Techniques Mathématiques en Assimilation de Données pour la Prévision Météorologique

Yann MICHEL

[1] Météo-France
Centre National de Recherches Météorologiques (CNRM-GAME)
Groupe de Modélisation pour l'Assimilation et la Prévision (GMAP)

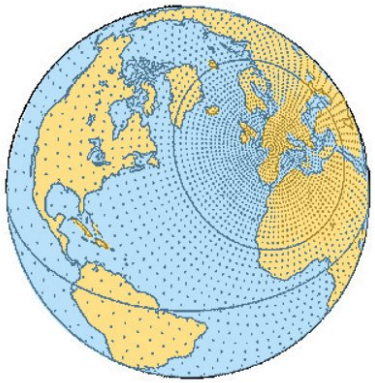
Journée "Mathématiques pour l'Ingénierie : de la Modélisation à la Résolution de Problèmes Concrets"

31 mars 2010

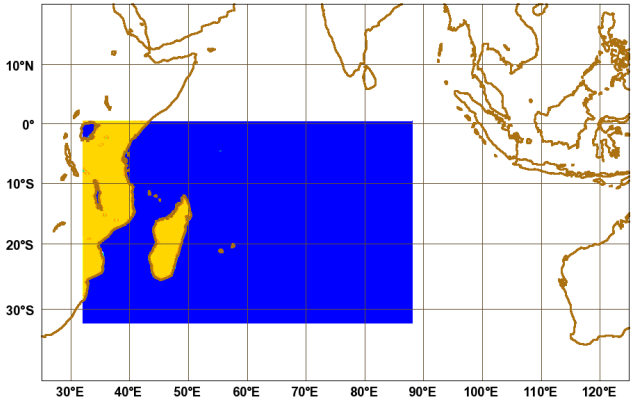
La Prévision Numérique du Temps (PNT) s'appuie sur deux domaines complémentaires, la modélisation et l'assimilation.

La modélisation numérique

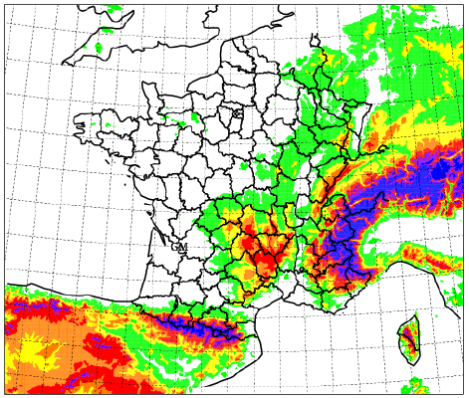
La modélisation numérique vise à développer des schémas numériques pour résoudre les équations gouvernant l'atmosphère sur une grille.
Météo-France s'appuie sur une hiérarchie de modèles aux résolutions spatio-temporelles de plus en plus fines et aux physiques de plus en plus complexes.



(a) Modèle global ARPEGE, équations primitives, T538C2.4



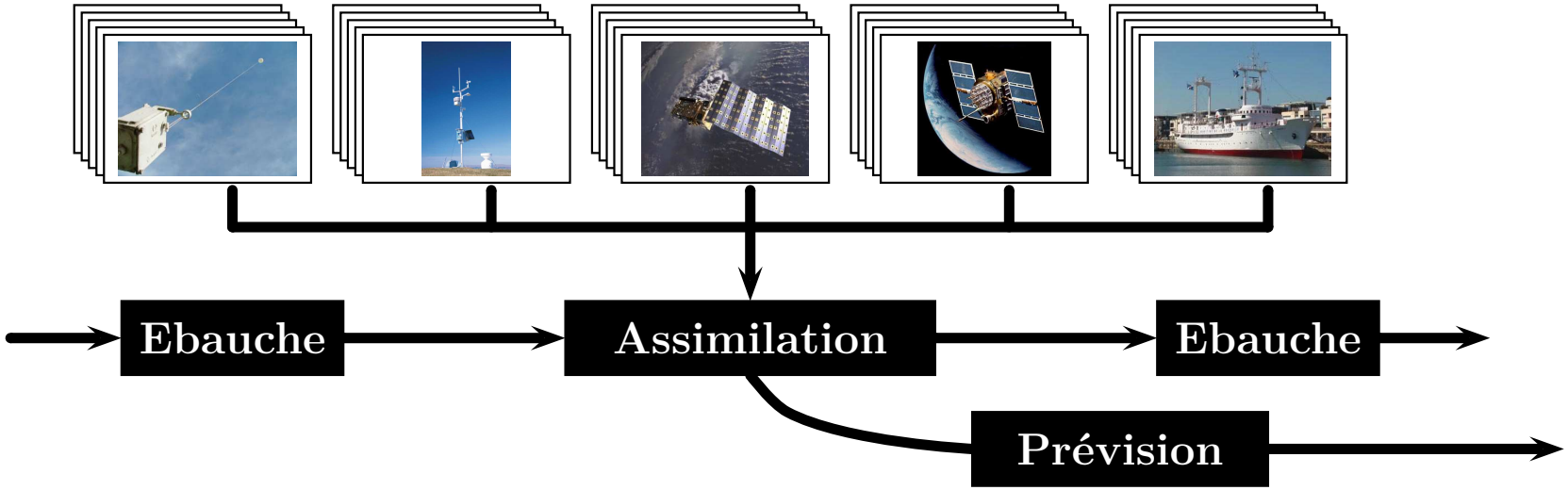
(b) Modèle régional ALADIN-Réunion à 9 km



(c) Modèle régional AROME-France, non-hydrostatique, 2.5 km

L'assimilation de données

L'assimilation de données (AD) vise à estimer les conditions initiales du modèle (et divers paramètres) en tenant compte des sources d'information disponibles: une prévision précédente (appelée ébauche) et des observations.



Les observations sont très diverses en nature (radiosondages, mesures sol, bateaux, avions de ligne, données GPS, radiances des satellites géostationnaires et défilants...)

Techniques Mathématiques en Assimilation de Données pour la Prévision Météorologique

1. Introduction
2. Techniques Mathématiques en Modélisation
3. Techniques Mathématiques en Assimilation de données
4. Perspectives

1 Introduction

2 Techniques Mathématiques en Modélisation

3 Techniques Mathématiques en Assimilation de données

4 Perspectives

Schémas numériques pour les modèles globaux spectraux

Par rapport à d'autres problèmes d'ingénierie, la spécificité de la météorologie provient de la géométrie sphérique, des contraintes fortes en coût de calcul et de la necessité de développer des paramétrisations physiques (très large gamme d'échelles spatiales et temporelles).

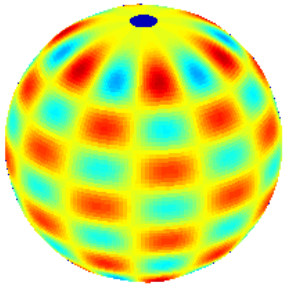


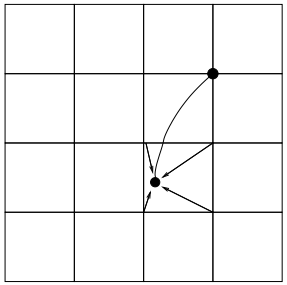
Figure: Y_{12}^{6}

Discrétisation en harmoniques sphériques

On décompose les champs météorologiques en harmoniques sphériques ($\sim$ séries de Fourier) $\sum_{l=0}^{+\infty} \sum_{m=-l}^{+l} C_l^m \cdot Y_l^m(\theta, \varphi)$

Schéma d'advection Semi-Lagrangien

Remplacement de l'advection (explicite) par une interpolation rétrograde, ce qui permet de s'affranchir de la condition de stabilité CFL.



Autre problème mathématique: condition aux limites bien posées (couplage) avec des modèles à aire limitée.

Schémas numériques en éléments finis

Exemple de MPAS (NCAR, Etats-Unis)

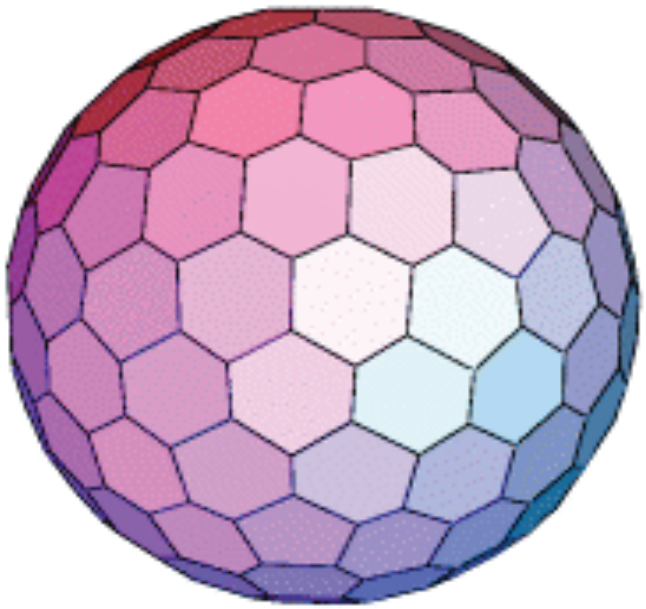
Avantages: Raffinement de maillage, couplage aisé entre modèles locaux et régionaux, adaptation aux architectures massivement parallèles, bonnes propriétés de conservation.

$$\partial_t u_1 + g\delta_{x_1} h + \frac{J}{\sqrt{3}}(u_{31} - u_{21}) = 0$$

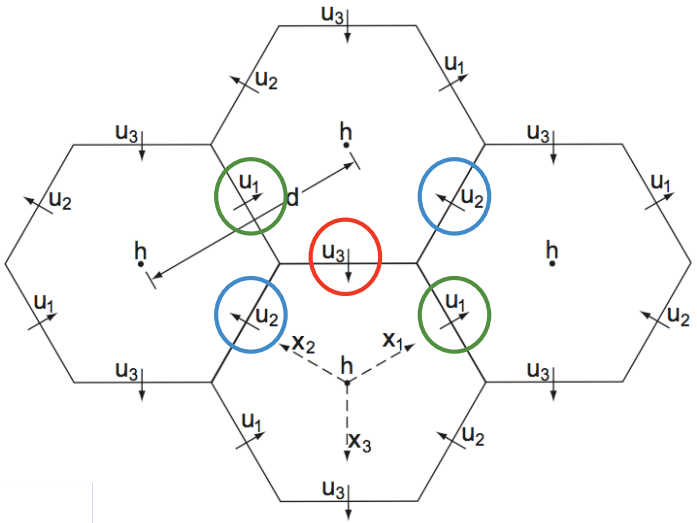
$$\partial_t u_2 + g\delta_{x_2} h + \frac{f}{\sqrt{3}}(u_{12} - u_{32}) = 0$$

$$\partial_t u_3 + g\delta_{x_3} h + \frac{f}{\sqrt{3}}(u_{23} - u_{13}) = 0$$

$$\partial_t h + \tfrac{2}{3}H(\delta_{x_1} u_1 + \delta_{x_2} u_2 + \delta_{x_3} u_3) = 0$$



(a) C-grille hexagonale sur la sphère



(b) Evaluation de la force de Coriolis

J. Klemp, B. Skamarock (NCAR), T. Ringler (Los Alamos National Laboratory), J. Thuburn (University of Exeter, UK)

1 Introduction

2 Techniques Mathématiques en Modélisation

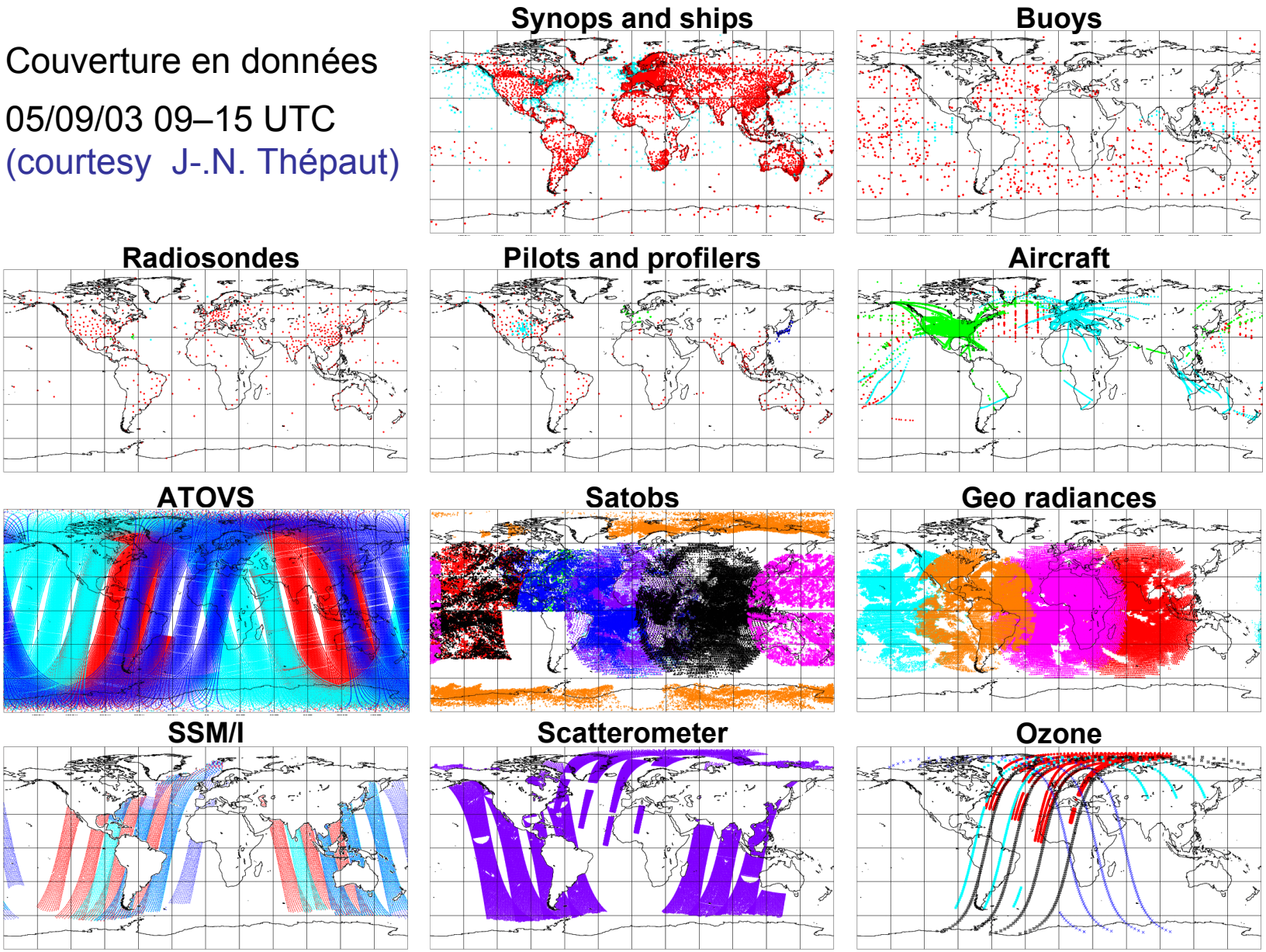
3 Techniques Mathématiques en Assimilation de données

4 Perspectives

Les observations

Couverture en données

05/09/03 09–15 UTC
(courtesy J-.N. Thépaut)



L'assimilation de données

L'assimilation de données recherche l'état $\mathbf{x}$ le plus probable de l'atmosphère étant donné une ébauche $\mathbf{x}_b$ et des observations $\mathbf{y}$

$$\mathbb{P}(\mathbf{x}|\mathbf{y}) \propto \mathbb{P}(\mathbf{y}|\mathbf{x})\mathbb{P}(\mathbf{x})$$

Dans le cadre Gaussien, ces probabilités s'écrivent:

$$\mathbb{P}(\mathbf{x}) \propto e^{-\frac{1}{2}(\mathbf{x}-\mathbf{x_b})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x_b})} \text{ et } \mathbb{P}(\mathbf{y}|\mathbf{x}) \propto e^{-\frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x}))}$$

où $\mathbf{B}$ et $\mathbf{R}$ sont les matrices de covariances des erreurs d'ébauche et d'observations. L'analyse $\mathbf{x}_a$ est le minimum de la fonction de coût J:

$$\begin{aligned}\mathbf{x}_a =& \text{Argmin } J \\ J(\mathbf{x}) =& \underbrace{\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)}_{J_b \text{ écart à l'ébauche}} + \underbrace{\frac{1}{2}(\mathbf{y}-\mathcal{H}(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x}))}_{J_o \text{ écart aux observations}}\end{aligned}$$

Après linéarisation ($\mathcal{H} \rightarrow \mathbf{H}$),

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}(\mathbf{y} - \mathcal{H}(\mathbf{x}_b))$$

Cadre mathématique du Filtre de Kalman.

L'assimilation de données variationnelle

L'approche variationnelle évite la résolution directe du système linéaire et préfère minimiser J

Avantages

- de nombreux algorithmes efficaces de minimisation existent (gradient-conjugué...),
- extension possible à la dimension temporelle par la méthode adjointe (4D-Var)

Inconvénients

- on doit spécifier la matrice $\mathbf{B}$,
- les opérateurs d'observations doivent être linéarisables
- algorithme de nature séquentielle (adaptation au massivement parallèle?)

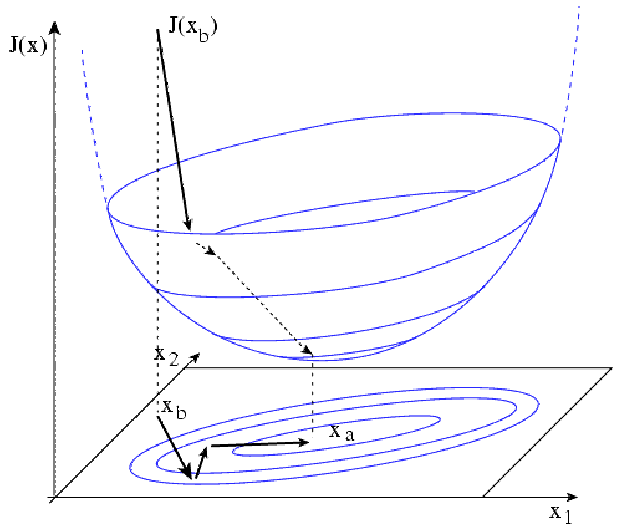


Figure: Minimisation itérative de J par descente du gradient (Figure due à F. Bouttier).

L'algorithme 4D-Var, inventé dans les années 80 (Le Dimet, Talagrand, Courtier...), est implémenté dans de nombreux centres opérationnels de PN. Conjointement à une assimilation massive des données satellite, il a conduit à une progression régulière des prévisions.

L'extension temporelle dans le 4D-Var

$$\delta_{\mathbf{x}} = \mathbf{x} - \mathbf{x}_b$$

$$\nabla J(\delta_{\mathbf{x}}) = \mathbf{B}^{-1}(\delta_{\mathbf{x}}) - \sum_{i=0}^{n} \mathbf{M}_1^{\mathrm{T}} \cdots \mathbf{M}_i^{\mathrm{T}} \mathbf{H}_i^{\mathrm{T}} {R_i}^{-1} (\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_b) - \mathbf{H}\delta_{\mathbf{x}})$$

La minimisation nécessite le calcul de ∇J, **gradient** de J par rapport à la variable de contrôle $\mathbf{x}$ qui est l'état initial. C'est l'utilisation de l'adjoint du modèle tangent-linéaire $\mathbf{M}$ qui permet de différencier J par rapport à l'état initial.

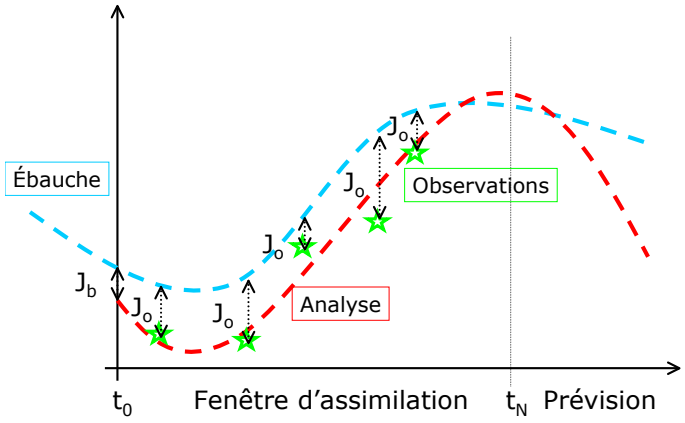


Figure: Illustration de la dimension temporelle dans le 4D-Var (F. Bouttier).

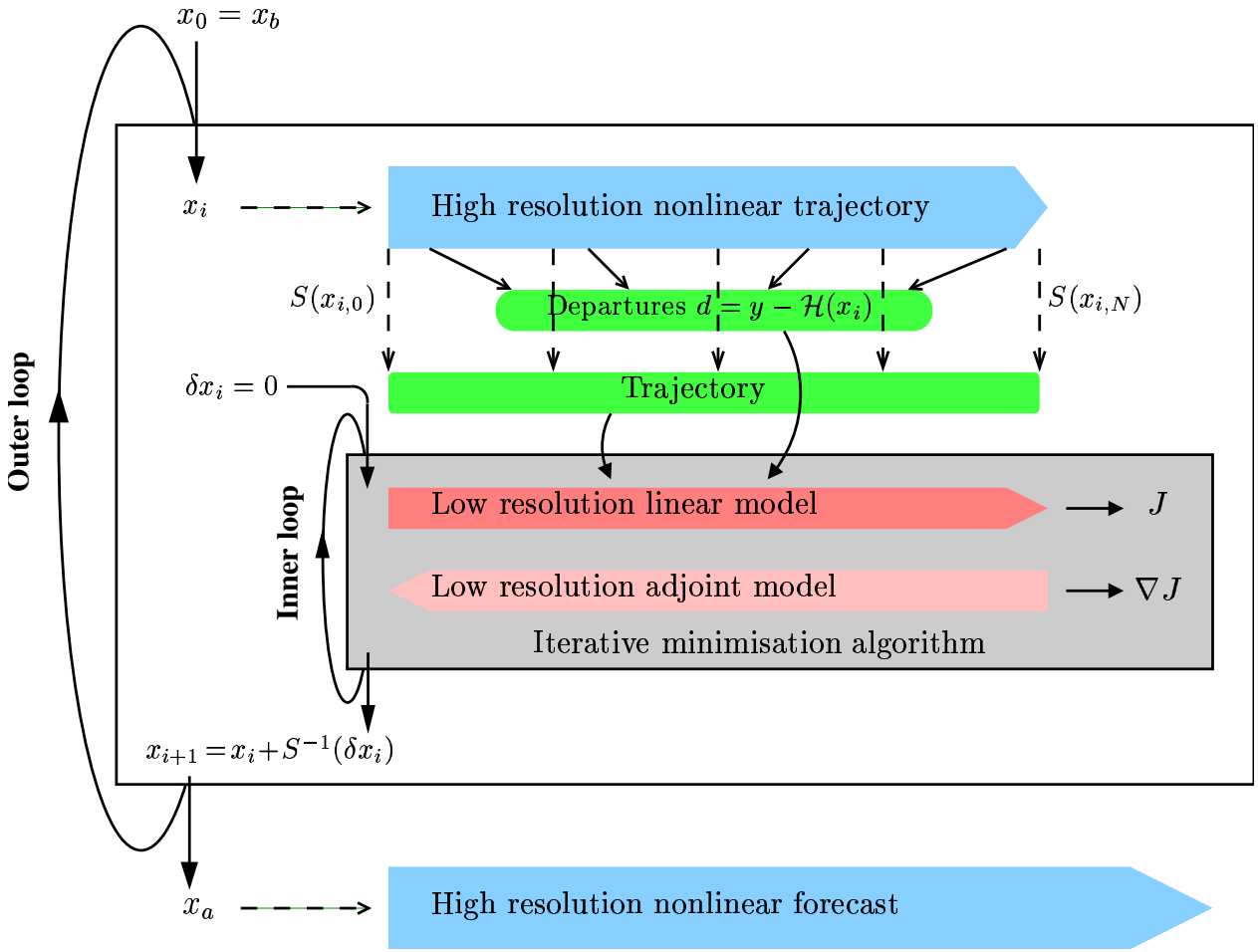


Figure: Traitement des non-linéarités par relinéarisation successives dans le 4D-Var opérationel au CEPMMT (Y. Tremolet)

La modélisation de la matrice $\mathbf{B}$

La matrice $\mathbf{B}$, de dimensions $10^7 \times 10^7$, est beaucoup trop grande pour être stockée et même estimée. Elle est décrite sous la forme d'opérateurs dans un changement de variable de contrôle $\mathbf{L} = \mathbf{B}^{\frac{1}{2}}$

$$\delta_{\mathbf{x}} = \mathbf{L}\mathbf{v}$$

$$J(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{v} + \frac{1}{2}\big(\mathbf{d} - \mathbf{HLv}\big)^{\mathrm{T}}\mathbf{R}^{-1}\big(\mathbf{d} - \mathbf{HLv}\big)$$

Avantage: amélioration du préconditionnement de la minimisation.
Il y a en général autant de formulations pour "$\mathbf{L}$" que de centres météorologiques (Bannister 2008).
$\mathbf{L}$ est une général une séquence d'opérateurs construit et calibrés pour représenter:

- les covariances croisées entre différentes variables, ou "balance"
- les autocovariances horizontales et verticales, qui propagent l'information spatialement.

Le problème est à la fois "physique" et "statistique".

La modélisation de la matrice **B**: l'exemple des corrélations bi-dimensionnelles

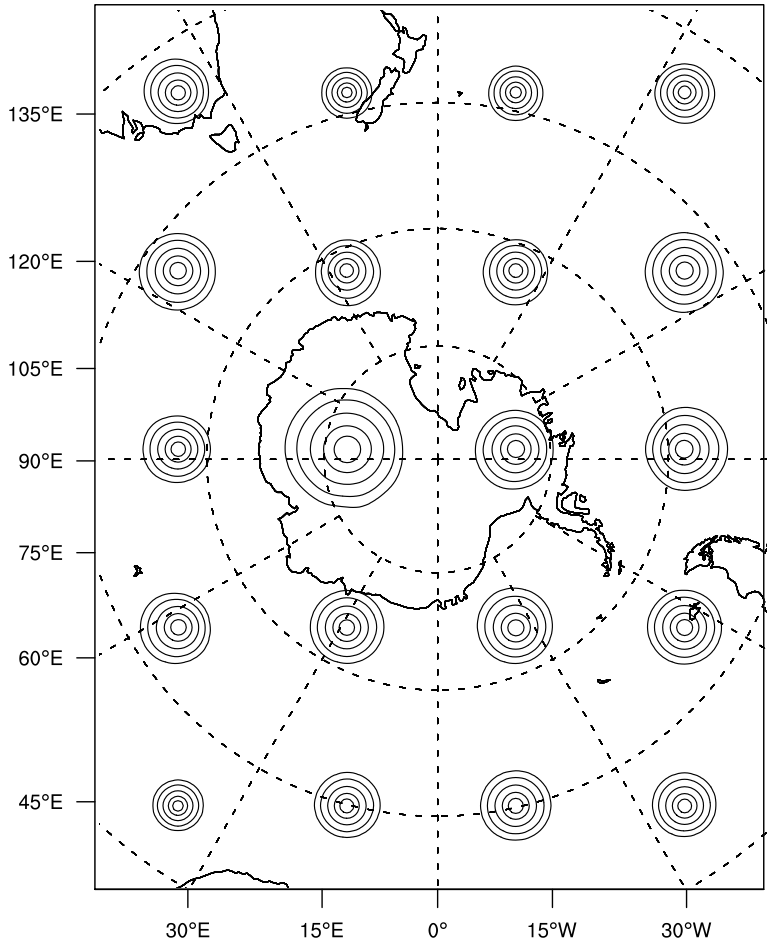


Figure: Corrélations horizontales obtenue avec AMPS pour la pression de surface non-balancée.

Utilisation d'une équation de diffusion ou de filtres récursifs

On cherche à exploiter le fait que la solution d'une équation de diffusion représente, sous la condition d'une normalisation appropriée, un opérateur de corrélation.

Diagonal-spectral; diagonal-ondelettes

D'autres modélisations peuvent être basées sur l'hypothèse de diagonalité dans une base (ou pseudo-base) d'ondelettes.

L'assimilation de données ensembliste

L'idée du filtre de Kalman d'ensemble est de représenter la loi recherchée par un échantillon discret (quelques membres). La matrice $\mathbf{B}$ est échantillonnée sur l'ensemble:

$$\mathbf{B} \simeq \overline{\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)\left(\mathbf{x}^f - \overline{\mathbf{x}^f}\right)^{\mathrm{T}}}$$

La propagation des pdf est assurée par l'évolution non-linéaire des membres.

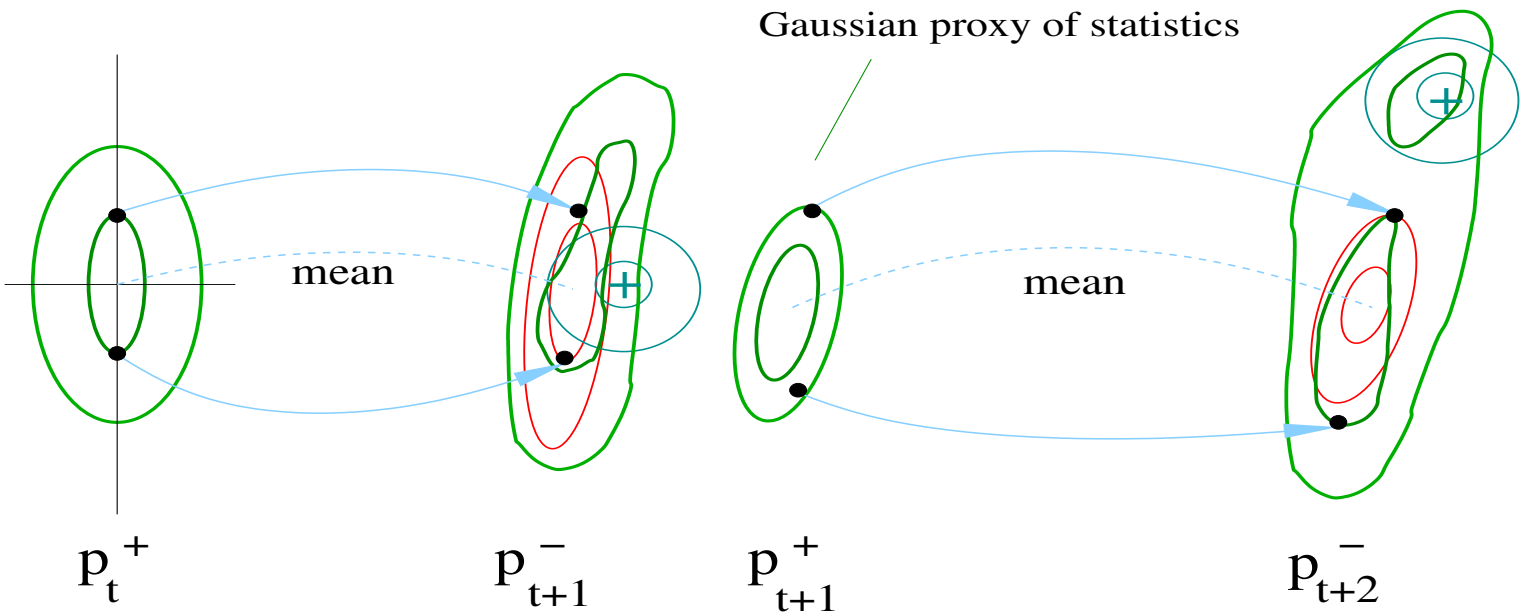
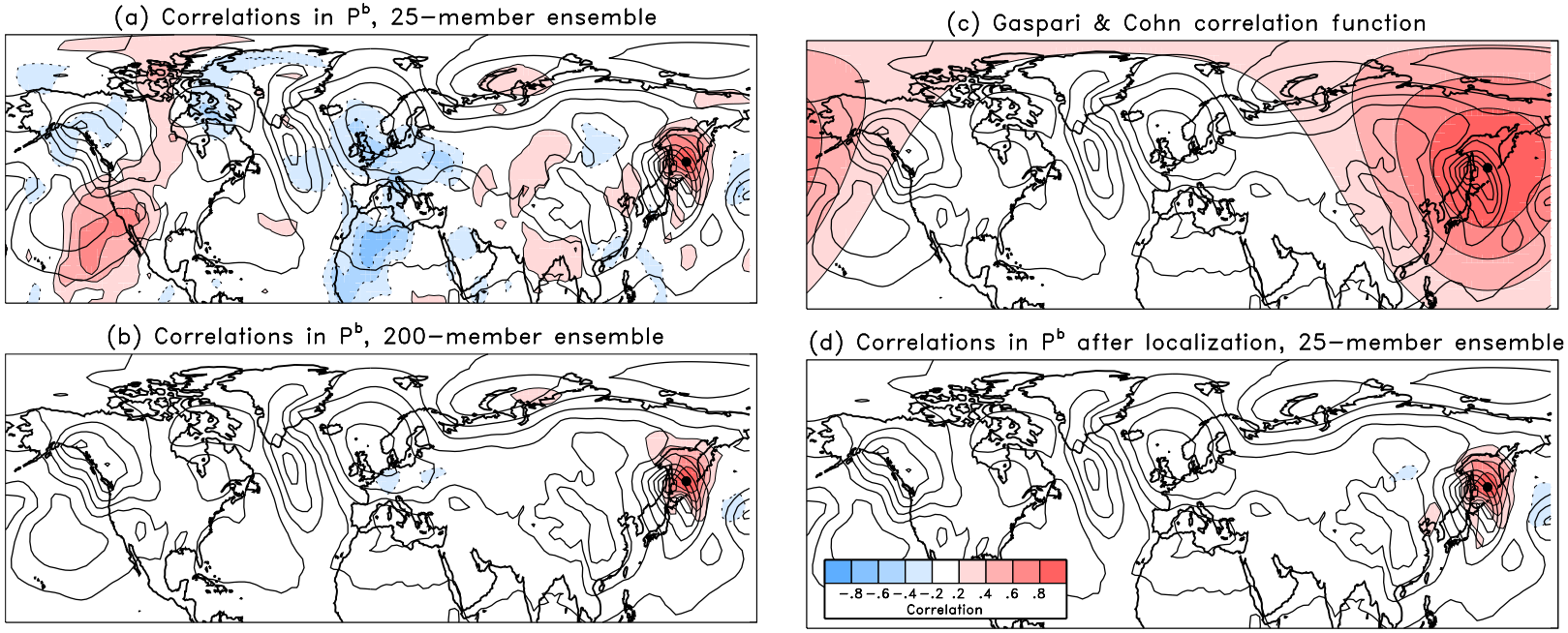


Figure: Schéma temporel d'un filtre de type Kalman d'Ensemble (Monte Carlo). Figure due à M. Bocquet (ENPC-EDF/CEREA)

L'assimilation de données ensembliste: bruit d'échantillonnage

Définition de méthodes objectives de filtrage des corrélations indésirables à longue distance dans **B** par produit de Schur ("tapering").



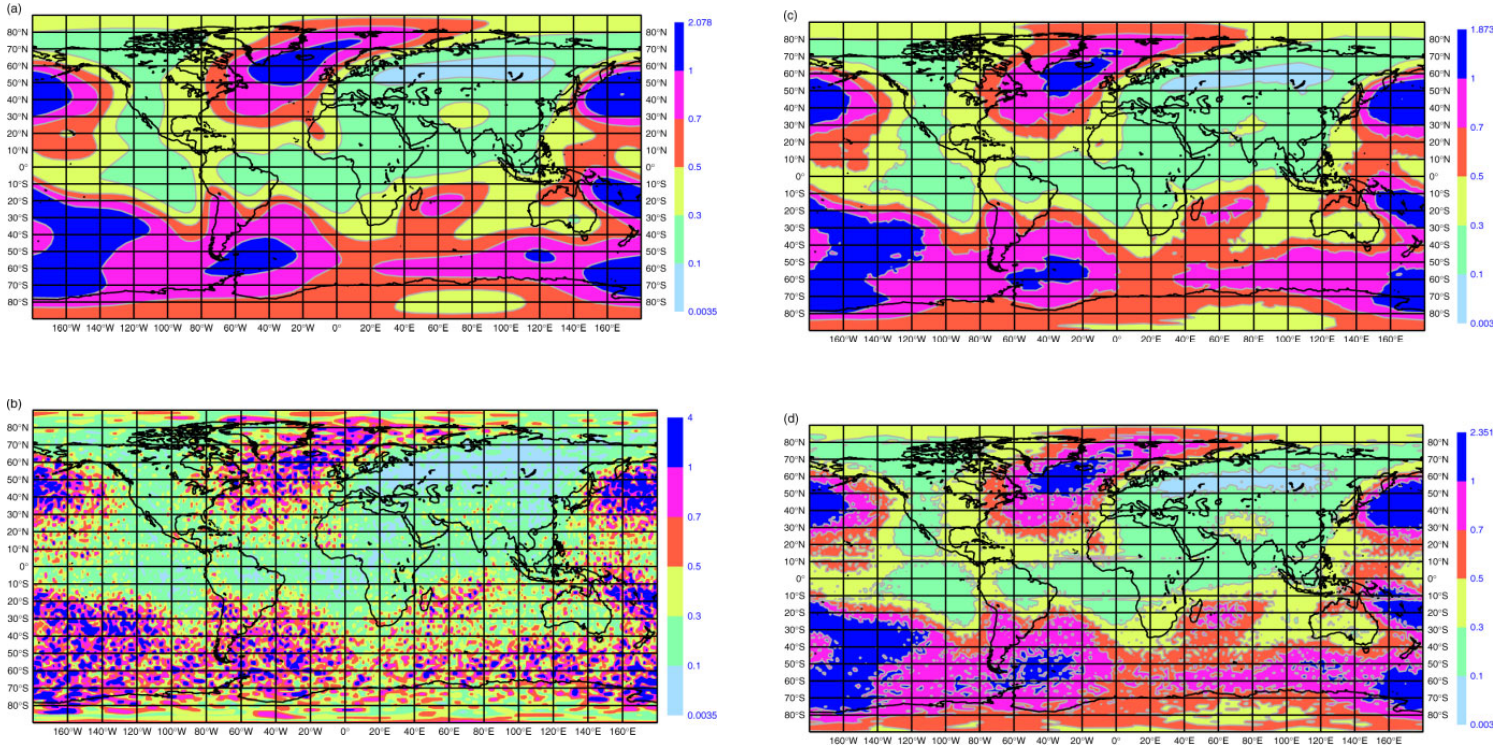
(a) Estimation brutes (25 membres en haut; 200 membres en bas)

(b) En haut: Une fonction de correlation. En bas: Correlation obtenue en multipliant par produit de Schur l'estimation de **B** avec 25 membres et cette fonction de correlation (haut).

Figure: Localisation des corrélations. Figure due à T. Hamill (CU/NOAA)

L'assimilation de données ensembliste: bruit d'échantillonnage

Définition de méthodes objectives de filtrage des variances d'erreurs d'ébauche grâce à une estimation du rapport signal/bruit.



(a) Haut: référence. Bas: Estimation brute avec 6 membres

(b) Haut: Filtrage de l'estimation obtenue avec 6 membres. Bas: Estimation brute avec 220 membres

Figure: Filtrage spatial des variances en tourbillon relatif à la surface. Figure due à L. Raynaud (CNRM/GMAP)

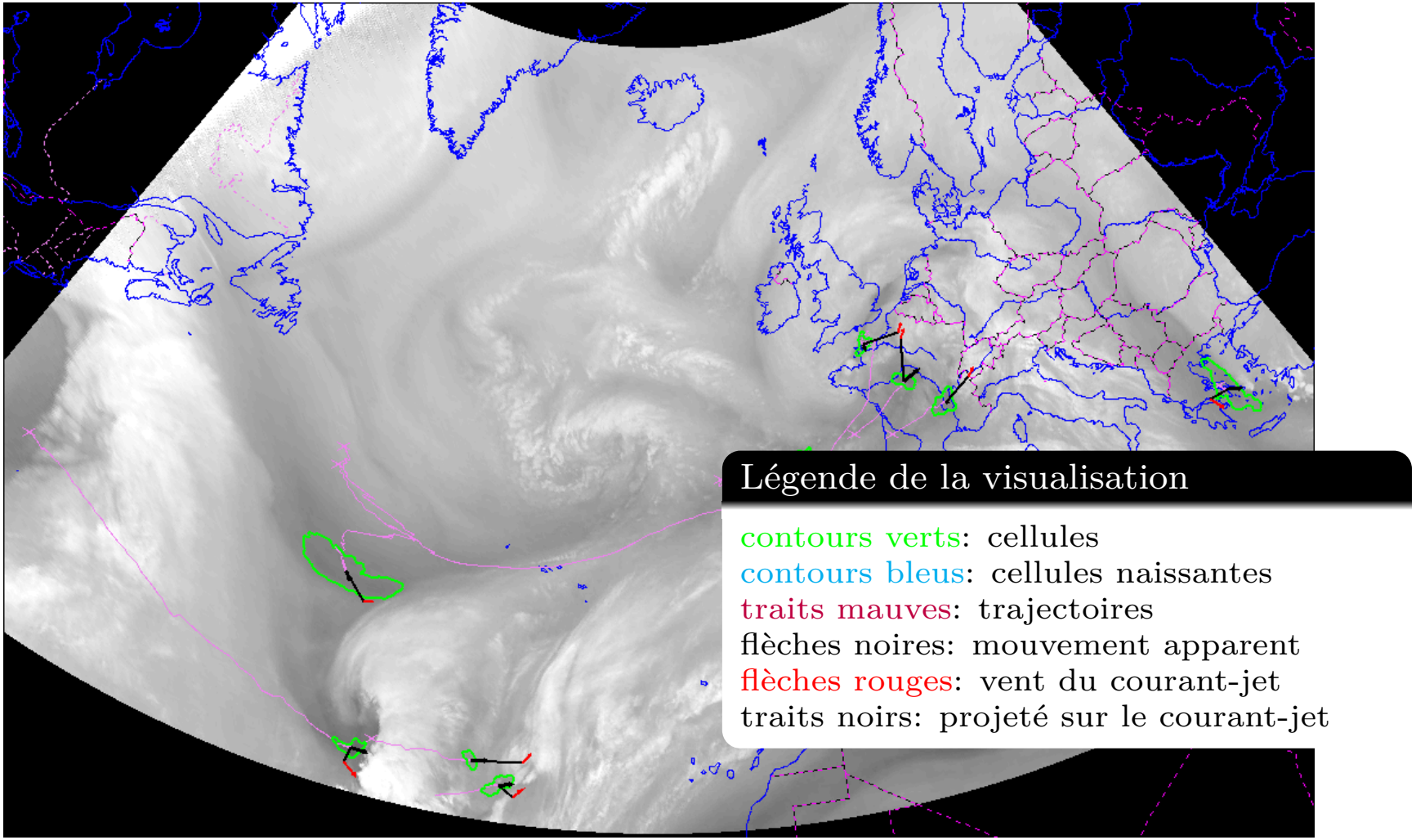


Figure: Antidote sur la tempête Xynthia (26-27 février 2010)

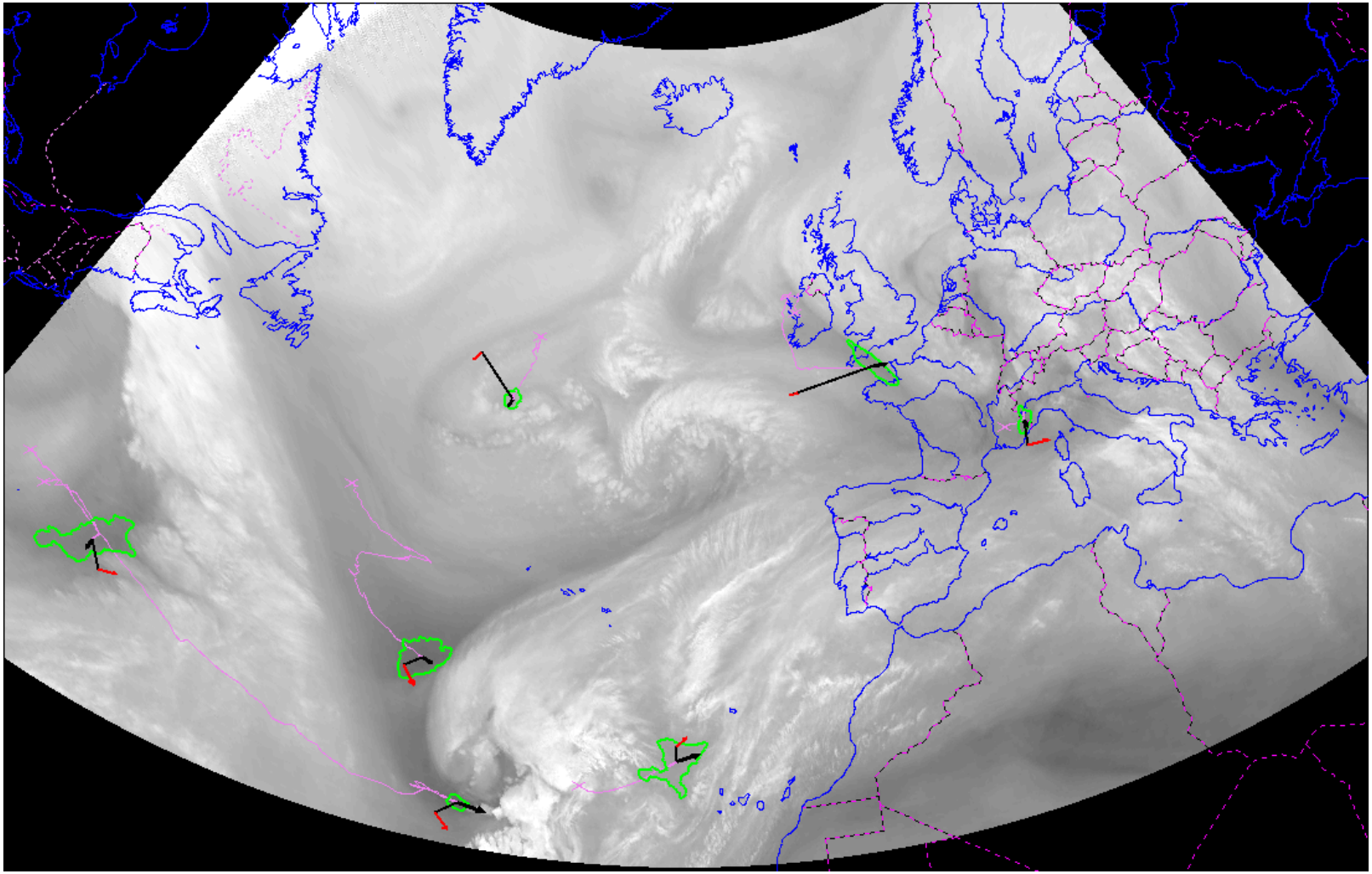


Figure: Antidote sur la tempête Xynthia (26-27 février 2010)

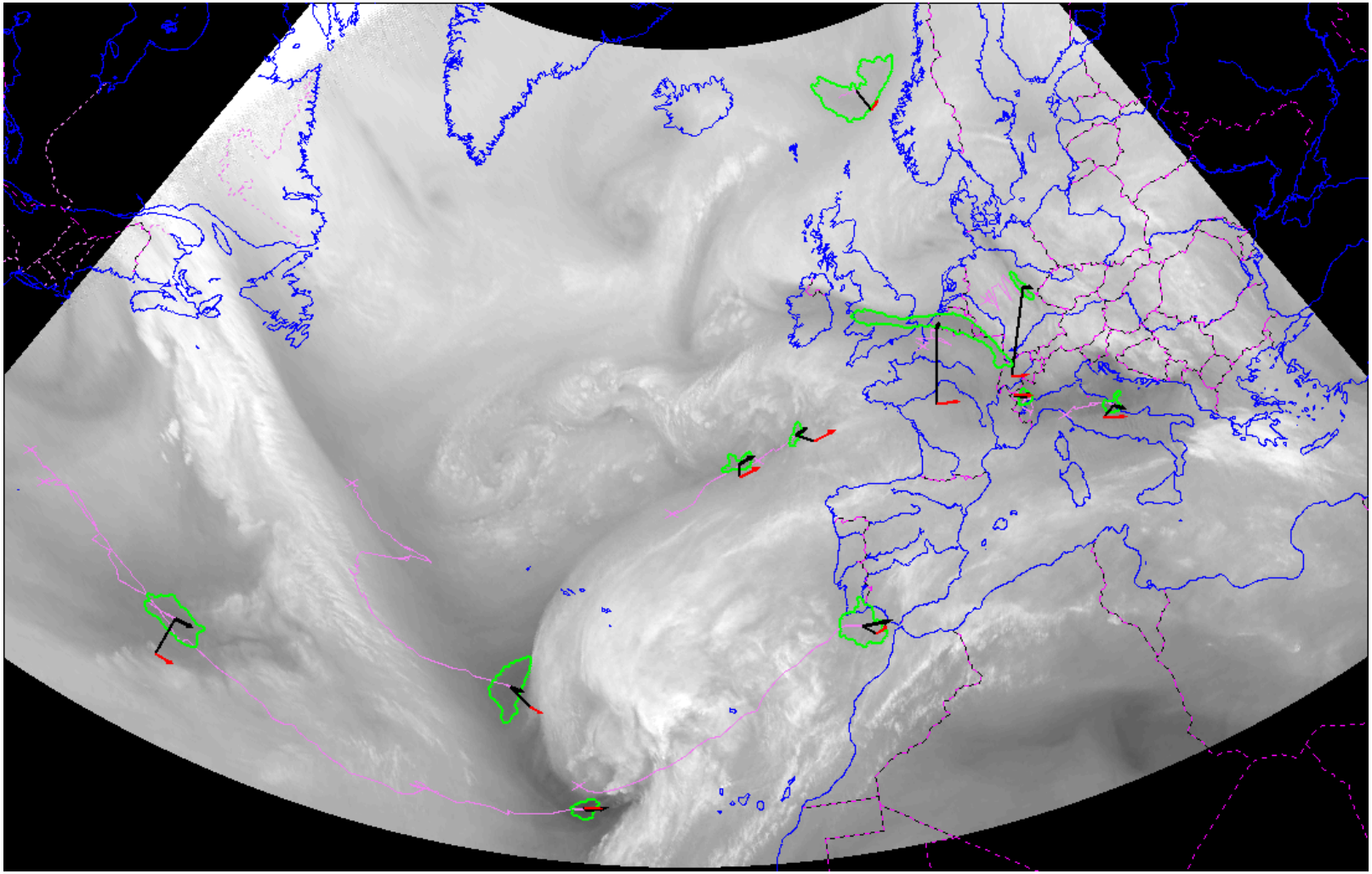


Figure: Antidote sur la tempête Xynthia (26-27 février 2010)

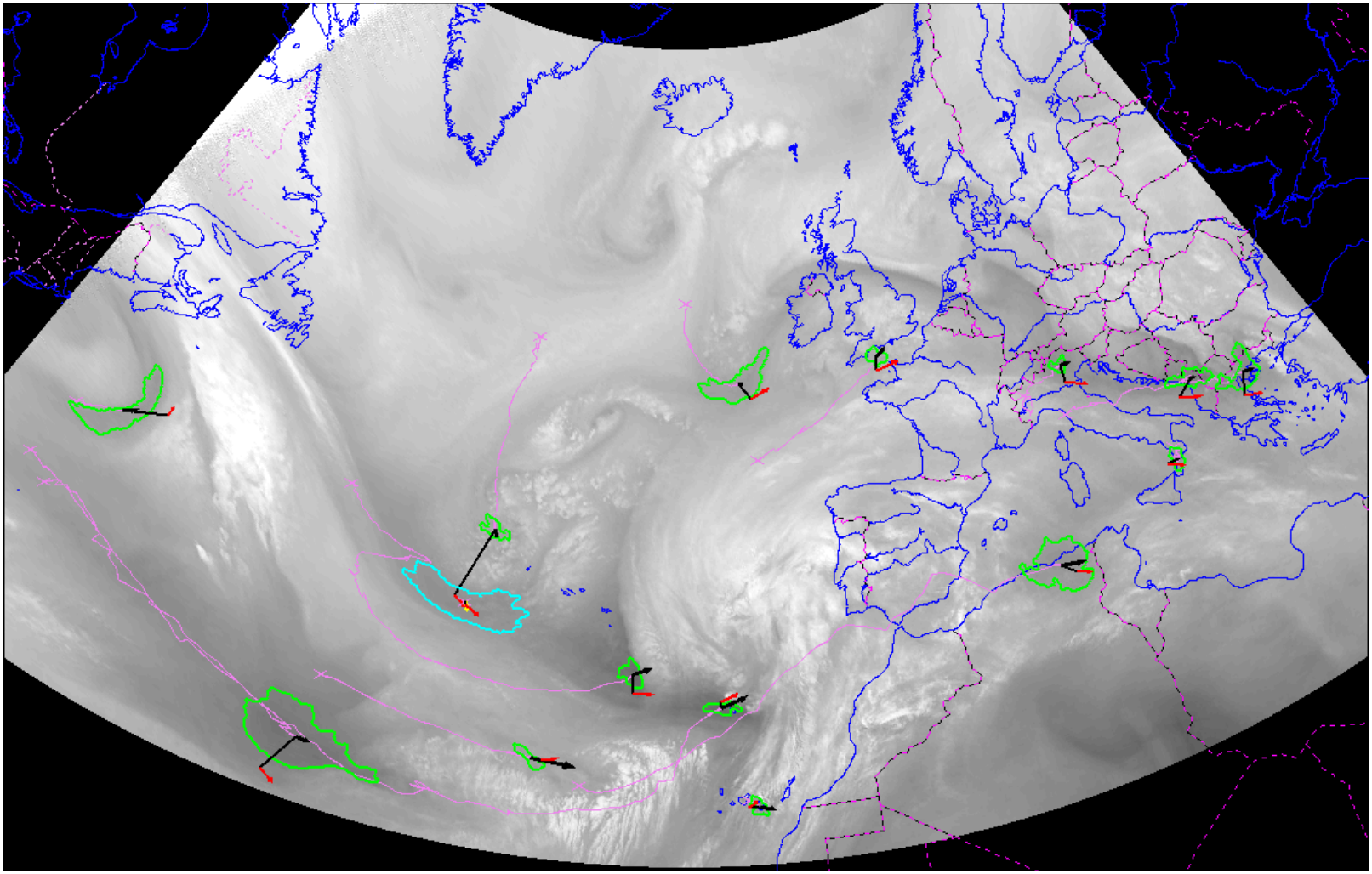


Figure: Antidote sur la tempête Xynthia (26-27 février 2010)

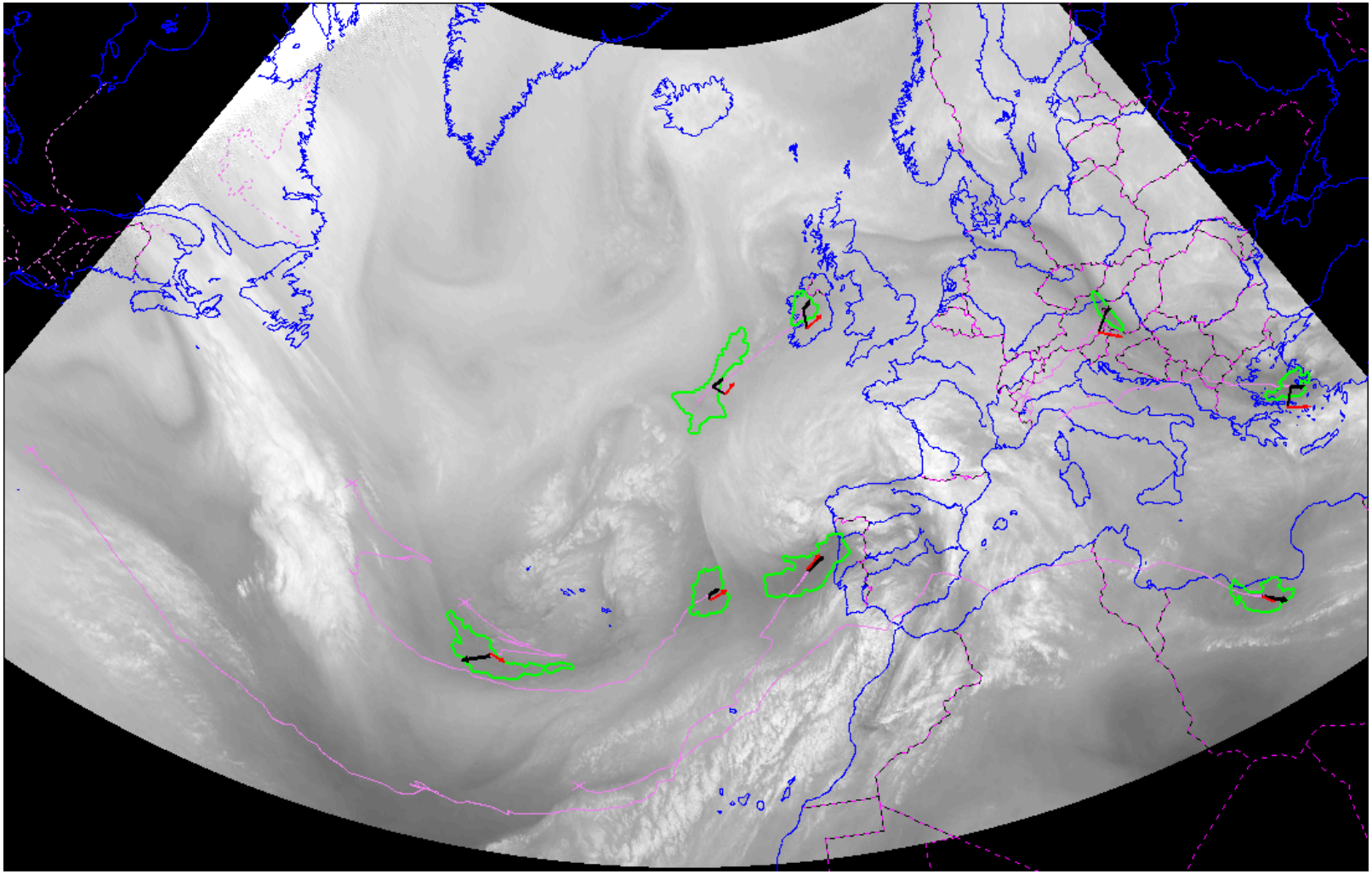


Figure: Antidote sur la tempête Xynthia (26-27 février 2010)

Assimilation de structures; Assimilation d'images

- Comment assimiler de l'information structurelle? (opérateur d'observation non-différentiable)
- Comment corriger des erreurs de positionnement spatio-temporel?

Modèle d'erreur en amplitude (classique)

$$\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}_a$$
$$\delta\mathbf{x}_a = BH^{\mathrm{T}}(HBH^{\mathrm{T}} + R)^{-1}(\mathbf{y} - H\mathbf{x}_b)$$

B est la matrice de covariance des erreurs d'ébauche.

Modèle d'erreur en déplacement

$$\mathbf{x}_a = \mathbf{x}_b\big(\mathbf{s} + \delta\mathbf{s}_a\big)$$
$$\delta\mathbf{s}_a = DH_d^{\mathrm{T}}(H_dDH_d^{\mathrm{T}} + R_d)^{-1}(\mathbf{y}_d - H_d\mathbf{x}_b)$$

D est la matrice de covariance des erreurs de positions dans l'ébauche.

1 Introduction

2 Techniques Mathématiques en Modélisation

3 Techniques Mathématiques en Assimilation de données

4 Perspectives

Perspectives: techniques mathématiques et prévision météo

Modélisation [conception de schémas numériques]

- Couplage (modèles régionaux, chimie-transport...)
- Adaptation aux nouvelles architectures matérielles
- Passage vers des méthodes en éléments finis?

Assimilation de données

- Modélisation et description des covariances d'erreur d'ébauche (matrice **B** inhomogène, hybridation avec EnKF, erreur d'échantillonnage de l'EnKF)
- Prise en compte de l'erreur modèle
- Propagation des covariances d'erreur d'ébauche, calcul des covariances d'erreur d'analyse (méthodes ensemblistes, prévision d'ensemble)
- Au-delà de l'hypothèse quasi-linéaire? (comment assimiler les données nuageuses à l'échelle convective)
- Au-delà de l'hypothèse Gaussienne? (filtres particulaires, morphing, assimilation de structures, assimilation d'images...)

Techniques Mathématiques en Assimilation de Données pour la Prévision Météorologique

Yann MICHEL

[1] Météo-France
Centre National de Recherches Météorologiques (CNRM-GAME)
Groupe de Modélisation pour l'Assimilation et la Prévision (GMAP)

Journée "Mathématiques pour l'Ingénierie : de la Modélisation à la Résolution de Problèmes Concrets"

31 mars 2010

1. Introduction
2. Techniques Mathématiques en Modélisation
3. Techniques Mathématiques en Assimilation de données
4. Perspectives