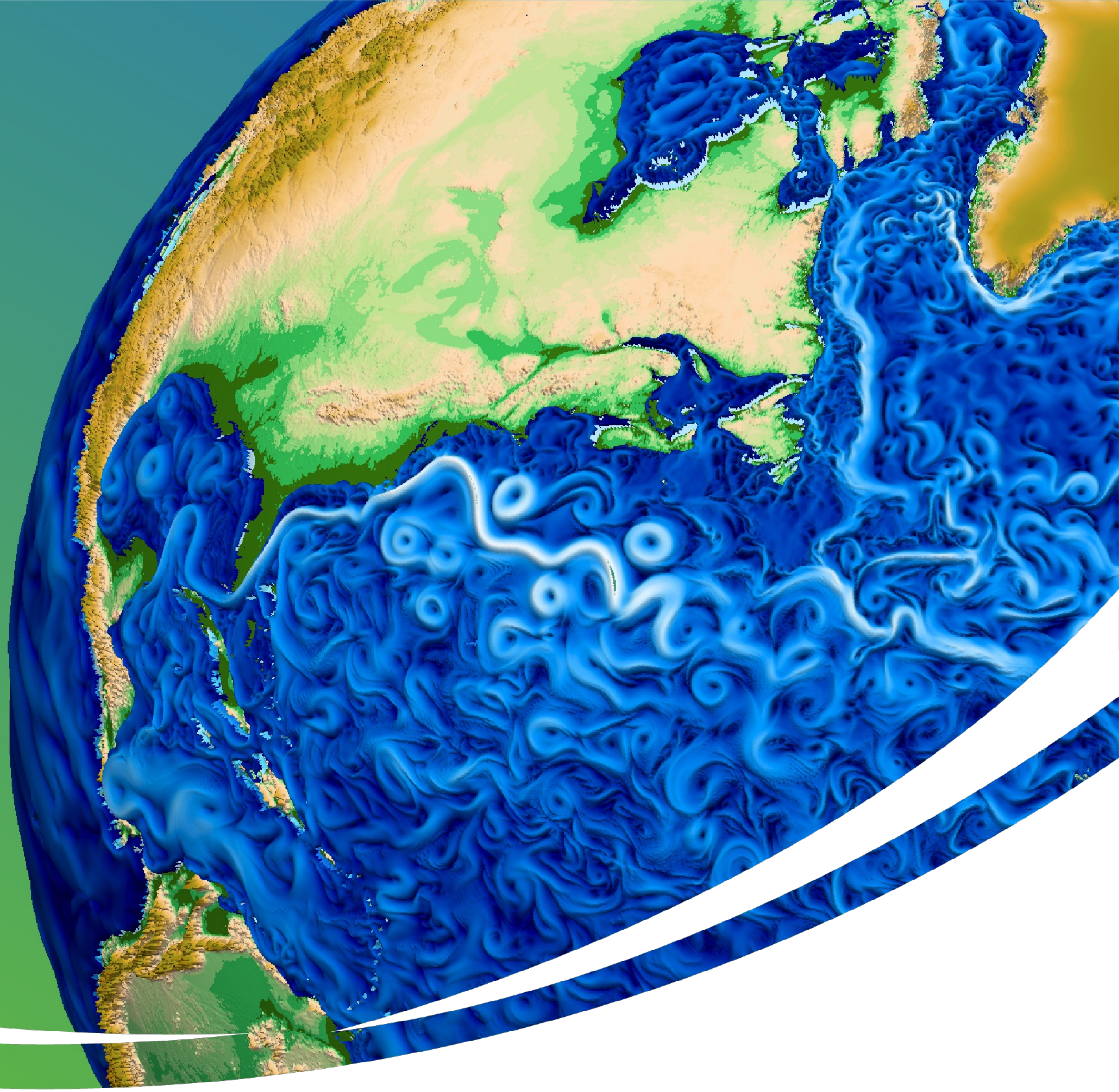
Reduced-Dimensional Neural Network Surrogate Construction for the E3SM Land Model

Khachik Sargsyan (SNL), Daniel Ricciuto (ORNL)



Sandia National Laboratories

Oak Ridge National Laboratory

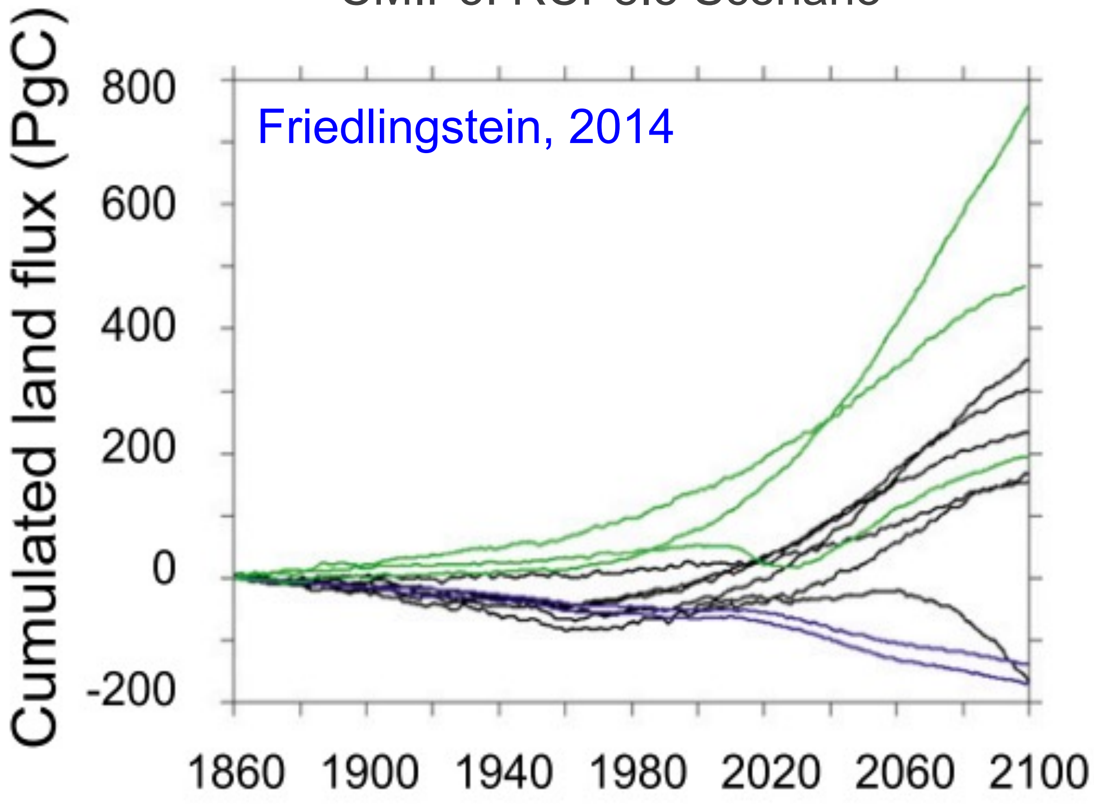
U.S. Department of Energy

FASTMath UQ Seminar
May 6, 2024

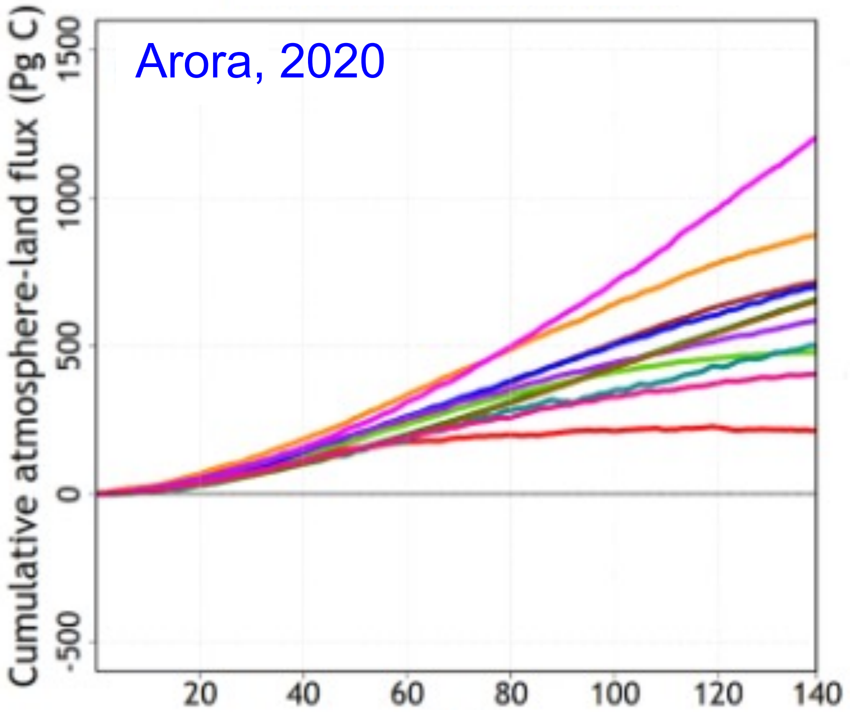
E3SM
Energy Exascale Earth System Model

Motivation: Uncertainties in Carbon Flux

CMIP5: RCP8.5 Scenario

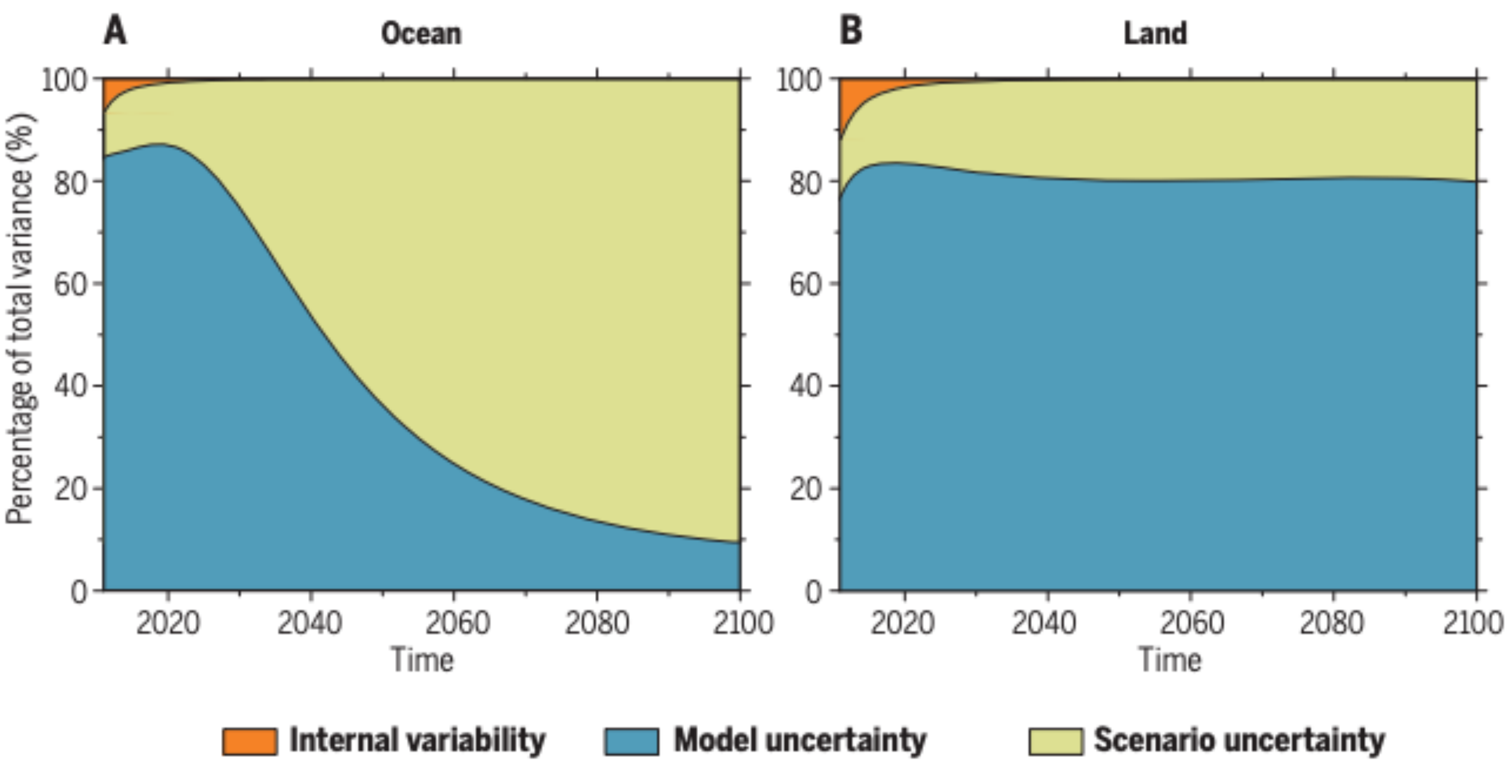


CMIP6: 1%/yr CO_2 incr. Scenario



Motivation: Model Uncertainty dominates for Land Model

Fig. 4. Ocean and land carbon cycle uncertainty. The percentage of total variance attributed to internal variability, model uncertainty, and scenario uncertainty in projections of cumulative global carbon uptake from 2006 to 2100 differs widely between (**A**) ocean and (**B**) land. The ocean carbon cycle is dominated by scenario uncertainty by the middle of the century, but uncertainty in the land carbon cycle is mostly from model structure. Data are from 12 ESMs using four different scenarios (94).



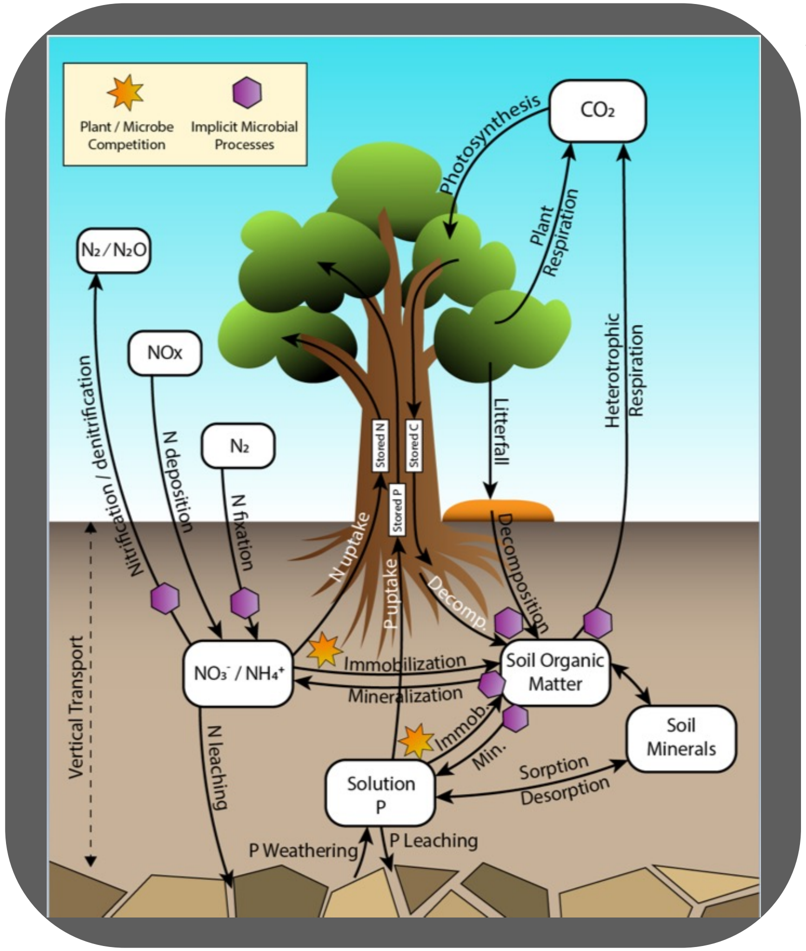
Bonan and Doney,
Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. Science, 2018

Overview: Surrogate-based Calibration of E3SM Land Model

- Land-surface model parametric uncertainty remains large
- High model expense → Need for model surrogates for sample-intensive studies, such as …
  - Global sensitivity analysis (forward UQ)
  - Model calibration (inverse UQ)
- Major challenges
  - Expensive model evaluation, small ensembles
  - High dimensional (spatio-temporal) outputs

- Reduced-dimensional, inexpensive surrogate construction via Karhunen-Loève expansions and Neural Networks (**KLNN surrogate**)
- Surrogate enables global sensitivity analysis and **Bayesian model calibration**

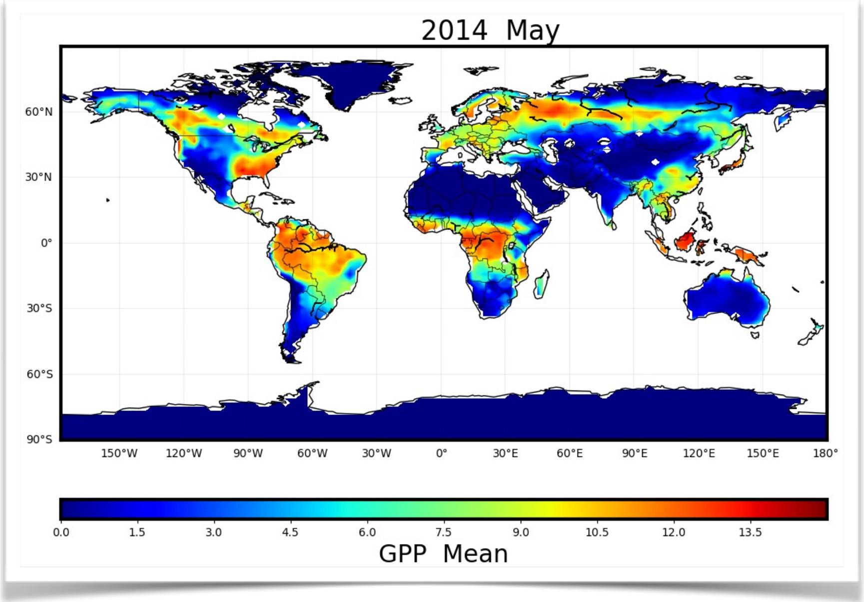
E3SM Land Model (ELM): focus on carbon and energy cycle



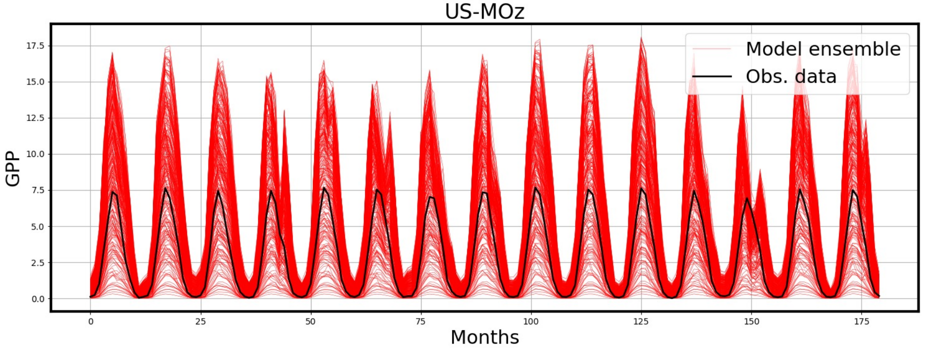
Satellite Phenology version used for this study

Quantity of Interest:
Gross primary productivity (GPP)…

… resolved in space, …

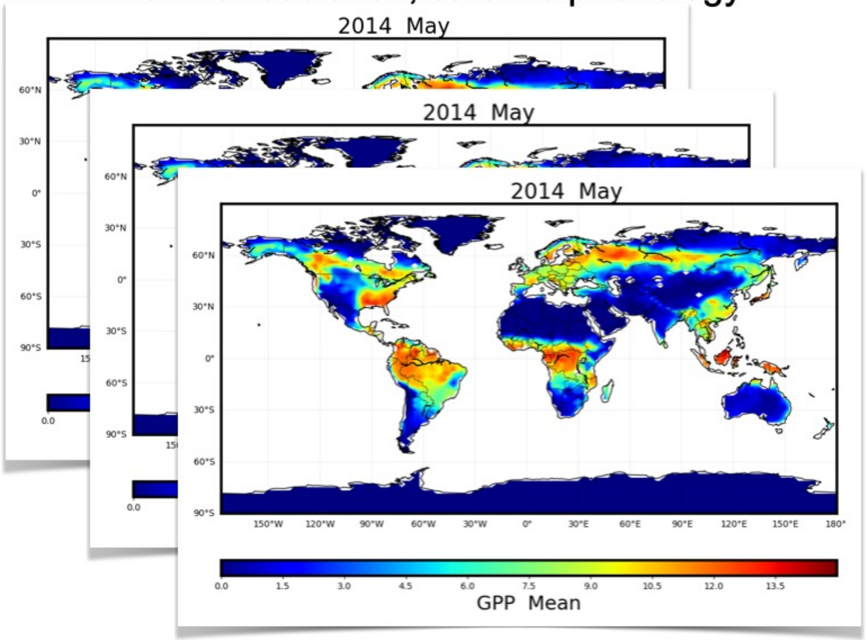


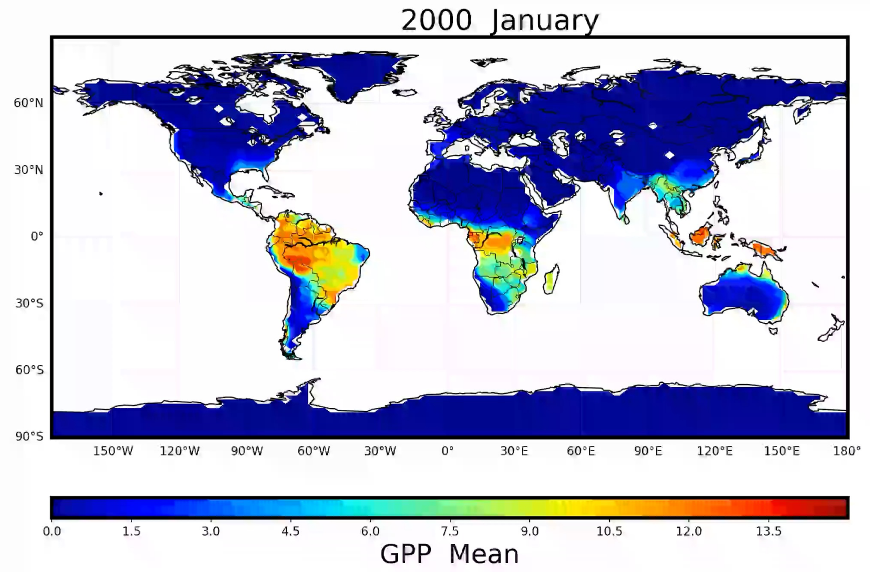
… and in time.

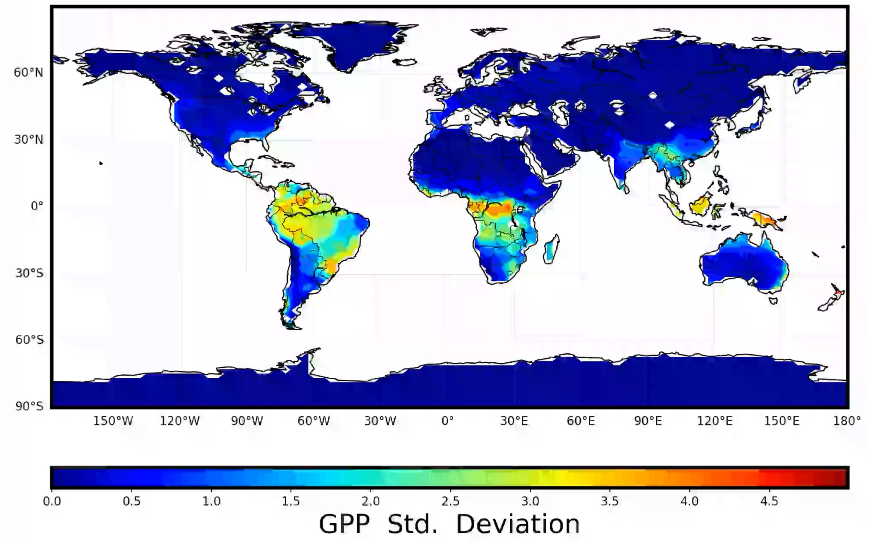


Model Ensemble (275 samples)

1.9x2.5 resolution, satellite phenology







Perturbed Parameters

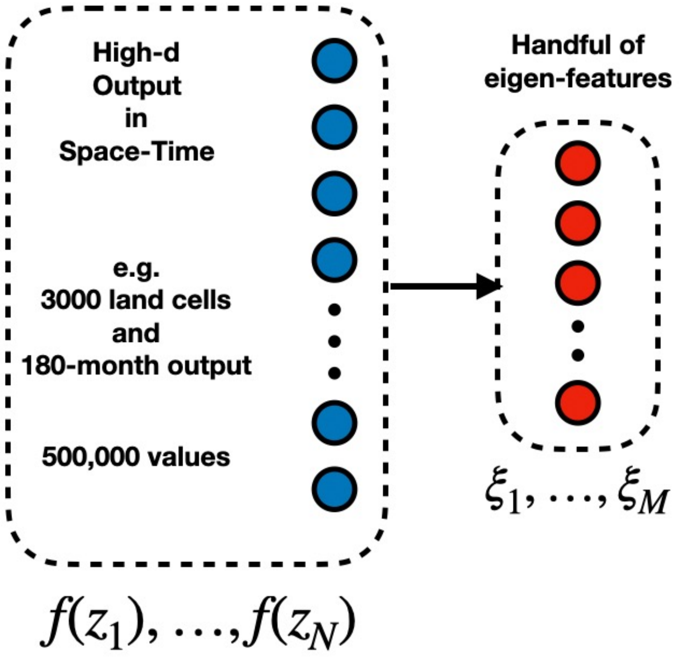
Parameter	Description	Min	Max
flnr	Fraction of leaf in in RuBisCO	0	0.25
mbbopt	Stomatal slope (Ball-Berry)	2	13
bbbopt	Stomatal intercept (Ball-Berry)	1000	40000
roota_par	Rooting depth distribution	1	10
vcmaxha	Activation energy for Vcmax	50000	90000
vcmaxse	Engropy for Vcmax	640	700
jmaxha	Activation energy for jmax	50000	90000
dayl_scaling	Day length factor	0	2.5
dleaf	Characteristic leaf dimension	0.01	0.1
xl	Leaf/stem orientation index	-0.6	0.8

Dimensionality Reduction via Karhunen-Loève Expansion

$$f(\lambda; z) \approx \bar{f}(z) + \sum_{m=1}^{M} \xi_m(\lambda)\sqrt{\mu_m}\phi_m(z)$$

Uncertain parameters (λ)   "Certain" conditions (z)

- Spatio-temporal model output $f(\lambda; z)$, where $z = (x, y, t)$
- Output field has large dimensionally $N = N_x \times N_y \times N_t$
- Eigenpairs $(\mu_m, \phi_m(z))$ are found via eigen-solve
- Analysis reduces to $M \ll N$ eigenfeatures $\xi_1, \ldots, \xi_m$
- Under the hood: this is essentially an SVD



KL is essentially a Singular Value Decomposition

KL $$f(\lambda^k; zi) - \bar{f}(z_i) \approx \sum_{m=1}^{M} \xi_m(\lambda^k)\sqrt{\mu_m}\phi_m(zi)$$

$$F_{ki} = \sum_{m=1}^{M} U_{km}\Sigma_{mm}V_{im}$$

SVD $$F = U\,\Sigma\,V^T$$

Karhunen-Loève expansion

-– is centralized (first subtract the mean)

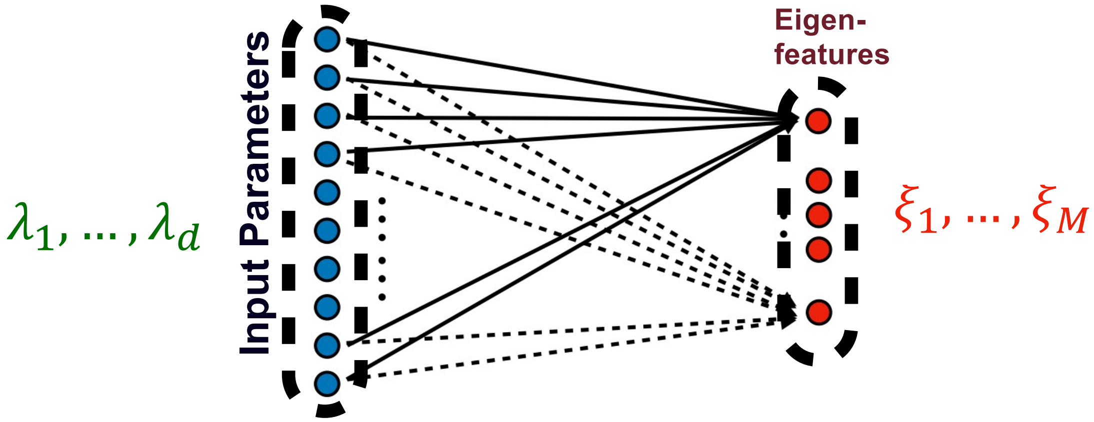
-– often comes with the continuous form

-– has random variable interpretation for the latent features (aka left singular vectors) ξ_m

KL+PC = reduced dimensional spatio-temporal surrogate

The goal is to construct a surrogate with respect to uncertain parameters λ, such that $f(\lambda; z_i) \approx f_s(\lambda; z_i)$ for all conditions z_i.

Instead of building surrogate for each individual z_i for $i = 1, \dots, N$,
we construct polynomial chaos (PC) surrogate for $\xi_1, \dots, \xi_M$ where $M \ll N$.



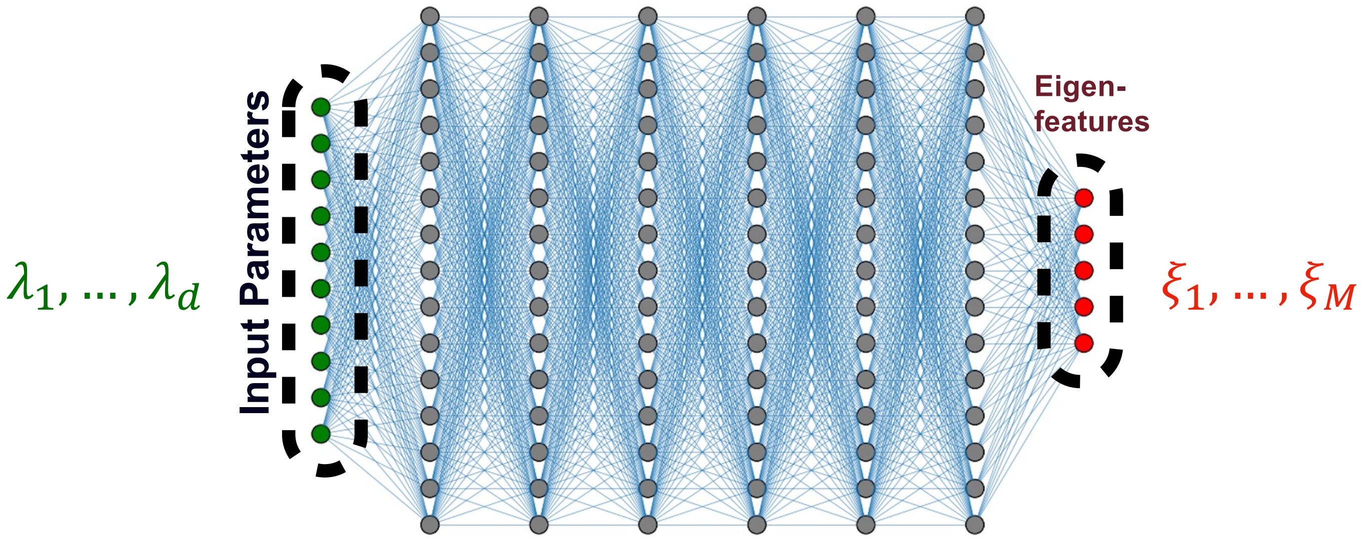
$$f(\lambda; z) \approx \bar{f}(z) + \sum_{m=1}^{M} \xi_m(\lambda) \sqrt{\mu_m} \phi_m(z)$$

$\xi_m^{PC}(\lambda)$

KL+NN = reduced dimensional spatio-temporal surrogate

The goal is to construct a surrogate with respect to uncertain parameters λ, such that $f(\lambda; z_i) \approx f_s(\lambda; z_i)$ for all conditions z_i.

Instead of building surrogate for each individual z_i for $i = 1, \dots, N$,
we construct neural network (NN) surrogate for $\xi_1, \dots, \xi_M$ where $M \ll N$.



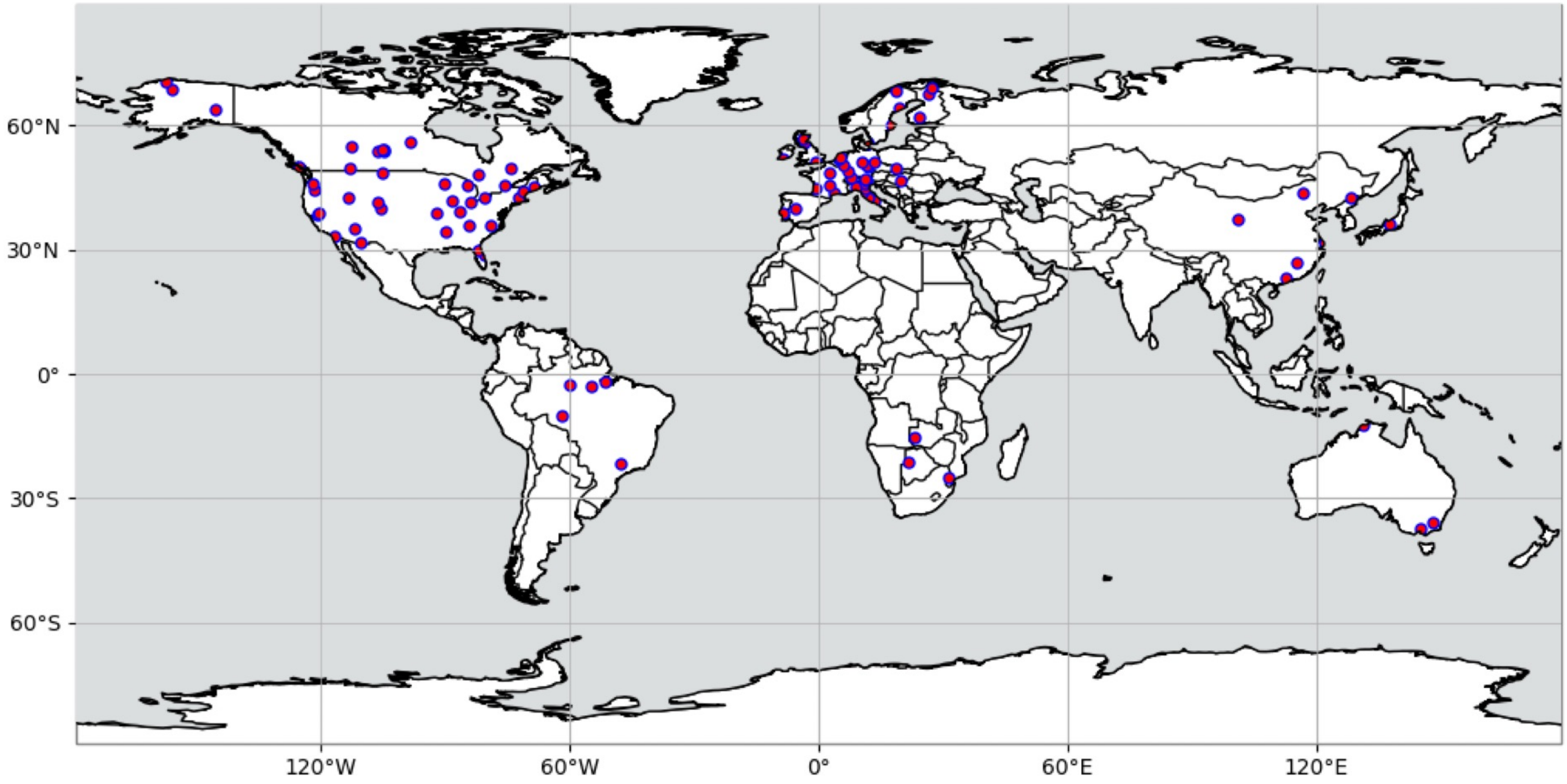
$$f(\lambda; z) \approx \overline{f}(z) + \sum_{m=1}^{M} \xi_m(\lambda) \sqrt{\mu_m} \phi_m(z)$$

$\xi_m^{NN}(\lambda)$

PC vs NN comparison

Polynomial Chaos	Simple regression, easy to train	GSA and variance decomposition, More interpretable
Neural Network	More flexible, highly customizable	Multiple outputs at once, More accurate (in theory)

96 FLUXNET sites

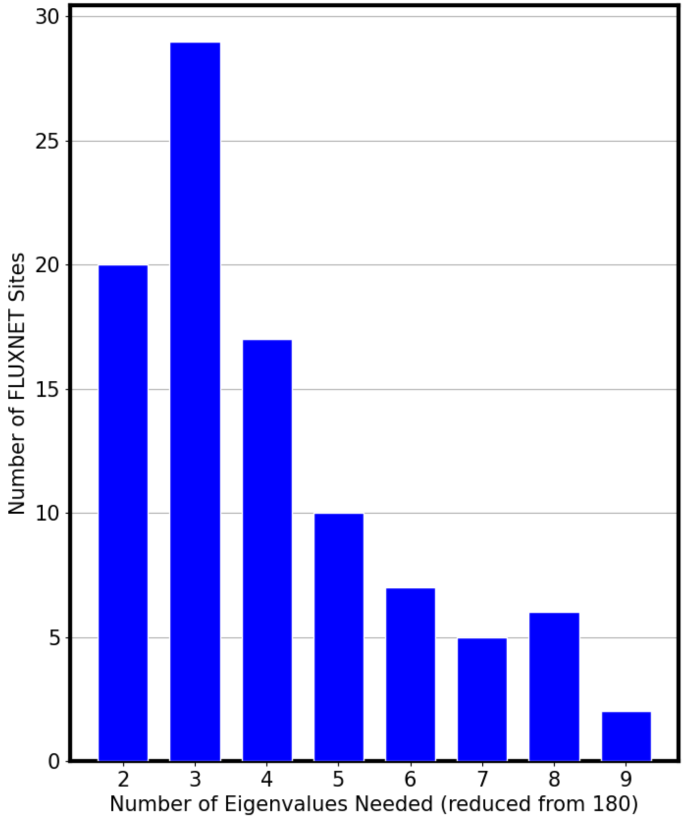


Several case studies

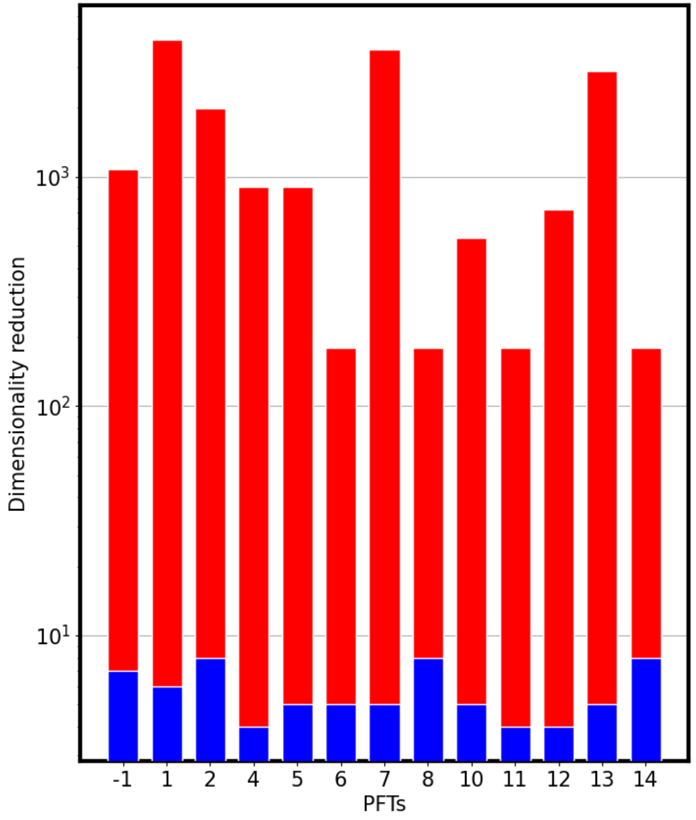
Time / Space	$N_t = 180$ Months (full 15 years)	$N_t = 12$ Months (average out interannual)	$N_t = 4$ Seasons (average out within seasons)	$N_t = 1$ (global time-average)
FLUXNET sites $N_x = 96$ (or group by PFTs)	F180	F12	F4	F1
Global 144x96 $N_x \cong 4000$ **vegetated cells** (or regional zoom)	G180	G12	G4	G1

Dimensionality reduction via KL

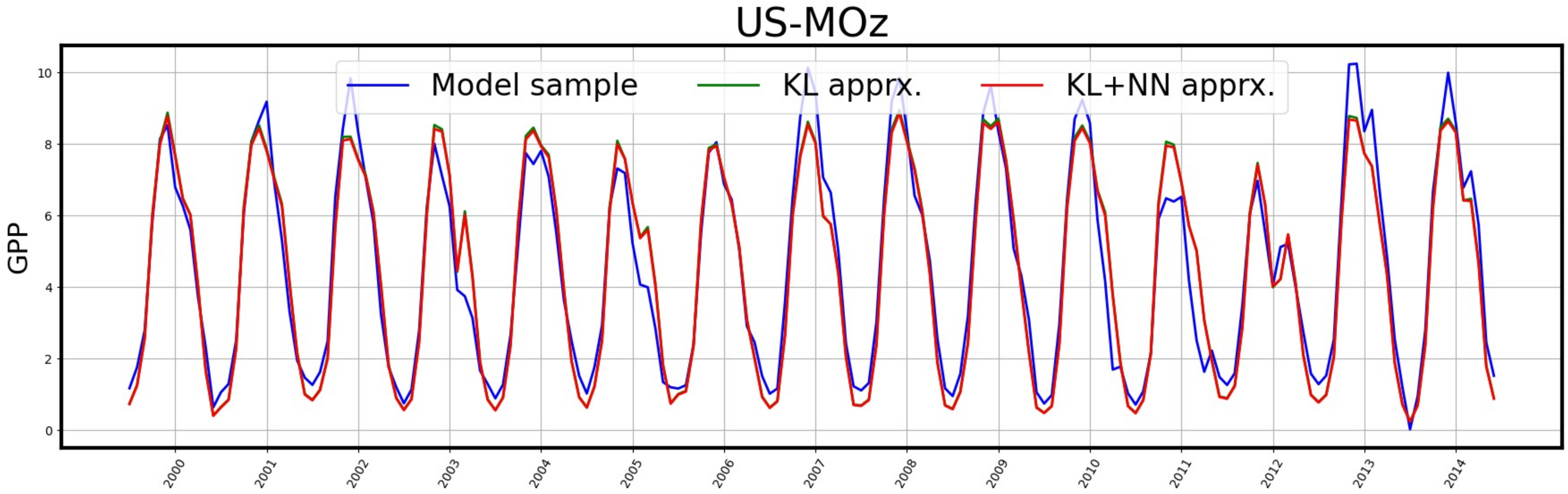
Per-site dimensionality reduction



Per-PFT dimensionality reduction

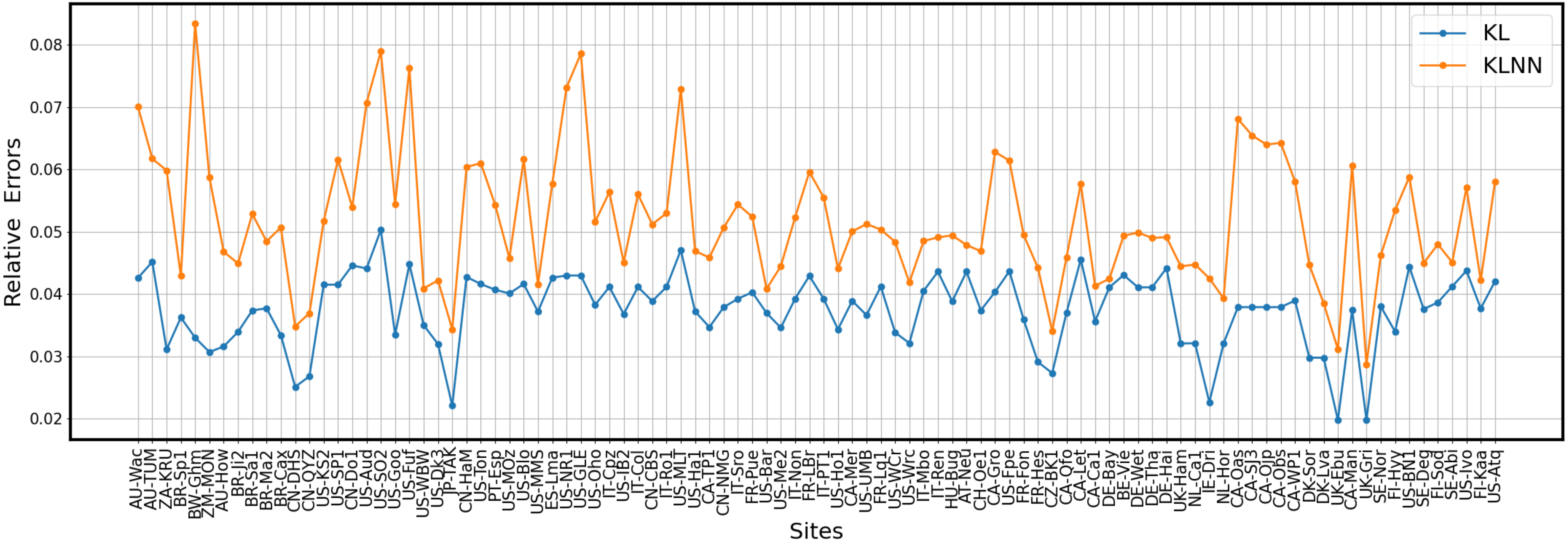


KL+NN a single training sample approximation

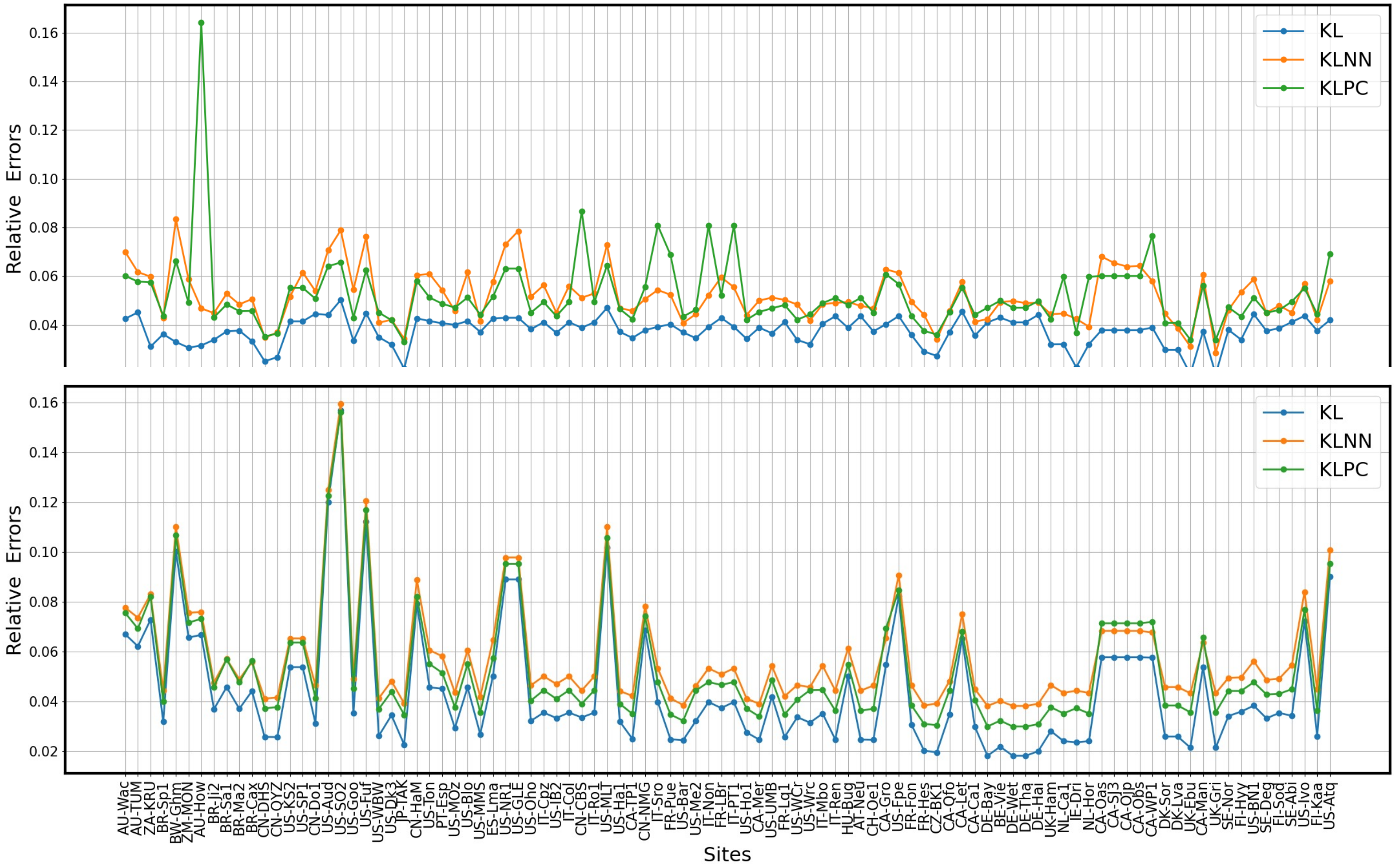


KL+NN surrogate performance

Instead of 96x180=**17280** surrogates, we build
a single NN surrogate in the reduced, **8**-dimensional latent space



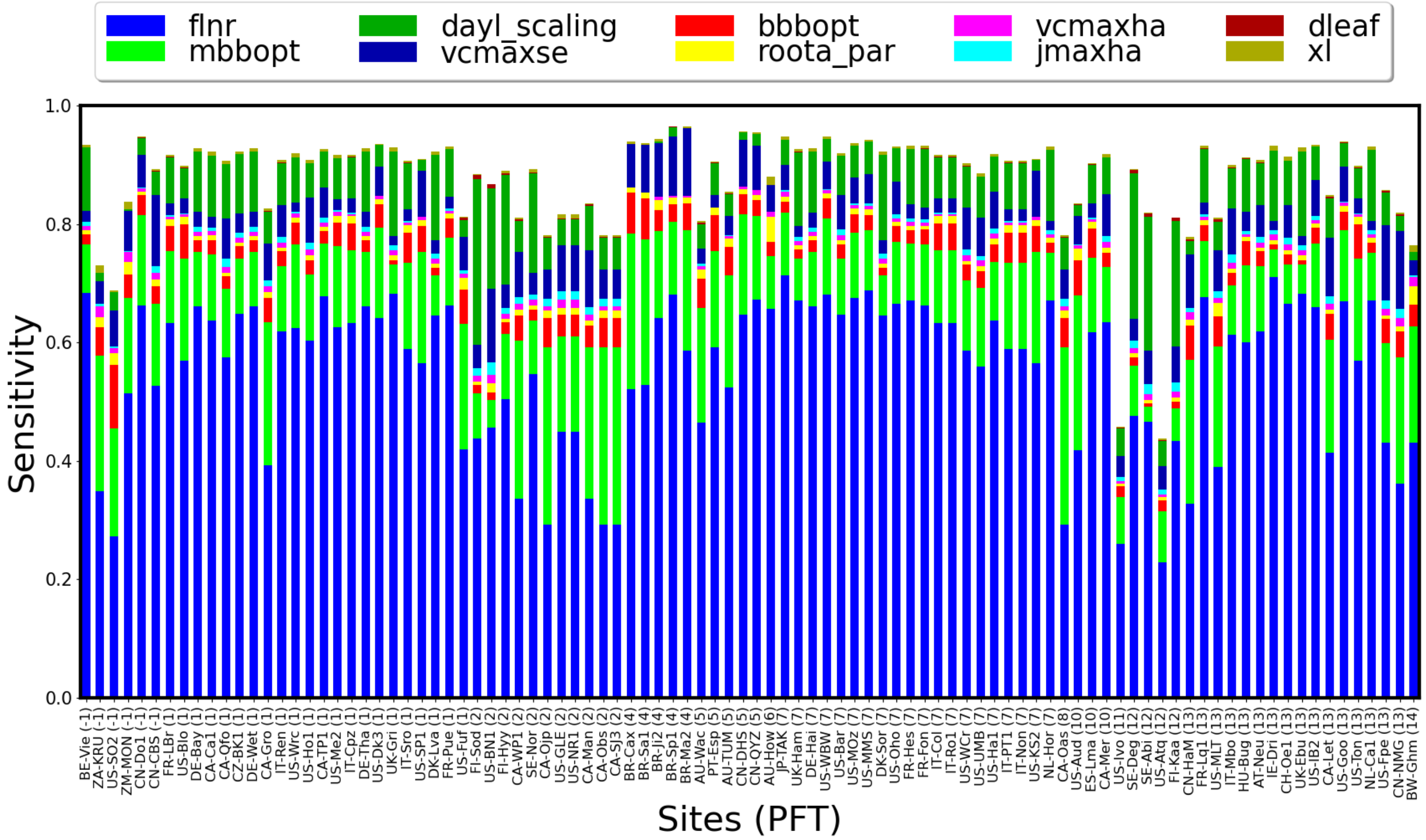
PC vs NN comparison



96 temporal surrogates with each 180 outputs

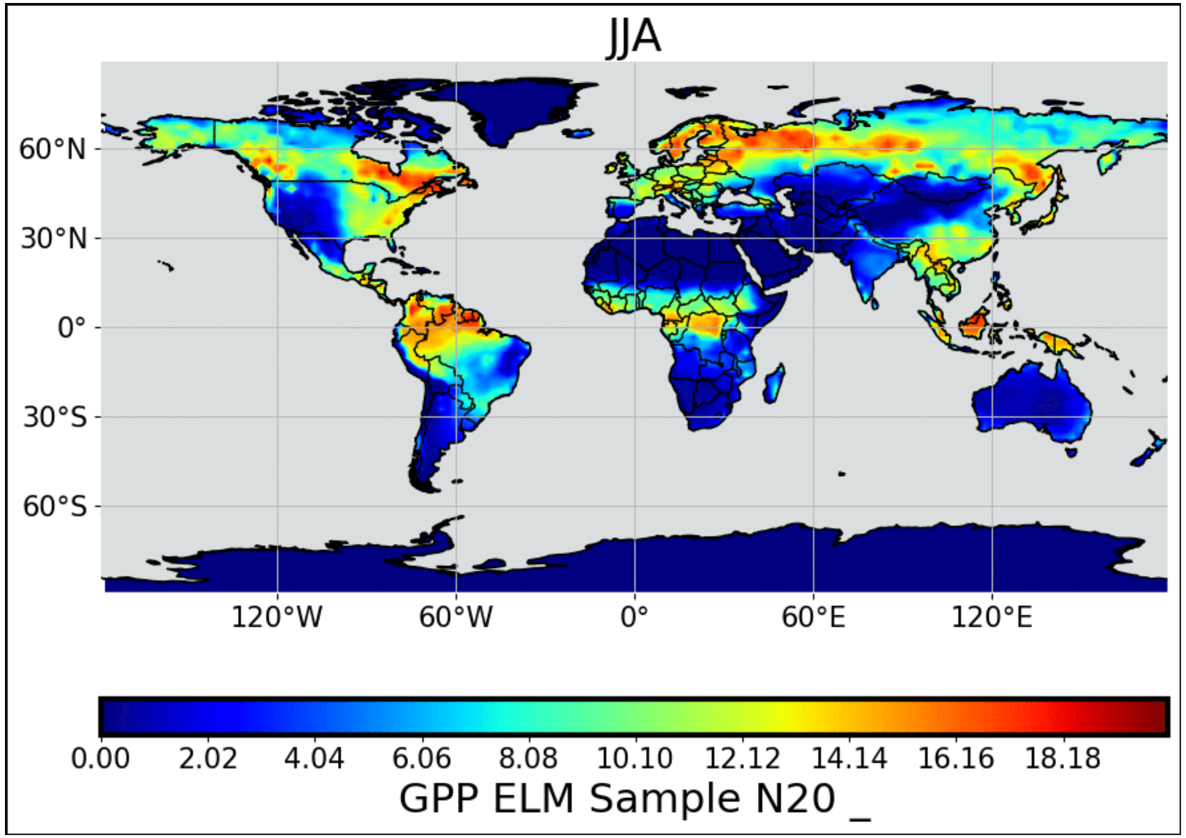
Single spatio-temporal surrogate with 96x180 outputs

Sensitivity at 96 FLUXNET sites: RuBisCO leaf fraction is the most impactful parameter

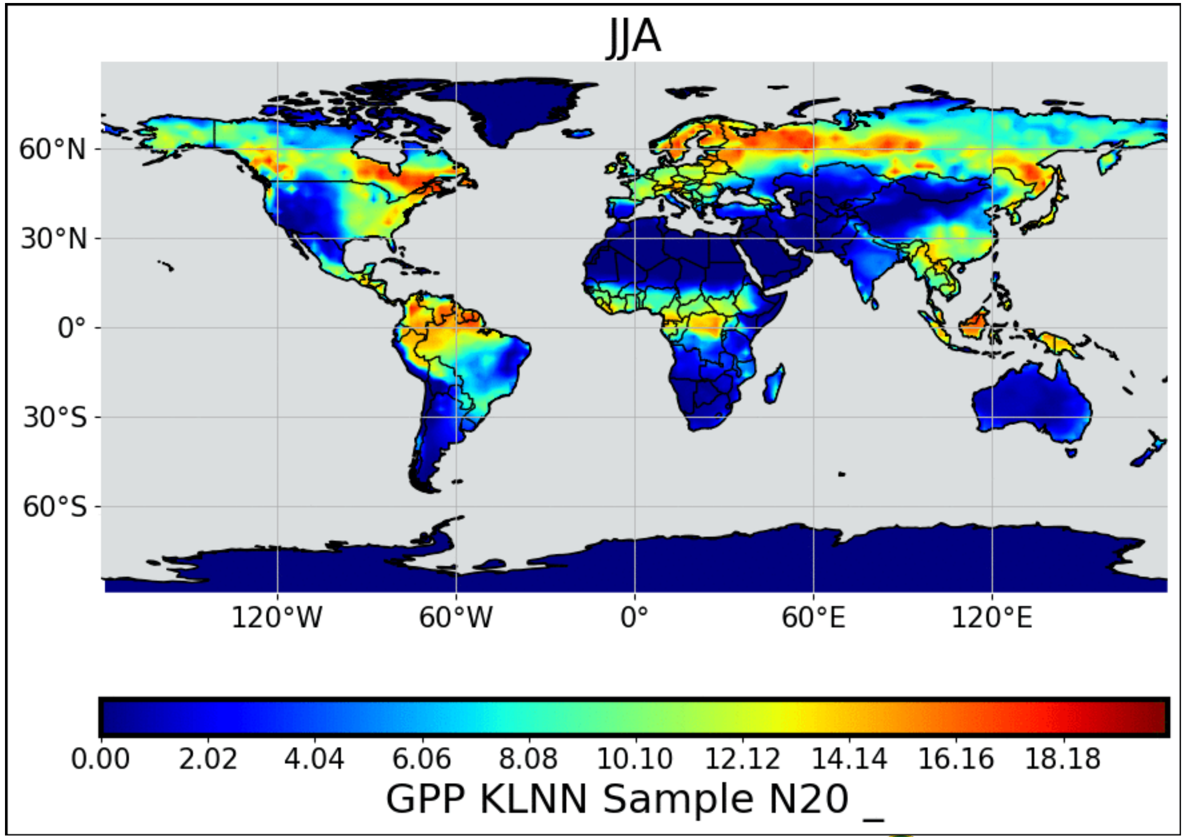


Dimensionality reduction from 4000 cells x 4 seasons = **16000** to **11**-dimensional latent space

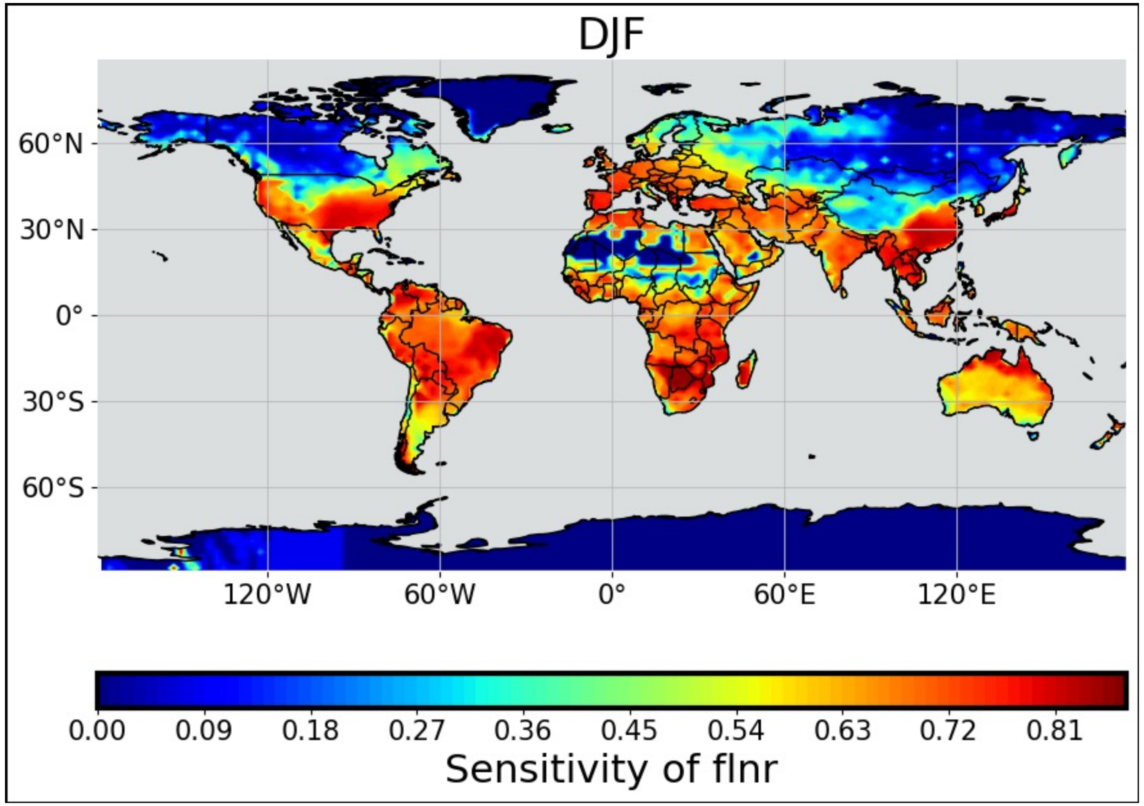
ELM Model Samples

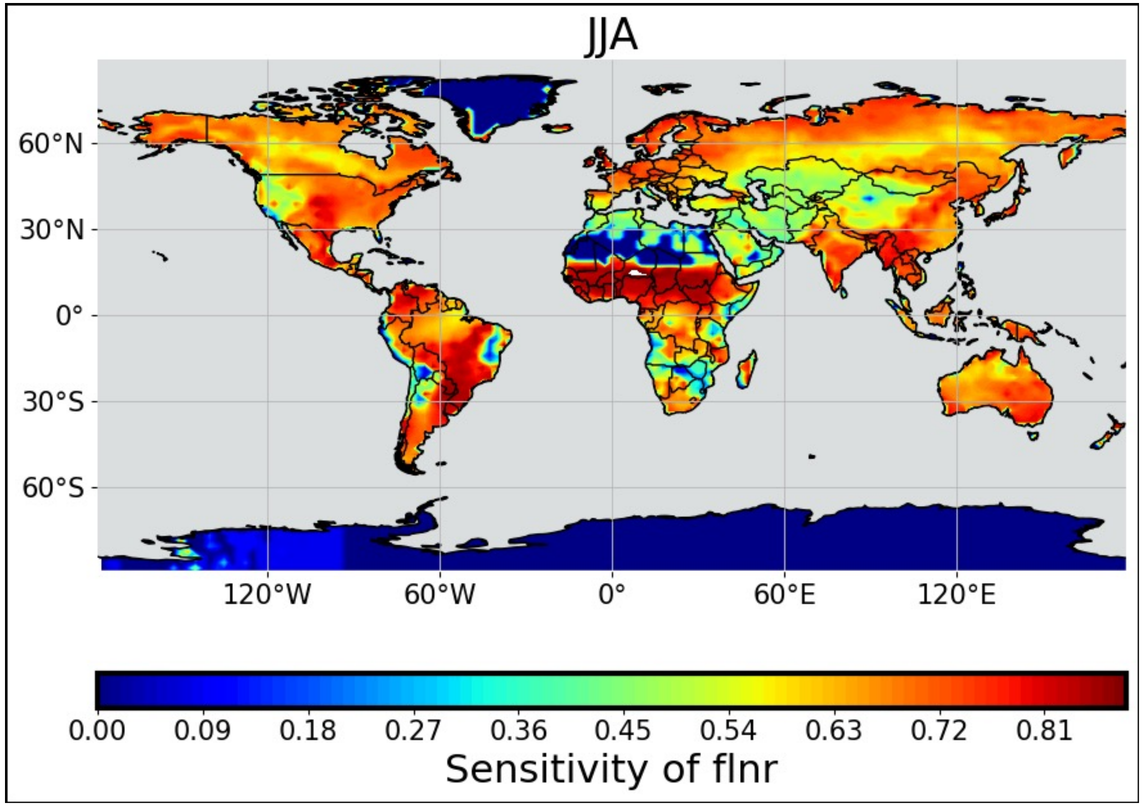


KLNN Surrogate Samples



fLNR sensitivity across the globe



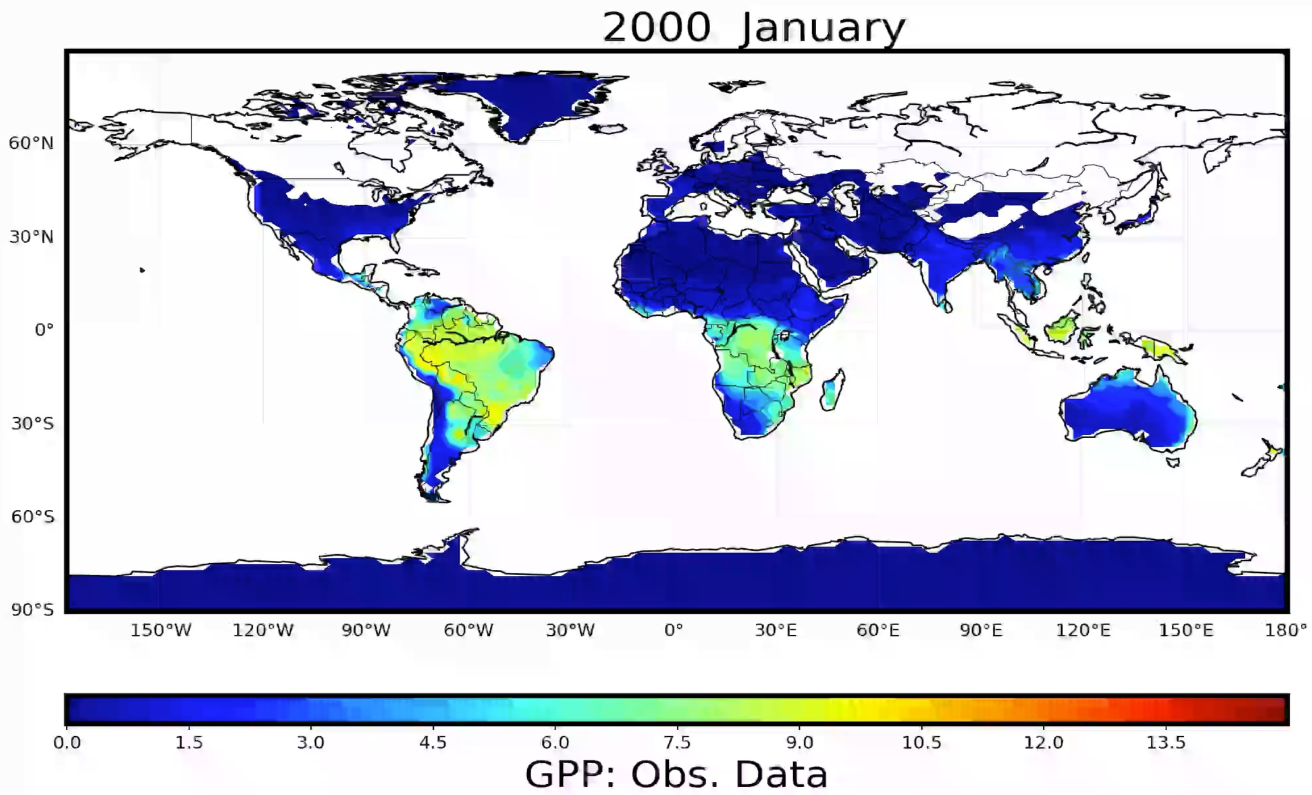


Surrogate-enabled Bayesian calibration

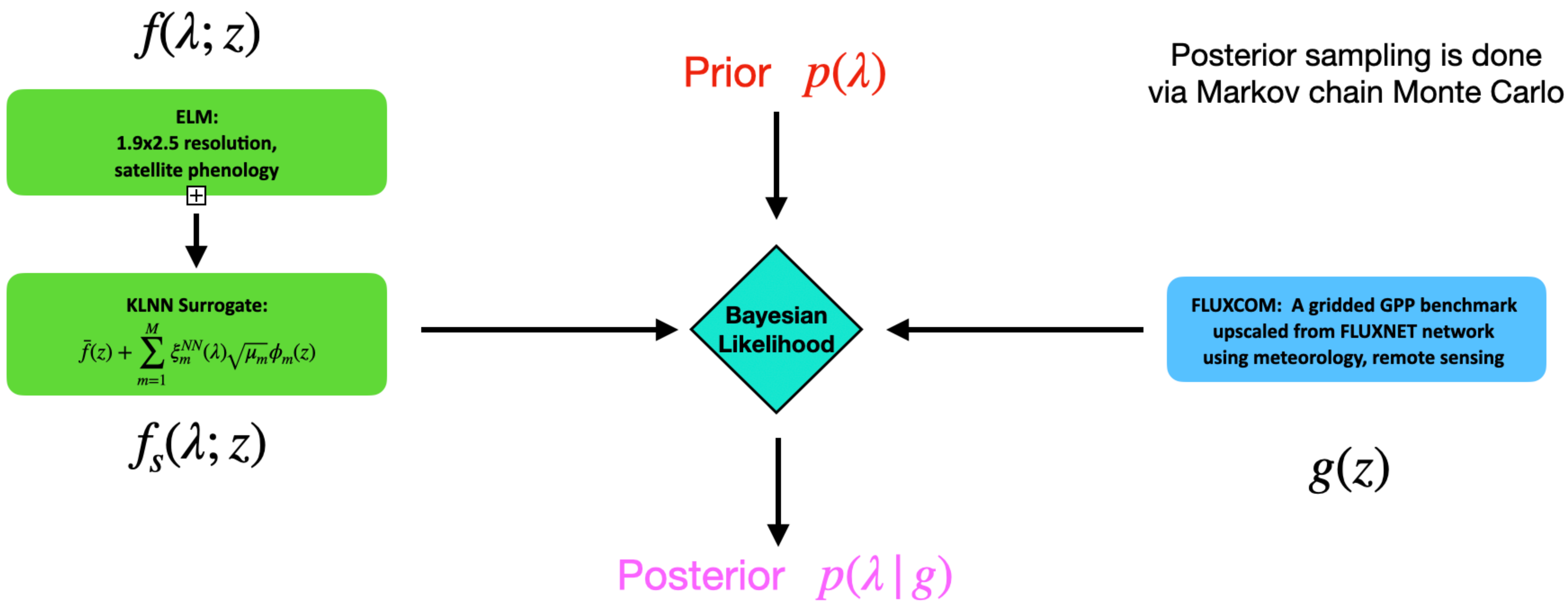
Reference Data

FLUXCOM: A gridded GPP benchmark upscaled from FLUXNET network using meteorology, remote sensing

https://www.fluxcom.org/



Bayes' formula $p(\lambda \mid g) \propto p(g \mid \lambda)\, p(\lambda)$



Bayesian Likelihood is constructed in the reduced space

Bayes' formula $p(\lambda|g) \propto p(g|\lambda)p(\lambda)$

KLNN surrogate:

$$f(\lambda; z) \approx \bar{f}(z) + \sum_{m=1}^{M} \xi_m^{NN}(\lambda)\sqrt{\mu_m}\phi_m(z)$$

Project observed data to the KL eigenspace:

$$g(z) \approx \bar{f}(z) + \sum_{m=1}^{M} \eta_m\sqrt{\mu_m}\phi_m(z)$$

Pointwise likelihood (naïve) :

$$L_g(\lambda) \equiv p(g|\lambda) \propto \exp\left(-\sum_{i=1}^{N} \frac{(g(z_i) - f(\lambda; z_i))^2}{2\sigma_i^2}\right)$$

Reduced likelihood :

$$L_g(\lambda) \equiv p(g|\lambda) \propto \exp\left(-\sum_{m=1}^{M} \frac{(\eta_m - \xi_m^{NN}(\lambda))^2}{2\sigma^2}\right)$$

Eigenfeatures ξ_m's are uncorrelated, zero-mean, unit variance,
hence iid gaussian likelihood is a much better assumption in the reduced space.

Likelihood in the reduced space is still Gaussian, but MVN

KLNN surrogate:

$$f(\lambda; z) \approx \bar{f}(z) + \sum_{m=1}^{M} \xi_m^{NN}(\lambda)\sqrt{\mu_m}\phi_m(z)$$

Project observed data to the KL eigenspace:

$$g(z) \approx \bar{f}(z) + \sum_{m=1}^{M} \eta_m\sqrt{\mu_m}\phi_m(z)$$

Pointwise likelihood (old) :

$$L_g(\lambda) \equiv p(g|\lambda) \propto \exp\left(-\sum_{i=1}^{N}\frac{(g(z_i) - f(\lambda; z_i))^2}{2\sigma_i^2}\right)$$

Data model (old) : i.i.d. Normal

$$g(z_i) = f(\lambda; z_i) + \sigma_i\epsilon_i$$

Latent-space likelihood (new) :

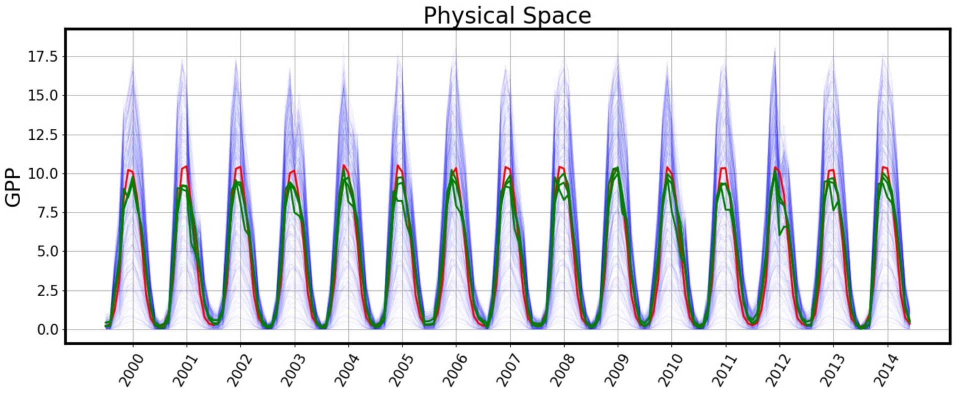
$$L_g(\lambda) \equiv p(g|\lambda) \propto \exp\left(-\sum_{m=1}^{M}\frac{(\eta_m - \xi_m^{NN}(\lambda))^2}{2\sigma^2}\right)$$

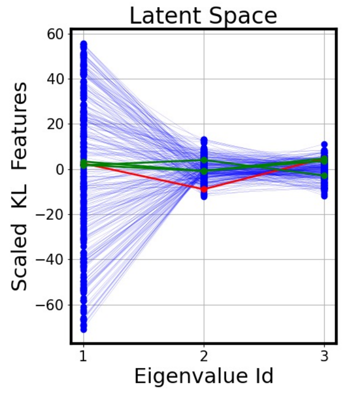
Data model (new) : MVN (physics-based)

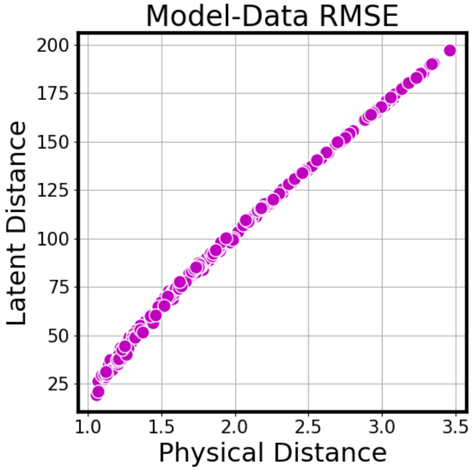
$$\eta_m = \xi_m^{NN}(\lambda) + \sigma\tilde{\epsilon}_m$$

$$g(z_i) = f(\lambda; z_i) + \sum_{m=1}^{M}\tilde{\epsilon}_m\sqrt{\mu_m}\phi_m(z_i)$$

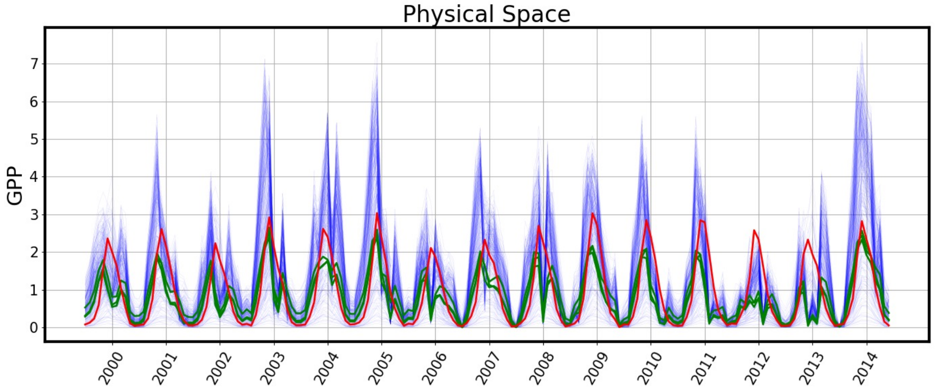
Latent space distance is well-correlated with the physical distance between model and data

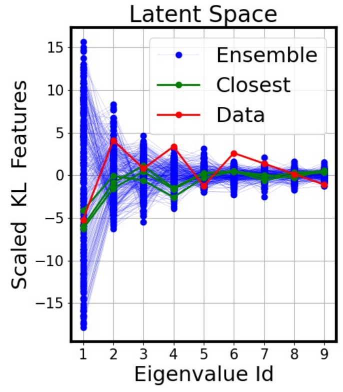


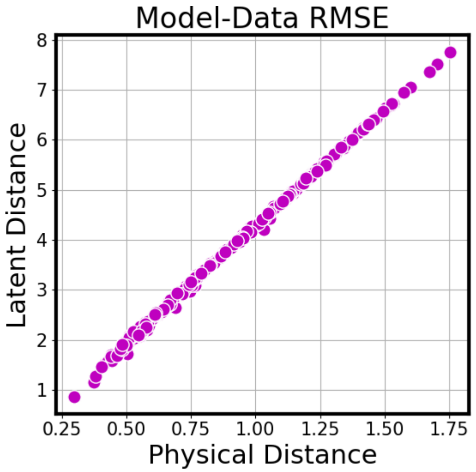




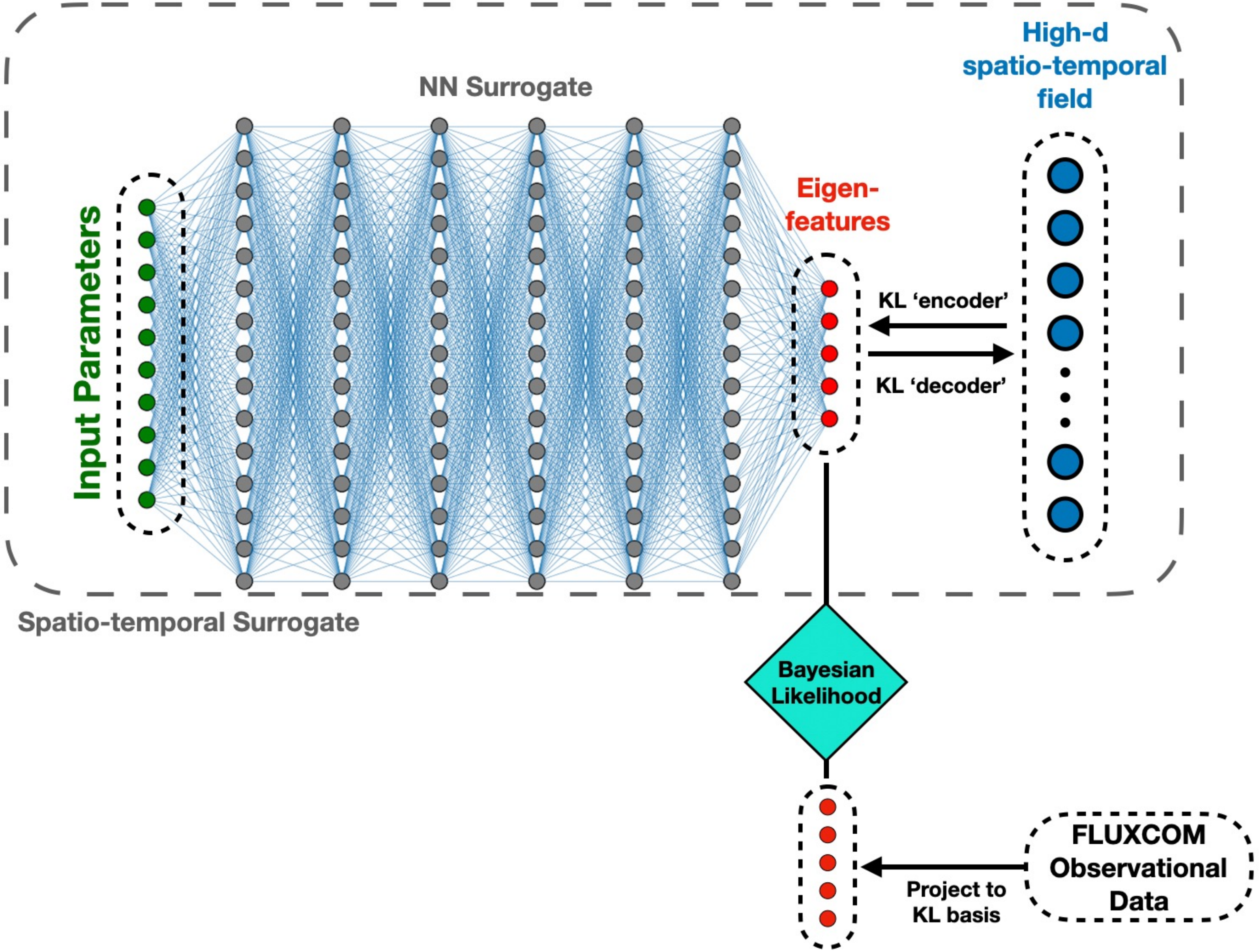
US-Ha1





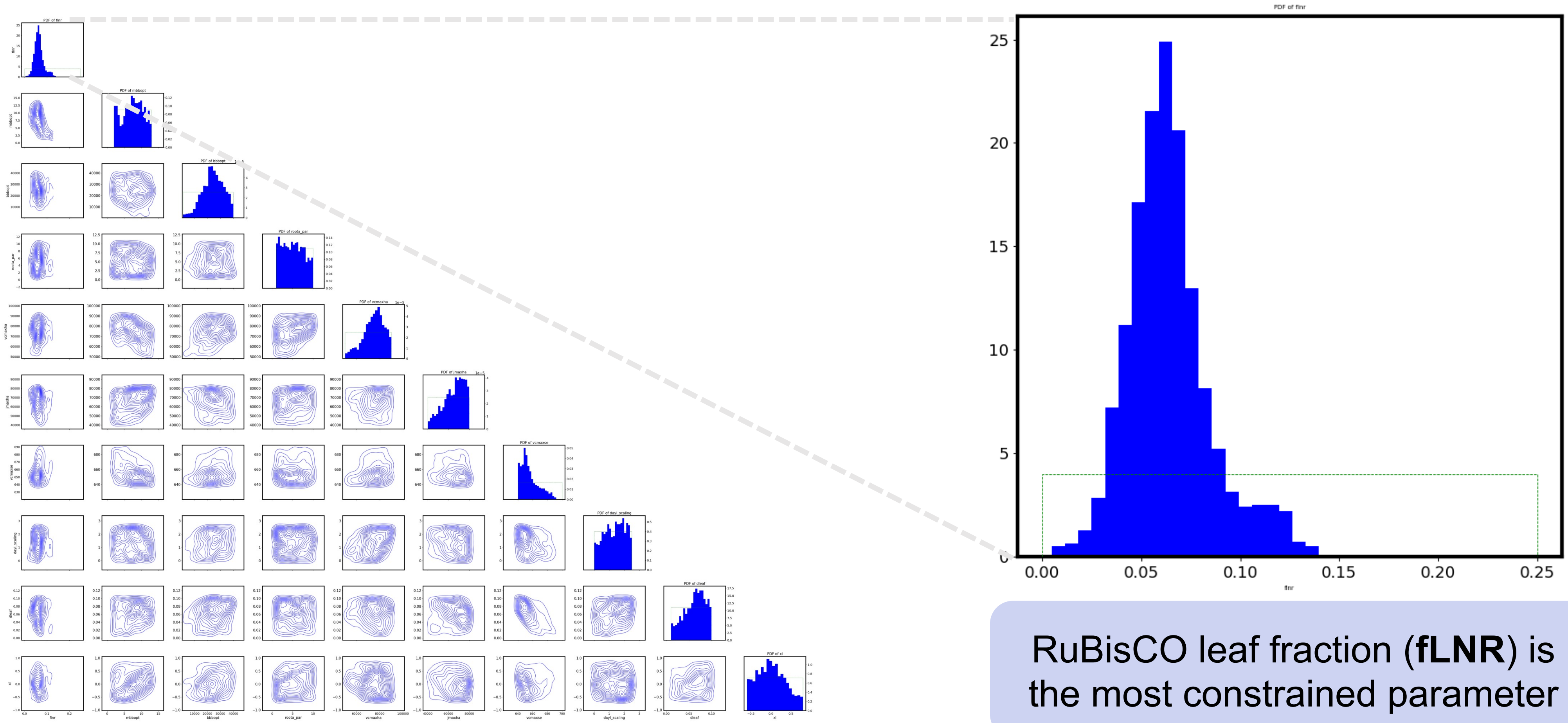


US-GLE



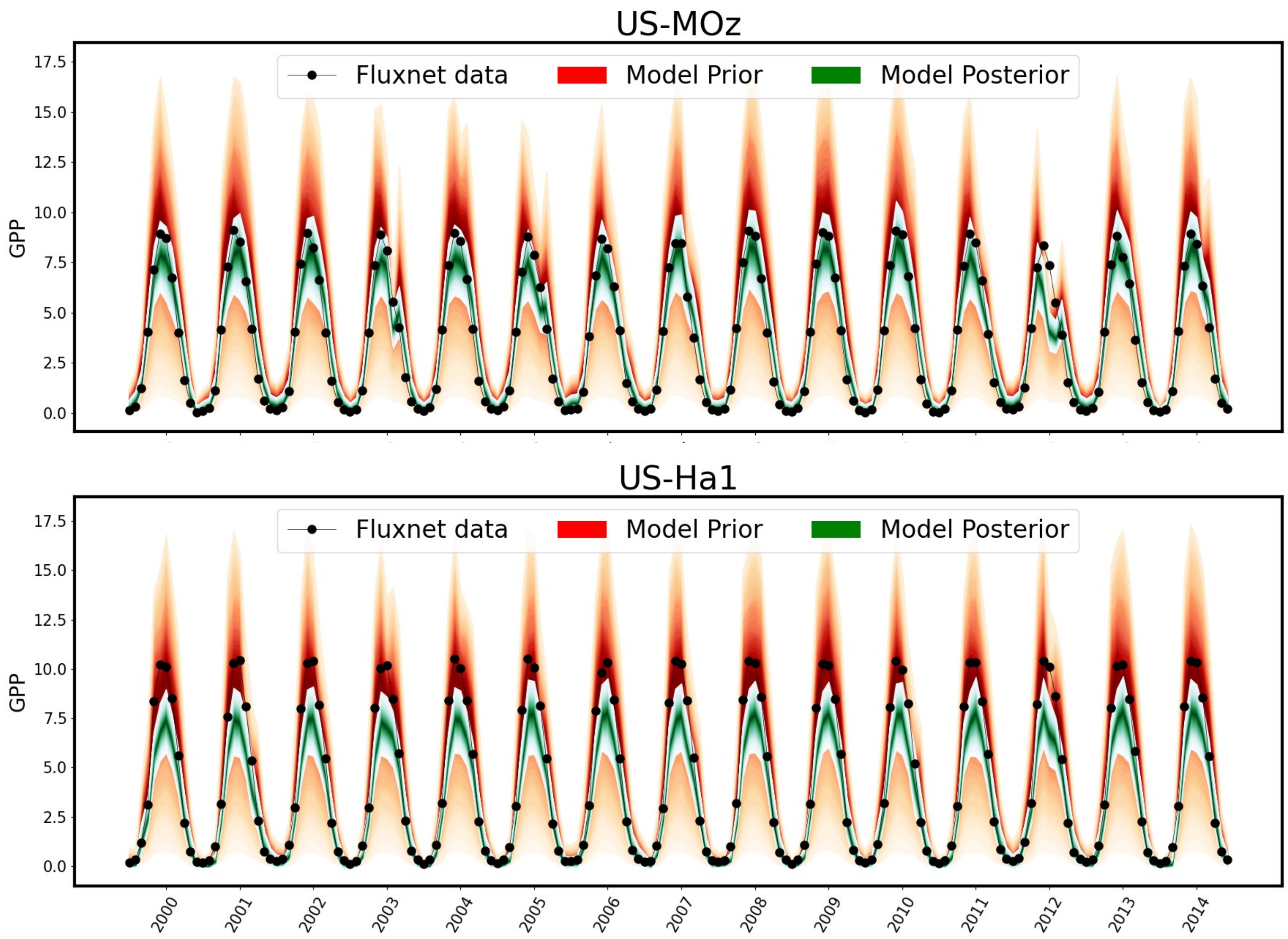
Surrogate-enabled calibration workflow incorporates both forward and inverse UQ tasks

Bayesian calibration enabled by KLNN surrogate



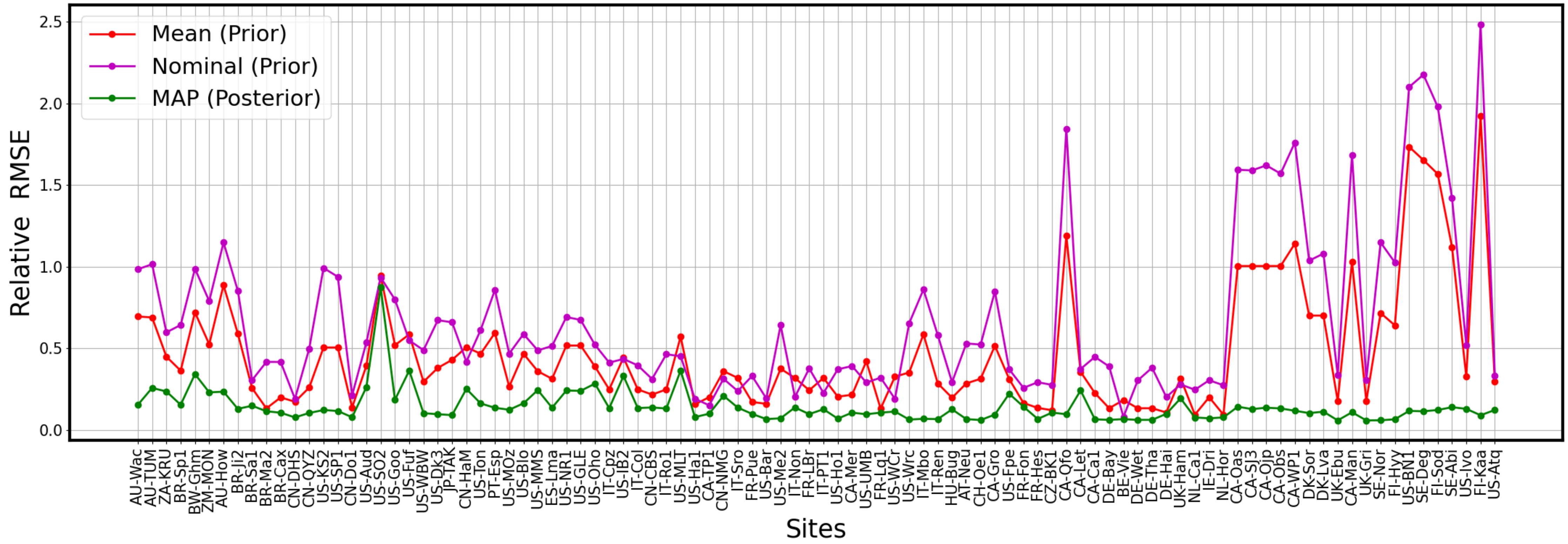
RuBisCO leaf fraction (**fLNR**) is the most constrained parameter

Time evolution of GPP at select FLUXNET sites



Calibration brings model prediction closer to reference data

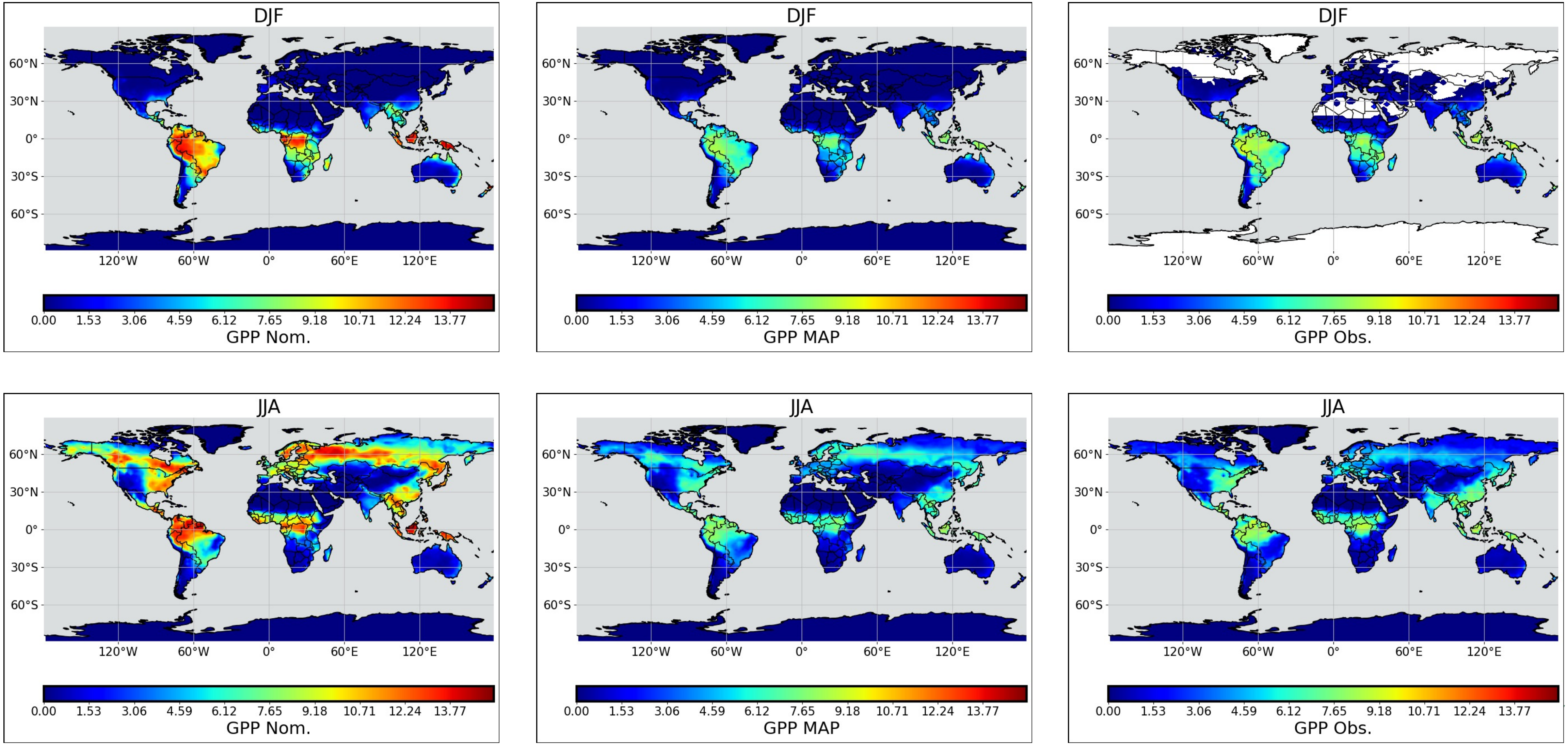
Site-specific parameters



Nominal parameter (prior) | Max a posteriori (MAP) | Reference data

Winter

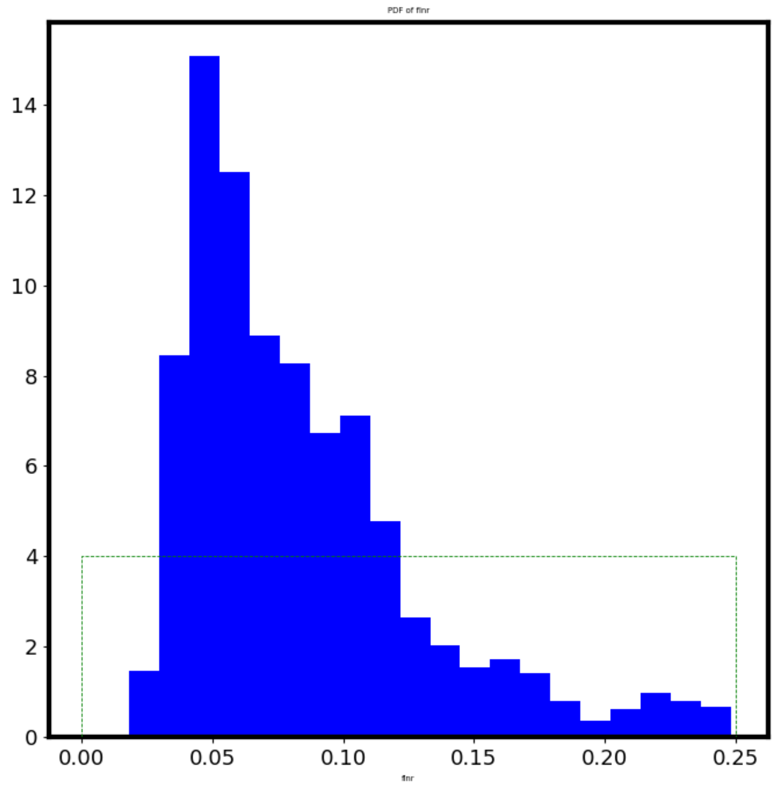
Summer



Two calibration regimes

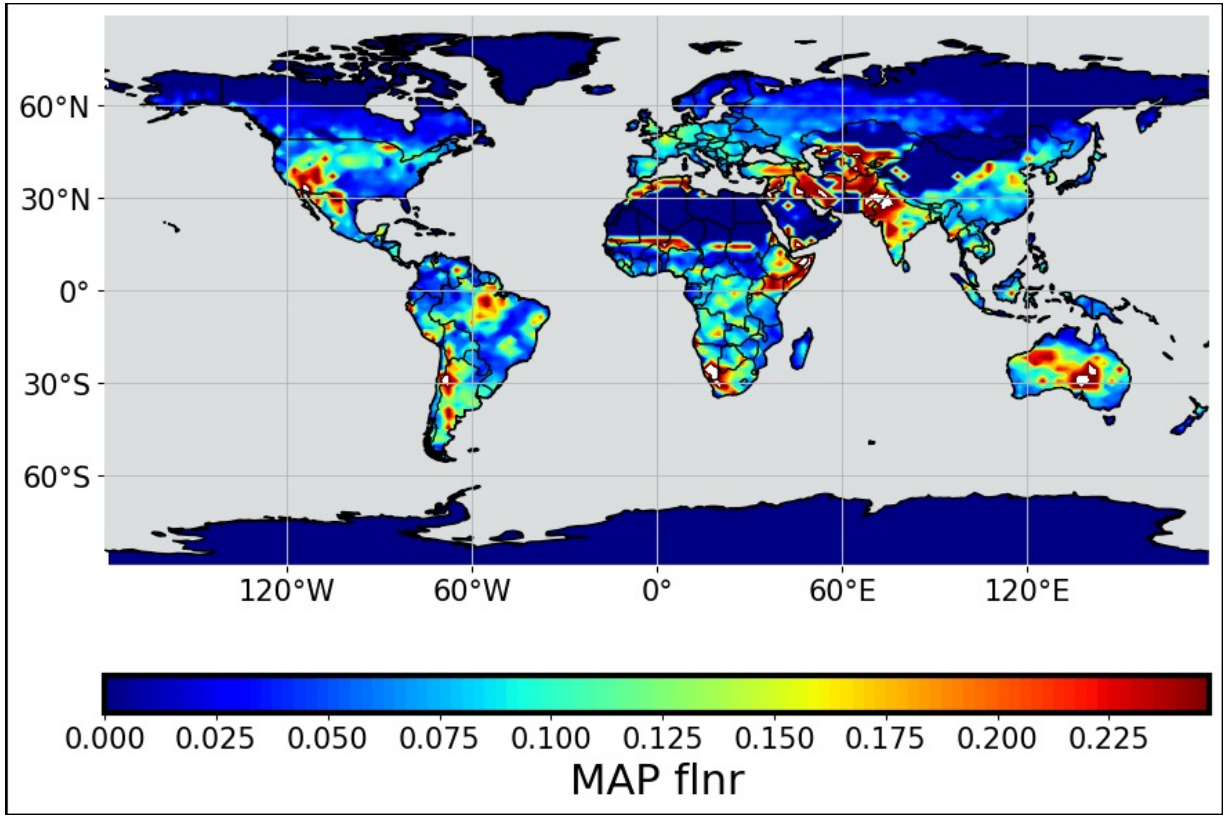
One global surrogate

Fixed global fLNR parameter

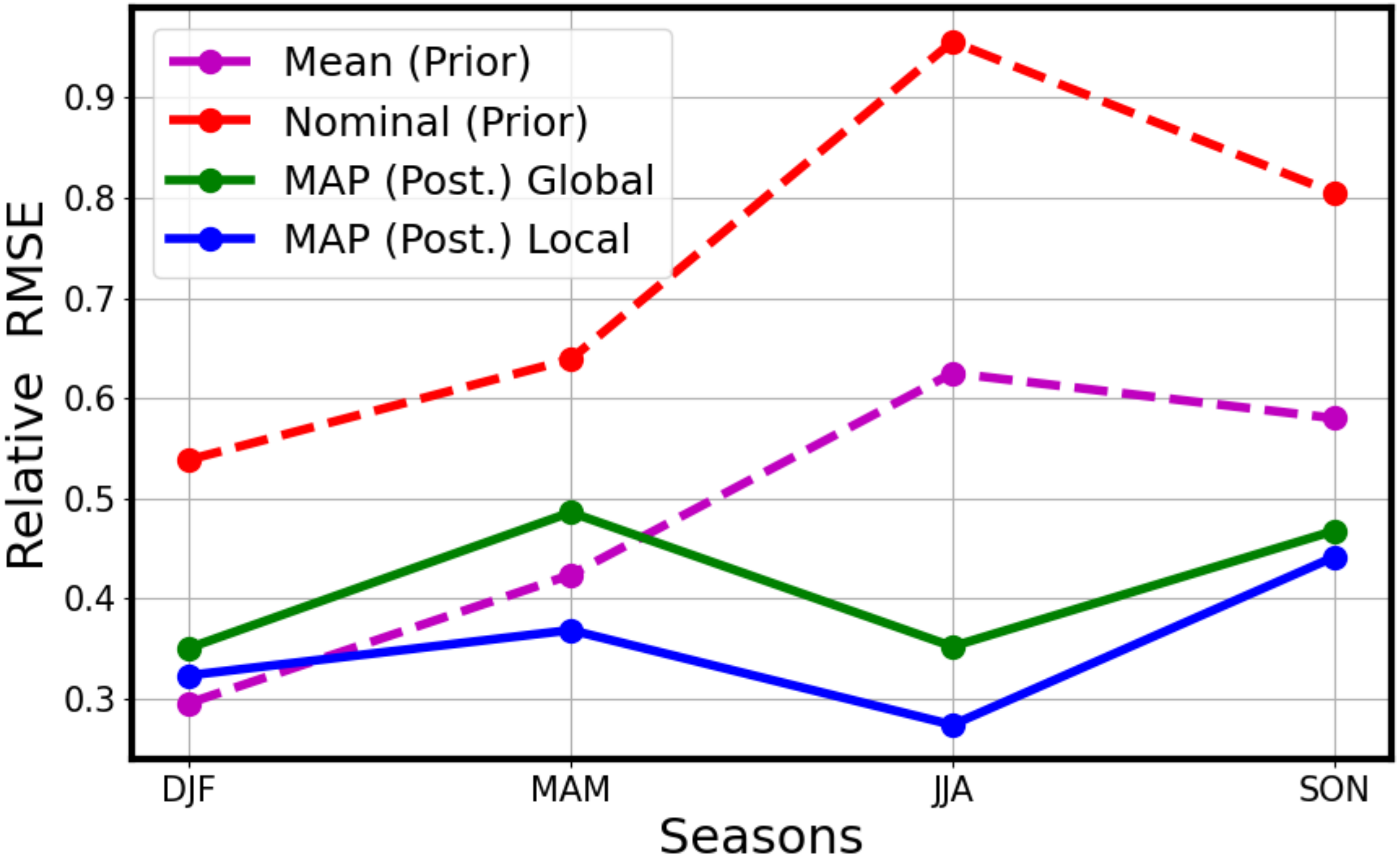


One surrogate per grid cell

Local fLNR parameter



Localized calibration works slightly better



Summary

- Karhunen-Loève (KL) decomposition reduces the spatio-temporal output dimensionality, taking advantage of correlations over space and time.

- Neural network (NN) surrogate in the reduced eigenspace leads to a spatio-temporal KLNN surrogate that is a small fraction of ELM cost.

- KLNN surrogate enables sampling based global sensitivity analysis and Bayesian calibration performed in the eigenspace.

Ongoing work:
- *Potential PFT-dependent reparameterization to improve model's ability to match reference data.*
- *Calibration with embedded model discrepancy to avoid overfitting.*

Additional Material

KL truncation relies on variance retention

$$f(\lambda; z) \approx \bar{f}(z) + \sum_{m=1}^{M} \xi_m(\lambda)\sqrt{\mu_m}\phi_m(z)$$

$$Var[f(z)] = \sum_{m=1}^{M} \mu_m {\phi^2}_m(z)$$

$$Var[f] = \sum_{m=1}^{M} \mu_m$$

$$M = \text{argmin}_{M'} \frac{\sum_{m=1}^{M'} \mu_m}{\sum_{m=1}^{\infty} \mu_m} > 0.99$$

Polynomial Chaos intro

- Our traditional tool for uncertainty representation and propagation
- Random variables represented as polynomial expansion of standard random variables, such as gaussian or uniform

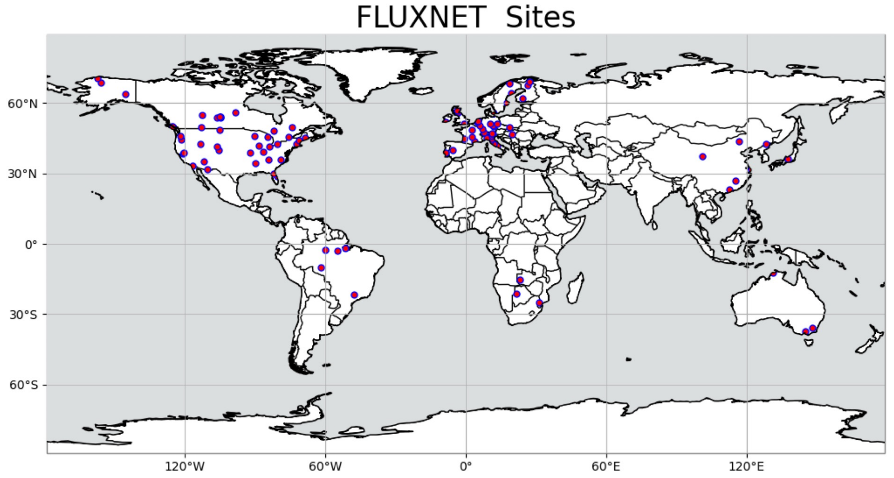
$$\xi = \sum_{k=1}^{K} c_k \, \psi_k(\eta)$$

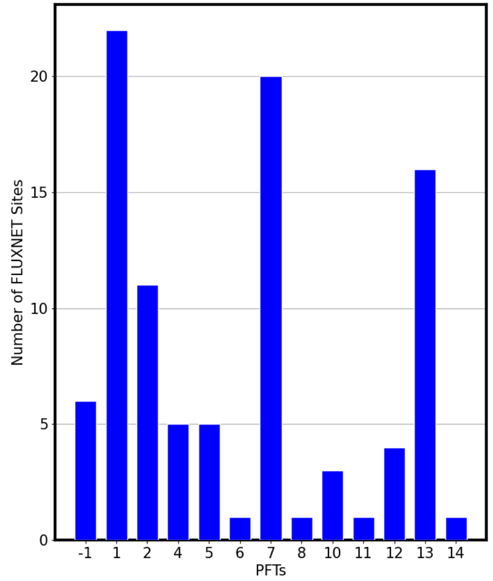
- Convenient for uncertainty propagation

$$f(\xi) = \sum_{k=0}^{K} f_k \, \psi_k(\eta)$$

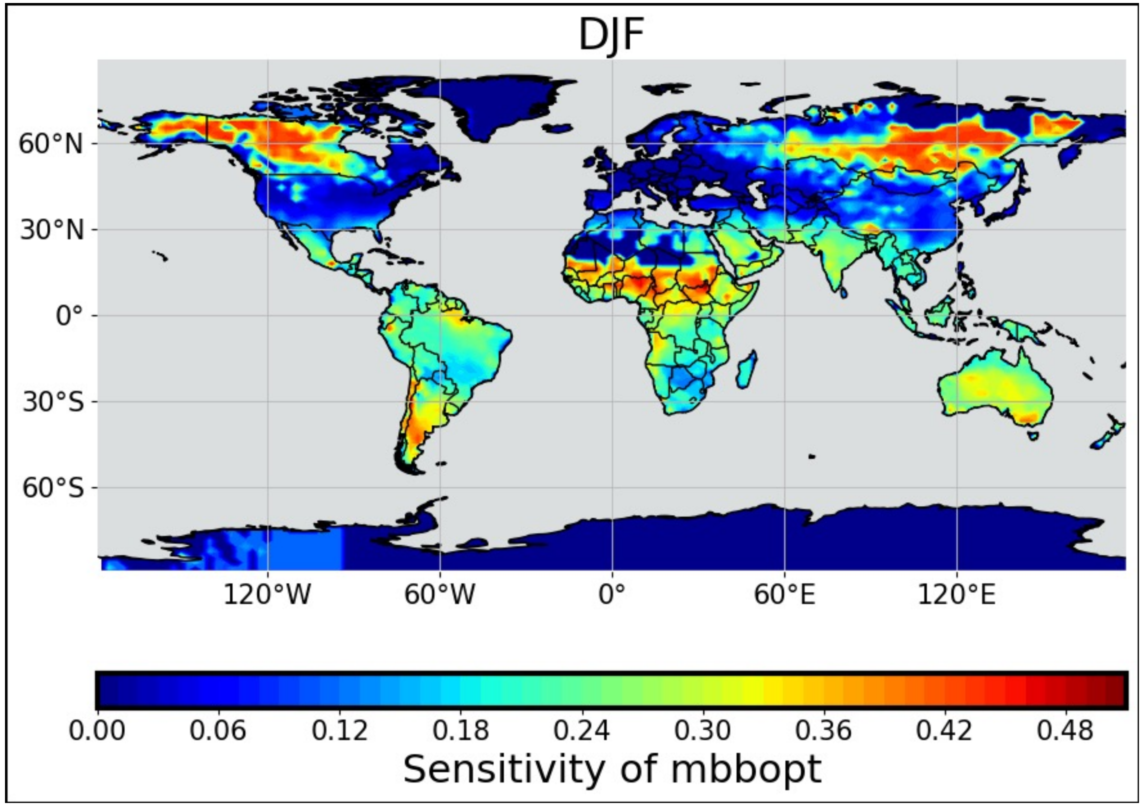
- Moment estimation
- Global Sensitivity Analysis (a.k.a. Sobol indices or variance-based decomposition)

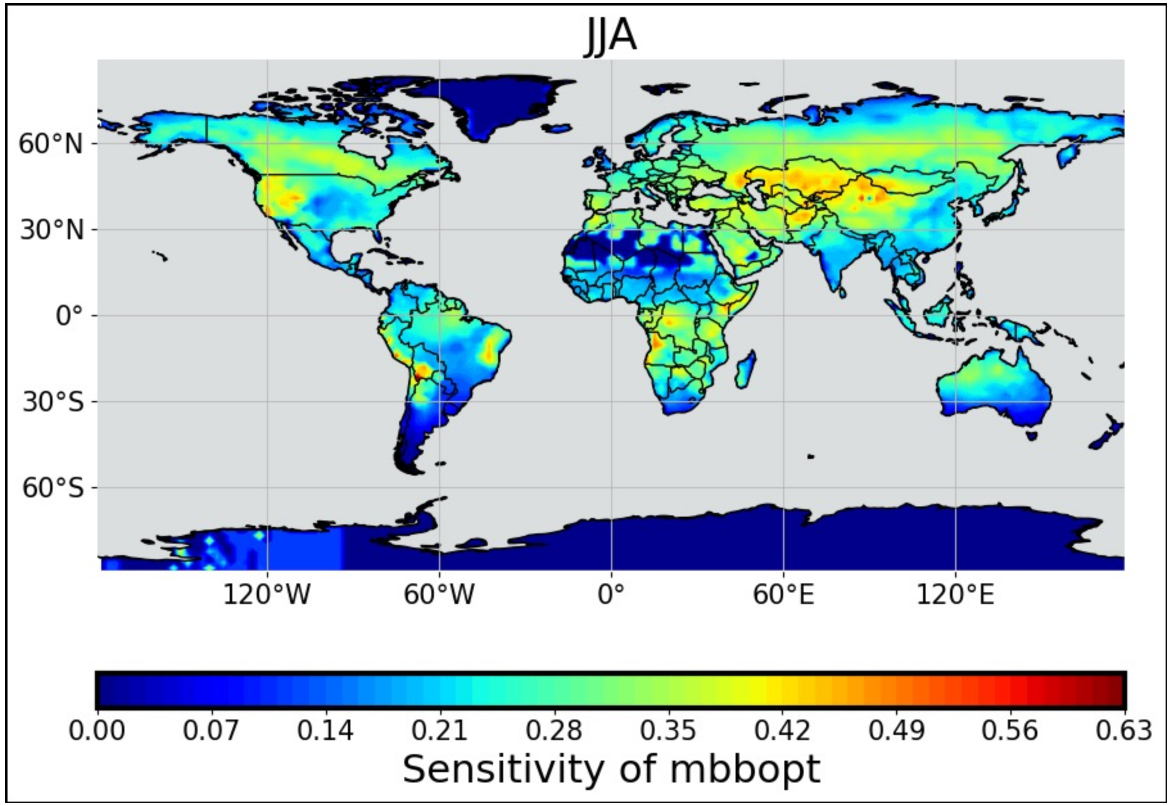
ID	PFT Name	Count
1	Boreal evergreen needleleaf tree	22
2	Temperate evergreen needleleaf tree	11
3	Boreal deciduous needleleaf tree	0
4	Tropical evergreen broadleaf tree	5
5	Temperate evergreen broadleaf tree	5
6	Tropical deciduous broadleaf tree	1
7	Temperate deciduous broadleaf tree	20
8	Boreal deciduous broadleaf tree	1
9	Broadleaf evergreen shrub	0
10	Temperate deciduous broadleaf shrub	3
11	Boreal deciduous broadleaf shrub	1
12	C3 arctic grass	4
13	C3 non-arctic grass	16
14	C4 grass	1
-1	Mixed	6





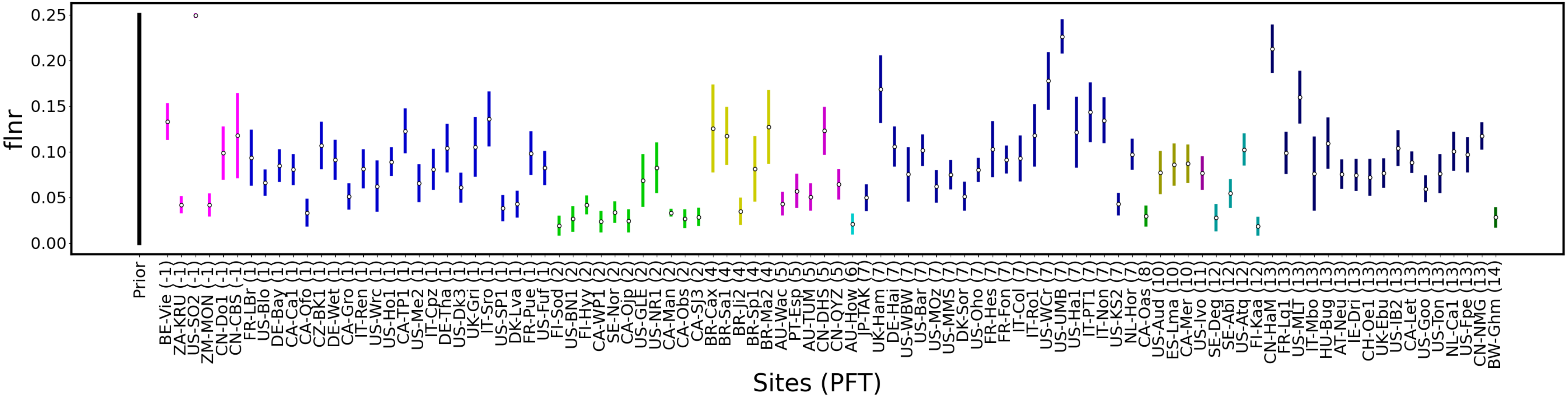
mbbopt sensitivity across the globe





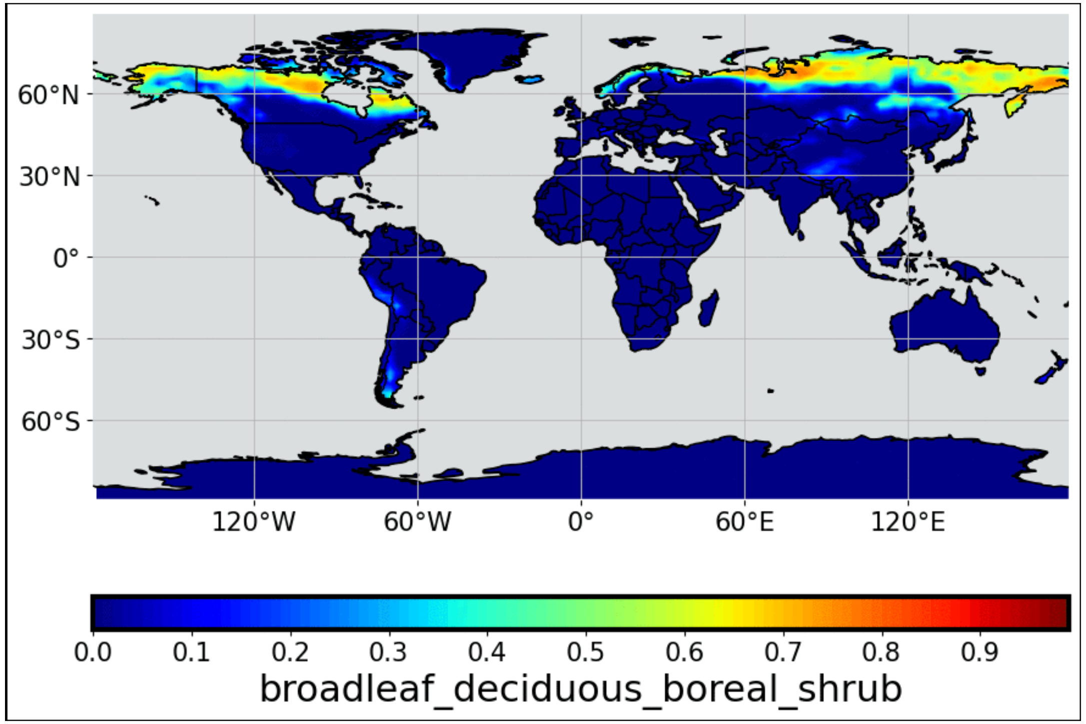
Local (site-specific) fLNR posterior PDFs

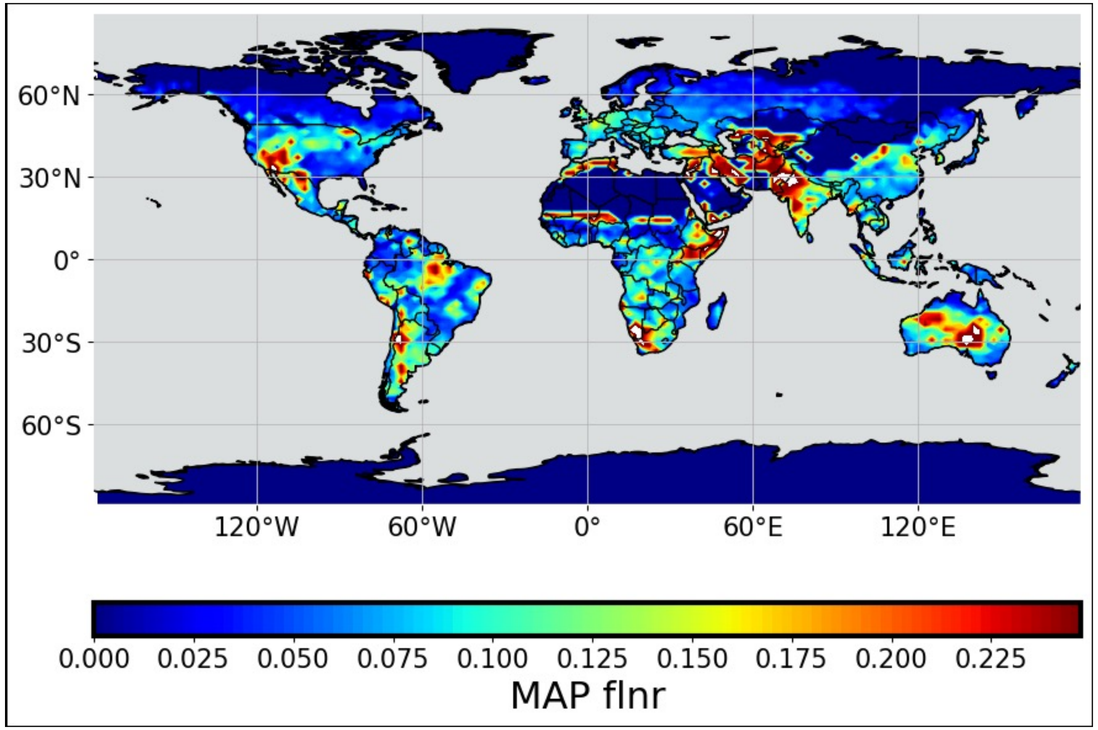
Grouped by PFTs



Correlate PFT fractions globally with best fLNR values

PFT Fractions for all PFTs





Correlate PFT fractions globally with best fLNR values