

# Simulateur d'apprentissage profond pour les modèles dynamiques : application visant à augmenter la taille des ensembles

Vikram Khade et Alain Beaulne

Conférence sur les données du gouvernement du Canada  
Tirer parti des données pour faire progresser l'innovation

*22 et 23 février 2023*



Environment and  
Climate Change Canada

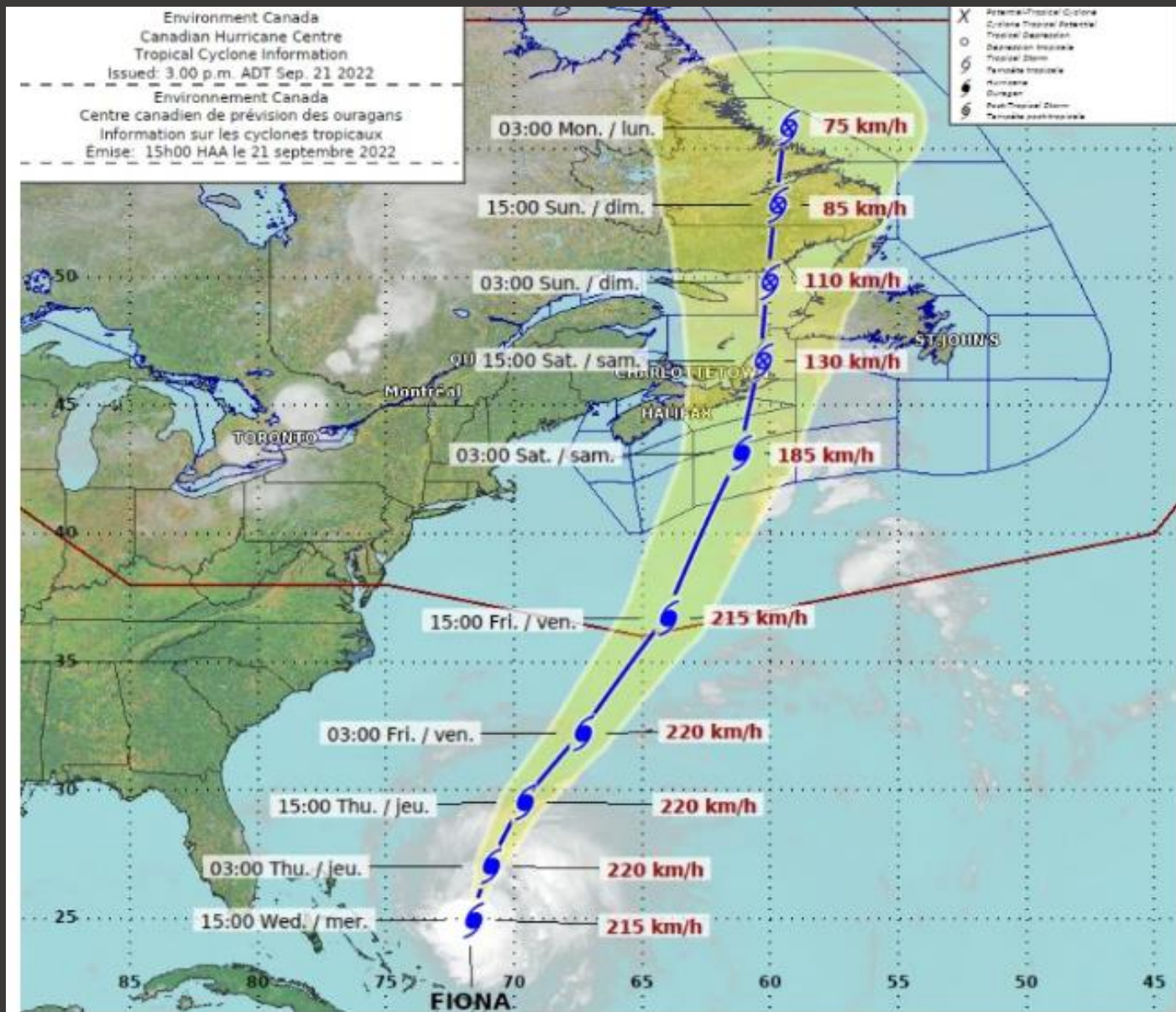
Environnement et  
Changement climatique Canada

# Aperçu

Dans cette présentation vous allez comprendre :

- pourquoi les prévisions atmosphériques/océaniques sont importantes;
- ce qu'est la prévision numérique du temps (PNT);
- le rôle des ressources informatiques utilisées par le Service météorologique du Canada (SMC), ECCO, dans la production de prévisions basées sur la PNT;
- que les prévisions sont intrinsèquement incertaines et que l'approche d'ensemble peut quantifier cette incertitude;
- certaines applications des ensembles, y compris la discipline mathématique de l'assimilation des données (en utilisant la technique du filtre de Kalman d'ensemble (EnKF));
- le rôle de la taille de l'ensemble dans un EnKF;
- les travaux proposés au SMC pour l'utilisation d'un simulateur basé sur l'apprentissage profond afin d'imiter la PNT en vue d'augmenter efficacement la taille d'un ensemble;
- une validation de principe de ces travaux proposés en utilisant un modèle très simplifié de l'atmosphère connu sous le nom de Lorenz-96.

# Prévision de la trajectoire de l'ouragan Fiona au moyen de la prévision numérique du temps (PNT)



Cette image montre le préavis de 5 jours pour l'ouragan Fiona émis par ECCC le 21 septembre 2022. La prévision est initialisée à 15 h.

La zone ombrée en vert montre l'enveloppe des trajectoires possibles d'ouragan telle que fournie par la PNT qui simule plusieurs **scénarios différents** (connus sous le nom d'**ensemble**).

L'ouragan Fiona s'est avéré être l'ouragan le plus intense à frapper le Canada. Il a frappé la Nouvelle-Écosse avec des vents de 169 km/h le samedi 24 septembre.

**Les organismes de secours gouvernementaux ont pu agir efficacement grâce à l'estimation précise de l'incertitude des prévisions réalisée à l'aide de la PNT d'ensemble**

### Crédits :

<https://globalnews.ca/news/9144648/hurricane-fiona-potentially-severe-event-atlantic-canada/>

Centre canadien de prévision des ouragans, SMC, ECCC  
[https://en.wikipedia.org/wiki/Hurricane\\_Fiona](https://en.wikipedia.org/wiki/Hurricane_Fiona)

# Qu'est-ce que la prévision numérique du temps (PNT)?

Des équations mathématiques, appelées équations différentielles, régissent le comportement dynamique de l'atmosphère (et des océans). L'évolution future de l'atmosphère est calculée en résolvant ces équations à l'aide de techniques numériques sur un ordinateur. Ces techniques, connues sous le nom de PNT, discrétisent le domaine mondial en utilisant une grille et les équations sont résolues sur cette grille. La forme discrétisée de ces équations est appelée modèle. Le GEM est le modèle utilisé par le SMC pour les prévisions météorologiques.

La taille de chaque cellule de la grille dépend de la résolution du modèle. Par exemple, une version du modèle GEM, utilisée par ECCO pour la modélisation d'ensemble mondiale, a une résolution de 39 km. La résolution étant finie, tout modèle n'est qu'une approximation des équations qui régissent l'atmosphère.

Le modèle ne connaît pas les processus physiques à l'intérieur d'une cellule donnée. Ces processus doivent être représentés à l'aide de *paramétrisations*. Il existe plusieurs paramétrisations possibles pour un même processus. Chacune d'entre elles donne lieu à une configuration légèrement différente du modèle. Par conséquent, des configurations différentes donnent lieu à des prévisions différentes.

Le délai de prévision peut aller de quelques heures à quelques jours selon l'application. Il est important que la prévision s'exécute dans un délai raisonnable pour être utile. Par exemple, si la prévision à 7 jours s'exécute en 7 jours sur l'ordinateur, cette prévision est essentiellement inutile.

# Qu'est-ce qu'un ensemble?

Un ensemble est composé de plusieurs membres (aussi connus sous le nom d'échantillons prélevés au moyen de la méthode de Monte-Carlo).

Chaque membre est une prévision différente basée sur un scénario unique. Chaque scénario est différent en ce qui concerne un ou plusieurs des éléments suivants :

(1) **Configuration du modèle** (p. ex., paramétrisations)

(2) **Conditions initiales** (comme la distribution globale de la température au moment initial de la prévision)

(3) **Conditions limites** (comme la température à la surface de la mer au cours de la période visée par les prévisions)

Les différents scénarios sont rendus nécessaires par le fait que les conditions initiales ne sont pas parfaitement connues. De même, les conditions limites sont également incertaines. Les membres de l'ensemble sont construits pour être cohérents avec l'incertitude des conditions initiales, des conditions limites et de la configuration du modèle.

**L'estimation de l'incertitude de la prévision s'améliore à mesure que la taille de l'ensemble (c'est-à-dire le nombre de scénarios) simulé augmente. Malgré l'utilisation d'ordinateurs très puissants, la taille des ensembles est relativement faible (quelques centaines de scénarios), car chaque simulation est gourmande en ressources informatiques.**



# Calcul de haute performance à ECCC

ECCC utilise de puissants superordinateurs pour réaliser la PNT. Une photo de ces superordinateurs en 2017 est présentée à droite. Ceux-ci ont été mis à jour à deux reprises depuis 2017.

Les superordinateurs actuels nommés Underhill et Robert étaient les 69<sup>e</sup> et 70<sup>e</sup> ordinateurs les plus puissants au monde en juin 2022!

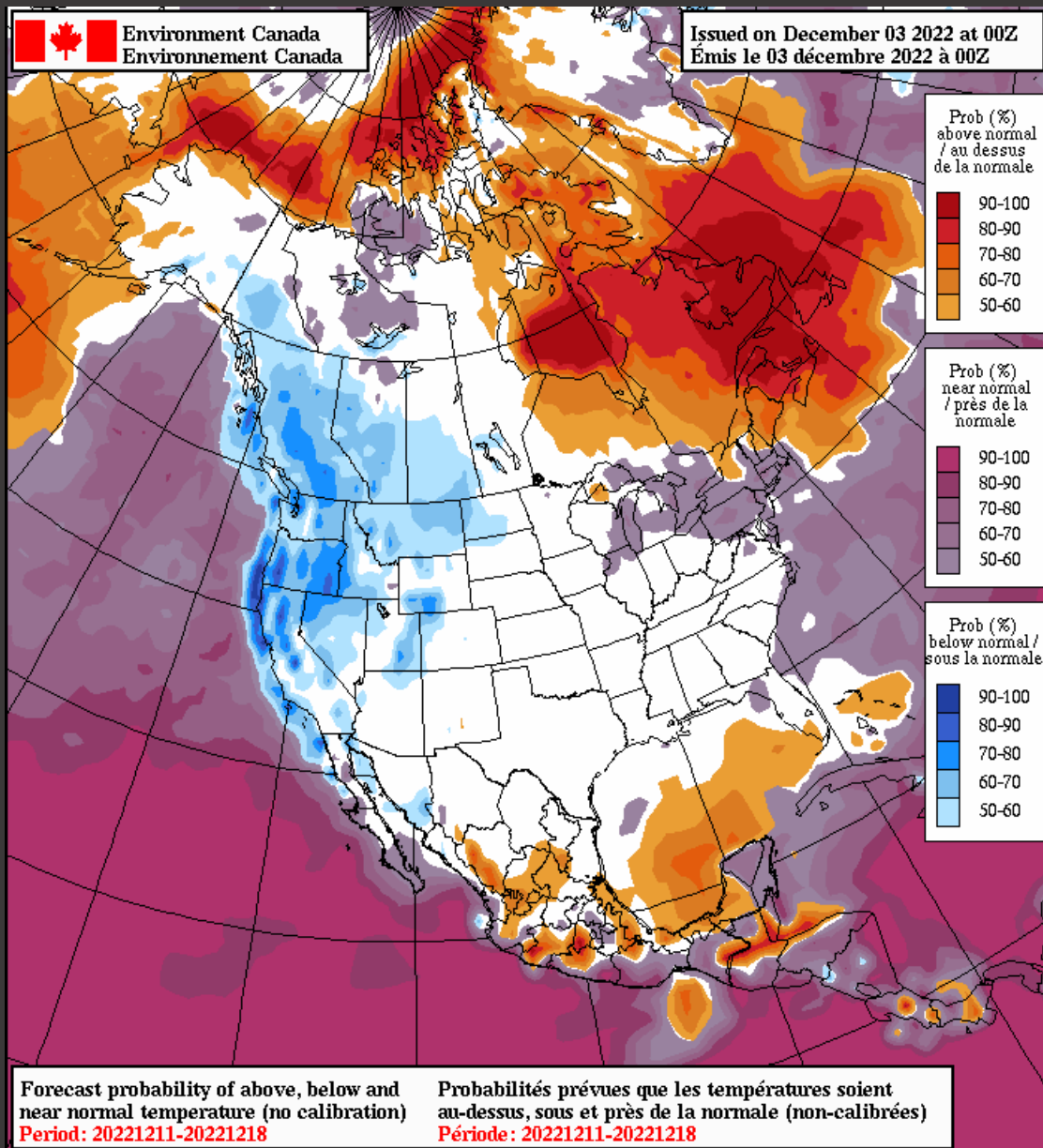


Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
69	<b>Underhill</b> - ThinkSystem SD650 V2, Xeon Platinum 8380 40C 2.3GHz, Infiniband HDR, Lenovo Shared Services Canada Canada	148,320	7.76	10.92	1,295
70	<b>Robert</b> - ThinkSystem SD650-N V2, Xeon Platinum 8380 40C 2.3GHz, Infiniband HDR, Lenovo Shared Services Canada Canada	148,320	7.76	10.92	1,295

## Référence :

<https://top500.org/lists/top500/list/2022/06/>  
<https://www.canada.ca/fr/services-partages/organisation/regroupement-centres-donnees/calcul-haute-performance.html>

# Un autre exemple d'ensemble



Les ensembles peuvent être utilisés pour calculer des probabilités. Plus le nombre de membres de l'ensemble est élevé, meilleure est l'estimation de la probabilité.

Le Système de prévision d'ensemble nord-américain (SPENA) combine les ensembles produits par le Service météorologique du Canada (SMC) et le National Weather Service (NWS) des États-Unis avec une taille combinée de 40 membres. La combinaison d'ensembles produits par différents organismes est une façon d'augmenter la taille de l'ensemble.

Cette figure montre les probabilités que les températures soient au-dessus, près ou en dessous de la normale pour la semaine du 11 au 18 décembre 2022.

Référence :  
[https://weather.gc.ca/ensemble/naefs/index\\_e.html](https://weather.gc.ca/ensemble/naefs/index_e.html)

# Coût de calcul associé à l'exécution de prévisions

À ECCC, une prévision d'ensemble à deux semaines d'avance, composée de 21 membres, est exécutée au moyen de  $800 \times 21$  CPU.

Le temps d'exécution de cette prévision est d'environ 1 heure. Les différents membres de l'ensemble sont exécutés en parallèle et donc la prévision d'ensemble se termine en ~ 1 heure.

L'augmentation de la taille de l'ensemble exige des ressources informatiques excessives puisque chaque membre supplémentaire de l'ensemble nécessite 800 CPU.

L'apprentissage profond (AP) est un outil réalisable sur le plan informatique pour augmenter la taille de l'ensemble.

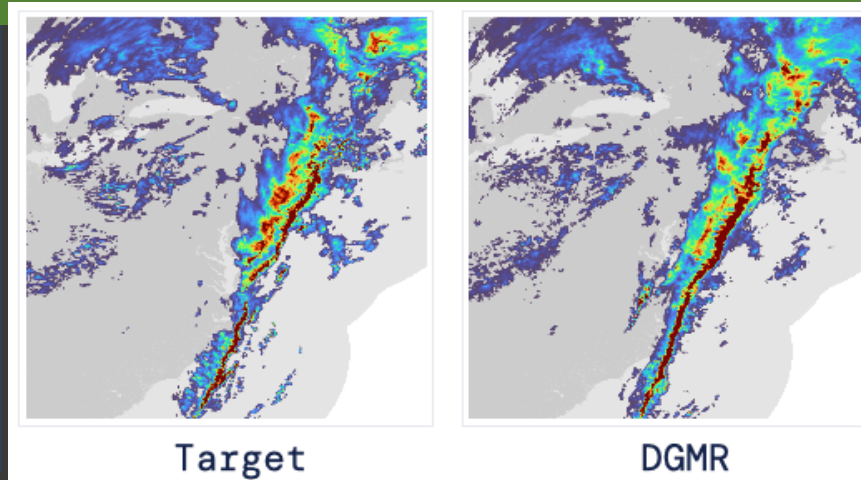
## Remerciements :

M. Xingxiu Deng, Ph. D.  
ECCC



# Apprentissage profond – une autre technique pour augmenter la taille d'un ensemble

DGMR est un simulateur basé sur l'AP élaboré par DeepMind pour prévoir les précipitations de pluie dans les 90 prochaines minutes. La figure montre la comparaison entre la prévision (DGMR) et les observations radar réelles (Target). La plupart des 56 prévisionnistes en météorologie ont estimé que DGMR était meilleur que les modèles de prévision numérique du temps existants!



Crédit :  
<https://www.technologyreview.com/2021/09/29/1036331/deepminds-ai-predicts-almost-exactly-when-and-where-its-going-to-rain/>

- L'apprentissage profond (AP) est devenu un outil puissant dans le domaine de l'intelligence artificielle pour résoudre plusieurs problèmes comme la reconnaissance d'images, la classification de textes, l'amélioration de vidéos, etc.
- Les données existantes sont utilisées pour entraîner les réseaux d'apprentissage profond à effectuer une tâche particulière. Le modèle entraîné est sauvegardé, puis utilisé sur des données non visibles. Le processus de formation est coûteux sur le plan des calculs informatiques, mais une fois le modèle entraîné, l'utilisation du modèle d'AP sur des données non visibles est très rapide.
- Un autre domaine d'application de l'AP est celui de l'entraînement de simulateurs afin qu'ils imitent le modèle de PNT. Dans cette approche, les archives (conditions initiales, données de prévision) du modèle de PNT sont utilisées pour former le simulateur d'AP.

# Weyn et coll. : Démonstration d'un apprentissage profond des conditions météo

JAMES | Modeling Earth Systems





## RESEARCH ARTICLE

10.1029/2020MS002109

### Key Points:

- A convolutional neural net (CNN) is developed for global weather forecasts on the cubed sphere
- Our CNN produces skillful global forecasts of key atmospheric variables at lead times up to 7 days
- Our CNN computes stable 1-year simulations of realistic atmospheric

## Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere

Jonathan A. Weyn<sup>1</sup> , Dale R. Durran<sup>1</sup> , and Rich Caruana<sup>2</sup>

<sup>1</sup>Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, <sup>2</sup>Microsoft Research, Redmond, WA, USA

Dans cet article, Weyn et coll. utilisent des réseaux neuronaux convolutifs (CNN) pour simuler un modèle de PNT pour deux semaines. Les CNN sont un type spécial de réseaux d'AP qui sont bien adaptés aux données ayant une structure spatiale.

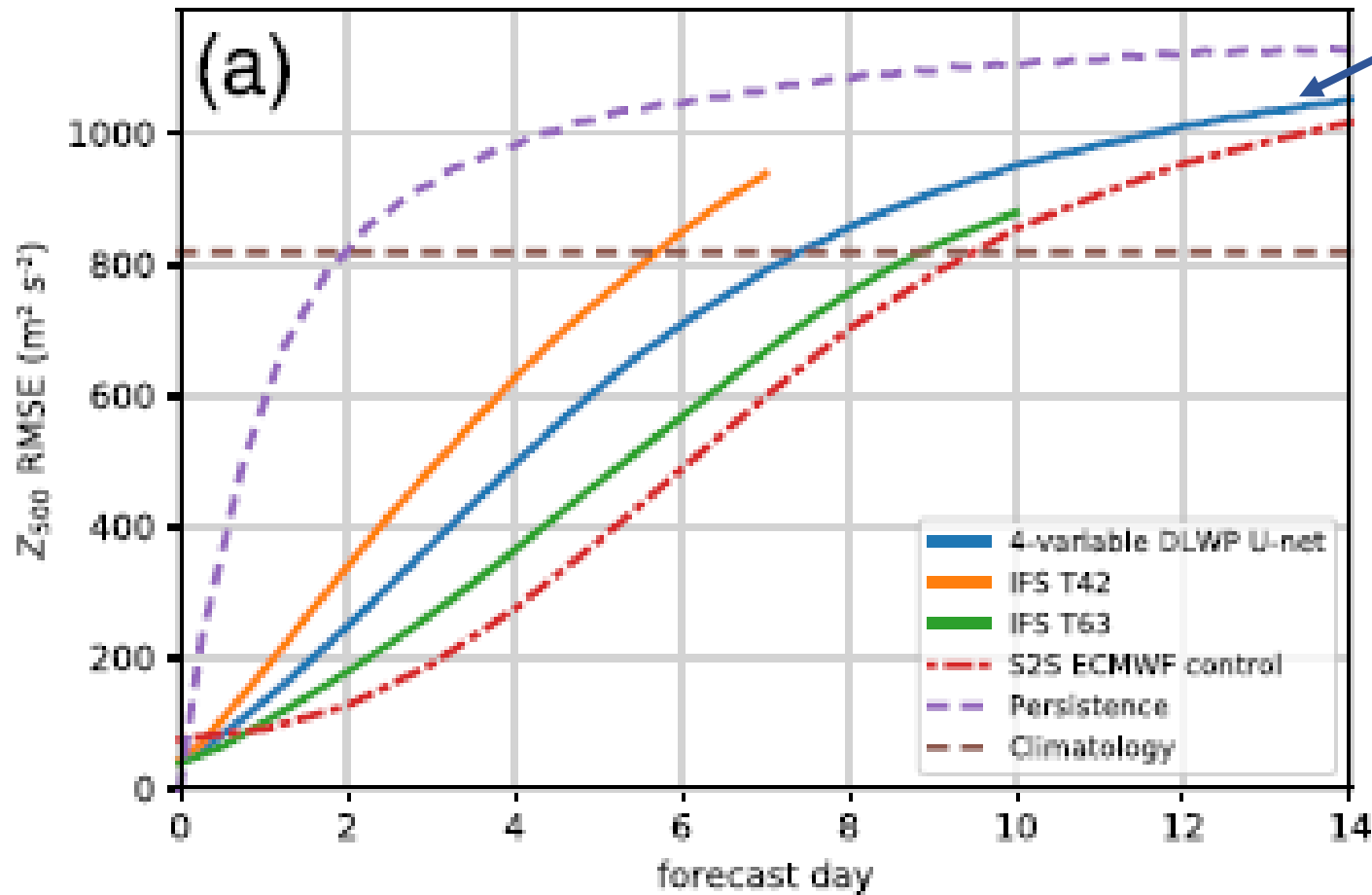
Données d'entraînement utilisées : ERA5, remaillées à 2 degrés (1979–2018)

Intrants :

- Hauteur géopotentielle à 500 hPa.
- Hauteur géopotentielle à 1000 hPa.
- Épaisseur géopotentielle à 300–700 hPa.
- Température à 2 m.
- Insolation du sommet de l'atmosphère.
- Masque terre-mer.
- Hauteur topographique.

Produit les champs mentionnés dans les **Intrants**.

# Weyn et coll. : Le modèle d'apprentissage profond surpasse le modèle T42



Le simulateur d'AP est plus performant que l'IFS T42 et Persistence.

IFS T42 : modèle CEPMMT IFS, 62 niveaux, 2,8 deg.

IFS T63 : modèle CEPMMT IFS, 137 niveaux, 1,9 deg.

S2S : modèle CEPMMT IFS infrasaisonnier à saisonnier  
16 km – 31 km.  
Entièrement couplé aux modèles océaniques et de glace de mer.

Le CEPMMT (Centre européen pour les prévisions météorologiques à moyen terme) est une organisation soutenue par la plupart des nations d'Europe. L'IFS (Integrated Forecasting System) est le modèle opérationnel du CEPMMT.

# Weyn et coll. : vitesse de calcul

Although our DLWP model lags the performance of a high-resolution operational NWP model by about 2–3 days of forecast lead time relative to climatology, it does have one significant advantage: computational speed. After a one-time computational cost of 2–3 days for training on a single NVidia Tesla V100 GPU, our DLWP model can produce a global 4-week forecast in less than two tenths of a second. At this speed, one could generate a 1,000-member ensemble of 1-month forecasts in about 3 min. In contrast, the full dynamical IFS model at approximately equivalent T63 horizontal resolution, run albeit somewhat inefficiently on a 36-core computing node, requires nearly 24 min to produce a single 4-week forecast, or about 16 days for the same 1,000-member ensemble forecast. Operationally, ECMWF, despite vast supercomputing resources,

## Processeur graphique unique Nvidia Tesla V100

*Entraînement* : 2 à 3 jours

*Prévision* : 1000 membres de l'ensemble, prévisions d'un mois en 3 minutes!

**Cela démontre que l'AP peut être utilisé de manière très efficace pour augmenter la taille de l'ensemble.**

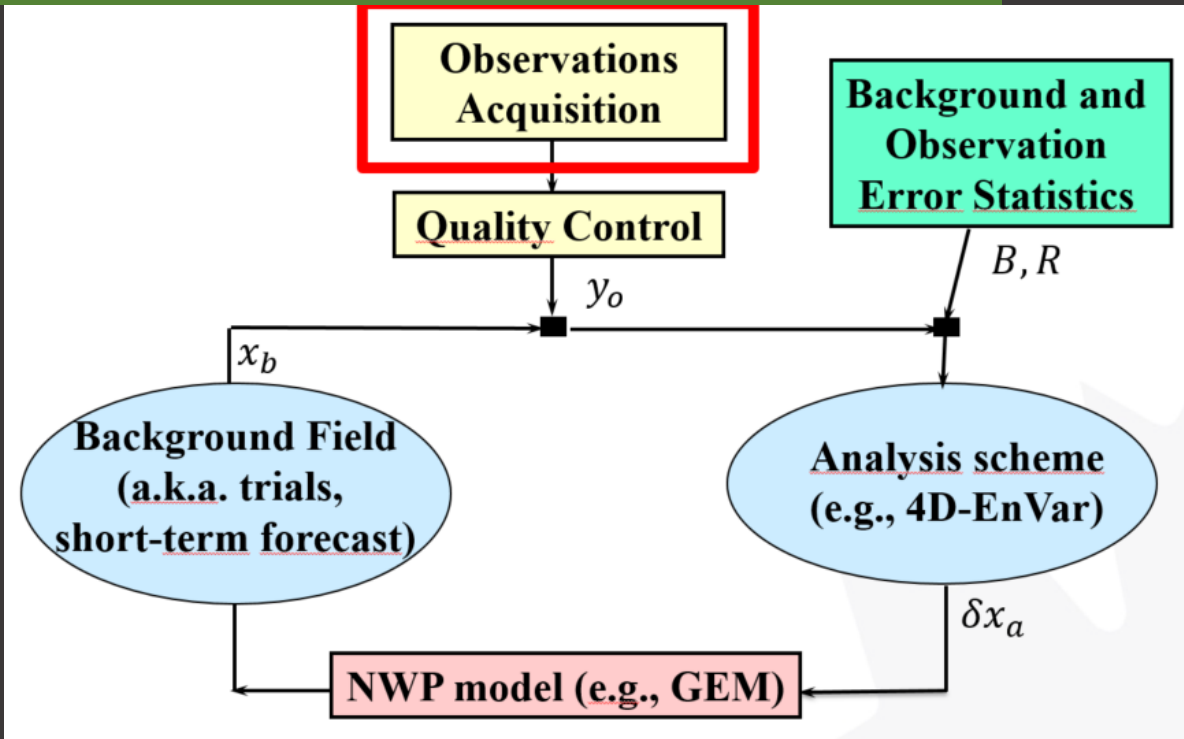
# Récapitulatif

Jusqu'à présent, nous avons appris que :

- Les ensembles de prévisions sont importants pour quantifier l'incertitude dans la prévision.
- Chaque membre de l'ensemble simule un scénario différent. Idéalement, nous aimerions simuler tous les scénarios possibles.
- Cependant, l'augmentation du nombre de scénarios est coûteuse en termes de ressources informatiques.
- Les simulateurs basés sur l'apprentissage profond (AP) sont apparus comme un outil permettant d'augmenter la taille de l'ensemble à un faible coût en termes de ressources informatiques.
- Dans les prochaines diapositives, nous allons découvrir l'importance de la taille de l'ensemble dans un filtre de Kalman d'ensemble (EnKF).
- L'EnKF est une technique d'assimilation de données qui combine l'estimation des conditions météorologiques actuelles fournies par le modèle de PNT avec les observations atmosphériques pour produire la meilleure image possible des conditions météorologiques actuelles.



# Assimilation de données dans la PNT



L'assimilation de données (AD) est une technique utilisée pour combiner différentes sources d'information avec leurs incertitudes afin d'obtenir la meilleure estimation possible des conditions météorologiques actuelles. Cette estimation est connue sous le nom d'analyse.

Les différentes sources d'information sont **l'estimation du modèle** de PNT (champ de fond) et les **observations** de l'atmosphère effectuées par divers instruments. L'incertitude de l'estimation du modèle est connue sous le nom de statistique d'erreur de fond. L'incertitude des observations est connue sous le nom de statistique d'erreur d'observation. Une connaissance précise de ces statistiques est très importante pour obtenir une analyse de haute qualité.

Une étape d'AD est effectuée toutes les 6 heures pour obtenir une analyse à partir de laquelle une prévision à court terme de 6 heures est émise, qui est ensuite utilisée comme estimation du modèle pour l'étape d'AD suivante. Une variété d'observations sont utilisées dans l'AD.

# Principaux réseaux d'observation utilisés dans les systèmes de PNT



Satellites géostationnaires



Satellites en orbite polaire



Micro-satellites de radio-occultation du GNSS



Aéronefs



Stations d'observation aérologiques



Stations de surface

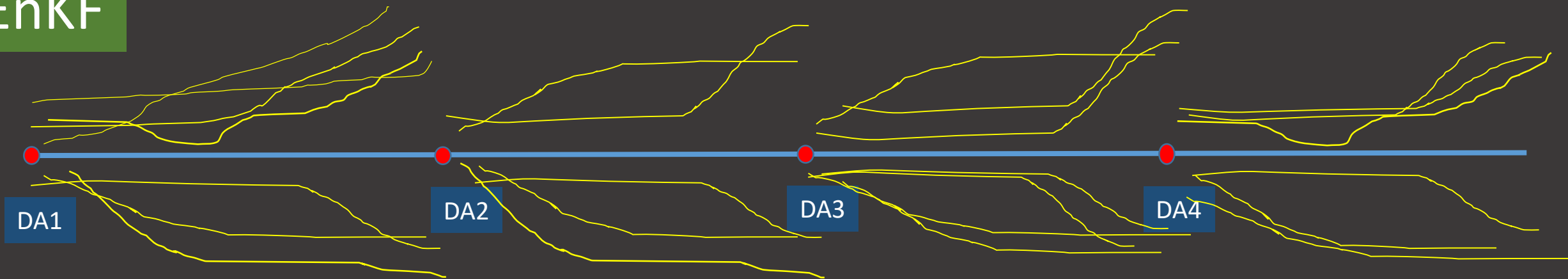


GNSS au sol



Bouées, dériveurs et navires

# EnKF



- L'EnKF (filtre de Kalman d'ensemble) est une technique d'assimilation de données qui utilise des échantillons de prévisions d'ensemble pour estimer l'incertitude de fond (covariance). La figure montre un dessin animé de l'EnKF. Les cycles d'AD se produisent toutes les 6 heures. Le point rouge montre l'analyse moyenne. Les lignes jaunes représentent les membres de l'ensemble qui sont déterminés en résolvant les équations de PNT sur l'ordinateur.
- Plus la taille de l'ensemble est grande, meilleure est l'estimation des statistiques d'erreur de fond.
- Comme déjà expliqué, il est très coûteux d'augmenter la taille de l'ensemble.
- Dans les diapositives suivantes, l'application d'un simulateur d'AP pour augmenter la taille de l'ensemble dans EnKF est démontrée en utilisant le modèle Lorenz-96
- Lorenz-96 (L96) est un modèle simplifié de l'atmosphère.
- Un simulateur d'AP est entraîné pour générer la prévision de L96. L'entraînement est effectué en utilisant les données (conditions initiales, prévision) générées par l'intégration numérique des équations de L96.

# Modèle de Lorenz 1996 à 40 dimensions (L96)

$$dx_i/dt = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F$$

Les  $x_i$  sont les composantes qui sont des variables météorologiques autour d'un cercle de latitude. Sur la droite, le premier terme est l'advection et le second terme est la diffusion.

$i = 1..40$  (Dimensionnalité =  $d = 40$ )

$F = 8$  (Forçage)

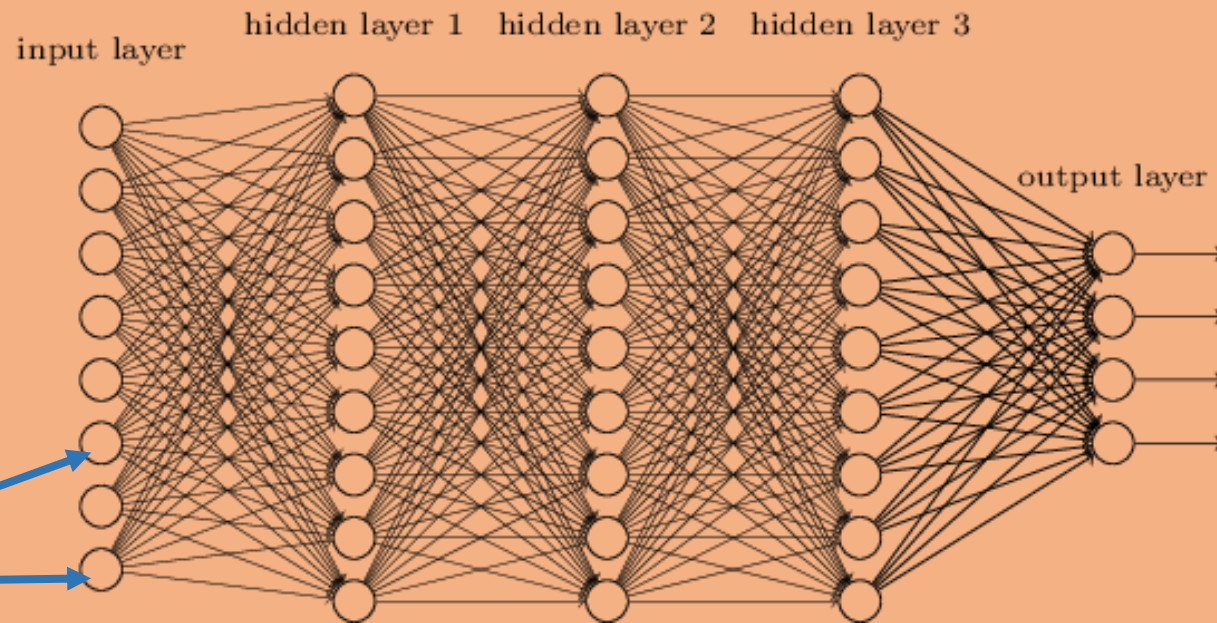
$x_0 = x_N$

0,05 unité de temps ~ 6 heures dans le modèle atmosphérique.

Ce modèle a été utilisé par de nombreux chercheurs comme banc d'essai pour tester de nouveaux schémas d'assimilation de données, des améliorations des techniques de DA, etc.

# Comment fonctionne l'apprentissage profond (AP)?

- Chaque connexion est associée à un poids.
- Les poids sont initialisés avec des valeurs aléatoires.
- Les prédicteurs des échantillons sont entrés et l'AP calcule la sortie.
- L'erreur entre cette sortie et la cible de l'échantillon est calculée.
- Cette erreur est propagée en retour pour mettre à jour les poids.
- Ce processus se poursuit jusqu'à ce que tous les échantillons soient utilisés.



Prédicteurs d'entrée

Neurones

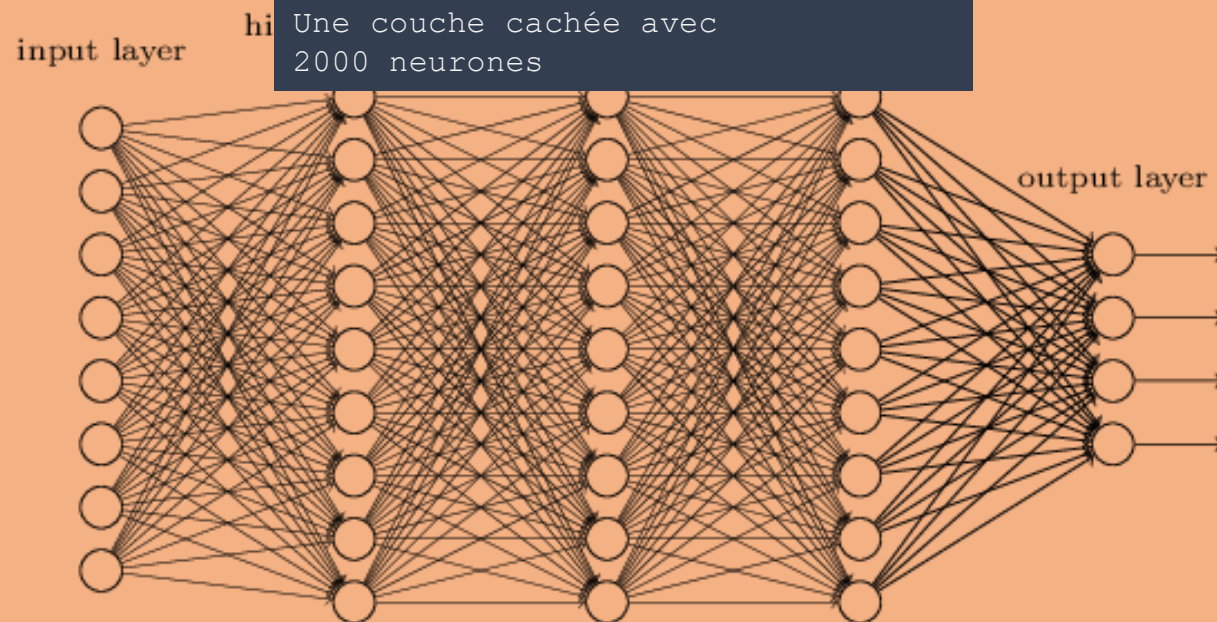
Prédicteurs de sortie (cibles)

Droits d'auteur de l'image :  
<http://neuralnetworksanddeeplearning.com/chap5.html>



# Entraînement de l'AP et validation pour L96

- $10^6$  prévisions sont générées en utilisant le modèle L96. Chaque prévision a un délai de 6 heures.
- Ces millions de conditions initiales sont situées dans diverses parties de l'espace de phase du modèle L96. Ces données (prédicteurs, prédicteurs) = (conditions initiales, prévision) sont appelées *échantillons*.
- Ces données sont ensuite utilisées pour entraîner le réseau AP.
- 70 % des données sont utilisées pour l'entraînement. Les 30% restants sont utilisés pour la validation.

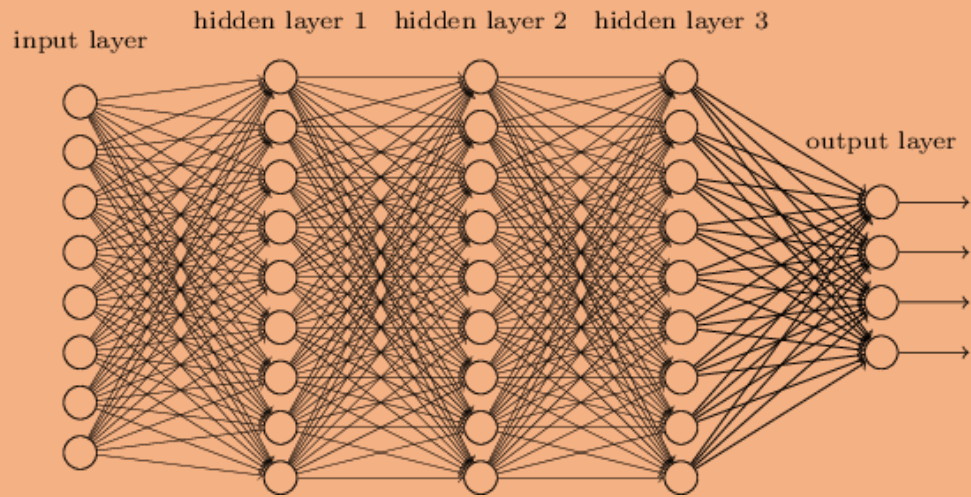


Conditions initiales à l'entrée

Prévision d'AP en sortie.

Activation sigmoïde

# L'apprentissage profond comporte de nombreux éléments mobiles!



L'entraînement d'un AP implique des essais et des erreurs pour régler les hyperparamètres.

Une époque utilise toutes les données pour l'entraînement.

Une époque est divisée en lots de 1000 chacun. (Le gradient est calculé sur un lot).

Les valeurs suivantes des hyperparamètres ont été jugées les meilleures pour ce simulateur.

Combien de couches cachées?	1 couche
Combien de neurones dans une couche cachée?	2000
Quelle fonction d'activation?	Sigmoïde
Taille du lot?	1000
Taux d'apprentissage?	0,01
Optimisation de la descente de gradient?	RMSprop

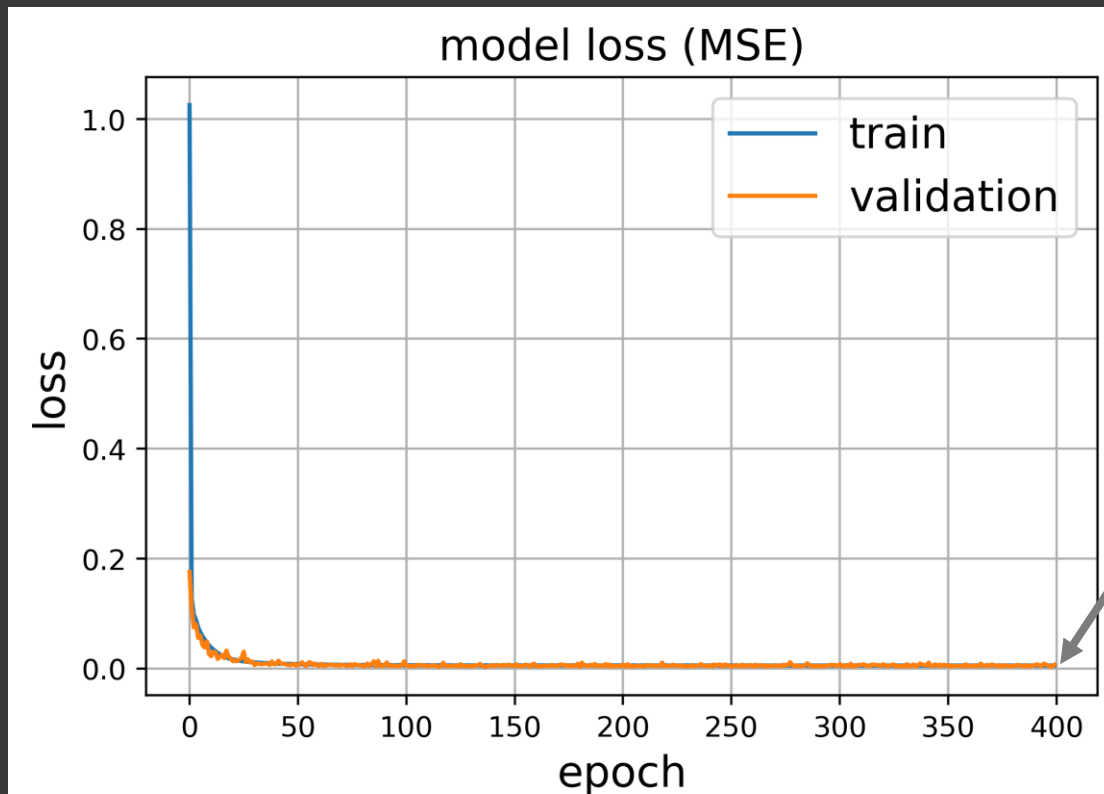
Des techniques plus avancées d'AP comme

CNN (Réseau neuronal convolutif)

RNN (Réseau neuronal récurrent)

ont beaucoup plus d'hyperparamètres réglables.

# Entraînement de l'AP



Petite erreur de simulation non nulle.

Les erreurs d'entraînement et de validation sont comparables - pas de surapprentissage!

Une fois l'entraînement terminé, le modèle est sauvegardé. Le modèle sauvegardé contient essentiellement toutes les pondérations optimisées.

Le modèle entraîné est utilisé pour lancer la prévision en utilisant 100 conditions initiales différentes qui ne font *pas* partie de l'ensemble d'entraînement, mais qui sont utilisées pour la validation.

# Expériences avec l'AP

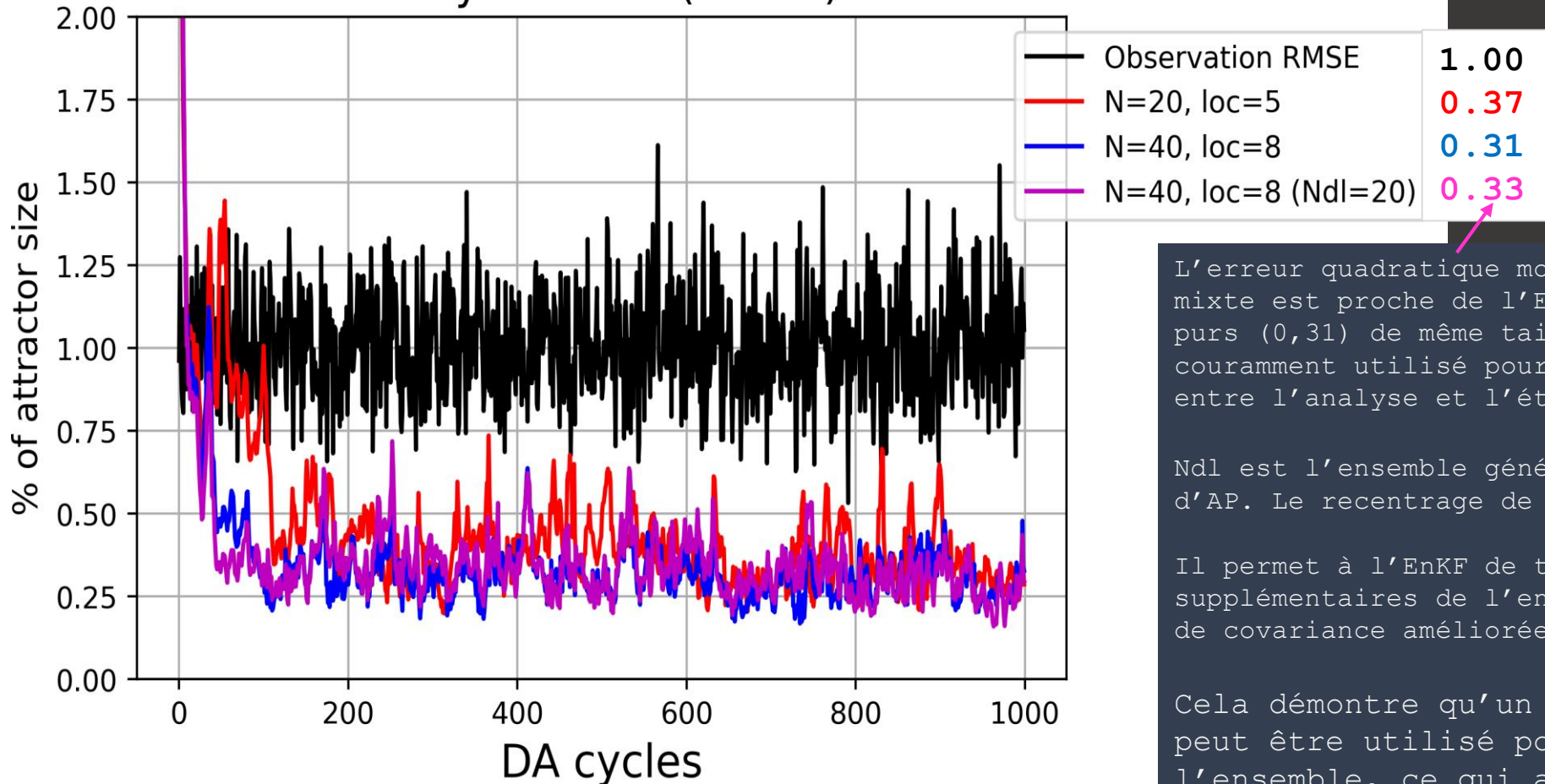
Trois expériences d'assimilation de données sont exécutées :

N=20, ensemble de modèles purs.

N=40, ensemble de modèles purs.

N=40, ensemble de modèles mixtes. Cette expérience utilise 20 membres d'ensembles de modèles et 20 membres d'ensembles de simulateurs. Ceux-ci sont ancrés (c.-à-d. recentrés) sur les 20 membres de l'ensemble de modèles purs à chaque cycle d'AD.

Analysis RMSE (m=20)



L'erreur quadratique moyenne (EQM) d'un ensemble mixte est proche de l'EQM d'un ensemble de modèles purs (0,31) de même taille. L'EQM est un paramètre couramment utilisé pour quantifier les différences entre l'analyse et l'état réel.

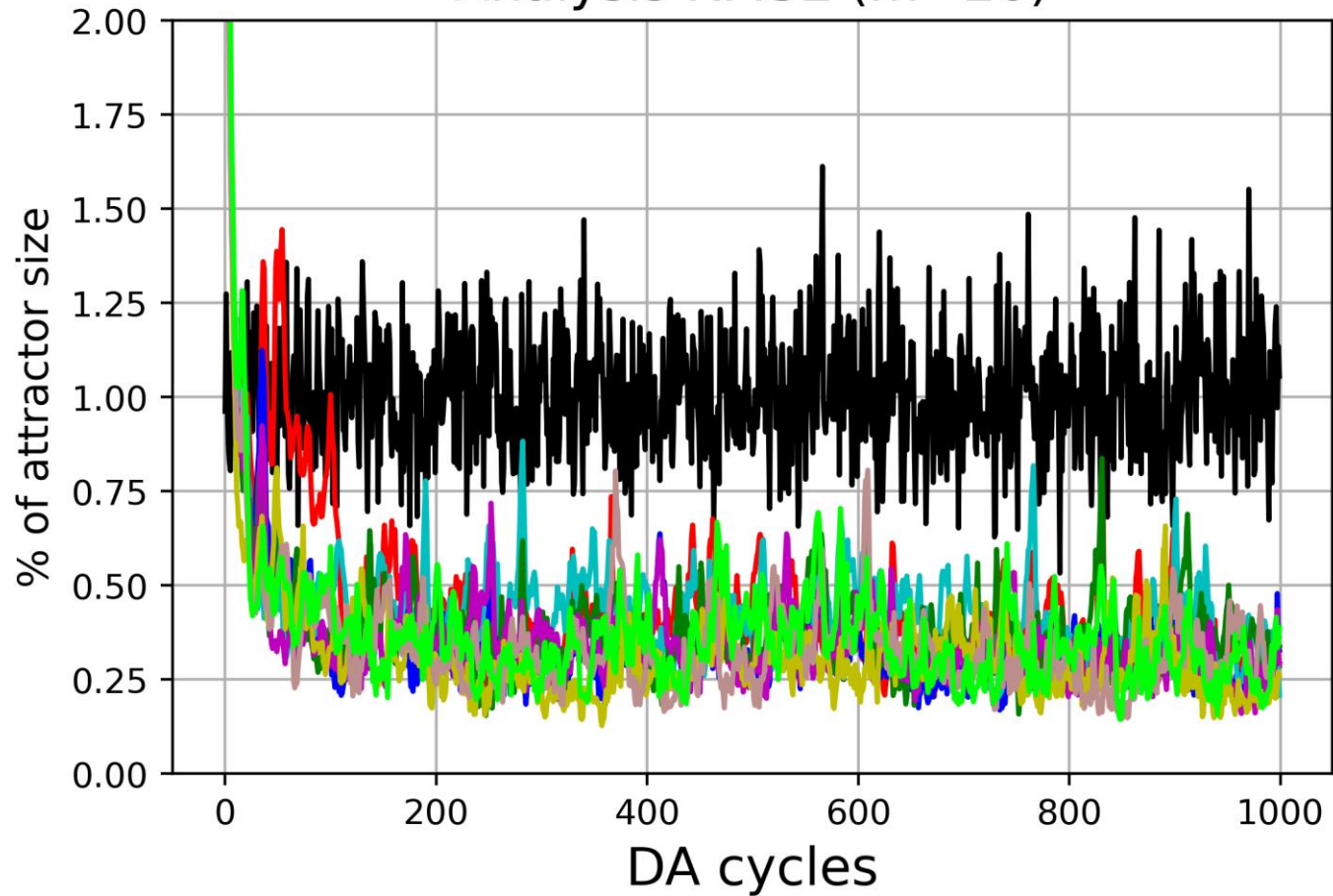
Ndl est l'ensemble généré à l'aide du simulateur d'AP. Le recentrage de Ndl corrige l'erreur moyenne.

Il permet à l'EnKF de tirer parti des 20 membres supplémentaires de l'ensemble grâce à des estimations de covariance améliorées.

Cela démontre qu'un simulateur basé sur l'AP peut être utilisé pour augmenter la taille de l'ensemble, ce qui augmente la performance et la précision de l'EnKF.

# Sensibilité à la taille des ensembles d'AP

Analysis RMSE (m=20)



— Observation RMSE	1.00
— N=20, loc=5	0.37
— N=40, loc=8	0.31
— N=25, loc=5 (Ndl=5)	0.41
— N=30, loc=5 (Ndl=10)	0.36
— N=40, loc=8 (Ndl=20)	0.33
— N=45, loc=8 (Ndl=25)	0.28
— N=50, loc=8 (Ndl=30)	0.32
— N=60, loc=8 (Ndl=40)	0.33

N=45 est le plus performant! Le simulateur d'AP peut être utilisé pour plus que doubler (N=45) la taille de l'ensemble.

Lorsque Ndl augmente, la correction de la moyenne devient moins efficace. Par conséquent, l'EQM commence à augmenter.



# Conclusion et travaux supplémentaires

- Un modèle dynamique chaotique peut être simulé au moyen d'un réseau d'AP si les données sont suffisantes.
- Dans le contexte de l'EnKF, l'ensemble mixte avec la correction appliquée au sous-ensemble d'AP par recentrage se rapproche de l'EQM de l'analyse de l'ensemble complet du modèle.
- Malgré l'erreur de simulation de chaque membre de l'ensemble, l'estimation améliorée de la covariance permet à l'ensemble mixte d'obtenir de bons résultats.
- Le réseau neuronal d'AP pourrait être remplacé par un réseau neuronal convolutif (CNN), car les CNN sont plus efficaces que les réseaux neuronaux d'AP.
- Ce travail fournit une validation de principe de l'application du simulateur d'AP à l'EnKF.
- L'objectif ultime de ce travail est de mettre en œuvre cette idée dans le système d'assimilation de données d'ECCC.
- Les travaux de mise au point d'un simulateur pour GEM (le modèle de prévision d'ECCC) sont en cours.
- Heureusement, les données de prévision d'ensemble passées de GEM sont archivées depuis

# Références

Côté, J., Gravel, S., Methot, A., Patoine, A., Roch, M., et Staniforth, A. (1998): The operational CMC-MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation, *Mon. Weather Rev.*, 126, 1373–1395, [https://doi.org/10.1175/1520-0493\(1998\)126<1373:TOCMGE>2.0.CO](https://doi.org/10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO);

Buehner, Mark. (2020) : Local Ensemble Transform Kalman Filter with cross validation. *Mon. Weather Rev.*, 148. 2265–2282. <https://doi.org/10.1175/MWR-D-19-0402.1>

Houtekamer, P.L. et Zhang, F. (2016) : Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Weather Rev.*, 144. 4489–4525.

Khade, V.M. (2022) : Using Deep Learning to increase the ensemble size in an EnKF with the recentering technique : experiments with the Lorenz 1996 model. *EnKF workshop, Balestrand, Norway, 30 May 2022*.

Lorenz, Edward (1996) : Predictability - A problem partly solved. *Seminar on Predictability, Vol. I, ECMWF*.

Seung-Jong Baek et coll., (2005) : Localized error bursts in estimating the state of spatiotemporal chaos. *Chaos* 14, 1042 (; <https://doi.org/10.1063/1.1788091>

Weyn, J. A., Durran, D. R., Caruana, R. (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002109.

<https://doi.org/10.1029/2020MS00210>

<https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/>

<http://neuralnetworksanddeeplearning.com/chap4.html>

<https://www.omnisci.com/technical-glossary/cpu-vs-gpu>