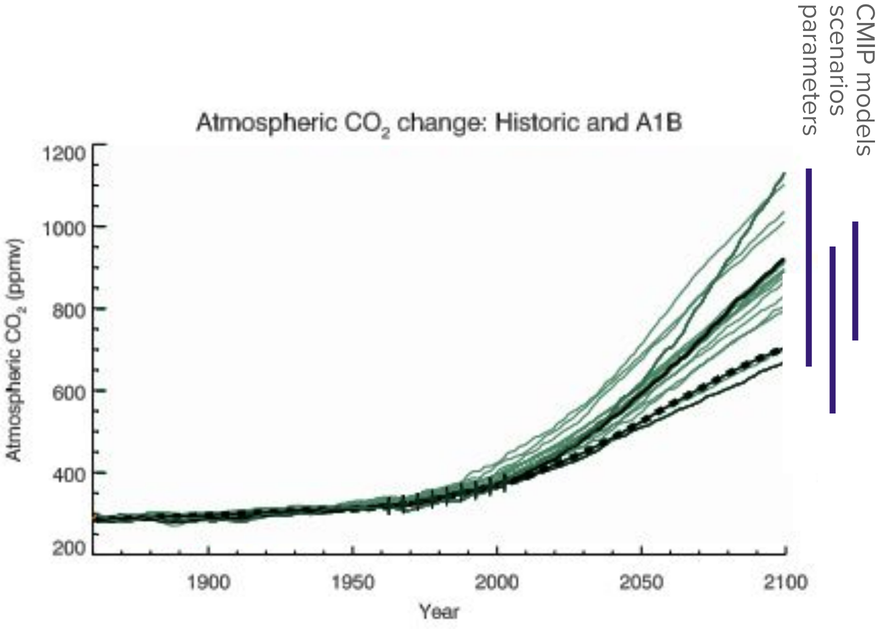
Bridging Physics and Data: Parameter Estimation and Emulation of Land Surface Models with Machine Learning

Nina Raoult
with contributions from LSCE, ECMWF, and the University of Exeter



Booth et al. (2012)

Funded by the European Union Destination Earth
Implemented by ECMWF esa EUMETSAT

JULES

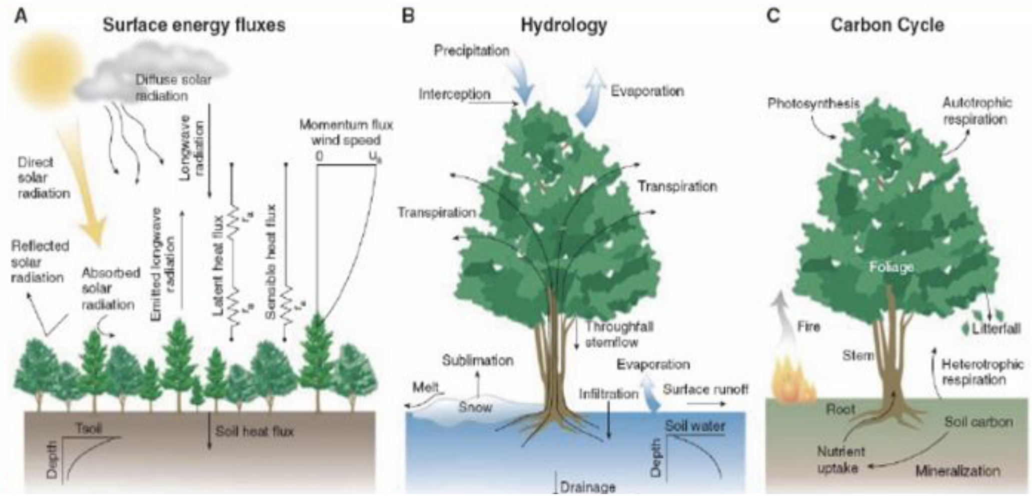
Joint UK Land Environment Simulator

ORCHIDEE

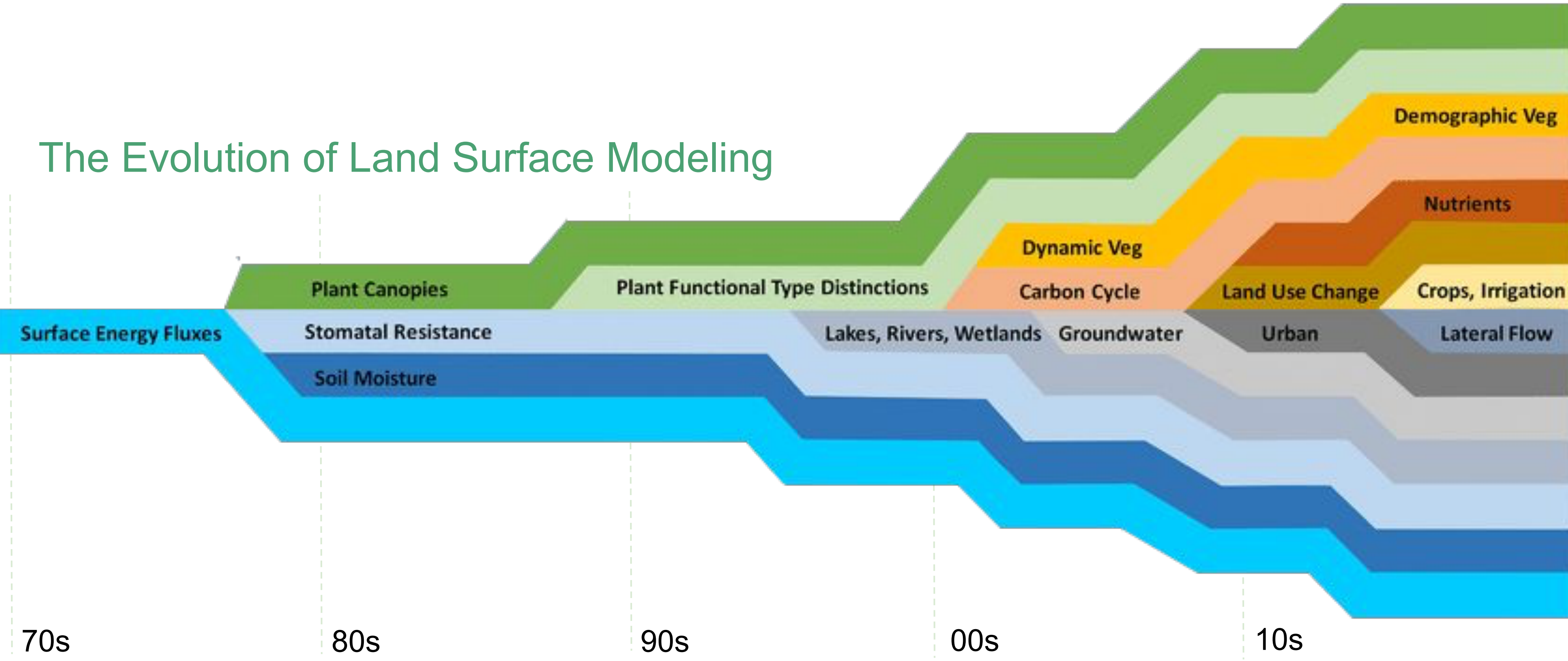
LAND SURFACE MODEL

ec Land

Solve for Energy / Water / Carbon / Nitrogen budgets

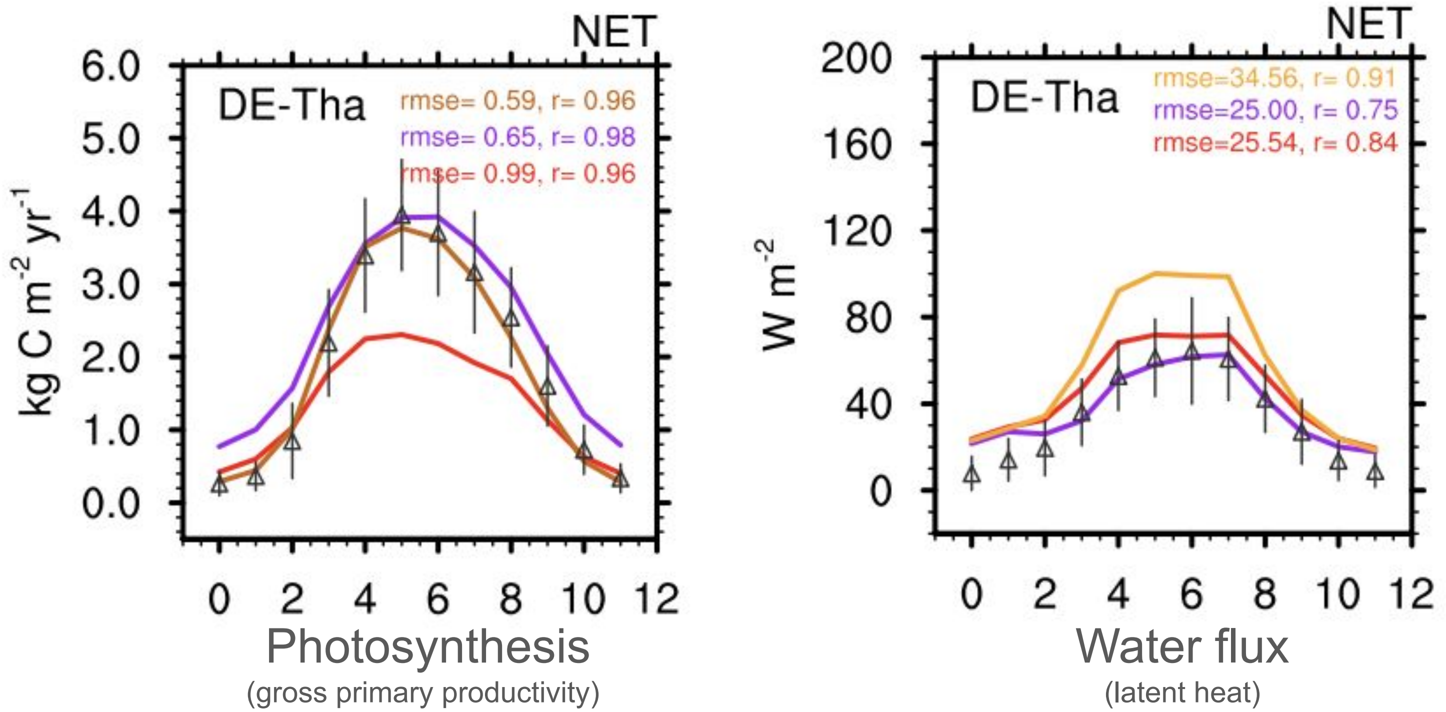


The Evolution of Land Surface Modeling



Fisher & Koven (2020)

Calibrating parameters can outperform model structural changes

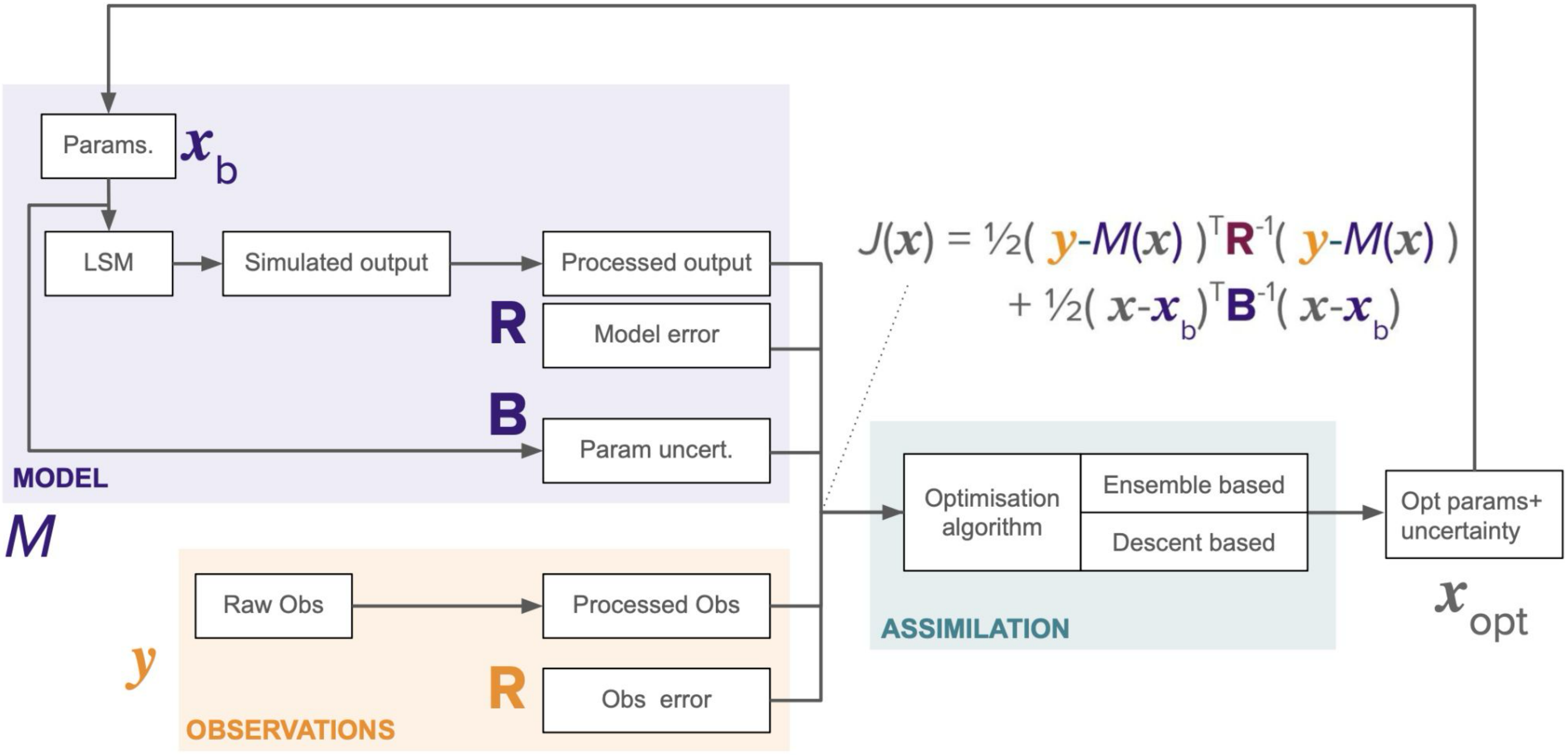


standard JULES

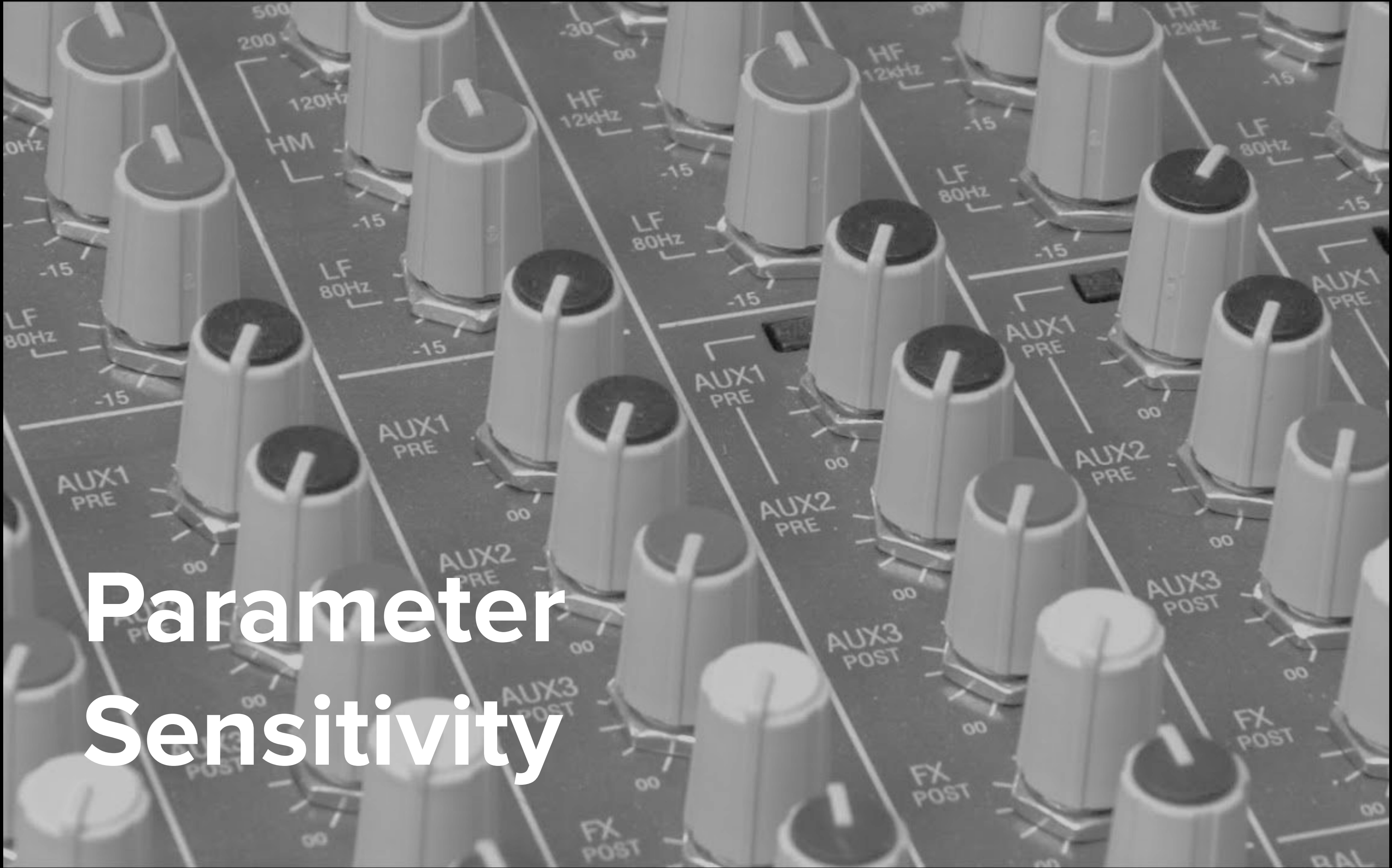
calibrated JULES

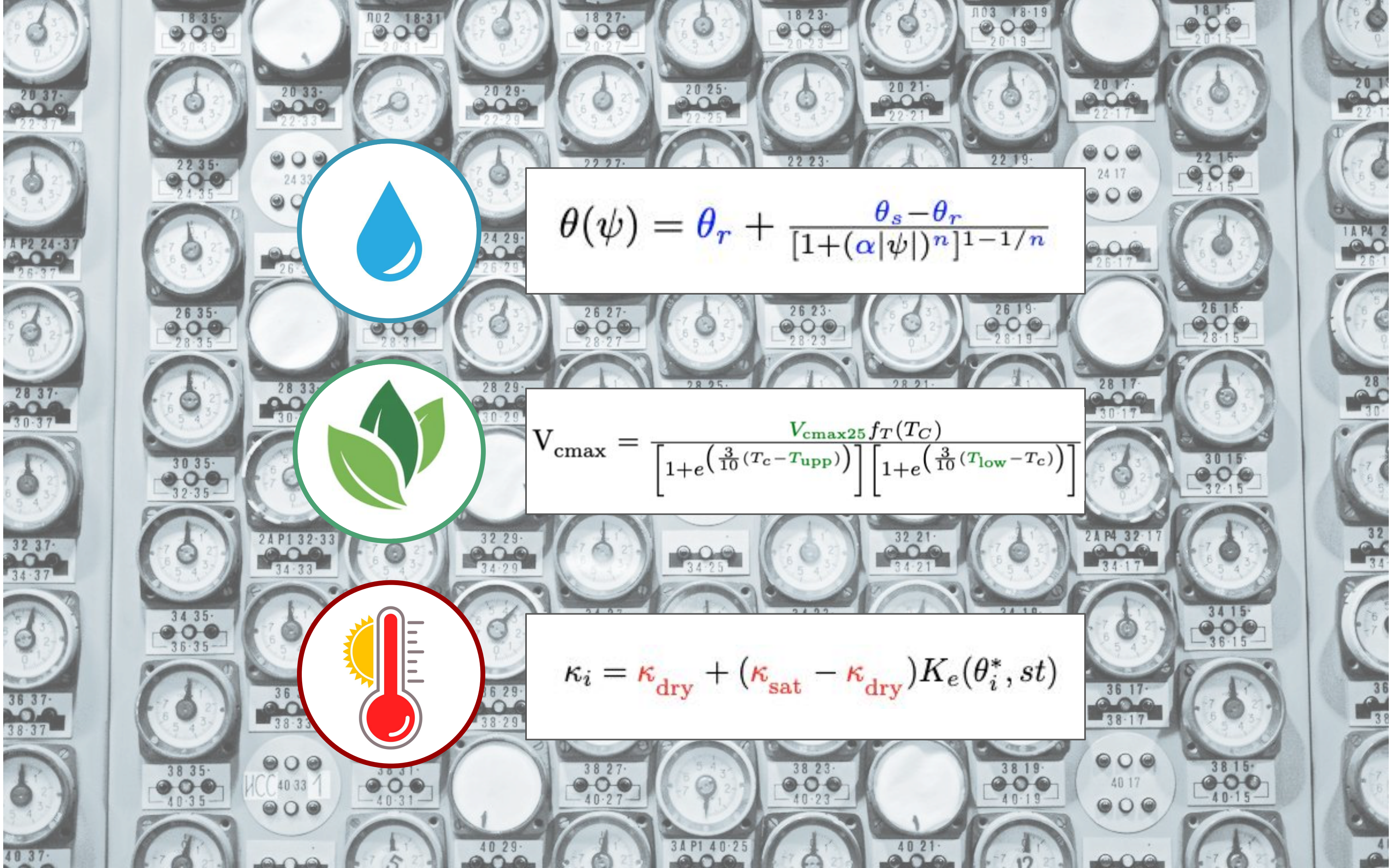
JULES with structural changes

A data assimilation framework allows us to get posterior uncertainty



Parameter Sensitivity



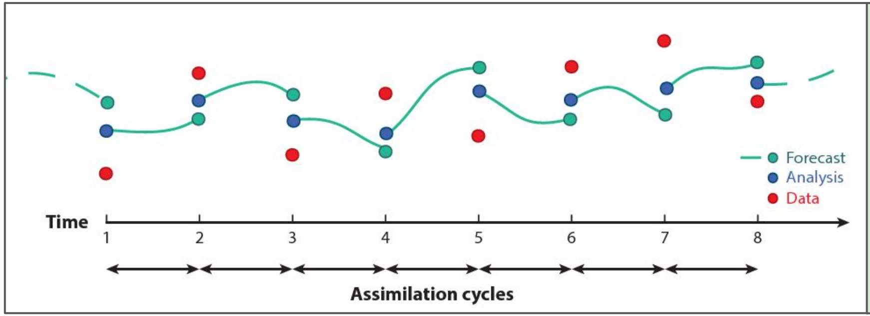


$$\theta(\psi) = \theta_r + \frac{\theta_s - \theta_r}{[1+(\alpha|\psi|)^n]^{1-1/n}}$$

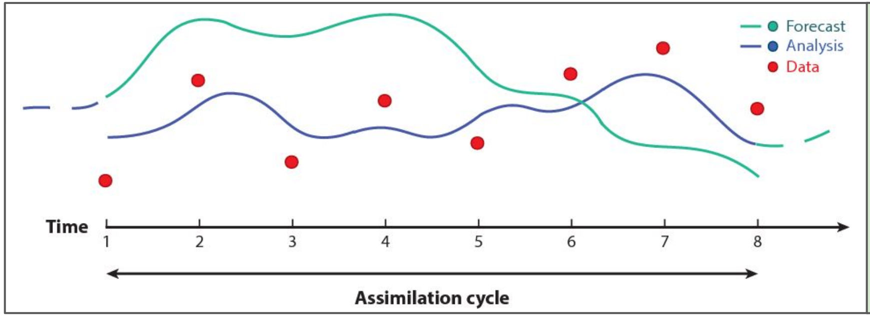
$$\mathrm{V}_{\mathrm{cmax}} = \frac{V_{\mathrm{cmax25}} f_T(T_C)}{\left[1+e^{\left(\frac{3}{10}(T_c - T_{\mathrm{upp}})\right)}\right]\left[1+e^{\left(\frac{3}{10}(T_{\mathrm{low}} - T_c)\right)}\right]}$$

$$\kappa_i = \kappa_{\mathrm{dry}} + (\kappa_{\mathrm{sat}} - \kappa_{\mathrm{dry}})K_e(\theta_i^*, st)$$

Parameters, by definition, are **time-invariant** favouring batched assimilation methods

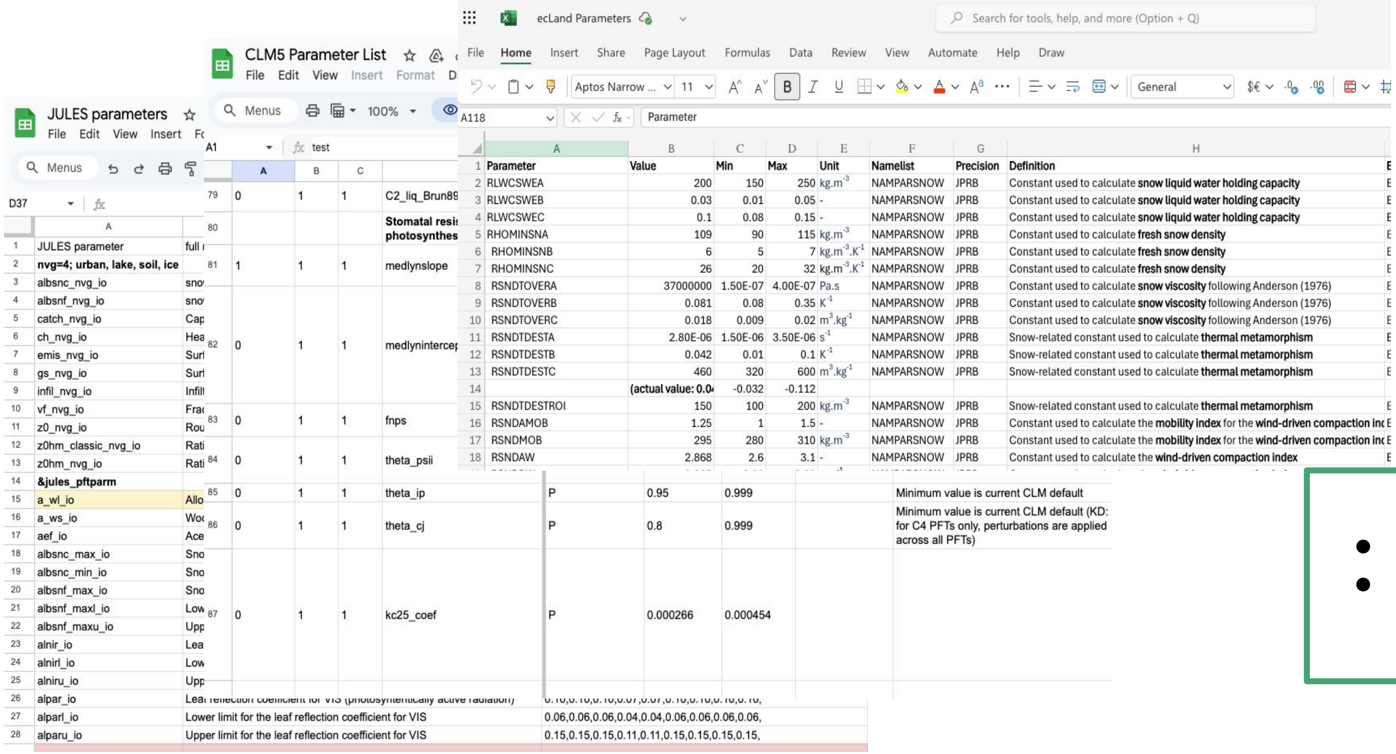


Sequential: one at a time
common in state estimation



Batched: full window
common in parameter estimation

Before we can do anything, we need to choose the parameters of interest and their prior ranges

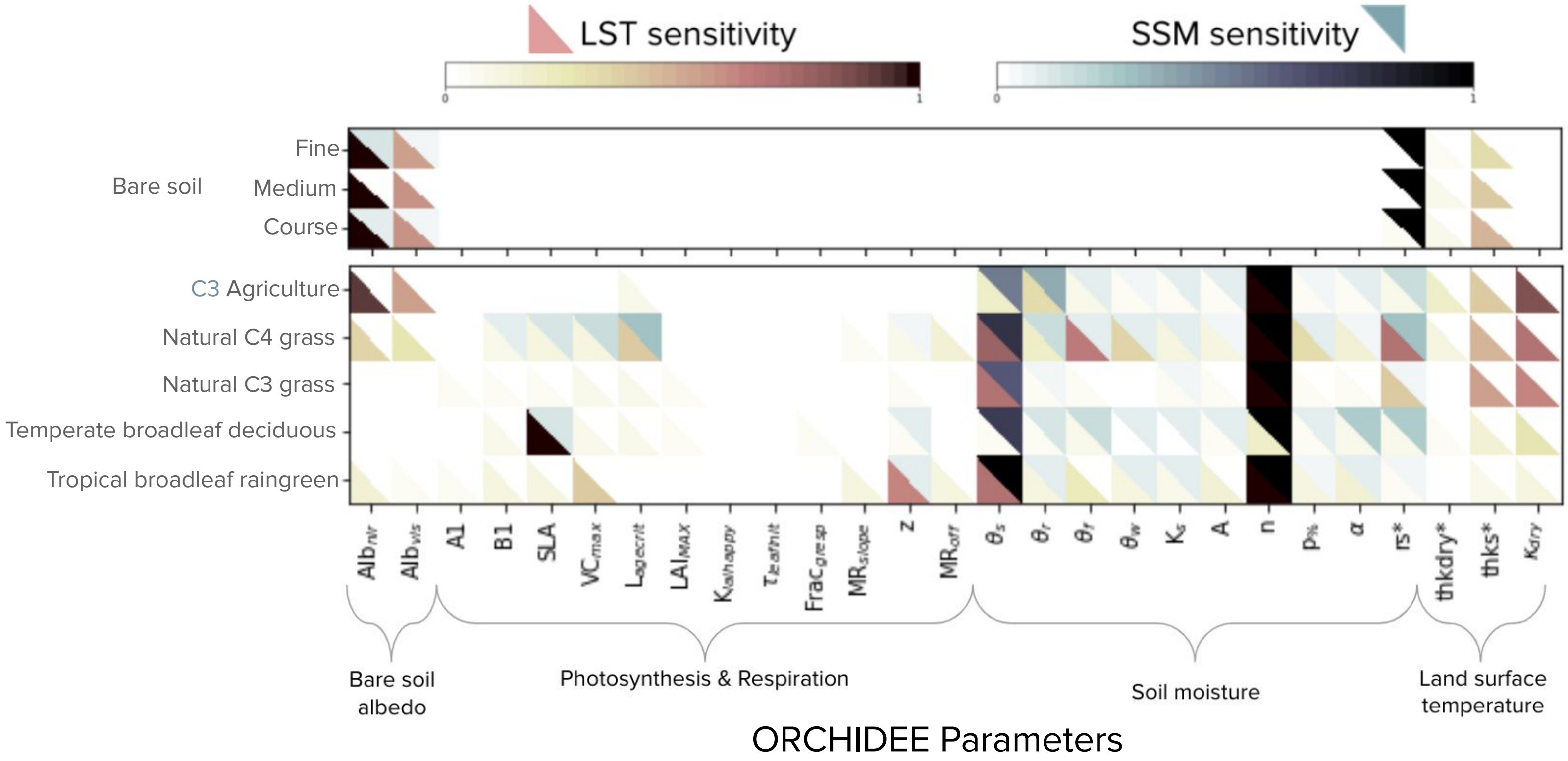


	Parameter	Value	Min	Max	Unit	Namelist	Precision	Definition
1	Parameter	Value	Min	Max	Unit	Namelist	Precision	Definition
2	RLWCSWEA	200	150	250	kg.m-3	NAMPARSNOW	JPRB	Constant used to calculate **snow liquid water holding capacity**
3	RLWCSWEB	0.03	0.01	0.05	-	NAMPARSNOW	JPRB	Constant used to calculate **snow liquid water holding capacity**
4	RLWCSWEC	0.1	0.08	0.15	-	NAMPARSNOW	JPRB	Constant used to calculate **snow liquid water holding capacity**
5	RHOMINSNA	109	90	115	kg.m-3	NAMPARSNOW	JPRB	Constant used to calculate **fresh snow density**
6	RHOMINSNB	6	5	7	kg.m-3.K-1	NAMPARSNOW	JPRB	Constant used to calculate **fresh snow density**
7	RHOMINSNC	26	20	32	kg.m-3.K-1	NAMPARSNOW	JPRB	Constant used to calculate **fresh snow density**
8	RSNDTOVERA	37000000	1.50E-07	4.00E-07	Pa.s	NAMPARSNOW	JPRB	Constant used to calculate **snow viscosity** following Anderson (1976)
9	RSNDTOVERB	0.081	0.08	0.35	K-1	NAMPARSNOW	JPRB	Constant used to calculate **snow viscosity** following Anderson (1976)
10	RSNDTOVERC	0.018	0.009	0.02	m3.kg-1	NAMPARSNOW	JPRB	Constant used to calculate **snow viscosity** following Anderson (1976)
11	RSNDTDESTA	2.80E-06	1.50E-06	3.50E-06	s-1	NAMPARSNOW	JPRB	Snow-related constant used to calculate **thermal metamorphism**
12	RSNDTDESTB	0.042	0.01	0.1	K-1	NAMPARSNOW	JPRB	Snow-related constant used to calculate **thermal metamorphism**
13	RSNDTDESTC	460	320	600	m3.kg-1	NAMPARSNOW	JPRB	Snow-related constant used to calculate **thermal metamorphism**
14		(actual value: 0.0	-0.032	-0.112				
15	RSNDTDESTROI	150	100	200	kg.m-3	NAMPARSNOW	JPRB	Snow-related constant used to calculate **thermal metamorphism**
16	RSNDAMOB	1.25	1	1.5	-	NAMPARSNOW	JPRB	Constant used to calculate the **mobility index** for the **wind-driven compaction in**
17	RSNDMOB	295	280	310	kg.m-3	NAMPARSNOW	JPRB	Constant used to calculate the **mobility index** for the **wind-driven compaction in**
18	RSNDAW	2.868	2.6	3.1	-	NAMPARSNOW	JPRB	Constant used to calculate the **wind-driven compaction index**

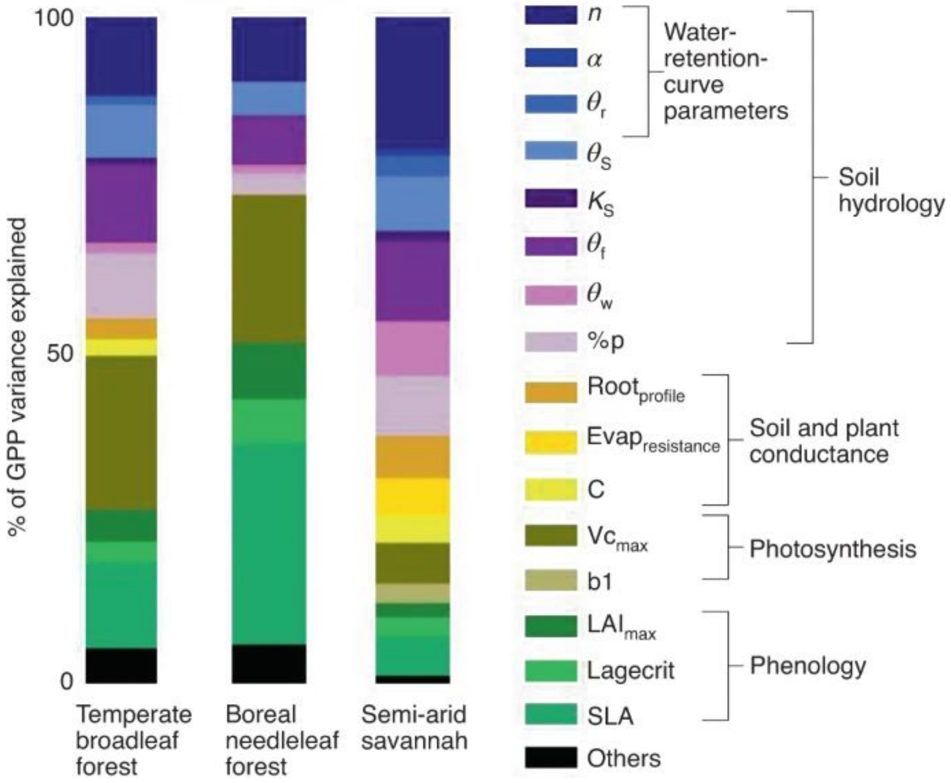
- 100s of parameters
- Global vs Spatialised
  (e.g., PFTs, soil type,...)

Kennedy et al. (2024)

Morris’ method allows us to identify to the most sensitive parameters and discount the least sensitive



Sobol’s method allows us to capture the **interactions** between the parameters

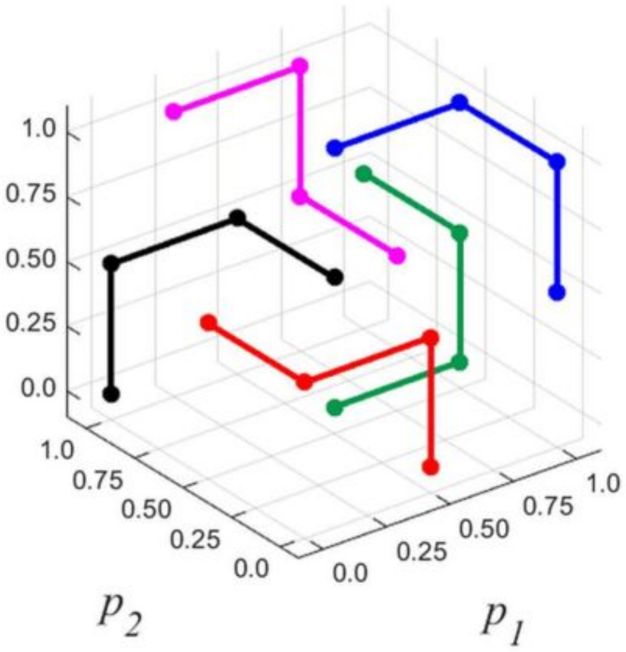


Novick et al. (2022)

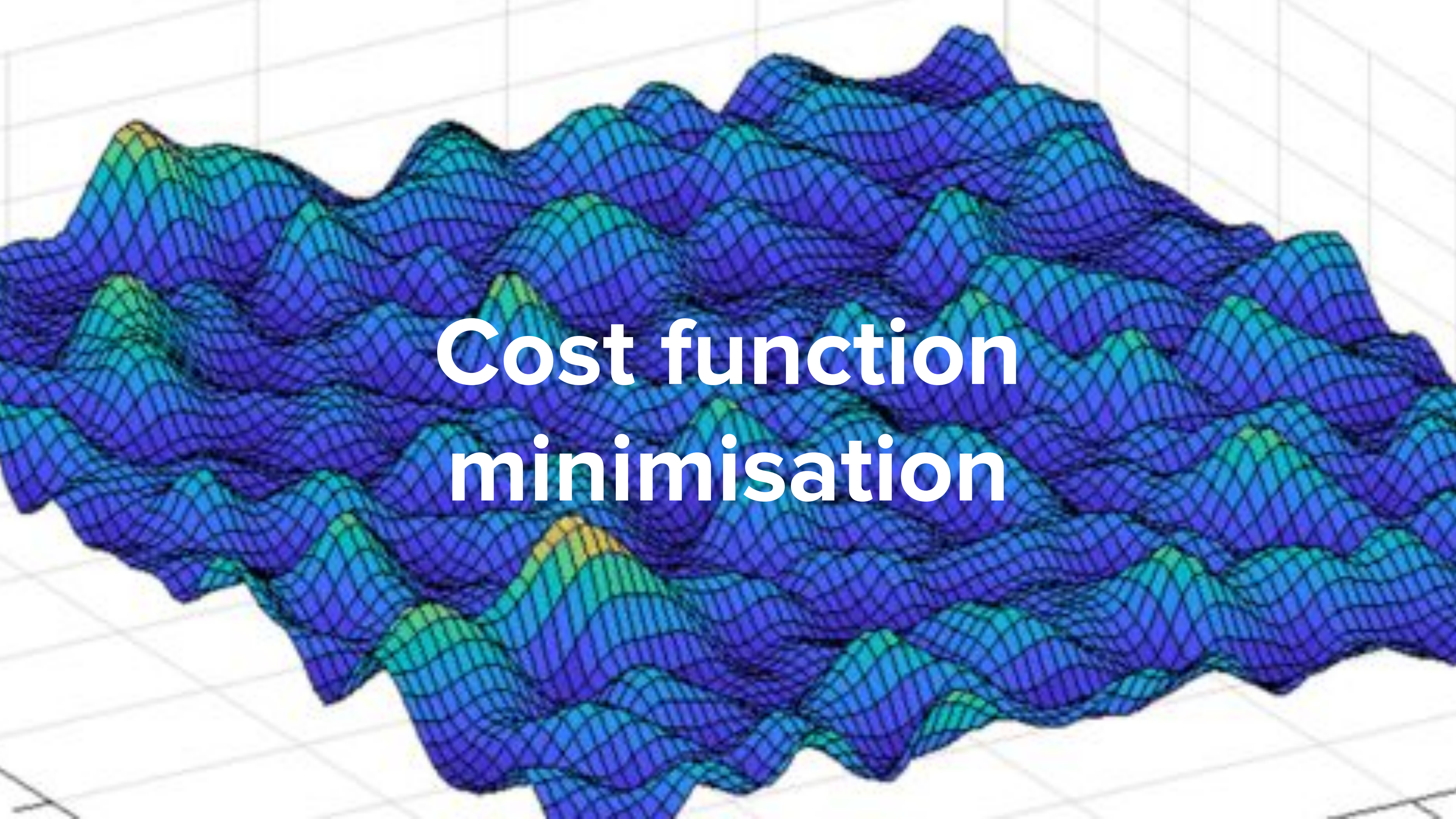
While these methods are powerful, they can require many costly model runs

Morris
O(100)

Sobol
O(10,000)

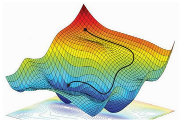


Cost function minimisation



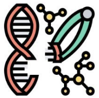
Different algorithm can be used to minimise the cost function, each with pros and cons

$$J(x) = \tfrac{1}{2}\underbrace{(\, y - M(x)\,)}_{\text{Difference between the model given the parameters and observations}}{}^{T}\mathbf{R}^{-1}(\, y - M(x)\,) + \tfrac{1}{2}(\, x - x_b)^{T}\mathbf{B}^{-1}\underbrace{(\, x - x_b)}_{\text{Difference between the parameters and their prior value}}$$



Gradient descent

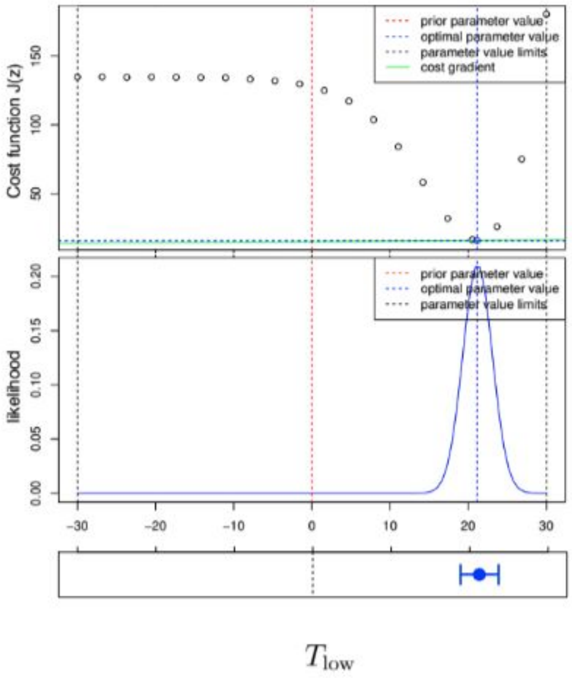
Pros: Deterministic and fast - often converging after few iterations.
Cons: Struggles with local minima, best results with tangent linear/adjoint.

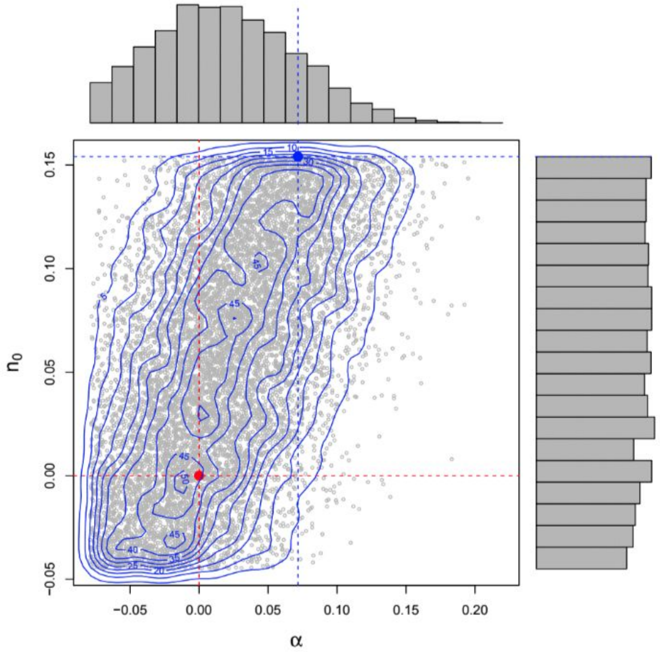


Ensemble methods

Pros: Explores global search space effectively.
Cons: Slower convergence compared to gradient-based methods.

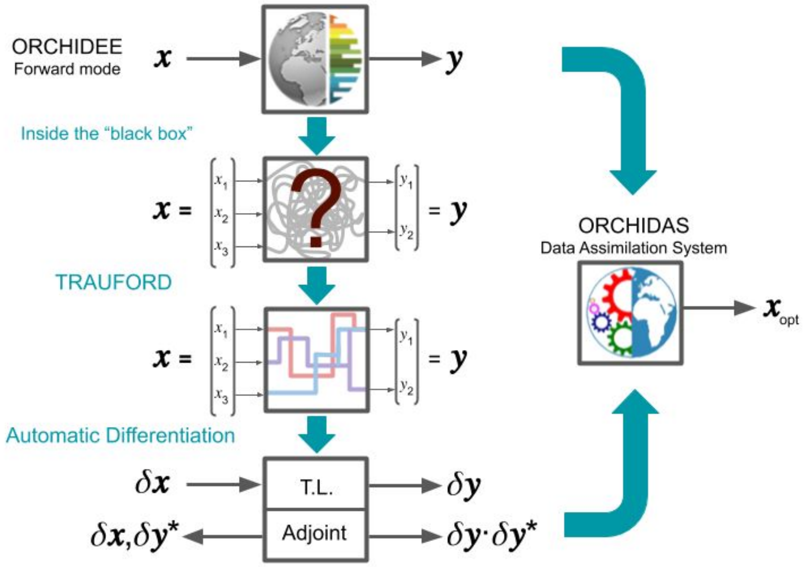
Using information about the curvature of parameter space at the optimum to derive posterior uncertainty





Raoult (2017)

For land surface models, adjoints are hard to maintain, so alternative methods are often used for gradient-based descent.



Finite differences

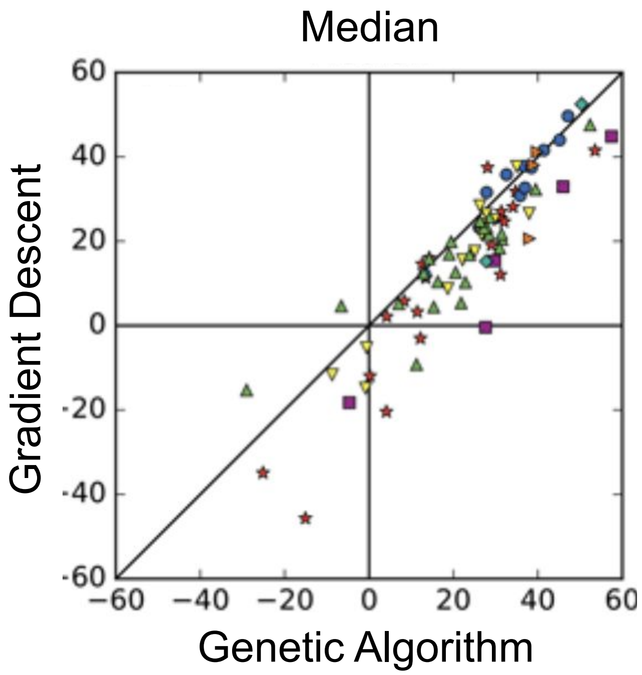
4DEnVar

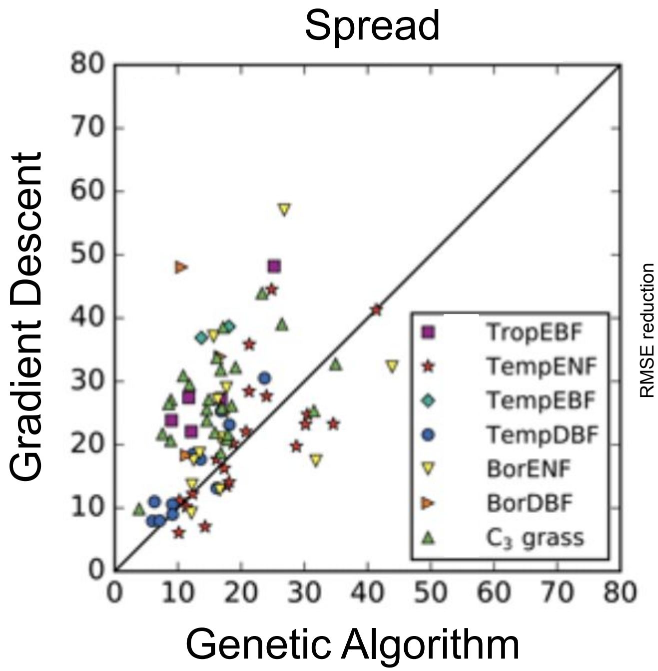
Emulate LSMs with NNs for fast, differentiable surrogates

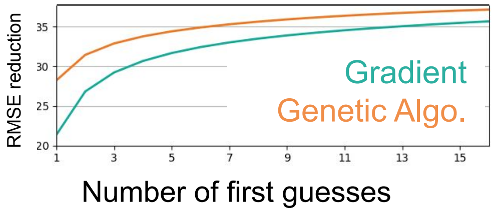
Use LLMs to translate legacy LSMs to modern, GPU-friendly languages

Beylat et al. (2024)
Douglas et al. (2025)

Ensemble approaches do not need gradient information



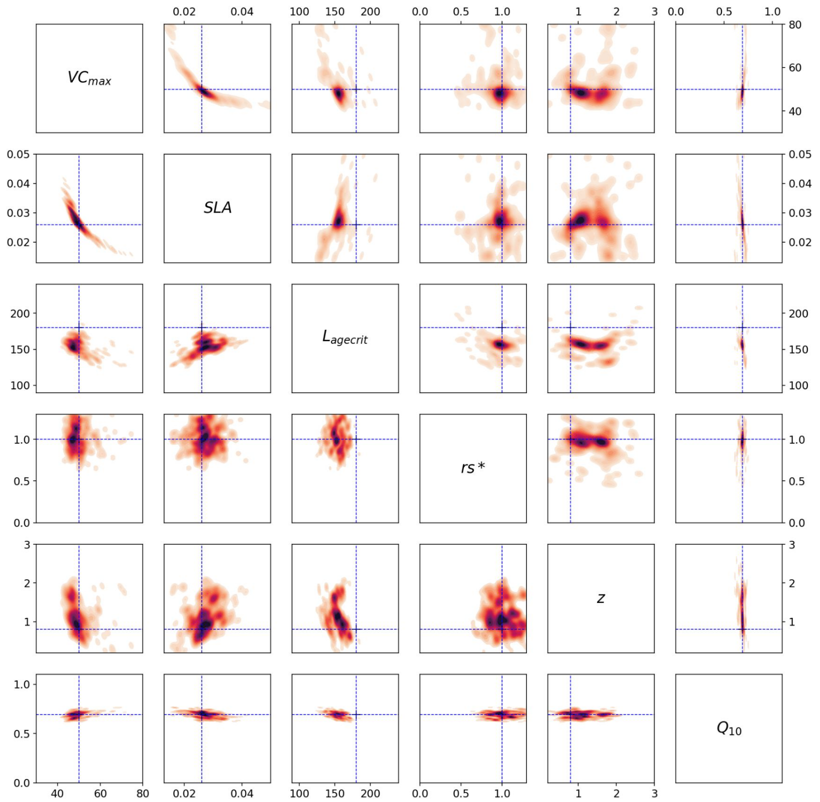




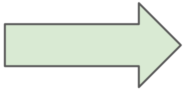
Median and spread in NEE / LE RMSE reductions for 16 first guesses

Bastrikov et al. (2018)

Equifinality is also a problem for both types of methods



Performing 200 **different optimisations** each with **different prior** values



need to validate against independent data

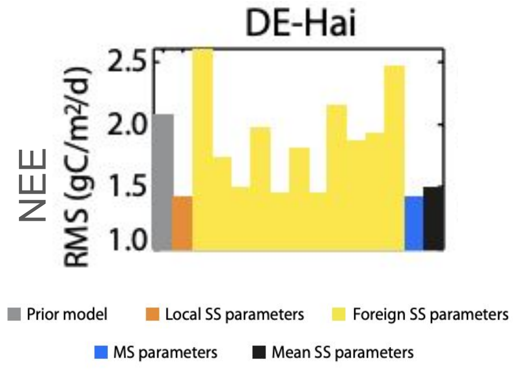
Raoult et al. (2024)


Parameter
estimation in action

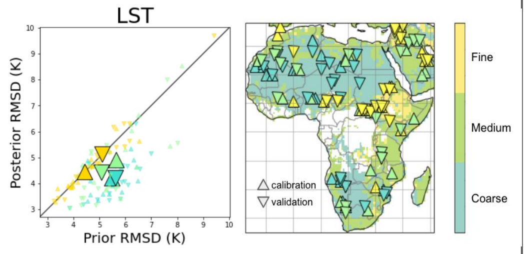
Examples of how parameter estimation is used practically in land surface modelling

- Deriving an operational set of parameters
- Identifying structural uncertainty in models
- Propagating uncertainty through the system

Since the land surface is heterogeneous, finding an operational set of parameters can be challenge



Single vs Multi site optimisation

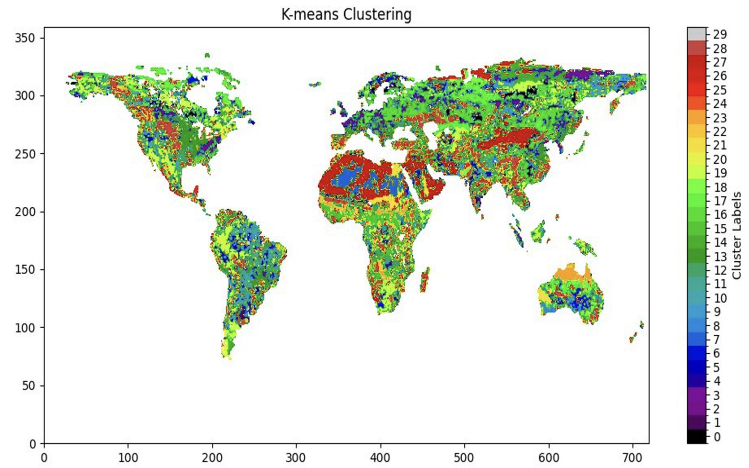


Representative pixels

Spatialised parameters often grouped by vegetation (PFT) or soil types, but are these representative enough?

Kuppel et al. (2012)

We can use clustering approaches to reduce dimensionality while capturing a greater spatial variability



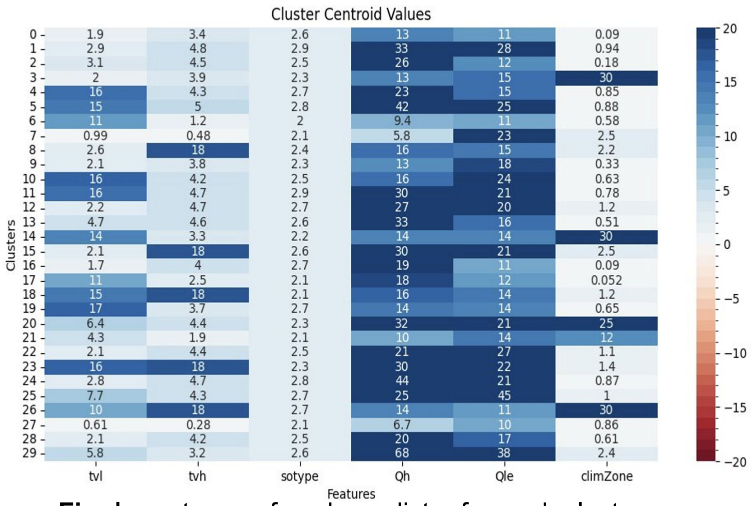


Fig: Importance of each predictor for each cluster

Predictors:

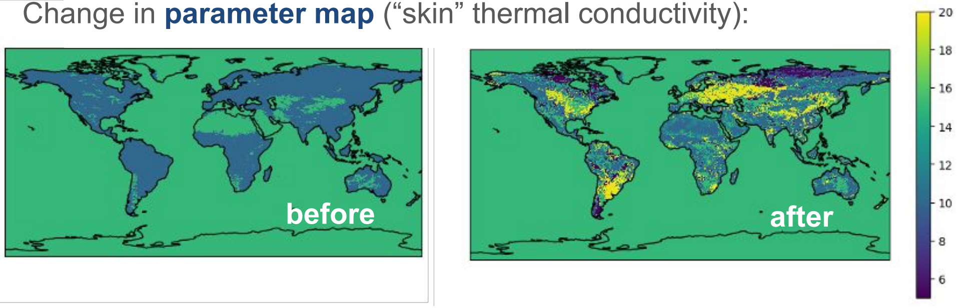
- Vegetation type (each for low and high vegetation),
- Soil type,
- RMSE (for **latent** and **sensible heat** fluxes)
- Climate (from Köppen–Geiger climate classification)

> optimisation run for a sub-sample of sites for each cluster

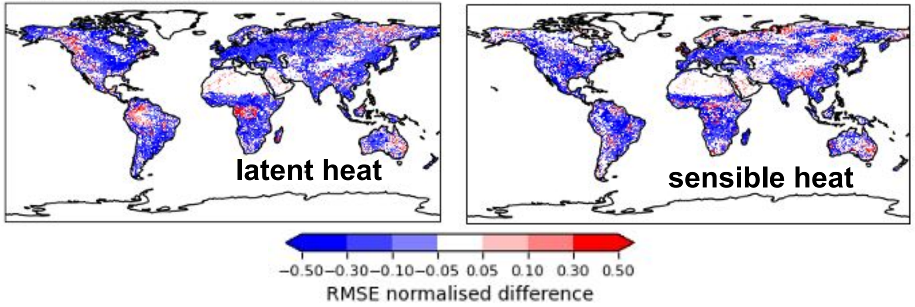
Arduini: ECMWF technical memo

Example: optimising against sensible and latent heat fluxes using clusters in an operational setting

Change in **parameter map** (“skin” thermal conductivity):

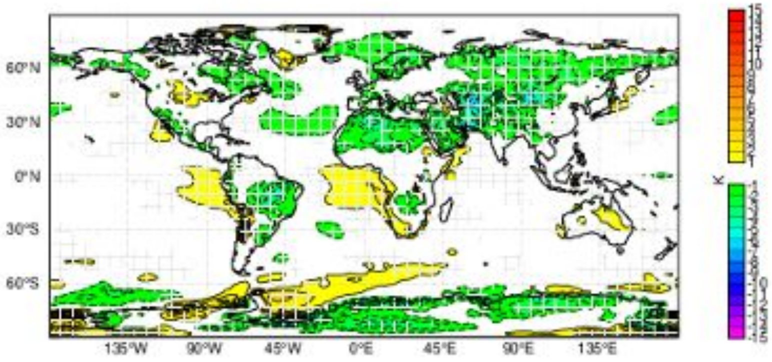


Change in **“offline” performance**



Impact on coupled forecast (compared to ERA5)

2m dew point



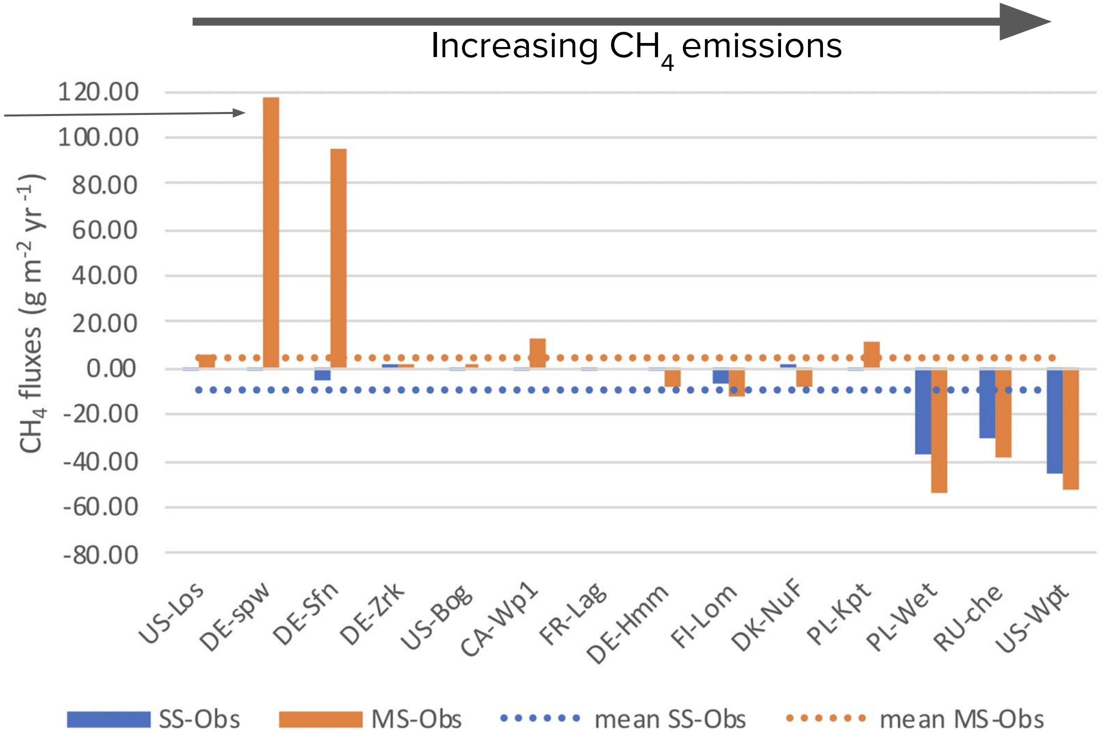
CTL: Bias = - 0.21 ; RMSE = 1.38

OPT: Bias = - 0.12 ; RMSE = 1.18

Arduini: ECMWF technical memo

Identifying structural errors: CH_4 emissions in northern peatlands

Parameter controlling plant transport efficiency most important

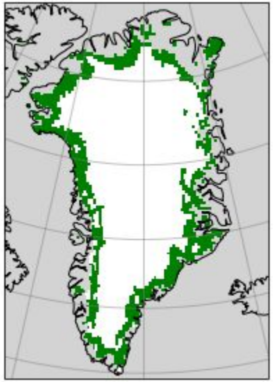


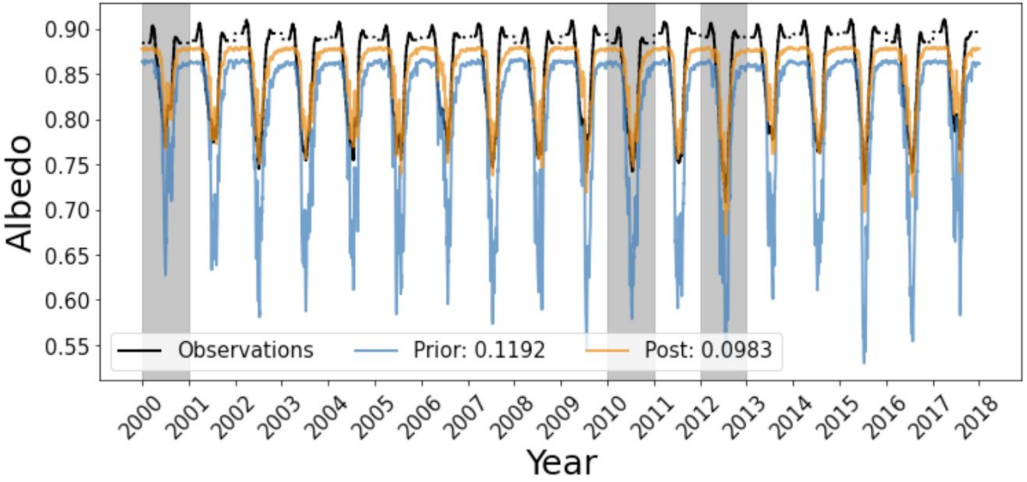
Not enough C available for methanogenesis

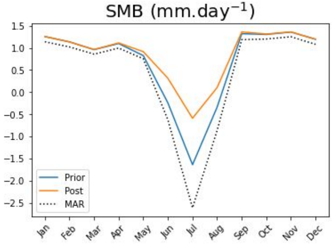
Differences in annual methane emissions between observed data (Obs), and simulations

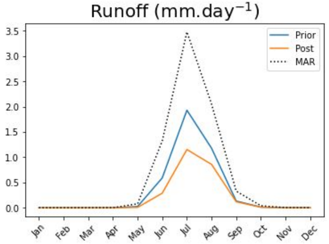
Salmon et al. (2022)

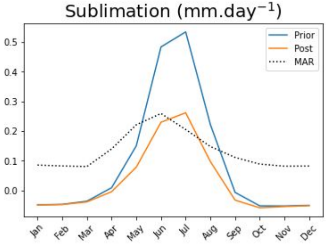
Identifying structural errors: Snow albedo over Greenland









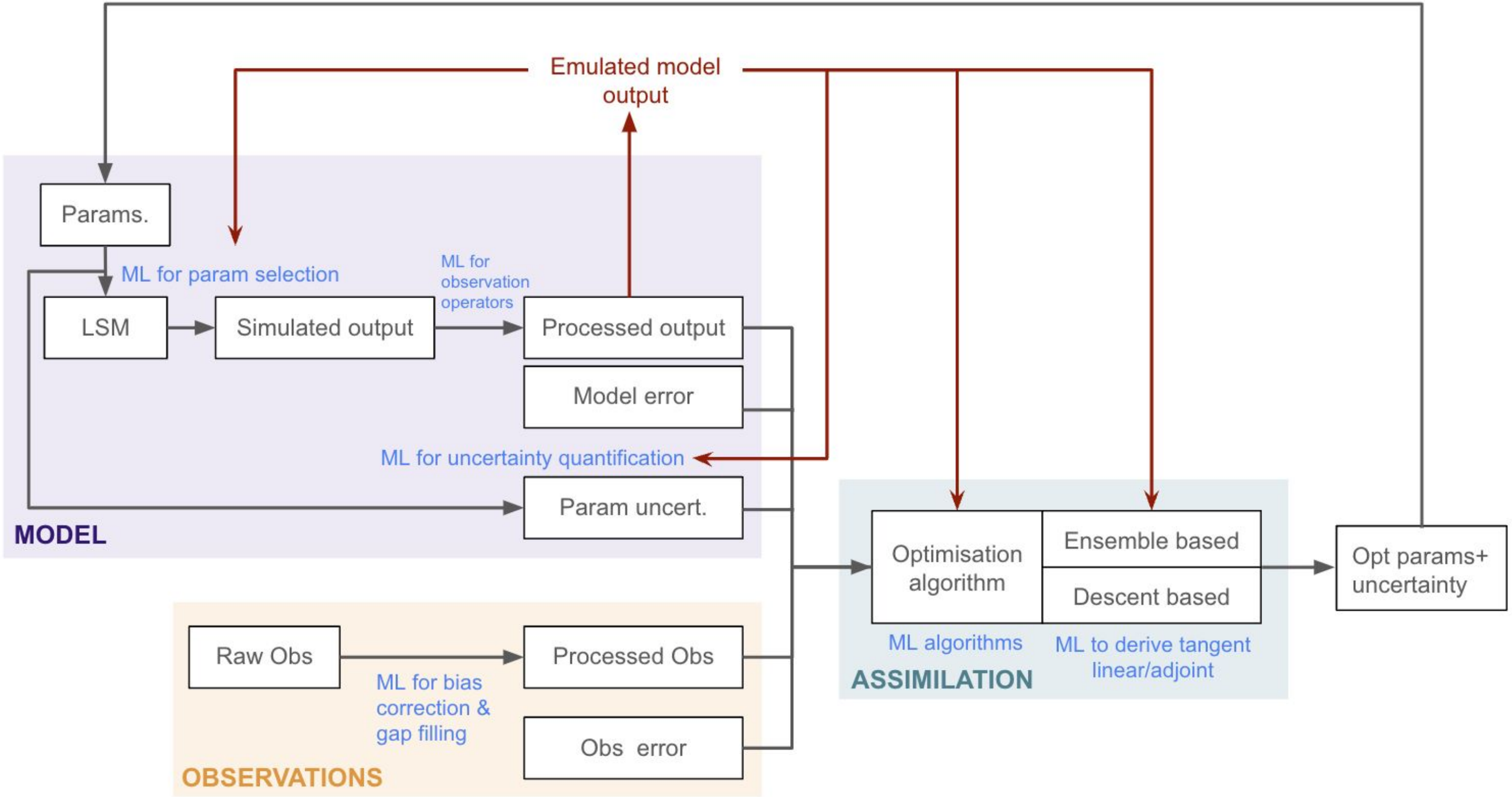


Albedo is improved, but the model-data of other processes are degraded

Raoult et al. (2023)

Machine learning

Machine learning can facilitate every aspect of the DA workflow

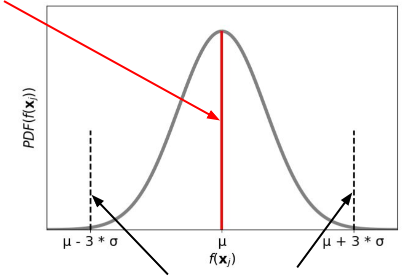


Gaussian Process emulators powerful tools for parameter estimation

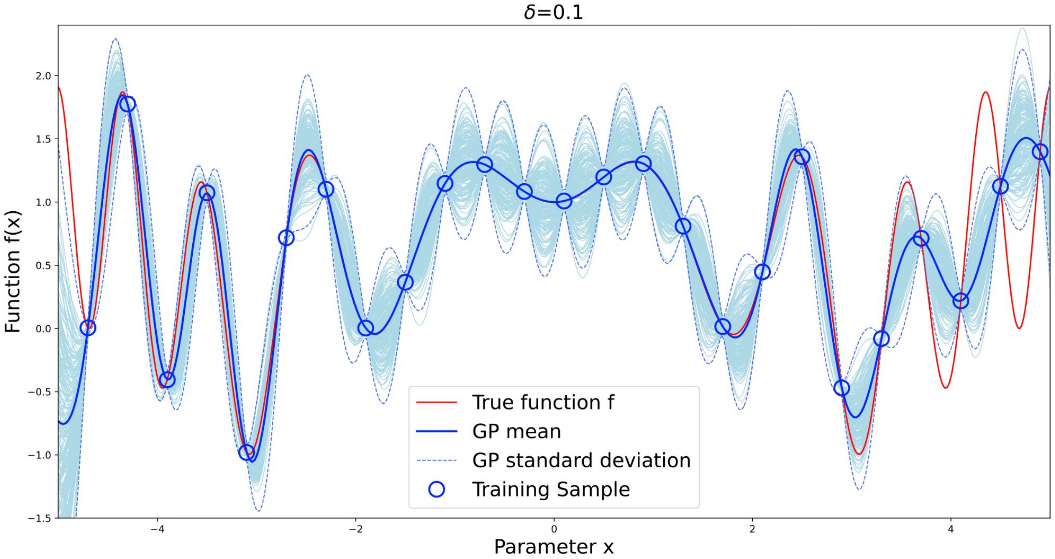
A Gaussian process defines a distribution over functions:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$$

The mean μ is the prediction of f(X)



The standard deviation σ is the uncertainty of our prediction.



$\sin(x^2)+\cos(x^2)$

mean function: m(x)=0

kernel function: squared exponential with length of decorrelation 0.1

Beylat et al. (prep)

Two are ways Gaussian processes are used in parameter estimation

$$I(\mathbf{x}) = \frac{|y(\mathbf{x}) - z|}{\sqrt{\sigma^2(\mathbf{x}) + \sigma^2_{\text{obs}} + \sigma^2_{\text{mod}}}}$$

History Matching
Rule out unlikely parameters using a implausibility function

$$\begin{aligned}\mathrm{EI}(\mathbf{x}) = (f_{\text{best}} - y(\mathbf{x}) - \xi)\,\Phi\Big(\frac{f_{\text{best}} - y(\mathbf{x}) - \xi}{\sigma(\mathbf{x})}\Big) \\ + \sigma(\mathbf{x})\,\phi\Big(\frac{f_{\text{best}} - y(\mathbf{x}) - \xi}{\sigma(\mathbf{x})}\Big)\end{aligned}$$

Bayesian Optimisation
Search for optimal parameters using a GP and acquisition function

Two are ways Gaussian processes are used in parameter estimation

Predicted metric

Target metric

$$I(\mathbf{x}) = \frac{|y(\mathbf{x}) - z|}{\sqrt{\sigma^2(\mathbf{x}) + \sigma^2_{\text{obs}} + \sigma^2_{\text{mod}}}}$$

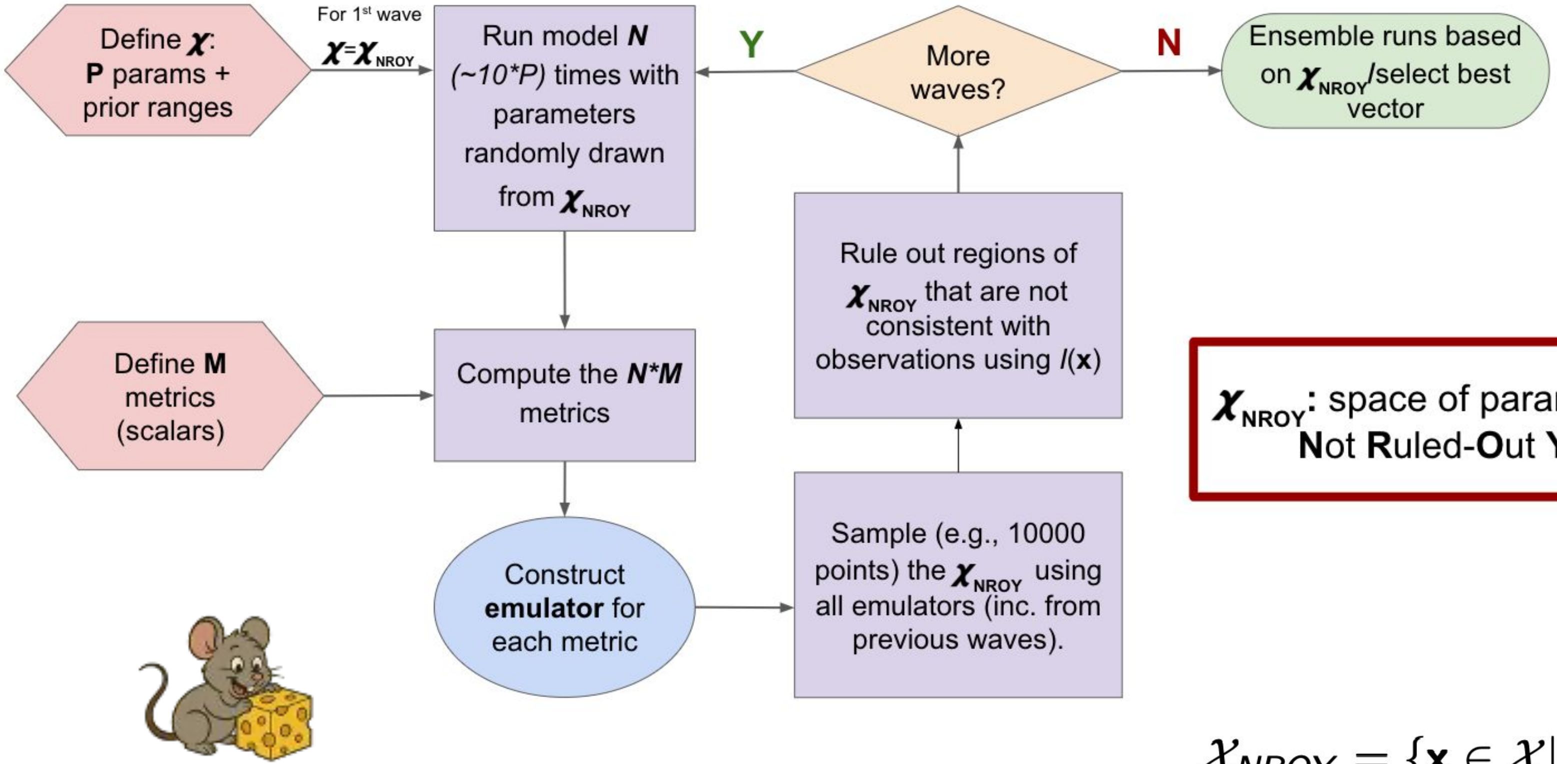
Uncertainty of the emulator

Error obs/model

History Matching
Rule out unlikely parameters using a implausibility function

Bayesian Optimisation
Search for optimal parameters using a GP and acquisition function

History Matching framework



$\boldsymbol{\chi}_{\mathrm{NROY}}$: space of parameters **N**ot **R**uled-**O**ut **Y**et

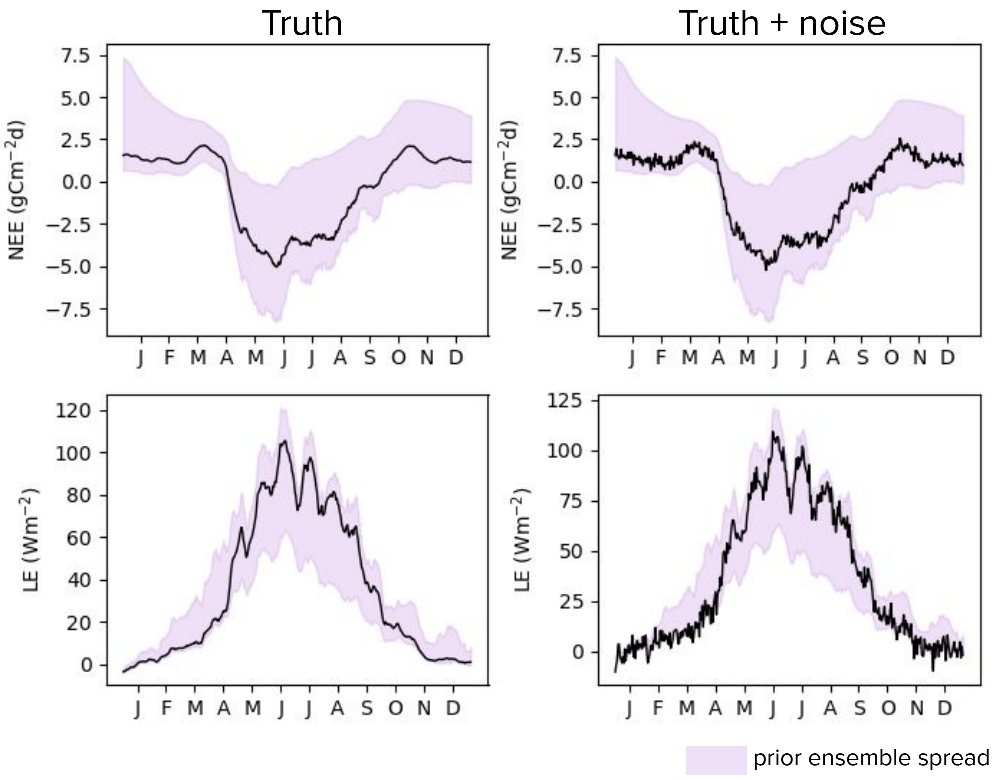
$$\mathcal{X}_{NROY} = \{\mathbf{x} \in \mathcal{X} | \mathcal{I}(\mathbf{x}) < 3^2\}$$

To investigate the potential of History Matching, we first set up twin experiment

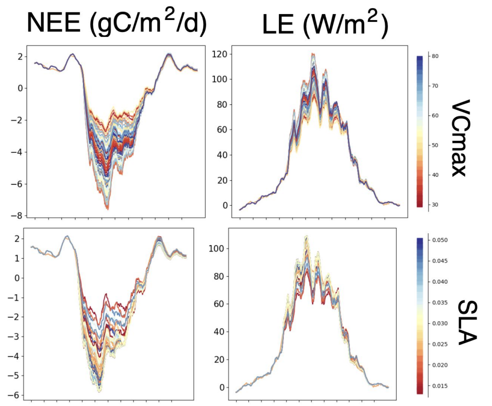
Site: FR-Fon (2005)

Goal: Recover default values

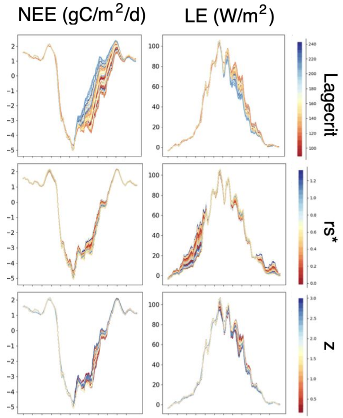
Model/observation errors set ~N(0,σ) where
σ = 0.1 x $\overline{\text{timeseries}}$



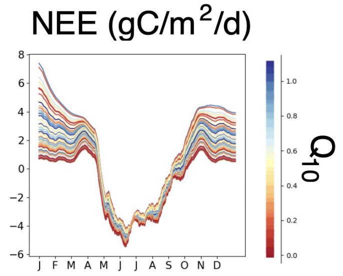
One of the advantages to History Matching is its ability to tune against multiple metrics



Amplitude



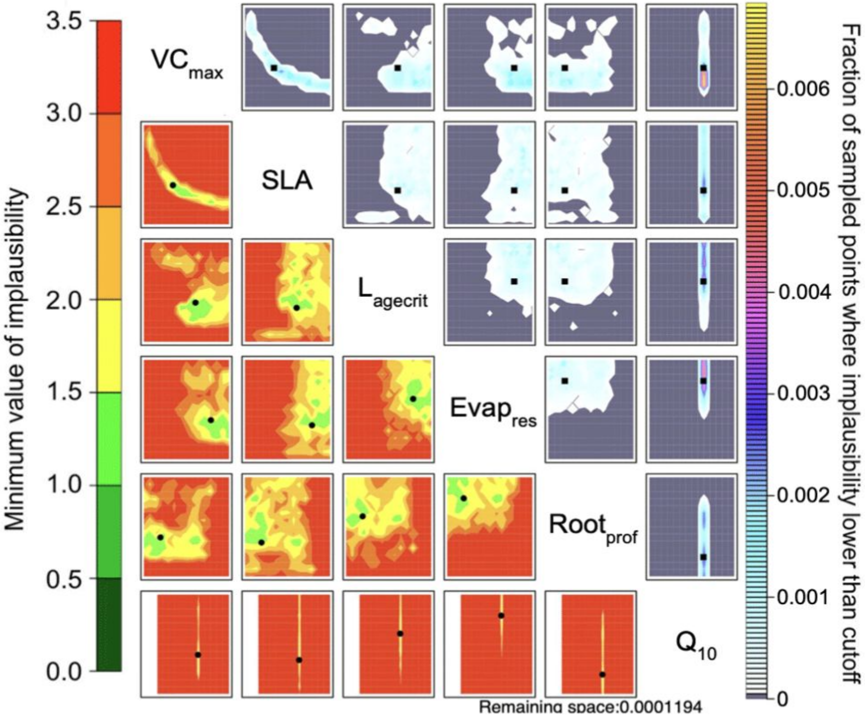
Spring & Autumn slopes

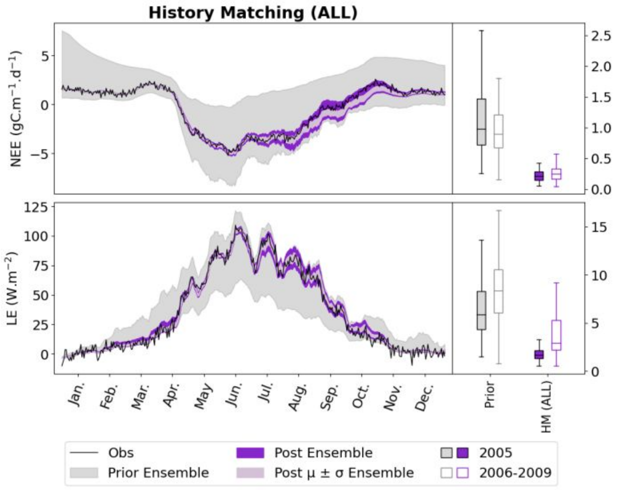


Initial carbon stocks

One at time parameter perturbation test helps give a sense of properties impacted by each parameter

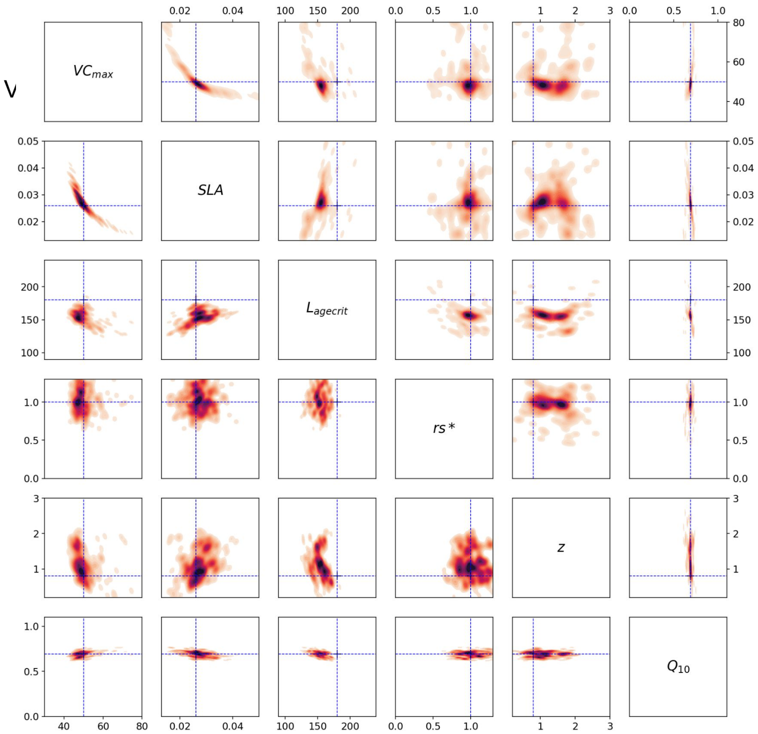
After 10 waves we have significantly reduced parameter space whilst retaining “true” values in the NROY

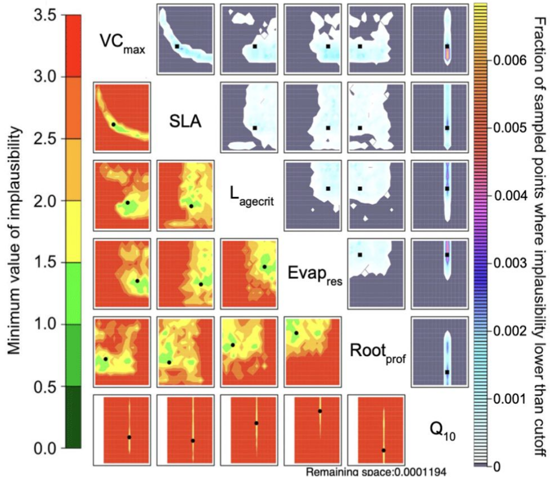




Comparing to results found using standard Bayesian Calibration (gradient based approach)

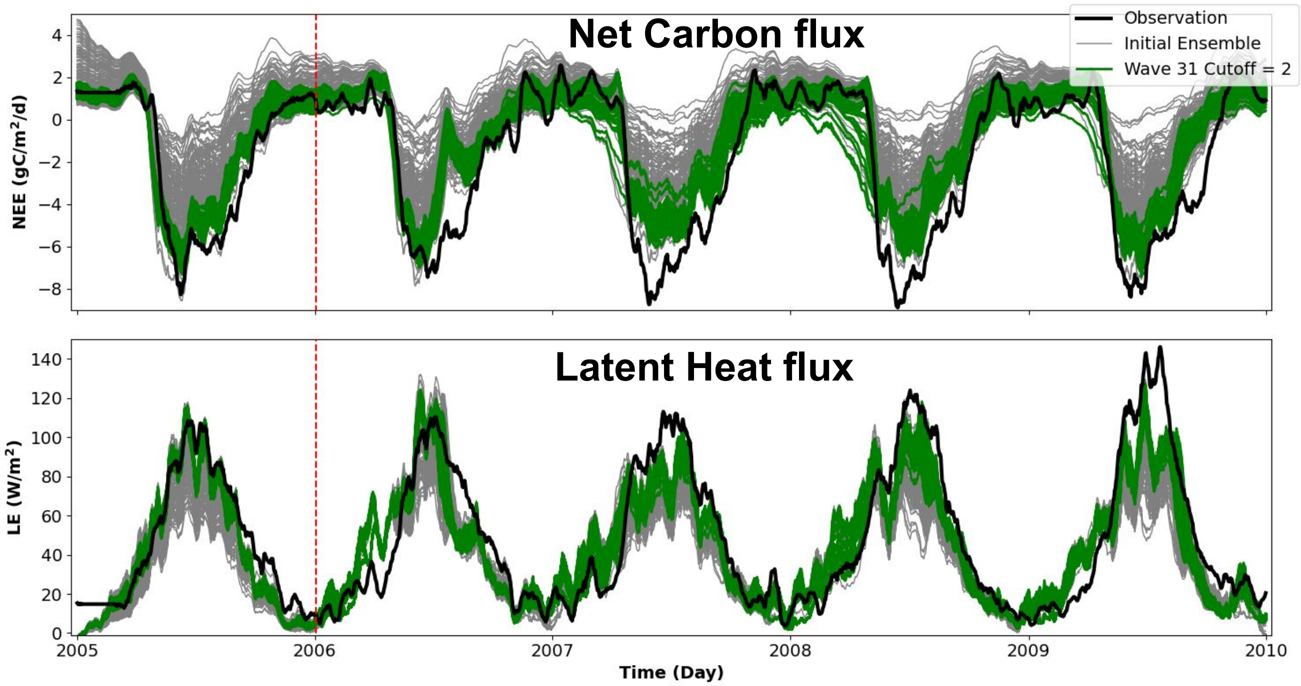
Performing 200 **different optimisations** each with **different prior** values (i.e., exploiting equifinality)

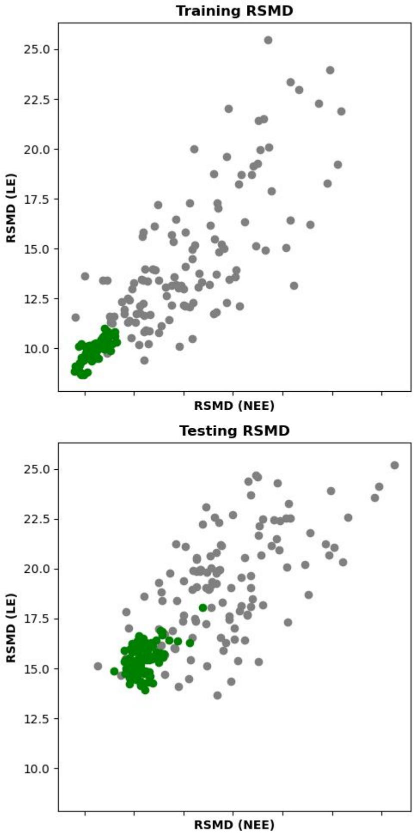




Raoult et al. (2024)

History matching is a great tool for uncertainty quantification





Beylat et al. (prep)

Two are ways Gaussian processes are used in parameter estimation

$$I(\mathbf{x}) = \frac{|y(\mathbf{x}) - z|}{\sqrt{\sigma^2(\mathbf{x}) + \sigma^2_{\text{obs}} + \sigma^2_{\text{mod}}}}$$

History Matching
Rule out unlikely parameters using a implausibility function

Emulator prediction

Best current observation

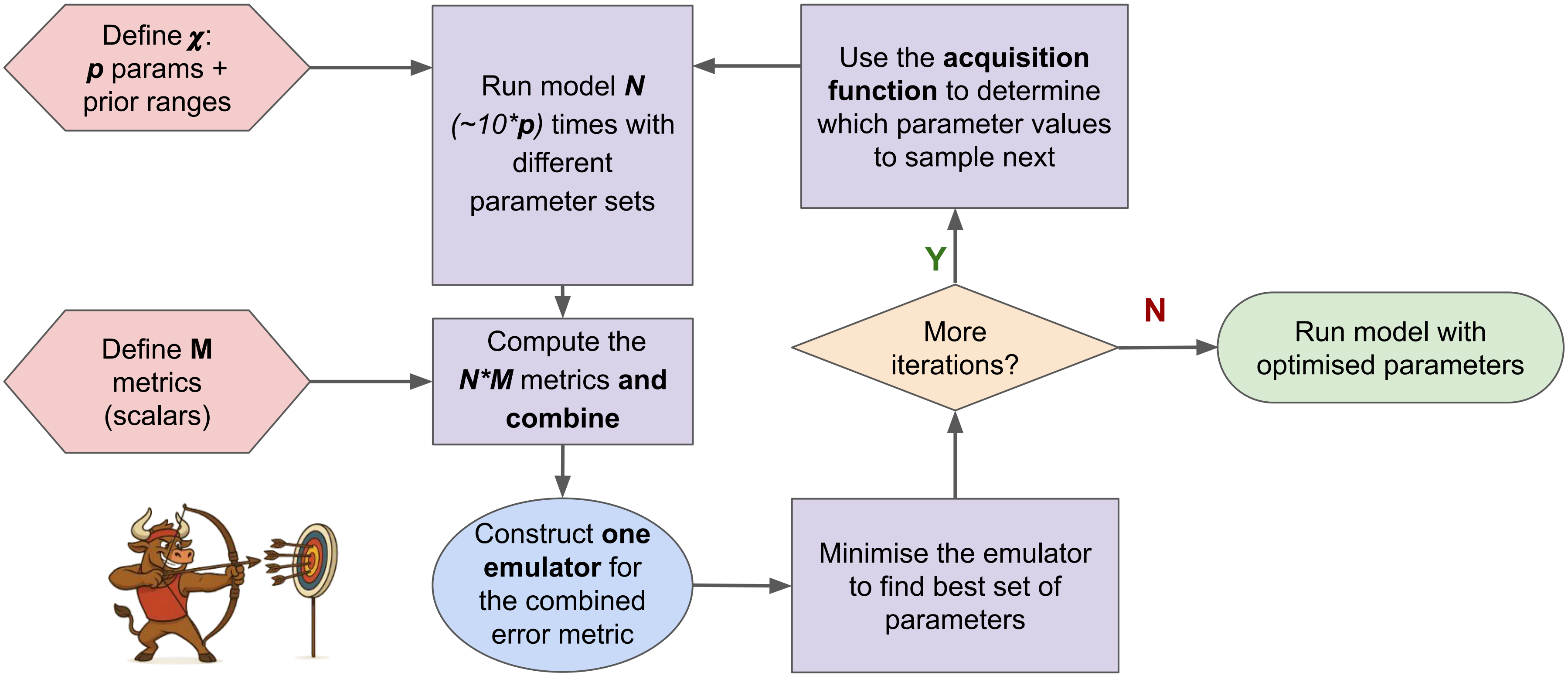
"jitter"

$$\mathrm{EI}(\mathbf{x}) = \left(f_{\text{best}} - y(\mathbf{x}) - \xi\right) \Phi\!\left(\frac{f_{\text{best}} - y(\mathbf{x}) - \xi}{\sigma(\mathbf{x})}\right) + \sigma(\mathbf{x})\, \phi\!\left(\frac{f_{\text{best}} - y(\mathbf{x}) - \xi}{\sigma(\mathbf{x})}\right)$$

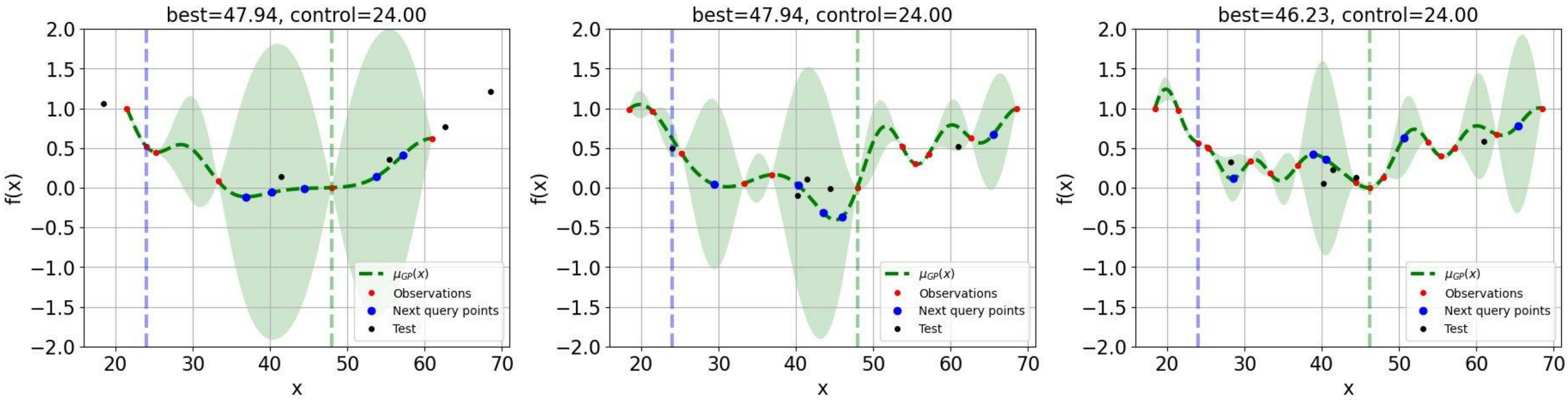
Emulator uncertainty

Bayesian Optimisation
Search for optimal parameters using a GP and acquisition function

Bayesian optimisation framework



Gaussian processes emulators can also be trained iteratively using an acquisition function



Iteration 0 Iteration 1 Iteration 2

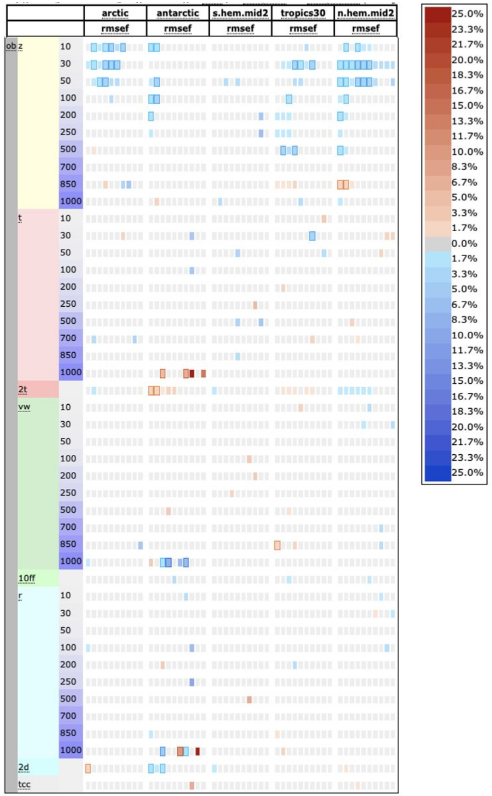
Expected improvement:

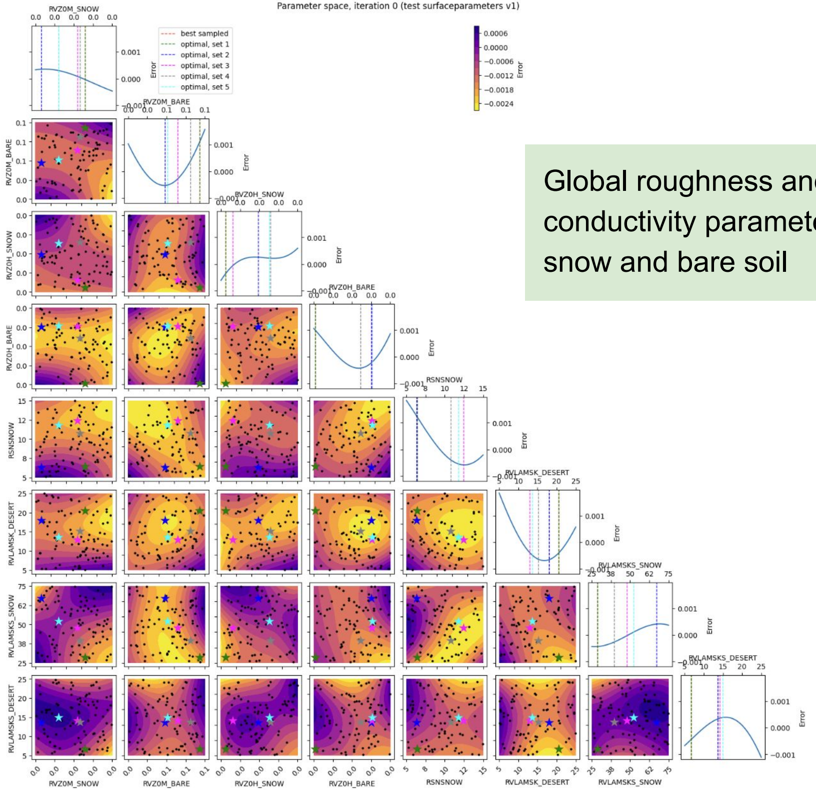
$$\mathrm{EI}(\mathbf{x}) = \left(f_{\text{best}} - y(\mathbf{x}) - \xi\right)\Phi\left(\frac{f_{\text{best}} - y(\mathbf{x}) - \xi}{\sigma(\mathbf{x})}\right) + \sigma(\mathbf{x})\,\phi\left(\frac{f_{\text{best}} - y(\mathbf{x}) - \xi}{\sigma(\mathbf{x})}\right).$$

exploration ⟵ ξ ⟶ exploitation

Suetzl & Van Niekerk: ECMWF technical memo

This approach is used at ECMWF, using Quaver scorecards as metrics



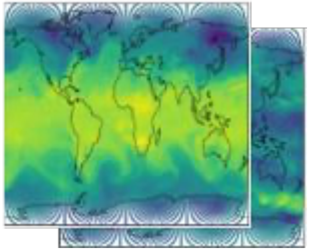


Global roughness and skin conductivity parameters for snow and bare soil

Emulating the land surface at ECMWF

The AIFS: ECMWF's data-driven weather forecasting systems

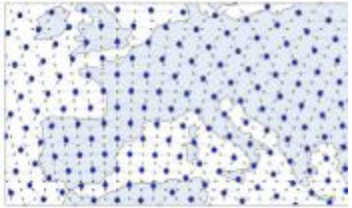
OPERATIONAL



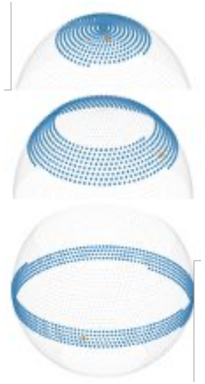
<u>Input</u>
Model state:
X(t), X(t-**6h**)

AIFS MODEL

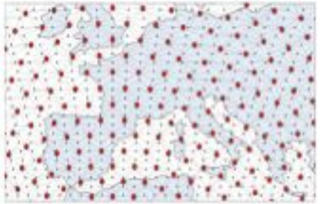
encoder → processor → decoder



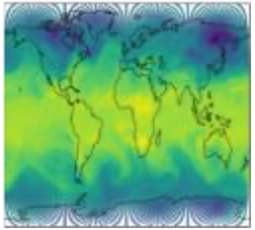
Graph attention



Transformer blocks and windowed attention (attention across regional bands). On a coarser grid than input grid



Graph attention



<u>Output</u>
Prediction:
X(t+**6h**)

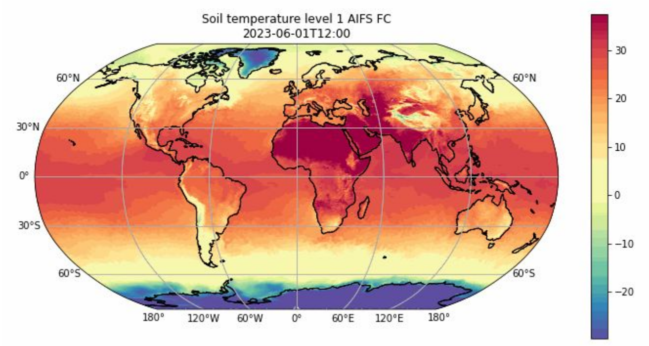
Training Scheme:

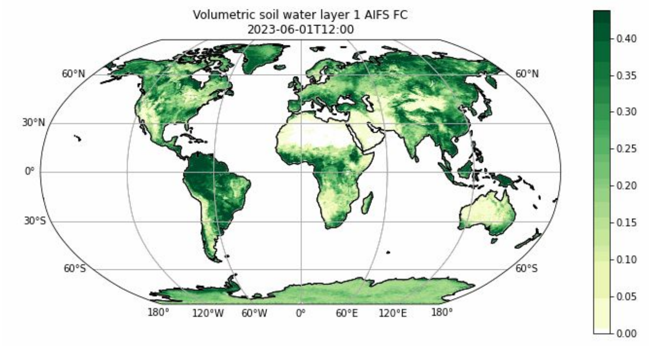
1. Train on ERA5
2. Fine-tune on operational analysis and lead times up to 72h

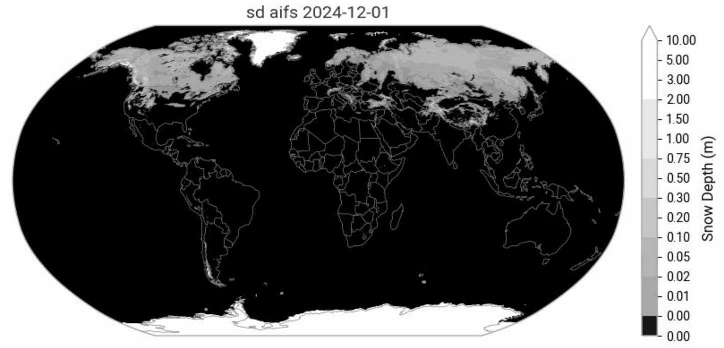
anemoi

Lang et al. (2024)

Land variables are included in the AIFS!

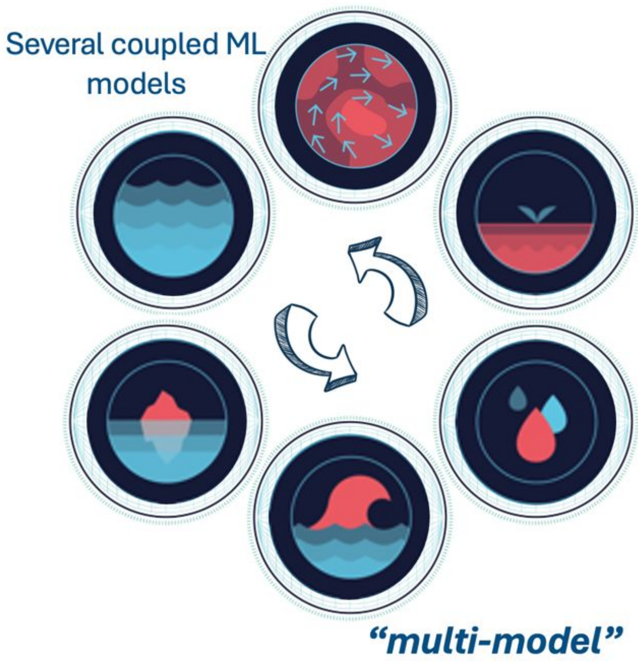


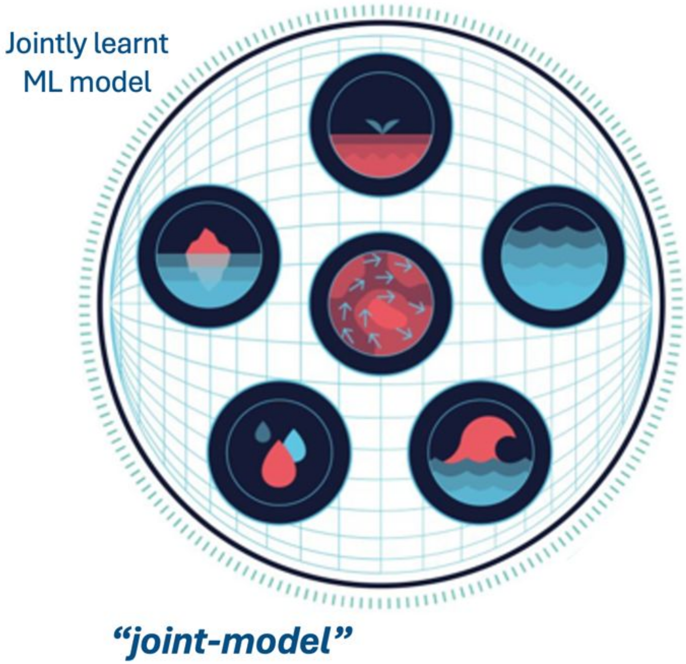




- Operational AIFSv1 has soil moisture and soil temperature (upper layers)
- Next version will have snow cover!
- Early experiments show AIFSsnow outperforming the physical model

As part of DestinE, we are building data-driven Earth System model

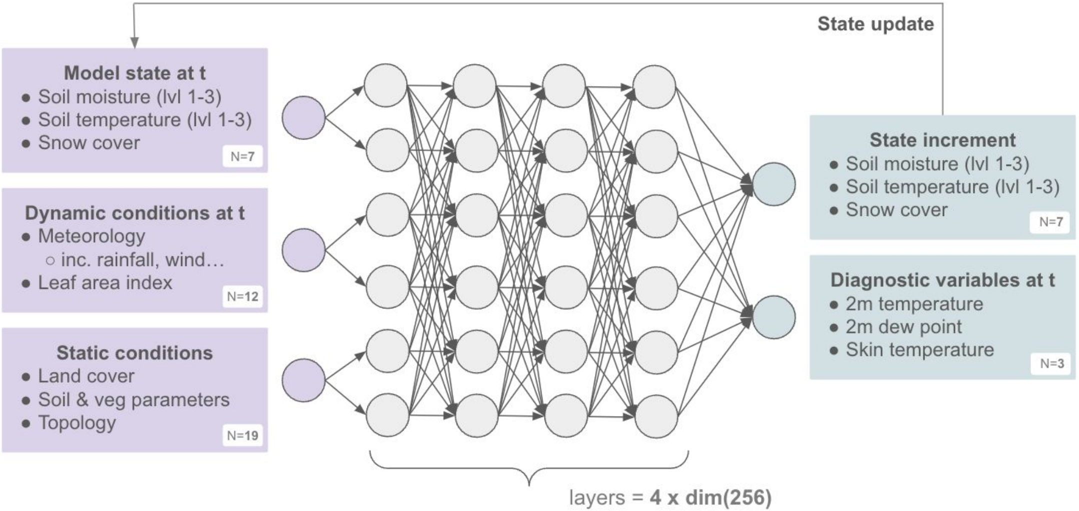


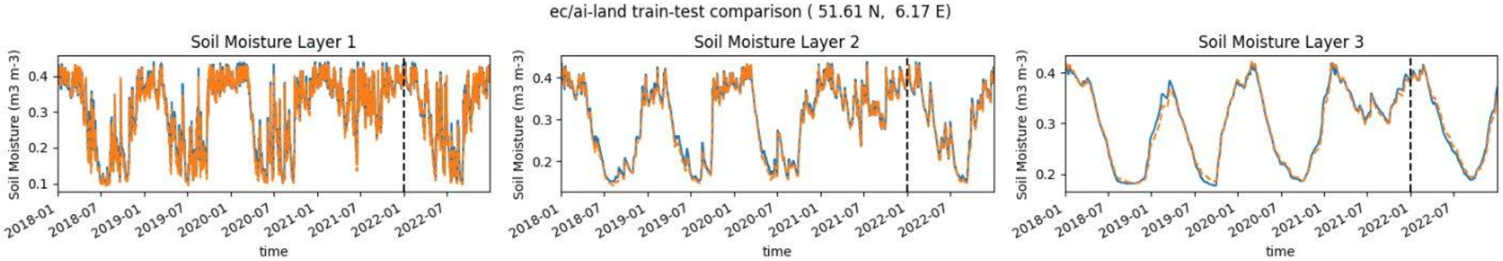


Stand-alone emulator: aiLand

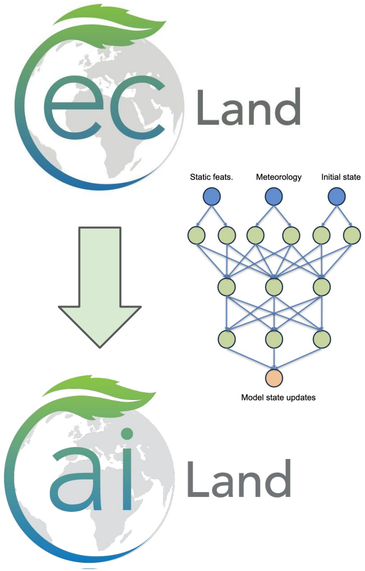
Exploit column model structure to train MLP

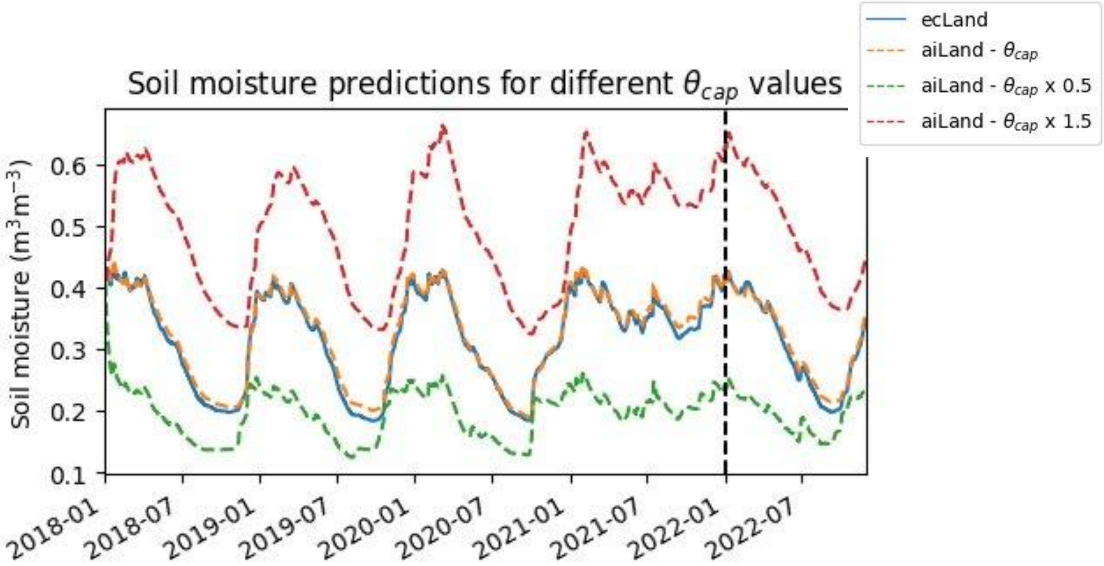
- Model resolution-agnostic
- Able to spatial variability
- Easy integration with observations





As we move to more sophisticated emulators of the land surface, can we use these for parameter estimation?





Next steps

	By fine tuning against **observations**, can we **learn biases** in the model?
	With more information on parameter sensitivities, can we use the emulator for **parameter estimation**?
	Can we exploit the **differentiability** of aiLand for land model **data assimilation**?
	How do we **couple** aiLand with the other **Earth System components** (physical or machine learnt)?

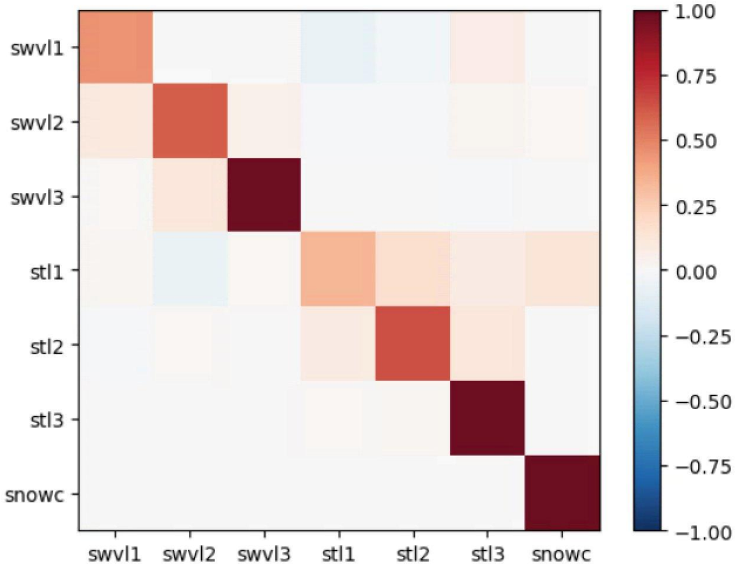


Fig: aiLand Jacobian

Take home messages:

- **Parameter uncertainty** is one of **large sources** of uncertainty in land surface models
- **Data assimilation** has been **shown to be a powerful tool** for reducing this uncertainty.
- **Machine learning** can **facilitate** parameter estimation by enhancing **computational efficiency** and replacing poorly represented processes.
- We can use **deep learning** to emulate land surface models

Nina Raoult, Natalie Douglas, Natasha MacBean, et al. Parameter Estimation in Land Surface Models: Challenges and Opportunities with Data Assimilation and Machine Learning. *ESS Open Archive* . October 08, 2024. DOI: 10.22541/essoar.172838640.01153603/v1

Model development

Parameter optimisation

ECMWF

Citations

Booth, B. B., Jones, C. D., Collins, M., Totterdell, I. J., Cox, P. M., Sitch, S., ... & Lloyd, J. (2012). High sensitivity of future global warming to land carbon cycle processes. *Environmental Research Letters*, *7*(2), 024002.

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2018MS001453.

Kennedy, D., Dagon, K., Lawrence, D. M., Fisher, R. A., Sanderson, B. M., Collier, N., ... & Luo, Y. (2024). One-at-a-time parameter perturbation ensemble of the community land model, version 5.1. *Authorea Preprints*.

Novick, K. A., Ficklin, D. L., Baldocchi, D., Davis, K. J., Ghezzehei, T. A., Konings, A. G., ... & Wood, J. D. (2022). Confronting the water potential information gap. *Nature Geoscience*, *15*(3), 158-164.

Beylat, S., Raoult, N., Bacour, C., Douglas, N., Quaife, T., Bastrikov, V., ... & Peylin, P. (2025). Towards the Assimilation of Atmospheric CO 2 Concentration Data in a Land Surface Model using Adjoint-free Variational Methods. *EGUsphere*, *2025*, 1-33.

Douglas, N., Quaife, T., & Bannister, R. (2025). Exploring a hybrid ensemble–variational data assimilation technique (4DEnVar) with a simple ecosystem carbon model. *Environmental Modelling & Software*, 106361.

Bastrikov, V., MacBean, N., Bacour, C., Santaren, D., Kuppel, S., & Peylin, P. (2018). Land surface model parameter optimisation using in situ flux data: Comparison of gradient-based versus random search algorithms (a case study using ORCHIDEE v1. 9.5. 2). *Geoscientific Model Development*, *11*(12), 4739-4754.

Raoult, N., Beylat, S., Salter, J. M., Hourdin, F., Bastrikov, V., Ottlé, C., & Peylin, P. (2024). Exploring the potential of history matching for land surface model calibration. *Geoscientific Model Development*, *17*(15), 5779-5801.

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., & Richardson, A. D. (2012). Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences*, *9*(10), 3757-3776.

Salmon, E., Jégou, F., Guenet, B., Jourdain, L., Qiu, C., Bastrikov, V., ... & Ziemblińska, K. (2022). Assessing methane emissions for northern peatlands in ORCHIDEE-PEAT revision 7020. *Geoscientific Model Development*, *15*(7), 2813-2838.

Lang, S., Alexe, M., Chantry, M., Dramsch, J., Pinault, F., Raoult, B., ... & Rabier, F. (2024). AIFS--ECMWF's data-driven forecasting system. *arXiv preprint arXiv:2406.01465*.

Raoult, N., Douglas, N., MacBean, N., Kolassa, J., Quaife, T., Roberts, A. G., ... & Zobitz, J. (2024). Parameter Estimation in Land Surface Models: Challenges and Opportunities with Data Assimilation and Machine Learning. *Authorea Preprints*.

CALMIP: https://aimesproject.org/cal-lmip/

ECMWF