The primal and dual forms of variational data assimilation in the presence of model error

Amal EL AKKRAOUI

Doctor of Philosophy

Department of Atmospheric and Oceanic Sciences

McGill University

Montréal, Québec

January, 2010

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

ABSTRACT

The aim of data assimilation is to find the optimal estimate of the state of the atmosphere at a given time using all the available observational data and the knowledge of the physical and dynamical laws that govern the system's motion. A variety of methods are used for this purpose, and most of them are based on statistical estimation theory, such as variational methods that are widely used in operational numerical weather prediction applications. Two forms can be used to solve the variational assimilation problem: the primal (3D/4D-Var) defined in the model space and the dual (3D/4D-PSAS) defined in the "observation" space. Both variants are theoretically equivalent *at* convergence and in the linear case, are expected to have similar convergence properties. In this thesis, the equivalence is confirmed in an operational setting for the three and four dimensional cases, and the convergence properties are studied. While results at convergence confirm the theoretical equivalence, the convergence of the dual method exhibits a spurious behaviour at the beginning of the minimization which leads to less probable states than the background state, and it takes a number of iterations to retrieve states of comparable probability to that of the background state. This is worrisome since operational implementations can only afford a limited number of iterations. Investigation of this problem showed that it could be avoided by using a minimization scheme, such as the minimum residual (Minres) algorithm, that monotonically decreases the norm of the gradient instead of the functional itself. The iterates of a dual minimization with Minres lead to increasingly probable states. A relationship is established showing that the primal functional is related to the value of the dual functional and the norm of its gradient. This holds for the incremental forms of both the three and four dimensional cases.

The intercomparison of the primal and dual forms is also examined in a two dimensional weak-constraint framework to account for model errors within the assimilation system. A dual form of the weak-constraint 4D-Var is formulated and results showed that both methods converge to the same solution and with similar convergence rates. As in the three and four dimensional cases, the dual algorithm is still sensitive to the choice of the minimization algorithm, and benefits from Minres properties to avoid the non-physical increments in the first iterations.

Singular vectors of primal and dual Hessians are used to improve the preconditioning of the minimization and to establish a connection between the Hessians which is key to cycling the dual Hessian to the next analysis window. This holds also in the weak-constraint case and the significantly lower dimension of the control variable in the dual case may be beneficial then. This is an attractive proposition as the length of the assimilation window is extended.

RÉSUMÉ

L'objectif de l'assimilation de données est de trouver une estimation optimale de l'état de l'atmosphère à un moment donné en utilisant toute l'information disponible à travers et les observations et les connaissances sur les lois dynamiques et physiques qui gouvernent l'atmosphère. Différentes méthodes sont utilisées à cette fin, dont la majorité sont basées sur les principes de l'estimation statistique. Les méthodes variationnelles en sont un exemple et sont actuellement implémentées dans les grands centres de prévision numérique du temps. Deux formes peuvent être utilisées pour résoudre le problème d'assimilation variationnelle : la primale (3D/4D-Var) qui est définie dans l'espace du modèle et la duale (3D/4D-PSAS) qui est définie dans l'espace des observations. Les deux variantes sont en théorie équivalentes à la convergence et dans le cas linéaire, et sont supposées avoir un comportement de convergence similaire. Dans cette thèse, l'équivalence est confirmée dans un cadre opérationnel, et les propriétés de la convergence étudiées pour les cas tri- et quadri -dimensionnels (3D et 4D). Alors que les résultats à la convergence confirment l'équivalence théorique, la convergence de la méthode duale présente un comportement étrange pendant les premières itérations de la minimisation, ce qui produit des états moins probables que l'ébauche. et cela prend quelques iterations avant de retrouver des états dont la probabilité est comparable à celle de l'ébauche. Ce comportement est inquiétant puisque les implémentations opérationnelles ne peuvent se permettre qu'un nombre limité d'itérations. L'examen de ce problème a montré qu'il peut être évité en utilisant des schemas de minimisations tels que les méthodes à résidu minimum (Minres) qui réduisent monotoniquement la norme du gradient au lieu de la fonction objective elle même. Ainsi, les itérés de la minimisation duale avec Minres conduisent à des états de plus en plus probables. Une relation est formulées montrant que la fonction

primale est liée à la valeur de la fonction duale ainsi que celle de la norme de son gradient. Cela est valide pour les formes incrémentales dans les cas 3D et 4D.

Aussi, la comparaison entre les formes primale et duale a-t-elle été effectuée dans le cadre d'un système bi-dimensionnel en contrainte faible pour tenir compte des erreurs modèle dans le processus d'assimilation. Une forme duale du 4D-Var contrainte faible a été formulée et les résultats montrent que les deux méthodes convergent à la même solution et avec un taux de convergence similaire. Comme dans les autres cas précédents, l'algorithme dual est encore sensible au choix du minimiseur, et profite des propriétés de Minres pour éviter de produire des incréments non physiques pendant les premières itérations.

Les vecteurs singuliers des Hessiennes des méthodes primales et duales sont utilisés pour améliorer le pré-conditionnement de la minimisation et pour établir un lien entre les Hessiennes, ce qui s'avère être déterminant dans la solution de cyclage de la Hessienne duale à la fenêtre d'assimilation suivante. Cette propriété reste aussi valide dans le cas de la contrainte faible où la dimension réduite de la variable de contrôle dans le cas dual peut être bénéfique. Cela est d'autant plus intéressant que la fenêtre d'assimilation est appelée à s'élargir.

STATEMENT OF ORIGINALITY

This is an original thesis and is entirely my own work with contributions from my supervisor Pierre Gauthier and two collaborators Simon Pellerin and Samuel Buis, whose contributions are stated next in the Contributions of Authors.
This thesis contains two published papers and one chapter in preparation for a journal submission, and to the best of my knowledge, it does not infringe upon anyone's copyright. Original realizations of the thesis include:

- Remapping of dual Hessian singular vectors permitted to carry information about a current Hessian to the next analysis window. Before this thesis, cycling the dual Hessian was an open question and one major limitation of the dual method in view of its operational use.
- The extension of the theoretical equivalence of the primal and dual methods to the nonlinear case using the incremental formulation properties.
- The use of Minres as a minimization algorithm : while this was found to be a solution to the dual convergence problem, is it an original use of Minimum Residual methods for variational data assimilation. Common methods in NWP centers are the Quasi-Newton or the Conjugate Gradient. To my knowledge, no NWP center uses Minres or any other minimization technique that controls the gradient norm instead of the functional.
- The new convergence criterion for the dual minimization is key to understanding the reality of minimizing two methods in two different spaces. With the appropriate termination criterion, the dual method converges to the same accuracy as the primal one with a comparable number of iterations.

CONTRIBUTIONS OF AUTHORS

Chapters 3 and 4 consist of papers published in the Quarterly Journal of the Royal Meteorological Society and present the research I have performed for my Ph.D. The first paper is co-authored by my co-supervisor Pierre Gauthier who played a normal supervisory role and performed editing on the manuscript text, and two collaborators : Simon Pellerin who helped with the technical part of the work and his implemented version of the 3D-PSAS was used for this paper, and Samuel Buis who helped at the beginning of the process that led to the implementation of a dual version of the operational 3D-Var of Environment Canada. The second paper is co-authored by my co-supervisor, Pierre Gauthier, who played a normal supervisory role in the research and performed editing on the manuscript text. Chapter 5 is based on a manuscript soon to be submitted to the Quarterly Journal of the Royal Meteorological Society, which will also be co-authored by my co-supervisor, Pierre Gauthier.

ACKNOWLEDGMENTS

It is a pleasure to thank the many people who made this thesis possible. First I would like to thank my two supervisors, Pierre Gauthier and Peter Bartello, for their support, guidance and enthusiasm. I am grateful to them for giving me the opportunity to work on this thesis. To Pierre Gauthier I owe my deepest gratitude for the endless support he provided me with from the very first day I started working on my thesis, and to Peter Bartello I am grateful for making his support always available in a number of ways. Pierre and Peter, I am very honored for studying under your supervisions.

I also want to thank Simon Pellerin for hours of discussions which have considerably helped to put this work on track, and Samuel Buis for his supervision when I was finding my way around the PALM coupler. Many other people from the data assimilation community have been of great help to achieve this work. I am heartily thankful to them for the nice discussions, in person or by e-mail, for their advice and the support: Olivier Talagrand, Marc Buehner, Liang Xu, Chris Paige, Didier Auroux, and Stephan Laroche. In the McGill department of Atmospheric and Oceanic Sciences, useful discussions with William Sacher and Kaoshen Chung have always been of great help. I am endlessly grateful to Michel Bourqui, David Straub and Bruno Tremblay for their support in everything. For their attention and their help in IT and all administrative matters, I am very grateful to Michael Havas, the former secretary Vaughn Thomassin, may he rest in peace, and the now retired Karin Braidwood.

I would like to thank my friends at McGill for their encouragements; the great fellows, namely Louis-Philippe Nadeau, Jean-François Lemieux and Jan Sedlácêk, and my closest friends Cybelle D'Amours, Nadia Fouladi, Mounir Riday, Houda Ba-Driss,

Hafid Soualhine, Nadia Zeroual, Babak Khorsandi, Badwi Mansour and of course my dearest Jaouad Mounaam.

Lastly, and most importantly, I am deeply indebted to my close family; my parents for their endless love, for all the sacrifices they made and for supporting me all the way long even in the hardest times. I am forever indebted to my sister Ilhame and my brother Abdeljalil, their love was my motivation and their great support was invaluable. To these four wonderful people, I dedicate this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

Why have meteorologists such difficulties in predicting the weather with any certainty? Why is it that showers and even storms seem to come by chance, so that many people think it is quite natural to pray for them, though they would consider it ridiculous to ask for an eclipse by prayer? [...] a tenth of a degree more or less at any given point, and the cyclone will burst here and not there, and extend its ravages over districts that it would otherwise have spared. If they had been aware of this tenth of a degree, they could have known it beforehand, but the observations were neither sufficiently comprehensive nor sufficiently precise, and that is the reason why it all seems due to the intervention of chance.

H. Poincaré, Science et Méthode, Paris, 1908
(translated by Dover Publ., 1952)

The need for understanding the atmospheric system has long been motivated by the human desire to predict the weather. The fundamental notions of numerical weather prediction were first stated by Vilhelm Bjerknes as early as 1904, but it was not until 1922 that Lewis F. Richardson formally proposed that weather could be predicted by solving numerically the equations of the physical laws that govern the atmospheric motion. Unfortunately, his experiment aiming to produce a 6 hours forecast of surface pressure by approximating these equations using finite differences failed (Lynch, 2006). After solving numerical instability issues encountered by Richardson, it was feasible to perform a first successful numerical prediction of the weather[1] (Charney *et al.*, 1950). Since then, and with the advent of electronic computers, the accuracy of numerical weather prediction (NWP) models has improved steadily as they became more complex and more comprehensive.

[1] This was a one-day forecast using a barotropic (one-layer) model.

However, as stated in Lewis *et al.* (2006), *"the atmosphere with its governing laws based on Newton's dynamics and associated thermodynamic laws, is an unforgiving system, i.e., a system where the errors in the initial conditions grow with a doubling time of 2-3 days depending on the scale of the phenomena"*. This means that having a numerical forecast model is certainly not sufficient to produce an accurate prediction of the atmospheric future evolution. Estimating as accurately as possible the *initial conditions* is then a critical step of the NWP process, for the uncertainty in a short-term forecast is highly related to errors in the initial conditions. However, part of the forecast error may be attributed to the model itself.

Information about the state of the atmosphere is based on some a priori estimate given by climatology or a numerical weather forecast which includes implicitly the basic laws governing the motion of the atmosphere. Observations on the other hand are obtained from measurements by instruments which are sampling the atmosphere irregularly. Moreover, the physics of the measurements and the instrument means that observations are not perfect and have limited accuracy. The forecast and the observations are two sources of information that can be combined to produce an estimate of the atmospheric state that is consistent with the dynamical laws and at the same time representative of the reality as observed. This process is what is referred to as "*data assimilation*".

Schematically, this can be seen as a feedback loop: forecast - observe - correct - forecast. As explained by O'Neill *et al.* (2004), this is part of every day life:

> *"You use a kind of data assimilation scheme if you sneeze whilst driving along the motorway. As your eyes close involuntarily, you retain in your mind a picture of the road ahead and traffic nearby, as well as a mental* model *of how the car will behave in the short time before you re-open your eyes and make a course correction".*

The objective observations come from several sources such as in-situ and satellite measurements. With over 10^7 collected data a day, one would hope this is sufficient to define a meaningful atmospheric state. However, the data are generally incomplete, redundant, geographically sparse, particularly for ground-based instruments, and often only indirectly related to the model variables (as is the case for satellite radiances). Furthermore, each data source has different error characteristics that depend on the specifics of each instrument.

To obtain a complete state of the atmosphere, it is necessary to introduce some *background* information provided by a short-term weather forecast. And certainly, this prior estimate has imperfections too. Therefore, when combining all these sources of data, it is important to take into account their respective errors and characteristics. As early as the 18^{th} century, Gauss proposed to " weight the data by the reciprocal of a measure of their error". Taking into consideration all the relative uncertainties in the system, an optimal combination would let the model bring the necessary consistency to the observational data on the one hand, and on the other hand, would allow the observations to correct the trajectory of the model, keeping it *on track*. It follows that the major question that all data assimilation schemes aim to answer is:

> *"how best can we use information from the model and the various available observational data, considering their respective errors, to produce an optimal estimate of the atmospheric state that is better than model or data alone?"* (Lewis *et al.*, 2006)

In meteorology, the main application of data assimilation is in NWP, where it is used operationally to obtain a good estimate of the current atmospheric state to initiate a forecast. Data assimilation is also useful for diagnostic studies of the atmosphere, in forecast verification, and for climate studies through reanalyses.

Over the past few decades, great improvements have been achieved in NWP performance and products (analyses and forecasts). This is mainly due to a better understanding of the physics and the dynamics of the atmosphere which has led to better forecast models. New generations of fast computers permit the integration of sophisticated physics in the models at higher resolutions and advances in assimilation methods (Rabier, 2005). The assimilation is now able to use more observations, particularly from satellite instruments, to obtain a more complete observation coverage. Furthermore, assimilation methods have evolved towards a framework in which observations over a period of time can be used to produce analyses that embeds the dynamical constraints of the atmosphere. Therefore, the assimilation process establishes a bridge between modeling and observing the atmosphere.

A brief history of data assimilation methods

NWP being primarily an initial value problem, the determination of the initial conditions for a forecast model is very important. A wide variety of assimilation techniques has been developed over the last 50 years or so for this purpose (Daley, 1991), but most of them rely basically on the same statistical estimation principles to take into account the relative accuracy of the observations and the background state to *weigh* the contributions from these different sources to produce the analysis (i.e. the initial conditions for the forecast model).

The earliest attempts by Richardson (1922) and Charney *et al.* (1950) used hand interpolations of the available observations to grid points. Interest grew in performing objective analysis through interpolation methods fitting data to grid points (Charney, 1951). Panofsky (1949) is credited for pioneering the first objective analysis based on two dimensional polynomial interpolation. It was followed by Gilchrist and Cressman (1954) who put forward an interpolation scheme based on minimizing mean square differences between observations and a quadratic polynomial

representing the state field, within a radius of influence of the closest grid point. It became quickly apparent that the available data are not enough to initialize the models. The use of prior information (called background) to supplement rather insufficient data was introduced by Bergthorsson and Döös (1955), Cressman (1959), and later on by Gandin (1965). A climatological state was initially used (Gandin, 1963), but it is more common nowadays to use a previous short-range forecast as a background.

The analysis state is basically produced by correcting the background term using weighted observation increments :

$$[Analysis] = [Background] + weight \times [Observation - background]$$

Early assimilation methods such as Cressman's analysis (Cressman, 1959) and successive corrections (Bergthorsson and Döös , 1955), introduced *arbitrary* weights at each individual grid point. The statistical estimation theory, on the other hand, provided the proper framework to establish what those weights should be. Hence, the Statistical Interpolation (SI) method proposed by Gandin (1963) employs a minimum variance estimation procedure that attempts to minimize the expected analysis error variance. The optimal weights are determined from this minimization and to make the problem tractable, a localized process of data selection is used that only retains a few observations in the vicinity of each grid point. The analysis equation is solved either grid point by grid point (e.g., Bergman 1979) or in small volumes (Lorenc 1981). The data selection is invoked to reduce the quantity of observations available locally to a number sufficiently small that the computational resources can handle, and also to avoid the numerical issues associated with the inversion of large matrices.

In order to avoid data selection and the local aspect of the SI, a global approach was presented for NWP applications by Lorenc (1986), where the observations are used *globally*. This is the variational form of the statistical estimation problem.

While it is also a minimum variance technique, the variational approach defines a cost function proportional to the square root of the distance between the analysis and both the background and the observations (Sasaki, 1970). This function is then minimized directly and globally, not grid point by grid point. The three dimensional variational data assimilation (3D-Var) uses a set of observations *at a single time* to look for the model state that best fits these observations, considering their relative accuracy and that of the background state. Although 3D-Var can be equivalent to the SI under some assumptions (Lorenc, 1981; 1986), it has a major advantage over the SI in that it provides a more natural framework to use observations that are only indirectly related to the atmospheric variables. In particular, satellite instruments measure quantities like radiances at different wavelengths which could now be assimilated in their raw form (Eyal, 1987; Derber and Wu, 1998).

In the earlier days of data assimilation, observations were available mainly at the synoptic and sub-synoptic times (i.e. every 3, 6, or 12 hours). Performing the assimilation at (or around) synoptic time made sense. However, as satellite observations are available on a near-continuous basis, it became important that data assimilation techniques should be able to extract information from observations distributed in time. The *Kalman Filter*, proposed by Kalman (1960) for engineering applications, is a sequential estimation method that can assimilate observations at any given time while retaining information gained from past observations. However, when the dimension of the model state is large, its full implementation is beyond the capacity of even the most powerful computers. Approximate forms have been proposed such as the Ensemble Kalman filter (Evensen, 1994; Houtekamer and Mitchell, 1998), the Reduced Rank Kalman Filter, RRKF, (Fisher, 1998) or the SEEK filter (Verron *et al.*, 1999). All these variants have the use of the underlying framework of Gaussian probability distributions in common.

Four-dimensional variational assimilation method (4D-Var) was suggested for meteorological applications by Sasaki in his PhD thesis (1958), but it awaited the introduction of adjoint methods to be feasible and practically affordable (Lewis and Derber, 1985; LeDimet and Talagrand, 1986; Talagrand and Courtier, 1987). 4D-Var is the temporal extension of 3D-Var for observations distributed over a period of time called the assimilation window. Using the nonlinear dynamical model *within* the assimilation process, the method seeks to obtain the model *trajectory* that best fits the data over the entire assimilation window by minimization of a cost function penalizing distances from the observations and from the background. The use of the model permits capturing some of the flow dependent information in the observations and in the error statistics (Fisher, 2001); the result being an analysis state that is more consistent dynamically with the flow evolution. 4D-Var is now used operationally in major NWP centres, such as ECMWF (Rabier et *al.*, 2000; Mahfouf and Rabier, 2000, Klinker et *et al.*, 2000), Météo-France (Janisková *et al.*, 1999; Gauthier and Thépaut, 2001; Desroziers *et al.*, 2003), and recently, 4D-Var was implemented at Environment Canada (Gauthier *et al.*, 2007), the UK Met Office (Rawlins *et al.*, 2007) and the Japan Meteorological Agency (Honda *et al.*, 2005).

The variational data assimilation problem can be cast in two forms : primal and dual, both solving the same problem but in two different spaces; the model space for the former and the observation space for the latter. The 3D and 4D-Var are two primal variational schemes, whereas the dual schemes are usually referred to as Physical-space Statistical Analysis System (3D/4D-PSAS), and were first introduced for NWP applications by Cohn *et al.* (1998). Theoretical equivalence, at convergence of the minimization and in the linear case, of these two formulations was shown in Courtier (1997) who also showed that the equivalence holds when the temporal extension is introduced to obtain the 4D-Var and 4D-PSAS. The interest for the dual form was first motivated by the reduced size of the problem, since the dimension of

the dual space is commensurate with the number of observations whereas the model space has the size of the number on degrees of freedom of the model which, for meteorological and oceanic applications is one to two orders of magnitude higher that the number of observations. The dual method is also very close in its form to the approach used in Statistical Interpolation but it uses the observations globally.

Due to the inherent limit of predictability of the atmosphere, changes in the initial conditions have a limited effect on the adjustment of the forecast to observations present at later times. Moreover, deficiencies in the dynamical model are also responsible in part for the misfit between the observations and the forecast. While errors in the background and the observations are being accounted for in data assimilation, the numerical model representing the evolution of the atmospheric/oceanic flow is usually assumed *perfect*. This is the assumption used in operational implementations of four dimensional variational data assimilation (4D-Var and 4D-PSAS). To account for model errors, the constraint that the analysis is entirely determined from a model integration is relaxed. This leads to the so-called weak-constraint 4D-Var, where the analysis solution is required to satisfy the model equations only approximately, not exactly (Sasaki, 1970; Derber, 1989, Trémolet 2006), which allows the assimilation to adjust the trajectory and compensate for the part of the discrepancy to the truth that is due to the model error.

Model errors in variational data assimilation

Being based on statistical estimation, assimilation products (analysis and forecast) rely directly on our knowledge about errors affecting all information sources : the observations and the background in the first instance. If the model is assumed to be perfect, errors in the initial conditions could explain all the forecast error. This is the perfect model assumption. However, the forecast error can also be attributed to

errors in the NWP model itself, associated with approximate numerical representations of the atmosphere, they contain errors due to lack of resolution, approximations in the representation of small-scale physics or boundary conditions, or inappropriate external forcing terms, etc. These errors have long been assumed small enough compared to the other errors and could be neglected. Nevertheless, many studies have shown that these errors do affect the quality of the analysis and the forecast if they are not properly accounted for in the assimilation process (Derber, 1989; Wergen, 1992; Zupanski, 1993; Bennett *et al.*, 1993 and Zupanski,1997, Bennett, 2002).

It has been proposed that the perfect model assumption be relaxed, in that the model can be considered as *imperfect*, which allows for a formulation of the assimilation problem that takes into account the uncertainties in the model along with the uncertainties both in the observations and in the background term. This is referred to as the *weak-constraint* formulation of variational data assimilation (Sasaki, 1969, 1970; Derber, 1989). In this framework, the 4D-Var is extended to correct for an additional term representing the model error and weighted by the model error covariances (Jazwinski, 1970; Derber, 1986). While in the strong-constraint 4D-Var, the analysis trajectory (the best possible fit to the time-distributed observations) is required to satisfy *exactly* the model equations that are imposed as a strong constraint, this trajectory is allowed, in the weak-constraint 4D-Var, to deviate from the model trajectory by an amount representing the importance of the model error.

Although the basic ideas behind the weak-constraint formulation are well established, accounting for model errors is still one of the challenges that the data assimilation community is facing; for the weak-constraint formulation is computationally very demanding on the one hand (Dee, 1995), and also no general form of the model error covariances is known yet. Furthermore, while model error includes a bias component and an unbiased one, most data assimilation techniques use the

underlying assumption of unbiased error. Therefore, one of the fundamental problems of the weak-constraint formulation is how to estimate and represent the model error, for both its systematic part (bias) and its unbiased part.

In data assimilation, a good estimation of error statistics is crucial. In the case of background errors, comparison to observations is used to test the consistency of the error statistics and to perform a recalibration until consistency is achieved (Hollingsworth and Lönnberg, 1986; Desrozier *et al.*, 2005). Given the number of degrees of freedom, not enough observations are available to characterize these error statistics. Hence, modeling hypotheses are introduced to reduce significantly the number of parameters to estimate. Model errors, on the other hand, are even more complex and difficult to approach as time correlations are an additional issue that is important and cannot be discarded.

Moreover, when accounting for model errors, the size of the control variable increases by two orders of magnitude. With 10^7 degrees of freedom in current NWP models, model error statistics as described by a covariance matrix $\mathbf{Q}$ would have of the order of $10^9 \times 10^9$ elements that need to be estimated.

> *"This is more than the total number of observations of the atmosphere taken since meteorological observations started in the 1940s. Since approximately* 6×10^6 *observations are available each day, it would take 250 million years to gather as many observations as there are parameters in* $\mathbf{Q}$". Trémolet (2006)

The size of the problem makes this formulation very expensive for any NWP center to consider, even with the computer power now available. To make the problem tractable requires that the representation of model error and its statistics be simplified as is the case for background errors.

As discussed earlier, using the same principles, two approaches can be taken to solve the variational assimilation problem. Courtier (1997) showed this to be true for 3D and 3D-Var, both having their "dual" equivalent. It is also shown that this can be extended to the weak-constraint 4D-Var as well. Its dual form offers the advantage of considerably reducing the dimension of the control variable. With the numbers given by Trémolet (2006), the dual control variable would be $\approx 6 \times 10^6$ instead of 10^9. The dimension of the observation space is indeed unchanged whether the assimilation is performed in the strong or weak-constraint framework.

Overview of the thesis

As stated earlier, the motivation for this thesis is accounting for model errors in data assimilation through the weak-constraint formulation. The dual form of this variational problem seems to be appropriate but few of its properties were properly documented since it was first proposed for operational applications (Cohn *et al.*, 1998). The first implementation of a 3D-PSAS scheme was adopted by NASA's Global Modeling Assimilation Office (GMAO) and the Naval Research Laboratory (NRL) (Cohn *et al.*, 1998; Daley, 2000), and was designed in an SI-based data assimilation system to compare global vs. local data selection as performed in the SI. In that sense, it employs a representation of error covariance statistics similar to that of the SI system, and uses a global conjugate gradient solver, preconditioned by a series of smaller SI-like problems (Cohn *et al.*, 1998). In Courtier (1997), it was shown that the 3D-PSAS and its four dimensional extension 4D-PSAS are theoretically equivalent to the 3D-Var and 4D-Var respectively, under the following assumptions : the linearity of the model and the observation operator[2] , the use of the *same* covariance

[2] The observation operator defines a model equivalent of the observational state at the observation locations

matrices, and more importantly, the equivalence is only valid *at convergence* of the minimization of the objective functions. Furthermore, using their own respective preconditioning as proposed by (Amodei, 1995), both 3D/4D-Var and 3D/4D-PSAS should, in theory, converge at the same rate and with the same overall cost (Courtier, 1997). Some convergence properties of PSAS were also examined by Louvel (2001) in a controlled context with an oceanic primitive equation model (synthetic observations and identity observation operator), and Auroux (2007) presented an extension of the dual method to some nonlinear situations.

In this thesis, the dual method is studied in depth, first in an operational context aiming to examine the extent to which the theoretical equivalence between 3D/4D-Var and 3D/4D-PSAS formulations still hold in a realistic and complex framework. The focus is on their convergence properties, their preconditioning and their formulation in the nonlinear case. Then, in a simplified framework, the effect of the choice of the minimization algorithm on the behavior of the dual approach is investigated, and finally, the weak-constraint formulation is examined both in the primal and dual case.

This thesis is organized as follows. The second chapter presents the theoretical framework of data assimilation methods. The presentation focuses on the schemes discussed and studied in the following chapters. The primal and dual formulations of variational data assimilation are presented in the three dimensional case, as well as their four dimensional extensions, both in the strong and weak-constraint frameworks.

In the third chapter, an intercomparison study is conducted between the primal and dual methods in a practical and operational context. In Courtier (1997), it was shown that with proper conditioning of the minimization problem, the two algorithms should have similar convergence rates and computational performances. This chapter examines this statement in the context of the variational data assimilation system of

the Meteorological Service of Canada to show the equivalence between the 3D-Var and the 3D-PSAS algorithms. The modularity of the operators is discussed and the emphasis is on the convergence properties of the dual method. The study revealed a particular discrepancy in the argument of Courtier (1997) that made the 3D-PSAS converge more slowly. Speed-up of the convergence can be achieved by using the approximate Hessian[3] of the problem to precondition the next assimilation. In this chapter, it is shown that this can easily be done in 3D-Var but that in 3D-PSAS, the control variable's morphology changes from one assimilation to the next. A remapping method is shown to make it possible to reuse the Hessian to precondition the next analysis. This is shown to be a practical and efficient preconditioning method.

In El Akkraoui *et al.* (2008), both methods are shown to be equivalent at convergence but the dual method exhibits a spurious behavior at the beginning of the minimization which leads to less probable states than the background state. This is a serious concern when using the dual method in operational implementations when only a finite number of iterations can be afforded. In the fourth chapter, we focus on this convergence problem. The 3D-Var minimizes a cost function that can be related to the probability that a given state is the true state. As the minimization reduces this cost function, each iterate has a higher probability of being the true state than the previous one. This is not so for 3D-PSAS as shown in El Akkraoui *et al.* (2008). In this chapter, it is shown that the 3D-PSAS will lead to an increasingly more probable state in so far as the minimization reduces the norm of the gradient, instead of the dual cost function itself. This can be achieved by using minimum residual methods which ensure a monotonic decrease of the norm of the gradient. This holds for the incremental forms of both the three and four dimensional cases. A

[3] The second derivative of the cost function

new convergence criterion is introduced based on the error norm in model space to make sure that, for the dual problem, the same accuracy is obtained in the analysis when only a finite number of iterations are completed.

In the fifth chapter, we examine the weak-constraint formulation, both in the primal and dual forms. Extending the equivalence of these forms to the case where model errors are accounted for is presented, and some issues related to the preconditioning of the minimization and the sensitivity to the linearity assumption of the operators are discussed.

The concluding chapter summarizes the results and findings of this thesis along with some perspectives on future work.

2 Overview of data assimilation methods

Data assimilation is a method whereby observations are combined with an a priori estimate of the system (the background) obtained from forecasts by a numerical model to produce an *optimal* estimate, the *analysis*, of the evolving state of the system. Thus, in a data assimilation process, there are three key components: a set of observations, a dynamical model that produces a forecast, taken as an a priori estimate, and an analysis scheme.

Being based on the principles of statistical estimation, the concept of error, error estimation and error modeling is central to the concept of data assimilation. The observations have errors arising from various sources: e.g. instrumental noise, environmental noise, sampling...etc. Dynamical models are imperfect too, with errors arising from the approximate physics and parameterizations, the discretization of continuum dynamics into a numerical model, as well as the inherent forecast errors resulting from the uncertainties in the initial conditions. A good estimation of the relative uncertainty of each source of data is a prerequisite to a good analysis, the aim being that, in the assimilation process, the information content of more accurate data should be given more weight.

This chapter is a brief introduction of how the data (observations and background) are melded in the data assimilation process, how the errors are accounted for, how their statistics are estimated and how the analysis is produced. The reader is introduced to some of the data assimilation techniques that will be used in the

following chapters, along with basic concepts, terminology, and some mathematical and statistical notations that will be used in this thesis.

2.1 The basic concepts of data assimilation

Data assimilation aims to find the *optimal* model representation of the atmospheric system, using all available data, and considering what is known about their errors. The word *optimal* refers to the statistical basis of this process. As discussed in Lorenc (1988), and in Rodgers (2000), the statistical estimation problem of data assimilation can be formulated using Bayesian theory.

2.1.1 Bayesian estimation

Let $\mathbf{x}$ be the unknown state of the atmosphere to be estimated, and $\mathbf{y}$ the vector of all the observations that contains information about $\mathbf{x}$. In the Bayesian framework, the unknown $\mathbf{x}$ is treated as a *random variable*. It is also assumed that all the prior information about this unknown is summarized in a *known* prior distribution $p(\mathbf{x})$ of $\mathbf{x}$. As explained in Lewis *et al.* (2006) :

> *"It is assumed that nature picks a value of* $\mathbf{x}$ *from the distribution* $p(\mathbf{x})$ *but decides to tease us by not disclosing her choice, thereby defining a* game. *In this game, we are only allowed to observe* $\mathbf{y}$ *whose conditional distribution* $p(\mathbf{y}|\mathbf{x})$ *is known."*

The conditional probability distribution $p(\mathbf{y}|\mathbf{x})$ is the probability of the observed state $\mathbf{y}$ being *true*, given $\mathbf{x}$.

The Bayesian theory states that the probability distribution that the model state $\mathbf{x}$ is the true value of the atmospheric state, given that $\mathbf{y}$ has been observed, is expressed as :

$$p(\mathbf{x}|\mathbf{y}) \quad = \quad \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \tag{2.1}$$

where $p(\mathbf{y})$ is the marginal distribution of $\mathbf{y}$, that represents our **a priori** knowledge about the true value of the observations. It is independent of $\mathbf{x}$ and therefore plays

the role of a normalization constant. The conditional probability $p(\mathbf{x}|\mathbf{y})$ is known as the **a posteriori** distribution of $\mathbf{x}$ given the observation $\mathbf{y}$. Therefore, the most probable state is obtained by finding the value of $\mathbf{x}$ that maximizes this probability (Lewis et *al.*, 2006). This is the state that we are interested in, and it corresponds to the *mode* of $p(\mathbf{x}|\mathbf{y})$, or the maximum likelihood estimate of $\mathbf{x}$. It is called the **analysis** state, and denoted $\mathbf{x}_a$.

The objective is then to combine the information in these two *approximately known* quantities ($p(\mathbf{x})$ and $p(\mathbf{y}|\mathbf{x})$) in order to obtain the *optimal estimate* of $\mathbf{x}$. In practice, these two probability distributions are not known, and are only *estimated* based on what is known about the errors affecting each term. Although the Bayesian framework allows arbitrary probability distributions, most data assimilation schemes often assume Gaussian distributions; a debatable assumption whose validity has received increasing interest over the past few years. Note that, as in most current NWP implementations, the error distributions all through this thesis are assumed Gaussian. In this context, the analysis state $\mathbf{x}_a$ is then a *minimum variance estimate*.

a− The background state

In the context of atmospheric data assimilation, the *prior* knowledge of the state of the atmosphere is called the *background state*, and denoted $\mathbf{x}_b$. It can be derived from a climatology or a previous forecast. Ideally, the background should be the best estimate of the state of the atmosphere prior to the use of current observational data. Within the process of successive assimilation cycles, using a previous analysis state as a background provides a good estimation of the prior information about the atmosphere as depicted by the model and by past observations.

It follows that the probability $p(\mathbf{x})$ in (2.1) represents our **a priori** knowledge of the state of the atmosphere contained within the background state. In the absence of any other information, the most probable state would have to be $\mathbf{x}_b$ as $p(\mathbf{x}|\mathbf{y}) \equiv p(\mathbf{x})$

in this case.

Let $\boldsymbol{\epsilon}_b = \mathbf{x} - \mathbf{x}_b$ denote the background error. It is assumed that the background state is *unbiased*; i.e. the mean of the error is $\mathbf{E}(\boldsymbol{\epsilon}_b) = 0$, and the covariance structure is given by the symmetric positive definite matrix $\mathbf{B} = \mathbf{E}[\boldsymbol{\epsilon}_b \boldsymbol{\epsilon}_b^T]$, called the **background error covariance matrix**. The properties of $\mathbf{B}$ will be discussed later in this chapter.

Assuming Gaussian distributions of errors, one can write:

$$p(\mathbf{x}) \quad = \quad \frac{1}{c_1} exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) \right] \tag{2.2}$$

where c_1 is a normalization constant. The exponent can be seen as a measure of the distance of $\mathbf{x}$ to the background $\mathbf{x}_b$, weighted by the inverse of the background error covariances, $\mathbf{B}^{-1}$.

b– The observed state and the observation operator

The global observing system provides numerous, diversified and nearly real-time collected data for NWP operational applications. However, the spatial data coverage is very sparse and irregular. This means that not all observations are collected *at* the predetermined model grid points. Besides, some observations are only indirectly related to the variables of the dynamical model.

To compare the observed values and the model representation of the system, some processing must be done to bring these two quantities to a *common ground*. The *observation operator*, represented by a function $\boldsymbol{\mathcal{H}}$, generates from the model state the equivalent of the observations. In practice $\boldsymbol{\mathcal{H}}$ is a nonlinear collection of interpolation operators from the model discretization (the model space) to the observation points (the observation space), and conversions from model variables to the observed parameters (Lorenc 1986; Pailleux 1990). $\boldsymbol{\mathcal{H}}$ can be as simple as a spatial

interpolation of the model variable to the observation location, but in most cases, more complex transformations are needed. For instance, observation of satellite radiances implies that $\boldsymbol{\mathcal{H}}$ will include a radiative transfer model to produce radiances in different wave bands associated with a given atmospheric profile. Thus, $\boldsymbol{\mathcal{H}}$ includes the physics of the measurement and acts as the bridge that links the model variables and the observations.

The *observation error*, $\boldsymbol{\epsilon}_o$, (i.e. the departure to the true representation of the system in the observation space) is defined as : $\boldsymbol{\epsilon}_o = \mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}_t)$. It represents errors in the observation process such as errors in the instruments, in the observation operator, representativeness errors[1] ...etc). As was the case for the background error, the observation error is assumed unbiased, $\mathbf{E}(\boldsymbol{\epsilon}_o) = 0$, and the covariances are represented by the **observation error covariance matrix**: $\mathbf{R} = \mathbf{E}[\boldsymbol{\epsilon}_o \boldsymbol{\epsilon}_o^T]$. Note that $\mathbf{R}$ is by construction a symmetric positive definite matrix defined in observation space. In practice, the observation error covariance matrix is obtained from instrument error estimates which, if independent means that $\mathbf{R}$ is often (but not always) taken as diagonal.

Using these notations, the conditional probability (in 2.1) of the observed state $\mathbf{y}$ being true, given the model state is $\mathbf{x} = \mathbf{x}_t$ can now be written as

$$p(\mathbf{y}|\mathbf{x}) \quad = \quad \frac{1}{c_2} exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}))\right] \tag{2.3}$$

where c_2 is a normalization constant. As in the case of the background state, the exponent can be seen as a measure of the distance, in observation space, of the state

[1] Since observations are point located, they may not be representative of the actual state of the atmosphere on the scale of the assimilation model (Swinbank *et al.*, 2003). Representativeness error is often associated with the subgrid scale variability.

$\mathbf{x}$ to the observations $\mathbf{y}$, weighted by the inverse of the observation error covariances, $\mathbf{R}^{-1}$.

In the Gaussian case, and using (2.2) and (2.3), (2.1) can be re-written as

$$p(\mathbf{x}|\mathbf{y}) = \frac{1}{c'} exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) - \frac{1}{2}(\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x}))\right] \quad (2.4)$$

where $c' = c_1 c_2 p(\mathbf{y})$ is independent of $\mathbf{x}$.

The mode of $p(\mathbf{x}|\mathbf{y})$ is the model state $\mathbf{x}$ with the maximum probability. It can be obtained by minimizing the objective functional defined by

$$J(\mathbf{x}) = -ln(p(\mathbf{x}|\mathbf{y})) \quad (2.5)$$

Therefore, the maximum likelihood estimate of the state of the atmosphere is obtained by minimizing the functional

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}[\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x})]^T\mathbf{R}^{-1}[\mathbf{y}-\boldsymbol{\mathcal{H}}(\mathbf{x})] \quad (2.6)$$

which can be seen as a measure of the combined distances of the model state to the observations and to the background state, weighted by the inverse of their respective error covariances. Through the minimization process, a fit of the model state to the observations is realized: the more accurate the observation, the closer the fit to the observations will be.

2.1.2 Modeling and estimation of the background error covariance matrix

Statistical estimation relies on the error statistics defined by $\mathbf{B}$ and $\mathbf{R}$. The background error covariance matrix is a key element of any assimilation system as it controls the way information is spread out to ensure that observations of one model variable produce dynamically consistent corrections in the other model variables.

Despite its importance, however, the way in which the background error covariance matrix is modeled in NWP assimilation systems is dominated by the compromises that must be made. This is due to the difficulties to estimate it fully as its size exceeds the volume of available observations. The state vector of a typical analysis system for NWP has a dimension around 10^7, with a total number of observations of the order of 10^6. Consequently, the background error covariance matrix contains roughly 10^{14} elements. There are simply too many unknown elements to estimate.

a- Modeling of B:

The error affecting a given term in the assimilation system ($\mathbf{x}_b$, say) is theoretically defined as the departure from its *true* counterpart ($\mathbf{x}_t$) : (i.e $\boldsymbol{\epsilon}_b = \mathbf{x}_t - \mathbf{x}_b$). In practice, the truth is unknown, and the error can only be estimated from a large number of realizations, the statistics of which are expected to converge to values that characterize the errors themselves.

As stated earlier, the background errors are the result of error growth in a short term forecast associated with the *initial uncertainty* in the initial conditions. In practice, given uncertainties of all *inputs* of the model, an ensemble of forecast at a given lead time ($\sim 6h$) is obtained each having errors on the model variables. Since the errors are assumed ergodic, each forecast difference can be interpreted as a realization of the stochastic process. The background error is then defined as the statistical mean of the second moments, assuming that the average error is removed.

$$< (\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T > \approx \mathbf{B} \tag{2.7}$$

For instance, for a model state defined as $\mathbf{x} = (\mathbf{u}^T, \mathbf{v}^T, \mathbf{T}^T)$, where $\mathbf{u}$ and $\mathbf{v}$ are the components of the wind field and $\mathbf{T}$ is the temperature field, the background

error covariance matrix is expressed as

$$\mathbf{B} = \begin{bmatrix} < \boldsymbol{\epsilon}_\mathbf{u} \boldsymbol{\epsilon}_\mathbf{u}^T > & < \boldsymbol{\epsilon}_\mathbf{u} \boldsymbol{\epsilon}_\mathbf{v}^T > & < \boldsymbol{\epsilon}_\mathbf{u} \boldsymbol{\epsilon}_\mathbf{T}^T > \\ < \boldsymbol{\epsilon}_\mathbf{v} \boldsymbol{\epsilon}_\mathbf{u}^T > & < \boldsymbol{\epsilon}_\mathbf{v} \boldsymbol{\epsilon}_\mathbf{v}^T > & < \boldsymbol{\epsilon}_\mathbf{v} \boldsymbol{\epsilon}_\mathbf{T}^T > \\ < \boldsymbol{\epsilon}_\mathbf{T} \boldsymbol{\epsilon}_\mathbf{u}^T > & < \boldsymbol{\epsilon}_\mathbf{T} \boldsymbol{\epsilon}_\mathbf{v}^T > & < \boldsymbol{\epsilon}_\mathbf{T} \boldsymbol{\epsilon}_\mathbf{T}^T > \end{bmatrix}$$

where the brackets represent a statistical mean.

The diagonal elements of the matrix contain variances, for each variable of the model, and the off-diagonal terms are the cross-covariances. This globally establishes a connection between any field variable at any given location (grid point) to field variables at other locations. That is, the underlying dynamics create a coupling between the error of model variables at different locations. These couplings are important as they determine how information from observations is spread spatially in a consistent way. For instance, observations of the wind field will provide information about the mass field due to the geostrophic balance constraint, and vice versa (Parrish and Derber, 1992; Derber and Bouttier, 1999). The background error covariance matrix is thus a crucially important quantity in any data assimilation system.

The size of the matrix makes it necessary to simplify the parameterize of the error statistics. A covariance model is used to reduce the number of degrees of freedom. Hollingsworth and Lönnberg (1986) assumed background error correlations to be homogeneous and isotropic while variances varied in the vertical and with latitude. Multivariate analysis assumed dynamical relationships between mass and winds and the "unbalanced" components are assumed to be totally uncorrelated (Parish and Derber, 1992; Derber and Bouttier, 1999; Gauthier *et al.*, 1998).

b- Estimation of the error statistics

Many methods have been proposed to estimate the background error statistics, but the focus here is only on those describing the stationary statistics of the background error. The "NMC" (now NCEP) method is based on differences between

forecasts ($\mathbf{x}_f$) of different lead times which verify at the same time (Parrish and Derber, 1992; Rabier *et al.*,1998). Lagged forecasts are taken as representative of forecast error caused by errors in the initial conditions. Typically, pairs of forecasts whose lengths differ by 24 hours are used, since this minimizes the chance of aspects of the model's diurnal cycle being incorrectly interpreted as background errors. The background error covariance matrix is then expressed as $\mathbf{B} \approx < \boldsymbol{\epsilon}_f \boldsymbol{\epsilon}_f^T >$, where $\boldsymbol{\epsilon}_f$ represents the forecast differences (e.g. $\boldsymbol{\epsilon}_f = (\mathbf{x}_{f(48h)} - \mathbf{x}_{f(24h)}) - \overline{(\mathbf{x}_{f(48h)} - \mathbf{x}_{f(24h)})}$). This gives the structure of the background errors and the amplitude is rescaled with respect to innovations to get proper representation of 6-h forecast error characteristics.

Another method to estimate the background error statistics is described in Houtekamer *et al.* (1996), and Fisher (2003a and 2003b), in which an ensemble of forecasts obtained by perturbing the observations in the analysis system is used. As explained in Fisher (2003b), the resulting analysis will be perturbed by an amount drawn from the distribution of analysis error. Besides, the background state will also be perturbed. By neglecting the effects of model error, the perturbation to the forecast will have the statistical characteristics of the background error.

2.2 Variational Data Assimilation

In variational data assimilation, the state of the atmosphere *at a given time* is estimated by minimizing an appropriate objective function (as is 2.6). In the Gaussian case, its minimum can be seen as the maximum likelihood estimate of the state of the atmosphere at the analysis time.

2.2.1 Three-dimensional variational assimilation: 3D-Var

The three dimensional variational assimilation scheme (3D-Var) solves the statistical estimation problem by minimizing the functional (2.6). The data assimilation problem is then a minimization of the least square error in the background and the

observations, all weighted by the inverse of the covariance matrices, over the assimilation period. The weight is inversely proportional to the magnitude of the errors affecting each term.

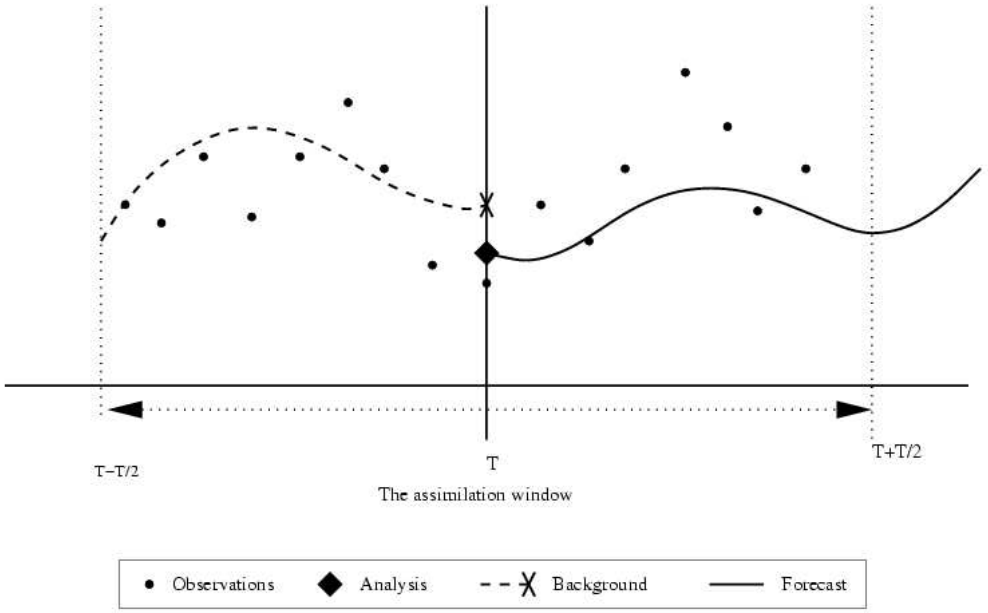


Figure 2–1: Schematic of the 3D assimilation. The analysis state is found by enriching the background state with information from the observations assumed to be all taken at the analysis time T.

The state vector can be written in terms of an increment (or a perturbation), $\boldsymbol{\delta}\mathbf{x}$, representing the misfit to the background state : $\boldsymbol{\delta}\mathbf{x} = \mathbf{x} - \mathbf{x}_b$. Using first order Taylor series, the nonlinear observation operator $\boldsymbol{\mathcal{H}}$ can be linearized in the vicinity of the background state such that

$$\boldsymbol{\mathcal{H}}(\mathbf{x}) \approx \boldsymbol{\mathcal{H}}(\mathbf{x}_b) + \mathbf{H}\delta\mathbf{x}$$

where $\mathbf{H}$ is the Jacobian of $\boldsymbol{\mathcal{H}}$. In 3D-Var, this has been found to be a relatively good approximation in so far as the linearization is made around a current estimate

of the state of the atmosphere, which ensures that the increment is small enough for the approximation to hold.

By introducing the *innovation vector* (i.e. the misfit between the observation state $\mathbf{y}$ and the background state at the observation locations), $\mathbf{y}' = \mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}_b)$, and re-arranging terms, the objective function of 3D-Var in (2.6) will now be defined with respect to the increment $\boldsymbol{\delta}\mathbf{x}$ as

$$J(\boldsymbol{\delta}\mathbf{x}) = \frac{1}{2}\boldsymbol{\delta}\mathbf{x}^T\mathbf{B}^{-1}\boldsymbol{\delta}\mathbf{x} + \frac{1}{2}(\mathbf{H}\boldsymbol{\delta}\mathbf{x} - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}\boldsymbol{\delta}\mathbf{x} - \mathbf{y}'). \tag{2.8}$$

This is the incremental[2] formulation of 3D-Var. The minimum of J is found by setting the gradient to zero, and therefore,

$$\begin{aligned} \boldsymbol{\nabla}_{\boldsymbol{\delta}_\mathbf{x}} J = & \left(\frac{\partial J}{\boldsymbol{\delta}\mathbf{x}}\right)^T \\ = & \mathbf{B}^{-1}\boldsymbol{\delta}\mathbf{x} + \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\boldsymbol{\delta}\mathbf{x} - \mathbf{y}') \\ = & (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\boldsymbol{\delta}\mathbf{x} - \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}' \\ = & 0 \end{aligned} \tag{2.9}$$

Two approaches can be used to solve $\boldsymbol{\nabla}_{\boldsymbol{\delta}_\mathbf{x}} J = 0$. From (2.9), one can directly formulate the analysis increment

$$\begin{aligned} \boldsymbol{\delta}\mathbf{x}_a = & \left[\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}' \\ = & \mathbf{K}\mathbf{y}' \end{aligned} \tag{2.10}$$

where $\mathbf{K}$ is the *gain* matrix and, in the linear and Gaussian case, can be written in two forms

$$\begin{aligned} \mathbf{K} = & \left[\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^T\mathbf{R}^{-1} \\ = & \mathbf{B}\mathbf{H}^T\left[\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T\right]^{-1} \end{aligned} \tag{2.11}$$

[2] The need for an incrementation formulation will be discussed when presenting the 4D-Var.

Given the large dimensions of the matrices involved, and the difficulty to calculate the inverse of a large matrix, $\mathbf{K}$ cannot be computed explicitly. In practice, finding the zero of the gradient can be approached as solving a linear system $\mathbf{A}\mathbf{x}-\mathbf{b}=0$, where $\mathbf{A}$ in this case is the Hessian (i.e., the second derivative) of the objective functional $\mathbf{A}=\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$, and $\mathbf{b}=\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}'$. Iterative minimization algorithms (such as the Limited-Memory Quasi-Newton or the Conjugate Gradient) can then be used to approach the minimum iteratively. The minimization can be stopped by limiting arbitrarily the number of iterations, or by requiring that the norm of the gradient decreases by a predefined amount during the minimization, which is an intrinsic measure of how much the analysis is closer to the optimum than the initial point of the minimization.

To speed up the convergence of the minimization, a change of variable $\mathbf{v}=\mathbf{B}^{-\frac{1}{2}}\delta\mathbf{x}$ is introduced (Lorenc, 1988; Gollub and Van Loan, 1996) as a preconditioner. The cost function becomes

$$J(\mathbf{v})=\frac{1}{2}\mathbf{v}^T\mathbf{v}+\frac{1}{2}(\mathbf{H}\mathbf{B}^{\frac{1}{2}}\mathbf{v}-\mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{B}^{\frac{1}{2}}\mathbf{v}-\mathbf{y}') \tag{2.12}$$

whose gradient is obtained by differentiating J with respect to $\mathbf{v}$

$$\boldsymbol{\nabla}_{\mathbf{v}}J=(\mathbf{I}_n+\mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{\frac{1}{2}})\mathbf{v}-\mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}' \tag{2.13}$$

The second derivative of J is the Hessian

$$\mathbf{J}''=\mathbf{I}_n+\mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{\frac{1}{2}} \tag{2.14}$$

whose diagonal elements are largely dominated by the identity matrix term $\mathbf{I}_n$ (where n is the size of the model space) when few observations are present (Fisher and Andersson, 2001). With increasing number of observations, this preconditioning is degrading its efficiency and approximate forms of the Hessian may be used to improve it.

Once the approximate solution $\mathbf{v}_a$ has been found by the minimization algorithm, the analysis increment and the analysis itself are obtained from

$$\boldsymbol{\delta}\mathbf{x}_a = \mathbf{B}^{\frac{1}{2}}\mathbf{v}_a \quad \text{and} \quad \mathbf{x}_a = \mathbf{x}_b + \boldsymbol{\delta}\mathbf{x}_a$$

Note that the preconditioning is an important step based on information available beforehand. A well conditioned problem will converge faster to the expected solution. The properties of the Hessian are critical in this matter. The condition number[3] of this matrix will determine the rate of convergence; the closer to 1, the faster the convergence. Thus, the preconditioning step aims to reduce the condition number of the Hessian. This will be discussed further in the next chapter.

The Hessian $\mathbf{J}''$ can also be related to the analysis error covariance matrix ($\mathbf{P_a} = \mathbf{E}[\boldsymbol{\epsilon}_a\boldsymbol{\epsilon}_a^T]$) in that it is its inverse

$$\mathbf{P}_a = \mathbf{J}''^{-1}. \tag{2.15}$$

This can simply be found by noticing that, from (2.10), the general form of the analysis state, and the analysis error are respectively $\mathbf{x}_a = \mathbf{K}\mathbf{y} + (\mathbf{I}_n - \mathbf{K}\mathbf{H})\mathbf{x}_b$, and $\boldsymbol{\epsilon}_a = \mathbf{K}\boldsymbol{\epsilon}_o + (\mathbf{I}_n - \mathbf{K}\mathbf{H})\boldsymbol{\epsilon}_b$. Using the definition of the gain matrix $\mathbf{K}$ and substituting terms, yields immediately $\mathbf{P}_a = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$.

Equation (2.15) is only true in the present case where the observation operator is linear and when error statistics are Gaussian, and under the assumption that background and observation errors are uncorrelated (Rabier and Courtier, 1992; Fisher and Courtier, 1995). This property ensures that minimizing J is a way of reducing the total analysis error variance and referring to (2.5) that the minimum is a more

[3] The condition number of a symmetric positive definite matrix is the ratio of the largest to the smallest eigenvalue of this matrix.

probable state since the analysis error is smaller than both the background and observation error.

At convergence, the 3D-Var gives an analysis that is the most probable state, according to all the information given by the observations, the background, and their assumed error statistics. Moreover, the smoothing properties of $\mathbf{B}$ ensure that the information in the analysis increments takes into account the balance properties of the model variables, as introduced in the formulation of $\mathbf{B}$.

However, the 3D-Var method presents some limitations. As shown in fig.2–1, all the observations collected within the assimilation window (usually a time interval of $[-3h, +3h]$) are assumed to be taken and valid *at the analysis time*. This can be problematic, for instance, for rapidly evolving atmospheric situations, where an early (late) observation may not be very representative of the atmospheric situation three hours later (before). To circumvent this limitation, a slightly modified version of the 3D-Var can be used, in which the innovations are obtained by comparing the observations against the forecast valid at the observation time. This approach is referred to as the First Guess at Appropriate Time (3D-FGAT) and can be seen as a half-way scheme between using all the observations at one fixed time as in 3D-Var, and properly accounting for the time dimension of the observations and of the dynamical reality as in 4D-Var. This will be discussed next.

2.2.2 Four-dimensional variational assimilation: 4D-Var

Unlike the 3D-Var, where no proper account is made of the observation time, 4D-Var is presented as a temporal extension of the 3D-Var for observations that are distributed in time. The basic concept is the same as 3D-Var provided that a nonlinear forecast model $\boldsymbol{\mathcal{M}}$ is used as part of the observation operator.

As shown in Fig.2–2, time integrations of the model $\boldsymbol{\mathcal{M}}$ are used to provide the model state at the time of the observations. More precisely, at each time step t_i, the

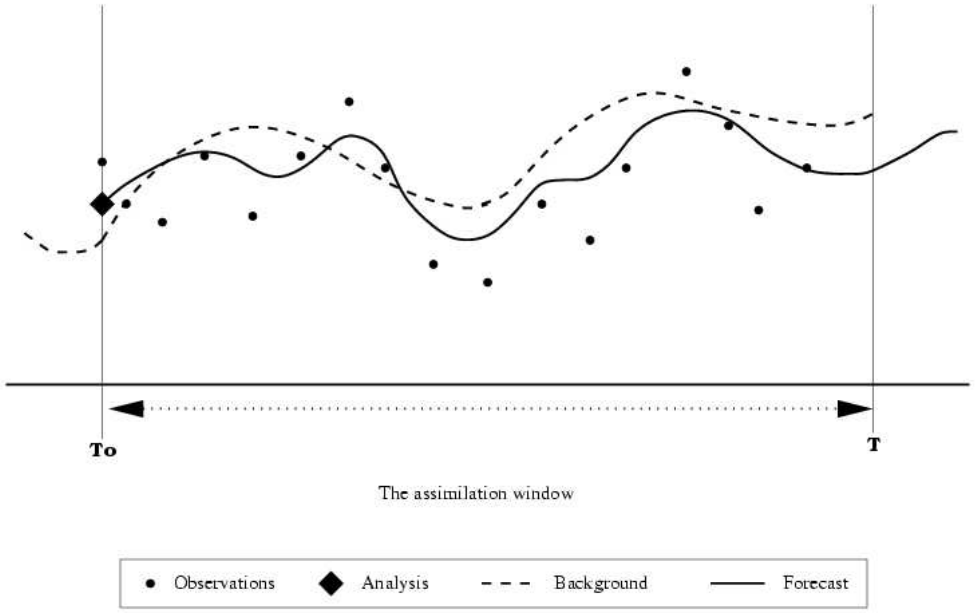


Figure 2–2: Schematic of the 4D-Var assimilation. The model trajectory is corrected to fit the time-distributed observations over the entire assimilation window.

model state $\mathbf{x}_i$ is related to the initial conditions by

$$\mathbf{x}_i = \boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0) \tag{2.16}$$

where $\boldsymbol{\mathcal{M}}_{i,0}$ is a forward integration of the model from time t_0 to time t_i. In the standard formulation of 4D-Var (Lewis and Derber, 1985; Le Dimet and Talagrand, 1986), the solution sought is the "trajectory" of the dynamical model that best fits a series of time-distributed observations $\{\mathbf{y}_i, i = 0..q\}$ over a time window (of 6h, 12h or 24h). The objective function of 4D-Var does not vary much from that of 3D-Var in (2.6):

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{i=0}^{q}(\boldsymbol{\mathcal{H}}_i(\mathbf{x}_i) - \mathbf{y}^0{}_i)^T \mathbf{R}_i{}^{-1}(\boldsymbol{\mathcal{H}}_i(\mathbf{x}_i) - \mathbf{y}^0{}_i) \tag{2.17}$$

where here the nonlinear observation operator as well as the observation error covariance matrix are defined for each time step t_i as $\boldsymbol{\mathcal{H}}_i$ and $\mathbf{R}_i$ respectively.

Substituting the dynamical constraint (2.16) into the objective function, the control variable (the trajectory $\mathbf{x}(t)$) is entirely defined by the initial conditions $\mathbf{x}_0$. This is consistent with the "perfect" model assumption used in this context; that is, the analysis state at the initial time can be integrated with the dynamical model to find the optimal analysis "trajectory". The minimization of the objective function

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2} \sum_{i=0}^{q} \left[\boldsymbol{\mathcal{H}}_i(\boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0)) - \mathbf{y}^0{}_i \right]^T \mathbf{R}_i{}^{-1} \left[\boldsymbol{\mathcal{H}}_i(\boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0)) - \mathbf{y}^0{}_i \right] \tag{2.18}$$

requires the computation of its gradient. From variational calculus, and using the increment $\boldsymbol{\delta}\mathbf{x}_0$ defined as previously ($\boldsymbol{\delta}\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_b$), one can write

$$\begin{aligned} \delta J(\mathbf{x}_0) = &\quad J(\mathbf{x}_0) - J(\mathbf{x}_0 - \delta\mathbf{x}_0) \\ \approx &\quad \boldsymbol{\delta}\mathbf{x}_0^T \boldsymbol{\nabla} J(\mathbf{x}_0) \end{aligned} \tag{2.19}$$

An expansion of $J(\mathbf{x}_0 - \delta\mathbf{x}_0)$ in a Taylor series is then performed, and only the leading terms are kept. For that, one needs to define a first-order approximation to the evolution of perturbations to the nonlinear forecast trajectory. This is given by the Tangent Linear Model (TLM) which is a linearized version of the nonlinear model dynamics (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Courtier *et al.*, 1994). The linearization is performed around the current background state, so that the perturbations evolve as

$$\mathbf{x}_i = \boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0) \approx \boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_b) + \mathbf{M}_{i,0} \boldsymbol{\delta}\mathbf{x}_0 \tag{2.20}$$

where $\mathbf{M}_{i,0}$ is a linear operator that transports the initial conditions $\boldsymbol{\delta}\mathbf{x}_0$ of the perturbation to time t_i, and is referred to as the *propagator* of the TLM. $\mathbf{M}_{i,0}$ is such that

$$\mathbf{M}_{i,0} = \mathbf{M}_{i,i-1} \cdots \mathbf{M}_{2,1} \mathbf{M}_{1,0} \tag{2.21}$$

where $\mathbf{M}_{i,i-1}$ describes the operator associated with the completion of a single time step of the TLM. The adjoint, $\mathbf{M}^T$, of the TLM corresponds to the transpose of the $\mathbf{M}_{i,0}$, and has a very useful property; it can be used as a "combination" of linear operators:

$$\mathbf{M}_{i,0}^T = \mathbf{M}_{1,0}^T \mathbf{M}_{2,1}^T \cdots \mathbf{M}_{i,i-1}^T \tag{2.22}$$

where $\mathbf{M}_{i,i-1}^T$ can be seen as a backward integration in time, from t_i to t_{i-1}. It follows that (2.19) can be written as

$$\delta J(\mathbf{x}_0) = \boldsymbol{\delta}\mathbf{x}_0^T \left[\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \sum_{i=0}^{q} \mathbf{M}_{i,0}^T \mathbf{H}_i^T {\mathbf{R}_i}^{-1} \left[\boldsymbol{\mathcal{H}}_i(\boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0)) - {\mathbf{y}^0}_i \right] \right] \tag{2.23}$$

where as before, $\mathbf{H}_i^T$ is the transpose of the Jacobian of the non linear observation operator. Using (2.19) and (2.23), the gradient of the objective function in (2.18) is then

$$\boldsymbol{\nabla}_{\mathbf{x}_0} J = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \sum_{i=0}^{q} \mathbf{M}_{i,0}^T \mathbf{H}_i^T {\mathbf{R}_i}^{-1} \left[\boldsymbol{\mathcal{H}}_i(\boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0)) - {\mathbf{y}^0}_i \right]. \tag{2.24}$$

Therefore, one integration of the full nonlinear model and its adjoint are required for each iteration of the minimization process. This is computationally expensive, since the total cost of the minimization can be evaluated as the cost of an integration of the nonlinear model for approximately 30 days. The incremental formulation introduced earlier in the case of 3D-Var, takes all its meaning in the 4D-Var context, it reduces the cost of the 4D-Var minimization since it uses linearized model integrations (Courtier *et al.*, 1994).

The incremental form of the objective functional and its gradient can then be written as:

$$J(\boldsymbol{\delta}\mathbf{x}_0) = \frac{1}{2} {\boldsymbol{\delta}\mathbf{x}_0}^T \mathbf{B}^{-1} \boldsymbol{\delta}\mathbf{x}_0 + \frac{1}{2} \sum_{i=0}^{q} \left[\mathbf{H}_i \mathbf{M}_{i,0} \boldsymbol{\delta}\mathbf{x}_0 - \mathbf{y}_i' \right]^T {\mathbf{R}_i}^{-1} \left[\mathbf{H}_i \mathbf{M}_{i,0} \boldsymbol{\delta}\mathbf{x}_0 - \mathbf{y}_i' \right] \tag{2.25}$$

$$\nabla_{\boldsymbol{\delta}\mathbf{x}_0} = \mathbf{B}^{-1}\delta\mathbf{x}_0 + \sum_{i=0}^{q} \mathbf{M}_{i,0}^T \mathbf{H}_i^T {\mathbf{R}_i}^{-1} \left[\mathbf{H}_i \mathbf{M}_{i,0} \boldsymbol{\delta}\mathbf{x}_0 - {\mathbf{y}'}_i\right] \tag{2.26}$$

where the innovation vector is defined, at each time step t_i, as $\mathbf{y}'_i = \mathbf{y}_i - \boldsymbol{\mathcal{H}}_i(\boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_b))$.

For the large-scale dynamics, 4D-Var adjusts the more energetic large-scale components first (Thépaut and Courtier, 1991; Tanguay *et al.*, 1995; Laroche and Gauthier, 1998), the Tangent Linear Model (TLM) and its adjoint can then be used at a coarser resolution with simplified physical parameterizations instead of the fully nonlinear model (Courtier *et al.*, 1994). The simplifications need to provide a reasonable approximation for the evolution of perturbations over a short period of time when compared to what can be obtained with the full model.

Using the TLM and its adjoint in the incremental formulation instead of the non linear model, the cost of one model integration is significantly reduced and the 4D-Var problem became tractable and could be used for operational NWP applications. However, another limitation arose: the accuracy of the analysis state is deteriorated by the linearizations of $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{M}}$. As the TLM is a linearization around a model trajectory, it has nonlinearities as the analysis increment $\boldsymbol{\delta}\mathbf{x}_0^a \equiv \boldsymbol{\delta}\mathbf{x}_0$ is of finite size. As discussed in Courtier *et al.* (1994), outer iterations are introduced by integrating the full nonlinear model with the initial conditions $\mathbf{x}_0^{(1)} = \mathbf{x}_b + \boldsymbol{\delta}\mathbf{x}_0^{(0)}$. The comparison to observations is reevaluated and the linearization $\mathbf{M}_{i,0}^{(1)}$ is refreshed. The minimization of the full nonlinear problem is then carried out through series of linear approximations, which is akin to what is often done to solve nonlinear systems of equations. Gauthier *et al.* (2007) describe how this was implemented in the operational 4D-Var assimilation system at the Meteorological Service of Canada.

As mentioned before, $\mathbf{B}$ is the background error covariance matrix at the initial time t_0. Fisher (2001) pointed out that in the 4D-Var case, the forecast error covariance matrix is expressed at each time t_i as $\mathbf{M}_{i,0}\mathbf{B}\mathbf{M}_{i,0}^T$ (see the discussion in Appendix D of Fisher and Andersson, 2001). Thus, the covariance matrix is implicitly evolved in 4D-Var using the dynamics of the TLM. As a consequence, both the

covariance matrix at the observation time and the analysis increments become flow dependent. Note also that for observations at the initial time t_0, the analysis increments are the same as for 3D-Var ($\mathbf{M}_{0,0} \equiv \mathbf{I}_n$). Therefore, at each assimilation cycle, the **initial** covariance matrix is a static flow-independent matrix $\mathbf{B}$. That is, the flow-dependent covariances are not propagated to the next cycle (Fisher, 2001). The cycling of 4D-Var can be done by using simplified forms of the Kalman filter such as the Reduced Rank Kalman Filter (Fisher, 1998; Fisher and Andersson, 2001). More recently, ensemble approaches have been used to provide flow-dependent background error covariances to 4D-Var (Berre *et al.*, 2009).

Following Courtier (1997), a compact form of (2.25) can be introduced as:

$$J(\boldsymbol{\delta}\mathbf{x}_0) = \frac{1}{2}\boldsymbol{\delta}\mathbf{x}_0^T\mathbf{B}^{-1}\boldsymbol{\delta}\mathbf{x}_0 + \frac{1}{2}(\mathbf{G}\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{G}\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{y}') \tag{2.27}$$

with $\mathbf{y}' = \begin{pmatrix} \mathbf{y}'_0 \\ \vdots \\ \mathbf{y}'_q \end{pmatrix}$, the new form of the observation operator is $\mathbf{G} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_{1,0} \\ \vdots \\ \mathbf{H}_i\mathbf{M}_{i,0} \\ \vdots \\ \mathbf{H}_q\mathbf{M}_{q,0} \end{pmatrix}$, and $\mathbf{R}$ is the block diagonal matrix of the $\mathbf{R}_i$. This compact form will be useful when introducing the physical-space and the weak-constraint formulations in the following sections.

The same preconditioning as in the 3D case can be used here, by introducing the change of variable $\mathbf{v} = \mathbf{B}^{-\frac{1}{2}}\boldsymbol{\delta}\mathbf{x}$, and the functional then becomes

$$J(\mathbf{v}) = \frac{\mathbf{1}}{\mathbf{2}}\mathbf{v}^{\mathbf{T}}\mathbf{v} + \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{G}\mathbf{B}^{\frac{\mathbf{1}}{\mathbf{2}}}\mathbf{v} - \mathbf{y}')^{\mathbf{T}}\mathbf{R}^{-\mathbf{1}}(\mathbf{G}\mathbf{B}^{\frac{\mathbf{1}}{\mathbf{2}}}\mathbf{v} - \mathbf{y}') \tag{2.28}$$

However, the Hessian being $\mathbf{J}" = \mathbf{B}^{-1} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}$, preconditioning with respect to $\mathbf{B}$ alone may not be as efficient as the 3D case (Andersson *et al.*, 2000).

One can then see that the operator $\mathbf{G}$ embeds most of the differences between the 4D-Var and 3D-Var objective functionals in (2.28) and (**??**). As explained earlier, the 4D-Var scheme is a generalized form of the 3D-Var; the time dimension and the flow dependencies being the major improvements to the three dimensional method. The first operational implementation of a 4D-Var system was completed in 1997 at ECMWF (Rabier et *al.*, 2000; Mahfouf and Rabier, 2000, Klinker et *et al.*, 2000). Another similar system was implemented at Météo-France in 2000 (Janisková *et al.*, 1999; Gauthier and Thépaut, 2001; Desroziers *et al.*, 2003). Recently, 4D-Var was also implemented at Environment Canada (Gauthier *et al.*, 2007), the UK Met Office (Rawlins *et al.*, 2007) and the Japan meteorological Agency (Honda *et al.*, 2005). The impact of adopting 4D-Var was qualified as substantial, resulting in improvements in NWP quality and accuracy (Rabier, 2005), which was mainly due to better consistency between the 4D-Var analyses and the dynamics of the atmosphere (Talagrand, 2003).

2.3 Variational Data Assimilation techniques: the dual form

2.3.1 3D-PSAS

The dual method is introduced here following Da Silva *et al.*(1995) and Courtier (1997). The minimum of the objective function of 3D-Var, J, corresponds to the analysis increment $\boldsymbol{\delta}\mathbf{x}_a$, which is written as

$$\boldsymbol{\delta}\mathbf{x}_a = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\mathbf{y}' \tag{2.29}$$

and can be found by minimizing (2.8) using descent algorithms. Another approach is to use a two-step algorithm to solve (2.29). This is the so-called dual approach, commonly called the PSAS (Physical-space Statistical Analysis System). Introducing $\mathbf{w}_a$ such that

$$\mathbf{w}_a = (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\mathbf{y}' \tag{2.30}$$

(2.29) becomes

$$\boldsymbol{\delta}\mathbf{x}_a = \mathbf{B}\mathbf{H}^T\mathbf{w}_a. \tag{2.31}$$

Moreover, to find $\mathbf{w}_a$, the auxiliary objective function

$$F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)\mathbf{w} - \mathbf{w}^T\mathbf{y}' \tag{2.32}$$

is minimized. The advantage of this dual form is that the dimension of the control variable, $\mathbf{w}$, corresponds to the number of observations. It was designed to solve the same data assimilation problem as the primal one, but in *observation space* (or physical space) rather than in *model space*. Therefore, it takes advantage of the fact that the number of observations used in atmospheric and oceanic data analysis is usually one order of magnitude less than the degrees of freedom of most NWP models to carry out the minimization in a smaller space than 3D-Var (Courtier, 1997).

The term $\mathbf{w}_a$ can also be seen as the increment term in observation space, and $\mathbf{B}\mathbf{H}^T$ as the representer term that maps the increments from physical to model space (Bennett *et al.*, 1993). Following Amodei (1995), a preconditioning with $\mathbf{R}^{-\frac{1}{2}}$ can be used by considering the change of variable $\mathbf{u} = \mathbf{R}^{\frac{1}{2}}\mathbf{w}$. The objective function and its gradient become:

$$F(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T(\mathbf{I}_m + \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}})\mathbf{u} - \mathbf{u}^T\mathbf{R}^{-\frac{1}{2}}\mathbf{y}' \tag{2.33}$$

$$\boldsymbol{\nabla}_{\mathbf{u}}F = (\mathbf{I}_m + \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}})\mathbf{u} - \mathbf{R}^{-\frac{1}{2}}\mathbf{y}' \tag{2.34}$$

so that the Hessian can be written as

$$\mathbf{F}"_u = \mathbf{I}_m + \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}} \tag{2.35}$$

The Physical-space Statistical Analysis System (PSAS) was developed and implemented first at the NASA's Data Assimilation Office (DAO)[4] in the late 90's, to replace the Statistical Interpolation scheme (SI), operational at that time. That is, the SI solves directly the linear system (2.30) by approximating the $\mathbf{HBH}^T$ term in the neighboring points of the observation locations (Daley, 1991). Thus, PSAS was presented as an advantage to the SI since it solves globally what the SI solves locally by using all available observations without any local data selection and by making unnecessary many of the simplifying approximations required by the SI. (Lorenc 1986, Da Silva *et al.* 1995, Cohn *et al.* 1998).

On the other hand, the main difference with respect to the 3D-Var case is that, as discussed earlier, the size of the control vector of the 3D-PSAS objective function is determined by the number of observations instead of the dimension of the model space. Solving the assimilation problem in a smaller space can be of interest for operational centres as they are limited in the overall memory storage and computation cost of analysis applications.

2.3.2 4D-PSAS

As 4D-Var is the temporal extension of 3D-Var, 4D-PSAS can also be seen as the dual formulation taking into account observation time and the flow-dependent properties of error covariances. Following the same process used to get the 3D-PSAS formulation, and using (2.27), 4D-PSAS is introduced as follows:

$$F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T(\mathbf{R} + \mathbf{GBG}^T)\mathbf{w} - \mathbf{w}^T\mathbf{y}' \tag{2.36}$$

[4] As of 2003, the DAO became the Global Modeling and Assimilation Office (GMAO).

where $\mathbf{w}$ is now a vector of the dimension of the total number of observations over the entire assimilation window, and $\mathbf{GBG}^T\mathbf{w}$ stands for the vector of the increments in observation space. As in the 3D case, both 4D-Var and 4D-PSAS can be preconditioned respectively with $\mathbf{B}^{\frac{1}{2}}$ and $\mathbf{R}^{-\frac{1}{2}}$, in which case they should have similar convergence rates (Courtier, 1997). Moreover, although the degrees of freedom is lower in the dual form, the preconditioning in the present case ensures that a similar number of iterations should be necessary. Hence, the total cost would be the same in the two cases.

It is also thought that 4D-PSAS can be, in some cases, a more interesting alternative to 4D-Var than 3D-PSAS would be for 3D-Var, while 3D-PSAS and 3D-Var seem to differ only in the way the problem is solved. This will be explained in more details in the next chapters.

2.3.3 Convergence properties

Courtier (1997) showed that, in the *linear case*, 3D/4D-PSAS are theoretically equivalent to 3D/4D-Var *at convergence* of the minimization of their objective functions. However, the incremental approach allows for the extension of the equivalence to the nonlinear case. Both approaches solve the same problem and their Hessians (2.14 and 2.35) have the same condition number (the ratio of largest to smallest eigenvalue); the smallest eigenvalue being 1. That is, if using their own preconditioning ($\mathbf{B}^{\frac{1}{2}}$ and $\mathbf{R}^{-\frac{1}{2}}$), both algorithms are expected to converge at the same rate (Courtier, 1997), with a comparable number of iterations and produce exactly the same analysis state. However, changing the preconditioning may alter this result.

2.4 The Weak-constraint formulation

2.4.1 The motivation

One of the main assumptions used in operational NWP implementations of variational schemes, is that the model equations "*perfectly*" describe the dynamics of the

atmosphere. This is known as the *perfect* model assumption, where any discrepancy between the model and the data is tacitly ascribed to errors in the background and in the observational state. However, the forecast error can also be attributed to imperfections in the NWP model itself. It is clear that model errors do affect the quality of the analysis and the forecast if not taken into account (Derber, 1989; Wergen, 1992; Zupanski, 1993; Bennett *et al.*, 1993; and Zupanski,1997, Bennett, 2002, Tsyrulnikov, 2005).

To address this issue, it has been proposed to relax the "perfect" model assumption by considering the model as *imperfect*, and therefore allowing the analysis to deviate from the exact model trajectory. Here, the analysis now refers to the model trajectory $x(t)$ over the whole assimilation interval. This is called the *weak constraint* approach, as opposed to the *strong constraint* framework seen earlier. In the latter, the analysis state was required to satisfy exactly the imposed model constraints as the whole trajectory is the result of a model integration from given initial conditions. In a weak-constraint formulation of the problem, the imposed constraints are satisfied only approximately, not exactly, depending on what is known about the model error. This approach was first introduced by Sasaki (Sasaki 1969, 1970), and explored for NWP by Derber (1989) and is now being considered for operational implementation (Trémolet, 2007).

Interest for the weak-constraint formulation lies beyond the correction for the model error term in the assimilation system. The *optimal length* of the assimilation time period (the assimilation window) is believed to be limited by the uncertainties in the model (Trémolet, 2007). Currently, most NWP centres use assimilation windows of the length of 6h, 12h or 24h at most. It has been argued (Pires *et al.*, 1996) that increasing this window gradually beyond the error doubling time of the system (usually about 3 days) significantly improves the quality of the assimilation state and of the forecast. Using a perfect model setting, Swanson *et al.* (1998) showed

that the state estimate saturates at an *optimal* assimilation period of $T_a = 10$ days, and that the predictability time scale in this case extends about 5 days beyond the case where the assimilation period is 1 day. In an attempt to mimic the behavior of NWP models, results with an *imperfect* model (where model errors were introduced) showed that the performance of the assimilation vary strongly with the growth rate of the model error (Swanson *et al.*, 1998). That is, for large, rapidly growing model error, extension of assimilation periods beyond 1-2 days results in a degradation in the quality of the assimilated state as well as in the forecast quality.

The weak-constraint formulation of variational assimilation is expected to allow longer assimilation windows as it accounts for model errors. In other words, an assimilation over a long window will rely on the model error term in a weak-constraint formulation to provide the appropriate temporal retention of information at all spatial scales: several days for synoptic scales, and a few hours for the smallest resolved scales (Fisher *et al.*, 2005).

2.4.2 Model error

NWP models only provide a discrete and reduced-complexity approximate representation of the atmosphere, which departs from the truth due to the presence of model error.

Model errors can be cast in the assimilation system in their *systematic* and/or *random* form. According to Dee and Da Silva (1998), these two parts of the error should be accounted for separately, otherwise the reduction of one may lead to an increase in the other.

- **Systematic errors** : (or bias), for which the mean error is non zero ($\mathbf{E}\left[\boldsymbol{\epsilon}(t)\right] \neq 0$), can be seen as an inherent, persistent deficiency in the model due to parameters not being adequately constrained by available observations or due to

the structure of the model being incapable of representing some specific phenomena. The effects of such errors often persist for a certain time, and can be detected when specific aspects of the model climatology differ from the actual climatology as derived from observations (Lewis *et al.* 2006).

- **Random errors**: in which case, $\mathbf{E}[\boldsymbol{\epsilon}(t)] = 0$. These errors are assumed to be stochastic variables that are unbiased (and usually uncorrelated in time), which allows for the definition of a Gaussian distribution together with a representation of error covariance matrix: $\mathbf{Q} = \mathbf{E}[\boldsymbol{\epsilon}(t)\boldsymbol{\epsilon}(t)^T]$, at a given time.

Following Dee and Da Silva (1998), since the errors in data assimilation are assumed *unbiased*, the procedure of taking into account model error is often a two-step process through which the bias is estimated first, removed, and then the unbiased component is estimated. However, this is difficult and relies on the availability of unbiased observations.

2.4.3 Weak-constraint 4D-Var

To account for model errors in a data assimilation system, a correction term is introduced (Jazwinski, 1970; Derber, 1989 and Cohn, 1997). The idea, is to add to the control vector a residual error correction term, which is added to the model forecast at every time step. Assuming the model is not perfect, (2.16) is modified to include correction terms ($\boldsymbol{\eta}_i$) at each time step t_i, such that:

$$\mathbf{x}_i = \boldsymbol{\mathcal{M}}_{i,i-1}(\mathbf{x}_{i-1}) + \boldsymbol{\eta}_i \tag{2.37}$$

which, in the incremental form, can be approximated as

$$\boldsymbol{\delta}\mathbf{x}_i = \mathbf{M}_{i,i-1}(\boldsymbol{\delta}\mathbf{x}_{i-1}) + \boldsymbol{\eta}_i \tag{2.38}$$

where the $\boldsymbol{\eta}_i$'s represent the model error terms, assumed to be zero in the perfect-model case. The objective function in (2.25) can then be extended in the absence of

systematic errors as follows:

$$J(\boldsymbol{\delta}\mathbf{x}_0, ..., \boldsymbol{\eta}_i, ...) = \frac{1}{2}\boldsymbol{\delta}\mathbf{x}_0^T\mathbf{B}^{-1}\boldsymbol{\delta}\mathbf{x}_0 + \frac{1}{2}\sum_{i=0}^{q}(\mathbf{H}_i\boldsymbol{\delta}\mathbf{x}_i - \mathbf{y}_i')^T\mathbf{R}_i^{-1}(\mathbf{H}_i\boldsymbol{\delta}\mathbf{x}_i - \mathbf{y}_i') + \frac{1}{2}\sum_{i=0}^{q}\boldsymbol{\eta}_i{}^T\mathbf{Q}_i^{-1}\boldsymbol{\eta}_i \tag{2.39}$$

where, as before, $\mathbf{B}$ is the background error covariance matrix, and $\mathbf{R}_i$ and $\mathbf{Q}_i$ are the observation and model error covariance matrices for each time t_i, respectively. These three types of errors are assumed to be *uncorrelated*. This assumption is commonly used but debatable because the background state is usually taken as a previous run of the same model and using the same type of data. The model error covariance matrix is also assumed to be uncorrelated in time, which is not realistic (Daley, 1991).

Note that the control vector here (with respect to which the functional J is minimized) is extended to include all the $\boldsymbol{\eta}_i$ terms. As well as the initial states, the model errors at every time are the control parameters that are sought to be determined. That is, over the assimilation time interval, it is the *full trajectory* $\mathbf{x}(t)$ that is adjusted to fit the observations, and not only the initial state as in the strong-constraint formulation (see Fig.2–3).

The variable $\boldsymbol{\eta}_i$ has the dimension of a three dimensional atmospheric state and represents the instantaneous model error. The overall size of the problem is now multiplied by the total number of time steps. In operational NWP systems, the size of the control vector would be about 10^9 (compared to about 10^7 for current operational applications of strong-constraint 4D-Var). Thus, $\mathbf{Q}$ would contain around $10^9 \times 10^9$ elements that are needed to be estimated. With only about 10^6 observations available each day, there is not enough information to estimate $\mathbf{Q}$.

Beside the high cost of the weak-constraint formulation, no general form of model errors nor their statistics is known so far, and many of the ideas proposed to overcome these limitations rely on important approximations and assumptions

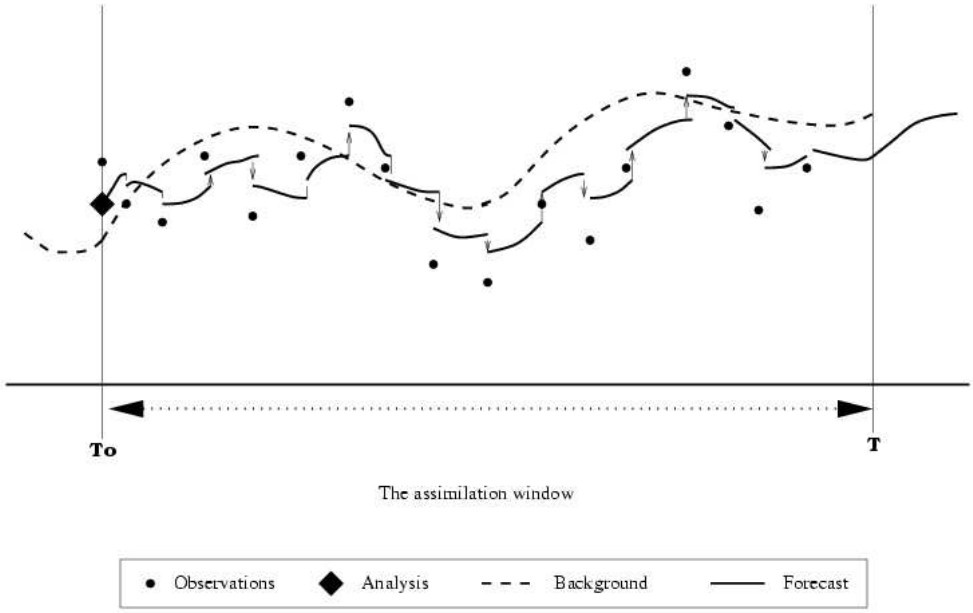


Figure 2–3: Schematic of the weak-constraint 4D-Var assimilation. The model trajectory is adjusted to fit as best as possible the observations. Besides, at each time step, the trajectory is corrected with an amount that compensates for the instantaneous model error.

that aim to reduce the size of the control vector or to simplify the modelling and representation of the model error covariance matrix (Daley, 1991).

Recent attempts have focused on reducing the size of the model error part of the control vector, either by controlling the correction term only in certain privileged directions (Vidard *et al.* 2001), or by using information provided by the analysis residual vector (Vidard *et al.* 2003), or by controlling only the systematic and time correlated part of the error (Griffith and Nichols, 2001).

Approximations have also been proposed for the estimation and modeling of the model error covariance matrix $\mathbf{Q}$. It can be assumed to have the same structure as the background error covariance matrix, or it can be constructed using the same statistical model used for $\mathbf{B}$ to tendencies instead of short term analysis increments.

That is, perturbations of model tendencies can be interpreted as possible realizations of the model error.

In a compact matrix form, (2.39) can be expressed as:

$$J(\mathbf{z}) = \frac{1}{2}\mathbf{z}^T\mathbf{D}^{-1}\mathbf{z} + \frac{1}{2}(\mathbf{S}\mathbf{z} - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{S}\mathbf{z} - \mathbf{y}') \tag{2.40}$$

where $\mathbf{z} = (\boldsymbol{\delta}\mathbf{x}_0, ..., \boldsymbol{\eta}_i, ...)$ is the new control vector, $\mathbf{D}$ stands for the block diagonal matrix consisting of $\mathbf{B}$ for the first block and the $\mathbf{Q}_i$ for the others,

$$\boldsymbol{\mathcal{D}} = \begin{pmatrix} \mathbf{B} & 0 & \cdots & 0 \\ 0 & \mathbf{Q}_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q}_q \end{pmatrix}$$

and $\mathbf{S}$ is the new *generalized observation operator*, such that

$$\mathbf{S}^T = [\mathbf{G}_0^T, \quad \mathbf{G}_1^T, \quad \cdots, \quad \mathbf{G}_n^T]$$

and

$$\mathbf{G}_j = [\mathbf{H}_j\mathbf{M}_{j,0}, \quad \mathbf{H}_j\mathbf{M}_{j,1}, \quad \cdots, \quad \mathbf{H}_j\mathbf{M}_{j,j}, \quad 0, \quad \cdots, \quad 0].$$

Note that $\mathbf{D}$ can also have off-diagonal terms in the case of time-correlated model errors. Again, we notice the similitude of the functional J in (2.40) to those of the strong-constraint 4D-Var (2.27), and the 3D-Var (2.6). The control vector, the observation operator and the background error covariance matrix have all been extended to account for the time tendency information in the observations first (4D-Var), and then to include the model error component into the assimilation system through the weak-constraint formulation.

2.4.4 Weak-constraint 4D-PSAS

Following the same process that led to the formulation of the 3D-PSAS functional, and using (2.40), it is easy to introduce the dual formulation in the four

dimensional weak-constraint case as follows:

$$F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T(\mathbf{R} + \mathbf{S}\mathbf{D}\mathbf{S}^T)\mathbf{w} - \mathbf{w}^T\mathbf{y}'. \tag{2.41}$$

Note that, in this context, the control vector, $\mathbf{w}$, is still defined in the observation space, and that unlike the primal variational formulation, the model error covariance matrix $\mathbf{Q}$ (in $\mathbf{D}$) is present in the algorithm in its direct (not inverse) form, so that the PSAS algorithm remains regular in the limit of vanishing model error.

As simple and concise this matrix form of the functional seems, the operators involved are very complex, and their implementation needs careful attention. A special care needs to be taken in testing the validity of the linearizations in the TLM and the adjoint model. This is very important since the linearity of the operators and the incremental formulation are the underlying assumptions that made this dual weak-constraint formulation possible.

As discussed earlier, the full weak-constraint 4D-Var is challenging both in its overall affordability, and in the estimation of its error statistics in $\mathbf{Q}$. This is where the dual formulation of variational assimilation comes into play, especially with respect to the affordability. Recall that the dual method solves the variational problem in the observation space; whose size *does not* change whether one uses the strong or weak-constraint formulation. That is, the dual framework would allow for accounting for model errors without any significant increase in computational efforts. Moreover, it is expected to be a more suitable choice when extending the length of the assimilation window.

When compared to the weak-constraint 4D-Var, the dual form uses the same operators and matrices, but in a different sequence. Provided that these operators are implemented in a modular way (i.e. they are independent of each other), it is possible to easily build a dual algorithm from a pre-existing primal algorithm. The modularity is key to the proper intercomparison of the primal and dual forms that

will be described in the next chapters.

2.5 Summary

In this chapter, the variational form of the statistical estimation problem of data assimilation was presented. In this framework, the variational problem can be cast in different forms : 3D or 4D, primal or dual, strong or weak-constraint. *Three dimensional* (3D) schemes are designed to provide an analysis state at a single time, assuming that all observations are "collected" at that specific time. In contrast, *four dimensional* (4D) schemes seek to benefit from the *time tendency information* in the observations by fitting observations at their respective time with the trajectory of the model. The primal form of the variational approach solves the problem in model space (i.e. the functional is defined with respect to a model state vector), whereas the dual form solves the *same* problem but in the observation space. Finally, the strong-constraint formulation assumes that the dynamical model is "perfect", whereas the weak-constraint formulation allows for model errors to be accounted for. All these methods are either equivalent (primal and dual) or are the extension of one another to a more general case (3D to 4D, and strong to weak-constraint) (see Table2–1). While the weak-constraint formulation of variational data assimilation is still an active field of research, the strong-constraint methods are currently used for meteorological (primal form) or oceanic applications (dual form). As discussed earlier,

	primal	dual
3D	3D-Var	3D-PSAS
4D Strong-Constraint	4D-Var	4D-PSAS
4D Weak-Constraint	weak-C 4D-Var	weak-C 4D-PSAS.

Table 2–1: The primal and dual variational assimilation methods.

weak-constraint 4D-Var could benefit from the dual form because of the reduction of the dimension of the assimilation problem. As a first step towards accounting for

model errors in primal and dual variational data assimilation (through the weak-constraint formulation), it is foremost crucial to build solid foundations for the dual method. The next chapter presents a thorough intercomparison of the 3D-Var and 3D-PSAS in an operational context to examine the extent to which the theoretical equivalence between these two methods still holds in a complex assimilation system. The convergence properties will be compared and the preconditioning will be examined. In chapter 4, the minimization algorithms are examined to assess their effect on the dual approach in 3D and 4D, and finally, the weak-constraint formulation is examined in chapter 5, both in the primal and dual case.

3 Intercomparison of the primal and dual formulations of variational data assimilation

In this chapter, the primal and dual forms of variational data assimilation are here investigated in an intercomparison study. With proper conditioning of the minimization problem, both algorithms are expected to have similar convergence rates and computational performances. This statement is examined here in the context of the variational data assimilation system of the Meteorological Service of Canada. The objective being to confirm the equivalence of the primal and dual methods and to examine their preconditioning and their convergence properties.

This Chapter is based on the following paper:

El Akkraoui A., P. Gauthier, S. Pellerin, and S. Buis, 2008 : *Intercomparison of the primal and dual formulations of variational data assimilation.* Q.J.R. Met. Soc., **134 :** 1015-1025.

Intercomparison of the primal and dual formulations of variational data assimilation

El Akkraoui A.[1], P. Gauthier[12], S. Pellerin[3], and S. Buis[4]

[1] Department of Atmospheric and Oceanic Sciences, McGill University, Canada

[2]Department of Atmospheric and Earth Sciences, Université du Québec à Montréal, Canada [3]Meteorological Research Division, Environment Canada, Canada

[4]European Centre for Research and Advanced Training in Scientific Computation (CERFACS), France

Abstract

Two approaches can be used to solve the variational data assimilation problem. The primal form corresponds to the 3D/4D-Var used now in many operational NWP centres. An alternative approach, called dual or 3D/4D-PSAS, consists in solving the problem in the dual of observation space. Both forms use the same basic operators so that once one method is developed, it should be possible to obtain easily the other provided these operators have a modular form. In Courtier (1997), it was shown that with proper conditioning of the minimization problem, the two algorithms should have similar convergence rates and computational performances. In presence of nonlinearities, the incremental form of 3D/4D-Var extends the equivalence to the so-called 3D/4D-PSAS. The first objective of this paper is to present results obtained with the variational data assimilation of the Meteorological Service of Canada to show the equivalence between the 3D-Var and the PSAS systems. This exercise has forced us to have a close look at the modularity of the operational 3D/4D-Var which then makes it possible to obtain the 3D-PSAS scheme. This paper then focuses on these two quadratic problems that show similar convergence rates.

However, the minimization of 3D-PSAS is examined more thoroughly as some parameters are shown to be determining elements in the minimization process. At last, preconditioning properties are studied and the Hessians of the two problems are shown to be directly related to one another through their singular vectors, which makes it possible to cycle the Hessian of the PSAS form.

3.1 Introduction

In recent years, data assimilation methods of increasing complexity have been developed to extract information from an ever increasing variety of instruments. Most instruments provide information that is indirectly related to the atmospheric variables like temperature, winds, pressure and humidity. Radiances coming from satellite instruments is one good example. As satellite data are now making the most important contribution to the global observation system, research in data assimilation seeks to develop new methods and improve existing ones to extract as much information as possible from this vast amount of data. Improving the efficiency of those complex data assimilation systems is also a concern in view of their operational use.

Compared to previous data assimilation methods like optimal interpolation (OI), variational methods offer a more natural framework to assimilate remotely-sensed measurements and can be extended to assimilate time-distributed data at their appropriate observation time within four dimensional schemes. Variational methods have now been implemented at several centres like the ECMWF (Rabier *et al.*, 2000), NCEP (Parish and Derber, 1992), UK Meteorological Office (Rawlins *et al.*, 2007), Météo-France (Gauthier and Thépaut, 2001) and Environment Canada (Gauthier *et al.*, 2007). The implementation can go in step by first implementing a 3D-Var which can be extended to 4D-Var (Courtier, 1997; Gauthier *et al.*, 2007). Referring to Courtier (1997), two approaches can be taken and can be shown to be only different algorithms to solve the same basic problem. The 3D/4D-Var, referred as the primal approach, uses a control variable defined in model space while the dual form uses a control variable defined in the dual of observation space. This approach was taken in the development of the Physical-space Statistical Analysis System (PSAS) (Cohn *et al.*, 1998) and has been adopted by NASA's Global Modeling Assimilation Office and the Naval Research Laboratory (Barker and Daley, 2000). Courtier (1997) pointed

out that the equivalence between the primal and dual approaches leads to two minimization problems for which the Hessian of the quadratic functionals have the same condition number. Using their own preconditioning as proposed by Amodei (1995) and Courtier (1997), both algorithms should give the same results at convergence and at approximately the same rate.

The objective of this paper is to investigate the equivalence of the two approaches in the framework of the operational variational data assimilation system of Environment Canada. In the second section, the equivalence of the two approaches will be cast in the context of the incremental approach, which then extends the equivalence to the nonlinear problem. The dual or primal approach could then be used indifferently to solve the quadratic problem associated with the inner loop. In section 3, results will be presented to confirm the equivalence at convergence and examine the convergence of the two algorithms. This will be shown in the context of an incremental 3D-Var using its respective dual form. When cycling the assimilation, it has been shown that using the Hessian approximation from the previous assimilation can help to precondition the next assimilation. As the dual form uses a control variable defined in the dual of observation space, it varies considerably from one analysis to the next according to the observation coverage. In section 4, the Hessian of the primal and dual forms are shown to have equivalent eigenvectors that can be mapped onto one another. This property is then used to map the Hessian of the dual problem to the next analysis and the impact on the convergence is examined when preconditioning the next inner loop problem. Section 5 will present a summary and conclusions with perspectives for applications.

3.2 Theoretical background

3.2.1 3D-Var

Variational assimilation estimates the state of the atmosphere at a given time by minimizing an appropriate objective function. As discussed in Lorenc (1988), based on a Bayesian approach, the probability distribution that the model state $\mathbf{x}$ (dimension n) is the true value of the atmospheric state given that $\mathbf{y}$ (the observation vector, dimension p) has been observed, can be expressed for Gaussian error distributions as :

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \\ &= \frac{1}{c_1} exp(-\frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x}))) \\ &\times \frac{1}{c_2} exp(-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)) \end{aligned} \tag{3.1}$$

where $\mathbf{B}$ and $\mathbf{R}$ are the background and observation covariance matrices respectively, $\mathbf{x}_b$ is the background term (usually taken as a short-term forecast), $\mathcal{H}$ is the non linear observation operator that maps the model variables into the dual of observation space, and c_1 and c_2 are constants. The best linear unbiased estimate of the state of the atmosphere at the analysis time is referred as the *analysis*, and is obtained by minimizing the function :

$$\begin{aligned} J(\mathbf{x}) &= -ln(p(\mathbf{x}|\mathbf{y})) \\ &= \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) \\ &+ \frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x})) \end{aligned} \tag{3.2}$$

This property ensures that minimizing J is a way of minimizing the variance of the analysis error and that the minimum is the most probable state.

In its incremental formulation (Courtier *et al.* 1994), the three dimensional variational scheme (3D-Var) operates a global minimization of its functional J with respect to an increment $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$, representing the misfit between the background

state $\mathbf{x}_b$ and the current model state $\mathbf{x}$. The incremental cost function is

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{y}') \tag{3.3}$$

where $\mathbf{H}$ is the Jacobian (or the Tangent linear) of the non-linear observation operator ($\mathcal{H}$), and the new term $\mathbf{y}' = \mathbf{y} - \mathcal{H}(\mathbf{x}_b)$ represents the "innovation" vector or the misfit between the observation state $\mathbf{y}$ and the background state in the physical space. In practice, a solution is found through iterative minimization algorithms (e.g., the Conjugate-Gradient or the Limited-Memory Quasi-Newton). In order to speed up the convergence of the minimization, a change of variable $\mathbf{v} = \mathbf{B}^{-\frac{1}{2}}\delta\mathbf{x}$ is introduced (Lorenc, 1988) as a preconditioner. The functional and its gradient in this case are :

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}(\mathbf{H}\mathbf{B}^{\frac{1}{2}}\mathbf{v} - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{B}^{\frac{1}{2}}\mathbf{v} - \mathbf{y}') \tag{3.4}$$

$$\nabla_{\mathbf{v}}J = (\mathbf{I}_n + \mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{\frac{1}{2}})\mathbf{v} - \mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}'. \tag{3.5}$$

The second derivative of J, the Hessian, is

$$\mathbf{J}'' = \mathbf{I}_n + \mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{\frac{1}{2}} \tag{3.6}$$

whose diagonal elements are largely dominated by the identity matrix in the model space ($\mathbf{I}_n$) when few observations are present (Fisher and Andersson, 2001). In variational analysis, the inverse of the Hessian ($\mathbf{J}''$) corresponds to the analysis error covariance matrix $\mathbf{A}$.

3.2.2 3D-PSAS

The physical-space or the dual formulation of variational methods was designed to solve the same data assimilation problem as the variational one, but in the dual of observation space rather than in the model space. The Physical-space Statistical Analysis System (PSAS) was developed and implemented first at NASA's Data Assimilation Office (DAO) in the late 90's (Da Silva *et al.*, 1995; Cohn *et al.*, 1998),

to replace the Optimal Interpolation scheme (OI), operational at that time. The 3D-PSAS formalism is introduced here following Da Silva *et al.*(1995) and Courtier (1997).

The 3D-Var analysis increment is obtained by setting $\nabla_{\delta \mathbf{x}_a} J = 0$ and can be written, in the linear case, as

$$\begin{aligned} \delta \mathbf{x}_a = & \quad (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}' \\ = & \quad \mathbf{B}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \mathbf{y}' \\ = & \quad \mathbf{B}\mathbf{H}^T \mathbf{w}_a. \end{aligned} \tag{3.7}$$

where $\mathbf{w}_a$ has the same dimension (p) as the observation vector $\mathbf{y}$ and subscript a refers to the *analysis* state. This shows that solving the 3D-Var problem is equivalent to solving, for $\mathbf{w}$, the linear system

$$(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)\mathbf{w} = \mathbf{y}'. \tag{3.8}$$

$\delta \mathbf{x}_a$ then satisfies $\delta \mathbf{x}_a = \mathbf{B}\mathbf{H}^T \mathbf{w}_a$, where $\mathbf{w}_a$ is the solution of (3.8). $\delta \mathbf{x}_a$ can be seen as the analysis increment term in the dual of observation space, and $\mathbf{B}\mathbf{H}^T$ as the representer matrix that maps the increments from observation to model space.

While the OI solves approximately the linear system (3.8) by selecting, for each model variable, a subset of few observations representing the most relevant data for determining the analysis increment, the dual approach proposes to solve (3.8) through a global minimization of the quadratic functional :

$$F(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)\mathbf{w} - \mathbf{w}^T \mathbf{y}'. \tag{3.9}$$

Following Amodei (1995), a preconditioning is used by considering the change of variable $\mathbf{u} = \mathbf{R}^{\frac{1}{2}} \mathbf{w}$. The objective function and its gradient become :

$$F(\mathbf{u}) = \frac{1}{2} \mathbf{u}^T (\mathbf{I}_p + \mathbf{R}^{-\frac{1}{2}} \mathbf{H}\mathbf{B}\mathbf{H}^T \mathbf{R}^{-\frac{1}{2}})\mathbf{u} - \mathbf{u}^T \mathbf{R}^{-\frac{1}{2}} \mathbf{y}' \tag{3.10}$$

$$\nabla_{\mathbf{u}} F = (\mathbf{I}_p + \mathbf{R}^{-\frac{1}{2}} \mathbf{H}\mathbf{B}\mathbf{H}^T \mathbf{R}^{-\frac{1}{2}})\mathbf{u} - \mathbf{R}^{-\frac{1}{2}} \mathbf{y}' \tag{3.11}$$

and the Hessian can be written as

$$\mathbf{F_u}" = \mathbf{I}_p + \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}} \tag{3.12}$$

where $\mathbf{I}_p$ is the identity matrix in the dual of observation space. The main difference with respect to the 3D-Var case is that the size of the control vector of the 3D-PSAS functional is determined by the number of observations (p) instead of the dimension of the model space (n). Therefore, it takes advantage of the fact that the number of observations used in atmospheric data analysis is usually one order of magnitude less than the degrees of freedom of most NWP models to carry out the minimization in a smaller space than 3D-Var. Moreover, PSAS was presented as an advantage to the OI since it solves globally what the OI solves locally by using all available observations without any local data selection and by making unnecessary many of the simplifying approximations required by the OI (Lorenc, 1986). This result was also confirmed in Gauthier *et al.* (1999), where the impact of data selection was shown to be significant in a single analysis increment.

3.2.3 The four dimensional case

Unlike the 3D-Var, where no account is made for the observation time, 4D-Var can be seen as the temporal extension of the 3D-Var for observations $\mathbf{y}_i$ distributed over a finite time interval (t_0, t_N). The basic concept is the same as 3D-Var provided that a nonlinear forecast model $\mathcal{M}$ is used to perform integrations in time, which implicitly leads to a flow-dependent background covariance matrix and analysis increments (Thépaut and Courtier, 1991). At each observation time t_i, the model state $\mathbf{x}_i$ is related to the initial state by $\mathbf{x}_i = \mathcal{M}(t_i, t_0)(\mathbf{x}_0)$. When the forecast model and the observation operator are both linear, the previous theoretical reasoning in the three dimensional case still holds, and a 4D- PSAS is easily formulated (Courtier, 1997). In the case of non-linearities, the incremental formulation uses the tangent linear and adjoint to approximate the time evolution of increments $\delta\mathbf{x}_i{}^{(l)}$ around a

reference trajectory ${\mathbf{x}_i}^{(l)}$, which needs to be updated regularly (Courtier *et al.*, 1994). The number of times that the trajectory is updated is called the number of outer loops and referred here by the superscript l.

Following Gauthier *et al.* (2007), the propagator of the Tangent linear model (TLM) is introduced $\delta{\mathbf{x}_i}^{(l)} = \mathbf{M}^{(l)}(t_i, t_o)\delta{\mathbf{x}_0}^{(l)}$, where $\mathbf{M}$ is the TLM of the nonlinear model $\mathcal{M}$, so that the incremental form of 4D-Var is a quadratic objective function that can be written as:

$$\begin{aligned} J_l(\delta{\mathbf{x}_0}^{(l)}) =& \frac{1}{2}[\delta{\mathbf{x}_0}^{(l)} - ({\mathbf{x}_0}^b - {\mathbf{x}_0}^{(l)})]^T\mathbf{B}^{-1} \\ & \times [\delta{\mathbf{x}_0}^{(l)} - ({\mathbf{x}_0}^b - {\mathbf{x}_0}^{(l)})] \\ & + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i{}^{(l)}\delta{\mathbf{x}_i}^{(l)} - {\mathbf{y'}_i}^{(l)})^T{\mathbf{R}_i}^{-1} \\ & \times (\mathbf{H}_i{}^{(l)}\delta{\mathbf{x}_i}^{(l)} - {\mathbf{y'}_i}^{(l)}) \end{aligned} \tag{3.13}$$

where, the observation operator and the observation error covariance matrix are $\mathbf{H}_i$ and $\mathbf{R}_i$, and the innovation vector is ${\mathbf{y'}_i}^{(l)} = \mathbf{y}_i - \mathcal{H}_i(\mathcal{M}(t_i, t_0)({\mathbf{x}^{(l)}}_i))$ for each observation time. In practice, the analysis state is found by performing first a number of inner iterations with a TLM and adjoint to minimize J_l followed by outer iterations, in which the full nonlinear model is integrated from the updated initial conditions $({\mathbf{x}_i}^{(l+1)} = \mathcal{M}(t_i, t_0)({\mathbf{x}_0}^{(l)} + \delta{\mathbf{x}_0}^{(l)}))$. This is meant to update the reference trajectory that defines the linearized and adjoint model and also to refresh the innovation vector consistently with the initial cost function to minimize. Let us now introduce the new definitions :

$$\begin{cases} {\xi_0}^{(l)} = \mathbf{B}^{-\frac{1}{2}}\delta{\mathbf{x}_0}^{(l)} \\ \bar{\xi}_l \quad = \mathbf{B}^{-\frac{1}{2}}({\mathbf{x}_0}^{(l)} - {\mathbf{x}_0}^b) \\ \mathbf{G}^{(l)} = (..., \mathbf{H}_i{}^{(l)}\mathbf{M}(t_i, t_0), ...) \end{cases}$$

where the $\mathbf{G}$ operator is consistent with the one introduced by Courtier (1997). The objective function can then be written as :

$$\begin{aligned} J_l({\xi_0}^{(l)}) = \quad & \frac{1}{2}({\xi_0}^{(l)} + \bar{\xi_l})^T({\xi_0}^{(l)} + \bar{\xi_l}) \\ & + \frac{1}{2}[\mathbf{G}^{(l)}\mathbf{B}^{\frac{1}{2}}{\xi_0}^{(l)} - \mathbf{y}'^{(l)}]^T\mathbf{R}^{-1} \\ & \times[\mathbf{G}^{(l)}\mathbf{B}^{\frac{1}{2}}{\xi_0}^{(l)} - \mathbf{y}'^{(l)}] \end{aligned} \tag{3.14}$$

where $\mathbf{R}$ is the blok diagonal matrix containing the $\mathbf{R}_i$'s, $\mathbf{y}'^{(l)}$ is the vector of the $\mathbf{y}'^{(l)}_i$. Setting the gradient of (3.14) to zero, we get

$$\begin{aligned} ({\xi_0}^{(l)})_a = \quad & (\mathbf{I}_n + \mathbf{B}^{\frac{1}{2}T}\mathbf{G}^{(l)T}\mathbf{R}^{-1}\mathbf{G}^{(l)}\mathbf{B}^{\frac{1}{2}})^{-1} \\ & \times (\mathbf{B}^{\frac{1}{2}T}\mathbf{G}^{(l)T}\mathbf{R}^{-1}\mathbf{y}'^{(l)} - \bar{\xi_l}) \\ = \quad & \mathbf{B}^{\frac{1}{2}T}\mathbf{G}^{(l)T}\mathbf{R}^{-\frac{1}{2}}(\mathbf{I}_p + \mathbf{R}^{-\frac{1}{2}}\mathbf{G}^{(l)}\mathbf{B}\mathbf{G}^{(l)T}\mathbf{R}^{-\frac{1}{2}})^{-1} \\ & \times (\mathbf{R}^{-\frac{1}{2}}\mathbf{y}'^{(l)} - (\mathbf{B}^{\frac{1}{2}T}\mathbf{G}^{(l)T}\mathbf{R}^{-\frac{1}{2}})^{-1}\bar{\xi_l}). \end{aligned}$$

Let us now define the $\mathbf{w}_a^{(l)}$ vector in the dual of observation space,

$$\begin{cases} ({\xi_0}^{(l)})_a &= \mathbf{B}^{\frac{1}{2}T}\mathbf{G}^{(l)T}\mathbf{R}^{-\frac{1}{2}}\mathbf{w}_a^{(l)} \\ \bar{\xi_l} &= \mathbf{B}^{\frac{1}{2}T}\mathbf{G}^{(l)T}\mathbf{R}^{-\frac{1}{2}}\bar{\mathbf{w}}_l, \end{cases}$$

so that

$$(\mathbf{I}_p + \mathbf{R}^{-\frac{1}{2}}\mathbf{G}^{(l)}\mathbf{B}\mathbf{G}^{(l)T}\mathbf{R}^{-\frac{1}{2}})\mathbf{w}_a^{(l)} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}'^{(l)} - \bar{\mathbf{w}}_l$$

It follows that the 4D-PSAS objective function can easily be defined within the incremental context as :

$$\begin{aligned} F(\mathbf{w}^{(l)}) = & \frac{1}{2}\mathbf{w}^{(l)^T}(\mathbf{I}_p + \mathbf{R}^{-\frac{1}{2}}\mathbf{G}^{(l)}\mathbf{B}\mathbf{G}^{(l)T}\mathbf{R}^{-\frac{1}{2}})\mathbf{w}^{(l)} \\ & - \mathbf{w}^{(l)^T}(\mathbf{R}^{-\frac{1}{2}}\mathbf{y}' - \bar{\mathbf{w}}_l). \end{aligned} \tag{3.15}$$

3.3 Equivalence

3.3.1 Modular implementation of data assimilation

Most data assimilation schemes rely basically on the same general concepts and use almost the same operators and matrices (the observation operator, the covariance error matrices, the tangent and adjoint linear models...etc); for instance, 3D-Var and

3D-PSAS differ only in the sequence in which these operators are used in (3.3) and (3.9). Therefore, it is important to proceed to a modularization of the assimilation algorithm that ensures these operators to be independent of each other and that any later change would only affect the specific operator under consideration regardless of its sequence in the algorithm. In Lagarde *et al.* (2001), this concept is used to propose a new approach PALM (Projet d'Assimilation par Logiciel Multi-Méthodes), where applications are created by first decomposing them into basic independent components that can then be re-gathered according to their sequence in the application algorithm. The PALM software is then in charge of controlling the data flow exchange between units (Input/Output), performing algebraic calculations and managing the parallelism if needed. The reader is referred to Buis *et al.* (2006) and the PALM team publications for more details[1] .

To conduct experiments on the intercomparison between the primal and dual algorithms of the operational assimilation system of Environment Canada, the first task was to ensure its modularity. This was first done in a feasibility study within the PALM framework (El Akkraoui, 2004), which highlighted the importance of having a modular decomposition in view of the implementation of the dual algorithm, and encouraged the development of a new modular version of the variational code. The use of this modularity is described in (Pellerin *et al.*, 2006) when presenting the coupling of the 3D-Var and the Global Environmental Multi-scale (GEM) model to obtain an incremental 4D-Var (Tanguay and Laroche, 2002). More details on the existing 3D/4D-Var configurations are presented in Gauthier *et al.* (2007) and Laroche *et al.* (2007).

Based on this modular version of the existing 3D/4D-Var, a 3D/4D-PSAS scheme was formulated using the same modular subroutines as described before.

[1] $http://www.cerfacs.fr/\sim palm/$

For sake of simplicity, the coupling of the independent components was done without the PALM coupler. All results shown in this paper are produced in this context.

3.3.2 Duality

Courtier (1997) showed that 3D/4D-PSAS are theoretically equivalent to 3D/4D-Var *at convergence* of the minimization of their objective functions in the *linear case*. However, the incremental approach allows for the extension of the equivalence to the nonlinear case. Both approaches solve the same problem and their Hessians (eq.3.6 and eq.3.12) have the same eigenvalue spectrum, possibly completed by some 1's, and the same condition number (the ratio of largest to smallest eigenvalue); 1 being the smallest eigenvalue. That is, if using their own preconditioning, $\mathbf{B}^{-\frac{1}{2}}$ for the primal and $\mathbf{R}^{\frac{1}{2}}$ for the dual formulation, both algorithms are expected to produce the same analysis state, to converge at the same rate and with a comparable number of iterations (Courtier, 1997). The objective of this section is to verify the equivalence of the primal and dual approaches using the operational three dimensional variational data assimilation system of Environment Canada. Experiments include all observations used operationally at the time.

The minimization of the 3D-Var and 3D-PSAS objective functions (Fig.3–1-a) shows that at the end of the minimization, solutions are such that $J(\mathbf{v}_a) = -F(\mathbf{u}_a)$ confirming the theoretical result (See Appendix A). As expected, the analysis increments at the end of the minimization (not shown) are the same. Louvel (2001) found similar results when conducting experiments using an oceanic primitive equation model. In such a controlled context (synthetic observations, identity observation operator and the number of degrees of freedom of the model is the same as the number of assimilated observations), numerical equivalence has been shown for both the increments and the required number of iterations, although the decrease of the two objective functions was not the same. However, in our operational experiments, 3D-PSAS requires more iterations than 3D-VAR to converge for the same convergence

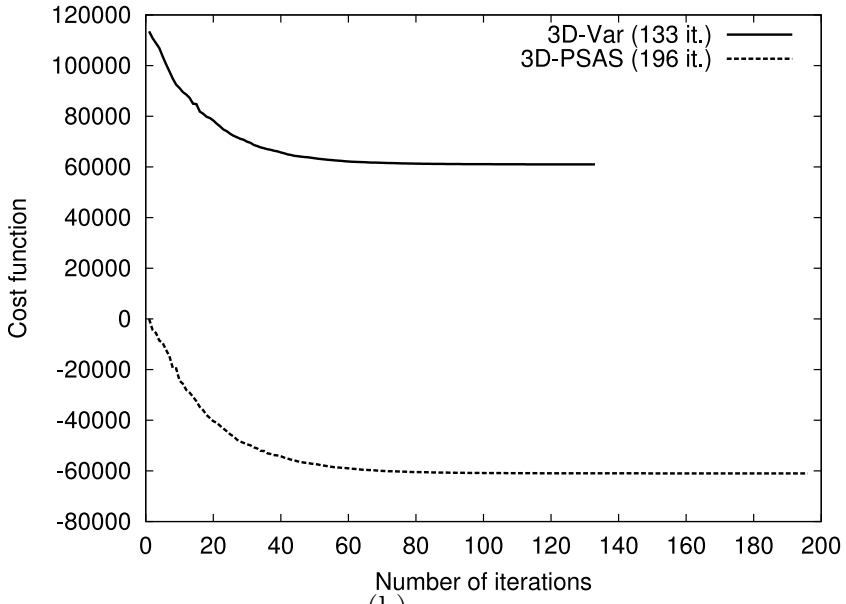
criterion : a two orders of magnitude decrease of the gradient norm with respect to its initial value.

What explains the additional number of iterations needed for the 3D-PSAS to converge is not clear yet. The convergence criterion used was chosen, for convenience, to be the same as that of 3D-Var. However, PSAS may need a more appropriate one. Fig.3–1-b shows the ratio of the gradient norms to their initial values for 3D-Var (solid line) and 3D-PSAS (dashed line), plotted in a logarithmic scale. As expected, both methods converge at similar rates as shown by the slopes of the gradient norms. However the norm of the 3D-PSAS first gradient increases surprisingly by a factor of ten with respect to its initial value, and it takes about 60 iterations to recover its initial value.

As mentioned earlier, the objective function of 3D/4D-Var is related to the analysis error in that its Hessian is the inverse of the analysis error covariance matrix ($J'' = \mathbf{A}^{-1}$), and considering the origin of this formulation (see Section 2), this ensures that, by construction, the minimization of 3D-VAR guarantees the analysis state to be more and more probable. However, this remark does not apply to the dual case as its objective function has no direct physical interpretation, and its minimization does only ensure the equivalence with 3D-VAR *at convergence*. In order to have an image of the PSAS convergence in the model space, the equivalent 3D-Var to the 3D-PSAS objective function at each iterate $\mathbf{u}_k$, where k is the iteration number, is calculated (Fig.3–2). That is, each iterate $\mathbf{u}_k$ is returned to the model space through the representer $\mathbf{BH}^T$ and the 3D-VAR objective function is evaluated for $\mathbf{v}_k = \mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}}\mathbf{u}_k$.

As shown in Fig.3–2, 3D-PSAS starts increasing the apriori probability (See 2.1) during the first iterations, before it decreases below its initial value (at the first iteration) only after about 60 iterations where the iterates gradually improve the analysis. One can also notice that 60 iterations are needed to obtain an iterate for which the norm of the gradient is about the same as that of the starting point of

(a)



(b)

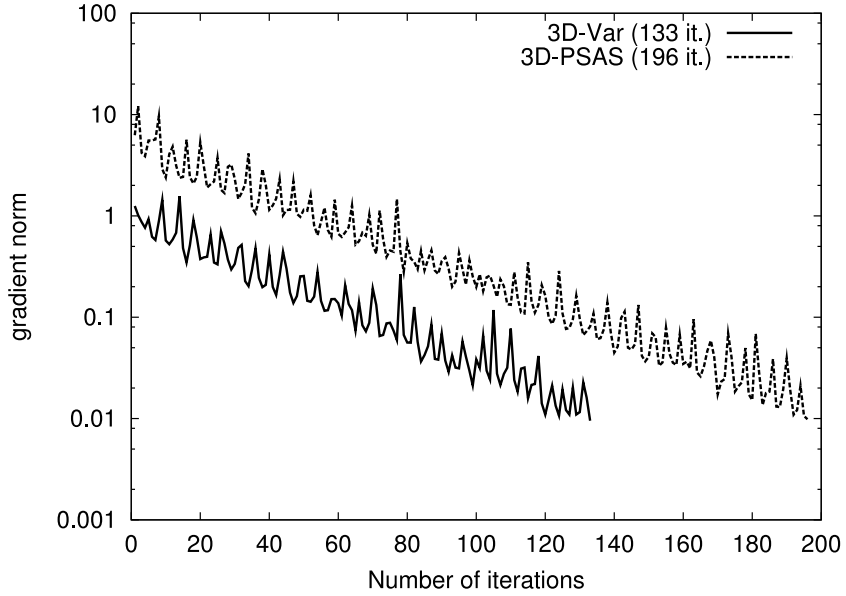


Figure 3–1: Objective function (a) and gradient norm (b) of 3D-Var (solid line) and 3D-PSAS (dashed line) with number of iterations. The gradient norm is plotted in a logarithmic scale.

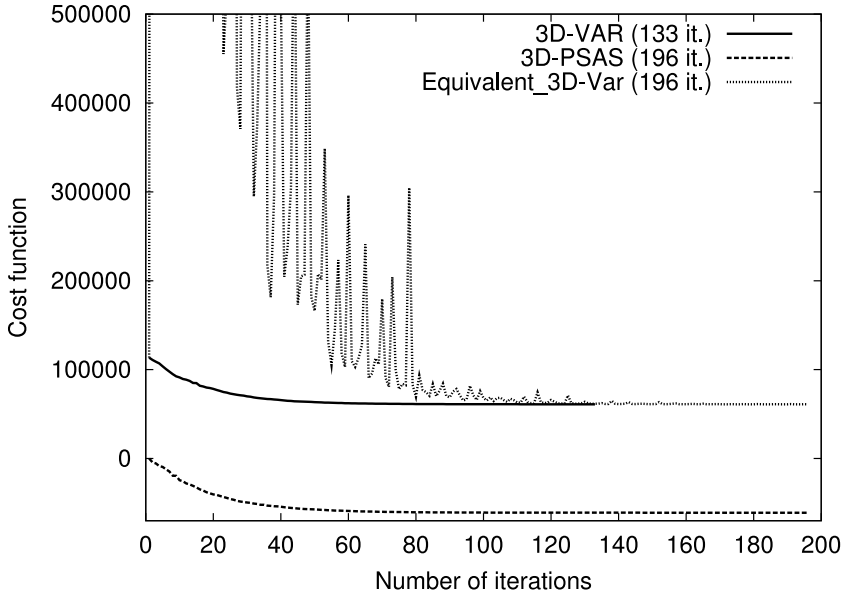


Figure 3–2: Objective function of 3D-Var (solid line), 3D-PSAS (dashed line) and the equivalent 3D-Var at PSAS iterates (dotted line) along the minimization. At each PSAS iteration k, the iterate $\mathbf{u}_k$ is brought to the model space through the representer $\mathbf{BH}^T$ and the 3D-Var objective function is calculated for $\mathbf{v}_k = \mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}}\mathbf{u}_k$.

the minimization, as shown in fig.3–1-b. In the context where a finite number of iterations is required, one may end up with a solution which could be worse than the initial point. This is a real concern for the use of PSAS in an operational setting.

3.4 Preconditioning

Preconditioning intends to accelerate the convergence by reducing the condition number of the objective function's Hessian. The optimal pre-conditioner for the 3D-Var is the square root of the Hessian in (3.6) which enables the convergence to occur in one iteration. Assuming that the Hessian is dominated by $\mathbf{B}$ due to the background errors, the observation term ($\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$) can be neglected and an approximate preconditioner ($\mathbf{B}^{\frac{1}{2}}$) is used. However, in the presence of very dense or particularly accurate observations for instance, this choice of preconditioning is not as efficient as required (Andersson *et al.*, 2000). For 3D-PSAS, the preconditioning

proposed by Amodei (1995) with $\mathbf{R}^{-\frac{1}{2}}$ may not be the best one either considering the sparseness of its Hessian (3.12). Alternative preconditionings are investigated here.

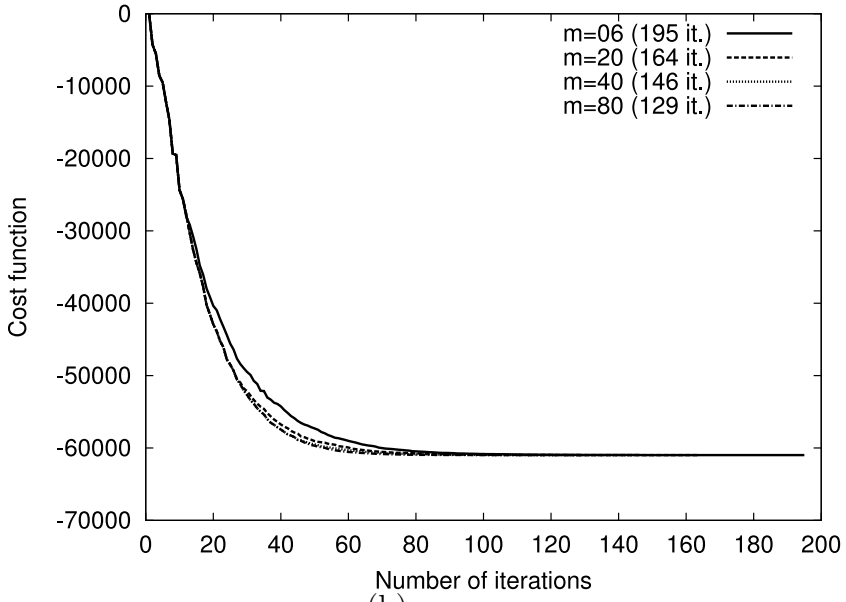
3.4.1 Preconditioning with an approximate Hessian

The Limited-Memory Quasi-Newton algorithm (LMQN) uses the BFGS to approximate the inverse Hessian that can be used to precondition the next analysis cycle. This approximation is performed sequentially at each iteration (k) by a set of m pairs of vectors $(\mathbf{y}_k, \mathbf{s}_k)$, where $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ is the step and $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ is the change in the gradient (Gilbert and Lemaréchal, 1989). The accuracy of the resulting approximate Hessian depends on the number (m) of the pairs of vectors which also determines the dimension of the approximate Hessian. Fig.3–3 shows the impact of increasing (m) on the convergence when the Hessian is estimated during the minimization and only the conventional preconditioning is used ($\mathbf{B}^{\frac{1}{2}}$ for 3D-Var and $\mathbf{R}^{-\frac{1}{2}}$ for 3D-PSAS).

For both algorithms, the number of iterations needed for the convergence drops by about 34% when increasing the number of pairs from 6 to 80 for the same stopping criterion. Note though that this gain requires storing an increasing number of pairs, which must be limited when the dimension of the control variable is large. Therefore, only a reasonable number of pairs can be taken; which in this case, can be set to $m = 40$ considering that most of the gain is already obtained.

In order to assess the accuracy of the resulting approximation, the approximate Hessian is used as a preconditioner for the same assimilation cycle rather than for the next one as it might be more realistic. The reason is that unlike the 3D-Var Hessian that can easily be re-used to precondition the next analysis, the 3D-PSAS Hessian is completely defined in the dual of observation space that depends on the assimilation period, and therefore cannot be re-used. This problem will be examined

(a)



(b)

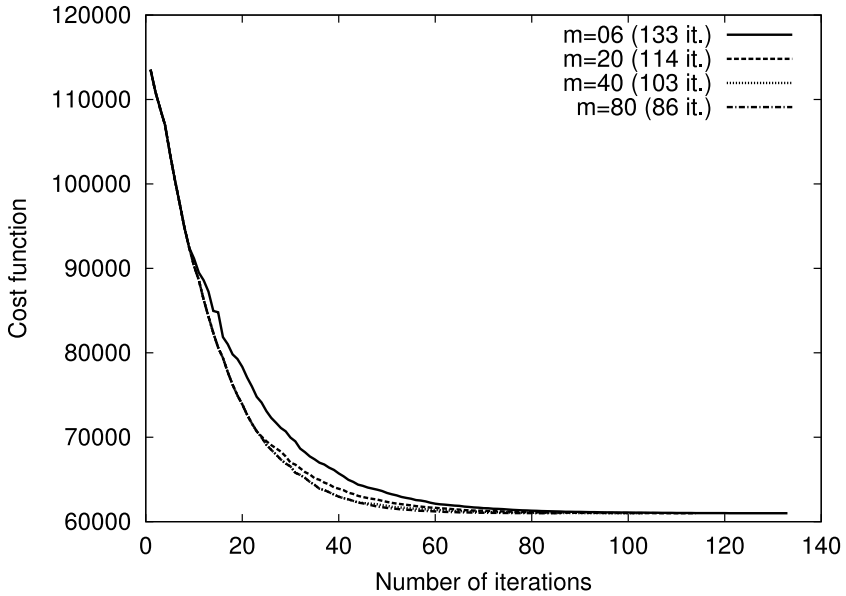


Figure 3–3: Sensitivity of the 3D-PSAS (a) and 3D-VAR (b) convergence to the number of pairs m used by the Quasi-Newton algorithm to approximate the Hessian (m is varying from 6 to 80).

later in this paper. The preconditioned objective functions of 3D-PSAS and 3D-Var for different numbers of pairs are shown in Fig.3–4-a and 3–4-b respectively. As in Fig.3–3, increasing the number of pairs is an efficient way to improve the convergence rate. This is also indicative of the dimension needed to approximate the Hessian.

3.4.2 Preconditioning with singular vectors

Another way of preconditioning the data assimilation problem is to use an approximate representation of the Hessian by its leading singular vectors. Ehrendorfer and Tribbia (1995) pointed out that the singular vectors of a covariance matrix provide the most efficient approximation in the sense that, for an approximation based on a given number of singular vectors, they account for a maximum fraction of the variance. For 3D-Var, this can easily be understood since the Hessian is the inverse of the analysis covariance matrix, whereas for 3D-PSAS, although its Hessian is built from the same basic operators as 3D-Var ($\mathbf{R}$, $\mathbf{B}$ and $\mathbf{H}$), no immediate physical meaning can be directly related to the functional $F(\mathbf{u})$. Still, duality properties discussed earlier in section (3) indicate a close relationship between the Hessians. We investigate here the conditioning of 3D-PSAS with its Hessian singular vectors. Hessian can be represented exactly by a set of all its singular vectors as :

$$\mathbf{J}" = \sum_{k=1}^{N} \lambda_k \nu_k \nu_k^T$$

where (λ_k, ν_k) are eigenpairs of $\mathbf{J}"$ determined using a Lanczos algorithm (Lanczos, 1950). Depending on the shape of the eigenvalues spectrum, this representation can be more or less computationally effective. Moreover, considering the large size of the matrices involved, it is only possible to compute a limited number of eigenpairs ($K < N$) which will be used to build the approximation :

$$\tilde{\mathbf{J}}" \approx \sum_{k=1}^{K} \lambda_k \nu_k \nu_k^T .$$

Fisher and Andersson (2001) pointed out that since the leading eigenvectors are large-scale patterns, they may be determined accurately at low horizontal resolution. Fisher and Courtier (1995) introduced the pre-conditioner P allowing for the definition of a new Hessian $\hat{\mathbf{J}"} = \mathbf{P}^{-T}\mathbf{J}"\mathbf{P}^{-1}$ with a smaller condition number, where

$$\mathbf{P}^{-1} = \mathbf{I} + \sum_{k=1}^{K}(\mu_k^{\frac{1}{2}} - 1)\nu_k\nu_k^T$$

This leads to an implicit representation of this new Hessian as follows :

$$\hat{\mathbf{J}"} = \sum_{k=1}^{K}\mu_k\lambda_k\nu_k\nu_k^T + \sum_{k=K+1}^{N}\lambda_k\nu_k\nu_k^T$$

where μ_k are, as in Fisher and Courtier (1995), some coefficients chosen such that $\mu_k\lambda_k < \lambda_{K+1}$; the condition number is reduced to $\kappa(\hat{\mathbf{J}"}) = \frac{\lambda_{K+1}}{\lambda_N}$.

An experiment was conducted in which the leading singular vectors of 3D-PSAS and 3D-Var Hessians were calculated using a Lanczos algorithm. Both methods were preconditioned with a set of the first 25, 50 and 75 singular vectors for $\mu_k = 1/\lambda_k$. In Fig.3–5 and Fig.3–6, the convergence of the objective functions and the reduction of their gradient norms show an important saving for both 3D-Var and 3D-PSAS since the convergence rate is increased with the number of singular vectors. With a small number of singular vectors ($K = 25$), the convergence is achieved with half of the required iterations.

It is then important to notice that the overall time required for the analysis does not reflect such a saving considering the additional cost of the computation of Hessian eigenpairs; a reasonable compromise is yet to be defined. One can also use a combined Lanczos-Conjugate-Gradient method that allows for the calculation of Hessian singular vectors in the course of the minimization process since there is a close connection between the gradient vectors of the CG and those generated by the Lanczos algorithm (Paige and Saunders, 1975; Fisher and Courtier, 1995). Thus the

number of iterations is comparable to the number of singular vectors, which makes this approach very useful in the context of operational processes.

When comparing the convergences in Fig.3–4, Fig.3–5 and Fig.3–6 for 3D-PSAS and 3D-Var, one can notice that both algorithms benefit from a refining of the global representation of their Hessians, either by increasing the number of pairs of the Quasi-Newton state vectors or by using their leading singular vectors.

3.4.3 Cycling the Hessian in PSAS

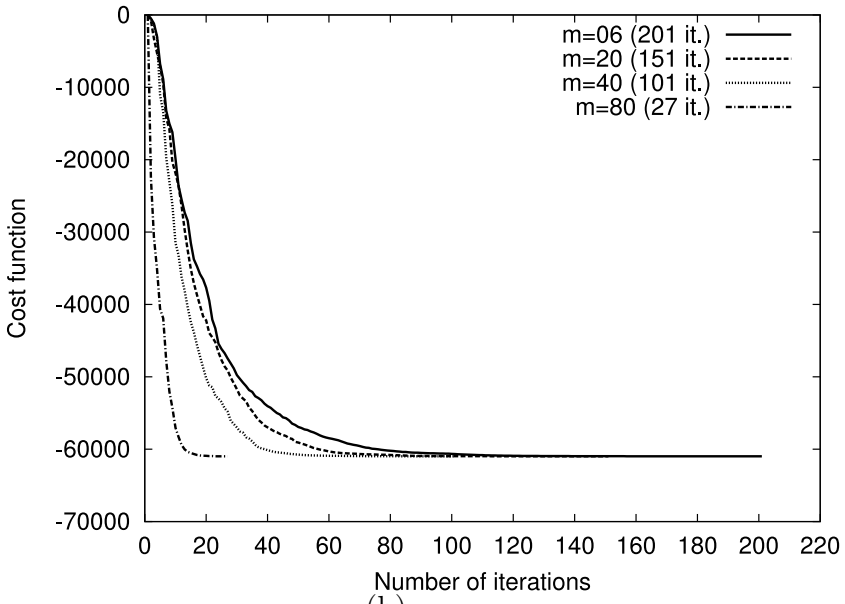
In sections 4.1 and 4.2, we examined the effect of approximating the Hessian in preconditioning the assimilation problem within the same analysis. In the context of an incremental variational assimilation, the approximation of the Hessian is also useful to precondition the next analysis or the next minimization after an outer loop. For 3D-Var, the Hessian and its singular vectors can easily be re-used due to the regularity of the control variable in model space, whereas for PSAS, the control variable, and consequently the Hessian, are completely defined in the dual space, and they change at every assimilation period due to the differences in the observations used. We introduce here, a way to overcome this limitation of PSAS by linking its Hessian singular vectors to those of 3D-Var and by operating a *remapping* of the approximate Hessian in the model space.

Relationship between model and the dual of observation space singular vectors

In section (3), we have seen the equivalence of 3D-PSAS and 3D-Var in terms of their convergence rate and their condition number as pointed out by Courtier (1997). It is possible to extend this equivalence to their Hessians singular vectors. By defining the operator $\mathbf{L} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}^{\frac{1}{2}}$, the Hessians of 3D-Var and 3D-PSAS in (3.6) and (3.12) can easily be written respectively as:

$$\mathbf{J}'' = \mathbf{I}_n + \mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{\frac{1}{2}} = \mathbf{I}_n + \mathbf{L}^T\mathbf{L} \tag{3.16}$$

(a)



(b)

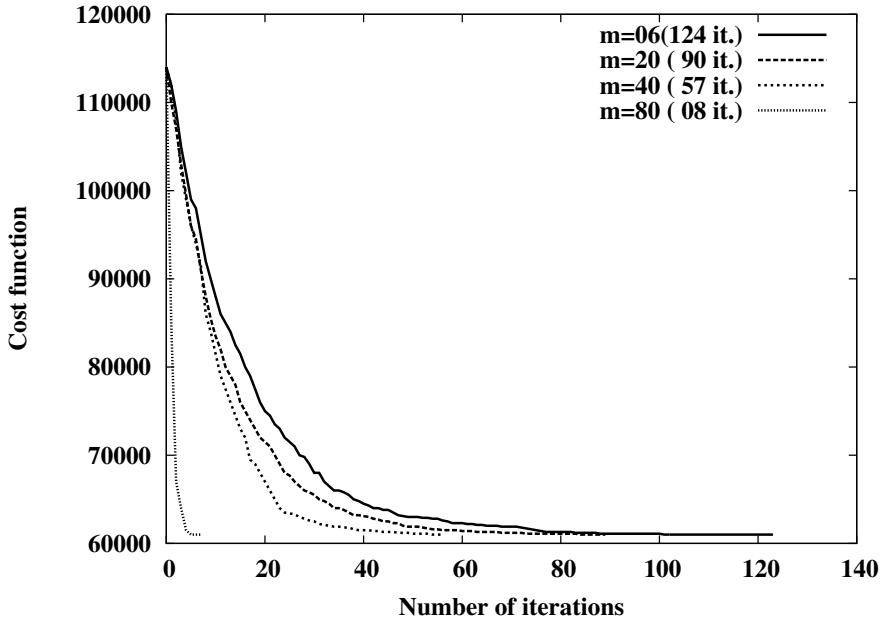


Figure 3–4: 3D-PSAS (a) and 3D-VAR (b) convergence with cycled Hessians for different number of pairs

$$\mathbf{F}'' = \mathbf{I}_p + \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}} = \mathbf{I}_p + \mathbf{L}\mathbf{L}^T. \tag{3.17}$$

Now, let us consider $(\mathbf{X}_i, \lambda_i)$ a set of orthonormal eigenpairs of $\mathbf{F}''$, for $i = 1, .., N$. One can show that

$$\mathbf{Y}_i = \frac{1}{\sqrt{\lambda_i - 1}}\mathbf{L}^T\mathbf{X}_i \tag{3.18}$$

form a set of orthonormal eigenvectors of $\mathbf{J}''$ for the same set of eigenvalues λ_i. Also, for each $(\mathbf{Y}_i, \lambda_i)$ an eigenpair of $\mathbf{J}''$, there is a unique corresponding eigenpair $(\mathbf{X}_i, \lambda_i)$ of $\mathbf{F}''$ that satisfies :

$$\mathbf{X}_i = \frac{1}{\sqrt{\lambda_i - 1}}\mathbf{L}\mathbf{Y}_i. \tag{3.19}$$

Relations (3.18) and (3.19) were confirmed by computing the first singular vectors of 3D-Var and 3D-PSAS and by comparing them when mapped with these relations. Linking Hessian singular vectors in both the model and the dual of observation space is useful for giving a physical meaning to PSAS Hessian through that given to 3D-Var singular vectors. Recall that the 3D-Var Hessian is closely related to the analysis covariance error matrix ($\mathbf{J}'' = \mathbf{A}^{-1}$). Also, following Cardinali *et al* (2004), this Hessian can be used to express the analysis sensitivity with respect to the observations or information content extracted from the available data. The sensitivity matrix is defined as follows: $\mathbf{S} = \mathbf{R}^{-1}\mathbf{H}\mathbf{A}^{-1}\mathbf{H}^T$. In the PSAS context, this information can then be retrieved since its Hessian is related to that of 3D-Var. $\mathbf{S}$ can be written as :

$$\begin{aligned} \mathbf{S} = &\quad \mathbf{R}^{-1}\mathbf{H}(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1} \\ = &\quad (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T \end{aligned}$$

which can be simplified by adding and subtracting an $\mathbf{R}$ in the second term:

$$\mathbf{S} = (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R} - \mathbf{R})$$

and finally we get :

$$\mathbf{S} = \mathbf{I} - (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}\mathbf{R}. \tag{3.20}$$

Following Cardinali *et al.* (2004), the term $(\mathbf{I}-\mathbf{S})$ represents the analysis sensitivity with respect to the background in the dual of observation space. This allows us to give a physical meaning to the inverse of PSAS Hessian. Moreover, considering relations (3.18) and (3.19), the analysis sensitivity can be calculated with the leading singular vectors of 3D-Var or equivalently those of 3D-PSAS.

Re-mapping of 3D-PSAS Hessian

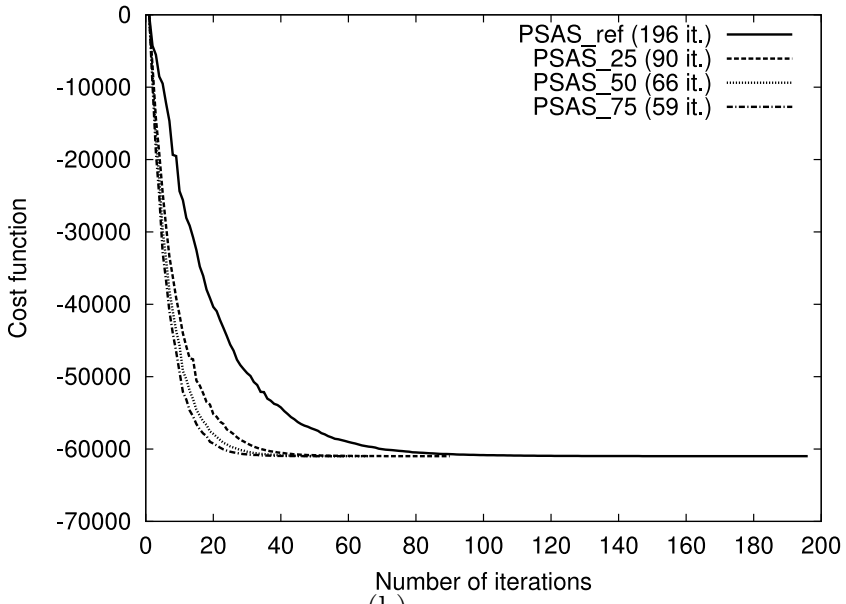
To cycle the PSAS Hessian, relations (3.18) and (3.19) link 3D-PSAS Hessian singular vectors to those of 3D-Var. It is now possible to calculate the singular vectors of 3D-PSAS Hessian in order to precondition the current analysis as in Fig.3–5 and Fig.3–6 (corresponding to time t_1), and at the same time use them as an approximation of those of the next cycle (time t_2).

This process consists in a first step in mapping a current given eigenpair of 3D-PSAS $(\mathbf{X}, \lambda)$ into the model space through the operator ${\mathbf{L}_1}^T$ (eq.3.18); the new vector can easily be exported towards the next window in the same way as in 3D-Var, and in a second step it consists in remapping this vector back into the dual of observation space using an updated operator $\mathbf{L}_2 = \mathbf{R}_2^{-\frac{1}{2}}\mathbf{H}_2\mathbf{B}^{\frac{1}{2}}$. $\mathbf{H}_2$ and $\mathbf{R}_2$ are the new observation operator and covariance matrix valid at time t_2. The approximate eigenpair for the next window would then be:

$$\tilde{\mathbf{X}} = \frac{1}{\lambda - 1}\mathbf{L}_2{\mathbf{L}_1}^T\mathbf{X}. \tag{3.21}$$

In Fig.3–7, the convergence of 3D-PSAS objective function is plotted for the 19 January 2004 at 06h (solid curve), along with a preconditioned function with current singular vectors as done in the previous section (dashed curve) and with singular vectors generated from the previous cycle at 00h and remapped using (3.21) (dotted curve). There is no significant difference in the two experiments, which states that not only 3D-PSAS Hessian singular vectors can be re-mapped in both the model and the dual of observation spaces, they also allow for cycling the Hessian from

(a)



(b)

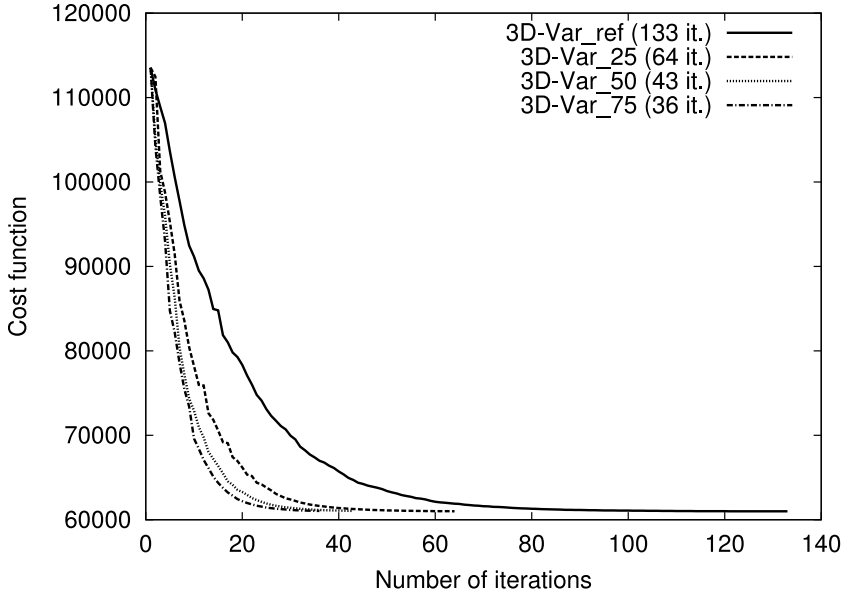
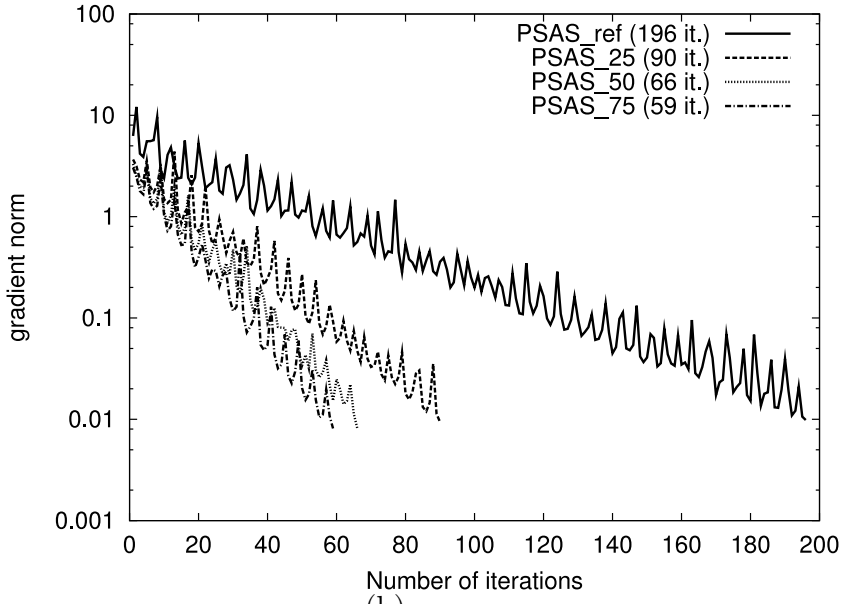


Figure 3–5: 3D-PSAS (a) and 3D-Var (b) objective functions convergence with preconditioned Hessian with 25, 50 and 75 leading singular vectors.

(a)



(b)

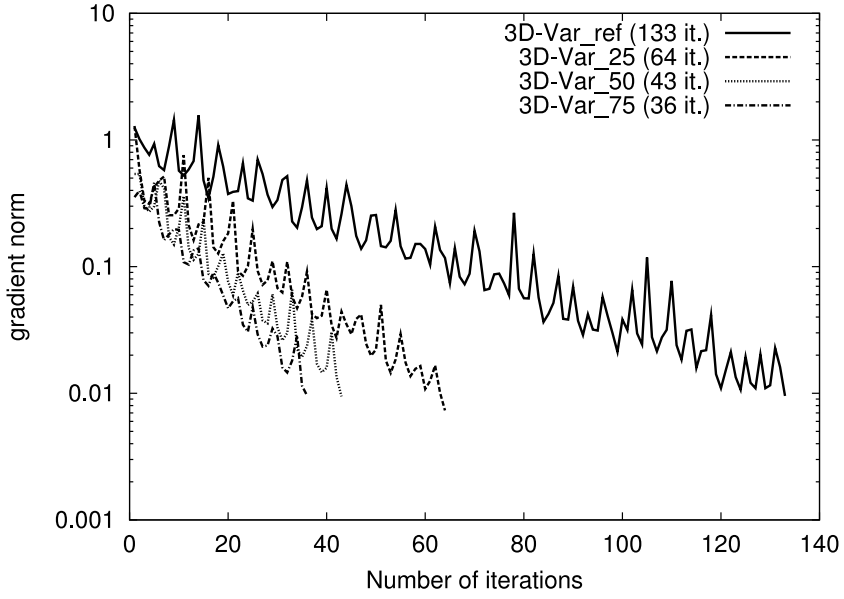


Figure 3–6: Gradient norms of 3D-PSAS (a) and 3D-Var (b) objective functions with preconditioned Hessians with 25, 50 and 75 leading singular vectors (logarithmic scale).

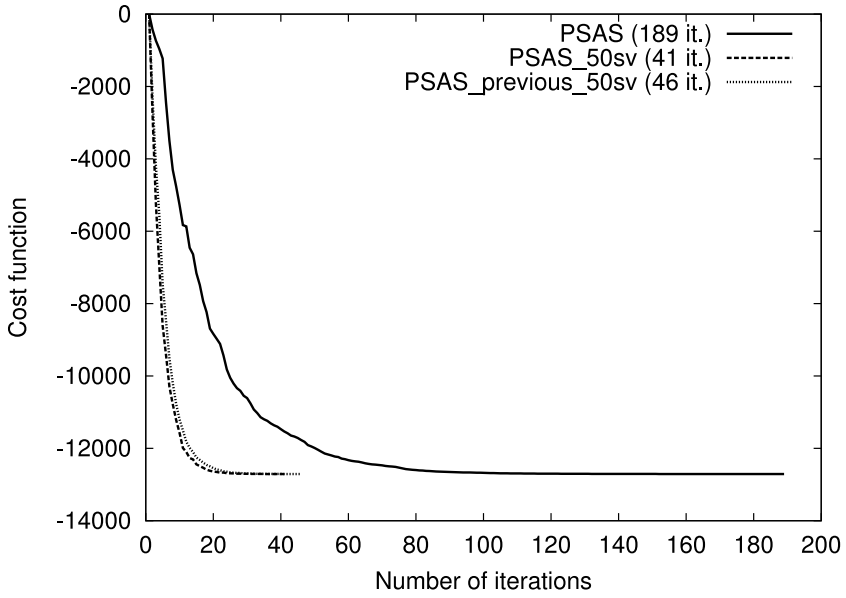


Figure 3–7: Cycling PSAS Hessian with previous assimilation window singular vectors for January 19, 2004 at 06h.

one assimilation to the next for an incremental PSAS which was an important issue when working in the dual space. This behavior was observed for several similar experiments. However, further experiments have though to be carried out in order to examine the sensitivity of eigenvectors to the observation network used.

3.5 Conclusion

Theoretical equivalence of 3D/4D-PSAS and 3D/4D-Var has been confirmed through an intercomparison in an operational context. Both algorithms have been shown to have the same convergence properties. However, the PSAS objective function has no physical meaning whereas that of 3D-Var can be related to the a posteriori probability density function. Our results show that the first iterates of the PSAS minimization lead to non-physical estimates and that it takes several iterates to obtain a model state that is more accurate than the background state. This is a cause of concern given that most operational implementations perform only a finite number of iterations. Equivalence has also been extended to the singular vectors of

Hessians, which was shown to have interesting applications to establishing a connexion between PSAS Hessian and the analysis sensitivities to the observations and/or to the background. This also makes it possible to re-map singular vectors in the model and the dual of observation space in order to precondition the PSAS problem when cycling the assimilation. This work has also highlighted the importance of preconditioning both dual and primal formulations well by choosing the appropriate approximation of the Hessian taking into account the balance between the affordable number of singular vectors or the number of pairs of the Limited-Memory Quasi Newton vectors for instance and the subsequent gain.

As discussed earlier, PSAS can be seen as an alternative to the 3D/4D-Var for solving the same analysis problem, with the same cost; and the equivalence between these two methods is still relevant in the non-linear case when using the incremental formulation. However, interest for PSAS lies beyond this equivalence since working in a reduced space (the dual of observation space) is particularly promising when accounting for model errors in a weak-constraint formulation, which is still unaffordable for a real-size weak-constraint 4D-Var application. As discussed in Courtier (1997), model errors can be accounted for in a PSAS context without any significant increase in the size of the four dimensional problem, which is of great interest provided model error covariances are specified. The dual approach can also be used to compute sensitivities with respect to observations as proposed by Langland and Baker (2004) and as mentioned in Pellerin *et al* (2006). This is currently examined in an ongoing study.

Acknowledgments

This work was funded in part by grant GR-500B-THORPEX from the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) to the project Impact of observing systems and predictability on forecasting extreme weather events in the short, medium and extended range: a Canadian contribution to THORPEX.

4 Convergence properties of the primal and dual forms of variational data assimilation

In the previous chapter, the minimization of the dual method was shown to exhibit a spurious behavior at the beginning of the minimization which leads to less probable states than the background state. Although the dual analysis result after convergence is equivalent to the primal one, this is a serious concern when using the dual method in operational implementations since only a finite number of iterations can be afforded. This chapter examines this convergence problem by comparing the sensitivity of the dual method to two minimization algorithms: the Conjugate Gradient (CG) and the MINimum RESidual (Minres). For this purpose, we chose to use a controlled frameword; a two dimensional turbulence model setting instead of a full operational context. By reducing the complexity of our system, we aim to monitor the origins of the dual problem more closely.

This Chapter is based on the following paper:

El Akkraoui A. and P. Gauthier, 2010 : *Convergence properties of the primal and dual forms of variational data assimilation.* Q.J.R. Met. Soc., **136**, Issue 646, 107-115.

Convergence properties of the primal and dual forms of variational data assimilation

El Akkraoui A.[1], and P. Gauthier[12]

[1] Department of Atmospheric and Oceanic Sciences, McGill University, Canada

[2]Department of Atmospheric and Earth Sciences, Université du Québec à Montréal, Canada

Abstract

The variational data assimilation problem can be solved in either its primal (3D/4D-Var) or dual form (3D/4D-PSAS). Both methods are equivalent at convergence but the dual method exhibits a spurious behaviour at the beginning of the minimization which leads to less probable states than the background state. This is a serious concern when using the dual method in operational implementations when only a finite number of iterations can be afforded. Two classes of minimization algorithms are examined in this paper: the Conjugate Gradient (CG) and the Minimum Residual (Minres) methods. While the conjugate gradient algorithms ensure a monotonic reduction of the cost function, those based on the minimum residual enforce instead a monotonic decrease of the norm of the gradient. In this paper, it is shown that when applied to the minimization of the dual problem, the minimum residual algorithms also lead to iterates for which their "image" in physical space leads to a monotonic decrease of the primal cost function. A relationship is established showing that the primal objective function is related to the value of the dual cost function and the norm of its gradient. This holds for the incremental forms of both the three and four dimensional cases. A new convergence criterion is introduced based on the error norm in model space to make sure that, for the dual problem, the same accuracy is obtained in the analysis when only a finite number of iterations is completed.

4.1 Introduction

Variational data assimilation methods have been widely used in meteorological and oceanographic applications to estimate the analysis state of the atmosphere or the ocean. Several NWP centres are currently using the four-dimensional variational scheme (4D-Var), such as the European Centre for Medium-range Weather Forecast (ECMWF; Rabier *et al.*,2000), the UK Met Office (Rawlins *et al.*, 2007), Météo-France (Gauthier and Thépaut, 2001) and Environment Canada (Gauthier *et al.*, 2007). The 4D-Var was made possible and practically affordable by using the adjoint methods introduced by Lewis and Derber (1985), LeDimet and Talagrand (1986) and Talagrand and Courtier (1987). When formulated in model-space, the variational data assimilation problem (3D/4D-Var) corresponds to the *primal form*. Introduced by da Silva *et al.* (1995) and Cohn *et al.* (1998), the Physical Space Assimilation system (PSAS) formulates instead the variational problem in observation space: this is often referred to as the 3D/4D-PSAS. Courtier (1997) showed the theoretical equivalence between the two forms, which are two formulations of the same problem. The PSAS is in fact the *dual form* of the primal variational problem. In El Akkraoui *et al.* (2008), this equivalence was examined and a relationship between the Hessians of the two problems was introduced. When the Hessian is represented by a finite number of singular vectors, those can be mapped to the singular vectors of the Hessian of the other problem.

In recent years, several studies were carried out to investigate ways to account for model error in the assimilation. This requires also to be able to specify and estimate model error covariances and to find effective ways to solve the resulting weak constraint problem (Vidard *et al.*, 2004; Griffith and Nichols, 2001; Trémolet, 2007). The interest for the weak constraint 4D-Var stems from the fact that model error imposes a limitation on the length of the assimilation window (Trémolet, 2007). Courtier (1997) showed that the weak constraint 4D-Var has also a dual form which,

at convergence, is also equivalent under some conditions. The dual method is particularly interesting for the weak constraint 4D-Var. In the primal form, the size of the control variable increases significantly in the weak constraint as it corresponds to the model trajectory over the whole assimilation window. However, the dual form is commensurate with the number of observations and the dimension of the problem does not experience a significant increase. The issues raised by El Akkraoui *et al.* (2008) regarding the convergence of the dual problem are then a source of concern for the development of a weak constraint 4D-Var in its dual form.

In variational assimilation, iterative methods are used, which require adequate preconditioning to speed up the convergence. Various preconditioning techniques have been developed altogether with more efficient minimization algorithms. The quasi-Newton (Gilbert and Lemaréchal, 1989) and the Conjugate Gradient (Hestenes and Stiefel, 1952) are widely used in variational data assimilation applications in NWP centers, although many other iterative methods for large sparse problems are proposed in the literature. The Conjugate Gradient and quasi-Newton algorithms (Navon and Legler, 1987) are based on a monotonic reduction of the functional. Another approach is to impose that the norm of the gradient be monotonically decreased: this leads to the Minimum residual (MINRES) method (Paige and Saunders, 1975), that can be extended to the generalized minimum residual (GMRES) proposed by Saad and Schultz (1986). All of these are basically Krylov subspace methods. The choice of the appropriate method to use is usually dictated by the prior knowledge of the problem to solve (i.e. sparseness, definite positiveness, spectrum distribution, etc.) and by the overall cost of the algorithms. An overview of Krylov subspace methods is given in Freund *et al.* (1992), Greenbaum (1997) and Saad (1996). The reader is also referred to Zhou and Walker (1994), Cullum and Greenbaum (1996) and Paige *et al.* (1995) for a comparison of two of these methods: the conjugate gradient and the minimum residual.

In El Akkraoui *et al.* (2008), the equivalence of the primal and dual forms was established within the framework of an operational 3D-Var. Although the two approaches lead to identical results at convergence, their results indicated that the dual form may exhibit unphysical states at the beginning of the minimization, which can add a significant number of iterations to reach convergence. This is a source of concern for operational implementations that are constrained to a limited number of iterations. The objective of this paper is to reexamine the minimization process in the dual space and to better understand this convergence problem.

The paper is organized as follows. A review of the primal and dual formulations of variational methods is presented in section 2, along with a description of the behaviour of the dual problem in the course of the minimization. In section 3, the Conjugate Gradient and the Minimum Residual methods are presented; convergence criteria for the primal and dual methods are examined and a stopping criterion is proposed based on the error norm in model space. The experiments and results are presented in section 4, where performing the minimization with Minres is shown to have a positive impact in the dual case, eliminating the convergence problem mentioned earlier for the first iterations in the 3D/4D-PSAS minimization. The new convergence criterion is shown to be more appropriate for the dual case.

4.2 3D-Var and 3D-PSAS

In variational data assimilation, the best linear unbiased estimate (or maximum likelihood) of the state of the atmosphere at the analysis time is obtained by minimizing a functional representing the distance of a model state $\mathbf{x}$ to the background state $\mathbf{x}_b$ and to observations $\mathbf{y}$, both weighted by their respective error covariances. As presented in Lorenc(1986) and Rodgers (2000), this functional is

$$J(\mathbf{x}) = -ln(p(\mathbf{x}|\mathbf{y})),$$

where $p(\mathbf{x}|\mathbf{y})$ is the conditional probability that the state vector $\mathbf{x}$ is true knowing that $\mathbf{y}$ has been observed. Higher values of $J(\mathbf{x})$ then correspond to less probable states.

When the background and observation error statistics are Gaussian, the incremental 3D-Var scheme has to minimize the functional

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{y}') \tag{4.1}$$

where $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices respectively, the increment $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$ represents the misfit between the background state $\mathbf{x}_b$ and the model state $\mathbf{x}$, $\mathbf{H}$ is the Jacobian (or Tangent linear) of the non-linear observation operator ($\mathcal{H}$) that maps the model variables into observation space, and the term $\mathbf{y}' = \mathbf{y} - \mathcal{H}(\mathbf{x}_b)$ represents the "innovation" vector or the misfit between the observation state $\mathbf{y}$ and the background state in physical space.

As shown by Courtier (1997) and discussed in El Akkraoui *et al* (2008), the variational data assimilation can be expressed either in its primal or dual form, where the latter solves the variational problem in the dual space instead of the model space in which the primal formulation is cast. The dual objective function to be minimized is :

$$F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)\mathbf{w} - \mathbf{w}^T\mathbf{y}'. \tag{4.2}$$

where its minimum ($\mathbf{w}_a$) satisfies $\delta\mathbf{x}_a = \mathbf{B}\mathbf{H}^T\mathbf{w}_a$. This guarantees that both formulations will give the same results at convergence. Courtier (1997) pointed out that the equivalence between the primal and dual formulations leads to two minimization problems for which the Hessians of the quadratic functionals have the same condition number.

Preconditioned formulations for both forms, with their own preconditioning, are obtained by introducing the changes of variable $\mathbf{v} = \mathbf{B}^{-\frac{1}{2}}\delta\mathbf{x}$ for 3D-Var and $\mathbf{u} = \mathbf{R}^{\frac{1}{2}}\mathbf{w}$

for 3D-PSAS, to get the functionals and their gradients, which are respectively

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T(\mathbf{I}_n + \mathbf{L}^T\mathbf{L})\mathbf{v} - \mathbf{v}^T\mathbf{L}^T\tilde{\mathbf{y}} + \frac{1}{2}\tilde{\mathbf{y}}^T\tilde{\mathbf{y}}, \tag{4.3}$$

$$\nabla_{\mathbf{v}}(J) = (\mathbf{I}_n + \mathbf{L}^T\mathbf{L})\mathbf{v} - \mathbf{L}^T\tilde{\mathbf{y}}, \tag{4.4}$$

and

$$F(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T(\mathbf{I}_m + \mathbf{L}\mathbf{L}^T)\mathbf{u} - \mathbf{u}^T\tilde{\mathbf{y}}, \tag{4.5}$$

$$\nabla_{\mathbf{u}}(F) = (\mathbf{I}_m + \mathbf{L}\mathbf{L}^T)\mathbf{u} - \tilde{\mathbf{y}}, \tag{4.6}$$

where $\tilde{\mathbf{y}} = \mathbf{R}^{-\frac{1}{2}}\mathbf{y}'$, and $\mathbf{L} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}^{\frac{1}{2}}$. This shows that the Hessians are

$$\mathbf{J}'' = \mathbf{I}_n + \mathbf{L}^T\mathbf{L}, \quad \mathbf{F}'' = \mathbf{I}_m + \mathbf{L}\mathbf{L}^T$$

where $\mathbf{I}_n$ and $\mathbf{I}_m$ are the identity matrices in model and observation space of order n and m respectively.

Since the dual objective function has no immediate physical interpretation, the *a posteriori* probability distribution, as measured by the primal functional $J(\mathbf{v})$, can be estimated by mapping each dual iterate $\mathbf{u}_k$ to physical space ($\mathbf{v}_k = \mathbf{L}^T\mathbf{u}_k$) and then evaluating $J(\mathbf{v}_k)$. The result of this process is shown in Figure 4–1, from El Akkraoui *et al.* (2008), where the iterates associated with the minimization of $F(\mathbf{u})$ were found to lead to less probable states than the background state at the beginning of the minimization. This is not desirable if the minimization has to stop after a finite number of iterations due to computational time constraints. The gradient of the 3D-PSAS functional shows also an increase of its norm at the first iteration as shown in Figure 4–2, from El Akkraoui *et al.* (2008), and it takes several iterations to come back to its original level. These results were obtained with a Limited Memory Quasi Newton (LMQN) for the minimization and a similar result (not shown) was found using a Conjugate Gradient (CG) algorithm.

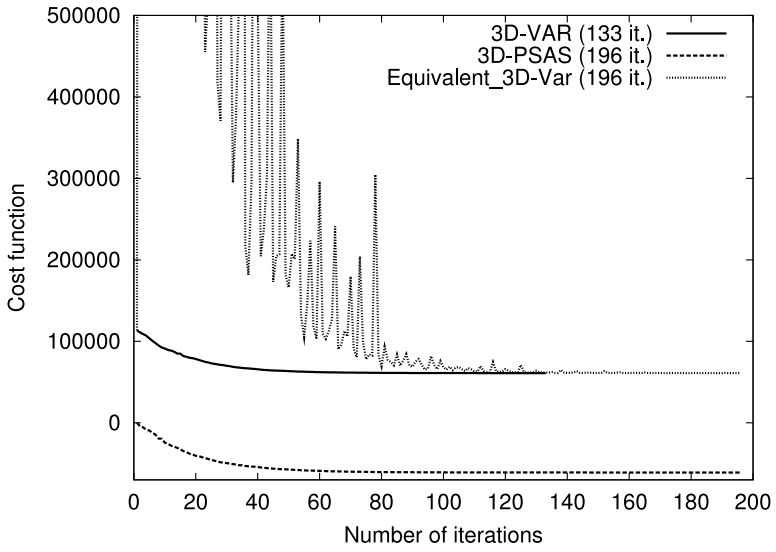


Figure 4–1: Objective function of 3D-Var (solid line), 3D-PSAS (dashed line) and the equivalent 3D-Var at PSAS iterates (dotted line) along the minimization. At each PSAS iteration k, the iterate $\mathbf{u}_k$ is brought to the model space through the operator $\mathbf{L}^T = \mathbf{B}^{\frac{1}{2}}\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}}$ and the 3D-Var objective function is calculated for $\mathbf{v}_k = \mathbf{L}^T\mathbf{u}_k$. (From El Akkraoui *et al.*, 2008)

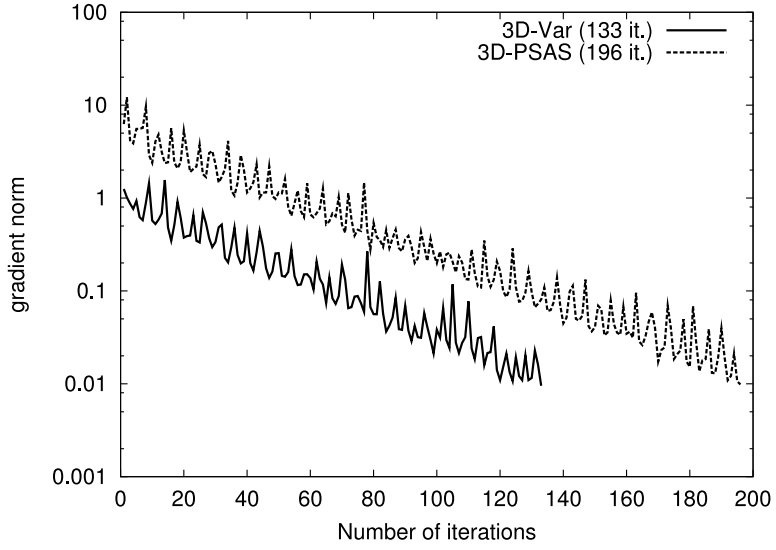


Figure 4–2: The ratio of the gradient norm at each iteration to the initial norm of 3D-Var (solid line) and 3D-PSAS (dashed line) with number of iterations, plotted in a logarithmic scale.(From El Akkraoui *et al.*, 2008)

It is shown in Appendix (B) that, using equations (4.3), (4.5), and when $\mathbf{v}_k = \mathbf{L}^T\mathbf{u}_k$, the primal function is related to the iterates of the dual problem by

$$J(\mathbf{v}_k) = \frac{1}{2}\|\nabla F(\mathbf{u}_k)\|^2 - F(\mathbf{u}_k). \tag{4.7}$$

The norm used here is the Euclidean norm $\|\mathbf{v}\|^2 = \mathbf{v}^T\mathbf{v}$.

As discussed earlier, the functional $J(\mathbf{x})$ corresponds to the *a priori* probability distribution function and reducing J then implies increasing the probability of $\mathbf{x}$ being the true value.

In view of (4.7), one can see that the behaviour shown in Figure 4–1 can be explained by the dominance of the gradient norm term in the first iterations. Since both the CG and the LMQN algorithms do not impose any constraint on the norm of the gradient, it would be more appropriate to consider other minimization algorithms that require the gradient norm to decrease monotonically. This leads to another class of minimization algorithms used to solve linear systems. In the following section, the Conjugate Gradient and the Minimum residual (Minres) methods are reviewed and compared.

4.3 Minimization algorithms

Solving large linear systems of equations is often cast as a minimization problem in which the gradient of a quadratic objective function will vanish when the solution of

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{4.8}$$

is reached. In the primal and dual assimilation problems, the matrix $\mathbf{A}$ is symmetric and positive definite and corresponds to the Hessians $\mathbf{J}$" and $\mathbf{F}$", and $\mathbf{b}$ to the terms $\mathbf{L}^T\tilde{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ respectively.

Different iterative methods for solving the linear system (4.8) are described in the literature. They can be regrouped in two classes: the Conjugate Gradients (CG) and the minimum residuals methods (MINRES). While the former is designed for symmetric positive definite matrices $\mathbf{A}$, the latter can be used for indefinite symmetric matrices as well. Some algorithmic and derivation aspects of both methods are presented in Appendix C. The reader is referred to Greenbaum (1997) for more details and for an overview of their convergence properties.

The following notations will be used hereafter. At each iteration k, the residual is defined as $\mathbf{r}_k = \mathbf{A}\mathbf{x}_k - \mathbf{b}$, which also represents the gradient of the objective function associated with (4.8), and the error is $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is the exact solution of (4.8). Here, $\|.\|$ and $\|.\|_{\mathbf{A}}$ will denote respectively the Euclidean 2-norm and the A-norm, where $\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x}$ and $\|\mathbf{x}\|^2_{\mathbf{A}} = \mathbf{x}^T\mathbf{A}\mathbf{x}$.

4.3.1 The Conjugate Gradient

A review of conjugate gradients and quasi-Newton methods for large-scale minimization in meteorology is presented in Navon and Legler (1987). At each iteration, the iterate is obtained by finding that point along the search direction that minimizes the cost function. To make the minimization efficient, this class of algorithms constructs its k^{th} iterate $\mathbf{x}_k$ as an element of the Krylov space $\mathbf{x}_0 + span\{\mathbf{r}_0, \mathbf{A}\mathbf{r}_0, ..., \mathbf{A}^{k-1}\mathbf{r}_0\}$ such that the objective function is minimized (Golub and Van Loan, 1996). This can also be viewed as an iterative reduction of the A-norm of the error $\|\mathbf{x}_k - \hat{\mathbf{x}}\|_{\mathbf{A}}$.

It can be shown that this norm is related to the residual by

$$\|\mathbf{x}_k - \hat{\mathbf{x}}\|_{\mathbf{A}} = \|\mathbf{r}_k\|_{\mathbf{A}^{-1}},$$

where $\|\mathbf{r}_k\|^2_{\mathbf{A}^{-1}} = \mathbf{r}_k^T\mathbf{A}^{-1}\mathbf{r}_k$. The stopping criterion of the CG algorithm is that the norm of the residual (or the gradient) relative to the initial residual norm is below

the convergence criterion specified by the parameter ϵ. In other words,

$$\xi_k^2 = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_0^T \mathbf{r}_0} \leq \epsilon. \tag{4.9}$$

It is important to remember that the actual criterion used to complete an iteration with those algorithms is *not* a reduction of the norm of the gradient but a reduction in the cost function. Figure 4–2 shows that even near the end of the minimization, significant increases in the norm of the gradient can be observed. Therefore, in practice, the stopping criterion is often set to ensure that the convergence criterion (4.9) is met for several consecutive iterations.

4.3.2 Minimum Residual algorithm : Minres

First introduced by Paige and Saunders (1975), Minres is a Lanczos-based iterative method for solving indefinite symmetric linear systems. Denoting $\mathcal{K}_k(\mathbf{A}, \mathbf{b})$ the k^{th} Krylov subspace of $\mathbf{A}$ and $\mathbf{b}$, $\mathcal{K}_k(\mathbf{A}, \mathbf{b}) := span\{\mathbf{b}, \mathbf{A}\mathbf{b}, ..., \mathbf{A}^{k-1}\mathbf{b}\}$, the minimum residual method namely consists on finding $\mathbf{x}_k \in \mathcal{K}_k(\mathbf{A}, \mathbf{b})$ such that $\|\mathbf{r}_k\|$ is reduced at each iteration, which amounts to imposing that the norm of the gradient is monotonically decreasing during the minimization. The reader is referred to Paige and Saunders (1975) and Choi (2006) for details on this algorithm.

It can be shown that the reduction of the residual norm is such that

$$\|\mathbf{r}_k\| = S_k \|\mathbf{r}_{k-1}\|$$

where S_k is constructed such that $0 < S_k \leq 1$. The monotonicity of $\|\mathbf{r}_k\|$ in Minres would then prevent the occurrence of the large increases in the norm of the gradients observed when minimizing the dual problem (4.5) as shown in Figure 4–2.

4.3.3 Convergence properties

Convergence behaviour of CG and Minres is completely determined by the spectrum of $\mathbf{A}$. A classical approach for the CG is to consider the upper bound of the

A-norm of the error based on the condition number of $\mathbf{A}$ (see Golub and Van Loan (1996) and Kaniel (1966)). The convergence rate of MINRES on the other hand is less straightforward to analyze. This is discussed in Kilmer and Stewart (1999).

Ideally, iterative processes should stop when the error norm falls below a threshold defined by the user. However, since the error is not known, most stopping criteria are built upon the estimation of the residual norm. In Barrett *et al.* (1994), a compilation of commonly used criteria is presented. We focus here on the following termination criterion (Dennis and Schnabel, 1983) :

$$R_1 = \frac{\|\mathbf{r}_k\|}{\|\mathbf{r}_o\|} \leq \epsilon, \quad \mathbf{r}_o = \mathbf{b}$$

which specifies the error bound : $\|\mathbf{e}_k\| \leq \epsilon \|\mathbf{A}^{-1}\| \|\mathbf{b}\|$.

As discussed in Greenbaum (1997) and Zhou and Walker (1994), the norms of the CG and Minres residuals, denoted $\mathbf{r}^c$ and $\mathbf{r}^m$ respectively, are connected at each iteration k by :

$$\frac{\|\mathbf{r}_k^m\|^2}{\|\mathbf{r}_k^c\|^2} = 1 - \frac{\|\mathbf{r}_k^m\|^2}{\|\mathbf{r}_{k-1}^m\|^2} \tag{4.10}$$

From this formula, and since both algorithms start with $\mathbf{x}_0 = 0$, that is $\mathbf{r}_0^m = \mathbf{r}_0^c = \mathbf{b}$, one can see that

$$\frac{\|\mathbf{r}_k^m\|}{\|\mathbf{r}_0^m\|} \leq \frac{\|\mathbf{r}_k^c\|}{\|\mathbf{r}_0^c\|}$$

at each iteration k. This suggests that Minres would need slightly less iterations than the CG to converge when using the criterion based on R_1 to terminate the minimization. However, relation (4.10) constrains the residual norms to not depart much from each other. That is, when Minres convergence is fast, the right hand side of (4.10) tends to 1 and $\|\mathbf{r}_k^m\|^2 \approx \|\mathbf{r}_k^c\|^2$ and when Minres is close enough to convergence (that is, the residual is very small, $\|\mathbf{r}_k^m\|^2 \approx 0$), then the CG residual is also small enough for the minimization to terminate.

4.3.4 Primal and dual convergence

Given the relationship (4.7) that links the dual objective function $F(\mathbf{u_k})$ to the primal function $J(\mathbf{v_k})$, we now establish a stopping criterion for the minimization of the dual problem that will result in the same accuracy as that obtained when minimizing the primal problem.

The Hessians of both formulations having the same spectrum (Courtier, 1997), they have the same convergence rate as shown in Figure 2 (the same slope in the gradient curves), which indicates that the two cases yield similar rates of reduction in the norm of the gradients. However, since the minimizations are performed in two different spaces, a proper termination criterion must be specified in each case to ensure that a similar accuracy in the analysis is achieved.

The termination criterion R_1 presented above can be interpreted in terms of the error in the analysis. This criterion means that the minimization should be stopped when the $\mathbf{A}^2$-norm of the error is such that

$$\frac{\|\mathbf{e}_k\|_{\mathbf{A}^2}}{\|\mathbf{e}_0\|_{\mathbf{A}^2}} \leq \epsilon$$

as it can easily be shown that $\|\mathbf{r}_k\|^2 = \|\mathbf{e}_k\|_{\mathbf{A}^2}$.

For the primal form, this error is $\mathbf{e}_k = \hat{\mathbf{v}} - \mathbf{v}_k$, which corresponds to the departure of the iterate $\mathbf{v}_k$ from the truth $\hat{\mathbf{v}}$. In the same way, the error defined in the dual space $\delta_k = \hat{\mathbf{u}} - \mathbf{u}_k$ will represent the departure of the iterates $\mathbf{u}_k$ from some *truth* in the dual space $\hat{\mathbf{u}}$, where $\hat{\mathbf{v}} = \mathbf{L}^T\hat{\mathbf{u}}$. It follows that the error in the dual space is related to the error in model space by the same operator $\mathbf{L}^T$, and we get

$$\|\mathbf{e}_k\|_{\mathbf{A}^2} = \|\mathbf{L}^T\delta_k\|_{\mathbf{A}^2}. \tag{4.11}$$

Since both the primal and dual formulations solve the same variational problem, the expected accuracy of the solution should be defined in the same reference space, say the model space. The dual minimization should then terminate not when the $\|\delta_k\|_{\mathbf{A}^2} \leq \epsilon$, but when the image of this error in the model space satisfies

$\|\mathbf{L}^T\delta_k\|_{\mathbf{A}^2} \leq \epsilon$. Thus, when comparing performances of the primal and dual algorithms, we introduce an appropriate termination criterion ($R2$) in the dual space such that :

- R_1 : stop the primal minimization when $\frac{\|\mathbf{r}_k\|}{\|\mathbf{r}_0\|} \leq \epsilon$.
- R_2 : stop the dual minimization when $\frac{\|\mathbf{L}^T\mathbf{r}_k\|}{\|\mathbf{L}^T\mathbf{r}_0\|} \leq \epsilon$.

Note that this new criterion implies that an additional integration of the adjoint model in $\mathbf{L}^T$ must be done just to test whether convergence has been reached. This could add a significant computing cost to 4D-PSAS for instance. However, if time is an issue, this may be done only at the end of the minimization to check how close we are to convergence.

4.4 Results

4.4.1 Experimental setting

The experiments are set up on a periodic β plane as results to be presented later will be extended to 4D-Var experiments. The model that will be used is the same two dimensional barotropic non-divergent model used by Tanguay *et al* (1995), and Laroche and Gauthier (1998), to solve the barotropic vorticity equation on the β-plane :

$$\frac{\partial \zeta}{\partial t} + J(\Psi, \zeta) + \beta v = f - D(\zeta),$$

where $\Psi = -U_0 y + \psi$ is the full stream function, with U_0 a large-scale East-West flow representing the mean zonal wind, and ψ represents the stream function. The vorticity is the Laplacian of the stream function, $\zeta = \nabla^2\psi$, and the horizontal wind components are $u = -\frac{\partial \psi}{\partial y}$, and $v = \frac{\partial \psi}{\partial x}$, J being the Jacobian, f a forcing term, and D, a linear dissipation operator. A detailed description of this model can be found in Tanguay *et al.* (1995), and Laroche and Gauthier (1998).

The assimilation is performed with perfect observations of wind components at each model grid point (and at every six time steps of the model for the 4D experiments). The observation error is assumed to be uncorrelated, thus the observation

error covariance matrix $\mathbf{R}$ is diagonal. The model state is the barotropic vorticity in Fourier space so that the observation operator converts vorticity into wind components, then performs an inverse Fourier transform. A background state is built from a previous run of the model for which the initial state has slightly been perturbed, and the background error covariance matrix is constructed such that $\mathbf{B} = \mathbf{\Sigma C \Sigma}$, where $\mathbf{\Sigma}$ is the diagonal matrix of the standard deviation and $\mathbf{C}$ represents homogeneous isotropic correlations (See Appendix B in Gauthier *et al.*, 1993). The correlation length is set to three grid points, which corresponds to approximately 300 km since our domain size corresponds roughly to 7000 km (Tanguay et *al.*, 1995). 3D/4D-var and 3D/4D-PSAS assimilation experiments were conducted with minimizations performed with both Minres and the Conjugate Gradient for comparison and stopped using the termination criterion R_1, for $\epsilon = 10^{-4}$.

4.4.2 Convergence Results

The objective functions of the three dimensional primal and dual methods are plotted in Figure 4–3, where the minimization is performed with the CG and then with Minres for comparison. Since the minima of the functionals must satisfy $J(\delta\mathbf{x}^*) = -F(\mathbf{w}^*)$ (See Appendix A), the dual functions in Figure 4–3 are plotted in their absolute value. Note that this relationship can be used as an *a posteriori* diagnostic of the minimization.

Objective functions curves plotted in Figure 4–3 indicate that for the primal case, Minres and the CG behave the same way but, in the dual case, the CG seems to perform better in the early stages of the minimization. By construction, the CG minimizes the $\mathbf{A}$-norm of the error, that is, the functional itself, whereas Minres minimizes the 2-norm of the residual, that is, the gradient norm, the functional is reduced monotonically, but this is not imposed in Minres.

Figure 4–4 shows the reduction of the norm of the residual to its initial value $\frac{\|\mathbf{r}_k\|}{\|\mathbf{r}_0\|}$ at each iteration k, using the CG and Minres respectively. As expected, the gradient

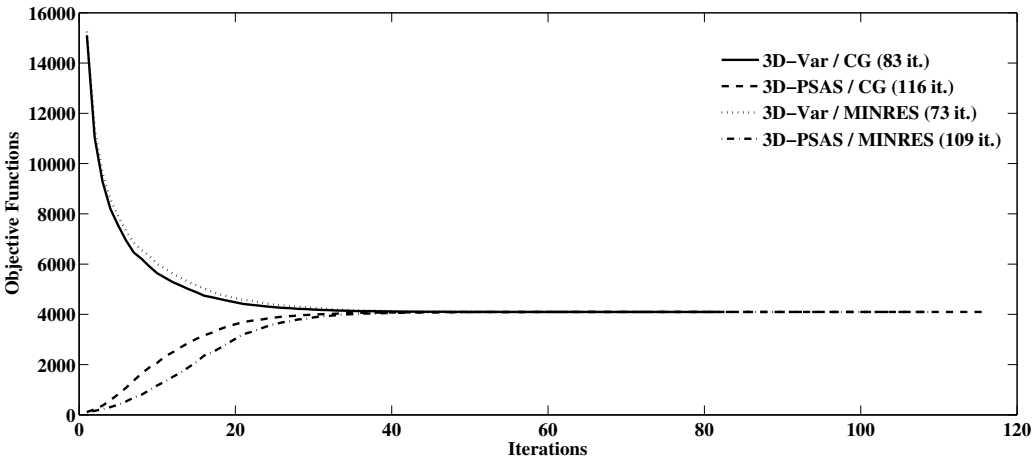


Figure 4–3: Objective function of 3D-Var performed with a CG (solid line) and Minres (dotted line), and the inverse of 3D-PSAS functional with the CG (dashed line) and minres (dash-dotted line).

norm is monotonically decreasing with Minres. Furthermore, for both primal and dual minimizations, Minres needs less iterations to reach the convergence criterion ($\frac{\|\mathbf{r}_k\|}{\|\mathbf{r}_0\|} \leq \epsilon$) than the Conjugate Gradient, which is consistent with the remark made in the previous section. As the stopping criterion is based on a reduction of the norm of the residual, Minres ensures that further iterations will not be increasing the norm of the gradient as it may happen when methods based on the conjugate gradient are used.

One notices however, that 3D-PSAS still requires more iterations than the 3D-Var in both minimization cases. As pointed out by Courtier (1997), both 3D-PSAS and 3D-Var are equivalent in terms of the overall cost, since they are expected to converge at the same rate and with a comparable number of iterations. As discussed in the previous section, performing the minimizations in two different spaces makes it necessary to adjust the stopping criterion for the dual case to be consistent with that in the model space. One way to do that is to map the dual residuals in the model space through the relation $\tilde{\mathbf{r}}_k = \mathbf{L}^T \mathbf{r}_k$; their norms are calculated in Figure 4–4 (the starred line).

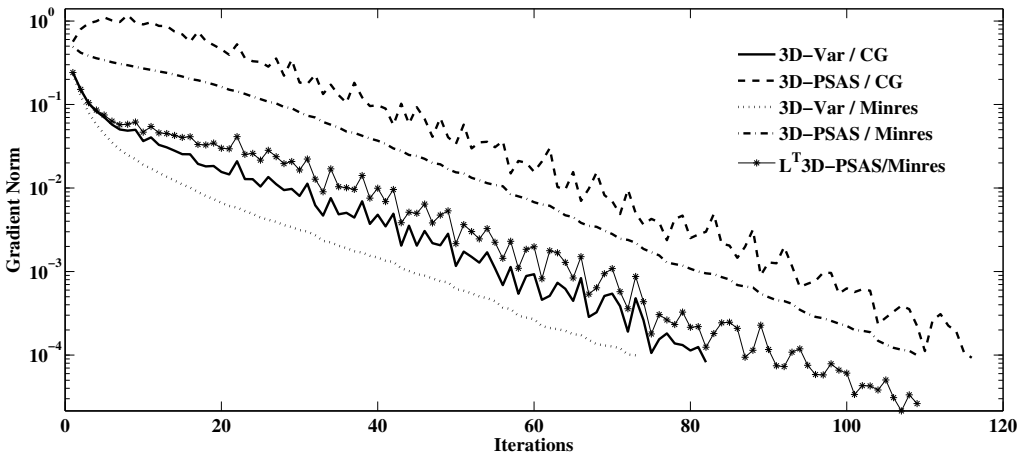


Figure 4–4: The CG and Minres residual norms of 3D-Var (solid and dotted lines), and 3D-PSAS (dashed and dash-dotted lines). The starred line represents the norm of the dual residuals in the model space.

Note that the dual residuals are

$$\mathbf{r}_k = (\mathbf{I}_m + \mathbf{L}\mathbf{L}^T)\mathbf{u}_k - \tilde{\mathbf{y}},$$

and their equivalents in model space can be written as $\tilde{\mathbf{r}}_k = (\mathbf{I}_n + \mathbf{L}^T\mathbf{L})\mathbf{L}^T\mathbf{u}_k - \mathbf{L}^T\tilde{\mathbf{y}}$, which represents the expression of the primal residuals calculated for $\mathbf{L}^T\mathbf{u}_k$. Comparing the curves of the norms of the equivalent dual residuals (starred line) and the primal residuals (solid line) in Figure 4–4, one can see that both curves are close suggesting that the image of the dual residual reduction in the model space is comparable to the reduction of the primal residuals, and that for the same stopping criterion in model space, when 3D-Var reaches convergence to the required tolerance, 3D-PSAS has technically also reached the same convergence in a comparable number of iterations. This is consistent with the statement by Courtier (1997) and confirms that considering the dual termination criterion proposed in the previous section is a better stopping criterion for 3D-PSAS.

The same experiment as in Figure 4–1 was performed to assess the impact of Minres on the equivalent of the dual minimization in the model space. Figure 4–5 shows the primal functional estimated for the dual iterates using the results of

eq. (4.7). In the CG case, this functional (dashed line) is clearly dominated by the gradient term in the RHS of (4.7) (solid line), which fully explains the spurious behaviour of the dual algorithm in the first iterations. This problem disappears when using Minres since the monotonic decrease of the dual gradient norm (dashed-dotted line) guaranties a similar decrease of the estimated primal functional (dotted-line with the circle marker). Moreover, this functional is quasi-identical to the original primal functional minimized with either Minres or the CG (solid line with the star marker), suggesting that at each iteration k, $J(\mathbf{v}_k) \approx J(\mathbf{L}^T\mathbf{u}_k)$, even during the first stage of the minimization. In Figure 4–6, the analysis increments of the assimilation are plotted for 3D-PSAS and 3D-Var after the first 10 iterations only, using Minres and the CG. As expected the Conjugate Gradient 3D-PSAS increments are degrading the analysis, whereas Minres increments are consistent with those of the 3D-Var.

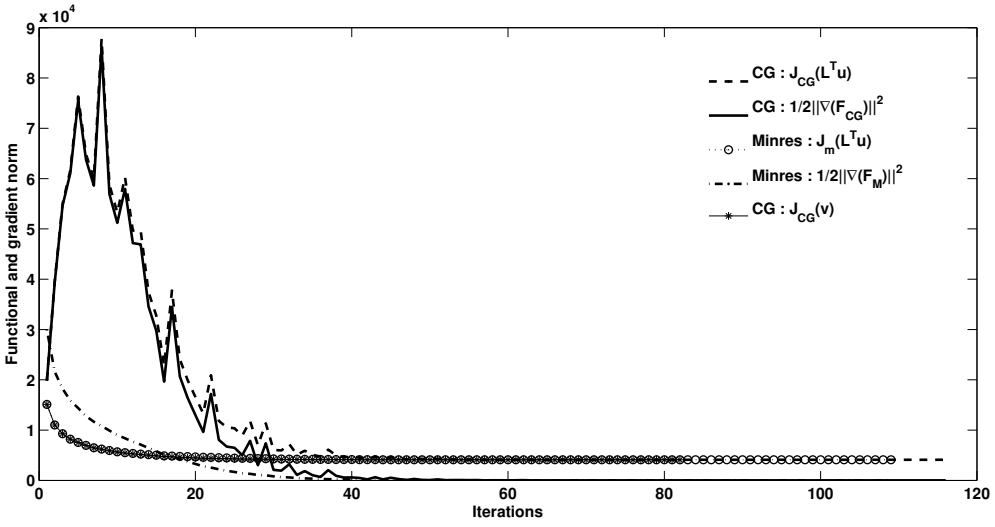


Figure 4–5: Three dimensional case : the primal functional estimated for the dual iterates using the formula in (4.7) for the CG (dashed line) and Minres (dotted-line with the circle marker). Also the term $\frac{1}{2}\|\nabla(F)\|^2$ is plotted for the CG (solid line), and Minres (dashed-dotted line), and finally, the original primal function calculated with the CG (solid line with the star marker) is plotted for comparison.

To extend the results of this work to the four dimensional case, a 4D-PSAS algorithm was implemented and similar experiments to those described earlier were

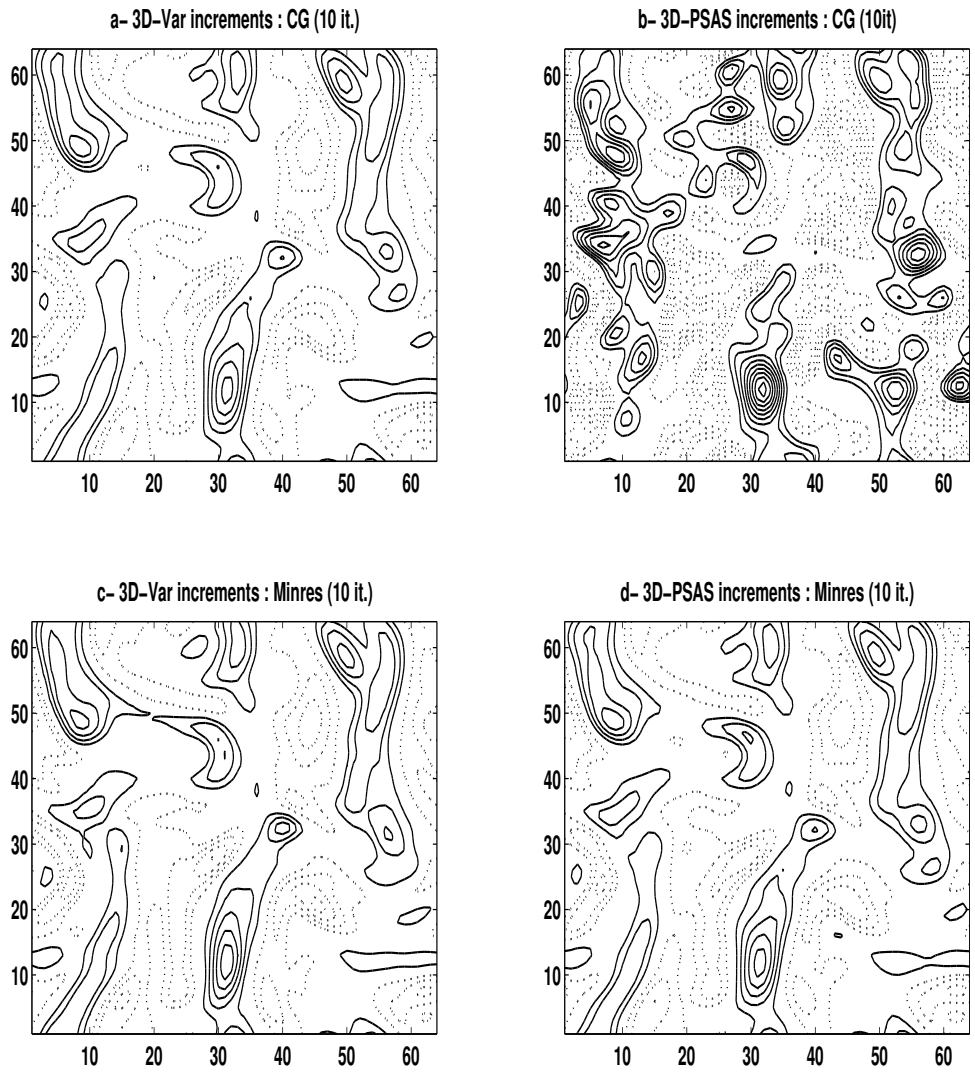


Figure 4–6: Vorticity increments after 10 iterations for a CG (top panels) and Minres (bottom panels) of the primal and dual minimizations. Equidistant contours of 0.2 units, with positive values in solid lines and negative values in dotted lines.

conducted in which the same question was examined : Would Minres be a better alternative to the CG in the dual case? The theoretical formulation of the 4D-PSAS is presented in Courtier (1997) and El Akkraoui *et al.* (2008). Note that the three dimensional formulations used for the 3D-Var/3D-PSAS can be extended to the incremental 4D-Var/4D-PSAS by including the tangent linear model (TLM) ($\mathbf{M}$) and its adjoint ($\mathbf{M}^T$) within the operators $\mathbf{L} = \mathbf{R}^{-\frac{1}{2}}\mathbf{HMB}^{\frac{1}{2}}$ and $\mathbf{L}^T = \mathbf{B}^{\frac{T}{2}}\mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-\frac{1}{2}}$.

Using the incremental formulation, these operators are linear and the same formula in (4.7) still holds in 4D. This leads to the results presented in Figure 4–7 where minimizations within the inner loop of 4D-Var and 4D-PSAS were performed with both the CG and Minres. The shape of the primal functional evaluated for the dual iterates is similar to the three dimensional case (Figure 4–5). The jumps in the dual gradient norm dominate when the CG is used, whereas the monotonic decrease of the gradient norm in Minres leads to a better behaviour in the reduction of this function.

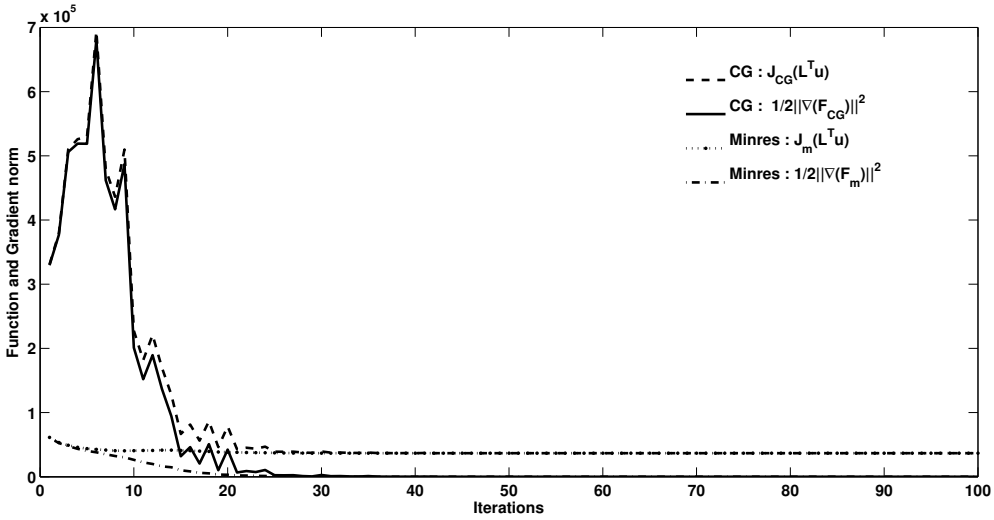


Figure 4–7: Four dimensional case : the primal functional estimated for the dual iterates using the formula in (4.7) for the CG (dashed line) and Minres (dotted-line). Also the term $\frac{1}{2}\|\nabla(F)\|^2$ is plotted for the CG (solid line), and Minres (dashed-dotted line).

It follows then that Minres is more appropriate for the dual minimization; the residual norms decrease monotonically, and the algorithm produces smoother corrections to the background state in the course of the minimization. Minres provides a nice response to the concern that needed to be resolved before considering the dual formulation for any operational implementation.

4.5 Conclusion

Convergence of the primal and dual formulations of variational data assimilation was investigated when two iterative minimization algorithms were used, the Conjugate Gradient and the Minimum Residual methods. The primary objective was to examine the dual problem discussed in El Akkraoui *et al.* (2008), where the dual minimization was shown to lead to less probable states than the background state at the beginning of the minimization process. Because only a finite number of iterations can usually be afforded in operational implementations, the spurious behaviour of the dual algorithm needed to be understood and resolved before going ahead with considering operational implementations of this method. In this paper, the use of the Minimum Residual algorithm is shown to be beneficial for the dual method, in that it guaranties a monotonic decrease of the gradient norm which appears to be the key to solving the dual problem and ensuring a smooth refinement of the solution even during the first iterations both in the three and four dimensional cases. Furthermore, it was shown that a new stopping criterion, based on the error norm in model space, should be used in the dual case to achieve the same accuracy in the analysis state with a comparable number of iterations (Courtier, 1997).

This paper is part of an ongoing work towards a weak-constraint variational assimilation, formulated in both the model space and the observation space. The dual framework is particularly interesting in this case since it decreases significantly the size of the variational problem. Many aspects of the weak-constraint formulation are still under study. Before considering the dual form of the weak-constraint 4D-Var,

it was important to address the convergence problem noted in El Akkraoui *et al.* (2008). This was the objective of the paper: to show that the dual form can be used with confidence in situations where the number of iterations is limited.

Acknowledgments

The authors would like to thank two anonymous reviewers for their comments and suggestions that improved the final version of the paper. This work was funded in part by Grant No.500-B from the Canadian Foundation for Climate and Atmospheric Sciences for the project on the *Impact of Observing Systems on Forecasting extreme weather in the short, medium and extended range: A Canadian contribution to THORPEX*. Additional support was provided by the Natural Sciences and Engineering Research Council (NSERC) of CANADA (Discovery Grant No. 357091).

5 The weak-constraint formulation of variational data assimilation: Intercomparison of the primal and dual forms

In this chapter, the weak-constraint formulation both in the primal and dual forms of variational data assimilation is examined. The extension of equivalence of these forms to the case where model errors are accounted for is presented, and some of the convergence and preconditioning properties are discussed. As in the previous chapter, the controlled 2D framework is used for the same reasons.

This Chapter is based on the following paper:

El Akkraoui A. and P. Gauthier: *The weak-constraint formulation of variational data assimilation: Intercomparison of the primal and dual forms.* In preparation. To be submitted to Q.J.R. Met. Soc.

5.1 Introduction

During the past few decades, many advances have been achieved in the field of data assimilation, which resulted in a substantial improvement of NWP products: the analyses and the forecasts. This is believed to be due to the synergy of efforts to improve the modeling and observing of the atmospheric system, whose increased complexity motivated the need to develop more advanced yet efficient data assimilation techniques (Rabier, 2005). Variational methods for instance have permitted the use of raw satellite radiance observations directly in the assimilation system rather than using retrievals of temperature and humidity profiles obtained from the data (Eyre, 1989; Thépaut, 2003). Furthermore, the proper inclusion of the time dimension in four dimensional variational data assimilation (4D-Var) permitted the treatment of the data at a near-optimal time. Consequently, the overall quality of the analysis has increased drastically, yet, new challenges arise constantly, as some sources of errors are still not accounted for.

Errors in NWP forecasts are mainly due to inaccurate initial conditions and to deficiencies in the dynamical model itself. Lorenz (1963) showed that, for a chaotic model, a small change in the initial conditions could lead to radically different solutions. Since then, much of the attention has focused on improving the initial conditions as the main source of error in weather forecasting. However, it has been argued in Orrell *et al.* (2001) that in the short term, it is the model error that dominates the forecast error, and that the effects of chaos lead to loss of predictability, but only over long time scales. This is corroborated by other studies showing the severe impact of the model deficiencies on the forecast quality (Boer, 1984; Dalcher and Kalnay, 1987; Bloom and Shubert, 1990; and Zupanski, 1993; Bennett, 2002).

Numerical models are based on a discrete version of model equations and cannot represent the atmospheric behavior exactly, for they only provide a discrete and approximate representation of the system. Hence, model errors arise due to all

the simplifications and approximations one makes to approach the complexity of the atmospheric system (i.e. resolution, approximate physics parametrization, inaccurate boundary conditions...etc). They can be random or systematic and lead to persistent bias (e.g. errors due to unresolved scales). Although little is known so far about the general form of the model error, techniques have been used to correct bias errors in the forecast using sequential and four-dimensional variational assimilation schemes (Dee and Da Silva, 1998; Derber, 1989), to estimate and take into account time-correlated stochastic errors (Daley, 1992; Zupanski, 1997).

In current NWP implementations of variational methods, the assimilation only accounts for the background and observation errors to find the *optimal* initial state for the forecast model; which implicitly ascribes the error in the forecast to the initial conditions. It is referred to as the *strong-constraint* approach (Sasaki, 1970) in which the dynamical model is assumed *perfect* and its equations describe exactly the atmospheric evolution. Accounting for model errors requires relaxing the perfect model assumption and the analysis is now required to satisfy the model equations only approximately, not exactly. This is referred to as the weak-constraint formulation, which leads to an assimilation process that corrects for the model error within the assimilation window while fitting the model trajectory to the data.

However, accounting for model errors is not a straightforward task, for these are unobserved quantities with unknown statistics. Moreover, the huge size of the assimilation problem in the weak-constraint case makes this approach very expensive for operational NWP applications, and simplifications must be made to tackle this problem (Trémolet, 2003). Recent attempts were focused on reducing the size of the model error part of the control vector, either by reducing the dimension of the subspace in which the model error is defined (Vidard *et al.* 2001), by using information provided by the analysis residual vector (Vidard *et al.* 2003), or by controlling only the systematic and time correlated part of the error (Griffith and Nichols, 2001). Trémolet (2003) considered the effect of using a reduced control variable where the

model error correction term can be assumed to be constant either on the whole assimilation window or on just a few sub-windows.

The cost of the weak-constraint formulation can be circumvented by carrying the assimilation into a space of lower dimension; the dual space (or the observation space), whose size is commensurate with the number of observations used in the assimilation process. In meteorology and oceanography, the number of degrees of freedom of the dynamical models (the size of the model space) is usually one to two orders of magnitude higher than the number of available observations.

As pointed out in Courtier (1997), the variational data assimilation can take two forms that are only different ways to solve the same problem. Those are referred to as the primal and dual forms of variational data assimilation, the primal form being the usual 3D/4D-Var as implemented at many NWP centres. The dual form (3D/4D-PSAS) looks at the problem from the observation space. Both forms solve the same assimilation problem, converge to the same solution, and with a comparable overall cost. Both approaches are based on the same basic operations and El Akkraoui et al. (2008) showed that it was indeed possible to "reconfigure" an operational 3D-Var system to obtain its dual form. Doing so, they were able to confirm the results announced in Courtier (1997): both algorithms converge to the same solution and the convergence rate is similar. However, their results also showed that the dual form exhibits a spurious behavior that augments the cost of the dual approach. El Akkraoui and Gauthier (2010) managed to explain the source of the problem and proposed a different approach to the minimization that resolved it. This could be related to the fact that, contrary to the primal problem, the dual functional does not have a direct relationship to the a posteriori probability distribution obtained from Bayesian theory.

The objective of this paper is to investigate the weak-constraint 4D-Var in both its primal and dual forms. This paper is organized as follows. In the second section,

a review of the primal form of weak-constraint variational assimilation is presented. Then , the formulation of the dual weak-constraint assimilation is examined in section 3, along with some aspects of the duality. In section 4, some intercomparison results are presented in which it is shown that the practical equivalence of the two algorithms can be extended to the weak-constraint case, and that the convergence properties of the dual method have the same characteristics as those observed for the strong constraint case. Some discussions and conclusions are presented in the last section of this paper.

5.2 Accounting for model errors in variational data assimilation

5.2.1 From strong to weak constraint formulation

Four dimensional variational data assimilation seeks to find the *optimal* model trajectory, $\mathbf{x}^a(t)$, that best fits the scattered representation of the atmospheric reality as depicted by a series of observations $\{\mathbf{y}_i, i = 1..p\}$, where the vector $\mathbf{y}_i$ represents all observations at time t_i, over the assimilation window. This is done by minimizing a functional that comprises a measure of the distance between the observation and their model equivalent on the one hand and, on the other, the distance between the *a priori* estimate and the current model state, both terms being weighted according to their respective accuracy. In the strong-constraint 4D-Var, the model equations are assumed to be a perfect representation of the true dynamical system and the time evolution $\mathbf{x}(t)$ of the dynamical system is described by the model discrete nonlinear equations

$$\mathbf{x}_i = \boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0) \tag{5.1}$$

where $\boldsymbol{\mathcal{M}}_{i,0}$ is a forward integration of the model from time t_0 to time t_i, and subscript i refers to the time t_i. This dynamical constraint, also referred to as the *perfect model assumption*, implies that the complete trajectory is entirely defined by the initial conditions $\mathbf{x}_0$. Therefore, instead of the whole trajectory, the objective function in

this case is only derived with respect to this initial state

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{i=0}^{q} \left[\boldsymbol{\mathcal{H}}_i(\boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0)) - \mathbf{y}_i\right]^T \mathbf{R}_i{}^{-1} \left[\boldsymbol{\mathcal{H}}_i(\boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0)) - \mathbf{y}_i\right] \tag{5.2}$$

where $\mathbf{B}$ and $\mathbf{R}_i$ are respectively the background and observation error covariance matrices, and $\boldsymbol{\mathcal{H}}_i$ the nonlinear observation operator that maps the model variables into the observation space.

To account for model errors, the perfect model assumption in (5.1) can be relaxed to add a correction term $\boldsymbol{\eta}_i$ to the model equations at each time step t_i

$$\mathbf{x}_i = \boldsymbol{\mathcal{M}}_{i,i-1}(\mathbf{x}_{i-1}) + \boldsymbol{\eta}_i \tag{5.3}$$

where each of the $\boldsymbol{\eta}_i$ vectors has the dimension of a three dimensional model state. The objective function in (5.2) can then be extended as

$$\begin{aligned} J(\mathbf{x}_0, \boldsymbol{\eta}) =& \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}\sum_{i=0}^{q} \left[\boldsymbol{\mathcal{H}}_i(\mathbf{x}_i) - \mathbf{y}_i\right]^T \mathbf{R}_i{}^{-1} \left[\boldsymbol{\mathcal{H}}_i(\mathbf{x}_i)) - \mathbf{y}_i\right] \\ &+ \frac{1}{2}\boldsymbol{\eta}^T \mathbf{Q}^{-1} \boldsymbol{\eta} \end{aligned} \tag{5.4}$$

where $\mathbf{Q}$ is the model error covariance matrix, and $\boldsymbol{\eta}^T = (\boldsymbol{\eta}_1^T ... \boldsymbol{\eta}_q^T)$ is the vector of the model error terms. This is the weak-constraint formulation of 4D-Var. The control vector here is extended to include all the $\boldsymbol{\eta}_i$ terms, so that the initial states, together with the model errors form the control parameters that must be determined by the assimilation. That is, over the assimilation window, it is the *full trajectory* $\mathbf{x}(t)$ that is adjusted to fit the observations, and not only the initial conditions as in the strong-constraint formulation.

To make the objective function quadratic, the incremental approach can be applied as in the strong-constraint case by using a Tangent Linear Model (TLM), $\mathbf{M}$, and its adjoint, $\mathbf{M}^T$, both defined in the vicinity of the current background state. Thus, only linear operators are involved, and the increments in this case evolve as

$\boldsymbol{\delta}\mathbf{x}_i = \mathbf{M}_{i,i-1}\boldsymbol{\delta}\mathbf{x}_{i-1} + \boldsymbol{\eta}_i$, where $\boldsymbol{\delta}\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_b(0)$, and $\mathbf{M}_{i,i-1}$ represents the integration of the TLM from time t_{i-1} to time t_i. The objective function becomes

$$J(\boldsymbol{\delta}\mathbf{x}_0, \boldsymbol{\eta}) = \frac{1}{2}\boldsymbol{\delta}\mathbf{x}_0{}^T\mathbf{B}^{-1}\boldsymbol{\delta}\mathbf{x}_0 + \frac{1}{2}\sum_{i=0}^{q}(\mathbf{H}_i\boldsymbol{\delta}\mathbf{x}_i - \mathbf{y}_i')^T\mathbf{R}_i^{-1}(\mathbf{H}_i\boldsymbol{\delta}\mathbf{x}_i - \mathbf{y}_i') + \frac{1}{2}\boldsymbol{\eta}^T\mathbf{Q}^{-1}\boldsymbol{\eta} \quad (5.5)$$

where $\mathbf{y}_i' = \mathbf{y}_i - \boldsymbol{\mathcal{H}}_i(\boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_b))$ are the innovation vectors. Iterative algorithms are then used to minimize the functional and obtain the analysis state, $(\boldsymbol{\delta}\mathbf{x}_0^a, \boldsymbol{\eta}^a)$. In the linear case, the analysis increment at each time step can then be written as

$$\boldsymbol{\delta}\mathbf{x}_i^a = \mathbf{M}_{i,0}\boldsymbol{\delta}\mathbf{x}_0^a + \sum_{j=1}^{i}\mathbf{M}_{i,j}\boldsymbol{\eta}_j^a \quad (5.6)$$

where the rightmost term represents the part of the forecast error that is only due to the model deficiencies accumulated from the initial time t_0 to time t_i. In matrix form, (5.6) can simply be written as

$$\boldsymbol{\delta}\mathbf{x}^a = \boldsymbol{\mathcal{N}}\boldsymbol{\delta}\mathbf{z}^a \quad (5.7)$$

or equivalently for each time step as $\boldsymbol{\delta}\mathbf{x}_i^a = \boldsymbol{\mathcal{N}}_i\boldsymbol{\delta}\mathbf{z}^a$, where

$$\boldsymbol{\mathcal{N}} = \begin{pmatrix} \boldsymbol{\mathcal{N}}_0 \\ \boldsymbol{\mathcal{N}}_1 \\ \boldsymbol{\mathcal{N}}_2 \\ \vdots \\ \boldsymbol{\mathcal{N}}_q \end{pmatrix} = \begin{pmatrix} \mathbf{I}_n & 0 & 0 & \cdots & 0 \\ \mathbf{M}_{1,0} & \mathbf{I}_n & 0 & \cdots & 0 \\ \mathbf{M}_{2,0} & \mathbf{M}_{2,1} & \mathbf{I}_n & & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \mathbf{M}_{q,0} & \mathbf{M}_{q,1} & \mathbf{M}_{q,2} & \cdots & \mathbf{I}_n \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\delta}\mathbf{z} = \begin{pmatrix} \boldsymbol{\delta}\mathbf{x}_0 \\ \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_q \end{pmatrix}.$$

where $\mathbf{I}_n$ is the identity matrix, and n the size of the three dimensional model state. The operator $\boldsymbol{\mathcal{N}}$ reconstructs the model trajectory from an initial state and an estimate of the model error terms at each time step. Following Courtier (1997), and using (5.6), the functional (5.5) and its gradient can be written in a compact form respectively as

$$J(\boldsymbol{\delta}\mathbf{z}) = \frac{1}{2}\delta\mathbf{z}^T\mathbf{D}^{-1}\boldsymbol{\delta}\mathbf{z} + \frac{1}{2}(\mathbf{S}\boldsymbol{\delta}\mathbf{z} - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{S}\boldsymbol{\delta}\mathbf{z} - \mathbf{y}') \quad (5.8)$$

$$\boldsymbol{\nabla}_{\boldsymbol{\delta}\mathbf{z}}J = \mathbf{D}^{-1}\boldsymbol{\delta}\mathbf{z} + \mathbf{S}^T\mathbf{R}^{-1}(\mathbf{S}\boldsymbol{\delta}\mathbf{z} - \mathbf{y}') \quad (5.9)$$

where $\mathbf{D} = \begin{pmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{Q} \end{pmatrix}$ is the matrix of the background and model error covariances and $\mathbf{S}$ is an operator designed such that $\mathbf{S}_i = \mathbf{H}_i \boldsymbol{\mathcal{N}}_i$, which can be seen as a generalization of the observation operator since $\mathbf{S}_i \boldsymbol{\delta z}^a = \mathbf{H}_i \boldsymbol{\delta x}_i^a$ is the analysis increment evaluated in the observation space.

At the minimum of the objective functional, the gradient vanishes, and the analysis result is

$$\boldsymbol{\delta z}^a = [\mathbf{D}^{-1} + \mathbf{S}^T \mathbf{R}^{-1} \mathbf{S}]^{-1} \mathbf{S}^T \mathbf{R}^{-1} \mathbf{y}' \tag{5.10}$$

The gain matrix here is $\mathbf{K} = [\mathbf{D}^{-1} + \mathbf{S}^T \mathbf{R}^{-1} \mathbf{S}]^{-1} \mathbf{S}^T \mathbf{R}^{-1}$, which can also be written as $\mathbf{K} = \mathbf{D}\mathbf{S}^T [\mathbf{R} + \mathbf{S}\mathbf{D}\mathbf{S}^T]^{-1}$. This compact form helps see that variational methods are just an extension of one another to a more general case. In the strong-constraint 4D-Var, no model error term is included so that the matrix $\mathbf{D}$ is simply the background error covariance matrix ($\mathbf{B}$) and $\mathbf{S}$ reduces to its first column, which performs forward integrations of the model from initial time to the observation times ($\mathbf{M}_{i,0}$), followed by the application of the observation operator $\mathbf{H}_i$. In 3D-Var, the dynamical model is not involved in the assimilation process, thus $\mathbf{S}$ is simply the observation operator $\mathbf{H}$.

5.2.2 The analysis error

The analysis error at each time step is defined as the departure of the analysis state, $\mathbf{x}_i^a$, to the truth, $\mathbf{x}_i^t$, or equivalently as

$$\boldsymbol{\epsilon}_i^a = \boldsymbol{\delta x}_i^a + \mathbf{x}_i^b - \mathbf{x}_i^t \tag{5.11}$$

Since the truth is not known, one usually tries to write this error in terms of known quantities in the assimilation: the background and observation error statistics. In the strong-constraint formulation, it is common to assume that the evolution of the true state is entirely described by the dynamical model equations, i.e. $\mathbf{x}_i^t = \boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0^t)$. However, this assumption is not valid if the model is not considered perfect. Hence,

in the weak-constraint formulation and in the linear case, one uses instead

$$\mathbf{x}_i^t \approx \boldsymbol{\mathcal{M}}_{i,0}(\mathbf{x}_0^t) + \sum_{j=1}^{i} \mathbf{M}_{i,j}\boldsymbol{\eta}_j = \boldsymbol{\mathcal{N}}_i \begin{pmatrix} \mathbf{x}_0^t \\ \boldsymbol{\eta} \end{pmatrix}$$

so that one can write $\mathbf{x}_i^b - \mathbf{x}_i^t = \boldsymbol{\mathcal{N}}_i \begin{pmatrix} \mathbf{x}_0^b \\ 0 \end{pmatrix} - \boldsymbol{\mathcal{N}}_i \begin{pmatrix} \mathbf{x}_0^t \\ \boldsymbol{\eta} \end{pmatrix} = \boldsymbol{\mathcal{N}}_i \begin{pmatrix} \boldsymbol{\epsilon}^b \\ -\boldsymbol{\eta} \end{pmatrix}$, where $\boldsymbol{\epsilon}^b$ is the background error. On the other hand, the analysis increment is $\boldsymbol{\delta}\mathbf{x}_i^a = \boldsymbol{\mathcal{N}}_i \mathbf{K}\mathbf{y}'$ and the innovations can be written as $\mathbf{y}' = \mathbf{y} - \boldsymbol{\mathcal{H}}(\mathbf{x}_b(t)) = \boldsymbol{\epsilon}^{obs} - \mathbf{S} \begin{pmatrix} \boldsymbol{\epsilon}^b \\ -\boldsymbol{\eta} \end{pmatrix}$.

It follows that the analysis error can be expressed in this case as

$$\boldsymbol{\epsilon}_i^a = \boldsymbol{\mathcal{N}}_i \left[\mathbf{K}\boldsymbol{\epsilon}^{obs} + (\mathbf{I} - \mathbf{K}\mathbf{S}) \begin{pmatrix} \boldsymbol{\epsilon}^b \\ -\boldsymbol{\eta} \end{pmatrix} \right] \tag{5.12}$$

Using both definitions of the gain matrix, and replacing terms, it is then possible to show that the analysis error covariance matrix at each time step is expressed as

$$\mathbf{P}_i^a = \boldsymbol{\mathcal{N}}_i [\mathbf{D}^{-1} + \mathbf{S}^T \mathbf{R}^{-1} \mathbf{S}]^{-1} \boldsymbol{\mathcal{N}}_i^T. \tag{5.13}$$

As mentioned earlier, the operators in the weak-constraint formulation are only extensions of those of the strong-constraint. Since $\boldsymbol{\mathcal{N}}_0$ reduces to the identity in the latter case, one can easily verify that the strong-constraint analysis error can be calculated directly from (5.13).

5.3 Dual formulation of variational data assimilation

For very large stochastic systems, such as the atmosphere or the ocean, an operational implementation of the weak-constraint formulation of 4D-Var is very expensive due to the enormous size of the assimilation problem that is augmented by the cost of estimating all the model errors. When accounting for model errors, the size of a weak-constraint 4D-Var is q-times the size of a strong-constraint 4D-Var. This is also expected to increase commensurately with the steady improvements in the numerical models and the use of higher resolutions. By carrying the assimilation

process into a smaller space (the observation space), it is possible to keep the weak-constraint formulation to a manageable size. In 4D-PSAS, the dual form of 4D-Var, the control vector is defined in observation space (Cohn *et al.*, 1998), and its size does not change, whether the model errors are accounted for or not.

The theoretical equivalence *at convergence* of the primal and dual forms is discussed in Courtier (1997). They both solve the same variational problem but in two different spaces; the model and observation space respectively. The equivalence is also subject to the assumption on the linearity of the operators (Courtier, 1997). However, this has recently been extended to the non linear case (Auroux, 2007; El Akkraoui *et al.*, 2008) and to the singular vectors of the Hessians.

5.3.1 Dual weak-constraint formulation

The step by step theoretical formulation of 3D and 4D-PSAS is presented in Courtier (1997), El Akkraoui *et al.* (2008), and El Akkraoui and Gauthier (2010). In the same way, the weak-constraint 4D-PSAS is derived directly from the primal form (weak-constraint 4D-Var) by using the two forms of the Gain matrix in the primal analysis state as

$$\boldsymbol{\delta}\mathbf{z}^a = \mathbf{D}\mathbf{S}^T \left[\mathbf{R} + \mathbf{S}\mathbf{D}\mathbf{S}^T\right]^{-1} \mathbf{y}'. \tag{5.14}$$

and solving, in observation space, the linear system $\left[\mathbf{R} + \mathbf{S}\mathbf{D}\mathbf{S}^T\right] \mathbf{w} = \mathbf{y}'$, the solution of which satisfies $\boldsymbol{\delta}\mathbf{z}^a = \mathbf{D}\mathbf{S}^T\mathbf{w}^a$. The control vector here, $\mathbf{w}$, has the size of the number of observations, which is the same both in the strong- and weak-constraint formulations.. It follows that the dual objective function and its gradient are respectively

$$F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T[\mathbf{R} + \mathbf{S}\mathbf{D}\mathbf{S}^T]\mathbf{w} - \mathbf{w}^T\mathbf{y}' \tag{5.15}$$

and

$$\boldsymbol{\nabla}_{\mathbf{w}} F = [\mathbf{R} + \mathbf{S}\mathbf{D}\mathbf{S}^T]\mathbf{w} - \mathbf{y}'. \tag{5.16}$$

Note that, in the dual case, the error covariance matrices are present in their direct form (not inverse) which implies that PSAS remains regular in the limit of vanishing model error. Once the minimization of the functional F has been completed, and the analysis result in the dual space ($\mathbf{w}^a$) found, it is straightforward to reconstruct the primal analysis increments using $\boldsymbol{\delta}\mathbf{x}_i^a = \boldsymbol{\mathcal{N}}_i\boldsymbol{\delta}\mathbf{z}^a = \boldsymbol{\mathcal{N}}_i\mathbf{D}\mathbf{S}^T\mathbf{w}^a$.

5.3.2 Duality and preconditioning

In El Akkraoui *et al.* (2008), it was shown that the equivalence of the primal and dual methods can be extended to nonlinear cases of four dimensional applications, and to their Hessian eigenvectors. The weak-constraint formulation is no exception. Using the compact forms, the primal and dual preconditioning is introduced by the change of variable $\boldsymbol{\xi} = \mathbf{D}^{-\frac{1}{2}}\boldsymbol{\delta}\mathbf{z}$ and $\mathbf{u} = \mathbf{R}^{\frac{1}{2}}\mathbf{w}$ respectively. The Hessians can then be written as

$$\mathbf{J}" = \mathbf{I}_{n\times q} + \mathbf{D}^{\frac{T}{2}}\mathbf{S}^T\mathbf{R}^{-1}\mathbf{S}\mathbf{D}^{\frac{1}{2}} \tag{5.17}$$

$$\mathbf{F}" = \mathbf{I}_m + \mathbf{R}^{-\frac{1}{2}}\mathbf{S}\mathbf{D}\mathbf{S}^T\mathbf{R}^{-\frac{1}{2}} \tag{5.18}$$

where n, q, and m refer respectively to the size of a 3D model state, the number of time steps in the assimilation period, and the number of assimilated observations. In the same way as proposed in El Akkraoui *et al.* (2008), the operator $\boldsymbol{\mathcal{L}}$ and its adjoint $\boldsymbol{\mathcal{L}}^T$ are defined such that $\boldsymbol{\mathcal{L}} = \mathbf{R}^{-\frac{1}{2}}\mathbf{S}\mathbf{D}^{\frac{1}{2}}$ and $\boldsymbol{\mathcal{L}}^T = \mathbf{D}^{\frac{T}{2}}\mathbf{S}^T\mathbf{R}^{-\frac{1}{2}}$. The Hessians can then be written as:

$$\begin{aligned} \mathbf{J}" &= \mathbf{I}_{n\times q} + \boldsymbol{\mathcal{L}}^T\boldsymbol{\mathcal{L}} \\ \mathbf{F}" &= \mathbf{I}_m + \boldsymbol{\mathcal{L}}\boldsymbol{\mathcal{L}}^T \end{aligned} \tag{5.19}$$

One can then easily see the duality between the two formulations. In Courtier (1997), it was shown that the Hessians of both formulations have the same spectrum, possibly completed by some 1's. This statement is still valid in the weak-constraint context. Furthermore, as in El Akkraoui *et al.* (2008), the expressions in (5.19) can be used to show that the singular vector of the Hessians are also related. One should bear in mind though that this equivalence is only valid under the assumption of linearity

of the observation operator and subject to the validity of the TLM and its adjoint. Depending on the modeling of the model error covariance matrix, the square root of $\mathbf{D}$ may not be easy to construct. Therefore, the dual Hessian is expressed in a more natural framework, for which the eigenvectors are easier to calculate. Although this preconditioning is more effective (Ehrendorfer and Tribbia, 1995), the calculations of the singular vectors in the weak-constraint case can be very expensive. One can use a combined Lanczos-Conjugate-Gradient method that allows for the calculation of Hessian singular vectors in the course of the minimization process since there is a close connection between the gradient vectors of the CG and those generated by the Lanczos algorithm (Paige and Saunders, 1975; Fisher and Courtier, 1995). Thus the number of iterations is comparable to the number of singular vectors, which makes this approach very useful in the context of operational processes.

5.4 Results

5.4.1 Experimental setting

The experiments are set up on a periodic β plane. The model used here is the same two dimensional barotropic non-divergent model used by Tanguay *et al* (1995), and Laroche and Gauthier (1998), to solve the barotropic vorticity equation on the β-plane :

$$\frac{\partial \boldsymbol{\zeta}}{\partial t} + \mathbf{J}(\boldsymbol{\Psi}, \boldsymbol{\zeta}) + \beta \mathbf{v} = \mathbf{f} - \mathbf{D}(\boldsymbol{\zeta}),$$

where $\boldsymbol{\Psi} = -U_0 \mathbf{y} + \boldsymbol{\psi}$ is the full stream function, with U_0 a large-scale East-West flow representing the mean zonal wind, and $\boldsymbol{\psi}$ represents the stream function. The vorticity is the Laplacian of the stream function, $\boldsymbol{\zeta} = \boldsymbol{\nabla}^2 \boldsymbol{\psi}$, and the horizontal wind components are $\mathbf{u} = -\frac{\partial \boldsymbol{\psi}}{\partial \mathbf{y}}$, and $\mathbf{v} = \frac{\partial \boldsymbol{\psi}}{\partial \mathbf{x}}$, $\mathbf{J}$ being the Jacobian, $\mathbf{f}$ a forcing term, and $\mathbf{D}$, a linear dissipation operator. A detailed description of this model can be found in Tanguay *et al.* (1995), and Laroche and Gauthier (1998).

The assimilation is performed with perfect observations of wind components at each model grid point and at every six time steps of the model. The observation

error is assumed to be uncorrelated, thus the observation error covariance matrix $\mathbf{R}$ is diagonal. The model state is the barotropic vorticity in Fourier space so that the observation operator converts vorticity into wind components, then performs an inverse Fourier transform. More details about this experimental setting is presented in El Akkraoui and Gauthier (2010).

The model error was specified in the β term of the barotropic vorticity equation. Twin experiments were conducted in which the "truth" and the observations are produced with a version of the model where $\beta = 0.4$, and the experiments are conducted with the "erroneous" model where $\beta = 0.5$. The error in the β term is supposed to mimic the effect of a random model error for which the assimilation seeks a correction estimate. Weak-constraint 4D-Var and 4D-PSAS assimilation experiments were conducted with minimizations performed with both the Minimum Residual (Minres) and the Conjugate Gradient (CG) methods for comparison (Hestenes and Stiefel, 1953; Paige and Saunders, 1975).

5.4.2 The model error covariance matrix

In this paper, the experiments are performed with a fixed model error covariance matrix related to the background error covariance term such that $\mathbf{Q}_i = \alpha\mathbf{B}$, for each time step i. For obvious reasons, this is a practical but very poor choice of the covariances since the solution that should lie in the subspaces spanned respectively by $\mathbf{B}$ and $\mathbf{Q}$, is here only allowed to span the same subspace as the background error term (Trémolet, 2007). That is, model error is restricted in the same directions as the initial condition increment, and only their relative amplitudes differ depending on the value of α (fixed here to $\alpha = 0.01$). We acknowledge the importance of modeling and estimating the model error covariances. Nonetheless, since the objective of this paper regards the algorithmic aspects of the dual form, this particular form of $\mathbf{Q}$ is sufficient for our purpose.

5.4.3 Implementation

Based on the assimilation system described in El Akkraoui and Gauthier (2010), the 4D-Var and 4D-PSAS were extended for the implementation of a weak-constraint version of each algorithm. It is important to stress here that the modularity of the operators being used is crucial to a smooth implementation of these methods. As discussed earlier, the primal and dual forms use the same operators and matrices, but in a different sequence. Once the basic operators extended from the strong to the weak-constraint, it is then possible to easily build both the primal and dual algorithms in the same setting. The primal method required the extension of the control vector to include the model errors at each time step, and thus the size of the working space was considerably increased. However, in the dual case, the working space is unchanged, and all the variables and matrices have either the size of the number of observations or the size of a three dimensional model state. Furthermore, a few changes were required to adapt the TLM and its adjoint to the weak-constraint case to construct the operators $\boldsymbol{\mathcal{N}}$ and $\boldsymbol{\mathcal{N}}^T$ embedded in the generalized observation operator and its adjoint, $\mathbf{S}$ and $\mathbf{S}^T$ respectively. $\mathbf{S}$ is implemented as a matrix-vector product, such that $\mathbf{SX} = \mathbf{X}'$, where $\mathbf{X}^T = (\mathbf{x}_0^T, \mathbf{x}_1^T, \cdots \mathbf{x}_q^T)$. That is $\mathbf{X}'_i = \mathbf{H}_i \boldsymbol{\mathcal{N}}_i \mathbf{X} = \mathbf{H}_{i-1} \boldsymbol{\mathcal{N}}_{i-1} \mathbf{x}_0 + \mathbf{H}_i \mathbf{x}_i$, and $\mathbf{S}_0 \mathbf{X} = \mathbf{H}_0 \mathbf{x}_0$.

On the other hand, $\mathbf{S}^T$ uses the integrations of the adjoint model that are now defined with respect to the new primal and dual adjoint variables, expressed respectively as

$$\begin{aligned} \delta\mathbf{x}_i^* &= \mathbf{M}_{i+1,i}^T \delta\mathbf{x}_{i+1}^* + \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{y}'_i \quad \text{with} \quad \delta\mathbf{x}_q^* = \mathbf{H}_q^T \mathbf{R}_q^{-1} \mathbf{y}'_q \\ \mathbf{w}_i^* &= \mathbf{M}_{i+1,i}^T \mathbf{w}_{i+1}^* + \mathbf{H}_i^T \mathbf{w}_i, \quad \text{with} \quad \mathbf{w}_q^* = \mathbf{H}_q^T \mathbf{w}_q \end{aligned}$$

The minimization of the objective functionals are then performed iteratively in exactly the same way as in the strong-constraint case.

5.4.4 Convergence properties

In the context on a weak-constraint formulation, experiments are conducted to assess first the equivalence of the primal and dual algorithms. In figure 5–1,

the analysis increments are plotted for the weak-constraint 4D-Var (top panel) and for the weak-constraint 4D-PSAS (bottom panel), and shows that both algorithms have converged to the same analysis state. The minimization of the functionals is

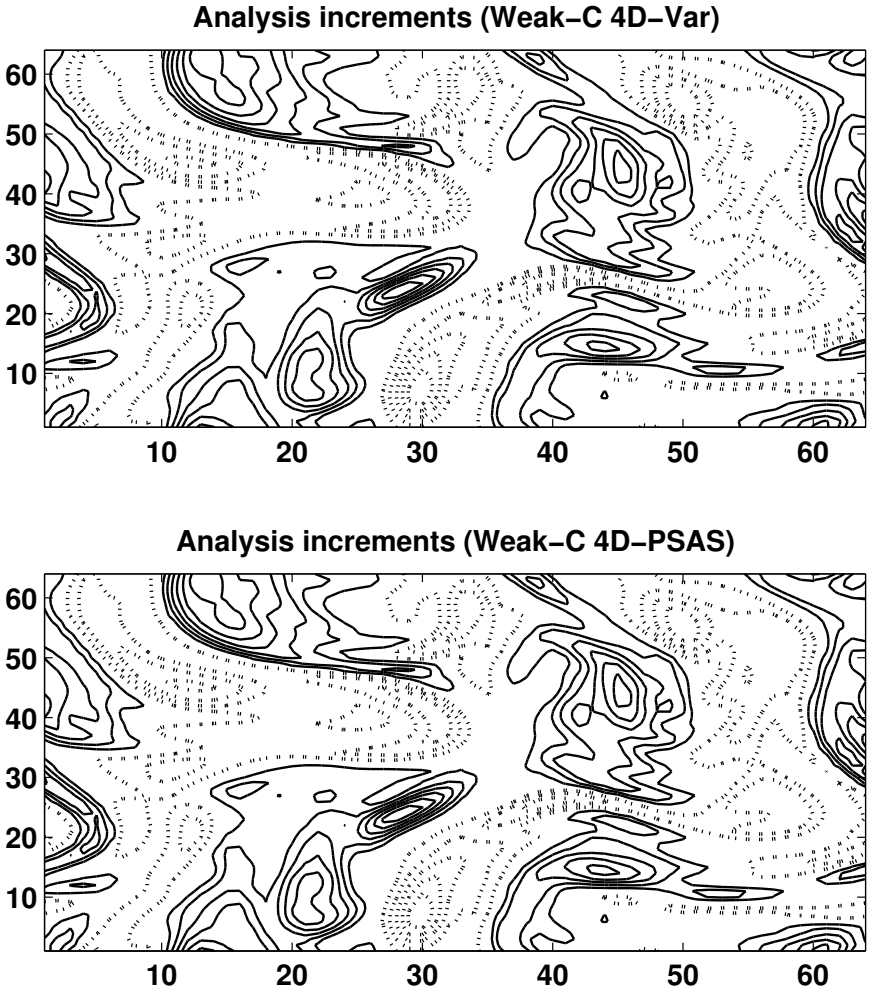


Figure 5–1: Analysis increments of the weak-constraint 4D-Var (top) and weak-constraint 4D-PSAS (bottom).

performed with Conjugate Gradient (CG) algorithms, and stopped when the ratio of the gradient norm to its initial value falls below a threshold of 10^{-5}. The reduction of the gradient norms is presented in figure 5–2. The primal and the dual methods

have the same convergence rate as shown by the slope of their gradient norms (solid and dashed lines respectively).

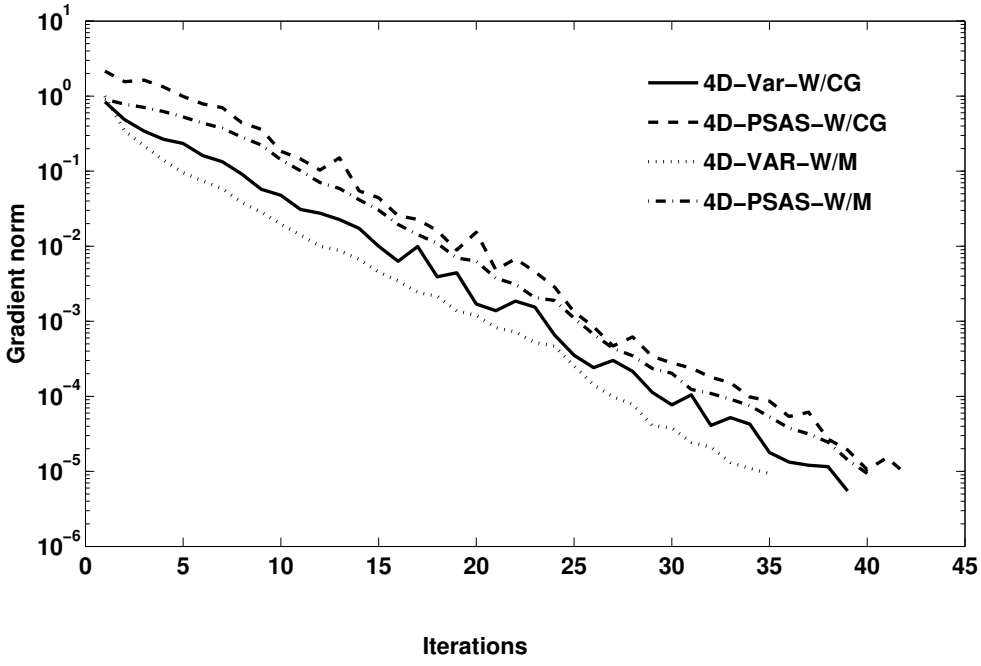


Figure 5–2: Reduction of the ratio of the gradient norm to its initial value for the weak-constraint 4D-Var, with the CG and Minres, and for the weak-constraint 4D-PSAS with the CG and Minres.

The equivalence of results *at convergence* and in the linear case was announced theoretically in Courtier (1997), and then confirmed in practice in the context of an operational three and four dimensional variational assimilation in El Akkraoui *et al.*, 2008. However, the same experiments highlighted the importance of bearing in mind the assumption used to formulate the dual algorithm. That is, the equivalence is valid *at convergence only*. During the primal minimization, the iterates are, by construction, gradually refining the analysis state; thus increasing the probability of having an analysis that is more accurate than the initial state. This property is intrinsic to the definition of the primal functional itself, $J = -ln(p)$, where p is the conditional probability of a model state being true knowing the observational

state. However, in the dual case, the functional has no direct physical meaning, which results in iterates for which their image in the primal space may degrade the analysis during the first iterations, before it starts to adjust and recover the same analysis as the primal algorithm. That is, the dual analysis is only reliable when the minimization is close enough to convergence. Therefore, it is expected that, in a weak-constraint framework, the first stage of the dual minimization can be challenging; the contributions of the model error terms, estimated iteratively by the minimization, may aggravate this behavior since they are directly included in the integrations of the model and its adjoint. This behavior is illustrated, in our context, in figure 5–3, which shows the curves of the minimization of the dual functional with the CG (dashed line), together with the primal functional evaluated for the dual iterates (i.e. the image of the dual minimization in the model space) using the CG (dotted line). As expected, during the first 5 iterations, the dual method produces iterates for which the image in the model space is unphysical, and are worse that the initial state that the minimization started with.

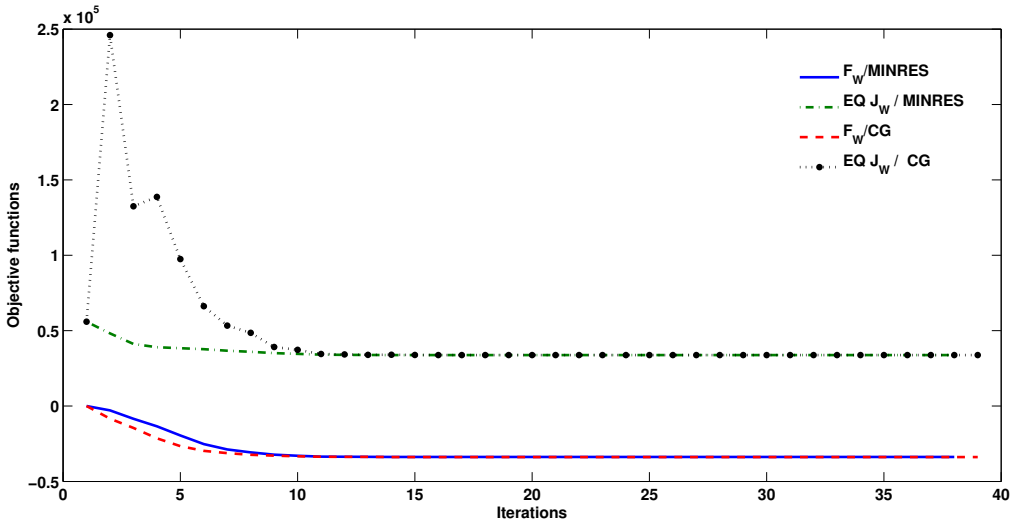


Figure 5–3: Minimization of the weak-constraint 4D-PSAS with the CG (dashed line) and Minres(solid line), and the primal functional evaluated for the dual iterates with the CG (dotted line) and Minres (dash-dotted line).

In El Akkraoui and Gauthier (2010), the choice of the minimization algorithm proved to be crucial in the dual case. It was shown that another algorithm can be used to avoid the dual behavior described above. By construction, the MINimum RESidual method (MINRES) minimizes the gradient norm (i.e. the residual of the linear system defining the gradient), instead of the functional itself as it is the case for the Conjugate Gradient method. Encouraging results were shown in El Akkraoui and Gauthier (2010) in the case of 3D and 4D-PSAS. In the present context, figure 5–3 also shows the curves of the minimization with Minres of the dual functional (solid line) and the primal functional evaluated for the dual iterates (dash-dotted line). When using Minres, the dual iterates improve gradually the analysis state even during the first iterations. The reason why the dual method finds its way from completely unphysical states to a converged state, which is the same as the primal method, can be explained by noticing that the formula found in El Akkraoui and Gauthier (2010) still holds in the weak-constraint case. At each iteration k, one can write

$$J(\boldsymbol{\mathcal{L}}^T\mathbf{u}_k) = \frac{1}{2}\|\boldsymbol{\nabla} F(\mathbf{u}_k)\|^2 - F(\mathbf{u}_k) \tag{5.20}$$

So that near convergence, the gradient norm is close to zero and $J(\boldsymbol{\mathcal{L}}^T\mathbf{u}_K) \approx -F(\mathbf{u}_K)$, which is consistent with the theoretical formulation *at convergence*: $J(\mathbf{v}^a) = -F(\mathbf{u}^a)$ (See Appendix in El Akkraoui *et al.*, 2008). It follows that, like in the 3D and 4D-PSAS case, the Minres method does solve the problem of the dual minimization in a weak-constraint framework.

5.5 Conclusion

In this paper, the equivalence of the primal and dual forms of variational data assimilation was examined in the case where model errors are accounted for within the assimilation process. Unlike the primal form, the upgrade of a strong-constraint 4D-PSAS to a weak-constraint 4D-PSAS can easily be done, and the size of the assimilation problem is still manageable. Results showed that while both methods

converge to the same solution, with the same convergence rate, and with comparable number of iterations, the dual method is still sensitive to minimization process. As in the 3D- and 4D-PSAS case, using the minimum residual (MINRES) method instead of the Conjugare Gradient provides a solution to the dual problem.

The interest for the weak-constraint formulation is motivated by the need for extending the length of the assimilation window, which is believed to be limited by the uncertainties in the model (Trémolet, 2007). The weak-constraint formulation of variational assimilation is expected to allow for extending the assimilation window which would result in a flow dependent background error covariances and would allow for the use of all relevant observations to optimally estimate the atmospheric state (Fisher *et al.*, 2005).

Accounting for model errors requires the proper estimation of the model error covariance matrix. In this paper, the choice of a fixed matrix proportional to the background error covariance matrix is justified by our focus only on the algorithmic aspects of the dual method. Other more realistic methods for estimating the $\mathbf{Q}$ matrix have been proposed, by using, for instance, statistics on the model error tendencies (Trémolet, 2007).

6 Conclusions

Courtier (1997) pointed out that 3D-Var, 4D-Var and the weak-constraint 4D-Var have similar forms, from which dual equivalent variants can be derived. This holds in the purely quadratic case but can easily be extended to the nonlinear case by using an incremental approach. In this thesis, the dual forms were shown to be equivalent to their primal forms as they converge to the same solution with similar convergence rates. However, the experiments also showed a spurious behaviour in the dual problems at the beginning of the minimization. Further investigation highlighted the importance of using a minimization scheme that monotonically decreases the norm of the gradient, instead of the dual functional itself. One important result of this thesis is a relationship between the primal $J(\mathbf{x})$ and the dual $F(\mathbf{w})$ objective functions, which is

$$J(\mathbf{x}) = \frac{1}{2}\|\boldsymbol{\nabla} F\|^2 - F(\mathbf{w})$$

with $\mathbf{x} = \mathbf{B}\mathbf{H}^T\mathbf{w}$. Conjugate Gradient algorithms do not ensure a reduction in the gradient norm $\|\boldsymbol{\nabla} F\|$ which then can lead to an increase of the functional J. The similarity of all variational problems ensures that this is the case of the dual forms of 3D-Var, 4D-Var and that of the weak-constraint.

The control variable of the dual problems corresponds to that of the observation space. Therefore, it changes in dimension and in nature, from one assimilation to the next, according to the observations to be assimilated. Experience with operational systems has shown that the approximate Hessian built during an analysis process

can be used to precondition the next assimilation period. In 3D/4D-Var, this can be implemented in a straightforward way because the control variable is the model state and its form remains unchanged.
To circumvent this difficulty, it was shown in this thesis that the Hessian of the primal and dual forms are related and in the dual case, it is possible to remap the Hessian from an assimilation with the new variable associated with the observation set of the next assimilation. Again, this holds for all the variational problems 3D/4D-var and the weak-constraint.

Working in different settings and with different algorithms stressed the importance of the modularity in the assimilation algorithm. Since all variational methods are only variants to solve the same problem, they can be built from the same basic operators and covariance matrices. Modularity guaranties the Independence of these building blocks from each other and from the algorithm used, and makes it easy to implement a 4D-PSAS for instance using an existing 4D-Var setting and vice versa. This is the key element that made it possible to adapt the same initial setting (3D-Var) to build the variety of methods used for the purpose of this work.

The results presented in this thesis therefore show that the dual from can be used with confidence in all cases. The objective though was to examine if the dual form of the weak-constraint 4D-Var is possible. As mentioned earlier, this thesis is part of the ongoing effort in the data assimilation community towards an affordable operational application of the weak-constraint variational assimilation. Many aspects of this formulation are still an active field of research. In the last part of this work, the dual framework is proposed as an alternative to the primal form since it benefits form the significantly lowed dimension of the observation space and permits the integration of the model error term in the assimilation system while keeping the problem to a manageable size.

Being able to correct for the part of the forecast error that is due to the model uncertainties is one of the reasons why the weak-constraint formulation is currently receiving considerable attention. As discussed earlier, a more important reason is the efforts to extend the length of the assimilation window, which would allow for more past observations to contribute to a better adjustment of the model analysis trajectory. With longer assimilation windows, the contribution of the model error in the forecast error growth may dominate and the initial conditions may no longer control the fit to the observations near the end of the window.

Accounting for model error requires the proper estimation of its statistics as described in the model error covariance matrix, which is still a major difficulty. In this thesis, the choice of a simple modeling of the statistics is only motivated by our focus on the algorithmic aspects of the dual method. The model error covariance matrix needs to be carefully estimated when assessing the improvements added by the weak-constraint formulation. Trémolet (2007) used statistics on the model error tendencies in a similar way as the lagged forecasts method used for the estimation of the background error covariance matrix. This was further examined in a recent study (Lindskog *et al.*, 2008).

Throughout this thesis, the dual form of variational data assimilation was shown to be adaptable to an operational implementation, in a strong or a weak-constraint framework. The equivalence of the primal and dual results gives the opportunity to choose the appropriate algorithm for each case scenario. As stated in Lewis *et al.* (2006), when the number of observations is lower than the number of degrees of freedom of the model, the assimilation problem is underdetermined and the dual space form is well-posed, while the primal form is ill-posed, and vice versa. In current NWP applications, the observation space is still smaller, and the ever continuing increase in the volume of available data is compensated by the use of higher model resolutions, and more and more sophisticated physics. This is particularly interesting in a

weak-constraint framework. More future work on the dual method; its convergence properties still need further investigations, and other minimization algorithms can also be tested. The work on weak-constraint formulation on the other hand, must focus primarily on the modeling and estimation of the model error covariance matrix.

Appendix A: Relationship between the minimum of the functionals of 3D-PSAS and 3D-Var

The objective functions of PSAS and 3D-Var can be shown to have the same minimum (with a different sign). Recall the gradient of PSAS functional:

$$\nabla_w F = (\mathbf{R} + \mathbf{HBH}^T)\mathbf{w} - \mathbf{y}' \tag{6.1}$$

At the minimum, the gradient vanishes and the analysis increment in the dual of observation space is :

$$\mathbf{w}_a = (\mathbf{R} + \mathbf{HBH}^T)^{-1}\mathbf{y}' \tag{6.2}$$

thus when replacing w_a in (**??**), the minimum of the F is :

$$F(\mathbf{w}_a) = -\frac{1}{2}{\mathbf{w}_a}^T\mathbf{y}' \tag{6.3}$$

We know that $\delta\mathbf{x}_a = \mathbf{BH}^T\mathbf{w}_a$ is the representer mapping of the increments in the model or the dual of observation space. The 3D-Var functional can then be written in terms of $\mathbf{w}_a$ as

$$\begin{aligned} J(\delta\mathbf{x}_a) =& \frac{1}{2}{\delta\mathbf{x}_a}^T\mathbf{B}^{-1}\delta\mathbf{x}_a + \\ & \frac{1}{2}(\mathbf{H}\delta\mathbf{x}_a - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}_a - \mathbf{y}') \end{aligned} \tag{6.4}$$

$$\begin{aligned} J(\mathbf{w}_a) =& \frac{1}{2}{\mathbf{w}_a}^T\mathbf{HBH}^T\mathbf{w}_a + \\ & \frac{1}{2}[\mathbf{HBH}^T\mathbf{w}_a - \mathbf{y}']^T\mathbf{R}^{-1}[\mathbf{HBH}^T\mathbf{w}_a - \mathbf{y}'] \end{aligned} \tag{6.5}$$

Using eq.6.2, one can show that $\mathbf{HBH}^T\mathbf{w}_a = \mathbf{y}' - \mathbf{R}\mathbf{w}_a$ which gives the result

$$J(\mathbf{w}_a) = \frac{1}{2}{\mathbf{w}_a}^T\mathbf{y}' \equiv -F(\mathbf{w_a}) \tag{6.6}$$

Appendix B: Evaluation of the primal functional associated with the dual iterates

Referring to (**??**) and (**??**), the state vectors are formulated in the model ($\mathbf{v}$) and observation ($\mathbf{u}$) space respectively. A given dual iterate $\mathbf{u}_k$ can be mapped in the model space through the operator $\mathbf{L}^T$ to get $\tilde{\mathbf{v}}_k = \mathbf{L}^T\mathbf{u}_k$, and the primal function is then expressed as

$$\begin{aligned} J(\tilde{\mathbf{v}}) =& \frac{1}{2}\mathbf{u}^T\mathbf{L}(\mathbf{I}_n + \mathbf{L}^T\mathbf{L})\mathbf{L}^T\mathbf{u}_k - \mathbf{u}^T\mathbf{L}\mathbf{L}^T\tilde{\mathbf{y}} + \frac{1}{2}\tilde{\mathbf{y}}^T\tilde{\mathbf{y}} \\ =& \frac{1}{2}\mathbf{u}^T(\mathbf{I}_m + \mathbf{L}\mathbf{L}^T)\mathbf{L}\mathbf{L}^T\mathbf{u}_k - \mathbf{u}^T\mathbf{L}\mathbf{L}^T\tilde{\mathbf{y}} + \frac{1}{2}\tilde{\mathbf{y}}^T\tilde{\mathbf{y}} \end{aligned}$$

Adding and subtracting the functional $F(\mathbf{u_k}) = \frac{1}{2}\mathbf{u}^T(\mathbf{I}_m + \mathbf{L}\mathbf{L}^T)\mathbf{u} - \mathbf{u}^T\tilde{\mathbf{y}}$ and rearranging terms, this can be rewritten as

$$\begin{aligned} J(\tilde{\mathbf{v}}) =& \frac{1}{2}((\mathbf{I}_m + \mathbf{L}\mathbf{L}^T)\mathbf{u} - \tilde{\mathbf{y}})^T((\mathbf{I}_m + \mathbf{L}\mathbf{L}^T)\mathbf{u} - \tilde{\mathbf{y}}) \\ & - \frac{1}{2}\mathbf{u}^T(\mathbf{I}_m + \mathbf{L}\mathbf{L}^T)\mathbf{u} + \tilde{\mathbf{y}}^T\mathbf{u} \end{aligned}$$

Using (**??**) and (**??**), this corresponds to

$$J(\tilde{\mathbf{v}}) = \frac{1}{2}\|\nabla_{\mathbf{u}}(F)\|^2 - F(\mathbf{u}_k).$$

Appendix C: The CG and MINRES algorithms

This is an outline of the derivation of the conjugate gradient and minimum residual algorithms. The reader is referred to Choi (2006) and Barrett *et al.* (1994) as well as to Hestenes and Stiefel (1952) and Paige and Saunders (1975) for more details.

To solve the linear system $\mathbf{Ax} = \mathbf{b}$, many iterative methods work first on the matrix $\mathbf{A}$ by applying some transformations, diagonalizations or decompositions. The Lanczos algorithm transforms a symmetric matrix $\mathbf{A}$ to a symmetric tridiagonal matrix with an additional row at the bottom

$$\underline{\mathbf{T}_k} = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \beta_3 & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \beta_k \\ & & & \beta_k & \alpha_k \\ & & & & \beta_{k+1} \end{bmatrix}$$

The square symmetric matrix of the first k rows of $\underline{\mathbf{T}_k}$ is denoted $\mathbf{T}_k$. The Lanczos algorithm iteratively builds a basis $\mathbf{V}_k = [\mathbf{v}_1 \cdots \mathbf{v}_k]$ of orthonormal vectors, so that :

$$\mathbf{AV}_k = \mathbf{V}_{k+1}\underline{\mathbf{T}_k}$$

In exact arithmetic, the columns of $\mathbf{V}_k$ are orthonormal and the process stops when $\beta_{k+1} = 0$, and we get :

$$\mathbf{AV}_k = \mathbf{V}_k\mathbf{T}_k$$

As explained by Choi (2006), each Lanczos step consists in finding $\mathbf{x}_k$ in the Krylov subspace $\mathbf{K}_k(\mathbf{A}, \mathbf{b})$ such that $\mathbf{x}_k = \mathbf{V}_k\mathbf{y}$, for some $\mathbf{y} \in \mathbf{R}^k$. It follows that $\mathbf{r}_k = \mathbf{b} - \mathbf{Ax}_k = \mathbf{V}_{k+1}(\beta_1\mathbf{e}_1 - \underline{\mathbf{T}_k}\mathbf{y})$, and all Lanczos-based methods attempt to make

$\beta_1 \mathbf{e}_1 - \underline{\mathbf{T}_k}\mathbf{y}$ small. The CG focuses on the first k equations, attempting to solve for $\mathbf{T}_k\mathbf{y} = \beta_1\mathbf{e}_1$ by applying the Cholesky decomposition of $\mathbf{T}_k$, while Minres works to minimize the 2-norm of $\beta_1\mathbf{e}_1 - \underline{\mathbf{T}_k}y$ by applying the QR decomposition to $\underline{\mathbf{T}_k}$.

The algorithms of the CG and MINRES are displayed in detail in Choi (2006). As discussed in Greenbaum (1997), for both algorithms, no explicit form of the matrix $\mathbf{A}$ is needed, and only few vectors need to be stored.

The version of MINRES used in this paper is provided by SOL, Stanford University under the terms of the OSI Common Public License (CPL) :

http://www.opensource.org/licenses/cpl1.0.php.

[2], [1], [3], [4], [6], [5], [7], [8], [9], [10], [11], [12], [13], [14], [15], [17], [16], [18], [19], [20], [21], [22], [23], [24], [26], [27], [29], [28], [25], [30], [31], [32], [33], [34], [36], [35], [37], [38], [39], [40], [42], [41], [43], [44], [45], [46], [47], [48], [50], [49], [51], [52], [53], [56], [54], [55], [58], [57], [59], [60], [61], [62], [63], [64], [65], [**?**], [66], [67], [68], [69], [70], [71], [72], [73], [74], [76], [75], [77], [78], [79], [80], [81], [82], [83], [84], [84], [85], [86], [87], [88], [89], [90], [91], [93], [92], [94], [95], [96], [97], [98], [99], [100], [102], [101], [103], [104], [105], [107], [106], [108], [109], [110], [111], [112], [113], [114], [115], [116], [118], [117], [120], [121], [119], [122], [123], [124], [125], [126], [127], [129], [128],

References

[1] L. Amodei. Approached solution for a data assimilation problem taking into account model errors (title originally in french). *C.r.Acad.sci.*, 321-II:1087–1094, 1995.

[2] E. Andersson, M. Fisher, R. Munro, and A. McNally. Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Q.J.R. Meteorol. Soc.*, 126:1455–1472, 2000.

[3] D. Auroux. Generalization of the dual variational data assimilation algorithm to a nonlinear layered quasi-geostrophic ocean model. *Inverse Problems*, 23:24852503, 2007.

[4] R. Barrett, M. Berry, T.F. Chan, J. Demmel, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods, 2nd Edition.* SIAM, Philadelphia, PA., 1994.

[5] A.F. Bennett. *Inverse Modeling of the Ocean and Atmosphere.* Cambridge University Press, 2002.

[6] A.F. Bennett, L. M. Leslie, C.R. Hagelberg, and P. E. Powers. Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method. *Mon. Wea. rev.*, 121, issue 6:1714–1729, 1993.

[7] K.H. Bergman. Multivariate analysis of temperatures and winds using optimum interpolation. *Mon. Wea. Rev.*, 107:1423–1444, 1979.

[8] P. Bergthorsson and B. Döös. Numerical weather map analysis. *Tellus*, 7:326–340, 1955.

[9] L. Berre, G. Desroziers, L. Raynaud, R. Montroty, and F. Gibier. Consistent operational ensemble variational assimilation. In *Proceedings of the 5th WMO International Sympisum on Data Assimilation, Melbourne, Australia, 5-9 October 2009, 196*, 2009.

[10] S. C. Bloom and S. Schubert. The influence of monte carlo estimates of model error growth on the gla oi assimilation system. Int. symp. on assimilation of observations in meteorology and oceanography. WMO, Geneva., 1990.

[11] G.J. Boer. A spectral analysis of predictability and error in an operational forecast system. *Mon. Wea. Rev.*, 112:1183–1197, 1984.

[12] F. Bouttier and P. Courtier. Data assimilation concepts and methods. ECMWF Meteorological Training Course, 1999.

[13] M. Buehner. Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational nwp setting. *Q. J. R. Meteorol. Soc.*, 131:1013–1044, 2005.

[14] S. Buis, A. Piacentini, and D. Dclat. Palm: a computational framework for assembling high-performance computing applications. *Concurrency Computa.: Pract.Exper*, 18:231–245, 2006.

[15] C. Cardinali, S. Pezzulli, and E. Andersson. Influence-matrix diagnostic of a data assimilation system. *Q.J.R. Meteorol. Soc*, 130 (B),603:2767–2786, 2004.

[16] J.G. Charney. *Dynamical forecasting by numerical process*. Compendium of meteorology. American Meteorological Society, Boston, MA, 1951.

[17] J.G. Charney, R. Fjørtoft, and J. von Newman. Numerical integration of the barotropic vorticity equation. *Tellus*, 2:237–254, 1950.

[18] S.-C. Choi. *Iterative methods for singular linear equations and least-square problems*. PhD thesis, Standford University, 2006.

[19] S.E. Cohn. An introduction to estimation theory. *J. Meteor. Soc. Japan*, 75 (1B):257–288, 1997.

[20] S.E. Cohn, A. DA Silva, J. Guo, M. Sienkiewicz, and D. Lamich. Assessing the effects of data selection with the dao physical-space statistical analysis system.

Mon. Wea. Rev, 126:2913–2926, 1998.

[21] P. Courtier. Dual formulation of four-dimensional variational assimilation. *Q.J.R. Meteorol. Soc.*, 123:2449–2461, 1997.

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

[23] G.P. Cressman. An operational onjective analysis system. *Mon. Wea. Rev.*, 87:367–374, 1959.

[24] J. Cullum and A. Greenbaum. Relations between galerkin and norm-minimizing iterative methods for solving linear systems. *SIAM J. Matrix Anal. Appl*, 17,issue 2:223–247, 1996.

[25] A. Da Silva, J. Pfaendtner, J. Guo, M. Sienkiewicz, and S. Cohn. Assessing the effects of data selection with dao's physical-space statistical analysis system. *Proceedings of the second international WMO symposium on assimilation of observations in meteorology and oceanography, Tokyo, 13-17 March 1995. WMO/TD*, 651:273–278, 1995.

[26] A. Dalcher and E. Kalnay. Error growth and predictability in operational ecmwf forecasts. *Tellus*, 39A:474–491, 1987.

[27] R. Daley. *Atmospheric data analysis*. Cambridge University Press, Cambridge., 1991.

[28] R. Daley. Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Wea. Rev.*, 120:1735–1746, 1992.

[29] R. Daley and E. Barker. The navdas sourcebook. *Naval Research Laboratory, NRL/PU/7530*, 00-418:151 pp, 2000.

[30] D. Dee. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, 123, no4:1128–1145, 1995.

[31] D. Dee and G. Gaspari. Development of anisotropic correlation models for atmospheric data assimilation. In *Preprints of the 11th conference on numerical weather prediction, Norfolk, VA, Amer. Meteor. Soc.*, pages 249–251, 1996.

[32] D. P. Dee and A. Da Silva. Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, 124:269–295, 1998.

[33] J.E. Dennis and R.B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice-Hall: Englewood Cliffs, New Jersey, USA. Reprinted as 'Classics in Applied Mathematics',SIAM: Philadelphia, USA, 1983.

[34] J. Derber. A variational continuous assimilation technique. *Mon. Wea. Rev.*, 117:2437–2446, 1989.

[35] J. Derber and F. Bouttier. A reformulation for the background error covariance in the ecmwf global data assimilation system. *Tellus*, 51A:195–222, 1999.

[36] J. C. Derber and W.-S Wu. The use of tovs cloud-cleared radiances in the ncep ssi analysis system. *Mon. Wea. Rev.*, 126:2287–2302, 1998.

[37] G. Desroziers. A coordinate change for data assimilation in spherical geometry of frontal structures. *Mon. Wea. Rev.*, 125:3030–3038, 1997.

[38] G. Desroziers, G. Hello, and J.-N. Thépaut. A 4d-var reanalysis of the fastex. *Q.J.R. Meteorol. Soc.*, 129:1301–1325, 2003.

[39] M. Ehrendorfer and J.J. Tribbia. Efficient prediction of covariances using singular vectors. In *Preprint Volume, 6th Intl. Meeting on Statistical Climatology, Galway*, pages 135–138, 1995.

[40] A. El Akkraoui. Mise en oeuvre d'une analyse variationnelle atmosphérique gérée par le coupleur palm: une étude de faisabilité. Internship report,Ecole Nationale de la Météorologie, Toulouse, N931, 2004.

[41] A. El Akkraoui and P. Gauthier. Convergence properties of the primal and dual forms of variational data assimilation. *Q.J.R. Meteorol. Soc.*, in press,

2010.

[42] A. El Akkraoui, P. Gauthier, S. Pellerin, and S. Buis. Intercomparison of the primal and dual formulations of variational data assimilation. *Q.J.R. Meteorol. Soc.*, 134:1015–1025, 2008.

[43] G. Evensen. sequential data assimilation with a nonlinear quasigeotrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, 99 (C5):10143–10162, 1994.

[44] J. R. Eyre. Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. *Q. J. R. Meteorol. Soc.*, 115:1001–1037, 1989.

[45] M. Fisher. Developments of a simplified kalman filter. *ECMWF Tech. Memo*, 260:16, 1998.

[46] M. Fisher. Assimilation techniques: 3dvar. ECMWF Meteorological Training Course, 2001.

[47] M. Fisher. Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. Technical Memo. 397, ECMWF, Reading, UK, 2003a.

[48] M. Fisher. Background error covariance modelling. In *Proceedings of seminar on recent developments in data assimilation for and ocean, 2-12 septembe 2003,*

ECMWF, reading, UK, pages 1–28, 2003b.

[49] M. Fisher and E. Andersson. Developments in 4d-var and kalman filtering. *ECMWF Tech. Memorandum*, 347:36 pages, 2001.

[50] M. Fisher and P. Courtier. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF Technical Memorandum,Reading U.K, 220, 28pp, 1995.

[51] 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.*, 131:3235–3246, 2005.

[52] R.W. Freund, G.H. Golub, and N.M. Nachtigal. Iterative solution of linear systems. *Acta Numerica*, 1:57–100, 1992.

[53] L.S. Gandin. *Objective analysis of meteorological fields, Gidrometerologicheskoe Isdatelstvo*. Leningrad. English translation by Israeli Program for Scientific Translations, Jerusalem, 1965.

[54] P. Gauthier, M. Buehner, and L. Fillion. Background-error statistics modeling in a 3d variational data assimilation scheme: estimation and impact on the analyses. In *Proceedings, ECMWF Workshop on diagnosis of data assimilation systems, 2-4 November 1998, Reading, UK*, 1998.

[55] P. Gauthier, C. Charette, L. Fillion, P. Koclas, and S. Laroche. Implementation of a 3d variational data assimilation system at the canadian meteorological centre. part i: The global analysis. *Atmos.-Ocean*, 37:103–156, 1999.

[56] P. Gauthier, P. Courtier, and P. Moll. Assimilation of simulated wind lidar data with a kalman filter. *Mon. Wea. Rev.*, 121:1803–1820, 1993.

[57] P. Gauthier, M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau. Extension of 3d-var to 4d-var: implementation of 4d-var at the meteorological service of canada. *Mon. Wea. Rev.*, 135,issue 6:2339–2354, 2007.

[58] P. Gauthier and J.-N. Thépaut. Impact of the digital filter as a weak constraint in preoperational 4dvar assimilation system of météo-france. *Mon. Wea. Rev.*, 129:2089–2102, 2001.

[59] J.C. Gilbert and C. Lemaréchal. Some numerical experiments with variable-storage quasi-newton algorithms. *Mathematical Programming*, 45:407–435, 1989.

[60] B. Gilchrist and G. Cressman. An experiment in objective analysis. *Tellus*, 6:309–18, 1954.

[61] G.H. Golub and C.F. Van Loan. *Matrix computations. Third edition.* The Johns Hopkins University Press, Baltimore and London., 1996.

[62] A. Greenbaum. *Iterative methods for solving linear systems*. SIAM, Philadelphia, 1997.

[63] A.K. Griffith and N.K. Nichols. Adjoint methods in data assimilation for estimating model error. *Flow, Turbulence and Combustion*, 65:469–488, 2001.

[64] T.M. Hamill, J.S. Whitaker, and C. Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Mon. Wea. rev.*, 129:2776–2790, 2001.

[65] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49:409–436, 1952.

[66] Y. Honda, M. Nishijima, K. Koizumi, Y. Ohta, K. Tamiya, Kawabata T., and T. Tsuyuki. A pre-operational variational data assimilation system for a non-hydrostatic model sat the japan meteorological agency: Formulation and preliminary results. *Q. J. R. Meteorol. Soc.*, 131:3465–3475, 2005.

[67] P.L. Houtekamer, L. Lefaivre, J. Derome, H. Ritchie, and H.L. Mitchell. A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, 124:1225–1242, 1996.

[68] P.L. Houtekamer and H.L. Mitchell. Data assimilation using an ensemble kalman filter technique. *Mon. Wea. Rev.*, 126:796–811, 1998.

[69] M. Janisková, J.-N. Thépaut, and J.-F. Geleyn. Simplified and regular physical parametrizations for incremental four-dimensional variational assimilation. *Mon. Wea. Rev.*, 127:2645, 1999.

[70] A.H. Jazwinski. *Stochastic processes and filtering theory.* Academic Press, New York, 1970.

[71] R. Kalman. A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.*, 82:35–45, 1960.

[72] S. Kaniel. Estimates for some computational techniques in linear algebra. *Math. Comp.*, 20:369–378, 1966.

[73] M. Kilmer and G.W. Stewart. Iterative regularization and minres. *SIAM J. Matrix Anal. Appl.*, 21(2):613–628, 1999.

[74] E. Klinker, F. Rabier, G. Kelly, and J.-F. Mahfouf. The ecmwf operational implementation of four-dimensional variational assimilation. iii: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteorol. Soc.*, 126:1191–1215, 2000.

[75] T. Lagarde, A. Piacentini, and O. Thual. A new representation of data assimilation methods: the palm flow charting approach. *Quarterly Journal of the Royal Met. Soc.*, 127:189–207, 2001.

[76] C. Lanczos. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Standards*, 45:255–282, 1950.

[77] R.H. Langland and N.L. Baker. Estimation of observation impact using the nrl atmospheric variational data assimilation adjoint system. *Tellus*, 56A:189–203, 2004.

[78] S. Laroche and P. Gauthier. A validation of the incremental formulation of 4d variational data assimilation in a nonlinear barotropic flow. *Tellus*, 50A:557–572, 1998.

[79] S. Laroche, P. Gauthier, M. Tanguay, S. Pellerin, and J. Morneau. Impact of the different components of 4d-var on the global forecast system of the meteorological service of canada. *Mon. Wea. Rev.*, 153, issue 6:2355–2364, 2007.

[80] F.X. LeDimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations - theoretical aspects. *Tellus*, 38A:97–110,

1986.

[81] J. Lewis and J. Derber. The use of adjoint equations to solve a variational adjustment problem with advection constraints. *Tellus*, 37:309–27, 1985.

[82] J. M. Lewis, S. Lakshmivarahan, and S. K. Dhall. *Dynamic data assimilation: a least square approach*. Cambridge University Press, Cambridge, 2006.

[83] A. Lorenc. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, 109:701–721, 1981.

[84] A. Lorenc. Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 112:1177–1194, 1986.

[85] E.N. Lorenz. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20:130–141, 1963.

[86] S. Louvel. Implementation of a dual variational algorithm for assimilation of synthetic altimeter data in the oceanic primitive equation model micom. *J. of Geophys. Res.*, 106,C5:9199–9212, 2001.

[87] P. Lynch. *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge: Cambridge University Press, 2006.

[88] J.-F Mahfouf and F. Rabier. The ecmwf operational implementation of four-dimensional variational assimilation. ii: Experimental results with improved

physics. *Q. J. R. meteorol. Soc.*, 126:1171–1190, 2000.

[89] I. M. Navon and D.M. Legler. Conjugate gradient methods for large scale minimization in meteorology. *Mon. Wea. Rev.*, 115:1479–1502, 1987.

[90] A. O'Neill, P.P. Mathieu, and C. Zehner. Making the most of earth observation with data assimilation. *ESA Bulletin*, 118:33–38, 2004.

[91] D. Orrell, L. Smith, J. Barkmeijer, and T. Palmer. Model error in weather forecasting. *Nonlin. Processes Geophys.*, 8:357–371, 2001.

[92] C. Paige, B. Parlett, and H. van der Vorst. Approximate solutions and eigenvalue bounds from krylov subspaces. *Numer. Lin. Appl.*, 29:115–134, 1995.

[93] C.C. Paige and M.A. Saunders. Solution of sparse indefinite systems of linear equations. *SIAM J. Numer. Anal.*, 12:617–629, 1975.

[94] J. Pailleux. A global variational assimilation scheme and its application for using tovs radiances. In *Proc. WMO International Symposium on Assimilation of observations in meteorology and oceanography, Clermont-Ferrand, france*, pages 325–328, 1990.

[95] H. Panofsky. Objective weather-map analysis. *J. Appli. Meteor.*, 6:386–392, 1949.

[96] D. Parish and J. Derber. The national meteorological center's spectral statistical analysis system. *Mon. Wea. Rev.*, 120:1747–1763, 1992.

[97] S. Pellerin, P. Gauthier, M. Tanguay, S. Laroche, and J. Morneau. A modular implementation of the operational 4d-var data assimilation at the meteorological service of canada. In *Proceedings of the* 4^{th} *WMO Symposium on assimilation of observations in meteorology and oceanography,Prague, Czech Republic, 18-22 April 2005 (available from the World Meteorological Organization)*, 2006.

[98] C. Pires, R. Vautard, and O. Talagrand. On extending the limits of variational assimilation in chaotic systems. *Tellus*, 48A:96–121, 1996.

[99] F. Rabier. Overview of global data assimilation developments in numerical weather-prediction centres. *Q. J. R. Meteorol. Soc.*, 131:3215–3233, 2005.

[100] F. Rabier and P. Courtier. Four-dimensional assimilation in the presence of baroclinic instability. *Q.J.R. Meteorol. Soc.*, 118:649–672, 1992.

[101] F. Rabier, H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons. The ecmwf operational implementation of four dimensional variational assimilation. part i: Experimental results with simplified physics. *Quart. J.R. Meteor. Soc.*, 126:1143–1170, 2000.

[102] F. Rabier, N. McNally, E. Andersson, P. Courtier, P. Unden, J. Eyre, A. Hollingsworth, and F. Bouttier. The ecmwf implementation of three-dimensional variational assimilation (3d-var). ii: Structure function. *Quart. J. Roy. Meteor. Soc.*, 124:1809–1830, 1998.

[103] F. Rawlins, S.P. Ballard, K.J. Bovis, A.M. Clayton, D. Li, G.W. Inverarity, A.C. Lorenc, and T.J. Payne. The met office global four-dimensional variational data assimilation scheme. *Quart. J.R. Meteor. Soc*, 133 issue 623:347–362, 2007.

[104] L.F. Richardson. *Weather prediction by numerical process.* Cambridge, 1922.

[105] C.D. Rodgers. Inverse methods for atmospheric sounding : theory and practice. *Series on atmospheric, oceanic, and planetary physics, World Scientific Ed., New York*, 2:238 pages, 2000.

[106] Y. Saad. *Iterative Methods for Sparse Linear Systems.* PWS Publishing Co., Boston, 1996.

[107] Y. Saad and M.H. Schultz. Gmres: a generalized minimal residual algorithm for solving non symmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 7, issue 3:856 – 869, 1986.

[108] Y. Sasaki. An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, 36:77–88, 1958.

[109] Y. Sasaki. Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis. *J. Meteor. Soc. Japan*, 47:115–124, 1969.

[110] Y. Sasaki. Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, 98:875–883, 1970.

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

[112] R. Swinbank, V. Shutyaev, and W.A. Lahoz. Data assimilation for the earth system. *NATO Science Series. Earth and Environmental Sciences*, 26, Kluwer Academic Publishers:377pp, 2003.

[113] O. Talagrand. Objective validation and evaluation of data assimilation. In *Proc. ECMWF seminar on recent developments in data assimilation for atmoshphere and ocean, 8-12 September 2003, Reading, UK*, pages 287–299, 2003.

[114] O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint vorticity equation. i: Theory. *Q.J.R. Meteor. Soc.*,

113:1321–1328, 1987.

[115] M. Tanguay, P. Bartello, and P. Gauthier. Four dimensional data assimilation with a wide range of scales. *Tellus*, 47A:974–997, 1995.

[116] M. Tanguay and S. Laroche. Grid-point response to the incremental strategy for variational applications. *Q. J. R. Meteorol. Soc.*, 128:385–397, 2002.

[117] J.-N. Thépaut. Satellite data assimilation in numerical weather prediction: an overview. In *Proceedings of ECMWF Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, ECMWF, Reading, UK 8-12 September*, pages 75–96, 2003.

[118] J.-N. Thépaut and P. Courtier. Four-dimensional variational assimilation using the adjoint of a multilevel primitive-equation model. *Q.J.R. Meteorol. Soc.*, 119:153–186, 1991.

[119] Y. Trémolet. Model error in variational data assimilation. In *Proc. ECMWF Seminar on.*

[120] Y. Trémolet. Accounting for an imperfect model in 4d-var. *Quart. J.R. Meteorol. Soc.*, 132:2483–2504, 2006.

[121] Y. Trémolet. Model-error estimation in 4d-var. *Q.J.R. Meteorol. Soc.*, 133,issue 626A:1267–1280, 2007.

[122] J. Verron, L. Gourdeau, D. Pham, R. Murtugudde, and A. Busalacchi. An extended kalman filter to assimilate altimeter data into a nonlinear numerical model of the tropical pacific ocean: method and violation. *Journal of Geophysical Research*, 104 (c3):5441–5458, 1999.

[123] A. Vidard, E. Blayo, F.-X Le Dimet, and A. Piacentini. 4d-variational data analysis with imperfect model. reduction of the size of the control. *J. Flow Turbulence Combustion*, 65:489–504, 2001.

[124] A. Vidard, F.-X. Le Dimet, and A. Piacentini. Optimal determination of nudging coefficient. *Tellus*, 55A:1–15, 2003.

[125] P. A. Vidard, A. Piacentini, and F.-X. Le Dimet. Variational data analysis with control of the forecast bias. *Tellus*, 56A:177–188, 2004.

[126] W. Wergen. Effect of model errors in variational assimilation. *Tellus*, 44A:297–313, 1992.

[127] L. Zhou and H.F. Walker. Residual smoothing techniques for iterative methods. *SIAM J. Sci. Comput.*, 15:297–312, 1994.

[128] D. Zupanski. A general weak constraint applicable to operational 4dvar data assimilation systems. *Mon. Wea. Rev.*, 125:2274–2292, 1997.

[129] M. Zupanski. Regional 4-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.*, 121:2396–2408, 1993.