UNIVERSITÉ DU QUÉBEC À MONTRÉAL

ÉTUDE SUR L'OBSERVABILITÉ DE L'ATMOSPHÈRE ET L'IMPACT DES OBSERVATIONS SUR LES PRÉVISIONS MÉTÉOROLOGIQUES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN SCIENCES DE L'ENVIRONNEMENT

PAR

CRISTINA LUPU

AOÛT 2010

REMERCIEMENTS

Cette thèse est le résultat de quatre années de travail à l'Université du Québec à Montréal. Je tiens remercier ici très chaleureusement toutes les personnes avec qui j'ai partagé cette période.

En premier lieu, je tiens à exprimer toute ma gratitude au Prof. Pierre Gauthier pour avoir encadré ma thèse. Les innombrables discussions que nous avons eues au cours de cette thèse m'ont permis d'apprécier son expertise en assimilation de données, son enthousiasme communicatif et son ouverture d'esprit. Je le remercie pour ses conseils avisés, pour sa disponibilité permanente et pour ses encouragements dans des moments difficiles tout au long de ces années. Un merci particulier pour avoir soigneusement corrigé mes textes et pour m'avoir appris la rédaction d'un texte scientifique.

Je tiens à remercier également les membres de mon comité de thèse, Messieurs Jean-Pierre Blanchet et Louis Garand, d'avoir contribué à l'encadrement de ma recherche de doctorat. Je tiens d'autre part à remercier les membres de la Division d'assimilation de données et de météorologie satellitaire d'Environnement Canada pour les échanges scientifiques, techniques et informatiques. Stéphane Laroche, Mark Buehner, Monique Tanguay, Jean-François Caron, Ahmed Mahidjiba, Pierre Koclas et Simon Pellerin m'ont apporté une aide précieuse. Merci également aux membres du Service Informatique à Dorval pour leur support et coopération.

Je tiens à remercier également les membres du jury qui ont accepté de réviser ma thèse. Merci au Prof. René Laprise d'avoir accepté de présider le jury de cette thèse, au Dr. Carla Cardinali, au Dr. Mark Buehner, et au Prof. Pierre Gauthier pour avoir accepté d'évaluer mon travail de thèse. Leurs remarques et recommandations m'ont permis d'apporter des améliorations à la version finale de la thèse.

Plus largement, ma reconnaissance va aussi à toute l'équipe du centre ESCER/UQAM: professeurs, collègues, personnel administratif et technique. Une pensée particulière à mon compagnon de bureau et ami Alexandru Stefanof, véritable encyclopédie vivante. Je le remercie pour ses encouragements et pour tout ce que j'ai appris lors de nos conversations quotidiennes. Je remercie aussi Emilia Diaconescu, Dorina Surcel, Cristina Stefanof, Ana

Berbeleac pour le soutien apporté tout au long de ma thèse et pour les bons moments passés ensemble au travail et en dehors.

Un merci tout particulier et chaleureux à mon mari, Gheorghe Milea, qui m'a soutenu toutes ces années et m'a donné la force et la motivation nécessaires pour terminer. Ce travail est aussi le résultat de son soutien et de sa confiance. Merci également à mes chers parents et à mon frère pour leurs encouragements.

TABLE DES MATIERES

LISTE DES FIGURES

LISTE DES TABLEAUX

LISTE DES SIGLES ET ACRONYMES

AIRS	Atmospheric Infra-Red Sounder
AMSU-A/B	Advanced Microwave Sounding Unit (A and B)
AMV	Atmospheric Motion Vectors
ATOVS	Advanced TIROS Operational Vertical Sounder
ATReC	Atlantic THORPEX Regional Campaign
CMC	Centre Météorologique Canadien
DFS	Degree of Freedom of Signal
ECMWF	European Centre for Medium range Weather Forecast
EnKF	Ensemble Kalman Filter
FASTEX	Fronts and Atlantic Storm-Track Experiment
GOES	Geostationary Operational Environmental Satellite
GEM	Global Environnemental Multi-échelle
IASI	Infrared Atmospheric Sounding Interferometer
IC	Information Content
MODIS	MODderate-resolution Imaging Spectroradiometer
MSC	Meteorological Service of Canada
NORPEX	North Pacific Experiment
NMC	National Meteorological Centre
NWP	Numerical Weather Prediction
OSEs	Observing Systems Experiments
RAOB	Radio soundings Observations
RRKF	Reduced-Rank Kalman Filter

SVs	Singular Vectors
SVD	Singular Value Decomposition
T-PARC	THORPEX Pacific-Asia Regional Campaign
UTC	Coordinated universal time
1D/3D/4D-VAR	One/Three/Four-dimensional variational data assimilation scheme

LISTE DES SYMBOLES

$^{\circ}$	Degré
$(.)^{-1}$	Opérateur inverse
$(.)^{*}$	Opérateur adjoint
$(.)^{T}$	Opérateur de transposition
$(.)_0$	Indice relatif au moment initial
$(.)_a$	Indice relatif à l'analyse
$(.)_b$	Indice relatif à l'ébauche
$(.)_o$	Indice relatif aux observations
$\langle .,. \rangle$	Produit scalaire
%	Pour cent
∇	Opérateur gradient d'une fonction
$\nabla\nabla^T$	Opérateur hessien d'une fonction
$\mathbf{A}$	Matrice de covariance d'erreurs d'analyse
$\tilde{\mathbf{A}}$	Matrice de covariance d'erreurs d'analyse *a posteriori*
$\mathbf{B}$	Matrice de covariance d'erreurs d'ébauche du 3D-Var
$\tilde{\mathbf{B}}$	Matrice de covariance d'erreurs d'ébauche *a posteriori*
$\tilde{\mathbf{B}}_x$	Matrice de covariances d'erreurs d'ébauche du 3D-Var adapté
C_1	Coefficient de projection de l'innovation sur la direction de sensibilité
C_2	La norme de la direction sensible dans l'espace d'observations

$\mathbf{d}_b^a$	Écarts de l'analyse à l'ébauche
$\mathbf{d}_a^o$	Différences entre l'analyse et les observations
$\mathbf{d}_b^o$	Différences entre le modèle de prévision et les observations
$\mathbf{D}$	Matrice de covariance de l'innovation
$\tilde{\mathbf{D}}$	Matrice de covariance de l'innovation *a posteriori*
$E[\cdot]$	Espérance mathématique
$\mathbf{f}$	Vecteur direction de sensibilité (dans l'espace de $\delta\mathbf{x}$)
H	Opérateur d'observations
$\mathbf{H}$	Opérateur d'observations linéarisé
$\mathbf{I}$	Matrice identité
J	Fonction coût
J_b	Partie de la fonction coût liée à l'ébauche
J_o	Partie de la fonction coût liée aux observations
$\mathbf{K}$	Matrice de gain
$\tilde{\mathbf{K}}$	Matrice de gain *a posteriori*
L	Longueur du domaine 1D-Var
L_b	Longueur de corrélation
p	Nombre total d'observations
$\mathbf{R}$	Matrice de covariances d'erreurs d'observations
$\tilde{\mathbf{R}}$	Matrice de covariances d'erreurs d'observations *a posteriori*
$\mathbf{S}$	Matrice d'influence

tr	Trace
$\mathbf{v}$	Vecteur direction de sensibilité normalisé (dans l'espace de $\delta\mathbf{x}$)
$\tilde{\mathbf{v}}$	Vecteur direction de sensibilité normalisé (dans l'espace de ξ)
$\mathbf{x}$	Vecteur de l'état de l'atmosphère
$\mathbf{x}_0$	Vecteur d'état de l'atmosphère à l'instant initial
$\mathbf{x}_a$	Vecteur analyse
$\mathbf{x}_a^*$	Vecteur analyse correspondant aux observations perturbées
$\mathbf{x}_b$	Vecteur de l'ébauche
$\mathbf{x}_t$	Vecteur de l'état vrai de l'atmosphère
$\mathbf{X}$	Matrice non aléatoire
$\mathbf{y}$	Vecteur des observations
$\mathbf{y}^*$	Vecteur des observations perturbées
$\mathbf{y}'$	Vecteur d'innovation
α	Amplitude de l'incrément d'analyse dans la direction sensible
$\delta(.)$	Opérateur de perturbation
$\delta\mathbf{x}$	Vecteur d'état incrémental
$\boldsymbol{\varepsilon}_a$	Vecteur des erreurs d'analyse
$\boldsymbol{\varepsilon}_b$	Vecteur des erreurs de l'ébauche
$\boldsymbol{\varepsilon}_m$	Vecteur des erreurs de mesure instrumentale
$\boldsymbol{\varepsilon}_o$	Vecteur des erreurs d'observations
$\boldsymbol{\varepsilon}_r$	Vecteur d'erreur de représentativité

∂	Opérateur de différentiation partielle
η	Coordonnée verticale
λ, φ	Longitude, latitude (rad)
ρ	Coefficient de corrélation
σ^2	Variance d'erreurs de prévision dans la direction de sensibilité
σ_b	Écart type des erreurs de prévision
σ_o	Écart type des erreurs d'observations
$\boldsymbol{\xi}$	Vecteur d'état préconditionné par $\mathbf{B}$
$\tilde{\boldsymbol{\xi}}$	Vecteur d'état préconditionné par $\tilde{\mathbf{B}}$
χ^2	Chi-carré

RÉSUMÉ

L'assimilation de données est une composante essentielle du système de prévision numérique du temps et consiste à trouver un état de l'atmosphère, l'analyse, qui est compatible avec les différentes sources d'observations, la dynamique de l'atmosphère et un état antérieur du modèle. Dans ce processus, il est important de bien caractériser l'erreur associée à chaque source d'information (observations, ébauche) afin de mieux décrire les conditions initiales.

La matrice de covariance des erreurs de prévisions joue un rôle clé dans le processus d'assimilation de données car elle détermine la nature de la correction apportée par l'analyse. Cette matrice étant trop grande pour être représentée explicitement, elle est modélisée sous la forme d'une suite d'opérateurs relativement simples. Les modèles de covariance d'erreurs de prévisions utilisées dans un 3D-Var sont généralement stationnaires et ne considèrent pas des variations dues à la nature de l'écoulement. En présence d'instabilité, une petite erreur dans les conditions initiales connaîtra une croissance rapide. Pour contrôler cette croissance d'erreur à courte échéance, il est nécessaire d'apporter des corrections à l'analyse dans des régions localisées selon une structure spatiale très particulière. L'assimilation adaptative 3D-Var considère une formulation différente des covariances d'erreur de prévision qui permet d'inclure les fonctions de structure basées sur des fonctions de sensibilité *a posteriori* et *a priori* définissant la structure de changements aux conditions initiales qui ont le plus d'impact sur une prévision d'échéance donnée. Dans le cadre de cette thèse, des fonctions de sensibilité sont introduites comme fonctions de structure dans l'assimilation 3D-Var. La définition d'une fonction de structure appropriée pour un système d'assimilation vise à simultanément concorder aux observations disponibles et améliorer la qualité des prévisions. L'observabilité des fonctions de structure par les observations est tout d'abord présentée et analysée dans le cadre plus simple d'une analyse variationnelle 1D (1D-Var) pour être ensuite introduite dans le 3D-Var d'Environnement Canada. L'amplitude de la correction est caractérisée par un seul paramètre défini par l'ensemble des observations disponibles. Les résultats montrent que si le rapport entre l'amplitude du signal et l'erreur d'observation est très faible, les observations ne sont pas en mesure de détecter les instabilités atmosphériques qui peuvent croître très rapidement. Dans cette perspective, l'assimilation pourra seulement extraire l'information contenue dans les structures atmosphériques déjà évoluées.

Dans un deuxième temps, nous présentons une nouvelle méthode permettant d'estimer l'impact des observations dans les analyses 3D/4D-Var basé sur *le Degrees of freedom for signal*. Le contenu en informations des observations est calculé en employant les statistiques *a posteriori*, à partir des écarts des observations à l'ébauche et à l'analyse. Les résultats montrent que le *DFS* estimé en utilisant les statistiques *a posteriori* est identique avec celui obtenu à partir des statistiques a priori. Ce diagnostic permet de comparer l'importance de différents types d'observations pour les expériences d'assimilation 3D et 4D-Var incluant toutes les observations assimilées opérationnellement. En particulier, cette étude s'intéresse à l'évaluation du réseau canadien d'observations et il est appliqué aux Observing System Experiments (OSEs) effectués à Environnement Canada pour les mois de janvier et février 2007.

Mots-clés: Assimilation de données, Dépendance à l'écoulement, 3D-Var, DFS, OSEs.

CHAPITRE I

INTRODUCTION

L'assimilation de données consiste à trouver un état de l'atmosphère, l'analyse, qui soit compatible avec les différentes sources d'observations, la dynamique de l'atmosphère et un état antérieur du modèle. Dans le processus d'assimilation de données, il est fondamental de bien caractériser l'erreur associée à chaque source d'information (ébauche, observations) afin de mieux caractériser les éléments du système d'analyse qui peuvent influencer la qualité des prévisions. Étant donné que l'état vrai de l'atmosphère n'est pas accessible, il est impossible d'obtenir directement des échantillons des erreurs d'ébauche et d'observation et ces erreurs doivent donc être estimées *a priori* à partir de statistiques. L'étude des innovations d'un jeu d'observations suffisamment grand et dense fournit des statistiques permettant de construire les matrices de covariance d'erreurs (Hollingsworth et Lönnberg, 1986). Toutefois, l'hypothèse que les erreurs d'observation ne sont pas corrélées spatialement est fondamentale car seule cette hypothèse permet de séparer l'information provenant de la matrice de covariances d'erreur d'observation $\mathbf{R}$ et d'ébauche $\mathbf{B}$.

D'autres méthodes permettent d'estimer les covariances d'erreur d'ébauche, par exemple la méthode du National Meteorological Centre (NMC) (Parrish et Derber, 1992). Cette méthode consiste à estimer les covariances d'erreur de prévision à partir des différences entre deux prévisions valides au même moment, mais à des échéances différentes. À Environnement Canada, les différences entre les prévisions à une échéance de 48-h et à 24-h respectivement sur une période de deux ou trois mois sont employées (Gauthier *et al.*, 1998).

Dans le 3D-Var, des hypothèses telles que la stationnarité, l'homogénéité et l'isotropie sont souvent imposées pour simplifier la modélisation des covariances d'erreur de prévision. Les systèmes d'assimilation de données doivent être capables de représenter, par exemple, les structures baroclines pour être en mesure de décrire correctement les phénomènes pour lesquels une petite erreur d'analyse conduit à des erreurs majeures de prévision. Dans ce

cadre, des recherches ont été réalisées pour améliorer l'estimation et la représentation des covariances d'erreurs de prévision et prendre en compte les variations dues à la nature de l'écoulement dans la matrice de covariance d'erreur de prévision.

Des études de sensibilités *a posteriori* (Rabier *et al.*, 1996; Klinker *et al.*, 1998) et *a priori* (Hello *et al.*, 2000) basées sur l'utilisation du modèle adjoint (Le Dimet et Talagrand, 1986) ont été développées afin de contrôler les erreurs sur les conditions initiales susceptibles de croître rapidement. Les fonctions de sensibilité *a posteriori* permettent de caractériser des corrections aux conditions initiales qui peuvent réduire significativement l'erreur de prévision à une échéance donnée (typiquement 24 ou 48 heures). L'erreur de prévision est définie par l'écart à une analyse de vérification et la fonction de sensibilité ne peut donc être calculée qu'*a posteriori*. Les fonctions de sensibilité *a priori* sont utilisées pour réduire la part des erreurs de prévision due aux conditions initiales sans attendre l'analyse de vérification. Pour évaluer la sensibilité d'une prévision aux changements aux conditions initiales, la fonction de sensibilité *a priori* est définie par rapport à un aspect particulier de la prévision à un temps ultérieur et on détermine ensuite la correction aux conditions initiales qui a le plus d'influence sur cet aspect particulier de la prévision. Cette idée peut être généralisée et l'emploi du modèle adjoint permet aussi le calcul des vecteurs singuliers qui représentent le sous-espace dans lequel des changements aux conditions initiales subissent le plus d'amplification sur une période donnée. Les vecteurs singuliers et les études de sensibilité aux conditions initiales permettent d'identifier pour chaque situation météorologique, les régions sensibles, où la moindre erreur sur les conditions initiales peut croître très rapidement et avoir un effet considérable sur la qualité de la prévision. Dans ces régions, il semblait naturel d'ajouter de nouvelles observations (observations ciblées) afin d'augmenter la précision de l'analyse et réduire de manière notable l'erreur de prévision. La technique de l'observation adaptative a été testée lors de différentes campagnes de mesures avec ciblage des observations : *Fronts and Atlantic Storm-Track Experiment* (FASTEX, Joly *et al.*, 1999), *North Pacific Experiment* (NORPEX, Langland *et al.*, 1999), 2003 Atlantic THORPEX Regional Campaign (ATReC, Langland, 2005) et tout récemment 2008 THORPEX Pacific-Asia Regional Campaign (T-PARC). Les résultats de ces expériences ont montré que l'impact de données ciblées n'est perçu que si de grandes régions sont échantillonnées et les systèmes d'assimilation variationnelle de données 3D-Var et même le

4D-Var ne peuvent extraire toute l'information apportée par ces données ciblées. La valeur des campagnes de ciblage des observations doit être reconsidérée dans le contexte où le déploiement des observations ciblées n'a qu'un impact faible mais néanmoins significatif sur la réduction des erreurs de prévision. Toutefois, l'information contenue dans les sensibilités devrait être considérée pour le déploiement des observations.

Hello et Bouttier (2001) ont introduit localement dans le 3D-Var les fonctions de sensibilité *a priori* comme fonction de structure. Leur méthode consistait à représenter les covariances d'erreur pour une sous-partie de la matrice **B** en suivant la procédure développée pour le filtre de Kalman de rang réduit (RRKF) par Fisher (1998). Essentiellement, cette approche traite différemment la partie de la variable de contrôle qui se projette dans l'espace décrivant la composante sensible. Ceci rejoint la formulation de modèles de covariance d'erreur de prévision basés sur les vecteurs singuliers (Fisher et Andersson, 2001; Buehner, 2004). Présentement, des approches hybrides considèrent dans l'assimilation variationnelle 4D-Var des modèles de covariance d'erreurs de prévision dépendant de l'écoulement obtenus par des méthodes d'ensembles (Buehner *et al.*, 2010 a,b; Berre *et al.*, 2009)

Au cours des dernières années, d'importants changements ont été introduits dans les systèmes d'assimilation de données des centres opérationnels de prévision numérique du temps afin de permettre une meilleure utilisation des données provenant d'un nombre croissant d'instruments embarqués sur satellites. Actuellement, les données satellitaires constituent la plus importante source de données utilisée pour l'analyse météorologique en vue de la prévision numérique du temps. Par exemple, à l'ECMWF les radiances satellitaires représentent 90% du nombre total d'observations assimilées. Le monitoring et le contrôle de données sont abordés par le contrôle systématique des écarts entre le modèle de prévision et les observations et entre l'analyse et les observations.

Les erreurs d'observation sont souvent censées être non corrélées et la matrice de covariance d'erreurs d'observations **R** est alors diagonale dans les schémas d'assimilation opérationnels. Si cette hypothèse paraît raisonnable pour des observations mesurées par des instruments différents, elle est moins évidente quand un jeu d'observations est obtenu par le même instrument de mesure ou quand une série temporelle de mesures d'une même station est utilisée dans un 4D-Var (Järvinen *et al.*, 1999). L'erreur d'observation renferme l'erreur

de mesure instrumentale, l'erreur due aux imperfections dans le modèle de transfert radiatif et les erreurs de représentativité. La variabilité sous maille de la variable observée, qui n'est pas résolue par le modèle, mais qui peut être mesurée par l'instrument, est à l'origine de l'erreur de représentativité. Dans le cas d'observations satellitaires, l'erreur d'observation est corrélée spatialement, temporellement, et spectralement puisqu'elles sont prises avec le même instrument et dans les mêmes conditions. Liu et Rabier (2002) ont montré que l'omission des corrélations d'erreur d'observation dans la matrice **R** pourrait être compensée par une augmentation des valeurs de l'écart type d'erreur d'observation ou encore en réduisant le volume de données.

L'estimation des erreurs d'observations peut être faite à partir de comparaisons entre les observations ou avec la méthode basée sur les innovations (Hollingsworth et Lönnberg, 1986). Plus récemment, Desroziers *et al.* (2005) ont développé des nouvelles approches qui permettent d'estimer la matrice de covariance d'erreur d'observations et la matrice de covariance d'erreur d'ébauche dans l'espace des observations à partir des écarts des observations à l'ébauche et à l'analyse.

L'estimation de l'impact des observations dans l'analyse et dans les prévisions à courte échéance est un enjeu important dans les centres de prévision numérique du temps. Différentes méthodes ont été développées et examinées dernièrement afin de quantifier le contenu en information des observations dans les analyses et dans les prévisions à courte échéance. Le *Degrees of freedom for signal* (ou *DFS*) est un diagnostique permettant d'évaluer le contenu en information apportée par les différents types d'observations dans l'analyse (Rodgers, 2000). En particulier, les observations issues de sondeurs infrarouges hyperspectraux dont les premiers embarqués sont AIRS (Atmospheric Infrared Sounder) et IASI (Infrared Atmospheric Sounding Interferometer) apportent un précieux contenu en information sur l'état de l'atmosphère mais comprennent une grande quantité des données. Un enjeu important est donc de mettre en place de méthodes de sélection de données permettant de faire un bon compromis entre le volume de données, la qualité de l'analyse et l'impact sur les prévisions. Un exemple de la façon dont le *DFS* a été appliqué pour la sélection de canaux de données simulées IASI est présenté dans Rabier *et al.* (2002). À ECMWF, Cardinali *et al.* (2004) ont utilisé la matrice d'influence pour évaluer la sensibilité

de l'analyse 4D-Var aux observations. Le contenu en information pour les différentes familles d'observations dans des régions géographiques diverses est estimé à partir des éléments diagonaux de la matrice d'influence. Fisher (2003) a examiné la possibilité d'évaluer le *DFS* exprimé comme la trace de la matrice d'influence pour le 4D-Var de l'ECMWF. La méthode de Girard (1987) permet également d'exprimer la trace d'une matrice très large par une procédure de randomisation. Une telle procédure a été également employée par Chapnik *et al.* (2006) à Météo-France.

L'impact des observations sur l'erreur de prévision à courte échéance est évalué avec la méthode de calcul de sensibilité aux observations (Baker et Daley, 2000; Doerenbecher et Bergot, 2001; Langland et Baker, 2004; Zhu et Gelaro, 2008; Cardinali, 2009). D'autre part, les Observing System Experiments (OSEs) sont également utilisées dans les centres de prévision numérique du temps pour diagnostiquer l'impact des observations dans un système de prévision. Ceci est obtenu à partir des OSEs dans lesquels certains types de données sont systématiquement ajoutés ou retirés du système d'assimilation (Kelly *et al.*, 2007). Il est à noter que l'impact de chaque type d'observation dépend de la méthode d'assimilation de données, du modèle de prévision et des erreurs des observations et de l'ébauche considérée dans le processus d'assimilation.

Le travail effectué dans cette thèse est présenté en cinq chapitres dont trois correspondent à des articles écrits en anglais qui seront soumis pour publication. Le présent chapitre introduit la problématique et le contexte général de la recherche et résume les objectifs et le plan de l'étude.

Le deuxième chapitre est constitué d'un article intitulé « *Observability of flow dependent structure functions for use in data assimilation* » par Cristina Lupu et Pierre Gauthier, et il sera soumis pour publication dans *Monthly Weather Review*. Tout d'abord, le système d'assimilation de données doit être capable de prendre en compte les structures baroclines pour être en mesure de décrire correctement les phénomènes pour lesquels une petite erreur d'analyse conduit à des erreurs majeures de prévision. Hello et Bouttier (2001) ont proposé ce qu'il est convenu d'appeler le 3D-Var adapté, basé sur une formulation différente des covariances d'erreur de prévision qui permet d'inclure l'information sur les instabilités atmosphériques contenue dans une fonction de sensibilité. Une variante du 3D-Var adapté a

été introduite par Lupu (2006) et considère que le modèle de covariance d'erreur de prévision du 3D-Var adapté inclut la composante sensible tout en retenant le modèle plus conventionnel du 3D-Var lorsque cette composante dite sensible est peu importante. En conséquence, le premier point vise à examiner des formulations différentes des covariances d'erreur de prévision, qui permettraient d'inclure à même le processus d'assimilation l'information sur les instabilités atmosphériques contenue dans une fonction de sensibilité. A ce titre, il faut réaliser que la caractérisation de ces structures instables dépend du modèle utilisé pour les générer et de la façon dont on définit la mesure de l'erreur de prévision. Dans un premier temps, nous présentons brièvement l'algorithme 3D-Var et sa version 3D-Var *adapté* proposé et validé par Lupu (2006). Dans un deuxième temps, le problème d'assimilation est formulé dans le sous-espace engendré par la fonction de sensibilité. Ceci nous conduira à la question d'observabilité des fonctions de structure qui sera formulée et discutée dans un contexte 1D-Var avec quelques exemples simples. Étant donné que les fonctions de sensibilité *a posteriori* définissent les changements aux conditions initiales qui réduisent significativement l'erreur de prévision à une échéance donnée, il parait logique de les inclure dans la matrice de covariance d'erreur de prévision du 3D-Var. Dans ce cas, l'amplitude de cette correction est déterminée en s'ajustant aux observations disponibles. Nous présentons différentes définitions des fonctions sensibles *a posteriori* qui seront introduites comme fonctions de structure dans le modèle de covariance du 3D-Var. Nous allons examiner si le 3D-Var adapté permet d'améliorer l'ajustement de l'analyse aux observations tout en ayant un impact sur la qualité de la prévision comparable à ce qui est obtenu avec les analyses de sensibilité *a posteriori*. Enfin, pour valider la méthode, l'incrément de l'analyse du 3D-Var normalisé est introduit comme fonction de structure dans le modèle de covariance du 3D-Var qui inclut cette seule composante sensible. Ceci a permis de vérifier dans ce cas que les observations permettent d'estimer correctement l'amplitude de l'incrément du 3D-Var. Par contre, tel n'est pas le cas si on utilise les analyses *a posteriori*. Ceci nous a conduit à définir la notion d'observabilité d'une fonction sensible qui permet d'établir si les observations sont en mesure de détecter une telle structure.

Les bases du troisième chapitre sont nées des questions qui ont émergé des études d'impact des observations dans les analyses et dans les prévisions. Ce chapitre est constitué d'un article intitulé « *Evaluation of the impact of observations on analyses in 3D and 4D-Var*

based on information content » par Cristina Lupu, Pierre Gauthier et Stéphane Laroche, et il sera soumis pour publication dans *Monthly Weather Review*. Dans ce chapitre, nous nous intéressons à évaluer l'impact des observations dans les analyses 3D and 4D-Var (Gauthier *et al.*, 1999, 2007). Nous proposons une nouvelle méthode qui permet l'estimation du *DFS* à partir des écarts des observations à l'ébauche et à l'analyse. Cette méthode sera illustrée et validée dans le cadre plus simple d'une analyse variationnelle 1D (1D-Var) et ensuite appliquée pour évaluer le contenu en informations des observations assimilées avec le 3D et 4D-Var.

Dans la suite de cette étude, le quatrième chapitre se concentre sur l'évaluation du *DFS* pour différentes familles d'observations à l'aide d'OSE (Observing System Experiments). Il est constitué d'un article intitulé « *Assessment of the impact of observations on analyses derived from observing systems experiments* » par Cristina Lupu, Pierre Gauthier et Stéphane Laroche, et il sera soumis pour publication dans *Monthly Weather Review*. Après une brève description des expériences OSEs réalisées par Laroche et Sarazzin (2010 a, b), nous allons mesurer l'impact des différents types de données dans les analyses 3D et 4D-Var en évaluant le contenu en information des observations sur différentes régions sur Amérique du Nord. Cette étude est complémentaire à celle de Laroche et Sarazzin (2010 a, b) qui ont réalisé les OSEs ayant servi de base à notre étude.

Le chapitre V résumera les principaux résultats obtenus dans ces études et quelques perspectives de recherche ouvertes par ce travail seront proposées.

BIBLIOGRAPHIE

Baker N. L. and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptative observation-targeting problem. *Q. J. R. Meteorol. Soc.,* **126**, 1431-1454.

Berre, L., G. Desroziers, L. Raynaud, R. Montroty and F. Gibier, 2009: Consistent operational ensemble variational assimilation, *Procedings of the CAWCR Workshop on Ensemble Prediction and Data Assimilation*, Melbourne, Australia.

Buehner, M., 2004: Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.*, **128**, 1-31.

_____, P. L. Houtekamer, C. Charette, H. L. Mitchell and B. He, 2010a: Intercomparaison of variational data assimilation and Ensemble Kalman filter for global deterministic NWP. Part I: Description and single observation experiments, *Mon. Weather Rev,* **138**, 1550-1566.

_____, P. L. Houtekamer, C. Charette, H. L. Mitchell and B. He, 2010b: Intercomparaison of variational data assimilation and Ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations, *Mon. Weather Rev,* **138**, 1567-1586.

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

——, 2009: Monitoring the observation impact on the short-range forecast. *Q. J. R. Meteorol. Soc.,* **135**, 239-250.

Chapnik B., G. Desroziers, F. Rabier and O. Talagrand, 2006: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Q. J. R. Meteorol. Soc.,* **132**, 543–565.

Desroziers G., L. Berre, B. Chapnik and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.,* **131**, 3385–3396.

Doerenbecher, A. and T. Bergot, 2001: Sensitivity to observations applied to FASTEX cases. *Nonlinear Processes in Geophysics*, **8** (6), 467-481.

Fisher, M., 1998: Development of a simplified Kalman Filter. *ECMWF Research Dept. Technical Memorandum,* no. **260**, 16 pp.

_____, and E. Andersson, 2001: Developments in 4D-Var and Kalman filtering. *ECMWF Research Dept. Technical Memorandum,* no. **347**.

_____, 2003: Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *Tech. Rep.*, **397**, ECMWF, 20pp.

Gauthier, P., M. Buehner and L. Fillion, 1998: Background-error statistics modelling in a 3D variational data assimilation scheme. Proceedings of the *ECMWF workshop on diagnosis of data assimilation systems*, Reading, UK, 131-145.

——, M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau, 2007: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Weather Rev.,* **135,** 2339–2354.

Girard, D., 1987: A fast Monte Carlo cross-validation procedure for large least squares problems with noisy data. *Tech. Rep.* **687-M**, IMAG, Grenoble, France, 22pp.

Hello, G., F. Lalaurette and J.-N. Thépaut, 2000: Combined use of sensitivity information and observations to improve meteorological forecasts: A feasibility study applied to the 'Christmas storm' case. *Q. J. R. Meteorol. Soc.*, **126**, 621-647.

—— and F. Bouttier, 2001: Using adjoint sensitivity as a local structure function in variational data assimilation. *Nonlinear Processes in Geophysics*, **8**, 347-355.

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

Joly, A., K. A. Browning, P. Bessemoulin, J. P. Cammas, G. Caniaux, J. P. Chalon, S. A. Clough, R. Dirks, K. A. Emanuel, L. Eymard, F. Lalaurette, R. Gall, T. D. Hewson, P. H. Hildebrand, D. Jorgensen, R. H. Langland, Y. Lemaître, P. Mascart, J. A. Moore, P. O. G. Persson, F. Roux, M. A. Shapiro, C. Snyder, Z. Toth and R. M. Wakimoto, 1999: Overview of the field phase of the Fronts and Atlantic Storm Track Experiment (FASTEX) project. *Q. J. R. Meteorol. Soc.*, **125**, 3131-3164.

Järvinen, H., E. Andersson and F. Bouttier, 1999: Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus*, **51A**, 468-487.

Kelly, G., J. N. Thépaut, R. Buizza and C. Cardinali, 2007: The value of observations. I: Data denial experiments for the Atlantic and the Pacific. *Q. J. R. Meteorol. Soc.*, **133**, 1803-1815.

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

Langland R. H., Z. Toth, R. Gelaro, I. Szynyogh, M. A. Shapiro, S. J. Majumdar, R. Morss, G. D. Rohaly, C. Velden, N. Bonds and C. H. Bishop, 1999: The North Pacific

Experiment (Norpex-98): Targeted observations for improved North American Weather Forecasts. *Bull. Am. Meteorol. Soc.*, **80**, 1363-1384.

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

——, 2005: Observation impact during the North-Atlantic TreC-2003. *Mon. Weather Rev.,* **133**, 2297-2309.

Laroche, S. and R. Sarrazin, 2010a: Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part I: contribution of radiosonde, aircraft and satellite data. *Atmos.-Ocean*, **48**, 10-25.

——, and R. Sarrazin, 2010b: Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part II: Sensitivity experiments. *Atmos.-Ocean*, **48**, 26-38.

Le Dimet F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97-110.

Liu, Z. Q. and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q. J. R. Meteorol. Soc*, **128**, 1367-1386.

Lupu, C., 2006: Impact d'un modèle de covariance d'erreur de prévision basé sur les fonctions de sensibilité dans un 3D-Var. *Mémoire de maîtrise en sciences de l'atmosphère, UQAM*, 138 pp.

Parrish, D. F. and J. C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Weather Rev.,* **120**, 1747-1763

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

_____, N. Fourrié, D. Chafai and P. Prunet, 2002: Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Q. J. R. Meteor. Soc.*, 128, 1011-1027.

Rodgers, C., 2000: Inverse Methods for Atmospheric Sounding Theory and Practice. World Scientific Publishing, London, 256 pp.

Zhu Y. and R. Gelaro 2008: Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Mon. Weather Rev.* **136**, 335–351.

CHAPITRE II

OBSERVABILITÉ DES FONCTIONS DE STRUCTURE DÉPENDANTES DE L'ÉCOULEMENT POUR UTILISATION EN ASSIMILATION DE DONNÉES

Ce chapitre, rédigé en anglais, est présenté sous la forme d'un article qui sera soumis pour publication dans la revue *Monthly Weather Review*.

L'assimilation de données joue un rôle important dans l'analyse des données atmosphériques, en particulier pour la prévision numérique du temps. Ce chapitre touche un aspect important de la méthodologie d'assimilation de données qui vise à corriger un état de référence en s'ajustant aux observations en tenant compte d'éléments dynamiques qui permettent le développement de systèmes météorologiques. Conséquemment, il est important d'une part de bien caractériser les structures atmosphériques qui peuvent conduire à une croissance de l'erreur de prévision et d'autre part d'incorporer cette information comme fonction de structure dans le système d'assimilation de données. La question d'observabilité des fonctions de structure sera tout d'abord formulée et discutée dans un contexte 1D-Var avec quelques exemples simples, pour être ensuite discutée dans le cadre du 3D-Var d'Environnement Canada.

OBSERVABILITY OF FLOW DEPENDENT STRUCTURE FUNCTIONS FOR USE IN DATA ASSIMILATION

by

Cristina Lupu [*,1] and Pierre Gauthier [1]

[1] *Department of Earth and Atmospheric Sciences*
Université du Québec à Montréal (UQAM)
Montréal (Québec), CANADA

Submitted to *Monthly Weather Review*

16 March 2010

[*] Corresponding author: *Department of Earth and Atmospheric Sciences*
Université du Québec à Montréal (UQAM)
P.O. Box 8888, Succ. Centre-ville
Montréal, Québec
CANADA H3C 3P8

Abstract

One of the objectives of data assimilation is to produce initial conditions that will improve the quality of forecasts. Studies on singular vectors and sensitivity studies have shown that small changes to the initial conditions can sometimes lead to exponential error growth. This has motivated research to include flow-dependent structures within the assimilation that would have the characteristics to correctly predict the growth or decay of meteorological systems. This relates to the characterization of precursors to atmospheric instability. In this paper, the observability of such structures by observations is discussed. Several studies have shown that deploying observations over regions where changes in the initial conditions may impact the forecast the most do not lead to the expected benefit. In this paper, it is shown that given the small magnitude of the signal to be detected, it is important to take into account the accuracy of the observations. If the signal-to-noise ratio is too low, observations cannot detect and characterize precursors to forecast error growths. From that perspective, the assimilation only has the possibility to extract information about evolved structures of error growth. Experiments with a simple 1D-Var system are presented and then, with an adapted 3D-Var system with different sensitivity structure functions is used. The results have been obtained by adapting the variational assimilation system of Environment Canada.

Keywords: 3D-Var, flow-dependent structure functions, observability.

2.1 INTRODUCTION

The accuracy of analyses produced by data assimilation systems depends on the precision of background and observation error covariances specified as input. The modelling and estimation of these covariances is critical for any data assimilation system in the context of numerical weather prediction (NWP). Algorithms like the three-dimensional variational data assimilation (3D-Var) produce analyses by blending together observations near the analysis time with a background state provided by a short-term numerical weather prediction. In this case, the background-error statistics are taken to be stationary and do not reflect the flow-dependency of error growth that depends on the particular meteorological situation. Flow-dependent covariances can be obtained from approximate forms of the Kalman filter like the ensemble Kalman filter (Evensen, 1994; Houtekamer and Mitchell, 2009).

Instabilities in atmospheric flows can be triggered by small perturbations to initial conditions and these can be characterized using adjoint methods that enable to trace back the source of errors in a forecast to errors in the analysis. Lacarra and Talagrand (1988) showed that it is possible to characterize the structure of perturbations to the initial conditions that would lead to the most significant growth over a finite period of time. Those correspond to the so-called *singular vectors* that define the unstable subspace containing those perturbations that will experience the most significant error growth. This has been the foundation of the design of ensemble prediction systems that aim to determine how errors in the analysis and the model will lead to forecast errors in the medium -range (Molteni *et al.*, 1996; Buizza *et al.*, 2007).

Since it is possible to characterize those regions where perturbations in the analysis can lead to important error growth, the next logical step was to use this information to deploy observations in those areas where a reduction in the analysis error could lead to the most important reduction of the forecast error. This is the basis of targeting methods, which use information from singular vector or sensitivity gradients to plan the deployment of adaptive observations. The FASTEX campaign (Joly *et al.*, 1999) was the first to test targeting methods and observations were deployed according to *sensitivity gradients*, that correspond

to a single singular vector. Other campaigns followed like NORPEX (Langland *et al.*, 1999), the 2003 Atlantic THORPEX Regional Campaign (ATReC) (Petersen and Thorpe, 2007; Langland, 2005) and recently, the 2008 THORPEX Pacific-Asia Regional Campaign (T-PARC). From all those campaigns, the conclusions are that the impact of observations deployed over *sensitive areas* in the extra-Tropics identified from singular vectors is, on average, about twice that of any other single observation, but the overall impact is small because of the large volume of data now assimilated (Langland, 2006; Kelly *et al.*, 2008; Buizza *et al.*, 2008; Cardinali *et al.*, 2008). These results bring us to reconsider the value of expensive observation campaigns for the sole purpose of assessing if targeted observations do lead to significant reduction of the forecast error. The current wisdom is that, if observations are to be deployed, it is then appropriate to take into account sensitivity information to do it. Particularly, this may be valuable for adaptive data selection for satellite data. Currently, due to limitations in the assimilation systems, a small fraction of the incoming volume of satellite data can be assimilated (Liu and Rabier, 2001). Adaptive data selection is now being considered to assess whether this results in improvements in the quality of forecasts.

Fisher and Andersson (2001) have proposed a reduced -rank Kalman filter (RRKF) that restricts the evolution of the forecast error covariances within an unstable subspace spanned by singular vectors. Their experimentation was thorough and went all the way to include the RRKF to provide the background error covariances for the ECMWF 4D-Var assimilation. This was found to lead to a positive but small impact on the resulting forecasts, which was not deemed significant enough to implement this approach in the ECMWF operational suite. Currently, hybrid approaches have been proposed in which ensemble methods are used to define a subspace that is appropriate to describe the evolved background error covariances (Buehner *et al.*, 2010a,b; Berre *et al.*, 2009). The preliminary results are very positive and this has sparked a renewed interest to include flow dependent background error covariances when cycling a 4D-Var assimilation system.

This paper has the objective to investigate some issues associated with the use of sensitivity information in the representation of background error covariances with a 3D-Var assimilation system. Hello and Bouttier (2001) did propose an approach through which *a priori* sensitivity information from a single singular vector was included within the

background error covariance matrix, denoted by $\mathbf{B}$. Their approach is called an adapted 3D-Var as it includes some flow dependency. The *a priori* sensitivity was used to deploy targeted data during the FASTEX campaign (Hello *et al.*, 2000). In the present paper, a variant of their algorithm is presented and tested both in a simple 1D-Var context and in 3D-Var. The rationale on which this study is based is the following.

The 24 or 48-hr forecast error can be evaluated by comparing with respect to a verifying analysis and the adjoint of the forecast model can be used to define the change in initial conditions that would reduce the forecast error. This is referred to as *key analysis errors*, a term coined by Klinker *et al.* (1998) and has been the object of several studies afterwards, (Laroche *et al.*, 2002; Langland *et al.*, 2002; Caron *et al.*, 2007a). This will be referred to as an *a posteriori* sensitivity function because it can only be obtained as a diagnostic of the origin of forecast error. In addition, a *priori* structure functions defined either as leading singular vectors (SVs) or from the gradient sensitivity vector method (Hello *et al.*, 2000) are tools that have been widely applied in sensitivity studies, particularly for the development of targeting techniques. In the gradient sensitivity vector method, the cost function can be defined with respect to a particular aspect of the forecast at a later time and then find out what are the changes to the initial conditions that will impact the most the forecast error growth. For example, taking the average of surface pressure of a 24-h forecast over an area of interest, one can then identify areas where changes in the current analysis could have a significant impact as defined by the sensitivity cost function. In Hello *et al.* (2000), this has been used to identify those regions where small changes to the initial conditions can be expected to lead to substantial changes in the forecast.

If *a posteriori* key analyses, as proposed by Klinker *et al.* (1998), do result in a dramatic reduction of forecast error, it would make sense to use those as structure functions within the $\mathbf{B}$ matrix so that observations would be used to define its amplitude. What was expected is that the amplitude of the key analysis would be recovered. However, this is not what happened. On second thought, the signal that was to be recovered being very small, this raised the question whether those structures could be detected at all by the observations, which contain some amount of observation error.

The paper is organized as follows. Section 2.2 briefly presents the formulation of the variational 3D-Var data assimilation system and its *adapted* 3D-Var version. Section 2.3 presents the assimilation in the subspace spanned by a single sensitive direction. A particular point concerns the observability of a structure function defined from *a posteriori* sensitivity. Results with a simple 1D-Var model are presented in section 2.4 to illustrate the new approach. Section 2.5 introduces different *a posteriori* sensitivity structures chosen for this work and results based on *adapted* 3D-Var experiments are described in section 2.5. Finally, section 2.6 summarizes the results and presents some conclusions.

2.2 FLOW-DEPENDENT STRUCTURE FUNCTIONS IN 3D-VAR: THE ADAPTED 3D-VAR

When representing the background error covariance matrix $\mathbf{B}$ in a subspace of low dimension with respect to that of the control variable, a regularization term can be added based on the usual 3D-Var covariances with homogenous and isotropic correlations. This can be done in different ways (Fisher, 1998; Hamill and Snyder, 2000). Here, a variant of the method of Hello and Bouttier (2001) is proposed.

2.2.1 3D variational assimilation

The three-dimensional variational (3D-Var) data assimilation used here has been developed at Environment Canada and is described in Gauthier *et al.* (1999, 2007). The basic objective of 3D-Var is to obtain the best estimate of the true atmospheric state at the analysis time. In its incremental form, the analysis increment is $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_b$ where $\mathbf{x}$ is the model state and $\mathbf{x}_b$, the background state and $\delta\mathbf{x}$ is obtained by minimizing the cost function

$$J(\delta\mathbf{x}) = J_b(\delta\mathbf{x}) + J_o(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{H}\delta\mathbf{x} - \mathbf{y}')^T\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{y}'), \qquad (1)$$

where $\mathbf{B}$ and $\mathbf{R}$ represent the background and observation error covariance matrices respectively, $\mathbf{y}' = \mathbf{y} - H(\mathbf{x}_b)$ is the innovation vector, $\mathbf{y}$ the observation vector and $\mathbf{H}$ is the linearized version of the observation operator H that maps the model state vector $\mathbf{x}$ to observation space. For sake of simplicity, it is assumed that there are no outer iterations. At

its minimum, (1) yields the analysis increment $\delta\mathbf{x}_a$ that is added to the background to obtain the analysis $\mathbf{x}_a$ defined as

$$\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\mathbf{y}' \tag{2}$$

where $\mathbf{K}$ stands for the Kalman gain matrix expressed as

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} . \tag{3}$$

In 3D-Var, the background error covariances are represented as a stationary matrix. Recently, the assimilation system of Environment Canada has been extended to 4D-Var (Gauthier *et al.* 2007), in which the background state is compared to the observations at the exact observation time. Moreover in 4D-Var, the background error statistics are implicitly evolved over the assimilation window, which makes them flow-dependent. This slightly relaxes the assumption of stationarity implicit in 3D-Var. In the context of the cycling process of any data assimilation system, it may be important to include a flow-dependent form for the background-error covariances to account for the evolved covariances from the previous assimilation (Fisher and Andersson, 2001; Buehner, 2010a,b).

2.2.2 Adapted 3D-Var approach

To account for anisotropic atmospheric flow, flow dependence can be included in $\mathbf{B}$. The approach at ECMWF has been to explicitly incorporate, within the background-error covariance matrix of 4D-Var, a flow-dependent component defined in a subspace spanned by the leading Hessian singular vectors. This is referred to as a reduced-rank Kalman filter (RRKF) (Fisher, 1998; Beck and Ehrendorfer, 2004). Results demonstrate that the impact of the RRKF is small when the number of Hessian singular vectors used is small compared to the dimension of phase space (Fisher and Andersson, 2001).

In the context of 3D-Var, Hello and Bouttier (2001) proposed to estimate the flow-dependent background-error covariances along a single sensitive direction. This approach uses the adjoint-based sensitivities to define the background-error covariance matrix along that component and the stationary background covariances for the remaining orthogonal subspace. As the spatial structure of the analysis increments is driven by the formulation of

the background error covariance, the result is that the analysis increment gives a representation of the sensitivity structure function and its amplitude is determined from the fit to the observations that project in that direction. Otherwise, the analysis increment gives a representation of the stationary background-error covariance matrix commonly used in 3D-Var.

A variant of this approach is proposed here to make corrections to the background along a single sensitive direction. This approach will be referred to as the *adapted 3D-Var*, for which the background-error covariance model embeds the structure functions as defined by sensitivity functions. The new background-error covariance matrix, $\tilde{\mathbf{B}}_x$ is composed of the original covariance matrix, $\mathbf{B}_h$, with homogeneous and isotropic error correlations to which an additional component is added in the direction spanned by the sensitivity function, $\mathbf{f}$. For any given sensitivity function $\mathbf{f}$, the corresponding *sensitivity structure function* $\mathbf{v}$ is defined as

$$\mathbf{v} = \frac{\mathbf{f}}{\langle \mathbf{f}, \mathbf{f} \rangle_{\mathbf{B}}^{1/2}},$$

where the inner product $\langle \mathbf{f}, \mathbf{f} \rangle_{\mathbf{B}} \equiv \mathbf{f}^T \mathbf{B}_h^{-1} \mathbf{f}$ has been used to normalize $\mathbf{f}$. The new covariance matrix is then

$$\tilde{\mathbf{B}}_x = \mathbf{B}_h + \sigma^2 \mathbf{v}\mathbf{v}^T \tag{4}$$

with σ^2 is the variance added to the background error in the sensitive direction. This assures that the 3D-Var behaves according to $\mathbf{B}_h$ in regions where the sensitivity function vanishes, but adopts the structure of the sensitivity function where it does not.

To formulate the background term in (1) requires the inverse of the covariance matrix $\tilde{\mathbf{B}}_x$. In Appendix A, it is shown that

$$\tilde{\mathbf{B}}_x^{-1} = \mathbf{B}_h^{-1/2T} \left(\mathbf{I} - \frac{\sigma^2}{(\sigma^2 + 1)} \left(\mathbf{B}_h^{-1/2} \mathbf{v} \right) \left(\mathbf{B}_h^{-1/2} \mathbf{v} \right)^T \right) \mathbf{B}_h^{-1/2}. \tag{5}$$

When introducing this in (1), it becomes

$$J\left(\tilde{\boldsymbol{\xi}}\right)=\frac{1}{2}\tilde{\boldsymbol{\xi}}^{T}\tilde{\boldsymbol{\xi}}+\frac{1}{2}\left(\mathbf{H}\mathbf{B}_{h}^{1/2}\mathbf{L}^{-1}\tilde{\boldsymbol{\xi}}-\mathbf{y}'\right)^{T}\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{B}_{h}^{1/2}\mathbf{L}^{-1}\tilde{\boldsymbol{\xi}}-\mathbf{y}'\right) \tag{6}$$

where $\mathbf{L}^{-1}=\mathbf{I}+\left(\sqrt{\sigma^{2}+1}-1\right)\tilde{\mathbf{v}}\tilde{\mathbf{v}}^{T}$ and $\tilde{\mathbf{v}}=\mathbf{B}_{h}^{-1/2}\mathbf{v}$. Details can be found in Appendix A. It can be seen that the standard 3D-Var is retrieved when $\sigma^{2}=0$.

The analysis increment can be expressed as $\delta\mathbf{x}_{a}=\tilde{\mathbf{K}}\mathbf{y}'$ where the gain matrix $\tilde{\mathbf{K}}$ is

$$\tilde{\mathbf{K}}=\tilde{\mathbf{B}}_{x}\mathbf{H}^{T}\left(\mathbf{R}+\mathbf{H}\tilde{\mathbf{B}}_{x}\mathbf{H}^{T}\right)^{-1}. \tag{7}$$

2.3 ASSIMILATION IN THE SUBSPACE SPANNED BY SENSITIVITIES

The motivation for introducing a sensitivity structure in the background covariance is for its potential to impact the most the forecast at a given lead time. In this section, we investigate the case where the background-error covariance contains only that flow-dependent structure. This is the limiting case that reflects the early rationale that comparison to observations would be used to determine the amplitude of the structure having the correct dynamics associated with error growth and at the same time agreeing with the available observations.

2.3.1 Use of a $\mathbf{B}$ *matrix confined within the subspace spanned by a single sensitive direction.*

Assuming that the σ term in (4) does not vanish, we are also interested in the limiting solution when the parameter σ increases. To present the argument, we will take $\mathbf{B}_{h}=0$ and the background-error covariance matrix is reduced to its component in the subspace spanned by the sensitivity function and (4) may be written is then $\tilde{\mathbf{B}}_{x}=\sigma^{2}\mathbf{v}\mathbf{v}^{\mathrm{T}}$. The analysis increment is then confined to that subspace and can be expressed as

$$\delta\mathbf{x}_{a}=\tilde{\mathbf{K}}\mathbf{y}'=\alpha\mathbf{v}. \tag{8}$$

Its amplitude α is then found to be

$$\alpha=\frac{(\mathbf{H}\mathbf{v})^{T}\mathbf{R}^{-1}\mathbf{y}'}{\sigma^{-2}+(\mathbf{H}\mathbf{v})^{T}\mathbf{R}^{-1}(\mathbf{H}\mathbf{v})}=\frac{\sigma^{2}C_{1}}{1+\sigma^{2}C_{2}}. \tag{9}$$

The coefficients $C_1 = (\mathbf{Hv})^T \mathbf{R}^{-1} \mathbf{y}'$ and $C_2 = (\mathbf{Hv})^T \mathbf{R}^{-1} (\mathbf{Hv})$ control the magnitude of the analysis increments and depend on several parameters such as the matrix $\mathbf{R}$ of observations error covariances, the estimated innovation, the volume of observations and also their locations with respect to the amplitude of the sensitive function $(\mathbf{Hv})$. Here, C_1 is the projection of the scaled innovation vector $\mathbf{R}^{-1/2}\mathbf{y}'$ onto the subspace spanned by the sensitivity function and C_2 is the norm of $\mathbf{R}^{-1/2}\mathbf{Hv}$ also scaled with respect to the observation error. If $\mathbf{R}^{-1/2}\mathbf{y}'$ happens to be orthogonal to $\mathbf{R}^{-1/2}(\mathbf{Hv})$, then $C_1 = 0$ and thus $\delta\tilde{\mathbf{x}}_a = \alpha\mathbf{v} = 0$. On the other hand, if $\mathbf{R}^{-1/2}\mathbf{y}'$ happens to be completely in the subspace spanned by the sensitive direction, then $C_1 \neq 0$. Moreover, it is important to point out that the amplitude of the analysis increment (9) will be small if the observation is located in areas of weak sensitivity ($\mathbf{v} \approx 0$) or if the observation value is similar to the background value, even if the observation is located inside an area of strong sensitivity. For large values of σ, the maximum amplitude is given by the ratio C_1 / C_2.

This particular case indicates that the value of α can be determined by a single observation: this would be the true value α_t if the observation were perfect. On the other hand, (9) indicates that when several observations are used, it is the average of the projection of the innovations onto the sensitivity structure that will define the amplitude. This raises the issue of whether the observations are able to detect a particular structure when observation error is present.

Finally, the information content, or *degrees of freedom per signal* (*DFS*), corresponds to

$$DFS = tr\left(\mathbf{H}\tilde{\mathbf{K}}\right) = \frac{\sigma^2 C_2}{1 + \sigma^2 C_2}$$

when $\tilde{\mathbf{B}}_x = \sigma^2 \mathbf{v}\mathbf{v}^T$. In the limit $\sigma \to \infty$, $DFS \to 1$ which indicates that when a single direction defines the analysis correction, the degrees of freedom can be reduced by at most 1. Moreover, the analysis increment can be expressed as

$$\delta\mathbf{x}_a = tr\left(\mathbf{H}\tilde{\mathbf{K}}\right)\frac{C_1}{C_2}\mathbf{v}.$$

As C_1 represents the projection of the innovations in the direction of $\mathbf{v}$, this shows that information will be added only if the observations can detect the sensitivity structure.

2.3.2 *Observability of a perturbation structure*

To quantify the agreement between the structure function in observation space and the existing observation network, a correlation coefficient ρ is defined as

$$\rho = \frac{(\mathbf{Hv})^T \mathbf{R}^{-1} \mathbf{y}'}{\left[(\mathbf{Hv})^T \mathbf{R}^{-1} (\mathbf{Hv})\right]^{1/2} \left[\mathbf{y}'^T \mathbf{R}^{-1} \mathbf{y}'\right]^{1/2}} = \frac{C_1}{(2C_2 J_o(0))^{1/2}} \tag{10}$$

where $\mathbf{Hv}$ is the sensitivity structure function in observation space, $\mathbf{y}'$ is the innovation vector, $\mathbf{R}$ is the observation error covariance matrix and $J_o(0)$ is the observation component of the cost function evaluated before the minimisation.

Small values of the correlation coefficient indicate that the structure function does not agree well with the innovation vector. The *observability of a sensitivity structure function* can then be defined as the correlation ρ given by (10). With the assumption that observation errors are uncorrelated, the covariance matrix $\mathbf{R}$ is diagonal and the correlation coefficient can be computed separately for each family of observations. This partition per observation type permits to reveal which data types project the most on a given structure function.

In particular, when a single observation is assimilated, the value of the correlation ρ will be equal to 1, unless either $\mathbf{Hv}$ or are $\mathbf{y}'$ exactly null.

In the next section, a simple 1D-Var example is used to illustrate this point.

2.4 EXAMPLE BASED ON 1D-VAR EXPERIMENTS

A simple one-dimensional (1D) univariate analysis system is used, which is very similar to the one used by Hello and Bouttier (2001) and by Bergot and Doerenbecher (2002). It consists of a circular domain with perimeter of 30,000 km. Within the incremental framework, the cost function is rewritten as in (1) which implies that the background is taken to be null and the observations replaced by the innovation departures $\mathbf{y}'$ with respect to the

background. The background-error covariance matrix $\mathbf{B}_h$ in physical space assumes isotropic error correlations in the guess field, with a length-scale of 300 km. The observation error covariances are assumed to be uncorrelated with the same observation error variance. Therefore, $\mathbf{R} = \sigma_o^2 \mathbf{I}$ where $\mathbf{I}$ is the identity matrix and σ_o^2 is the observation error variance. The sensitive function is represented using simple trigonometric functions as

$$\mathbf{f}(x) = \frac{1}{2}\exp\left(-\frac{1}{2}\left(\frac{x - L/2}{L_b}\right)^2\right)\cos\left(4\left(\frac{x - L/2}{L_b}\right)\right), \tag{11}$$

where L is the length of the circular domain and L_b = 600 km is the correlation length-scale for $\mathbf{f}$.

The experiments will first try to assess the extent to which a signal of given amplitude can be detected by observations for different levels of observation error. This would correspond to *a posteriori sensitivity functions* often used to trace back the key analysis errors that can explain forecast error at a given lead time (Klinker *et al.*, 1998; Laroche *et al.*, 2002). So we know after the fact what should be the structure of the correction to the analysis that would impact the forecast the most. The objective is then to use the *a posteriori sensitivity* as a *structure function* and find out if the analysis will recover the correct amplitude. In the computation of *a posteriori sensitivities* no constraint is imposed to have the analysis increment close to the observations.

Taking the background state to be zero and the true state $\mathbf{x}_t = \alpha_t \mathbf{v}$, the background error of this particular realization is then $\varepsilon_b = -\alpha_t \mathbf{v}$. On the other hand, the observation is such that $\mathbf{y} = \mathbf{y}_t + \boldsymbol{\varepsilon}_o = \alpha_t \mathbf{H}\mathbf{v} + \boldsymbol{\varepsilon}_o$ and the background-error covariance matrix is taken as $\tilde{\mathbf{B}}_x = \sigma^2 \mathbf{v}\mathbf{v}^T$. The innovation is then $\mathbf{y} - \mathbf{H}\mathbf{x}_b = \alpha_t \mathbf{H}\mathbf{v} + \boldsymbol{\varepsilon}_o$ so that

$$\alpha = \frac{\sigma^2 C_2}{1 + \sigma^2 C_2}\left(\alpha_t + \frac{(\mathbf{H}\mathbf{v})^T \mathbf{R}^{-1} \boldsymbol{\varepsilon}_o}{(\mathbf{H}\mathbf{v})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{v})}\right).$$

which expresses the signal (α_t) with respect to the observation error projected along $\mathbf{v}$. The variance of the *noise* is $\sigma_\alpha^2 = 1/\left((\mathbf{H}\mathbf{v})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{v})\right)$ and the signal-to-noise ratio is α_t / σ_α.

To illustrate the impact of the observation error, *a posteriori sensitivities* have been sampled to generate the observations used in the assimilation. Implicitly, it is assumed then that the amplitude of the sensitivity function is below the level of the background error but greater than that of the observation error so that it can be detected by observations. Assuming $\alpha_t = 2$, Table 2.1 gives the values of the coefficients C_1 and C_2, and the correlation coefficient ρ for three experiments in which the observation is first taken to be the truth and then when random observation error is added with variance $\sigma_o^2 = 1$ and 4, respectively. In all three cases, experiments were done with 10, 20 and 40 observations. With perfect observations, the amplitude is recovered and the correlation is very close to 1. Adding observation error dramatically reduces the correlation. With $\sigma_o^2 = 1$, the correlation decreases with the number of observations. Increasing the observation error to a level that compares with the signal, there is no correlation at all between the analysis increment and the sensitivity function. Fig. 2.1 shows the amplitude of the analysis increment as a function of σ for these experiments. When $\sigma_o^2 = 1$, the amplitude is less than the actual value which reflects the fact that the presence of random error "blurs" the signal.

As studied in Hello and Bouttier (2001) and Bergot and Doerenbecher (2002), a single observation should be enough to determine the amplitude of the correction, provided this observation projects onto the sensitivity structure ($C_1 \neq 0$). However, if observation error is present, then the analysis would also fit the observation error. This would happen when $\sigma^2 \to \infty$. With more observations and with observation error, all observations would agree on what α should be; but if the observation error is above the signal, α would vary randomly and the overall fit should yield a value near zero. This is what Fig. 2.1 indicates.

In those previous experiments, the background-error covariance used a sensitivity structure function that corresponded exactly to what was observed. However, the sensitivities are computed under a number of assumptions and this may result in differences in sensitivities present in the atmosphere and detected by the observations from those being computed with a given numerical model (referred to as $\mathbf{v}_t$). In a second set of experiments, different structure functions were used in the background term and to generate the observations. In these experiments, the observations were generated by introducing a phase

shift in the structure function. The results are shown in Fig. 2.2. When the observations are sampled from $\mathbf{v}=\mathbf{v}_t$, Fig. 2.2-a shows that the adapted 1D-Var does recover the right amplitude. In Fig. 2.2-b, $\mathbf{v}\neq\mathbf{v}_t$ and the observations have contradicting views on what the amplitude of $\mathbf{v}$ should be, and with several observations, the net result is the average of the individual contributions and the amplitude of the analysis increment is then very small. Fig. 2.2 also shows, in red, the analysis increment obtained with the standard 1D-Var (homogeneous correlations). It shows that with several observations, the standard 1D-Var is able to capture the signal. By opposition, the adapted 1D-Var is able to reconstruct the correct increment from a single observation provided $\mathbf{v}=\mathbf{v}_t$ and there is no observation error. If not, the adapted 1D-Var with several observations tends to yield an increment that is of small amplitude.

In summary, a 1D-Var example was used to show that, in presence of observation error and with several observations, the adapted 1D-Var can recover the signal, provided the observation error is smaller than the signal. If the sensitivity structure functions differ from those present in the atmosphere and detected by the observations, the adapted 1D-Var will underestimate the amplitude of the signal. In the next section, an adapted 3D-Var was implemented within the variational system of Environment Canada and experiments have been carried out to test if the added flow-dependent structure function manages to improve the forecasts.

2.5 RESULTS WITH 3D-VAR USING DIFFERENT DEFINITIONS FOR THE STRUCTURE FUNCTIONS

A posteriori sensitivities have a structure and amplitude that result in a significant reduction in the forecast error. However, they are not constrained to fit the observations at the initial time (Isaksen *et al.*, 2005). In this section, sensitivity structure functions are defined as normalized *a posteriori* sensitivities. The object is then to investigate the extent to which the assimilated observations can recover the amplitude of the *a posteriori sensitivity*. There are different ways to define the *a posteriori sensitivities*. The sensitivities depend on the metric used to measure the forecast error (e.g., dry energy norm, Hessian norm). The definition of norm may also involve the area over which it is computed. When the sensitivity function is

computed globally, this defines a global sensitivity function. *Local sensitivity functions* can be also calculated to identify the source of forecast error only over a local area (Hello and Bouttier, 2001). By computing the forecast error over a limited area, the local sensitivity function focuses in changes in the initial conditions that will impact that specific area at a given lead time. In Caron *et al.* (2007b), the computation of sensitivity functions was done by imposing also a nonlinear balance constraint using a potential vorticity (PV) inversion method. Finally, it is important to remember that the sensitivities, as for singular vectors, depend on the resolution and configuration of the adjoint model (e.g., simplified physics, vertical extent and resolution). This leads to several possibilities to consider as potential sensitivity structure functions.

Several experiments in which different definitions of the sensitivity functions were used as structure functions in an adapted 3D-Var based on the operational 3D-Var of Environment Canada (Gauthier *et al.*, 1999, 2007). Experiments involving winter cases documented by Caron *et al.* (2007a) will be discussed. *Key analysis errors* were estimated for four 3D-Var analyses: January 6 and January 27, 2003 at 12 UTC, January 19, 2002 at 00 UTC and February 06, 2002 at 12 UTC. Those cases were associated with cases of severe weather over North America.

For all cases, *a posteriori sensitivities* were computed in different ways to minimize the 24-h forecast error as measured with respect to a verifying analysis. The method employed is explained in Laroche *et al.* (2002) and Caron *et al.* (2007a). Four types of structure functions will be considered in our study:

- a *global* sensitivity, for which the error is measured globally,
- a *local* sensitivity, for which the error is measured over an area on the East coast of North America,
- an hemispheric sensitivity function computed over the latitudinal band 25-90N.
- a PV-bal sensitivity function, for which the control variable is potential vorticity which constrains the sensitivity to be more dynamically balanced.

All cases used the dry energy norm at initial and end time. As already mentioned, the analysis increment (9) has the direction of the sensitivity structure function and the amplitude that best fits the observations. Table 2.2 summarizes the correlation associated with different observation types for all four cases. The results show poor correlations between the observations and the sensitive functions in observation space. This indicates that in the limiting case where $\sigma \to \infty$, the adapted 3D-Var could not be expected to improve the forecast as much as the key analyses do. This is true for all cases considered here.

2.5.1 A test case

For each case documented in Caron *et al.* (2007a), regular 3D-Var global analyses were performed using the full set of observations assimilated operationally at Environment Canada and the background state is the same that was used in the 3D-Var system operational at the time. For each case, the analysis increment is as close as possible to the true atmospheric state in a root-mean-square sense. To test the adapted 3D-Var, analysis increments obtained from 3D-Var analyses were normalized with respect to the norm $\langle \mathbf{f}, \mathbf{f} \rangle_{\mathbf{B}}$ defined in section 2.2 and used as structure function $\mathbf{v}$ in $\mathbf{B} = \sigma^2 \mathbf{v}\mathbf{v}^T$. In that case, the objective was to test whether this limiting case of the adapted 3D-Var could recover the amplitude of the analysis increment, knowing that this structure does have the ability to fit the observations. Fig. 2.3 shows the estimated amplitude of the analysis increment calculated from (9) as a function of the parameter σ for different families of observational data. The estimated amplitude of analysis increment increases rapidly with σ and saturates for each data types at a value corresponding to the ratio C_1 / C_2. This example indicates that the *adapted* 3D-Var analysis increment recovers the amplitude of the sensitive function when σ is sufficiently large. To assess the agreement between the analysis increment for each case and the observation network, the correlation coefficients have been computed for all data types and the results are summarized in Table 2.3. The correlation coefficients values are shown for each family of observations and for all observations combined indicate good agreement for all cases.

2.5.2 *Application in an adapted 3D-Var context*

The *adapted* 3D-Var is closer to the observations than the 3D-Var for observed storm cases documented in Caron *et al.* (2007). A measure of the fit to the observations is given by the observation component of the cost function, J_o. Following Caron *et al.* (2007), the relative difference in J_o is examined individually for each family of observations (radio-soundings RAOB, aircraft data AIREP, surface and ship data SURFC, radiances data from satellite ATOVS, and wind vectors derived from satellite data AMV) and for the combined set of observations combined (TOTAL), in the form

$$\Delta J_o = \frac{J_o(\mathbf{x}^{Ad,3D}) - J_o(\mathbf{x}^{3D})}{J_o(\mathbf{x}^{3D})} \tag{12}$$

where $J_o(\mathbf{x}) = \frac{1}{2}(\mathbf{y}-\mathbf{Hx})^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Hx})$ measures the distance between the model state $\mathbf{x}$ and the observations $\mathbf{y}$. A positive value ($J_o(\mathbf{x}^{Ad,3D}) > J_o(\mathbf{x}^{3D})$ means that that the *adapted* 3D-Var analyses are further away from the observations than the corresponding operational analysis and, conversely, a negative value ($J_o(\mathbf{x}^{Ad,3D}) < J_o(\mathbf{x}^{3D})$ means that the *adapted* 3D-Var analyses fit the observations better than 3D-Var.

As demonstrated by Caron *et al.* (2007), adjoint sensitivity structures from the CMC energy-norm-based key analysis error algorithm manage to minimize short-range (24 h) forecast errors, but depart more from the observations than the original 3D-Var analysis. The percentage of improvement or degradation of the fit to the observations is shown in Fig. 2.4 when *a posteriori* sensitivities are used as structure functions in the *adapted* 3D-Var. The results show that the experiments with different ways to define sensitivity functions can lead to quite different results. However, for the cases presented here the results are approximately neutral.

2.5.3 *Experiments with a pseudo-inverse defined in a subspace spanned by a finite number of singular vectors*

For true flow-dependent background error covariances, a *priori* structure functions would be defined either as a finite number of singular vectors (Leutbecher, 2003) or from sensitivity

gradient which identifies those structures that would impact the most the forecast as measured by a given metric (Hello and Bouttier, 2001). Here, the question asked is to know whether those structures can both fit the observations and reduce the forecast error.

Singular vectors (SVs) are the perturbations with the largest amplification rate over a given time interval. A set of 60 singular vectors was calculated for 18 cases of December 2007 for a time interval of forty-eight hours. The singular vectors are calculated using the total energy norm at initial and final times. The first singular vector (SV1) has the largest singular value that is much larger than the others. The correlation coefficients values in Table 2.4 suggest that the SV1 at initial and final time do not correlate well with the innovation vector. The pseudo-inverse is the perturbation with the largest impact on the forecast error (Mahidjiba *et al.*, 2007) obtained by expressing the forecast error at 48-h as a linear combination of a finite number of singular vectors. Using here the set of 60 singular vectors, this leads to a correction to the initial conditions that does reduce the forecast error represented in a subspace that truly represents the growing modes of forecast error. However, the last column of Table 2.4 indicates that this pseudo-inverse is not well correlated with the observations either: the correlation coefficient is no better than that of the first singular vector alone. We therefore conclude that the structures defined by singular vectors are not well correlated with observations, which means that they are not observable given the level of error in the observations being assimilated.

2.6 SUMMARY AND CONCLUSION

The argument presented in this paper is that structures that can explain a substantial part of future error growth have small amplitude and the signal is often below the level of observation error. In other words, the signal-to-noise ratio is too low for them to be detected by observations. This is an important issue for the use of flow-dependent structures that could be related to precursors of error growth. Several experiments have been performed to include known *a posteriori* sensitivities as structure functions within a so-called adapted 3D-Var. The results obtained by Caron *et al.* (2007-a,b) and in our own experiments indicate that *a posteriori* key analyses do succeed to reduce significantly the forecast error, but tend to pull the analysis further away from the observations than the reference analysis they were

correcting. The experiments with an adapted 3D-Var manage to correct the analysis with the structure of the key analysis under the constraint that the resulting analysis is close to the observations. The results indicate that the analysis is then close to the observations, but this does not improve significantly the quality of the forecast. Pushing this to the limit where the bulk of the forecast error variance is put in the direction of the sensitivity structure function, it was expected that one would recover the amplitude of the *a posteriori* sensitivity (or key analysis). This was not the case. A close study of this limiting case indicated that the retrieved amplitude is determined by the correlation of the structure function with the innovations. With a single observation, one recovers the projection of the innovation in that direction; but adding more observations results in very small amplitude, as the correlation of the innovation vector with the image, in observation space, of the sensitivity structure function is small.

These results are important and more thought is needed on how to include information about precursors in the analysis. An element that needs to be considered is that the analysis may have to wait for the instability to develop above the signal-to-noise ratio for the observations to be able to detect it and properly correct the initial conditions. In a sense, this would indicate that evolved covariances obtained from a Kalman filter as obtained from an Ensemble Kalman filter (Houtekamer and Mitchell, 2009) should be better observable than covariances represented in a subspace spanned by singular vectors. However, evolved singular vectors could be a good prospect. This will be the object of future work.

ACKNOWLEDGEMENTS

Authors would like to deeply thank to Dr. Jean-François Caron who provided the *a posteriori* sensitivity functions used in this study. Stimulating discussions with Drs. Mark Buehner and Ahmed Mahidjiba were very helpful during the course of this study. They kindly provided the singular vectors and the pseudo-inverses used in this study. Environment Canada provided the computing facilities and technical assistance for the use of their assimilation system.

This work has been funded mostly by Grant 500-B of the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) for the project on the *Impact of Observing Systems on Forecasting Extreme Weather in the short, medium and extended range: A Canadian contribution to THORPEX*, with additional support from Discovery Grant No.357091 of the Natural Sciences and Engineering Research Council (NSERC) of Canada.

APPENDIX A: FORMULATION OF THE ADAPTED 3D-VAR

Adding a sensitive component to $\mathbf{B}_h$ led to

$$\tilde{\mathbf{B}}_x = \mathbf{B}_h + \sigma^2 \mathbf{v}\mathbf{v}^T$$

where $\langle \mathbf{v}, \mathbf{v} \rangle_{\mathbf{B}} \equiv \mathbf{v}^T \mathbf{B}_h^{-1} \mathbf{v} = 1$. Using the Sherman-Morrison formula (Gollub and van Loan, 1996), the inverse of $\tilde{\mathbf{B}}_x$ is found to be

$$\tilde{\mathbf{B}}_x^{-1} = \mathbf{B}_h^{-1/2T} \left(\mathbf{I} - \frac{\sigma^2}{(\sigma^2 + 1)} \left(\mathbf{B}_h^{-1/2} \mathbf{v} \right) \left(\mathbf{B}_h^{-1/2} \mathbf{v} \right)^T \right) \mathbf{B}_h^{-1/2} \tag{A.1}$$

and the 3D-Var cost function (1) can be rewritten as

$$J(\delta \mathbf{x}) = \frac{1}{2} \delta \mathbf{x}^T \mathbf{B}_h^{-1/2T} \left(\mathbf{I} - \frac{\sigma^2}{(\sigma^2 + 1)} (\mathbf{B}_h^{-1/2} \mathbf{v})(\mathbf{B}_h^{-1/2} \mathbf{v})^T \right) \mathbf{B}_h^{-1/2} \delta \mathbf{x} + \frac{1}{2} (\mathbf{H} \delta \mathbf{x} - \mathbf{y}')^T \mathbf{R}^{-1} (\mathbf{H} \delta \mathbf{x} - \mathbf{y}'),$$

Defining the change of variables, $\boldsymbol{\xi} = \mathbf{B}_h^{-1/2} \delta \mathbf{x}$ and $\tilde{\mathbf{v}} = \mathbf{B}_h^{-1/2} \mathbf{v}$ yields that

$$\begin{aligned} J(\boldsymbol{\xi}) &= \frac{1}{2} \boldsymbol{\xi}^T \left(\mathbf{I} - \frac{\sigma^2}{\sigma^2 + 1} \tilde{\mathbf{v}} \tilde{\mathbf{v}}^T \right) \boldsymbol{\xi} + \frac{1}{2} \left(\mathbf{H} \mathbf{B}_h^{1/2} \boldsymbol{\xi} - \mathbf{y}' \right)^T \mathbf{R}^{-1} \left(\mathbf{H} \mathbf{B}_h^{1/2} \boldsymbol{\xi} - \mathbf{y}' \right) \\ &= J_b(\boldsymbol{\xi}) + J_o(\boldsymbol{\xi}) \end{aligned} \tag{A.2}$$

So defined, the sensitivity structure function is such that

$$\tilde{\mathbf{v}}^T \tilde{\mathbf{v}} = \mathbf{v}^T \mathbf{B}_h^{-1/2} \mathbf{B}_h^{-1/2} \mathbf{v} = \mathbf{v}^T \mathbf{B}_h^{-1} \mathbf{v} = 1.$$

In terms of these new variables, we have that

$$\begin{aligned} \left(\mathbf{I} - \frac{\sigma^2}{\sigma^2 + 1} \tilde{\mathbf{v}} \tilde{\mathbf{v}}^T \right) &= \left(\mathbf{I} + \frac{\left(\sqrt{\sigma^2 + 1} - \sigma^2 - 1 \right)}{\sigma^2 + 1} \tilde{\mathbf{v}} \tilde{\mathbf{v}}^T \right)^2, \\ &= \mathbf{L}^T \mathbf{L} \end{aligned}$$

with.

$$\mathbf{L} = \mathbf{I} + \left(\frac{\sqrt{\sigma^2 + 1} - \sigma^2 - 1}{\sigma^2 + 1} \right) \tilde{\mathbf{v}}\tilde{\mathbf{v}}^T = \mathbf{L}^T . \tag{A.3}$$

This allows to introduce another change of variable $\tilde{\boldsymbol{\xi}} = \mathbf{L}\boldsymbol{\xi}$ so that $\boldsymbol{\xi} = \mathbf{L}^{-1}\tilde{\boldsymbol{\xi}}$. The inverse of $\mathbf{L}$ is found to be

$$\mathbf{L}^{-1} = \mathbf{I} + \left(\sqrt{\sigma^2 + 1} - 1 \right) \tilde{\mathbf{v}}\tilde{\mathbf{v}}^T$$

so that (A.2) is finally expressed as

$$J\left(\tilde{\boldsymbol{\xi}}\right) = \frac{1}{2}\tilde{\boldsymbol{\xi}}^T\tilde{\boldsymbol{\xi}} + \frac{1}{2}\left(\mathbf{H}\mathbf{B}_h^{1/2}\mathbf{L}^{-1}\tilde{\boldsymbol{\xi}} - \mathbf{y}'\right)^T \mathbf{R}^{-1}\left(\mathbf{H}\mathbf{B}_h^{1/2}\mathbf{L}^{-1}\tilde{\boldsymbol{\xi}} - \mathbf{y}'\right). \tag{A.4}$$

Its gradient is readily found to be

$$\nabla_{\tilde{\boldsymbol{\xi}}} J = \tilde{\boldsymbol{\xi}} + \mathbf{L}^{-T}\mathbf{B}_h^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{H}\mathbf{B}_h^{1/2}\mathbf{L}^{-1}\tilde{\boldsymbol{\xi}} - \mathbf{y}'\right).$$

REFERENCES

Beck A. and Ehrendorfer, M., 2004: Singular-vector-based covariance propagation in a quasigeostrophic assimilation system. *Mon. Weather Rev.*, **133**, 1295-1310.

Bergot T. and A. Doerenbecher, 2002: A study on the optimization of the deployment of targeted observations using adjoint-based methods. *Q. J. R. Meteorol. Soc.*, **128**, 1689–1712.

Berre, L., G. Desroziers, L. Raynaud, R. Montroty and F. Gibier, 2009: Consistent operational ensemble variational assimilation, *Procedings of the CAWCR Workshop on Ensemble Prediction and Data Assimilation*, Melbourne, Australia.

Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell and B. He, 2010a: Intercomparaison of variational data assimilation and Ensemble Kalman filter for global deterministic NWP. Part I: Description and single observation experiments, *Mon. Weather Rev,* **138**, 1550-1566.

_____, P. L. Houtekamer, C. Charette, H. L. Mitchell and B. He, 2010b: Intercomparaison of variational data assimilation and Ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations, *Mon. Weather Rev,* **138**, 1567-1586.

Buizza, R., J.-R. Bidlot, N. Wedi, M. Fuentes, M. Hamrud, G. Holt and F. Vitart, 2007: The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Q. J. R. Meteorol. Soc.*, **133**, 681-695.

_____, C. Cardinali, G. Kelly and J.-N. Thépaut, 2008: The value of observations. Part II: The value of observations located in singular vector-based target areas. *Q. J. R. Meteorol. Soc.*, **133**, 1817–1832.

Cardinali, C., R. Buizza, G. Kelly, M. Shapiro and J-N. Thépaut, 2008: The value of observations. Part III: influence of weather regimes on targeting, *Q. J. R. Meteorol. Soc.*, **133**, 1833–1842.

Caron, J.-F., Yau M. K, Laroche S. and Zwack P., 2007a: The characteristics of key analysis errors. Part I: Dynamical balance and comparison with observations. *Mon. Weather Rev.*, **135**, 249–266.

_____, Yau M. K, Laroche S. and Zwack P., 2007b: The characteristics of key analysis errors. Part II: The importance of the PV corrections and the impact of balance. *Mon. Wea. Rev.*, **135**, 267–280.

Evensen, G., 1994: Sequencial data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143-10162.

Fisher, M., 1998: Development of a simplified Kalman Filter. *ECMWF Research Dept. Technical Memorandum,* no. **260**, 16 pp.

_____ and E. Andersson, 2001: Developments in 4D-Var and Kalman filtering. *ECMWF Research Dept. Technical Memorandum,* no. **347**.

Gauthier, P., C. Charette, L. Fillion, P. Koclas and S. Laroche, 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmos.-Ocean,* **37,** 103–156.

_____, Tanguay M., S. Laroche, S.Pellerin, and J. Morneau, 2007: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Weather Rev.*, **135**, 2339–2354.

Golub, H. G. and Van Loan C. F., 1983: *Matrix computations*. The John Hopkins University Press, Baltimore, U.S.A., 476 pp.

Hamill. T. and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Mon. Weather Rev.*, **128**, 2905-2919.

Hello, G., Lalaurette F.and Thépaut J.-N., 2000: Combined use of sensitivity information and observations to improve meteorological forecasts: A feasibility study applied to the 'Christmas storm' case. *Q. J. R. Meteor. Soc.*, **126**, 621-647.

_____ and Bouttier F., 2001: Using adjoint sensitivity as a local structure function in variational data assimilation. *Nonlinear Processes in Geophysics*, **8**, 347-355.

Isaksen, L., M. Fisher, E. Andersson, and J. Barkmeijer, 2005: The structure and realism of sensitivity perturbations and their interpretation as "Key analysis errors." *Q. J. R. Meteor. Soc.,* **131,** 3053–3078.

Joly, A., K. A. Browning, P. Bessemoulin, J. P. Cammas, G. Caniaux, J. P. Chalon, S. A. Clough, R. Dirks, K. A. Emanuel, L. Eymard, F. Lalaurette, R. Gall, T. D. Hewson, P. H. Hildebrand, D. Jorgensen, R. H. Langland, Y. Lemaître, P. Mascart, J. A. Moore, P. O. G. Persson, F. Roux, M. A. Shapiro, C. Snyder, Z. Toth and R. M. Wakimoto, 1999: Overview of the field phase of the Fronts and Atlantic Storm Track Experiment (FASTEX) project. *Q. J. R. Meteorol. Soc.*, **125**, 3131-3164.

Kelly, G., J.-N. Thépaut, R. Buizza and C. Cardinali, 2008: The value of observations. Part I: the value of observations taken over the oceans, *Q. J. R. Meteorol. Soc.*, **133**, 1803–1815.

Klinker, E., Rabier F. and Gelaro R., 1998: Estimation of key analysis errors using the adjoint technique. *Quart. J.R. Meteor. Soc.*, **124**, 1909-1933.

Lacarra, J. F. and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. *Tellus*, **40A**, 81-95.

Langland R. H., Z. Toth, R. Gelaro, I. Szynyogh, M. A. Shapiro, S. J. Majumdar, R. Morss, G. D. Rohaly, C. Velden, N. Bonds and C. H. Bishop, 1999: The North Pacific Experiment (Norpex-98): Targeted observations for improved North American Weather Forecasts. *Bull. Am. Meteorol. Soc.*, **80**, 1363-1384.

——, M. A. Shapiro and R. Gelaro, 2002: Initial condition sensitivity and error growth in forecasts of the 25 January 2000 east coast snowstorm. *Mon. Weather Rev.*,**130**, 957-974.

——, 2005: Observation impact during the North-Atlantic TreC-2003. *Mon. Weather Rev.*, **133**, 2297-2309.

——, 2006: Issues in targeted observing. *Q. J. Roy. Meteorol. Soc.*, **131**, 3409–3425.

Laroche, S., M. Tanguay, A. Zadra and J. Morneau, 2002: Use of adjoint sensitivity analysis to diagnose the CMC Global analysis performance: A case study. *Atmos.- Ocean,* **40** (4), 423-443.

Leutbecher, M., 2003: A Reduced Rank Estimate of Forecast Error Variance Changes due to Intermittent Modifications to the Observing Network, *J. Atmos. Sci.*, 729–742.

Liu, Z. Q. and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q. J. R. Meteorol. Soc*, **128**, 1367-1386.

Mahidjiba, A., M. Buehner and A. Zadra, 2007: Excitation of Rossby-wave trains: optimal growth of forecast errors. *Meteorologische Zeitschrift*, **16**, 665-673.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteor. Soc.*, **122,** 73–120.

Petersen, G. N. and A. J. Thorpe, 2007: The impact on weather forecasts of targeted observations during A-TreC, *Q. J. R. Meteorol. Soc.*, **133**, 417–431.

a) $\mathbf{y}' = 2(\mathbf{Hv})$	Nb obs.	C_1	C_2	ρ
	10 obs.	1.29	0.64	0.99
	20 obs.	1.96	0.97	0.99
	40 obs.	2.26	1.13	1.
b) $\mathbf{y}' = 2(\mathbf{Hv}) + \varepsilon_o$ $\sigma_o^2 = 1$	10 obs.	0.95	0.64	0.38
	20 obs.	1.15	0.97	0.22
	40 obs.	1.48	1.13	0.20
c) $\mathbf{y}' = 2(\mathbf{Hv}) + \varepsilon_o$ $\sigma_o^2 = 4$	10 obs.	0.89	0.64	0.17
	20 obs.	0.89	0.97	0.11
	40 obs.	0.87	1.13	0.08

Table 2.1: Coefficients C_1 and C_2 and correlation coefficient ρ computed using 1D-Var assimilation system for three experiments. a) perfect observation, b) $\sigma_o^2 = 1$, c) $\sigma_o^2 = 4$. In each case, experiments were done with 10, 20 and 40 observations.

Structure functions	Obs. type	Correlation coefficient ρ			
		January 27, 2003	January 06, 2003	February 06, 2002	January 19, 2002
GLOBAL FUNCTION	RAOB	0.01	0.02	0.03	-0.01
	AIREP	0.00	0.02	-0.01	-0.01
	AMV	0.02	0.01	0.02	0.02
	SURFC	0.14	0.11	0.19	0.04
	ATOVS	0.13	0.11	0.07	0.12
	TOTAL	0.05	0.05	0.05	0.03
LOCAL FUNCTION	RAOB	-0.01	0	-0.01	-0.02
	AIREP	-0.03	-0.01	-0.03	-0.03
	AMV	0	0.03	-0.04	0
	SURFC	0.04	-0.03	0	0.02
	ATOVS	0.05	0.01	0.06	0.02
	TOTAL	0	0	0	-0.01
HEMISPHERIC FUNCTION	RAOB	0.00	0.02	0.01	0.01
	AIREP	-0.05	0.02	-0.02	-0.03
	AMV	-0.05	-0.08	-0.02	0.02
	SURFC	0.12	0.1	0.16	0.08
	ATOVS	0.08	0.07	0.07	0.04
	TOTAL	0.03	0.04	0.04	0.02
PV-BAL FUNCTION	RAOB	0.01	0	0.01	0
	AIREP	-0.03	0.01	-0.03	0
	AMV	-0.04	-0.08	-0.03	0.01
	SURFC	0.04	-0.06	0.21	0.06
	ATOVS	0.09	0.08	0.08	0.05
	TOTAL	0.03	-0.01	0.06	0.02

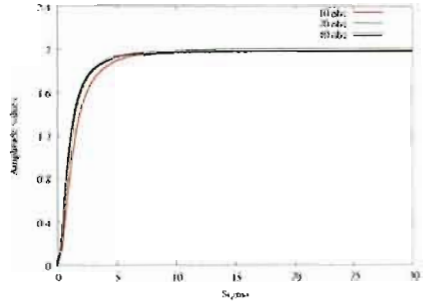
Table 2.2: Correlation coefficient computed for different data types and for all observations combined. Different sensitivity functions from the key analysis error algorithm are used: GLOBAL (initial corrections that minimized the 24-h forecast error over the globe); LOCAL (initial corrections that minimized the 24-h forecast error over an area on the East coast of North America); HEMISPHERIC (initial corrections over the latitudinal band 25N-90N); PV-BAL (balanced initial corrections over the latitudinal band 25N-90N). Cases shown are a) January 27, 2003, b) January 06, 2003, c) February 06, 2002, d) January 19, 2002.

Obs. type	Correlation coefficient ρ			
	January 27, 2003	January 06, 2003	February 06, 2002	January 19, 2002
RAOB	0.73	0.76	0.77	0.76
AIREP	0.73	0.73	0.73	0.72
AMV	0.68	0.72	0.72	0.73
SURFC	0.69	0.74	0.75	0.76
ATOVS	0.59	0.58	0.71	0.65
TOTAL	0.71	0.73	0.75	0.74

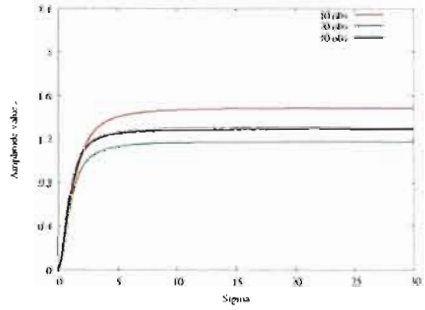
Table 2.3: Correlation coefficient computed for different data types and for all observations combined. The 3D-Var analyses are used as sensitivity function to adapt the background error covariance matrix for 4 case studies: a) January 27, 2003; b) January 06, 2003, c) February 06, 2002, d) January 19, 2002.

Date	Obs. type	Correlation coefficient ρ		
		SV no. 1 Initial time	SV no. 1 Final time	Pseudo-inverse
2007120100	TOTAL	0.0098	0.0067	0.0169
2007120212	TOTAL	0.0140	-0.0179	-0.0011
2007120400	TOTAL	-0.0187	-0.0211	-0.0034
2007120512	TOTAL	0.0022	-0.0020	0.0124
2007120700	TOTAL	0.0159	0.0020	-0.0033
2007120812	TOTAL	0.0019	0.0212	0.0062
2007121000	TOTAL	-0.0029	-0.0151	0.0040
2007121112	TOTAL	0.0054	0.0148	0.0096
2007121300	TOTAL	0.0125	-0.0241	-0.0028
2007121412	TOTAL	0.0224	-0.056	0.0209
2007121600	TOTAL	0.0125	0.0235	0.0234
2007121712	TOTAL	0.0041	0.0465	-0.0064
2007121900	TOTAL	0.0119	-0.0097	-0.0010
2007122012	TOTAL	0.0067	0.0217	0.0047
2007122200	TOTAL	0.0103	-0.0084	-0.0053
2007122312	TOTAL	0.0099	-0.0068	0.0110
2007122500	TOTAL	-0.0020	-0.0065	-0.0059
2007122612	TOTAL	-0.0086	0.0056	-0.0117

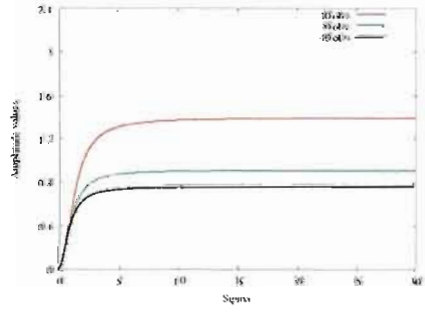
Table 2.4: Correlation coefficient computed for all data types for 18 cases of December 2007. The first singular vector at initial and final time and the pseudo-inverse are used as structure function.



a)



b)



c)

Figure 2.1: Variation of the amplitude increment for different values of associated with the three experiments for which $\mathbf{B} = \sigma^2 \mathbf{v}\mathbf{v}^T$ and the observation error is a) $\sigma_o^2 = 0$, b) $\sigma_o^2 = 1$, and c) $\sigma_o^2 = 4$. In each case, experiments were done with 10, 20 and 40 observations.

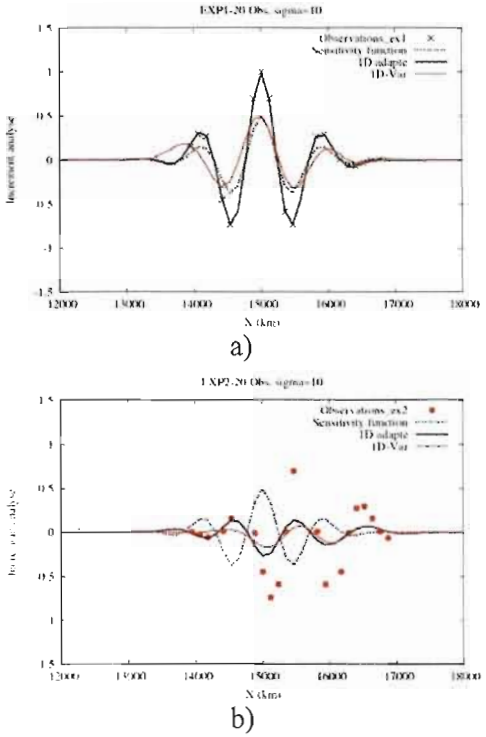


Figure 2.2: Analysis increments obtained with an adapted 1D-Var (in black), and a standard 1D-Var (in red). The sensitivity function used in the analysis is also shown (dotted line). The observations are shown as crosses or dots for the two experiments. In all cases, the adapted 1D-Var used $\sigma = 10$: a) the observations were generated by sampling the sensitivity structure function used in the analysis, b) the observations were generated from a function corresponding to a different structure function.

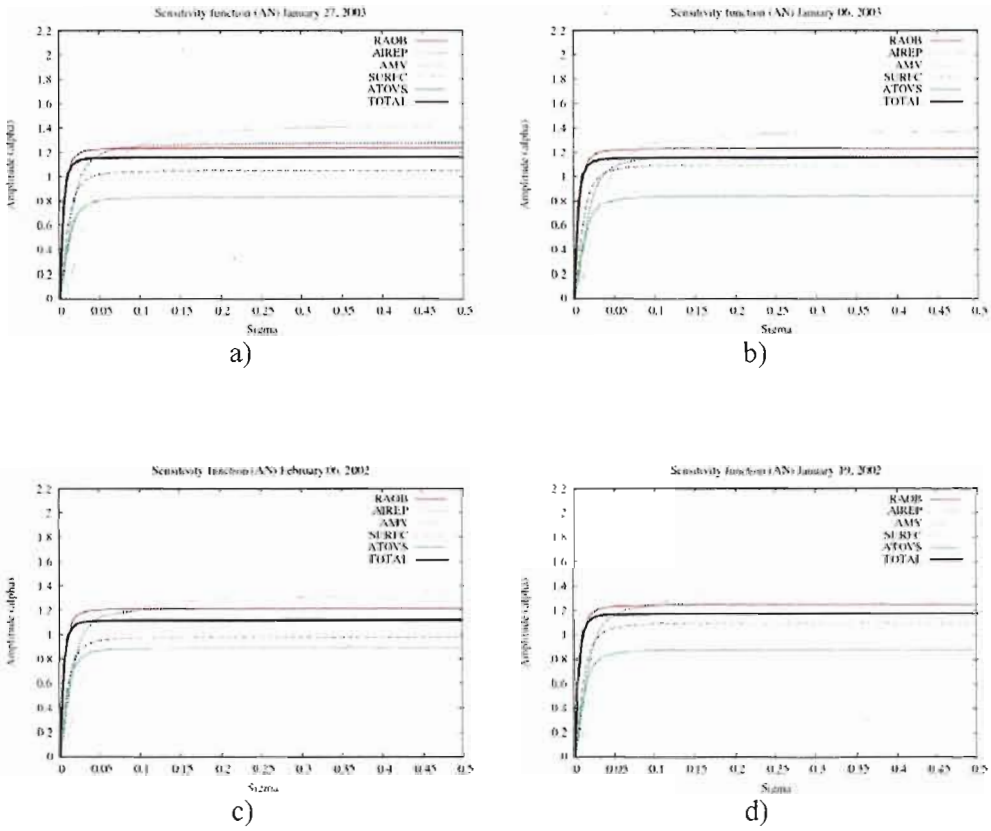


Figure 2.3: Amplitude of the analysis increment as a function of parameter σ for different families of observational data: radio soundings (RAOB), aircraft data (AIREP), wind vectors derived from satellite data (AMV), surface and ship data (SURFC), radiance data from satellite (ATOVS) and all observations combined (TOTAL). The 3D-Var analyses are used as sensitivity function to adapt the background error covariance matrix for 4 case studies: a) January 27, 2003; b) January 06, 2003, c) February 06, 2002, d) January 19, 2002.

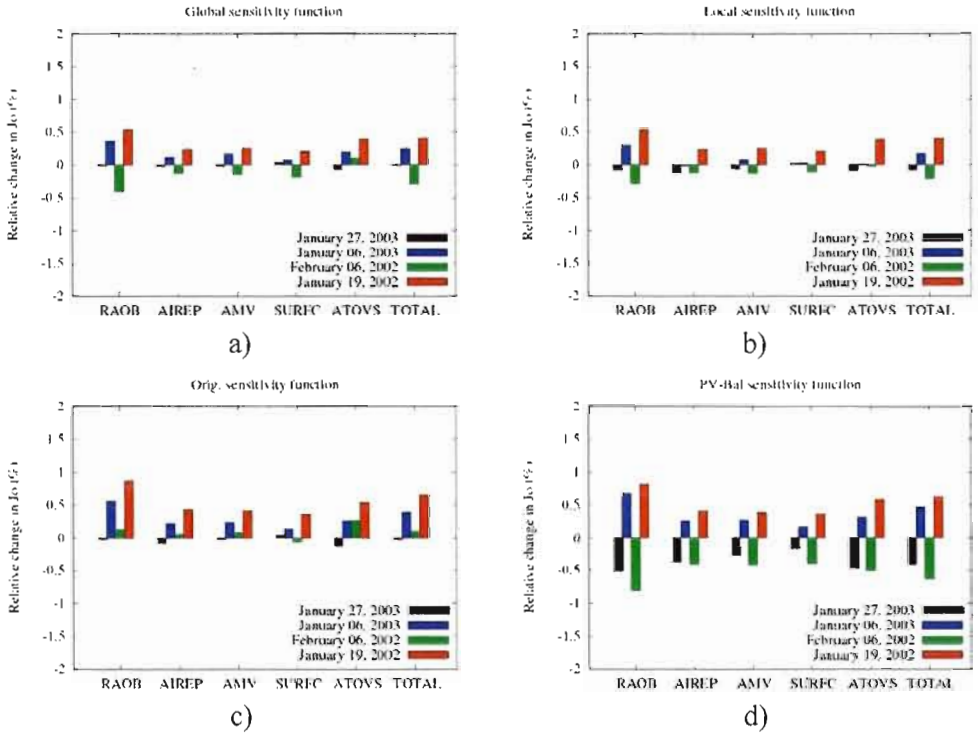


Figure 2.4: Relative change in the global fit to the observations for different families of observational data. The *a posteriori* sensitivity functions are used as structure functions in the *adapted* 3D-Var for 4 case studies. A positive value means that the *adapted* 3D-Var analyses are farther away from the observations than the operational analysis and a negative value means that the *adapted* 3D-Var analyses fit the observations values better than 3D-Var.

CHAPITRE III

EVALUATION DE L'IMPACT DES OBSERVATIONS DANS LES ANALYSES 3D AND 4D-VAR BASÉ SUR LE CONTENU EN INFORMATION

Ce chapitre, rédigé en anglais, est présenté sous la forme d'un article qui sera soumis pour publication dans la revue *Monthly Weather Review*.

Il décrit une méthode pratique qui permet l'estimation du *Degrees of freedom for signal* (*DFS*) à partir des statistiques *a posteriori*. La méthode ne requiert pas la consistance des statistiques d'erreur et l'impact des observations dans les analyses est estimé à partir d'écarts des observations à l'ébauche et à l'analyse. En premier lieu dans l'article, la méthode proposée est introduite et validée dans le cadre plus simple d'une analyse variationnelle 1D (1D-Var) pour être ensuite appliquée pour évaluer le *DFS* des observations assimilées avec le 3D et 4D-Var d'Environnement Canada.

EVALUATION OF THE IMPACT OF OBSERVATIONS ON ANALYSES IN 3D- AND 4D-VAR BASED ON INFORMATION CONTENT

by

Cristina Lupu [*,1], Pierre Gauthier [1] and Stéphane Laroche [2]

[1] *Department of Earth and Atmospheric Sciences*
Université du Québec à Montréal (UQAM)
Montréal (Québec), CANADA

[2] *Meteorological Research Division*
Environment Canada
Dorval (Québec), CANADA

Submitted to *Monthly Weather Review*

8 February 2010
(revised July 21, 2010)

[*] Corresponding author: *Department of Earth and Atmospheric Sciences*
Université du Québec à Montréal (UQAM)
P.O. Box 8888, Succ. Centre-ville
Montréal, Québec
CANADA H3C 3P8

Abstract

The Degrees of Freedom for Signal (*DFS*) is used in data assimilation applications to measure the self-sensitivity of analysis to different observation types. This paper describes a practical method to estimate the *DFS* of observations from *a posteriori* statistics. The method does not require the consistency of the error statistics in the analysis system and it is shown that the observational impact on analyses can be estimated from observation departures with respect to analysis or the forecast. This method is first introduced to investigate the impact of a complete set, or subsets, of observations on the analysis for idealized 1D-Var analysis experiments and then applied in the framework of the 3D and 4D-Var assimilation schemes developed at Environment Canada.

Keywords: Data assimilation, 3D-Var, Degrees of freedom for signal, Information content.

3.1 INTRODUCTION

Recently, new methods have been developed to quantify the impact of the observations on the assimilation and on the ensuing forecast. For the analysis, this can be achieved by evaluating the information content expressed in terms of Degrees of Freedom for Signal (*DFS*) (Purser and Huang, 1993; Rodgers, 2000; Rabier *et al.*, 2002). For any data assimilation system, this diagnostic quantifies information brought by any given type of observations and is useful to assess the relative impact of the different types of observations being assimilated. With the increasing number of datasets used by modern data assimilation systems, such as the hyperspectral infrared sounders AIRS (Atmospheric Infrared Sounder) and IASI (Infrared Atmospheric Sounding Interferometer), it is important to know the information content associated with the radiance measurements which permits to reduce the volume of data associated with these new instruments. An example of how this diagnostic was applied for channel selection procedure is presented in Rabier *et al.* (2002) in the context of IASI simulated data by evaluating the impact of the different channels on the analysis.

Proposed by Cardinali *et al.* (2004), the influence matrix gives a measure of how much any given observation impacted the analysis. They used this approach to estimate the information content supplied by different types of observational data to analyses produced by the ECMWF 4D-Var system. The sensitivity of the analysis to observations then showed that about 25% of the information was provided by ground-based observing systems and 75% by satellite systems. This approach also allows the partial influence of observational subsets to be examined based on geographical area and observation type. Another method was applied with the ARPEGE 4D-Var system of Météo-France (Chapnik *et al.*, 2006). It is based on a method proposed in Girard (1987) in which perturbations to both the background and the observations are introduced to measure the sensitivity of the resulting analysis given the uncertainty in both the observations and the background. In any data assimilation system, the impact on the analysis depends critically on the observation and background error statistics used in the assimilation.

In numerical weather prediction, one is interested to know the impact of the observation on the forecast made from the analysis. Traditionally, the observation impact on forecasts has been obtained from Observing System Experiments (OSEs) in which selected datasets are systematically added or removed from the assimilation system (e.g. Kelly *et al.*, 2007). Using the OSEs, the impact of various observation network configurations can be assessed by comparing forecast scores from experiments that use different observation scenarios. This approach is expensive and only provides a global view of the impact of observations. Recently, adjoint-based sensitivities with respect to observations have also been proposed to assess the observation impact on short-range forecasts without carrying out data-denial experiments (Baker and Daley, 2000; Langland and Baker, 2004; Zhu and Gelaro, 2008; Cardinali, 2009). Zhu and Gelaro (2008) showed that the adjoint-based method provides accurate assessments of the forecast sensitivity with respect to most of the observations assimilated. Gelaro and Zhu (2009) and Cardinali (2009) have recently applied adjoint-based impact calculations to results from OSEs to show that the two methods provide complementary information.

The objective of this paper is to propose a simple approach that permits to easily evaluate the information content associated with observations used in any data assimilation system directly from observation departures from the analysis and forecast, a natural by-product of the assimilation process. The emphasis is then on the impact of observations on the analysis only, not the forecasts. Following Desroziers *et al.* (2005), observation departures from analyses and forecasts can be used to make diagnostics about the consistency of the observation and background-error statistics used in the assimilation. If these error statistics are sub-optimal, they showed that this information can be used to recalibrate the error statistics to meet the χ^2 optimality criteria. The methods are presented in Desroziers and Ivanov (2001) and Chapnik *et al.* (2006). What they show is that observation departures with respect to the background and the analysis are directly related to the observation, background and analysis error covariances. Based on this, they showed that any inconsistency between those diagnostics and the *a priori* error statistics used in the assimilation can be used to recalibrate the observation and background error statistics. As pointed out in Chapnik *et al.* (2006), these relationships show that they can provide an estimate of the information content, provided the error statistics are consistent. What we show in this paper is that these

relationships provide a reliable estimate of the information content as evaluated with the perturbation method of Girard (1987).

An analytic derivation is presented to show how the *DFS* can be evaluated from the *a posteriori* statistics. Section 2 describes the methodology for computing the information content brought by the observations. Based on the results of Desroziers *et al.* (2005), it is shown that the information content brought in by the data assimilation system can be estimated from observation departures from the analysis and forecast, even when the expected statistics of innovation vector differ from those specified in the assimilation system. A unique aspect of the method proposed here is that it does not require the consistency of the error statistics in the analysis system. In section 3, results obtained with the simplified 1D-Var assimilation scheme are presented and discussed. In section 4, this is applied to results from 3D and 4D-Var to show how the impact of observations depends on the assimilation method. Those results were obtained from analyses produced with the 3D-Var and 4D-Var systems of Environment Canada (Gauthier *et al.*, 1999, 2007). Finally, summary and conclusions are given in section 5.

3.2 ESTIMATION OF INFORMATION CONTENT BROUGHT BY THE OBSERVATIONS

Consider a data assimilation scheme that provides an optimal analysis $\mathbf{x}_a$

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H\mathbf{x}_b), \tag{1}$$

where $\mathbf{x}_b$ is the background state, $\mathbf{y}$ is a vector of observational data, H is the nonlinear observation operator while

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1} \tag{2}$$

is the optimal Kalman gain matrix expressed in terms of the background error covariance matrix $\mathbf{B}$, the observation error covariance matrix $\mathbf{R}$ and $\mathbf{H}$ the tangent linear model of H, linearized in the vicinity of $\mathbf{x}_b$.

The Degrees of Freedom for Signal (*DFS*) is used in data assimilation applications to measure the self-sensitivity of analysis to different observation types (Rodgers, 2000). The

DFS is the image in observation space of the trace of the derivative of the analysis with respect to observations, i.e.,

$$DFS = tr\left\{\frac{\partial(\mathbf{H}\mathbf{x}_a)}{\partial \mathbf{y}}\right\}, \tag{3}$$

where $tr\{.\}$ denotes trace of $\{.\}$. In the linear case, (1) and (3) imply that

$$DFS = tr\left\{\frac{\partial(\mathbf{H}\mathbf{x}_a)}{\partial \mathbf{y}}\right\} = tr(\mathbf{K}^T\mathbf{H}^T) = tr(\mathbf{H}\mathbf{K}) \tag{4}$$

Because of the size of the matrices involved, the evaluation of the *DFS* using (4) is not straightforward. Moreover, because the Kalman gain matrix is not readily available in a variational scheme, Cardinali *et al.* (2004) compute an estimate of $tr(\mathbf{HK})$ using the leading singular vectors of the Hessian of the cost function provided by the Lanczos' Conjugate gradient algorithm. Fisher (2003) applied numerical methods for calculating the trace of functions of large matrix to the problem. Another approach is based on a randomization technique proposed by Chapnik *et al.* (2006). In their study, the trace of $\mathbf{HK}$ is evaluated from simple consistency diagnostics introduced by Desroziers *et al.* (2005).

Desroziers *et al.* (2005) developed a set of diagnostics in observation space based on combinations of differences between the observation and the background ($\mathbf{d}_b^o = \mathbf{y} - H(\mathbf{x}_b)$), the observation and the analysis ($\mathbf{d}_a^o = \mathbf{y} - H(\mathbf{x}_a)$) and differences between the background and analysis ($\mathbf{d}_b^a = H(\mathbf{x}_a) - H(\mathbf{x}_b) \equiv \mathbf{d}_b^o - \mathbf{d}_a^o$), the image of the analysis increment in observation space. From these quantities, it is possible to diagnose *a posteriori* observation, background and analysis-error statistics in observation space. The mean diagnostics are the following:

$$E\left[\mathbf{d}_a^o(\mathbf{d}_b^o)^T\right] = \tilde{\mathbf{R}} = \mathbf{R}\mathbf{D}^{-1}\tilde{\mathbf{D}} \tag{5-a}$$

$$E\left[\mathbf{d}_b^a(\mathbf{d}_b^o)^T\right] = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{\mathrm{T}} = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-1}\tilde{\mathbf{D}}, \tag{5-b}$$

$$E\left[\mathbf{d}_b^a(\mathbf{d}_a^o)^T\right] = \mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^{\mathrm{T}} = \mathbf{H}\mathbf{K}\tilde{\mathbf{D}}\mathbf{D}^{-1}\mathbf{R}\ , \tag{5-c}$$

$$E\left[\mathbf{d}_b^o(\mathbf{d}_b^o)^T\right] = \tilde{\mathbf{D}} = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{\mathrm{T}} + \tilde{\mathbf{R}}, \tag{5-d}$$

where $E[\ ]$ is the statistical expectation operator, $\mathbf{D} = \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}$ is the *a priori* innovation covariance, $\tilde{\mathbf{D}} = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}$ is the estimated covariance from innovations and $\mathbf{A}$ is the analysis error covariance. The diagnosed observation and background error covariance in observation space are respectively $\tilde{\mathbf{R}} = E\left[\mathbf{d}_a^o(\mathbf{d}_b^o)^T\right]$ and $\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T = E\left[\mathbf{d}_b^a(\mathbf{d}_b^o)^T\right]$.

It is important to stress that $\mathbf{d}_b^a$ and $\mathbf{d}_a^o$ are related to the innovation vector by

$$\mathbf{d}_b^a = H(\mathbf{x}_a) - H(\mathbf{x}_b) \cong \mathbf{H}\delta\mathbf{x}_a = \mathbf{H}\mathbf{K}\mathbf{d}_b^o, \tag{6-a}$$

$$\mathbf{d}_a^o = \mathbf{y} - H(\mathbf{x}_b + \delta\mathbf{x}_a) \cong (\mathbf{I} - \mathbf{H}\mathbf{K})\mathbf{d}_b^o = \mathbf{R}\mathbf{D}^{-1}\mathbf{d}_b^o\ . \tag{6-b}$$

An expression for the *DFS* can also be derived from these two expressions. The statistical expectation of the outer product of $\hat{\mathbf{d}}_b^a = \mathbf{R}^{-1/2}\mathbf{d}_b^a$ with $\hat{\mathbf{d}}_a^o = \mathbf{R}^{-1/2}\mathbf{d}_a^o$ is

$$\begin{aligned} E\left[\hat{\mathbf{d}}_a^o(\hat{\mathbf{d}}_b^a)^T\right] &= E\left[\mathbf{R}^{-1/2}\mathbf{d}_a^o(\mathbf{d}_b^a)^T\mathbf{R}^{-T/2}\right] \\ &= E\left[\mathbf{R}^{-1/2}\mathbf{R}\mathbf{D}^{-1}\mathbf{d}_b^o(\mathbf{d}_b^o)^T\mathbf{K}^T\mathbf{H}^T\mathbf{R}^{-T/2}\right] \\ &= E\left[\mathbf{R}^{T/2}\mathbf{D}^{-1}\tilde{\mathbf{D}}\mathbf{K}^T\mathbf{H}^T\mathbf{R}^{-T/2}\right] \end{aligned} \tag{7}$$

and therefore,

$$tr\left\{E\left[\hat{\mathbf{d}}_a^o(\hat{\mathbf{d}}_b^a)^T\right]\right\} = tr\left\{E\left[\mathbf{R}^{T/2}\mathbf{D}^{-1}\tilde{\mathbf{D}}\mathbf{K}^T\mathbf{H}^T\mathbf{R}^{-T/2}\right]\right\} \tag{8}$$

By using the property that the statistical expectation and the trace operator commute, i.e. $tr\{E[.]\} = E\{tr[.]\}$ and $tr(\mathbf{a}\mathbf{b}^T) = \mathbf{b}^T\mathbf{a}$ for any two vectors $\mathbf{a}$ and $\mathbf{b}$, (8) reduces to:

$$tr\left\{E\left[\hat{\mathbf{d}}_a^o(\hat{\mathbf{d}}_b^a)^T\right]\right\} = E\left[tr(\hat{\mathbf{d}}_a^o\hat{\mathbf{d}}_b^{aT})\right] = E\left[\mathbf{d}_b^{aT}\mathbf{R}^{-1}\mathbf{d}_a^o\right] = E\left\{tr\left[\mathbf{D}^{-1}\tilde{\mathbf{D}}\mathbf{K}^T\mathbf{H}^T\right]\right\}. \tag{9}$$

Case with consistent error statistics

When the sample covariance matches the prescribed innovation covariance ($\tilde{\mathbf{D}} = \mathbf{D}$), (9) provides an estimation of the information content extracted from observations by an analysis scheme (3D/4D-Var). The globally estimated trace of $\mathbf{HK}$ for all observation types is the total *DFS* then given by

$$DFS^{Globe} = tr(\mathbf{H}\mathbf{K}) = E\left[\mathbf{d}_b^{aT}\mathbf{R}^{-1}\mathbf{d}_a^o\right] \tag{10}$$

Equation (10) gives a simple and efficient way to estimate the *DFS* for an optimal assimilation scheme because only by-products of the data assimilation scheme are necessary.

For many observation types, the observation error covariance matrix $\mathbf{R}$ can be reasonably assumed to be diagonal, and that the observation error is not correlated. There are of course limitations to this assumption but it remains reasonable to a certain extent. Reliable estimates of the information content can be obtained for any subset of the observational data assuming that the observation error of that subset is uncorrelated with that of the other observations. In that case, the partial *DFS* of the k^{th} subset ($\mathbf{y}_k = \mathbf{\Pi}_k^o \mathbf{y}$) extracted from the full observation vector by means of the projection operator $\mathbf{\Pi}_k^o$, is given by:

$$DFS_k^{Globe} = tr(\mathbf{\Pi}_k^o \mathbf{H} \mathbf{K} \mathbf{\Pi}_k^{oT}) = E\left[(\mathbf{d}_b^o)_k^T \mathbf{R}_k^{-1} (\mathbf{d}_a^o)_k\right] \tag{11}$$

Case with inconsistent error statistics

These results hold insofar as the innovation error statistics $\tilde{\mathbf{D}}$ are consistent with those specified in the assimilation, namely that $\tilde{\mathbf{D}} = E\left[\mathbf{d}_b^o \mathbf{d}_b^{oT}\right] = \mathbf{D} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$, so the *a priori* $\mathbf{D}$ and *a posteriori* $\tilde{\mathbf{D}}$ term in (9) cancel out. However, as pointed out by Desroziers *et al.* (2005), if they differ, the diagnosed covariance matrices in (5-a) and (5-b) may be seen as some adjusted covariance estimates. The *a posteriori* Kalman gain matrix is now defined as

$$\tilde{\mathbf{K}} = \tilde{\mathbf{B}}\mathbf{H}^T (\tilde{\mathbf{R}} + \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T)^{-1} = \tilde{\mathbf{B}}\mathbf{H}^T \tilde{\mathbf{D}}^{-1} . \tag{12}$$

Therefore, the estimate of $tr(\mathbf{H}\tilde{\mathbf{K}})$ from the *a posteriori* statistics is

$$\tilde{DFS} = tr(\mathbf{H}\tilde{\mathbf{K}}) = tr\left[\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T \tilde{\mathbf{D}}^{-1}\right] = tr\left[\mathbf{H}\mathbf{B}\mathbf{H}^T \mathbf{D}^{-1}\right] = tr(\mathbf{H}\mathbf{K}) = DFS \tag{13}$$

where $\tilde{\mathbf{D}}^{-1}$ denotes the pseudo-inverse of $\tilde{\mathbf{D}}$. A generalization of the usual inverse matrix (Golub and van Loan, 1996) must be used here because $\tilde{\mathbf{D}}$ may be singular. It follows that the information content can be determined either from the *a posteriori* statistics or from the *a priori* statistics.

A more interesting form can be obtained by introducing (5-b) in (13). Using then the properties that the trace and expectation operators commute and that $\mathbf{X}E\left[(.)\right]=E\left[\mathbf{X}(.)\right]$ for any non-random matrix $\mathbf{X}$, then leads to the following result:

$$\begin{aligned} \tilde{DFS} &= tr\left[\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T\tilde{\mathbf{D}}^{-1}\right]=tr\left[E\left\{\mathbf{d}_b^a\mathbf{d}_b^{oT}\right\}\tilde{\mathbf{D}}^{-1}\right] \\ &= E\left\{tr\left[\mathbf{d}_b^a\mathbf{d}_b^{oT}\tilde{\mathbf{D}}^{-1}\right]\right\}=E\left\{\mathbf{d}_b^{oT}\tilde{\mathbf{D}}^{-1}\mathbf{d}_b^a\right\} \end{aligned} \tag{14}$$

In other words, the *DFS* associated with any assimilation system can be directly obtained from $\tilde{\mathbf{D}}=E\left[\mathbf{d}_b^o(\mathbf{d}_b^o)^T\right]$ and (14).

The equivalence established here states that the *DFS* evaluated using diagnostics of $E\left[\mathbf{d}_b^a(\mathbf{d}_b^o)^T\right]=\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T$ and $E\left[\mathbf{d}_b^o(\mathbf{d}_b^o)^T\right]=\tilde{\mathbf{D}}$ yields the same results as if a perturbation method was used to evaluate the *DFS* associated with the *a priori* error statistics. This is the method proposed by Chapnik et al. (2006). Inspection of (5-a, b) indicates that $\tilde{\mathbf{R}}=\mathbf{R}\mathbf{D}^{-1}\tilde{\mathbf{D}}$ and $\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T=\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{D}^{-1}\tilde{\mathbf{D}}$ differs from their a priori definition by the same factor, $\mathbf{D}^{-1}\tilde{\mathbf{D}}$. When using those *a posteriori* definitions, those factors cancel out to retrieve the same *DFS* as would be obtained using the *a priori* error statistics.

A difficulty remains however, since (14) requires that $\tilde{\mathbf{D}}$ be inverted, which is not immediate as it embeds both the observation error and the background error. The latter cannot be assumed to be uncorrelated which makes $\tilde{\mathbf{D}}$ non-diagonal. However, an alternative approach can be taken to simplify the computation. The analysis sensitivity matrix, introduced in Cardinali *et al.* (2004), being $\mathbf{S}=\mathbf{K}^T\mathbf{H}^T$, it can also can be defined with respect to the *a posteriori* statistics. Using (5-a) and (5-c), it is easily shown that

$$\tilde{\mathbf{R}}^{-1}(\mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^T)^T=\tilde{\mathbf{R}}^{-1}(\mathbf{H}\mathbf{K}\tilde{\mathbf{D}}\mathbf{D}^{-1}\mathbf{R})^T=\mathbf{K}^T\mathbf{H}^T=\mathbf{S}.$$

and consequently,

$$\tilde{DFS}=tr(\tilde{\mathbf{K}}^T\mathbf{H}^T)=tr\left[\tilde{\mathbf{R}}^{-1}(\mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^T)^T\right]=tr(\mathbf{K}^T\mathbf{H}^T)=DFS \tag{15}$$

Substituting (5-c) into (15), the *a posteriori* $\tilde{DFS}$ can be rewritten as

$$\begin{aligned} \tilde{DFS} &= tr\left(\tilde{\mathbf{K}}^T\mathbf{H}^T\right) = tr\left\{\tilde{\mathbf{R}}^{-1}E\left[\mathbf{d}_b^o\mathbf{d}_a^{oT}\right]^T\right\} = tr\left\{E\left[\tilde{\mathbf{R}}^{-1}\mathbf{d}_a^o\mathbf{d}_b^{aT}\right]\right\} \\ &= E\left\{tr\left[\tilde{\mathbf{R}}^{-1}\mathbf{d}_a^o\mathbf{d}_b^{aT}\right]\right\} = E\left\{\mathbf{d}_b^{aT}\tilde{\mathbf{R}}^{-1}\mathbf{d}_a^o\right\} \end{aligned} \tag{16}$$

This has the same form as (10) except that the estimated observation error covariance matrix $\tilde{\mathbf{R}}$ is to be used. This matrix is possibly non-diagonal full matrix and, in general, may not be symmetric and contain cross-correlations due to the presence of background error in its estimate, as indicated by (5-a). To calculate the generalized inverse, $\tilde{\mathbf{R}}^{-1}$, a Singular Value Decomposition (SVD) of the matrix $\tilde{\mathbf{R}}$ can be used by decomposing $\tilde{\mathbf{R}} = \mathbf{U}\Lambda\mathbf{V}^{\mathrm{T}}$, where $\mathbf{U}$ and $\mathbf{V}$ denotes the matrices formed by the left ($\mathbf{U}$) and right ($\mathbf{V}$) singular vectors while Λ is a diagonal matrix defined by the singular values. In that case, $\tilde{\mathbf{R}}^{-1} = \mathbf{V}\Lambda^{-1}\mathbf{U}^{\mathrm{T}}$ and the *DFS* in (16) can be evaluated at the cost of a few dot products. This would also be the approach to take to evaluate $\tilde{\mathbf{D}}^{-1}$ to compute the *DFS* using (14).

For many observation types like radiosondes and ground-based instruments, the observation error is uncorrelated between distinct observations. We then introduce the assumption that $\tilde{\mathbf{R}}$ can be approximated as a block-diagonal matrix, each being of the form $\tilde{\mathbf{R}}_k \cong \tilde{\sigma}_o^2(k)\mathbf{I}_k$ where $\tilde{\sigma}_o^2(k)$ is the diagnosed observation error variance associated with the k^{th} observation type. This is justified when the observation error is expected to be uncorrelated for observations coming from independent instruments. This is the case for several observation types such as radiosondes and ground-based instruments but may not be valid for measurements from satellite instruments. As stated in Talagrand (1999), the approach for computing the *a posteriori* covariances cannot provide any new information about $\mathbf{R}$ and $\mathbf{B}$ without imposing an external hypothesis to disentangle the observation and background error embedded within the innovation error statistics. An important fact is that only the observation error variances are extracted from the diagnosed statistics $\tilde{\mathbf{R}}$ by assuming that the observation error is uncorrelated. This is where the evaluation of the *DFS* using (16) shows a clear advantage over (14): the matrix to be inverted can be assumed to be diagonal. This remains to be verified however.

In the next section a simple system 1D-Var is used to investigate the extent to which this assumption is a reasonable one in an idealized context in which ensemble of analyses can be generated.

3.3 APPLICATION TO 1D-VAR DATA ASSIMILATION SYSTEM

Using the methodology presented in the previous section we discuss the estimation of the *DFS* with a simplified 1D-Var assimilation scheme. The 1D domain contains N=256 points uniformly distributed over a circle of latitude (approximately at 41° latitude) with perimeter of 30,000 km. The true background-error covariance matrix $\mathbf{B}_t$ in physical space assumes isotropic error correlations is defined as

$$B_t(i,j) = \sigma_{b(t)}^{(i)} \sigma_{b(t)}^{(j)} \exp(-r_{ij}^2 / 2L_t^2) \tag{17}$$

where $\sigma_b^{(i)}$, $\sigma_b^{(j)}$ are the true background standard deviation of component i and j of $\mathbf{B}_t$ respectively ($\sigma_{b(t)}^2 = 1.$), r_{ij} is the Euclidean distance between points i and j and L_t is the true horizontal length-scale taken to be 300 km. In our experiments, we consider three different values of the background correlation length (300, 500 and 1000 km) in the *a priori* background error statistics. The observing system is fixed to be 60 observations at every other three-grid point. The observations are simulated by adding Gaussian random noise to the truth and the innovation vector $\mathbf{y}'$ is defined as $\mathbf{y}' = \mathbf{y} - H(\mathbf{x}_b) \cong \boldsymbol{\varepsilon}_o - \mathbf{H}\boldsymbol{\varepsilon}_b$, where $\boldsymbol{\varepsilon}_b$ and $\boldsymbol{\varepsilon}_o$ represent the errors in the background state and the observations respectively. Every observation is taken directly as a value at a grid point and all the observations have the same error variance. Therefore, $\mathbf{R}_t$ is defined as $\mathbf{R}_t = \sigma_{o(t)}^2 \mathbf{I}$, with $\mathbf{I}$, the identity matrix and $\sigma_{o(t)}^2 = 4$, the true observation error variance. In this context, it is possible to repeat the analysis for a number of realizations based on the true observation and background error which may differ from the *a priori* statistics used in the assimilation. Based on the true error statistics, an ensemble of 2000 analyses was produced to estimate the *a posteriori* error statistics.

3.3.1 *Estimation of the off-diagonal terms in the observation error covariance*

The first experiment is to examine whether the *a posteriori* estimate of observation error covariance can be assumed to be diagonal and their importance for the definition of $\tilde{\mathbf{R}}$ is discussed in this section.

The non-diagonal elements of $\tilde{\mathbf{R}}$ were estimated using (5-a) assuming the observation error variance to be the same for all the 60 observations used in this experiments. Moreover, the error covariance is assumed to be identical when the distance between the observations is the same. The observation-error covariance $\tilde{\mathbf{R}}(i,j)$ between components i and j as a sample mean is given by:

$$\tilde{\mathbf{R}}(i,j) = \overline{(\mathbf{d}_a^o)_i (\mathbf{d}_b^o)_j^T}\,, \tag{18}$$

where the overbar represent the sample mean for the whole ensemble of 2000 analyses. With consistent error statistics, the observation and background error variances are perfectly known, i.e., the specified values are $\sigma_o^2 = \sigma_{o(t)}^2 = 4$, $\sigma_b^2 = \sigma_{b(t)}^2 = 1$ but different values for the horizontal length-scale L_c =300 km, 500 km and 1000 km were used. For all cases, the magnitude of the off-diagonal elements in the observation-error covariances is very small compared with those of the diagonal components of each element of $\tilde{\mathbf{R}}$. Figure 3.1 shows a representation of $\tilde{\mathbf{R}}(i,j)$ as a function of distance r_{ij} between points i and j. The examination of the off-diagonal elements in the observation-error covariance matrix reveals small values (below 10%). This shows that the diagnosed observation-error covariance matrix $\tilde{\mathbf{R}}$ may be considered diagonal ($\tilde{\mathbf{R}} \cong \tilde{\sigma}_o^2 \mathbf{I}$).

3.3.2 *Degrees of freedom for signal (DFS)*

In a second set of experiments, the *DFS* is evaluated using the *a posteriori* statistics and compared with that obtained using the perturbation method (Girard, 1987). Results are also shown when the *a posteriori* diagnostics are evaluated using either (14) or (16). Finally, the *DFS* is estimated using (16) but retaining only the estimated diagonal elements of $\tilde{\mathbf{R}}$. The objective of this experiment is to show that the information content estimated from the *a*

posteriori and *a priori* statistics concur. One has to keep in mind that the *DFS* estimated is a reflection of the error statistics used in the assimilation. The *DFS* estimated from the true statistics gives what would be obtained if the error statistics of the assimilation were consistent with the estimation based on observation departures from the background state and the analysis.

In general, the direct evaluation of $tr(\mathbf{HK})$ is not straightforward because the Kalman gain matrix is not explicitly available in a variational data assimilation system. However, the calculation of this trace in a simplified 1D-Var context considered here can be done. In particular, assuming that $\mathbf{R}$ and $\mathbf{B}$ are the covariances used in the assimilation, the theoretical *DFS* can be evaluated as:

$$DFS_{ANALYTIC} = tr(\mathbf{HK}) = tr\left[\mathbf{HBH}^T(\mathbf{R}+\mathbf{HBH}^T)^{-1}\right] \tag{19}$$

in which $\mathbf{K}$ is the gain matrix.

For more complex systems, Girard (1987) proposed a randomization method to approximate the trace of a matrix only known as a composition of operators. A practical method that requires a random perturbation of the vector of observations was introduced in Desroziers and Ivanov (2001) and was employed in Chapnik *et al.* (2006). It can be shown that a randomized estimation of $tr(\mathbf{HK})$ where $\mathbf{K}$ is based on the specified $\mathbf{R}$ and $\mathbf{B}$ covariances, that were used in the analysis is given by

$$DFS_{GIRARD} = tr(\mathbf{HK}) = (\mathbf{y}^* - \mathbf{y})^T\mathbf{R}^{-1}(\mathbf{Hx}_a^* - \mathbf{Hx}_a), \tag{20}$$

where $\mathbf{Hx}_a^*$ and $\mathbf{Hx}_a$ contain the analysis increments obtained from perturbed and unperturbed observations, respectively. The observations are perturbed by adding small perturbations $\boldsymbol{\varepsilon}_o = \mathbf{R}^{1/2}\boldsymbol{\xi}$ to the original set of observations $\mathbf{y}^* = \mathbf{y} + \mathbf{R}^{1/2}\boldsymbol{\xi}$, where $\boldsymbol{\xi}$ is a vector of random numbers with zero mean and unit variance.

In our study, the argument we propose is that the *DFS* can be computed directly from observation departures from the analysis and forecast. Relying on expressions (13) and (15), the *DFS* can be also evaluated from (14) using the *a posteriori* statistics:

$$\tilde{DFS}^{(1)}_{\text{APOSTERIORI}} = tr\left[\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T\tilde{\mathbf{D}}^{-1}\right] = E\left[\mathbf{d}_b^{oT}\tilde{\mathbf{D}}^{-1}\mathbf{d}_b^a\right] \tag{21}$$

or, equivalently, using (16),

$$\tilde{DFS}^{(2)}_{\text{APOSTERIORI}} = tr\left[\tilde{\mathbf{R}}^{-1}(\mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^T)^T\right] = E\left[\mathbf{d}_b^{aT}\tilde{\mathbf{R}}^{-1}\mathbf{d}_a^o\right] \tag{22}$$

The question then becomes which *a posteriori* relation should be used? In particular, for (21), the inversion of $\tilde{\mathbf{D}}$ may be complicated by the fact that $\tilde{\mathbf{D}}$ may be singular. By replacing the *a posteriori* observation-error covariance matrix $\tilde{\mathbf{R}}$ by a diagonal matrix $\tilde{\mathbf{R}} \cong \tilde{\sigma}_o^2\mathbf{I}$, in this case, (22) simplifies to:

$$\tilde{DFS}_{\text{DIAG}} = E\left[\mathbf{d}_b^{aT}\tilde{\mathbf{R}}^{-1}\mathbf{d}_a^o\right] \cong E\left[\frac{\mathbf{d}_b^{aT}\mathbf{d}_a^o}{\tilde{\sigma}_o^2}\right] \tag{23}$$

In the following experiment, the *DFS* has been estimated from 2,000 analyses. This is compared with the *DFS* computed with Girard's method (20) and the *DFS* calculated using the *a posteriori* statistics as introduced in (21)-(23).

Table 3.1 shows the estimates of *DFS* obtained with the true background and observation errors. In this case, the estimated $\mathbf{K}$ is equal to the true Kalman gain matrix. The *a posteriori* estimate of *DFS* is similar (within 0.1% accuracy) with that found from Girard's method and in good agreement with the analytic value. Since the *DFS* is a function of $\mathbf{B}$, the horizontal correlations have an influence as can be seen in the results of Table 3.1. They indicate that the *DFS* tends to decrease when the horizontal length-scale increases.

The second set of experiments is similar to the previous one except that the observation error variance is now underestimated and taken to be $\sigma_o^2 = 2.25$. The results, shown in Table 3.2, are similar to that of Table 3.1. Similarly, Table 3.3 presents the results obtained when both the background and observation error variances are underestimated ($\sigma_b^2 = 0.25$ and $\sigma_o^2 = 2.25$, respectively). In both experiments, the *DFS* calculations using the full estimate of the *a posteriori* observation-error covariance matrix $\tilde{\mathbf{R}}$ give similar results to that obtained using the randomized Girard method. In addition, the *DFS* results considering only the diagonal elements of $\tilde{\mathbf{R}}$ give also a good approximation of the information content of observations.

The conclusions from these experiments are now summarized. When the a priori error statistics differ from those estimated from observation departures, the estimated observation-error covariance matrix might show cross-correlations due in part to the presence of background error in its estimate. In this study, the non-diagonal elements of $\tilde{\mathbf{R}}$ were shown to be small, so that the diagnosed $\tilde{\mathbf{R}}$ matrix can be approximated as a diagonal matrix. The idealized experiments with the 1D-Var show that it is possible to obtain the appropriate value for the *DFS* from *a posteriori* statistics. The results indicate in all experiments that the information content estimated from the *a posteriori* and *a priori* statistics provide quite similar results. A simple method has been introduced in which the estimated observation error covariances are assumed to be diagonal. The results obtained are also found to be in good agreement with the method proposed by Girard (1987) and Chapnik *et al.* (2006).

3.4 EVALUATION OF THE INFORMATION CONTENT IN 3D-VAR AND 4D-VAR

In this section, the diagnostics introduced in the previous section are used to evaluate the *DFS* from the 3D-Var and 4D-Var data assimilation systems of the Meteorological Service of Canada (MSC). The 3D-Var and 4D-Var experiments used in this study are those described in Laroche and Sarrazin (2010-a, b). The 3D-Var and 4D-Var systems have been cycled over the period 21 December 2006 to 28 February 2007 using a 6-h assimilation window. All diagnostics exclude the first 11 days, the spin-up period of the analysis. The incremental 4D-Var is used (Gauthier *et al.*, 2007) in which the analysis increment is calculated at a lower horizontal resolution (~170 Km). The 4D-Var analysis is obtained after two outer loops by interpolating this lower resolution analysis increment to the same grid (~35 Km) as the background state before adding the two. The subsets of observations assimilated in either 3D or 4D-Var during winter 2006-2007 include radiosoundings (RAOB), aircraft data (AI), surface and ship data (SF), wind profiler data (PR), atmospheric motion vectors from geostationary satellites (AMV) and radiances from polar-orbiting satellite (AMSU-A and AMSU-B) and from geostationary satellites (GOES-East and GOES-West). A summary is given in Table 3.4.

3.4.1. A posteriori diagnostics and consistency checks

The variational data assimilation formulation relies on a number of hypotheses on the background and observation-error statistics. The validity of these hypotheses is an important factor in determining the optimality of the analysis. The chi-square χ^2 diagnostic can be used to check if the sample covariances of innovations in a region, or for a given observing system, are very different from what has been prescribed. For data assimilation χ^2 is defined as

$$\chi^2 = \mathbf{d}_o^{bT}\mathbf{D}^{-1}\mathbf{d}_o^b,$$

and its expected value is $E[\chi^2] = tr(\mathbf{D}^{-1}\tilde{\mathbf{D}})$. Assuming that $\mathbf{D} = \tilde{\mathbf{D}}$, then $E[\chi^2] = p$ where p is the total number of observations used in the analysis. In 3D-Var/4D-Var, χ^2 can be obtained from the value of the cost function at minimum, which is

$$E[J_{\min}] = E[J(\mathbf{x}_a)] = \frac{1}{2}E[\mathbf{d}_o^{bT}\mathbf{D}^{-1}\mathbf{d}_o^b] = \frac{1}{2}E[\chi^2] = \frac{p}{2} \tag{24}$$

Eq. (24) provides a simple diagnostic to check the global consistency of an assimilation algorithm. In Table 3.5, the average over January and February 2007 of the estimated values of χ^2 in 3D-Var and 4D-Var systems are shown and compared to the number of observations. The expected value of χ^2 / p is less than 1, which implies that either the background or observation error variances, or both, have been overestimated. In fact, in both 3D-Var and 4D-Var systems the observation-error variances for many satellite radiances are over-estimated on purpose by using the innovation variance as an estimate for the observation-error variance and consequently, the value for chi-square decrease.

The consistency diagnostic has been calculated for the observation and background-error covariances as in (5-a) and (5-b). Results confirm the overestimation of error statistics for most observation types. However, the objective was not to recalibrate the error statistics and to improve the observation and background-error variances. The objective is to show that *a posteriori* diagnostics can be used to estimate the information content associated with different subsets of observations. In the previous section, it has already been shown that the

DFS can be estimated either from the randomized Girard's scheme or from *a posteriori* diagnostics.

3.4.2. Computation of DFS in MSC's 3D-Var and 4D-Var

Figure 3.2 presents estimates of the total *DFS* averaged over January-February 2007 in the MSC 3D-Var and 4D-Var data assimilation systems for the following regions: the Globe, Northern Hemisphere (20N-90N), Tropics (20S-20N) and Southern Hemisphere (90S-20S). These results indicate that the *DFS* for 3D-Var is larger than for 4D-Var over all regions. This is surprising as 4D-Var is expected to extract more information from satellite data than 3D-Var. This will warrant further investigation. On average, the innovations in 4D-Var are smaller than those of 3D-Var, indicating that the sum of the observation and background error covariances is smaller. In particular, it would be interesting to recalibrate the background and observation error statistics and assess the impact this may have on the observation impact. This is beyond the scope of the present study.

The *DFS* for the main data types was also evaluated. The partial sum DFS_k^{Region} represents the relative importance of the k^{th} subset of observations over a region. For instance, if the region is the whole globe, it is defined as

$$DFS_k^{Globe}(\%) = 100 \cdot \frac{DFS_k^{Globe}}{DFS^{Globe}}$$

and represents the ratio of the DFS_k^{Globe} obtained from a particular subset of observations to the total DFS^{Globe} extracted from all observations. Expressed as a percentage, it then represents the relative contribution of any subset of observations to the global *DFS*. More generally, for a particular region, DFS_k^{Region} of different observation types can be written as:

$$DFS_k^{Region}(\%) = 100 \cdot \frac{DFS_k^{Region}}{DFS^{Globe}}$$

In fact, one could aggregate the observations not only by type but also in different ways, assuming that the observation errors are uncorrelated between subsets, and also that there is enough data to insure statistical significance.

Figure 3.3 shows the *DFS* percentages in the 3D-Var and 4D-Var by observation type over the globe. Results show that the most important observations in terms of information content in the analyses are radiosonde and brightness temperature data types (AMSU-A/B) followed by aircraft data. Different results have been obtained at the European Centre for Medium-range Weather Forecasts (ECMWF: Cardinali *et al.,* 2004) where satellite observations (AMSU-A, HIRS and SSMI) contribute more to the *DFS* than conventional observations. The MSC 3D/4D-Var relies on a smaller number of satellite data as compared to ECMWF. It is also observed that radiosonde, wind profiler, aircraft and AMSU-B data have more relative impact in 4D-Var than in 3D-Var. In the Northern Hemisphere, the largest *DFS* are obtained for radiosonde and aircraft data while satellite radiances are dominant in the Southern Hemisphere. Again, this may be indicative of a need to do a recalibration of the error statistics.

For any selected subset of data, the observation influence (OI), is defined as the *DFS* normalized by the number of observations:

$$OI(\%) = 100 \cdot \frac{DFS_k^{Region}}{p_k}.$$

Figure 3.4 shows the impact of individual observations in both 3D-Var and 4D-Var. We note that the observation influence is larger for the radiosonde data in both data assimilation systems. All other data types show a much smaller impact *per* observation. We also note that the AMSU-B data have a mean influence larger than the AMSU-A data. Information in AMSU-B data is with respect to humidity while the AMSU-A's channels are sensitive to high tropospheric and low stratospheric temperature variations.

Figure 3.5 shows the $DFS_k^{Globe}(\%)$ for different AMSU-A/B channels. The number of assimilated radiance channels in our system is seven from an AMSU-A instrument (channels 4-10) and four from an AMSU-B instrument (channels 2-5). In particular, the weighting functions of channels 9 and 10 from AMSU-A peak around 50-100 hPa and a fraction of their weighting function is above the model top. We note that a large part of the *DFS* is coming from stratospheric AMSU-A channel 10. In 4D-Var, AMSU-A channel 9 sounding in the high atmosphere has also a relatively large negative *DFS* partly due to the fact that the

observation-errors are misrepresented and may be also biased as it is sensitive to a region near the model lid. The method proposed in this paper assumes that observation departures are unbiased which may not be exactly verified in the results obtained from an operational system.

Figure 3.6 shows the information content, for the main data types in the 3D-Var and 4D-Var as a function of the observation time within the assimilation window. The regions represented here are the Northern and Southern hemispheres. The results suggest that radiosonde and surface pressure data have the largest *DFS* near the middle of the assimilation window as most of the data are available at the synoptic time. On the other hand, the satellite-data are roughly evenly spread across the assimilation window but have the largest *DFS* at the end of the assimilation window. A negative impact is observed for AMSU-A in both the 3D and 4D-Var results. This may be a signature of a model bias which experiences a spin-up that influences mostly humidity. However, the negative impact is not seen in the Southern Hemisphere which may be indicative that the assimilation systems draws too much towards the conventional observations at the expense of a good fit to satellite data. This would require further experimentation to identify the cause of this negative impact.

Contrary to results obtained with adjoint sensitivity of forecast error to observations (Morneau *et al.,* 2006), observations at the end of the assimilation window are more constraining on forecast error growth than those at the beginning of the assimilation window and 4D-Var is able to capture that but not 3D-Var. What the *DFS* measures is the impact on the analysis and the results show here indicate that there is no significant difference in the impact of observations as a function of observation time between the results from 3D and 4D-Var.

3.5 CONCLUSIONS

As described in this paper, there are a number of approaches that have recently been used to evaluate the value of observations in data assimilation systems. The *DFS* is used in data assimilation applications to indicate the self-sensitivity of analysis to different observation types. In this paper, a new method to assess the information content of observation on

analyses is presented and applied to calculate the *DFS* of a complete set, or subsets, of observations in the MSC's 3D-Var and 4D-Var systems. Based on the results of Desroziers *et al.* (2005), it is shown that the information content brought in by the data assimilation system can be estimated from observation departures from the analysis and the background state. The main point made in this paper is that even though the error statistics may not be consistent, the observation departures can still be used to measure the information content in observations associated with the *a priori* error statistics used in the assimilation. These *a posteriori* estimates were inspired by the results of Desroziers *et al.* (2005). It was shown here that by introducing the additional assumption that the observation error is uncorrelated, then the method is easily applicable as a diagnostic of the results produced by any data assimilation system. One has to be aware that it is implicitly assumed that the observation departures are unbiased which may not be verified. A simplified 1D-Var assimilation system was used to test the validity of the methods and the results then confirmed that the estimates obtained agree with a method proposed by Girard (1987). With error statistics differing from the true ones, it was shown that the *a posteriori* estimates of the observation error is reasonably diagonal which justifies the hypothesis made on the *a posteriori* estimate of the observation error covariances.

The diagnostics developed in this work were then applied to results obtained from MSC's 3D-Var and 4D-Var assimilation systems. The partition by observation types allows diagnosing the relative influence on the analysis of different observing systems. In terms of *DFS*, the results suggest that radiosondes are the most influential data type of the global observing system, followed by brightness temperature data types (AMSU-A/B) and aircraft data. It is worth mentioning that the largest observation influence is provided by radiosonde and AMSU-B data. It has already been shown that the *DFS* is useful to evaluate the sensitivity of the analysis to different channels for a particular radiometer. The estimation of the *a posteriori* error standard-deviations for satellite radiances indicate that the errors are generally over-estimated in the MSC's 3D-Var and 4D-Var assimilation schemes. It is however planned to more carefully investigate the *a posteriori* estimation of the observation-error variance for radiometers channels sounding in the high atmosphere where *a priori* satellite radiance errors were overestimated to a larger extent than other error variances.

As shown in the paper, the estimation of the *DFS* for any assimilation system can be easily implemented and estimated *a posteriori* from the proposed diagnostics based only on by-products from the data assimilation system. The results shown in the paper indicate some deficiencies in the current estimate of the error statistics used in the assimilation. Future work will have to be done to recalibrate the error statistics to reflect changes brought to the system. These diagnostics will be used to evaluate the information content of a complete set, or subsets, of observations on the 4D-Var data assimilation scheme that was implemented operationally in 2008. The number of ATOVS and AMVs observations was increased in the new system and new observation types as AIRS, SSM/I (clear sky radiances) and QuickScat Seawinds are now assimilated. We should also mention that calculations of information content are dependent on the error covariance matrices $\mathbf{B}$ and $\mathbf{R}$. This could be exploited in the evaluation of information content of the observations by using different formulations for $\mathbf{B}$. The simplicity of the method stems from the assumption that the observation error is uncorrelated. A particular challenge is to consider observation error correlation structure in variational data assimilation. The major impact of including observation-error correlation structure in data assimilation algorithms is the inversion of a non-diagonal full observation error covariance matrix. However, this issue is important to improve the assimilation of satellite data.

ACKNOWLEDGMENTS

The authors would like to thank Mr. Pierre Koclas of the Meteorological Service of Canada who helped us in the utilization of the observation database to build the diagnostics in observation space used in this study. Environment Canada provided partial financial support for this study on top of the computing facilities and technical assistance for the use of their assimilation system. They also would like to thank Drs. Mark Buehner from Environment Canada and Carla Cardinali from ECMWF for their comments and discussions which contributed to significant improvements to the paper. Comments from an anonymous reviewer are also acknowledged and helped to improve the paper as well.

This work has been funded in part by Grant 500-b of the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) for the project on the *Impact of Observing*

Systems on Forecasting Extreme Weather in the short, medium and extended range: A Canadian contribution to THORPEX, with additional support from Discovery Grant 357091 of the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES

Baker N. L. and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptative observation-targeting problem. *Q. J. R. Meteorol. Soc.*, **126**, 1431-1454.

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

——, 2009: Monitoring the observation impact on the short-range forecast. *Q. J. R. Meteorol. Soc.*, **135**, 239-250.

Chapnik B., G. Desroziers, F. Rabier and O. Talagrand, 2006: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Q. J. R. Meteorol. Soc.*, **132**, 543–565.

Desroziers G. and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q. J. R. Meteorol. Soc.*, **127**, 1433–1452.

Desroziers G., L. Berre, B. Chapnik and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.*, **131**, 3385–3396.

Fisher M., 2003: Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *ECMWF Tech. Memo.*, **397**, pp. 18.

Gauthier, P., C. Charette, L. Fillion, P. Koclas and S. Laroche, 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmos.-Ocean*, **37**, 103–156.

——, M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau, 2007: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Wea. Rev.*, **135**, 2339–2354.

Gelaro, R. and Y. Zhu, 2009: Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179-193.

Golub, G. H. and C. F. Van Loan, 1996: Matrix computations. The Johns Hopkins University Press, Baltimore, 728 pp.

Girard, D., 1987: A fast Monte Carlo cross-validation procedure for large least squares problems with noisy data. Tech. Rep. 687-M, IMAG, Grenoble, France, 22pp.

Kelly, G., J. N. Thépaut, R. Buizza and C. Cardinali, 2007: The value of observations. I: Data denial experiments for the Atlantic and the Pacific. *Q. J. R. Meteorol. Soc.*, **133**, 1803-1815.

Laroche, S. and R. Sarrazin, 2010a: Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part I: contribution of radiosonde, aircraft and satellite data. *Atmos.-Ocean*, **48**, 10-25.

——, and R. Sarrazin, 2010b: Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part II: Sensitivity experiments. *Atmos.-Ocean,* **48**, 26-38.

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

Morneau, J., S. Pellerin, S. Laroche and M. Tanguay, 2006: Estimation of adjoint sensitivity gradients in observation space using the dual (PSAS) formulation of the Environment Canada operational 4D-Var. Proceedings of the *Second THORPEX International Science Symposium*, 4-8 December 2006, Landshut, Germany, *WMO/Technical Document*, **1355**, 162-163.

Purser, R. J. and H. L. Huang, 1993: Estimating effective data density in a satellite retrieval or an objective analysis. *J. Appl. Meteorol.*, **32**, 1092-1107.

Rabier, F., N. Fourrié, D. Chafai and P. Prunet, 2002: Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Q. J. R. Meteor. Soc.*, **128**, 1011-1027.

Rodgers, C., 2000: Inverse Methods for Atmospheric Sounding Theory and Practice. World Scientific Publishing, London, 256 pp.

Talagrand, O., 1999: A posteriori evaluation and verification of the analysis and assimilation algorithms. In *Proceedings of Workshop on Diagnosis of data assimilation systems* (November 1998), ECMWF, Reading, UK, pp.17-28.

Zhu Y. and R. Gelaro 2008: Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Mon. Weather Rev.* **136**, 335–351.

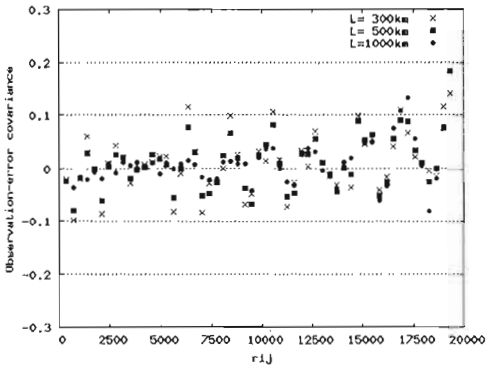


Figure 3.1: Off-diagonal terms in the observation error covariance as function of distance r_{ij} between points *i* and *j*.

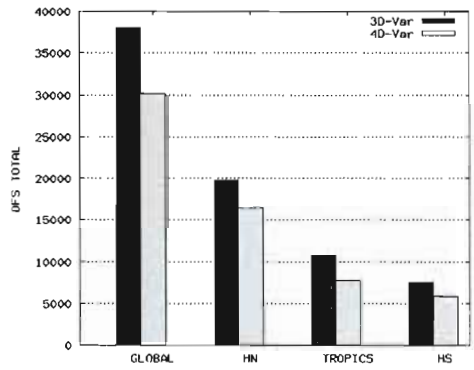


Figure 3.2: The total *DFS* two-month averaged (January-February 2007) in the MSC 3D-Var and 4D-Var analysis over the four regions: all globe, over the northern hemisphere (20N-90N), over the tropics (20S-20N) and over the southern hemisphere (90S-20S).

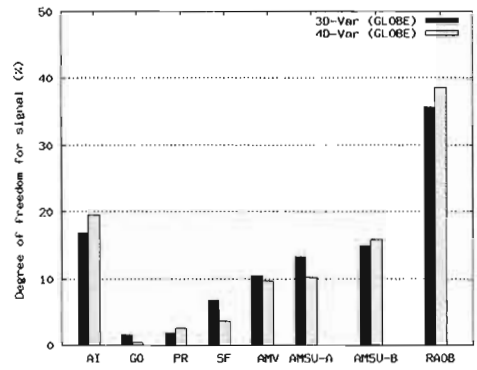


Figure 3.3: *DFS* two-month averaged (January-February 2007) in the MSC 3D/4D-Var analyses for the eight data types over globe.

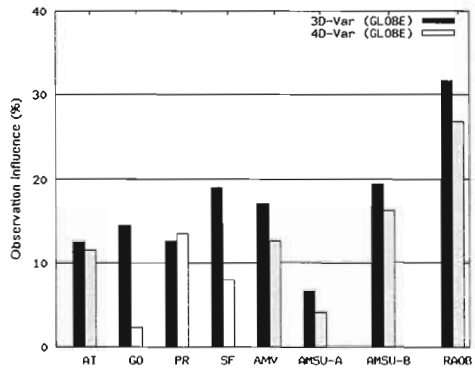


Figure 3.4: Observation influence in the MSC 3D/4D-Var analyses for the eight data types over globe.

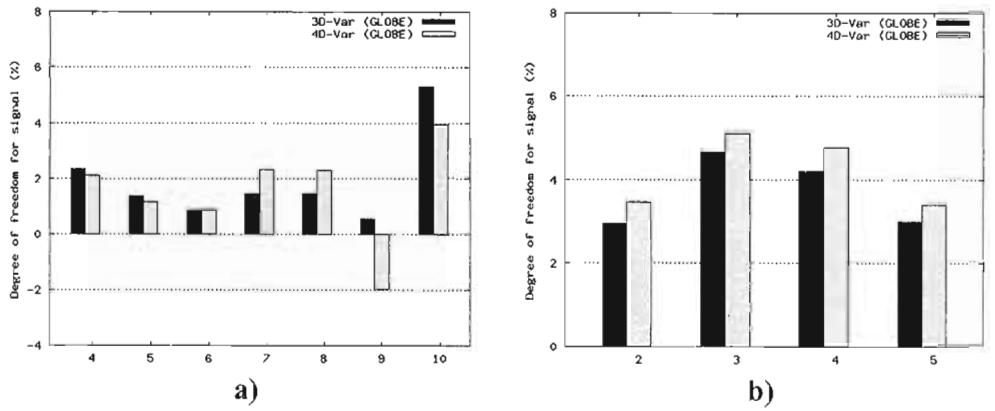


Figure 3.5: *DFS* two-month averaged over all globe in the MSC 3D/4D-Var analyses for each channel of (a) AMSU-A and (b) AMSU-B.

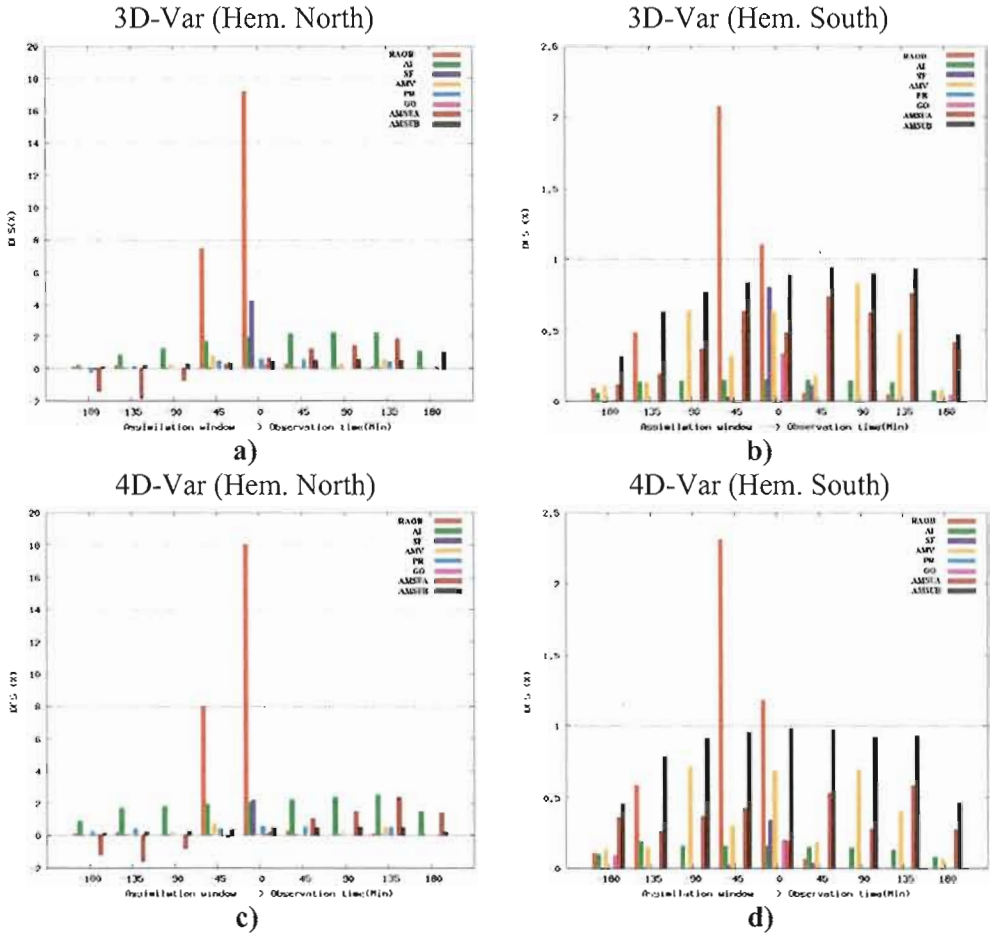


Figure 3.6: Degree of freedom for signal for the main data types in the 3D-Var and 4D-Var data assimilation systems as a function of observation time relative to the assimilation window. The observing platforms are color-coded and given in the legend.

L(Km)	$DFS_{ANALYTIC}$	DFS_{GIRARD}	$\tilde{DFS}^{(1)}_{APOST}$	$\tilde{DFS}^{(2)}_{APOST}$	$\tilde{DFS}_{DIAG}$
300	11.03	10.88	10.81	10.80	10.70
500	9.50	9.37	9.21	9.20	9.07
1000	7.34	7.08	6.79	6.79	6.75

Table 3.1: *DFS* estimate values as function of background correlation length-scale (L_c). $DFS_{ANALYTIC}$ as calculated from the prescribed statistics, DFS_{GIRARD} as computed with Girard's method, $DFS^{(1)}_{APOST}$, $DFS^{(2)}_{APOST}$ and $\tilde{DFS}_{DIAG}$ as obtained from eq. (21)-(23). The *a priori* values are perfectly known $\sigma_o^2 = \sigma_{o(t)}^2 = 4.$, $\sigma_b^2 = \sigma_{b(t)}^2 = 1.$

L(Km)	$DFS_{ANALYTIC}$	DFS_{GIRARD}	$\tilde{DFS}^{(1)}_{APOST}$	$\tilde{DFS}^{(2)}_{APOST}$	$\tilde{DFS}_{DIAG}$
300	16.44	16.21	16.11	16.21	15.18
500	13.39	13.19	12.99	13.01	11.25
1000	9.75	9.24	8.89	8.88	7.01

Table 3.2: Same as in Table 3.1 but for the experiment with $\sigma_b^2 = \sigma_{b(l)}^2 = 1$ and an underestimated value of observation error variance ($\sigma_o^2 = 2.25$).

L(Km)	$DFS_{ANALYTIC}$	DFS_{GIRARD}	$\tilde{DFS}^{(1)}_{APOST}$	$\tilde{DFS}^{(2)}_{APOST}$	$\tilde{DFS}_{DIAG}$
300	5.73	5.65	5.61	5.65	5.84
500	5.26	5.19	5.10	5.11	5.70
1000	4.43	4.36	4.17	4.17	5.35

Table 3.3: Same as in Table 3.1 but for the experiment with both the observation and background error variances underestimated ($\sigma_o^2 = 2.25$ and $\sigma_b^2 = 0.25$, respectively).

Observing Network	Variables			Thinning
radiosonde/dropsonde	U, V, T, (T-Td), Ps			28 levels
Surface report	T, (T-Td), Ps, (U, V over water)			1 report/6h
Aircraft	U, V, T			1° x 1° x 50 hPa
ATOVS		Ocean	Land	250 km x 250 km
	AMSU-A	Ch. 3-10	Ch. 6-10	
	AMSU-B	Ch. 2-5	Ch. 3-4	
Water vapor channel	Ch. 3 (6.7 μm)			2° x 2°
Geostationary AMV	U, V (IR, WV, VI channels)			1.5° x 1.5°
MODIS AMV	U, V			1.5° x 1.5°
Wind Profiler	U, V			750 m (vertical)

Table 3.4: List of observations assimilated in 3D and 4D-Var assimilation systems of the Environment Canada during winter 2006-2007.

Assimilation method	$\chi^2 = 2E[J_{min}]$	p	$\frac{\chi^2}{p}$
3D-Var	149899.34	265412	0.56
4D-Var	148744.75	265538	0.56

Table 3.5: Comparison between estimated values of chi square χ^2 and the number of observation p in 3D-Var and 4D-Var averaged for a two-month winter period (1 January to 28 February 2007).

CHAPITRE IV

ÉVALUATION D'IMPACT DES OBSERVATIONS DANS LES ANALYSES CONTROL ET OSE

Ce chapitre, rédigé en anglais, est présenté sous la forme d'un article qui sera soumis pour publication dans *Monthly Weather Review*.

Cette étude s'intéresse à l'évaluation du réseau canadien d'observations et il est appliqué aux Observing System Experiments (OSEs) effectués au MSC pour les mois de janvier et février 2007. Le *DFS* est utilisé comme outil diagnostique pour quantifier l'impact de différents types d'observations dans les analyses. Dans un premier temps, nous présentons brièvement la méthode d'estimation du *DFS* à partir des écarts d'observations à l'ébauche et à l'analyse ainsi que les expériences OSEs effectués. Ensuite, les résultats du *DFS* pour les expériences d'assimilation 3D and 4D-Var qui considèrent toutes les observations disponibles (control) ainsi que pour les OSEs sont comparées et discutées. En particulier, nous allons montrer que nos résultats concordent qualitativement aux ceux obtenu par Laroche and Sarrazin (2010 a, b). Des discussions et conclusions sont élaborées à la fin de cet article.

ASSESSMENT OF THE IMPACT OF OBSERVATIONS ON ANALYSES DERIVED FROM OBSERVING SYSTEM EXPERIMENTS

by

Cristina Lupu [*,1], Pierre Gauthier [1] and Stéphane Laroche [2]

[1] *Department of Earth and Atmospheric Sciences*
Université du Québec à Montréal (UQAM)
Montréal (Québec), CANADA

[2] *Meteorological Research Division*
Environment Canada
Dorval (Québec), CANADA

Submitted to *Monthly Weather Review*

[*] Corresponding author: *Department of Earth and Atmospheric Sciences*
Université du Québec à Montréal (UQAM)
P.O. Box 8888, Succ. Centre-ville
Montréal, Québec
CANADA H3C 3P8

ABSTRACT

Observing System Experiments (OSEs) can help quantify the impact of different observation types on specific numerical weather prediction system. Methods based on Degree of Freedom for Signal (*DFS*) have been implemented to diagnose the impact of observations on analyses itself. This paper describes the use of *DFS* as a diagnostic tool to estimate the amount of information brought by subsets of observations in the context of OSEs. In particular, this study is interested in the evaluation of the Canadian observing network and is applied to OSEs performed at Meteorological Service of Canada for January and February 2007. The relative values of the main observing networks over North America derived from *DFS* calculations are compared with those from the OSEs. The results show that removing some observation types from the assimilation system influence the effective weight of the remaining assimilated observations. The response of the remaining observations when a given set of observations is denied is illustrated comparing *DFS* calculations with the observations impact estimated from observing system experiments.

Keywords: Degrees of freedom for signal, Observing System Experiments.

4.1 INTRODUCTION

Of particular importance in developing data assimilation systems is to quantify the actual impact of different observation networks within the assimilation process. The value of observations in data assimilation systems has been obtained by evaluating the information content of observations or Degrees of Freedom for signal (*DFS*) (Rodgers, 2000; Rabier *et al.*, 2002; Cardinali *et al.*, 2004). Others methods that diagnose the impact of assimilated observations on a given analysis or forecast include analysis sensitivity (Rodgers, 2000; Cardinali *et al.*, 2004) and adjoint-based procedures (Baker and Daley, 2000; Langland and Baker, 2004; Zhu and Gelaro, 2008; Cardinali, 2009). Recent diagnostics work by Desroziers *et al.* (2005) showed how simple consistency diagnostics can be obtained for covariances of observation, background and analysis errors in observation space. Lupu *et al.* (2010) showed that *DFS* calculations could be obtained from the diagnosed covariance matrices estimated as in Desroziers *et al.* (2005).

An Observing System Experiment (OSE) is a traditional approach to estimate the impact of a specific observing network on a numerical weather prediction system. An OSE is composed of two experiments, both covering the same period. In the first experiment, (control), all the observations operationally available are used. In the second experiment selected data sets are systematically removed from the assimilation procedure in order to assess the degradation in quality of a model forecast when that observation type is denied. Gelaro and Zhu (2009) and Cardinali (2009) have compared adjoint-based impact calculations against results from OSEs. Despite some basic differences between the adjoint-based and OSE techniques, the general conclusions of these studies were that the two approaches provide unique, and complemen-tary, information.

The main objectives of this paper are to examine the variability of the *DFS* when subsets of observations are denied as in OSEs, as well as to compare the impact estimates from both techniques (i.e. OSEs and *DFS*). The results from the OSEs carried out by Laroche and Sarrazin (2010 a, b) are used. In these OSEs, the forecast impacts of radiosonde and aircraft data over North America, as well as the impact of satellite data over the North Pacific Ocean

in both 3D-Var and 4D-Var contexts (Gauthier *et al.*, 1999, 2007) were examined. The *DFS* calculated from a posteriori statistics (Lupu *et al.*, 2010) is used here to assess the impacts of the observing systems on analyses of the various OSEs. Although OSEs are used to estimate data impact in a forecasting system, whereas the *DFS* calculations are used to assess the impacts of various observing systems on analyses, we investigate in this work whether *DFS* calculations show some agreement with results obtained from OSEs.

Section 2 outlines the methodology to estimate the *DFS* from observation departures of the analysis and forecast. Section 3 consists of a brief summary of the OSEs used in this study. In section 4, the information content of observations is evaluated for both MSC's 3D and 4D-Var control experiments and for a number of OSEs to estimate how the results vary with the observation coverage. Section 5 presents quantitative comparison of *DFS* in OSEs experiments. Section 6 briefly compares our results obtained using *DFS* diagnostics with those obtained in data impact studies by Laroche and Sarrazin (2010 a, b). Section 7 gives a summary and conclusions drawn from this study.

4.2 COMPUTATION OF DFS FROM A POSTERIORI STATISTICS

The *DFS* can be used to evaluate the impact of observations on the analysis (Rodgers, 2000; Rabier *et al.* 2002; Cardinali *et al.*, 2004; Chapnik *et al.*, 2006). It is defined as the trace of the partial derivative of the analysis in observation space to the observations:

$$DFS = tr\left\{\frac{\partial(\mathbf{H}\mathbf{x}_a)}{\partial \mathbf{y}}\right\} \tag{1}$$

where $tr(\,)$ denotes trace of $(\,)$, $\mathbf{x}_a$ represents the analysis, $\mathbf{y}$ is a vector of observational data and $\mathbf{H}$ is the tangent linear operator of H. For an optimal case, the analysis can be written as:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H\mathbf{x}_b) \tag{2}$$

where $\mathbf{x}_b$ is the background state, $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$ is the Kalman gain matrix, $\mathbf{B}$ is the background-error covariance matrix and $\mathbf{R}$ is the observation-error covariance matrix. In a linear framework (1) and (2) imply that

$$DFS = tr\left\{\frac{\partial(\mathbf{H}\mathbf{x}_a)}{\partial \mathbf{y}}\right\} = tr(\mathbf{K}^T\mathbf{H}^T) = tr(\mathbf{H}\mathbf{K}) \tag{3}$$

This diagnostic quantifies the gain in information brought by the observations on analyses and may be also applied for a particular subset of observations as long as they are not correlated with the rest of the observations.

In this study the calculation of *DFS* can be performed using the diagnosed covariance matrices estimated as in Desroziers *et al.* (2005). Their diagnostics are based on combinations of differences between observation and analysis, observation and background, and differences between the background and analysis, and can be summarized as follows:

$$E\left[\mathbf{d}_a^o(\mathbf{d}_b^o)^T\right] = \tilde{\mathbf{R}} = \mathbf{R}\mathbf{D}^{-1}\tilde{\mathbf{D}} \tag{4-a}$$

$$E\left[\mathbf{d}_b^a(\mathbf{d}_b^o)^T\right] = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{\mathrm{T}} = \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-1}\tilde{\mathbf{D}} \tag{4-b}$$

$$E\left[\mathbf{d}_b^a(\mathbf{d}_a^o)^T\right] = \mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^{\mathrm{T}} = \mathbf{H}\mathbf{K}\tilde{\mathbf{D}}\mathbf{D}^{-1}\mathbf{R} \tag{4-c}$$

$$E\left[\mathbf{d}_b^o(\mathbf{d}_b^o)^T\right] = \tilde{\mathbf{D}} = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{\mathrm{T}} + \tilde{\mathbf{R}} \tag{4-d}$$

Here, the innovation vector $\mathbf{d}_b^o$ is the departure between observations $\mathbf{y}$ and their background counterparts $H(\mathbf{x}_b)$, $\mathbf{d}_b^a$ is the difference between analysis and background in observation space, and $\mathbf{d}_a^o$ is the difference between observation and analysis in observation space. Moreover, $\mathbf{D} = \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}$ is the *a priori* innovation covariance, $\tilde{\mathbf{D}} = \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T + \tilde{\mathbf{R}}$ is the sample covariance of innovations and $E[\ ]$ is the statistical expectation operator. $\tilde{\mathbf{R}}$ is the diagnosed observation-error matrix and $\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T$ and $\mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^T$ are the diagnosed background- and analysis-error covariance in observation space respectively.

An estimate of *DFS* can be computed either from the *a posteriori* statistics or from the *a priori* statistics. Defining the *a posteriori* Kalman gain matrix $\tilde{\mathbf{K}} = \tilde{\mathbf{B}}\mathbf{H}^T(\tilde{\mathbf{R}} + \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T)^{-1}$ and using (4-c), the estimate of $tr(\tilde{\mathbf{K}}^T\mathbf{H}^T)$ from the *a posteriori* statistics is:

$$\begin{aligned} D\tilde{F}S = tr(\tilde{\mathbf{K}}^T\mathbf{H}^T) &= tr\left[\tilde{\mathbf{R}}^{-1}(\tilde{\mathbf{A}}\mathbf{H}^T)^T\right] \\ &= tr\left[\tilde{\mathbf{R}}^{-1}(\mathbf{H}\mathbf{K}\tilde{\mathbf{D}}\mathbf{D}^{-1}\mathbf{R})^T\right] \end{aligned}$$

$$= tr(\mathbf{K}^T \mathbf{H}^T) = DFS \tag{5}$$

In operational systems, a major issue with the estimation of *DFS* using (5) is that the matrices involved are too large to be stored explicitly. Substituting (4-c) into (5) and using the properties that the trace and expectation commute and also $\mathbf{X}E[(.)] = E[\mathbf{X}(.)]$ for any nonrandom matrix $\mathbf{X}$, the *a posteriori* $\tilde{\text{DFS}}$ can be further rewritten into:

$$\begin{aligned} \tilde{DFS} = tr(\tilde{\mathbf{K}}^T \mathbf{H}^T) &= tr\left\{ \tilde{\mathbf{R}}^{-1} E\left[\mathbf{d}_b^a \mathbf{d}_a^{oT} \right]^T \right\} \\ &= tr\left\{ \tilde{\mathbf{R}}^{-1} E\left[\mathbf{d}_a^o \mathbf{d}_b^{aT} \right] \right\} \\ &= tr\left\{ E\left[\tilde{\mathbf{R}}^{-1} \mathbf{d}_a^o \mathbf{d}_b^{aT} \right] \right\} \\ &= E\left\{ tr[\tilde{\mathbf{R}}^{-1} \mathbf{d}_a^o \mathbf{d}_b^{aT}] \right\} \\ &= E\left[\mathbf{d}_b^{aT} \tilde{\mathbf{R}}^{-1} \mathbf{d}_a^o \right] \end{aligned} \tag{6}$$

Relation (6) gives a simple and efficient way to estimate the *DFS* for any assimilation scheme because only by-products of the data assimilation scheme are used. A unique aspect of this formulation is that does not require the consistency of the error statistics in the analysis system. When the sample covariance matches the prescribed innovation covariance ($\tilde{\mathbf{D}} = \mathbf{D}$), (6) reduces to:

$$\tilde{DFS} = tr(\mathbf{HK}) = E\left[\mathbf{d}_b^{aT} \mathbf{R}^{-1} \mathbf{d}_a^o \right] \tag{7}$$

It must be stressed that the equality between the *DFS* based on the trace of the full matrix product $\mathbf{HK}$ and the *DFS* based on the *a posteriori* quantities (eq. 6) holds when the complete diagnosed $\tilde{\mathbf{R}}$ matrix is used. We may take into account this off-diagonal covariance by forming an eigenvalue decomposition of $\tilde{\mathbf{R}}$, so the *DFS* in (6) can be evaluated at the cost of a few dot products. Note that even when the specified $\mathbf{R}$ matrix is diagonal, the estimated observation error covariance matrix $\tilde{\mathbf{R}}$ can show non-zero cross-correlations due to misspecification of other aspect of $\mathbf{R}$ and $\mathbf{B}$, as indicated by (4-a). Previous study by Lupu *et al.* (2010) pointed out that the off-diagonal covariances are relatively small. Another way

is to describe $\tilde{\mathbf{R}}$ by a diagonal approximation $\tilde{\mathbf{R}} \cong \tilde{\sigma}_o^2 \mathbf{I}$ where $\tilde{\sigma}_o^2$ the diagnosed observation-error variance, calculated for each subset of observations operationally assimilated at MSC.

Approximating $\tilde{\mathbf{R}}$ by a diagonal matrix we can deduce from (6)

$$D\tilde{F}S = E\left[\mathbf{d}_b^{aT}\tilde{\mathbf{R}}^{-1}\mathbf{d}_a^o\right] \cong E\left[\frac{\mathbf{d}_b^{aT}\mathbf{d}_a^o}{\tilde{\sigma}_o^2}\right] \tag{8}$$

In this study, the assessment of observation value with respect to analyses through OSEs is performed comparing the information content or *DFS* calculated as in (8), obtained with and without the subset of data of interest.

4.3 SUMMARY OF THE OSES CARRIED OUT AT MSC

A series of OSEs that used the standard data denial method was performed using the MSC's 3D-Var and 4D-Var data assimilation systems (Laroche and Sarrazin, 2010 a, b). The experiments covered the two-month period of January and February 2007. The observation types operationally assimilated at MSC in winter 2006-2007 are the radiosondes data (RAOB), aircraft reports (AI), surface and ship data (SF), wind profiler data (PR), atmospheric motion vectors from geostationary satellites (AMV) and radiances from polar-orbiting satellite (AMSU-A and AMSU-B) and from geostationary satellites (GOES-East and GOES-West). Each series of OSEs systematically removed different observation types from the operational system: radiosonde and wind profiler over North America (NO_RAOB), aircraft reports over the North America (NO_AIRCRAFT) and satellite data over the North Pacific Ocean (NO_SAT). Two additional experiments were conducted using 4D-Var data assimilation system: (NO_ASCENT/DESCENT) which excludes aircraft data between the ground and 350 hPa, and the combined (NO_RAOB+NO_AIRCRAFT) which excludes radiosonde, wind profiler and aircraft data over North America. These OSEs are used to test the relevance of the different existing components of the observing system over North America. The NO_AIRCRAFT and NO_ASCENT/DESCENT experiments allow to assess the relative value of aircraft measurement profiles located over major airports in North America and the last experiment NO_RAOB+NO_AIRCRAFT will thus enable us to assess

the joint impact of these observing networks over North America. Figure 4.1 shows the areas where the observations are denied over the North America. The Canadian Arctic, Canada and continental United States regions are chosen to examine the impact of observation on 3D/4D-Var analyses through *DFS*.

4.4 OBSERVATION IMPACT ESTIMATED FROM DFS IN OSES

The aim of this section is to assess the impacts of various observing systems on analyses during two-month winter period in terms of information content or *DFS*. In the following, we compare the *DFS* results for different data types obtained from control experiments, which include all observations, with those from OSEs. In fact, the removal of any observing systems from the assimilation system will produce a distinct experiment that differs from the others in terms of number of observations that are assimilated. Consequently, OSEs can change the analysis constraints on the remaining data and can alter the outcome of the assimilation. In this context it is important to understand how the absence of an observing system affects the information content supplied by different types of observations to an analysis.

We first discuss the impact of removing RAOB and PR data (NO_RAOB) or AI reports (NO_AIRCRAFT) over North America on analyses over four regions covering the North America. The averaged *DFS* over the two-month period for the various subsets of observations is presented in Figure 4.2 for each OSE experiments over North America. The observing system removed in a given OSE is plotted with zero value. Results indicate that AI data are as informative as RAOB data in the 3D-Var CTRL experiment, while other observations have much less impact. However, the information content is greater for the AI data in 4D-Var CTRL experiment, which indicate that 4D-Var better exploit the aircraft data that are distributed all over the assimilation time window. When focusing on AMSU-A radiances, primarily sensitive to the atmospheric temperature profile, we note a negative value of *DFS* estimate over the North America in the control experiment. The examination of *DFS* for different AMSU-A channels indicates that this comes from channels 9 and 10, which peak in the stratosphere just below the model lid (i.e. 10 hPa). At these levels, model errors in the short-range forecast are important and may contribute to degrade the analysis. Note that observation-error variance of each channel was estimated *a posteriori* independently from

diagnostics of Desroziers *et al.* (2005). We found that the *a priori* observation-errors variances for channels of AMSU-A are generally overestimated in both MSC 3D and 4D-Var assimilation systems.

Results in Fig. 4.2 indicate that the removal of RAOB and PR observations over North America affects the relative *DFS* of several observing systems. The relative change in the *DFS* of an individual data type k inside a particular region is defined here as the normalized difference between the DFS_k^{Region} of OSE experiment and DFS_k^{Region} of the control experiment:

$$\frac{\Delta(DFS_k^{Region})}{DFS_{all\_obs}^{Region}(ctrl)}(\%) = 100 \cdot \frac{DFS_k^{Region}(OSE) - DFS_k^{Region}(ctrl)}{DFS_{all\_obs}^{Region}(ctrl)} \tag{9}$$

where $DFS_{all\_obs}^{Region}$ refer to the total *DFS* of the control experiment.

For the AI data over North America, the relative *DFS* increases by 6.5% with respect to the control when RAOB and PR data are removed in 3D-Var. Similarly, we note an increase in the relative *DFS* of radiosonde and profilers data of 5% and 1.7% respectively when AI data are removed in 3D-Var. When RAOB and PR data are removed over North America, the relative *DFS* for AI and AMSU-A data in 4D-Var experiments increase by 3.2% and 12.9% respectively with respect to the control experiment. Removing AI data over North America leads to a larger increase of relative *DFS* of RAOB and PR data in 4D-Var (5.8% and 5.4% respectively). The contributions from AMSU-A data, which have small negative *DFS* in the control experiment, become positive when RAOB and PR data are removed, but remain small overall in both 3D-Var and 4D-Var.

To further explore and understand the impact of the removal of the data over the North America, we examine the impact of observations on analyses in different regions. As pointed out by Laroche and Sarrazin (2010 a), for a given region, the impact on analyses depends on the accuracy of the data provided by the observational network and the ability of the data assimilation scheme to extract the information from these observations. Figure 4.3 presents *DFS* results over the Canadian Arctic (Figs. 4.3-a and 4.3-b), Canada (Figs. 4.3-c and 4.3-d) and continental United States (Figs. 4.3- e and 4.3-f). *DFS* values per observation type for 3D-Var control experiment show that RAOB is the most informative data over the Canadian

Arctic. Other satellite observations (i.e. AMSU-A and AMSU-B) have less impact on the analyses with values of 51.1 and 5.0 respectively, compared to 189.9 for RAOB data. Without any RAOB and PR data over the North America, the *DFS* associated to AMSU-A and AMV data increases by 29.7% and 7.1%, respectively (Fig. 4.3-a). In the 4D-Var control experiment (Fig. 4.3-b) the *DFS* for AMSU-A and AMV data increases by 50.1% and 13.1% respectively, without RAOB and PR data. 4D-Var seems superior to 3D-Var to exploit the satellite data over this region. The results for the NO_AIRCRAFT experiments using both 3D-Var and 4D-Var schemes are closer to the control experiment for all observations types.

Over Canada, the RAOB is the main contributor to the *DFS* in both 3D-Var and 4D-Var CTRL experiments. In the NO_RAOB experiment with 3D-Var, the *DFS* for AI and AMSU-A data increases by 5.1% and 25.8% respectively, as compared to the CTRL experiment (Fig. 4.3-c). The difference between NO_RAOB and CTRL experiments with 4D-Var is even more noticeable for AMSU-A, for which the *DFS* increases by 42.1% (Fig. 4.3-d). The results for the NO_AIRCRAFT experiments show that the *DFS* for RAOB data increases by 4.1% in 3D-Var and 5.8% in 4D-Var.

Over the continental United States, the *DFS* for AI with are larger than the *DFS* for RAOB data in both 3D-Var and 4D-Var CTRL experiments. The comparison of NO_RAOB and CTRL experiments over continental United States (Figs. 4.3-e and 4.3-f) reveals that the *DFS* for AI data increases by 7.7% in 3D-Var and by 3.3% in the 4D-Var. Finally, it can be seen that the removal of AI data affects the *DFS* of RAOB and PR data. The removal of AI data increases the *DFS* of RAOB by 5.9% and the *DFS* of PR data by 2.4% in 3D-Var. Those values are larger when the 4D-Var is used (6.7% and 7.4% respectively).

Figure 4.4 shows results obtained using 4D-Var for the same denial experiments and observations subsets as Fig. 4.3, and for two additional denial experiments (NO_ASCENT/DESCENT, NO_RAOB_NO_AIRCRAFT). The results are presented over the Canadian Arctic (Fig. 4.4-a), Canada (Fig. 4.4-b) and continental United States (Fig. 4.4-c). Not surprisingly, without aircraft report below 350 hPa over North America, the *DFS* associated with AI data decreases. Results show that over Canada and continental United States, the AI ascent/descent reports alone account for roughly 50% of the impact of all the AI data. In addition, the increase of *DFS* for the other data types assimilated is much weaker

than when all the AI data are denied. Without RAOB, PR and AI data, the *DFS* associated with AMSU-A and AMV data over the Canadian Arctic increase, by respective values of 54.1% and 13.1%.

The impact of subsets of the global observing system on the analyses 3D/4D-Var over North America has been evaluated. The largest *DFS* over this region are clearly for radiosonde and aircraft data. Removal radiosonde and aircraft data over North America affects the relative *DFS* of several observing systems. Over Canada and Canadian Arctic, the radiosonde is the main contributor to the *DFS* in both 3D-Var and 4D-Var while over the continental United States the *DFS* of aircraft data is dominant.

The results presented here show that removing some observation types from the assimilation system influences the effective weight of the remaining assimilated observations. We have found that 4D-Var seems superior to 3D-Var to exploit the satellite data in absence of RAOB data. The changes observed in *DFS* calculations for different data types over different regions reveal that some of the remaining observations may compensate by having more impact on the analyses. In the next section, the aim is to quantify the reduction in the total *DFS* resulting from the removal of different subsets of observations in OSEs experiments and to estimate the compensation supplied by the assimilated observations in analyses.

4.5 INTERDEPENDENCY OF OBSERVING SYSTEMS

The Degrees of freedom for signal (*DFS*) is used for estimating the value of observations in a data assimilation system. In this section, we show that the *DFS* can also be useful for assessing the complementarity and redundancy of observing networks. This can be achieved by examining the percentages of *DFS* for different observing systems k estimated for a given region with respect to all other observations assimilated in the control experiment:

$$DFS_k^{Region}(\%) = 100 \cdot \frac{DFS_k^{Region}}{DFS_{all\_obs}^{Region}} \tag{10}$$

and the fractional impact due to the removal of the observing system k with respect to the control experiment:

$$F_{no\_k}^{Region}(\%) = 100 \cdot \frac{DFS_{no\_k}^{Region}(OSE) - DFS_{all\_obs}^{Region}(ctrl)}{DFS_{all\_obs}^{Region}(ctrl)} \qquad (11)$$

where $DFS_{no\_k}^{Region}(OSE)$ and $DFS_{all\_obs}^{Region}(ctrl)$ are the total *DFS* estimated respectively for OSE without observing system k and for control run in the various regions. Note that the exclusion of part of the observations from the data assimilation system generally leads to a decrease in total *DFS,* so that the numerator of (11) is generally a negative number. Relation (11) provides a measure of the change (typically a reduction) in total *DFS* resulting from the removal of observing system k from the system. It is interesting to use relations (10) and (11) to give quantitative comparisons between *DFS* in the various OSEs. Figure 4.5 shows the averaged values of $F'^{Region}_{no\_k} = \left|F^{Region}_{no\_k}\right|$ and DFS_k^{Region} during January-February 2007 for two observation sets denied over the North America: radiosonde and wind profiler (RAOB+PR) and aircraft data (AI).

Figures 4.5a-b show averaged values of DFS_k^{Region} and $F'^{Region}_{no\_k}$ during January and February 2007 for the RAOB and PR data over four regions (North America, Canadian Arctic, Canada and continental US) obtained with 3D-Var (Fig. 4.5a) and 4D-Var (Fig. 4.5b), respectively. Over all regions, the values of DFS_k^{Region} are larger than those of $F'^{Region}_{no\_k}$. Note that the difference between these two values is related to the fact that the remaining data types compensate for the loss of RAOB and PR data. This compensation is more important over Canada and Canadian Arctic regions, where RAOB data are the most informative data source. However it is also worth noting that with 4D-Var, the remaining data types compensate better for the removal of RAOB data. Over the continental United States and North America the *DFS* of RAOB data is smaller mainly because in these regions the AI data are at least as informative as RAOB data. This explains why in these regions the compensation by other data types is less significant.

Figures 4.5c-d show average values of DFS_k^{Region} and $F'^{Region}_{no\_k}$ during January and February 2007 for the AI data over the same regions as Figs. 4.5a-b obtained with 3D-Var (Fig. 4.5c) and 4D-Var (Fig. 4.5d), respectively. Results show that the *DFS* for AI data is dominant over the continental United States mainly due to the larger number of AI data over

this region. In contrast, over the Canadian Arctic, where the analysis essentially relies on radiosonde network, the relative *DFS* of AI data is small. With 4D-Var, these values are more important, because the 4D-Var scheme better handles asynoptic observations such as AI data.

The results presented thus far, based on relative *DFS* calculations, showed the impact of RAOB, PR and AI data in the data assimilation system itself. The fractional impact can be used to illustrate the responses of the remaining observations when a given set of observations is denied.

4.6 COMPARISON OF OBSERVATION IMPACTS ESTIMATED FROM OSES AND DFS CALCULATIONS

In this section, we examine some results from the OSEs presented in Laroche and Sarrazin (2010 a, b) for short-range forecasts. In particular, we assess how the forecast impacts from the OSEs agree well with the observation impacts deduced from the *DFS* diagnostics presented in the previous sections. In the OSE context, the forecast impact (FI) of individual data type over a region of interest can be defined as:

$$FI(\%) = 100 \cdot \frac{RMS_{no\_k} - RMS_{ctrl}}{RMS_{ctrl}} \tag{12}$$

where RMS_{no_k} is the root mean-squared forecast error for a given OSE, RMS_{ctrl} is the one of the control model run. A positive FI score indicates that the forecast quality is improved when the denied data set is assimilated.

First, we examine forecast impacts when RAOB, PR and AI data are omitted over North America in both 3D and 4D-Var systems. Figure 4.6 shows the FI for the 500-hPa geopotential heights for 12-h forecast for the NO_RAOB experiment (Fig. 4.6a) and NO_AIRCRAFT experiment (Fig. 4.6b) over four geographical areas. Note that the vertical scale in 4.6b is approximately 5 times small than in 4.6a. The short term forecast impact of AI data over the continental United States is larger than from RAOB data, while it is the opposite over the Canadian Arctic. In addition, the impact in the 4D-Var experiments is smaller by about 5% with respect to the 3D-Var experiments.

Not surprisingly, the results for the NO_RAOB experiment in Fig. 4.6a demonstrate that the greatest forecast impacts are in the Canadian Arctic and Canada regions. Consistent with all other results presented herein, based on *DFS* calculations, the removal of AI provides very little forecast impact over the Canadian Arctic, as shown in Fig. 4.6b. The *DFS* percentages for RAOB and PR data over the various regions of North America (Fig. 4.5a,b) are more homogenous than the corresponding FIs (Fig. 4.6a). However, the variation of the *DFS* percentages and FIs from one region to another agree better for the AI data (Fig. 4.5c,d and Fig. 4.6b). This indicates that the results from the *DFS* calculations are not always consistent with those from the OSEs. The main difference between the two methods is that the *DFS* measures the influence of the data in the analysis while the OSEs assess the forecast skill provided by the data. Since the forecast skill depends primarily on atmospheric structures that grow most rapidly in time, data sets that capture best these structures in the analysis will provide the most benefit to forecasts. This cannot be measured by the *DFS*. Methodologies that use adjoint models to estimate the observation sensitivities to short-range forecast skill (e.g. Langland and Baker, 2004; Cardinali, 2009) are more suitable to assess the forecast impact of observations. However, as pointed out by Zhu and Gelaro (2008), OSEs and adjoint-based procedures provide unique, but complementary, information about the impact of observations on forecasts. This is also true for the *DFS* and OSE methodologies.

In the following, we examine the effect of the weather regime on the *DFS* diagnostics in a similar manner to that in Laroche and Sarrazin (2010b). They showed that the flow regime in January 2007 was significantly different of the one that prevailed in February 2007, which has a noticeable effect on the forecast impacts over the North American continent. Table 1 show the averages values of $DFS_{RAOB+PR}^{Region}$ and DFS_{AI}^{Region} respectively estimated with the 3D-Var and 4D-Var scheme for both months individually as well as for the two-month period over the Canadian Arctic, Canada and continental United States. The results for January and February estimated separately indicate that the $DFS_{RAOB+PR}^{Region}$ of RAOB and PR data is larger in January than in February over the Canada and Canadian Arctic, whereas it does not change over the continental US. The *DFS* seems also sensitive to the weather regimes that prevail during the period under investigation. For example, over Canada, the difference in the *DFS* can be as large as 5% (8%) with 3D-Var (4D-Var) due to the change in the weather regime.

Similarly, the DFS_{AI}^{Region} of AI data is larger in January than in February over the Canada and continental US. Overall, the difference in the *DFS* of aircraft data is more important in the 4D-Var context.

4.7 CONCLUSIONS

This study focused on the impacts of various observing systems on analyses (CTRL and OSEs) during two-month winter period in terms of information content or *DFS* calculated from *a posteriori* statistics. A particular aspect of this work is that the results from OSEs presented by Laroche and Sarrazin (2010 a, b) have been used to investigate how the impact of observations on analyses may vary depending on the observation environment.

The *DFS* is an attractive method for estimating the value of observations in a data assimilation system. The procedure is computationally inexpensive and can be readily applied because it only combines quantities available after the analysis. A single data assimilation experiment that includes all operational observations assimilated is necessary and sufficient to estimate the *DFS* of any subset of observations with respect to all other observations assimilated simultaneously. Also, the method can be readily integrated into an operational analysis system. The impact of subsets of the global observing system on the analyses 3D/4D-Var over North America has been evaluated. Results indicate that the quality of analyses over this region relies on the radiosonde and aircraft data. Removal radiosonde and aircraft data over North America affects the relative *DFS* of several observing systems over North America. The impacts of observations on analyses in different regions have been explored in detail. Over Canada and Canadian Arctic, the radiosondes are the main contributor to the *DFS* in both 3D-Var and 4D-Var while over the continental United States the *DFS* of aircraft data is dominant. We noted over Canada and continental United States a negative *DFS* partly due to the channel sounding in the high troposphere. It is also found that the *DFS* seems sensitive to the weather regimes that prevail during the period under investigation.

In addition, a detailed comparison between *DFS* and fractional impact was performed to assess the interdependency of observing networks. Results show that over all regions of

North America the values of *DFS* are larger than those obtained for the fractional impact. The difference between these values is atributed to the fact that the remaining data types compensate for the loss of denial data. Consequently, for the RAOB and PR data this compensation is more important over the Canada and Canadian Arctic regions, where these data are the most informative data source. Likewise, for the AI data the compensation is more important over the continental United States and North America.

Some interesting conclusions have been drawn based on the forecast impact (FI) of individual data type over a particular region. In particular, it was demonstrated that on the short range forecast *DFS* and OSEs provide a similar qualitative picture of improvement due to radiosonde or aircraft observations. However, the variation of the *DFS* percentages and FIs from one region to another agree better for the aircraft data. The main difference between the two tools is that the *DFS* measures the influence of the data in the analysis while the OSEs assess the forecast skill provided by the data. Furthermore, OSEs and adjoint-derived measures of observation impact appear to are more suitable to assess the forecast impact of observations as described in Gelaro and Zhu (2009) and Cardinali (2009). Our study suggests that *DFS* show some agreement with results obtained from OSEs despite differences in the way observation impacts are measured in the two approaches,

ACKNOWLEDGEMENTS

The authors would like to thank Mr. Pierre Koclas of the Meteorological Service of Canada who helped us in the utilization of the observation data base to build the diagnostics in observation space used in this study. Environment Canada provided the computing facilities, and technical assistance for the use of their assimilation system.

This work has been funded mostly by Grant 500-B of the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) for the project on the *Impact of Observing Systems on Forecasting Extreme Weather in the short, medium and extended range: A Canadian contribution to THORPEX*, with additional support from Discovery Grant No.357091 of the Natural Sciences and Engineering Research Council (NSERC) of Canada.

REFERENCES

Baker N. L. and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptative observation-targeting problem. *Q. J. R. Meteorol. Soc.*, **126**, 1431-1454.

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

——, 2009: Monitoring the observation impact on the short-range forecast. *Q. J. R. Meteorol. Soc.*, **135**, 239-250.

Chapnik B., G. Desroziers, F. Rabier and O. Talagrand, 2006: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Q. J. R. Meteorol. Soc.*, **132**, 543–565.

Desroziers G., L. Berre, B. Chapnik and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.*, **131**, 3385–3396.

Gauthier, P., C. Charette, L. Fillion, P. Koclas and S. Laroche, 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmos.-Ocean,* **37,** 103–156.

——, M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau, 2007: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Wea. Rev.*, **135,** 2339–2354.

Gelaro, R. and Y. Zhu, 2009: Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179-193.

Kelly, G., J. N. Thépaut, R. Buizza and C. Cardinali, 2007: The value of observations. I: Data denial experiments for the Atlantic and the Pacific. *Q. J. R. Meteorol. Soc.*, **133**, 1803-1815.

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

Laroche, S. and R. Sarrazin, 2010a: Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part 1: contribution of radiosonde, aircraft and satellite data. *Atmos.-Ocean*, **48**, 10-25.

——, and R. Sarrazin, 2010b: Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part II: Sensitivity experiments. *Atmos.-Ocean*, **48**, 26-38.

Lupu C., P. Gauthier and S. Laroche, 2010: Evaluation of the impact of observations on analyses in 3Dand 4D-Var based on information content. To be submitted to *Mon. Weather Rev.*

Rabier, F., N. Fourrié, D. Chafai and P. Prunet, 2002: Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Q. J. R. Meteor. Soc.*, 128, 1011-1027.

Rodgers, C., 2000: Inverse Methods for Atmospheric Sounding Theory and Practice. World Scientific Publishing, London, 256 pp.

Zhu Y. and R. Gelaro 2008: Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI) analysis system. *Mon. Weather Rev.* **136**, 335–351.

Region	3D-Var			4D-Var		
	$DFS_{RAOB+PR}$ (%)			$DFS_{RAOB+PR}$ (%)		
	January	February	Jan.-Feb.	January	February	Jan.-Feb.
Can. Arctic	62.96	60.55	61.77	88.08	60.44	72.09
Canada	66.54	63.05	64.85	77.64	71.80	74.87
US Cont.	49.58	49.7	49.34	48.75	49.85	49.27
	DFS_{AI} (%)			DFS_{AI} (%)		
Can. Arctic	2.75	3.23	2.98	3.98	3.08	3.46
Canada	36.17	28.34	32.39	42.15	31.83	37.27
US Cont.	50.62	48.46	49.57	54.34	52.45	53.45

Table 4.1: Averages values of *DFS* with 3D-Var and 4D-Var for January, February and during January-February 2007 for three regions (Canadian Arctic, Canada and continental United States).

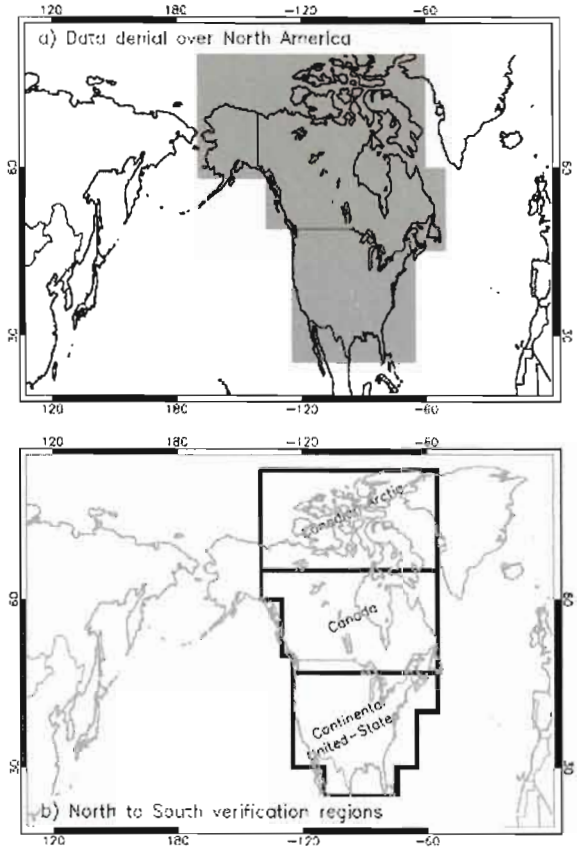


Figure 4.1: Areas (in grey) where (a) profiling observations (radiosonde, aircraft and wind profiler data are denied over North America. (b) The Canadian Arctic, Canada and continental United States regions chosen to examine the impact of observation on analyses.

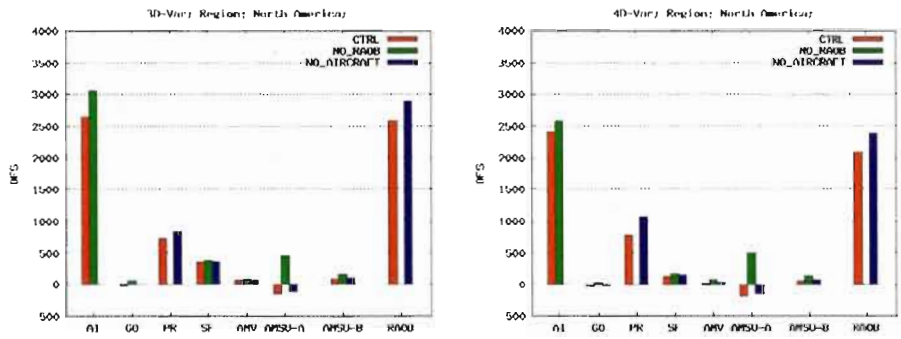


Figure 4.2: North America data denial experiments. Averages values of *DFS* for eight families of observational data (see text for description) in the control experiment (red bars), in the NO_RAOB (green bars) and NO_AIRCRAFT (blue bars) experiments inside the North America region. Results with 3D-Var are in the left panel while those with 4D-Var are in the right panel.

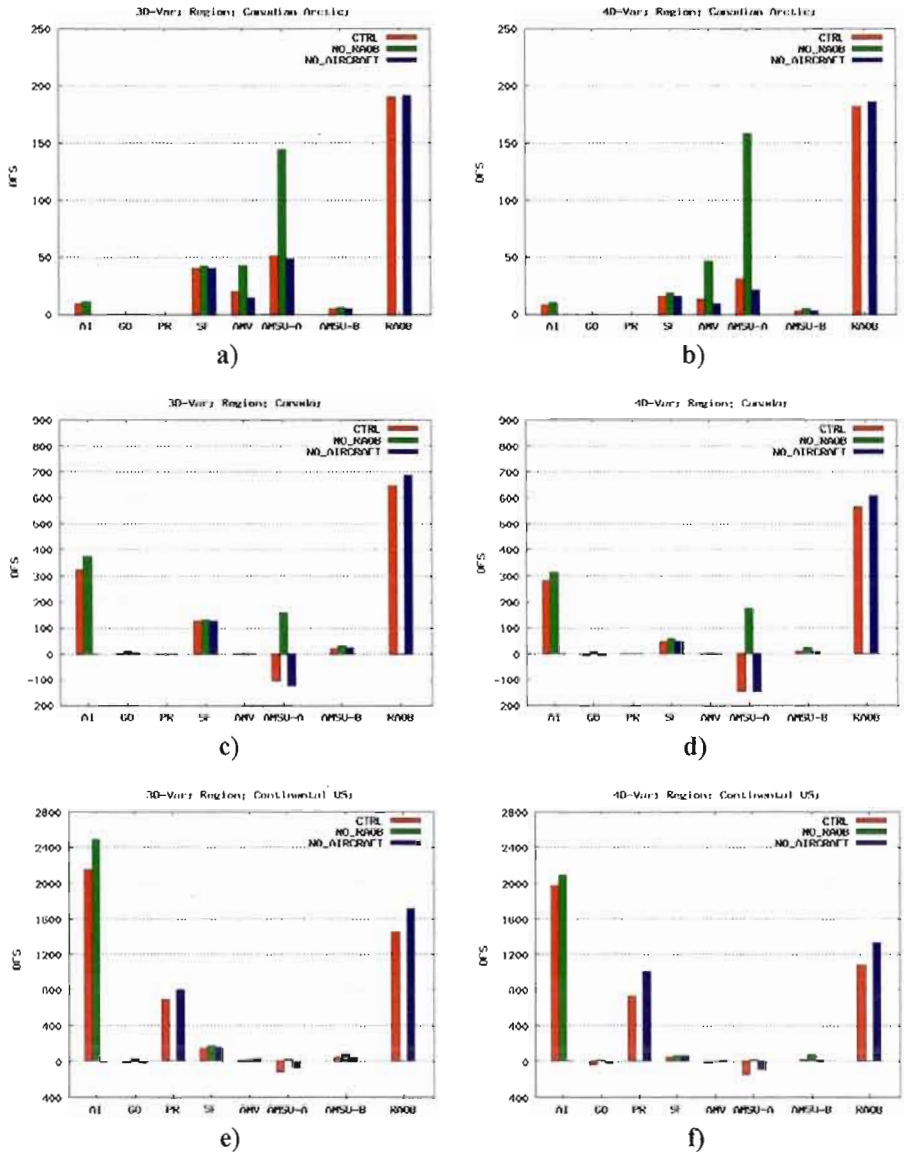
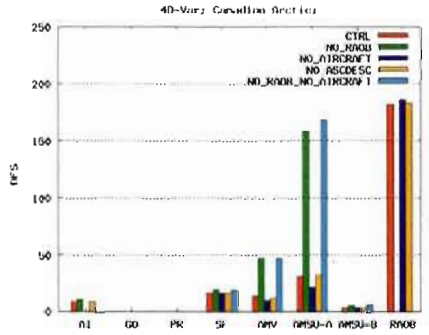
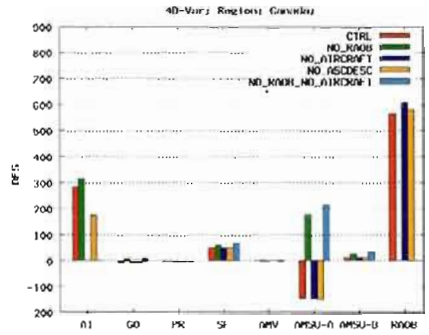


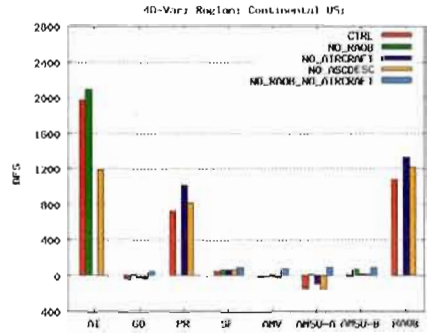
Figure 4.3: Same as Fig. 4.2 but over: a) Canadian Arctic, c) Canada and e) continental United States for the experiments with 3D-Var. Results with 4D-Var are in b) Canadian Arctic, d) Canada and f) continental United States. Experiments shown for each region include, from left to right, the control simulation and denials of radiosonde and wind profiler (NO_RAOB) and aircraft data (NO_AIRCRAFT) over the North America.



a)



b)



c)

Figure 4.4: North America 4D-Var data denial experiments. Averages values of *DFS* for eight families of observational data (see text for description) for the experiments CTRL (red bars), NO_RAOB (green bars), NO_AIRCRAFT (blue bars), NO_ASCENT-DESCENT (orange bars) and NO_RAOB_NO_AIRCRAFT (bars) inside a) Canadian Arctic, b) Canada and c) continental United States regions.

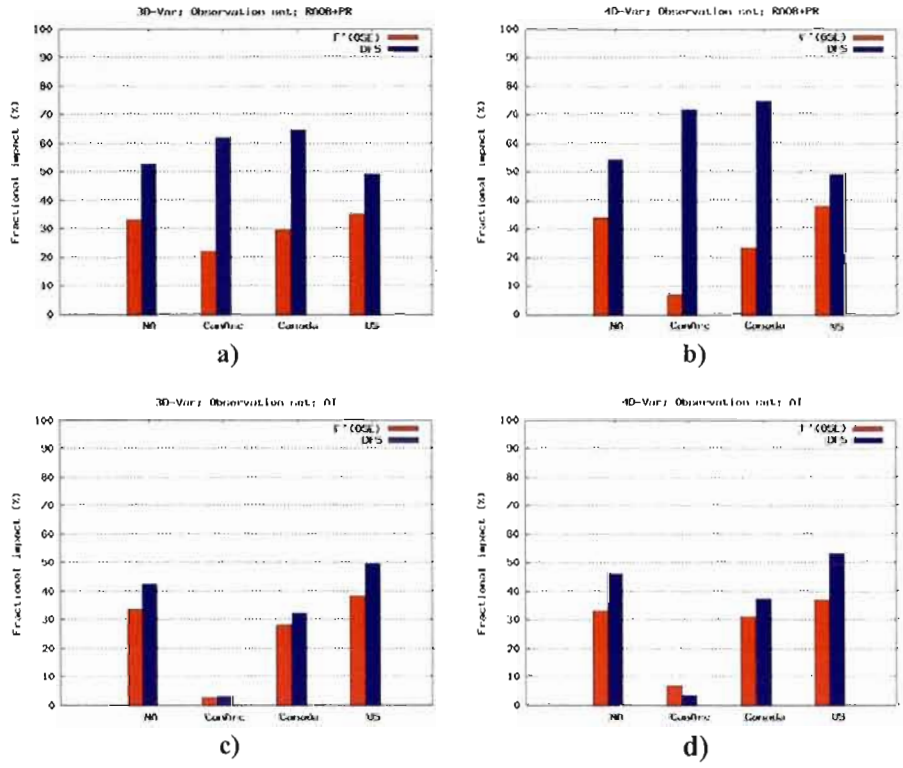


Figure 4.5: Averages values of $F^{Region}_{no\_k}$ and DFS^{Region}_{k} during January-February 2007 for two observation sets (k=RAOB+PR; AI;) over the different regions (North America, Canadian Arctic, Canada and Continental US). Results with 3D-Var are in the left panel while those with 4D-Var are in the right panel.

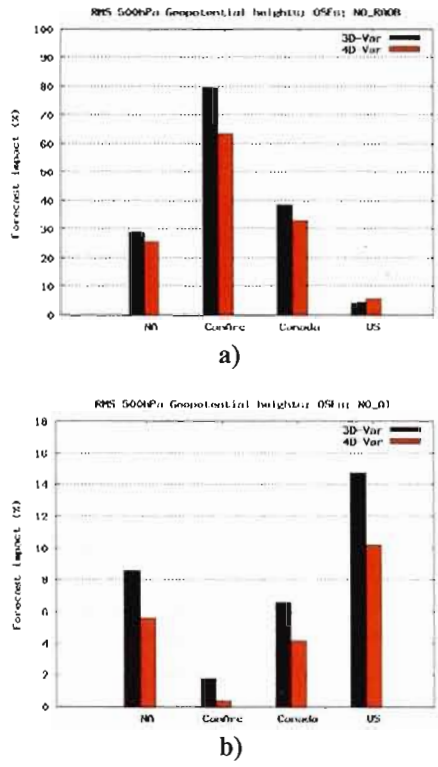


Figure 4.6: Forecast impact (%) for 500 hPa geopotential heights of the OSEs experiments (NO_RAOB, NO_AI) at 12 hours forecast period over four geographical areas.

CHAPITRE V

CONCLUSION

5.1 CONCLUSIONS PRINCIPALES DE LA RECHERCHE

L'assimilation de données joue un rôle important dans l'analyse des données atmosphériques, en particulier pour la prévision numérique du temps. Les vecteurs singuliers et les études de sensibilité aux conditions initiales permettent d'identifier, pour chaque situation météorologique, les régions sensibles, où la moindre erreur sur les conditions initiales peut croître très rapidement et avoir un effet considérable sur la qualité de la prévision. De ce constat a émergé l'idée qu'il est important que l'information sur les fonctions de sensibilité définissant la structure des changements aux conditions initiales qui ont le plus d'impact sur une prévision d'échéance donnée, soit considérée par le système d'assimilation de données. D'autre part, dans ces régions, il semblait naturel d'ajouter de nouvelles observations (observations ciblées) afin d'augmenter la précision de l'analyse et réduire de manière notable l'erreur de prévision.

La première partie de la thèse a eu comme objectif d'examiner des formulations différentes des covariances d'erreur de prévision, qui permettraient d'inclure à même le processus d'assimilation l'information sur les instabilités atmosphériques contenue dans une fonction de sensibilité. Ceci a été examiné dans le cadre de l'assimilation adaptative 3D-Var d'Environnement Canada. Essentiellement, le 3D-Var adapté propose un modèle de covariance d'erreur de prévision qui inclut la composante sensible tout en retenant le modèle plus conventionnel lorsque cette composante dite sensible est peu importante. La définition d'une fonction de structure appropriée pour un système d'assimilation vise à simultanément concorder aux observations disponibles et améliorer la qualité des prévisions.

Un des volets importants de cette recherche était d'évaluer l'observabilité des fonctions de structure par les observations. Nous avons montré que les fonctions de sensibilités qui peuvent expliquer la croissance de l'erreur de prévision dans une région particulière de l'espace au sens d'une norme convenablement choisie sont caractérisées par une faible amplitude et, autrement dit, le rapport entre le signal et le bruit est faible pour qu'elle soit observable.

La comparaison de l'analyse de sensibilité avec les observations montre qu'elle est plus loin des observations que l'analyse standard. Le 3D-Var adapté apporte des corrections à l'analyse selon la structure spatiale de la fonction sensible introduite et améliore légèrement l'ajustement de l'analyse aux observations. Toutefois, la prévision issue de l'analyse du 3D-Var adapté n'est pas améliorée par rapport à celle du 3D-Var conventionnel. Dans le cas limite où le modèle de covariance d'erreur de prévision considère seulement la composante sensible, nous avons montré que l'amplitude de la fonction sensible est déterminée par le coefficient de projection de l'innovation sur la direction de sensibilité.

Les différentes méthodes de télédétection satellitaire mises en œuvre ces dernières années ont permis d'augmenter le volume des données collectées par les satellites et actuellement, les données satellitaires constituent la plus importante source de données utilisée pour l'analyse météorologique en vue de la prévision numérique du temps. Étant donné qu'une faible fraction de l'ensemble des données est actuellement assimilée dans les systèmes d'assimilation des centres de NWP, il est important d'améliorer notre capacité d'extraire et d'assimiler plus de l'information contenue dans les observations satellitaires. Dans ce contexte, l'estimation de l'impact des données d'observation dans l'analyse et dans les prévisions à courte échéance est un enjeu important dans les centres de prévision numérique du temps.

Nous avons proposé une nouvelle méthode permettant d'estimer l'impact des observations dans les analyses et comparer l'importance de différents types d'observations. Le contenu en informations des observations est calculé en employant les statistiques *a posteriori*, à partir des écarts des observations à l'ébauche et à l'analyse. Les résultats montrent que le *DFS* estimé en utilisant les statistiques *a posteriori* est identique à celui obtenu à partir des statistiques *a priori*.

Les diagnostiques *a posteriori* permettent d'obtenir les statistiques d'erreur qui rencontrent les propriétés de consistance du système d'assimilation. En conséquence, les matrices de covariance d'erreur d'observations et d'erreur d'ébauche dans l'espace des observations, peuvent être estimées à partir des écarts des observations à l'ébauche (O-P) et à l'analyse (O-A) et des écarts de l'analyse à l'ébauche (A-P). Ces diagnostiques ont été appliqués premièrement dans l'assimilation 1D-Var pour l'estimation des variances d'erreurs d'observation ($\widetilde{\sigma}_o^2$) et d'ébauche ($\widetilde{\sigma}_b^2$) pour un jeu de 60 observations. Comme (A-P) et (O-A) dépendent eux-mêmes de la variance d'erreur d'observations et de l'ébauche, Desroziers *et al.* (2005) suggèrent d'utiliser de méthodes itératives permettant d'estimer les vraies valeurs de la variance d'erreurs d'observation et de l'ébauche. Les expériences 1D-Var ont montré que les itérations sur la variance d'erreur d'observation permettent d'estimer la vraie valeur de la variance d'erreur d'observation seulement dans le cas où la variance d'erreur de prévision est correctement spécifiée dans l'analyse et vice-versa. Dans le cas où la variance d'erreur de prévision est surestimée (sous-estimée), les itérations sur la variance d'erreur d'observation conduiront à une valeur sous-estimée (surestimée) de la variance d'erreurs d'observation. Les termes hors diagonaux de la matrice *a posteriori* de covariance d'erreur d'observation ont des petites valeurs et celle-ci est approximée par une matrice diagonale, facile à inverser. Les expériences 1D-Var réalisées ont permis de montrer que le *DFS* estimé à partir des diagnostiques *a posteriori* est le même avec celui estimé comme introduit par Girard (1987).

Après avoir validé la méthode d'estimation du *DFS* à partir des statistiques diagnostiquées, elle a été appliquée pour évaluer le *DFS* des observations assimilées avec le 3D et 4D-Var d'Environnement Canada. La méthode permet de quantifier l'impact des observations dans les analyses pour chaque type d'observations, par type d'instrument, par région géographique ou niveau vertical. Les résultats ont montré que les radiosondes, les données satellitaires AMSU-A/B et les données d'avion contribuent principalement à la qualité des analyses 3D/4D-Var d'Environnement Canada. Comme les systèmes d'assimilation sont en évolution continue, le travail peut se poursuivre pour évaluer l'impact des observations dans l'analyse opérationnelle 4D-Var qui assimile actuellement plus

d'observations ATOVS et AMVs, ainsi que nouvelles types des données comme AIRS, SSM/I et QuickScat Seawinds.

5.2 PERTINENCE DE LA RECHERCHE, CONTRIBUTION À L'AVANCEMENT DES CONNAISSANCES ET ORIGINALITÉ

De prévisions météorologiques de qualité sont demandées partout dans le monde pour répondre aux besoins de la société dans de nombreux domaines. Certains cas d'échec de la prévision peuvent être expliqués par des faiblesses de l'analyse, ce qui suggère qu'il faut d'une part repenser les méthodes d'assimilation pour être en mesure d'extraire l'information la plus significative pour le développement de systèmes météorologiques, et d'autre part améliorer le réseau d'observation. L'assimilation de données étant un processus constant de comparaison de prévisions obtenues d'un modèle de prévision contre des observations de tout genre, elle fournit également de l'information sur les faiblesses du modèle lui-même et permet de l'améliorer. C'est ainsi qu'on peut améliorer les analyses et prévisions atmosphériques. Des efforts sont également consacrés pour quantifier l'impact de différents types de données sur la qualité des analyses et des prévisions du temps.

L'originalité principale de la thèse consiste en : i) la quantification de l'observabilité des différentes fonctions de structure par les observations; ii) l'estimation de l'impact des observations dans les analyses 3D/4D-Var basé sur les statistiques *a posteriori;* iii) l'utilisation du *DFS* comme outil diagnostique dans les OSEs pour l'évaluation du réseau canadien d'observations.

La présente thèse s'appuie sur les travaux de chercheurs d'Environnement Canada qui ont fourni les facilités informatiques pour l'utilisation du système d'assimilation de données 3D/4D-Var (Gauthier *et al.*, 1999, 2007), ainsi que les fonctions de sensibilité *a posteriori* (Caron *et al.*, 2007a, b) calculés à partir du système d'analyse de sensibilité pour le système de prévisions (Laroche *et al.* 2002) et *a priori* (Mahidjiba *et al.*, 2007). J'ai développé la méthodologie et j'ai accompli les implantations nécessaires pour l'utilisation du 1D et 3D-Var adaptés. Pour la deuxième partie de la thèse, j'ai développé la méthodologie pour le calcul du *DFS* et j'ai réalisé tous les calculs diagnostiques nécessaires pour quantifier

l'impact des observations dans les analyses (contrôle et OSEs). Ceci a demandé la mise en place des outils permettant la manipulation de la base de données des observations. Les expériences OSEs avec le 3D et 4D-Var pour les deux mois d'hiver du 2007 ont été réalisées par Laroche et Sarazzin (2010 a, b).

La connaissance des techniques adaptatives ainsi que les expériences effectuées au cours de cette thèse ont montré que l'introduction des fonctions de structure dépendant de l'écoulement est délicate et que leur impact dans un système d'assimilation dépend de plusieurs éléments qu'il faut prendre en compte. Les outils développés au cours de cette thèse pourraient se révéler utiles pour comparer l'importance de différents types d'observations dans l'assimilation 3D et 4D-Var incluant tous les types d'observations assimilées opérationnellement.

5.3 LIMITES DE LA RECHERCHE

Comme nous l'avons déjà mentionné, dans le processus d'assimilation de données, il est fondamental de bien caractériser l'erreur associée à chaque source d'information (ébauche, observations) afin de mieux caractériser les éléments du système d'analyse qui peuvent influencer la qualité des analyses et des prévisions.

Les corrélations d'erreurs d'observations ne sont pas connues et leurs spécifications dans la matrice $\mathbf{R}$ est un enjeu important. Dans les schémas d'assimilation opérationnels, la matrice de covariance d'erreurs d'observations $\mathbf{R}$ est considérée diagonale pour plus de simplicité et pour réduire le coût de calculs. La variance de l'erreur de d'observation peut être estimée en utilisant la méthode basée sur les innovations (Hollingsworth et Lonnberg, 1986) ou à partir des diagnostiques introduits par Desroziers *et al.* (2005), mais il est généralement difficile d'estimer les corrélations d'erreurs d'observations. Bien que cette hypothèse soit raisonnable pour des observations mesurées par des instruments différents, elle est moins évidente dans le cas d'observations satellitaires. Dans ce cas, l'erreur d'observation est corrélée spatialement, temporellement, et spectralement (Bormann *et al.*, 2003). Par exemple, les vents AMV dérivés des satellites géostationnaires présentent des corrélations spatiales jusqu'à 800 km avec des différences insignifiantes entres les différents satellites, canaux et

niveaux verticaux considérés. Dans la pratique, l'omission des corrélations d'erreur d'observation dans la matrice $\mathbf{R}$ est compensée par une augmentation des valeurs de l'écart type d'erreur d'observation (Liu et Rabier, 2002). Ceci suggère qu'en négligeant les corrélations d'erreur d'observation, le contenu en information des observations n'est pas extrait de façon optimale.

Dans le troisième chapitre, la nouvelle méthode permettant d'estimer le *DFS* pour les différents types d'observation demande l'inversion de la matrice de covariance d'erreurs d'observations *a posteriori* $\tilde{\mathbf{R}}$. Pour notre étude, cette matrice a été approximée avec une matrice diagonale, facile à inverser et une décomposition en valeurs et vecteurs propres pour la matrice $\tilde{\mathbf{R}}$ prenait en compte les covariances. Toutefois, pour améliorer l'assimilation de données satellitaire, plus d'études sont nécessaires pour considérer les corrélations d'erreurs d'observations dans la matrice $\mathbf{R}$.

BIBLIOGRAPHIE

Bormann, N., C. Delsol et Thépaut J.-N., 2006: Challenges and plans for the assimilation of Atmospheric Motion Vectors. *Proceedings of the Procedings of Eighth International Winds Workshop.* Beijing, China.

Desroziers G., L. Berre, B. Chapnik et P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.,* **131**, 3385–3396.

Gauthier, P., M. Buehner et L. Fillion, 1998: Background-error statistics modelling in a 3D variational data assimilation scheme. Proceedings of the *ECMWF workshop on diagnosis of data assimilation systems*, Reading, UK, 131-145.

——, M. Tanguay, S. Laroche, S. Pellerin et J. Morneau, 2007: Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Mon. Weather Rev.,* **135,** 2339–2354.

Girard, D., 1987: A fast Monte Carlo cross-validation procedure for large least squares problems with noisy data. *Tech. Rep.* **687-M**, IMAG, Grenoble, France, 22pp.

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

Laroche, S., et R. Sarrazin, 2010a: Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part I: contribution of radiosonde, aircraft and satellite data. *Atmos.-Ocean.*

——, et R. Sarrazin, 2010b: Impact study with observations assimilated over North America and the North Pacific Ocean on the MSC global forecast system. Part II: Sensitivity experiments. *Atmos.-Ocean.*

Liu, Z. Q., et F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q. J. R. Meteorol. Soc*, **128**, 1367-1386.

Mahidjiba, A., M. Buehner et A. Zadra, 2007: Excitation of Rossby-wave trains: optimal growth of forecast errors. *Meteorologische Zeitschrift*, **16**, 665-673.