Universidade de Lisboa

Faculdade de Ciências

Departamento de Engenharia Geográfica, Geofísica e Energia

Diagnosis and impact of time variations of regionally averaged background error covariances

Maria José Correia Monteiro

Doutoramento em Ciências Geofísicas e da Geoinformação
(Meteorologia)

2011

Universidade de Lisboa

Faculdade de Ciências

Departamento de Engenharia Geográfica, Geofísica e Energia

Diagnosis and impact of time variations of regionally averaged background error covariances

Maria José Correia Monteiro

Doutoramento em Ciências Geofísicas e da Geoinformação
(Meteorologia)

Tese orientada por:
Prof. Carlos Alberto L. Pires
Dr. Loïk Berre

2011

Acknowledgements

First I would like to express my gratitude to my supervisors. Loïk Berre provided a wise and patient guidance, an endless motivation and his friendship and Carlos Pires backed enthusiatically this work on data assimilation since its beginning.

This investigation was done within the framework of the research programme of the international High Resolution Numerical Weather Prediction project ALADIN. The long term commitment between Instituto de Meteorologia, I.P. (IM) and the ALADIN project laid the fruitfull ground that made this work possible. For their carefull and enlightened management, I adress my sincere thanks to Adérito Vicente Serrão, Jean-François Geleyn, Piet Termonia and Claude Fischer.

Many tasks were performed during stays at Météo-France/CNRM/GMAP under the ALADIN flat-rate funding. I take this opportunity to present my warmest compliments to the French colleagues who friendly hosted me in their working environment. Furthermore, Eric Sevault, Guillaume Beffrey, Veronique Mathiot, Pierre Brousseau, Ludovic Auger, Francis Pouponneau, Ryad El Khatib, Jean-Daniel Gril, Jean-Antoine Maziejewski and Eric Gerbier provided me precious scientific and technical support. I will always remember with gladness the time shared in Toulouse with them and also with other ALADIN community colleagues.

At IM, I value the understanding of my colleagues of the Numerical Weather Prediction (NWP) group for freed me from some of the team regular duties. José Carlos Monteiro

and Fernando Oliveira, from the computers division, were very kind in providing me local technical support. Besides, I have to thank Victor Prior for believing that this work could bring an added value to our service. Vanda Costa, as Coordinator of the NWP group activities, Teresa Abrantes, as Head of the Department of Meteorology and Climatology, and Pedro Viterbo, as Head of Research, also played an important personal and institutional support.

I would have been ungrateful in here to not remember some ECMWF colleagues that always volunteered on my cries for help with their timeless expertise in severe technical issues, in particular, Manuel Fuentes and Fernando Ii.

Finally, I acknowledge those who have had a crucial role on many different situations along this project but whose name is not mention here, many colleagues and friends, my brother and, specially, my closest family which gave me the needed space and time to fulfill this long term wish, many times under its own deprivation.

I dedicate this publication to my parents.

Abstract

Current data assimilation systems aim at produce the best possible estimate of the atmospheric state to be used as initial conditions by a forecasting model. This estimate is called the model analysis. In the formalism of the Three-Dimensional Variational Data Assimilation (3D-Var), the analysis should minimize a quadratic cost function that measures the distance between the estimate and the available information, namely the observations and the background, taking into account their respective precisions. Observation error covariances contain information about errors in the observation process (measurement and representativeness uncertainties). Background error covariances are used to filter and propagate the observational information in a spatial multivariate way. They compose the gain matrix that determines how the analysis increments (that is to say, the analysis corrections to the background) are obtained from the innovations (i.e., the differences between observations and background at the observation locations). However, since the true atmospheric state can not be exactly known, background error covariances (in particular) can only be estimated.

Nowadays, Numerical Weather Prediction (NWP) centers often use an off-line specification of background error covariances. This relies on running an ensemble of perturbed assimilation cycles during a past period, to simulate the error evolution during the successive analysis/forecast steps. Moreover, as the full background error covariance matrix is far too large to be handled explicitly, sparse covariance models are usually employed, and they are calibrated by using ensemble forecast perturbations. These sparse covariance models are often based on simplifications such as horizontal homo-

geneity and temporal stationarity. This is the case of the ALADIN/France regional system which has been used in this study.

The impact of seasonal and daily variations are studied. Results indicate that specifying background error covariances corresponding to a one-day sliding average (preceding each each analysis date) has a positive impact on the short-range forecast quality, compared to the currently operational static covariances. This positive impact arises to a large extent from the update of the monthly component of covariance variations. The update of the daily component contributes to additional positive impacts, which are however relatively localized and modest during this period. These impacts are illustrated by case studies for humidity during an anticyclonic situation, and for wind during a cyclonic event. These results support the idea to consider an on-line updated specification of background error covariances.

Keywords: limited area model, 3D-Var, ensemble assimilation cycling, time-dependent background error covariance estimates, operational application

Resumo

Em previsão numérica do tempo (PNT), um sistema de assimilação de dados tem como objetivo produzir a melhor estimativa possível do estado da atmosfera num dado instante utilizando toda a informação relevante disponível. Esta estimativa é conhecida como "análise" e visa fornecer condições iniciais de qualidade a um modelo de previsão numérica. Como informação relevante consideram-se tipicamente as observações e uma estimativa conhecida *a priori* que normalmente corresponde a uma previsão de curto alcance, obtida por integração do modelo numérico a partir de uma análise anterior. As observações são de dois tipos: as observações *in situ* (como são as observações de altitude resultantes de radiossondagens e as observações de superfície obtidas a partir de uma rede de estações assentes sobre a superfície terrestre) e por deteção remota (como são as radiações medidas a partir de satélites geostacionários ou de órbita polar e ainda as refletividades medidas por radares meteorológicos). No formalismo da assimilação variacional tri-dimensional conhecido como 3D-Var, a análise resulta da minimização de uma função custo quadrática que mede a distância da análise às observações e à estimativa *a priori* tomando em consideração a precisão com que estas fontes de informação são conhecidas. Em particular, é utilizada a covariância dos erros das observações. Esta covariância contêm a informação sobre os erros de medição e de representatividade que afetam as observações. É também utilizada a covariância dos erros da estimativa *a priori* com a função de filtrar e propagar, espacialmente e de forma multivariada, a informação observada. Estas duas entidades, a matriz de covariância dos erros das observações e a matriz de covariância dos erros da estimativa *a priori*, compõem a chamada "matriz ganho" da equação da análise. Esta equação determina

como é que as correções à estimativa *a priori* (i.e., os incrementos da análise) são obtidas a partir das inovações (i.e., as diferenças entre as observações e a estimativa *a priori* nos locais das observações). Contudo, dado que o estado exato da atmosfera não pode ser conhecido, também as covariâncias dos erros apenas podem ser estimadas com base em hipóteses que só podem ser confirmadas, ou não, *a posteriori* em amostras independentes daquelas onde se calibram as estatísticas dos erros. Este é um dos principais desafios científicos da assimilação de dados na atualidade.

Os centros modernos de PNT usam frequentemente uma especificação da covariância dos erros da estimativa *a priori* que é obtida por processos independentes do ciclo de assimilação em tempo não-real. Dado que não varia ao longo do ciclo de assimilação, diz-se tratar-se de uma especificação de tipo "estático". Além disso, recorrem a técnicas de *ensemble* baseadas na perturbação do ciclo de assimilação durante um determinado período de tempo (passado) para simular a evolução dos erros ao longo dos diversos passos da análise e da previsão (por dizer respeito a estados passados da atmosfera, esta especificação é vulgarmente designada de "climatológica"). Como a matriz da covariância dos erros da estimativa *a priori* é demasiado grande para ser manipulada de forma explícita, os centros de PNT também empregam modelos que envolvem apenas matrizes esparsas de covariâncias, os quais são calibrados a partir de *ensembles* de previsões perturbadas usando metodologias mais ou menos simplificadas e eventualmente sub-optimizadas. Muito frequentemente, estes modelos são baseados em simplificações, como são as da homogeneidade espacial e a da estacionaridade dos erros, que inibem a representação da dependência da covariância com a evolução da situação meteorológica. Em particular, esta inibição é intrínseca à atual versão operacional do modelo de área limitada ALADIN/France utilizado neste estudo.

Esta tese em Assimilação de Dados tem como objetivo principal demonstrar, num contexto real de PNT, a importância duma representação adequada das variações temporais da covariância do erro da estimativa *a priori* que é utilizada num sistema de área limitada. Tendo como base o sistema numérico de área limitada ALADIN/France, este

estudo evidencia claramente o efeito positivo resultante da utilização de uma especificação dessas estimativas em tempo-(*quasi*) real. A simulação dos erros da estimativa *a priori* foi efetuada por técnicas de *ensemble* sob a hipótese de que o modelo de previsão é perfeito, isto é, de que não existe qualquer erro do modelo (determinista ou estocástico) e para qualquer membro do *ensemble*. Os erros assim estimados foram utilizados para calibrar o modelo de covariâncias esparsas utilizadas no sistema ALADIN/France.

A primeira fase do trabalho, descrita por Monteiro & Berre (2010), visou a generalização da hipótese de estacionaridade utilizada no cálculo das covariâncias dos erros da estimativa *a priori* (previsões a 6 horas), através do diagnóstico da sua variação temporal. Com base nessas estimativas, foram igualmente calculadas escalas integrais de correlação horizontal e vertical. Foi efetuado o diagnóstico da variação temporal nas estimativas da covariância em função da mudança da estação do ano (inverno versus verão), do dia (em conexão com a mudança da situação sinóptica) e da hora (relacionada com o ciclo diurno). Para tal consideraram-se dois períodos de um mês, um no inverno e outro no verão, e em cada um destes períodos construiram-se dois sistemas de *ensembles* de 6 membros, os quais resultaram da perturbação independente do mesmo conjunto de observações. Esta duplicação permitiu confirmar a robustez das estatísticas. De uma forma geral, verificou-se que no verão as variâncias dos erros tendem a ser superiores às que se observam no inverno (para os parâmetros da temperatura, humidade relativa e vento). Em particular, no verão registaram-se correlações que são mais estreitas horizontalmente e mais extensas verticalmente e que estão relacionadas com o aumento da atividade convetiva na região geográfica em estudo (em comparação com a que se observa no inverno). No que respeita às variações diárias verificou-se (no período de inverno em estudo) que o desvio padrão do erro *a priori* é em geral maior quando a situação sinóptica tem maior instabilidade baroclínica (por exemplo, em situações de desenvolvimento depressionário) particularmente no inverno, sendo acompanhada de redução do comprimento de correlação horizontal do erro no caso da humidade e da temperatura. Além disso, observou-se que as variações diárias da variância do error são por vezes mais acentuadas do que as variações sazonais (se compararmos covariâncias

para uma situação anticiclónica de inverno com as de um caso depressionário de inverno, por exemplo). Finalmente, verificou-se que as variações das covariâncias ao longo de um ciclo diurno são significativas na camada limite da atmosfera no período de verão: os desvios padrão tendem a ser maiores para as previsões válidas às 18 UTC, quando a escala do comprimento da correlação horizontal é reduzida (especialmente para os parâmetros da humidade e da temperatura) e as funções de correlação vertical são extensas, sendo este um efeito esperado, dado o aumento da atividade convetiva que acontece com o aquecimento da superfície terrestre durante a tarde, sobre o domínio geográfico em causa.

Na segunda parte do trabalho, foi estudado o efeito da variação sazonal e diária das matrizes de covariância do erro *a priori* sobre a qualidade das previsões do sistema ALADIN/France ao longo de um mês de inverno. Para tal realizaram-se três experiências separadas. Na primeira consideraram-se as covariâncias calculadas a partir de uma média mensal (estática) ao longo de um período passado (esta experiência reproduz as condições do sistema operacional em que a covariância climatológica é referente a um período de outono); na segunda, as covariâncias foram calculadas a partir de uma média mensal (estática) sobre o período de tempo em estudo (inverno); na terceira, as covariâncias foram calculadas a partir de uma média diária móvel (precedente a cada data da análise). Os resultados indicaram que a especificação das covariâncias dos erros da estimativa *a priori* correspondente a uma média diária móvel tem um efeito positivo na qualidade da previsão de curto prazo (para alcances entre as 12 e 18 h, para a temperatura e vento e até às 24 h para a humidade) quando comparada com a qualidade resultante de uma especificação estática climatológica (a utilizada no atual sistema operacional). Este efeito positivo resulta principalmente da atualização sazonal (do mês de outono para o mês de inverno) das covariâncias. A atualização diária contribui também para um efeito positivo adicional mas este é mais localizado e modesto. O efeito da variação diária é ilustrado através de dois casos de estudo, verificados durante o mês de inverno: uma situação anticiclónica em que a melhoria é nítida sobre a humidade e um evento ciclónico em que a melhoria é mais evidente sobre

o campo do vento. Neste caso, regista-se uma melhoria da previsão extremos espaciais do vento a prazos até às 36 h.

Em conclusão, os resultados obtidos sustentam a ideia de que é adequado implementar uma especificação em tempo-(*quasi*) real das covariâncias do erro da estimativa *a priori*, em susbtituição da atual aproximação estática climatológica e, tal como foi visto na primeira parte do trabalho, um *ensemble* de 6 membros já permite obter estimativas diárias robustas a um baixo custo computacional, apesar das muitas melhorias metodológicas que ainda podem ser feitas.

Palavras chave: modelo de area limitada, 3D-Var, assimilação cíclica de *ensemble*, estimativas da covariância do erro da estimativa *ensemble* com dependência temporal, aplicação operacional

List of acronyms and abbreviations

ALADIN Aire Limited Adaptation Dynamique Développment InterNational . . 2

ARPEGE Action de Recherche Petite Echelle Grande Echelle 16

AUT07S Impact experiment of Autumn 2007 monthly-averaged covariances 49

BLUE Best Linear Unbiased Estimator 5

CANARI Code d'Analyse Necessaire a ARPEGE pour ses Rejects et son Initialisation 4

ECMWF European Centre for Medium-range Weather Forecasts 50

EnS1 First ALADIN/France Ensemble for summer 28

EnS2 Second ALADIN/France Ensemble for summer 28

EnW1 First ALADIN/France Ensemble for winter 28

EnW2 Second ALADIN/France Ensemble for winter 28

HIRLAM High Resolution Local Area Model 9

IFS Integrated Forecasting System 18

LAM Limited Area Model 2

LBC Lateral Boundary Conditions 49

NMC National Meteorological Center 13

NWP Numerical Weather Prediction 1

OI Optimal Interpolation 4

PBL Planetary Boundary Layer 40

PNT Previsão Numérica do Tempo vii

REnDA Regional Ensemble Data Assimilation 15

RMSE Random Mean Squared Error 54

WIN08D Impact experiment of winter 2008 daily-averaged covariances 50

WIN08S Impact experiment of winter 2008 monthly-averaged covariances . . 50

3D − Var	Three-Dimensional Data Assimilation	7
4D − Var	Four-Dimensional Data Assimilation	10

List of symbols

$\mathbf{x}^t$	True model state	11
$\mathbf{y}^o$	Observations vector	5
$\mathbf{B}$	Background error covariance matrix	5
$\mathbf{B}_u$	Background error covariance matrix in the space of control variables	20
$\mathbf{C}_{q_u}$	Auto-covariance error matrix for unbalanced specific humidity	20
$\mathbf{C}_{(T,p_s)_u}$	Auto-covariance error matrix for unbalanced mass	20
$\mathbf{C}_{\eta_u}$	Auto-covariance error matrix for unbalanced divergence	20
$\mathbf{C}_{\zeta}$	Auto-covariance error matrix for vorticity	20
$\mathbf{C}_{\zeta,\eta}$	Cross-covariance matrix between vorticity and divergence	19
$\mathbf{C}_{\zeta,\zeta}$	Auto-covariance matrix of vorticity	19
$\mathcal{F}$	Low pass filter	12
H	Observations operator	12
$\mathbf{H}$	Linearized observation operator	5
$\mathbf{H}^T$	$\mathbf{H}$ associated adjoint operator	7
$\mathbf{I}$	Identity matrix	12
J	Cost function	7
J_b	Background cost function	7
J_o	Observations cost function	7
$\mathbf{K}$	Gain matrix	5
$\mathcal{K}$	Balance operator	20
$K^\star$	Maximum horizontal wave number	29
L_x	Zonal domain length	29
L_y	Meridional domain length	29
$\mathcal{L}_z$	Horizontal correlation length scale for teh vertical level z	30
M	Forecast model operator	29
$\mathbf{M}$	Linearized model operator	12
M	Linear regression matrix	19
N	Linear regression matrix	19
P	Linear regression matrix	19
Q	Linear regression matrix	19

$\mathbf{R}$	Observation error covariance matrix	5
R	Linear regression matrix	19
S	Linear regression matrix	19
$\mathbf{T}$	Temperature increment	18
$(\mathbf{T}, \mathbf{P}_s)_u$	Unbalanced mass increment	19
$\mathbf{U}$	Linearized operator that transforms model to control variables error	8
χ	Space of the transformed analysis increment	8
η	Divergence increment	18
η_u	Unbalanced divergence increment	19
γ_z	Normalized spectral density of the variance at level z	21
$\rho(z, z')$	Vertical correlations between vertical levels z and z'	30
$\sigma_b(z)$	Background error standard deviation at level z	29
$\sigma_b(z)^2$	Background error variance at level z	29
ε^a	Analysis perturbation	14
ε^f	Forecast perturbation	14
ε^o	Observation perturbation	14
ε^m	Model perturbation	14
ζ	Vorticity	18

Contents

Chapter 1

Introduction

The goal of current data assimilation systems is to produce the best possible estimate of the atmospheric state, using all relevant available information. This estimate is called the analysis, and it is used to provide accurate initial conditions to a Numerical Weather Prediction (NWP) model. Available information is typically made of observations and a background field, which corresponds to a short-range forecast launched from a previous analysis. Observation types are very diverse, corresponding to either *in situ* measurements (such as radiosondes and surface network data) or remote sensing data (such as radiances from geostationary and polar-orbiting satellites). These different information sources are combined during the analysis step, and they are weighted by their error covariances. More precisely, the role of background error covariances is to spatially filter and propagate observed information. However, estimating and modelling these error covariances is one the main scientifical challenges of data assimilation.

In particular, since the true atmospheric state is never exactly known, ensemble techniques based on perturbed data assimilation cycles are nowadays employed to simulate analysis and forecast errors. Moreover, assumptions of temporal stationarity and horizontal homogeneity are often used to model background error covariances, whereas these simplifications are known to prevent the representation of the dependence of background error covariances with respect to the meteorological situation.

The main goal of the present data assimilation study is to diagnose time variations of regionally averaged background error estimates and to study their impact in a real NWP context. The ALADIN/France system from Météo-France is the experimental framework for the studies here described.

The thesis structure is as follows. In chapter 2 the principles of data assimilation and of the 3D-Var formulation are presented. In chapter 3, the basic concepts of the ALADIN 3D-Var formalism for background error covariance modelling is introduced, as well as the error covariance calibration using the ensemble method. A diagnosis of time variations of background error covariances is exposed in chapter 4. The impact of these time variations on the ALADIN/France limited area model (LAM) is presented in chapter 5. Finally, some conclusions and a future outlook for this working area are given in chapter 6.

Chapter 2

Analysis equations and data assimilation cycling

2.1 Data assimilation as an inverse estimation problem

The problem of projecting an instantaneous atmospheric state on the forecasting model space, using a set of observations irregularly located, can be explained under the view of the mathematical discrete inversion theory (see e.g. Tarantola (1987) or Menke (1989)): it is generally classified as an undetermined problem, since the number of degrees of freedom that define the model state is usually bigger than the number of independent observations available; the undetermination is usually solved by use of *a priori* knowledge on the atmospheric flow.

In modern NWP systems, the *a priori* knowledge on the atmospheric flow comes from a short-range forecast (called the background, which is typically a 6 h forecast), from the knowledge of the statistical structure and amplitude of the observation and background errors, and from approximations of known balanced features of the meteorological parameters involved.

The practical problem of finding the projection which gives the best possible estimate

a model can have in a certain moment, forces us to some abstractions. Firstly, the atmospheric reality is far more complex than what a model state can represent, for instance due to its finite spatial resolution; therefore, the best possible atmospheric representation a model state can give is an approximation of the true atmospheric conditions. It is called the true model state. Secondly, due to the undetermined nature of the projection process, we realize that the model state that we seek is an estimate - the best possible - of the true model state. It is called the analysis. Finally, we look at the background as a previously determined best model state estimate (known before using the observations at the analysis time).

Since the middle of the 20th century, several mathematical methods have been used by different authors to solve the practical problem of the atmospheric objective analysis estimation. Two different approaches, either deterministic or probabilistic, have been considered.

In the deterministic estimation theory, the inversion problem is solved by looking for the parameter estimates that minimise the distance to their true value. This distance is defined using a meaningful norm. One particular case of this approach is the least-square analysis estimate (or the minimum variance estimate), where the distance is measured by a squared (or Euclidian) norm [1] .

Using the Bayesian probabilistic approach, Lorenc (1986) proposed a generalised formulation of the atmospheric analysis, and derived previously used methods. The methods discussed included variational techniques, successive corrections (Bergthósson & Döos (1955), Cressman (1959)), and optimal interpolation (OI) (Gandin 1963) amongst others. Some of these methods are still valuable for low cost applications such as the surface parameter analysis CANARI (Taillefer 2002) [2] .

[1] It should be noted that this norm specification is particularly suitable for the case when the parameters to estimate have probability distribution functions not too dispersive around their average value, as it happens with Gaussian distributed values.

[2] As it is described in appendix A, some applications of these methods have been recently used for

In particular, assuming that the inverse of the analysis problem, the so-called direct problem [3] , can be represented by a linear approach of a generalised interpolation and that the background and observation errors have a Gaussian distribution function, it is possible to prove in a simple way (see e.g. Desroziers (2001)) that the probabilistic approach, usually called of maximum-likehood, can be reduced to the simplest deterministic least-square approach (i.e the minimum error variance estimate).

2.2 The BLUE analysis equation and the role of B

In this work, we assume that the analysis is a linear combination of the background and the observations, that the background and the observation errors are unbiased, not mutually correlated, and have a Gaussian probability distribution function. Besides, we consider a linear approach of the generalised interpolation. Under these conditions both the probabilistic and the deterministic methods give the same solution which is called the Best Linear Unbiased Estimator (BLUE) analysis. The BLUE analysis $\mathbf{x}^a$ is the minimum squared error unbiased estimate, and its equation is given by:

$$\mathbf{x}^a = (\mathbf{I} - \mathbf{KH})\mathbf{x}^b + \mathbf{K}\mathbf{y}^o \tag{2.1}$$

where $\mathbf{x}^b$ is the background n-dimensional model state vector; $\mathbf{y}^o$ is the observational p-dimensional vector;

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{2.2}$$

is called the gain matrix with n x p dimension; $\mathbf{R}$ and $\mathbf{B}$ are the observation and background error covariance matrices with p x p and n x n dimensions, respectively; $\mathbf{H}$ is the observation operator assumed here to be linear. $\mathbf{H}$ is the operator which projects the model state onto the observation state. It can correspond for instance to spatial interpolations from model grid points to observation points. In the case of

different purposes at the Portuguese Meteorological Service.

[3] Knowing the true model state we want to obtain the original set of observations that reproduced the atmosphere conditions.

satellite radiances, $\mathbf{H}$ contains a radiative transfer model; $(.)^T$ is the transpose of $(.)$.

The equation (2.1) is of particular interest along this work when it takes the form:

$$\delta\mathbf{x} = \mathbf{x}^a - \mathbf{x}^b = \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) = \mathbf{K}\mathbf{d} \tag{2.3}$$

where the difference $\delta\mathbf{x}$ is the n-dimensional vector called the analysis increment. In this equation, $\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^b)$ is a p-dimensional vector called the innovation.

It can be shown (Daley 1991) that the analysis increment, that is, the analysis correction to the background, can be seen as resulting from a two-step process of the innovation information, corresponding respectively to a filtering step and to a propagation step, if we write the previous equation under the form

$$\delta\mathbf{x} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{H}\mathbf{B}\mathbf{H}^T)(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}. \tag{2.4}$$

The factor responsable for the filtering is $(\mathbf{H}\mathbf{B}\mathbf{H}^T)(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$, which corresponds to the analysis increments produced at the observation points. This filter, that is proportional to $\mathbf{B}$, penalizes under a differential way the different components of the innovation vector; the components less penalized are those associated to large background error variances (e.g. Berre (2001)). Typically, if the background error variance is large compared to observation error variances in $\mathbf{R}$, the analysis increment will be also large, which means that the analysis will closely fit the observation values. The factor responsable for the propagation is $\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$, which implies that the analysis increments at the observation points are propagated (or interpolated under a muti-variate way) to the model grid points. Moreover we see that this propagation is proportional to $\mathbf{B}$. Typically, if spatial correlation functions of background errors specified in $\mathbf{B}$ are large-scale, analysis increments at the observation points will be propagated spatially to a large extent. We then verify that the information of $\mathbf{B}$ is used to filter and propagate spatially under a multi-variate way the meteorological information contained at the observation points.

2.3 3D-Var formulation

The variational method allows the BLUE estimate to be recovered for the above mentioned working hypoteses. In the formalism of the Three-Dimensional Variational Data Assimilation (3D-Var), the analysis $\mathbf{x}^a$ should minimize a quadratic function that measures the distance between the estimate and the available information, the observations and the background, taking into account their respective precisions. This function is called the cost function. It is represented by J. Under the assumed hypotheses it is written as (Le Dimet & Talagrand (1986) , Courtier *et al.* (1998)) [4]:

$$J = J_b + J_o = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x})) \tag{2.5}$$

where $\mathbf{x}$ is the estimate n-dimensional model state vector. And the observation operator is now noted $H(.)$, because it can be non linear here. J_b and J_o are the background and the observation terms of J, respectively.

Knowing that $\mathbf{B}$ and $\mathbf{R}$ are positive definite matrices [5] , the cost function is convex and its minimum is obtained when its derivative reaches zero (see for instance Bocquet (2009)). That is, when

$$\nabla_x J = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}^o - H(\mathbf{x})) = 0 \tag{2.6}$$

where $\mathbf{H}$ is the observation operator linearized around the background state, and $\mathbf{H}^T$ is the associated adjoint operator, which corresponds to the transpose of $\mathbf{H}$.

The vanishing of gradient (2.6) is obtained for $\mathbf{x} = \mathbf{x}^a$, and it can be shown that this corresponds to the BLUE equation of the analysis state $\mathbf{x}^a$:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}^b)) \tag{2.7}$$

[4] Along this work we will follow as much as possible the notation by Ide *et al.* (1997).

[5] A symmetric matrix $\mathbf{B}$ is defined to be positive definite if, for any vector $\mathbf{x} \neq 0$, the scalar $\mathbf{x}^T\mathbf{B}\mathbf{x} > 0$.

There is an immediate practical advantage of implementing 3D-Var in operations compared with previously used methods: the computation of $\mathbf{K}$, which obliges to the manipulation of big matrices or to an *a priori* data selection, is avoided since the solution is found globally for all data sources, though an iterative process [6] where several calculations of equations (2.5) and (2.6) are performed (Bouttier & Courtier 1999).

An incremental formulation of (2.5) is also possible, with the advantage of rendering the cost function a quadratic function of the increment which has a single minimum:

$$J = J_b + J_o = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{d} - \mathbf{H}\delta\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\delta\mathbf{x}) \tag{2.8}$$

where $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$ is the model state increment (which is a n-dimensional vector) so that at the minimum the resulting analysis increment $\delta\mathbf{x}^a$ added to the background $\mathbf{x}^b$ to provide the analysis $\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}^a$; and the remaining symbols keep their previous meaning. The gradient of this cost function is:

$$\nabla_{\delta x} J = \mathbf{B}^{-1}\delta\mathbf{x} - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\delta\mathbf{x}) \tag{2.9}$$

In its general formulation, the minimization by the incremental method is still an iterative procedure. However, in the incremental 3D-Var a fast convergence of the minization can be achieved by use of a change of variable such that $\chi = \mathbf{U}\delta\mathbf{x}$ where $\mathbf{U}$ is defined so that $\mathbf{B}^{-1} = \mathbf{U}^T\mathbf{U}$ (Cholesky transformation).

The expression of the background cost function J_b, in particular, can be expressed as (we refer here Derber & Bouttier (1999) and Gustafsson *et al.* (2001) for a deeper comprehension on this topic):

$$J_b = \frac{1}{2}\chi^T\chi \tag{2.10}$$

In this way, in the space of the transformed analysis increment, χ, the background error covariance matrix becomes the identity matrix and the incremental formulation of the

[6] like those of the conjugate gradient or quasi-Newton methods

total cost function and its gradient are written as follows, respectively:

$$J = \frac{1}{2}\chi^T\chi + \frac{1}{2}(\mathbf{d} - \mathbf{H}\mathbf{U}^{-1}\chi)^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{U}^{-1}\chi) \tag{2.11}$$

$$\nabla_\chi J = \chi - \mathbf{U}^{-T}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{U}^{-1}\chi) \tag{2.12}$$

This formulation converges more quickly than (2.8), because the second derivative of J (i.e. its Hessian) is much closer to the identity matrix, which means that the Hessian is better conditioned. As a final remark we should notice that at the initial iteration, $\mathbf{x} = \mathbf{x}^b$ so $\delta\mathbf{x} = \chi = 0$ and that at the final iteration the analysis increments are given by $\delta\mathbf{x} = \mathbf{U}^{-1}\chi$.

While this 3D-Var approach has been introduced firstly for global models (e.g. Courtier *et al.* (1998)), it has been used more recently for limited area models such as HIRLAM (Gustafsson *et al.* 2001) and ALADIN/France (Fischer *et al.* 2005). It is the ALADIN/France 3D-Var system which will be used in our study.

2.4 Data assimilation cycling and other analysis methods

Figure 2.1 is an illustration of the principle of data assimilation cycling. At time t_1, a 6 hour data assimilation is done by combining a set of observations valid in the ± 3 h time interval, centered at the analysis time, with a background corresponding to a 6 h forecast issued from the previous analysis valid at time t_0. Once the analysis $\mathbf{x}^a_{t_1}$ is calculated, a 6 h forecast can be run to provide a background $\mathbf{x}^f = M(\mathbf{x}^a_{t_1})$ valid at time t_2, to be used for the calculation of the analysis valid at time t_2.

The two fundamental equations describing a given analysis/forecast step are thus (when considering the case where the observation operator is linear):

$$\mathbf{x}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{x}^b + \mathbf{K}\mathbf{y}^o \tag{2.13}$$

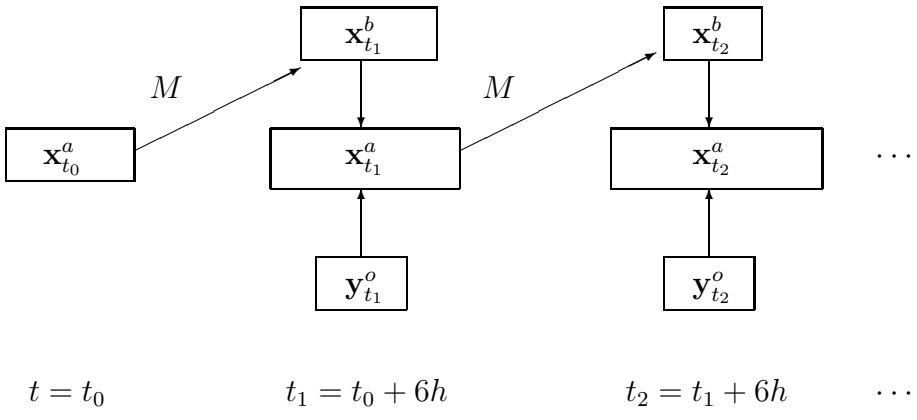


Figure 2.1: Schematic illustration of the data assimilation cycling: $\mathbf{y}^o_{t_i+1}$ and $\mathbf{x}^a_{t_i+1}$ represent, respectively, the observation vector and the analysis field at instant $t_{i+1} = t_i + 6h$; M represents the operator that corresponds to the 6 h integration provided by the forecast model.

$$\mathbf{x}^f = M(\mathbf{x}^a) \tag{2.14}$$

where M is the non-linear forecast model. As will be discussed in the next chapter, these two equations of analysis and forecast states lead to similar equations for the analysis and forecast errors, respectively.

Although this will not be discussed extensively in the present manuscript, there are other analysis techniques than 3D-var in order to calculate the BLUE analysis estimation. One of these techniques corresponds to four-dimensional variational assimilation (4D-Var), which is a generalization of 3D-Var that takes better account of the temporal dimension of the data assimilation problem, by including the forecast model M as a part of the observation operator (e.g Le Dimet & Talagrand (1986), and Rabier *et al.* (1998b)). Other analysis techniques correspond to Ensemble Kalman filter approaches, whose different variants are reviewed in Evensen (2003) for instance. Ensemble Kalman filters are not based on the minimization of a cost function (in contrast with variational approaches), but they rely more explicitly on the gain matrix formula (equation (2.2)), which is estimated by using ensemble forecast perturbations. Nowadays, these two

different approaches tend to converge towards each other, in the sense that ensemble forecast perturbations can also be used to calibrate and model the $\mathbf{B}$ matrix used in variational assimilation.

Chapter 3

Background error covariance modelling and calibration

Background error covariance specification relies on two main ingredients. A first step is to use techniques which simulate the error evolution during the successive analysis/forecast steps of the assimilation cycling. A second step is to use these simulated error data to calibrate a sparse covariance model.

3.1 Error simulation techniques for calibration of error covariances

3.1.1 Error equations and their simulation

In data assimilation, background error covariance estimates have the crucial role of filtering and propagating the observed information from the observation locations to the grid points of the atmospheric model. However, background error covariances can only be estimated approximately because, in particular, the true atmospheric model state is never exactly known.

The background error $\mathbf{e}^b$ is defined as the difference between the background state $\mathbf{x}^b$ and the true state $\mathbf{x}^t$:

$$\mathbf{e}^b = \mathbf{x}^b - \mathbf{x}^t \tag{3.1}$$

There are two fundamental equations which govern the evolution of analysis and background errors. We will assume that $H(.)$ and $M(.)$ are linear for the sake of simplicity (see also e.g Desroziers *et al.* (2009) for the weakly non-linear case), i.e. $H(.) = \mathbf{H}$ and $M(.) = \mathbf{M}$. It may be noticed that (2.7) can be re-written as $\mathbf{x}^a = (\mathbf{I} - \mathbf{KH})\mathbf{x}^b + \mathbf{K}\mathbf{y}^o$. Moreover, the true state $\mathbf{x}^t$ can also be written as $\mathbf{x}^t = (\mathbf{I} - \mathbf{KH})\mathbf{x}^t + \mathbf{KH}\mathbf{x}^t$. This implies that the equation of the analysis error $\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}^t$ is:

$$\mathbf{e}^a = (\mathbf{I} - \mathbf{KH})\mathbf{e}^b + \mathbf{K}\mathbf{e}^o \tag{3.2}$$

where $\mathbf{e}^o$ is the observation error and $\mathbf{I}$ is the identity matrix and at the same time, $\mathbf{d} = \mathbf{e}^o - \mathbf{H}\mathbf{e}^f$. The error $\mathbf{e}^f$ of the forecast $\mathbf{x}^f$ launched from the analysis $\mathbf{x}^a$ is given by:

$$\mathbf{e}^f = \mathbf{M}\mathbf{e}^a + \mathbf{e}^m \tag{3.3}$$

where $\mathbf{e}^m$ is the 6 hour accumulated model error. Note that the six-hour forecast $\mathbf{x}^f$ is used as a background $\mathbf{x}^b$ for the next analysis cycle so that $\mathbf{e}^b$ becames equal to $\mathbf{e}^f$.

As outlined by Fisher (2003), there are two basic ways by which we can try to calibrate background error covariance models: either by use of the available observed information (innovation statistics), or by using error simulation techniques.

An example of the first approach is the observational method, used for instance by Hollingsworth & Lönnberg (1986). They have used the innovation statistics over a data-dense radiosonde network (over the North of America) to diagnose the statistics of background errors, assuming that observation errors are uncorrelated with background errors and spatially uncorrelated. It can be shown (e.g. Berre & Desroziers (2010)) that this amounts to applying a low-pass filter $\mathcal{F}$ to estimate $\mathbf{HBH}^T$ from innovation covariances, namely: $\overline{\mathbf{d}\mathbf{d}^T} \simeq \overline{(\mathbf{e}^o - \mathbf{H}\mathbf{e}^b)} \ \ \overline{(\mathbf{e}^o - \mathbf{H}\mathbf{e}^b)^T} \simeq \mathbf{HBH}^T + \mathbf{R}$; that is:

$$\mathbf{HBH}^T \simeq \mathcal{F}(\mathbf{HBH}^T + \mathbf{R}) \tag{3.4}$$

One main disadvantage of the observational method is that these estimates are not representative of areas with a low data-density observation network; besides, the estimates can only be found for observed quantities and not directly for the model variables.

The second approach relies most often on the generation of forecast difference fields (as a surrogate of error fields) on the model grid for the model variables. An immediate advantage of such an approach is that the error statistics will be available on the whole model domain; a difficulty corresponds to the fact that these quantities can differ in some aspects from the statistics of background error. The most known example of this approach is the NMC method (Parrish & Derber 1992). In this method, the surrogate quantity whose statitics are used to simulate the background errors is the difference between two forecasts valid at the same analysis time, but for different forecast ranges (such as 12 and 36 hours, typically). One of the main ingredients of this NMC method is the accumulation of 4 sucessive analysis increments during the 24 hour differing period. Such an approach can be seen as relying on the analysis increment equation to simulate the analysis error equation (3.2) (after noticing that $\mathbf{y}^o - H(\mathbf{x}^b) \simeq \mathbf{e}^o - \mathbf{He}^b$):

$$\mathbf{e}^a \simeq \delta\mathbf{x} = -\mathbf{KHe}^b + \mathbf{Ke}^o \tag{3.5}$$

which indicates that the operator $(\mathbf{I} - \mathbf{KH})$ applied to $\mathbf{e}^b$ in (3.2) is replaced by the operator $-\mathbf{KH}$. This analysis increment is then evolved by the forecast model $\mathbf{M}$, to approximate (3.3):

$$\mathbf{e}^b \simeq \mathbf{M}\delta\mathbf{x} \tag{3.6}$$

by assuming that $\mathbf{e}^a \simeq \delta\mathbf{x}$ and that $\mathbf{e}^m \simeq 0$.

This is not an expensive method, which justifies its popularity. However, it has two conceptual drawbacks in particular. Firstly, the error variance will be under-estimated in poorly-observed regions because $\mathbf{K} \simeq \mathbf{KH} \simeq 0$ and thus $\delta\mathbf{x} \simeq 0$, whereas $\mathbf{e}^a \simeq \mathbf{e}^b \gg 0$

in these regions, i.e. $\mathbf{e}^a \gg \delta\mathbf{x}$ (see e.g. Belo Pereira & Berre (2006)). Secondly, since the forecast length of the two forecasts used to compute the differences is longer than the forecast range used as background (typically a 6 h forecast), the error spatial structures are likely to be broader in the vertical and in the horizontal than those of the exact background errors (Fisher 2003).

Another methodology was proposed by Fisher (2003), based on a conceptualization by Houtekamer *et al.* (1996). In this approach, called analysis-ensemble method, an ensemble of perturbed analysis/forecat steps is calculated, by using observation and background perturbations that are consistent with associated errors. In common with the NMC method, this has the advantage of producing global fields of model variables over the model grid. Moreover, and contrarily to the NMC method, it produces reasonable estimates of the background error in data-sparse regions in particular.

As shown in Berre *et al.* (2006), for instance, this is related to the representation of error equations (3.2) and (3.3) through corresponding equations for the analysis perturbations ε^a and for the forecast perturbations ε^f (where perturbations are constructed as random differences from the corresponding unperturbed (real) data):

$$\varepsilon^a = (\mathbf{I} - \mathbf{KH})\varepsilon^b + \mathbf{K}\varepsilon^o \tag{3.7}$$

$$\varepsilon^f = \mathbf{M}\varepsilon^a + \varepsilon^m \tag{3.8}$$

where ε^o are observation perturbations which are randomly drawn from the specified observation error covariance matrix $\mathbf{R}$, while ε^a, ε^b and ε^f are perturbations implicitly generated by the cycled application of equations (3.7) and (3.8). ε^m are model perturbations which are either neglected ($\varepsilon^m \simeq 0$), or drawn from an estimate of the model error covariance matrix (often denoted by $\mathbf{Q}$). Because model error covariances are poorly known in practice, the perfect-model assumption (i.e. $\varepsilon^m \simeq 0$) is often used in such ensemble analysis simulations. In order to correct the associated underestimation of background error variances, *a posteriori* diagnostics (see e.g. Desroziers *et al.*

(2005)) are usually applied to inflate background error standard deviation estimates by a factor 2 typically. This is the approach which will be used in our study, while model error representation is also an area of active research (e.g. Raynaud *et al.* (2011)).

We should note furthermore that in the ensemble technique used in this work we are trying to get a new and better representation of $\mathbf{B}$ by calibrating it using the perturbation of successive analyses and forecasts obtained from a reference system that uses the original, thought distinct, $\mathbf{B}$ matrix. This mechanism is an approximation that introduces imbalances and inconsistencies in the assimilation algorithm, and we say that the reference $\mathbf{B}$ matrix is used under a sub-optimal way. However, as will be seen, this procedure is still a reasonable first correction to the algorithm and is a commonly used technique in other contexts as well.

3.1.2 Implementation of an ensemble of perturbed ALADIN analysis and forecasts

In this section, we will describe more precisely the practical implementation of the ALADIN ensemble assimilation experiments. The Regional Ensemble Data Assimilation (REnDA) [1] used for the investigation here described was set up as a 6-perturbed member ensemble of the actual reference 3D-Var ALADIN/France data assimilation scheme. Figure 3.1 represents the first two 6 hour networks of its cycling for a generic member $[k], k = 1, \cdots, 6$.

For the very first date, the background is the same for all the members; it is an unperturbed 6 h ALADIN/France forecast. During the next data assimilation cycles, a 6 h model integration from the actual perturbed analysis - an implicitly perturbed background - is used for the new analysis step. For each REnDA member, the perturbed analysis is obtained by combination of the perturbed background with randomly perturbed observations. These observation pertubations have the same statistical characteristics as the prescribed observation error since they are randomly drawn from the specified observation error covariance matrix $\mathbf{R}$. Lateral boundary conditions for the 6

[1] by analogy with the abbreviation used at Isaken *et al.* (2007)

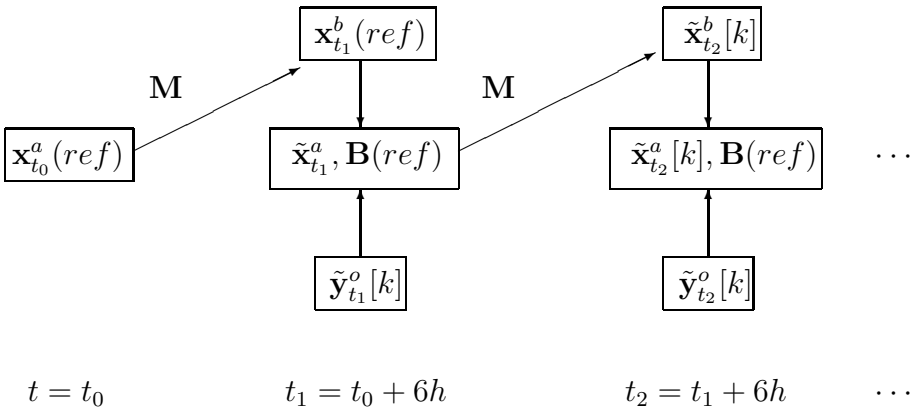


Figure 3.1: Schematic illustration of the data assimilation member $[k]$ assimilation cycling networks which are replicated by the 6 members of the REnDA: $\mathbf{y}^o_{t_i+1}$ and $\mathbf{x}^a_{t_i+1}$ represent, respectively, the observation vector and the analysis field at instant $t_{i+1} = t_i + 6h$; $\mathbf{M}$ represents the (linear) operator that corresponds to the 6 h integration provided by the forecast model, that is the background $\mathbf{x}_b$ on the next assimilation cycle (or network). The symbol $\tilde{(.)}$ represents a perturbed quantity, ref refers the operational reference system.

hour ALADIN/France integration are ARPEGE forecasts starting from a corresponding ARPEGE ensemble data set.

The reference background statistics, represented by $B(ref)$ in Figure 3.1, is used in every 6-hour assimilation cycle. It is a climatological ensemble-based background error covariance matrix, calculated over a three-week autumnal period, from the 9th of September to the 4th of October 2007.

Let us consider the model state evolution of an ensemble member which will be referred by the index k (see Figure 3.1) and compare it with the evolution of the unperturbed system. In a first step, at the analysis time t_i, the difference between the perturbed and unperturbed analysis is equal to $\varepsilon^a_i = \tilde{\mathbf{x}}^a_i[k] - \mathbf{x}^a_i$. Starting from a perturbed analysis $\tilde{\mathbf{x}}^a_i[k]$, we can produce a 6 h forecast, considering $\tilde{\mathbf{x}}^b_{i+1}[k] = \mathbf{M}\tilde{\mathbf{x}}^a_i[k]$, whose difference

with the unperturbed forecast $\mathbf{x}^b_{i+1} = \mathbf{M}\mathbf{x}^a_i$ is

$$\varepsilon^b_{i+1} = \tilde{\mathbf{x}}^b_{i+1}[k] - \mathbf{x}^b_{i+1} = \mathbf{M}\tilde{\mathbf{x}}^a_i[k] - \mathbf{M}\mathbf{x}^a_i = \mathbf{M}\varepsilon^a_{i+1} \tag{3.9}$$

In a second step, at the analysis time t_{i+1}, we can combine the 6 h forecast of member $[k]$ with its specific set of (independently perturbed) observations, $\tilde{\mathbf{y}}^o_{i+1}[k] = \mathbf{y}^o_{i+1} + \varepsilon^o_{i+1}[k]$, where $\varepsilon^o_{i+1}[k] \sim \mathcal{N}(0, \mathbf{R})$ are Gaussian random draws of $\mathbf{R}$, and produce a new perturbed analysis field:

$$\tilde{\mathbf{x}}^a_{i+1}[k] = \tilde{\mathbf{x}}^b_{i+1}[k] + \mathbf{K}(\tilde{\mathbf{y}}^o_{i+1}[k] - H(\tilde{\mathbf{x}}^b_{i+1}[k])) \tag{3.10}$$

whereas the equation of the unperturbed analysis is:

$$\mathbf{x}^a_{i+1} = \mathbf{x}^b_{i+1} + \mathbf{K}(\mathbf{y}^o_{i+1} - H(\mathbf{x}^b_{i+1}))$$

The difference between the two last analysis equations leads us to the following expression for the analysis perturbation at time t_{i+1}:

$$\varepsilon^a_{i+1} = \varepsilon^b_{i+1} + \mathbf{K}(\varepsilon^o_{i+1} - \mathbf{H}\varepsilon^b_{i+1}) \tag{3.11}$$

Thus, equations such as (3.9) and (3.10) illustrate the idea that ensemble perturbations are obtained by running perturbed data assimilation cycles, based on explicitly perturbed observations and implicit background perturbations.

3.2 Background error covariance modelling in ALADIN

Once these simulated error data have been produced, they are usually not employed to compare directly a full covariance matrix $\mathbf{B}$ as such, because the size of $\mathbf{B}$ is much too large to be stored in memory. On the other hand, these simulated error data can be used to calibrate a sparse covariance model of $\mathbf{B}$.

The analysis component of the ALADIN/France data assimilation system is based on the 3D-Var method, and it takes advantage of scientific achievements of the

ARPEGE/IFS system (Fischer *et al.* 2005). As explained in section 2.1, it relies on the incremental approach. Besides, a preconditioning of the variational problem is applied such that the minimization of the cost function is done through the change of variable that makes its Hessian closer to the identity matrix.

In the 3D-Var system of ALADIN/France, the background-error covariance model formulation is detailed in Berre (2000), as a bi-Fourier counterpart (to a limited area) of the global, spherical harmonic formalism proposed by Courtier *et al.* (1998) and Derber & Bouttier (1999). The original (global) multivariate formulation, based on the model variables of vorticity, divergence, temperature and the (logarithm of) surface pressure, is extended also to humidity. As detailed in Gustafsson *et al.* (2001) for instance, the change of variable $\mathbf{U} = \mathbf{B}^{-1/2}$ presented in section 2.3 can be designed as an operator which transforms the background error $\mathbf{e}^b$ into a variable $\mathbf{e}^{b'} = \mathbf{U}\mathbf{e}^b$ whose covariance matrix is the identity matrix:

$$\overline{\mathbf{e}^{b'}(\mathbf{e}^{b'})^T} = \overline{\mathbf{U}\mathbf{e}^b(\mathbf{U}\mathbf{e}^b)^T} = \mathbf{U}\overline{\mathbf{e}^b(\mathbf{e}^b)^T}\mathbf{U}^T = \mathbf{B}^{-1/2}\mathbf{B}\mathbf{B}^{-T/2} = \mathbf{I} \tag{3.12}$$

where $\overline{(.)}$ represents the ensemble average over a number of identic independent cases.

3.2.1 Cross-covariances and multivariate regressions

The model state increment vector $\delta\mathbf{x}$ is written as the block vector

$$\delta\mathbf{x} = \begin{pmatrix} \zeta \\ \eta \\ (\mathbf{T}, \mathbf{P}_s) \\ \mathbf{q} \end{pmatrix}$$

where the increment representation δ is dropped from the notation for the sake of simplicity at the different increment fields that are just mentioned by using its common physical names, so that ζ is called vorticity; η is divergence; $(\mathbf{T}, \mathbf{P}_s)$ corresponds to the temperature and the logarithm of surface pressure, and $\mathbf{q}$ is the specific humidity.

A first step to achieve the appropriate conditioning for the $\mathbf{B}$ matrix in ALADIN/France is to transform the background error of the full [2] model variables (ζ,η,($\mathbf{T}, \mathbf{P}_s$),$\mathbf{q}$) into a set of variables (ζ,η_u,($\mathbf{T}, \mathbf{P}_s$)$_u$,$\mathbf{q}_u$) which are mutually uncorrelated. This can be done by using linear regression operators in order to extract cross-covariances between the full model variables (ζ,η,($\mathbf{T}, \mathbf{P}_s$),$\mathbf{q}$):

$$\begin{aligned} \zeta &= \zeta \\ \eta &= \mathsf{M}\zeta + \eta_u \\ (T, P_s) &= \mathsf{N}\zeta + \mathsf{P}\eta_u + (\mathbf{T}, \mathbf{P}_s)_u \\ q &= \mathsf{Q}\zeta + \mathsf{R}\eta_u + \mathsf{S}(\mathbf{T}, \mathbf{P}_s)_u + \mathbf{q}_u \end{aligned} \tag{3.13}$$

where $\mathsf{M}, \mathsf{N}, \mathsf{P}, \mathsf{Q}, \mathsf{R}$ and S are linear regression matrices which are deduced from cross-covariances and auto-covariances, as shown for M as an example:

$$\mathsf{M} = \mathbf{C}_{\zeta,\eta}\mathbf{C}^{-1}_{\zeta,\zeta} \tag{3.14}$$

where $\mathbf{C}_{\zeta,\eta}$ is the cross-covariance matrix between vorticity and divergence, and $\mathbf{C}_{\zeta,\zeta}$ is the auto-covariance matrix of vorticity.

These regression operators are often called balance operators, because they represent couplings such as geostrophic mass/wind balance in particular. The regression residuals are often named unbalanced variables, because they can be seen as residual terms of balance equations. Moreover, it can be shown easily that the resulting regression residuals are uncorrelated with vorticity, and that they are also mutually uncorrelated:

$$\begin{aligned} \overline{\eta_u \zeta^T} &= \overline{(\eta - \mathsf{M}\zeta)\zeta^T} \\ &= \overline{\eta\zeta^T} - \mathsf{M}\overline{\zeta\zeta^T} \\ &= \overline{\eta\zeta^T} - \overline{\eta\zeta^T}\ \overline{(\zeta\zeta^T)^{-1}}\ \overline{\zeta\zeta^T} \end{aligned}$$

[2] which correspond to the addition of the physically balanced and the unbalanced parts of the model fields increments

$$
\begin{aligned}
&= \overline{\eta\zeta^T} - \overline{\eta}\,\overline{\zeta^T} \\
&= 0
\end{aligned} \tag{3.15}
$$

The balance operators are calculated in spectral bi-Fourier space in order to represent the scale dependence of these multivariate couplings, with for instance a lesser degree of geostrophy for small horizontal scales (e.g.(Derber & Bouttier 1999), (Berre 2000)). Cross-covariances between different wave vectors are neglected, which corresponds to an assumption that cross-covariances are horizontally homogeneous.

The mutual uncorrelation between vorticity and regression residuals implies that the background error covariance matrix $\mathbf{B}_u$ in the space of (ζ,η_u,($\mathbf{T}, \mathbf{P}_s$)$_u$,$\mathbf{q}_u$) is block-diagonal:

$$
\mathbf{B}_u = \begin{pmatrix}
\mathbf{C}_\zeta & 0 & 0 & 0 \\
0 & \mathbf{C}_{\eta_u} & 0 & 0 \\
0 & 0 & \mathbf{C}_{(T,p_s)_u} & 0 \\
0 & 0 & 0 & \mathbf{C}_{q_u}
\end{pmatrix}
$$

where e.g. $\mathbf{C}_\zeta$ represents now the auto-covariance matrix of ζ (previously represented as $\mathbf{C}_{\zeta,\zeta}$) for the sake of simplicity.

Moreover, the regression equations presented in equation (3.13) can be written in matricial form as corresponding to the application of a balance operator $\mathcal{K}$:

$$
\mathcal{K} = \begin{pmatrix}
I & 0 & 0 & 0 \\
\mathsf{M} & I & 0 & 0 \\
\mathsf{N} & \mathsf{P} & I & 0 \\
\mathsf{Q} & \mathsf{R} & \mathsf{S} & I
\end{pmatrix}
$$

This implies that the background error covariance matrix $\mathbf{B}$ in the space of the full model variables (ζ,η,($\mathbf{T}, \mathbf{P}_s$),$\mathbf{q}$) is written as follows:

$$
\mathbf{B} = \mathcal{K}\mathbf{B}_u\mathcal{K}^T = \mathbf{U}^{-1}(\mathbf{U}^{-1})^T \tag{3.16}
$$

where $(.)^T$ denotes the transpose of the matrix (.), and $\mathbf{U}^{-1} = \mathcal{K}\mathbf{B}_u^{1/2}$.

These last three equations illustrate the fact that $\mathbf{B}$ is modelled as the product of sparse operators. This is crucial because otherwise the size of the full covariance matrix $\mathbf{B}$ is so huge (it is equal to the square of the model dimension) that it can not be stored and handled explicitly.

3.2.2 Horizontal and vertical auto-covariances

Spatial auto-covariances contained in $\mathbf{B}_u$ are modelled using an assumption of horizontal homogeneity and isotropy. This assumption allows the matrix $\mathbf{B}_u$ to be represented as a block-diagonal matrix in bi-Fourier space, with zero covariances between different wave vectors (e.g. Berre (2000)).

On the other hand, auto-covariances in $\mathbf{B}_u$ are allowed to vary with height and horizontal scale. This enables the representation of the increased broadness of horizontal correlation functions with height and of the increased sharpness of vertical correlations for small horizontal scales (e.g. Rabier *et al.* (1998a)).

These different aspects of spatial auto-covariances can be studied by using the following decomposition of the covariance between two spectral coefficients at levels z and z', respectively:

$$\overline{q_{mn}^{z} q_{mn}^{z'}{}^{*}} = \sigma_z \sigma_{z'} \sqrt{\gamma_z(k^*)\gamma_{z'}(k^*)}\, r_{k^*}(z, z') \tag{3.17}$$

where q_{mn}^{z} and $q_{mn}^{z'}$ result from the bi-Fourier expansion; m and n are wave numbers in the zonal and meridional direction respectively; σ_z, σ'_z are the averaged standard-deviations at levels z, z'; γ_z, γ'_z are the normalized spectral densities of the variance at those levels; and $r_{k^*}(z, z')$ is the vertical correlation between levels z, z' for the total horizontal wave number $k^* = L_y \sqrt{(\frac{m}{L_x})^2 + (\frac{n}{L_y})^2}$ (where L_x and L_y are the zonal and meridional domain lengths, respectively). $\overline{(.)}$ is the ensemble average. We should note here that since $q(\mathbf{h}, z)$ values are real numbers, q_{mn}^{z} is a complex number such that

$q^z_{-m-n} = {q^z_{mn}}^*$, where $(.)^*$ denotes the conjugate operator.

3.3 The issue of temporal variations in the current homogeneous model

The current homogeneous covariance model implies that horizontal variations of covariances are not represented. Although this will not be explored in our studies, it should be mentioned that investigations are also ongoing in order to represent horizontal heteorogeneities in alternative sparse covariance models. For instance, horizontal variations of background error standard deviations can be represented in grid point space, possibly after applying spatial filtering techniques to reduce sampling noise effects (e.g Raynaud *et al.* (2009)). Moreover, horizontal variations of background error correlations can be represented in wavelet space (e.g Fisher (2003), Deckmyn & Berre (2005)) for instance. Correlation heterogeneities can also be modelled either by using recursive spatial filters (Purser *et al.* 2003) or diffusion techniques (Weaver & Courtier 2001), or by including ensemble perturbations as control variables of the variational minimization (e.g. Lorenc (2003)).

Regarding the current homogeneous covariance model framework, it can be shown that the application of the horizontal homogeneity assumption amounts to calculating horizontal averages of covariances over the regional domain of interest (e.g Berre (2000)). In addition to this spatial average, background error covariances are usually temporally averaged over a few-week past period (see e.g. Fisher (2003) and Belo Pereira & Berre (2006)), during which an ensemble of perturbed assimilation cycles has been run off-line.

Once these spatially and temporally averaged covariances have been calculated, they are usually specified to be static in the covariance model of 3D-Var. This implies that potential temporal variations of background error covariances, such as those related to the weather situation dependence, are not represented in the 3D-Var system in this case.

Note that this issue of temporal variations can already be raised in the current homogeneous covariance model framework. This is particularly the case for the regionally averaged covariances used in limited area models such as ALADIN/France. This is related to the fact that a regional spatial averaged is expected to be more prone to temporal variations than a global spatial average.

The goal of the present work is thus to diagnose the time variability of regionally-average covariances (which corresponds to chapter 4), and also to examine their impact in the ALADIN/France 3D-Var system (which corresponds to 5).

Chapter 4

A diagnostic study of time variations of regionally averaged background error covariances

This chapter corresponds to the follwoing publication, with the addition of a few notes:

Maria Monteiro[1] and Loïk Berre[2] (2010): A diagnostic study of time variations of regionally averaged background error covariances, *Journal of Geophysical Research*, **115**, D23203, doi:10.1029/2010JD014095

[1]CPPN, Instituto de Meteorologia I.P., Lisbon, Portugal
[2]CNRM/GMAP, Météo-France, Toulouse, France

Abstract

In variational data assimilation systems, background error covariances are often estimated from a temporal and spatial average. For a limited area model such as the Aire Limited Adaptation Dynamique Développment InterNational (ALADIN)/France, the spatial average is calculated over the regional computation domain, which covers western Europe. The purpose of this study is to revise the temporal stationarity assumption by diagnosing time variations of such regionally averaged covariances. This is done through examination of covariance changes as a function of season (winter versus summer), day (in connection with the synoptic situation), and hour (related to the diurnal cycle), with the ALADIN/France regional ensemble Three-Dimensional Variational analysis (3D-Var) system. In summer, compared to winter, average error variances are larger, and spatial correlation functions are sharper horizontally but broader vertically. Daily changes in covariances are particularly strong during the winter period, with larger variances and smaller-scale error structures when an unstable low-pressure system is present in the regional domain. Diurnal variations are also significant in the boundary layer in particular, and, as expected, they tend to be more pronounced in summer. Moreover, the comparison between estimates provided by two independent ensembles indicates that these covariance time variations are estimated in a robust way from a 6-member ensemble. All these results support the idea of representing these time variations by using a real-time ensemble assimilation system.

4.1 Introduction

The analysis of meteorological fields, obtained by data assimilation systems, can be seen as a weighted combination of observations and of a model forecast (called the background). The weights are determined by specified spatial covariances of observation errors and of background errors. As discussed by Daley (1991), background error covariances are used to filter and propagate spatially observed information.

In the regional operational Aire Limited Adaptation Dynamique Développment InterNational (ALADIN)/France system, a variational scheme (Fischer *et al.* 2005) is used for data assimilation. Background error covariances are estimated by using an ensemble assimilation approach (Berre *et al.* 2006) and are calculated from a temporal and spatial average. In this ALADIN/France system, the spatial average is calculated over the regional computation domain, which covers western Europe. The associated homogeneity assumption implies that 3-D spatial covariances can be calculated and represented as sparse block-diagonal matrices in spectral space (with one block for each wave vector). The temporal average is usually calculated off-line over a period of several consecutive weeks (Belo Pereira & Berre 2006, e.g.). A stationarity assumption is thus introduced when using these time-averaged covariances in the assimilation scheme. This means that time variations corresponding to changes related to the season, the synoptic situation, and the diurnal cycle are not represented, whereas they are likely to be significant.

Therefore, current research efforts are devoted to the relaxation of these homogeneity and stationarity assumptions. With respect to the first assumption (i.e. homogeneity), a wavelet representation of background error correlations is being developed (Deckmyn & Berre (2005); Lindskog *et al.* (2007); see also http://www.ecmwf.int/publications/library/do/references/list/14092007), to represent associated spatial variations. The goal of the present paper is to revise the other hypothesis (i.e. stationarity). As a first step, this will be examined in the current operational homogeneous framework. Later on, once a heterogeneous covariance formulation is available, it is planned to extend the study to time variations in the wavelet framework (as evoked in Pannekoucke *et al.* (2008) and Berre *et al.* (2009) in a spherical wavelet context).

Compared with a global homogeneous framework, in which covariances are averaged over the whole globe, diagnosing time variations is likely to be more relevant in the current regional homogeneous context. This corresponds to the idea that time

variations are expected to be more pronounced in the regional average over western Europe than in the global average over the whole sphere.

The paper is structured as follows. The experimental framework is described in section 4.2. The seasonal variation of covariances is studied in section 4.3 by comparing covariances in winter and summer periods. Day-to-day covariance changes in the winter period, related to the synoptic situation, are examined in section 4.4. Section 4.5 shows results corresponding to diurnal cycle effects in the summer period. Conclusions and perspectives are summarized in section 4.6.

4.2 Experimental framework

4.2.1 Error simulation with the ALADIN/France 3D-Var system

The present study has been carried out in the context of the ALADIN/France Three-Dimensional Variational analysis (3D-Var) system, which is based on a 6 h assimilation period and which is described in Fischer *et al.* (2005). The ALADIN model is the limited area model (LAM) counterpart of the Action de Recherche Petite Echelle Grande Echelle/Integrated Forecasting System (ARPEGE/IFS) (Geleyn *et al.* 1995) global system (Horányi *et al.* 1996). The ALADIN/France model is spectral (based on a bi-Fourier representation of the fields), it has a 10 km horizontal resolution and 60 vertical levels, and its lateral boundary conditions are provided by the ARPEGE model. To avoid the propagation of spurious gravity waves, a digital filter initialization method (Lynch & Huang 1992) is applied after the 3D-Var analysis step. As illustrated in Figure 4.7, the ALADIN/France computation domain covers France, the Iberian Peninsula, and part of the surrounding countries. The vertical model levels are hybrid pressure terrain-following coordinates, and their approximate correspondence with standard pressure levels is shown in Figure 4.1.

ALADIN/France background errors are estimated here by using an error simulation

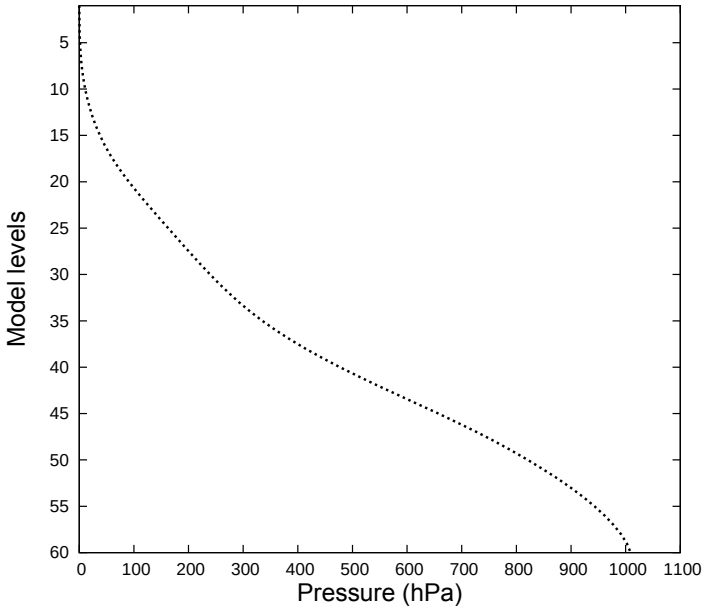


Figure 4.1: Standard pressure values of vertical levels in the ALADIN/France model.

technique as done, for example, by Houtekamer *et al.* (1996), Fisher (2003) and Belo Pereira & Berre (2006). The experimental framework is based on a regional ensemble 3D-Var system, made of six members. Each member can be seen as a perturbed ALADIN/France assimilation experiment. For each member, observations are explicitly perturbed in accordance with the specified observation error covariance matrix. Moreover, the background of a given member is implicitly perturbed through the previous analysis perturbations, evolved by the model during the 6 h forecast period. The combination of perturbed observations and of the perturbed background provides a new perturbed analysis, and so on. During the forecast step, each member is also coupled to a perturbed global run, provided by the ARPEGE ensemble assimilation system ((Berre *et al.* 2007, 2009, see also http://www.ecmwf.int/publications/library/do/references/list/14092007)), which has been operational at Météo-France since 2008. This coupling to a global ensemble allows the effect of lateral boundary errors to be represented in the regional ensemble. This perturbed assimilation cycling has been run over two monthly periods, one in winter (12 February to 13 March 2008) and the other in summer (3 July to 2 August 2008).

To evaluate the robustness of ensemble covariance estimates, a second set of 6-member

regional and global ensembles has been run independently during each of these two periods. The two winter ensemble data sets are denoted by EnW1 and EnW2, and the two summer ensemble data sets are denoted by EnS1 and EnS2. As discussed by Berre *et al.* (2007) in the context of standard deviation estimates (see their section 3), the difference between the two independent ensemble estimates is a measure of the level of sampling noise, induced by the finite size of the ensemble. This measure of sampling noise can be compared to estimated covariance values and to the amplitude of their temporal variations, to evaluate to which extent such ensemble estimates are robust.

4.2.2 Diagnostics of regionally averaged background error covariances

Covariance estimates have been calculated and diagnosed for the four main historical variables of the ALADIN/France model, namely vorticity, divergence, temperature and specific humidity. In particular, the horizontally averaged background error variance $\sigma_b(z)^2$ has been calculated at each vertical level z for these variables, together with the associated standard deviation $\sigma_b(z)$.

As shown in Berre (2000, section 3.a, and also equation (A3)), this horizontally averaged background error variance can also be written as the sum of variance contributions from different horizontal wave numbers $k^\star$:

$$\sigma_b(z)^2 = \sum_{k^\star=0}^{K^\star} p(k^\star, z). \tag{4.1}$$

where $k^\star$ is a two-dimensional total wave number defined by:

$$k^\star = L_d \sqrt{\left(\frac{m}{L_x}\right)^2 + \left(\frac{n}{L_y}\right)^2}$$

and where m and n are wave numbers in the zonal and meridional directions, respectively, while L_x and L_y are the corresponding domain lengths. L_d is an arbitrary scaling factor which defines the wavelength that corresponds to $k^\star = 1$. In the case of ALADIN/France, $L_d = L_y = 2780km$ is used and the maximum horizontal wave

number is $K^\star = 149$.

The contribution of $k^\star$ to the variance $\sigma_b(z)^2$ is $p(k^\star, z)$, and its formula is given by:

$$p(k^\star, z) = \frac{2\pi L_x L_y}{L_d^2} q(k^\star, z) k^\star,$$

where $q(k^\star, z)$ is the isotropic average of modal variances $\overline{|e^b_{mnz}|^2} = \overline{e^b_{mnz} e^{b*}_{mnz}}$ of background error spectral coefficients e^b_{mnz} associated to $k^\star$:

$$q(k^\star, z) = \frac{1}{2\pi} \int \overline{|e^b_{mnz}|^2} d\theta,$$

where (m, n) are wave vectors such that $L_d \sqrt{(\frac{m}{L_x})^2 + (\frac{n}{L_y})^2} = k^\star$, and $\theta = tan^{-1}(\frac{n/L_y}{m/L_x})$.

Modal variances $\overline{|e^b_{mnz}|^2}$ are calculated from the 6-member ensemble either for a given date (in the case of section 4.4), or from a temporal average over a monthly period (in the case of section 4.3). In the first case, the overbar corresponds to an ensemble average over the 6 members. In the second case, the overbar corresponds to an ensemble time average over the 6 member and over all days of the monthly period (corresponding to 180 realizations). Note that equation (4.1) can also be written as follows (see, e.g., the first equation of Remark 3 in the work of (Berre 2000, p. 664)): $\sigma_b(z)^2 = \sum_{m,n} \overline{|e^b_{mnz}|^2}$.

Equation (4.1) is used to evaluate the contribution of different horizontal scales to the horizontally averaged variance and to its temporal variations. This variance decomposition is also used to interprete changes in horizontal correlation length scales. As shown, for example, by Berre (2000), the horizontal correlation length scale $\mathcal{L}_z$, for each vertical level z, can be calculated from spectral variances as

$$\mathcal{L}_z = \frac{L_d}{2\pi} \sqrt{\frac{2 \sum_{m,n} \overline{|e^b_{mnz}|^2}}{\sum_{m,n} k^{\star 2} \, \overline{|e^b_{mnz}|^2}}} \tag{4.2}$$

This means that for a given vertical level, the length scale is inversely proportional to the variance spectrum times the squared wave number $k^\star$. In particular, when the increase in the variance spectrum occurs predominantly for large wave numbers, $\mathcal{L}_z$

decreases.

Another typical statistical quantity to be diagnosed corresponds to vertical correlations $\rho(z, z')$ between vertical levels z and z', which can be calculated as follows from associated vertical covariances $cov(z, z')$:

$$\rho(z, z') = \frac{cov(z, z')}{\sigma_b(z)\sigma_b(z')} = \frac{\sum_{m,n} \overline{e^b_{mnz} e^{b*}_{mnz'}}}{\sigma_b(z)\sigma_b(z')}$$

Both auto- and cross-covariances have been calculated and examined. It has been found that temporal variations in cross-covariances reflect changes affecting standard deviations to a large extent. Therefore, for the sake of conciseness, the present study focuses mostly on standard deviations and spatial auto-correlations.

4.3 Seasonal variation of covariances

4.3.1 Standard deviations and spectral decomposition of variance

Vertical profiles of standard deviations and associated variance spectra are examined in this section. They are shown in Figures 4.2 and 4.3, respectively, for 6 h forecasts valid at 1800 UTC. For a given season and model parameter, the curves obtained for the two independent ensembles are practically identical (not shown). This indicates that these average seasonal estimates are quite robust, as expected from the large temporal and spatial sample size. Therefore, all illustrations in this section are shown for the first ensemble of each season.

As shown in Figure 4.2, standard deviations are larger in summer than in winter. This is particularly noticeable for specific humidity (Figure 4.2a), which reflects the larger water content in summer than in winter, induced by the increased air temperature. For all variables, the increase of variance occurs mostly in the troposphere (below level 27 roughly, which is around 200 hPa (see Figure 4.1)), and for temperature in particular, it is more pronounced in the boundary layer. This is likely to correspond to the oc-

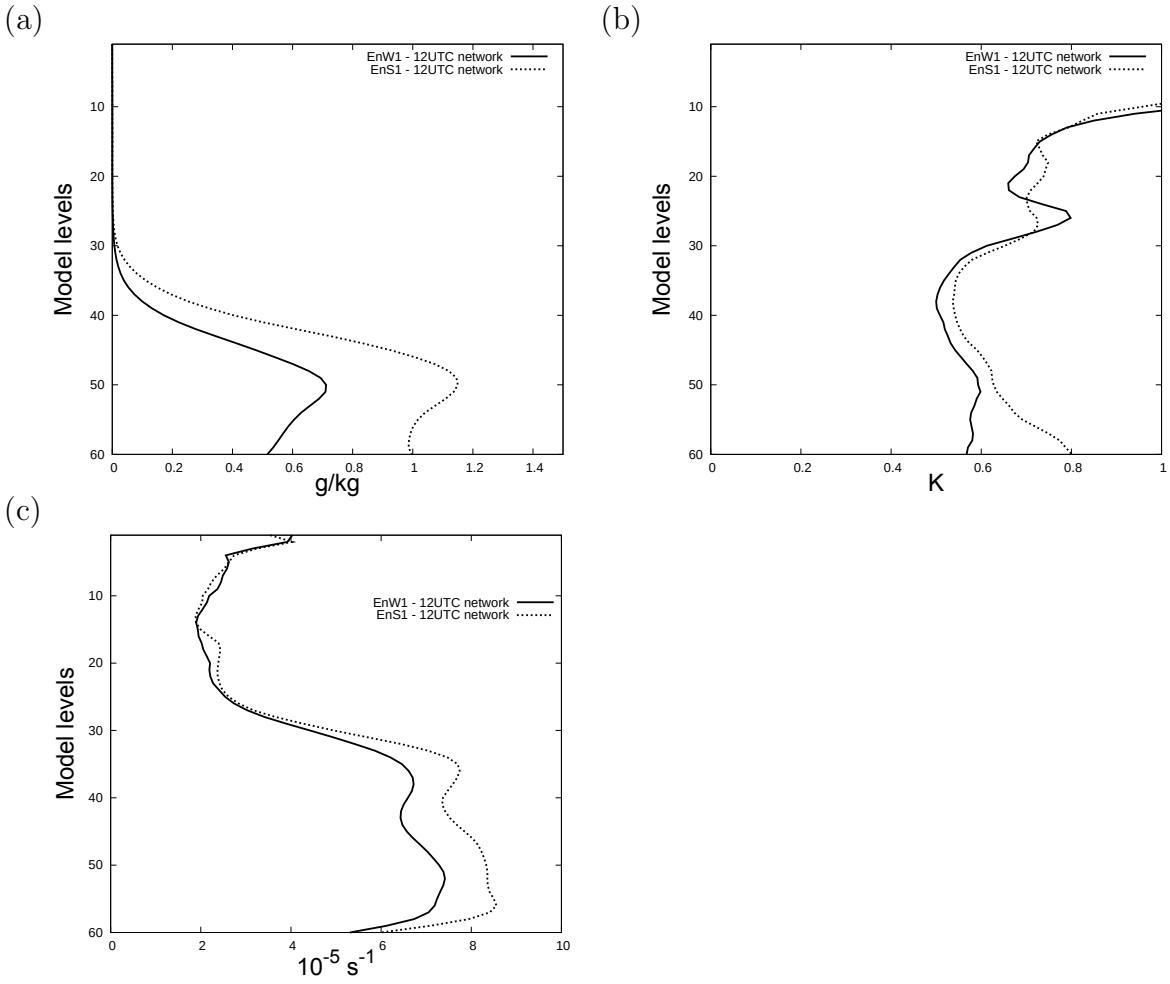


Figure 4.2: Vertical profiles of horizontally averaged standard deviations of ALADIN/France 6 h forecast errors (issued from the 1200 UTC network and valid at 1800 UTC), estimated by the first ensemble in winter (solid line) and in summer (dashed line), for (a) specific humidity, (b) temperature and (c) vorticity.

curence of small-scale convective events, which are more frequent in the summer period.

This interpretation is supported by variance spectra (Figure 4.3), which indicate, for temperature (Figure 4.3b) and vorticity (not shown), that the increase of variance is larger in the small-scale part of the spectrum (i.e. for wave numbers larger than 10 typically). For specific humidity (Figure 4.3a), the increase of variance is quite visible for all scales (and in particular for small and intermediate wave numbers). This corresponds to the increase of water vapour content in summer, which contributes to an amplitude increase of error structure at all acales.

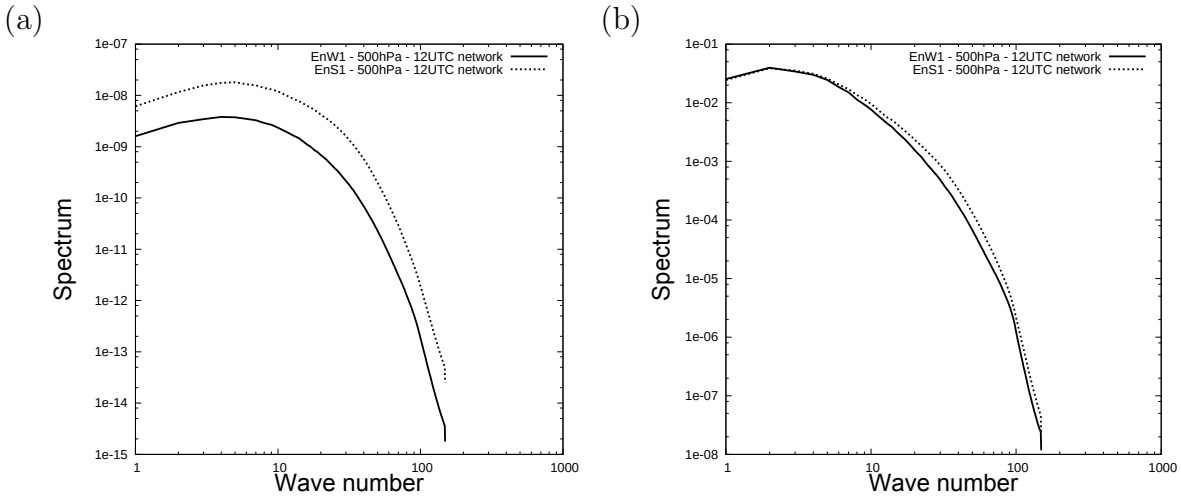


Figure 4.3: Horizontal variance spectra of ALADIN/France 6 h forecast errors (issued from the 1200 UTC network and valid at 1800 UTC) at 500hPa, estimated by the first ensemble in winter (solid line) and in summer (dashed line), for (a) specific humidity and (b) temperature.

These significant seasonal variations in a regional model are interesting to contrast with previous studies in global models. For instance, Derber & Bouttier (1999) made an extensive comparison between globally averaged covariances of five periods. As discussed near the end of their section 2, no significant seasonal differences were found, because the global spatial average does not distinguish between the Northern and Southern Hemispheres. Conversely, in the case of a LAM such as ALADIN/France, seasonal changes in the covariances are visible, as expected from seasonal variations affecting the regional domain.

4.3.2 Horizontal and vertical correlations

Figure 4.4 shows the vertical profiles of horizontal correlation length scales $\mathcal{L}_z$ of background errors (see equation (4.2)). It reflects the classical increase of length scale with height (e.g. Rabier *et al.* (1998a), Berre (2000)), which corresponds to the stronger contribution of large-scale phenomena in altitude. Moreover, tropospheric length scales tend to be smaller in summer than in winter, in accordance with increased small-scale variances found in Figure 4.3a. This is particularly marked for temperature

(Figures 4.3b and 4.4b) and specific humidity (Figures 4.3a and 4.4a), while this is somewhat less pronounced for vorticity and divergence (not shown). This means that horizontal correlation functions tend to be sharper (i.e. less broad) in summer than in winter.

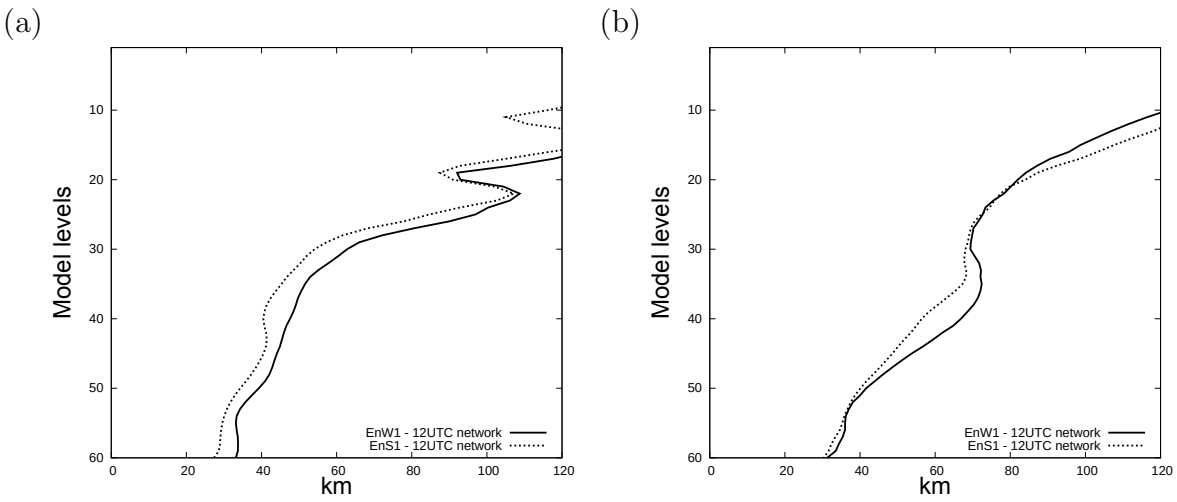


Figure 4.4: Vertical profiles of horizontal length scales of 6 h ALADIN/France forecast errors (issued from the 1200 UTC network and valid at 1800 UTC), estimated by the first ensemble in winter (solid line) and in summer (dashed line), for (a) specific humidity and (b) temperature. Length-scale values larger than 120 km (at high levels) are not shown.

Figure 4.5 show that the situation is different for vertical correlations. Seasonal changes are particularly pronounced for divergence, with broader negative lobes in summer than in winter. This is likely to reflect increased vertical couplings in summer convective situations. Similar changes are found for other variables (not shown).

Vertical cross covariances have been examined too (not shown). They reflect changes in the vertical profiles of variance to a large extent (see Figure 4.2). Vertical cross correlations remain relatively similar between the two seasons.

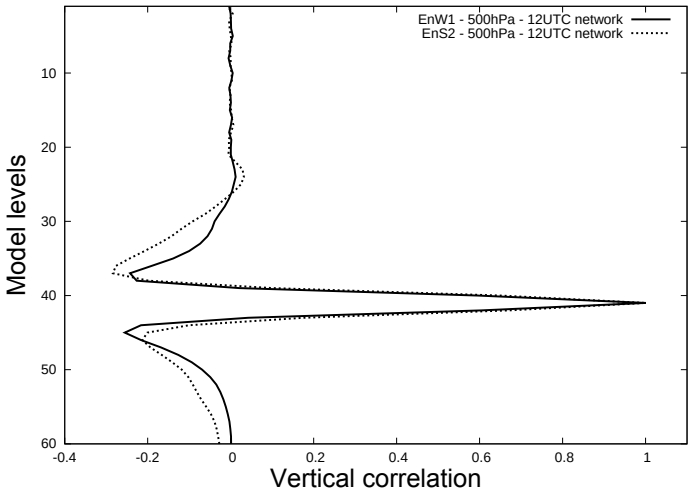


Figure 4.5: Vertical autocorrelations functions of ALADIN/France 6 h forecast errors at 500 hPa (issued at 1200 UTC network, and valid at 1800 UTC), estimated by the first ensemble in winter (solid line) and in summer (dashed line), for divergence.

4.4 Day-to-day variations of covariances (winter period)

After the previous study of seasonal variations, it is natural to examine daily variations of covariances in order to study the influence of the varying synoptic situation within the analysed method. Here this is done essentially for the winter season, for which day-to-day variations of error covariances are particularly pronounced. Values from 500 and 975 hPa are shown for forecasts valid at 1800 UTC. Relevant conclusions are equally valid for the other assimilation networks, namely 0000, 0600 and 1800 UTC.

4.4.1 Standard deviations and spectral decomposition of variance

Figure 4.6 shows the day-to-day variation of background error standard deviations, for each of the two seasons, and for each of the two independent ensembles. It can be seen that standard deviation values provided by the two independent ensembles are in close agreement (for a given season). This indicates that these regionally averaged

time-dependent standard deviations are estimated in a robust way from a 6-member ensemble. It may also be noticed that the robustness looks larger for a small-scale variable such as vorticity (Figure 4.6c) than for a larger-scale variable like temperature (Figure 4.6b). This is related to the associated larger number of spatially independent error realizations for vorticity over the ALADIN/France domain, which strengthens the robustness of the spatially averaged variances estimates. Small scale variable is a sharpe horizontal autocorrelation.

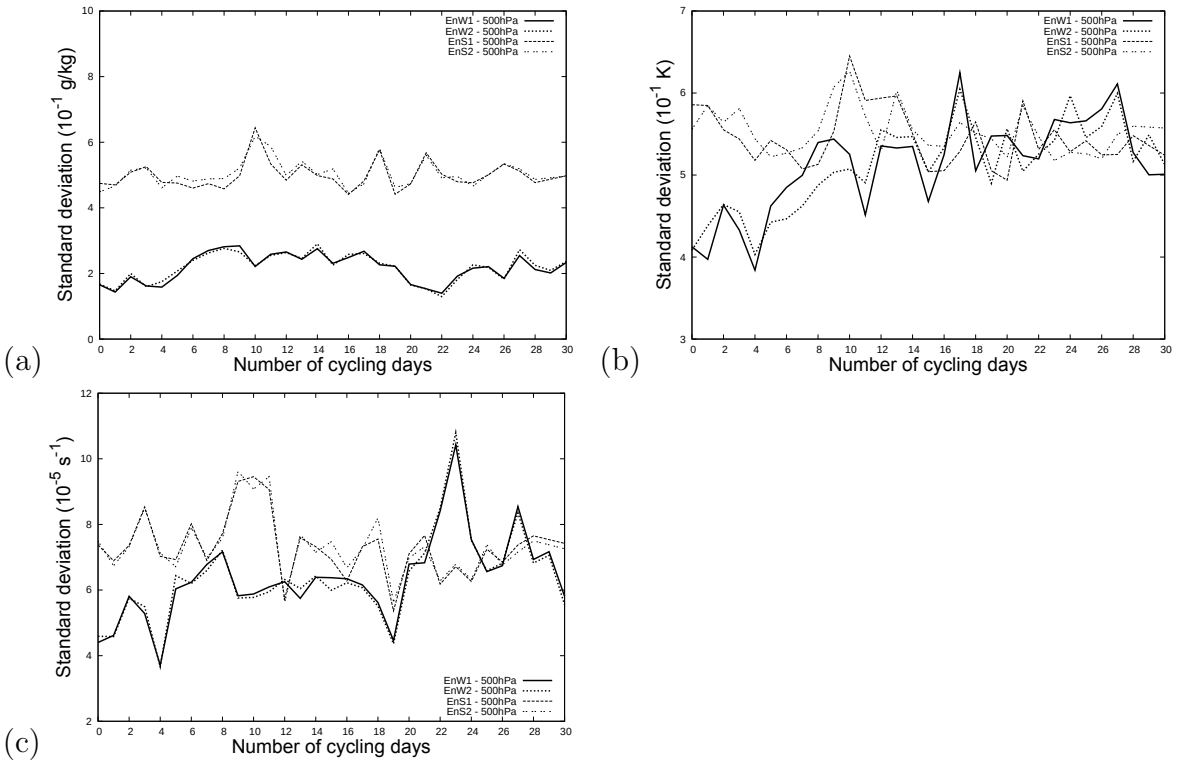
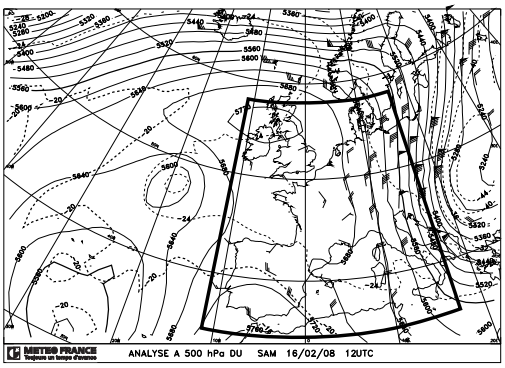


Figure 4.6: Time evolution of horizontally averaged standard deviations of ALADIN/France 6 h forecast errors at 500 hPa (issued from the 1200 UTC network and valid at 1800 UTC), for each of the two independent ensembles and for each of the two periods, EnW1 and EnW2 in winter, and EnS1 and EnS2 in summer, for (a) specific humidity, (b) temperature and (c) vorticity.

One of the first obvious features corresponds to larger humidity error values in summer than in winter (Figure 4.6a), in accordance with what was found in section 4.3 about seasonal variations. This standard deviation increase in summer remains valid

(a) 

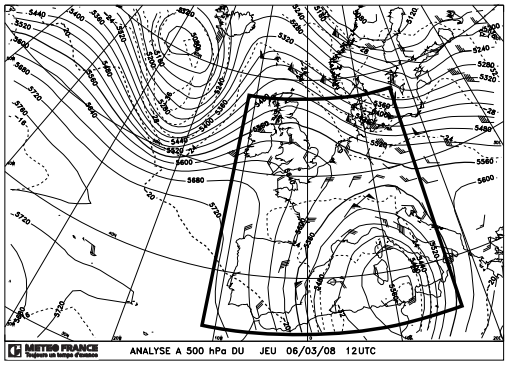
(b) 

Figure 4.7: Two contrasted weather regimes in February/March 2008 over western Europe, in terms of geopotential field (in meters) at 500 hPa and associated wind field: (a) winter anticyclonic flow on 16 February at 1200 UTC; (b) winter short upper air cutoff flow on 6 March at 1200 UTC. The bold line corresponds to the lateral boundaries of the ALADIN/France domain.

for each day of the two periods. The situation is somewhat different for the other variables. Figures 4.6b and 4.6c show that the error increase visible in Figure 4.2 for summer (compared to winter) hides a temporal variability illustrated by Figure 4.6. In particular, there are days in winter for which error standard deviations are larger than some error standard deviations in summer. This is noticeable, for instance, on winter day 23 (6 March 2008), for which vorticity (Figure 4.6c) and divergence (not shown) standard deviations reach a well-marked maximum. This error peak corresponds to the presence of an unstable cutoff low over the ALADIN/France domain (Figure 4.7b). Conversely, the smallest variance values of wind and temperature variables in winter occur on day 4 (16 February 2008). This coincides well with the stable anticyclonic situation illustrated in Figure 4.7(a). In other words, the day-to-day variation of error standard deviations for wind and temperature reflects expected dependencies with respect to the synoptic situation.

Figure 4.8 shows the vertical profiles of standard deviations for each of the two dates of interest (shown in Figure 4.7) and for each of the two ensembles. Figure 4.8, which illustrates an example of day-to-day variation, is interesting to compare with Figure 4.2, which shows the seasonal variation of time-averaged standard deviations. For

specific humidity (not shown), the seasonal variation appears to be stronger than the example of day-to-day variation. This reflects the dependence on the basic water vapour content of the air mass, which is more related to the season than to synoptic variations within a given season. On the other hand, the daily variation visible in

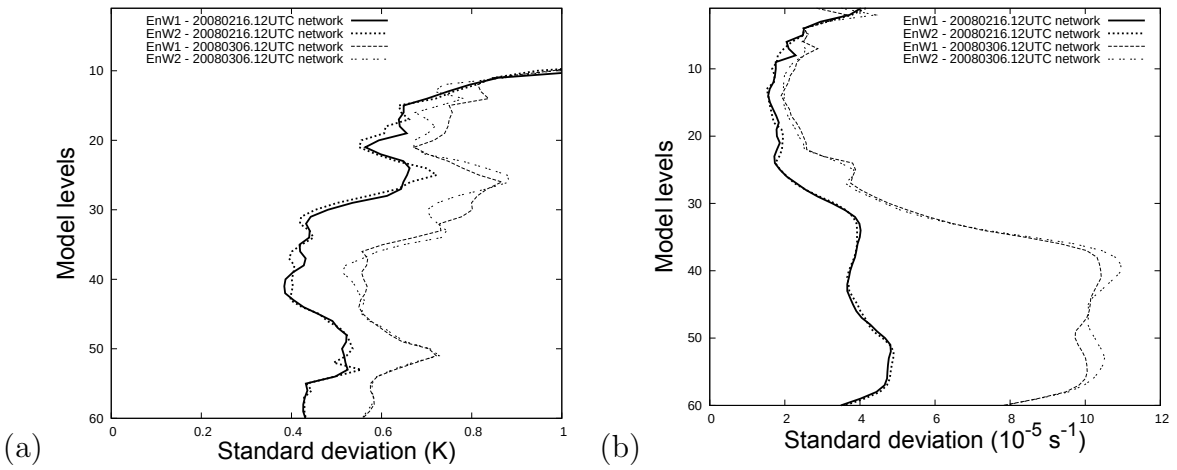


Figure 4.8: Vertical profiles of horizontally averaged standard deviations of ALADIN/France 6 h forecast errors (issued from the 1200 UTC network and valid at 1800 UTC), estimated by two independent ensembles (EnW1 and EnW2) on 16 February and 6 March 2008, for (a) temperature and (b) vorticity.

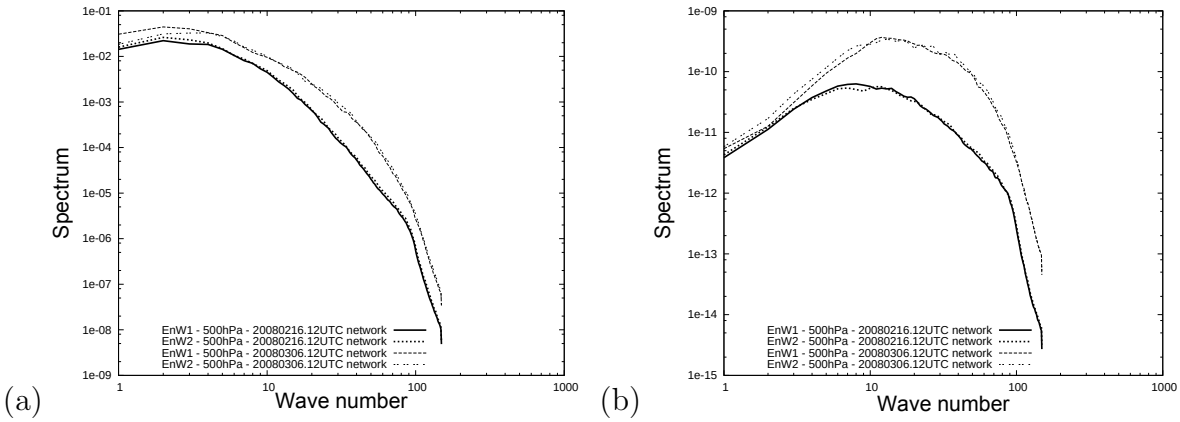


Figure 4.9: Horizontal variance spectra of ALADIN/France 6 h forecast errors at 500 hPa (issued from the 1200 UTC network and valid at 1800 UTC), estimated by two independent ensembles (EnW1 and EnW2) on 16 February and 6 March 2008, for (a) temperature and (b) vorticity.

Figure 4.8 for temperature and wind is stronger than the seasonal variation in Figure 4.2. In particular, we see in Figure 4.8 that in the cyclonic unstable case (compared to the anticyclonic situation) there is a tropospheric standard daviation increase by a factor of around 1.5 for temperature and 2.5 for vorticity.

Figure 4.9 shows how these day-to-day variations of error amplitudes change as a function of horizontal scale. It appears that the variance increase is not uniform spectrally. It tends to be relatively stronger for an intermediate range of wave numbers, typically between 5 and 70. This is particularly visible for vorticity (Figure 4.9b) and implies a change in the shape of the variance spectrum, in which the contribution of intermediate wave numbers is strengthened. This has some consequences on horizontal correlations, which is shown in section 4.4.2.

4.4.2 Horizontal and vertical correlations

Figure 4.10 illustrates the vertical profiles of horizontal length scales (eq. 4.2) for each winter date of Figure 4.7 and for each of the two independent ensembles. In accordance with changes found in the variance spectra (Figure 4.9), horizontal length scales are smaller in the cyclonic situation than in the anticyclonic case. This is particularly visible for tropospheric humidity (Figure 4.10a) and temperature (Figure 4.10b).

Vertical correlations have been examined too, as illustrated for divergence (Figure 4.11a) and vorticity (Figure 4.11b) at 975 hPa. Vertical correlation functions tend to be broader in the cyclonic situation (than in the anticyclonic case), which reflects stronger vertical couplings induced by vertical instabilities. Representing such day-to-day changes of vertical correlations in the data assimilation scheme would thus allow observed information to be vertically prolongated to a larger extent in the unstable case.

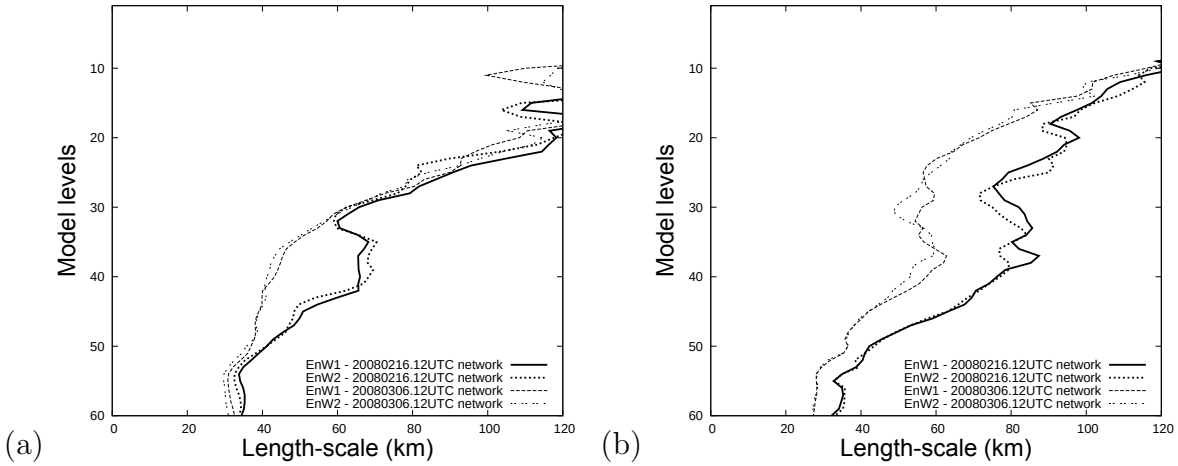


Figure 4.10: Vertical profiles of horizontal length scales for 6 h ALADIN/France forecast errors (issued from the 1200 UTC network and valid at 1800 UTC) estimated by two independent ensembles (EnW1 and EnW2) on 16 February and 6 March 2008, for (a) specific humidity and (b) temperature.

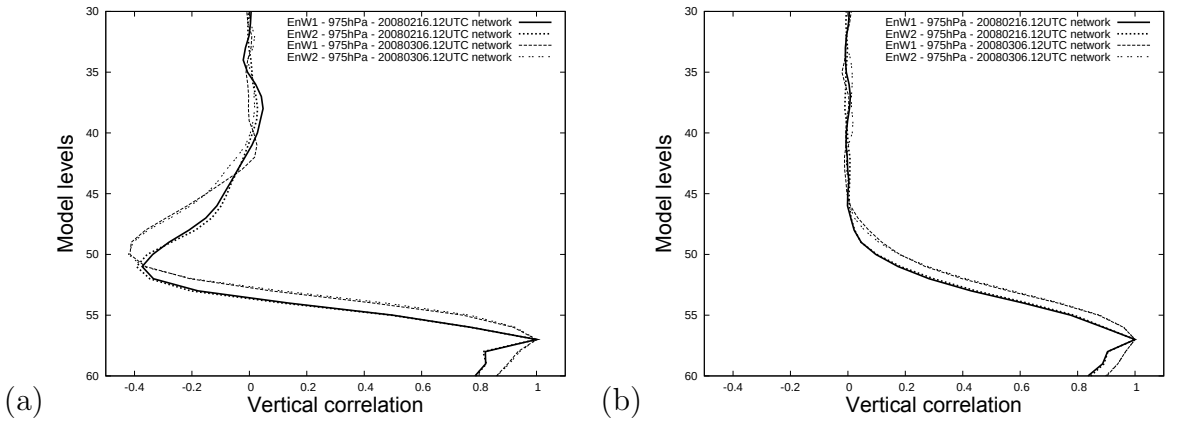


Figure 4.11: Vertical autocorrelation functions of ALADIN/France 6 h forecast errors at 975 hPa (issued from the 1200 UTC network and valid at 1800 UTC), estimated by two independent ensembles (EnW1 and EnW2) on 16 February and 6 March 2008, for (a) divergence, and (b) vorticity.

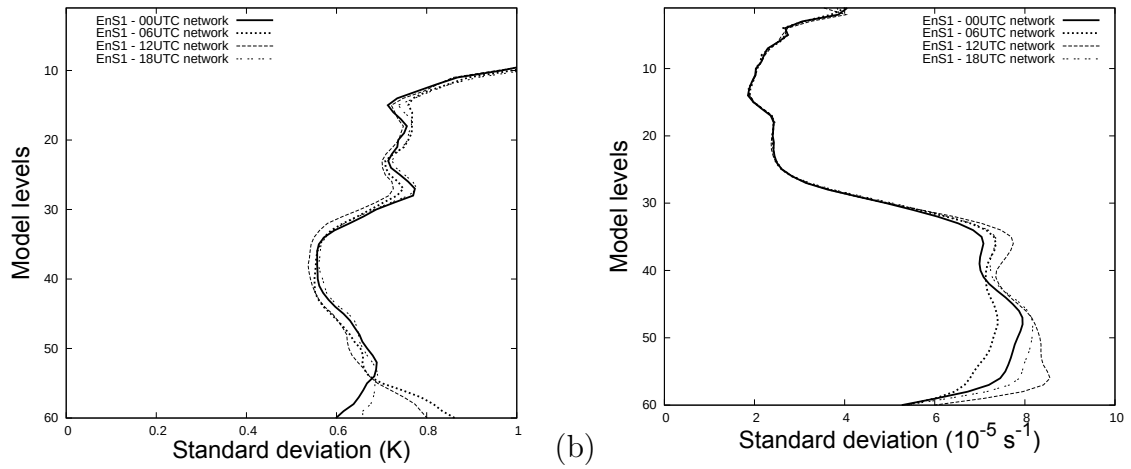
4.5 Diurnal cycle variations (summer period)

Although seasonal and day-to-day variations appear to be significant, it remains to be seen to which extent covariances change as a function of hour within a given day. These potential diurnal variations are studied in this section.

4.5.1 Standard deviations and spectral decomposition of variance

Figure 4.12 shows vertical profiles of standard deviations in the summer period for the four analysis networks: 0000, 0600, 1200 and 1800 UTC. These profiles were plotted using seasonal averages (as was done in section 4.3) to provide a first synthetic view of diurnal variations. One of the most pronounced diurnal variations is for temperature in the planetary boundary layer (PBL), with larger standard deviations for forecasts valid at 1200 and 1800 UTC (issued from the 0600 and 1200 UTC networks, respectively). This can be associated to the increased convective activity for these afternoon and evening hours (e.g. Yang & Slingo (2001)). Diurnal variations are also visible for the other parameters, with a tendency to have the largest standard deviation values at 1800 UTC.

Associated variance spectra at 975 hPa are shown for temperature (Figure 4.13a) and vorticity (Figure 4.13b), for forecasts valid at 0600 and 1800 UTC, respectively. The increase of variance for forecasts valid at 1800 UTC tends to be larger for intermediate wave numbers, between 5 and 70 typically, although this is somewhat



(a) (b)

Figure 4.12: Vertical profiles of horizontally averaged standard deviations of 6 h ALADIN/France forecast errors in summer from each of the four daily networks for (a) temperature and (b) vorticity.

parameter-dependent.

The full time evolution of standard deviations at 975hPa is shown in Figure 4.14, for each of the two periods and for each of the two independent ensembles. Diurnal variations are visible in both seasons, but they have a larger amplitude in summer, as expected. For such levels in the PBL, diurnal changes have a larger amplitude in summer than day-to-day variations, whereas it is the reverse in winter. This indicates that a representation of diurnal variations is particularly important in summer for the PBL.

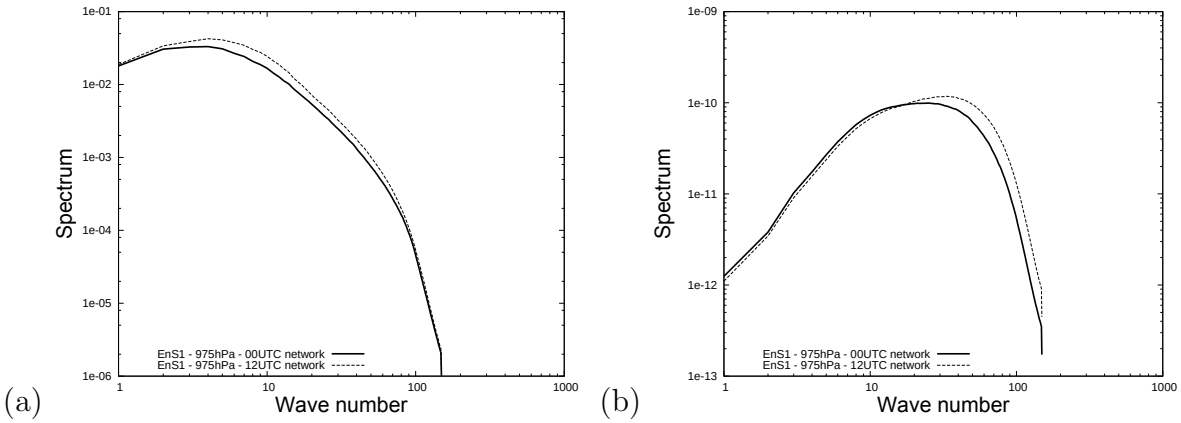


Figure 4.13: Horizontal variance spectra as function of the total horizontal wave number of 6 h ALADIN/France forecast errors at 975 hPa estimated by the first summer ensemble at 0000 UTC (solid line, for forecast fields valid at 0600 UTC) and at 1200 UTC (dashed line, for forecast fields valid at 1800 UTC) for the model parameters (a) temperature and (b) vorticity.

4.5.2 Horizontal and vertical correlations

Figure 4.15 illustrates the full time evolution of horizontal correlation length scales at 975 hPa. It appears that diurnal variations are significant not only for standard deviations (as shown by Figure 4.14), but also for correlation length scales. This Figure 4.15 shows also that length-scale variations are parameter-dependent and that they are influenced by other factors than the diurnal cycle. For instance, length-scale differences

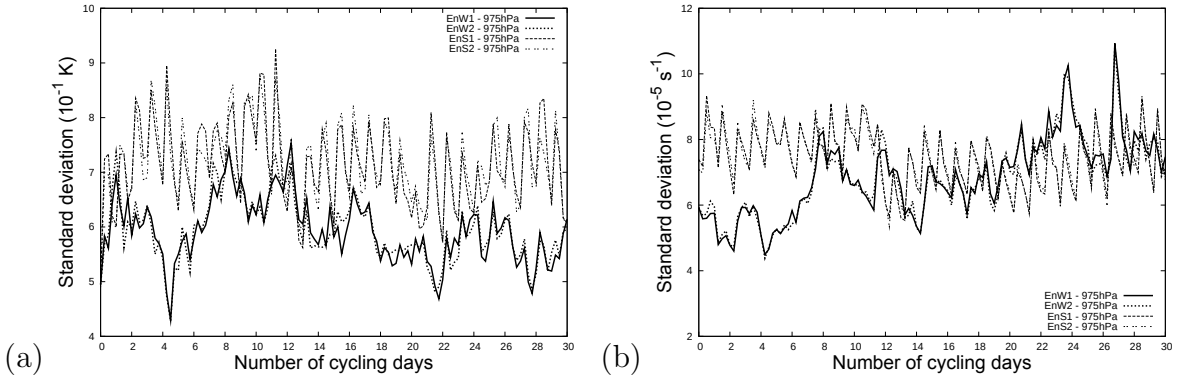


Figure 4.14: Full time evolution of standard deviation at 975 hPa (including all four daily networks) for (a) temperature and (b) vorticity.

between the two periods (winter and summer) are larger at the beginning of the two periods (days 0-15 typically) for specific humidity (Figure 4.15a). In contrast, for divergence (Figure 4.15b), they are larger at the end of the two periods (days 17-30 typically).

Seasonally-averaged vertical profiles of length scales are shown in Figure 4.16 for the two daily networks in summer. It can be seen that, for humidity and temperature in particular, the smallest length-scales occur for forecasts valid at 1800 UTC. As shown by Figure 4.17, vertical correlation functions tend to be slightly broader for forecasts valid at 1800 UTC This is likely to reflect stronger vertical couplings associated to convection in particular, although this time-averaged diurnal change is relatively small in the studied period.

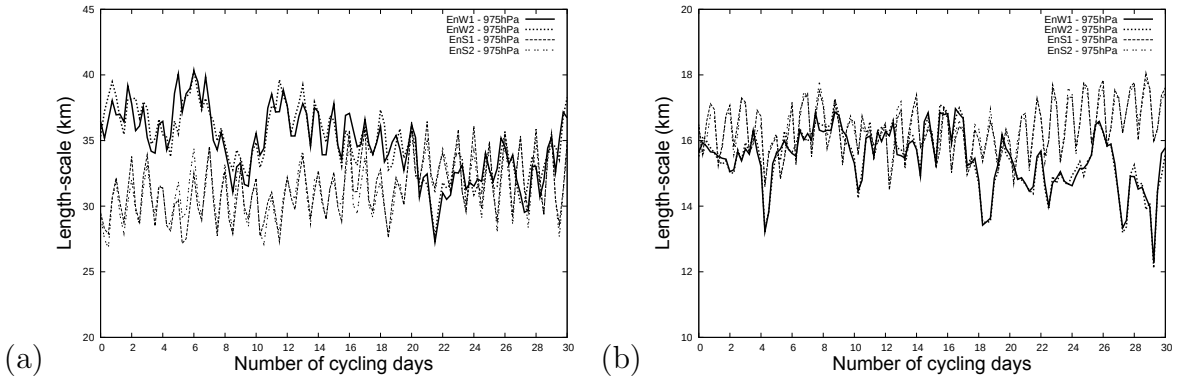


Figure 4.15: Full time evolution of horizontal length scale at 975 hPa (including all four daily networks) for (a) specific humidity and (b) divergence.

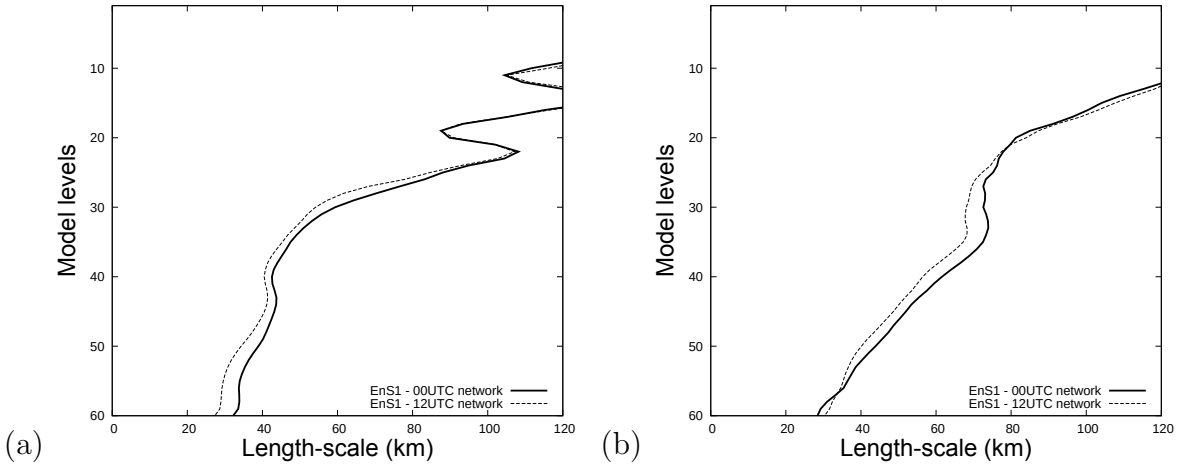


Figure 4.16: Vertical profiles of horizontal length scales of ALADIN/France 6 h forecast errors in summer from two daily networks, 0000 UTC (for forecast fields valid at 0600 UTC) and 1200 UTC (for forecast fields valid at 1800 UTC) for (a) specific humidity and (b) temperature.

(a)

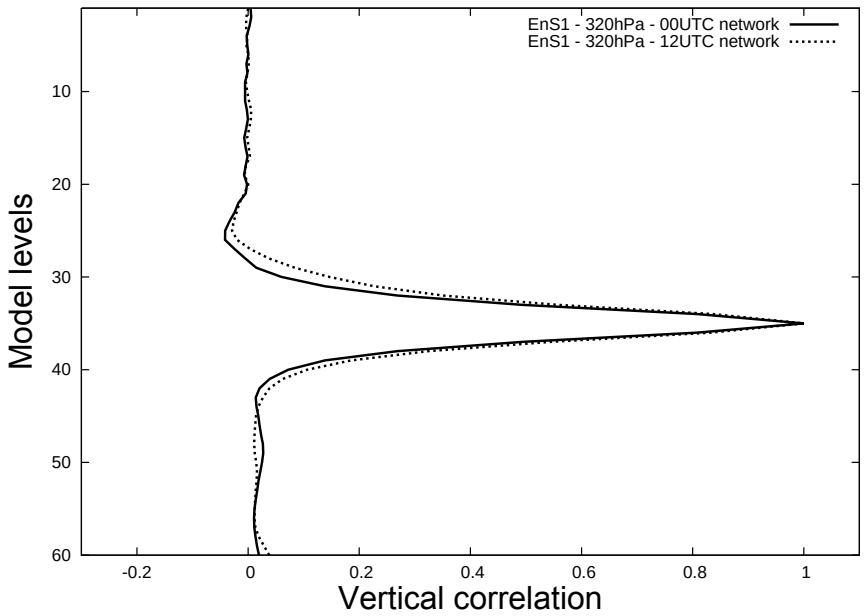


Figure 4.17: Vertical autocorrelations of ALADIN/France 6 h forecast errors at 320 hPa in summer for specific humidity from two daily networks: 0000 UTC (for forecast fields valid at 0600 UTC) and 1200 UTC (for forecast fields valid at 1800 UTC).

4.6 Conclusions

Background error covariances are often estimated by using spatial homogeneity and temporal stationary assumptions. In the case of the ALADIN/France LAM, the homogeneity assumption implies that covariances are calculated from a spatial average over the regional computation domain, which covers western Europe. However, because of its regional character, this spatial average is likely to be more prone to temporal variations than spatial averages over the whole globe. Therefore, the purpose of this study is to revise the temporal stationarity assumption, by diagnosing time variations of such regionally averaged covariances. This was done with the ALADIN/France regional ensemble 3D-Var system by examining variations with season (winter versus summer), day (in connection with the synoptic situation), and hour (related to the diurnal cycle).

It appears that error variances tend to be larger in summer than in winter, in particular for specific humidity, vorticity, divergence, and also for temperature in the boundary layer. Also in summer, correlation functions tend to be sharper horizontally in the troposphere, whereas they tend to be broader vertically (in comparison with winter). These three features reflect implications of increased convective activity in summer.

Day-to-day changes are even more pronounced than seasonal variations, if one compares covariances for a winter anticyclonic situation and for a winter low pressure case. In the latter case (cutoff situation), error standard deviations are much larger in accordance with the instability of the situation, and horizontal correlation length-scales are much reduced for humidity and temperature in particular.

Finally, diurnal variations are also found to be significant, in particular in the boundary layer and in summer. Error standard deviations tend to be larger for forecasts valid at 1800 UTC. It is also at this hour that horizontal length scales are the smallest ones and vertical correlation functions are broadest, in accordance with expected effects of convective activity in the afternoon.

Moreover, all these time-dependent covariance estimates appear to be robust with a 6-member ensemble, according to the comparison between estimates of two independent ensembles. Therefore, these results support the idea of representing these time variations by using a real-time ensemble assimilation system.

A first perspective of this work is thus to implement such time-dependent covariance estimates in the ALADIN/France 3D-Var and to carry out impact studies. Another natural continuation is to extend this study of time variations to heterogeneous covariance estimates, provided by a wavelet formulation, for instance. It is expected that spatial filtering properties of these wavelets will help to provide robust space- and time-varying covariance estimates.

Chapter 5

An impact study of updating background error covariances in the ALADIN/France data assimilation system

This chapter was prepared to be submitted for publication:

Loïk Berre[1] and Maria Monteiro[2]: An impact study of updating background error covariances in the ALADIN/France data assimilation system

[1]CNRM/GMAP, Météo-France, Toulouse, France

[2]CPPN, Instituto de Meteorologia I.P., Lisbon, Portugal

Abstract

The operational ALADIN/France 3D-Var system is based on static background error covariances calculated off-line during a few-week past period. In this study, the impact of an on-line updated specification of background error covariances is evaluated in the ALADIN/France system. This evaluation is done by comparing three experiments, respectively based on (i) covariances calculated from a monthly average over a past period, (ii) covariances calculated from a monthly average over the period of study, and (iii) covariances calculated from a sliding daily average over the period of study. Firstly, it is shown that updating the monthly average of error covariances has a positive impact on the short-range forecast quality. This is related to the specification of covariances which are more representative of average weather regimes at play during the period of study. Secondly, additional positive impacts of a daily update of error covariances are also visible, although they tend to be somewhat localized and modest during this period. These impacts are illustrated by case studies for humidity during an anticyclonic situation, and for wind during a cyclonic event. These results support the idea to consider an on-line updated specification of background error covariances.

5.1 Introduction

Usual data assimilation techniques for Numerical Weather Prediction (NWP) rely on a combination of observed information and of a background, which corresponds to a short-range forecast. These two information sources are weighted by their respective error covariances, and classically (Daley 1991) it can be shown that the role of background error covariances is to spatially filter and propagate observed information. In practice, these error covariances are however difficult to estimate, for instance because the true atmospheric state is never exactly known. Moreover, the size of the covariance matrix is too large to allow for explicit storage and evolution of covariances.

Due to their important role during the analysis step, research efforts are still ongoing to

increase the realism of these background error covariance estimates. This includes the use of error simulation techniques (e.g. Houtekamer *et al.* (1996)), for instance which can rely on the addition and evolution of perturbations representative of errors existing and cycled in the data assimilation system. In addition to these error simulation techniques, research efforts are devoted to the relaxation of assumptions of temporal stationarity and horizontal homogeneity which have been often used in background error covariance modelling (e.g. Fisher (2003)).

The present study is carried out in the framework of the ALADIN/France 3D-Var regional data assimilation system. The currently operational version of this system is based on a static and horizontally homogeneous covariance model, and the present paper focuses on the impact of relaxing the static covariance assumption in this 3D-Var system. This study has been conducted after diagnosing temporal variations of associated covariances (Monteiro & Berre 2010, reproduced in chapter 4), which indicates for instance that seasonal and daily covariance changes are significant, and that they are related to weather situation variations.

The structure of this chapter is as follows. The experimental framework is described in section 5.2. The impact of seasonal variations is presented in section 5.3, while section 5.4 is about the impact of daily variations. Conclusions are discussed in section 5.5.

5.2 Experimental framework

The operational ALADIN/France system is based on a local version of the regional ALADIN model (Horányi *et al.* 1996) and on a 3D-Var data assimilation system (Fischer *et al.* 2005), with boundary conditions provided by the Météo-France global ARPEGE system.

The regional ALADIN model is spectral (based on a bi-Fourier representation of the fields) and results from a limited area counterpart of the ARPEGE/IFS global system (Geleyn *et al.* 1995). The version used in this study is based on a 10 km horizontal

resolution and 60 hybrid vertical levels over a domain that covers France, the Iberian Peninsula, part of surrounding countries and of the Mediterranean Sea. The 3D-var ALADIN/France data assimilation system consists on 6-hourly assimilation cycling (at the main synoptic hours). It uses surface and upper air conventional observations over land and over sea (e.g. SYNOP, BUOY, TEMP and PILOT) and also remote sensing observational data (such as AMSU-A and -B, HIRS, MHS, AMV, SEVIRI, AIRS and IASI data).

The ALADIN/France background error covariance model is detailed in Berre (2000). It is based on horizontally homogeneous but scale-dependent covariance estimates. These covariances are usually calculated off-line from a few-week average of forecast perturbation covariances, which are obtained by running an ensemble of perturbed assimilation cycles (e.g. (Houtekamer *et al.* 1996), (Fisher 2003), (Belo Pereira & Berre 2006)). This ensemble assimilation method is based on the explicit addition of observation perturbations to real observations in order to simulate the effect of observation errors. It also relies on implicit background perturbations which are provided by the previous analysis perturbations, and explicit model perturbations may also be added to represent model error contributions. The ALADIN/France ensemble assimilation cycle is coupled to the global ARPEGE ensemble assimilation system Berre *et al.* (2007), which allows lateral boundary condition (LBC) errors to be simulated through the use of perturbed ARPEGE boundary conditions. Apart from these LBC perturbations, the current version of the ALADIN/France ensemble assimilation system is nevertheless based on a perfect model assumption (i.e. no explicit model perturbations are added), and resulting background error variance estimates are increased by a factor 2 typically, to account for unrepresented model error contributions.

The version of the ALADIN/France 3D-Var which has been operational in 2008 is based on covariance estimates which have been temporally averaged off-line over a three-week period in Autumn 2007 (9 September - 4 October 2007) from a 6-member ensemble. This operational run will be refered to by using the acronym AUT07S

(where the final letter S refers to the use of static covariances). In order to evaluate the impact of relaxing this static covariance approach, this operational version will be compared with two other experimental versions during a one-month winter period (13 February - 14 March 2008).

The first experimental run is based on covariance estimates which are temporally averaged off-line over the winter period of study (13 February - 14 March 2008) from a 6-member ensemble. This experiment will be denoted by WIN08S. The comparison between WIN08S and AUT07S allows for the evaluation of the impact of monthly variations affecting covariance estimates during the winter 2008 period compared to the autumn 2007 period.

The second experimental run is based on covariance estimates which are temporally averaged in a sliding way over each day preceding each analysis network. For instance, for the analysis calculated on 13 February at 0000 UTC, background error covariances are estimated from a temporal average over samples of 6h perturbed forecasts produced from perturbed analyses calculated on 12 February at 0000, 0600, 1200 and 1800 UTC. In order to compensate the decrease of sample size when using such 4-network averages (instead of a few-week average), a 12-member ensemble has been used to estimate these daily covariances (instead of a 6-member ensemble in WIN08S and AUT07S) in a total of 48 realizations. The experiment using these daily covariances will be denoted by WIN08D (where the final letter D refers to the daily approach). The comparison between WIN08D and WIN08S allows the impact of daily changes of covariances to be evaluated.

The impact of monthly and daily variations of covariances will be assessed in particular by calculating root-mean squared errors and biases of AUT07S, WIN08S and WIN08D of forecasts up to 48 h with respect to TEMP and SYNOP observations, and also with respect to the ECMWF analysis.

5.3 Impact of monthly covariance variations

5.3.1 Impact on the forecast quality

As described in the previous section, comparing experiments AUT07S and WIN08S allows for the evaluation of the forecast quality impact of monthly covariance variations between the two considered few-week periods used for covariance estimation.

The impact on the 12 hour forecast average quality is illustrated in Figures 5.1 and 5.2. While the impact is fairly neutral above 300 hPa, positive impacts of WIN08S are noticeable (although modest) in the middle and low troposphere, for both temperature (Figure 5.1) and wind (Figure 5.2).

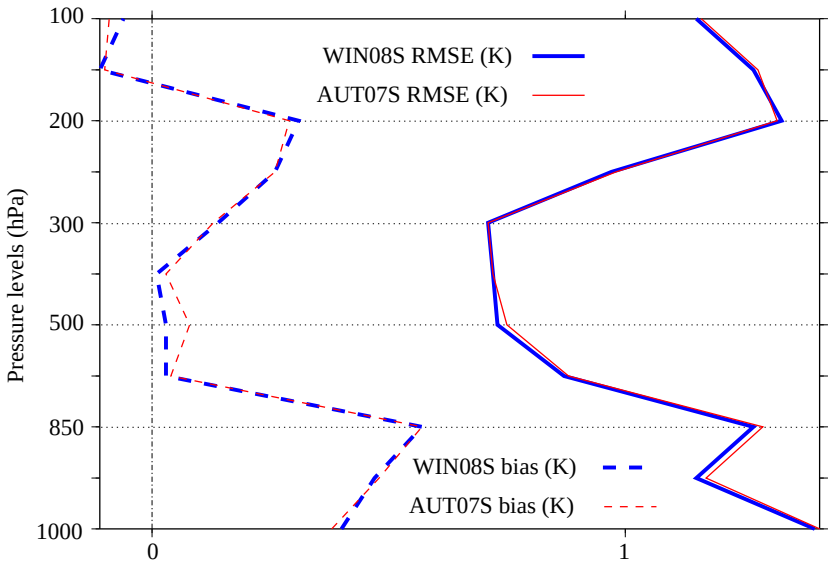


Figure 5.1: Vertical profiles of statistics of the departures from radiosonde observations of 12 h ALADIN/France temperature forecasts (valid at 1200 UTC, on the period 20080213-20080313), for different monhtly-averaged background error covariances.

Examination of associated temporal variations are shown for 500 hPa temperature in Figure 5.3 and for 500 hPa wind in Figure 5.4. These Figures indicate that improvements are relatively frequent over the whole period. This is also supported for instance by Figure 5.5, which corresponds to the impact on 24 hour forecasts of

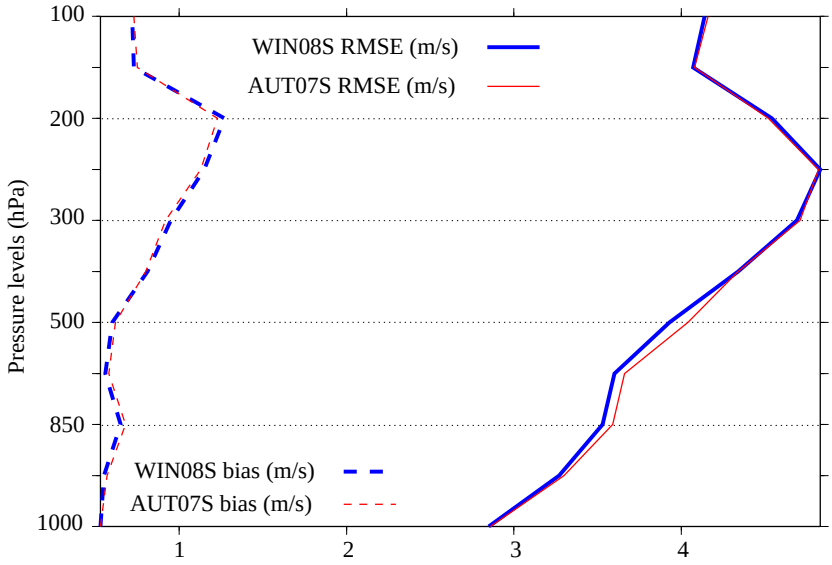


Figure 5.2: Vertical profiles of statistics of the departures from radiosonde observations of 12 h ALADIN/France wind forecasts (valid at 1200 UTC, on the period 20080213-20080313), for different monhtly-averaged background error covariances.

relative humidity at 700 hPa.

These results suggest that using covariance estimates averaged during the one-month period of study has a positive impact on the short-range forecast quality. The impact is rather neutral beyond the 24 hour forecast range (not shown). This is likely to be related to the influence of the lateral boundary conditions, which are identical in the two experiments, and whose impact on the regional forecast increases when the forecast range is larger.

5.3.2 Changes in analysis fit and in vertical correlations

In order to understand the positive impact of WIN08S on short range forecasts, the analysis fit to observations has been compared between experiments WIN08S and AUT07S. A general increase of the analysis fit to observations can be noticed in WIN08S compared to AUT07S. This is illustrated in Figure 5.6 for the analysis fit to TEMP observations of temperature at 500 hPa. This suggests that the short range

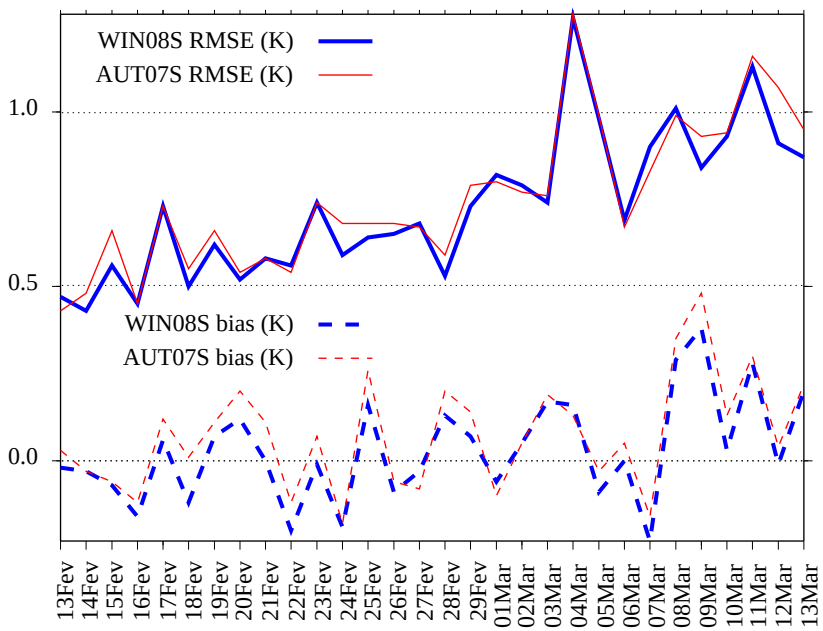


Figure 5.3: Time evolution of statistics of the departures from radionsonde observations of 12 h ALADIN/France 500 hPa temperature forecasts (valid at 1200 UTC, on the period 20080213-20080313), for different monthly-averaged background error covariances.

positive impact arises from an increased analysis fit to observations (believing that the observations are of good quality).

These analysis fit changes between WIN08S and AUT07S arise from differences in the background error covariances that are specified respectively in these experiments. While specified standard deviations and horizontal correlations of background errors are similar in the two experiments (not shown), vertical correlation diagnostics indicate that the increased analysis fit is likely to be connected with the use of sharper vertical correlations in WIN08S compared to AUT07S. The increased sharpness of vertical correlations in WIN08S is illustrated at 500 hPa in Figure 5.7 for specific humidity (top panel) and for divergence (bottom panel). These vertical correlation differences reflect the fact that vertical mixing processes were relatively less prevailing during the winter 2008 period than in the autumn 2007 time interval.

These results support the idea that using background error covariances that are repre-

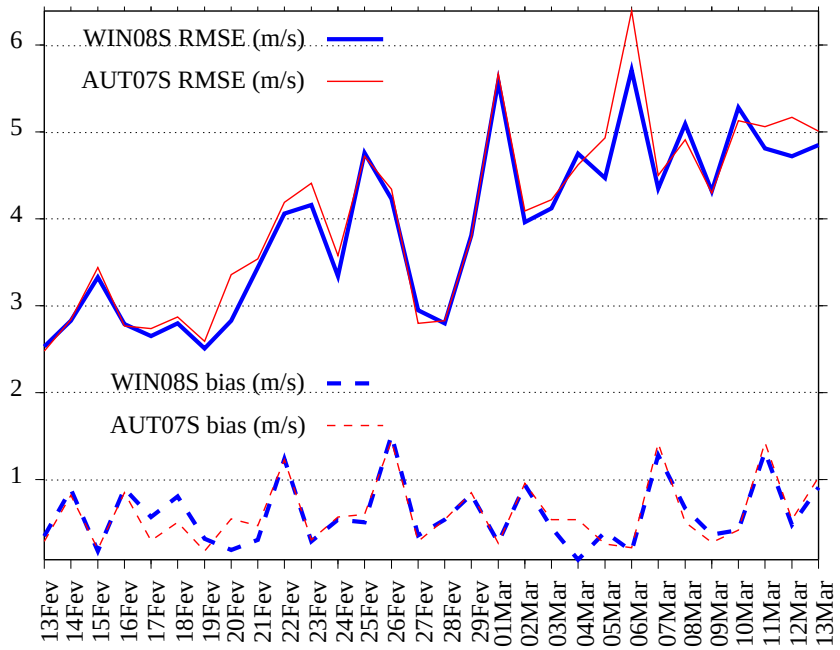


Figure 5.4: Time evolution of statistics of the departures from radiosonde observations of 12 h ALADIN/France 500 hPa wind forecasts (valid at 1200 UTC, on the period 20080213-20080313), for different monthly-averaged background error covariances.

sentative of prevailing weather regimes during the impact study can be beneficial to the short range forecast quality.

5.4 Impact of daily covariance variations

As described in the section 5.2, comparing experiments WIN08S and WIN08D allows for the evaluation of the forecast quality impact of daily covariance variations within the considered February-March 2008 period.

5.4.1 Global impact

Examination of time-averaged forecast scores indicates that the average impact of daily variations of covariances is nearly neutral for the considered period. This is illustrated in Figure 5.8 for the time series of surface pressure RMSE of 6 hour forecasts with respect to SYNOP data. It appears that the RMSE values are very close between

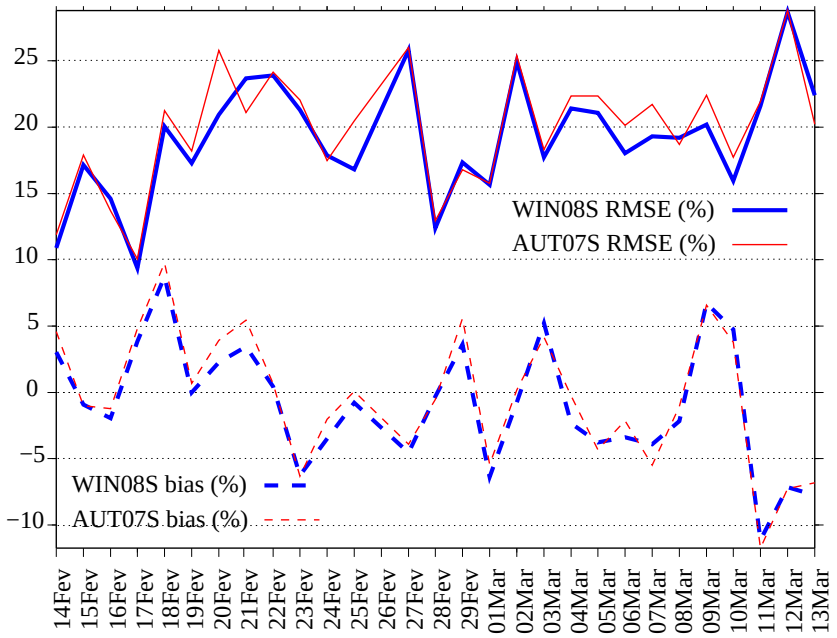


Figure 5.5: Time evolution of statistics of the departures from radiosonde observations of 24 h ALADIN/France 700 hPa relative humidity forecasts (valid at 0000 UTC, on the period 20080214-20080313), for different monthly-averaged background error covariances.

the two experiments, although a tendency to have local positive impacts of WIN08D (compared to WIN08S) can be noticed, for instance at the beginning of the period, and also during the March part of the period.

To some extent, the global neutral impact is expected in the sense that daily covariances in WIN08D are relatively similar on average to those in WIN08S, since they are calculated from the same February-March period. On the other hand, one could also expect that the local impact of WIN08D may be daily-varying and somewhat larger for dates during which daily covariances are relatively different from the one-month average used in WIN08S. This expectation tends to be supported by the somewhat larger positive impact of WIN08D visible in Figure 5.8 at the beginning of the period, which is anticyclonic, and in the March part which is cyclonic.

This kind of local positive impacts will be further illustrated by two case studies.

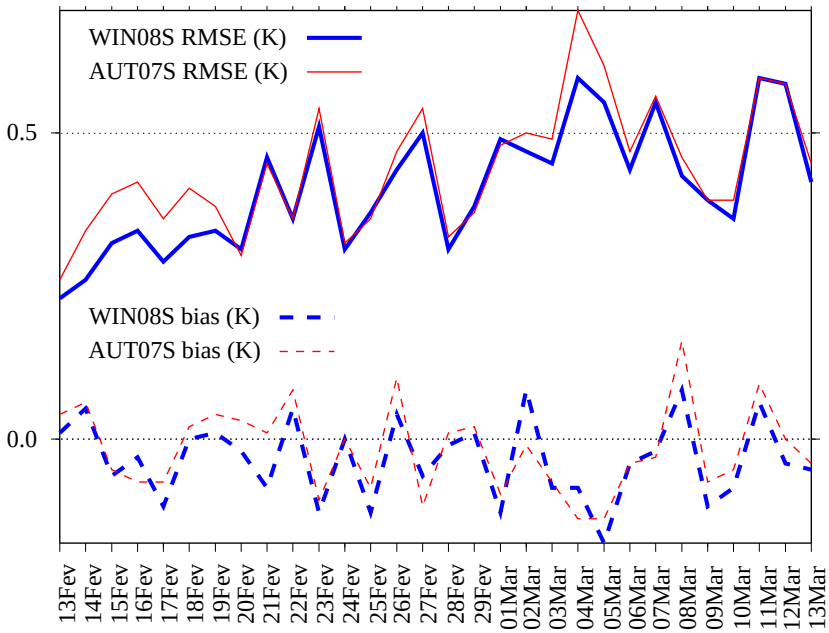


Figure 5.6: Time evolution of statistics of the departures from radionsonde observations of ALADIN/France 500 hPa temperature analysis (valid at 0000 UTC, on the period 20080213-20080313), for different monthly-averaged background error covariances.

5.4.2 Impact on humidity during the anticyclonic period

Figure 5.9 corresponds to the time evolution of horizontally averaged background error standard deviation estimates of specific humidity near 500 hPa which are specified in WIN08D. It can be noticed in particular that there is a relatively large contrast between small values (around 0.15 g/kg) in the first six days of the period (during the anticyclonic period) and larger values (up to 0.25-0.3 g/kg) during the next days (during a transition situation to the cyclonic period). The static horizontally averaged background error standard deviation of WIND08S is also illustrated to contrast with the daily flutuations of the on-line approach.

This is related to the fact that the first six days correspond to a winter anticyclonic situation with relatively cold and dry air, while the next days are affected by a zonal regime with a warmer and moister atmosphere. This dependence of moisture error standard deviations on the weather situation is also visible for instance in the time

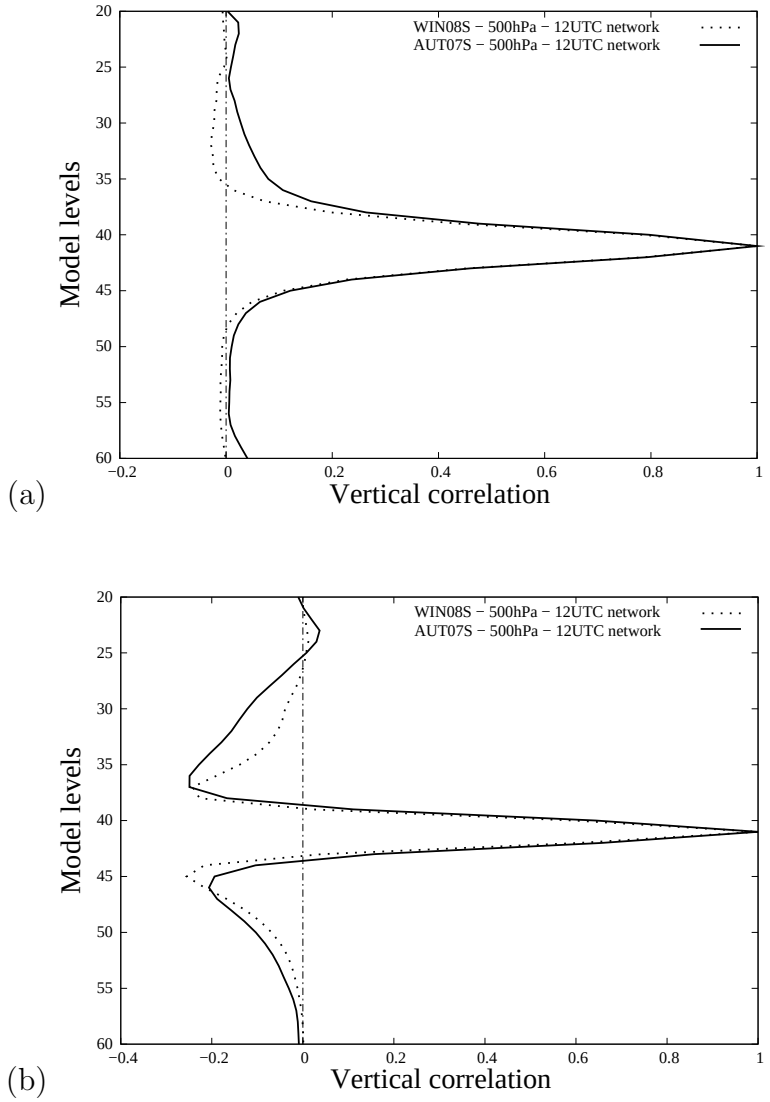


Figure 5.7: Vertical autocorrelation functions of 6 h ALADIN/France 500 hPa forecast errors (issued from the 1200 UTC network and valid at 1800 UTC), for :(a) specific humidity and (b) divergence.

series of 12 hour forecast RMSE of relative humidity at 500 hPa. This is shown in Figure 5.10, which corresponds to a zoom over the first 10 days over the period, in order to illustrate changes between the first 5-6 anticyclonic days (with values around 15 %) and the next dates (with values up to 30 %).

It can also be noticed in Figure 5.10 the 12 hour forecast RMSE is smaller for WIN08D than in WIN08S during the anticyclonic 6 day beginning of the period. This suggests

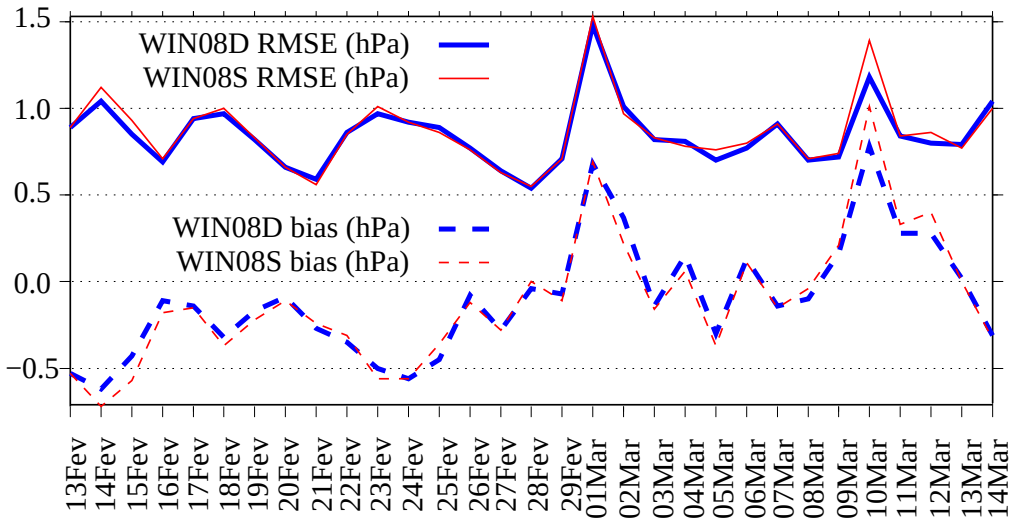


Figure 5.8: Time evolution of statistics of the departures from surface observations of 6 h ALADIN/France mean sea level pressure forecasts (valid at 0600 UTC), for different time averaged background error covariances.

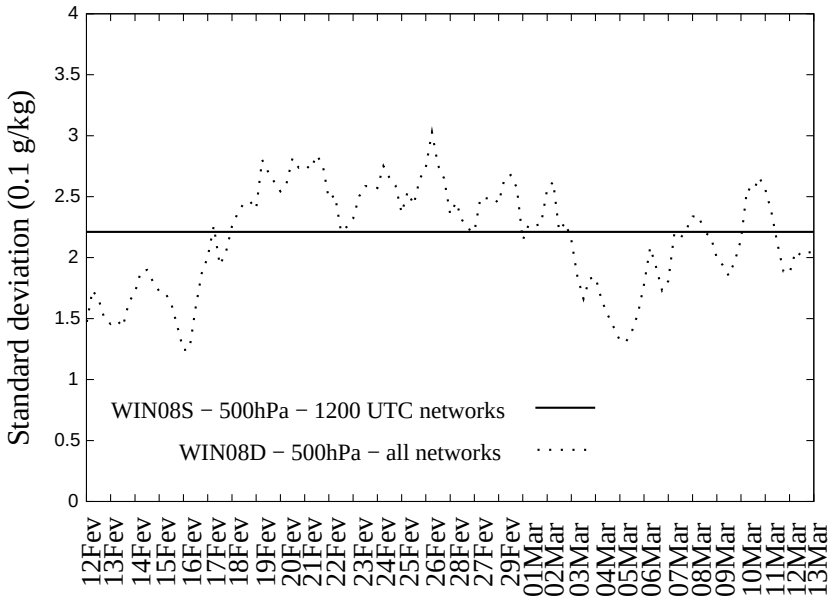


Figure 5.9: Time evolution of horizontally-averaged standard deviation of background errors for specific humidity near 500 hPa.

that this may be a good example to illustrate local positive impacts of representing daily variations of background error standard deviations.

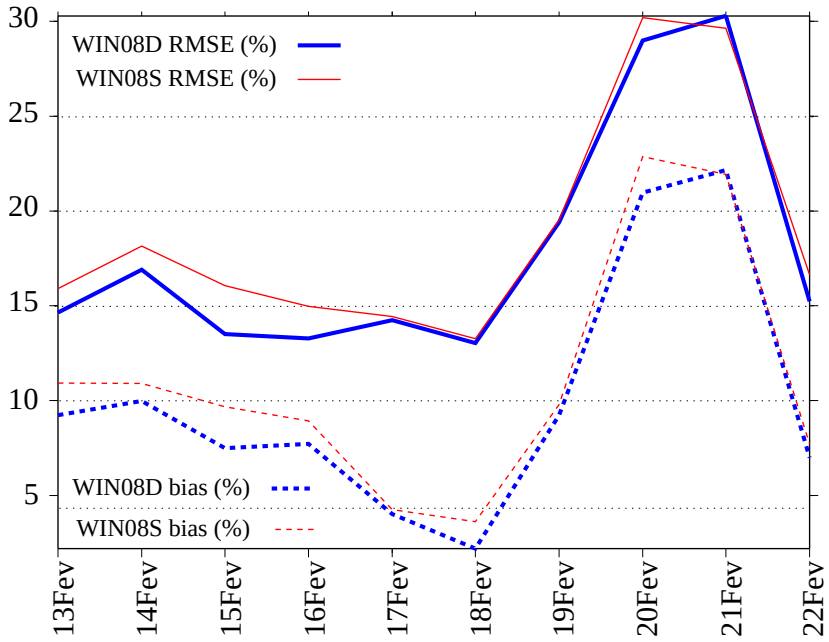
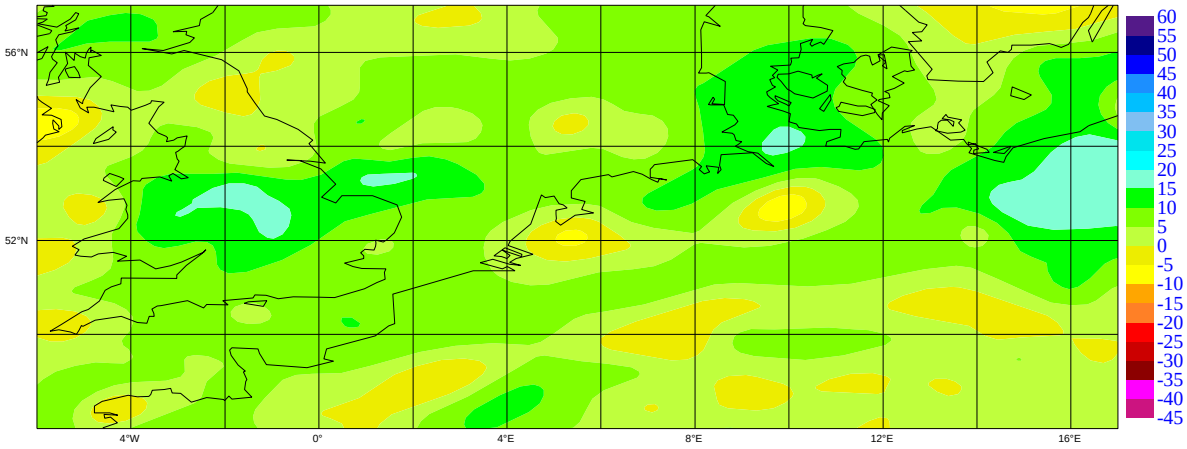


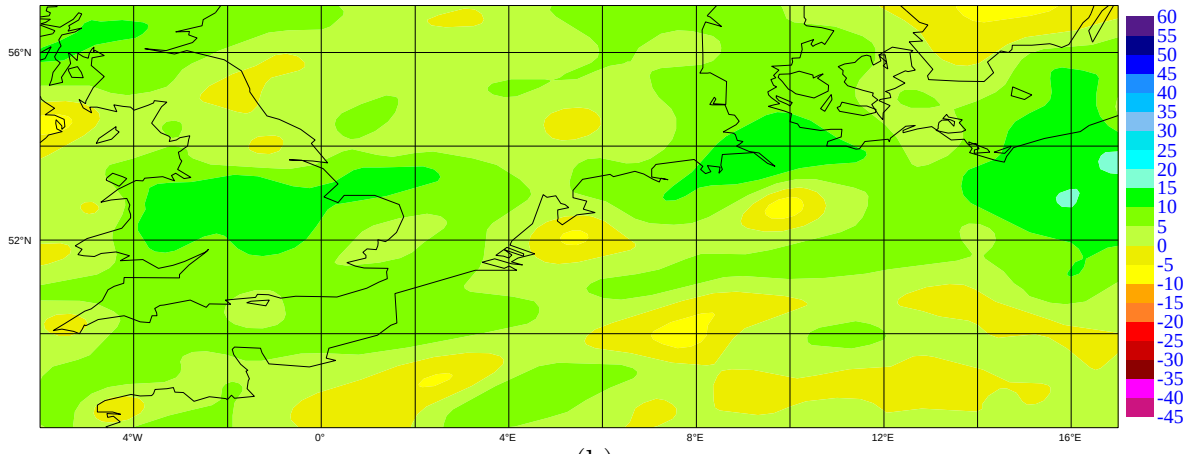
Figure 5.10: Time evolution of statistics of the departures from radiosonde observations of 12 h ALADIN/France 500 hPa relative humidity forecast (valid at 1200 UTC), for different time-averaged background error covariances. Zoom over the anticyclonic period.

To examine this more in detail, respective analysis increments are plotted in Figure 5.11 (for the 13th of February 2008). While the spatial structure of increments is relatively similar between the two experiments, it can be seen that the amplitude of increments is smaller in WIN08D than in WIN08S. This is a direct effect of specifying smaller daily-varying humidity background error standard deviations in WIN08D than the monthly-averaged values used in WIN08S, as shown in Figure 5.9. This shows that specifying daily-varying humidity background error variances that are consistent with the weather situation and with the associated water content of the air mass can be beneficial to the forecast quality.

This tends to be supported by Figures 5.12 and 5.13 which correspond to differences between ALADIN/France 6 hour forecasts and the ECMWF analysis taken here as a reference. It can be seen in Figure 5.12 that the amplitude of departures is smaller in WIN08D than in WIN08S, for instance over Denmark, England, and also in the



(a)

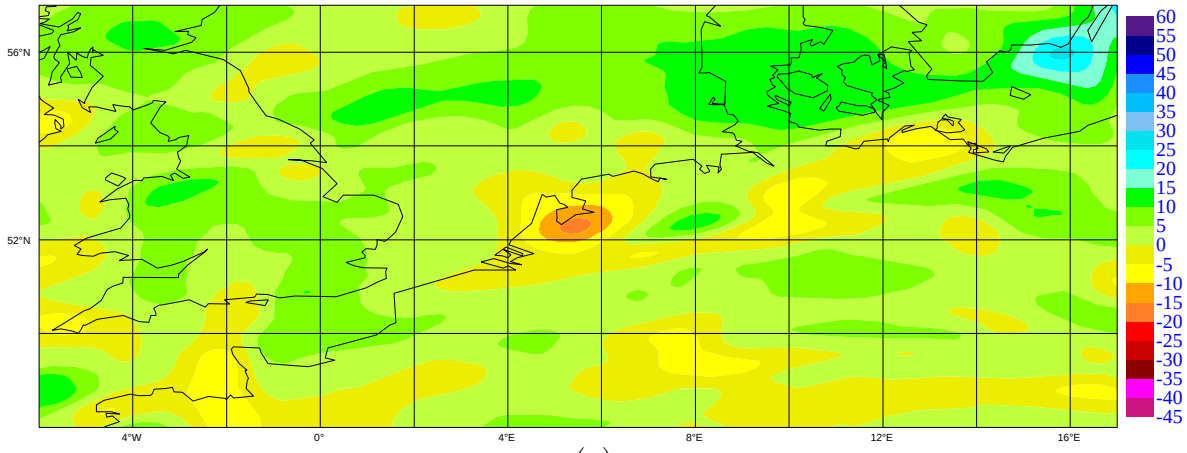


(b)

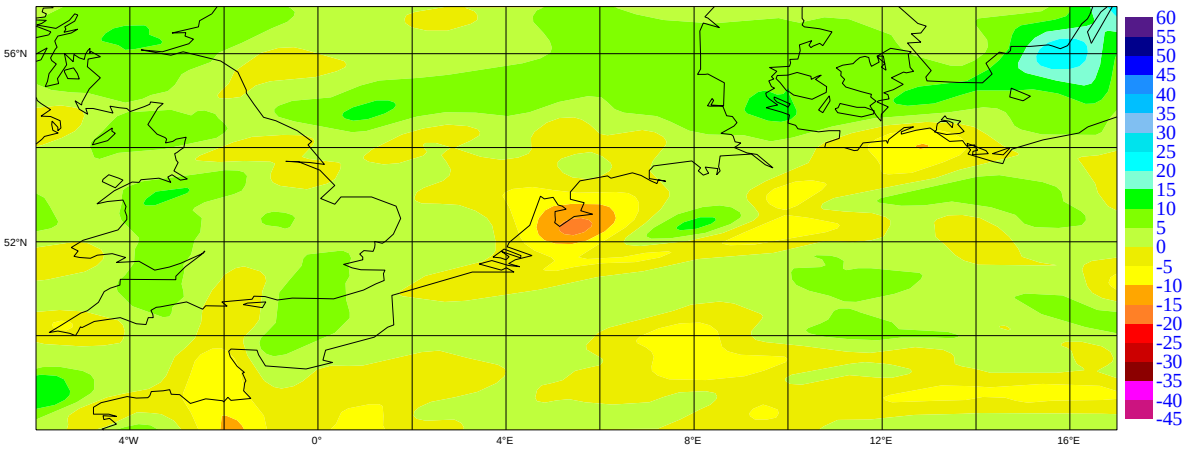
Figure 5.11: Spatial distribution of analysis increments of ALADIN/France relative humidity at 500 hPa, on the 20080213 at 0000 UTC, for the experiments: (a) WIN08S and (b) WIN08D. Zoom over the north-east part of the domain.

surrounding oceanic areas. This coincides well with areas where the amplitude of analysis increments has been reduced in WIN08D compared to WIN08S.

These effects are also visible in Figure 5.13, which indicate that the amplitude of departures from the ECMWF analysis is noticeably reduced during the anticyclonic first five days.



(a)



(b)

Figure 5.12: Spatial distribution of departures from ECMWF analysis of ALADIN/France 500hPa relative humidity analysis on the 20080213 at 0000 UTC, for the experiments: (a) WIN08S and (b) WIN08D. Zoom over the north-east part of the domain.

5.4.3 Impact on wind during the cyclonic period

Figure 5.14 corresponds to the time-averaged RMSE vertical profile of 24 hour forecast with respect to radiosondes taken as a reference. An average slight positive impact of WIN08D can be noticed at 850 hPa in particular. Examination of time series indicates that this kind of average slight positive impact tends to be more pronounced during the March cyclonic period. This is illustrated in Figure 5.15, which corresponds to 36 hour forecast RMSE for wind at 850 hPa, zoomed over the March cyclonic period. The

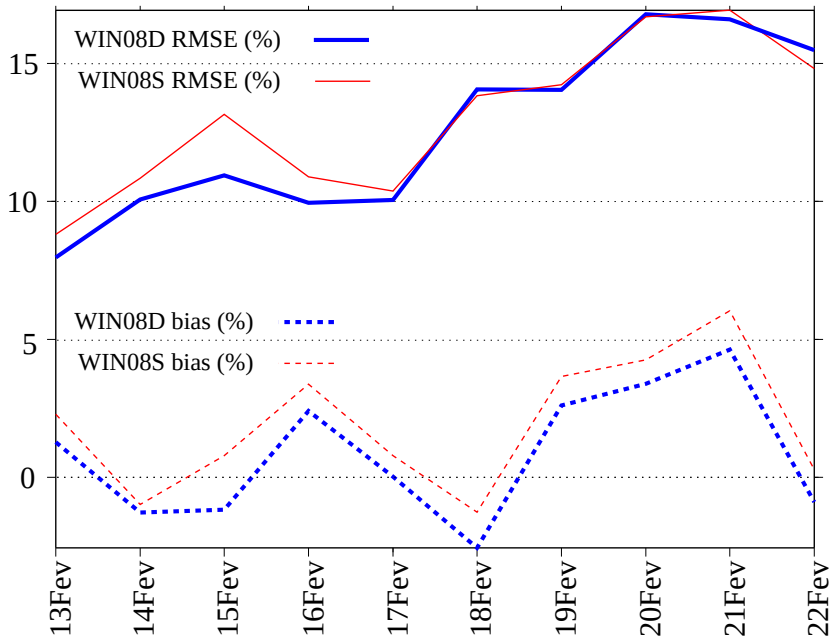


Figure 5.13: Time evolution of statistics of the departures from ECMWF analysis of ALADIN/France 500 hPa relative humidity analysis (valid at 0000 UTC), for different time-averaged background error covariances. Zoom over the anticyclonic period.

positive impact of WIN08D is particularly noticeable for the 36 hour forecast valid on 11 March at 1200 UTC, which has been launched from the analysis calculated on 10 March at 0000 UTC. It is interesting to investigate whether this may correspond to specific daily changes in the background error covariances specified in WIN08D.

To examine this, an example of respective analysis increments of WIN08S and of WIN08D is plotted in Figure 5.16 for zonal wind at 850 hPa. This Figure 5.16 corresponds to increments produced on 10 March at 0000 UTC, zoomed over the southern part of France and surrounding areas. It can be seen for instance that the amplitude of increments is much larger for WIN08D than for WIN08S near the South-East coast of France. Figure 5.17 shows that this is consistent with larger specified vorticity background error standard deviations in WIN08D than in WIN08S for this date, which belongs to the cyclonic part of the period of study. Figure 5.18 indicates that this leads to smaller departures with respect to the ECMWF analysis (taken as a reference here) for the WIN08D analysis than for the WIN08S analysis.

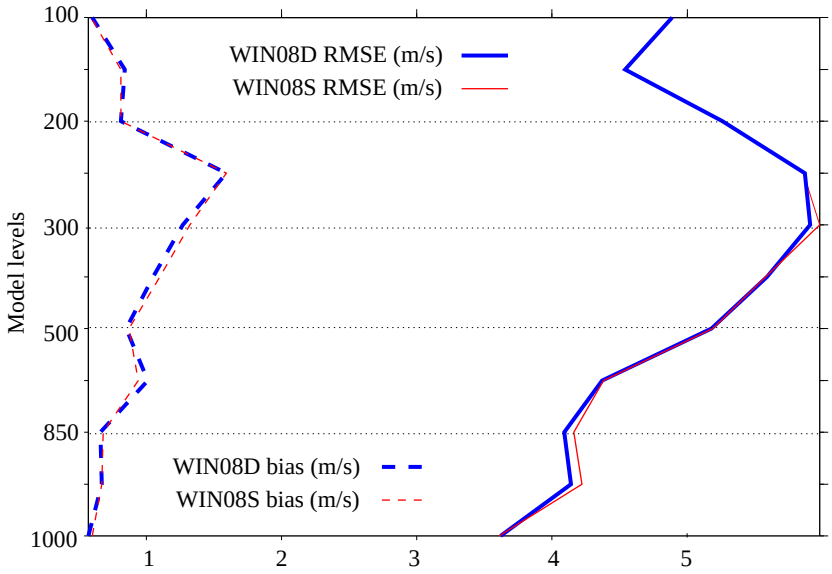


Figure 5.14: Vertical profiles of statistics of the departures from radiosonde observations of 24 h ALADIN/France wind forecasts (valid at 0000 UTC, on the period 20080213-20080313), for different time-averaged background error covariances.

This is visible for instance in the area corresponding to the dark blue departure structure that is elongated meridionally near the middle of Figure 18, for instance with maximum values of 8-10 m/s in WIN08S reduced to 6-8 m/s in WIN08D.

This is a case that illustrates the mechanism through which specifying daily-varying background error standard deviations in cyclonic situations can lead to daily-varying analysis increment amplitudes and potentially more realistic analysis estimates.

5.5 Conclusions

In this study, the impact of temporally updating specified background error covariances has been studied for the ALADIN/France 3D-Var system during a one-month winter period. This has been carried out by comparing an operational 3D-Var version, which uses covariances that are estimated off-line from a few-week average over September 2007, with two experimental versions in which updated background error covariance

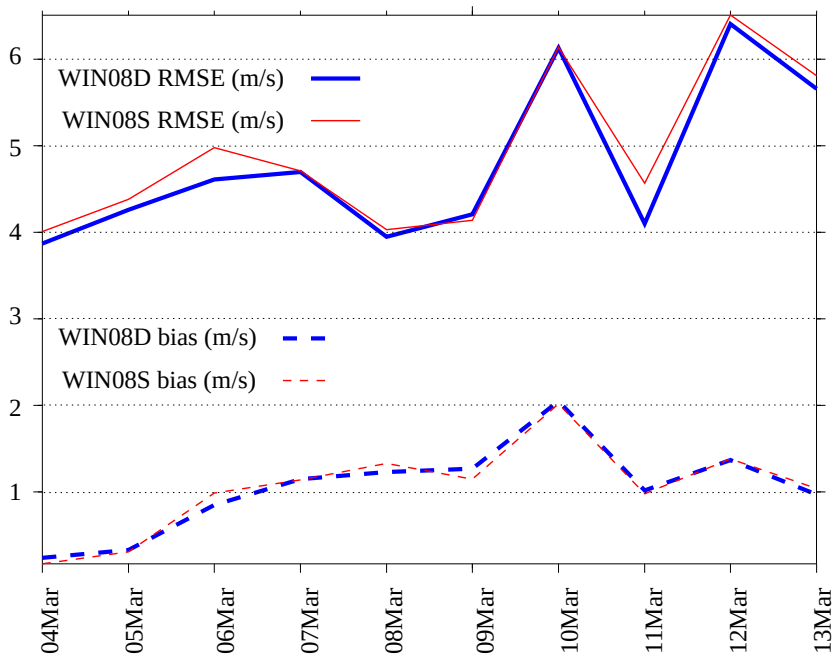


Figure 5.15: Time evolution of statistics of the departures from radiosonde observations of 36 h ALADIN/France 850hPa wind forecasts (valid at 1200 UTC), for different time-averaged background error covariances. Zoom over the cyclonic period.

estimates are specified.

The first experimental version uses specified background error covariances estimated from a one-month average corresponding to the period of study. This allows for using time-averaged covariances that are consistent with average weather regimes prevailing over the period of study. Results indicate that using such updated covariances has a positive impact on the short-range forecast quality of temperature, wind and humidity. This is connected to an increased analysis fit to observations, which arises from sharper vertical correlations for the updated winter covariances than in the operational September covariances.

The second experimental version uses specified background error covariances estimated from one-day sliding averages preceding each analysis date. Compared to the first experimental version, this allows also daily covariance variations (in

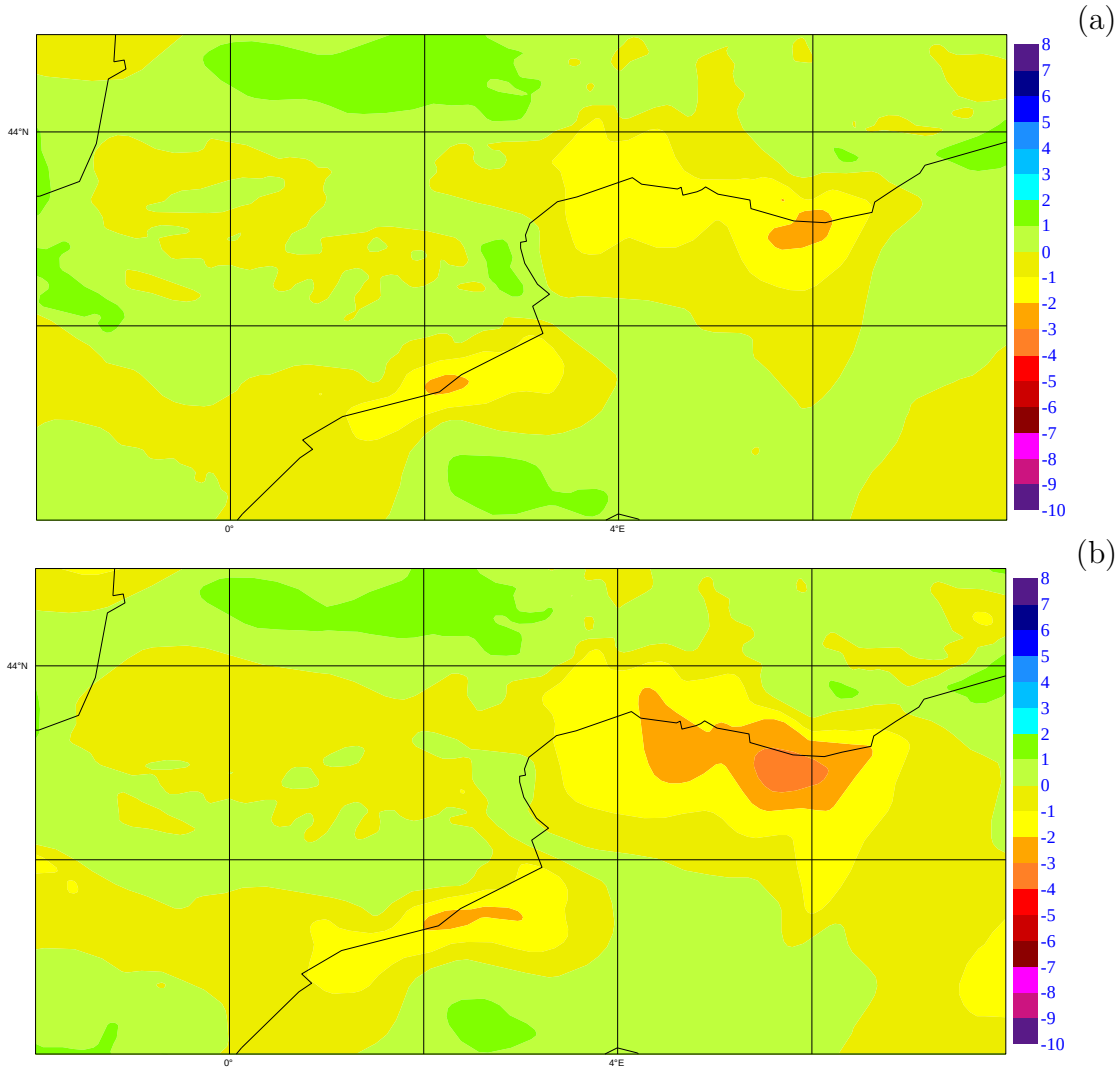


Figure 5.16: Spatial distribution of analysis increments of ALADIN/France zonal wind at 850 hPa, on the 20080310 at 0000 UTC, for the experiments: (a) WIN08S and (b) WIN08D. Zoom over the central part of the domain, at the south coast of France.

addition to monthly variations) to be taken into account in the 3D-Var system. Results indicate that, during our period of study, these daily variations have more modest and localized positive impacts than the monthly variation impact which has been studied by comparing the operational and first experimental version. Case studies have been shown to illustrate local positive effects of daily-varying background error standard deviations for humidity during an anticyclonic situation and for wind during a cyclonic period. For temperature there were no significant improvements.

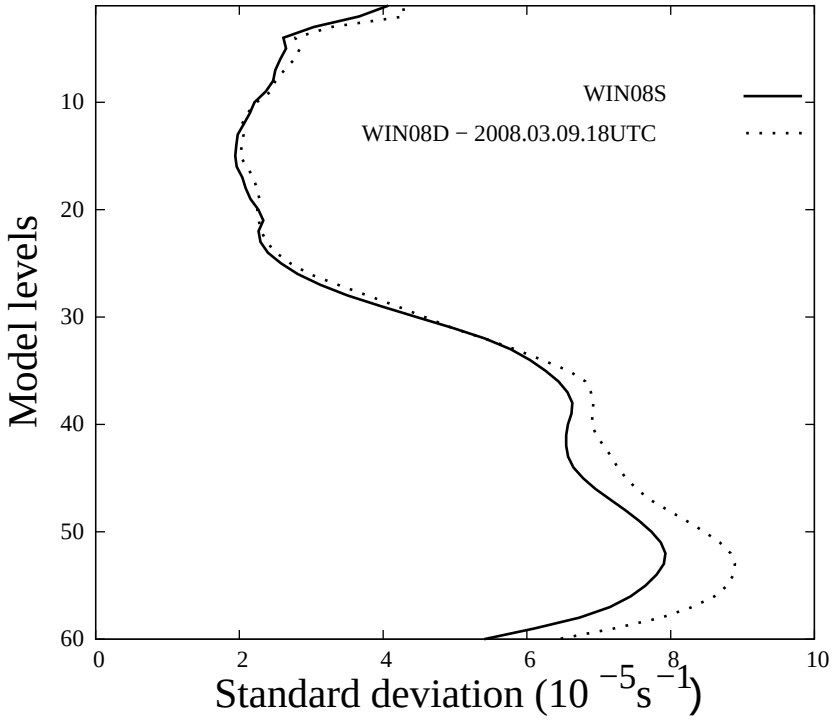


Figure 5.17: Vertical profiles of temporal and spatially averaged standard deviation of 6 h ALADIN/France vorticity estimated forecast errors (issued from the 1200 UTC network and valid at 1800 UTC), used on the 20080310 assimilation at 0000 UTC.

Moreover, it may be underlined that the daily-varying experimental version also has a better forecast quality than the operational version, in a similar way as for the off-line monthly-updated experimental version compared also to the operational version. This implies that the on-line daily-varying covariance approach may be envisaged for operational applications, as a replacement to the current off-line static approach (corresponding to the calibration over September 2007). This would allow both monthly and daily variations of background error covariances to be represented.

While this kind of impact studies could be carried out during other periods, there are also other aspects that could be examined. For instance, the impact of diurnal variations of covariances could be evaluated by using a larger ensemble than in the current study. Moreover, while the present study has focused on temporal variations within the currently operational homogeneous covariance framework, the effect of horizontal

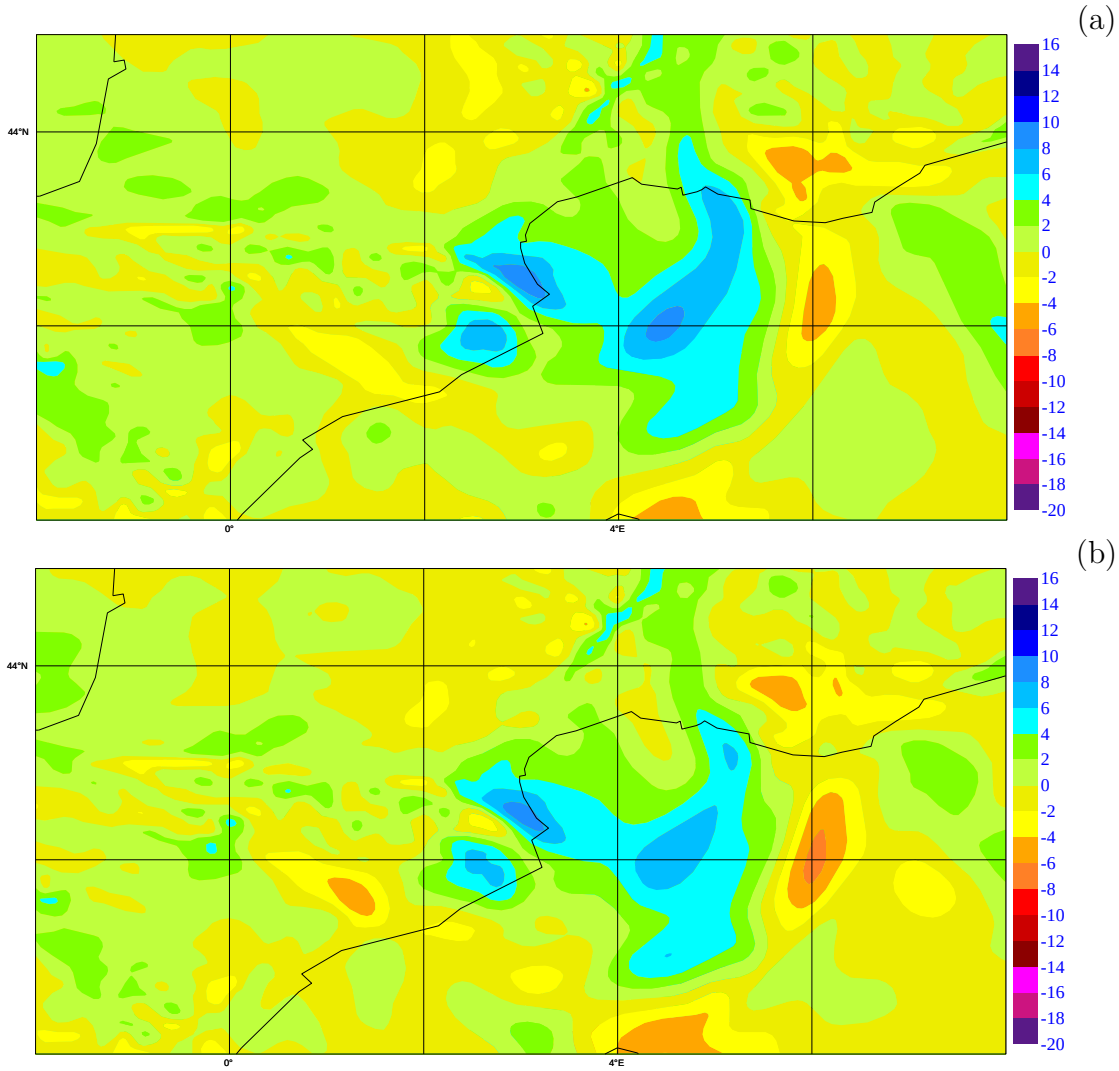


Figure 5.18: Spatial distribution of departures from ECMWF analysis of ALADIN/France 850 hPa zonal wind analysis on the 20080310 at 00UTC, for the experiments: (a) WIN08S and (b) WIN08D. Zoom over the central part of the domain, at the south coast of France

heterogeneities could also be studied in addition.

Chapter 6

Conclusions and future outlook

Time variations of regionally averaged background error estimates have been diagnosed in a robust way, using an off-line ensemble technique. Their impact on the ALADIN/France forecasting system was studied, foreseeing a possible operational implementation of an on-line specification of time-dependent background error covariance estimates. Historically, background error covariances of the ALADIN/France system are estimated using spatial homogeneity and temporal stationarity assumptions. However, while the homogeneity assumption implies the calculation of covariances from a spatial average over the regional computational model domain, this regional average is expected to be prone to temporal variations.

The first part of our study consisted on the revision of the temporal stationarity assumption, by diagnosing time variations of regionally averaged covariances. ALADIN/France background errors were estimated by using an ensemble technique based on perturbations of the data assimilation cycling. In our technique, the perfect model assumption was used and therefore only observations were perturbed explicitly (the background was perturbed implicitly). Time variations of covariance estimates resulted from different time averages of spectral covariances of the simulated background errors. The diagnosis of time variations of covariance estimates, was done by examining their variations with season (winter versus summer), day (in connection with the synoptic situation), and hour (related to the diurnal cycle). Larger error

variances were found when changing from winter to summer; sharper horizontal and broader vertical correlation functions were found to be related to the increased convective activity in summer (when compared to winter conditions). Concerning day-to-day variation of the covariances, it was seen that during a winter season changes were done in accordance with the instability of the situation: error standard deviations were generally much larger and horizontal correlation length scales were much reduced (for humidity and temperature in particular) when passing from an anticyclonic to a low-pressure situation. Besides, these day-to-day changes were sometimes more pronounced than seasonal variations. Finally, it was also found that diurnal variations were significant for the boundary layer during the summer period: error standard deviations tend to be larger, horizontal length scales tend to be smaller and vertical correlation functions tend to be broader for forecasts valid at 18UTC, in accordance with expected effects of convective activity in the afternoon.

In the second part of our work, two particular specifications of the diagnosed covariance estimates have been tested on the ALADIN/France system for a winter period. Their impact was analysed through the quality assessment of the respective forecasts in comparison with the actual operational products, and also through the examination of the relation between the analysis increments and the covariance changes on a case by case situation. In the first specification, the background error covariance estimates were obtained from a one-month average, corresponding to the winter period of study; in the second specification, these estimates were obtained from a one-day sliding average preceding each analysis date, representing the daily covariance variations in addition to monthly variations. It was found that seasonally updated covariances, from an autumnal to a winter period, have a positive impact on the short-range forecast quality of temperature, wind and humidity (up to 24 h for humidity and up to 18 h for temperature and wind). The diagnosis of the covariances for the autumnal and winter periods has shown furthermore that this is connected to an increased analysis fit to observations, which arises from sharper vertical correlations in winter (when compared to autumnal vertical correlations, as

expected from the seasonal variation diagnosis done in the first part of our study). Moreover, it was found that the daily variations have a modest and localized positive impact when compared with the monthly variation impact. Positive effects of daily-varying background error standard deviations were found, in particular for humidity during an anticyclonic situation and for wind (up to 36 h) during a cyclonic period.

A first continuation of this work is thus to implement an on-line daily-varying covariance approach as a replacement of the current off-line static approach. As seen in the first part of the work, robust daily-varying covariance estimates are obtained from an ensemble with a small number of members (6, typically), at a reasonable low cost: there is a significant variability in the statistics structure linked to the meteorological situation over the model domain. This would allow both monthly and daily variations of background error covariances to be represented.

In addition to these extensions of the sparse covariance model, the ensemble simulation technique is also considered to be generalized. This is related in particular to the representation of model error contributions, as a replacement of the current perfect model assumption. This could be achieved for instance by adding random draws of a model error covariances estimate, which can be derived by comparing ensemble assimilation estimates with innovation-based estimates (e.g.Raynaud *et al.* (2011)).

Another aspect which would deserve further studies in the future corresponds to the four-dimensional aspect of the data assimilation problem, for instance within the currently considered 6 hour window. 4D-Var and ensemble techniques may also be useful in this area. For instance, Liu *et al.* (2008) and Buehner *et al.* (2010) recently suggested the possibility to use four-dimensional ensemble trajectories within the minimization, as an alternative to tangent linear and adjoint models. These new applications of ensmeble and variational approaches will certainly continue to be explored in the future.

Appendix A

Objective analysis applications at the Portuguese Meteorological Service

At the Portuguese Meteorological Service several methods of objective analysis have been used for the development of operational applications to support surveillance activities. A short chronologic characterization of such low-cost applications takes place in here as an example of valuable implementations of the first methods of objective analysis. This description only mention activities related to the actual group of NWP, founded in 1993, although some of these local applications have already been discontinued.

In 1994, a successive corrections scheme (see for instance Daley (1991)) was implemented, for a limited area which included the east part of Europe, Iberian Peninsula and the adjacent Atlantic Ocean (Monteiro 1994). The objective analysis built from this scheme was the support for the surface synoptic diagnosis of the mean sea level pressure field done each 6-hour for weather surveillance purposes. It was based on surface conventional data over land and over sea. At running time, the first available ECMWF short-range forecast, valid at the analysis hour (tipically a 6 h or a 12 h long forecast), was used as initial estimate of an iterative scheme that lead to the analysis solution. Its variance was known *a priori*. The influence of each observation information on each gridpoint was given by an analytic formula, as a function of the distance to the gridpoint, after Barnes (1964). This function was then weighted by an *a priori* factor that was a function of the forecast error variance and of the observation instrumental error variance.

Later, in 1997, a limited area upper air univariate optimal interpolation scheme was also created and validated (Monteiro 1997). The initial motivation for its implementation was to provide information for diagnosis purposes on the upper air levels of the troposphere, however it was used only in testing mode. The domain covered by this scheme included the east and north parts of Europe, Iberian Peninsula and the adjacent Atlantic Ocean. It provided the analysis for the parameters of geopotential, temperature, relative humidity and the zonal and meridional wind components at standard pressure levels. As background used the 6 h ECMWF forecasts;

as observational data it used radiosondes over land. The background error estimates were calculated by the observational method (Hollingsworth & Lönnberg 1986) from the innovation statistics (observation - background), during a time period of a few months over the analysis domain, and were modulated using a series of Bessel functions.

In 2001, after the entrance of Portugal to the ALADIN project (http://www.cnrm.meteo.fr/aladin/), the surface analysis scheme CANARI, recently documented by Taillefer (2002), was installed and validated over the Portuguese mainland domain used to run the ALADIN model, as described by Oliveira (2001). This application is based on a multivariate optimal interpolation method which means that several observational parameters are used to perform the analysis of each model field. It provides, in particular, the analysis of the parameters of 2 m temperature, 2 m relative humidity, 10 m wind, mean sea level pressure, sea surface temperature and snow depth. It is assumed that the background error has a Gaussian probability distribution and that the different error fields follow the geostrofic and hydrostatic balances. Moreover, the separability of the horizontal and vertical components of its spatial correlation is assumed. In the implementation done by Oliveira (2001) surface observations over land and over sea were used. As background, an ARPEGE analysis was dynamically adapted to the ALADIN/Portugal geometry (12,7 km). The validation of this application was carried out on an assimilation framework, through the assessment of the impact that the CANARI analysis had on the ALADIN forecasts up to 48 h when used as initial conditions for the ALADIN model integrations. The results did not show improvements on the ALADIN/Portugal quality when compared with a normal dynamical adaptation initialization directly from an ARPEGE analysis, besides a slightly improvement on temperature fields at low levels and at 2 m. Further experiences and studies were then suggested, among then the calibration of background errors, by tunning a statistics stretching factor. In December 2007, after a second implementation effort and validation (carried our under a subjective way on a case studies base), the CANARI application was set up into operational mode for the analysis of 2 m temperature and 2 m relative humidity, according with Lopes (2008).

This application is used now just for surface diagnostic purposes, however, it is expected to have a valuable role near future by providing a surface assimilation once set up a new 3D-var assimilation system to ALADIN/Portugal.

Bibliography

Barnes, S. (1964). A Technique for Maximizing details in Numerical Weather Map Analysis. *J. Appl. Meteor.*, **3**, 395–409.

Belo Pereira, M. & Berre, L. (2006). The Use of an Ensemble Approach to Study the Background Error Covariances in a Global NWP Model. *Monthly Weather Review*, **134**, 2466–2489.

Bergthósson, P. & Döos, B.R. (1955). Numerical Weather Map Analysis. *Tellus*, **7**, 329–340.

Berre, L. (2000). Estimation of Synoptic and Mesoscale Forecast Error Covariances in a Limited-Area Model. *Monthly Weather Review*, **128**, 644–667.

Berre, L. (2001). *REPRÉSENTATIONS DES COVARIANCES SPATIALES DES ERREURS DE PRÉVISION POUR UNE ASSIMILATION VARIATIONNELLE DANS UN MODÈLE ATMOSPHÉRIQUE À AIRE LIMITÉE*. Ph.D. thesis, UNIVERSITÉ TOULOUSE III - PAUL SABATIER.

Berre, L. & Desroziers, G. (2010). Filtering of background error variances and correlations by local spatial averaging: A review. *Monthly Weather Review*, **138**, 3693–3720.

Berre, L., Ştefănescu, S.E. & Belo Pereira, M. (2006). The representation of the analysis effect in three error simulation techniques. *Tellus*, **58A**, 196–209.

Berre, L., Pannekoucke, O., Desroziers, G., Ştefănescu, S.E., Chapnik, B. & Raynaud, L. (2007). A variational assimilation ensemble and the spatial filtering of its error covariances: increase of sample size by local spa-

tial averaging. In *Proceedings of ECMWF workshop on flow-dependent aspects of data assimilation*, 151–168, 11-13 June 2007, available on line at: http://www.ecmwf.int/publications/library/do/references/list/14092007.

Berre, L., Desroziers, G., Raynaud, L., Montroty, R. & Gibier, F. (2009). Consistent operational ensemble variational assimilation. In *Extended abstract of Fifth WMO International Symposium on Data Assimilation*, Paper N. 196.

Bocquet, M. (2009). Introduction aux principes et méthodes de l'assimilation de données en géophysique. Notes de cours du Master M2 OACOS et de l'École Nationale Supérieure des Techniques Avancées ParisTech et l'École des Ponts ParisTech. Paris, France, CEREA, École des Ponts ParisTech, Révision 1.22.

Bouttier, F. & Courtier, P. (1999). Data assimilation concepts and methods. Lecture Series, European Centre for Medium-Range Weather Forecasts, Reading, UK.

Buehner, M., Houtekamer, P.L., Charette, C., Mitchell, H.L. & He, B. (2010). Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. *J. Appl. Meteor.*, **138**, 1550–1566.

Courtier, P., Andersson., E., Heckley, W., Pailleux, J., Vasiljevic, D. & Hamrud, M. (1998). The ECMWF implementation of three dimensional variational assimilation (3D-Var). I: Formulation. *Quartely Journal of the Royal Meteorological Society*, **124**, 1783–1807.

Cressman, G. (1959). An Operational Objective Analysis System. *Monthly Weather Review*, **87**, 367–374.

Daley, R. (1991). *Atmospheric Data Analysis*. Cambridge University Press, 460 pp.

Deckmyn, A. & Berre, L. (2005). A Wavelet Approach to Representing Background Error Covariances in a Limited-Area Model. *Monthly Weather Review*, **133**, 1279–1294.

Derber, J. & Bouttier, F. (1999). A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**, 195–221.

Desroziers, G. (2001). Optimal Estimation in Meteorology. ALATNET Seminar Notes, Gourdon, France.

Desroziers, G., Berre, L., Chapnik, B. & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quartely Journal of the Royal Meteorological Society*, **131**, 3385–3393.

Desroziers, G., Berre, L., Chabot, V. & Chapnik, B. (2009). A Posteriori Diagnostics in an Ensemble of Perturbed Analyses. *Monthly Weather Review*, **137**, 3420–3436, Issue 10.

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation. In *Proceedings of ECMWF seminar on recent developments in data assimilation*, 221–257, Reading, UK.

Fischer, C., Montmerle, T., Berre, L., Auger, L. & Ştefănescu, S.E. (2005). An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. *Quartely Journal of the Royal Meteorological Society*, **131**, 3477–3492.

Fisher, M. (2003). Background error covariance modelling. In *Proceedings of ECMWF seminar on recent developments in data assimilation*, 45–63, Reading, UK.

Gandin, L.S. (1963). *Atmospheric Data Analysis*. Izd. Leningrad (in Russian), English translation from Russian, Israel Program for Scientific Translations, Jerusalem, 1965.

Geleyn, J.F., Bazile, E., Bougeault, P., Déqué, D., Ivanovici, V., Joly, A., Labbé, L., Piédelièvre, J., Piriou, J. & Royer, J. (1995). Atmospheric parametrization schemes in Météo-France's ARPEGE NWP model. In *Proceedings of Seminar on physical parametrizations in numerical models*, 385–402, ECMWF, Reading, UK.

Gustafsson, N., Berre, L., S. Hörnquist, S., Huang, X.Y., Lindskog, M., avascues, Mogensen, K.S. & Thorsteinsson, S. (2001). Three-dimensional variational data assimi-

lation for a limited area model. Part I: General formulation and the background error constraint. *Tellus*, **53**, 425–446, Issue 4.

Hollingsworth, A. & Lönnberg, P. (1986). The statistical strutcture of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136.

Horányi, A., Ihász, I. & Radnóti, G. (1996). ARPEGE/ALADIN: A numerical weather prediction model for central-Europe with the participation of the hungarian meteorological service. *IDŐJÁRÁ*, **100**, 227–301.

Houtekamer, P.L., Lefaivre, L., Derome, J., Ritchie, H. & Mitchell, H.L. (1996). A System Simulation Approach to Ensemble Prediction. *Monthly Weather Review*, **124**, 1225–1242.

Ide, K., Courtier, P., Ghil, M. & Lorenc, A.C. (1997). Unified Notation for Data Assimilation: Operational, Sequential and Variational. *Journal of the Meteorologcial Society of Japan*, **75**, 181–189.

Isaken, L., Fisher, M. & Berner, J. (2007). Use of analyis ensembles in estimating flow-dependent background error variance. In *Proceedings of ECMWF workshop on flow-dependent aspects of data assimilation*.

Le Dimet, F.X. & Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, **38A**, 97–110.

Lindskog, M., Vignes, O., Gustafsson, N., Landelius, T. & Thorsteinsson, S. (2007). Background errors in HIRLAM variational data assimilation. In *Proceedings of ECMWF workshop on flow-dependent aspects of data assimilation*, 113–123, 11-13 June 2007, available on line at: http://www.ecmwf.int/publications/library/do/references/list/14092007.

Liu, C., Xiao, Q. & Wang, B. (2008). An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Monthly Weather Review*, **136**, 3363–3373.

Lopes, N.R.P. (2008). Pré-operacionalização do CANARI CY32T1. Tech. Rep. 3/2008, Centro de Processamento e Previsão Numérica. Instituto de Meteorologia, Lisboa.

Lorenc, A.C. (1986). Analysis methods for numerical weather prediction. *Quartely Journal of the Royal Meteorological Society*, **112**, 1177–1194.

Lorenc, A.C. (2003). The potential of the ensemble Kalman filter for NWP - A comparison with 3D-Var. *Quartely Journal of the Royal Meteorological Society*, **129**, 3183–3203.

Lynch, P. & Huang, X.Y. (1992). Initialization of the HIRLAM Model Using a Digital Filter. *Monthly Weather Review*, **120**, 1019–1034.

Menke, W. (1989). *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, Inc., USA.

Monteiro, M. (1994). Análise Objectiva de Campos Meteorológicos. Método das Correcções Sucessivas. Tech. Rep. 10/94, Divisão de Previsão Numérica, Processamento e Arquivo. Instituto de Meteorologia, Lisboa.

Monteiro, M. (1997). *O Método de Interpolação Óptima em Meteorologia. Aplicação à Anàlise Univariada em Altitude numa Área Limitada*. Dissertação para a obtenção do grau de Mestre em Ciências Geofísicas, Universidade Clássica de Lisboa - Faculdade de Ciências, Lisboa.

Monteiro, M. & Berre, L. (2010). A diagnostic study of time variations of regionally averaged background error covariances. *Journal of Geophysical Research*, **115**.

Oliveira, H.M.G.S. (2001). *Instalação do Programa CANARI e Utilização da sua Análise no Modelo de Área Limitada ALADIN*. Relatório de Estágio Profissionalizante em Ciências Geofísicas - Especialidade: Meteorologia, Universidade de Lisboa - Faculdade de Ciências, Lisboa.

Pannekoucke, O., Berre, L. & Desroziers, G. (2008). Background error correlation length-scales and their sampling statistics. *Quartely Journal of the Royal Meteorological Society*, **133**, 497–508.

Parrish, D.F. & Derber, J.C. (1992). The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, **120**, 1747–1763.

Purser, R., Wu, W.S., Parrish, D.F. & Roberts, N.M. (2003). Numerical aspects of the application of recursive filter to variational statistical analysis, Part II: spatially inhomogeneous and anisotropic general covariances. *Monthly Weather Review*, **131**, 1535–1548.

Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A. & Bouttier, F. (1998a). The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Quartely Journal of the Royal Meteorological Society*, **124**, 1809–1829.

Rabier, F., McNally, A. & Courtier, P. (1998b). Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quartely Journal of the Royal Meteorological Society*, **124**, 1861–1887, Issue 550. July 1998 Part B.

Raynaud, L., Berre, L. & Desroziers, G. (2009). Optimal filtering of ensemble-based background-error variances. *Quartely Journal of the Royal Meteorological Society*, **135A**, 1177–1199.

Raynaud, L., Berre, L. & Desroziers, G. (2011). Accounting for model error in the Météo-France ensemble data assimilation system. *Quartely Journal of the Royal Meteorological Society*, Article first published online : 22 AUG 2011.

Taillefer, F. (2002). CANARI (based on ARPEGE cycle CY25T1 (AL25T1 for ALADIN)). GMAP/CNRM Technical Documentation. Météo-France, Toulouse, France.

Tarantola, A. (1987). *Inverse Problem Theory*. Elsevier, USA.

Weaver, A. & Courtier, P. (2001). Correlation modelling on the sphere using a generalized diffusion equation. *Quartely Journal of the Royal Meteorological Society*, **127**, 1815–1846, Article first published on line : 19 DEC 2006.

Yang, G.Y. & Slingo, J. (2001). The diurnal cycle in the tropics. *Monthly Weather Review*, **129**, 784–801.