Data assimilation for Earth system models; challenges and ML remedies

Tomas Landelius and colleagues

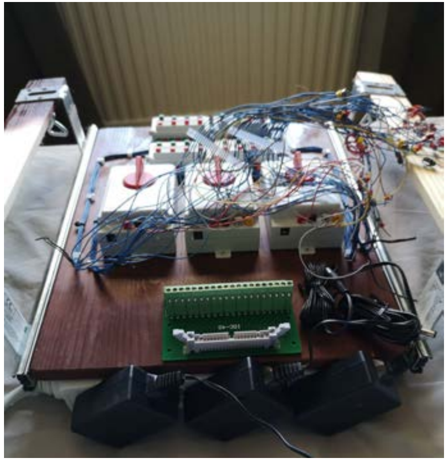
Outline

- My hobbies and DA background
- DA in operational NWP
- Machine learning and Variational methods
- Probabilistic DA with generative models
- Sequential Monte Carlo or KF in latent space
- Foundation models and DA
- Liouville, Fokker-Planck and DA on quantum computers?
- ML NWP projects at SMHI

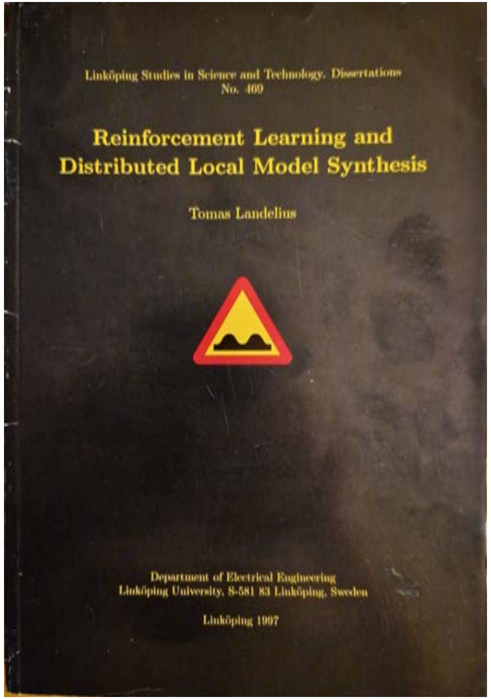
$ whoami

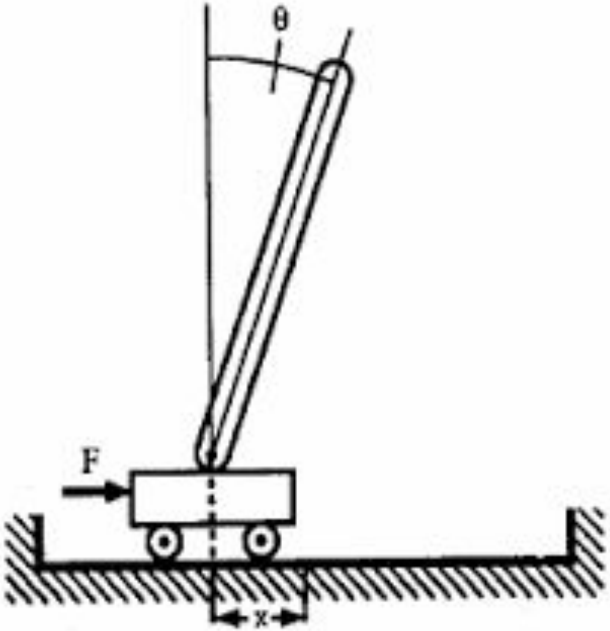
Total Immersion™

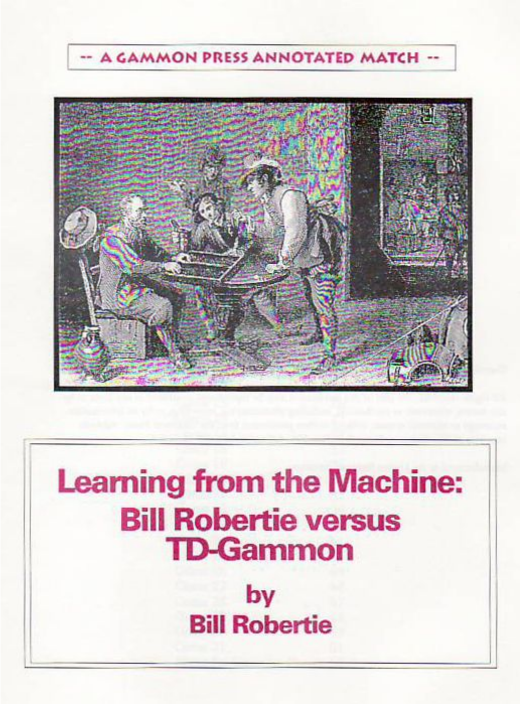
FLORAVÄKTARNA



PhD from LiU (1997)





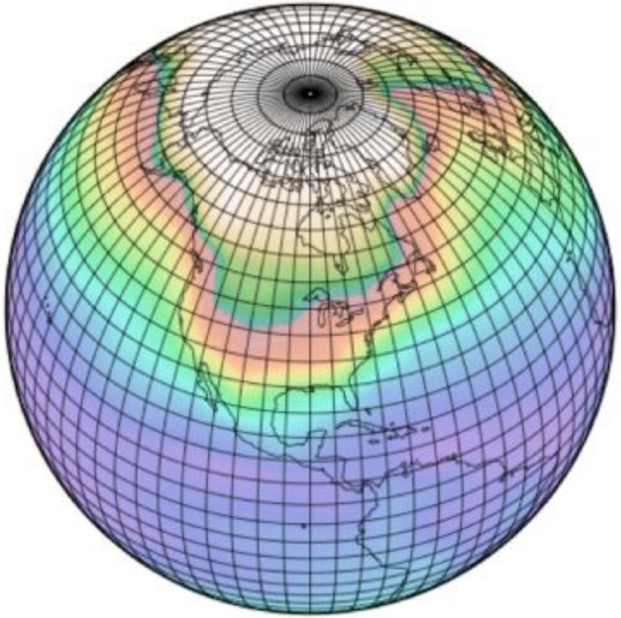


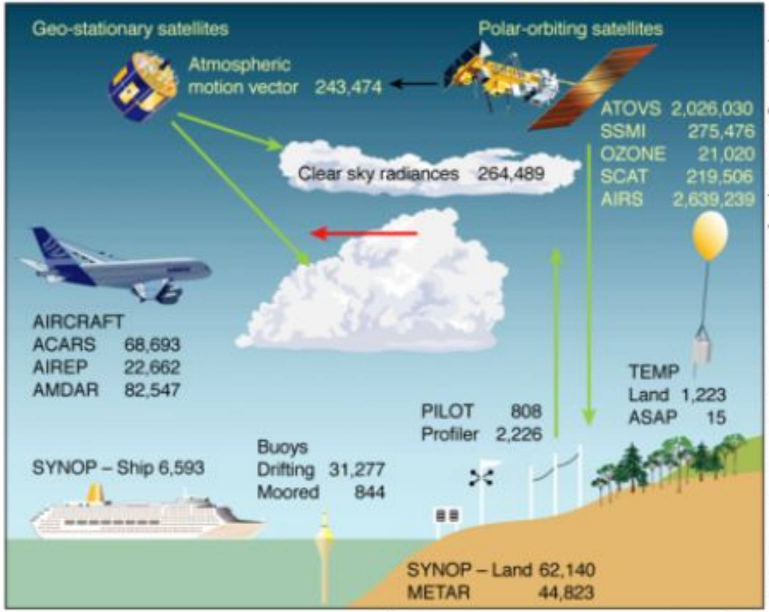
https://www.idi.ntnu.no/emner/it3105/materials/rl/cartpole-brownlee-2005.pdf

Reinforcement learning with local LQR

$$
\begin{aligned}
r(x_k, y_k) &= x_k^T P(x_k) x_k + y^T R(x_k) y_k \\
&= x_k^T \sum \alpha_i(x_k) P_i \, x_k + y_k^T \sum \alpha_i(x_k) R_i \, y_k
\end{aligned}
$$

DA in operational NWP





http://images.slideplayer.com/25/7834014/slides/slide_8.jpg

$$x \sim N(x_b, B) \qquad y \sim N(H(x_b), R)$$

$$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$$

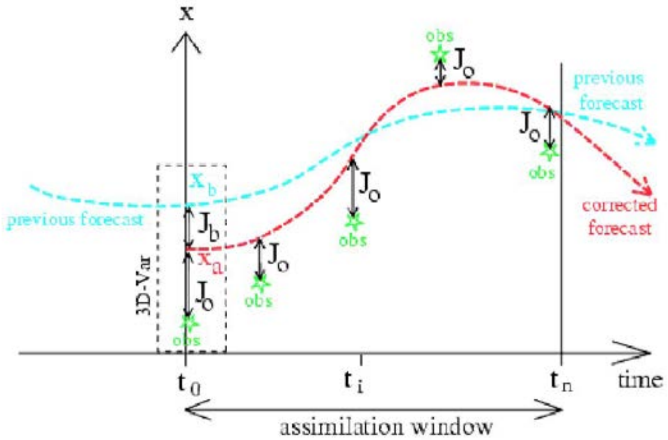
3D- and 4D-Var

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H[\mathbf{x}])^T\mathbf{R}^{-1}(\mathbf{y}-H[\mathbf{x}])$$

$$= J_b(\mathbf{x}) + J_o(\mathbf{x})$$

$$J(\mathbf{x}) = (\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \sum_{i=0}^{n}(\mathbf{y}_i-H_i[\mathbf{x}_i])^T\mathbf{R}_i^{-1}(\mathbf{y}_i-H_i[\mathbf{x}_i])$$

$$\forall i,\quad \mathbf{x}_i = M_{0\rightarrow i}(\mathbf{x})$$



EnVar, EnKF or hybrid methods

$$\mathbf{x}^a(t) = \mathbf{x}^f(t) + \mathbf{K}[\mathbf{y}^o - \mathcal{H}\mathbf{x}^f(t)],$$

$$\mathbf{K} = \mathbf{P}^f\mathcal{H}^{\mathrm{T}}(\mathcal{H}\mathbf{P}^f\mathcal{H}^{\mathrm{T}} + \mathbf{R})^{-1},$$

$$\mathbf{x}^f(t+1) = \mathcal{M}[\mathbf{x}^a(t)].$$

$$\mathbf{P}_{\text{hybrid}} = \beta\mathbf{P}^s + (1-\beta)\mathbf{P}^f$$

$$\mathbf{P}^f\mathcal{H}^{\mathrm{T}} \equiv \frac{1}{N_{\text{ens}}-1}\sum_{i=1}^{N_{\text{ens}}}(\mathbf{x}_i^f - \overline{\mathbf{x}^f})(\mathcal{H}\mathbf{x}_i^f - \overline{\mathcal{H}\mathbf{x}^f})^{\mathrm{T}},$$

$$\mathcal{H}\mathbf{P}^f\mathcal{H}^{\mathrm{T}} \equiv \frac{1}{N_{\text{ens}}-1}\sum_{i=1}^{N_{\text{ens}}}(\mathcal{H}\mathbf{x}_i^f - \overline{\mathcal{H}\mathbf{x}^f})(\mathcal{H}\mathbf{x}_i^f - \overline{\mathcal{H}\mathbf{x}^f})^{\mathrm{T}},$$

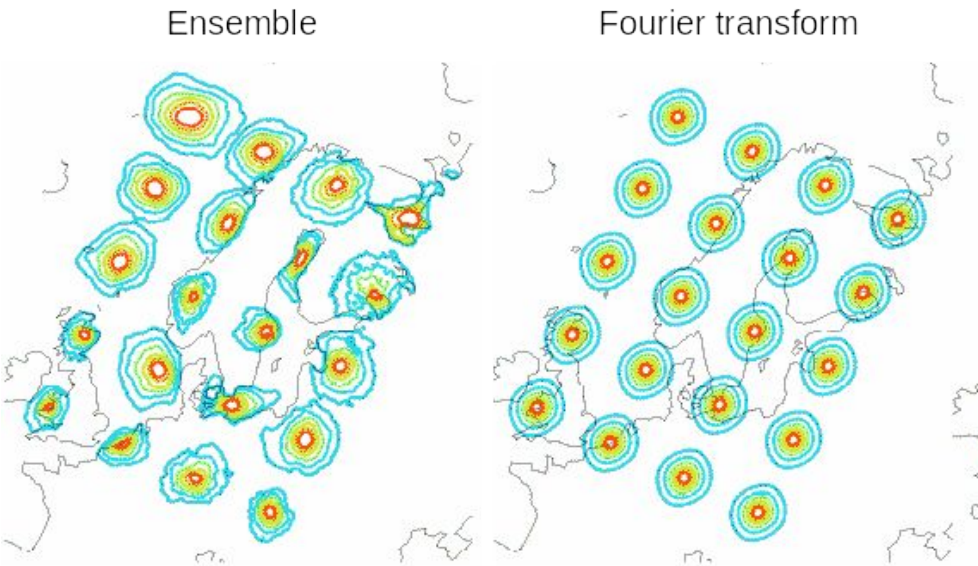
Dealing with O(10^16)

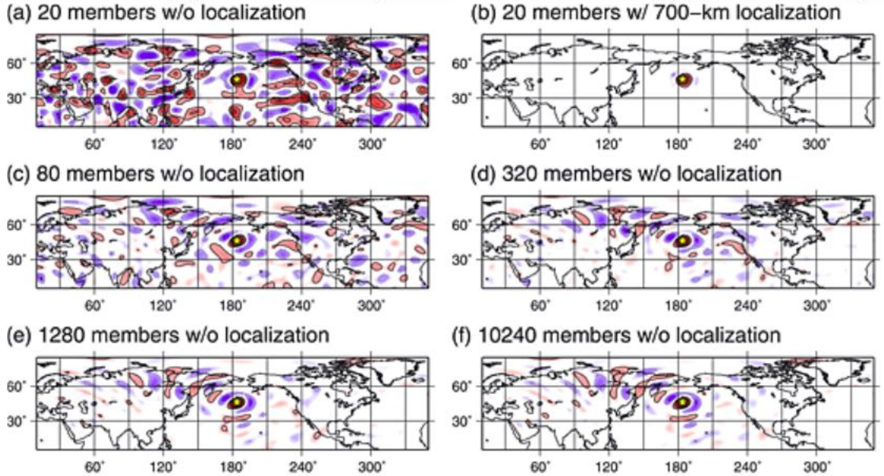
Diagonalize with spectral transform

$$\mathbf{B} = \mathbf{L}^{\mathrm{T}}\boldsymbol{\Sigma}^{\mathrm{T}}\mathbf{C}\boldsymbol{\Sigma}\mathbf{L}$$

EnKF with localization

$$\mathbf{K} = [\rho \circ (\mathbf{P}^f \mathcal{H}^{\mathrm{T}})][\rho \circ (\mathcal{H}\mathbf{P}^f \mathcal{H}^{\mathrm{T}}) + \mathbf{R}]^{-1}$$





DA challenges and remedies?

High dimensional state space

- MLWP models that can generate O(10^6) ensemble members
- Latent space DA

Nonlinear dynamics (clouds, snow, …)

- Monte-Carlo filtering
- Koopman operators

Non-Gaussian errors

- Transport methods; Diffusion models, Normalizing flows, …

MLWP for limited area models

Probabilistic Weather Forecasting with Hierarchical GNN (LiU, SMHI, DMI, MS)

Stretched grid and LAM using ECMWF Anemoi (deterministic so far)

Py4Cast (Météo France)

AICON LAM (DWD)

NVIDIA StormCast (US)

NVIDIA CorrDiff downscaling (Taiwan)

YingLong (Academy, China)

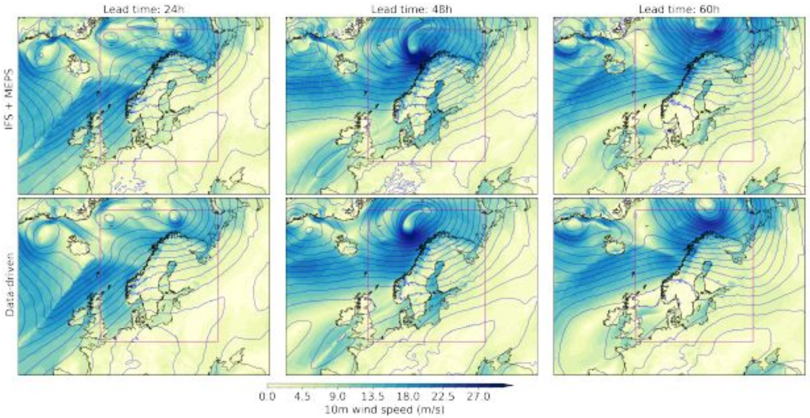
MetMamba (PERSKY Tech, China)

… probably more

Performance at km scale?

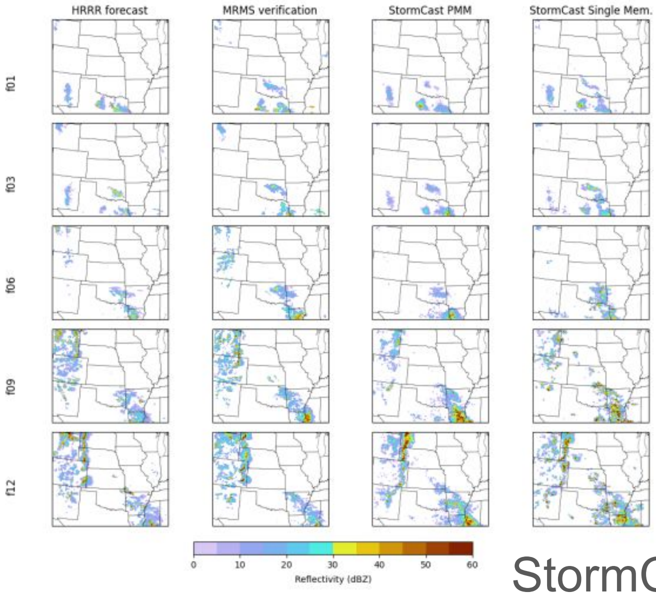
REGIONAL DATA-DRIVEN WEATHER MODELING WITH A GLOBAL STRETCHED-GRID

Thomas Nils Nipen [a*] Håvard Homleid Haugen [a*] Magnus Sikora Ingstad [a*]
Even Marius Nordhagen [a*] Aram Farhad Shafiq Salihi [a*] Paulina Tedesco [a*]
Ivar Ambjørn Seierstad [a*] Jørn Kristiansen [a] Simon Lang [b] Mihai Alexe [c] Jesper Dramsch [c]
Baudouin Raoult [b] Gert Mertes [b] Matthew Chantry [b]



Kilometer-Scale Convection Allowing Model Emulation using Generative Diffusion Modeling

Jaideep Pathak[*1], Yair Cohen[*1], Piyush Garg[*1], Peter Harrington[*2], Noah Brenowitz[1], Dale Durran[1,3], Morteza Mardani[1], Arash Vahdat[1], Shaoming Xu[†4], Karthik Kashinath[1], Michael Pritchard[1]



StormCast

ML and Variational DA

ML 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 given by an ML model we can calculate the gradients $\frac{\partial M(x)}{\partial x}$

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

PREDICTABILITY LIMIT OF THE 2021 PACIFIC NORTHWEST HEATWAVE FROM DEEP-LEARNING SENSITIVITY ANALYSIS

A PREPRINT

P. Trent Vonich [*1,2] and Gregory J. Hakim [†1]
[1]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA
[2]Air Force Institute of Technology, Wright-Patterson AFB, OH, USA

June 10, 2024

Dynamics with auto differentiation via Pytorch / JAX

https://doi.org/10.5194/egusphere-2022-943
Preprint. Discussion started: 28 September 2022

Pace v0.1: A Python-based Performance-Portable Implementation of the FV3 Dynamical Core

Johann Dahm[1,*], Eddie Davis[1,*], Florian Deconinck[1,*], Oliver Elbert[1,*], Rhea George[1,*], Jeremy McGibbon[1,*], Tobias Wicky[1,*], Elynn Wu[1,*], Christopher Kung[2], Tal Ben-Nun[3], Lucas Harris[4], Linus Groner[5], and Oliver Fuhrer[1,6]

[*]These authors contributed equally to this work and are listed in alphabetical order.
[1]Allen Institute of Artificial Intelligence, Seattle, U.S.A.
[2]Global Modeling and Assimilation Office, NASA, Greenbelt MD, U.S.A.
[3]Department of Computer Science, ETH Zurich, Zurich, Switzerland
[4]Geophysical Fluid Dynamics Laboratory, NOAA, Princeton NJ, U.S.A.
[5]Swiss National Supercomputing Centre (CSCS), ETH Zurich, Lugano, Switzerland
[6]Federal Institute of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

FVM Code adaptation approach : Built on GT4Py + DaCe

GridTools 4python

Python Domain Specific Language (DSL) for Weather and Climate HPC code generation

GT4Py : GridTools for Python

+ Portable across CPU and GPU (Nvidia, AMD) architectures
+ Modularization of the code (dycore, physical packages) and OOP (Object Oriented Programming)
+ Used by ICON (Exclaim), COSMO, and NOAA (FV3GFS)

DaCe : Data Centric Parallel Programming

- Generating high-performance code for parts out of GT4Py
- DaCeML : Merging AI and Physics based models
  - Model inference using ONNX
  - Bindings with Pytorch

Towards a Portable Model for All-scale Predictions (PMAP) : FVM

Loïc Maurin, Fabrice Voitus

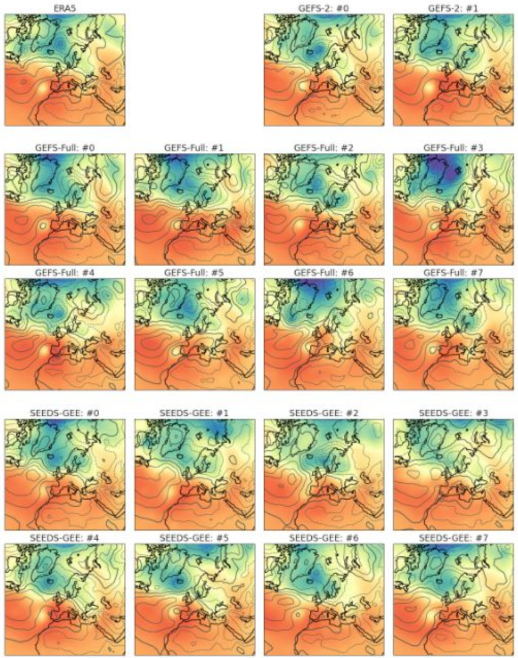
CNRM/GMAP/ALGO, Météo-France

April 18th, 2024

B 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

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



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

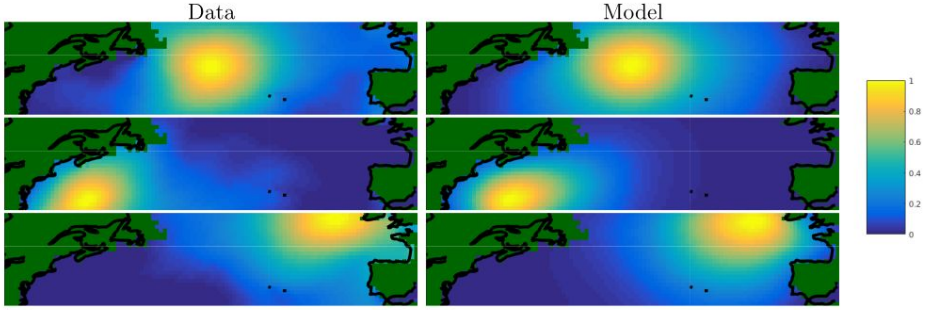
John Anderson
Google Research
janders@google.com

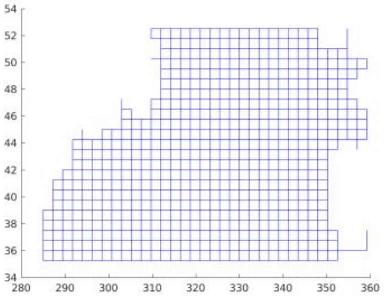
Or model B using SPDEs (So’s talk on message passing)

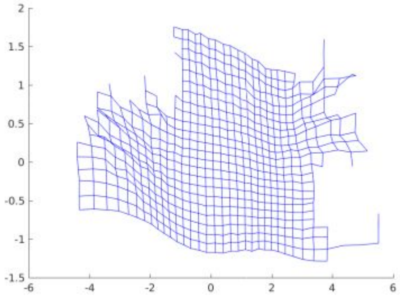
Deformed SPDE models, with an application to spatial modeling of significant wave height

Anders Hildeman, David Bolin, and Igor Rychlik

Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Sweden

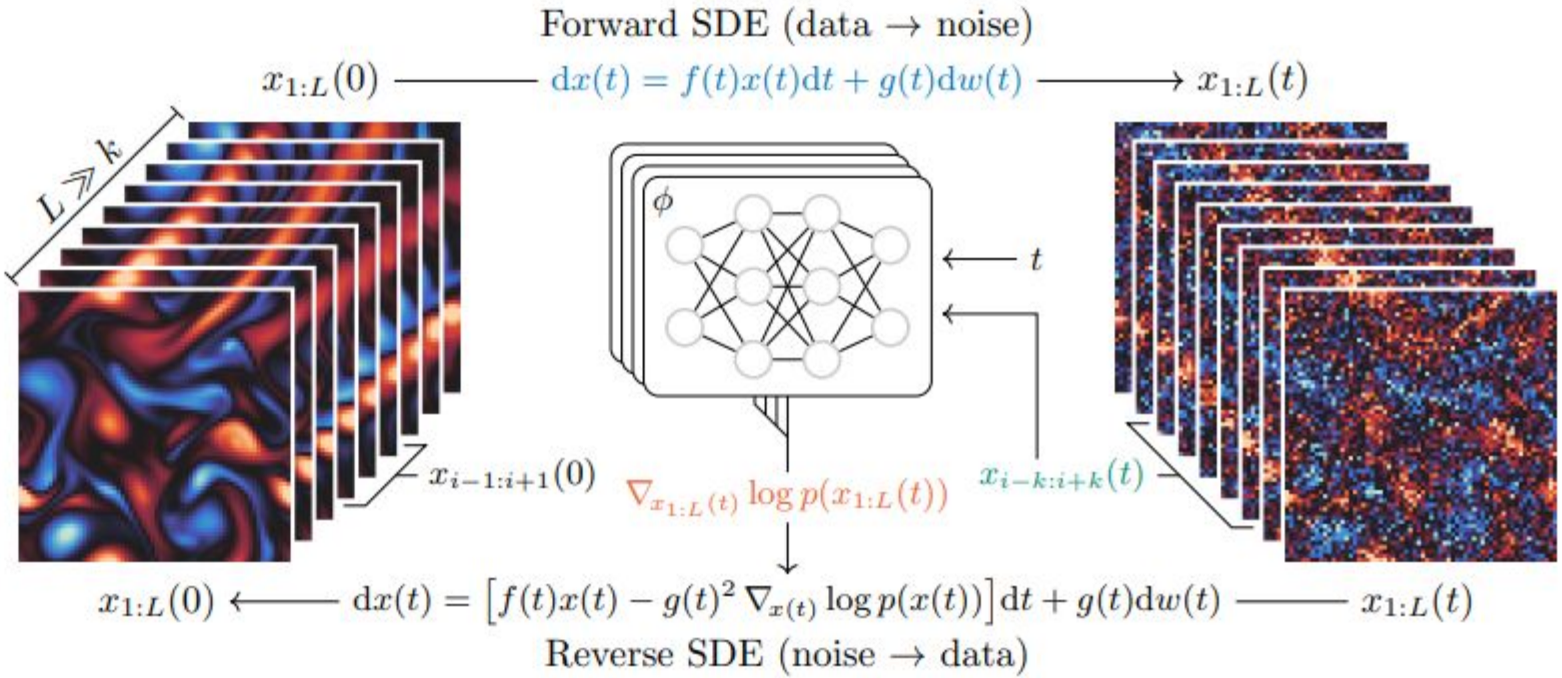






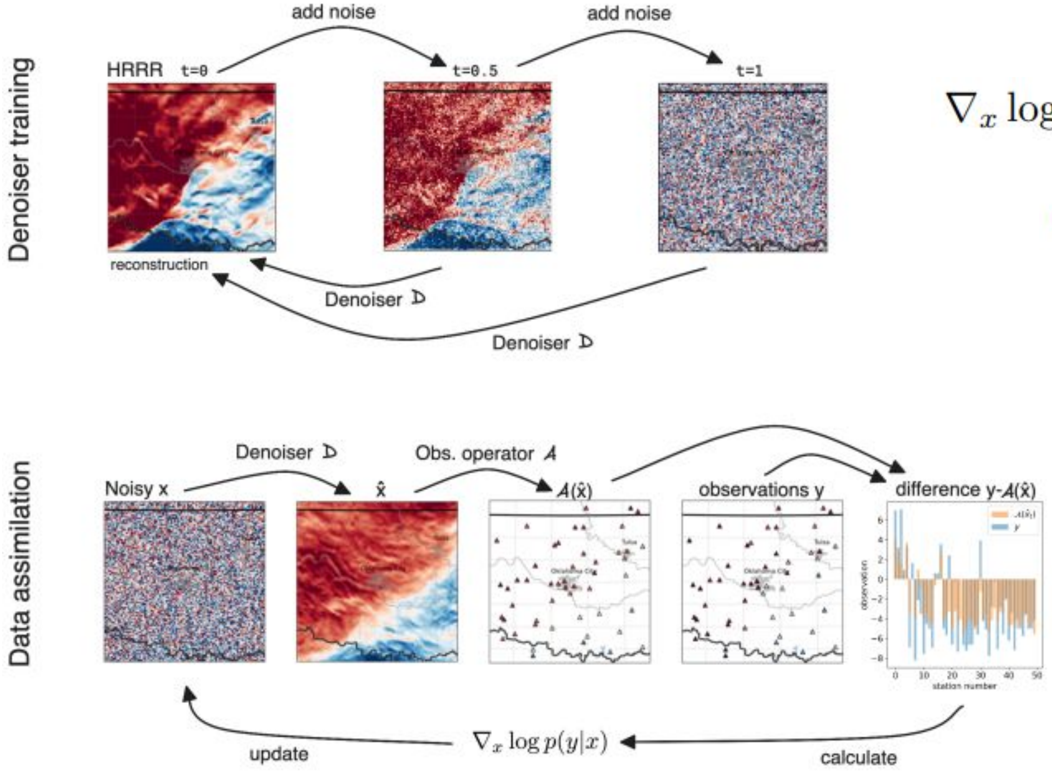
Probabilistic DA with generative models

Score-based diffusion models



https://proceedings.neurips.cc/paper_files/paper/2023/file/7f7fa581cc8a1970a4332920cdf87395-Paper-Conference.pdf

Score based DA (more general challenge from So)



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

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

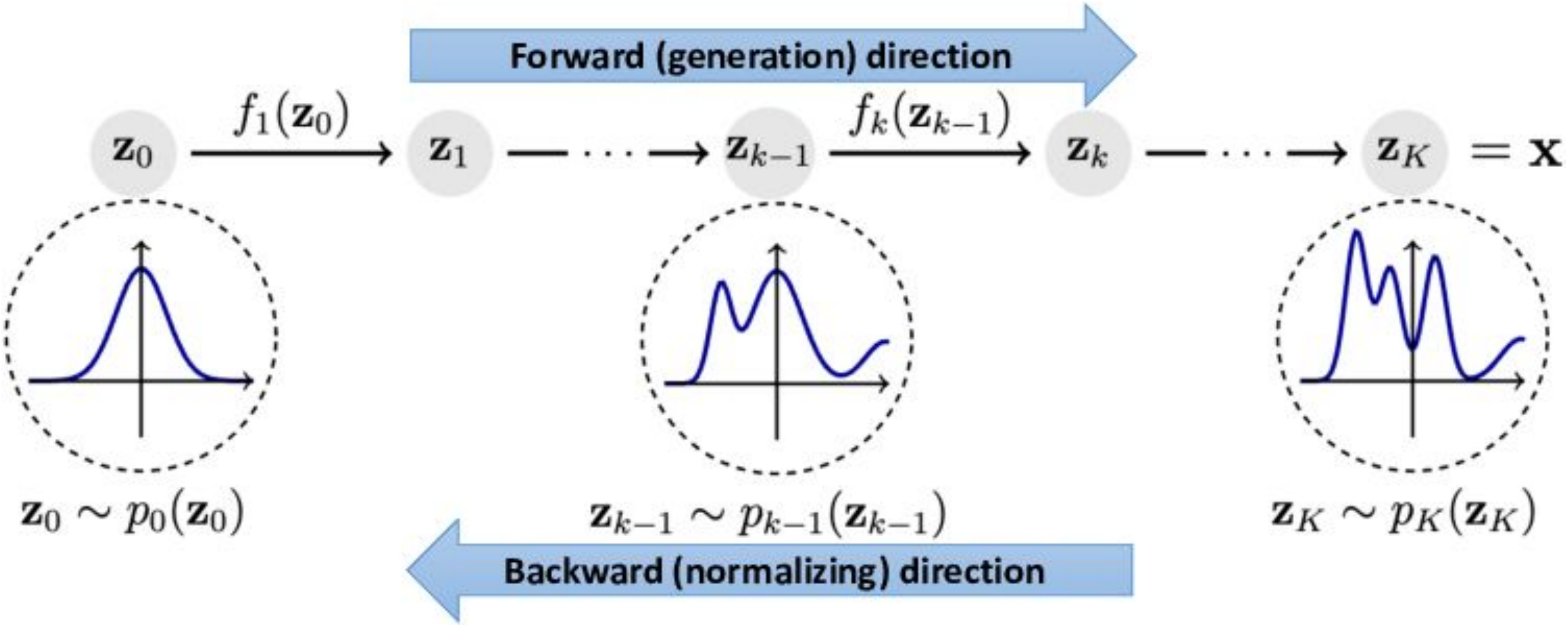
Generative Data Assimilation of Sparse Weather Station Observations at Kilometer Scales

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

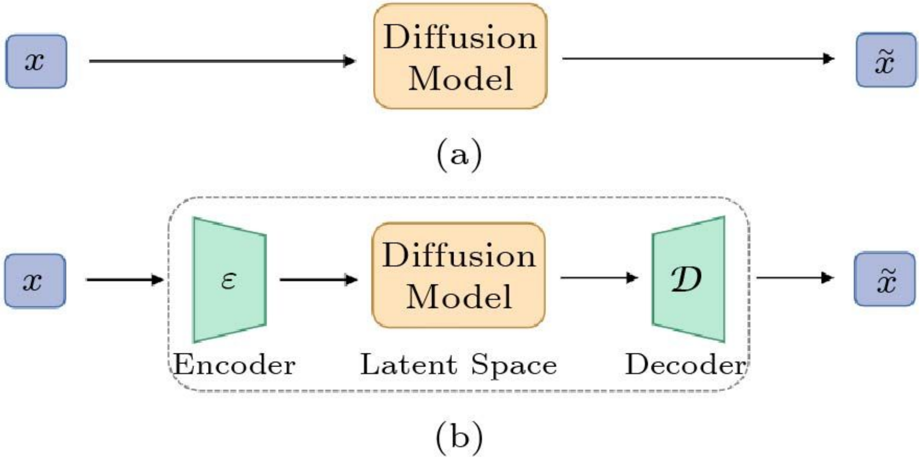
Gaussian errors via transport methods



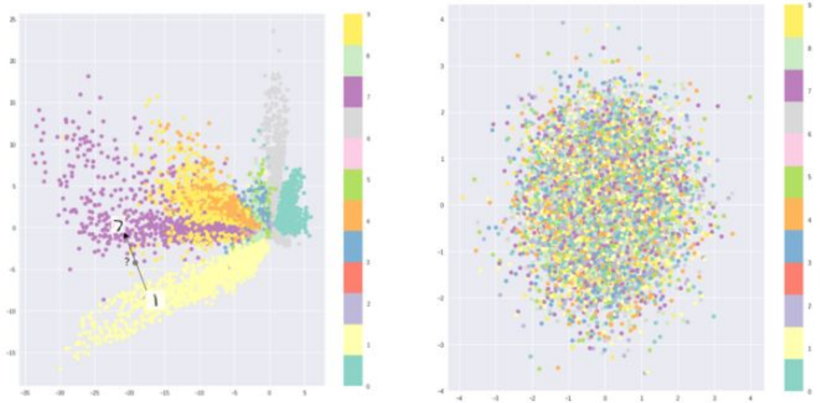
https://www.researchgate.net/figure/Normalizing-flow-distribution-transformation-20-The-forward-direction-transforms-a_fig1_373016632

Combine autoencoder with 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” SMC (nonlinear with Gaussian noise)

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

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

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

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

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

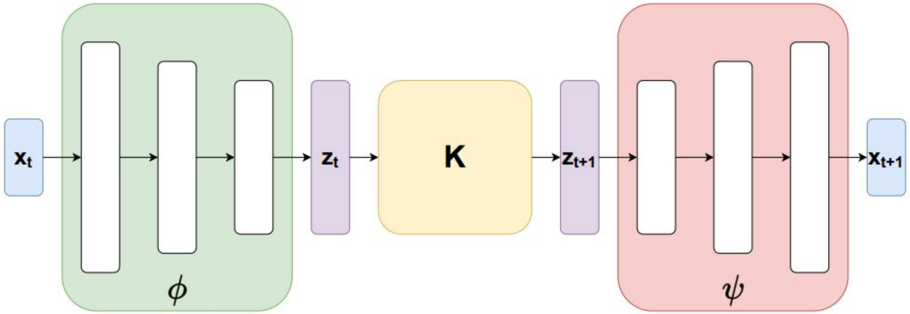
Koopman operator and Kalman filter in latent space

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

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

Neural operators to model infinite dimensional Koopman operator

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



https://arxiv.org/pdf/2309.05317

Foundation models and DA

Foundation Models

Foundation Models provide task-independent but widely applicable machine learning models for a domain or system (Bommasani, 2021).

Bommasani et al., CoRR abs/2108.07258, 2021

Big players are interested

IBM and NASA (Prithvi WxC)

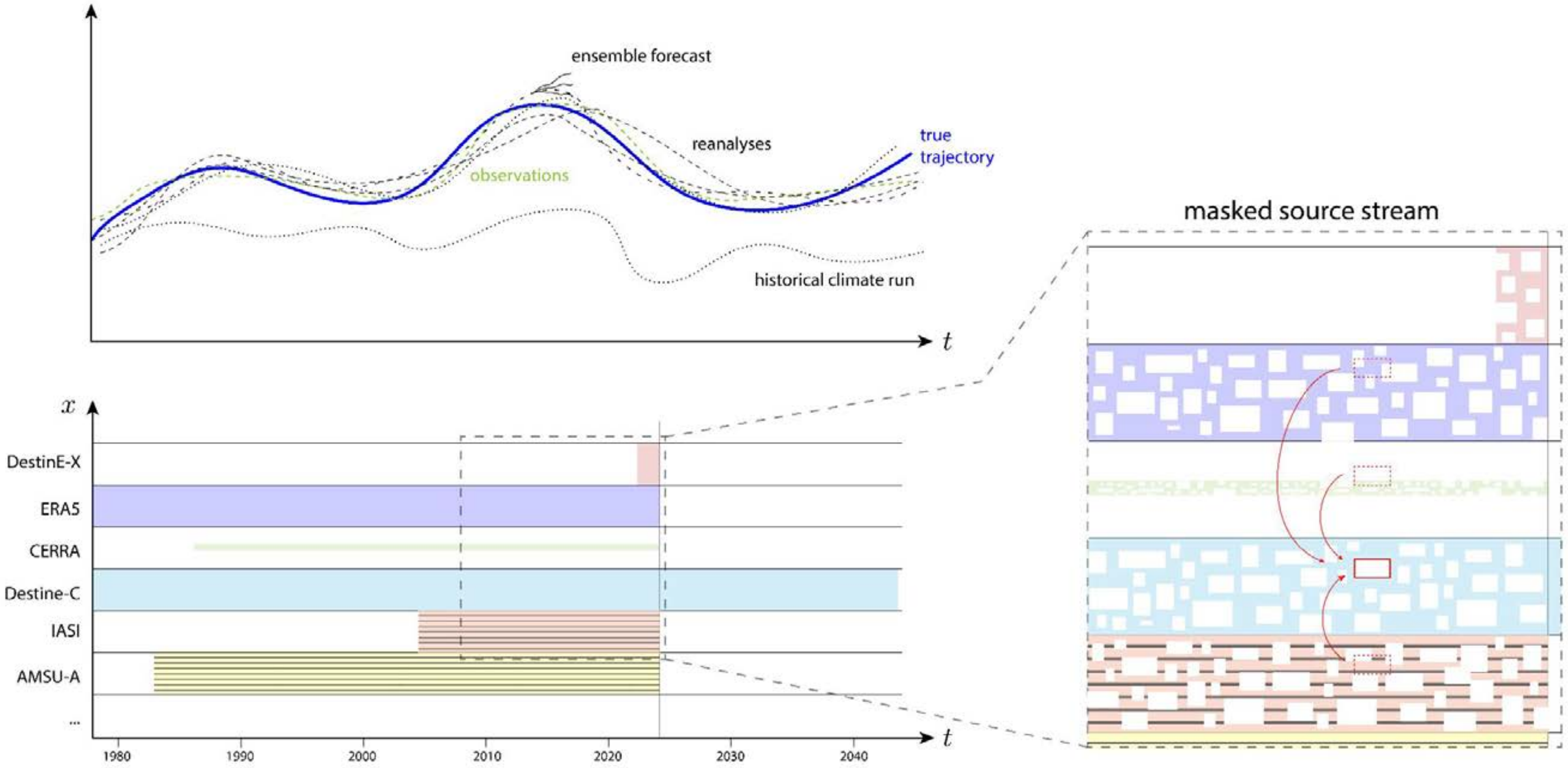
Microsoft (ClimaX and Aurora)

Google (TimesFM, not spatio-temporal)

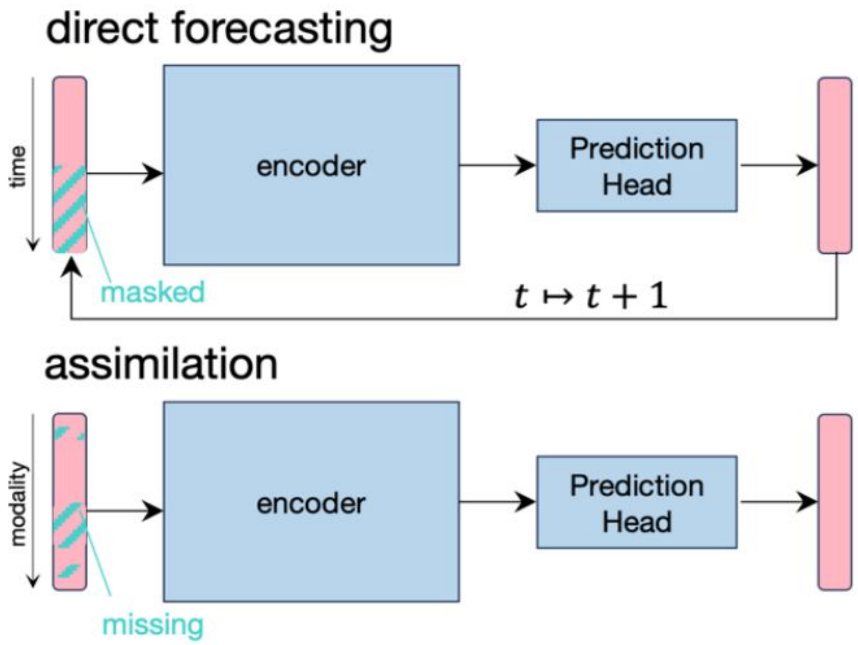
ECMWF EU project Weather Generator

(not so big) Jua (raised $16M)

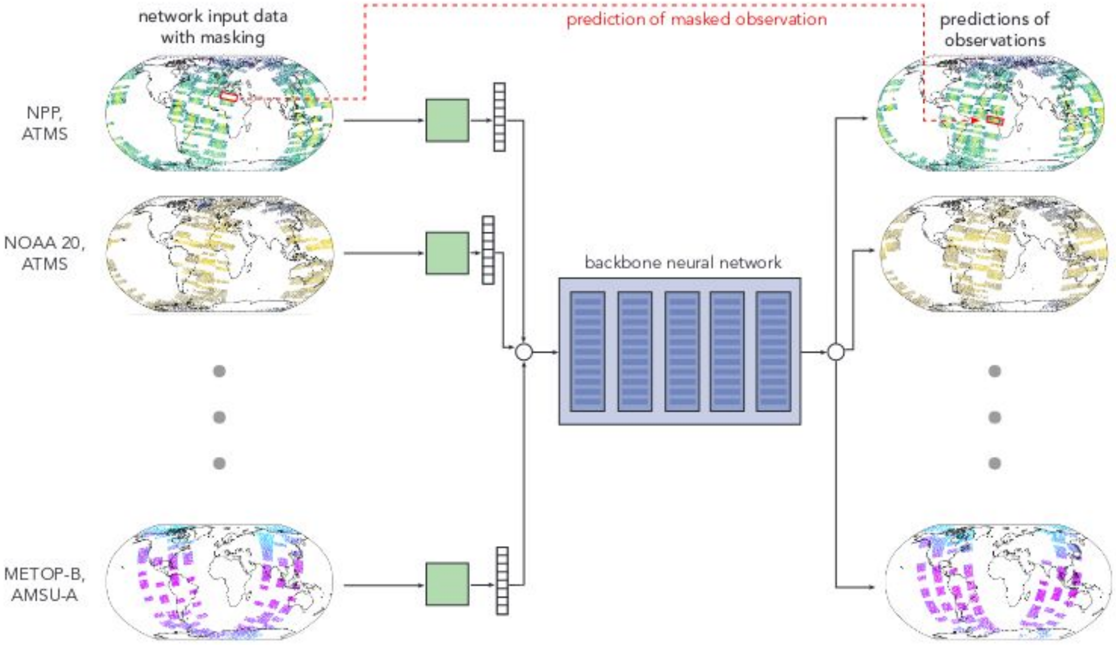
Self supervised learning

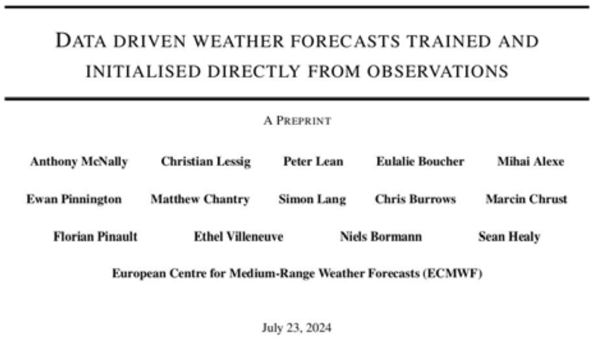


Mask out in time (forecast) or modality (assimilation)



Obs only AtmoRep and a new take on data assimilation





DATA DRIVEN WEATHER FORECASTS TRAINED AND INITIALISED DIRECTLY FROM OBSERVATIONS

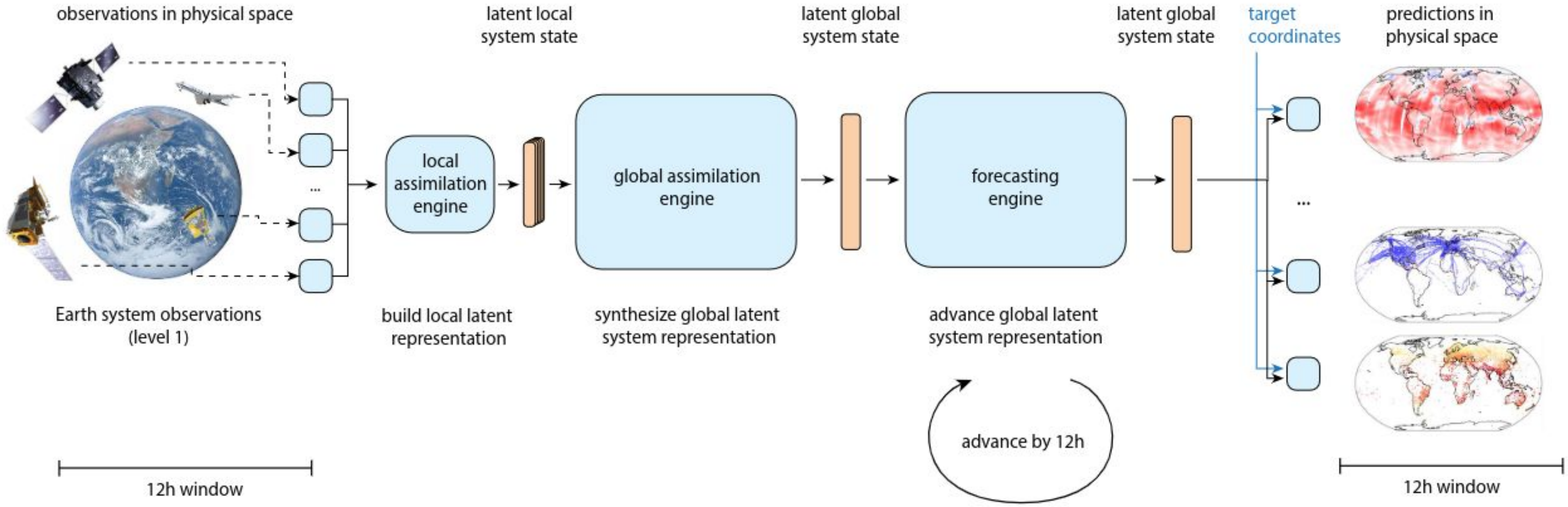
A PREPRINT

Anthony McNally, Christian Lessig, Peter Lean, Eulalie Boucher, Mihai Alexe, Ewan Pinnington, Matthew Chantry, Simon Lang, Chris Burrows, Marcin Chrust, Florian Pinault, Ethel Villeneuve, Niels Bormann, Sean Healy

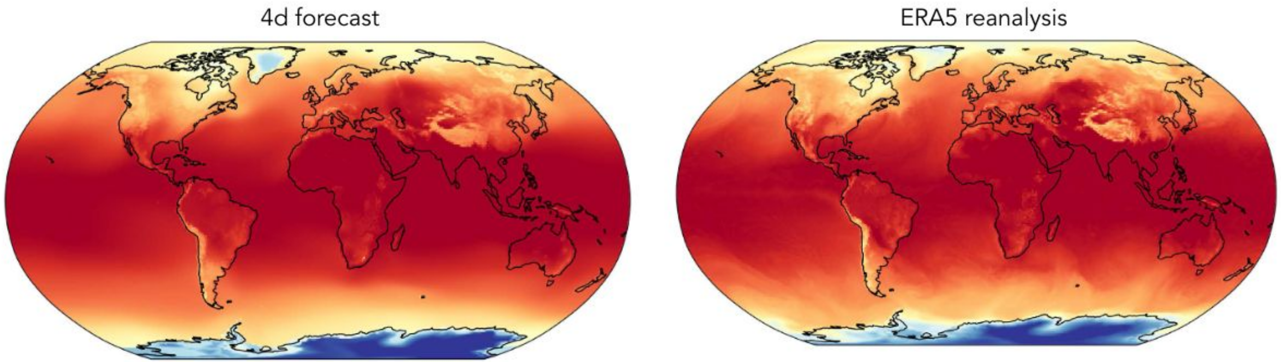
European Centre for Medium-Range Weather Forecasts (ECMWF)

July 23, 2024

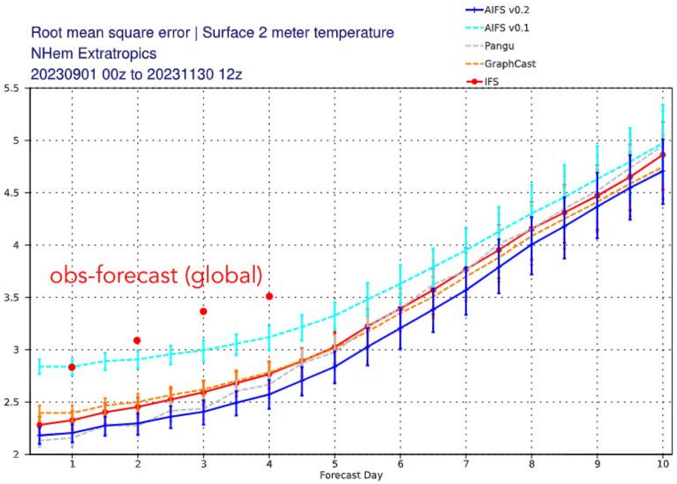
Representational DA - end of DA as we know it?



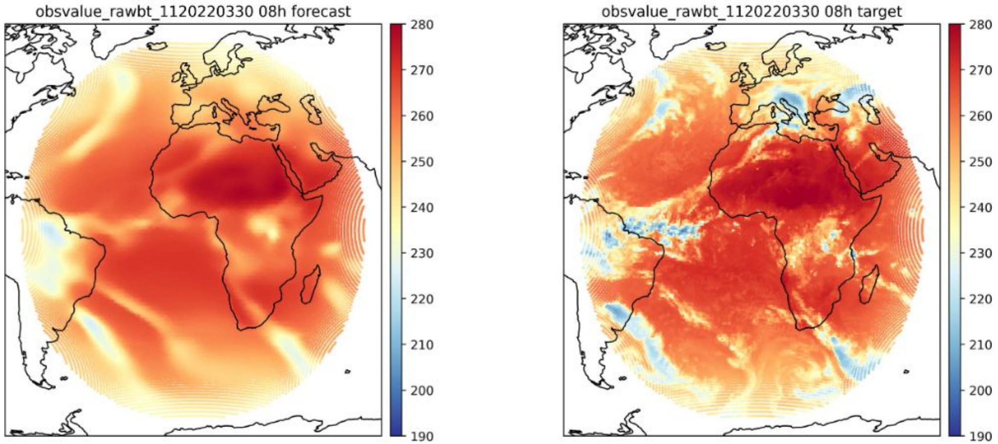
Examples



4d forecast
ERA5 reanalysis



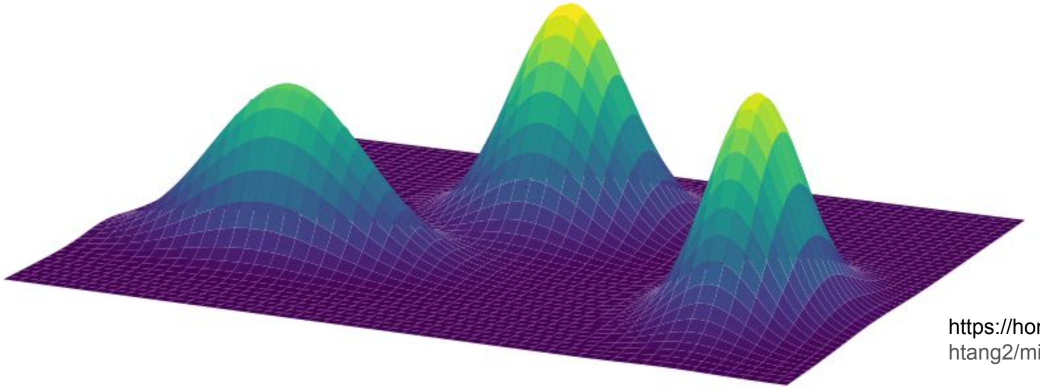
Root mean square error | Surface 2 meter temperature
NHem Extratropics
20230901 00z to 20231130 12z
AIFS v0.2
AIFS v0.1
Pangu
GraphCast
IFS
obs-forecast (global)
Forecast Day



obsvalue_rawbt_1120220330 08h forecast
obsvalue_rawbt_1120220330 08h target

Liouville, Fokker-Planck and DA on quantum computers?

Model the evolution of a parameterized pdf, e.g. GMM?



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

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

Possible to model both p and f with neural networks? How to train?
No need for Monte-Carlo. Can evaluate p(x) directly.

Liouville, Fokker-Planck 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

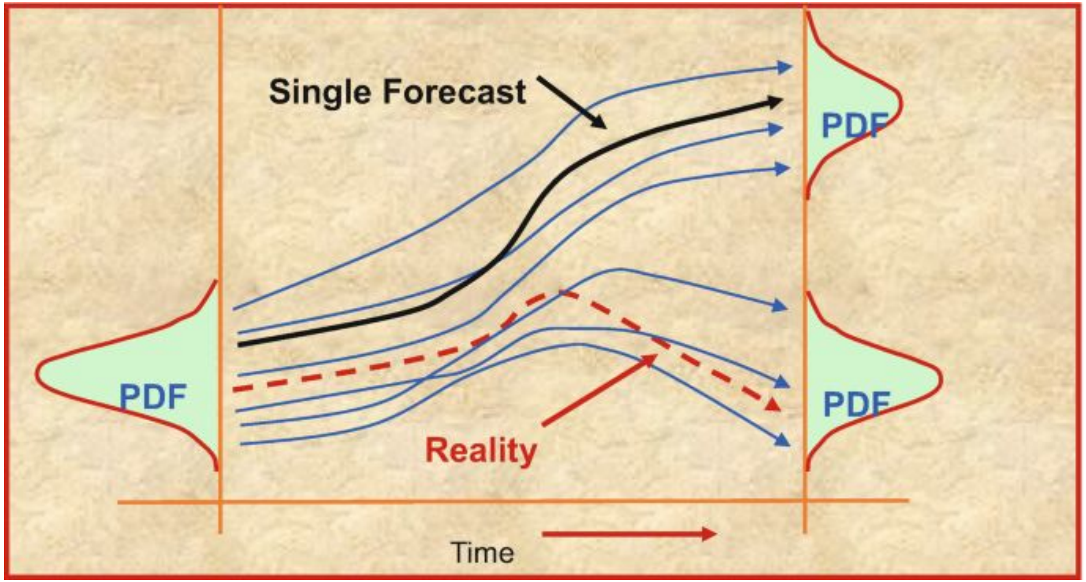
$$\frac{\delta}{\delta t}|\Psi\rangle = \frac{1}{i\hbar}H|\Psi\rangle$$

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.

Data assimilation with Perron-Frobenius operator

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

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

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

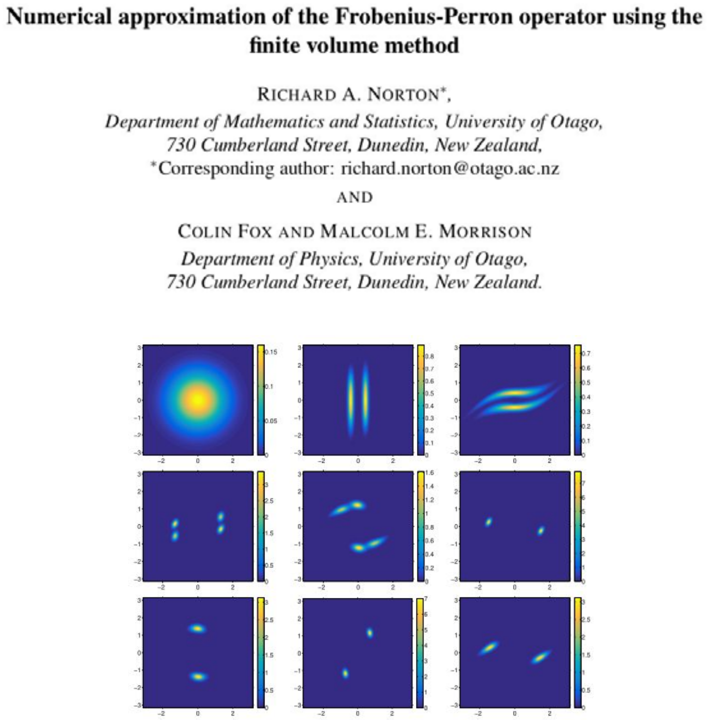
$$\rho(x_k|\mathbf{z}_{1:k-1}) = S(t_k - t_{k-1})\rho(x_{k-1}|\mathbf{z}_{1:k-1}).$$ S is Perron-Frobenius op.

Approximating with FVM (solve on quantum computer?)

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.

Time to get back to reality…

Ongoing/upcoming ML NWP projects at SMHI

CAISA towards coupled atmosphere - sea/ice data assimilation (ML obop)

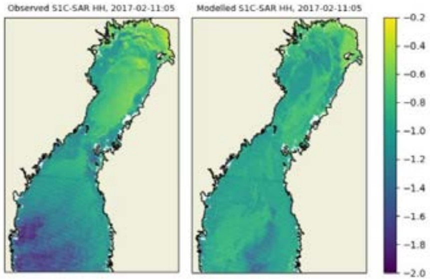
OWGRE improved NWP for renewable energy (MLWP, generative)

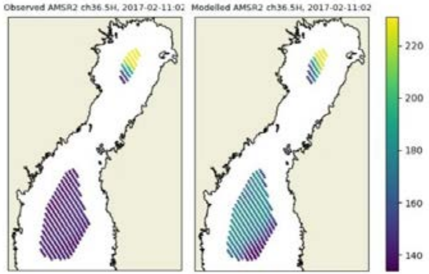
Destination Earth Programme

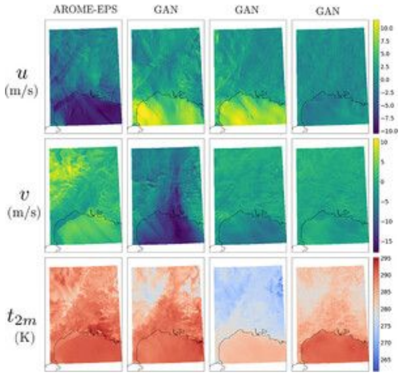
- DE_330 On-demand Extremes Digital Twin (MLWP, generative)
- DE_371 Machine Learning for Earth system Digital Twins (generative, interp)

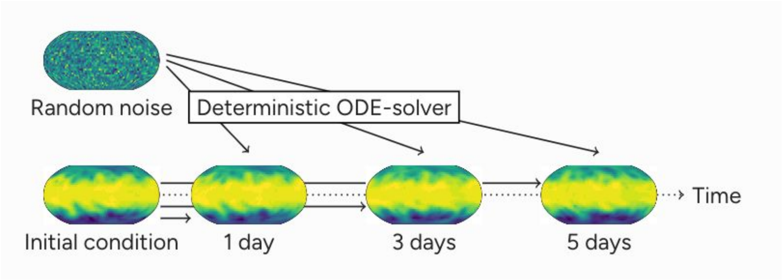
ECMWF ML pilot project (MLWP, generative)

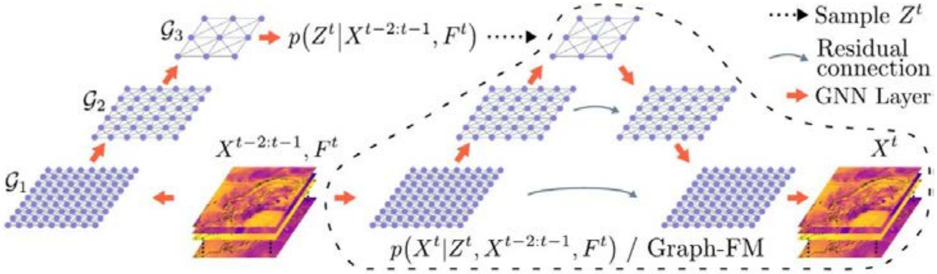
EU Horizon WeatherGenerator (Foundation models, generative)

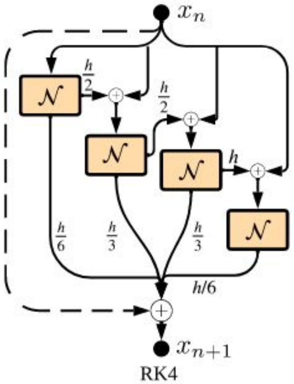

Observed S1C-SAR HH, 2017-02-11:05
Modelled S1C-SAR HH, 2017-02-11:05
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
-1.6
-1.8
-2.0

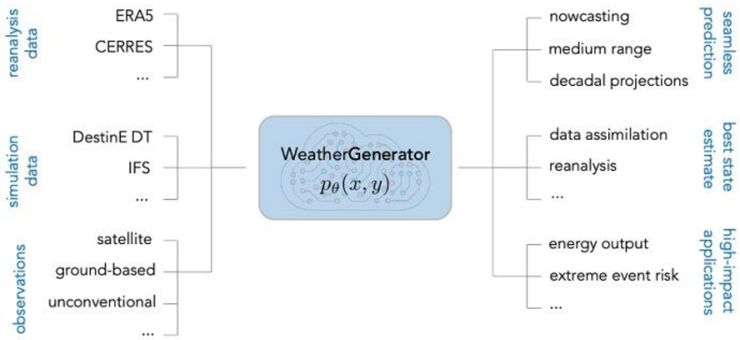

Observed AMSR2 ch36.5H, 2017-02-11:02
Modelled AMSR2 ch36.5H, 2017-02-11:02
220
200
180
160
140


AROME-EPS
GAN
GAN
GAN
u
(m/s)
v
(m/s)
t_{2m}
(K)


Random noise
Deterministic ODE-solver
Initial condition
1 day
3 days
5 days
Time


$\mathcal{G}_3$
$p(Z^t|X^{t-2:t-1}, F^t)$
Sample Z^t
Residual connection
GNN Layer
$\mathcal{G}_2$
$X^{t-2:t-1}, F^t$
X^t
$\mathcal{G}_1$
$p(X^t|Z^t, X^{t-2:t-1}, F^t)$ / Graph-FM


x_n
$\mathcal{N}$
$\frac{h}{2}$
$\frac{h}{2}$
h
$\frac{h}{6}$
$\frac{h}{3}$
$\frac{h}{3}$
$h/6$
x_{n+1}
RK4


reanalysis data
ERA5
CERRES
...
simulation data
DestinE DT
IFS
...
observations
satellite
ground-based
unconventional
...
WeatherGenerator
$p_\theta(x, y)$
nowcasting
medium range
decadal projections
seamless prediction
data assimilation
reanalysis
...
best state estimate
energy output
extreme event risk
...
high-impact applications