

Assimilation de données pour la prévision numérique du temps

Benjamin Ménétrier, IRIT, Toulouse, France
(benjamin.menetrier@irit.fr)
(gergaud@enseeiht.fr)



- 30 septembre 2019

Plan

- 1 Prédiction numérique du temps
- 2 Observations de l'atmosphère
- 3 Assimilation de données
- 4 Défis techniques
- 5 Résumé

Plan

- 1 Préviation numérique du temps
- 2 Observations de l'atmosphère
- 3 Assimilation de données
- 4 Défis techniques
- 5 Résumé

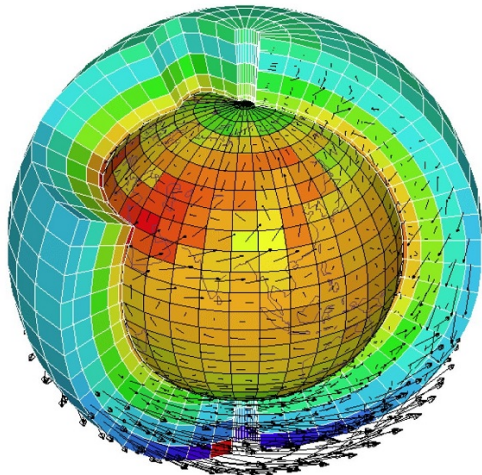
La prévision numérique du temps (PNT)

- En météorologie opérationnelle : la prévision numérique du temps fournit des résultats bruts à post-traiter et à expertiser.
- Comment ça marche ?

La prévision numérique du temps (PNT)

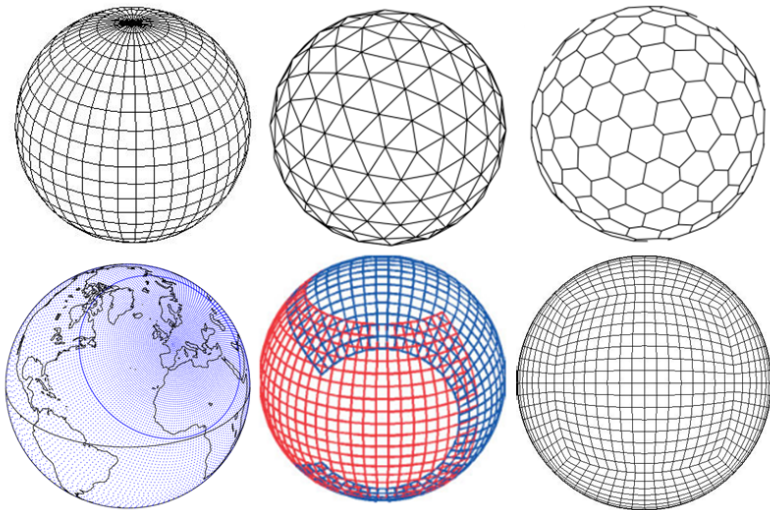
- En météorologie opérationnelle : la prévision numérique du temps fournit des résultats bruts à post-traiter et à expertiser.
- Comment ça marche ?

- Variables pronostiques (vent, température, humidité, pression, hydrométéores, etc...) définies à chaque maille de la grille.
- Grâce aux équations de la physique, le modèle propage les variables de l'instant t à l'instant $t + \Delta t$: **Équations aux Dérivées Partielles.**



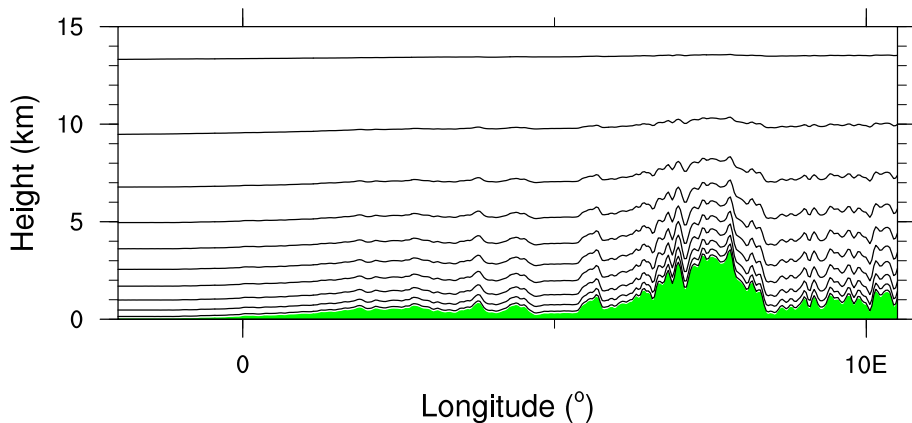
Discrétisation horizontale

Une grande variété de grilles horizontales...

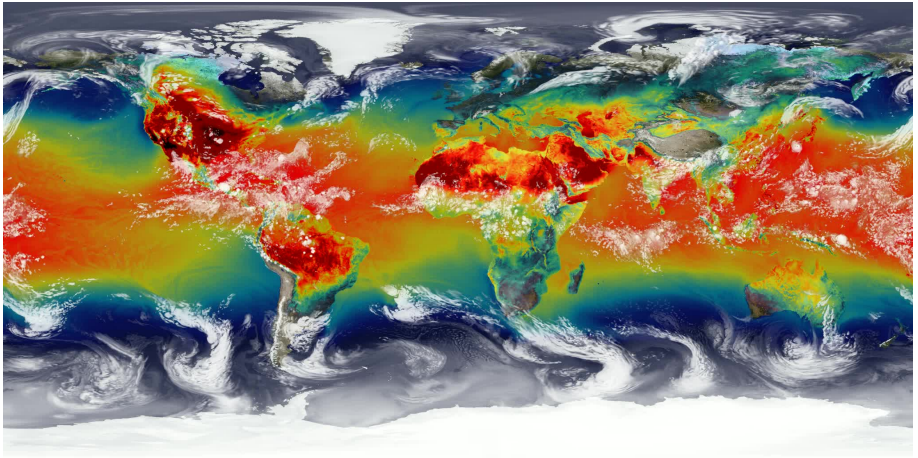


Discrétisation verticale

Coordonnée hybride : suit le terrain près du sol, puis se relaxe vers des niveaux pression.



Simulation GEOS-5 (NASA)



Température de surface (couleur)
et rayonnement IR au sommet de l'atmosphère (blanc)

Source : <https://svs.gsfc.nasa.gov>

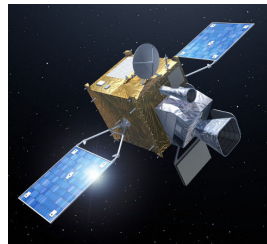
Plan

- 1 Prévion numérique du temps
- 2 Observations de l'atmosphère**
- 3 Assimilation de données
- 4 Défis techniques
- 5 Résumé

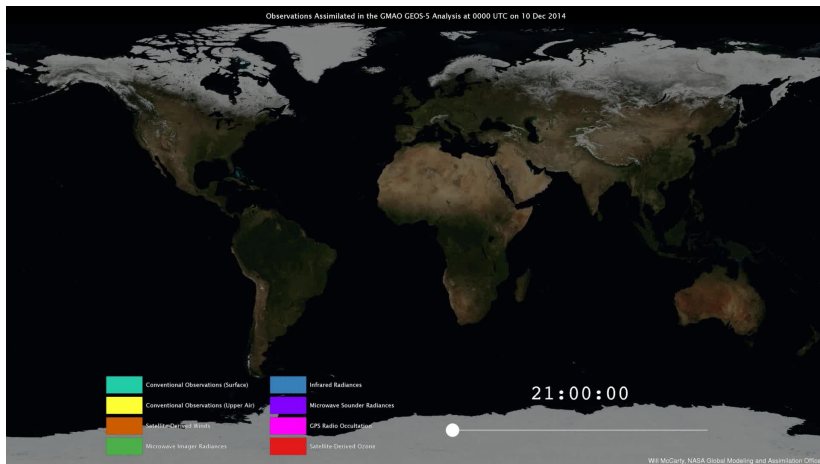
Observations disponibles



- Radiosondages
- Stations de surface
- Bouées
- Données d'avions
- Satellites



Diversité des observations



Données assimilées pendant une fenêtre de 6 h

Source : <https://svs.gsfc.nasa.gov>

Insuffisances des observations

- Répartition des observations très hétérogène
- Observations indirectes des variables pronostiques du modèle (ex. : un satellite ne mesure pas une température mais une luminance pour une certaine bande spectrale...)

Modèle :

- $\sim 10^6$ points de grille
- $\sim 10^2$ niveaux
- ~ 10 variables pronostiques

Total : 10^9 valeurs à initialiser

Observations (fenêtre de 6 h) :

- $\sim 3 \cdot 10^5$ obs. conventionnelles
- $\sim 5 \cdot 10^5$ obs. avions
- $\sim 9 \cdot 10^7$ obs. satellitaires

Total : $9 \cdot 10^7$ obs. disponibles
mais : $7 \cdot 10^6$ obs. utilisables
après contrôle de qualité / écrémage

Insuffisances des observations

- Répartition des observations très hétérogène
- Observations indirectes des variables pronostiques du modèle (ex. : un satellite ne mesure pas une température mais une luminance pour une certaine bande spectrale...)

Modèle :

- $\sim 10^6$ points de grille
- $\sim 10^2$ niveaux
- ~ 10 variables pronostiques

Total : 10^9 valeurs à initialiser

Observations (fenêtre de 6 h) :

- $\sim 3 \cdot 10^5$ obs. conventionnelles
- $\sim 5 \cdot 10^5$ obs. avions
- $\sim 9 \cdot 10^7$ obs. satellitaires

Total : $9 \cdot 10^7$ obs. disponibles
mais : $7 \cdot 10^6$ obs. utilisables
après contrôle de qualité / écrémage

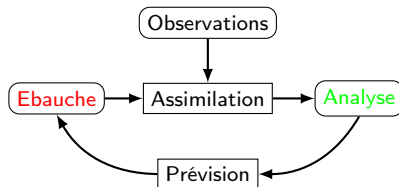
Problème d'initialisation largement sous-déterminé...

Plan

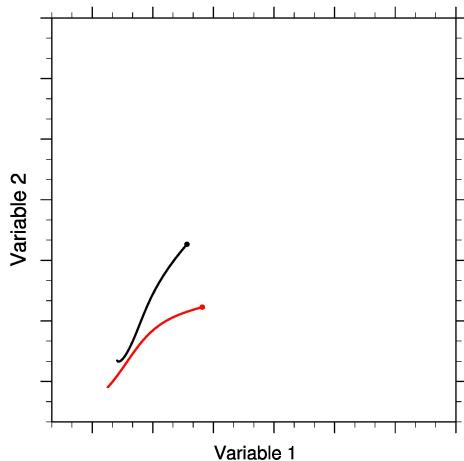
- 1 Prévion numérique du temps
- 2 Observations de l'atmosphère
- 3 Assimilation de données**
- 4 Défis techniques
- 5 Résumé

Assimilation de données

- Ensemble des techniques permettant une initialisation efficace des modèles numériques.
- Deux sources d'information principales :
 - des observations
 - une "ébauche" (= prévision récente à courte échéance)
- Les observations permettent de corriger l'ébauche pour se rapprocher de la vérité (sans jamais l'atteindre !).
- L'état corrigé (analyse) initialise la prévision suivante.

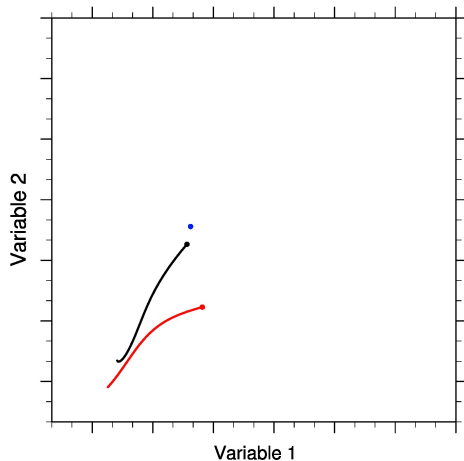


Divergence du modèle chaotique

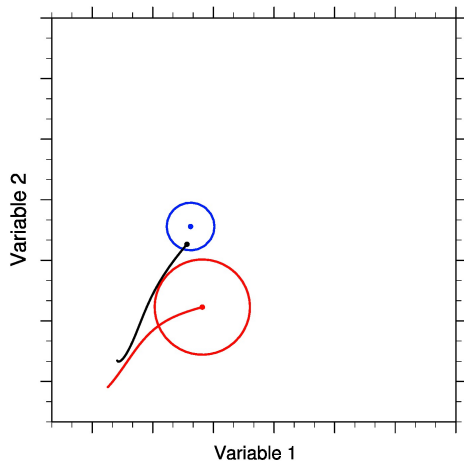


La prévision (en rouge) a une erreur d'initialisation, et diverge donc inéluctablement de la vérité (en noir).

Nouvelle observation

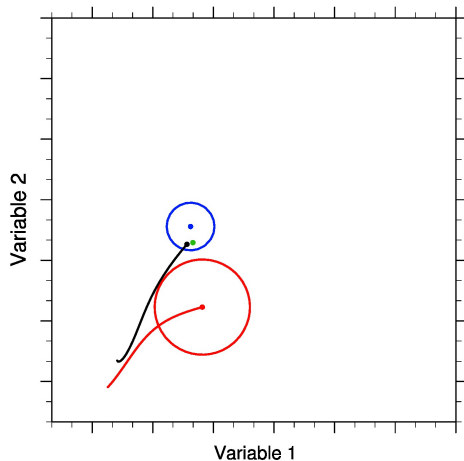


Une observation (en bleu) est disponible, mais elle n'est pas égale à la vérité (en noir).



Chaque source d'information a son incertitude, représentée par un cercle de dispersion :

- erreur d'ébauche (cercle rouge)
- erreur d'observations (cercle bleu)



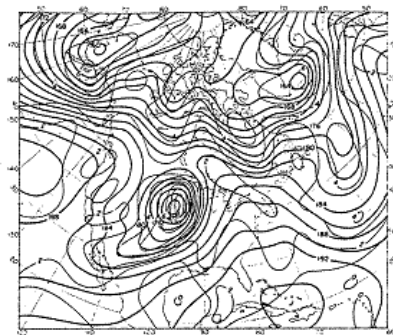
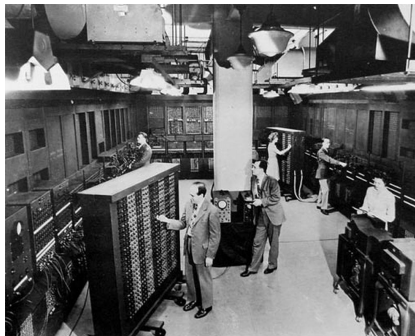
L'analyse est l'état le plus probable obtenu en tenant compte de l'incertitude des deux sources d'information :
Optimisation en grande dimension

Plan

- 1 Prévion numérique du temps
- 2 Observations de l'atmosphère
- 3 Assimilation de données
- 4 Défis techniques**
- 5 Résumé

Calcul intensif et prévision météorologique

Dès les débuts du calcul intensif, la prévision numérique du temps est une application critique :



Calculateur ENIAC et première prévision numérique en 1950.

- 270 points de grille (résolution 700 km)
- 24 h nécessaires pour calculer une prévision à 24 h d'échéance (incluant le temps de chargement des cartes perforées)

Calcul intensif et prévision météorologique

Premier ordinateur à Météo-France en 1992 : le Cray 2



Calculateurs Cray 1 (banquette incluse) et Cray 2

- Machine la plus rapide à sa sortie : 4 processeurs vectoriels
- En 2012, une tablette avait les mêmes performances pour le benchmark LINPACK.

Calcul intensif et prévision météorologique

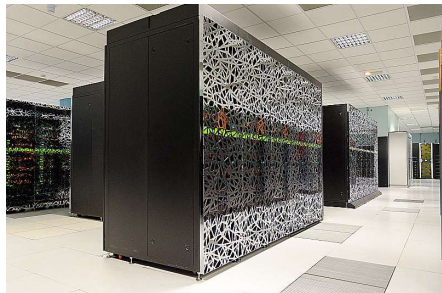
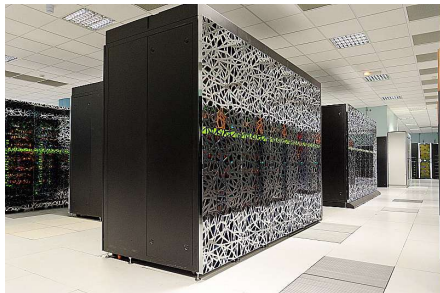
Depuis 2014 : calculateur Bull à Météo-France



Calculateur Bull DLC B710

Calcul intensif et prévision météorologique

Depuis 2014 : calculateur Bull à Météo-France



Calculateur Bull DLC B710 (x 2 : Météopole et ECA-Montaudran)

- En 2016 : 145 440 cœurs de calcul
- 111^{ème} et 113^{ème} places mondiales en juin 2019
- Serveur de stockage : 71,7 Po, pour 338 millions de fichiers.
- Renouvellement : x 5 en 2020.

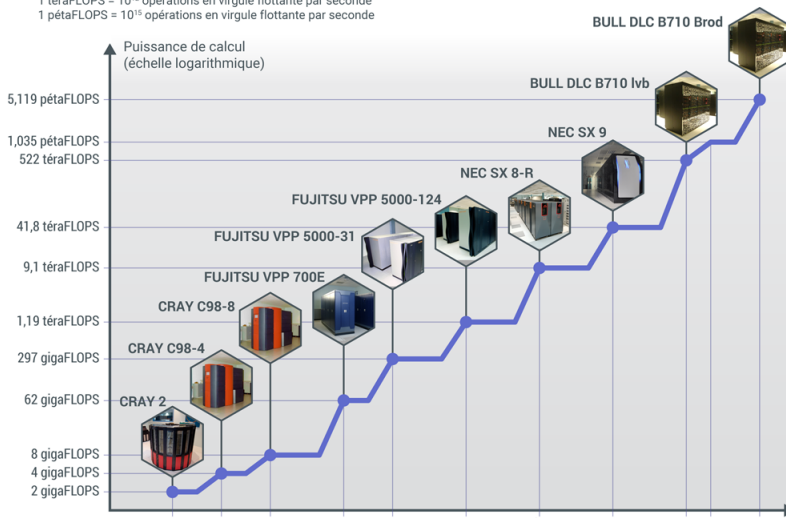
Calcul intensif et prévision météorologique

Progression depuis 1992 : $\times 2\,500\,000$

1 gigaFLOPS = 10^9 opérations en virgule flottante par seconde

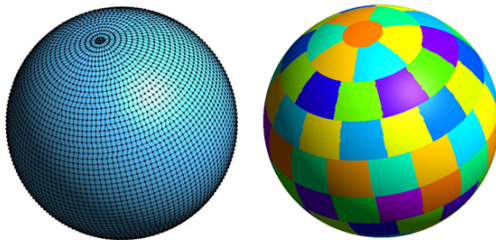
1 téraFLOPS = 10^{12} opérations en virgule flottante par seconde

1 pétaFLOPS = 10^{15} opérations en virgule flottante par seconde



Défis pour l'assimilation de données

- Exécution sur plusieurs cœurs : décomposition de la grille



- Travail partagé entre cœurs mais communications nécessaires.
- Deux problèmes fondamentaux :
 - **Déséquilibre de charge** : si certains cœurs ont plus de travail que les autres, ces derniers doivent les attendre.
 - **Temps de communication** : les communications entre cœurs prennent du temps... parfois plus que les calculs !
- L'algorithme d'assimilation doit être pensé **dès la théorie** avec ces contraintes en tête.

Plan

- 1 Préviation numérique du temps
- 2 Observations de l'atmosphère
- 3 Assimilation de données
- 4 Défis techniques
- 5 **Résumé**

- L'assimilation de données est un prérequis capital pour initialiser les modèles de prévision numérique du temps.

Résumé

- L'assimilation de données est un prérequis capital pour initialiser les modèles de prévision numérique du temps.
- Malgré le nombre important d'observations ($\sim 10^7$), le système est sous-déterminé ($\sim 10^9$ variables).

- L'assimilation de données est un prérequis capital pour initialiser les modèles de prévision numérique du temps.
- Malgré le nombre important d'observations ($\sim 10^7$), le système est sous-déterminé ($\sim 10^9$ variables).
- La correction de l'ébauche par les observations prend en compte les incertitudes de ces deux sources d'information.

- L'assimilation de données est un prérequis capital pour initialiser les modèles de prévision numérique du temps.
- Malgré le nombre important d'observations ($\sim 10^7$), le système est sous-déterminé ($\sim 10^9$ variables).
- La correction de l'ébauche par les observations prend en compte les incertitudes de ces deux sources d'information.
- De nombreux défis théoriques restent d'actualité :
 - Covariances d'erreur d'ébauche
 - Et aussi : traitement des non-linéarités, systèmes couplés, covariances d'erreur d'observation, calibration des incertitudes...

- L'assimilation de données est un prérequis capital pour initialiser les modèles de prévision numérique du temps.
- Malgré le nombre important d'observations ($\sim 10^7$), le système est sous-déterminé ($\sim 10^9$ variables).
- La correction de l'ébauche par les observations prend en compte les incertitudes de ces deux sources d'information.
- De nombreux défis théoriques restent d'actualité :
 - Covariances d'erreur d'ébauche
 - Et aussi : traitement des non-linéarités, systèmes couplés, covariances d'erreur d'observation, calibration des incertitudes...
- Les défis techniques sont également importants :
 - Exécution sur des calculateurs massivement parallèles
 - Rapidité des entrées/sorties
 - Temps d'exécution imposé par les contraintes opérationnelles

- Équation aux Dérivées Partiels (Partial Differential Equation)
- Algèbre linéaire numérique
- Calcul différentiel
- Optimisation
- Probabilités, statistiques
- Calcul Haute Performance (High Performance Computing)

- 10 cours/TD
- 3 TP : EDP et algèbre linéaire (rendu noté)
- 3 TP : Optimisation et algèbre linéaire (rendu noté)