Learning from earth system observations: machine learning or data assimilation?

Alan Geer

European Centre for Medium-range Weather Forecasts

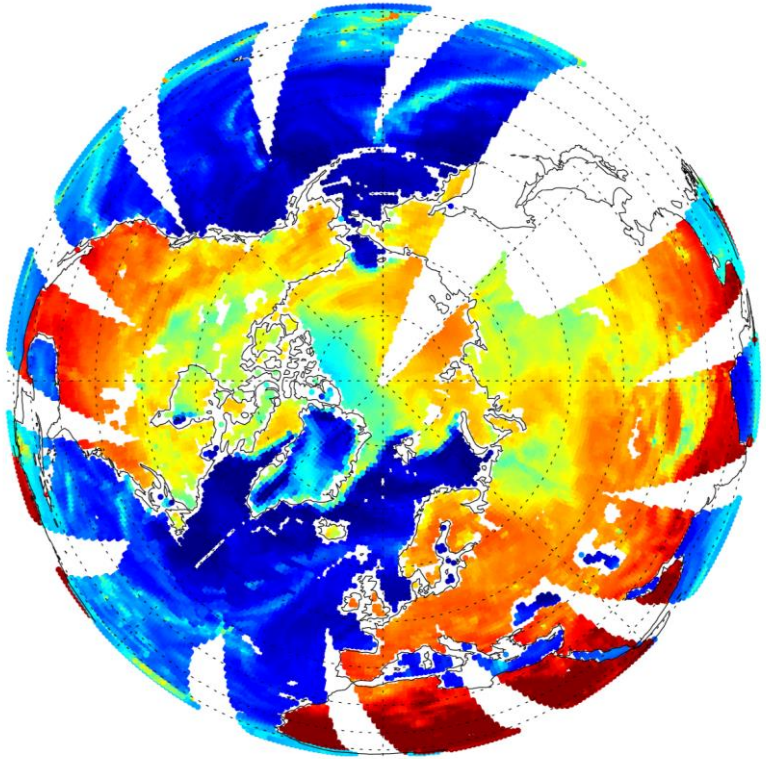
alan.geer@ecmwf.int

Thanks to: Massimo Bonavita, Sam Hatfield, Patricia de Rosnay, Peter Dueben, Peter Lean

ECMWF

Physical forward model

Satellite observations



SSMIS F-17 channel 13 (19 GHz, v)
Microwave brightness temperatures
3rd December 2014

$$y = h \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_{\dots} \end{pmatrix}$$

Forward function

Geophysical variables

Atmospheric temperature, water vapour, wind, cloud, precipitation

Skin and substrate temperature and moisture

Ocean wind, waves, foam

Sea-ice

Snowpack

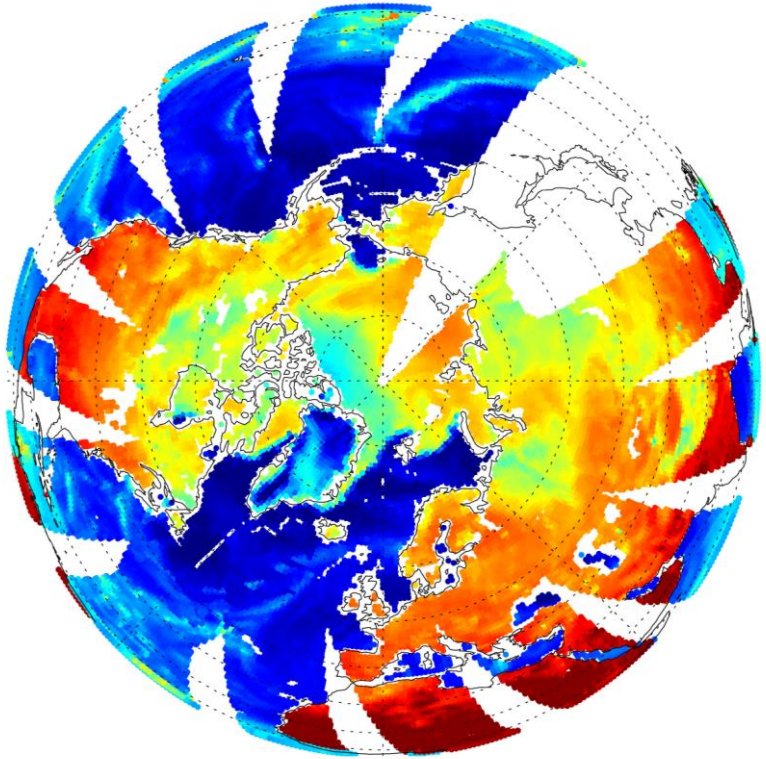
Ice

Vegetation

Soil

Physical forward model

Satellite observations



SSMIS F-17 channel 13 (19 GHz, v)
Microwave brightness temperatures
3rd December 2014

Geophysical variables

Equations & parameters – where sometimes knowledge is quite uncertain

$$y = h\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_{\dots} \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_{\dots} \end{bmatrix}\right)$$

Forward function

Gas spectroscopy

Scattering from hydrometeors

Cloud and precipitation micro and macro-structure

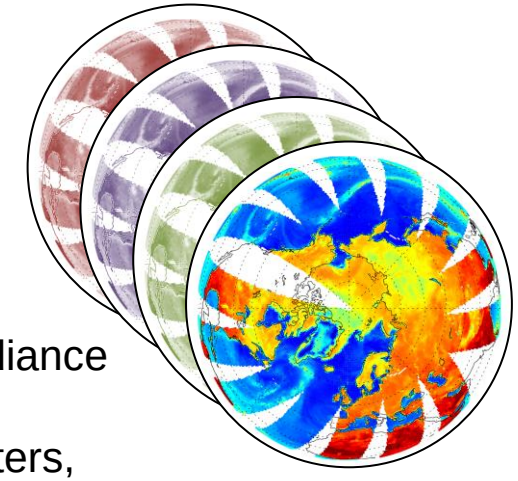
Snow / ice grain size and structure

Ocean foam coverage

4D coupled earth system assimilation

Observations

Forward model: observation operator and geophysical models



Satellite radiance sensors, scatterometers, radars

GPSRO, AMVs, scatterometer, aircraft, sonde, met. stations, ships, buoys …

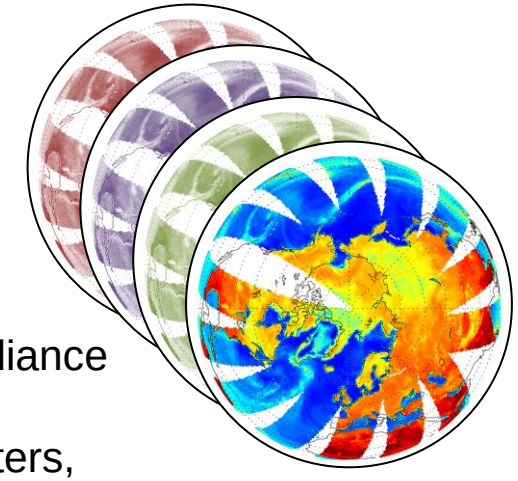
Earth system model

$$y = h\left(m\left(\begin{bmatrix} x_{\text{atmos}} \\ x_{\text{land}} \\ x_{\text{snow}} \\ x_{\text{seaice}} \\ x_{\text{ocean}} \\ x_{\dots} \end{bmatrix}, \begin{bmatrix} w_{\text{atmos}} \\ w_{\text{land}} \\ w_{\text{snow}} \\ w_{\text{seaice}} \\ w_{\text{ocean}} \\ x_{\dots} \end{bmatrix}\right), \begin{bmatrix} w_{\text{atmos}} \\ w_{\text{land}} \\ w_{\text{snow}} \\ w_{\text{seaice}} \\ w_{\text{ocean}} \\ x_{\dots} \end{bmatrix}\right)$$

4D coupled earth system assimilation

Observations

Forward model: observation operator and geophysical models



Satellite radiance sensors, scatterometers, radars

GPSRO, AMVs, scatterometer, aircraft, sonde, met. stations, ships, buoys …

Simplify: in this talk, h() includes observation operator and geophysical model

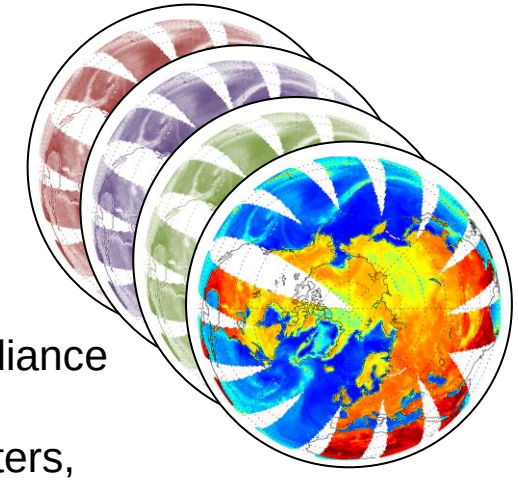
$$y = h\left(\begin{bmatrix} x_{\text{atmos}} \\ x_{\text{land}} \\ x_{\text{snow}} \\ x_{\text{seaice}} \\ x_{\text{ocean}} \\ x_{\dots} \end{bmatrix}, \begin{bmatrix} w_{\text{atmos}} \\ w_{\text{land}} \\ w_{\text{snow}} \\ w_{\text{seaice}} \\ w_{\text{ocean}} \\ w_{\dots} \end{bmatrix}\right)$$

Geophysical state

Model parameters, often shared with observation operators

Machine learning

Observations (‘labels’)



Satellite radiance sensors, scatterometers, radars

GPSRO, AMVs, scatterometer, aircraft, sonde, met. stations, ships, buoys …

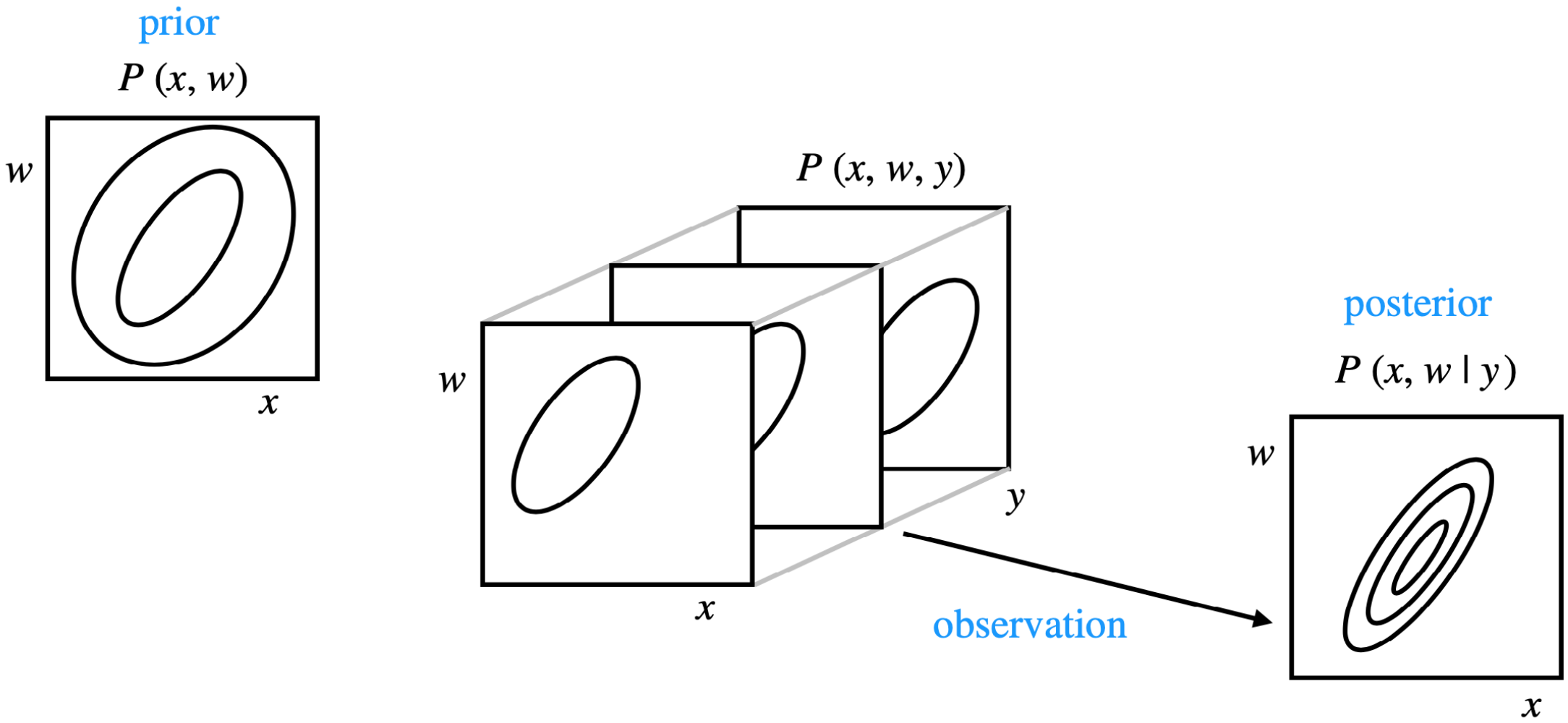
Empirical model

$$y = h\left(\begin{bmatrix} x_{\text{atmos}} \\ x_{\text{land}} \\ x_{\text{snow}} \\ x_{\text{seaice}} \\ x_{\text{ocean}} \\ x_{\dots} \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_{\dots} \end{bmatrix}\right)$$

Geophysical state (‘features’)

Machine learning parameters, e.g. neural network weights

Bayes' theorem



Cost function for variational DA

Assume Gaussian errors (error standard deviation σ)
and for clarity here simplify to scalar variables
and ignore any covariance between observation, model or state error

Prior (background)

$$J(x,w) = \underbrace{\frac{(y-h(x,w))^2}{(\sigma^y)^2}}_{J^y} + \underbrace{\frac{(x^b-x)^2}{(\sigma^x)^2}}_{J^x} + \underbrace{\frac{(w^b-w)^2}{(\sigma^w)^2}}_{J^w}$$

DA | Cost function | Observation term | Prior knowledge of state | Prior knowledge of model

Cost / loss function equivalence of ML and variational DA

Assume Gaussian errors (error standard deviation σ)
and for clarity here simplify to scalar variables
and ignore any covariance between observation, model or state error

ML **Loss function**

Basic loss function

Feature error?

Weights regularisation

$$J(x,w) = \underbrace{\frac{(y-h(x,w))^2}{(\sigma^y)^2}}_{J^y} + \underbrace{\frac{(x^b-x)^2}{(\sigma^x)^2}}_{J^x} + \underbrace{\frac{(w^b-w)^2}{(\sigma^w)^2}}_{J^w}$$

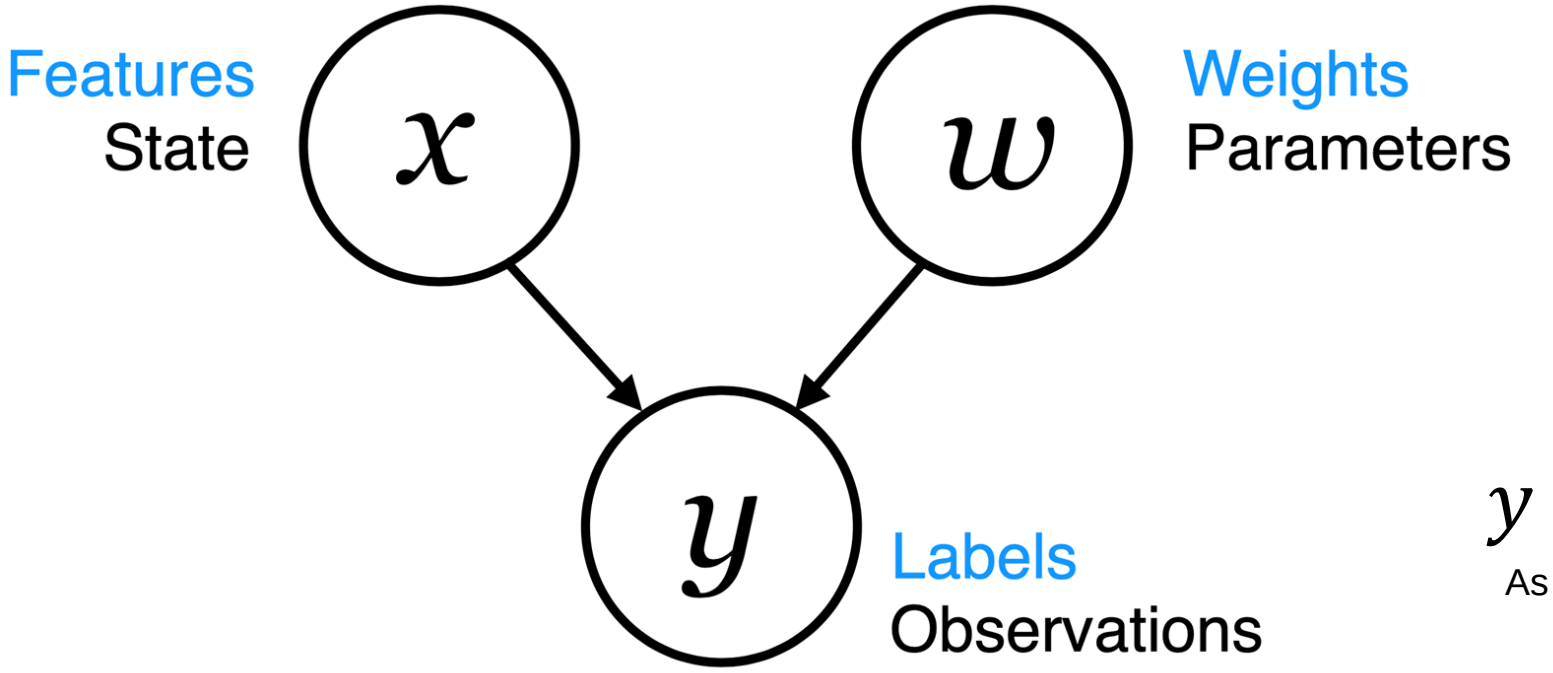
DA **Cost function**

Observation term

Prior knowledge of state

Prior knowledge of model

Bayesian equivalence of ML and DA



$$y = h(x, w)$$

As a Bayesian network

Geer (2021) https://doi.org/10.1098/rsta.2020.0089 or preprint: https://doi.org/10.21957/7fyj2811r

Bocquet et al. (2020) https://arxiv.org/abs/2001.06270

Abarbanel et al. (2018) https://doi.org/10.1162/neco_a_01094

Hsieh and Tang (1998) https://doi.org/10.1175/1520-0477(1998)079%3C1855:ANNMTP%3E2.0.CO;2

Goodfellow et al. (2016) https://www.deeplearningbook.org

Why do we need to learn models from observations?

ECMWF

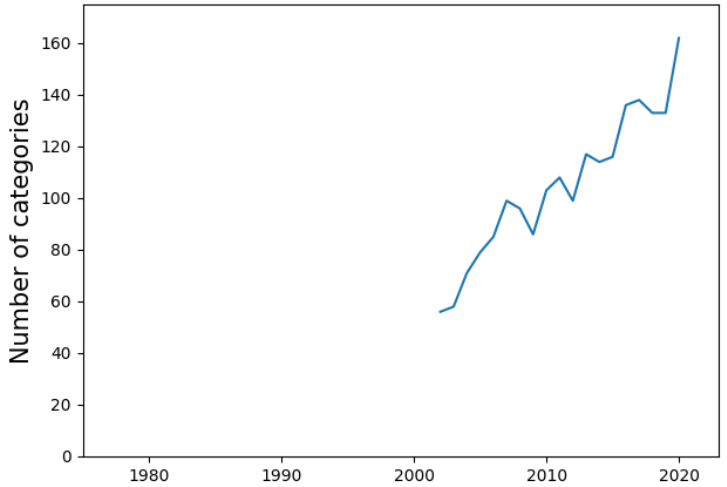
Observation challenges: increasing variety, data quality and calibration issues

Smallsat revolution

Constellations of short-lived satellites

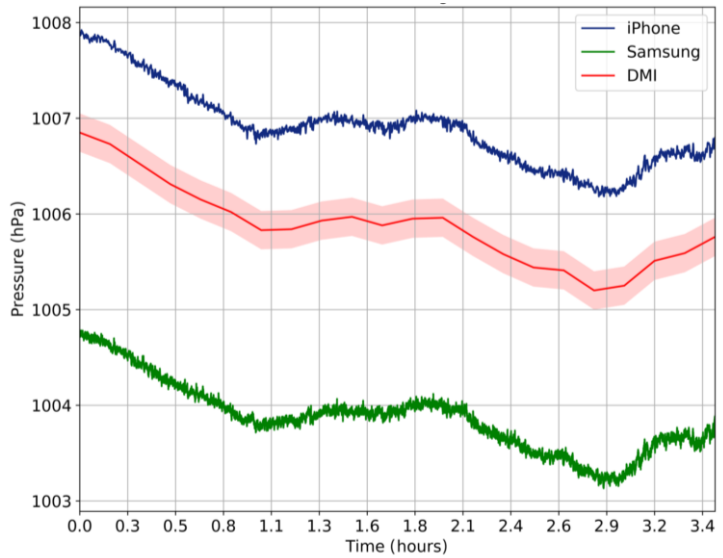
Variety

Variety of observation types at ECMWF (thanks to Peter Lean)



IoT

Pressure timeseries



One phone

Barometer

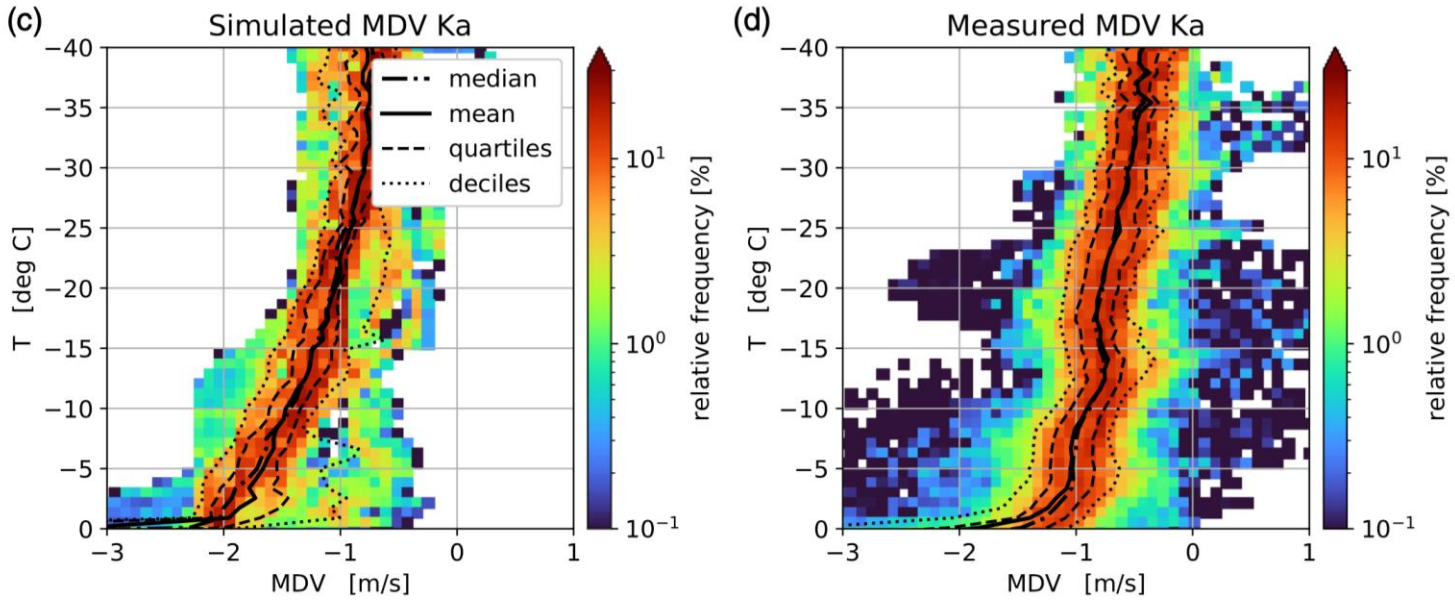
Another phone

From: Hintz et al. (2019), Collecting and processing of barometric data from smartphones for potential use in numerical weather prediction data assimilation, Met. Appl., https://doi.org/10.1002/met.1805

General challenge: make better use of campaign and research observations

Field campaign data underly many models and parametrisations: use them in a more optimal way, e.g. data assimilation

Mean doppler velocity from 2-month TRIPEx campaign vs. ICON-LEM simulations

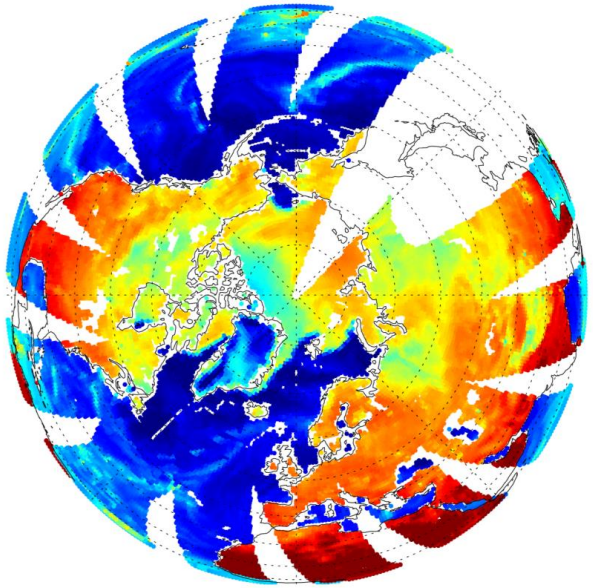


From: Ori et al. (2020) Evaluation of ice particle growth in ICON using statistics of multi-frequency Doppler cloud radar observations, QJRMetS, https://doi.org/10.1002/qj.3875

Assimilation of observations that do not have fully-known physical forward models

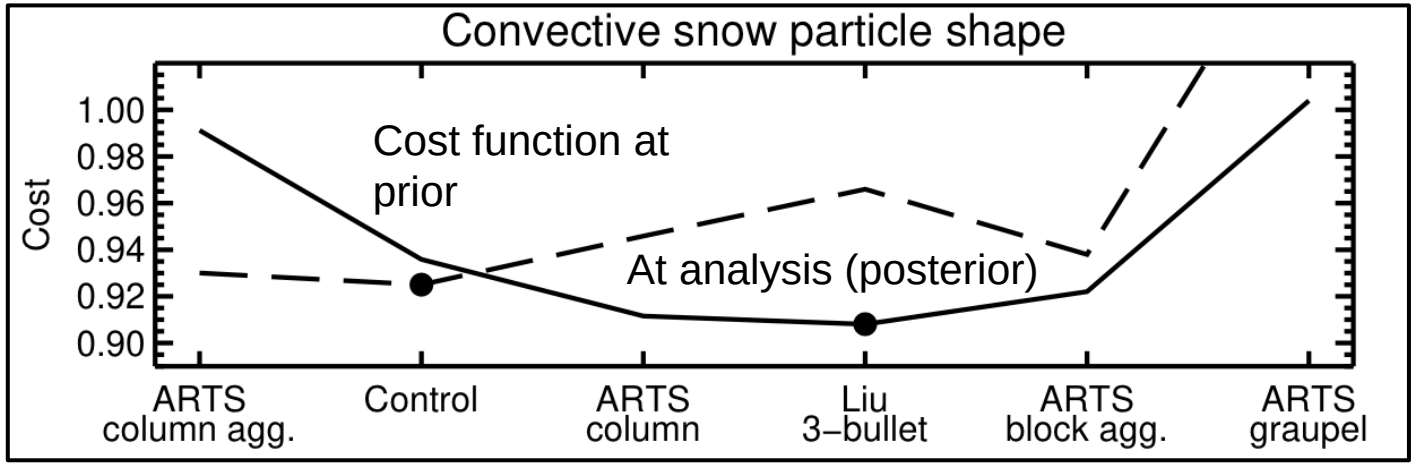
All-surface coupled data assimilation

Observations with joint sensitivities to atmosphere and to complex surfaces(e.g. sea-ice, snow, desert, vegetation)



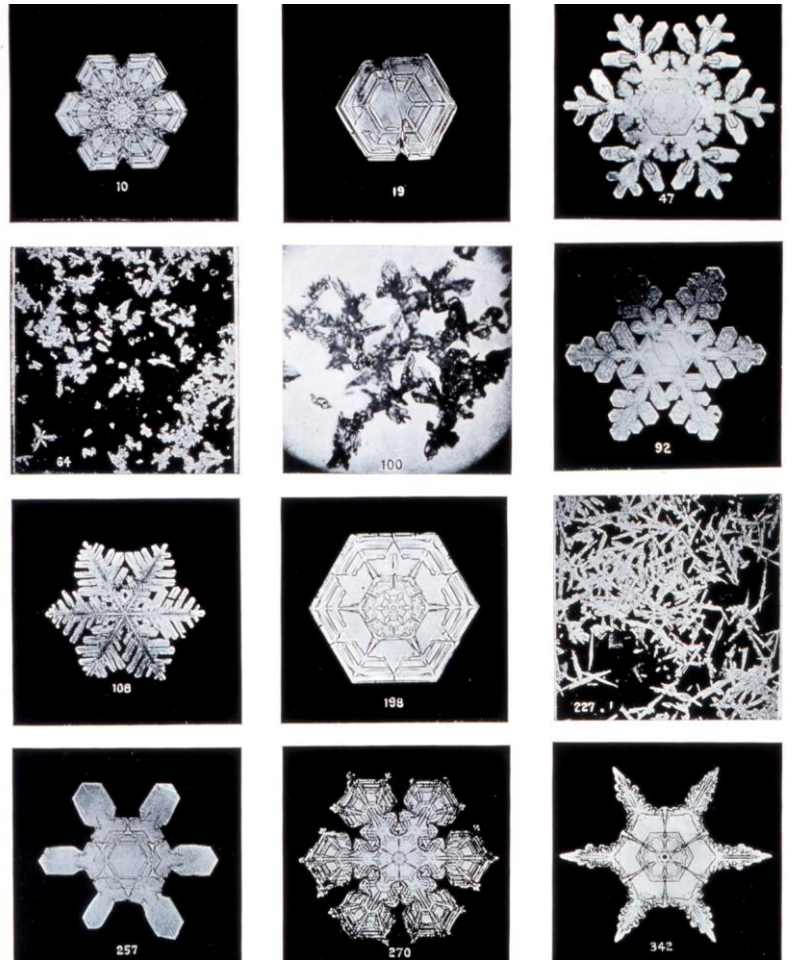
All-sky atmospheric data assimilation

Estimating cloud microphysical assumptions through parameter estimation (Geer, 2021, https://doi.org/10.5194/amt-2021-50)



Forecast model parameter & structural uncertainty

- Truncation of model resolution and dynamical processes therefore needing parametrisations of diffusion, turbulence, gravity waves etc.
- The water problem: phase changes
  - Cloud and preciptiation processes – microscopic scales
  - Sea-ice
- Throughout the earth system – many processes or boundary conditions that are not modeled from physical first principles and/or are strongly heterogeneous in time and space
  - Biosphere, lithosphere, aerosols ...

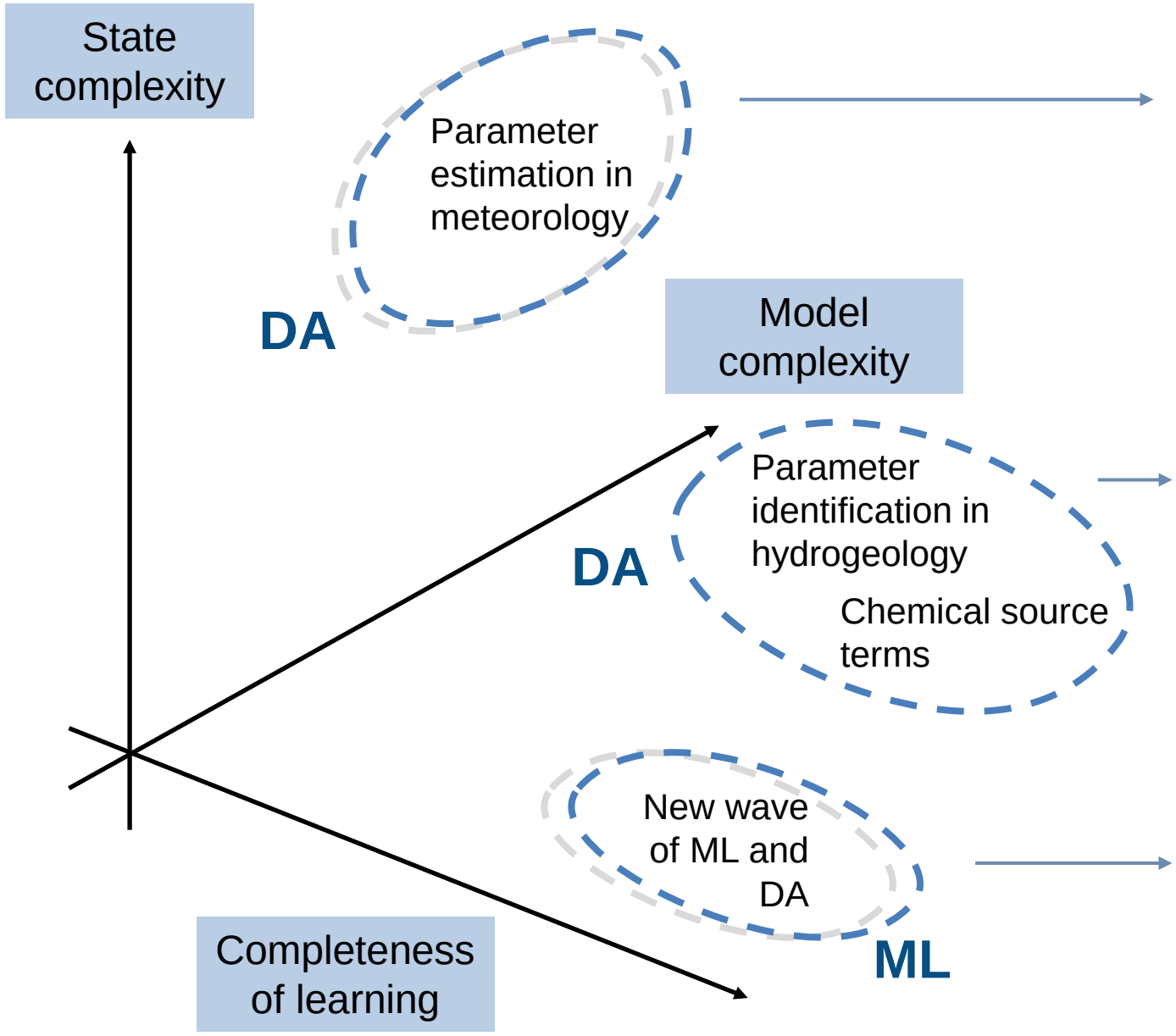


Wilson Bentley (1902, http://www.photolib.noaa.gov)

Model learning spectrum: from parameter estimation to machine learning

ECMWF

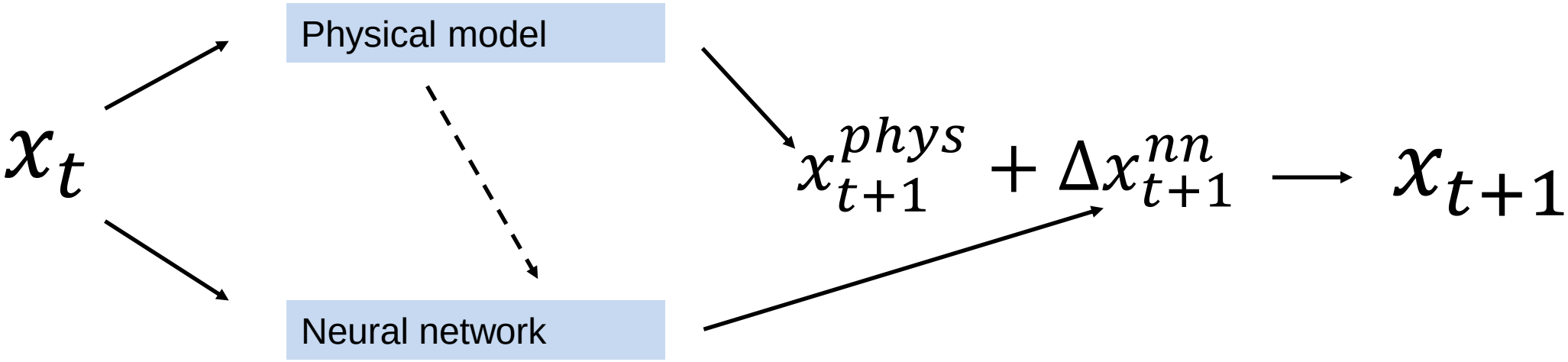
Model learning from observations



- Simultaneous state and parameter estimation:
  - Autoconversion parameter learnt in a GCM (Kotsuki et al., 2020) https://doi.org/10.1029/2019JD031304
  - Roughness length learnt in local area model (Ruckstuhl and Janjić, 2020) https:/doi.org/10.1175/MWR-D-19-0233.1
- Pure parameter estimation, e.g:
  - Groundwater modeling, e.g. hydraulic conductivity (Zhou et al., 2014). https://doi.org/10.1016/j.advwatres.2013.10.014
  - CO_2 source term estimation (Peylin et al., 2013) https://doi.org/10.5194/bg-10-6699-2013
- Model (and sometimes) state estimation in simple models:
  - ML of KS equation (Pathak et al., 2018) https:/doi.org/10.1103/PhysRevLett.120.024102
  - Use DA to infer ODE representation of Lorenz models (Bocquet et al., 2019) https://doi.org/10.5194/npg-26-143-2019

Model correction from observations

- DA-based:
  - Weak-constraint 4D-Var
  - Bias correction (e.g. VarBC)
- ML-based:

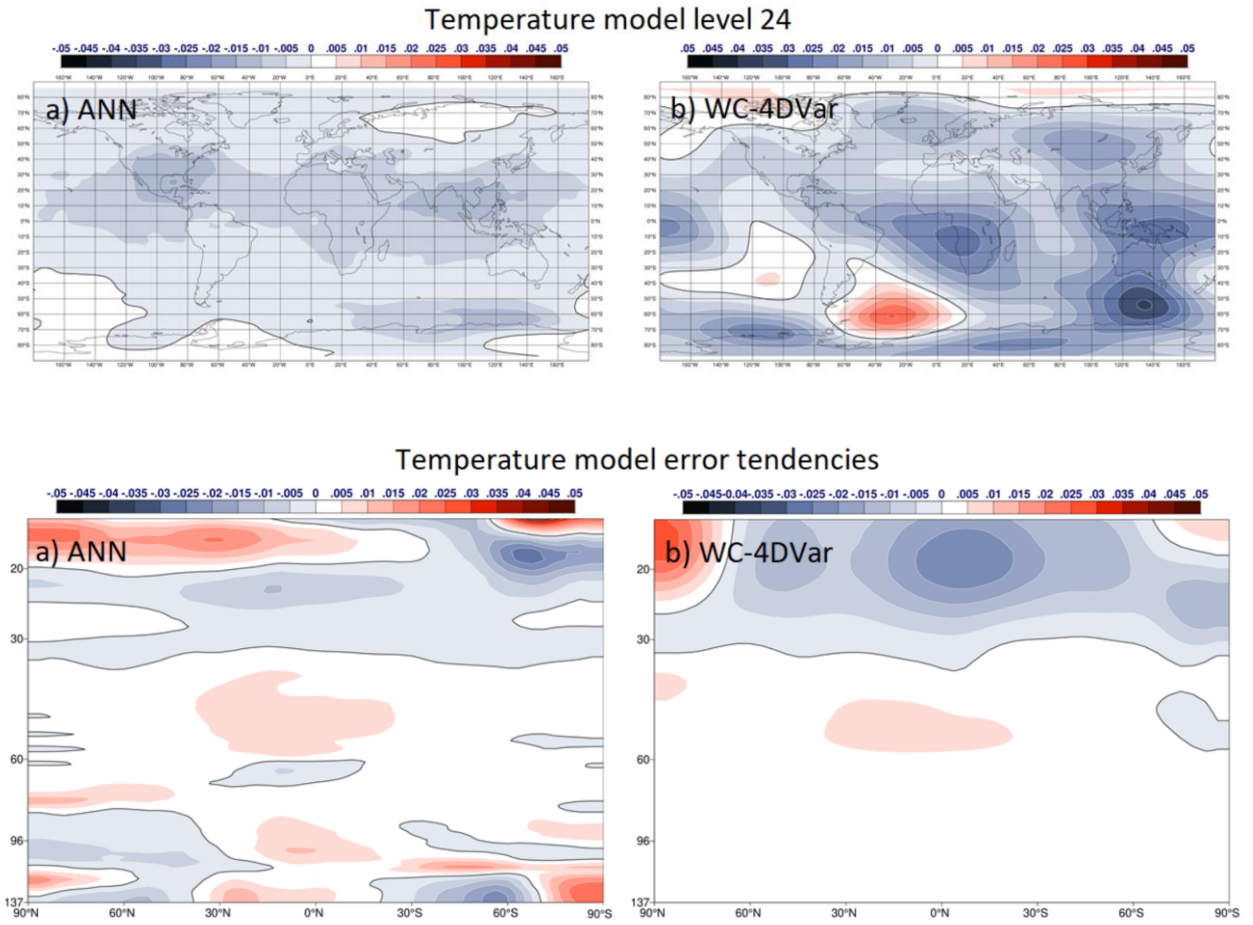


e.g.
Physics guided neural networks (lake temperature): Karpatne et al. (2018) https://arxiv.org/abs/1710.11431
Uresolved scale learning: Brajard et al. (2020) https://arxiv.org/abs/2009.04318

Model error correction using ML – Bonavita and Laloyaux (2020, submitted to JAMES, https://www.essoar.org/doi/abs/10.1002/essoar.10503695.1)

Offline ML – learn weights w of model for model error tendency

$$J(x,w) = (y - h(x,w))^2$$



Weak-constraint 4D-Var, done online (w is a field of model error tendencies)

$$J(x,w) = \ldots + \frac{(w^b - w)^2}{(\sigma^w)^2}$$

Physically constrained machine learning

```
def net_u(self, x, t):
    u = self.neural_net(tf.concat([x,t],1), self.weights, self.biases)
    return u

def net_f(self, x,t):
    u = self.net_u(x,t)
    u_t = tf.gradients(u, t)[0]
    u_x = tf.gradients(u, x)[0]
    u_xx = tf.gradients(u_x, x)[0]
    f = u_t + u*u_x - self.nu*u_xx

    return f

self.loss = tf.reduce_mean(tf.square(self.u_tf - self.u_pred)) + \
            tf.reduce_mean(tf.square(self.f_pred))
```

Neural network

Gradients of the network

Burger's equation

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} - v\frac{\partial^2 u}{\partial x^2} = 0$$

Custom loss function

https://github.com/maziarraissi/PINNs

Raissi, Maziar, Paris Perdikaris, and George Em Karniadakis. "Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations." arXiv preprint arXiv:1711.10561 (2017)

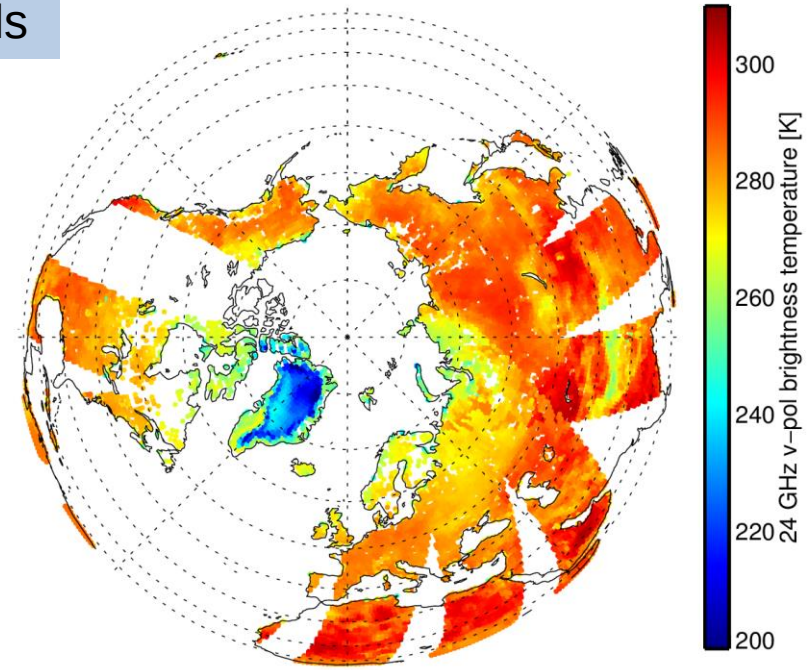
An ML example: microwave land surface observation operator

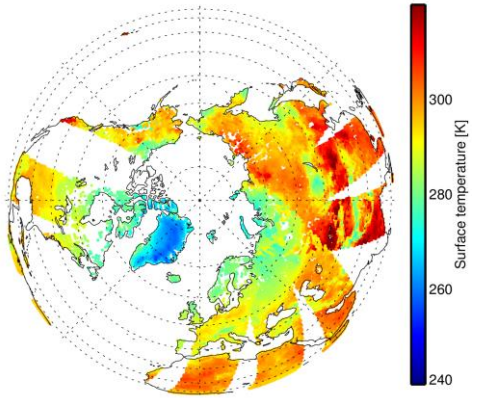
ECMWF

Datasets

AMSR2 24GHz v-pol observations

Labels

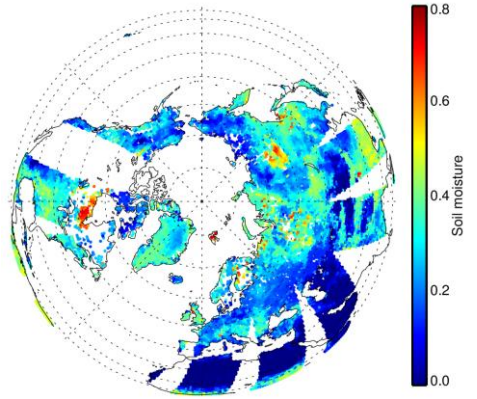




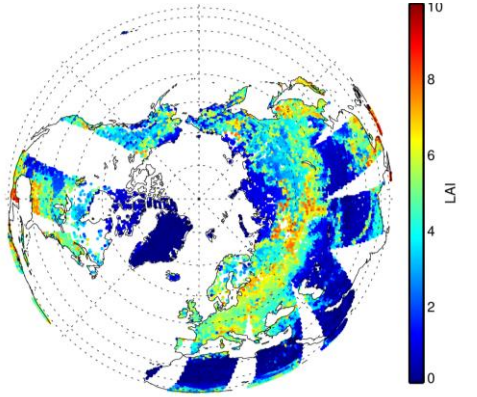
10 possible predictors for the brightness temperature, extracted from the IFS

Skin temperature

Features



Soil moisture

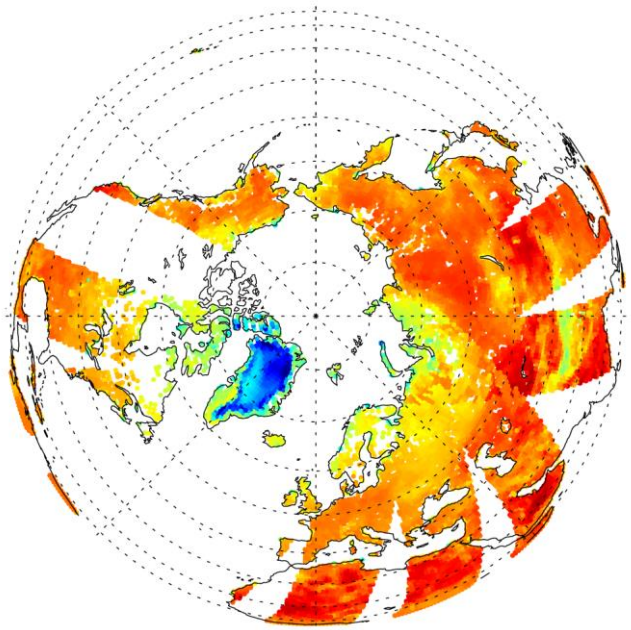


Leaf area index

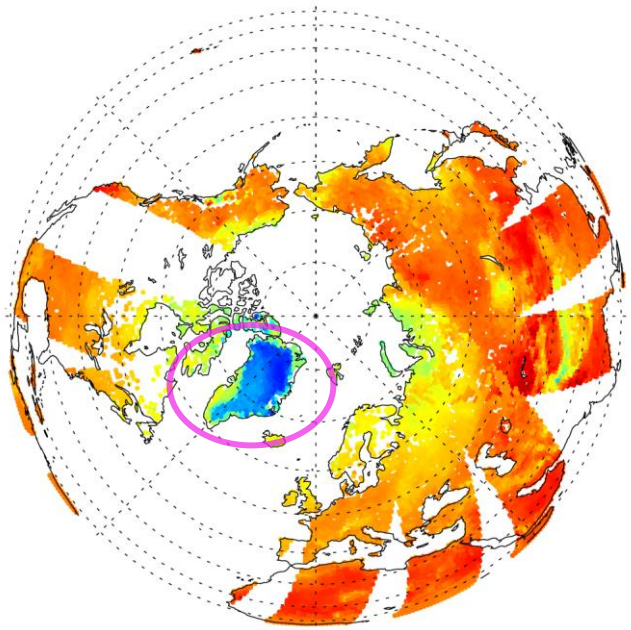
+ orography, snow depth, snow density, integrated water vapour, cloud, rain and snow water contents

Results (ability to fit training dataset)

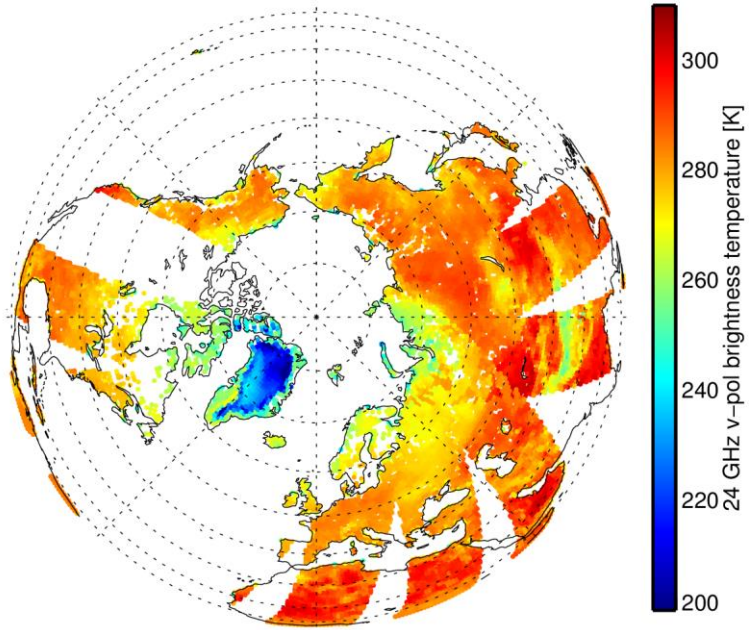
Observations



ML predicted



Physically-based simulation produced within IFS



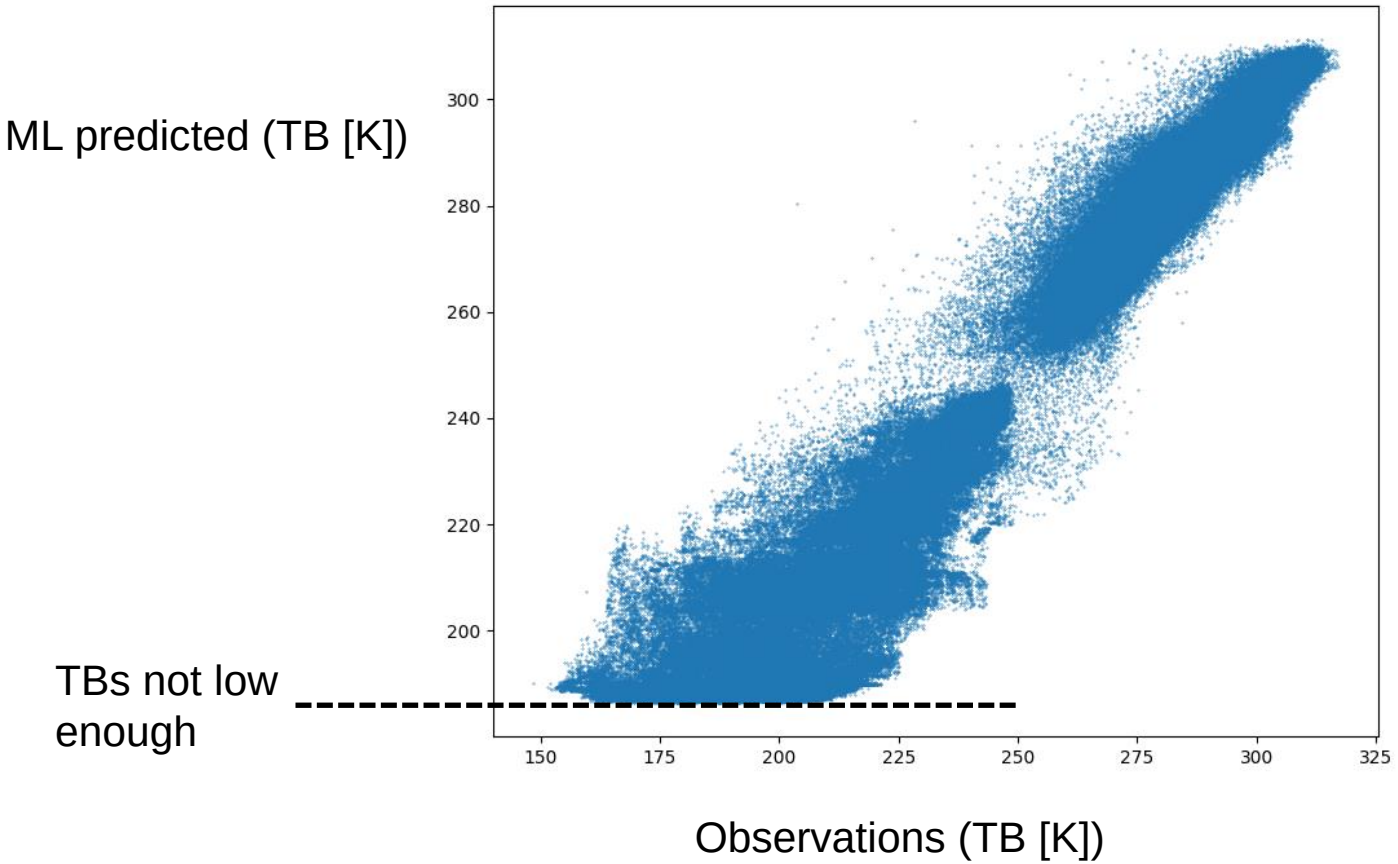
2 hidden layers fully-connected NN, 10-6-1: 183 trainable parameters

Error: ≈8K

RTTOV for atmosphere, dynamical emissivity retrieval for surface emissivity

Error: ≈2K

Results (ability to fit training dataset)



Insufficient information in the known inputs (features)

- Need ML methods to explore the latent space (e.g. GAN, autoencoders)
- Need DA to retrieve the latent variables (auxiliary control variables)

Need a blend of data assimilation and ML

Applications for ML within data assimilation

- Model learning from observations:
  - Observation operators and parameterisations wherever physically-based operators are lacking
  - Bias correction and model error estimation

Blend of data assimilation and ML

- Many other ML applications in DA apart from direct model learning:
  - Acceleration
    - Emulators
    - Error covariance modelling
    - ‘First guess by ML’
  - Automatic differentiation for variational DA tangent-linear and adjoint operators
    - Gravity wave drag scheme emulated by ML and then ML used as TL/adjoint: Hatfield, Chantry et al. (https://www.essoar.org/doi/abs/10.1002/essoar.10506310.1)

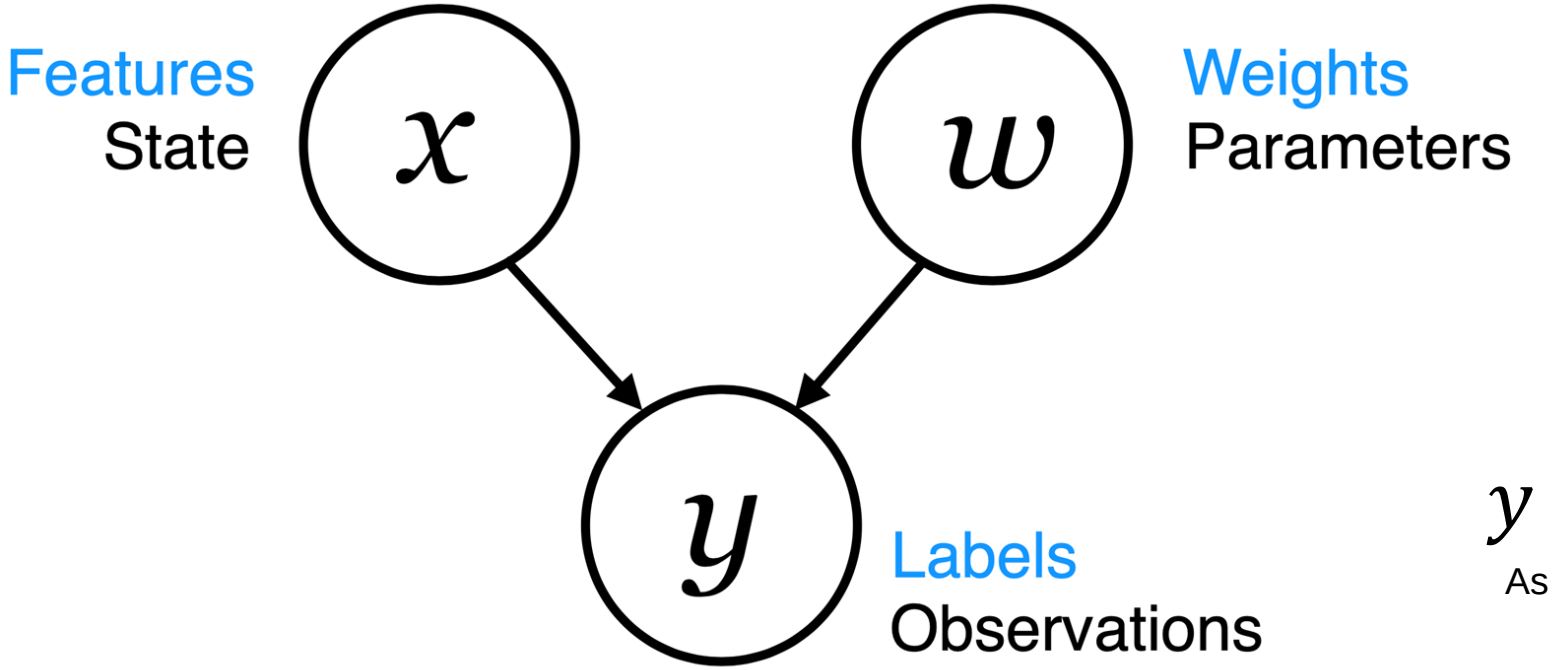
- See also Rosella Arcucci (today), Massimo Bonavita (tomorrow) and several other talks

Bayesian view

From up here there is no ‘ML’ or ‘DA’, just information… ☺

ECMWF

Bayesian equivalence of ML and DA



$$y = h(x, w)$$

As a Bayesian network

Geer (2021) https://doi.org/10.1098/rsta.2020.0089 or preprint: https://doi.org/10.21957/7fyj2811r

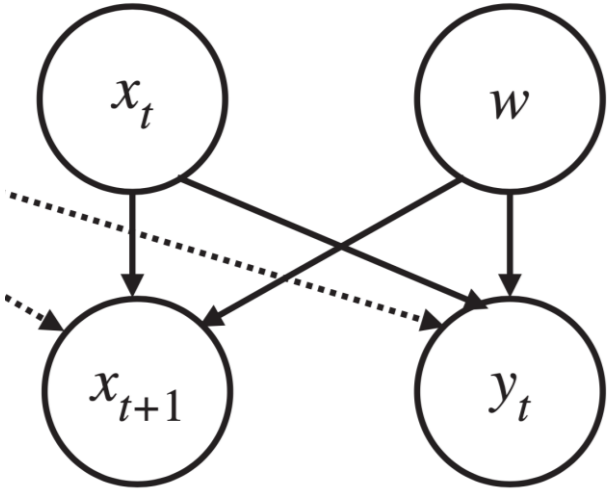
Bocquet et al. (2020) https://arxiv.org/abs/2001.06270

Abarbanel et al. (2018) https://doi.org/10.1162/neco_a_01094

Hsieh and Tang (1998) https://doi.org/10.1175/1520-0477(1998)079%3C1855:ANNMTP%3E2.0.CO;2

Goodfellow et al. (2016) https://www.deeplearningbook.org

Adding the time dimension



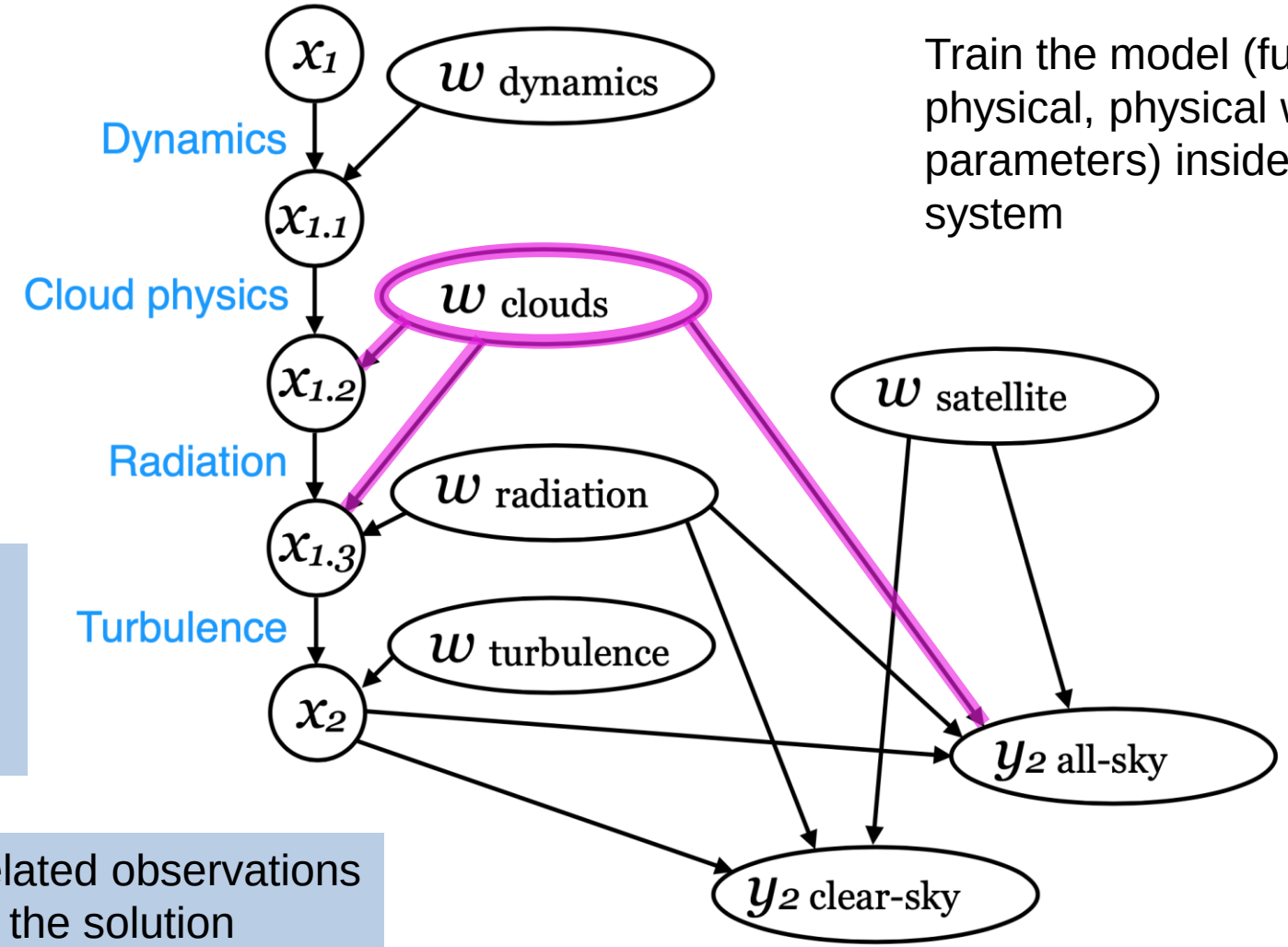
Assume x_{t+1} depends only on initial state x_t and parameters w

$$P(x_{t+1}, w, x_t | y_t) = \underbrace{P(x_{t+1} | w, x_t)}_{\text{Forward model}} \underbrace{P(x_t, w | y_t)}_{\text{Bayes' theorem}}$$

The fundamental equations leading to ML (e.g. RNN, reservoir computing), 4D-Var, EnKF, KF etc.

$$P(x_{t+1}, w | y_t) = \underbrace{\int P(x_{t+1}, w, x_t | y_t)\, \mathrm{d}x_t}_{\text{'Marginalisation'}}$$

Inside an atmospheric model



Train the model (fully empirical, part-physical, physical with unknown parameters) inside the data assimilation system

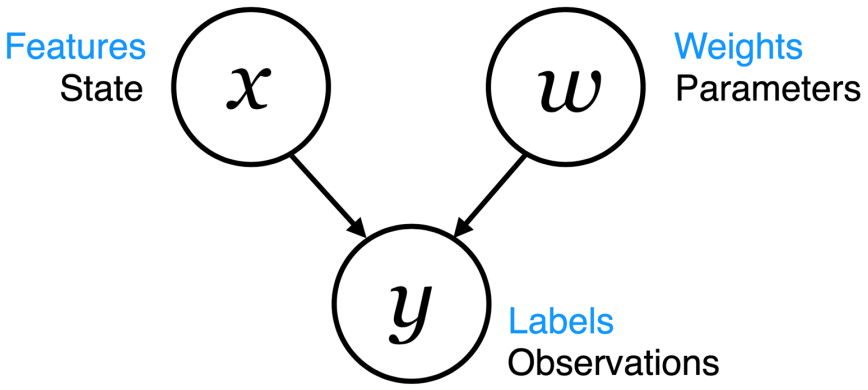
Prior knowledge across the whole system (w) constrains the solution

Apparently unrelated observations constrain the solution

Backup slides

ECMWF

Bayesian networks: representing the factorisation of joint probability distributions



1. Factorise in two different ways using the chain rule of probability

$$P(y,x,w) = P(x|w,y)P(w|y)P(y)$$

$$P(y,x,w) = P(y|x,w)P(x|w)P(w)$$

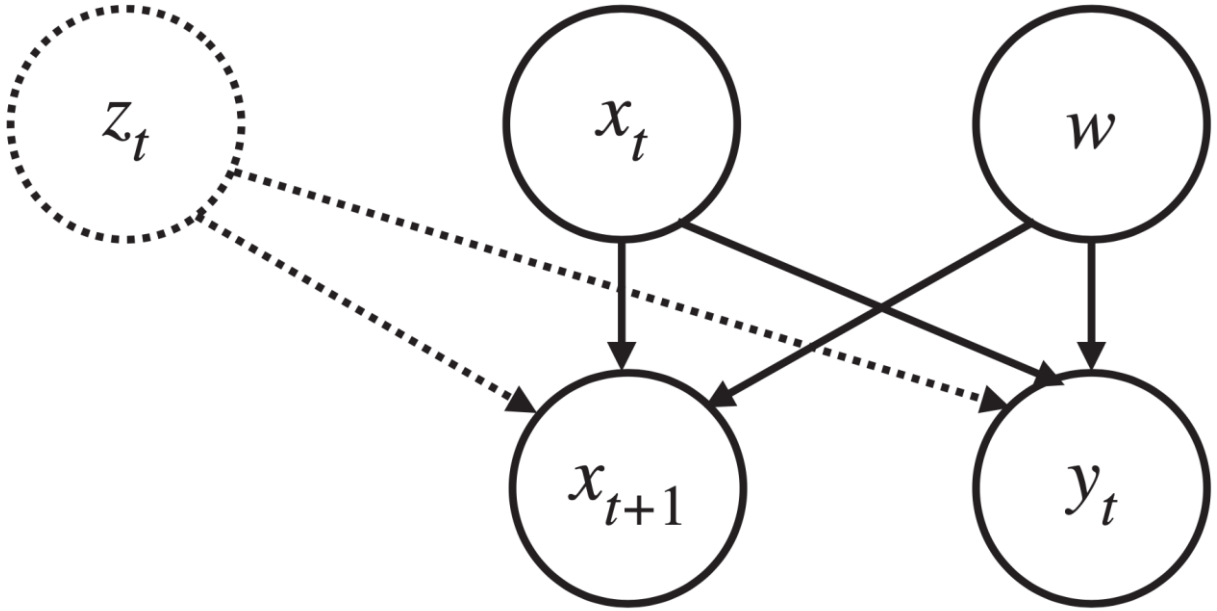
2. Equate the two right hand sides and rewrite

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

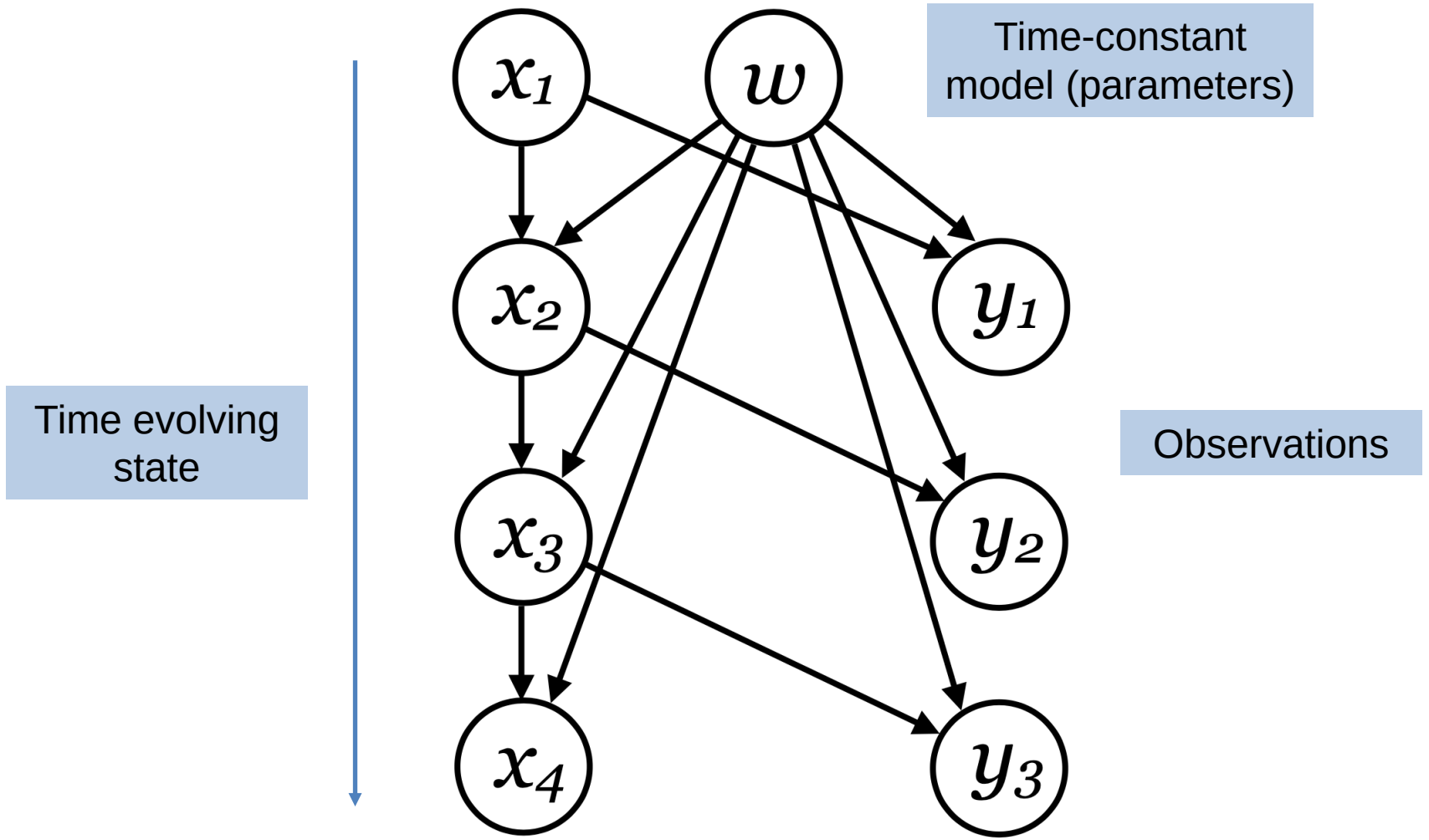
3. Rewrite by putting back the joint distributions of x,w: Bayes' rule

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

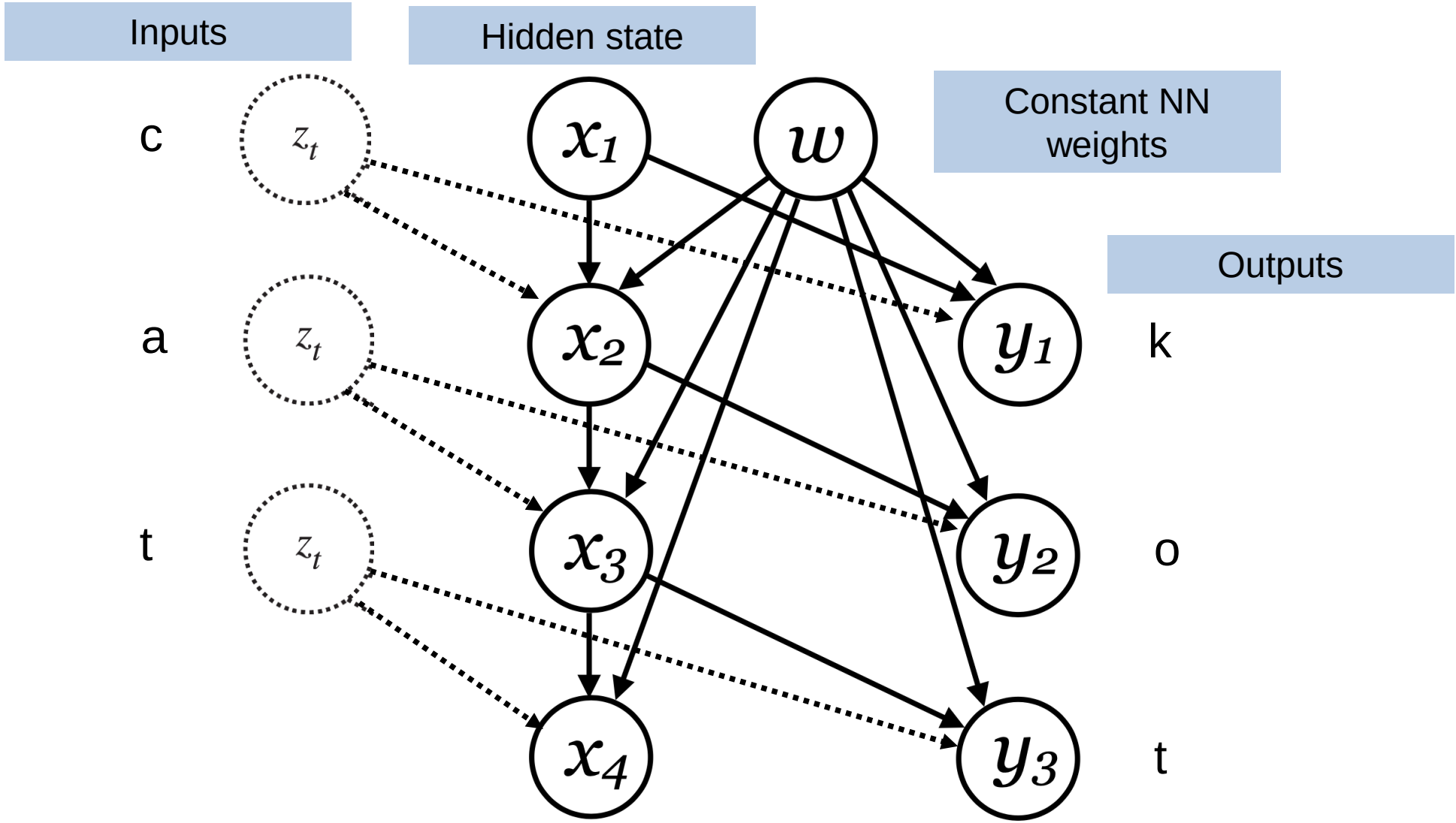
Recursive neural network ≈ cycled data assimilation



Time evolution of state – cycled data asimilation



Recursive neural network



Feedforward neural network - example