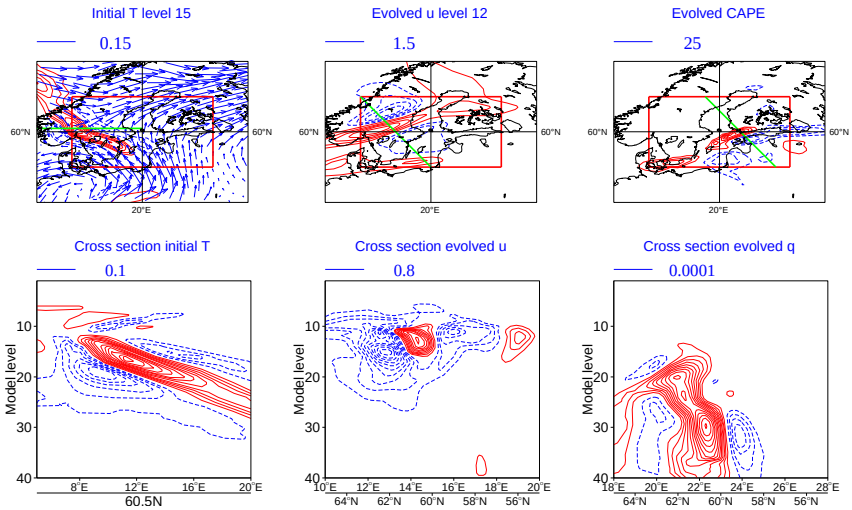

Figure 2.3: Leading TE-SVs for August 21 at 18 UTC, initial temperature and background wind at model level 15 (top left) cross section initial temperature (bottom left), evolved u-component wind at level 12 (top middle), cross section u-component wind (bottom middle), Evolved CAPE (top right) and cross section specific humidity (bottom right). The vertical cross sections are along the green line in the upper plot. The red box indicates the target area for the SVs. Positive perturbations (solid red), negative perturbations (dashed blue). The evolved fields are valid at 12h forecast time. The contour interval is given by the number above each panel.

will mainly project on those CAPE-SVs that have a large positive bias in the mean CAPE field. Being linear combinations of CAPE-SVs that all have a positive bias in the CAPE field the scale of the CAPE-field of the SCAPE-SVs is larger than the scale of the CAPE-field of the CAPE-SVs.

2.3.3 Vertical structures

Figures 2.6, 2.7 and 2.8 show the average vertical energy distribution inside the projection domain of the TE-SVs, the CAPE-SVs and the SCAPE-SV respectively. For the TE-SVs and the CAPE-SVs the average vertical energy was based on the leading nine SVs. For each day each average field is scaled such that the maximum value is 1. The upper right plots show the energy distribution inside the projection domain as a function of the analysis time for temperature, kinetic energy and specific humidity in terms of TE. For the TE-SVs on average 80% of the energy is in the temperature field at initial time. The height where this initial temperature maximum occurs varies between 250 and 450 hPa (model levels

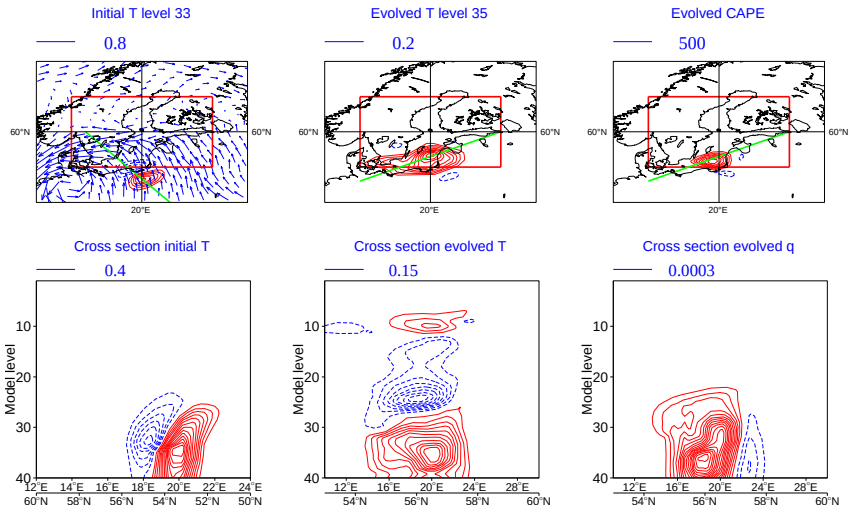

Figure 2.4: Leading CAPE-SV for August 21 at 18 UTC, initial temperature and background wind at model level 33 (top left) cross section initial temperature (bottom left), evolved temperature at level 35 (top middle), cross section temperature (bottom middle), Evolved CAPE (top right) and cross section specific humidity (bottom right). The vertical cross sections are along the green line in the upper plot. The red box indicates the target area for the SVs. Positive perturbations (solid red), negative perturbations (dashed blue). The evolved fields are valid at 12h forecast time. The contour interval is given by the number above each panel.

11–17). At evolved time almost all energy is in the kinetic energy (note that the evolved kinetic energy values are divided by 5 in the upper right plot) between 250 hPa and 350 hPa (model levels 11–15). We use a dry total energy norm at initial and final time and so there is no role of the specific humidity field in the SV computation. In the fields that contribute to the total energy the TE-SVs show little variability as a function of time.

For the CAPE-SV on average 70% of the energy is in the temperature field at initial time with the remaining 30% in the kinetic energy. These temperature perturbations are always between 850–950 hPa. For SCAPE-SVs these ratio's are 80% and 20% for temperature and kinetic energy on average respectively. There is also much more variability in the height of the initial temperature perturbation. Especially in periods with low singular values the initial temperature field of the SCAPE-SV tends to be higher in the troposphere compared to the CAPE-SVs.

At evolved time the SCAPE-SVs have more energy in the temperature field compared to the CAPE-SV. The energy in this evolved temperature field is relatively constant as a function of time compared to the energy in the evolved specific

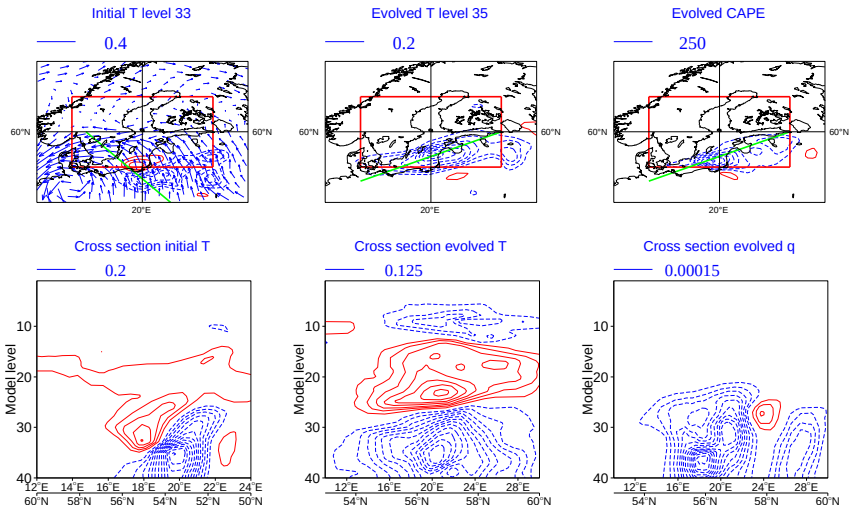

Figure 2.5: SCAPE-SVs for August 21 at 18 UTC, initial temperature and background wind at model level 33 (top left) cross section initial temperature (bottom left), evolved temperature at level 35 (top middle), cross section temperature (bottom middle), Evolved CAPE (top right) and cross section specific humidity (bottom right). The vertical cross sections are along the green line in the upper plot. The red box indicates the target area for the SVs. Positive perturbations (solid red), negative perturbations (dashed blue). The evolved fields are valid at 12h forecast time. The contour interval is given by the number above each panel.

humidity field and the evolved kinetic energy. Note that wind perturbations at final time have no impact on the CAPE values. These wind perturbations are however an effective way to create moisture perturbations at final time, which is important in creating CAPE. Both CAPE and SCAPE-SVs mainly affect the specific humidity around 850 hPa (model level 30) with little variability in the vertical structure compared to the evolved kinetic energy and temperature.

The results presented here all used a dry total energy norm at intial time (no specific humidity perturbations). We performed experiments with SVs with specific humidity perturbations included at initial time using different values for the coefficient in the TE-norm, which measures the energy content in the specific humidty humidity field (see Barkmeijer et al. (2001)). What we observed was that part of the energy was in the specific humidity field at initial time, as expected, but the structures of the temperature and wind field were equal to SVs computed without allowing specific humidity perturbations. This is probably related to the exclusion of important diabatic physics in the tangent linear model.

We selected four time periods to examine in more detail the large fluctuations in the evolved specific humidity and the vertical extent of the evolved temperature

field for the SCAPE-SVs (see caption Fig. 2.9 for the definition of the four time periods).

In the first and third period the evolved specific humidity field is dominant in terms of TE. In addition this third period is characterized by a large vertical extent in the evolved temperature field and high singular values. The second and fourth period are characterized by low values for the evolved specific humidity field. Figure 2.9 shows the horizontally averaged evolved perturbation of temperature and specific humidity inside the projection domain averaged over the four periods. For all periods we observe that the CAPE-SVs increase the temperature and specific humidity of the parcel, defined as the average condition in the lowest 50 hPa (lowest 5 model levels). This effect is the strongest for the third period with the highest singular values. Furthermore the specific humidity and lapse rate are increased above the lowest 50 hPa. The height over which the lapse rate is increased strongly depends on the synoptic situation. During the second period the lapse rate is increased in a very shallow region around 850 hPa while for the third period the lapse rate over the entire layer from 950 hPa and 400 hPa is increased. This dependence of the vertical temperature profile on the synoptic situation will be difficult to account for in cost functions defined at single model levels or in the planetary boundary layer such as used in Xu et al. (2001). The heating and moistening of the parcel will, in most cases, reduce the convective inhibition and will therefore decrease the lifting needed for a parcel to reach its level of free convection. Therefore the CAPE-SVs are beneficial to all three necessary ingredients for deep convection mentioned in the introduction.

From Fig. 2.8 we observe that the evolved specific humidity field of the SCAPE-SVs in terms of TE have the largest impact around 850 hPa (model level 30), while in Fig. 2.9 the maximum occurs at 950 hPa (model level 35). The effect can be seen most clearly for the third period in Fig. 2.9. This is caused partly because the model levels represent less mass in the lower troposphere with the result that perturbations at lower levels have less impact on the TE. The same effect is observed in the evolved temperature field. Furthermore at higher model levels there are regions with both positive and negative contributions to the specific humidity field such that the average specific humidity perturbation is smaller while in terms of energy there is a maximum (see Figure 2.5).

2.4 Linearity test twin experiments

To test the potential usefulness of the tangent linear model we performed twin experiments for the “Finnish case” similar to the nonlinear test in Vukićević and Errico (1993). A necessary condition for the tangent linear model to be potentially useful is that the error growth in the nonlinear run behaves linearly according to a predefined error measure. Figure 2.10 shows the CAPE perturbation at initial and final time using the leading CAPE-SV as initial condition perturbation. The leading CAPE-SV was multiplied by 0.1 resulting in a perturbation with

maximum values for temperature of 0.58 K and for wind of 0.38 m/s. As we used a dry TE-norm to constain the initial perturbations, the CAPE perturbation over the north of Poland at initial time is determined completely by the temperature field. At final time the specific humidity is the dominant field in terms of energy and the CAPE perturbation is now located over the Baltic sea. As shown in figure 2.10 the result of adding or subtracting the CAPE-SV gives very similar CAPE perturbations in the nonlinear runs indicating that the tangent linear assumption is valid if CAPE is used as verification measure. The tangent linear model however overestimates the amplitude of the CAPE perturbation.

Although these results look promising we noticed that already after 1 hour small scale noise with low amplitude is present in the boundary layer for the entire integration domain in the twin experiment runs for the temperature and wind field. We noticed that vertical integrals such as CAPE are rather insensitive to this small scale noise for reasonably sized initial perturbations, but similarity indices between the full 3D fields are low. When the amplitude of the initial condition perturbation is decreased the noise starts to interfere too much with the SV growth when twin experiments are performed leading to lower similarity indices. Similar behavior has been observed by Vukićević and Errico (1993). They conclude that the dominant mechanism for very small perturbation evolution is the forcing of gravity waves for the regions and periods where the convection is sensitive to $2\Delta t$ oscillations. We will investigate the loss of linearity of HIRLAM for these cases in a follow-up paper. Note that the loss of linearity for the small amplitude twin experiment does not imply that the results of the tangent linear model presented here are invalid but might actually reveal problems with the nonlinear model itself.

2.5 Conclusions and discussion

The feasibility of using the singular vector technique to create initial condition perturbations for short-range ensemble predictions has been investigated. For this a new final time norm based on CAPE has been introduced. We have studied the properties of two types of CAPE-SVs and compared their structure with TE-SVs for a two week summer period in 2007 during which a severe mesoscale convective system in the south west of Finland was observed. Both types of CAPE-SVs show a clear increase in the singular values indicating that small perturbations can give rise to large CAPE perturbations during the mesoscale convective event. During this period there is a daily pattern in the singular values with lower values for singular vectors starting from the 00 and 18 UTC analysis. TE-SVs show no increase in the singular values for this period and there is no daily pattern.

During the two week period all SCAPE-SVs increase CAPE at final time by increasing the specific humidity and temperature of the parcel and at the same time increasing the lapse rate above the parcel. The increase in moisture content and temperature in the low troposphere will also decrease the absolute value

of CIN and therefore reduce the lift needed for parcels to reach their LFC. So, although the CAPE-SVs do not provide the lift needed to reach the LFC, most convection schemes will be easier triggered in runs that are perturbed with CAPE-SVs. Therefore adding CAPE-SVs to the initial condition is beneficial to all three necessary ingredients for development of deep convection mentioned in the introduction (Doswell III, 1987). Another useful property of both types of CAPE-SVs for short-range EPS is that their activity mainly takes place in the lower troposphere. This in contrast to TE-SVs with OT=12h which predominantely remain in the high troposphere (above 500 hPa). Note that although the specific humidity component has the largest contribution in terms of TE in the evolved CAPE-SVs there is also perturbation growth in the kinetic energy component. We plan to explore the impact of this different behaviour in the context of a short-range ensemble systems (GLAMEPS, Iversen et al. (2007)) focussing particularly on the performance with respect to weather parameters.

For certain vertical temperature and humidity profiles parcels will never reach their LFC. In those cases there are no infinitesimal perturbations that will produce CAPE and therefore the CAPE operator acts as a projection operator for these points. This also happens for cost functions based on precipitation (Mahfouf and Bilodeau, 2007) where only levels where supersaturation is diagnosed in the background state contribute to the norm. One solution to avoid this projection is to change the lower integration limit in the CAPE integral to the initial parcel height instead of the LFC. This amounts to substracting the energy needed to reach the LFC (the convective inhibition) from the CAPE. Optimal perturbation based on this norm will decrease the absolute value of CIN for those specific humidity and temperature profiles with zero CAPE and both decrease the absolute value of CIN and increase CAPE for profiles with positive CAPE values.

We have shown that the CAPE-norm can be used in the method of Xu et al. (2001) to create optimal perturbations with a cost function that measures the mean CAPE inside a projection domain. Short-range (ensemble) forecasting has to deal with many forecast aspects. Several indices have been developed to evaluate the convective and severe weather potential of the atmosphere such as the Showalter index, lifted index, total totals, total energy index, severe weather threat, bulk Richardson number, helicity index, vorticity generation potential, etc. (Blanchard, 1998) and Peppler (1988)). For indices that depend differentiably on the model state the same method can be used to derive a norm and create optimal perturbations. These optimal perturbations are an interesting alternative for short-range EPS based on singular vectors. First of all, each ensemble member can be constructed to focus on a particular forecast aspect by using a norm derived from a related weather index. Note that we can easily adjust the optimization time and the resolution in the adjoint integration, e.g. we can choose short optimization times and higher resolutions when norms are used that focus on mesoscale aspects of the forecasts. Secondly, when several indices are being used to compute the optimal perturbations, the impact of each optimal perturbation can be judged on their impact on ensemble skill scores. This information

could be used to gradually improve the norms used to create the initial condition perturbations.

Acknowledgements

We would like to thank Nils Gustafsson, Toon Moene, Ole Vignes and Xiaohua Yang for support during the implementation of the singular vector code in HIRLAM. We thank Theo Opsteegh, Sander Tijm and Wim Verkley for carefully reading earlier versions of the manuscript and two anonymous reviewers for their useful comments.

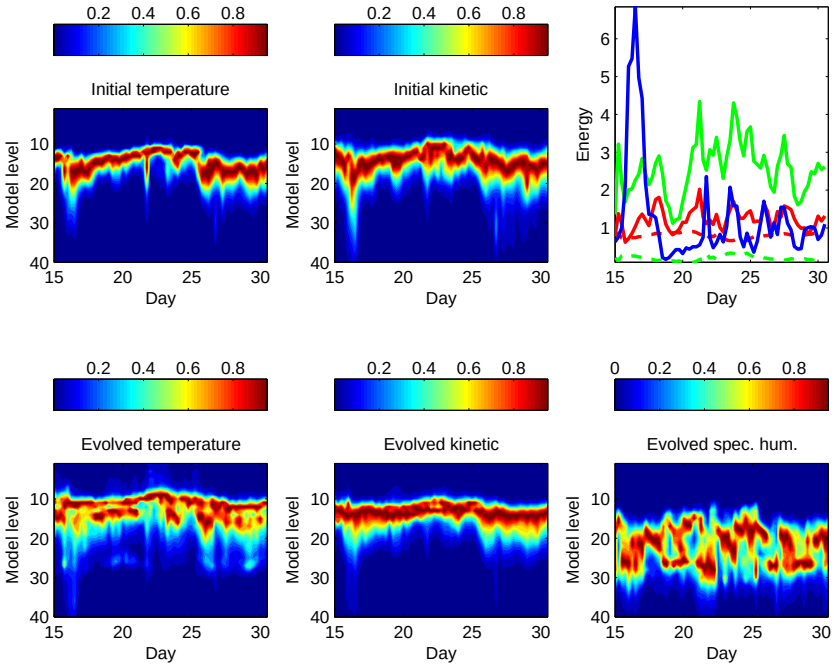

Figure 2.6: Scaled average energy distribution of the TE-SV as a function of analysis time. Initial temperature (top left) initial kinetic (top middle), evolved temperature (bottom left), evolved kinetic (bottom middle) and evolved specific humidity (bottom right). Each field is scaled such that the maximum is 1. The total energy inside the projection domain for each field is shown in the upper right with temperature (red) kinetic energy (green) and specific humidity (blue). Values at initial time (dashed) values at evolved time (solid). The evolved kinetic energy is divided by 5 in the upper right plot.

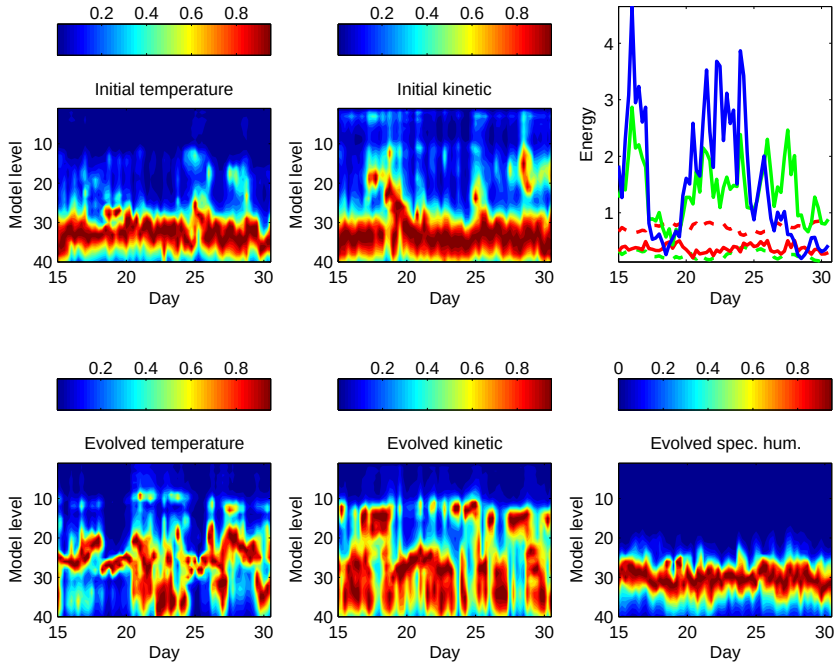

Figure 2.7: Scaled average energy distribution of the CAPE-SV as a function of analysis time. See caption figure 2.6. The evolved specific humidity energy is divided by 2 in the upper right plot.

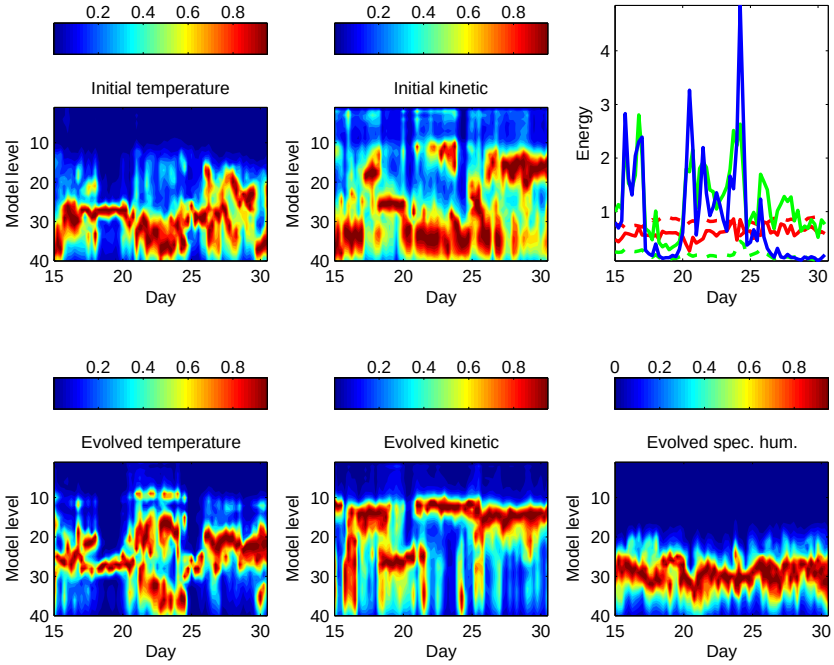

Figure 2.8: Scaled energy distribution of the SCAPE-SV as a function of analysis time. See caption figure 2.6. The evolved specific humidity energy is divided by 2 in the upper right plot.

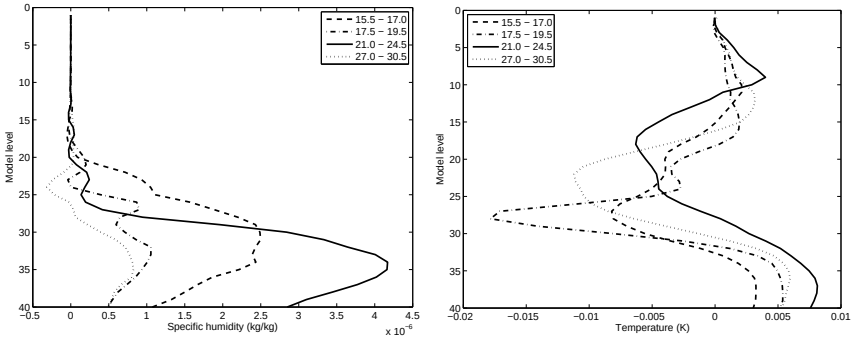

Figure 2.9: Horizontally averaged perturbation specific humidity field (left) and perturbation temperature field (right) inside the projection domain for the SCAPE-SVs averaged over four time periods. Period 1: August 15, 06 UTC to August 17, 00 UTC (dash) period 2: August 17, 06 UTC to August 19, 06 UTC (dash-dotted) period 3: August 21, 00 UTC to August 24, 06 UTC (solid) and period 4: August 27, 00 UTC to August 30, 6 UTC (dotted).

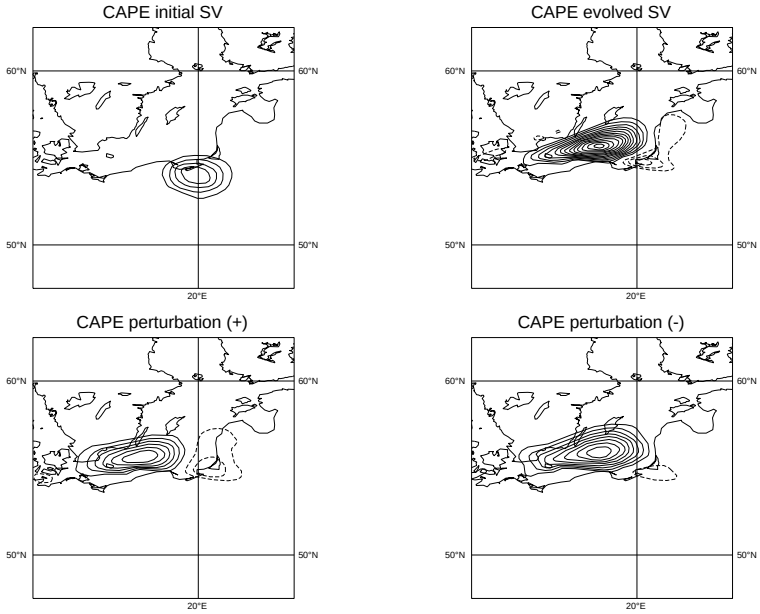

Figure 2.10: CAPE perturbations at initial time (top left) and final time in the tangent linear model (top right) and in the two nonlinear runs (bottom left and bottom right). The CAPE values in the (-) run are multiplied with -1 for ease of comparison. The contour interval is 30 J/kg in all panels.

CHAPTER 3

OPTIMAL LINEARIZATION TRAJECTORIES FOR TANGENT LINEAR MODELS

We examine differential equations where nonlinearity is a result of the advection part of the total derivative or the use of quadratic algebraic constraints between state variables (such as the ideal gas law). We show that these types of nonlinearity can be accounted for in the tangent linear model by a suitable choice of the linearization trajectory. Using this optimal linearization trajectory we show that the tangent linear model can be used to reproduce the exact nonlinear error growth of perturbations for more than 200 days in a quasi geostrophic model and more than (the equivalent of) 150 days in the Lorenz 96 model. We introduce an iterative method, purely based on tangent linear integrations, that converges to this optimal linearization trajectory.

The main conclusion from this paper is that this iterative method can be used to account for nonlinearity in estimation problems without using the nonlinear model. We demonstrate this by performing forecast sensitivity experiments in the Lorenz 96 model and show that we are able to estimate analysis increments that improve the two day forecast using only four backward integrations with the tangent linear model.

This text is published as: Optimal linearization trajectories for tangent linear models, Q. J. R. Meteorol. Soc., (Stappers and Barkmeijer, 2012)

3.1 Introduction

The use of tangent linear and, in particular, adjoint models has been very useful in several applications in numerical weather prediction (NWP) (see Errico, 1997, 2003; Errico and Ehrendorfer, 2007, for an overview). For example at the European Centre for Medium-Range Weather Forecasts (ECMWF) these linear models play a crucial role in the computation of initial condition perturbations used in the ensemble prediction system (Leutbecher and Palmer, 2008) and in their 4D-VAR data assimilation system (Courtier et al., 1994). One of the major limitations to the application of linear models is that the results are useful only when the linear approximation is valid (Errico, 1997). By this we mean that the difference between two runs of the nonlinear model can be described by the associated linearized version of the nonlinear model. To achieve this, great effort is taken to develop linearized models which capture as many as possible features of the full nonlinear model (Janisková et al., 1999). Despite these efforts the use of tangent linear and adjoint models is restricted to "short" time spans. The time span for which the TL-model can be considered accurate will be referred to as the tangent linear regime.

The duration of the tangent linear regime depends on many factors. Typically the difference between two nonlinear forecasts is compared with the linear forecast by a scalar index, and it is said that the TL-assumption is violated when the index has reached a threshold value. So the measure which is employed to compare forecast fields is already important in the definition of the TL-regime. But also the size of the initial condition perturbation, the orientation of the perturbation, the background trajectory around which the TL-model is linearized and the physical processes taken into account in the TL-model play a role. Another issue which influences the usefulness of linear models is whether we are considering forecast problems, where error growth is determined by the singular value spectrum of the propagator, or estimation problems that are typically characterized by the reciprocal of the singular value spectrum. In general the spectrum of the reciprocal of the singular values attains higher values (Reynolds and Palmer, 1998) and therefore the usefulness of the TL-model in estimation problems is shorter. This effect becomes even more pronounced by the fact that the typical size of perturbations used in backward integrations is larger than in forward mode.

In this forward mode the tangent linear assumption is generally believed to be valid for 2-3 days at the synoptic scale. However, see Gilmour et al. (2001) who argue that 1 day is perhaps a better estimate. On the cloud resolving scale the tangent linear assumption probably holds for much shorter time periods on the order of 1.5 hours (Hohenegger and Schär, 2007). As models reach higher resolutions the validity of the TL assumption is therefore a major concern. We will show that for bilinear systems the usefulness of the TL-model can be greatly extended by modifying the linearization trajectory and therefore the one of the major limitations for using tangent linear models can be eliminated.

In section 3.2 the definition of bilinear differential equations is given. In sec-

tion 3.3 we show that for bilinear systems there is an optimal linearization trajectory such that if the tangent linear model is linearized around this trajectory the perturbation growth in the tangent linear model is equal to the nonlinear perturbation growth. Knowing that such a trajectory exist we show in section 3.4 that there is an iterated map based purely on TL-integrations that converges to this linearization trajectory. In section 3.5 we show how the iterative method can be used in forecast sensitivity experiments using the inverse of the TL-model. In section 3.6 the experimental results using a quasi geostophic model (Marshall and Molteni, 1993, described in appendix 3.A) and the Lorenz 96 model (Lorenz, 1996, described in appendix 3.B) are given. In the discussion in section 3.7 the prospects for using the method in realistic NWP models and a method to regularize the error growth in the tangent linear model are discussed. The conclusions are given in section 3.8.

To keep the notation simple we use the convention that lower case variables are perturbations (also referred to as increments) to upper case variable, e.g. $\mathbf{x}$ is a perturbation to the state vector $\mathbf{X}$.

3.2 Bilinear differential equations

In this section some terms are defined that will be used throughout the rest of the paper. Let $\mathbf{U}, \mathbf{V}, \mathbf{W}$ and $\mathbf{X}$ be elements of a vector space $\mathbb{H}$.

Definition 1 (Bilinear map)**.** A map q

$$\begin{aligned} q : \mathbb{H} \times \mathbb{H} &\to \mathbb{H} \\ (\mathbf{U}, \mathbf{V}) &\mapsto \mathbf{W} = q(\mathbf{U}, \mathbf{V}), \end{aligned}$$

is called bilinear if q is linear in both arguments.

Definition 2 ((Anti)symmetric bilinear map)**.** A bilinear map s will be called symmetric if for any $\mathbf{U}, \mathbf{V} \in \mathbb{H}$

$$s(\mathbf{U}, \mathbf{V}) = s(\mathbf{V}, \mathbf{U}). \tag{3.1}$$

A bilinear map a will be called antisymmetric if

$$a(\mathbf{U}, \mathbf{V}) = -a(\mathbf{V}, \mathbf{U}). \tag{3.2}$$

Note that for any bilinear map q there is a unique decomposition

$$q(\mathbf{U}, \mathbf{V}) = s(\mathbf{U}, \mathbf{V}) + a(\mathbf{U}, \mathbf{V}), \tag{3.3}$$

where

$$s(\mathbf{U}, \mathbf{V}) = \frac{1}{2}\left(q(\mathbf{U}, \mathbf{V}) + q(\mathbf{V}, \mathbf{U})\right)$$

is symmetric and

$$a(\mathbf{U},\mathbf{V}) = \frac{1}{2}\left(q(\mathbf{U},\mathbf{V}) - q(\mathbf{V},\mathbf{U})\right)$$

is antisymmetric.

Definition 3 (Bilinear differential equation)**.** A differential equation will be called bilinear if it is of the form

$$\dot{\mathbf{X}} = q(\mathbf{X},\mathbf{X}) + b(\mathbf{X}) + c, \tag{3.4}$$

where q is a bilinear map, b is a linear map and c is a forcing. If $\mathbf{X}$ is finite dimensional this is an ordinary differential equation (ODE) while if $\mathbf{X}$ represents a (collection of) space and time dependent field(s) this is a partial differential equation (PDE). For PDE's the mappings q, b and forcing c are allowed to depend on space and time explicitly.

Definition 4 (Bilinear differential algebraic equation)**.** A differential algebraic equation (Brenan et al., 1996) will be called bilinear if it is of the form

$$e(\mathbf{X},\dot{\mathbf{X}}) + d(\dot{\mathbf{X}}) = q(\mathbf{X},\mathbf{X}) + b(\mathbf{X}) + c, \tag{3.5}$$

where q and e are bilinear maps, d and b are linear maps and c is a forcing.

Example 1 (Barotropic vorticity equation)**.** The barotropic vorticity equation (bve) is

$$\begin{aligned}
\frac{\partial \eta}{\partial t} &= -J(\psi,\eta) \\
0\frac{\partial \psi}{\partial t} &= \nabla^2\psi + f - \eta \\
0\frac{\partial \mathbf{u}}{\partial t} &= \mathbf{u} - \mathbf{k}\times\nabla\psi
\end{aligned}$$

where the first equation is a prognostic equation for the absolute vorticity η, the second and third equation are algebraic constraints (diagnostic equations) for the stream function ψ and the two dimensional velocity $\mathbf{u}$ respectively (hence the zeros in front of the time derivatives), f is the Coriolis parameter, $\mathbf{k}$ is the vertical unit vector and J is an antisymmetric bilinear map defined as

$$J(\psi,\eta) = \frac{\partial\psi}{\partial x}\frac{\partial\eta}{\partial y} - \frac{\partial\psi}{\partial y}\frac{\partial\eta}{\partial x}. \tag{3.6}$$

If we define $\mathbf{X} = (\eta,\psi)^T$ we see that the bve is a bilinear partial differential algebraic equation in the state vector $\mathbf{X}$ with $e = 0$ and $d = \mathrm{diag}(I,0)$ and the velocity is a "post-processed" variable. Alternatively the equation for $\frac{\partial\eta}{\partial t}$ can be written as $\frac{\partial\eta}{\partial t} = -\mathbf{u}\cdot\nabla\eta$ in which case the state vector should be defined as $\mathbf{X} = (\eta,\psi,\mathbf{u})^T$.

Example 2 (Momentum equation)**.** The momentum equation in a uniform rotating coordinate frame is (Pedlosky, 1987)

$$\rho\frac{\partial \mathbf{u}}{\partial t} + \rho(\mathbf{u}\cdot\nabla)\mathbf{u} + 2\mathbf{\Omega}\times\rho\mathbf{u} = -\nabla p - \rho\nabla\phi + \mathcal{F}(\mathbf{u}),$$

where $\mathbf{u}$ is the three dimensional velocity vector, p is pressure, ρ is density, $\mathbf{\Omega}$ is the angular rotation vector, ϕ is the potential that represents conservative body forces, including gravity, $\mathcal{F}$ represents nonconservative (frictional) forces. The prognostic equation for $\mathbf{u}$ is a trilinear differential equation due to the term $\rho(\mathbf{u}\cdot\nabla)\mathbf{u}$ caused by the advection part of the total derivative. It is however easy to transform the trilinear equation to a bilinear differential algebraic equation by augmenting the state vector with the momentum density $\mathbf{p} = \rho\mathbf{u}$

$$\begin{aligned}\rho\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{p}\cdot\nabla)\mathbf{u} + 2\mathbf{\Omega}\times\mathbf{p} &= -\nabla p - \rho\nabla\phi + \mathcal{F}(\mathbf{u})\\ 0\frac{\partial \mathbf{p}}{\partial t} &= \rho\mathbf{u} - \mathbf{p}.\end{aligned}$$

Alternatively the momentum density vector field $\mathbf{p}$ can be considered as the prognostic variable

$$\begin{aligned}\frac{d\mathbf{p}}{dt} + 2\mathbf{\Omega}\times\mathbf{p} &= -(\nabla\cdot\mathbf{u})\mathbf{p} - \nabla p - \rho\nabla\phi + \mathcal{F}(\mathbf{u})\\ 0\frac{\partial \mathbf{u}}{\partial t} &= \mathbf{p} - \rho\mathbf{u},\end{aligned}$$

where we used the mass continuity equation.

Example 3 (Equation of state)**.** The equation of state for an ideal gas can be formulated as an algebraic constraint as

$$0\frac{\partial p}{\partial t} = p - \rho RT. \tag{3.7}$$

Examples 1 and 2 illustrate that in fluid dynamics bilinearity is typically a result of the advection part of the total derivative. Example 3 shows that another source for bilinearity is the use of algebraic constraints between state variables such as the ideal gas law. Example 2 further illustrates that it is easy to reduce multilinear systems (see appendix 3.D) to bilinear systems by augmenting the state vector.

Notation 1 (Nonlinear integrations)**.** Integrations with a nonlinear model starting from an initial condition $\mathbf{X}_0$ are denoted by

$$\mathbf{X}(t) = \mathcal{M}(t, 0, \mathbf{X}_0). \tag{3.8}$$

Then, by definition, the exact increment trajectory for a given perturbation $\mathbf{x}_0$ of an initial condition $\mathbf{X}_0$ is given by

$$\mathbf{x}(t) \equiv \mathcal{M}(t, 0, \mathbf{X}_0 + \mathbf{x}_0) - \mathcal{M}(t, 0, \mathbf{X}_0). \tag{3.9}$$

Notation 2 (Tangent linear integrations)**.** Integrations with the tangent linear model starting with an initial condition perturbation $\mathbf{x}_0$ are denoted by

$$\hat{\mathbf{x}}(t) = \mathbf{M}_{\mathbf{X}}(t, 0)\mathbf{x}_0, \tag{3.10}$$

where $\mathbf{M}_{\mathbf{X}}(t, 0)$ is known as the propagator and $\mathbf{X}$ is the trajectory around which the tangent linear model is linearized.

3.3 Optimal linearization trajectories

In this section we derive the tangent linear model for the general form of bilinear system and show how to modify the linearization trajectory to obtain an exact correspondence between the nonlinear time evolution and the corresponding tangent linear evolution of perturbations.

Consider the general form of a bilinear differential equation

$$\dot{\mathbf{X}} = q(\mathbf{X}, \mathbf{X}) + b(\mathbf{X}) + c, \tag{3.11}$$

where $\mathbf{X} \in \mathbb{H}$, q is a bilinear mapping, b is a linear mapping and c is a forcing and the mappings q and b and forcing c are allowed to explicitly depend on space and time. Solutions (trajectories in $\mathbb{H}$) of (3.11) are denoted as $\mathbf{X}(t)$. The time evolution of a perturbed run $\mathbf{X}(t) + \mathbf{x}(t)$ is given by

$$\dot{\mathbf{X}} + \dot{\mathbf{x}} = q(\mathbf{X} + \mathbf{x}, \mathbf{X} + \mathbf{x}) + b(\mathbf{X} + \mathbf{x}) + c. \tag{3.12}$$

Now using the bilinearity of q, the linearity of b and (3.11) to eliminate $\dot{\mathbf{X}}$ we obtain

$$\dot{\mathbf{x}} = \underbrace{q(\mathbf{X}, \mathbf{x}) + q(\mathbf{x}, \mathbf{X}) + b(\mathbf{x})}_{\mathbf{J}(\mathbf{X})\mathbf{x}} + q(\mathbf{x}, \mathbf{x}). \tag{3.13}$$

Here we used the linearity of q in both arguments and the linearity of b to define an operator

$$\mathbf{J}(\mathbf{X})\mathbf{x} \equiv q(\mathbf{X}, \mathbf{x}) + q(\mathbf{x}, \mathbf{X}) + b(\mathbf{x}). \tag{3.14}$$

For finite dimensional systems $\mathbf{J}(\mathbf{X})$ is the Jacobian of (3.11) evaluated along the trajectory $\mathbf{X}(t)$. In the tangent linear approximation the bilinear term $q(\mathbf{x}, \mathbf{x})$ is neglected and the system

$$\dot{\hat{\mathbf{x}}}^1 = \mathbf{J}(\mathbf{X})\hat{\mathbf{x}}^1 \tag{3.15}$$

is known as the tangent linear model. We use a hat to indicate that this is only an approximation to the true evolution $\mathbf{x}$ and the reason to add the superscript 1 will become apparent later.

The key observation in this section is that the exact time evolution of perturbations in (3.13) can also be written as

$$\dot{\mathbf{x}} = \mathbf{J}(\mathbf{X} + \mathbf{x}/2)\mathbf{x}, \tag{3.16}$$

i.e. we obtain the exact time evolution of perturbations if the tangent linear model is linearized around the trajectory $\mathbf{X}+\mathbf{x}/2$ instead of $\mathbf{X}$. The trajectory $\mathbf{X}+\mathbf{x}/2$ will be referred to as the optimal linearization trajectory. The previous results can be generalized to bilinear partial differential algebraic equations (BPDAE). Let

$$e(\mathbf{X}, \dot{\mathbf{X}}) + d(\dot{\mathbf{X}}) = q(\mathbf{X}, \mathbf{X}) + b(\mathbf{X}) + c, \tag{3.17}$$

substitution of $\mathbf{X} \to \mathbf{X}+\mathbf{x}$, using (3.17) and retaining only terms linear in $\mathbf{x}$ gives the TL-model

$$e(\hat{\mathbf{x}}^1, \dot{\mathbf{X}}) + e(\mathbf{X}, \dot{\hat{\mathbf{x}}}^1) + d(\dot{\hat{\mathbf{x}}}^1) = 2s(\mathbf{X}, \hat{\mathbf{x}}^1) + b(\hat{\mathbf{x}}^1).$$

where we used definition 2 to write $q(\mathbf{X}, \hat{\mathbf{x}}^1) + q(\hat{\mathbf{x}}^1, \mathbf{X}) = 2s(\mathbf{X}, \hat{\mathbf{x}}^1)$. It is easy to see that the neglected bilinear terms $e(\mathbf{x}, \dot{\mathbf{x}})$ and $q(\mathbf{x}, \mathbf{x})$ are recovered if the TL-model is linearized around the trajectory $\mathbf{X} + \mathbf{x}/2$. An important difference with the previous result is that to integrate the TL-model both the trajectory $\mathbf{X}$ and the tendencies $\dot{\mathbf{X}}$ are required if $e \neq 0$.

Integrations with the tangent linear model linearized around a trajectory $\mathbf{X}+\mathbf{x}/2$ starting from an initial condition $\mathbf{x}_0$ will be denoted by

$$\mathbf{x}(t) = \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t, 0)\mathbf{x}_0. \tag{3.18}$$

Note that, although the tangent linear model is used to propagate the increment $\mathbf{x}_0$, this is not a linear mapping from $\mathbf{x}_0$ to $\mathbf{x}(t)$ due to the dependence of the linearization trajectory on $\mathbf{x}_0$. In appendix 3.C we discuss how to preserve bilinearity when higher than first order integration schemes are used to integrate (3.11) and show that bilinearity is preserved if a finite dimensional representation of the state vector $\mathbf{X}$ is obtained by truncating the coordinate vector w.r.t. a time independent orthonormal basis.

3.4 Iterative relinearization

In section 3.3 we observed that, for a given initial condition perturbation, there is an optimal linearization trajectory for the TL-model such that the tangent linear predictions become exactly equal to the nonlinear predictions. In this section we introduce an iterative method purely based on integrations with the tangent linear model that converges to this optimal trajectory. Section 3.5 shows how this

iterative method can be used to update the linearization trajectory in forecast sensitivity experiments without using the nonlinear model.

We have seen that for bilinear systems

$$\mathbf{x}(t) = \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t,0)\mathbf{x}_0, \tag{3.19}$$

describes the exact time evolution of perturbations. For a given initial condition perturbation $\mathbf{x}_0$ and a trajectory $\mathbf{X}(t)$ this equation can be written as a map $\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}$ that maps increment trajectories to increment trajectories

$$\mathbf{x} = \mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\mathbf{x}), \tag{3.20}$$

with $\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\mathbf{x}) = \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t,0)\mathbf{x}_0$, i.e. the trajectory $\mathbf{x}(t)$ is a fixed point of $\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}$. If for a fixed time interval $[0,T]$ there is a constant $0 < q < 1$ and a suitable metric d on the space of increments defined on the interval $[0,T]$ such that $d(\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\mathbf{x}), \mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\mathbf{y})) \leq qd(\mathbf{x},\mathbf{y})$ then $\mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}$ is known as a contraction mapping. The Banach fixed point theorem then guarantees that the fixed point $\mathbf{x}$ is unique and more over the iterated map

$$\hat{\mathbf{x}}^k = \mathbf{T}^{\mathbf{X}}_{\mathbf{x}_0}(\hat{\mathbf{x}}^{k-1}), \tag{3.21}$$

converges to this fixed point. This suggests that given an estimate of the trajectory $\hat{\mathbf{x}}^{k-1}(t)$ the tangent linear model can be integrated in the form

$$\hat{\mathbf{x}}^k(t) = \mathbf{M}_{\mathbf{X}+\hat{\mathbf{x}}^{k-1}/2}\mathbf{x}_0, \tag{3.22}$$

where the superscripts indicate the iteration number. With $\hat{\mathbf{x}}^0(t) = 0$, the first iteration $k = 1$ is equal to a standard TL-integration (as given by (3.15)) and gives a trajectory $\hat{\mathbf{x}}^1$. During the second iteration we integrate the tangent linear model with a modified trajectory $\mathbf{X} + \hat{\mathbf{x}}^1/2$ etc. Alternatively the iteration can be started with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ which has the advantage that the time derivatives in the TL-model become exact at $t = 0$. In the experiments both methods are compared.

In appendix 3.D an analysis of (3.22) for multilinear models is given and we show that, independent of the order of the nonlinearities in the nonlinear model, at convergence (3.22) always gives better predictions of the time evolution of perturbations than the standard TL-model (3.15). In particular the bilinear terms are exactly taken into account. In section 3.6.2 we examine the rate of convergence for the iterated map (3.21) for the QG and the L96 model.

Remark 1 (Radius of convergence)**.** *Iterated maps can exhibit a finite radius of convergence even though there is a fixed point valid for all t. Therefore, even though the fixed point trajectory $\mathbf{x}(t)$ is valid for all t, this does not imply that the iterated map* (3.22) *converges to this fixed point. As an example consider the system $\dot{X} = -2tX^2$ with $X(0) = 1$. The solution is given by the Witch of Agnesi $X(t) = \frac{1}{t^2+1}$. The Picard iteration $X^k(t) = X(0) + \int_0^t -2s(X^{k-1}(s))^2 ds$, with*

$X^0(t) = 1$ converges to the Taylor series of $X(t)$ but because $X(t)$ has poles at $t = \pm i$ the Picard iteration only converges to the fixed point for $|t| < 1$.

3.5 Estimation using the inverse TL-model

The estimation problem considered in this paper is: given a forecast starting from an analysis $\mathbf{X}_0$

$$\mathbf{X}_T = \mathcal{M}(T, 0, \mathbf{X}_0), \tag{3.23}$$

and an analysis $\mathbf{X}_T + \mathbf{x}_T$ valid at time T. Can we determine an analysis increment $\mathbf{x}_0$ such that

$$\mathbf{X}_T + \mathbf{x}_T = \mathcal{M}(t, 0, \mathbf{X}_0 + \mathbf{x}_0). \tag{3.24}$$

These types of experiments are known as forecast sensitivity experiments and have been studied by Rabier et al. (1996); Klinker et al. (1998); Pu et al. (1997a,b). If the tangent linear assumption is valid we expect

$$\mathbf{x}_T = \mathbf{M}_{\mathbf{X}}(T, 0)\mathbf{x}_0, \tag{3.25}$$

and therefore we can obtain estimates of $\mathbf{x}_0$ from

$$\hat{\mathbf{x}}_0 = \mathbf{M}_{\mathbf{X}}^{-1}(T, 0)\mathbf{x}_T. \tag{3.26}$$

Besides giving estimates for $\mathbf{x}_0$ the integration with the inverse tangent linear model can be used to produce estimates of the complete trajectory $\mathbf{x}(t)$. The result from section 3.4 therefore suggests to use this method iteratively

$$\hat{\mathbf{x}}^k(t) = \mathbf{M}^{-1}_{\mathbf{X}+\hat{\mathbf{x}}^{k-1}/2}(T, t)\mathbf{x}_T, \tag{3.27}$$

with $\hat{\mathbf{x}}^0(t) = 0$ or $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$. This defines an iterated map on the space of increments trajectories

$$\hat{\mathbf{x}}^k = \mathbf{S}^{\mathbf{X}}_{\mathbf{x}_T}(\hat{\mathbf{x}}^{k-1}), \tag{3.28}$$

where $\mathbf{S}^{\mathbf{X}}_{\mathbf{x}_T}(\hat{\mathbf{x}}^{k-1}) = \mathbf{M}^{-1}_{\mathbf{X}+\hat{\mathbf{x}}^{k-1}/2}\mathbf{x}_T$. In section 3.6.3 the convergence rate of the iterated map (3.28) is investigated for the L96 model.

3.6 Applications

3.6.1 Indices

To highlight different aspects of the optimal linearization trajectories and the iterative relinearization method the exact time evolution of the perturbations

$\mathbf{x}(t)$ and the corresponding tangent linear evolution $\hat{\mathbf{x}}^k(t)$ are compared using five indices l^k, α^k, R^k and d^k and R_d^k. The similarity index $l^k(t)$ is defined as

$$l^k(t) = \frac{\left(\mathbf{x}(t), \hat{\mathbf{x}}^k(t)\right)}{\|\mathbf{x}(t)\| \, \|\hat{\mathbf{x}}^k(t)\|}, \tag{3.29}$$

the angle $\alpha^k(t)$ is given by

$$\alpha^k(t) = \operatorname{acos}\left(l^k(t)\right), \tag{3.30}$$

the relative norm $R^k(t)$ is given by

$$R^k(t) = \frac{\|\hat{\mathbf{x}}^k(t)\|}{\|\mathbf{x}(t)\|}, \tag{3.31}$$

the error norm $d^k(t)$ by

$$d^k(t) = \|\mathbf{x}(t) - \hat{\mathbf{x}}^k(t)\|, \tag{3.32}$$

and the relative error norm $R_d(t)$ by

$$R_d^k(t) = \frac{\|\mathbf{x}(t) - \hat{\mathbf{x}}^k(t)\|}{\|\mathbf{x}(t)\|}. \tag{3.33}$$

For the QG-model the values of d^k, l^k, R^k and R_d^k are determined using the kinetic energy inner product. For the L96 model the Euclidean inner product is used. In the context of twin experiments values of $l = 0.7$, corresponding to an angle $\alpha = 45°$, are commonly used to indicate that the TL assumption is violated (e.g. Gilmour et al. (2001)).

We will say that $\hat{\mathbf{x}}^k(t)$ is more similar to $\mathbf{x}$ than $\hat{\mathbf{x}}^{k-1}(t)$ at time t if $\alpha^k(t) < \alpha^{k-1}(t)$ or, equivalently, if $l^k(t) > l^{k-1}(t)$. We say that $\hat{\mathbf{x}}^k(t)$ is closer to $\mathbf{x}$ than $\hat{\mathbf{x}}^{k-1}(t)$ at time t if $d^k(t) < d^{k-1}(t)$ or, equivalently, if $R_d^k(t) < R_d^{k-1}(t)$.

3.6.2 Iterative relinearization

In this section the rate of convergence of the iterated map $\mathbf{T}$ is examined in a quasi geostrophic model (described in appendix 3.A) and the Lorenz 96 model (described in appendix 3.B).

QG model

In figure 3.1 we show the 2 day forecast difference of the stream function at 500 hPa (upper left). The initial condition for the control run and the perturbed run are 100 days apart and therefore we may assume that they are uncorrelated (see also figure 3.9). The size of the perturbations used in these experiments is therefore much larger than typical analysis increments. In the upper right we show the forecast of the standard TL-model $\hat{\mathbf{x}}^0(t) = 0$.

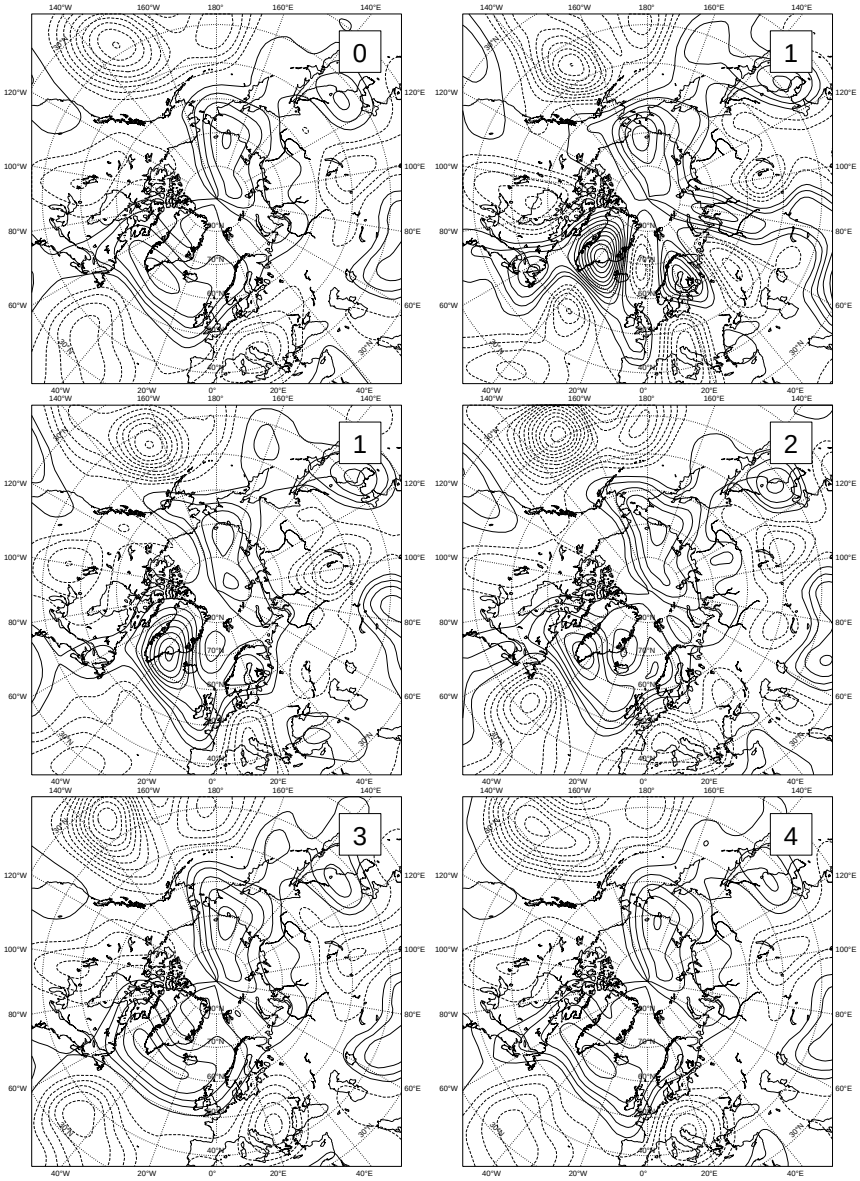

Figure 3.1: Stream function perturbation at 500 hPa after 2 days using the non-linear model $\mathbf{x}$ (top left), the standard tangent linear model with $\hat{\mathbf{x}}^0(t) = 0$ (top right) and the iterative relinearization method with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ for iteration 1 (middle left) to iteration 4 (bottom right). The initial condition for the perturbed run and the control run are 100 days apart. The contour interval is $1 \cdot 10^{-3}\ \Omega a^2$ in all panels (with a and Ω the average radius and the angular velocity of the Earth, respectively). Positive values (solid) negative values (dashed).

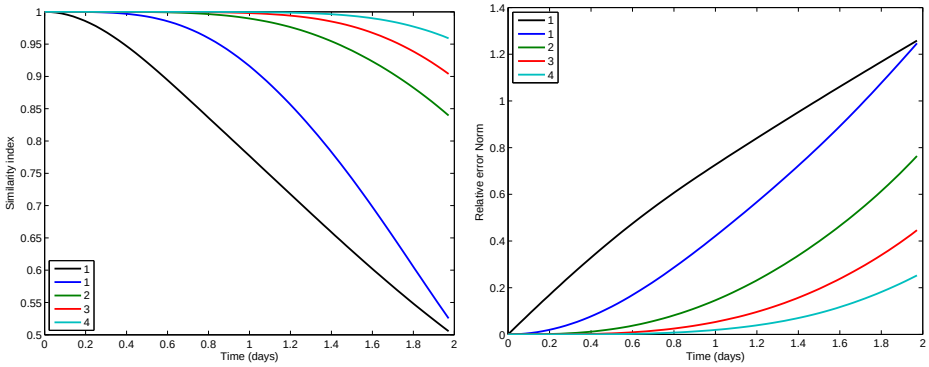

Figure 3.2: Similarity index $l^k(t)$ (left) and relative error norm $R_d^k(t)$ (right) as a function of time and iteration number for the QG model. Shown are average values for 20 experiments. The black line is the standard TL-model with $\hat{\mathbf{x}}^0(t) = 0$. The colored lines are iteration 1 to 4 with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$. The control run and perturbed run are two days apart.

The other panels show the iterative method for 4 iterations with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$. Both the standard TL-integration ($l^1 = 0.55$) and the first iteration with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ ($l^1 = 0.73$) differ substantially from the truth, with large differences north of 60°N. In the first iteration with $\mathbf{x}^0 = 0$ there is a wave pattern over the North Atlantic ocean which is absent in the first iteration with $\mathbf{x}^0 = \mathbf{x}_0$. At subsequent iterations all positive and negative cells are gradually moved to their correct location and with the correct amplitude. At iteration 2 to 4 we have $l^2 = 0.90$, $l^3 = 0.95$ and $l^4 = 0.99$ respectively, indicating that the iterative method converges quickly with the largest improvement when going from iteration 1 to 2.

Figure 3.2 shows the similarity index l^k and the relative error norm R_d^k as a function of time and iteration number. The solid black line refers to the standard TL-model with $\hat{\mathbf{x}}^0(t) = 0$. The colored lines show the iterative relinearized results for 4 iterations with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$. The control run and the perturbed run are 2 days apart. From the standard TL-integration we see that the duration of the TL-regime is slightly larger than 1 day. Especially in the short range it is beneficial to use $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ because the derivatives in the TL-model become exact at $t = 0$. In figure 3.2 (right) this can be seen for example from the relative error norm where $\dot{R}_d(0) \neq 0$ when the standard TL-model is used. Observe that the iterative method adds approximately 0.5 days to the usefulness of the TL-model at each iteration.

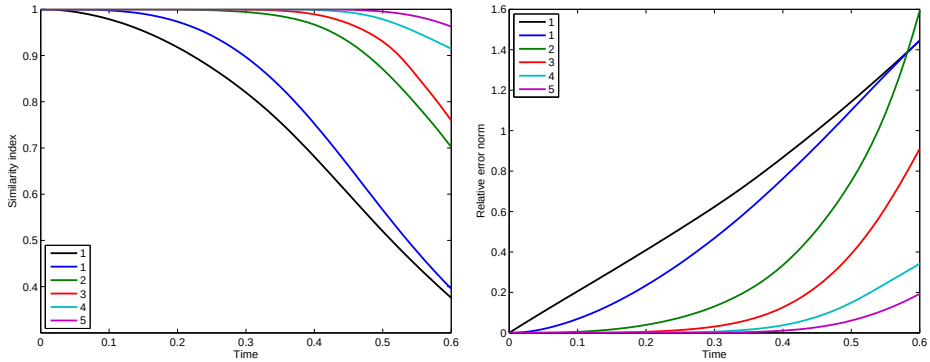

Figure 3.3: Similarity index $l^k(t)$ (left) and relative error norm $R_d^k(t)$ (right) as a function of time for the L96 model. Shown are average results for 50 experiments using random initial condition perturbations with $\|\mathbf{x}_0\| = 10$. The black line is the result for iteration 1 with $\hat{\mathbf{x}}^0(t) = 0$ the colored lines for iteration 1 to 5 with $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$

L96 model

Figure 3.3 shows the similarity index and relative error norm (average over 50 experiments) as a function of time and iteration index for the L96 model. All experiments start with a random initial condition perturbation with $\|\mathbf{x}_0\| = 10$. Such an initial condition amplitude is approximately equal to the size of 12 hour forecast differences (see figure 3.10). From the first iteration using $\hat{\mathbf{x}}^0(t) = 0$ (black) we see that the duration of the tangent linear regime is slightly larger than 1.5 days (0.3 time units). Using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_0$ this can be extended to 2 days. The iterative linearization method converges to the true increment at subsequent iterations. For a 2 day forecast (0.4 time units) of the order of 4 iterations are required to converge to the true time evolution of the increment, with the largest improvements when going from iteration 1 to 2. For longer lead times more iterations are needed. This is related to the fact that the TL-model produces large increments beyond the duration of tangent linear regime (see also figure 3.7). Therefore the corrections $\hat{\mathbf{x}}^1/2$ used in the second iteration are actually deteriorating the linearization trajectory at the end. As a result of this the second iteration is further away from the truth at the end of the optimization window, even though it is more similar to the truth. In section 3.7.2 we discuss a method to regularize this behaviour without affecting the fixed point of the iterated map.

3.6.3 Estimation using the inverse TL-model

Here we examine the iterated map (3.28) from section 3.5 The action of $\mathbf{M}_{\mathbf{X}}^{-1}$ on a vector $\mathbf{x}_T$ is obtained by integrating the tangent linear model backwards in time. In the L96 model the fourth order Runge Kutta (RK4) scheme is used to propagate the state. Theoretically the backwards integration requires the use of the inverse integration scheme (which will be an implicit scheme) to ensure $\mathbf{M}_{\mathbf{X}}\mathbf{M}_{\mathbf{X}}^{-1} = \mathbf{I}$. Here the adjoint of the RK4 scheme is used to integrate the TL-model backwards in time. In the L96 model we find experimentally that the angle between $\mathbf{M}_{\mathbf{X}}^{-1}\mathbf{M}_{\mathbf{X}}\mathbf{x}_0$ and $\mathbf{x}_0$ is of the order $O(10^{-3})$ degrees and the relative norm $\|\mathbf{M}_{\mathbf{X}}^{-1}\mathbf{M}_{\mathbf{X}}\mathbf{x}_0\|/\|\mathbf{x}_0\| - 1 \approx O(10^{-6})$ for an optimization time of 0.6 time units (3 days). So it appears that $\mathbf{M}_{\mathbf{X}}^{-1}\mathbf{M}_{\mathbf{X}}$ is close to the identity operator. We conclude that the adjoint RK4 scheme can be used for the inverse integrations.

Figure 3.4 shows the result when we iteratively solve (3.27) using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$. Even though the estimate from the first iteration differs substantialy from the truth with $l_1 = 0.4$, the method quickly converges and the subsequent iterations are more similar and closer to the truth. Approximately 4 iterations are required to obtain an almost perfect estimate. Note that during the inverse integration we also obtain the corrections needed for the next iterations. Therefore the computational cost is equal to four TL-integrations (backwards). This cost should be compared to the alternative of solving this estimation problem in terms of a cost function minimization (e.g. 4D-VAR) where a single inner loop iteration already involves two linear integrations (1 adjoint and 1 tangent linear integration). For comparison figure 3.4 also includes the result when the standard TL-model, i.e. $\hat{\mathbf{x}}^0(t) = 0$, is used to propagate the increment backwards in time (the black line). If $l = 0.7$ is used as threshold value then the gain of using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$ in the first iteration is 0.13 time units (0.65 days). From the time evolution of the error norm (figure 3.4 right) we see that this gain is mainly a result of the fact that, by using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$, the time derivatives at $t = 0.4$ become exact in the TL-model and thus $\dot{d}_1 = 0$ at $t = 0.4$. In particular for large perturbations we therefore expect to benefit from using $\hat{\mathbf{x}}^0(t) = \mathbf{x}_T$.

From figure 3.4 it is also clear that for long optimization windows the estimated increment at $t = 0$ from the first iteration becomes uncorrelated with the true increment. As a result, the nonlinear forecast starting from $\mathbf{X}_0 + \hat{\mathbf{x}}_0^1$ bears low similarity to the truth (dashed lines in figure 3.4). For long windows therefore, the nonlinear model starting from $\mathbf{X}_0 + \hat{\mathbf{x}}_0^1$ can not be used to update the linearization trajectory. In a forthcoming paper applying optimal linearization trajectories in the context of 4D-VAR we will show that also in 4D-VAR it is better to update the linearization trajectory using the TL-model.

If we use $l < 0.7$ to indicate the breakdown of the TL-assumption, figure 3.4 indicates that the tangent linear assumption linearized around the control run is valid for 0.15 time units (i.e. from $t = 0.4$ to $t = 0.25$). This should be compared with the forward integration in figure 3.3 where the value 0.7 is reached after 0.3 time units. The duration of the TL-regime is shorter for inverse integrations.

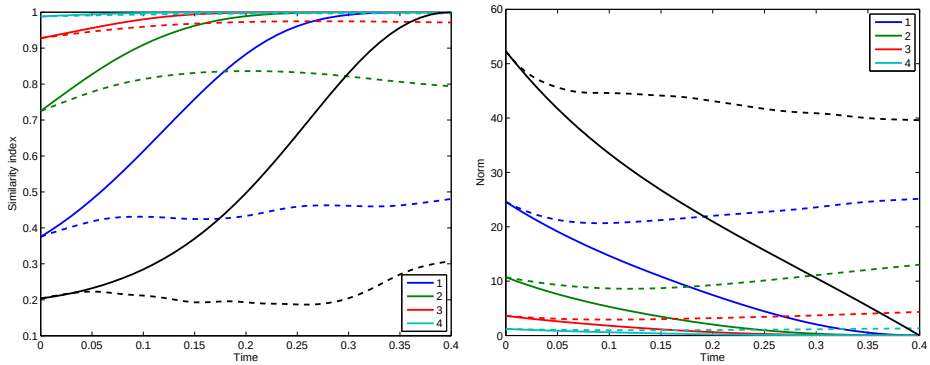

Figure 3.4: Similarity index (left) and error norm $d_k = \|\mathbf{x} - \hat{\mathbf{x}}^k\|$ (right) as a function of time for the inverse TL-model (solid) and the NL-model starting from $\mathbf{X}_0 + \hat{\mathbf{x}}_0^k$ (dashed). Shown are average results over 50 experiments with an OT=0.4 and random initial condition perturbations with norm $\|\mathbf{x}_0\| = 10$. The black line is the first iteration with $\mathbf{x}^0(t) = 0$, i.e the standard TL-model

Partly this is a result of the fact that error growth in the backward integration is characterized by the reciprocal singular value spectrum and these values are larger than the singular values (see figure 3.6). Another reason is that typically $\|\mathbf{x}_T\| > \|\mathbf{x}_0\|$ and therefore the backward integration is started with larger initial conditions. The idea of using the inverse of the TL-model has been studied in Pu et al. (1997a) using a method called the quasi inverse. They reversed the sign of the dissipation terms in the TL-model as a form of regularization. As will be discussed in section 3.7.2 on the regularized prediction experiments, there is no need for bilinear systems to add regularization when the optimal linearization trajectory is used. Therefore the amount of regularization should depend on how close we are to the optimal linearization trajectory. If the linear term b in the nonlinear model (3.11)) is a purely dissipative term, i.e. $(\mathbf{X}, b(\mathbf{X})) < 0$, then the TL-model can be integrated in the form

$$\dot{\hat{\mathbf{x}}}^k = \mathbf{J}(\mathbf{X} + \hat{\mathbf{x}}^{k-1}/2)\hat{\mathbf{x}}^k + \alpha b(\hat{\mathbf{x}}^{k-1} - \hat{\mathbf{x}}^k) \tag{3.34}$$

The choice $\alpha = 2$ amounts to reversing the sign of the dissipation terms (see also equation (3.13)) during the first iteration. However at subsequent iterations, at locations in space and time where the solution has converged, the unmodified TL is used.

3.6.4 Identification of bilinear systems

For bilinear systems the time evolution of the increment $\hat{\mathbf{x}}(t)$ in the TL-model linearized around the trajectory $\mathbf{X} + \mathbf{x}/2$ given by

$$\hat{\mathbf{x}}(t) = \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t, 0)\mathbf{x}_0, \tag{3.35}$$

is equal to the time evolution according to the nonlinear model: $\hat{\mathbf{x}}(t) - \mathbf{x}(t) = 0$. Therefore a necessary condition for the model $\mathcal{M}$ to be a bilinear system is that the error norm (or equivalent the relative error norm) is zero:

$$d(t) = R_d(t) = 0. \tag{3.36}$$

However, numerical integrations will be subject to round off error leading to nonzero values for d and R_d. To highlight different aspects, the time evolution of perturbations is examined in terms of the angle α (3.30) and the relative norm R (3.31). Note that $\alpha = 0$ and $R = 1$ if and only if $R_d = d = 0$. In the following sections we study the behaviour of R_d, α and R in the QG and L96 model.

QG model

Figure 3.5 shows the relative error norm R_d, the relative norm R and angle α as a function of time for the QG model for 10 experiments. The control run is obtained by integrating the nonlinear model for 300 days. Continuing the integration for another 300 days yields the perturbed run. The trajectory for the second experiment starts using the final condition of the previous perturbed run and so forth. Due to the long integration times the initial condition for the tangent linear model is given by the difference between two uncorrelated state vectors on the model attractor and is therefore larger than typical analysis increments. For these large amplitude perturbations the TL approximation is valid for 1 day.

The 10 experiments show exponential growth of the relative norm R after day 210. Before day 210 both $\alpha \approx 0$ and $R \approx 1$ and we conclude that the TL-model can be used for lead times shorter than 210 days. The time evolution of the relative error norm R_d (figure 3.5, left) shows no signal at day 210. Instead it merely indicates exponential growth beyond day 10 with an exponent of 0.148 day^{-1} (standard deviation 0.005 in 10 experiments) corresponding to an error doubling time of $\tau_d = 4.7$ days. Note that this is longer than the error doubling time based on linearization of the TL-model around a control run, which gives a Lyapunov exponent of $\sigma = 0.254$ (with standard deviation 0.014 in 10 experiments) and a corresponding error doubling 2.7 days. This is in agreement with other studies, e.g. Swanson et al. (1998), where an approximate value of 3 days is given. The increase of the error doubling time when we linearize around the average trajectory of the control and perturbed run is consistent with Hoskins et al. (2000) who determined singular vector growth using different linearization

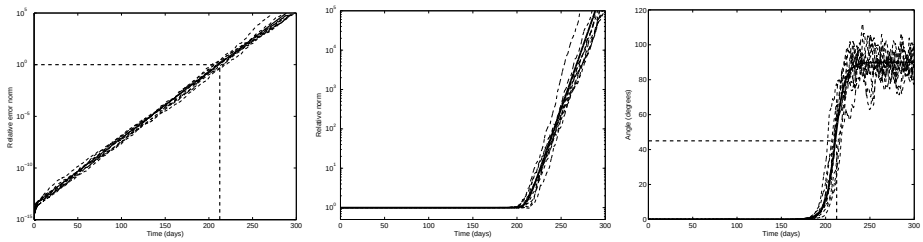

Figure 3.5: Relative error norm R_d (left), relative norm R (middle) and angle α (right) for the QG model with individual experiments dashed. The solid line in the left panel is the estimate R_d using (3.37) with $R_d(0) = 2.9 \cdot 10^{-14}$ and $\sigma = 0.148$. In the middle panel and right panel the solid lines are estimates using the assumption that the error vector is perpendicular to $\mathbf{x}$.

trajectories in the TL-model. They found that the dominant factor for singular vector growth is the dynamic structure of the linearization trajectory and, in particular, its smoothness.

From figure 3.5 we see that the time evolution of the relative error norm R_d is approximately exponential beyond day 10. This suggests that we can model the time evolution of R_d for $t > 10$ days by

$$\dot{R}_d = \sigma R_d. \tag{3.37}$$

The values of σ and $R_d(0)$ are estimated using linear least squares on the experimental values of $\ln(R_d(t))$. The solid line in figure 3.5 (left) show the predictions of this model with the estimated values $R_d(0) = 2.9 \cdot 10^{-14}$ (standard deviation $2.2 \cdot 10^{-14}$ in 10 experiments) and $\sigma = 0.148$ (standard deviation 0.005 in 10 experiments). With the additional assumption that the error vector $\hat{\mathbf{x}} - \mathbf{x}$ is perpendicular to $\mathbf{x}$ the modelled time evolution of R_d can be used to predict values of the angle α and the relative norm R (solid lines in the middle and right panel of figure 3.5). We emphasize that the solid lines in the middle and right panel of figure 3.5 are not fitted to the experimental data but purely a result of the geometric assumption that the error vector $\hat{\mathbf{x}} - \mathbf{x}$ is perpendicular to $\mathbf{x}$. Experimentally we find that the angle between $\mathbf{x}$ and $\mathbf{x} - \hat{\mathbf{x}}$ is 89.6° with a standard deviation of 8.1°.

With the assumption that the error vector $\hat{\mathbf{x}} - \mathbf{x}$ is perpendicular to $\mathbf{x}$ the condition $\alpha = 45°$ is equivalent to the condition $R_d = 1$. Setting $R_d = 1$ in the error growth model gives the estimate

$$t_{\mathrm{p}} = \frac{-1}{\sigma} \ln R_d(0) = 212 \pm 3.8 \text{ days}. \tag{3.38}$$

This estimate is plotted in the left panel of figure 3.5. The same estimate is obtained from $\alpha = 45$ and $R = \sqrt{2}$. Note that in the absence of round-off error

$\mathbf{x} = \hat{\mathbf{x}}$ and as such there is no reason to prefer the NL over the TL integration. Therefore these results also put a predictability limit on the nonlinear model due to round-off error of 212 days.

L96 model

In the L96 model we obtain the estimates $\hat{R}_d(0) = 4.0 \cdot 10^{-16}$ and $\sigma = 0.233\ \mathrm{day}^{-1}$ equivalent to an error doubling time of 2.97 days. The error doubling times are higher than estimates based on the Lyapunov exponent (2.1 days) consistent with the reduced growth of singular vectors for smooth trajectories in Hoskins et al. (2000). The figures for R_d, α and R are similar to the results for the QG-model (not shown) For the L96 model the TL-model can be used for $t_\mathrm{p} = \frac{-1}{\sigma} \ln R_d(0) =$ 152 days.

3.7 Discussion

3.7.1 Prospects for using the method in NWP

We demonstrated the advantage of using the optimal linearization trajectories in the context of two simple bilinear models. Although the analysis in appendix 3.D shows that, independent of the order of the nonlinearities in the nonlinear model, the iteratively relinearized TL-model always gives better results at convergence, to get an exact correspondence between the TL and the NL model, the NL-model has to be bilinear. In example 2 it was shown that it is possible to transform multilinear systems to bilinear systems by augmenting the state vector.

There are other situations where apparent "infinite" order nonlinearities can be transformed to bilinear terms. Let $\dot{X} = e^{\alpha X}$ and define $Y = e^{\alpha X}$ then $\dot{X} = Y$ and $\dot{Y} = \alpha Y^2$ which is a bilinear system. One difference between the reduction of multilinear systems (appendix 3.D) to bilinear system in example 2 is that in this case the newly introduced variable Y has to be a prognostic variable because the algebraic constraint $0\dot{Y} = Y - e^X$ is not bilinear and can therefore not be used. Similary it can be shown that $\dot{X} = \sin(X)$ (Define $Y = \sin(X)$ and $Z = \cos(X)$), $\dot{X} = \ln X$ (define $Y = \ln X$) and $\dot{X} = X^\alpha$ (define $Y = X^{\alpha-1}$) can be written as bilinear systems. Although this does not show that realistic NWP models can be formulated as bilinear systems, it illustrates that both multilinear models and models that contain "infinite" order nonlinearites can be written as a bilinear system and demonstrates that the class of bilinear systems is very general. In a forthcoming paper we show that the restriction to bilinear systems can be lifted if the tangent linear model is linearized around an ensemble of trajectories simultaneously.

3.7.2 Regularized relinearization in the L96 model

The TL-model produces large increments for long lead times (see figure 3.7). This will deteriorate the linearization trajectory for the next iterations. In principle, this can be solved by increasing the dissipation in the TL-model, however, in that case the solution would no longer converge to the true solution during the iterative process. Here we propose to add a term $\alpha(\hat{\mathbf{x}}^{k-1} - \hat{\mathbf{x}}^k)$ to the TL-model leading to

$$\dot{\hat{\mathbf{x}}}^k = \mathbf{J}(\mathbf{X} + \hat{\mathbf{x}}^{k-1}/2)\hat{\mathbf{x}}^k + \alpha(\hat{\mathbf{x}}^{k-1} - \hat{\mathbf{x}}^k) \tag{3.39}$$

So dissipation is added to the model but at the same time the previous iteration is used as a forcing in the TL-model. At convergence of the algorithm $\hat{\mathbf{x}}^{k-1} = \hat{\mathbf{x}}^k$ and the added term becomes zero, i.e. the added term does not modify the fixed point of the iterated map $\mathbf{T}$ (3.21). In general α could be an operator (see also section 3.6.3), here we only discuss the situation where α is a scalar.

Using $\hat{\mathbf{x}}^0 = 0$ the first iteration is given by

$$\hat{\mathbf{x}}^1(t) = \mathbf{M}_{\mathbf{X}}(t,0)e^{-\int_0^t \alpha dt'}\mathbf{x}_0 \tag{3.40}$$

Where $\mathbf{M}_{\mathbf{X}}(t,0)$ is the propagator for the TL-model with $\alpha = 0$. If $\mathbf{M}_{\mathbf{X}} = \mathbf{UDV}^T$ is the singular value decomposition of $\mathbf{M}_{\mathbf{X}}$ we obtain

$$\hat{\mathbf{x}}^1(t) = \mathbf{U}(\mathbf{D}e^{-\int_0^t \alpha dt'})\mathbf{V}^T\mathbf{x}_0 \tag{3.41}$$

So the added term has no impact on the singular vectors but it changes the singular value spectrum. Let $\sigma_{\max}(t)$ denote the leading singular value of $\mathbf{M}_{\mathbf{X}}(t,0)$. By choosing α such that

$$\int_0^t \alpha dt' > \ln \sigma_{\max}(t) \tag{3.42}$$

we conclude that $||\hat{\mathbf{x}}^1(t)|| < ||\mathbf{x}_0||$ for all $\mathbf{x}_0$. In figure 3.6 we show the leading singular value as a function of the optimization time and the value for α when α is kept constant during the optimization window $\alpha = \frac{1}{t}\log\sigma_{\max}(t)$.

Figure 3.7 shows the impact of the added term by examining the norm $\|\hat{\mathbf{x}}^k\|$ as a function of time for $\alpha = 0$ and $\alpha = 8$. The iterative method still convergences to the true solution but in a more controlled manner. At the first iteration the norm decreases monotonically as expected. At subsequent iterations the forcing ensures that we still converge to the true solution.

In NWP models we know that at each grid point in the integration domain the density ρ, absolute temperature T, pressure p and the specific humidity q are all positive quantities. Tangent linear integrations do not respect these types of constraints and therefore it is possible that in the linearization trajectory $\mathbf{X}+\mathbf{x}^{k-1}/2$ some of these variables are negative. We therefore suggest to use a projection operator $\mathbf{P}$ that sets negative values of ρ, T, p and q to zero and integrate the tangent linear model in the form

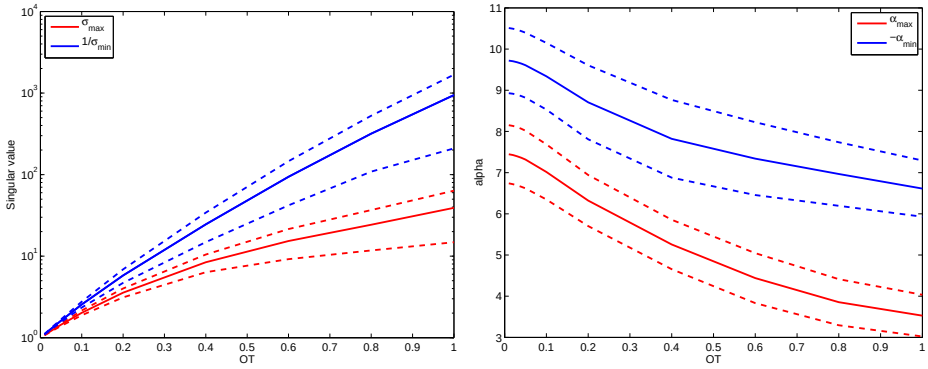

Figure 3.6: Average leading singular value $\sigma_{\max}$ and reciprocal of the trailing singular value $1/\sigma_{\min}$ (left) and corresponding mean values of α for the regularized prediction $\alpha_{\max}$ and regularized estimation $\alpha_{\min}$ values as a function of optimization time (right) for the Lorenz model. In both plots the range indicates the standard deviation in 50 experiments.

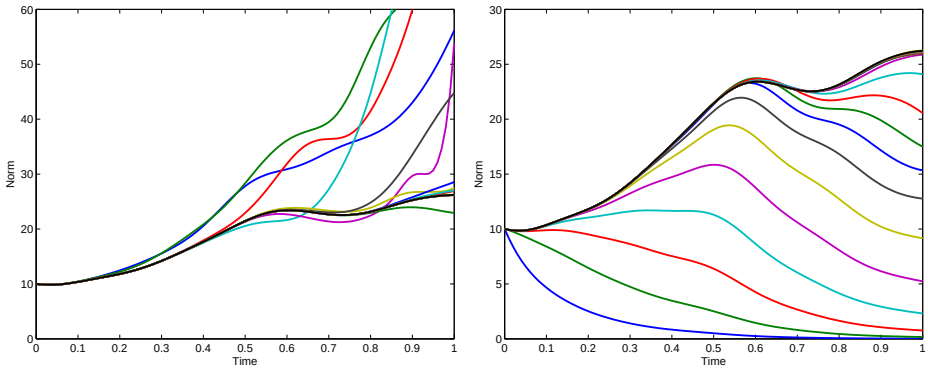

Figure 3.7: The norm $\|\hat{\mathbf{x}}^k\|$ as a function of time without regularization (left) and with regularization (right) using $\alpha = 8$. The black line is $\|\mathbf{x}\|$. The colored lines are the values for $\|\hat{x}^k\|$. Initial condition perturbation is random with norm $\|\mathbf{x}_0\| = 10$

$$\hat{\mathbf{x}}^k(t) = \mathbf{M}_{\mathbf{P}(\mathbf{X}+\mathbf{x}^{k-1}/2)}\mathbf{x}_0 \tag{3.43}$$

Being solutions of the nonlinear model the trajectories $\mathbf{X}$ and $\mathbf{X}+\mathbf{x}$ do not contain negative values for ρ, T and q. At convergence of the iterated map the linearization trajectory $\mathbf{X}(t)+\mathbf{x}(t)/2$ is the average of $\mathbf{X}(t)+\mathbf{x}(t)$ and $\mathbf{X}(t)$ and therefore the linearization trajectory does not contain negative values for ρ, T, p and q, i.e. the projection operater does not modify the fixed point of the iterated map but ensures that during the iterations only "physically consistent" trajectories are used.

3.7.3 Identification of multilinear system

In section 3.6.4 we introduced a necessary condition (3.36) for a nonlinear model to have at most bilinear terms. Here we illustrate that this condition can be used to detect higher order multilinearities.

Consider the L96 model with modified dissipation

$$\dot{\mathbf{X}} = q(\mathbf{X},\mathbf{X}) - \mathbf{X} + \mathbf{F} + \alpha(1 - \frac{\|\mathbf{X}\|^2}{\|\mathbf{F}\|^2})\mathbf{X} \tag{3.44}$$

Where $q(\mathbf{X},\mathbf{X})$ is given in appendix 3.B, $\mathbf{F} = 8$ and $\alpha \geq 0$. For $\alpha = 0$ we recover the L96 model and dissipation is linear. For $\alpha = 1$ the dissipation is a purely trilinear term and dependent on the total energy in the system. The factor $\|\mathbf{F}\|^2$ is introduced to ensure that the (unstable) steady state solution $\mathbf{X}^* = \mathbf{F}$ for the case $\alpha = 0$ is also a (unstable) steady state for $\alpha \neq 0$. For $\alpha \neq 0$ the additional steady state solutions are $\mathbf{X}^* = -\frac{\mathbf{F}}{2}\left(1 \pm \sqrt{1-4/\alpha}\right)$. For $0 < \alpha < 4$ the last expression gives two complex conjugate steady state solutions which can not be reached if we start with a real valued initial condition. The time derivative of the total energy is

$$\begin{aligned}\dot{E} &= -\|\mathbf{X}\|^2 + (\mathbf{X},\mathbf{F}) + \alpha(1 - \frac{\|\mathbf{X}\|^2}{\|\mathbf{F}\|^2})\|\mathbf{X}\|^2 \\ &\leq -\|\mathbf{X}\|^2 + \|\mathbf{X}\|\|\mathbf{F}\| + \alpha(1 - \frac{\|\mathbf{X}\|^2}{\|\mathbf{F}\|^2})\|\mathbf{X}\|^2\end{aligned}$$

For points outside the sphere with radius $\|\mathbf{F}\|$ we therefore have $\dot{E} < 0$ and we conclude that all trajectories eventually enter this ball and cannot escape afterwards.

We expect that for nonzero values of α we have $\|\mathbf{x}(t) - \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t,0)\mathbf{x}_0\| > 0$ and this is indeed what we observe (see figure 3.8). This shows that nonbilinearity can be identified purely based on the model output and might be useful in realistic NWP models where analysing the code to determine nonbilinearity might be prohibitive.

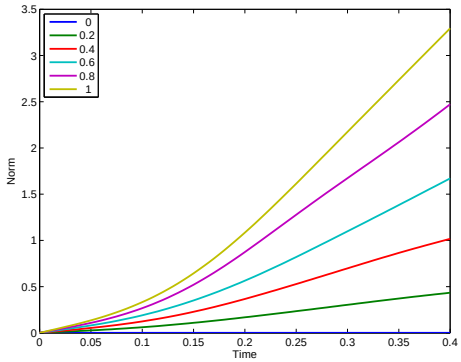

Figure 3.8: Error norm $\|\mathbf{x}(t) - \mathbf{M}_{\mathbf{X}+\mathbf{x}/2}(t,0)\mathbf{x}_0\|$ as a function of time for the trilinear Lorenz model for $\alpha \in \{0, 0.2, 0.4, 0.6, 0.8, 1\}$. Shown are average results for 50 experiments

3.8 Conclusions

The nonlinearities in fluid dynamics as a result of the advection part of the total derivative and the use of algebraic constraints such as the ideal gas law give rise to bilinear differential equations. We have shown that for bilinear systems there exists an optimal linearization trajectory for the tangent linear model, such that the TL-model predicts the exact time evolution of the perturbations. Using a quasi geostrophic model and the Lorenz 96 model we showed that, when the optimal linearization trajectory is used, the tangent linear model can be used for more than 200 days in a quasi geostropic model and more than 150 days in the Lorenz 96 model. For bilinear systems therefore one of the major limitations to the application of linear models mentioned in the introduction can be eliminated by linearizing around the optimal linearization trajectory.

We introduced an iterative method that, based purely on TL-integrations, converges to this optimal linearization trajectory. We showed that the optimal linearization trajectory is a fixed point of this iterative method and using prediction experiments in the QG and L96 model we showed that the iterative method converges to the fixed point. In the discussion we introduced a method to regularize the error growth in the TL-model without affecting the fixed point of the iteration.

The main conclusion from this paper is that this iterative method can be used in estimation problems to account for nonlinearity without using the nonlinear model. In particular when long windows are used in forecast sensitivity experiments the estimated increment at $t = 0$ will be uncorrelated with the true increment and the nonlinear model can not be used to update the linearization trajectory. Using forecast sensitivity experiments in the Lorenz 96 model where

we iteratively use the inverse of the tangent linear model we showed that the iterative method can be used for long windows and converges quickly. Typically 4 iterations (computation cost equal to 4 integrations with the linear model) are needed to find the optimal corrections for a two day forecast. In a forthcoming paper we will show that the same ideas can be used in multi incremental 4D-VAR.

Acknowledgements

We would like to thank Wim Verkley, Theo Opsteegh and two anonymous reviewers for carefully reading earlier versions of the manuscript.

3.A Quasi-geostrophic model

Marshall and Molteni (1993) introduced a spectral three-level quasi-geostrophic (QG) model with global domain and pressure as the vertical coordinate. The model is truncated at wave number 21 and the model levels are at 200 (level 1), 500 (level 2) and 800 hPa (level 3). The model integrates the system

$$\begin{aligned}
\frac{\partial q_1}{\partial t} &= -J(\psi_1, q_1) - D_1(\psi_1, \psi_2) + S_1 \\
\frac{\partial q_2}{\partial t} &= -J(\psi_2, q_2) - D_2(\psi_1, \psi_2, \psi_3) + S_2 \\
\frac{\partial q_3}{\partial t} &= -J(\psi_3, q_3) - D_3(\psi_2, \psi_3) + S_3,
\end{aligned}$$

where q_i is potential vorticity (PV), ψ_i streamfunction, D_i are linear operators that represent dissipative terms, S_i are constant PV sources and J the Jacobian of a two dimensional field. We refer to Marshall and Molteni (1993) for a complete description of the model.

Figure 3.9 shows the norm $\|X(t) - X(t - \delta t)\|$ as a function of δt averaged over 1 year for the quasi geostrophic model. In Bengtsson et al. (2008, their figure 3) a similar picture is shown for the RMSE of the geopotential height at 500 hPa for the ECMWF model but based on analyses instead of forecasts. If the trend due to seasonal variability is removed in the ECMWF model the RMSE reaches a maximum of 110.8m and the RMSE of analyses one day apart is 61m, i.e. at 1 day the error is already half of the value reached for large δt. The QG model the L96 model (figure 3.10) and the ECMWF model therefore show similar behaviour in this respect. Both in the QG model and the L96 model the growth of the error norm saturates at $\delta t = 10$ days.

3.B Lorenz 96 model

Lorenz (1996) introduced a simple system of the form (3.50) with $q^i_{i-1,i+1} = 1$,

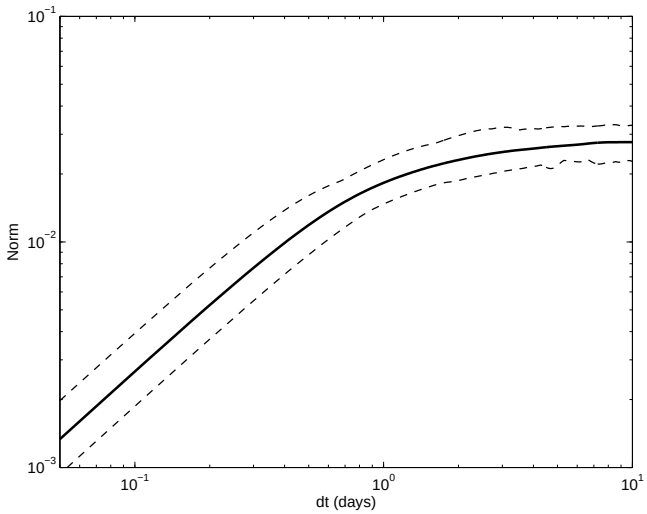

Figure 3.9: The norm $\|\mathbf{X}(t) - \mathbf{X}(t - \delta t)\|$ as a function of δt averaged over 1 year for the quasi geostrophic model. The solid line is the average the dashed lines are the maximum and minimum value that occurred during the 1 year period.

$q^i_{i-2,i-1} = -1$, all other $q^i_{jk} = 0$, $b^i_j = -\delta^i_j$ the Kronecker delta and $c^i = F$. Giving the system

$$\dot{X}^i = X^{i-1}(X^{i+1} - X^{i-2}) - X^i + F, \tag{3.45}$$

where the dimension of the state vector is N and the cyclic convention $X^{i+N} = X^i$ is used. We will use the vector notation

$$\dot{\mathbf{X}} = q(\mathbf{X}, \mathbf{X}) - \mathbf{X} + \mathbf{F}. \tag{3.46}$$

The nonlinear term conserves the total energy $\frac{1}{2}\|\mathbf{X}\|^2$, i.e. $(\mathbf{X}, q(\mathbf{X}, \mathbf{X})) = 0$. The linear term $-\mathbf{X}$, representing mechanical or thermal dissipation, decreases the total energy $-(\mathbf{X}, \mathbf{X}) < 0$ while the constant term $\mathbf{F}$, representing external forcing prevents the total energy from decaying to zero. We imagine that $\mathbf{X}$ represents some atmospheric variable around a latitude circle and X^i is the value at longitude $360i/N$. In all simulations we use $N = 40$ and $\mathbf{F} = 8$. If 1 time unit in the model is identified with 5 days the error doubling time of the model is 2.1 days (Lorenz and Emanuel, 1998).

Figure 3.10 shows the norm $\|\mathbf{X}(t) - \mathbf{X}(t - \delta t)\|$ as a function of δt average over 1 year. This should be compared with figure 3.9 for the QG-model. The forecast error norm saturates after day 10 in both models. The straight line in figure 3.10 is the estimated bound $\sqrt{2}\|\frac{\mathbf{F}}{2}\|$ which can be derived as follows.

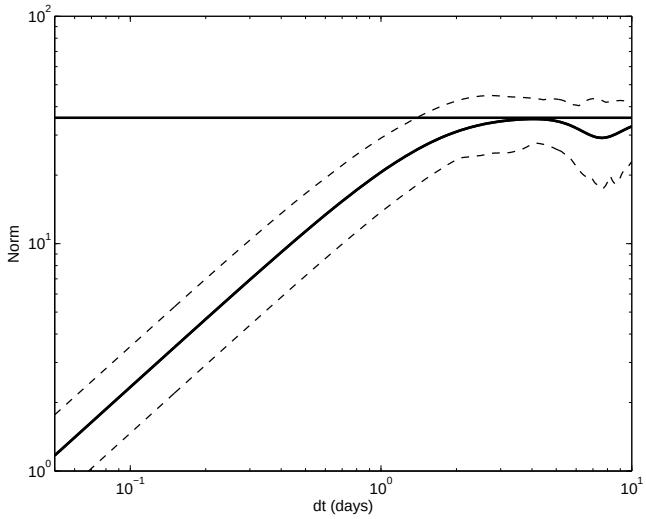

Figure 3.10: The norm $\|\mathbf{X}(t) - \mathbf{X}(t - \delta t)\|$ as a function of δt averaged over 1 year on a loglog scale for the L96 model. The solid line is the average and the dashed lines are the maximum and minimum value of the norm that occurred during the 1 year period. The straight line is the estimated value $\sqrt{2}\|\frac{\mathbf{F}}{2}\|$ based on the sphere $\mathcal{C}$ given by (3.48). For ease of comparison with figure 3.9 the time axis is scaled such that 1 time unit is 5 days.

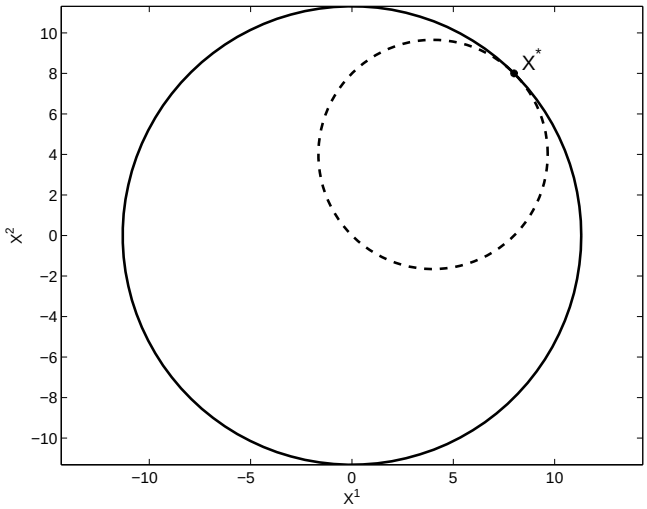

Figure 3.11: The boundary of ball $\mathcal{B}$ (solid) and the sphere $\mathcal{C}$ (dashed) and the steady state solution $\mathbf{X}^*$ for $N = 2$ and $F = 8$. Trajectories that start in $\mathcal{B}$ can not cross the boundary of $\mathcal{B}$. This figure is not equivalent to a cross section through the X^1 X^2 plane of the system with $N = 40$ because the center of $\mathcal{C}$ will not be contained in this cross section. In particular the sphere $\mathcal{C}$ will appear much smaller in such cross sections

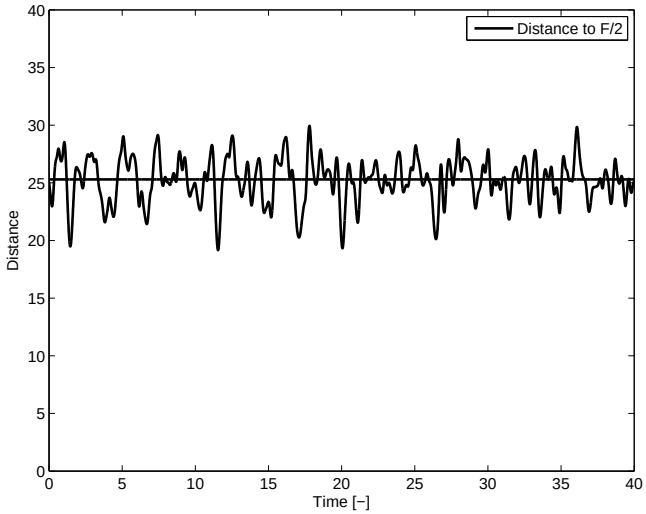

Figure 3.12: Distance to the center of the sphere $\mathcal{C}$ as a function of time. The straight line is the radius of $\mathcal{C}$.

The time evolution of the total energy $\frac{1}{2}\|\mathbf{X}\|^2$ is given by

$$\begin{aligned}\frac{d\frac{1}{2}\|\mathbf{X}\|^2}{dt} &= -\|\mathbf{X}\|^2 + (\mathbf{X},\mathbf{F}) \\ &\leq (-\|\mathbf{X}\| + \|\mathbf{F}\|)\,\|\mathbf{X}\|,\end{aligned}$$

where we used the Cauchy-Schwarz inequality. If we define the closed ball

$$\mathcal{B} = \{\mathbf{X} \mid \|\mathbf{X}\| \leq \|\mathbf{F}\|\}, \tag{3.47}$$

then for all $\mathbf{X} \notin \mathcal{B}$ we have $\frac{d\frac{1}{2}\|\mathbf{X}\|^2}{dt} < 0$. For all $\mathbf{X}$ on the boundary of $\mathcal{B}$ we have $\frac{d\frac{1}{2}\|\mathbf{X}\|^2}{dt} \leq 0$. So all trajectories that start in the interior of $\mathcal{B}$ at $t = 0$ remain in this interior for $t > 0$. Note that the steady state solution $\mathbf{X}^* = \mathbf{F}$ is on the sphere.

The time derivative of the energy can also be written as

$$\frac{d\frac{1}{2}\|\mathbf{X}\|^2}{dt} = -\|\mathbf{X} - \frac{\mathbf{F}}{2}\|^2 + \|\frac{\mathbf{F}}{2}\|^2. \tag{3.48}$$

Therefore there is a sphere $\mathcal{C}$ with radius $R = \|\frac{\mathbf{F}}{2}\|$ and center $\mathbf{C} = \frac{\mathbf{F}}{2}$ on which the time derivative of the total energy is zero. Again note that the steady solution $\mathbf{X}^*$ is on this sphere (see figure 3.11). Trajectories that start in the interior of $\mathcal{B}$

stay in the interior for $t > 0$ and therefore the energy of the state is bounded as $T \to \infty$. This is only possible if either the state asymptotically approaches $\mathcal{C}$ or by crossing the surface of the sphere indefinitely. In either case this implies that the dynamics of the system takes place "near" the surface of the sphere $\mathcal{C}$. This is indeed what we observe, see figure 3.12.

Assume now that the state vectors for large δt are uncorrelated and on the sphere $\mathcal{C}$. By symmetry considerations the expected value for the angle between two vectors associated with two random points on a $N-1$ dimensional sphere is $\pi/2$ (see Borel E. (1914) where it is shown that for large N the probability density function tends to a normal distribution with mean $\pi/2$ and standard deviation $1/\sqrt{N}$) and therefore the expected distance between two random points on the sphere $\mathcal{C}$ is $\sqrt{2}R$. This estimate is shown in figure 3.10. Given the simplicity of the arguments that were used in the derivation this is remarkably good estimate of the asymptotic behaviour of the forecast error norm.

Before each experiment we started from a random point on the sphere $\mathcal{C}$ and integrate for 100 days (20 time units) to allow the system to reach the attractor. All integrations were performed using a Runge Kutta order 4 scheme with a time step of 0.01.

3.C Bilinearity preserving finite dimensional representations and time discretizations

If $\mathbf{e}_i$ is a complete time independent orthonormal basis of the phase space $\mathbb{H}$ w.r.t. an inner product $\langle \cdot, \cdot \rangle$ we can write $\mathbf{X}(t) = \sum_i X^i(t)\mathbf{e}_i$. Using the bilinearity of q and linearity of b, (3.11) can be written as

$$\dot{X}^j \mathbf{e}_j = X^j X^k q(\mathbf{e}_j, \mathbf{e}_k) + X^j b(\mathbf{e}_j) + c, \tag{3.49}$$

where we use the convention that there is an implied summation over a repeated upper and lower index in a single term. Taking the inner product of this equation with $\mathbf{e}_i$ gives the time evolution of the coordinates $X^i(t)$

$$\dot{X}^i = q^i_{jk} X^j X^k + b^i_j X^j + c^i, \tag{3.50}$$

where $q^i_{jk} = \langle \mathbf{e}_i, q(\mathbf{e}_j, \mathbf{e}_k) \rangle$, $b^i_j = \langle \mathbf{e}_i, b(\mathbf{e}_j) \rangle$ and $c^i = \langle \mathbf{e}_i, c \rangle$. We see that if the coordinate vector is truncated at a certain index N the truncated system is bilinear (e.g. if $\mathbf{e}_i$ is a spherical harmonic basis). Therefore the time evolution of the coordinates X^i w.r.t. a time independent truncated orthonormal basis is given by a bilinear differential equation and the optimal linearization trajectory can be obtained by adding the coordinates.

3.C.1 Integration schemes

The Euler forward scheme propagates the state vector as

$$\mathbf{X}_{k+1} = \mathbf{X}_k + hf(t_k, \mathbf{X}_k), \tag{3.51}$$

where h is the time step. If $f(t_k, \mathbf{X}_k) = q(\mathbf{X}_k, \mathbf{X}_k) + b(\mathbf{X}_k) + c$ then the highest order nonlinear term in the map from $\mathbf{X}_k$ to X_{k+1} is bilinear and therefore the time discretization by the integration scheme preserves the bilinearity of the underlying differential equation. This is no longer true if higher order integration schemes are used. For these schemes the value that is used to evaluate the right hand side of the differential equation at intermediate time steps needs to be stored in the linearization trajectory. In the tangent linear integration these values from the linearization trajectory should then be used in the evaluation of the right hand side of the tangent linear model.

3.D Multilinear systems

Definition 5 (Multilinear map)**.** A map $q_n(\mathbf{X}_1, \dots \mathbf{X}_n)$ is called multilinear if it is linear in each argument.

Definition 6 (Symmetric Multilinear map)**.** For a given multilinear map q_n we define a symmetric map s_n by

$$s_n(\mathbf{X}_1, \dots \mathbf{X}_n) = \frac{1}{n!} \sum_{\text{permutations}} q_n(\mathbf{X}_1, \dots, \mathbf{X}_n), \tag{3.52}$$

where the summation is over all possible permutations of the arguments $\mathbf{X}_1, \dots \mathbf{X}_n$.

Consider the general form of a multilinear system with at most N-th order multilinearities.

$$\dot{\mathbf{X}} = \sum_{n=0}^{N} q_n(\underbrace{\mathbf{X}, \dots, \mathbf{X}}_{n \text{ times}}), \tag{3.53}$$

where q_0 is the forcing term in the model. Substitution of $\mathbf{X} \to \mathbf{X} + \mathbf{x}$ and using (3.53) gives

$$\dot{\mathbf{x}} = \sum_{n=0}^{N} q_n(\mathbf{X} + \mathbf{x}, \dots, \mathbf{X} + \mathbf{x}) - q_n(\mathbf{X}, \dots, \mathbf{X}).$$

Using definition 6 and Newton's binomium this can be written as

$$\dot{\mathbf{x}} = \sum_{n=1}^{N} \sum_{k=1}^{n} \binom{n}{k} s_n(\underbrace{\mathbf{X}, \dots, \mathbf{X}}_{n-k \text{ times}}, \underbrace{\mathbf{x}, \dots, \mathbf{x}}_{k \text{ times}}). \tag{3.54}$$

The sum over k starts from $k = 1$ because the terms with only capital $\mathbf{X}$'s are cancelled. The summation over n starts from $n = 1$ because the constant term

is cancelled. Retaining only the terms linear in $\mathbf{x}$ (terms with $k = 1$) gives the tangent linear model

$$\dot{\hat{\mathbf{x}}}^1 = \sum_{n=1}^{N} n s_n(\underbrace{\mathbf{X}, \ldots, \mathbf{X}}_{n-1 \text{ times}}, \hat{\mathbf{x}}^1). \tag{3.55}$$

If we iteratively relinearized the TL model around the trajectory $\mathbf{X} + \hat{\mathbf{x}}^{k-1}/2$ we get at convergence of the algorithm a unique increment $\hat{\mathbf{x}}^*$ that satisfies

$$\dot{\hat{\mathbf{x}}}^* = \sum_{n=1}^{N} n s_n(\underbrace{\mathbf{X} + \hat{\mathbf{x}}^*/2, \ldots, \mathbf{X} + \hat{\mathbf{x}}^*/2}_{n-1}, \hat{\mathbf{x}}^*). \tag{3.56}$$

Using Newton's binomium this can be written as

$$\dot{\hat{\mathbf{x}}}^* = \sum_{n=1}^{N} n \sum_{k=0}^{n-1} \binom{n-1}{k} s_n(\underbrace{\mathbf{X}, \ldots, \mathbf{X}}_{n-1-k}, \underbrace{\frac{\hat{\mathbf{x}}^*}{2}, \ldots, \frac{\hat{\mathbf{x}}^*}{2}}_{k}, \hat{\mathbf{x}}^*).$$

Shifting the summation over k with 1 gives

$$\dot{\hat{\mathbf{x}}}^* = \sum_{n=1}^{N} \sum_{k=1}^{n} 2^{1-k} n \binom{n-1}{k-1} s_n(\underbrace{\mathbf{X}, \ldots, \mathbf{X}}_{n-k \text{ times}}, \underbrace{\hat{\mathbf{x}}^*, \ldots, \hat{\mathbf{x}}^*}_{k \text{ times}}),$$

which can also be written as

$$\dot{\hat{\mathbf{x}}}^* = \sum_{n=1}^{N} \sum_{k=1}^{n} k 2^{1-k} \binom{n}{k} s_n(\underbrace{\mathbf{X}, \ldots, \mathbf{X}}_{n-k \text{ times}}, \underbrace{\hat{\mathbf{x}}^*, \ldots, \hat{\mathbf{x}}^*}_{k \text{ times}}).$$

Therefore we have for the exact time evolution the coefficients

		k				
		1	2	3	4	5
	1	1	-	-	-	-
	2	2	1	-	-	-
n	3	3	3	1	-	-
	4	4	6	4	1	-
	5	5	10	10	5	1

while at convergence we get for the relinearized TL model

		k				
		1	2	3	4	5
	1	1	-	-	-	-
	2	2	1	-	-	-
n	3	3	3	$1 \cdot \frac{3}{4}$	-	-
	4	4	6	$4 \cdot \frac{3}{4}$	$1 \cdot \frac{4}{8}$	-
	5	5	10	$10 \cdot \frac{3}{4}$	$5 \cdot \frac{4}{8}$	$1 \cdot \frac{5}{16}$

The normal TL-model has only nonzero values in the first column. Therefore we see that the relinearized model takes into account all linear terms but also all terms quadratic in the perturbation $\mathbf{x}$. For terms higher than quadratic in $\mathbf{x}$ the relinearized model multiplies the exact coefficient with a factor $k2^{1-k}$. This is a number between 0 and 1 and therefore is always closer to the exact coefficient then setting the coefficient to zero as is done in the standard TL-model. We therefore conclude that the relinearization iteration will always give better approximations than the standard TL-model at convergence of the algorithm.

CHAPTER 4

GAUSSIAN QUADRATURE 4D-VAR

A new incremental 4D-VAR algorithm is introduced. The algorithm does not require the computationally expensive integrations with the nonlinear model in the outer loops. Nonlinearity is accounted for by modifying the linearization trajectory of the observation operator based on integrations with the tangent linear model. This allows us to update the linearization trajectory of the observation operator in the inner loops at negligible computational cost. As a result the distinction between inner and outer loops is no longer necessary.

The key idea on which the proposed 4D-VAR method is based is that by using Gaussian quadrature it is possible to get an exact correspondence between the nonlinear time evolution of perturbations and the time evolution in the tangent linear model. It is shown that J-point Gaussian quadrature can be used to derive the exact adjoint based observation impact equations and furthermore that it is straightforward to account for the effect of multiple outer loops in these equations if the proposed 4D-VAR method is used. The method is illustrated using a three level quasi geostrophic model and the Lorenz 1996 model.

This text is published as: Gaussian quadrature 4D-VAR, Q. J. R. Meteorol. Soc., (Stappers and Barkmeijer, 2013)

4.1 Introduction

Strong constraint 4D-VAR can be considered as an unconstrained nonlinear least squares problem (Lawless et al., 2005b). The most common method to minimize the 4D-VAR cost function is an iterative method known as incremental 4D-VAR (Courtier et al., 1994) which is structurally similar to Gauss-Newton methods (Lawless et al., 2005b,a; Fisher et al., 2009). In incremental 4D-VAR a distinction is made between inner and outer loops (Trémolet, 2004) and the method uses

- High resolution nonlinear models in the outer loops
- Low resolution linear models in the inner loops

The high resolution nonlinear model integrations are needed to update the linearization trajectories for the tangent linear and adjoint model and are responsible for a significant fraction (≈40%) of the total computational cost of 4D-VAR (Fisher et al., 2009). As explained by Trémolet (2007b) due to the resolution differences (in particular the temporal resolution) between the linear and nonlinear integrations gravity waves propagate at different phase speeds. As a result the low resolution increment determined in the inner loops can have the exact opposite effect when integrated using the high resolution nonlinear model in the outer loop. This can cause a positive feedback loop resulting in the divergence of the combined inner/outer loop formalism. We refer to Trémolet (2007b) for a more detailed description of the positive feedback loop in the European Centre for Medium-Range Weather Forecasts (ECMWF) implementation of incremental 4D-VAR.

Recently it has been shown by Stappers and Barkmeijer (2012) that the tangent linear model of a quasi geostrophic model can be used for very long lead times (more than 200 days) if the trajectory around which the tangent linear model is linearized is modified. The method made use of the fact that the nonlinearity in the QG-model has a specific form called bilinearity. In this paper we will first show that the method of Stappers and Barkmeijer (2012) can be generalized to arbitrary nonlinear models for which an exact TL-model exists by using Gaussian quadrature. This result is subsequently used to write the incremental 4D-VAR equations as a fixed point iteration in the space of increment trajectories.

Several papers have appeared on adjoint based observation impact that show that higher than first order approximations of the forecast error reduction by the analysis are required (Gelaro et al., 2007; Errico, 2007). In Trémolet (2007a) it is shown that first order approximations will typically overestimate the true reduction of forecast error by a factor of two. These higher order estimates are based on Taylor expansion of the forecast error measure. The resulting algorithms still use (linear) adjoint models to propagate the gradients backwards in time but multiple integrations of the adjoint model linearized around different trajectories are used (typically the nonlinear trajectories starting from the analysis and the background). Using numerical quadrature formulas Daescu and Todling (2009) derived

higher order estimates based on nonlinear forecasts using initial conditions that are interpolated between the analysis and background. Due to the chaotic nature of the atmosphere however the nonlinear forecasts from these interpolations between the background and the analysis will become "uncorrelated" to both the background and the analysis for sufficient long lead times. As such the Gaussian quadrature formulates as proposed by Daescu and Todling (2009) will not improve the estimates of the observation impact for long lead times. Here we show that by evaluating the tangent linear propagator using J-point Gaussian quadrature during the TL-integrations the estimates of the observation impact can be improved. To demonstrate the advantage of the proposed incremental 4D-VAR method we show that it is straightforward to account for the effect of multiple outer loops in 4D-VAR in the adjoint based observation impact equations.

The organisation of the paper is as follows. In section 4.2 some notation and definitions that will be used throughout the rest of the paper are introduced. Section 4.3 shows that the exact nonlinear time evolution of perturbations can be obtained as an integral over the propagator. In section 4.4 it is shown how Gaussian quadrature can be used to approximate this integral over the TL propagator. Section 4.5 introduces an iterative method that can be used to get an exact correspondence between the nonlinear time evolution of perturbations and the integrations with the TL-model using Gaussian quadrature. In section 4.6 the incremental 4D-VAR of Courtier et al. (1994) is summarized. In section 4.7 the new 4D-VAR method is introduced by writing the equations as a fixed point iteration in the space of increment trajectories. Section 4.8 shows that in the proposed 4D-VAR method the distinction between inner and outer loops can be eliminated. In section 4.9 we show that Gaussian quadrature TL-model can be used in the adjoint based observation impact equations and show that it is possible to account for the effect of using multiple outer loops in 4D-VAR. The conclusions are given in section 4.10.

4.2 Notation

Integrations with the nonlinear model

$$x_i = \mathcal{M}_{t_{i-1}\to t_i}(x_{i-1}), \tag{4.1}$$

will also be written as $x_i = \mathcal{M}_i(x_{i-1})$. The corresponding tangent linear integrations are written as

$$\delta x_i = \mathrm{M}_i \delta x_{i-1}. \tag{4.2}$$

Given a sequence of states $x_0, \ldots, x_{N-1}$ at times $t_0, \ldots, t_{N-1}$ we define four dimensional vectors by

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix} \quad \mathbf{y}^o = \begin{bmatrix} y_0^o \\ y_1^o \\ \vdots \\ y_{N-1}^o \end{bmatrix} \quad \delta\mathbf{x} = \begin{bmatrix} \delta x_0 \\ \delta x_1 \\ \vdots \\ \delta x_{N-1} \end{bmatrix}. \tag{4.3}$$

The 4D-vector $\mathbf{x}$ will also be referred to as a trajectory. $\mathbf{y}^o$ is the observation vector (distributed in time) and $\delta\mathbf{x}$ is a vector of increments. We write

$$\mathbf{x} = \mathcal{M}(x_0), \tag{4.4}$$

for the nonlinear integration that produces the trajectory $\mathbf{x}$ starting from the initial condition x_0. Tangent linear integrations are written as

$$\delta\mathbf{x} = \mathbf{M}_{\mathbf{x}}\delta x_0, \tag{4.5}$$

where $\delta\mathbf{x}$ is a 4D-vector of increments and $\mathbf{x}$ is the trajectory around which the model is linearized. The propagation of increments to a specific time t_f is written as

$$\delta x^f = \mathrm{M}_{\mathbf{x}}\delta x_0. \tag{4.6}$$

The nonlinear observation operator acts on the 4D-vector $\mathbf{x}$ as

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{N-1} \end{bmatrix} = \begin{bmatrix} \mathcal{H}_0(x_0) \\ \mathcal{H}_1(x_1) \\ \vdots \\ \mathcal{H}_{N-1}(x_{N-1}) \end{bmatrix}, \tag{4.7}$$

and is written as

$$\mathbf{y} = \mathcal{H}(\mathbf{x}). \tag{4.8}$$

We define the linearized observation operator and observation error covariance w.r.t. these 4D-state vectors.

$$\mathbf{H} = \mathrm{diag}(\mathrm{H}_0, \mathrm{H}_1, \cdots, \mathrm{H}_{N-1}) \tag{4.9}$$

$$\mathbf{R} = \mathrm{diag}(\mathrm{R}_0, \mathrm{R}_1, \cdots, \mathrm{R}_{N-1}). \tag{4.10}$$

The observation operator linearized around the trajectory $\mathbf{x}$ will be written both as $\mathbf{H}_{\mathbf{x}}$ and $\mathbf{H}(\mathbf{x})$. The analysis and background are denoted by x^a and x^b respectively. The analysis and background trajectories by $\mathbf{x}^a$ and $\mathbf{x}^b$.

Definition 7 (Multilinear map)**.** A map $z = q_n(x_1, \ldots, x_n)$ is called multilinear if it is linear in each argument. A multilinear map of n variables will also be referred to as an n-linear map. In the following 1-linear maps are simply called linear and 2-linear maps will be called bilinear. Multilinear maps with $n > 1$ are also called nonlinear, multilinear maps with $n > 2$ are called nonbilinear.

Definition 8 (Symmetric Multilinear map)**.** For a given n-linear map q_n we define a symmetric map s_n by

$$s_n(x_1, \ldots, x_n) = \frac{1}{n!} \sum_{\text{permutations}} q_n(x_1, \ldots, x_n) \tag{4.11}$$

where the summation is over all possible permutations of the arguments $x_1, \ldots, x_n$. Note that if $x_i = x$ for all $i = 1, \ldots, n$ we have

$$s_n(x, \ldots, x) = q_n(x, \ldots, x). \tag{4.12}$$

4.3 Fundamental theorem for line integrals over vector fields

We show that the fundamental theorem for line integrals over vector fields can be used to write the exact nonlinear time evolution of perturbations in terms of an integral over the propagator. Suppose we have two states x^b_{i-1} and x^a_{i-1} at time t_{i-1}. Let $g : \mathbb{R}^N \to \mathbb{R}$ be a scalar valued function defined on the phase space of the model and C a smooth curve parameterized by $r(\beta)$ with $r(0) = x^b_{i-1}$ and $r(1) = x^a_{i-1}$ then the fundamental theorem of calculus for line integrals states that

$$\int_C \nabla g \cdot dr = g(r(1)) - g(r(0)) = g(x^a_{i-1}) - g(x^b_{i-1}).$$

This identity is independent of the path from x^b_{i-1} to x^a_{i-1}. The generalization of the fundamental theorem to a vector valued function $\mathcal{M}_i(x_{i-1})$ is

$$\int_C \mathrm{M}_i(r)dr = \mathcal{M}_i(x^a_{i-1}) - \mathcal{M}_i(x^b_{i-1}), \tag{4.13}$$

where M_i is the tangent linear propagator and $\mathcal{M}_i$ is the nonlinear propagator. If the path is parameterized as

$$r(\beta) = x^b_{i-1} + \beta(x^a_{i-1} - x^b_{i-1}) = x^b_{i-1} + \beta\delta x^a_{i-1},$$

we have $dr = \delta x^a_{i-1} d\beta$ and therefore

$$\int_C \mathrm{M}_i(r)dr = \left(\int_0^1 \mathrm{M}_i(x^b_{i-1} + \beta\delta x^a_{i-1})d\beta \right) \delta x^a_{i-1}. \tag{4.14}$$

The exact time evolution of the increment δx^a_{i-1} is given by

$$\delta x^a_i = x^a_i - x^b_i = \mathcal{M}_i(x^a_{i-1}) - \mathcal{M}_i(x^b_{i-1}).$$

Using (4.13) and (4.14) this can be written as

$$\delta x_i^a = \left(\int_0^1 \mathrm{M}_i(x_{i-1}^b + \beta \delta x_{i-1}^a) d\beta \right) \delta x_{i-1}^a. \tag{4.15}$$

For a given time t_i equation (4.15) shows that it is possible to get the exact value of the increment δx_i^a using the tangent linear model but an infinite number of evaluations of the action of the propagator on the vector δx_{i-1}^a are required. Note, however, that the integral over β in (4.15) can be approximated to any desired accuracy by quadrature formulas

$$\delta x_i^a = \sum_{j=1}^{J} \alpha_j \mathrm{M}_i(x_{i-1}^b + \beta_j \delta x_{i-1}^a) \delta x_{i-1}^a. \tag{4.16}$$

In the next section we show that Gaussian quadrature is the natural choice for the nonlinearities typically encountered in fluid dynamical systems.

Remark 2. *We emphasize that evaluating the right hand side in* (4.16) *only requires that the TL-model is available in the form of computer code and as such is applicable for realistic NWP models. The J evaluations of the propagator only differ in the linearization state around which the propagator is evaluated.*

Remark 3. *In NWP the dynamical system $\mathcal{M}_i(x_{i-1})$ usually comprises terms associated with the "dynamics" and terms associated with "physics". As shown in Stappers and Barkmeijer (2012) bilinear terms (such as advection terms in simple models) only require a single quadrature point to be represented exactly. Physical parameterizations might require more quadrature points. Therefore to reduce the computational cost in NWP it would be advantageous the let J depend on whether we are evaluating linear versions of physical parameterization schemes or dynamics. On the other hand note that the J evaluations of $\alpha_j \mathrm{M}_i(x_{i-1}^b + \beta_j \delta x_{i-1}^a) \delta x_{i-1}^a$ can be executed in parallel and require only one communication step to determine the summation over j at each time step.*

4.4 The Gaussian quadrature tangent linear model

We demonstrate that Gaussian quadrature (GQ) is the natural choice for the α_j's and β_j's in (4.16). In particular we show that for bilinear systems only a single linearization trajectory is required as initially shown in Stappers and Barkmeijer (2012).

Consider the general form for a nonlinear propagator over a single time step

$$x_i = \mathcal{M}_i(x_{i-1}) = \sum_{n=0}^{N} q_n(\underbrace{x_{i-1}, \ldots, x_{i-1}}_{n \text{ times}}). \tag{4.17}$$

The propagator will have this form for models where the nonlinearities are caused by products between state variables in the ordinary differential equations (e.g. quasi geostrophic models and the Lorenz 63 and 96 models) and an explicit integration scheme is used. If a simple Euler forward scheme is used for the time integration then the propagator from time t_{i-1} to t_i in (4.17) will only contain terms up to $N = 2$. For higher order integration schemes the propagator will contain higher order nonlinearities even if the ordinary differential equation only contains quadratic nonlinearities. We refer to Stappers and Barkmeijer (2012, appendix C) for a short discussion on these higher order nonlinearities introduced by the integration scheme in relation to the linearization trajectory for the TL model.

Suppose that at time t_{i-1} we have the value for the background x^b_{i-1} and the analysis x^a_{i-1}. We denote the corresponding trajectories by $\mathbf{x}^b$ and $\mathbf{x}^a$ respectively and let $\delta\mathbf{x}^a \equiv \mathbf{x}^a - \mathbf{x}^b$. Both the background trajectory and the analysis trajectory satisfy equation (4.17)

$$x^b_i = \mathcal{M}_i(x^b_{i-1}) = \sum_{n=0}^{N} q_n(\underbrace{x^b_{i-1}, \ldots, x^b_{i-1}}_{n \text{ times}}) \tag{4.18}$$

and

$$x^a_i = \mathcal{M}_i(x^a_{i-1}) = \sum_{n=0}^{N} q_n(\underbrace{x^a_{i-1}, \ldots, x^a_{i-1}}_{n \text{ times}}) \tag{4.19}$$

Substitution of $x^a_{i-1} = x^b_{i-1} + \delta x^a_{i-1}$ and $x^a_i = x^b_i + \delta x^a_i$ gives

$$x^b_i + \delta x^a_i = \sum_{n=0}^{N} q_n(\underbrace{x^b_{i-1} + \delta x^a_{i-1}, \ldots, x^b_{i-1} + \delta x^a_{i-1}}_{n \text{ times}}). \tag{4.20}$$

Using (4.18) to eliminate x^b_i gives

$$\delta x^a_i = \sum_{n=0}^{N} q_n(\underbrace{x^b_{i-1} + \delta x^a_{i-1}, \ldots, x^b_{i-1} + \delta x^a_{i-1}}_{n \text{ times}}) - \sum_{n=0}^{N} q_n(\underbrace{x^b_{i-1}, \ldots, x^b_{i-1}}_{n \text{ times}}). \tag{4.21}$$

Using identity (4.12) and subsequently using the symmetry and Newton's binomial theorem allows us to write this as

$$\delta x^a_i = \sum_{n=1}^{N} \sum_{k=1}^{n} \binom{n}{k} s_n(\underbrace{x^b_{i-1}, \ldots, x^b_{i-1}}_{n-k \text{ times}}, \underbrace{\delta x^a_{i-1}, \ldots, \delta x^a_{i-1}}_{k \text{ times}}). \tag{4.22}$$

The sum over k starts from $k = 1$ because the s_n terms with only x^b_{i-1}'s are cancelled. The summation over n starts from $n = 1$ because the q_0 term is

cancelled. Retaining only the terms linear in δx^a_{i-1} (terms with $k = 1$) gives the tangent linear model

$$\delta x_i = \sum_{n=1}^{N} n s_n(\underbrace{x^b_{i-1}, \ldots, x^b_{i-1}}_{n-1 \text{ times}}, \delta x_{i-1}) \equiv \mathrm{M}_i(x^b_{i-1})\delta x_{i-1}, \tag{4.23}$$

Here we drop the superscript a to indicate that this is only an approximation for δx^a_i. For short lead times $\delta x_i \approx \delta x^a_i$. If the state around which the TL-propagator is evaluated is modified as $\delta x_i = \mathrm{M}_i(x^b_{i-1} + \beta\delta x^a_{i-1})\delta x_{i-1}$ we get

$$\delta x_i = \sum_{n=1}^{N} n s_n(\underbrace{x^b_{i-1} + \beta\delta x^a_{i-1}, \ldots, x^b_{i-1} + \beta\delta x^a_{i-1}}_{n-1 \text{ times}}, \delta x_{i-1}) \tag{4.24}$$

Equation (4.23) is a special case of (4.24) with $\beta = 0$. Using Newton's binomial theorem (4.24) can be written as

$$\delta x_i = \sum_{n=1}^{N} n \sum_{k=0}^{n-1} \binom{n-1}{k} s_n(\underbrace{x^b_{i-1}, \ldots, x^b_{i-1}}_{n-k-1 \text{ times}}, \underbrace{\beta\delta x^a_{i-1}, \ldots, \beta\delta x^a_{i-1}}_{k \text{ times}}, \delta x_{i-1}). \tag{4.25}$$

redefining the summation dummy index $k \to k - 1$ gives

$$\delta x_i = \sum_{n=1}^{N} \sum_{k=1}^{n} \beta^{k-1} n \binom{n-1}{k-1} s_n(\underbrace{x^b_{i-1}, \ldots, x^b_{i-1}}_{n-k \text{ times}}, \underbrace{\delta x^a_{i-1}, \ldots, \delta x^a_{i-1}}_{k-1 \text{ times}}, \delta x_{i-1}), \tag{4.26}$$

using $n\binom{n-1}{k-1} = k\binom{n}{k}$ this can also be written as

$$\delta x_i = \sum_{n=1}^{N} \sum_{k=1}^{n} k\beta^{k-1} \binom{n}{k} s_n(\underbrace{x^b_{i-1}, \ldots, x^b_{i-1}}_{n-k \text{ times}}, \underbrace{\delta x^a_{i-1}, \ldots, \delta x^a_{i-1}}_{k-1 \text{ times}}, \delta x_{i-1}). \tag{4.27}$$

Comparing (4.22) and (4.27) we see that setting $\beta = 1/2$ ensures that $k\beta^{k-1} = 1$ for $k = 1$ and $k = 2$. Therefore if the nonlinear model is bilinear the time evolution of perturbations integrated with the TL-model with $\beta = 1/2$ is equal to the nonlinear time evolution of the perturbation $\delta x_i = \delta x^a_i$. This results was also obtained in Stappers and Barkmeijer (2012) by considering bilinear systems directly.

The key observation in this section is that by allowing multiple evaluations of M_i at each time step of the TL integration nonbilinearities can be taken into account. This can be seen as follows. Suppose we allow for J evaluations of $\mathrm{M}_i(x_{i-1})\delta x_{i-1}$ during the TL integration

$$\delta x_i = \sum_{j=1}^{J} \alpha_j \mathrm{M}_i(x_{i-1}^b + \beta_j \delta x_{i-1}^a)\delta x_{i-1}. \tag{4.28}$$

Using (4.27) this can be written as

$$\delta x_i = \sum_{j=1}^{J}\sum_{n=1}^{N}\sum_{k=1}^{n} \alpha_j k \beta_j^{k-1} \binom{n}{k} s_n(\underbrace{x_{i-1}^b, \ldots, x_{i-1}^b}_{n-k \text{ times}}, \underbrace{\delta x_{i-1}^a, \ldots, \delta x_{i-1}^a}_{k-1 \text{ times}}, \delta x_{i-1}) \tag{4.29}$$

Comparing (4.29) with the exact time evolution of the perturbation (4.22) we see that if we choose α_j and β_j such that

$$\sum_{j=1}^{J} \alpha_j k \beta_j^{k-1} = 1, \tag{4.30}$$

for $k = 1, \ldots, 2J$ the TL-model using J-point Gaussian quadrature (4.29) takes correctly into account all s_n with $n \leq 2J$. For any $k \neq 0$ we have the identity

$$1 = \int_0^1 k x^{k-1} dx. \tag{4.31}$$

Applying J-point Gaussian quadrature (Stoer and Bulirsch, 2002) to the right hand side gives

$$1 = \int_0^1 k x^{k-1} dx \approx \sum_{j=1}^{J} \frac{w_j}{2} k \left(\frac{1}{2}x_j + \frac{1}{2}\right)^{k-1}. \tag{4.32}$$

This equation is exact if $k \leq 2J$. Here x_j are roots of the J-th Legendre polynomial $P_J(x)$ and

$$w_j = \frac{2}{(1 - x_j^2)\left[P_J'(x_j)\right]^2}. \tag{4.33}$$

Comparing (4.32) with (4.30) we see that the solutions to (4.30) are given by

$$\alpha_j = \frac{w_j}{2}, \tag{4.34}$$

and

$$\beta_j = \frac{1}{2} + \frac{1}{2}x_j. \tag{4.35}$$

Table 4.1 gives the exact values of α_j and β_j for $J = 1$ to $J = 5$. For $J > 5$ tables for Gaussian quadrature (Abramowitz and Stegun, 1964) together with (4.34) and (4.35) can be used.

Table 4.1: Solutions of equation (4.30) for $J = 1$ to $J = 5$

J	β	α
1	$\frac{1}{2}$	1
2	$\frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{1}{3}}$	$\frac{1}{2}$
3	$\frac{1}{2}$	$\frac{8}{18}$
	$\frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{3}{5}}$	$\frac{5}{18}$
4	$\frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{15-2\sqrt{30}}{35}}$	$\frac{18+\sqrt{30}}{72}$
	$\frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{15+2\sqrt{30}}{35}}$	$\frac{18-\sqrt{30}}{72}$
5	$\frac{1}{2}$	$\frac{64}{225}$
	$\frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{35-2\sqrt{70}}{63}}$	$\frac{322+13\sqrt{70}}{1800}$
	$\frac{1}{2} \pm \frac{1}{2}\sqrt{\frac{35+2\sqrt{70}}{63}}$	$\frac{322-13\sqrt{70}}{1800}$

Example 4 (Lorenz 96 model)**.** In this example we determine the "quality" of the J-point Gaussian quadrature TL-model as a function of the number of quadrature points by examining the time evolution of the similarity index between the true perturbation δx_i^a obtained as the difference between two nonlinear integrations and the prediction by the J-point Gaussian quadrature TL-model δx_i

$$l_i = \frac{(\delta x_i^a, \delta x_i)}{\|\delta x_i^a\| \, \|\delta x_i\|}, \tag{4.36}$$

where $(\cdot, \cdot)$ is the Euclidean inner product. From Stappers and Barkmeijer (2012) we know that, because the Lorenz 1996 is a bilinear model, the TL-model reproduces the nonlinear time evolution of perturbations exactly using only a single quadrature point. To introduce higher order nonlinearities we performed a nonlinear change of variables in the Lorenz model (see appendix 4.A). Figure 4.1 shows the impact of the number of quadrature points on the similarity index l_i for 3 time units (The equivalent of a 15 day forecast if 1 time unit in the model is identified with 5 days as suggested by Lorenz and Emanuel (1998)) in the tangent linear model derived from this modified Lorenz model.

If $l_i = 0.7$ is used to indicate violations of the TL-assumption we see from the figure that the TL-model linearized around the background is valid for 0.05 time units (6 hours). If a single quadrature point is used this can be extended to approximately 0.4 time units (2 days) With every extra quadrature point the applicability of the J-point GQ TL-model is extended with 4 days. The 5-point GQ TL-model can be used for more than 15 days.

In principle the modified Lorenz 96 model would require an infinite number of quadrature points in the GQ TL-integrations to get an exact correspondence

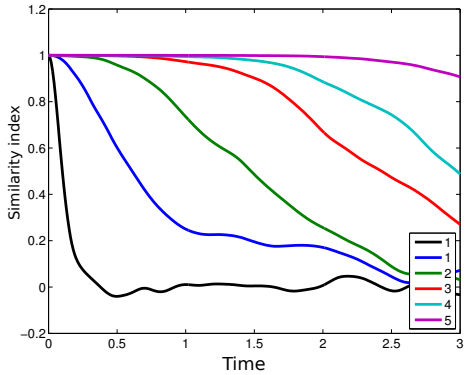

Figure 4.1: Similarity index as a function of time for the modified Lorenz 96 model using the standard TL-model linearized around the background trajectory (black) and the proposed method for $J = 1$ to $J = 5$ (colors). The experiment used random initial condition perturbations with $\|\mathbf{x}_0\| = 10$. Shown are average values over 50 experiments.

between the GQ TL integrations and the nonlinear time evolution of perturbations. From example 4 however, we see that a small number of quadrature points is sufficient to capture these "infinite" order nonlinearities.

Notation 3 (Gaussian quadrature TL integrations)**.** Integrations with the Gaussian quadrature tangent linear model from t_0 to t_N are written as

$$\delta x_N = \widetilde{\mathrm{M}}_{\delta\mathbf{x}^a} \delta x_0. \tag{4.37}$$

The subscript $\delta\mathbf{x}^a$ is the incremental correction to the background trajectory $\mathbf{x}^b$. These TL integrations produce increments for all times t_0 to t_N. We denote integrations where the complete trajectory is stored as

$$\delta\mathbf{x} = \widetilde{\mathbf{M}}_{\delta\mathbf{x}^a} \delta x_0. \tag{4.38}$$

In the case that sufficient quadrature points are used and the incremental corrections to the background linearization trajectory are obtained as the difference between two nonlinear integrations $\delta\mathbf{x}^a = \mathcal{M}(x_0^a) - \mathcal{M}(x_0^b)$ we have

$$\delta\mathbf{x}^a = \widetilde{\mathbf{M}}_{\delta\mathbf{x}^a} \delta x_0^a. \tag{4.39}$$

Note that, although the tangent linear model is used to propagate the increment, this is not a linear map from δx_0^a to $\delta\mathbf{x}^a$ due to the dependence of $\widetilde{\mathbf{M}}_{\delta\mathbf{x}^a}$ on the linearization trajectory correction $\delta\mathbf{x}^a$.

4.5 Contractions and fixed points

The propagator (4.39) depends on the increment trajectory $\delta\mathbf{x}^a$ used to modify the background. For a given initial condition increment δx_0^a we can therefore interpret (4.39) in a slightly different way by writing

$$\delta\mathbf{x}^a = \mathbf{T}(\delta\mathbf{x}^a), \tag{4.40}$$

where $\mathbf{T}(\delta\mathbf{x}_a) = \widetilde{\mathbf{M}}_{\delta\mathbf{x}^a}\delta x_0^a$, i.e. we consider the increment trajectory correction as the independent variable. The operator $\mathbf{T}$ maps increment trajectories to increment trajectories and from (4.40) we see that the increment trajectory $\delta\mathbf{x}^a$ is a fixed point of this map.

If for a fixed time interval $[0, T]$ and a given initial condition perturbation δx_0^a there is a constant $0 < q < 1$ and a suitable metric d on the space of increments defined on the interval $[0, T]$ such that for any pair of increment trajectories $\delta\mathbf{x}$, $\delta\mathbf{z}$ we have $d(\mathbf{T}(\delta\mathbf{x}), \mathbf{T}(\delta\mathbf{z})) \leq qd(\delta\mathbf{x}, \delta\mathbf{z})$ then $\mathbf{T}$ is known as a contraction mapping. The Banach fixed point theorem then guarantees that the fixed point $\delta\mathbf{x}^a$ is unique and moreover the iterated map

$$\delta\mathbf{x}_k = \mathbf{T}(\delta\mathbf{x}_{k-1}), \tag{4.41}$$

converges to the fixed point $\delta\mathbf{x}^a$. The necessary and sufficient conditions for the existence of a suitable metric d (depending on $\mathcal{M}$, $\mathbf{x}^b$ and δx_0^a) such that (4.41) is a contraction requires further research. In Stappers and Barkmeijer (2012) the convergence of (4.41) was demonstrated numerically for the quasi geostrophic (QG) model of Marshall and Molteni (1993) and the Lorenz 96 model for several cases.

4.6 Incremental 4D-VAR (Gauss-Newton estimation)

In strong constraint 4D-VAR the analysis x^a is the state vector that minimizes the cost function

$$J(x) = \tfrac{1}{2}\|x - x^b\|^2_{\mathrm{B}^{-1}} + \tfrac{1}{2}\|\mathbf{y}^o - \mathcal{G}(x))\|^2_{\mathbf{R}^{-1}}, \tag{4.42}$$

where B and $\mathbf{R}$ are covariance matrices of background and observation error, $\mathbf{y}^o$ is the vector with measurements (distributed in time) and $\mathcal{G} = \mathcal{H} \circ \mathcal{M}$ is a generalized nonlinear observation operator that involves integrations with the nonlinear model. The incremental formulation of 4D-VAR, introduced by Courtier et al. (1994) and for example used at ECMWF, is an iterative method that can be regarded as a Gauss-Newton method (Fisher et al., 2009). At iteration k, for a given guess of the analysis x_{k-1}^g, the nonlinear cost function (4.42) is approximated by

a quadratic cost function in an increment δx_k by linearizing the nonlinear observation operator around the guess. The innovation vector at outer loop iteration k is defined as

$$\mathbf{d}_{k-1} = \mathbf{y}^o - \mathcal{G}(x^g_{k-1}). \tag{4.43}$$

and the following quadratic cost function is minimized in incremental 4D-VAR

$$J^k_g(\delta x_k) = \tfrac{1}{2}\|x^g_{k-1} + \delta x_k - x^b\|^2_{\mathrm{B}^{-1}} + \tfrac{1}{2}\|\mathbf{d}_{k-1} - \mathbf{G}_{\mathbf{x}^g_{k-1}}\delta x_k\|^2_{\mathbf{R}^{-1}}, \tag{4.44}$$

where $\mathbf{G}_{\mathbf{x}^g_{k-1}}$ denotes the linearized generalized observation operator. The solution of this linear least squares problem is given by

$$\delta x^g_k = \mathrm{K}_{\mathbf{x}^g_{k-1}}\mathbf{d}_{k-1}. \tag{4.45}$$

where the gain matrix $\mathrm{K}_{\mathbf{x}^g_{k-1}}$ is given by

$$\mathrm{K}_{\mathbf{x}} = \left(\mathbf{G}^T_{\mathbf{x}}\mathbf{R}^{-1}\mathbf{G}_{\mathbf{x}} + \mathrm{B}^{-1}\right)^{-1}\mathbf{G}^T_{\mathbf{x}}\mathbf{R}^{-1}. \tag{4.46}$$

The new guess $x^g_k = x^g_{k-1} + \delta x^g_k$ is used to update the linearization trajectory and the innovation vector $\mathbf{d}_k = \mathbf{y}^o - \mathcal{G}(x^g_k)$. In the first outer loop the background trajectory $\mathbf{x}^g_0 = \mathbf{x}^b$ is used.

Remark 4. *The main reason to introduce the incremental formulation of 4D-VAR was to reduce the computational cost by running the linear models at lower resolution (Courtier et al., 1994). Therefore x^g_{k-1} and δx^g_k are elements of vector spaces with different dimensions. In expressions like $x^g_{k-1} + \delta x^g_k$ it is implicitly assumed that δx^g_k is first mapped to the high resolution vector space before taking the vector sum. Similarly when writing $\mathbf{G}_{\mathbf{x}^g_{k-1}}$ it is implicitly assumed that the linearization trajectory is first mapped to the low resolution vector space used in the TL-model.*

4.7 Incremental 4D-VAR as a fixed point iteration

The gradient of the 4D-VAR cost function (4.42) at the point x is

$$\mathbf{g}(x) = \mathrm{B}^{-1}(x - x^b) - \mathbf{G}^T_{\mathbf{x}}\mathbf{R}^{-1}\left(\mathbf{y}^o - \mathcal{G}(x)\right).$$

Let $\mathcal{H}_i$ and H_i denote the nonlinear and linear observation operator at a specific time t_i respectively. Then $\mathcal{H}_i(x_i)$ can be written as

$$\mathcal{H}_i(x_i) = \mathcal{H}_i(x^b_i) + \left(\int_0^1 \mathrm{H}_i(x^b_i + \alpha\delta x_i)d\alpha\right)\delta x_i, \tag{4.47}$$

with $\delta x_i = x_i - x_i^b$. This integral can be approximated by Gaussian quadrature. Equation (4.47) holds for any time t_i. We write

$$\mathcal{H}(\mathbf{x}) = \mathcal{H}(\mathbf{x}^b) + \widetilde{\mathbf{H}}_{\delta\mathbf{x}}\delta\mathbf{x},$$

for the observation operator acting on 4D-vectors. We write $\widetilde{\mathbf{G}}_{\delta\mathbf{x}} = \widetilde{\mathbf{H}}_{\delta\mathbf{x}}\widetilde{\mathbf{M}}_{\delta\mathbf{x}}$ for the linearized generalized observation operator where Gaussian quadrature is used in both $\mathbf{H}$ and $\mathbf{M}$. Using Gaussian quadrature the gradient of the nonlinear 4D-VAR cost function can be written as a function of the increment $\delta x = x - x^b$ as

$$\mathbf{g}_b(\delta x) = \mathrm{B}^{-1}\delta x - \mathbf{G}^T_{\mathbf{x}^b+\delta\mathbf{x}}\mathbf{R}^{-1}\left(\mathbf{d} - \widetilde{\mathbf{G}}_{\delta\mathbf{x}}\delta x\right), \tag{4.48}$$

where $\mathbf{g}_b(\delta x) = \mathbf{g}(x^b + \delta x)$ and $\mathbf{d} = \mathbf{y}^o - \mathcal{G}(x^b)$. Setting the gradient equal to zero gives

$$\delta x = \widetilde{\mathrm{K}}_{\delta\mathbf{x}}\mathbf{d}, \tag{4.49}$$

with

$$\widetilde{\mathrm{K}}_{\delta\mathbf{x}} = \left(\mathrm{B}^{-1} + \mathbf{G}^T_{\mathbf{x}^b+\delta\mathbf{x}}\mathbf{R}^{-1}\widetilde{\mathbf{G}}_{\delta\mathbf{x}}\right)^{-1}\mathbf{G}^T_{\mathbf{x}^b+\delta\mathbf{x}}\mathbf{R}^{-1}. \tag{4.50}$$

The estimated increment can be propagated with the QG TL-model. This gives

$$\delta\mathbf{x} = \widetilde{\mathbf{M}}_{\delta\mathbf{x}}\widetilde{\mathrm{K}}_{\delta\mathbf{x}}\mathbf{d} \equiv \widetilde{\mathbf{K}}_{\delta\mathbf{x}}\mathbf{d}. \tag{4.51}$$

Solutions $\delta\mathbf{x}^a$ (i.e. the analysis trajectory is $\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}^a$) of equation (4.51) are points where the gradient of the nonlinear cost function is zero. At this point no approximations have been made other than the assumption that sufficient quadrature points are used in the linear observation operator and tangent linear model. In (4.51) the increment $\delta\mathbf{x}$ still appears in the corrections to the linearization trajectories and this is therefore an implicit nonlinear equation for the increment $\delta\mathbf{x}$. As such it can not be used directly to determine the solution $\delta\mathbf{x}^a$. For a given innovation vector $\mathbf{d}$ and background trajectory $\mathbf{x}^b$ observe that (4.51) can be considered as a mapping from increment trajectories to increment trajectories

$$\delta\mathbf{x} = \mathbf{S}^{\mathbf{d}}(\delta\mathbf{x}), \tag{4.52}$$

where $\mathbf{S}^{\mathbf{d}}(\delta\mathbf{x}) = \widetilde{\mathbf{K}}_{\delta\mathbf{x}}\mathbf{d}$. The increment trajectory $\delta\mathbf{x}$ is a fixed point of the operator $\mathbf{S}^{\mathbf{d}}$. This suggests the following iterative procedure $\delta\mathbf{x}_k = \mathbf{S}^{\mathbf{d}}(\delta\mathbf{x}_{k-1})$,

$$\delta\mathbf{x}_k = \widetilde{\mathbf{K}}_{\delta\mathbf{x}_{k-1}}\mathbf{d}, \tag{4.53}$$

i.e. we use the previous estimate for the linearization trajectory corrections to obtain the current estimate. We point out that the GQ tangent linear integrations in $\widetilde{\mathbf{K}}$ are performed using J-point Gaussian quadrature while the adjoint integrations are performed around a single analysis trajectory $\mathbf{x} = \mathbf{x}^b + \delta\mathbf{x}$.

Remark 5. *If we compare the incremental 4D-VAR estimation* (4.45) *with the proposed method* (4.51) *we see that both methods account for nonlinearity by modifying the trajectory around which the generalized observation operator is linearized. There is however a conceptual difference. In* (4.45) *it is assumed that the new guess* x_k^g *is "closer" to the true analysis than the background and therefore we linearize the observation operator around the nonlinear forecast starting from the new guess at the next outer loop. In the proposed method* (4.53) *we improve the estimate of the incremental correction* $\delta\mathbf{x}_{k-1}$*, i.e. even if the background plus the estimated increment at* $t = 0$ *is not closer to the truth we can still improve the evaluations of the tangent linear propagator and linear observation operator at points in space and time close to the measurements. At subsequent iterations this information will spread in space and time. The proposed 4D-VAR method therefore seems to have better convergence properties. This can be seen, for example, in the forecast sensitivity experiments in Stappers and Barkmeijer (2012). There it is shown that in case of a single full state measurement at the end of the assimilation window the estimated increment at* $t = 0$ *becomes uncorrelated with the true increment for sufficiently long assimilation window. The new guess will therefore not improve the background and the nonlinear forecast starting from this new guess will not improve the linearization trajectory. Therefore the ordinary incremental 4D-VAR method will diverge in this case. The proposed method, however, does improve the linearization trajectory at the end of the window and these improvements of the linearization trajectory propagate backwards in time at subsequent outer loop iterations starting from the measurement time. After sufficient iterations the complete trajectory in the assimilation window is improved.*

Remark 6. *Another important difference between GQ 4D-VAR and the incremental 4D-VAR formulation is that in GQ 4D-VAR the innovation vector is not modified in the outer loops and therefore at each outer loop iteration the analysis increment is related to innovation vector as*

$$\delta\mathbf{x}_k = \widetilde{\mathbf{K}}_{\delta\mathbf{x}_{k-1}}\mathbf{d}.$$

Only $\delta\mathbf{x}_{k-1}$ *used in the linear models that define* $\widetilde{\mathbf{K}}$ *changes during the outer loops. Note that although linear models are used to determine the action of the gain matrix we should be careful to interpret this as a linear relation between* $\mathbf{d}$ *and* $\delta\mathbf{x}_k$ *for* $k > 1$ *because of the dependence of the linearization trajectory on* $\mathbf{d}$*. In the context of adjoint based observation impact studies the most important advantage of the proposed method is that at each iteration the adjoint* $\widetilde{\mathbf{K}}^T_{\delta\mathbf{x}_{k-1}}$ *of the gain matrix can be used to translate the impact of analysis increments* $\delta\mathbf{x}_k$ *to the impact of the observation* $\mathbf{y}^o$*. This will be used in section 4.9 to derive the adjoint based observation impact equations.*

Remark 7. *The analysis equation is*

$$x^a = x^b + \widetilde{\mathrm{K}}_{\delta\mathbf{x}^a}(\mathbf{y}^o - \mathcal{G}(x^b)). \tag{4.54}$$

Subtracting the truth x^t from both sides, defining $\epsilon_a = x^a - x^t$, $\epsilon_b = x^b - x^t$ and $\epsilon_o = \mathbf{y}^o - \mathbf{y}^t$ and making the perfect model assumption $\mathbf{y}^t = \mathcal{G}(x^t)$ gives

$$\epsilon_a = \epsilon_b + \widetilde{\mathrm{K}}_{\delta\mathbf{x}^a}(\mathcal{G}(x^t) + \epsilon_o - \mathcal{G}(x^b)). \tag{4.55}$$

We define error trajectories by $\boldsymbol{\epsilon}_b = \mathbf{x}^b - \mathbf{x}^t$ and $\boldsymbol{\epsilon}_a = \mathbf{x}^a - \mathbf{x}^t$ and we have

$$\boldsymbol{\epsilon}_b = \mathbf{x}^b - \mathbf{x}^t = \mathbf{x}^b - \mathbf{x}^a + \mathbf{x}^a - \mathbf{x}^t = -\delta\mathbf{x}^a + \boldsymbol{\epsilon}_a. \tag{4.56}$$

Typically $\boldsymbol{\epsilon}_a$ will be small compared to $\boldsymbol{\epsilon}_b$ and we can approximate $\mathcal{G}(x^t) - \mathcal{G}(x^b) = -\widetilde{\mathbf{G}}_{-\boldsymbol{\epsilon}_b}\epsilon_b$ by $-\widetilde{\mathbf{G}}_{\delta\mathbf{x}^a}\epsilon_b$. Using this approximation the analysis error covariance becomes

$$\overline{\epsilon_a\epsilon_a^T} = (\mathbf{I} - \widetilde{\mathrm{K}}_{\delta\mathbf{x}^a}\widetilde{\mathbf{G}}_{\delta\mathbf{x}^a})\overline{\epsilon_b\epsilon_b^T}(\mathbf{I} - \widetilde{\mathrm{K}}_{\delta\mathbf{x}^a}\widetilde{\mathbf{G}}_{\delta\mathbf{x}^a})^T + \widetilde{\mathrm{K}}_{\delta\mathbf{x}^a}\overline{\epsilon_o\epsilon_o^T}\widetilde{\mathrm{K}}_{\delta\mathbf{x}^a}^T. \tag{4.57}$$

We write $\mathrm{A} = \overline{\epsilon_a\epsilon_a^T}$, $\mathrm{B} = \overline{\epsilon_b\epsilon_b^T}$ and $\mathbf{R} = \overline{\epsilon_o\epsilon_o^T}$ and assume that B and $\mathbf{R}$ in the gain matrix K are equal to the true covariance matrices. It can be shown that with these assumptions the analysis error covariance simplifies to

$$\widetilde{\mathrm{A}}_{\mathbf{x}^b+\delta\mathbf{x}} = \widetilde{\mathrm{W}}_{\delta\mathbf{x}}\mathrm{A}_{\mathbf{x}^b+\delta\mathbf{x}}^{-1}\widetilde{\mathrm{W}}_{\delta\mathbf{x}}^T, \tag{4.58}$$

where $\widetilde{\mathrm{W}}_{\delta\mathbf{x}} = \left(\mathrm{B}^{-1} + \mathbf{G}_{\mathbf{x}^b+\delta\mathbf{x}}^T\mathbf{R}^{-1}\widetilde{\mathbf{G}}_{\delta\mathbf{x}}\right)^{-1}$ and

$$\mathrm{A}_{\mathbf{x}} = \left(\mathrm{B}^{-1} + \mathbf{G}_{\mathbf{x}}^T\mathbf{R}^{-1}\mathbf{G}_{\mathbf{x}}\right)^{-1}, \tag{4.59}$$

is the standard 4D-VAR analysis error. Note in particular that if $\mathcal{G}$ is linear we have $\widetilde{\mathrm{W}}_{\delta\mathbf{x}} = \mathrm{A}_{\mathbf{x}^b+\delta\mathbf{x}}$ and therefore the analysis error covariance is equal to the strong constraint 4D-VAR analysis error covariance.

4.8 Merging of inner and outer loops

The proposed method does not use the nonlinear model in the outer loops. In this section it is shown that this introduces the possibility to account for nonlinearity in the inner loops of 4D-VAR and the distinction between inner and outer loops is no longer necessary.

Suppose we have a guess for the linearization trajectory correction $\delta\mathbf{x}_{k-1}$ and a guess for the current increment δx_k at $t = 0$. Here k is the inner loop index and we define $\delta\mathbf{x}_0 = \mathbf{0}$. Based on these estimates we can compute the value of the cost function $J_b(\delta x_k)$ and the gradient

$$g_b(\delta x_k) = \mathrm{B}^{-1}\delta x_k + \mathbf{G}_{\mathbf{x}^b+\delta\mathbf{x}_{k-1}}^T\mathbf{R}^{-1}\left(\widetilde{\mathbf{G}}_{\delta\mathbf{x}_{k-1}}\delta x_k - \mathbf{d}\right)$$

and the minimization algorithm can determine the next value of the guess. The key observation here is that the computation of the gradient requires the computation of $\widetilde{\mathbf{G}}_{\delta\mathbf{x}_{k-1}}\delta x_k$ which involves integrations with the Gaussian quadrature

TL-model and therefore at no additional cost we obtain $\delta\mathbf{x}_k = \widetilde{\mathbf{M}}_{\delta\mathbf{x}_{k-1}}\delta x_k$. At the next inner loop iteration the new guess δx_{k+1} and the new linearization trajectory correction $\delta\mathbf{x}_k$ are used.

Remark 8. *In the proposed method the linearization trajectory updates in the inner loops are based on the low resolution GQ TL integrations. For a given estimate of the increment δx_k it might be beneficial to replace the update*

$$\delta\mathbf{x}_k = \widetilde{\mathbf{M}}_{\delta\mathbf{x}_{k-1}}\delta x_k, \tag{4.60}$$

by using the high resolution nonlinear model

$$\delta\mathbf{x}_k = \mathcal{M}(x^b + \delta x_k) - \mathbf{x}^b, \tag{4.61}$$

every N inner loop iterations to ensure "consistency" between the small and large scales. The innovation is still defined w.r.t. the background after the update.

Example 5 (Merging inner and outer loops (QG model))**.** The setup of the experiment is similar to the experiments in Thépaut and Courtier (1991). First the ECMWF analysis of 3 May 2011, 12 UTC and the 24h ECMWF forecast verifying at the same time are obtained from the Meteorological Archival and Retrieval System (MARS) and, to reduce possible transient effects, both states are integrated for 48h with the QG model of Marshall and Molteni (1993). The final state of the 48h-integration starting from the ECMWF analysis is considered to be the "truth" while the final state starting from the 24h forecast is used as the background. A 48h forecast starting from the "truth" is used to produce the "truth" trajectory and we assume that there is a complete state vector "measurement" x^t at final time. A limited memory quasi-Newton method (M1QN3) (Gilbert and Lemaréchal, 1989) is used for the minimization. In the first iteration the background is used as guess. The cost function contains no background term and the observation error covariance matrix is equal to the identity matrix. At each inner loop iteration k we compute the angle $\alpha_k = \mathrm{acos}(l_k)$ (the arccosine of the similarity index and the relative norm $\|\delta x_k\|/\|\delta x^a\|$, between the true increment δx^a and the estimated increment δx. Here $\|\cdot\|$ is the kinetic energy norm.

Table 4.2: Comparison of the different gradient calculations

Method	$g_b(\delta x_k)$
4D-VAR	$\mathbf{G}^T_{\mathbf{x}^b+\delta\mathbf{x}_0}\ \mathbf{R}^{-1}\big(\widetilde{\mathbf{G}}_{\delta\mathbf{x}_0}\ \delta x_k - \mathbf{d}\big)$
GQ 4D-VAR	$\mathbf{G}^T_{\mathbf{x}^b+\delta\mathbf{x}_{k-1}}\ \mathbf{R}^{-1}\big(\widetilde{\mathbf{G}}_{\delta\mathbf{x}_{k-1}}\delta x_k - \mathbf{d}\big)$
Exact	$\mathbf{G}^T_{\mathbf{x}^b+\delta\mathbf{x}_k}\ \mathbf{R}^{-1}\big(\widetilde{\mathbf{G}}_{\delta\mathbf{x}_k}\ \delta x_k - \mathbf{d}\big)$

In table 4.2 we define $\delta\mathbf{x}_0 = \mathbf{0}$, i.e. we have $\widetilde{\mathbf{G}}_{\delta\mathbf{x}_0} = \mathbf{G}_{\mathbf{x}^b}$. The 4D-VAR gradient calculation therefore uses the background trajectory as linearization tra-

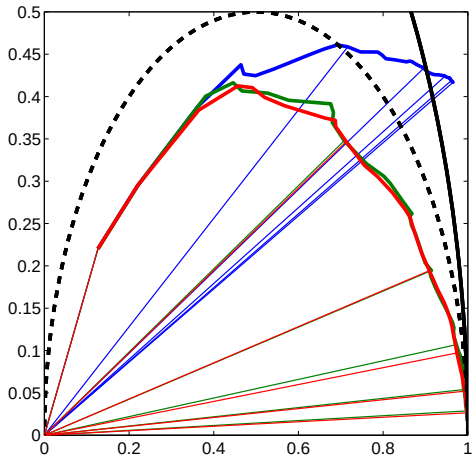

Figure 4.2: Estimated increments in the (x, y)-plane (51 iterations) for the forecast sensitivity experiment with the quasi-geostrophic model (48h assimilation window) using the 4D-VAR gradient calculation (blue) the proposed gradient calculation (green) and the exact gradient calculation (red). The straight lines are plotted every 10 iterations. The solid black line is the unit circle. The dashed black line is a circle with radius 0.5.

jectory at each inner loop iteration. Note that the exact gradient calculation requires integrations with the nonlinear model.

From figure 4.2 we conclude that if the background trajectory is fixed the estimated increment does not get more similar or closer to the truth after approximately 40 iterations. If, however, the linearization trajectory is updated in the inner loops at each iteration we get closer and more similar to the truth at negligible additional computational cost. The speed of convergence is roughly equal to the speed of using the exact gradient calculation.

Remark 9. *In the proposed method the integration* $\widetilde{\mathbf{G}}_{\delta\mathbf{x}_{k-1}}\delta x^k$ *produces the linearization trajectory correction for the next iteration. This correction is determined before the adjoint integration at iteration* k *and therefore an alternative gradient calculation is given by*

$$g_b(\delta x_k) = \mathbf{G}^T_{\mathbf{x}^b+\delta\mathbf{x}_k}\mathbf{R}^{-1}\left(\widetilde{\mathbf{G}}_{\delta\mathbf{x}_{k-1}}\delta x_k - \mathbf{d}\right), \tag{4.62}$$

which should be compared with gradient calculations in table 4.2.

4.9 Adjoint based observation impact

In this section we illustrate the advantage of using the J-point Gaussian quadrature TL-model and GQ 4D-VAR by deriving the equations for adjoint based observation impact for nonlinear models.

Let x^f be a nonlinear forecast and x^t an analysis verifying at the same time. In adjoint based observation studies x^t is considered to be the true state and the forecast error is defined by (Langland and Baker, 2004)

$$e(x^f) = \|x^f - x^t\|_{\mathrm{C}}^2, \tag{4.63}$$

where C is a symmetric weighting matrix. The impact of assimilating observations can then be defined as

$$\delta e = e(x_a^f) - e(x_b^f), \tag{4.64}$$

where x_a^f and x_b^f are forecasts starting from the analysis and the background respectively. Using a sufficient number of quadrature points we have

$$x_a^f = x_b^f + \widetilde{\mathrm{M}}_{\delta\mathbf{x}^a}\delta x^a, \tag{4.65}$$

where $\delta\mathbf{x}^a$ is the difference between the analysis trajectory and the background trajectory and δx^a is the value of the difference at $t = 0$ and therefore δe can be written as

$$\begin{aligned}
\delta e &= e(x_a^f) - e(x_b^f) \\
&= \|x_a^f - x^t\|_{\mathrm{C}}^2 - \|x_b^f - x^t\|_{\mathrm{C}}^2 \\
&= \|x_b^f - x^t + \widetilde{\mathrm{M}}_{\delta\mathbf{x}^a}x^a\|_{\mathrm{C}}^2 - \|x_b^f - x^t\|_{\mathrm{C}}^2 \\
&= \left(\delta x^a, \widetilde{\mathrm{M}}_{\delta\mathbf{x}^a}^T\mathrm{C}(\widetilde{\mathrm{M}}_{\delta\mathbf{x}^a}\delta x^a + 2(x_b^f - x^t))\right) \\
&= \left(\delta x^a, \widetilde{\mathrm{M}}_{\delta\mathbf{x}^a}^T\mathrm{C}(x_a^f + x_b^f - 2x^t)\right).
\end{aligned}$$

Except for the assumption that J-point Gaussian quadrature is sufficient to approximate the nonlinearities in the nonlinear model, no approximations have been made in this derivation. If GQ 4D-VAR is used then at the final outer loop iteration K we have (if the inner and outer loops are merged then K is the index of the last inner loop)

$$\delta x^a = \delta x_K = \widetilde{\mathrm{K}}_{\delta\mathbf{x}_{K-1}}\mathbf{d}.$$

This gives for the observation impact

$$\delta e = \left(\mathbf{d}, \widetilde{\mathrm{K}}_{\delta\mathbf{x}_{K-1}}^T\widetilde{\mathrm{M}}_{\delta\mathbf{x}^a}^T\mathrm{C}(x_a^f + x_b^f - 2x^t)\right), \tag{4.66}$$

where $(\cdot,\cdot)$ is the Euclidean inner product in observation space. This equation is the exact equation for adjoint based observation impact in the sense that the

adjoint based change in δe is equal to the actual decrease in δe. Note however that this equation is still subject to the interpretation difficulties mentioned in Errico (2007). In particular if we write (4.66) in the form

$$\delta e = (\mathbf{d}, \mathbf{g}), \tag{4.67}$$

then $\mathbf{g}$ depends on $\mathbf{d}$ through the dependence of x_a^f on $\mathbf{d}$ and therefore these equations will not give the exact impact for a subset of measurements $i \in S$ by computing

$$\delta e = \sum_{i \in S} d_i g_i. \tag{4.68}$$

See also Gelaro et al. (2007) where it is shown that this dependence of $\mathbf{g}$ on $\mathbf{d}$ has little impact on the relative ordering of the different observing systems in the NASA GEOS-5 atmospheric data assimilation system.

Based on numerical quadrature equations Daescu and Todling (2009) derived higher order adjoint based observation impact equations. The midpoint rule gives the following expression (their equation 41)

$$\delta e_2^{(a+b)/2} = \left(\mathbf{d}, \mathrm{K}^T \mathrm{M}^T_{\mathbf{x}_{(a+b)/2}} \mathrm{C}(2x^f_{(a+b)/2} - 2x^t)\right),$$

where $\mathbf{x}_{(a+b)/2}$ is the trajectory from a nonlinear forecast starting from the midpoint between the analysis and the background and $x^f_{(a+b)/2}$ is the value at verification time of this forecast. This expression bears some resemblance with the expressions derived here if the nonlinear model is bilinear in the sense that a single adjoint integration is used around a modified linearization trajectory to propagate the gradients backward in time. Furthermore, at least for short lead times we have

$$\mathbf{x}_{(a+b)/2} \approx \frac{1}{2}(\mathbf{x}^b + \mathbf{x}^a) \equiv \mathbf{x}^b + \delta \mathbf{x}^a/2.$$

It is clear however that even for bilinear systems the expression derived in Daescu and Todling (2009) will no longer be exact because the forecast $x^f_{(a+b)/2}$ will become uncorrelated to both the forecast from the analysis and the forecast from the background for sufficiently long lead times. This affects both the linearization trajectory used in the adjoint model and the vector $2x^f_{(a+b)/2} - 2x^t$ that is being propagated by the adjoint model. Also note that using $\mathbf{x}_{(a+b)/2}$ as linearization trajectory requires an extra nonlinear integration to determine the adjoint observation impact while $\frac{1}{2}(\mathbf{x}^b + \mathbf{x}^a)$ can be computed at negligible computational cost.

The difference between the two approaches can perhaps best seen by considering the vector that is propagated by the adjoint model (neglecting the weighting matrix C) when a single Gaussian quadrature point is used. In our case this gives

$$x_a^f + x_b^f - 2x^t = 2\mathcal{M}(x^b) + \widetilde{\mathrm{M}}_{\delta \mathbf{x}^a} \delta x^a - 2x^t, \tag{4.69}$$

while, if we define $\delta\mathbf{z} = \mathcal{M}(x^b + \delta x^a/2) - \mathcal{M}(x^b)$, Daescu and Todling (2009) propose

$$\begin{aligned} 2x^f_{(a,b)/2} - 2x^t &= 2\mathcal{M}(x^b + \delta x^a/2) - 2x^t \\ &= 2\mathcal{M}(x^b) + 2\widetilde{\mathrm{M}}_{\delta\mathbf{z}}\delta z - 2x^t \\ &= 2\mathcal{M}(x^b) + \widetilde{\mathrm{M}}_{\delta\mathbf{z}}\delta x^a - 2x^t, \end{aligned} \tag{4.70}$$

where we use the GQ TL-model to rewrite the nonlinear forecast starting from $x^b + \delta x^a/2$. The two approaches therefore differ in the linearization correction for the GQ TL-model. For short lead times we have $\delta\mathbf{z} \approx \delta\mathbf{x}^a/2$. The method of Daescu and Todling (2009) multiplies the increment corrections with a factor 1/2, however because both $\delta\mathbf{z}$ and $\delta\mathbf{x}$ are small the difference between the two approaches will be small. The effect of using multiple outer loops in 4D-VAR on the observation impact has been studied by Trémolet (2008). The equations derived in that paper are complicated because the previous outer loop has an impact on both the linearization trajectory and the innovation vector of the current outer loop. In the proposed method only the linearization trajectory used for the linear models that define K change and as a result it is straightforward to define the exact effect of the outer loops

$$\delta e^k = \left(\mathbf{d}, \widetilde{\mathrm{K}}^T_{\delta\mathbf{x}^a_{k-1}} \widetilde{\mathrm{M}}^T_{\delta\mathbf{x}^a_k} \mathrm{C}(x^{k,f}_a + x^f_b - 2x^t)\right), \tag{4.71}$$

where $x^{k,f}_a$ is the nonlinear forecast starting from the estimated analysis $x^b + \delta x^a_k$ at outer (or inner) loop k.

4.10 Conclusions

We have shown that the exact time evolution of perturbations given by a multilinear nonlinear model can be described exactly by using J-point Gaussian quadrature in the tangent linear model integrations. Based on this result we have introduced the Gaussian quadrature 4D-VAR algorithm that has four advantages over the incremental formulation of Courtier et al. (1994):

1. The updates of the linearization trajectory in the outer loops to account for nonlinearity are based on the tangent linear model instead of the nonlinear model. This will reduce the cost of 4D-VAR as the high resolution nonlinear integrations in the outer loops are no longer needed. Furthermore, as the nonlinear model is no longer needed to update the linearization trajectory, there is no resolution difference between the inner and outer loops and the positive feedback loop caused by the phase speed differences of gravity waves resulting in divergence of the ECMWF incremental 4D-VAR implementation can not occur.

2. The distinction between inner and outer loops is no longer necessary. Gaussian quadrature 4D-VAR allows to update the trajectory in the inner loops at negligible additional computational cost. The heuristic choice of the number of inner loops in each outer loop is thereby avoided.

3. For long assimilation windows the tangent linear assumption is not valid. The information from measurements at the end of the assimilation window can therefore not be used to estimate increments at the beginning of the window using linear models and strong constraint incremental 4D-VAR can not be used to estimate increments at $t = 0$ in these cases. Incremental 4D-VAR will therefore diverge if there are only measurement at the end of assimilation window if the assimilation window is sufficient long. Gaussian quadrature 4D-VAR however will still be able to improve the linearization trajectory at the end of the assimilation window and propagate this information backwards in time at subsequent iterations of the algorithm. The proposed method therefore seems to have better convergence properties but more research is required to put this on a solid mathematical footing.

4. The innovation vector is not modified in the (outer) loops. As a result of this it is straightforward to derive the exact adjoint based observation impact equations and to account for multiple (outer) loops.

Acknowledgements

We like to thank two referees whose suggestions have greatly improved this document and Dr. Wim Verkley for showing that identity (4.31) can be used to prove the relation between the Gaussian quadrature coefficients and the coefficients used for the Gaussian quadrature in the TL-model.

4.A Modified Lorenz 96 model

Lorenz (1996) introduced the system

$$\dot{x}^i = x^{i-1}(x^{i+1} - x^{i-2}) - x^i + F, \tag{4.72}$$

where the dimension of the state vector is N and the cyclic convention $x^{i+N} = x^i$ is used. We will use the vector notation

$$\dot{\mathbf{x}} = s(\mathbf{x}, \mathbf{x}) - \mathbf{x} + \mathbf{F}. \tag{4.73}$$

If we define $q(\mathbf{y}, \mathbf{x}) = y^{i-1}(x^{i+1} - x^{i-2})$ the symmetric bilinear map s is given by

$$s(\mathbf{x}, \mathbf{y}) = \tfrac{1}{2}\left(q(\mathbf{y}, \mathbf{x}) + q(\mathbf{x}, \mathbf{y})\right).$$

The nonlinear term conserves the total energy $\frac{1}{2}\|\mathbf{x}\|^2$, i.e. $(\mathbf{x}, s(\mathbf{x},\mathbf{x})) = 0$. The linear term $-\mathbf{x}$, representing mechanical or thermal dissipation, tends to decreases the total energy $-(\mathbf{x},\mathbf{x}) \leq 0$ while the constant term $\mathbf{F}$, representing external forcing prevents the total energy from decaying to zero. We imagine that $\mathbf{x}$ represents some atmospheric variable around a latitude circle and x^i is the value at longitude $360i/N$. In all simulations we use $N = 40$ and $\mathbf{F} = 8$. If 1 time unit in the model is identified with 5 days the error doubling time of the model is 2.1 days (Lorenz and Emanuel, 1998).

The time evolution of the total energy $\frac{1}{2}\|\mathbf{x}\|^2$ is given by

$$\frac{d\frac{1}{2}\|\mathbf{x}\|^2}{dt} = -\|\mathbf{x}\|^2 + (\mathbf{x},\mathbf{F}) \leq (-\|\mathbf{x}\| + \|\mathbf{F}\|)\,\|\mathbf{x}\|, \tag{4.74}$$

where we used the Cauchy-Schwarz inequality. Define the closed ball

$$\mathcal{B} = \{\mathbf{x} \mid \|\mathbf{x}\| \leq \|\mathbf{F}\|\}. \tag{4.75}$$

The interior and boundary of $\mathcal{B}$ will be denoted by $\mathrm{int}(\mathcal{B})$ and $\partial\mathcal{B}$ respectively. For all $\mathbf{x} \notin \mathcal{B}$ we have $\frac{d\frac{1}{2}\|\mathbf{x}\|^2}{dt} < 0$. Furthermore for all points $\partial\mathcal{B} - \{\mathbf{F}\}$ we have $\frac{d\frac{1}{2}\|\mathbf{x}\|^2}{dt} < 0$. While for $\{\mathbf{F}\}$ we have $\frac{d\frac{1}{2}\|\mathbf{x}\|^2}{dt} = 0$. For a trajectory to escape from $\mathcal{B}$ (forward in time) it would have to intersect $\partial\mathcal{B}$ and at the intersection point we should have $\frac{d\|\mathbf{x}\|^2}{dt} > 0$. This is a contradiction with the previous results and therefore if $\mathbf{x}(0) \in \mathrm{int}(\mathcal{B})$ then $\mathbf{x}(t) \in \mathrm{int}(\mathcal{B})$ for all $t > 0$.

Knowing that if $\mathbf{x}(0) \in \mathrm{int}(\mathcal{B})$ then $\mathbf{x}(t) \in \mathrm{int}(\mathcal{B})$ for $t > 0$ we define the map $\mathbf{z} : \mathrm{int}(\mathcal{B}) \to \mathbb{R}^N$ as

$$\mathbf{z} = \frac{1}{\|\mathbf{F}\| - \|\mathbf{x}\|}\mathbf{x}. \tag{4.76}$$

This is a bijection, the inverse map is given by $\mathbf{x} : \mathbb{R}^N \to \mathrm{int}(\mathcal{B})$

$$\mathbf{x} = \frac{\|\mathbf{F}\|}{1 + \|\mathbf{z}\|}\mathbf{z}. \tag{4.77}$$

Using $\frac{d}{dt}\|\mathbf{x}\| = \frac{(\mathbf{x},\dot{\mathbf{x}})}{\|\mathbf{x}\|}$ and taking the time derivative of (4.76) gives

$$\dot{\mathbf{z}} = \frac{1}{\|\mathbf{F}\| - \|\mathbf{x}\|}\dot{\mathbf{x}} + \left(\frac{1}{\|\mathbf{F}\| - \|\mathbf{x}\|}\right)^2 \frac{(\mathbf{x},\dot{\mathbf{x}})}{\|\mathbf{x}\|}\mathbf{x}. \tag{4.78}$$

The rhs of the Lorenz 96 model can be written as

$$\dot{\mathbf{x}} = \left(\frac{\|\mathbf{F}\|}{1 + \|\mathbf{z}\|}\right)^2 s(\mathbf{z},\mathbf{z}) - \frac{\|\mathbf{F}\|}{1 + \|\mathbf{z}\|}\mathbf{z} + \mathbf{F}, \tag{4.79}$$

and

$$(\mathbf{x},\dot{\mathbf{x}}) = -\|\mathbf{x}\|^2 + (\mathbf{x},\mathbf{F}) = \frac{\|\mathbf{F}\|}{1 + \|\mathbf{z}\|}\left((\mathbf{z},\mathbf{F}) - \frac{\|\mathbf{F}\|}{1 + \|\mathbf{z}\|}\|\mathbf{z}\|^2\right). \tag{4.80}$$

Using $\frac{1}{\|\mathbf{F}\|-\|\mathbf{x}\|} = \frac{\|\mathbf{z}\|+1}{\|\mathbf{F}\|}$ gives

$$\dot{\mathbf{z}} = \frac{\|\mathbf{F}\|}{1+\|\mathbf{z}\|} s(\mathbf{z},\mathbf{z}) - \mathbf{z} + \frac{\|\mathbf{z}\|+1}{\|\mathbf{F}\|}\mathbf{F} + \frac{\|\mathbf{z}\|+1}{\|\mathbf{F}\|}(\mathbf{z},\mathbf{F})\frac{\mathbf{z}}{\|\mathbf{z}\|} - \|\mathbf{z}\|\mathbf{z}. \tag{4.81}$$

Using $\frac{\partial\|\mathbf{z}\|}{\partial\mathbf{z}} = \frac{\mathbf{z}}{\|\mathbf{z}\|}$ it is straightforward to derive the TL-model.

CHAPTER 5

CONCLUSIONS AND FURTHER RESEARCH

Both variational data assimilation and ensemble prediction systems based on singular vectors make use of tangent linear and adjoint models. One of the major limitations of the application of these linear models is that the results are useful only when the linear approximation is valid. In particular for high resolution models, where nonlinearities play a more prominent role, the assumption that the growth of small perturbations can be described by linear models becomes problematic. There are concerns that the incremental 4D-VAR approach will be defeated by these nonlinearities when the resolution of the model is increased to the km scale. This thesis examines the possibilities of using the incremental 4D-VAR approach for high resolution models by accounting for nonlinearities in the inner loop of 4D-VAR. This chapter summarizes the main conclusions from this thesis and gives suggestions for further research.

5.1 Conclusions

Chapter 2 studies the properties of singular vectors that maximize the 12 hour convective available potential energy perturbation (CAPE-SVs). Their structure is compared with TE-SVs for a two week summer period in 2007 during which a severe mesoscale convective system in the south west of Finland was observed. We compare the structure of CAPE-SVs that maximize the sum of squared CAPE perturbations in a projection domain with the structure of SVs that maximize the mean CAPE in a projection domain (SCAPE-SVs). Both types of CAPE-SVs show a clear increase in the singular values indicating that small perturbations

can give rise to large CAPE perturbations during the mesoscale convective event. During the period there is a daily pattern in the singular values with lower values for singular vectors starting from the 00 and 18 UTC analysis. TE-SVs show no increase in the singular values for this period and there is no daily pattern.

During the two week period all SCAPE-SVs produce large values of CAPE at final time by increasing the specific humidity and temperature of air parcels and at the same time increasing the lapse rate above the parcels, consistent with physical considerations. The increase in moisture content and temperature in the lower troposphere will also decrease the absolute value of the convective inhibition (CIN) and therefore reduce the lift needed for parcels to reach their level of free convection (LFC). So, although the CAPE-SVs do not provide the lift needed to reach the LFC, most convection schemes are triggered more easily in runs that are perturbed with CAPE-SVs. Therefore adding CAPE-SVs to the initial condition is beneficial to all three necessary ingredients for the development of deep convection mentioned in Doswell III (1987). Another useful property of both types of CAPE-SVs for short range EPSs is that their activity mainly takes place in the lower troposphere. This in contrast to TE-SVs with OT=12h which predominantely remain in the higher troposphere (above 500 hPa). In short range EPSs we are mainly interested in creating sufficent ensemble spread near the surface. We conclude that linear combinations of these (S)CAPE-SVs could be more useful than TE-SVs in short range EPSs. However, more research is needed to establish this.

In chapter 3 a new method to integrate tangent linear models is introduced and it is shown that, by modifying the linearization trajectory, linear models can be used to simulate exactly the nonlinear time evolution of perturbations. In chapter 3 this is illustrated for the nonlinearities typically encountered in "toy models" of the atmosphere such as nonlinearities caused by advection terms (bilinear nonlinearities). Using a quasi geostrophic model the usefulness of the tangent linear model is extended from 2 days to more than 200 days by modifying the trajectory around which the linear model is linearized. In chapter 4 this result is generalized to nonlinear models with higher order nonlinearities by using Gaussian quadrature. In both chapters the idea is to use incremental corrections to the linearization trajectories and it is shown that this incremental correction is a fixed point of a mapping from increment trajectories to increment trajectories. Numerically it is shown that there exists a fixed point iteration in the space of increment trajectories, purely based on linear integrations, that converges to the optimal incremental trajectory correction. Chapter 3 shows how this result can be used to solve inverse problems as fixed point iterations in the space of increment trajectories using forecast sensitivity experiments. Chapter 4 generalizes this result to the incremental formulation of 4D-VAR. One of the main attractive features of the method is that its implementation is relatively straightforward even in highly

Mathematically the linear model describes the nonlinear time evolution of the perturbations exactly. The 200 days is a result of the finite machine precision used in the calculations.

complex NWP models. The proposed incremental 4D-VAR algorithm has several advantages over the current 4D-VAR algorithm:

- The update of the linearization trajectory in the outer loop is based on the tangent linear model. This reduces the cost of 4D-VAR as the high resolution nonlinear integrations are no longer required.
- The distinction between inner and outer loops becomes obsolete. The heuristic choice of the number of inner loops in each outer loop is thereby avoided.
- Based on the forecast sensitivity experiments, the proposed method seems to have better convergence properties. In particular with long windows in forecast sensitivity experiments the estimated increment at $t = 0$ is uncorrelated with the true increment and the nonlinear model can not be used to update the linearization trajectory. Using forecast sensitivity experiments in the Lorenz 96 model the iterative method is illustrated for long windows and it exhibits good converge properties.
- The innovation vector is not modified in the outer loop. This allows us to derive the exact adjoint based observation impact equations even in cases where multiple outer loops are used.

5.2 Further research

We summarize the main points that need to be addressed before operational implementation of ensemble prediction systems based on CAPE-SVs and variational data assimilation methods based on Gaussian quadrature are considered.

1. The structure of the CAPE-SVs is consistent with physical consideration. However, the tangent linear model that was used to calculate the CAPE-SVs only contained vertical diffusion as physical parameterization. To what extent the growth of the CAPE-SVs is hampered by the limited use of physics is not known. To further exploit their use in convective situations it would be useful to continue experimentation in a model environment where the linearized physics includes more physical parameterizations of mesoscale processes.
2. The strong nonlinearities at the convective scale might deteriorate the optimal performance of techniques that rely on the tangent linear assumption (Hohenegger and Schär, 2007) such as the singular vector calculation in this thesis. It would be interesting to investigate whether the results of chapter 2 can be combined with the results of chapter 3 and chapter 4 to calculated singular vectors that show maximum nonlinear growth.
3. In realistic NWP models the tangent linear model is integrated at a lower resolution and typically simplified physical parameterizations are used. The

impact of these approximations on the iterative methods from chapter 3 and 4 should be assessed.

4. Several NWP centres now take model uncertainty into account in the data assimilation process, i.e. the model is only imposed as a weak constraint. It should be examined whether the method can be used in these weak constraint formulations of the data assimilation problem.

5. A numerical or theoretical examination is required to find out for which optimization problems with nonlinear constraints the proposed alternative to the Gauss-Newton method gives better convergence properties and to identify the strengths and weaknesses of the method.

6. In chapter 4 it is shown that Gaussian quadrature between the background and analyis trajectory in the linear model can be used to obtain the exact time derivative of the perturbations. In ensemble methods we have a collection of background trajectories available and modern ensemble DA methods typically produce a collection of "analyses" and forecasts. It would be interesting to investigate whether the Gaussian quadrature method between a single background and analysis trajectory can be extended to cases where we have collections of background and analysis trajectories.

7. It can be shown that the y-intercept of the relative error norm in figure 3.5 should be proportional to the machine precision used in the calculations. It is remarkable that we can treat the model as a black box but can still infer the machine precision that was used in the calculations. It would be interesting to investigate whether this can be used as a diagnostic tool in complex models.

BIBLIOGRAPHY

M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions.* Dover publications, 1964.

E. Andersson, M. Fisher, E. Hólm, L. Isaksen, G. Radnoti, and Y. Trémolet. Will the 4D-Var approach be defeated by nonlinearity? Tech. Memo 479, ECMWF, 2005.

J. Badger and B.J. Hoskins. Simple initial value problems and mechanisms for baroclinic growth. *J. Atmos. Sci.*, 58(1):38–49, 2001.

J. Barkmeijer. Local skill prediction for the ecmwf model using adjoint techniques. *Mon. Weather Rev.*, 121(4):1262–1268, 1993.

J. Barkmeijer. Adjoint diagnostics for the atmosphere and ocean. In *Seminar on Diagnosis of Forecasting and Data Assimilation Systems, 7 - 10 September 2009*, pages 287–300, 2010.

J. Barkmeijer, R. Buizza, TN Palmer, K. Puri, and J.F. Mahfouf. Tropical singular vectors computed with linearized diabatic physics. *Q. J. R. Meteorol. Soc.*, 127(572):685–708, 2001.

L.K. Bengtsson, L. Magnusson, and E. Källén. Independent estimations of the asymptotic variability in an ensemble forecast system. *Mon. Weather Rev.*, 136: 4105–4112, 2008.

D.O. Blanchard. Assessing the vertical distribution of convective available potential energy. *Weather Forecast*, 13(3):870–877, 1998.

Borel E. *Introduction géométrique à quelques théories physiques.* Gauthier-Villars, 1914.

N.E. Bowler, A. Arribas, K.R. Mylne, K.B. Robertson, and S.E. Beare. The mogreps short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.*, 134 (632):703–722, 2008.

K.E. Brenan, S.L. Campbell, and L.R. Petzold. *Numerical solution of initial-value problems in differential-algebraic equations*. Society for Industrial Mathematics, 1996.

P. Courtier, J.N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4D-var using an incremental approach. *Q. J. R. Meteorol. Soc.*, 120:1367–1387, 1994.

D.N. Daescu and R. Todling. Adjoint estimation of the variation in model functional output due to the assimilation of data. *Mon. Weather Rev.*, 137(5): 1705–1716, 2009.

C.A. Doswell III. The distinction between large-scale and mesoscale contribution to severe convection: A case study example. *Weather Forecast*, 2(1):3–16, 1987.

C.A. Doswell III and E.N. Rasmussen. The effect of neglecting the virtual temperature correction on CAPE calculations. *Weather Forecast*, 9(4):625–629, 1994.

J. Du et al. Ncep short-range ensemble forecast (sref) system upgrade in 2009. In *Preprints, 19th Conf. on Numerical Weather Prediction and 23rd Conf. on Weather Analysis and Forecasting, Omaha, Nebraska, Amer. Meteor. Soc. Weather and Forecasting*, 2009.

R.M. Errico. What is an adjoint model? *Bull. Am. Meteorol. Soc.*, 78(11): 2577–2591, 1997.

R.M. Errico. The Workshop on Applications of Adjoint Models in Dynamic Meteorology. *Bull. Am. Meteorol. Soc.*, 84:795–798, 2003.

R.M. Errico. Interpretations of an adjoint-derived observational impact measure. *Tellus*, 59A(2):273–276, 2007.

R.M. Errico and M. Ehrendorfer. Adjoint Model Applications in Dynamic Meteorology. *Bull. Am. Meteorol. Soc.*, 88(12):1979–1981, 2007.

R.M. Errico, K.D. Raeder, and L. Fillion. Examination of the sensitivity of forecast precipitation rates to possible perturbations of initial conditions. *Tellus*, 55A(1):88–105, 2003.

R.M. Errico, K.D. Raeder, and M. Ehrendorfer. Singular vectors for moisture-measuring norms. *Q. J. R. Meteorol. Soc.*, 130(598):963–988, 2004.

L. Fillion and S. Bélair. Tangent Linear Aspects of the Kain–Fritsch Moist Convective Parameterization Scheme. *Mon. Weather Rev.*, 132(10):2477–2494, 2004.

M. Fisher, M. Leutbecher, and G.A. Kelly. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, 131(613):3235–3246, 2005.

M. Fisher, J. Nocedal, Y. Trémolet, and S.J. Wright. Data assimilation in weather forecasting: a case study in PDE-constrained optimization. *Optimization and Engineering*, 10(3):409–426, 2009.

R. Gelaro, Y. Zhu, and R.M. Errico. Examination of various-order adjoint-based approximations of observation impact. *Meteorol. Z.*, 16(6):685–692, 2007.

J.C. Gilbert and C. Lemaréchal. Some numerical experiments with variable-storage quasi-Newton algorithms. *Mathematical programming*, 45(1):407–435, 1989.

I. Gilmour, L.A. Smith, and R. Buizza. Linear regime duration: Is 24 hours a long time in synoptic weather forecasting? *J. Atmos. Sci.*, 58:3525–3539, 2001.

J. Gleick. *Chaos: Making a new science*. Vintage, 1987.

E. Hágel and A. Horányi. The ARPEGE/ALADIN limited area ensemble prediction system: the impact of global targeted singular vectors. *Meteorol. Z.*, 16 (6):653–663, 2007.

C. Hohenegger and C. Schär. Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Am. Meteorol. Soc.*, 88(11):1783–1793, 2007.

V. Homar, D.J. Stensrud, J.J. Levit, and D.R. Bright. Value of human-generated perturbations in short-range ensemble forecasts of severe weather. *Weather Forecast*, 21(3):347–363, 2006.

B.J. Hoskins, R. Buizza, and J. Badger. The nature of singular vector growth and structure. *Q. J. R. Meteorol. Soc.*, 126(566):1565–1580, 2000.

T. Iversen, I.-L. Frogner, K. Sattler, S. Ivatek-Sahdan, J. Barkmeijer, H. Feddersen, E. Hágel, and R. Mladek. Glameps - the hirlam/aladin grand limited area model ensemble prediction system. In *ECMWF Workshop on Ensemble Prediction, 7–9 November 2007*, pages 95–106. ECMWF, 2007.

M Janisková. Physical processes in adjoint models: potential pitfalls and benefits. In *Proceedings of the ECMWF seminar on Recent developments in data assimilation for atmosphere and ocean*, pages 179–192, Reading, UK, September 2003. ECMWF.

M. Janisková, J. N. Thépaut, and J. F. Geleyn. Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. *Mon. Weather Rev.*, 127(1):26–45, 1999.

R.H. Johns and C.A. Doswell III. Severe local storms forecasting. *Weather Forecast*, 7(4):588–612, 1992.

E. Klinker, F. Rabier, and R. Gelaro. Estimation of key analysis errors using the adjoint technique. *Q. J. R. Meteorol. Soc.*, 124(550):1909–1933, 1998.

R.H. Langland and N.L. Baker. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, 56A(3):189–201, 2004.

A. S. Lawless, S. Gratton, and N. K. Nichols. An investigation of incremental 4d-var using non-tangent linear models. *Q. J. R. Meteorol. Soc.*, 131(606): 459–476, 2005a.

A.S. Lawless, S. Gratton, and N.K. Nichols. Approximate iterative methods for variational data assimilation. *Int. J. Numer. Meth. Fluids*, 47(10-11):1129–1135, 2005b.

M. Leutbecher and T.N. Palmer. Ensemble forecasting. *J. Comput. Phys.*, 227 (7):3515–3539, 2008.

M. Lindskog, D. Dee, Y. Trémolet, E. Andersson, G. Radnóti, and M. Fisher. A weak-constraint four-dimensional variational analysis system in the stratosphere. *Q. J. R. Meteorol. Soc.*, 135(640):695–706, 2009.

A.C. Lorenc. Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc.*, 114 (479):205–240, 1988.

E.N. Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on Predictability*, volume 1, pages 1–18. ECMWF, 1996.

E.N. Lorenz and K.A. Emanuel. Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, 55(3):399–414, 1998.

J.-F. Mahfouf and B. Bilodeau. Adjoint sensitivity of surface precipitation to initial conditions. *Mon. Weather Rev.*, 135(8):2879–2896, 2007.

J. Marshall and F. Molteni. Toward a dynamical understanding of planetary-scale flow regimes. *J. Atmos. Sci.*, 50(12):1792–1818, 1993.

T.N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza. Singular vectors, metrics, and adaptive observations. *J. Atmos. Sci.*, 55(4):633–653, 1998.

J. Pedlosky. *Geophysical fluid dynamics*. Springer, 1987. ISBN 0387963871.

R.A. Peppler. *A review of static stability indices and related thermodynamic parameters*. Illinois State Water Survey climate and meteorology section, 1988.

Z.X. Pu, E. Kalnay, J.C. Derber, and J.G. Sela. Using forecast sensitivity patterns to improve future forecast skill. *Q. J. R. Meteorol. Soc.*, 123(540):1035–1053, 1997a.

Z.X. Pu, E. Kalnay, J. Sela, and I. Szunyogh. Sensitivity of forecast errors to initial conditions with a quasi-inverse linear method. *Mon. Weather Rev.*, 125 (10):2479–2503, 1997b.

F. Rabier, E. Klinker, P. Courtier, and A. Hollingsworth. Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc.*, 122(529):121–150, 1996.

C.A. Reynolds and T.N. Palmer. Decaying singular vectors and their impact on analysis and forecast correction. *J. Atmos. Sci.*, 55(19):3005–3023, 1998.

D.M. Schultz and P.N. Schumacher. The use and misuse of conditional symmetric instability. *Mon. Weather Rev.*, 127(12):2709–2732, 1999.

R.J.J. Stappers and J. Barkmeijer. Properties of singular vectors using convective available potential energy as final time norm. *Tellus*, 63A(3):373–384, 2011.

R.J.J. Stappers and J. Barkmeijer. Optimal linearization trajectories for tangent linear models. *Q. J. R. Meteorol. Soc.*, 138(662):170–184, 2012.

R.J.J. Stappers and J. Barkmeijer. Gaussian quadrature 4D-VAR. *Q. J. R. Meteorol. Soc.*, 2013. doi: 10.1002/qj.2056.

D.J. Stensrud, J.W. Bao, and T.T. Warner. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Weather Rev.*, 128(7):2077–2107, 2000.

J. Stoer and R. Bulirsch. *Introduction to numerical analysis.* Springer Verlag, 2002.

K. Swanson, R. Vautard, and C. Pires. Four-dimensional variational assimilation and predictability in a quasi-geostrophic model. *Tellus*, 50A(4):369–390, 1998.

M. Tanguay, P. Bartello, and P. Gauthier. Four-dimensional data assimilation with a wide range of scales. *Tellus*, 47A(5):974–997, 1995.

J.N. Thépaut and P. Courtier. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc.*, 117(502):1225–1254, 1991.

Y. Trémolet. Diagnostics of linear and incremental approximations in 4D-Var. *Q. J. R. Meteorol. Soc.*, 130(601):2233–2251, 2004.

Y. Trémolet. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.*, 132(621):2483–2504, 2006.

Y. Trémolet. First-order and higher-order approximations of observation impact. *Meteorol. Z.*, 16(6):693–694, 2007a.

Y. Trémolet. Incremental 4D-Var convergence study. *Tellus*, 59A(5):706–718, 2007b.

Y. Trémolet. Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus*, 60A(5):964–978, 2008.

P. Undén, L. Rontu, H. Järvinen, P. Lynch, J. Calvo, G. Cats, J. Cuxart, K. Eerola, C. Fortelius, J.A. Garcia-Moya, et al. Hirlam-5 scientific documentation. Technical report, SMHI, 2002.

T. Vukićević and R.M. Errico. Linearization and adjoint of parameterized moist diabatic processes. *Tellus*, 45A(5):493–510, 1993.

M. Xu, D.J. Stensrud, J.W. Bao, and T.T. Warner. Applications of the adjoint technique to short-range ensemble forecasting of mesoscale convective systems. *Mon. Weather Rev.*, 129(6):1395–1418, 2001.

X. Yang. Analysis blending using spatial filter in grid-point model coupling. *Hirlam Newsletter*, 48:49–55, 2005.

SAMENVATTING

Dit proefschrift is gericht op het verbeterd gebruik van waarnemingen in hoge resolutie numerieke weersverwachtingsmodellen en verbeterde karakterisaties van de onzekerheid in de weersverwachtingen van deze modellen, met name voor gevaarlijk weer situaties.

Wiskundig gezien is het maken van een weersverwachting een beginwaardeprobleem. Als gevolg van het chaotisch gedrag van de atmosfeer kunnen kleine verstoringen van de beginconditie in numerieke weersverwachtingsmodellen snel groeien (het bekende vlindereffect). Voor het maken van accurate weersverwachtingen is het daarom essentieel om een goede schatting van de huidige toestand van de atmosfeer te hebben. Het assimileren van waarnemingen in het weermodel is dan ook een van de belangrijkste onderdelen van het maken van een weersverwachting.

De huidige, meest succesvolle, wiskundige technieken om de optimale toestand van de atmosfeer te schatten op de synoptische schaal (1000 km) maken gebruik van gelineariseerde modellen van de atmosfeer dynamica en van de operator die de model toestand op de waarnemingen afbeeldt. Door het gebruik van deze lineaire modellen hebben deze data assimilatietechnieken moeite met het assimileren van hoge resolutie waarnemingen die vaak sterk niet lineair afhangen van de model toestand, zoals b.v. radarwaarnemingen. Voor het maken van hoge resolutie korte termijn weersverwachtingen bevatten deze waarnemingen echter cruciale informatie over snel ontwikkelende systemen zoals onweer. Met name voor deze situaties wordt in dit proefschrift een nieuwe data assimilatietechniek ontwikkeld die beter geschikt lijkt voor hoge resolutie modellen met waarnemingen die sterk niet lineair afhangen van de modeltoestand.

Als gevolg van de toename van het aantal waarnemingen en het verbeteren van de data assimilatie algoritmes is de schatting van de toestand van de atmosfeer in de loop der jaren sterk verbeterd. De toestand zal echter nooit exact bekend zijn. Om de invloed van deze beginconditieonzekerheid op de weersverwachtingen te karakteriseren, zijn verschillende weerscentra overgestapt op het maken van kansverwachtingen. Bij het maken van deze kansverwachtingen wordt het weermodel

een aantal keer (tientallen) geïntegreerd met net iets andere begincondities (om de onzekerheid in het model te karakteriseren, wordt bij iedere integratie vaak ook een net iets ander model gebruikt). Het idee is dat de spreiding in deze weersverwachtingen een indicatie geeft over de voorspelbaarheid van het weer. Een van de meest succesvolle manieren om geschikte begincondities te bepalen voor de kansverwachtingen, op de synoptische schaal, maakt gebruikt van singuliere vectoren. Deze singuliere vectoren zijn de linear snelst groeiende beginconditieverstoringen over een gegeven optimalisatie tijdsvenster gemeten in een bepaalde norm. Als een eerste aanzet tot het ontwikkelen van hoge resolutie kansverwachtingssystemen worden in dit proefschrift de eigenschappen van singuliere vectoren bestudeerd die maximale impact hebben op de convectief beschikbare potentiële energie.

Data assimilatie

Het is algemeen bekend dat voor lineaire modellen met Gaussisch verdeelde fouten de beste schatting van de toestand wordt gegeven door het Kalman filter. Door het grote aantal vrijheidsgraden in weermodellen is het propageren van de covariantiematrix in de tijd, wat nodig is in het Kalman filter, niet mogelijk. In weermodelen worden daarom variationele methoden gebruikt. Deze produceren dezelfde schatting als het Kalman filter aan het eind van het assimilatie tijdsvenster echter zonder de noodzaak om de covariantiematrix in de tijd te propageren. Voor de typische dimensies van de toestandsvector in weermodellen zijn deze variationele methoden vele malen efficiënter.

In het data-assimilatie systeem wordt periodiek, typisch iedere 6 of 12 uur, een nieuwe optimale schatting van de toestand van de atmosfeer gemaakt op basis van miljoenen waarnemingen verdeeld, onregelmatig in ruimte en tijd, over de aarde. Het bepalen van de optimale beginconditie komt neer op het minimaliseren van een kostfunctie. Deze kostfunctie “meet” hoe ver de modeltoestand van de waarnemingen is verwijderd in het 6 of 12 uurs assimilatie tijdsvenster. Het data-assimilatie probleem is uitdagend om een aantal redenen:

1. Er zijn miljarden onbekende parameters (temperatuur, druk, vocht, windrichting en windsnelheid) met tijdstappen van 30 minuten in het 12-uurs tijdsvenster. Het aantal waarnemingen is dus bijna 3 ordegroottes kleiner dan het aantal te schatten parameters. Wiskundig gezien is het assimilatie probleem slecht gesteld. Om de gesteldheid van het probleem te verbeteren wordt extra informatie in de vorm van een korte termijn weersverwachting gebruikt. Verder wordt doorgaans aangenomen dat het niet-lineaire weermodel de atmosferische dynamica exact beschrijft met als gevolg dat het vastleggen van de beginconditie ook de toekomstige toestanden vastlegt. Met andere woorden, de modeldynamica wordt opgelegd als een sterke beperking om zo het aantal onbekenden te reduceren.

2. Er is een strikt tijdsbudget beschikbaar om het probleem op te lossen. Weersverwachtingen worden gemaakt met een 12-uurlijkse cyclus. Van deze 12 uur zijn er slechts 2 beschikbaar om het data-assimilatie probleem op te lossen.

3. De kostfunctie is niet-convex en de beperkingen zijn niet-lineair.

4. Het evalueren van de beperkingen doormiddel van een model integratie kost veel rekenkracht en kan slechts enkele malen per 12 uur worden gedaan.

Om het data-assimilatie probleem efficiënt op te lossen wordt een incrementele methode gebruik waarbij gebruikt wordt gemaakt van zowel gelineariseerde versies van het niet-lineaire atmosfeermodel als van de operator die de modeltoestand aan de observaties relateert (de observatieoperator). Het grote voordeel van deze aanpak is dat het deelprobleem nu reduceert tot een standaard lineair kleinste kwadraten probleem waarvoor efficiënte oplossingsmethoden bestaan. Het oplossen van dit lineaire probleem gebeurt met iteratieve methoden we noemen dit binnenlus. Aangezien niet-lineaire effecten wel degelijk belangrijk zijn, wordt het minimaliseren van de kostfunctie een aantal malen herhaald waarbij het niet-lineaire model steeds opnieuw wordt gelineariseerd (de buitenlus) rond (hopelijk) steeds betere oplossingen. Deze aanpak om niet-lineaire kleinste kwadraten problemen op te lossen staat ook bekend als de Gauss-Newton methode.

Er zijn een aantal problemen met het gebruik van deze Gauss-Newton methode in het data-assimilatie probleem. Ten eerste is bekend dat het gecombineerde binnen- en buitenlus systeem kan divergeren als er te veel buitenlus iteraties worden gedaan. Dit wordt (deels) veroorzaakt door het feit dat, om rekentijd te besparen, het lineaire model op een lagere resolutie wordt geïntegreerd dan de niet-lineaire integraties die worden gebruikt om de linearisatiebaan aan te passen. Dit zorgt er voor dat zwaartekrachtsgolven verschillende voortplantingssnelheden hebben in beide integraties. Hierdoor hebben de geschatte incrementen soms precies het tegenovergestelde effect dan verwacht werd wanneer deze worden geïntegreerd met het niet-lineaire model. Er ontstaat zo een positieve terugkoppel lus die zorgt voor de divergentie. Ten tweede, met het toenemen van de rekenkracht van supercomputers neemt ook de resolutie van de weermodellen toe waardoor steeds meer processen op kleine schaal kunnen worden beschreven. Met name ook processen die gevaarlijk weer opleveren zoals diepe convectie. Door het verhogen van de resolutie worden echter ook niet-lineaire effecten steeds belangrijker. Hier komt nog bij dat de relatie tussen de metingen van kleinschalige fenomenen (denk aan radarmetingen) en de modeltoestand vaak sterk niet-lineair is. Dit alles zorgt er voor dat er serieuze zorgen zijn dat de Gauss-Newton aanpak, die zo succesvol is gebleken op de synoptische schaal, niet meer bruikbaar is voor hoge resolutie weermodellen.

In dit proefschrift wordt eerst aangetoond dat, voor niet-lineaire modellen waar de niet-lineariteit een gevolg is van producten tussen toestandsvariabelen, het mogelijk is om met het lineaire model de niet-lineaire groei van begin conditie

verstoringen exact te beschrijven wanneer een geschikte keuze word gemaakt voor de linearisatiebaan van het lineaire model. Deze linearisatiebaan is een dekpunt van een afbeelding. Op basis van deze observatie wordt een iteratieve methode voorgesteld om dit dekpunt te bepalen. Numeriek wordt aangetoond dat voor simpele modellen deze iteratieve methode convergeert naar de optimale linearisatiebaan. Dit resultaat wordt vervolgens generaliseerd naar meer algemene niet lineaire modellen waar gebruik wordt gemaakt van Gaussische quadrature formules. Dit leidt vervolgens tot een alternatief algoritme om variationele data-assimilatie problemen op te lossen. Het voorgestelde algoritme heeft een aantal voordelen ten opzichte van het Gauss-Newton algoritme. Ten eerste is het bruikbaar voor zeer lange assimilaties vensters in situatie waar het Gauss-Newton algoritme divergeert. Met het nieuwe algoritme is het bijvoorbeeld mogelijk om in een quasi-geostroof model een optimale beginconditie te schatten op basis van waarnemingen die twee dagen in de toekomst liggen. Ten tweede maakt de methode geen gebruik van het niet-lineaire model voor het minimaliseren van de niet lineaire kostfunctie en dit zorgt voor een aanzienlijke besparing in de rekentijd indien, zoals gebruikelijk, de lineaire modellen worden geïntegreerd op een lagere resolutie. Het feit dat het niet-lineaire model niet meer nodig is impliceert ook dat het algoritme gemakkelijker te analyseren is. Dit blijkt bijvoorbeeld uit het feit dat het mogelijk is om de impact van de waarnemingen op de voorspellingen exact te kwantificeren. Een bijkomend voordeel is dat de techniek relatief eenvoudig te implementeren is als de gelineariseerde versies van het niet-lineaire model beschikbaar zijn, zefs in complexe weermodellen. Met de “klassieke” linearisatiebanen kan het lineaire model typisch voor ongeveer 2 dagen worden gebruikt om de groei van beginconditie verstoringen te benaderen. Met de in dit proefschrift geïntroduceerde methode is het mogelijk om met het lineaire model de foutengroei exact te beschrijven voor meer dan 200 dagen.

Kansverwachtingen en CAPE singuliere vectoren

Een van de kerntaken van het KNMI is het uitgeven van waarschuwingen voor gevaarlijk weer. Voor het maken van de korte termijn weersverwachtingen maakt het KNMI o.a. gebruik van het weermodel HIRLAM. Dit model wordt verder gebruikt door Denemarken, Estland, Finland, Ierland, IJsland, Litouwen, Noorwegen, Spanje en Zweden voor het maken van de operationele weersverwachting. Zoals al aangeven, zijn verschillende weerscentra in de wereld overgestapt op het maken van kansverwachtingen. In dit proefschrift is gekeken naar de mogelijkheid om een kansverwachtingssysteem binnen HIRLAM op te zetten om een indruk te krijgen van de voorspelbaarheid van één type gevaarlijk weer, namelijk diep convectieve systemen. Voor het ontstaan van diepe convectie is het noodzakelijk dat de atmosfeer voldoende instabiel is, dat wil zeggen dat de temperatuur snel genoeg afneemt met de hoogte en er dient voldoende vocht aanwezig te zijn, zodanig dat luchtpakketjes die opstijgen en daardoor afkoelen snel verzadigd raken. De

latente energie die vrijkomt bij deze condensatie is een extra warmtebron en zorgt er voor dat de luchtpakketjes een hogere temperatuur houden dan de omgeving. Een derde eis is de aanwezigheid van externe verstoringen die zorgen dat luchtpakketjes de convectieve remming (engels: convective inhibition) kunnen overwinnen. De eerste twee vereisten kunnen worden uitgedrukt in de convectief beschikbare potentiële energie (CAPE). Als een eerste aanzet tot het opzetten van een korte termijn hoge resolutie kansverwachtingssysteem worden in dit proefschrift singuliere vectoren geïntroduceerd die maximale impact hebben op de 12-uurs CAPE verwachting en de eigenschappen van deze vectoren worden beschreven.

De singuliere vectoren zijn iedere zes uur berekend voor een periode van twee weken in augstus 2007. Gedurende deze periode was er een zware onweersbui boven zuidwest Finland die door verschillende operationele HIRLAM modellen slecht (of zelfs niet) werd voorspeld. Uit experimenten die in dit proefschrift zijn beschreven, blijkt dat tijdens de onweersbui de 12-uurs CAPE verwachtingen voor zuidwest Finland relatief gevoelig waren voor kleine verstoringen van de beginconditie. Dit suggereert dat de “slechte” HIRLAM voorspellingen mogelijk een gevolg waren van fouten in de beginconditie.

Het blijkt dat de CAPE singuliere vectoren de convectief beschikbare potentiële energie in de verwachting vergroten door de temperatuur en de specieke luchtvochtigheid van de luchtpakketjes te verhogen en de verticale temperatuurgradiënt boven de luchtpakketjes te versterken, consistent met wat op basis van fysische overwegingen verwacht mag worden. De beginconditieverstoringen die maximale impact hebben op de 12-uurs CAPE verwachting zijn voornamelijk temperatuur verstoringen rond 850 hPa. De CAPE singuliere vectoren bevinden zich relatief laag in de troposfeer in vergelijking met de meer gangbare totale energie singuliere vectoren die worden gebruikt in kansverwachtingssystemen voor de middellange termijn. Bij het maken van korte termijn hoge resolutie weersverwachtingen zijn we voornamelijk geïnteresseerd in het creëren van voldoende spreiding in het kansverwachtingssysteem aan het aardoppervak. De verwachting is dan ook dat korte termijn kansverwachtingssystemen gebaseerd op CAPE singuliere vectoren beter presteren dan systemen gebaseerd op totale energie singuliere vectoren, maar meer onderzoek is nodig om dit te bevestigen.

DANKWOORD

Ik wil graag een aantal mensen bedanken die direct of indirect een bijdrage hebben geleverd aan de totstandkoming van dit proefschrift.
Op de eerste plaats zijn dat mijn co-promotor Jan Barkmeijer en KNMI-collega Wim Verkley. Met hen heb ik dagelijks geluncht en regelmatig overleg gehad over de inhoud van de artikelen. De vele discussies en het commentaar van beide op mijn artikelen heeft de kwaliteit van dit proefschrift aanzienlijk verbeterd. Zeker ook in de laatste fase van mijn onderzoek heb ik veel gehad aan hun goede raad. Bedankt!
Mijn promotor Henk Dijkstra is pas in een later stadium bij deze promotie betrokken. Ik wil hem graag bedanken voor het optreden als promotor en voor zijn commentaar op het proefschrift. Theo Opsteegh wil ik graag bedanken voor zijn commentaar op de artikelen.
In de beginfase van mijn AIO-schap kwam ik dagelijks bij Toon Moene langs met vragen over Fortran en het HIRLAM model. Toon bedankt voor de Fortran adviezen en het wegwijs maken in het HIRLAM model.
I would like to thank Nils Gustafsson (SMHI/Met.no), Xiaohua Yang (DMI) and Ole Vignes (Met.no) for their help and support with the implementation of the singular vector calculation in HIRLAM. The working visit to the Norwegian meteorological institute has been very helpful.
In het laatste jaar van mijn AIO-schap heb ik mijn rookpauzes kunnen delen met een goede vriend uit mijn studententijd. Kasper bedankt voor de gezellige pauzes en de vele discussies over van alles en nog wat.
Ik dank alle vrienden uit Venlo, oud-huisgenoten uit Wageningen en Utrecht en kamergenoten, collega's en mede AIO's van het KNMI voor de gezellig tijd.
Tenslotte dank ik mijn ouders, broer, zussen en de aanhang. Zonder de zorg en opvoeding van mijn ouders zou ik nooit terecht zijn gekomen waar ik nu ben. Bedankt voor alles!

Roel

CURRICULUM VITAE

Roel Stappers werd geboren op 31 mei 1976 te Horst en groeide op in het Limburgse dorpje Grubbenvorst. In 1995 haalde hij zijn VWO-diploma aan het stedelijk college Den Hulster in Venlo en begon in datzelfde jaar aan de opleiding Landbouwtechnische Wetenschappen aan de Universiteit van Wageningen. Hier specialiseerde hij zich in de meet-, regel- en systeemtechniek en studeerde in 2001 cum laude af.

Na een korte tijd als assistent in opleiding bij de vakgroep meet-, regel- en systeemtechniek volgde hij de bacheloropleiding Natuur- en Sterrenkunde en vervolgens 1 jaar de masteropleiding Theoretische Natuurkunde aan de Universiteit van Utrecht. In december 2006 begon hij met zijn promotieonderzoek aan de Universiteit van Utrecht en het KNMI dat heeft geleid tot dit proefschrift. Sinds 1 januari 2012 is hij postdoc aan de Universiteit van Reading.

ACRONYMS

3D-VAR	Three dimensional variational data assimilation
4D-VAR	Four dimensional variational data assimilation
BPDAE	Bilinear partial differential algebraic equation
AD	Adjoint
CAPE	Convective available potential energy
CIN	Convective inhibition
DA	Data assimilation
ECMWF	European centre for medium-range weather forecasts
EPS	Ensemble prediction system
GEOS	Goddard earth observing system
GLAMEPS	Grand limited area ensemble prediction system
GQ	Gaussian quadrature
HIRLAM	High resolution limited area model
L96	Lorenz 96 model
LFC	Level of free convection
NCEP	National centers for environmental prediction
NL	Nonlinear
NWP	Numerical weather prediction
PDF	Probability density function
QG	Quasi-geostrophic
RK4	Runge-Kutta order 4
RMSE	Root mean square error
SCAPE-SV	Singular vector that maximizes the mean CAPE perturbation
SREPS	Short-range ensemble prediction system
SREF	Short-range ensemble forecasting
SV	Singular vector
TL	Tangent linear
TE	Total energy