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
Quantitative methods, e.g., data assimilation

Earth system observations

Geophysical analyses

Model development

Improved forecasts and geophysical products

Where physical equations are not known, often qualitative methods, including trial and error
What are the current observational challenges?
Data

Exponential data growth?

About 57,300,000 results

Real earth system observations – actually quite precious

Satellite microwave radiance types assimilated in ERA5

The ERA5 global reanalysis, Hersbach et al., 2020, QJRMS,
https://doi.org/10.1002/qj.3883
Observation trends at ECMWF – from Peter Lean

- Move to hyperspectral IR
- Golden age of microwave
- New contributors in the global observing system
- Future: smallsats and constellations

Mostly driven by satellite data:

- Observation volume increasing (but not exponentially)
- Variety may be a bigger problem
IoT challenges? Example: Smartphone observations

Smartphone pressure tendency (circles) vs. 1h precipitation

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

See also https://www.rmets.org/sites/default/files/2019-10/Hintz_RMetS_Presentation.pdf

IoT challenges: variety, uncertainty, data protection
Mean doppler velocity from 2-month TRIPEX campaign vs. ICON-LEM simulations


These kind of highly heterogeneous observations need to be brought into the global observing system
Real observations: sparse, uncertain, **ambiguous**

Satellite observations

**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

SSMIS F-17 channel 13 (19 GHz, v)
Microwave brightness temperatures
3\textsuperscript{rd} December 2014
Real observations: sparse, uncertain, **ambiguous**

Satellite observations  \[ \begin{array}{c} y = h \\ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ \vdots \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ \vdots \end{bmatrix} \end{array} \]

**Equations & parameters** – where sometimes knowledge is quite uncertain

- Gas spectroscopy
- Scattering from hydrometeors
- Cloud and precipitation micro and macro-structure
- Snow / ice grain size and structure

SSMIS F-17 channel 13 (19 GHz, v)
Microwave brightness temperatures
3\(^{rd}\) December 2014
Varied, sparse, uncertain, ambiguous observations

→ Bayesian inverse methods
→ Data assimilation
The forward and inverse problem

Forward model

\[ y = h(x, w) \]

Observations

Geophysical state

Model parameters

Inverse model

No unique solution: ill-posed

The best that observations can do is to provide a statistical improvement in our knowledge of \( x \) and \( w \).
The inverse problem solved by Bayes theorem

\[ P(x, w | y) = K (y, P(y|x, w), P(x, w)) \]

Observations
Geophysical state
Model parameters

Bayes theorem

Probabilistic equivalent of the forward model \( h() \)

(Posterior) Probability of \( x \) and \( w \) given \( y \)
Prior probability of \( x \) and \( w \)
Cost function for variational DA

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

$$J(x, w) = \frac{(y - h(x, w))^2}{(\sigma_y)^2} + \frac{(x^b - x)^2}{(\sigma_x)^2} + \frac{(w^b - w)^2}{(\sigma_w)^2}$$

- $J_y$: Observation term
- $J_x$: Prior knowledge of state
- $J_w$: Prior knowledge of model

variety, ambiguity, sparsity
Data assimilation for numerical weather prediction: (NWP)

Blending short forecasts and new observational data in a statistically optimal way
4D-Variational DA with a perfect model assumption

Geophysical state

Background model trajectory

Analysis model trajectory

Improved forecast

DA adjusts initial state to better fit observations

Observations

Time
But models are not perfect…

• Within the fluid systems (atmosphere / ocean):
  – Truncation of model resolution and dynamical processes (e.g. 8 km currently) therefore needing parametrisations of diffusion, turbulence, gravity waves etc.
  – The water problem: phase changes
    • Cloud and precipitation 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](http://www.photolib.noaa.gov))
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
  - CO₂ 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
Bayesian equivalence of ML and DA

As a Bayesian network

\[ y = h(x, w) \]

Geer (2020)
Bocquet et al. (2020)
Abarbanel et al. (2018)
Goodfellow et al. (2016)

https://doi.org/10.21957/7fyj281lr
https://doi.org/10.1162/neco_a_01094
https://doi.org/10.1175/1520-0477(1998)079%3C1855:ANNMTP%3E2.0.CO;2
https://www.deeplearningbook.org
Cost / loss function equivalence of ML and variational DA

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

ML Loss function

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

DA Cost function

Observation term

Prior knowledge of state

Prior knowledge of model

Weights regularisation
Inside an atmospheric model & data assimilation timestep

One model time-step

Inside an atmospheric model & data assimilation timestep

One model time-step
Learning an improved model of cloud physics (ML or DA)

We want to train a model against observations, but we cannot directly observe gridded intermediate states $x_{1.1}$ and $x_{1.2}$ ... or more precisely model tendencies ...
Inside an atmospheric model

... so train the model inside the data assimilation system
Offline physical parameter search

- Cost function between ECMWF model and SSMIS observations
- 6 parameters from RTTOV-SCATT radiative transfer model

Issues:

- Cost function has multiple minima
- Cost function is based on mapped mean and skewness, not the squared error (errors are highly non-Gaussian)
- Some aspects are quite ambiguous (e.g. snow mixing ratio versus particle shape)

See also e.g. Posselt and Bishop (2012) https://doi.org/10.1175/MWR-D-11-00242.1
Ultimate goal: state and model learning, direct from observations, in real time

→ But first, more achievable ideas
Learning model error

\[ x_t \]

\[ x_{t+1}^{\text{phys}} + \Delta x_{t+1}^{\text{nn}} \rightarrow x_{t+1} \]

- Physical model
- Neural network

e.g.
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)

$$J(x, w) = \ldots + \frac{(w^b - w)^2}{(\sigma w)^2}$$
Where can ML techniques help improve DA applications in the earth sciences?

- Hybrid DA – ML (e.g. error learning)

- ML as a statistical accelerator or surrogate model for inverse methods:
  - In Monte-Carlo inverse frameworks (e.g. ensemble DA, particle filtering, Markov Chain Monte Carlo)
  - In variational frameworks (e.g. 4D-Var)
    - Accelerate the differentiation of the cost function (upcoming slide)

- ML as a modelling component where physical models are lacking
  - E.g. observation operators for surface-sensitive observations

- Dozens of other possibilities (see: this workshop + NOAA AI workshop)
Automatic differentiation for variational DA tangent-linear and adjoint operators: Hatfield, Chantry et al. (ongoing work)

- Train a NN on IFS non-ographic gravity wave drag scheme (NOGWD)
- Differentiate the NN via the chain rule to get:
  - TL (tangent-linear)
  - Adjoint (== back-propagation)
- Use the NN TL and adjoint in the 4D-Var (but keep the physical nonlinear NOGWD scheme)
  - Successful 2-month run of IFS cycling 4D-Var completed

\[ \varepsilon = \langle |H\delta x - h(x + \delta x) - h(x)| \rangle \]

\( \varepsilon \) = H\( \delta x \) - h(x + \delta x) - h(x)

- Tangent linear model
- Nonlinear model
Could DA techniques and the Bayesian approach help improve ML in earth sciences?

- Can data normalisation be replaced with a more quantitative physical description of label and feature uncertainty?
  - In the same way DA quantifies observation and background error matrices

- Use physical layers:
  - As an output layer, physical observation operators can move from gridded geophysical states to observation space
  - Physical constraint layers to take advantage of prior knowledge where it exists (e.g. known physical laws)

Bayesian perspective: treat all variables as probabilistic and represent any prior knowledge probabilistically too

Bayesian perspective: constrain solution using prior physical knowledge (embedded in physical models)
Satellite bias correction using ML

- Application: SSMIS instrument solar-dependent calibration anomalies
- Offline training of a bias model in observation space
  - Learn bias between observations and model as a function of satellite-solar orbital parameters.
- Problems: many latent variables, orbit changes, model changes
Satellite bias correction

- Dynamic NN approach:
  - Re-train each day on yesterday’s biases
  - Learn bias as a function of orbit angle
- This is essentially an offline, nonlinear version of VarBC, which is a standard bias correction technique for DA

![Daily binned bias](image1)

**Dynamic NN model predicted**

![Dynamic NN model predicted](image2)
Sequential learning

• Batch learning
  – relearn on everything, i.e. past and new data, each time?
  – take pre-trained network and run training on new data only

• Sequential learning
  – e.g. Online sequential Extreme Learning Machine (OS-ELM, Liang et al., 2006) [https://doi.org/10.1109/TNN.2006.880583](https://doi.org/10.1109/TNN.2006.880583)
  – e.g. Forecasting daily streamflow using OSELM (Lima, Cannon, Hsieh, 2016) [https://doi.org/10.1016/j.jhydrol.2016.03.017](https://doi.org/10.1016/j.jhydrol.2016.03.017)

• How close are these techniques to the overarching Bayesian framework, i.e. to data assimilation?
Bayesian equivalence of ML and DA → hybridise DA ML
→ just Bayesian inverse methods
Summary

• Observational challenges
  – Real observations are diverse, sparse, ambiguous, uncertain but precious – we are not overwhelmed!
  – Data assimilation arose exactly to deal with diverse, sparse and indirect observations

• Model learning from real observations
  – Hybrid DA – ML to supersedes existing approaches?

• DA to benefit ML?
  – More physical, more Bayesian (e.g. explicit error representations, physical constraint layers)

• ML to benefit DA?
  – Surrogate models for speeding up inference
  – Learned model components where physical models are unavailable
Extras
Observation operators using ML

- Fast observation operators (e.g. radiative transfer operators RTTOV and CRTM) are already ‘ML’ – just using linear regression or heuristic model components
  - Train on physical reference models
  - Train directly between model background and observations

- E.g. Soil moisture
  - Single hidden layer NN trained between SMOS observations and model soil moisture (H-TESSEL or IFS)
  - A version of this is used in the operational ECMWF forecasting system

Time evolution of state

Time evolving state

Time-constant model (parameters)

Observations