Harnessing machine learning for high resolution data assimilation

Tomas Landelius presenting work done by others

Workshop on data assimilation: initial conditions and beyond.
Part of ECMWF's 50th anniversary celebrations. Bonn, Germany 9-10 April 2025

Data assimilation at km to hectometric scale - a challenge

SCIENTIFIC CHALLENGES OF CONVECTIVE-SCALE NUMERICAL WEATHER PREDICTION

Jun-Ichi Yano, Michał Z. Ziemiański, Mike Cullen, Piet Termonia, Jeanette Onvlee, Lisa Bengtsson, Alberto Carrassi, Richard Davy, Anna Deluca, Suzanne L. Gray, Victor Homar, Martin Köhler, Simon Krichak, Silas Michaelides, Vaughan T. J. Phillips, Pedro M. M. Soares, and Andrzej A. Wyszogrodzki

Numerical weather prediction models are increasing in resolution and becoming capable of explicitly representing individual convective storms, but we do not yet know if it is the improved resolution that is leading to better forecasts.

Quarterly Journal of the Royal Meteorological Society

RMetS

REVIEW ARTICLE | Open Access

The hectometric modelling challenge: Gaps in the current state of the art and ways forward towards the implementation of 100-m scale weather and climate models

Humphrey W. Lean, Natalie E. Theeuwes, Michael Baldauf, Jan Barkmeijer, Geoffrey Bessardon, Lewis Blunn, Jelena Bojarova, Ian A. Boutle, Peter A. Clark, Matthias Demuzere, Peter Dueben, Inger-Lise Frogner, Siebren de Haan, Dawn Harrison, Chiel van Heerwaarden, Rachel Honnert, Adrian Lock, Chiara Marsigli, Valéry Masson, Anne Mccabe, Maarten van Reeuwijk, Nigel Roberts, Pier Siebesma, Petra Smolíková, Xiaohua Yang ... See fewer authors

First published: 07 October 2024 | https://doi.org/10.1002/qj.4858 | Citations: 3

SMHI

Rymdstyrelsen
Swedish National Space Agency

Workshop
on
Perspectives of data assimilation on hecto-metric scales

In Memoriam of Nils Gustafsson

Tofta, Gotland, Sweden, 10- 12 September, 2024

Loïk Berre (remote), Jelena Bojarova, Pau Escribà, Elias Holm,
Elias Holm, Heikki Järvinen (remote), Tomas Landelius, Magnus Lindskog,
Kristian Mogensen, Patrick Samuelsson, Michael Tjernström, Ole Vignes, Tomas
Wilhelmsson, Xiaohua Yang

Identified obstacles (worse at hectometric scale)

- Breakdown of linearity and Gaussian assumptions
- Spin-up problems due to imbalances in the analysis
- Proper localization of ensemble based error matrices
- Need for coupled surface-atmosphere assimilation
- Treatment of correlated observation errors
- Non-convex optimization and local minima
- Assimilation on all resolved scales
- …

https://www.vecteezy.com/vector-art/6652439

DA challenges and some ML remedies?

High dimensional state space

- Latent space DA

Nonlinear dynamics (clouds, snow, …)

- Auto differentiation
- Monte-Carlo filtering
- Koopman operators

Non-Gaussian errors

- Transport methods; Diffusion models, Normalizing flows, …

ML and Variational DA

ML or NWP model with gradients for 4D-Var

$$J(x_0,\eta) = (x_0 - x_b)^T B^{-1}(x_0 - x_b) + \sum_{i=0}^{N}(y_i - H(x_i))^T R^{-1}(y_i - H(x_i)) + (\eta - \eta_b)^T Q^{-1}(\eta - \eta_b)$$

$$x_i = M(x_{i-1}) + \eta$$

When M is a differentiable model we can calculate the gradients $\dfrac{\partial M(x)}{\partial x}$

and find the optimum $\operatorname{argmin}_{(x_0,\eta)}\ J(x_0,\eta)$

EXPLORING THE USE OF MACHINE LEARNING WEATHER MODELS IN DATA ASSIMILATION

A PREPRINT

Xiaoxu Tian*
NOAA/NCEP/EMC
College Park, MD 20740

Daniel Holdaway
NOAA/NCEP/EMC
College Park, MD 20740

Daryl Kleist
NOAA/NCEP/EMC
College Park, MD 20740

Unphysical TL from ML -
does it matter?

Geophysical Research Letters

Research Letter | Open Access

FuXi-En4DVar: An Assimilation System Based on Machine Learning Weather Forecasting Model Ensuring Physical Constraints

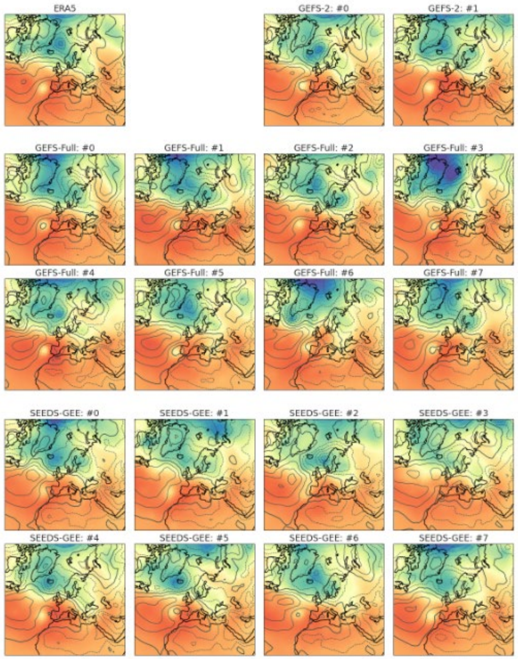
Yonghui Li, Wei Han, Hao Li, Wansuo Duan, Lei Chen, Xiaohui Zhong, Jincheng Wang, Yongzhu Liu, Xiuyu Sun

First published: 14 November 2024 | https://doi.org/10.1029/2024GL111136 | Citations: 2

Error matrix from MLWP ensemble or generative model

Given K samples $\mathcal{E}^K = (\boldsymbol{v}^1, \boldsymbol{v}^2, \cdots, \boldsymbol{v}^K)$ from $p(\boldsymbol{v})$, we construct an easy-to-sample *conditional* distribution

$$\hat{p}(\boldsymbol{v}) = p(\boldsymbol{v}; \mathcal{E}^K),$$



Also MLPP WP2 & DE 371 projects

SEEDS: Emulation of Weather Forecast Ensembles with Diffusion Models

Lizao Li
Google Research
lizaoli@google.com

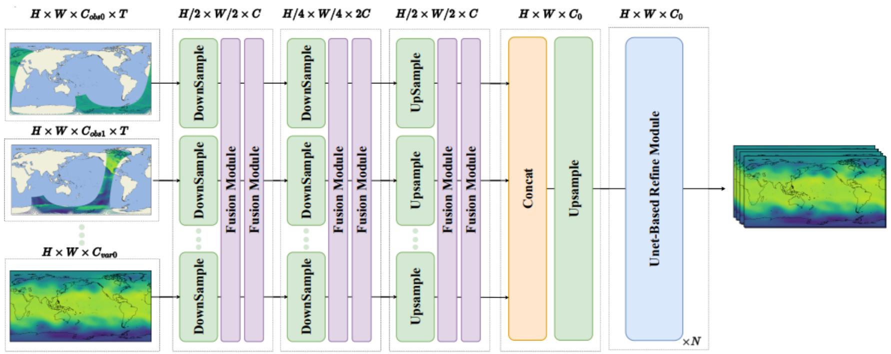
Robert Carver*
Google Research
carver@google.com

Ignacio Lopez-Gomez*
Google Research
ilopezgp@google.com

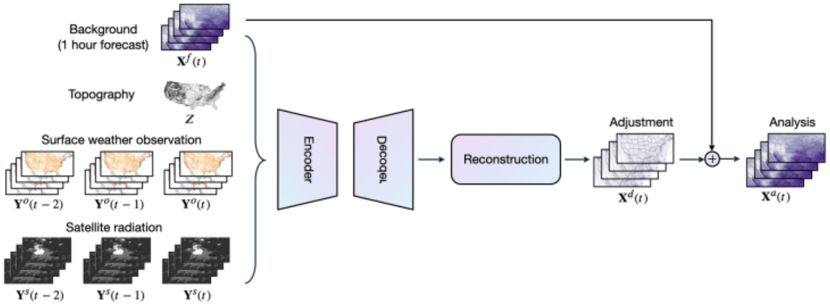
Fei Sha†
Google Research
fsha@google.com

John Anderson
Google Research
janders@google.com

Learning the analysis: other examples (none at km scale)



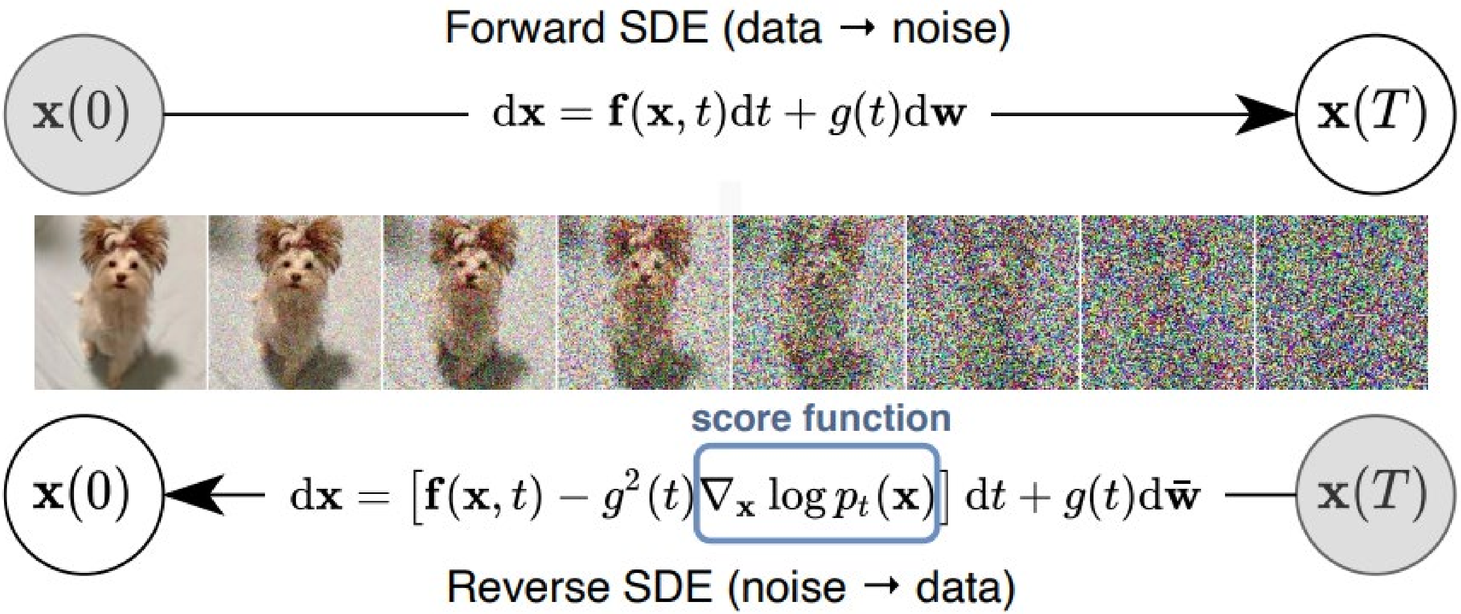
FuXi-DA,
https://arxiv.org/pdf/2408.05472



Microsoft ADAF,
arxiv.org/pdf/2411.16807

Probabilistic DA with generative models

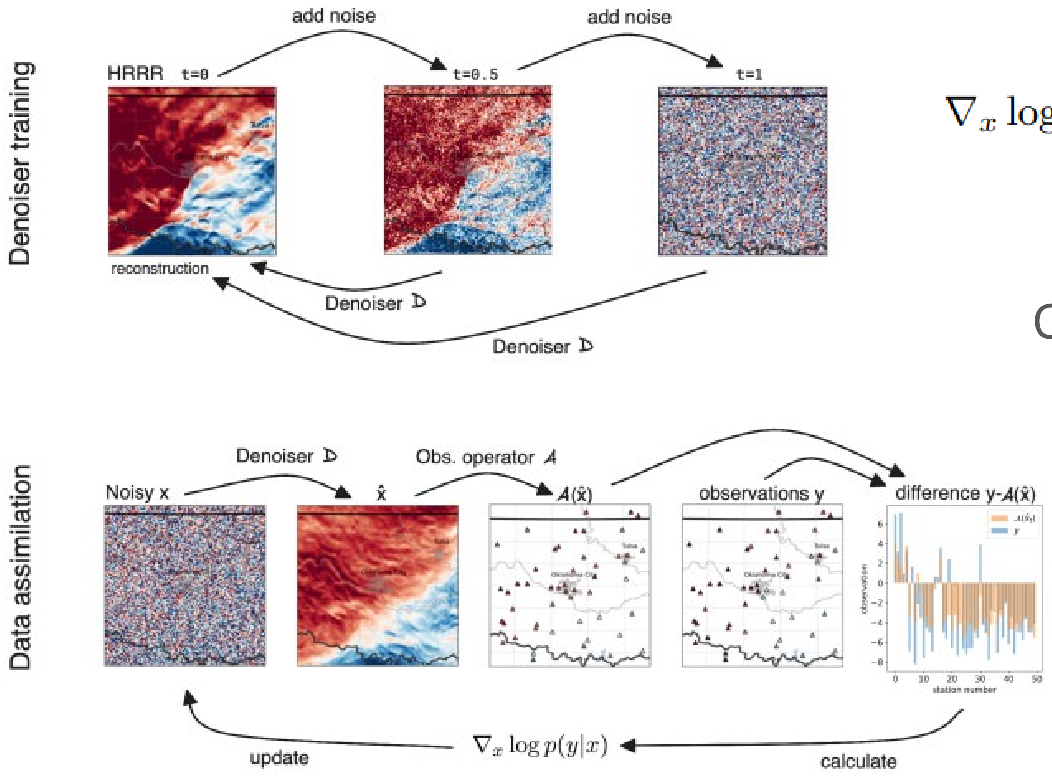
Score based diffusion (t is diffusion time)



$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_t \Big\{ \lambda(t) \mathbb{E}_{\mathbf{x}(0)} \mathbb{E}_{\mathbf{x}(t)|\mathbf{x}(0)} \Big[\left\| \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log p_{0t}(\mathbf{x}(t) \mid \mathbf{x}(0)) \right\|_2^2 \Big] \Big\}$$

Song et al., 2021 https://arxiv.org/pdf/2011.13456

Score based data assimilation



$$\nabla_x \log p(x(t)|y) = \nabla_x \log p(x(t)) + \nabla_x \log p(y|x(t))$$

$$p(y|x(t)) = \mathcal{N}\left(y|\mathcal{A}\left(\hat{x}\right), \Sigma_y(t)\right)$$

Challenge: only known for t=0

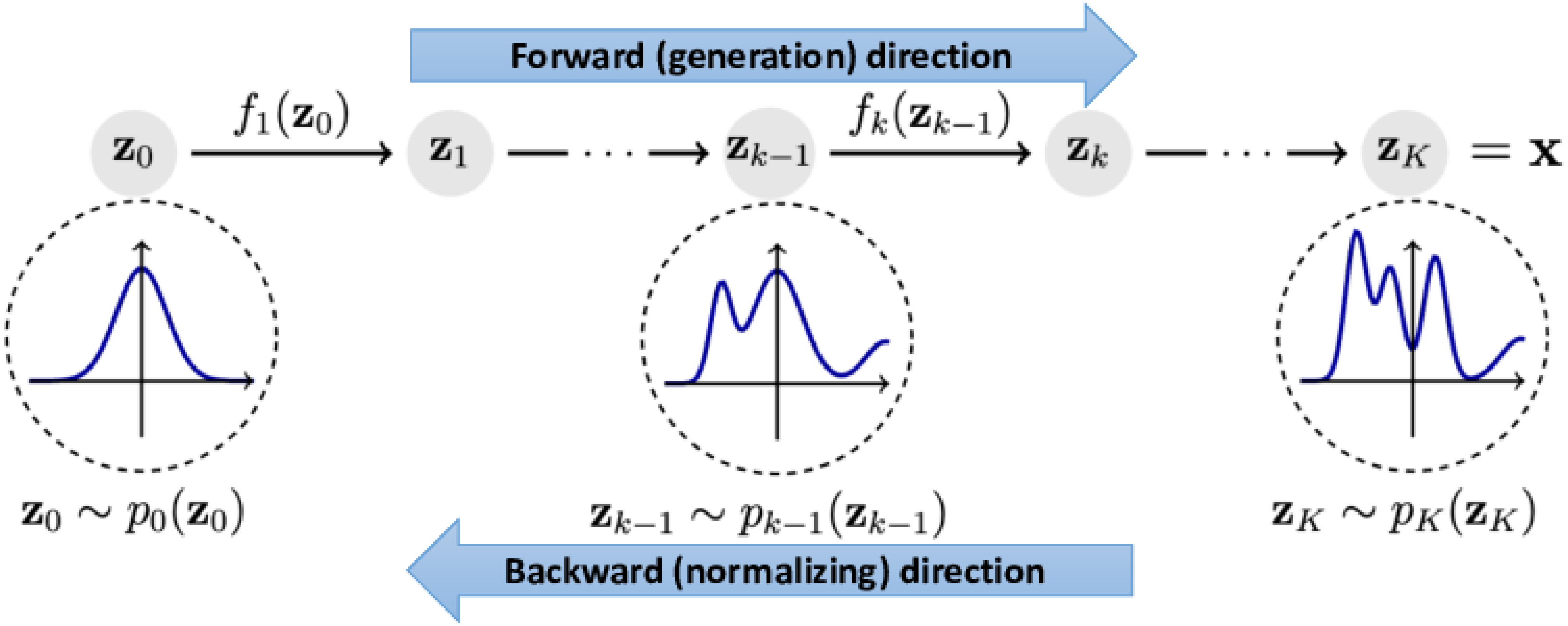
Generative Data Assimilation of Sparse Weather Station Observations at Kilometer Scales

Peter Manshausen[1,2], Yair Cohen[1], Jaideep Pathak[1], Mike Pritchard[1,3], Piyush Garg[1], Morteza Mardani[1], Karthik Kashinath[1], Simon Byrne[1], Noah Brenowitz[1]

[1]NVIDIA, Santa Clara, CA, USA
[2]University of Oxford, Oxford, UK
[3]University of California Irvine, Irvine, CA, USA

Data assimilation in a latent space

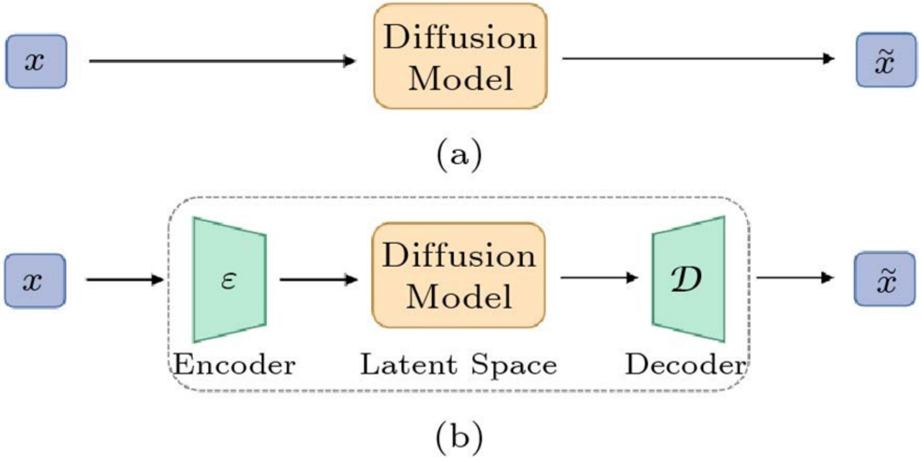
Gaussian errors via transport methods

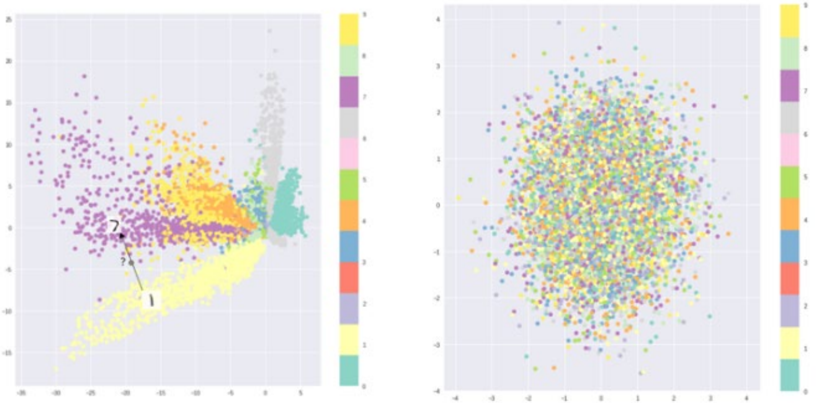


https://www.researchgate.net/figure/Normalizing-flow-distribution-transformation-20-The-forward-direction-transforms-a_fig1_373016632

Combine autoencoder with transport; normflow/diffusion

$$x(t+1) = f(x(t)) + e(t)$$





$$z(t+1) = g(z(t)) + n(t)$$

https://jcst.ict.ac.cn/supplement/1bab19cc-8f20-413f-86ec-84d7acdeb207

https://www.jeremyjordan.me/variational-autoencoders/

“Optimal” or other SMC in low dimensional latent space

$$\hat{p}^N(x_t|y_{1:t}) = \sum_{i=1}^{N} w_t^i \delta_{x_t^i}(x_t)$$

$$\tilde{w}_t = p(y_t|x_{t-1}^i) w_{t-1}^i$$

$$x_t = f(x_{t-1}) + v_t$$

$$x_t|x_{t-1} \sim \mathcal{N}(f(x_{t-1}), Q), \quad y_t|x_t \sim \mathcal{N}(Cx_t, R)$$

$$p(y_t|x_{t-1}) = \mathcal{N}(y_t|Cf(x_{t-1}), R + CQC^T)$$

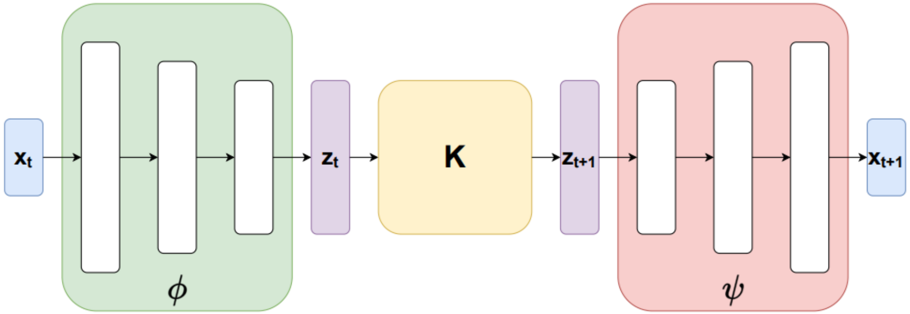
Koopman operator and EnKF in latent space (inf dim…)

$$\mathbf{x}_{t+1} = F(\mathbf{x}_t)$$

$$\mathcal{K}f(\mathbf{x}_t) \triangleq (f \circ F)(\mathbf{x}_t) = f(\mathbf{x}_{t+1})$$

Neural operators to model infinite dimensional Koopman operator

$$\mathcal{K}f(\mathbf{x}_t) = f(\mathbf{x}_{t+1}) = \sum_{1 \leq i \leq d} b_i \phi_i(\mathbf{x}_t)$$



TRANSACTIONS ON SIGNAL PROCESSING

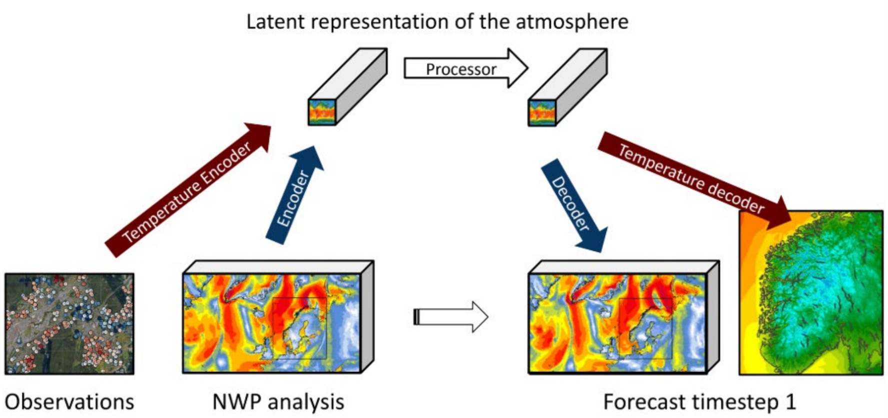
Neural Koopman prior for data assimilation

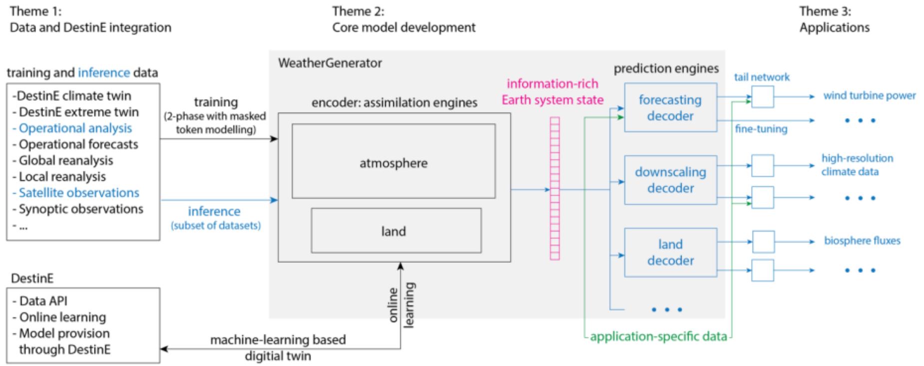
Anthony Frion, *Student Member, IEEE*, Lucas Drumetz, *Member, IEEE*, Mauro Dalla Mura, *Senior Member, IEEE*, Guillaume Tochon, Abdeldjalil Aïssa El Bey

https://arxiv.org/pdf/2309.05317

Augmenting existing models with observations

Ex: Anemoi multiple encoders and the WeatherGenerator





Adding extra observations via multiple encoders

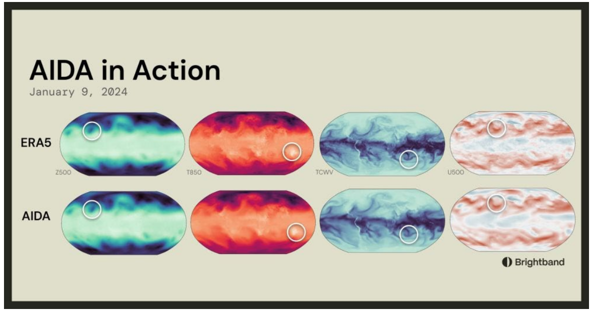
Mixing NWP and observations via representational learning

Met Norway Bris work in Anemoi

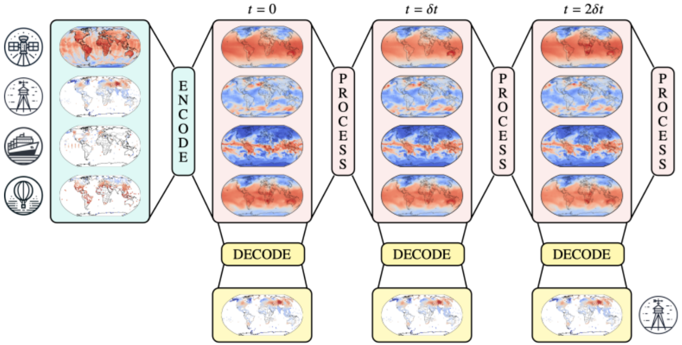
Direct observation forecasting
DA into “representation” space

Some current flavours

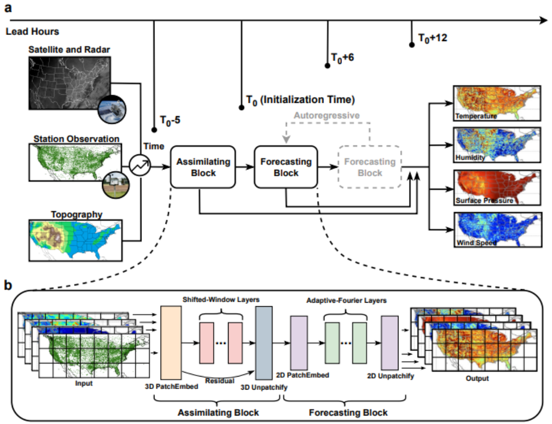
Brightband AIDA



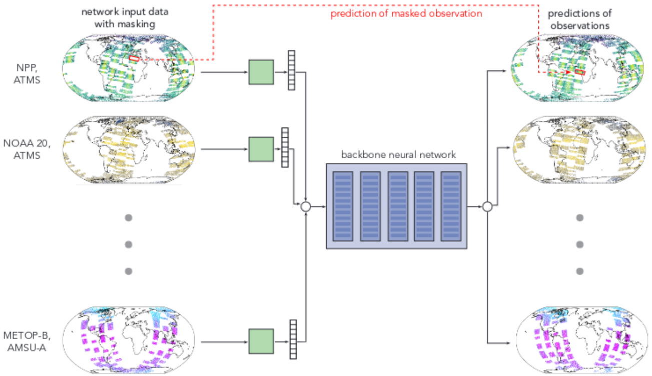
Aardvark



MS OMG-HD



ECMWF AI-DOP



Some are more “observation-only” than others…

How are these trained?

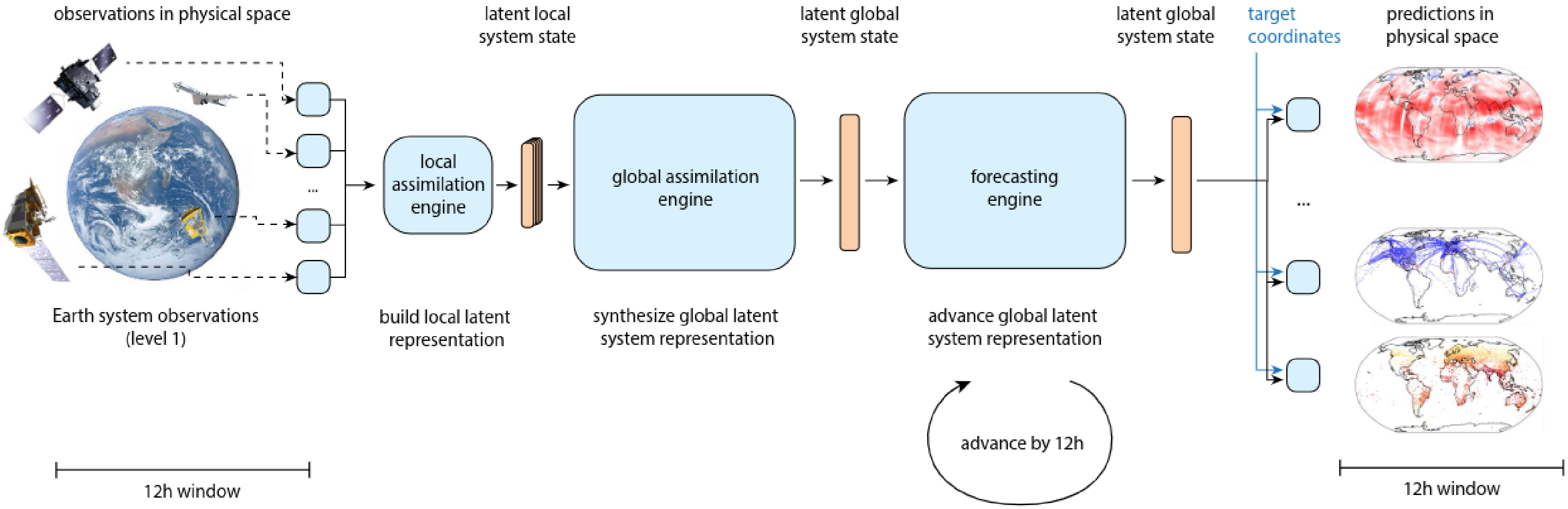
Brightband AIDA - No forecast. Presented at AMS in January. How?

MS OMG-HD: Assimilation (grid) - Forecast. Pretrain Assimilation with RTMA

Aardvark: Encode - Process (grid) - Decode. Pretrain Processor with ERA5

AI-DOP: Assimilation - internal rep (no grid) - Forecast. Self supervised training.

Representational “DA” - end of DA as we know it?

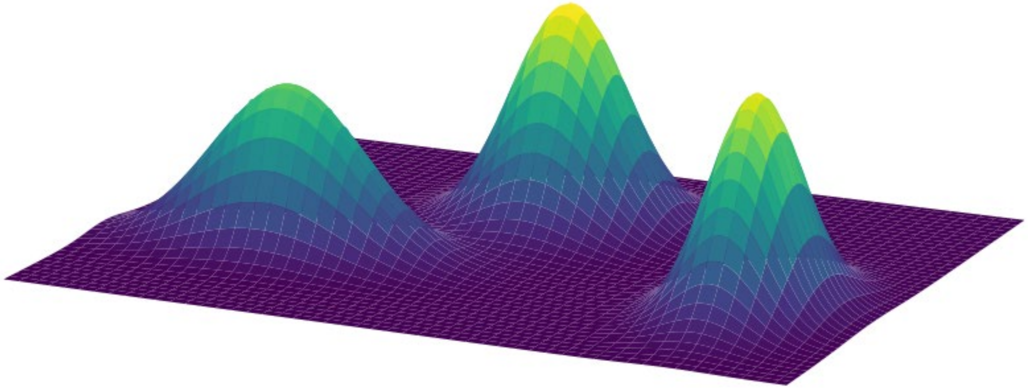


Beyond the horizon…

Liouville, Fokker-Planck and DA on quantum computers?

Model the evolution of a parameterized pdf?

https://homepages.inf.ed.ac.uk/
htang2/mini-asr/gmm/index.html



$$p(x(t), w(t)),\ \ w(t+1) = f(w(t))$$

Model both p and f with neural networks.
No need for Monte-Carlo. Can evaluate p(x) directly.
Relates to “information geometry” (https://arxiv.org/pdf/1808.08271)

Fokker-Planck, Liouville and Perron-Frobenius

$$\frac{dx(t)}{dt} = f(x(t)) + \eta$$

$$\frac{\partial}{\partial t}\rho(x,t) = -\nabla_x \cdot [f(x)\rho(x,t)] + \frac{D}{2}\nabla_x^2[\rho(x,t)]$$

$$\frac{\partial}{\partial t}\rho(x,t) = H\rho(x,t)$$

Modelling the evolution of the probability distribution

Schrödinger

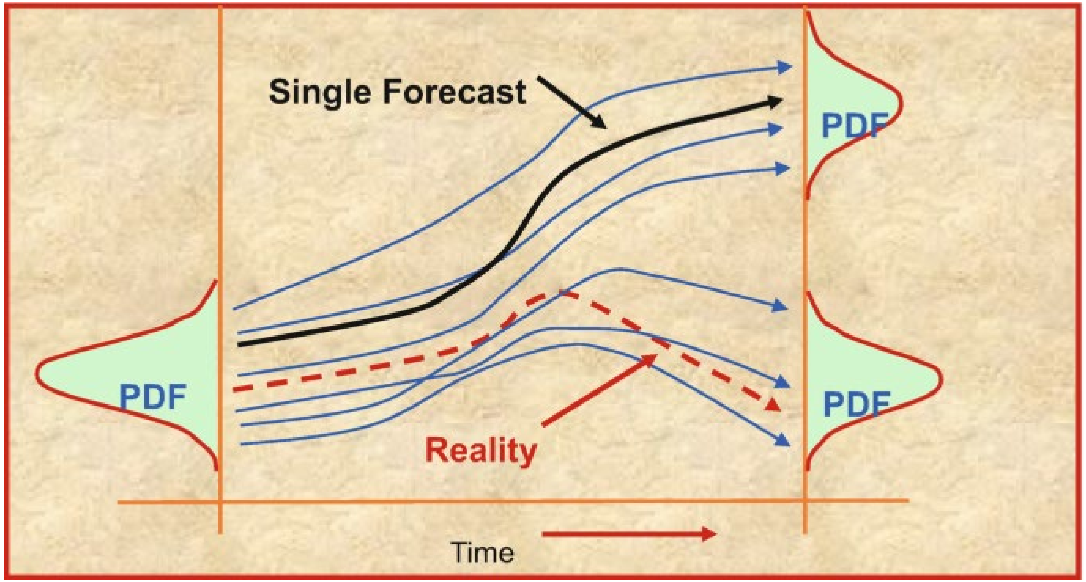
$$\frac{\delta}{\delta t}|\Psi\rangle = \frac{1}{i\hbar}H|\Psi\rangle$$

von Neumann

$$\frac{\delta\rho}{\delta t} = \frac{1}{i\hbar}[H,\rho]$$

Liouville

$$\frac{\delta\rho}{\delta t} = -\{\rho, H\}$$



https://media.springernature.com/lw685/springer-static/image/chp%3A10.1007%2F978-3-642-39925-1_13/MediaObjects/316607_1_En_13_Fig2_HTML.png

Simulate evolution on quantum computer

Algorithm: Quantum simulation

Inputs: (1) A Hamiltonian $H = \sum_k H_k$ acting on an N-dimensional system, where each H_k acts on a small subsystem of size independent of N, (2) an initial state $|\psi_0\rangle$, of the system at $t = 0$, (3) a positive, non-zero accuracy δ, and (3) a time t_f at which the evolved state is desired.

Outputs: A state $|\tilde{\psi}(t_f)\rangle$ such that $|\langle\tilde{\psi}(t_f)|e^{-iHt_f}|\psi_0\rangle|^2 \geq 1 - \delta$.

Runtime: $O(\text{poly}(1/\delta))$ operations.

Approximating Perron-Frobenius operator with FVM

PF op is a solution to the continuity equation.

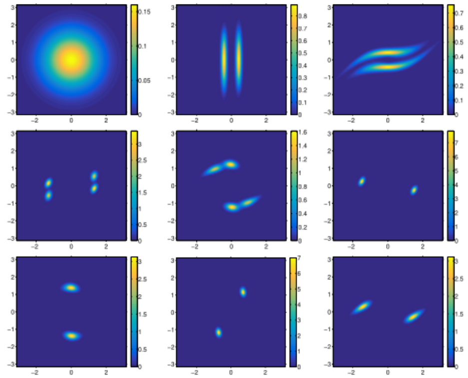
FVM methods preserve continuity
i.e. total probability.

Markov operator by satisfying
Courant-Friedrichs-Lewy (CFL) condition.

Solve linear equation system, e.g. with
HHL algorithm on a quantum computer.

Numerical approximation of the Frobenius-Perron operator using the finite volume method

RICHARD A. NORTON*,
Department of Mathematics and Statistics, University of Otago,
730 Cumberland Street, Dunedin, New Zealand,
*Corresponding author: richard.norton@otago.ac.nz
AND
COLIN FOX AND MALCOLM E. MORRISON
Department of Physics, University of Otago,
730 Cumberland Street, Dunedin, New Zealand.



Data assimilation with Perron-Frobenius operator

$$\rho(x_k|\mathbf{z}_{1:k}) = \frac{\rho(z_k|x_k)\rho(x_k|\mathbf{z}_{1:k-1})}{\rho(z_k|\mathbf{z}_{1:k-1})}$$

$$\rho(x_k|\mathbf{z}_{1:k-1}) = \int \rho(x_k|x_{k-1},\mathbf{z}_{1:k-1})\rho(x_{k-1}|\mathbf{z}_{1:k-1})\,\mathrm{d}x_{k-1},$$

$$\rho(z_k|\mathbf{z}_{1:k-1}) = \int \rho(z_k|x_k)\rho(x_k|\mathbf{z}_{1:k-1})\,\mathrm{d}x_k.$$

$$\rho(x_k|\mathbf{z}_{1:k-1}) = S(t_k - t_{k-1})\rho(x_{k-1}|\mathbf{z}_{1:k-1}).$$

Here S is Perron-Frobenius op.

Time to get back to reality…

Conclusions?

ML could offer remedies to some high-res DA challenges

Score based DA offers great flexibility and a probabilistic setting (watch out for Louppe et al.!)

Not much coupled multi-scale (time and space) DA so far

ML optimization often results in local minima - important?

DA as we know it is challenged by “direct observation prediction”

WP on quantum computers is the hot topic when ECMWF celebrates 100 years?

Congratulations to a new era of successful collaborations!

ECMWF ML Pilot Project

Destination Earth ML people

Anemoi