Deep Learning simulator for dynamical model : Application to increase ensemble size

Vikram Khade and Alain Beaulne

Government of Canada Data Conference
Leveraging Data to Advance Innovation

22-23 February 2023

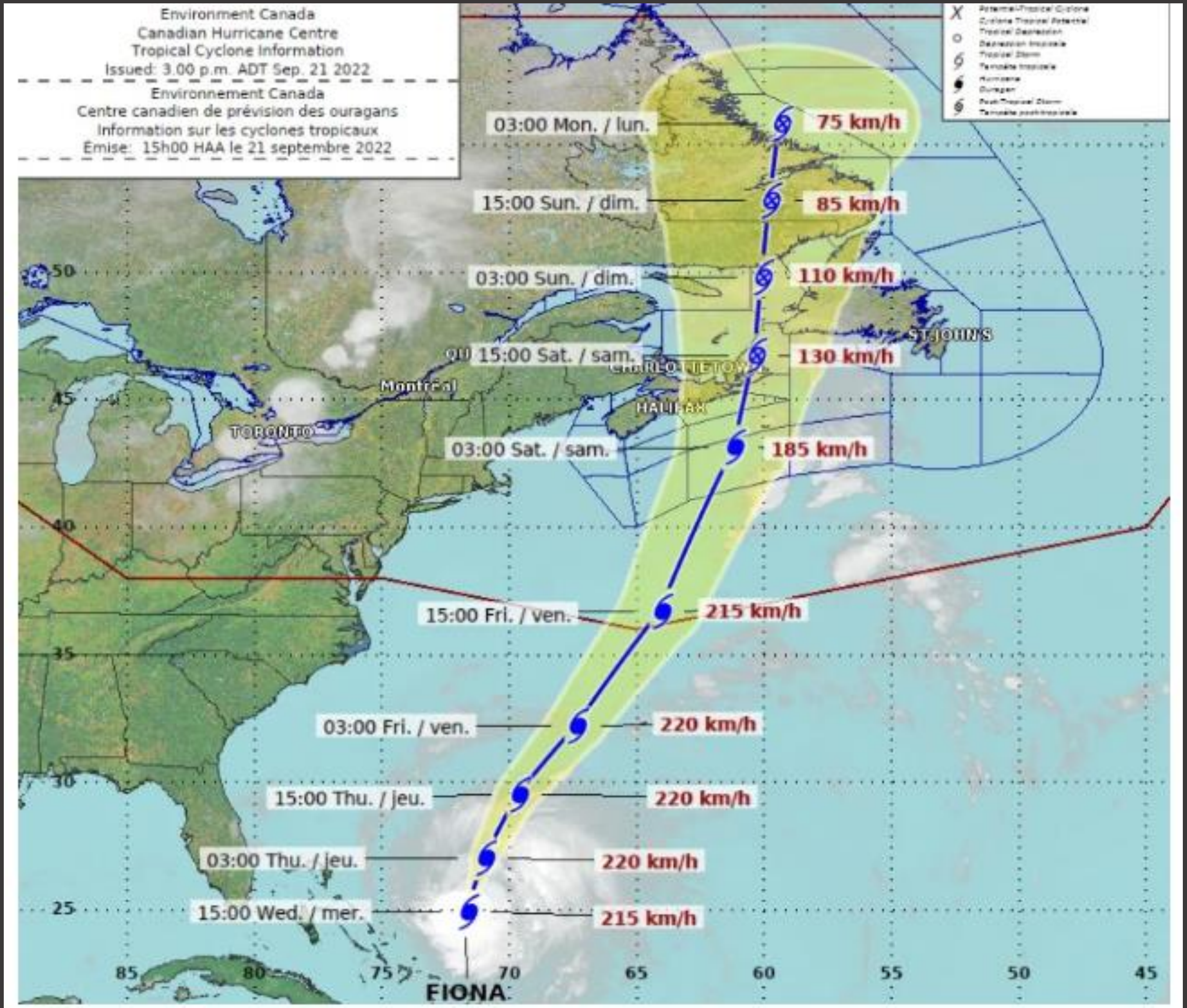
Environment and Climate Change Canada
Environnement et Changement climatique Canada

Outline

In this presentation we will understand,

- Why forecasts of the atmosphere/ocean are important.
- What is Numerical Weather Prediction (NWP).
- The role of computational resources used by the Meteorological Service of Canada (MSC), ECCC in producing NWP based forecasts.
- That forecasts are inherently uncertain and that the ensemble approach can quantify this uncertainty.
- Some applications of ensembles including the mathematical discipline of Data Assimilation (using the Ensemble Kalman Filter(EnKF) technique).
- The role of ensemble size in an EnKF.
- Proposed work at MSC to use a Deep Learning based simulator to mimic NWP to efficiently increase the ensemble size.
- A proof-of-concept of this proposed work using a highly simplified model of the atmosphere known as Lorenz-96.

Hurricane Fiona track forecast using Numerical Weather Prediction (NWP)



This picture shows the 5 day lead time forecast for hurricane Fiona issued by ECCC on Wednesday, 21 Sept., 2022. The forecast is initialized at 15:00.

The green shading shows the envelope of possible hurricane tracks as given by NWP which simulates several **different scenarios** (known as an **ensemble**).

Hurricane Fiona proved to be the most intense hurricane to hit Canada. It hit Nova Scotia with 169 km/h winds on Saturday, 24 September.

The government relief agencies could act effectively thanks to the accurate estimation of the forecast uncertainty done using ensemble NWP.

Credits :
https://globalnews.ca/news/9144648/hurricane-fiona-potentially-severe-event-atlantic-canada/
Canadian Hurricane Centre, MSC, ECCC
https://en.wikipedia.org/wiki/Hurricane_Fiona

What is Numerical Weather Prediction (NWP) ?

Mathematical equations known as differential equations govern the dynamic behavior of the atmosphere (and oceans). The future evolution of the atmosphere is calculated by solving these equations using numerical techniques on a computer. These techniques known as NWP discretizes the global domain by using a grid and the equations are solved on this grid. The discretized form of these equations is known as the model. GEM is the model used by MSC for weather forecasting.

The size of each cell in the grid depends on the resolution of the model. For example, one version of GEM model, used by ECCC for global ensemble modelling, has a resolution of 39 km. Since resolution is finite, any model is only an approximation to the equations governing the atmosphere.

The model is not aware of the physical processes within any given cell. These processes have to be represented using *parametrizations*. There are several parametrizations possible for the same process. Each one gives rise to a slightly different configuration of the model. Consequently, different configurations give rise to different forecasts.

The forecast lead time can range from a few hours to a few days depending on the application. It is important for the forecast to execute in reasonable time to be useful. For example, if the 7 day forecast executes in 7 days on the computer, this forecast is essentially useless.

What is an ensemble ?

Ensemble is composed of several members (also known as Monte Carlo samples).

Each member is a different forecast based on an unique scenario. Each scenario is different in terms of one or more of the following :

(1) Model configuration (eg. parametrizations).

(2) Initial conditions (such as the global distribution of the temperature at the initial time of the forecast.)

(3) Boundary conditions (such as sea surface temperature over the forecast period).

Different scenarios are necessitated by the fact that the initial conditions are not known perfectly. Similarly, the boundary conditions are also uncertain. The ensemble members are constructed to be consistent with the uncertainty in initial conditions, boundary conditions and the model configuration.

The estimation of uncertainty in the forecast improves as the ensemble size (i.e. number of scenarios) simulated increases. In spite of very powerful computers, relatively small ensemble size (few 100s) is possible because each simulation is computationally intensive.

High Performance Computing at ECCC

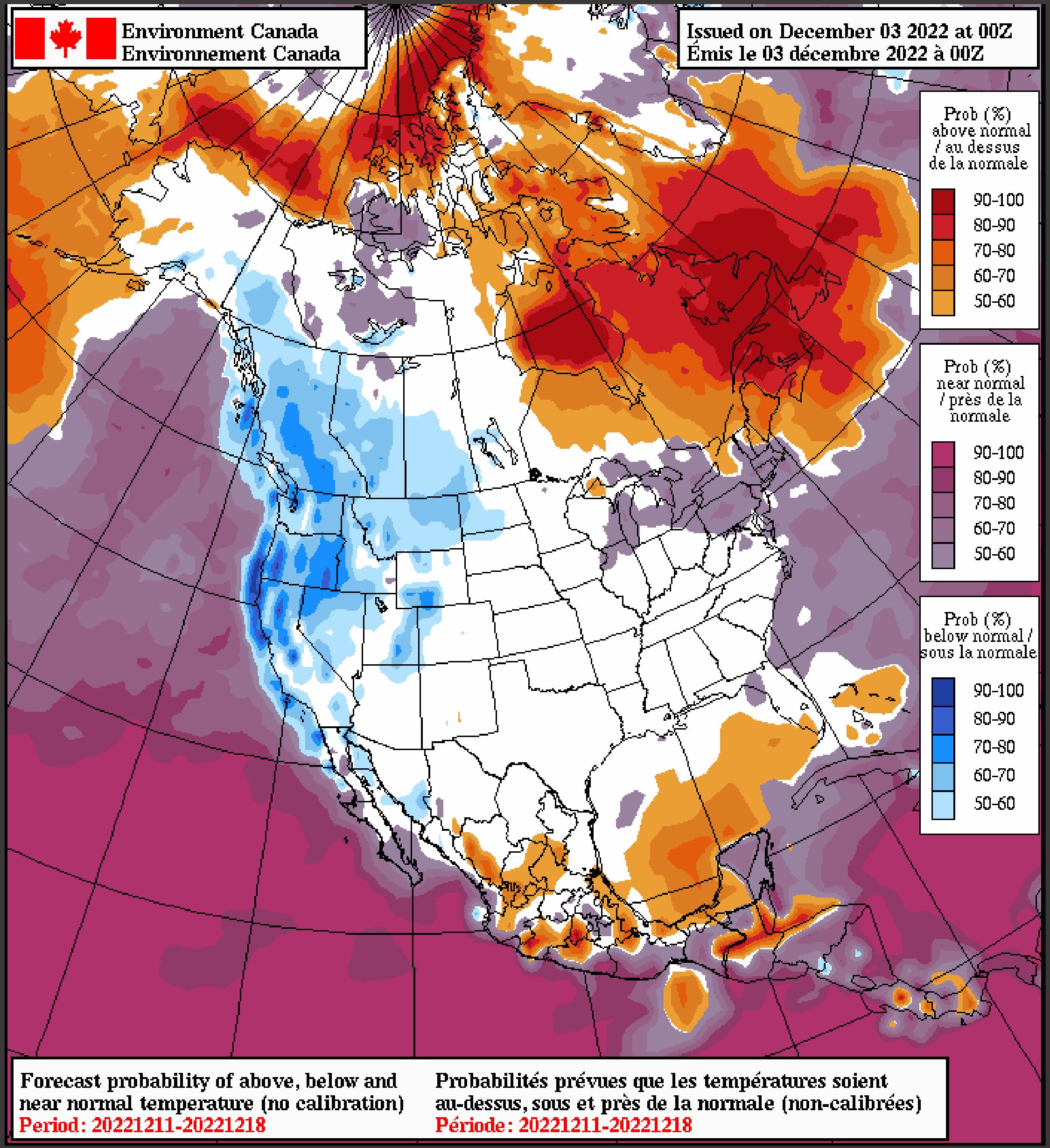
ECCC uses powerful supercomputers to carry out NWP. A picture of these supercomputers in 2017 is shown on the right. These have been updated twice since 2017.

The current supercomputers named Underhill and Robert were the 69th and 70th most powerful computers globally in June 2022 !

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
69	**Underhill** - 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	**Robert** - 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

Reference :
https://top500.org/lists/top500/list/2022/06/
https://www.canada.ca/en/shared-services/corporate/data-centre-consolidation/high-performance-computing.html

Another example of ensemble



Ensembles can be used to calculate probabilities. Higher the number of ensemble members better the estimate of the probability.

The North American Ensemble Forecast System (NAEFS) combines the ensembles produced by the Meteorological Service of Canada (MSC) and the United States National Weather Service (NWS) with a combined size of 40 members. Combining ensembles generated by different agencies is one way of increasing the ensemble size.

This figure shows probabilities of temperature being above, near or below normal for the week of 11-18 December, 2022.

Reference :
https://weather.gc.ca/ensemble/naefs/index_e.html

Computational cost of running forecast

At ECCC, a 2 week lead time ensemble forecast consisting of 21 members is executed using 800x21 CPUs.

The execution time for this forecast is about 1 hour. The different ensemble members are run in parallel and hence the ensemble forecast finishes in ~ 1 hour.

Increasing ensemble size is computationally expensive since each additional ensemble member needs 800 CPUs.

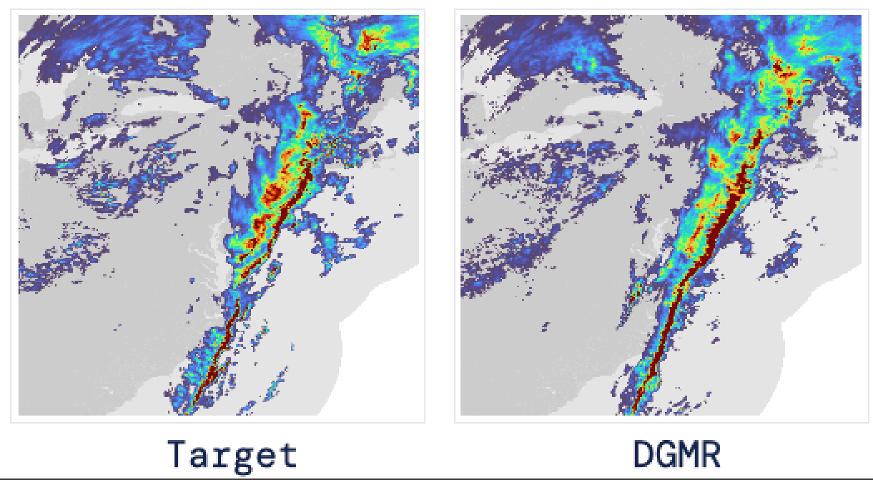
Deep Learning (DL) is a computationally feasible tool to increase the ensemble size.

Acknowledgements :

Dr. Xingxiu Deng
ECCC

Deep Learning – another technique to increase ensemble size

DGMR is a DL based simulator developed by DeepMind to forecast rainfall in the next 90 minutes. The figure shows the comparison between the forecast (DGMR) and the actual radar observations (Target). Most of the 56 weather forecasters rated DGMR as better than the extant numerical weather prediction models !



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

- Deep Learning (DL) has emerged as a powerful tool in the field of Artificial Intelligence to tackle several problems like image recognition, text classification, video enhancement etc.

- Existing data is used to train Deep Learning networks to perform a particular task. The trained model is saved and then used on unseen input. The training process is computationally costly but once trained, using the DL model on unseen data is very fast.

- Another area of DL application is that of training simulators to mimic the NWP model. In this approach archived (initial conditions, forecast data) of the NWP model is used to train the DL simulator.

- Though the training is expensive, the trained model executes ensemble forecasts very fast. This has been demonstrated by researchers.

Weyn et. al. : Demonstrates Deep Learning of weather

JAMES | Journal of Advances in 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[1], Dale R. Durran[1], and Rich Caruana[2]

[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, [2]Microsoft Research, Redmond, WA, USA

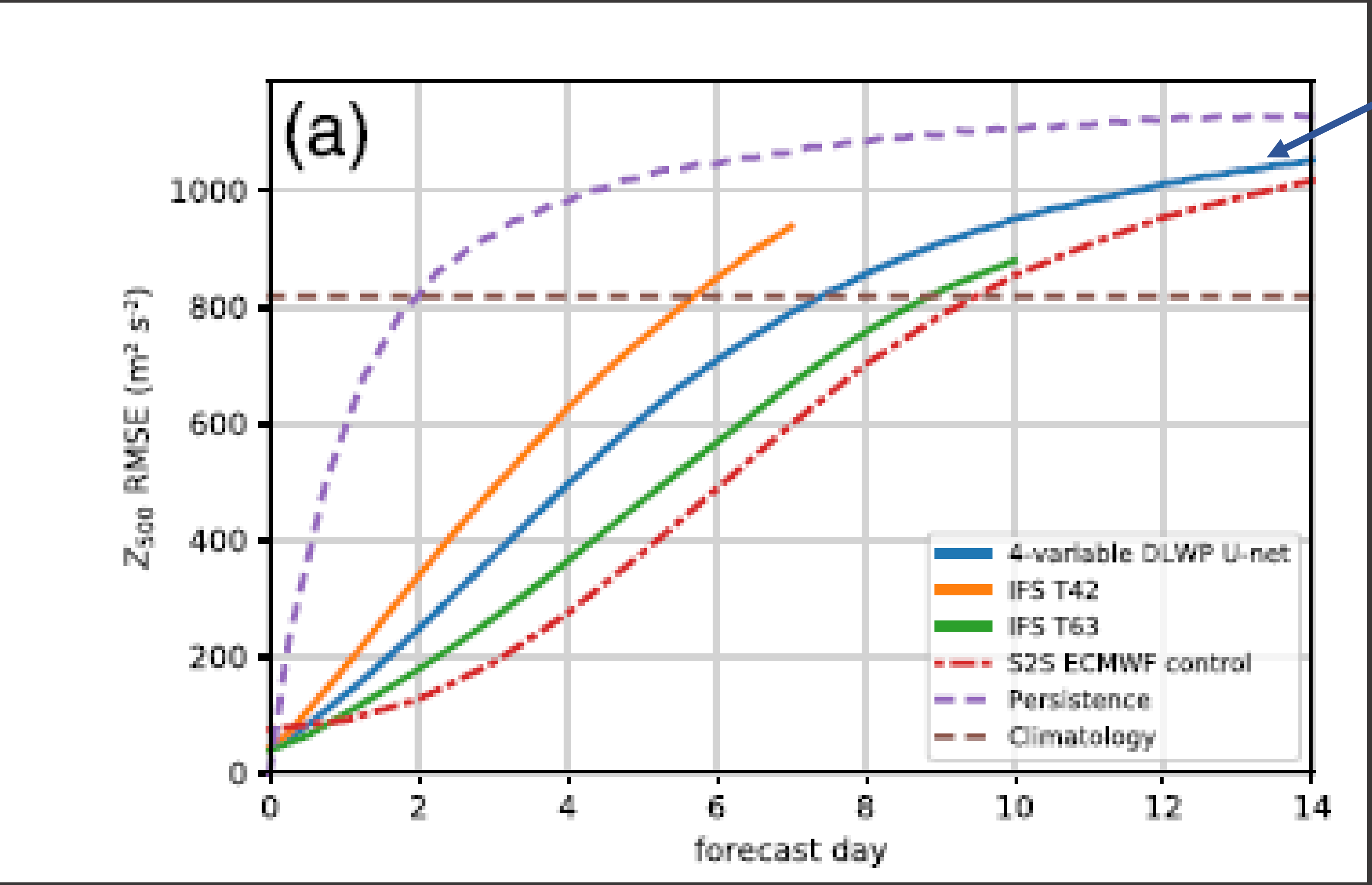
In this paper CNN (Convolutional Neural Networks) are used by Weyn et. al. to simulate a NWP model for 2 week forecast. CNN are special type of DL networks which are well suited to data having spatial structure.

Training data used : ERA5, regridded to 2 degree (1979-2018)

Inputs : Geopotential height at 500 hPa.
Geopotential height at 1000 hPa.
300-700 hPa geopotential thickness.
2m temperature.
ToA insolation.
Land-sea mask.
Topographic height.

Outputs the fields mentioned in Inputs.

Weyn et. al. : Deep learning model outperforms T42 model



DL simulator performs better than IFS T42 and Persistence.

IFS T42 : ECMWF IFS model, 62 levels, 2.8 deg.

IFS T63 : ECMWF IFS model, 137 levels, 1.9 deg.

S2S : ECMWF IFS subseasonal to seasonal model
16 km – 31 km.
Fully coupled to ocean and sea ice models.

ECMWF (European Centre for Medium-Range Weather Forecast) is an organization supported by most nations of Europe. IFS (Integrated Forecasting System) is the operational model of ECMWF.

Weyn et. al. : Computational speed

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,

Single Nvidia Tesla V100 GPU

Training : 2-3 days

Forecasting : 1000 ensemble members, 1-month forecasts in 3 minutes !

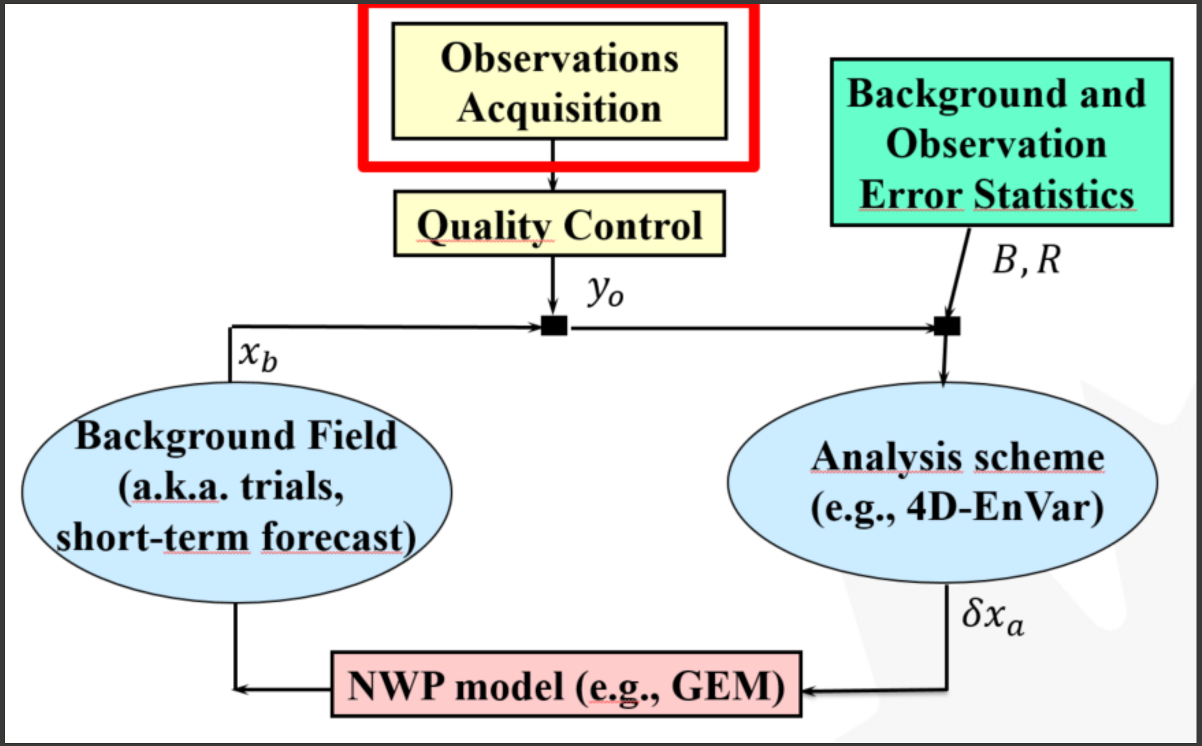
This demonstrates that DL can be used very efficiently to increase the ensemble size.

Recap

So far we have learnt that,

- Ensembles of forecasts are important to quantify the uncertainty in the forecast.
- Each ensemble member simulates a different scenario. Ideally we would like to simulate all possible scenarios.
- However, increasing the number of scenarios is computationally expensive.
- Deep Learning (DL) based simulators have emerged as a tool to increase the ensemble size at a low computational cost.
- In the next few slides we will learn about the importance of ensemble size in an Ensemble Kalman Filter (EnKF).
- EnKF is a Data Assimilation technique which combines the estimate of current weather as given by the NWP model with atmospheric observations to produce the best possible picture of the current weather.

Data Assimilation in NWP



Data Assimilation (DA) is a technique used to combine different sources of information along with their uncertainties to obtain the best possible estimate of the current weather condition. This estimate is known as the analysis.

The different sources of information are the NWP **model estimate** (background field) and **observations** of the atmosphere made by various instruments. The uncertainty in the model estimate is known as the Background error statistics. The uncertainty in the observations is known as the Observation error statistics. An accurate knowledge of these statistics is very important to obtain a high quality analysis.

A DA step is carried out every 6 hours to obtain an analysis from which a short-term 6 hour lead time forecast is issued, which is then used as a model estimate for the next DA step. A variety of observations are used in DA.

Credit :
Judy St-James and Thomas Milewski
Data Assimilation and Quality Control Section
ECCC

Main Observing Networks used in NWP Systems

Geostationary satellites

Polar-orbiting satellites

Micro satellites
GNSS Radio Occultation

Aircraft

Upper Air sites

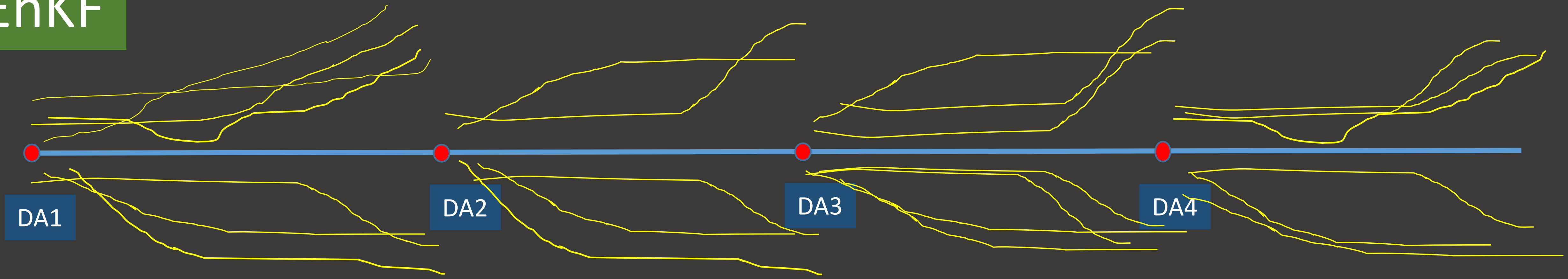
Surface stations

Ground-based GNSS

Buoys, drifters and ships

Credit :
Judy St-James and Thomas Milewski
Data Assimilation and Quality Control Section
ECCC

EnKF



- The EnKF (Ensemble Kalman Filter) is a Data Assimilation technique which uses ensemble forecast samples to estimate the background uncertainty (covariance). The figure shows a cartoon of the EnKF. The DA cycles happen every 6 hours. The red dot shows the mean analysis. The yellow lines represent the ensemble members which are determined by solving the NWP equations on the computer.

- Larger the ensemble size better the estimate of the background error statistics.

- As already explained, it is very costly to increase the ensemble size.

- In the following slides the application of a DL simulator to increase the ensemble size in EnKF is demonstrated using the Lorenz-96 model.

- Lorenz-96 (L96) is a simplified model of the atmosphere.

- A DL simulator is trained to generate the forecast of L96. The training is done using the (initial condition, forecast) data generated by numerical integration of L96 equations.

40 dimensional Lorenz 1996 model (L96)

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

The x_i are the components which are weather variables around a latitude circle. On the rhs, the first term is advection and the second term is diffusion.

$i = 1…40$ (Dimensionality = d = 40)

$F = 8$ (Forcing)

$x_0 = x_N$

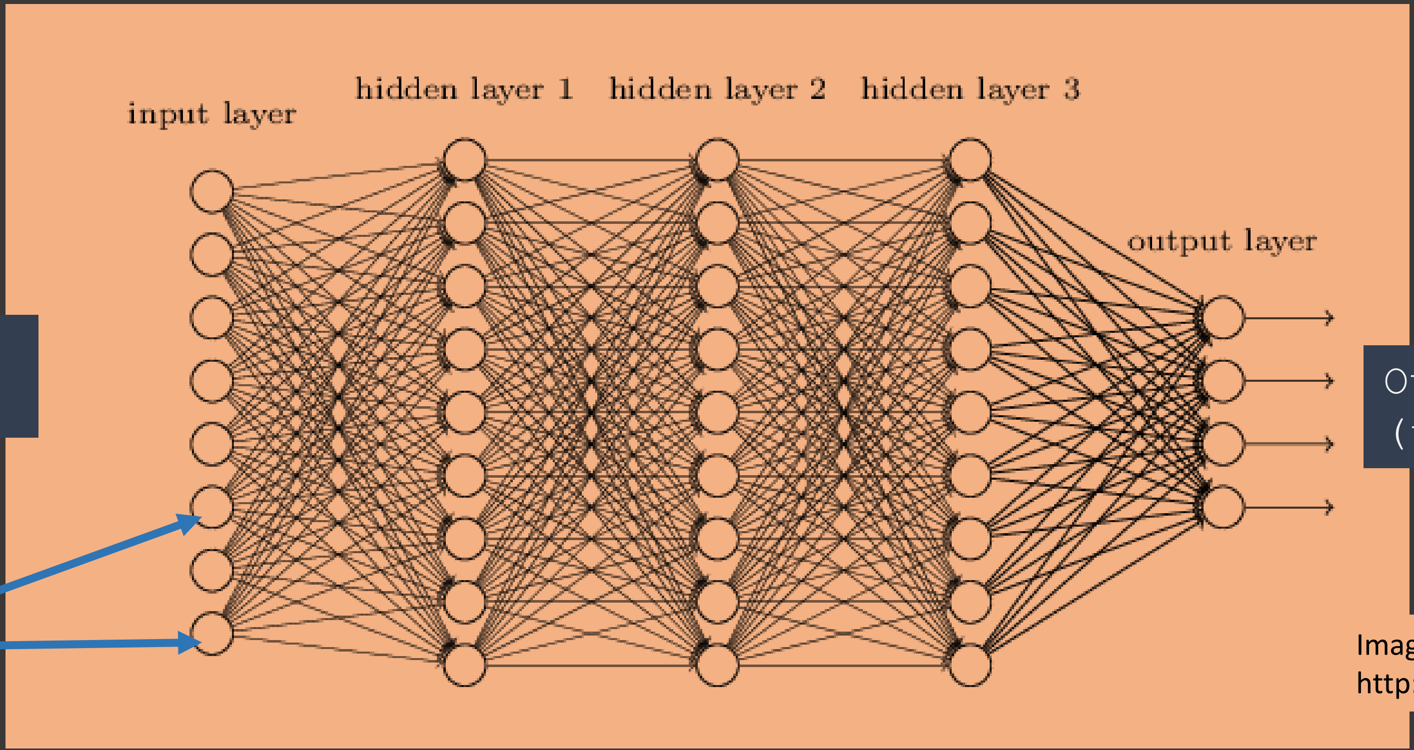
0.05 time units ~ 6 hours in atmospheric model.

This model has been used by many researchers as a test bed to try out new data assimilation schemes, improvements to DA techniques etc.

How does Deep Learning (DL) work ?

- Each connection has a weight associated with it.
- The weights are initialized with random values.
- The predictors of samples are inputted and the DL calculates the output.
- The error between this output and sample target is calculated.
- This error is back propagated to update the weights.
- This process is continued till all the samples are used.

Input predictors



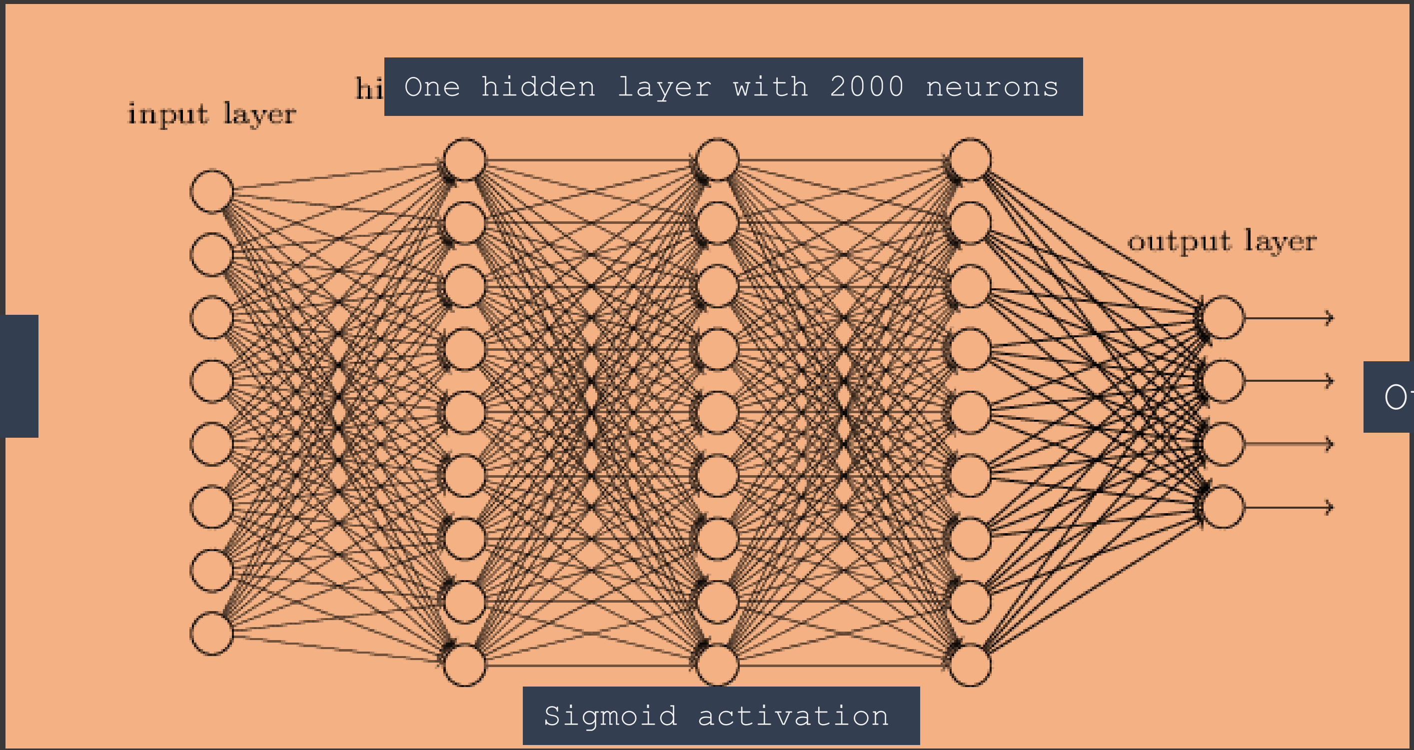
Neurons

Output predictands (targets)

DL training and validation for L96

- 10^6 forecasts are generated using L96 model. Each forecast has 6 hours lead time.
- These million initial conditions are located in various parts of the phase space of the L96 model. These data (predictors, predictand) = (initial condition, forecast) are known as *samples*.
- These data are then used to train the DL network.
- 70% of the data are used for training. The remaining 30% are used for validation.

Input initial conditions

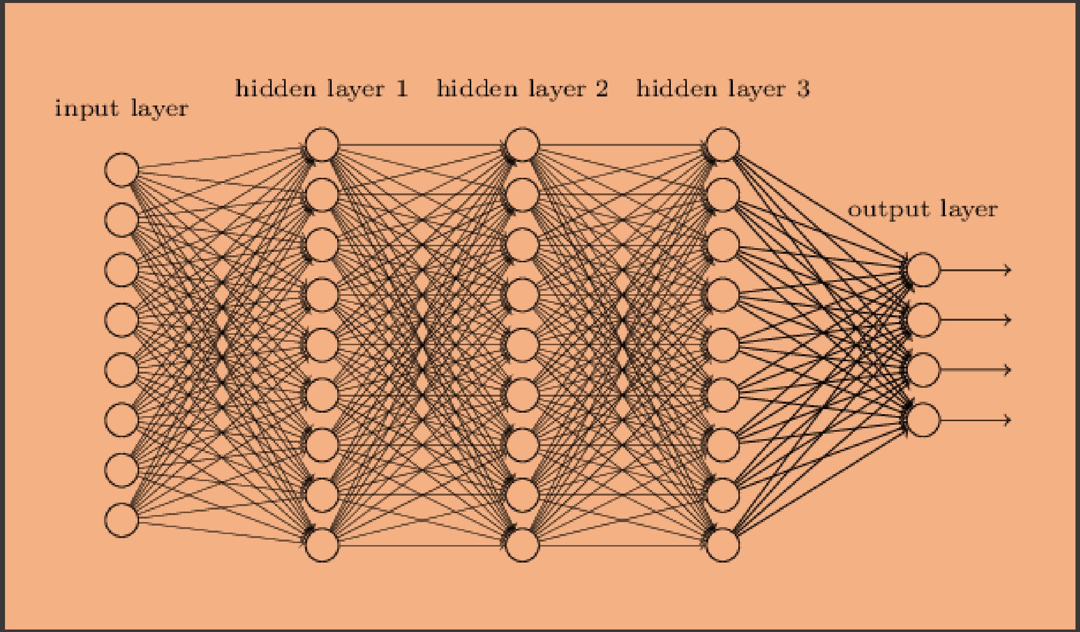


One hidden layer with 2000 neurons

Sigmoid activation

Output DL forecast.

Deep Learning has many moving parts !



Training a DL involves some trial and error to tune the hyperparameters.

One epoch uses all the data for training.

One epoch is divided into batches of 1000 each (The gradient is calculated over one batch).

The following hyperparameters values were found to be best for this simulator,

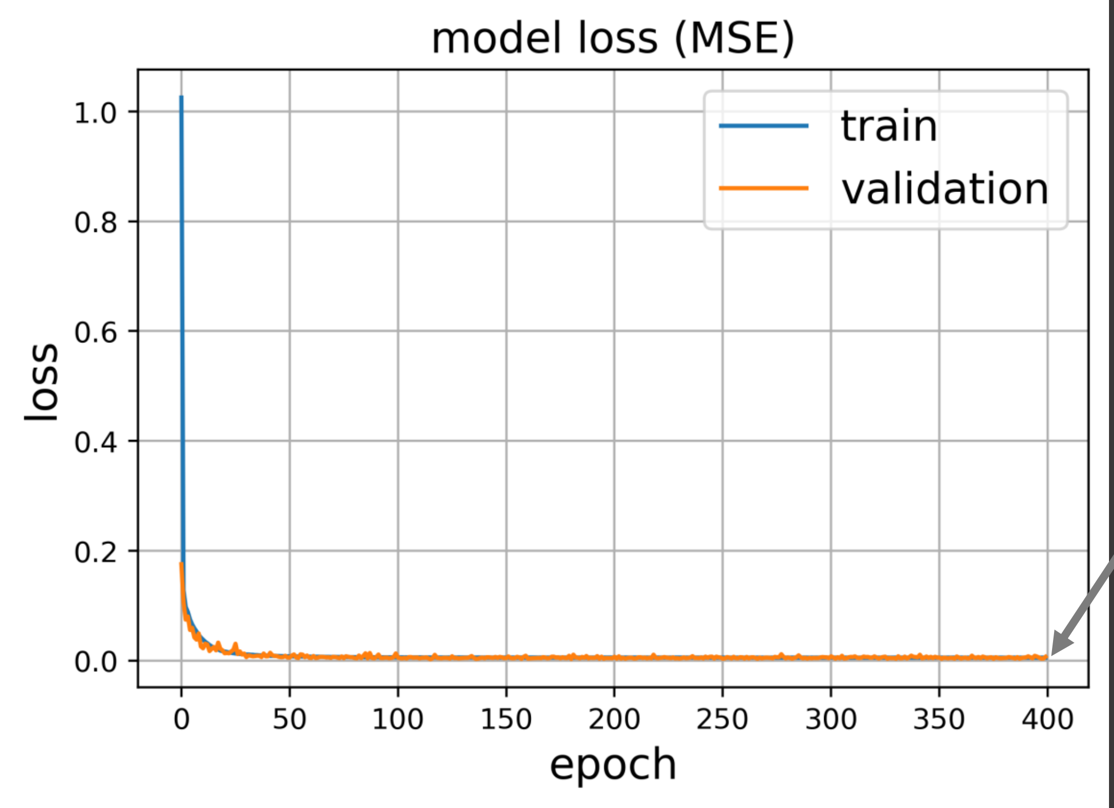
How many hidden layers ?	1 layer
How many neurons in hidden layer ?	2000
Which activation function ?	Sigmoid
Batch size ?	1000
Learning rate ?	0.01
Gradient descent optimization ?	RMSprop

More advanced DL techniques like

CNN (Convolutional neural network)
RNN (Recurrent neural network)

have many more tunable hyperparameters.

DL training



Small non-zero simulation error.

Training and validation errors are comparable – no overfitting !

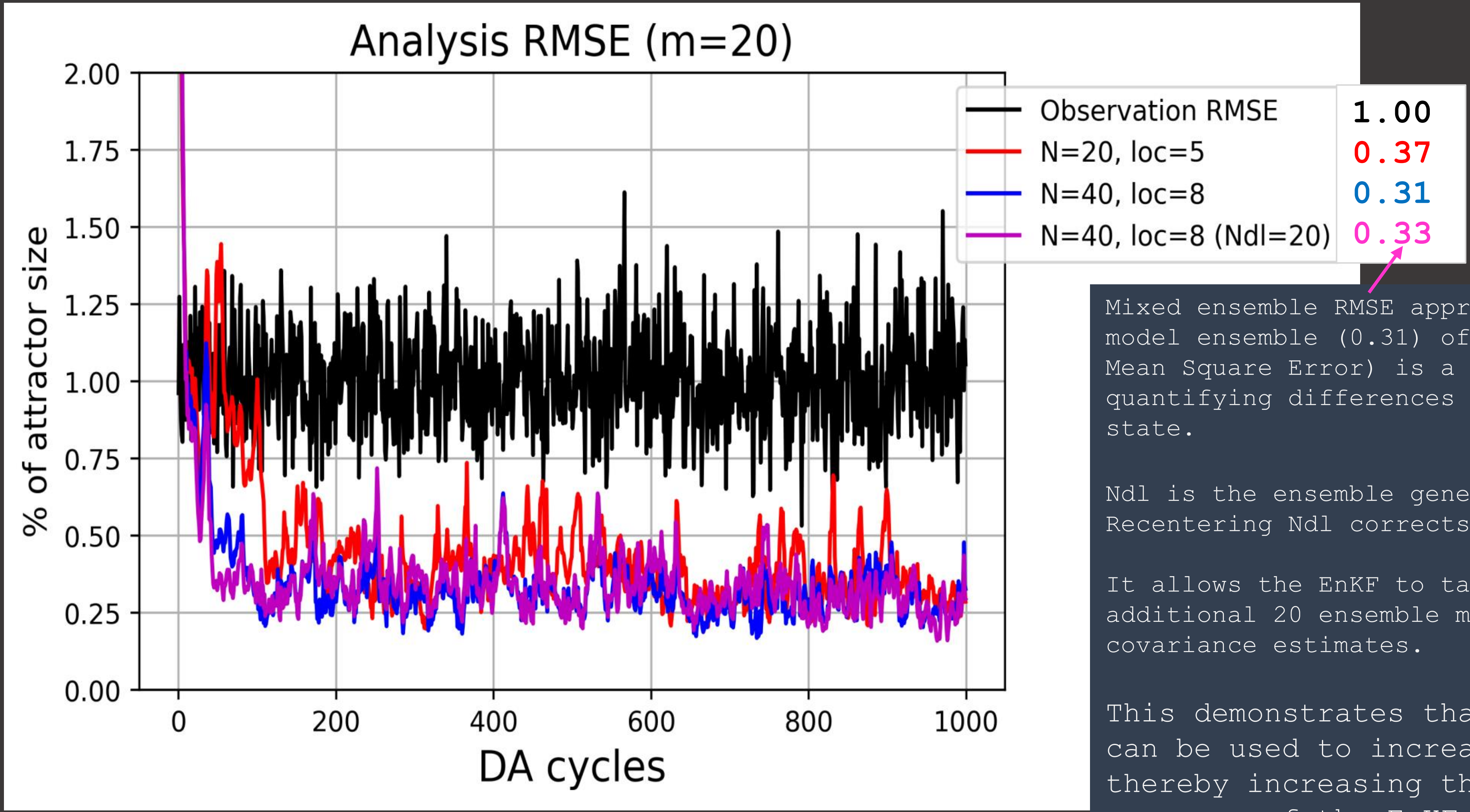
After the training is completed the model is saved. The saved model basically contains all the optimized weights.

The trained model is used to launch forecast using 100 different initial conditions which are *not* from the training set, but are used for validation.

DA experiments

Three Data assimilation experiments are run :

N=20, pure model ensemble.
N=40, pure model ensemble.
N=40, mixed model ensemble. This experiment uses 20 model ensemble members and 20 simulator ensemble members. These are anchored (i.e. re-centered) on the 20 pure model ensemble members at every DA cycle.



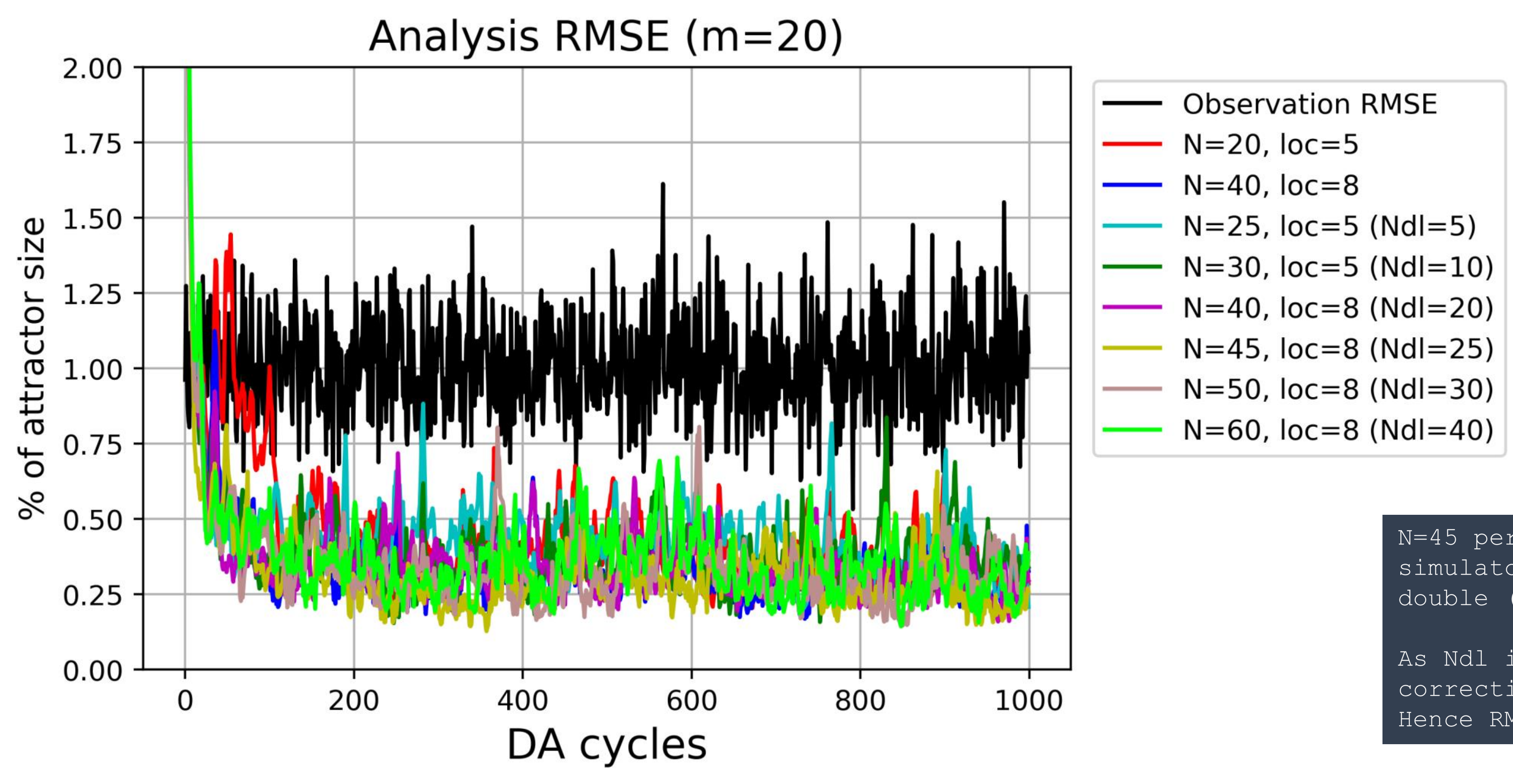
Mixed ensemble RMSE approximates the RMSE of a pure model ensemble (0.31) of the same size. RMSE (Root Mean Square Error) is a commonly used metric for quantifying differences between analysis and the true state.

Ndl is the ensemble generated using DL simulator. Recentering Ndl corrects the mean error.

It allows the EnKF to take advantage of the additional 20 ensemble members through improved covariance estimates.

This demonstrates that a DL based simulator can be used to increase the ensemble size thereby increasing the performance and accuracy of the EnKF.

Sensitivity to DL ensemble size



1.00
0.37
0.31
0.41
0.36
0.33
0.28
0.32
0.33

N=45 performs the best ! The DL simulator can be used to more than double (N=45) the size of ensemble.

As Ndl increases the mean correction becomes less effective. Hence RMSE starts increasing.

Conclusion and further work

- A chaotic dynamical model can be simulated using a DL network given enough data.
- In the context of the EnKF, the mixed ensemble with correction applied to the DL sub-ensemble through recentering approximates the analysis RMSE of the full model ensemble.
- In spite of the simulation error in each ensemble member the improved covariance estimate allows the mixed ensemble to perform well.
- The DL neural network could be replaced by a Convolutional Neural Network (CNN) since CNNs are more efficient than DL neural networks.
- This work provides a proof-of-concept of application of DL simulator to EnKF.
- The ultimate objective of this work is to implement this idea in the ECCC Data Assimilation system.
- Work to develop a simulator for GEM (ECCC’s forecast model) is underway.
- Fortunately, the past ensemble forecast data from GEM has been archived since 2015. This data (known as GEPS) will be used for training. It is well known that amount of data plays the most important role in DL. Consequently, the most important question is whether an accurate enough simulator for GEM can be trained using about 6 years of ensemble data.

References

Côté, J., Gravel, S., Methot, A., Patoine, A., Roch, M., and 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;

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. and 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. al, (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