Opportunities for Machine Learning to Accelerate Progress in Environmental Modelling

Marc Deisenroth

10 October 2024

ELLIIT Focus Period Symposium on Machine Learning for Climate Science, Linköping, Sweden

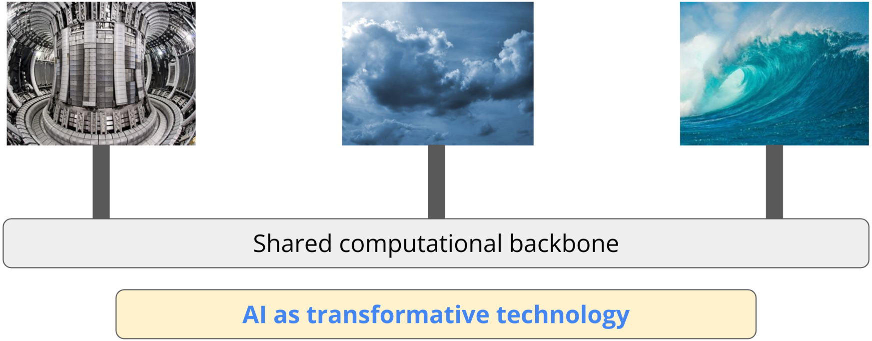
Challenging problems in environment & sustainability

- Challenging (computational) problems in environment and sustainability
  - Modelling and forecasting of very complex, high-dimensional processes (e.g., weather, oceans)
  - “System of systems”, e.g., ecosystems, nuclear fusion
  - Data problems: large, scarce, multimodal

Challenging problems in environment & sustainability

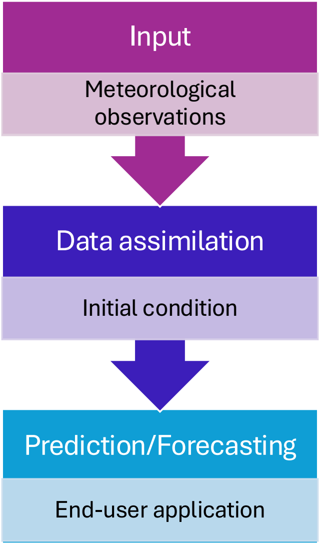
- Challenging (computational) problems in environment and sustainability
  - Modelling and forecasting of very complex, high-dimensional processes (e.g., weather, oceans)
  - "System of systems", e.g., ecosystems, nuclear fusion
  - Data problems: large, scarce, multimodal
- Solving some of these problems can have a positive effect on people and planet
  - Early-warning systems (e.g., natural disasters, tipping points)
  - Clean energy (e.g., nuclear fusion)

Scientific simulation as shared computational backbone



- Seemingly disjoint areas, such as nuclear fusion, weather, and ocean modelling, share a computational backbone
- AI as transformative technology within this backbone, e.g., to improve and accelerate scientific simulation

Typical NWP workflow (amended from (Schultz et al. 2021))



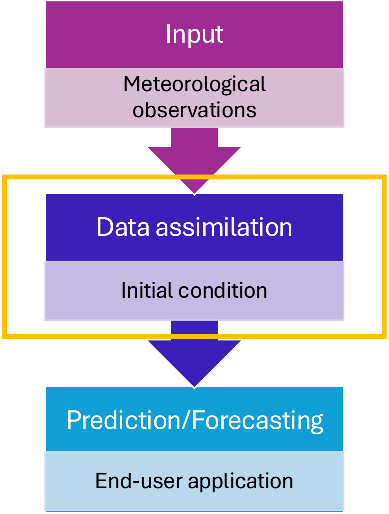
- Traditionally, numerical simulations and solvers are key for NWP
- Progress has been slow with this approach
- AI can be used to significantly accelerate progress (better and faster predictions)

Roadmap

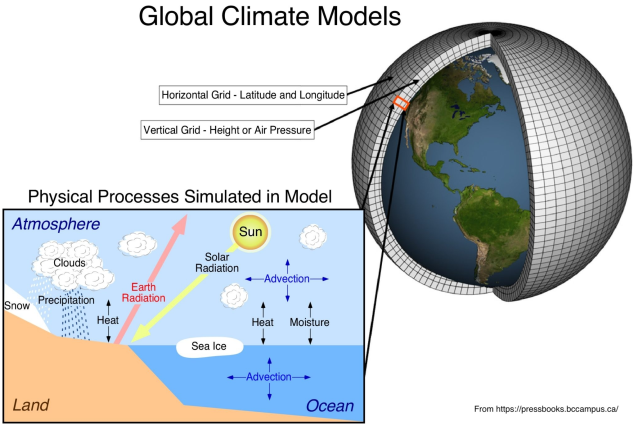
1. Data Assimilation
2. Modelling and Prediction
3. Conclusion
4. What's next?

Data Assimilation

ML for data assimilation



Data assimilation



- Global atmospheric state (e.g., on lat/lon grid)
- Some sparse, noisy observations (e.g., satellite, ground sensors, weather balloons)
- Objective: Infer an updated atmospheric state given observations
- Spatio-temporal inference problem

Formulation

- **Goal:** Update an (unobserved) atmospheric state by incorporating sparse, noisy observations

Formulation

- **Goal:** Update an (unobserved) atmospheric state by incorporating sparse, noisy observations
- Describe global (unknown) atmospheric state as a latent variable $Z(t)$

Formulation

- **Goal:** Update an (unobserved) atmospheric state by incorporating sparse, noisy observations
- Describe global (unknown) atmospheric state as a latent variable $Z(t)$
- Typically, $Z(t)$ evolves in space and time according to a (stochastic) partial differential equation ((S)PDE)

Formulation

- **Goal:** Update an (unobserved) atmospheric state by incorporating sparse, noisy observations
- Describe global (unknown) atmospheric state as a latent variable $Z(t)$
- Typically, $Z(t)$ evolves in space and time according to a (stochastic) partial differential equation ((S)PDE)
- Observations Y_t at some locations

Formulation

- **Goal:** Update an (unobserved) atmospheric state by incorporating sparse, noisy observations
- Describe global (unknown) atmospheric state as a latent variable $Z(t)$
- Typically, $Z(t)$ evolves in space and time according to a (stochastic) partial differential equation ((S)PDE)
- Observations Y_t at some locations
- Bayesian inference problem:

$$\underbrace{p(Z_t|Y_t)}_{\substack{\text{updated}\\ \text{weather state}}} \propto \underbrace{p(Z_t)}_{\substack{\text{current}\\ \text{weather state}}} \underbrace{p(Y_t|Z_t)}_{\substack{\text{observation}\\ \text{model}}}$$

Formulation

- **Goal:** Update an (unobserved) atmospheric state by incorporating sparse, noisy observations
- Describe global (unknown) atmospheric state as a latent variable $Z(t)$
- Typically, $Z(t)$ evolves in space and time according to a (stochastic) partial differential equation ((S)PDE)
- Observations Y_t at some locations
- Bayesian inference problem:

$$\underbrace{p(Z_t|Y_t)}_{\text{updated weather state}} \propto \underbrace{p(Z_t)}_{\text{current weather state}} \underbrace{p(Y_t|Z_t)}_{\text{observation model}}$$

  - Idea: use some form of Kalman filter to solve it

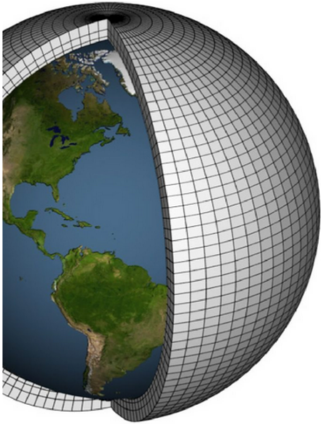
Formulation

- **Goal:** Update an (unobserved) atmospheric state by incorporating sparse, noisy observations
- Describe global (unknown) atmospheric state as a latent variable $Z(t)$
- Typically, $Z(t)$ evolves in space and time according to a (stochastic) partial differential equation ((S)PDE)
- Observations Y_t at some locations
- Bayesian inference problem:

$$\underbrace{p(Z_t|Y_t)}_{\text{updated weather state}} \propto \underbrace{p(Z_t)}_{\text{current weather state}} \underbrace{p(Y_t|Z_t)}_{\text{observation model}}$$

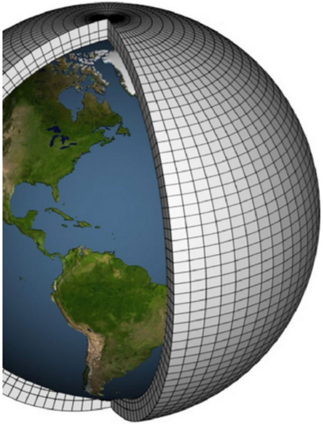
  - Idea: use some form of Kalman filter to solve it
  - Challenge: State space is huge ($\mathcal{O}(10^9)$) ⏵⏵⏵ Compute/memory issues

Gaussian Markov random fields



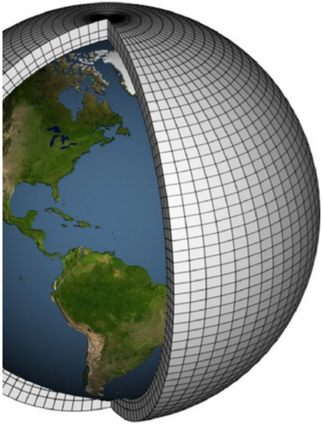
- Exploit grid structure (Lindgren et al., 2011)

Gaussian Markov random fields



- Exploit grid structure (Lindgren et al., 2011)
- Spatially discretise SPDE (at grid nodes)
- Discretised SPDE is a Gaussian Markov random field (GMRF)

Gaussian Markov random fields



- Exploit grid structure (Lindgren et al., 2011)
- Spatially discretise SPDE (at grid nodes)
- Discretised SPDE is a Gaussian Markov random field (GMRF)
- Efficient solvers exist, e.g., INLA (Rue et al., 2009)
- Get Gaussian posterior on marginals of global weather state

Challenges

- GMRF approach works only for linear SPDEs
- Scalablity: GMRF approach does not scale beyond 10^6 many state dimensions

Challenges

- **GMRF approach works only for linear SPDEs**
- Scalablity: GMRF approach does not scale beyond 10^6 many state dimensions

Iterated INLA for data assimilation (Anderka et al.; UAI 2024)

Goal: Extend GMRF approach to nonlinear SPDEs

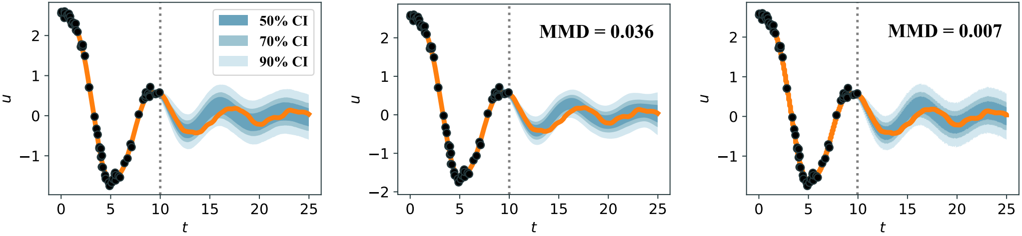
Iterated INLA for data assimilation (Anderka et al.; UAI 2024)

Goal: Extend GMRF approach to nonlinear SPDEs

Key idea

1. Iteratively linearise the dynamical model (SPDE) in time
2. Discretise linearised SPDE in space
3. Use INLA for inference in linearised model

Results



- Left: Ground truth
- Centre: Ensemble Kalman smoother
- Right: Iterated INLA
- Significant improvement over commonly-used approaches for DA

Challenges

- GMRF approach works only for linear SPDEs
- Scalablity: GMRF approach does not scale beyond 10^6 many state dimensions

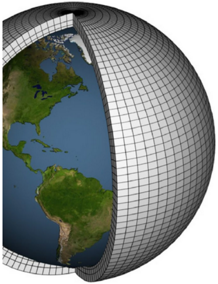
Challenges

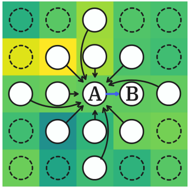
- GMRF approach works only for linear SPDEs
- **Scalablity: GMRF approach does not scale beyond** 10^6 **many state dimensions**

Scalable data assimilation via message passing (Key, Takao, Giles, Deisenroth; Climate Informatics 2024)

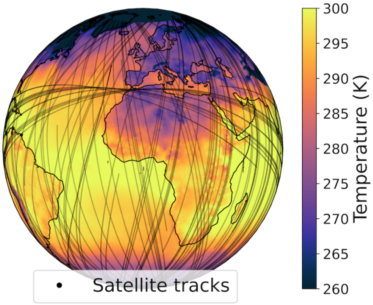
Key idea

- Exploit graph structure of problem for iteratively propagating information to local neighbours via message passing (loopy belief propagation)
- Long-distance information shared via multi-resolution grids
- Light-weight computations that can be run in parallel



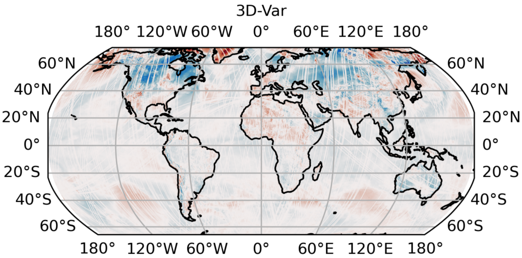


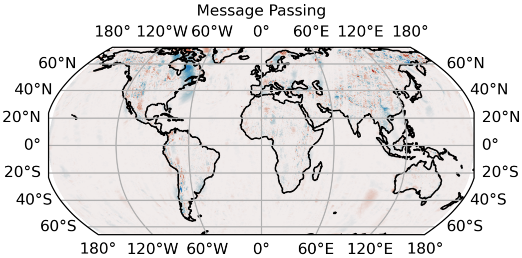
Result: High-resolution surface temperature



- Ground truth: UK Met Office Unified Model at $\approx 10\,\text{km}$ resolution
- $2,500 \times 2,500$ grid ⏵⏵⏵ 3.75M grid points
- 8% of grid has observations (satellite tracks; black lines in figure)

Result: High-resolution surface temperature (2)

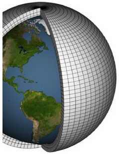


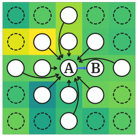


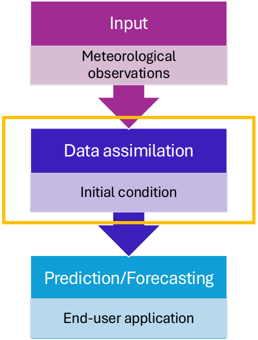
- Message passing significantly outperforms 3D-Var
- Compute time approx. 2 min
- GMRF not applicable
- Overall: Promising paradigm for scalable data assimilation

Re-cap

- Sheer dimensionality of data assimilation causes issues
- GMRF exploits graph structure of the problem, but only works for linear SPDEs and is limited by the size of the graph
- Generalise to nonlinear SPDEs by linearising the nonlinear dynamics
- Scale to larger graphs by using a message-passing paradigm

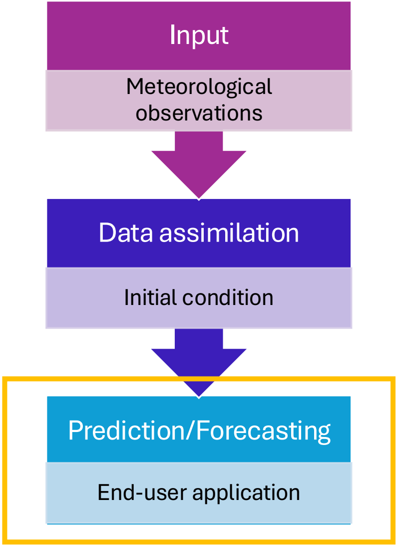






Modelling and Prediction

ML for data modelling and prediction



Data-driven models for forecasting

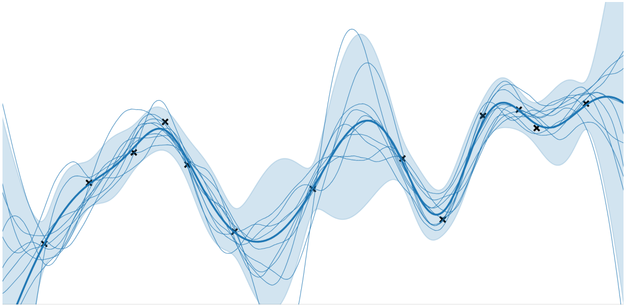
What do we (ideally) want from data-driven forecasting models?
(Focus on regression)

Data-driven models for forecasting

What do we (ideally) want from data-driven forecasting models?
(Focus on regression)

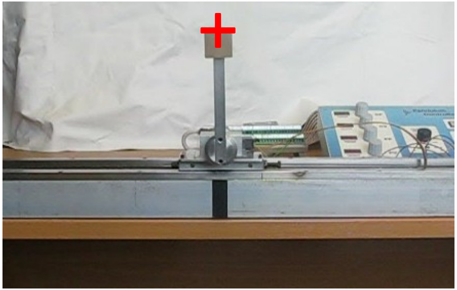
- Flexibility: can model a huge class of functions
- Uncertainty quantification: Equip predictions with meaningful error bars
- Incorporation of prior information
- Scalability to large datasets
- Interpretability
- ...

Gaussian processes for regression



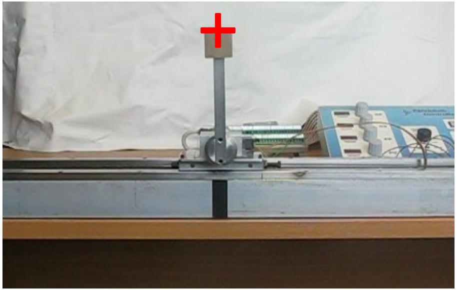
- Flexible (non-parametric) model
- Error bars
- Opportunities to incorporate prior knowledge
- Reasonably interpretable
- Excellent choice for small, low-dimensional datasets

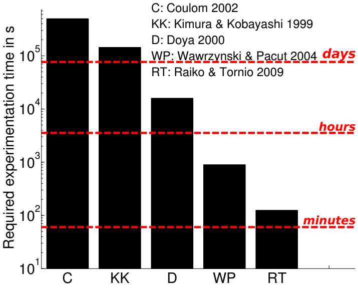
Example: Learning to control a robot (Deisenroth & Rasmussen; ICML 2011)



- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch

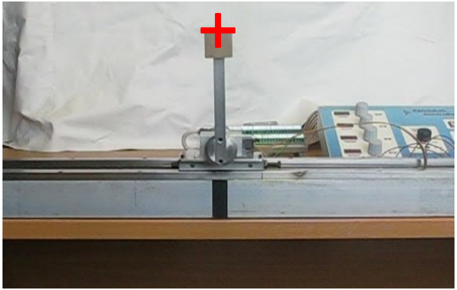
Example: Learning to control a robot (Deisenroth & Rasmussen; ICML 2011)

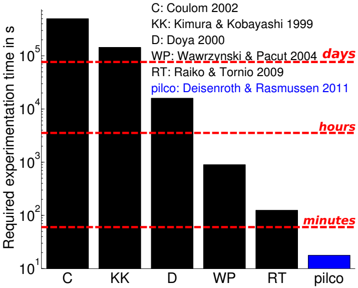




- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch

Example: Learning to control a robot (Deisenroth & Rasmussen; ICML 2011)

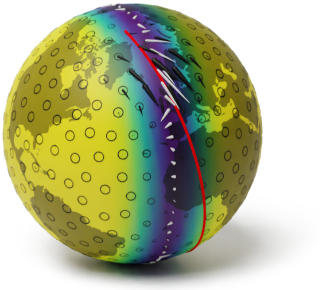




- Swing up and balance a freely swinging pendulum on a cart
- No knowledge about nonlinear dynamics ⏩ Learn from scratch
- Unprecedented learning speed compared to state of the art

Can we use Gaussian processes for environmental modelling and forecasting?

Incorporating underlying geometry

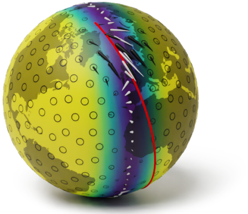


- When modelling global weather, we can build the underlying geometry into a Gaussian process
- We then make predictions on a sphere (Earth)

Gausian processes on Riemannian manifolds (Hutchinson et al.; NeurIPS 2021)

Goal:

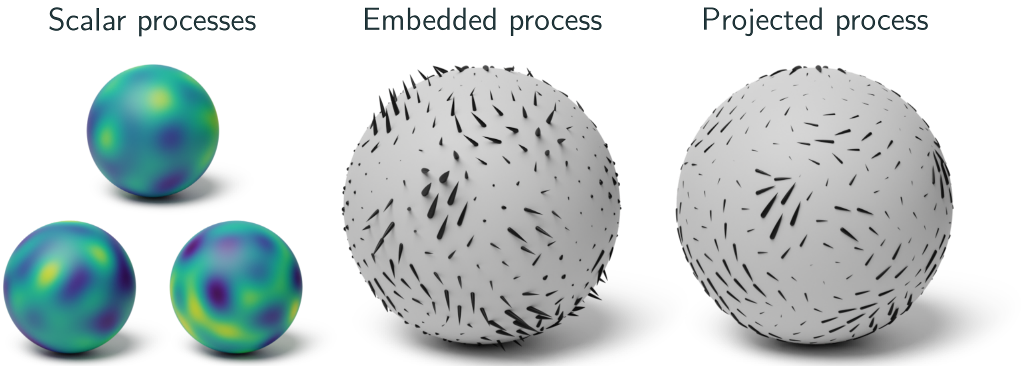
- Define Gaussian vector fields on Riemannian manifolds to make vector-valued predictions that themselves lie on a manifold
- Predictions independent of local coordinate system



Key idea: Projected kernels

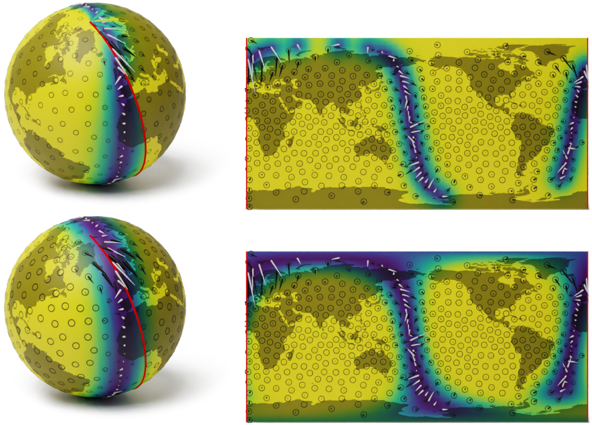
1. Embed manifold in higher-dimensional Euclidean space
2. Construct vector-valued GP in Euclidean space
3. Project GP onto tangent space of manifold

Construction of the projected process



- Three identical scalar GPs (left) are placed on manifold
- Construct vector-valued GP in ambient Euclidean space (centre)
- Project onto tangent space of sphere (right)

Results: Wind measurements along Aeolus satellite trajectory



- Top: Standard Euclidean GP trained on wind measurements in $\mathbb{R}^3$
- Bottom: GP with manifold kernel on $\mathbb{S}^2$

Limits of Gaussian processes

- So far, datasets have been fairly small and were of low dimensionality
- Datasets in environment and sustainability are typically not small and low dimensional; they can be vast and high dimensional

  ⏵⏵⏵ Standard Gaussian processes cannot be applied

Limits of Gaussian processes

- So far, datasets have been fairly small and were of low dimensionality
- Datasets in environment and sustainability are typically not small and low dimensional; they can be vast and high dimensional

  ⏵⏵⏵ Standard Gaussian processes cannot be applied

Challenges in environmental modelling

1. **Scalability** (many data points)
2. High dimensionality

Actually sparse variational Gaussian processes (Cunningham et al.; AISTATS 2023)

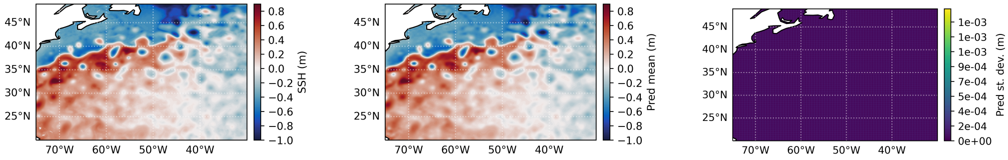
Key idea

Project GP onto a set of compactly supported B-spline basis

Benefits

- Admit use of sparse linear algebra (speed up matrix operations; small memory footprint)
- Allows for use of a large basis / inducing variables ($\gg 10,000$)
- Efficiently model fast-varying spatial phenomena with short length scales

Result



(a) Ground truth.

(b) Predictive mean.

(c) Predictive standard deviation.

- Real-world data from the eNATL60 ocean model over the Gulfstream at $1/60^\circ$ grid resolution (2M training data points; 10,000 basis functions; training in $< 2\,\text{min}$)
- Predict at a regular grid with $1/12^\circ$ resolution
- Predictive mean closely matches ground truth

Challenges in environmental modelling

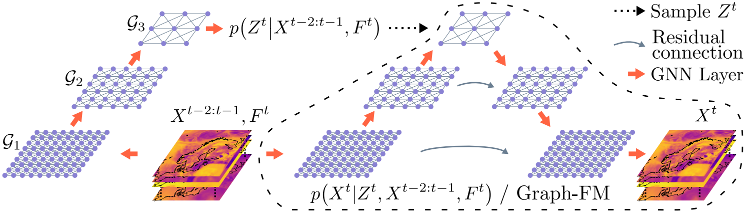
1. Scalability (many data points)
2. **High dimensionality**

Encode-process-decode

General approach:

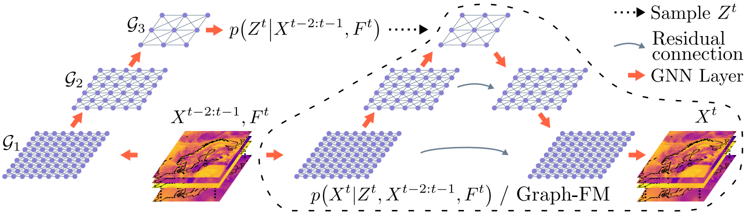
1. Find lower-dimensional embedding of high-dimensional data
2. Work with embedded data (e.g., forecasting)
3. Project back into original data space

Probabilistic weather forecasting with hierarchical GNNs (Oskarsson et al.; arXiv:2406.04759)



- Graph-based ensemble via combination of latent-variable model with hierarchical graph neural network (GNN)

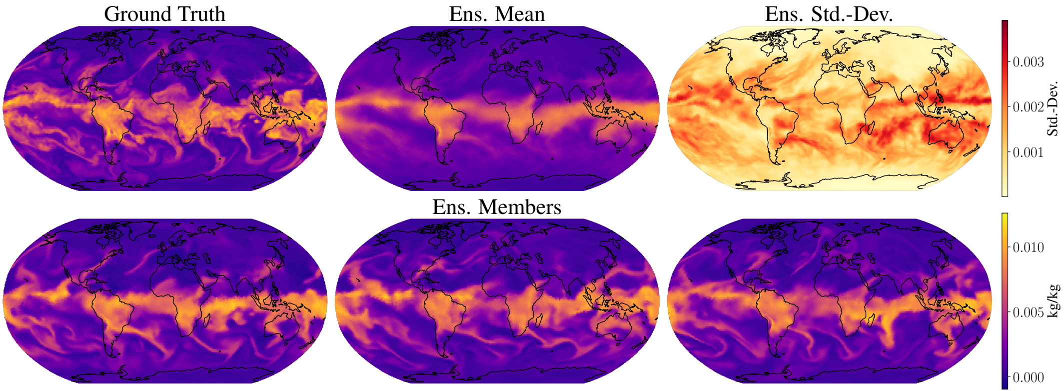
Probabilistic weather forecasting with hierarchical GNNs (Oskarsson et al.; arXiv:2406.04759)



- Graph-based ensemble via combination of latent-variable model with hierarchical graph neural network (GNN)
- Distribution modelled in lower-dimensional latent space
- Sampled forecasts are spatially coherent

$$\underbrace{p(X_t|X_{t-1}, X_{t-2}, F_t)}_{\text{predict in data space}} = \int \underbrace{p(Z_t|X_{t-1}, X_{t-2}, F_t)}_{\text{predict in latent space}} \underbrace{p(X_t|Z_t, X_{t-1}, X_{t-2}, F_t)}_{\text{map back to data space}} \, dZ_t$$

Example forecasts (10 days ahead)

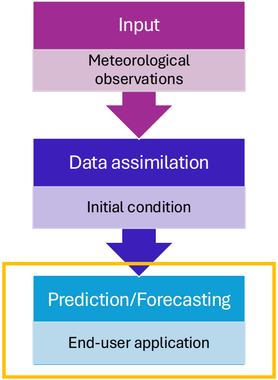


- Example ensemble forecast for specific humidity at 700 hPa (q700)
- Calibrate error bars via conformal prediction (Gopakumar et al.; arXiv:2408.09881)

Re-cap

Modelling and forecasting are challenging problems in environmental systems

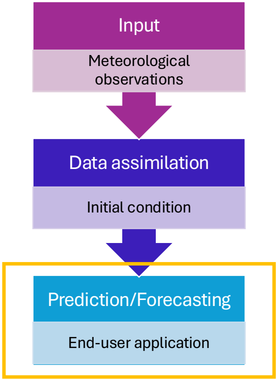
- Gaussian processes good in low dimensions
  - Flexible (non-parametric)
  - Incorporation of underlying geometric properties
  - Error bars
  - Scale GPs to large datasets



Re-cap

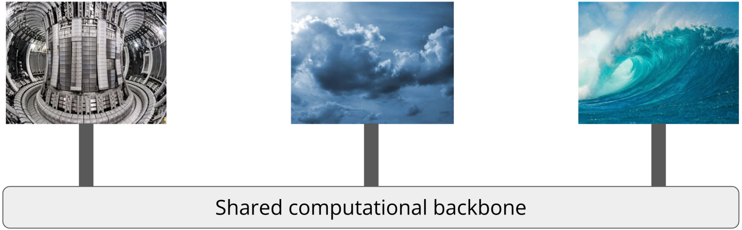
Modelling and forecasting are challenging problems in environmental systems

- Gaussian processes good in low dimensions
  - Flexible (non-parametric)
  - Incorporation of underlying geometric properties
  - Error bars
  - Scale GPs to large datasets
- Different approach for high-dimensional problems
  - Hierarchical graph neural networks + latent variables
  - Conformal prediction ⏵⏵⏵ Meaningful error bars

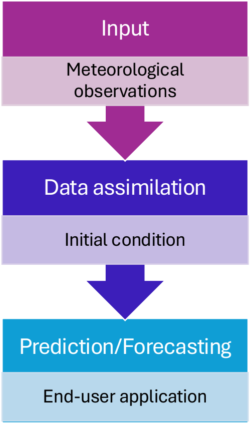


Conclusion

Summary



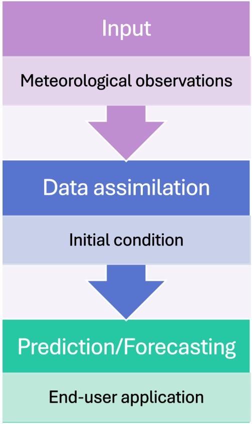
AI as transformative technology



- Challenging problems in environmental modelling, but scientific simulation is a shared computational backbone where AI can play a transformative role
- AI for data assimilation and forecasting within the traditional NWP workflow

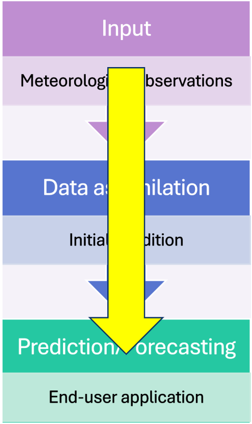
What's next?

End-to-end forecasts



End-to-end forecasts

- End-to-end forecasting: Go straight from observations to forecasts
- AtmoRep (Lessig et al., 2023)
- Aardvark (Vaughan et al., 2024)



Weather generator

- European effort, coordinated by ECMWF
- Multi-task end-to-end forecasting: from nowcasting to climate
- Multi-modal data (re-analysis, simulation, observations)
- Single model

Physics-infused machine learning

- Purely data-driven forecasting systems has shown promise
- Incorporation of domain knowledge (e.g., physics, geometry) into data-driven models should improve things (?)
  - Less data hungry, faster training
  - Better extrapolation (e.g., data-sparse regions, climate)
  - More interpretable ⯈⯈⯈ Easier to operationalize

References

- Anderka, R., Deisenroth, M. P., and Takao, S. (2024). **"Iterated INLA for State and Parameter Estimation in Nonlinear Dynamical Systems".** In: *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*.
- Cunningham, H. J., de Souza, D. A., Takao, S., van der Wilk, M., and Deisenroth, M. P. (2023). **"Actually Sparse Variational Gaussian Processes".** In: *Proceedings of the Conference on Artificial Intelligence and Statistics (AISTATS)*.
- Deisenroth, M. P. and Rasmussen, C. E. (2011). **"PILCO: A Model-Based and Data-Efficient Approach to Policy Search".** In: *Proceedings of the International Conference on Machine Learning (ICML)*.

References ii

- Gopakumar, V., Gray, A., Oskarsson, J., Zanisi, L., Pamela, S., Giles, D., Kusner, M., and Deisenroth, M. P. (2024). **"Uncertainty Quantification of Pre-Trained and Fine-Tuned Surrogate Models using Conformal Prediction".** In: *arXiv:2408.09881*.
- Hutchinson, M., Terenin, A., Borovitskiy, V., Takao, S., Teh, Y. W., and Deisenroth, M. P. (2021). **"Vector-valued Gaussian Processes on Riemannian Manifolds via Gauge Independent Projected Kernels".** In: *Advances in Neural Information Processing Systems (NeurIPS)*.
- Key, O., Takao, S., Giles, D., and Deisenroth, M. P. (2024). **"Scalable Data Assimilation with Message Passing".** In: *Climate Informatics*.
- Lessig, C., Luise, I., Gong, B., Langguth, M., Stadtler, S., and Schultz, M. (2023). **"AtmoRep: A Stochastic Model of Atmosphere Dynamics Using Large Scale Representation Learning".** In: *arXiv:2308.13280*.

- Lindgren, F., Rue, H., and Lindström, J. (2011). **"An Explicit Link between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach".** In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4, pp. 423–498.
- Oskarsson, J., Landelius, T., Deisenroth, M. P., and Lindsten, F. (2024). **"Probabilistic Weather Forecasting with Hierarchical Graph Neural Networks".** In: *arXiv:2406.04759*.
- Rue, H., Martino, S., and Chopin, N. (2009). **"Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations".** In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, pp. 319–392.

References iv

- Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., and Stadtler, S. (2021). **"Can Deep Learning Beat Numerical Weather Prediction?"** In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379 (2194), p. 20200097.
- Vaughan, A., Markou, S., Tebbutt, W., Requeima, J., Bruinsma, W. P., Andersson, T. R., Herzog, M., Lane, N. D., Chantry, M., Hosking, J. S., and Turner, R. E. (2024). **"Aardvark Weather: End-to-end Data-driven Weather Forecasting".** In: *arXiv:2404.00411*.