IA et prévision numérique du temps: que reste-t-il aux météorologistes ?
AI and weather forecasting: What's left for the meteorologists?

Marc Bocquet[†],

with illustrations from collaborations with

Alban Farchi[†], Quentin Malartic[†],
Massimo Bonavita[§], Patrick Laloyaux[§], Marcin Chrust[§],
Tobias Finn[†], Charlotte Durand[†], Fabio d'Andrea[‡],
Julien Brajard[*], Alberto Carassi[‖], Laurent Bertino[*].

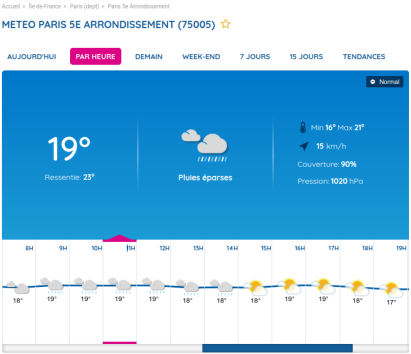
[†] CEREA, École des Ponts and EDF R&D, Île-de-France, France
[‡] LMD, ENS, Paris
[§] ECMWF, Reading, United Kingdom,
[*] NERSC, Bergen, Norway,
[‖] University of Bologna, Italy.

Outline

Numerical weather prediction (NWP)

- NWP has been framed as a *mathematical* problem with an initial condition in 1904 by Vilhelm Bjerknes, and envisioned as a *numerical computational* problem by Lewis Fry Richardson in 1922.

- This program has been implemented starting in the 1950's by e.g., Jule Gregory Charney.



- Current *medium-range* weather forecasts typically predict temperature, pressure, humidity, precipitation and hydrometeors over a range of *15 days*.

- Critical for most human activities: agriculture, energy production and distribution, transportation, extreme and dangerous natural hazards.

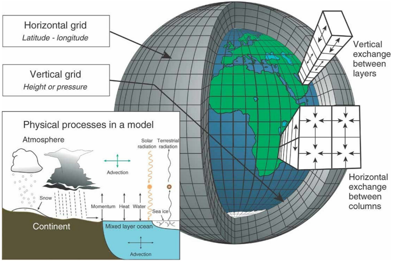
- Important NWP centres in Europe: The UK MetOffice, Deutscher Wetterdienst (DWD), Météo-France, etc and the most skilled worldwide: *The European Centre for Medium-Range Weather Forecasts (ECMWF)*.

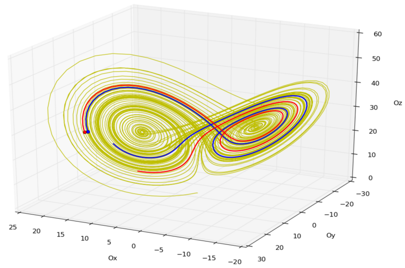
The fundamental obstacles: chaos and model error

▶ Geofluids (atmosphere, ocean, ice, etc.) are almost always dynamically *chaotic*: as dynamical systems, they are extremely *sensitive to their initial condition*. This has been discovered by Edward Lorenz [Lorenz 1963].

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x)$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho x - y - xz$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z,$$

with $\sigma, \rho, \beta = 10, 28, 8/3$.



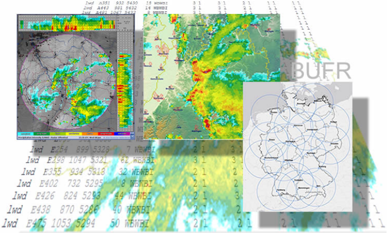


▶ There are *stochastic* and *systematic model errors* in the physical processes, but also in the *numerical representation*, and in the *forcings*.

▶ Hence, the trajectory of the atmosphere state *constantly need to be adjusted* to compensate for its sensitivity to the initial condition and its drift via model error.

The solutions: observations and computations

▶ Getting more and more *computational power* to run high-resolution models and carry out *deterministic* and *ensemble* forecasts.



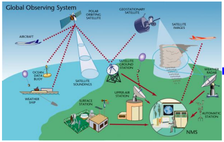
▶ Getting more and more *observations* as often as often as possible (radiosondes, synoptic stations, planes/boats/buoys, satellite, radar, lidar, etc.)

▶ But these data are *heterogeneous* and need to be *optimally combined* to yield the most accurate possible prediction over the state of the geophysical system.

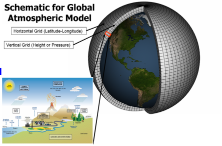
⟶ we need data assimilation!

Outline

Data assimilation (DA) in the geosciences



Data assimilation
best combines
observations and models



Expanded from numerical weather prediction to the climate science/geosciences:

- Oceanography
- Atmospheric chemistry
- Climate prediction and assessment
- Glaciology, sea-ice.
- Hydrology and hydraulics
- Geology
- Space weather
- and many other fields

Data assimilation: an inference problem

- Inference is the process of taking a decision based on limited information.

- Information comes from
  - an approximate knowledge about the laws (if any) governing the time evolution of the dynamical system,
  - imperfect (partial, noisy, indirect) observations of this system, as illustrated before.

- Sequential inference is the problem of updating our knowledge about the system each time a new batch of observations becomes available.

First ingredient: the dynamical model

- The physics and dynamics are described as PDEs and constitutive laws and as discretised into a discrete stochastic dynamical system,

$$\mathbf{x}_k = M(\mathbf{x}_{k-1}, \mathbf{p}) + \boldsymbol{\eta}_k.$$

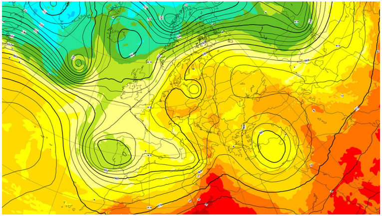
- $\mathbf{x}_k \in \mathbb{R}^{N_x}$ and $\mathbf{p} \in \mathbb{R}^{N_\mathrm{p}}$ are the model state and parameter vectors respectively.

- $M : \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ is usually a nonlinear, possibly chaotic, map.

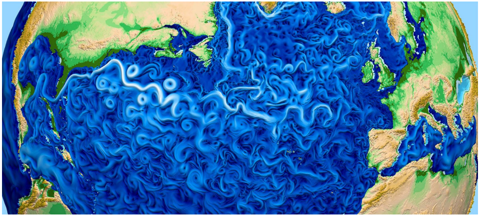
- $\boldsymbol{\eta}_k \in \mathbb{R}^{N_x}$ is the model error, represented as a stochastic additive term (more general representations are possible). Misspecified $\mathbf{p}$ is also a source of model error.

First ingredient: the dynamical model

- In the geosciences:
  - The state space dimension is huge (up to 10^9 for operational systems, up to 10^7 for research systems). A big data problem with costly models to integrate.
  - Numerical models (i.e. implementation of M) are often computationally very costly.
  - The unstable dynamics of chaotic geofluids has implicit consequences on the design of DA algorithms: One key reason why we use sequential inference.



ECMWF IFS: Geopotential at 500hPa
and temperature at 850hPa



E3SM Earth system model

- The model of the ECMWF is known as the Integrated Forecasting System (IFS).

Second ingredient: the observations

- Noisy observations, $\mathbf{y}_k \in \mathbb{R}^{N_y}$, are available at discrete times and are related to the model state vector through

$$\mathbf{y}_k = H(\mathbf{x}_k) + \boldsymbol{\epsilon}_k,$$

with $H : \mathbb{R}^{N_x} \to \mathbb{R}^{N_y}$ being the (generally nonlinear) observation operator mapping from the model to the observational space.

- The observation error, $\boldsymbol{\epsilon}_k$, is represented as a stochastic term. It accounts for the instrumental error, for deficiencies in the formulation of H, and for the representation error.

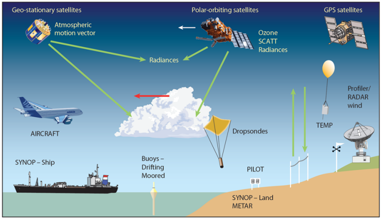
- The *representation error* arises from the presence of unresolved scales and represents their effect on the resolved scales – it is ubiquitous in physical science and inherent to the discretisation procedure [Janjić et al. 2018].

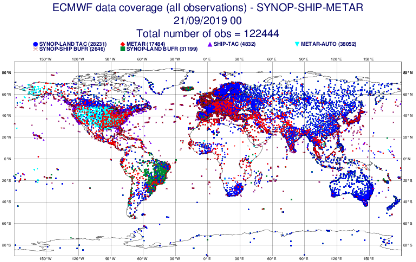
- Very often $N_y \ll N_x$, i.e. the amount of available data is insufficient to fully describe the system.

Second ingredient: the observations

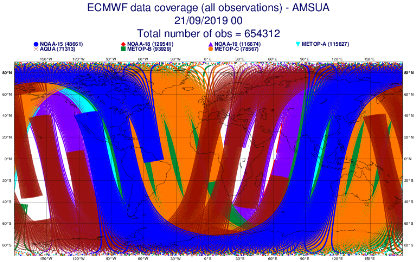
- In the geosciences: The observation space dimension is huge (up to 10^7 for operational systems, up to 10^6 for research systems). A big data problem.

- The Earth observations gather measurements of many sources: conventional, space-borne, ground-based remote sensing.





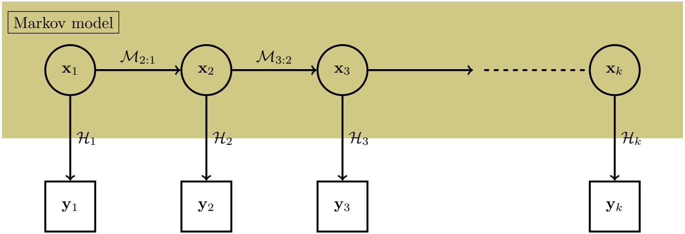
Conventional observations coverage used at ECMWF



AMSUA observations used at ECMWF

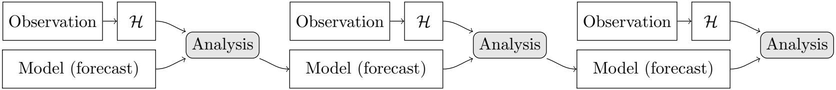
...and one to bind them all: hidden Markov model

▶ Considering the states and observations as random variables, the dynamical model, together with the observation model, define a hidden Markov model:

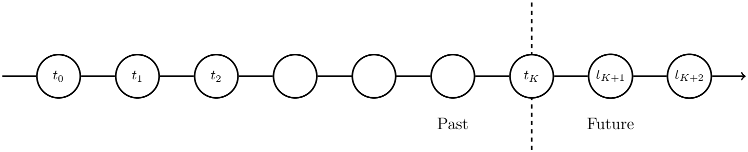


▶ This is an inverse problem: Estimate the state $\mathbf{x}$ given the observation $\mathbf{y}$.

▶ Data assimilation for forecasting chaotic geofluids: sequential schemes



Main goals of data assimilation



- Recall $\mathbf{x}_{K:0} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_K\}$, $\mathbf{y}_{K:0} = \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_K\}$:
  - Prediction: Estimate $\mathbf{x}_k$ for $k > K$, knowing $\mathbf{y}_{K:0}$,
  - Filtering: Estimate $\mathbf{x}_K$, knowing $\mathbf{y}_{K:0}$,
  - Smoothing: Estimate $\mathbf{x}_{K:0}$, knowing $\mathbf{y}_{K:0}$.
- Less formal names:
  - hindcasting, nowcasting and forecasting,
  - reanalysis,
  - parameter estimation.

The adjoint problem

▶ Typical cost function met in data assimilation (least squares problem):

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\left\|\mathbf{y} - H(\mathbf{x})\right\|^2_{\mathbf{R}^{-1}} + \frac{1}{2}\left\|\mathbf{x} - \mathbf{x}^{\mathrm{b}}\right\|^2_{\mathbf{B}^{-1}}.$$

▶ If $\mathbf{x}$ is a high-dimensional vector and H a large and complex nonlinear geophysical model, we should compute the gradient of $\mathcal{J}$ with respect to $\mathbf{x}$ to minimise it efficiently:

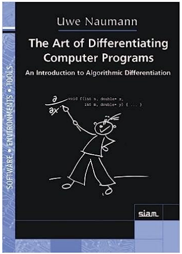
$$\nabla_{\mathbf{x}}\mathcal{J}(\mathbf{x}) = -\left[\mathbf{H}\right]^{\top}\mathbf{R}^{-1}\left(\mathbf{y} - H(\mathbf{x})\right) + \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^{\mathrm{b}}\right),$$

with $\mathbf{H} = H'$.

▶ We need to be able to compute $[\mathbf{H}]^{\top}$, i.e. tangent linear and adjoint.

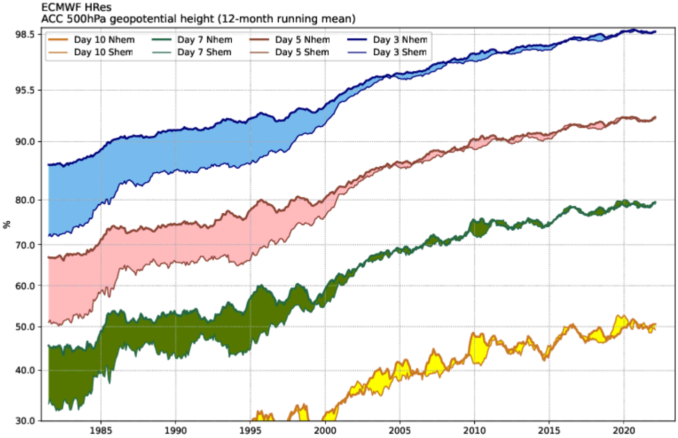
▶ But H may be a Fortran/C/C++ code consisting of millions of lines!

▶ This is a fundamental and difficult computer science problem known as automatic differentiation.



Weather forecasting with data assimilation

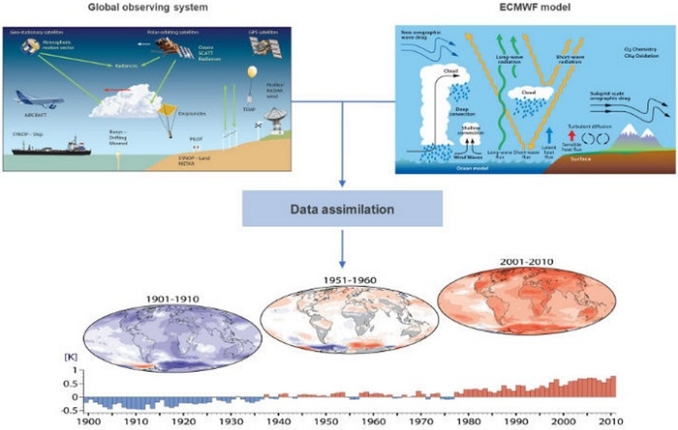
▶ Constant progress due to *more* observations, better and *finer* models, and the improvements of *data assimilation techniques*, the silent revolution [Magnusson et al. 2013; Bauer et al. 2015].



An incredibly useful product of NWP: Reanalysis

▶ Using data assimilation methods with past observations and state-of-the-art models allows to *reconstruct the weather* decades in the past.
⟶ Critically useful for all the geosciences and climate research community, the insurance companies, energy production companies (nuclear, wind, solar), transport/carriers, etc.

▶ ERA-5 from the ECWMF is currently the most accurate reanalysis for the synoptical scale (1979-2018) [Hersbach et al. 2020].



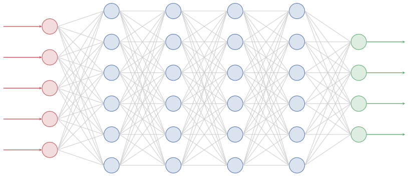
Outline

Learning data-driven models from dense and perfect observations

▶ A typical (supervised) machine learning problem: given observations $\mathbf{y}_k$ of a system, derive a *surrogate model* of that system from the loss function:

$$\mathcal{J}(\mathbf{p}) = \sum_{k=1}^{K} \left\| \mathbf{y}_k - M\left(\mathbf{p}, \mathbf{y}_{k-1}\right) \right\|^2$$

▶ The surrogate model to be learned M depends on a *set of coefficients* $\mathbf{p}$ (*e.g.*, the weights and biases of a neural network (NN)).



▶ This requires *dense* and*perfect* observations of the physical system.

▶ In the geosciences, observations are usually sparse and noisy: we need data assimilation!

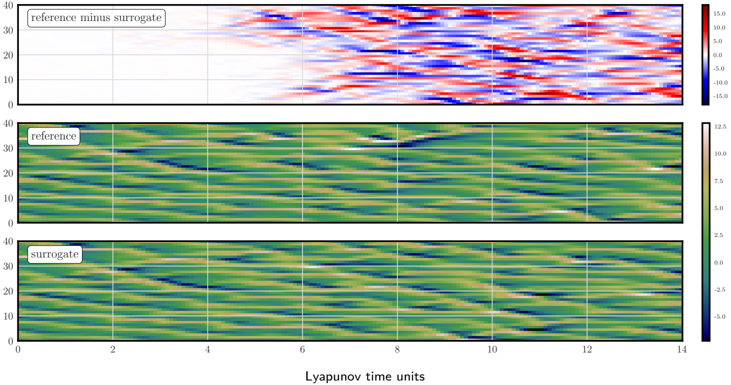
Example of learning the dynamics of toy models – 1

▶ Inferring the dynamics from dense & noiseless observations of an almost-identifiable model

The Lorenz 96 model (40 variables)

$$\frac{\mathrm{d}x_n}{\mathrm{d}t} = (x_{n+1} - x_{n-2})x_{n-1} - x_n + F,$$

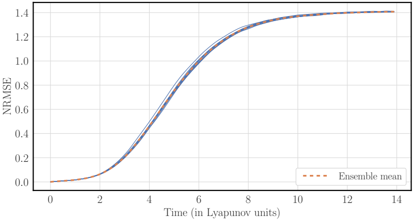
Surrogate model based on an RK2 scheme.

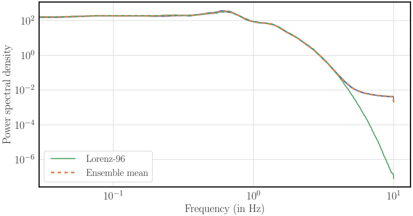


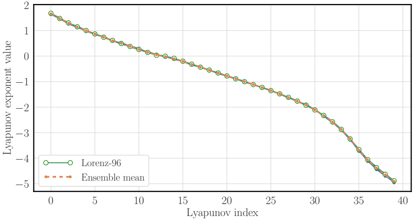
Example of learning the dynamics of toy models – 2

- Very good reconstruction of the long-term properties of the model (L96 model).

  - Approximate scheme
  - Fully observed
  - Significantly noisy observations $\mathbf{R} = \mathbf{I}$
  - Long window $K = 5000$, $\Delta t = 0.05$
  - EnKS with $L = 4$
  - 30 EM iterations





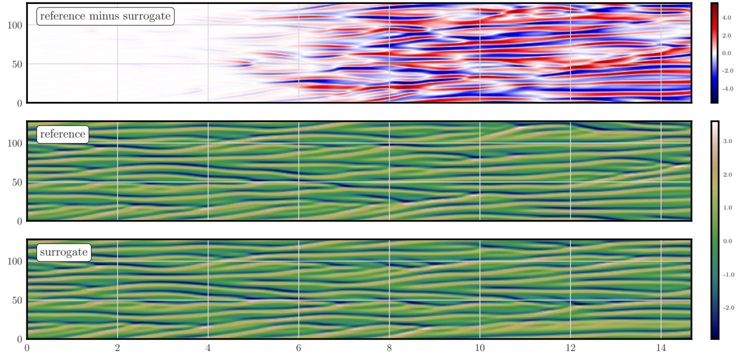


Example of learning the dynamics of toy models – 3

▶ Inferring the dynamics from dense & noiseless observations of a non-identifiable model

The Kuramoto-Sivashinski (KS) model (continuous, discr. into 128 var.).

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4}.$$



Lyapunov time units

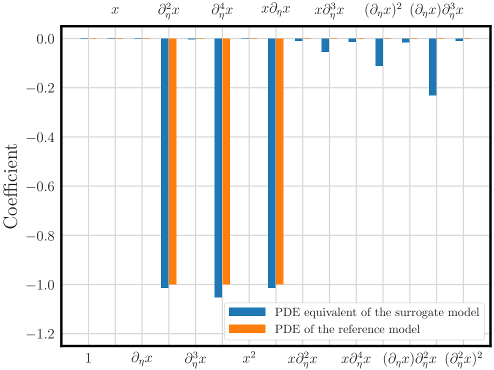
Example of learning the dynamics of toy models – 4

▶ Inferring the dynamics from dense & noiseless observations of a non-identifiable model

The Kuramoto-Sivashinski (KS) model (continuous, discr. into 128 var.).

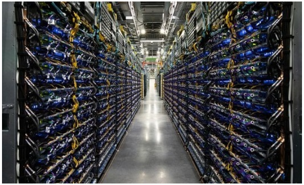
$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^4 u}{\partial x^4}.$$

▶ Explainable AI:



Why does AI surrogate models matter?

▶ AI/ML models run on CPU, GPU and TPU and can hence be very fast.



▶ Thanks to deep learning, new sparse representations of data that yield better, systematic and numerically affordable optimisations.

▶ AI models coded with comprehensive and convenient deep learning libraries (Tensorflow/Keras, PyTorch/Lightening, Julia/Flux, etc.) powered by Google, Facebook, Apache, Nvidia, etc.

TensorFlow Keras flux PyTorch Lightning julia

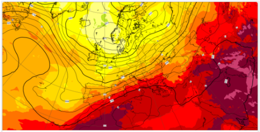
▶ Building codes on these libraries or languages *automatically provides the tangent linear and adjoint of the code*, practically solving one of the old issue if traditional data assimilation.

▶ AI/ML models should also be able *to generate and build large ensemble of forecasts*, hence with better uncertainly estimation and potential detection of extreme events.

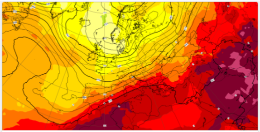
GAFA's data driven (synpotic scale) meteorological models

- Nvidia's FourcastNet [Pathak et al. 2022; Kurth et al. 2022]
- Google/Deepmind's Graphcast [Lam et al. 2022]
- Huawei's Pengu-Weather [Bi et al. 2023]
- ClimaX' Microsoft [Nguyen et al. 2023]
- FengWu [Chen et al. 2023]

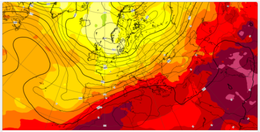
- Forecasts (geopotential at 500hPa, temperature at 850 hPa) from July 22, 00UTC to July 24, 15UTC



500 hPa geopotential height and 850 hPa temperature

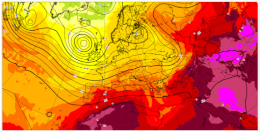


(FOURCAST machine learning model): 500 hPa geopotential height and 850 hPa temperature

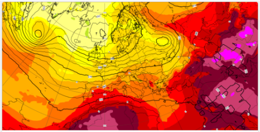


(PANGU machine learning model): 500 hPa geopotential height and 850 hPa temperature

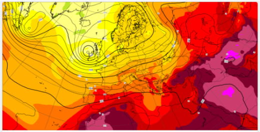
- Forecasts (geopotential at 500hPa, temperature at 850 hPa) from July 22, 00UTC to July 29, 12UTC



500 hPa geopotential height and 850 hPa temperature



(FOURCAST machine learning model): 500 hPa geopotential height and 850 hPa temperature



(PANGU machine learning model): 500 hPa geopotential height and 850 hPa temperature

- Most of them achieve performances similar to that of the ECMWF's IFS!
- Some of them become more accurate than the IFS in some regimes!?

Can we do the same? Data-driven meteorological model from ERA-5 reanalysis

▶ True model: A selection of ERA-5 fields in 1979-2018 at 5.625°.[a]

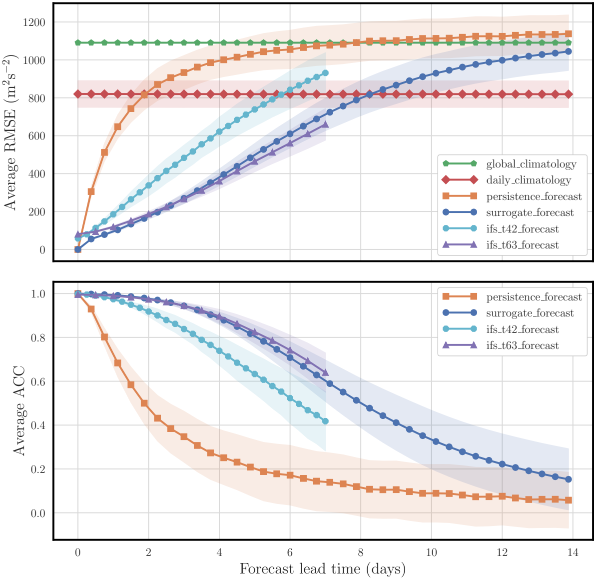
▶ DL model: Residual NN at the same resolution.

▶ Forecast skill score of the geopotential at 500hPa as a function of the forecast lead time.[b]

▶ η has also successfully been tested with DA.

[a] [Rasp et al. 2020]

[b] [Bocquet et al. 2023]



What are the limitations of these AI weather models?

- What about their *deep neural architectures*?

Model	Variables	Levels	Resolution	Architecture	Learned from
FourCastNet	20	5	variable	Vision transformers	ERA-5
PanguWeather	4	13(+4)	variable	Vision transformers	ERA-5
GraphCast	6	7(+5)	variable	Graph neural network	ERA-5
ClimaX	6	7(+3)	variable	Vision transformers	ERA-5
FenGu	5	37(+4)	variable	Vision transformers	ERA-5

Their architecture might *not be as crucial* as their authors sometimes claim (idiosyncrasy of the ML community?).[1]

- What are their limitations?
  - Only a few of these research teams made their code public (NVIDIA, Huawei, & Google over the summer'23) for now,
  - The models' outperformance vs the IFS is questionable because of their *smoothing effect vs a high-resolution IFS*, but they recently became aware of the caveat [Rasp et al. 2023],
  - Most importantly *they all rely on the ERA-5* (hence the IFS, data assimilation, decades of expertise acquired by the European meteorologists).

- How can they/we do genuinely better than the IFS?

$\longrightarrow$ Learn from the massive (sparse and noisy) observations!

[1] [Bocquet 2023; Lguensat 2023]

Outline

Machine learning for the geosciences with sparse and noisy observations

- A rigorous *Bayesian* formalism for this problem[2] yields the cost function:

$$\mathcal{J}(\mathbf{p}, \mathbf{x}_{0:K}) = \frac{1}{2}\sum_{k=0}^{K} \|\mathbf{y}_k - H(\mathbf{x}_k)\|^2_{\mathbf{R}_k^{-1}} + \frac{1}{2}\sum_{k=1}^{K} \|\mathbf{x}_k - M(\mathbf{p}, \mathbf{x}_{k-1})\|^2_{\mathbf{Q}_k^{-1}} + \dots$$

- This resembles a typical *weak-constraint 4D-Var* cost function!

- The solution is a *state trajectory*, and a *stochastic model*:

$$\mathbf{x}_k = M(\mathbf{p}, \mathbf{x}_{k-1}) + \sqrt{\mathbf{Q}_k \mathrm{d}t} \cdot \boldsymbol{\epsilon}_k .$$

- Machine learning limit

If the physical system is fully and directly observed, i.e. $\mathbf{H}_k \equiv \mathbf{I}$, and if the observation errors tend to zero, i.e. $\mathbf{R}_k \to \mathbf{0}$, then the observation term in the cost function is completely frozen and imposes that $\mathbf{x}_k \simeq \mathbf{y}_k$, so that, in this limit, $\mathcal{J}(\mathbf{p}, \mathbf{x}_{0:K})$ becomes

$$\mathcal{J}(\mathbf{p}) = \frac{1}{2}\sum_{k=0}^{K} \|\mathbf{y}_k - M(\mathbf{p}, \mathbf{y}_{k-1})\|^2_{\mathbf{Q}_k^{-1}} - \ln p(\mathbf{y}_0, \mathbf{p}).$$

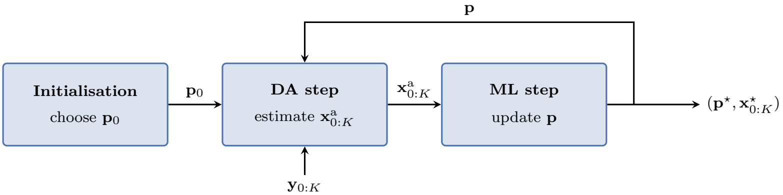
[2] [Bocquet et al. 2019; Bocquet et al. 2020; Brajard et al. 2020] in the wake of [Hsieh et al. 1998; Abarbanel et al. 2018]

Machine learning for the geosciences with sparse and noisy observations

- We need to minimise this cost function on both states and parameters:

$$\mathcal{J}(\mathbf{x}_{0:K}, \mathbf{p}) = \frac{1}{2}\sum_{k=0}^{K} \left\| \mathbf{y}_k - H(\mathbf{x}_k) \right\|^2_{\mathbf{R}_k^{-1}} + \frac{1}{2}\sum_{k=1}^{K} \left\| \mathbf{x}_k - M(\mathbf{p}, \mathbf{x}_{k-1}) \right\|^2_{\mathbf{Q}_k^{-1}} + \dots .$$

- *DA* is used to estimate the state and then *AI/ML* is used to estimate the model (*coordinate descent*):[3]



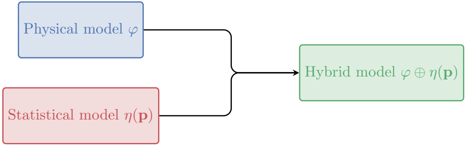
DA: 4D-Var, WC 4D-Var, EnKS, IEnKS, etc. and *ML*: neural networks.

This DA standpoint is remarkable as it allows for *noisy and partial observations* of the physical system.

[3] [Bocquet et al. 2020; Brajard et al. 2020]

Hybrid physical/AI model

- The *hybrid* physical/statistical model is a combination of the proxy φ–model and a NN η–model:
  - Even though geophysical models are not perfect, they are sometimes already quite good (especially in NWP)!
  - Instead of building a surrogate model from scratch, we use the DA-ML framework to build a *hybrid* surrogate model, with a physical part and a statistical part:[4]



  - In practice, the statistical part is trained to learn the *error* of the physical model.
  - In general, it is easier to train a correction model than a full model: we can use *smaller NNs* and *less training data*.

[4] [Farchi et al. 2021b; Brajard et al. 2021; Farchi et al. 2021a; Farchi et al. 2023].

Example of hybrid modelling with a deep learning subgrid parametrisation

- Marshall-Molteni[5] 3-layer intermediate QG model: Learning subgrid scale parametrisation at *loop order–1* to perform more accurate forecasts at low resolution (LR) from high resolution simulations (HR).[6]

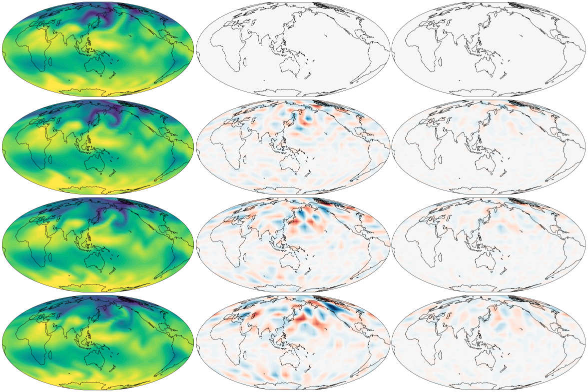
Truth φ_{HR} | Error with φ_{LR} | Error with $\varphi_{\mathrm{LR}} \oplus \eta_{\mathrm{HR}\to\mathrm{LR}}$

Time $t = 0$

$t = 24$h

$t = 48$h

$t = 72$h



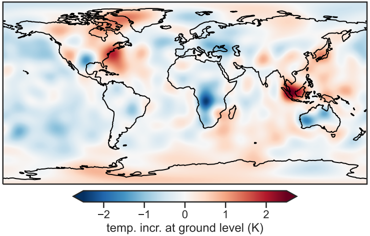
- $\varphi_{\mathrm{LR}} \oplus \eta_{\mathrm{HR}\to\mathrm{LR}}$ has also successfully been tested with DA.

[5] [Marshall et al. 1993]

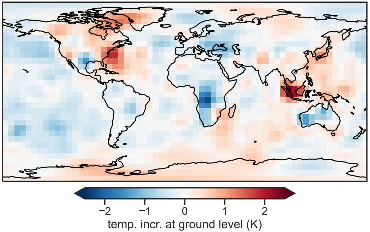
[6] [Malartic et al. 2022]

Experiments with the IFS

- We want to develop a *model error correction* for the operational IFS [Farchi et al. 2023].
- Offline experiments rely on preliminary work by [Bonavita et al. 2020], using the *operational analyses* produced by ECMWF between 2017 and 2020.
- Our NN is trained to predict the *analysis increments*, which are available every 12 hours.
- Focus on *large-scale model errors*: we use the data at a low spectral resolution (T21), interpolated in Gaussian grid with 32×63 nodes (moderate oversampling).

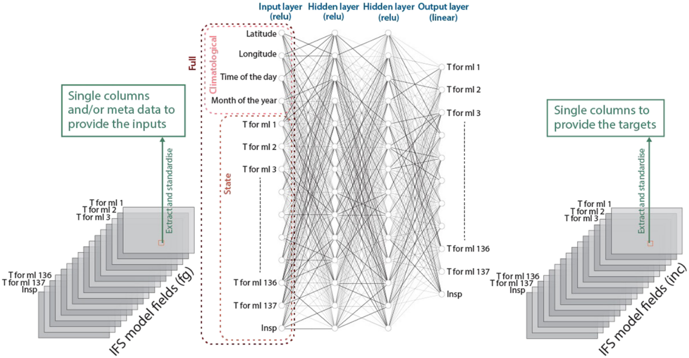


$\rightarrow$



Neural network architecture

▶ *Column-based NN* for temperature and logarithm of surface pressure:

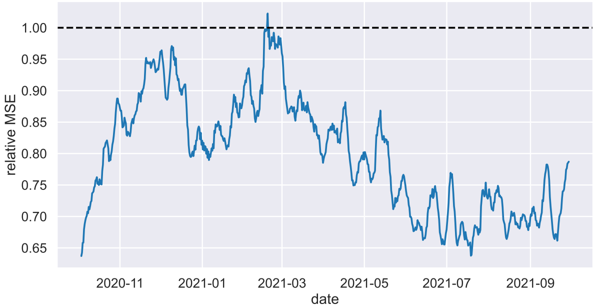


▶ In the architecture above, the number of parameters is small ($\sim 7 \times 10^4$) compared to the dimension of the control vector.

▶ Currently tested architecture include extra predictors such as sin/cos of lat, lon, time_of_day and day_of_year. The number of parameters is still small (up to $\sim \times 10^7$) compared to the dimension of the control vector.

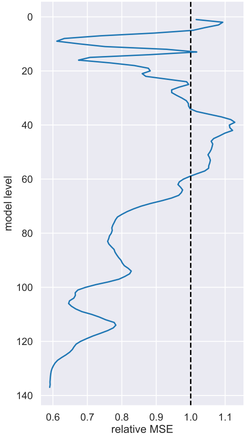
Offline performance of the neural network for tlnsp

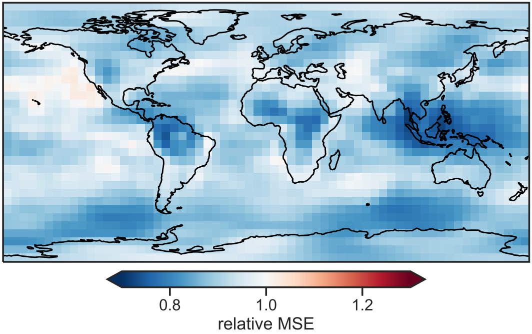
- Relative MSE (normalised by the MSE of the zero prediction) over the validation data.
- Overall, the NN predicts approx. *20% of the analysis increments*.
- The increments are more predictable in summer than in the winter.



Offline performance of the neural network for tlnsp

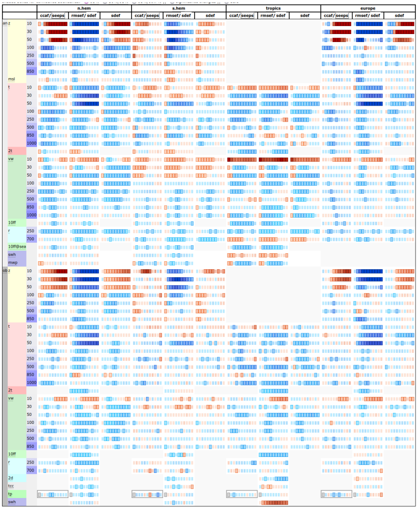
- Vertical profile and map of MSE over the validation data.
- The lower levels are in general more predictable.





First set of online experiments with the IFS

- The trained NNs are inserted into the IFS (cycle 48R1) and *trained online* with our NN 4D-Var.
- *Scorecard* of NN 4D-Var vs WC 4D-Var, for a three-month experiment in winter 2020/2021.



- The forecasts are compared to observations and the operational analysis.
- Significant improvements for the *geopotential*, especially in the southern hemisphere.
- Degradation of the winds in the tropics.

Outline

Conclusions–1

- Main messages:
  - State-of-the-art AI models of the weather seem as accurate as the IFS, albeit much faster!
  - However, they are all learned from the ERA-5 dataset generated from *observations*, *data assimilation*, and the *IFS*!
  - Hence, models/modellers are still dearly needed . . . for now.
  - But one may envision building an AI data-driven model *purely from observation*.
  - Because these data are sparse and noisy, *combining DA+ML* will be needed!

Conclusions–2

- Perspectives:
  - AI/ML techniques can be applied to other geofluid physics such as sea-ice surrogate modelling: Schmidt Futures/VESRI/SASIP project[7],
  - Learning *generative models* for stochastic-like dynamical behaviour (SASIP project),
  - Other loss function criteria based on reliability and resolution for ensemble forecasting (ERA-5).

- Theoretical and applicative challenge:
  - Develop a similar approach but for *convective-scale meteorology*,
  - Towards more accurate prediction of *natural disasters* (flash floods, tornadoes, hailstorm) at finer scale,
  - Data assimilation+AI+strongly nonlinear dynamics+no ERA-5-like dataset = hard but rewarding problem.

ERC Synergy proposal **DAMLEAP** selected for step 3
[KU, ENPC, INP Toulouse, Uni. Bologna]

[7] [Finn et al. 2023; Durand et al. 2023]

References I

[1] H. D. I. Abarbanel, P. J. Rozdeba, and S. Shirman. "Machine Learning: Deepest Learning as Statistical Data Assimilation Problems". In: *Neural Computation* 30 (2018), pp. 2025–2055.

[2] P. Bauer, A. Thorpe, and G. Brunet. "The quiet revolution of numerical weather prediction". In: *Nature* 525 (2015), pp. 47–55.

[3] K. Bi et al. "Accurate medium-range global weather forecasting with 3D neural networks". In: *Nature* 619 (2023), pp. 533–538.

[4] M. Bocquet. "Surrogate modelling for the climate sciences dynamics with machine learning and data assimilation". In: *Front. Appl. Math. Stat.* 9 (2023).

[5] M. Bocquet et al. "Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization". In: *Foundations of Data Science* 2 (2020), pp. 55–80.

[6] M. Bocquet et al. "Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models". In: *Nonlin. Processes Geophys.* 26 (2019), pp. 143–162.

[7] M. Bocquet et al. "Higher-order correction to deep-learning surrogate model of meteorology from data assimilation". In: *In preparation* (2023).

[8] M. Bonavita and P. Laloyaux. "Machine Learning for Model Error Inference and Correction". In: *J. Adv. Model. Earth Syst.* 12 (2020), e2020MS002232.

[9] J. Brajard et al. "Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model". In: *J. Comput. Sci.* 44 (2020), p. 101171.

[10] J. Brajard et al. "Combining data assimilation and machine learning to infer unresolved scale parametrisation". In: *Phil. Trans. R. Soc. A* 379 (2021), p. 20200086.

[11] K. Chen et al. "FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Day". In: *arXiv* (2023).

[12] C. Durand et al. "Data-driven surrogate modeling of high-resolution sea-ice thickness in the Arctic". In: *EGUsphere* 2023 (2023), na–na.

[13] A. Farchi et al. "A comparison of combined data assimilation and machine learning methods for offline and online model error correction". In: *J. Comput. Sci.* 55 (2021), p. 101468.

[14] A. Farchi et al. "Online Model Error Correction With Neural Networks in the Incremental 4D-Var Framework". In: *J. Adv. Model. Earth Syst.* 15 (2023), e2022MS003474.

[15] A. Farchi et al. "Using machine learning to correct model error in data assimilation and forecast applications". In: *Q. J. R. Meteorol. Soc.* 147 (2021), pp. 3067–3084.

References II

[16] T. S. Finn et al. "Deep learning of subgrid-scale parametrisations for short-term forecasting of sea-ice dynamics with a Maxwell elasto-brittle rheology". In: *The Cryosphere* 17 (2023), pp. 2965–2991.

[17] H. Hersbach et al. "The ERA5 global reanalysis". In: *Q. J. R. Meteorol. Soc.* 146 (2020), pp. 1999–2049.

[18] W. W. Hsieh and B. Tang. "Applying Neural Network Models to Prediction and Data Analysis in Meteorology and Oceanography". In: *Bull. Amer. Meteor. Soc.* 79 (1998), pp. 1855–1870.

[19] T. Janjić et al. "On the representation error in data assimilation". In: *Q. J. R. Meteorol. Soc.* 144 (2018), pp. 1257–1278.

[20] T. Kurth et al. "FourCastNet: Accelerating Global High-Resolution Weather Forecasting using Adaptive Fourier Neural Operators". In: *arXiv* (2022).

[21] R. Lam et al. "GraphCast: Learning skillful medium-range global weather forecasting". In: *arXiv* (2022).

[22] R. Lguensat. "Les nouveaux modèles de prévision météorologique basés sur l'intelligence artificielle : opportunité ou menace ?" In: *La Météorologie* 121 (2023), pp. 11–15.

[23] E. N. Lorenz. "Deterministic nonperiodic flow". In: *J. Atmos. Sci.* 20 (1963), pp. 130–141.

[24] L. Magnusson and E. Källén. "Factors influencing skill improvements in the ECMWF forecasting system". In: *Mon. Wea. Rev.* 141 (2013), pp. 3142–3153.

[25] Q. Malartic et al. "Data assimilation and machine learning subgrid parametrisation in an advanced quasi-geostrophic model". In: (2022).

[26] J. Marshall and F. Molteni. "Toward a Dynamical Understanding of Planetary-Scale Flow Regimes". In: *J. Atmos. Sci.* 50 (1993), pp. 1792–1818.

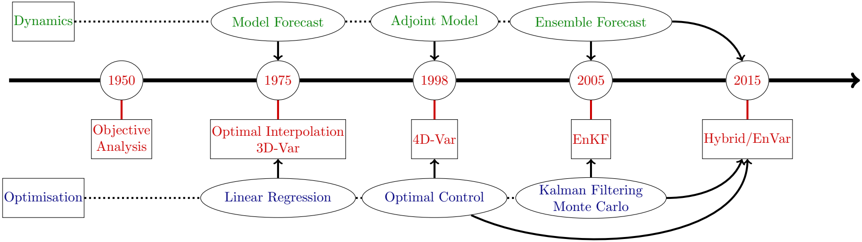
[27] T. Nguyen et al. "ClimaX: A foundation model for weather and climate". In: *arXiv* (2023).

[28] J. Pathak et al. "FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators". In: *arXiv* (2022).

[29] S. Rasp et al. *WeatherBench 2: A benchmark for the next generation of data-driven global weather models*. 2023. arXiv: 2308.15560 [physics.ao-ph].

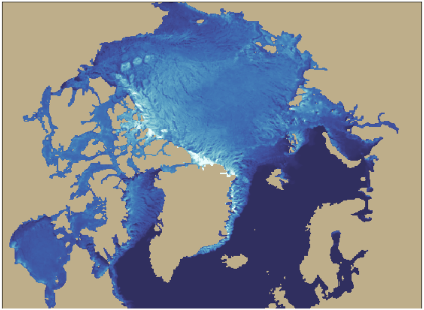
[30] S. Rasp et al. "WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting". In: *J. Adv. Model. Earth Syst.* 12 (2020), e2020MS002203.

- Introduction of mathematical methods in operational numerical weather prediction:



- Using increasingly complex mathematical methods and increasingly resolved high-dimensional models.

2018-01-01 03:00:00



Complex dynamics in sea-ice:

- Multifractality
- Anisotropy
- Stochasticity
- (mildly) chaotic

Two NN types:

- Unet (multiscale approach)
- ResNet (residual NN)

With partial convolutions and SE blocks.

Inputs: sea-ice thickness from NeXtSIM + ERA5
Forcings: 10m air velocity, 2m air temperature and sea surface temperature
For several past timesteps
Outputs: 12h sea-ice thickness evolution

SCHMIDT FUTURES