

SCA-7216
Introduction à l'assimilation de données

Pierre Gauthier
Département des Sciences de la Terre et de l'atmosphère, UQAM
Site Web pour le cours: <http://people.sca.uqam.ca/~gauthier>
Courriel: gauthier.pierre@uqam.ca
Local: PK-7210

Horaire de cours et évaluation

- **Horaire**
 - o Lundi: 10:30-12:00
 - o Jeudi: 10:30-12:00
- **Évaluation**
 - o Travaux pratiques 30%
 - (programmation Scilab)
 - Accès à Scilab (logiciel est gratuit et disponible pour Windows, Macintosh et Linux)
 - Site Web Scilab: (www.scilab.org, SCILAB 5.5.2)
 - o Examen No.1 (6 février 2017): 20%
 - o Examen No.2 (13 mars 2017): 20%
 - o Examen final (20 avril 2017): 30%
- **Semaine de relâche: 27 février au 3 mars**
SCA-7216 Assimilation de données

Références
APPLICATIONS AUX ÉCOULEMENTS GÉOPHYSIQUES

Kalnay, E., 2003: *Atmospheric Modeling, data assimilation and predictability*. Cambridge University Press, 341 pages.

Evensen, G., 2009: *Data assimilation: the ensemble Kalman filter*. Ed. Springer (SpringerLink), 307 pages.

Daley, R., 1991: *Atmospheric data analysis*. Cambridge University Press, Atmospheric and Space Science Series, 457 pages.

Rodgers, R.D., 2000: *Inverse Methods for Atmospheric Sounding: theory and practice*. World Scientific Series On Atmospheric and Planetary Physics, vol.2, 238 pages.

SCA-7216 Assimilation de données

Références: STATISTIQUES

Maybeck, P.S., 1979: *Stochastic models, estimation and control: vol. 1 et 2*. Academic Press, Mathematics in science and engineering, 712 pages.

von Storch, H. et F.W. Zwiers, 1999: *Statistical analysis in climate research*. Cambridge University Press, 484 pages

Rodgers, R.D., 2000: *Inverse Methods for Atmospheric Sounding: theory and practice*. World Scientific Series On Atmospheric and Planetary Physics, vol.2, 238 pages.

Tarantola, A., 2005: *Inverse problem theory and methods for model parameter estimation*. SIAM, Philadelphia, USA, 342 pages.

SCA-7216 Assimilation de données

Programme du cours

- ASSIMILATION DE DONNÉES ET LE PROBLÈME INVERSE
 - Assimilation de données et sa relation avec la prévision numérique: caractérisation des conditions initiales.
 - Problème inverse: comment reconstruire l'état instantané de l'atmosphère lorsque le nombre d'observations est insuffisant.
- RAPPEL DE CONCEPTS STATISTIQUES DE BASE
 - Variable aléatoire.
 - Densité de probabilité.
 - Espérance mathématique.
 - Variance et covariance d'erreur. Corrélation.
 - Régression linéaire univariée et multivariée
- MÉTHODE D'INTERPOLATION STATISTIQUE: CAS UNIVARIÉ
 - Estimateur de variance minimale.
 - Algorithme d'interpolation statistique: cas univarié.
 - Variances et covariances d'erreur de prévision.
 - Impact de l'erreur sur la spécification des statistiques d'erreur de prévision

SCA-7216 Assimilation de données

Programme du cours (suite)

- MÉTHODE D'INTERPOLATION STATISTIQUE: CAS MULTIVARIÉ
 - Inclusion de contraintes dynamiques dans les covariances d'erreur de prévision.
 - Opérateurs d'observations ou modèles direct.
 - Formulation variationnelle: le 3D-Var.
- ASSIMILATION QUADRI-DIMENSIONNELLE VARIATIONNELLE: LE 4D-VAR.
 - Assimilation variationnelle continue d'observations distribuée dans le temps.
 - Structures dynamiques implicites des covariances d'erreur de prévision du 4D-Var.
 - Notion de vecteur-singulier et sa relation aux éléments précurseurs à l'instabilité.
- EXTENSION DE L'INTERPOLATION STATISTIQUE AU CAS 4D: LE FILTRE DE KALMAN
 - Approche séquentielle à l'assimilation 4D: le filtre de Kalman.
 - Le filtre de Kalman d'ensemble

SCA-7216 Assimilation de données

Nature des observations et leur utilisation

SCA-7216 Assimilation de données

Principaux réseaux d'observation utilisés dans les systèmes de PNT

Micro-satellites GPS-RO
(COSMIC (4), GRAS, TERRASAR, GRACE)

Satellites orbites polaires
(NOAA-15,18,19, NPP; METOP-A,B; DMSP-F16,17,18; AQUA, TERRA)

Satellites géostationnaires
(GOES-E/W, Meteosat-E/W; MTSAT)

Avions
(BUFR, AIREP, AMDAR, ADS)

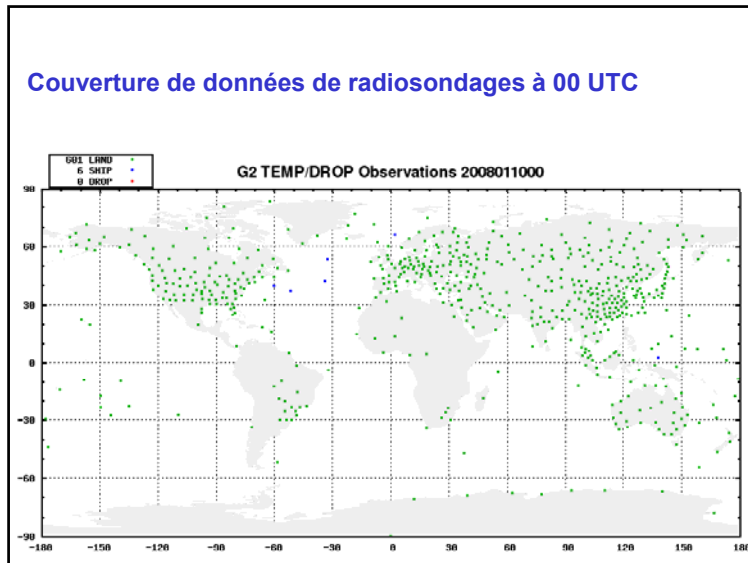
Profileurs de vent
(NOAA network)

Sites aérologiques
(TEMP, PILOT, DROP)

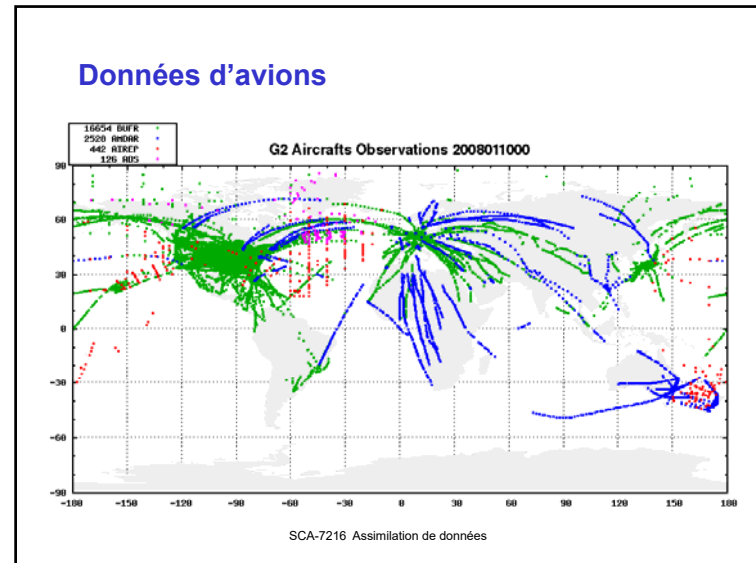
Bouées, Navires

Stations de surface
(SYNOP, ASYNOP, METAR, GR, GPS)

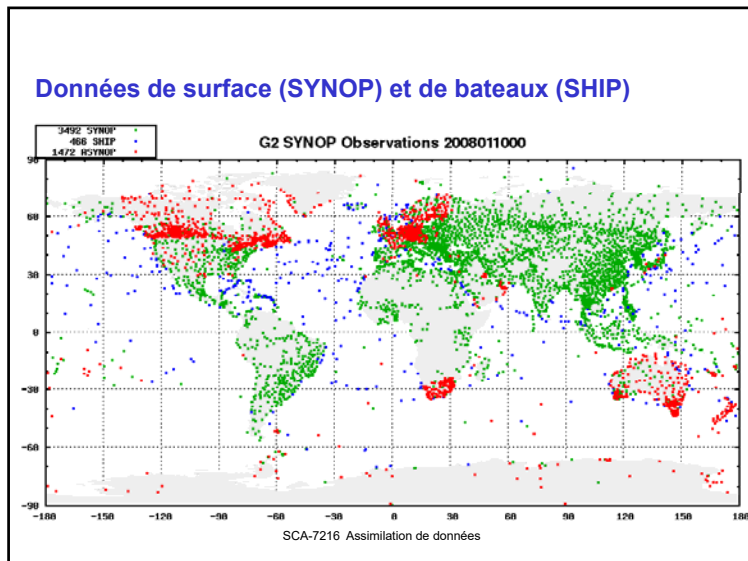
Couverture de données de radiosondages à 00 UTC



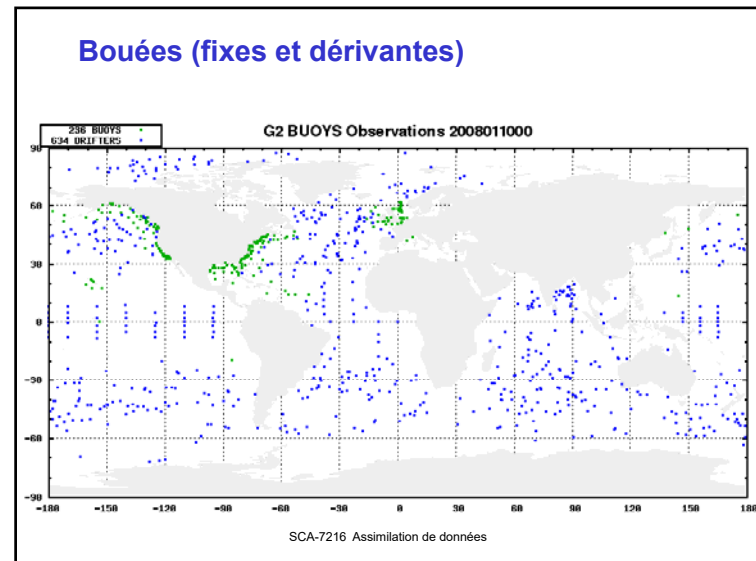
Données d'avions



Données de surface (SYNOP) et de bateaux (SHIP)

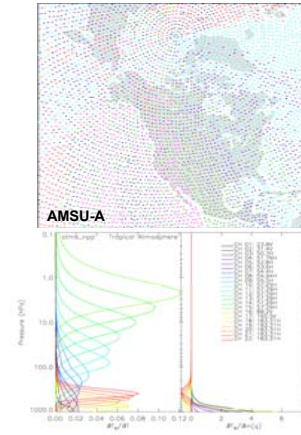


Bouées (fixes et dérivantes)

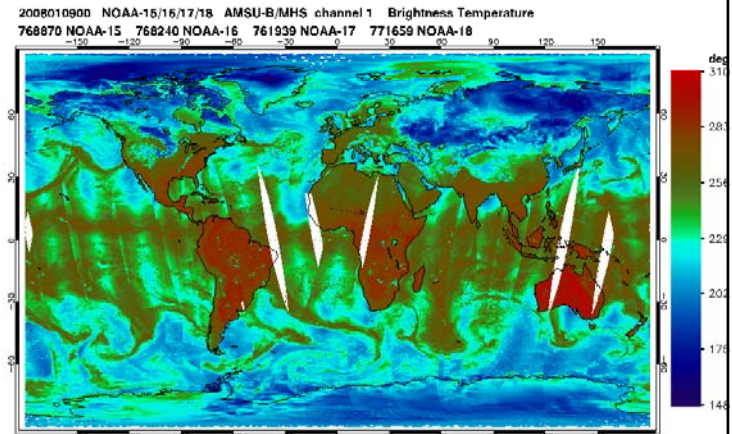


Observations satellitaires

- Présentement, les services météorologiques reçoivent des données de 18 satellites polaires et 5 géostationnaires, donnant une couverture globale tout les 6h.
- Radiances micro-ondes de AMSU-A, AMSU-B, MHS, ATMS, SSMIS, et infra-rouge de AIRS, IASI, Cris, CSR sont les principales sources de données des satellites pour la PNT.
- Ces radiances apportent de l'information sur la température et l'humidité ayant une bonne résolution horizontale, mais une résolution verticale plutôt grossière.

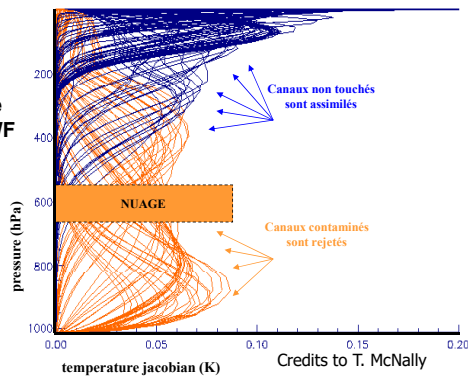


Couverture horizontale de données de radiance ATOVS AMSU-b (pour une période de 6-h): données reçues



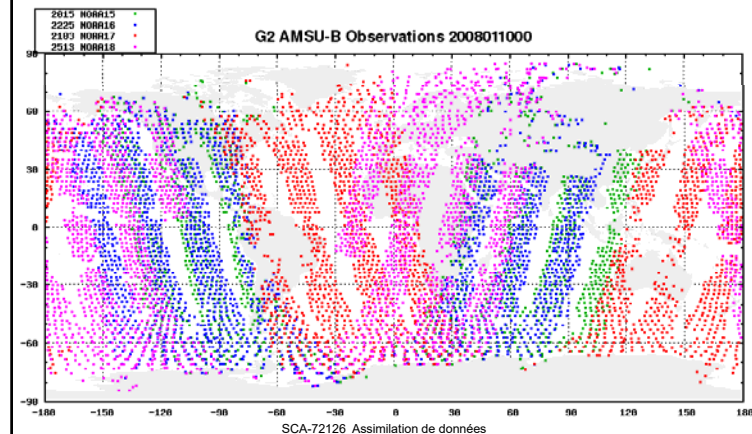
Détection de nuages

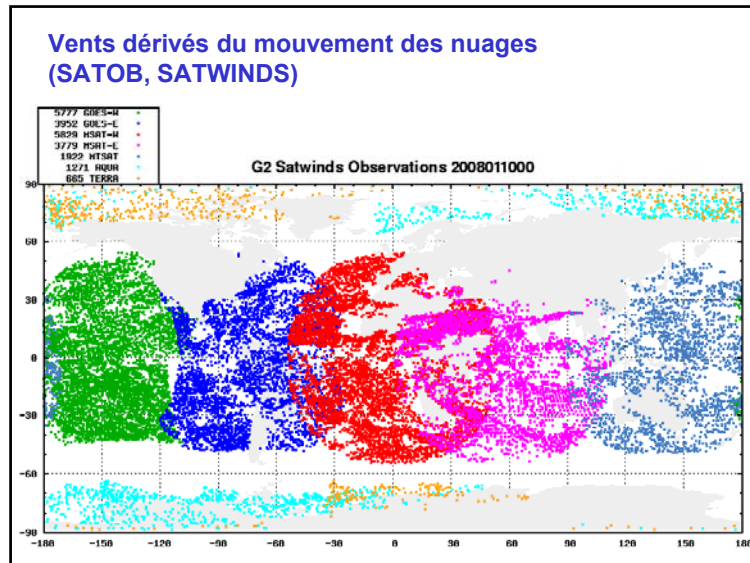
Information sur la base d'un seul canal: ECMWF scheme (McNally & Watts, 2001)



Atelier du CEPMMT sur l'assimilation des sondeurs hyperspectraux en prévision numérique du temps

Couverture horizontale de données de radiance ATOVS AMSU-b (pour une période de 6-h)

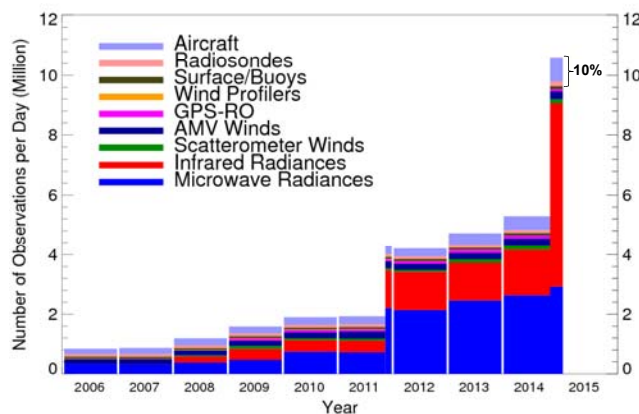




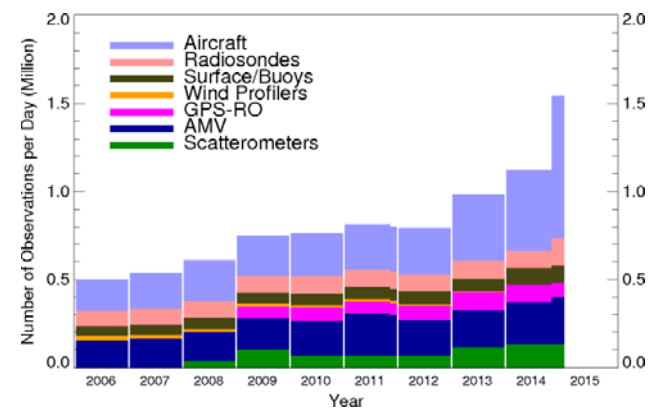
Principaux traitements des observations avant leur assimilation

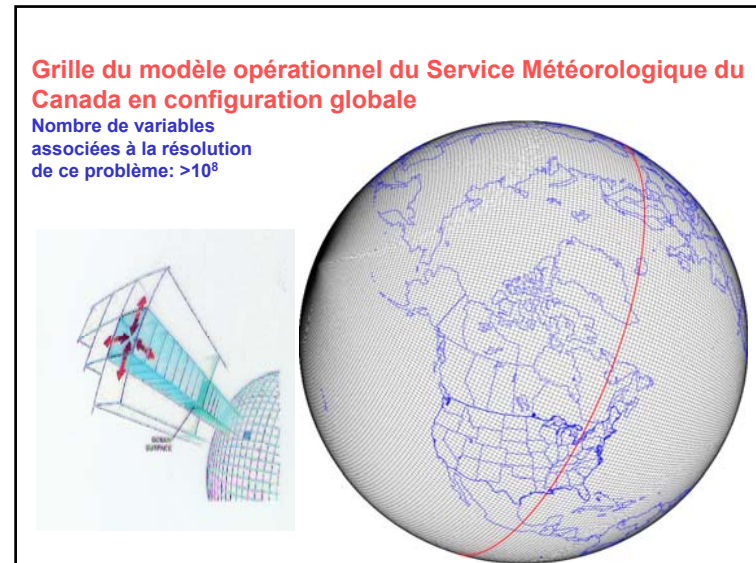
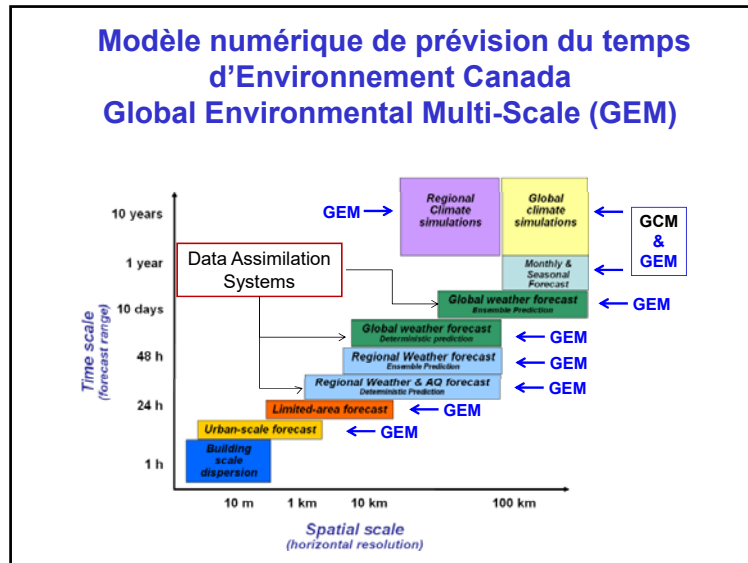
- **Extraction des données**
 - Provenant de diverses organisations.
- **Correction de biais**
 - Les données bien calibrées *a priori*, tels les radiosondages et GPS-RO, jouent un rôle important d'observations de référence.
- **Contrôle de qualité**
 - Traitement important des données puisque de grandes erreurs dans les observations peuvent compromettre les prévisions. Il y a trois étapes de contrôle de qualité:
- **Sélection de données**
 - La densité spatio-temporelle des observations doit être réduite afin d'éviter les corrélations d'erreur d'observation, ainsi que pour s'adapter aux limites des superordinateurs. C'est une des raisons qui expliquent l'accroissement constant des besoins pour des ordinateurs toujours plus puissants.

Nombre d'observations assimilées quotidiennement dans le SGPD



Nombre d'observations assimilées quotidiennement dans le SGPD





Problème d'estimation

- Variable d'état d'un modèle atmosphérique: composantes du vent, température, humidité etc. définies sur tous les niveaux du modèle et à chaque point de grille
- Nombre de paramètres à estimer:

$$\# \text{ (variables)} \times \# \text{ (niveaux)} \times \# \text{ (points de grille)}$$

$$= 4 \times 60 \times (800 \times 600)$$

$$\sim 10^8$$
- Nombre d'observations utilisées $\sim 10^6$
- Assimilation de données vise à corriger une prévision en utilisant les observations disponibles

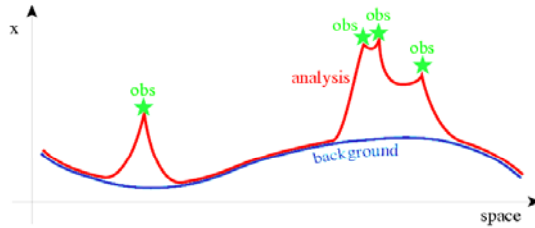
SCA-7216 Assimilation de données

Historique des méthodes d'assimilation

- **Utilisation d'un estimé *a priori* de l'état de l'atmosphère**
 - Climatologie, prévision numérique
- **Insertion directe**
 - remplacement de la valeur du champ d'essai (ou ébauche) par la valeur observée
- **Méthode de Cressman**
 - introduction de poids en fonction de la distance pour corriger les points avoisinants
- **Méthode des corrections successives**
(réf., Daley, 1991)

SCA-7216 Assimilation de données

Schématisation des méthodes d'insertion directe



SCA-7216 Assimilation de données

Difficultés de l'assimilation

- **Tenir compte de la précision relative du champ d'essai et de celles des observations**
 - Observations n'ont pas toutes la même précision
 - Champ d'essai possède une certaine qualité
- **Contraintes de balance dynamique**
 - Modifications de l'état atmosphérique causent un déséquilibre
 - Générations de transients dans les première heures d'intégration
- **Caractérisation des poids optimaux**

SCA-7216 Assimilation de données

Estimation statistique: cas univarié

X_b = prévision (ou champ d'essai)	$\varepsilon_b = (X_b - X_t)$:	erreur de prévision
X_o = observation	$\varepsilon_o = (X_o - X_t)$:	erreur d'observation
X_a = analyse	$\varepsilon_a = (X_a - X_t)$:	erreur d'analyse
et X_t = état réel		

Hypothèses sous-jacentes à l'interpolation statistique

- Erreur d'observation et de prévision sont non biaisées et non -corrélées
 $\langle \varepsilon_b \rangle = \langle \varepsilon_o \rangle = 0, \langle \varepsilon_o \varepsilon_b \rangle = 0$
- variances d'erreur σ_b^2 et σ_o^2 sont connues

Meilleur estimé linéaire non-biaisé (Best Linear Unbiased Estimate) (BLUE)

- Analyse

$$X_a = X_b + \lambda(X_o - X_b)$$

- Variance d'erreur d'analyse

$$J(\lambda) = \langle (X_a - X_t)^2 \rangle = \sigma_b^2(1-\lambda)^2 + \lambda^2\sigma_o^2$$

- Estimé de λ minimisant la variance d'erreur d'analyse

$$\lambda^* = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2} \Rightarrow \frac{1}{\sigma_a^2} = \frac{1}{\sigma_b^2} + \frac{1}{\sigma_o^2}$$

$$\Rightarrow \sigma_a^2 < \min \{ \sigma_b^2, \sigma_o^2 \}$$

SCA-7216 Assimilation de données

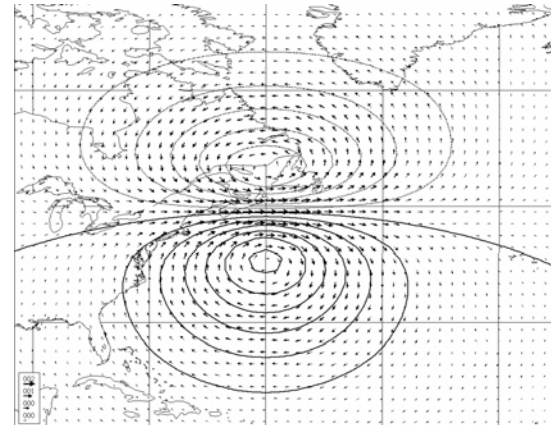
Interpolation statistique (ou méthode de Gauss-Markov)

(Gandin, 1963; Rutherford, 1973; Schlatter, 1977; Daley, 1991)

- Définitions**
- $\mathbf{y} \in \mathbb{R}^m$: vecteur d'observation ($m \sim 10^6$)
 - $\mathbf{X} \in \mathbb{R}^n$: vecteur d'état du modèle ($n \sim 10^8$)
 - $\mathbf{H}: \mathbb{R}^n \rightarrow \mathbb{R}^m$: Opérateur d'observation
 - $\mathbf{B} = \langle \varepsilon_b, \varepsilon_b^T \rangle$: covariances d'erreur de prévision
 - $\mathbf{R} = \langle \varepsilon_o, \varepsilon_o^T \rangle$: covariances d'erreur d'observation

SCA-7216 Assimilation de données

Incréments de vents et de géopotiel obtenus en réponse à une seule observation de la composante zonale du vent localisée à 265 hPa dans les extra-tropiques Nord (45N-60W) (Gauthier *et al.*, 1998)



Opérateurs d'observation

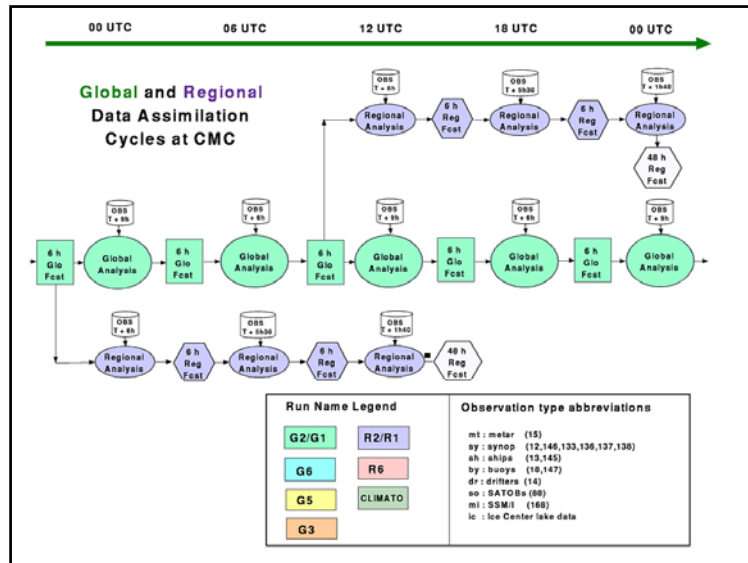
- **Problème d'inversion est mal posé**
 - Exemple: reconstruction d'un profil de température et d'humidité à partir d'une observation de radiance
 - Algorithmes d'inversion utilisés pour effectuer les recouvrements utilisent un champ de référence qui comble les lacunes de l'observation
 - Profils résultants sont hautement dépendants de cet état de référence (e.g., climatologie)
- **1D-Var: inversion utilisant la prévision du modèle**
 - $\langle \varepsilon_b, \varepsilon_o^T \rangle \neq 0$

SCA-7216 Assimilation de données

Cycle d'assimilation

- **Analyse contient l'information extraite des observations utilisées**
- **Prévision utilise cette analyse comme conditions initiales**
 - Transport de l'information dans le temps et l'espace
- **Champ d'essai cumule l'information extraite de l'ensemble des observations passées**

SCA-7216 Assimilation de données



Applications de l'assimilation de données

- **Conditions initiales (analyse) pour la prévision numérique du temps**
- **Réanalyses**
 - Utilisées pour étudier l'évolution du climat
- **Assimilation de constituants chimiques**
 - Chimie de l'atmosphère (ozone, méthane)
- **Diagnostics de modèle**
 - comparaison directe contre les observations (e.g., radiances)

SCA-7216 Assimilation de données