On the Use of ML for Modelling Land Surface Dynamics

Nuno Carvalhais

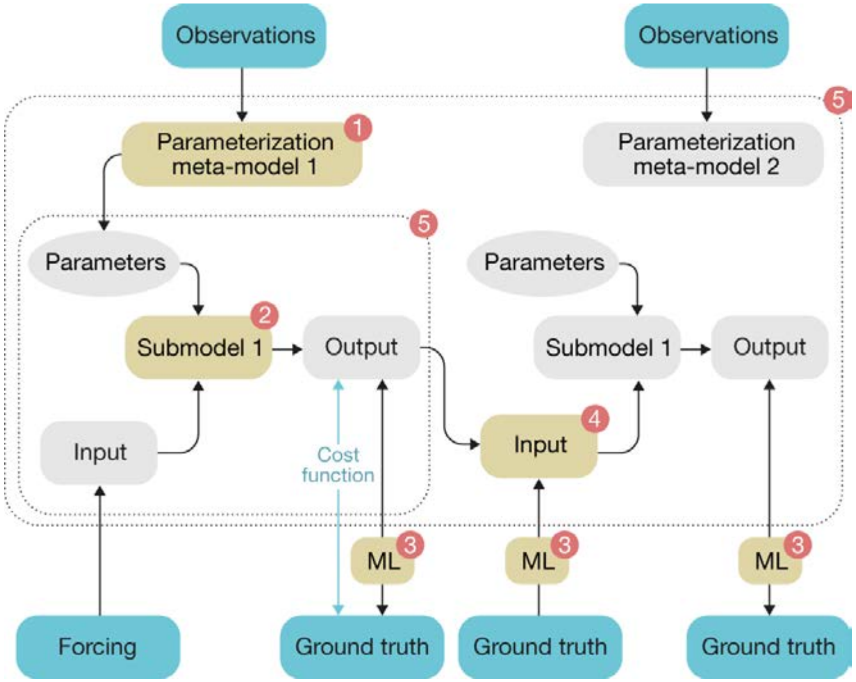
Shanning Bao, Rackhun Son, Ranit De, Reda El Ghawi, Christian Reimers,
Lazaro Alonso, Sujan Koirala, Markus Reichstein, ...

[ncarvalhais@bgc-jena.mpg.de]

Max Planck Institute
for Biogeochemistry

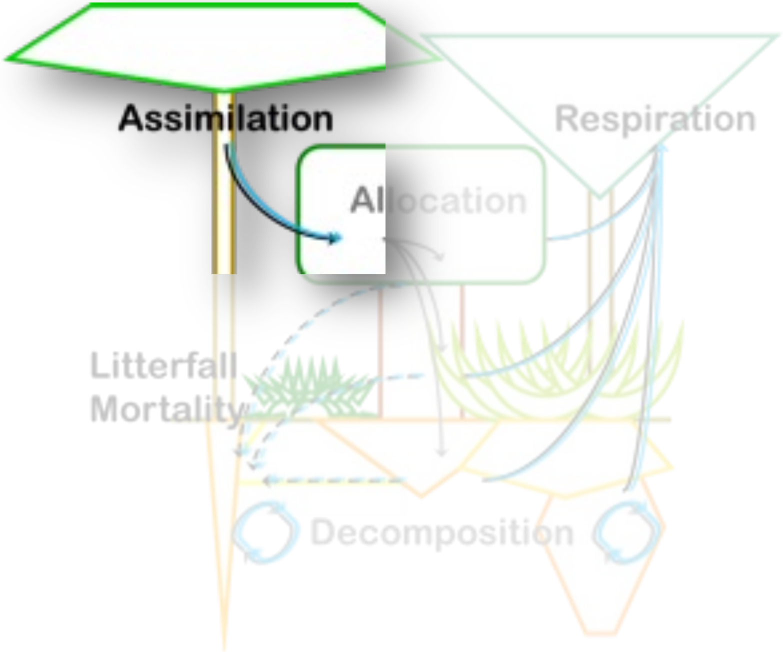
ellis
European Laboratory for Learning and Intelligent Systems

Perspective...



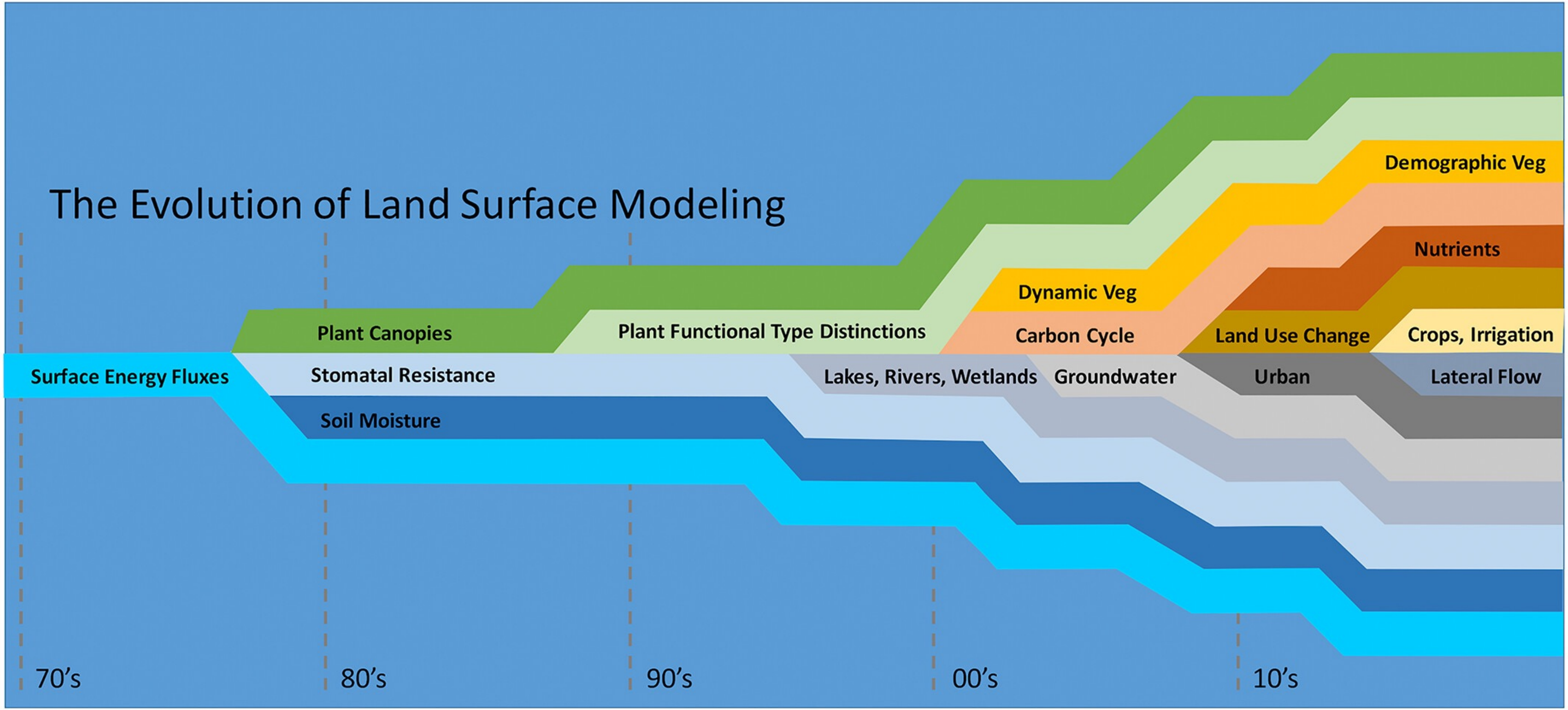
1. Improving parameterizations
2. Replacing a ‘physical’ sub-model with a machine learning model
3. Analysis of model–observation mismatch
4. Constraining submodels
5. Surrogate modelling or emulation

[Reichstein et al., 2019]

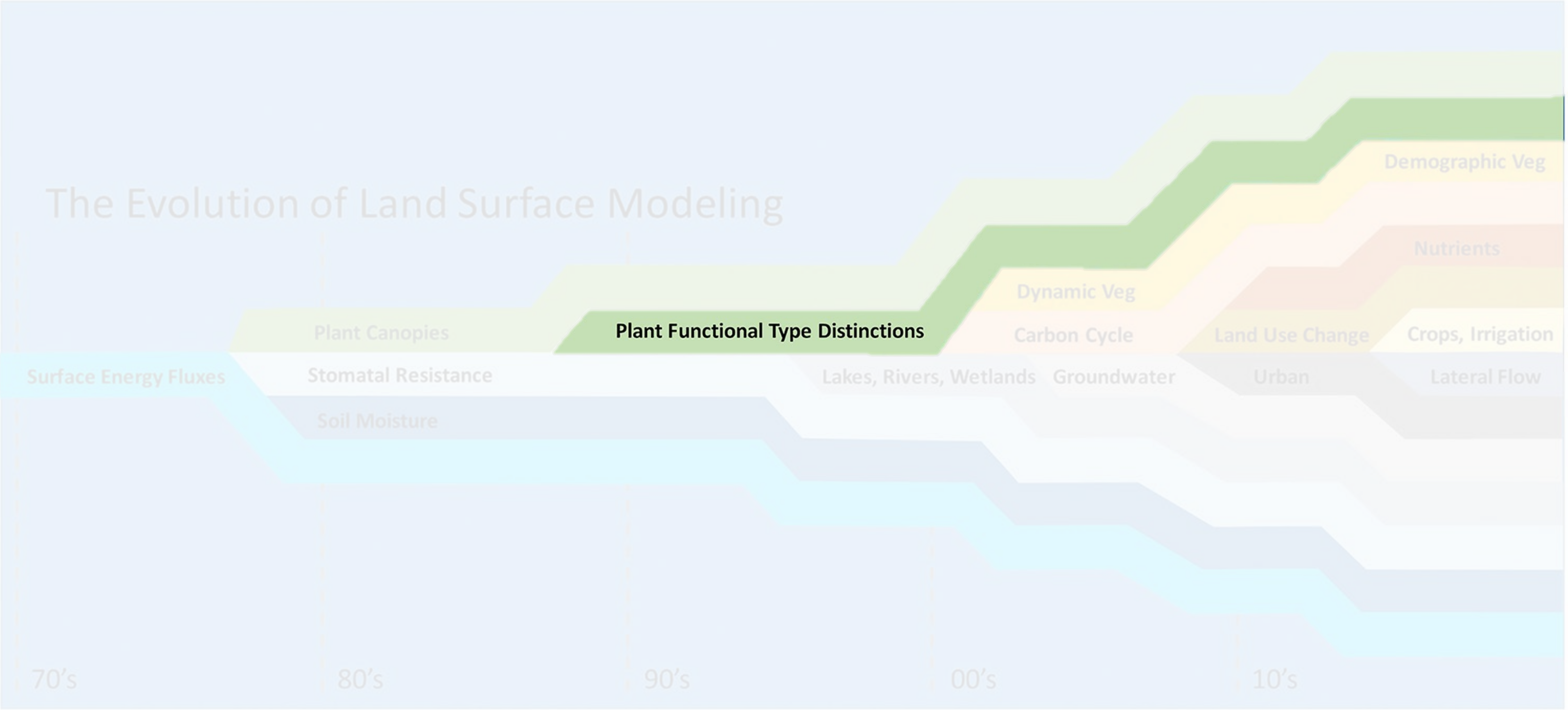


[Bao et al., JAMES, 2024]

PARAMETERIZATIONS

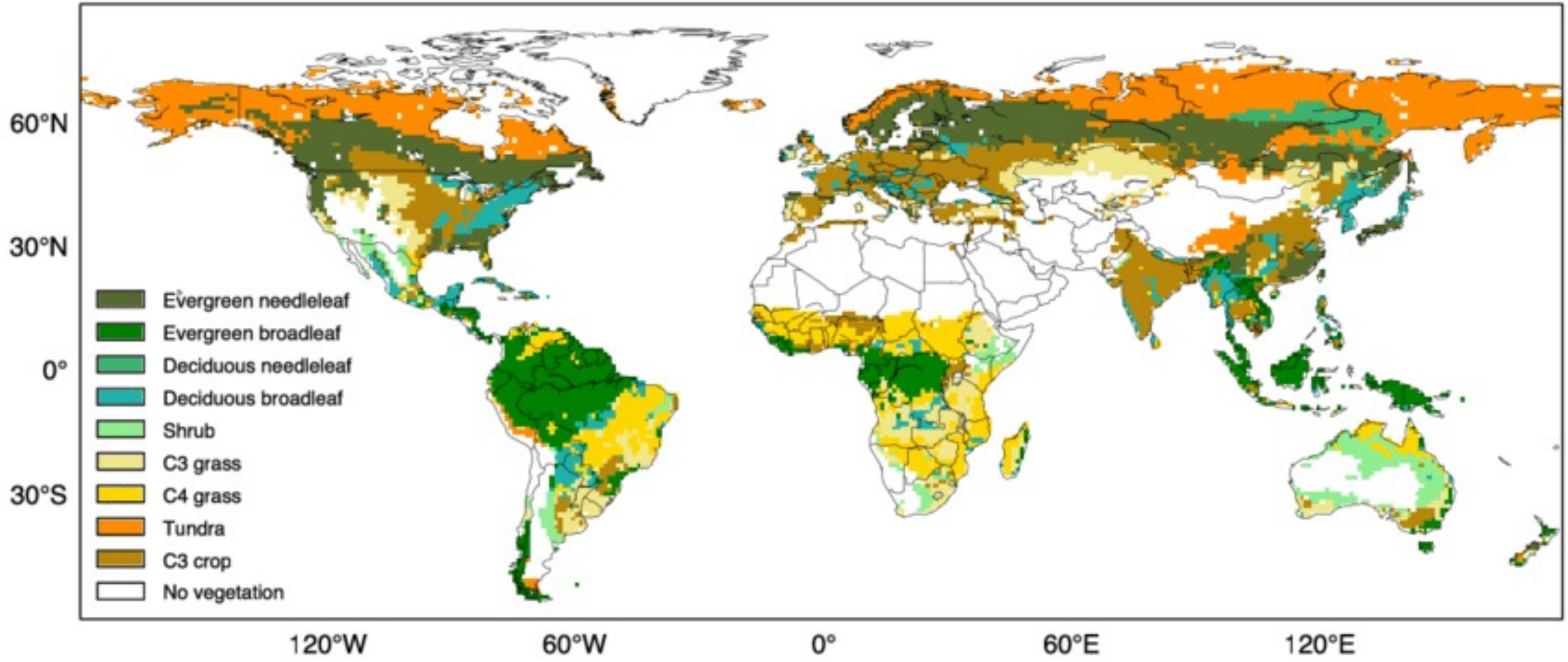


[Fisher and Koven, 2020]



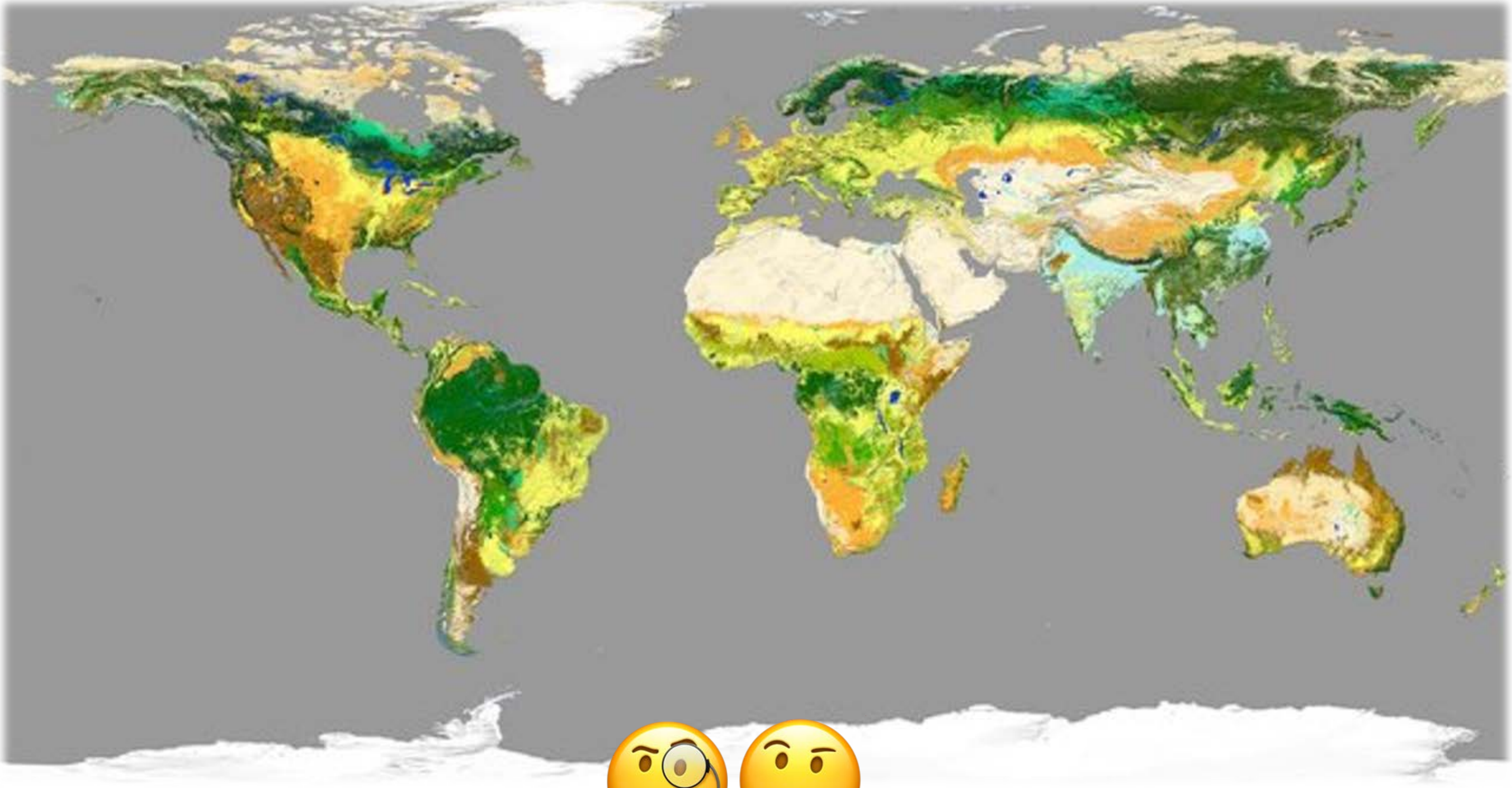
[Fisher and Koven, 2020]

$$\frac{\partial C}{\partial t} = f(X, \theta)$$



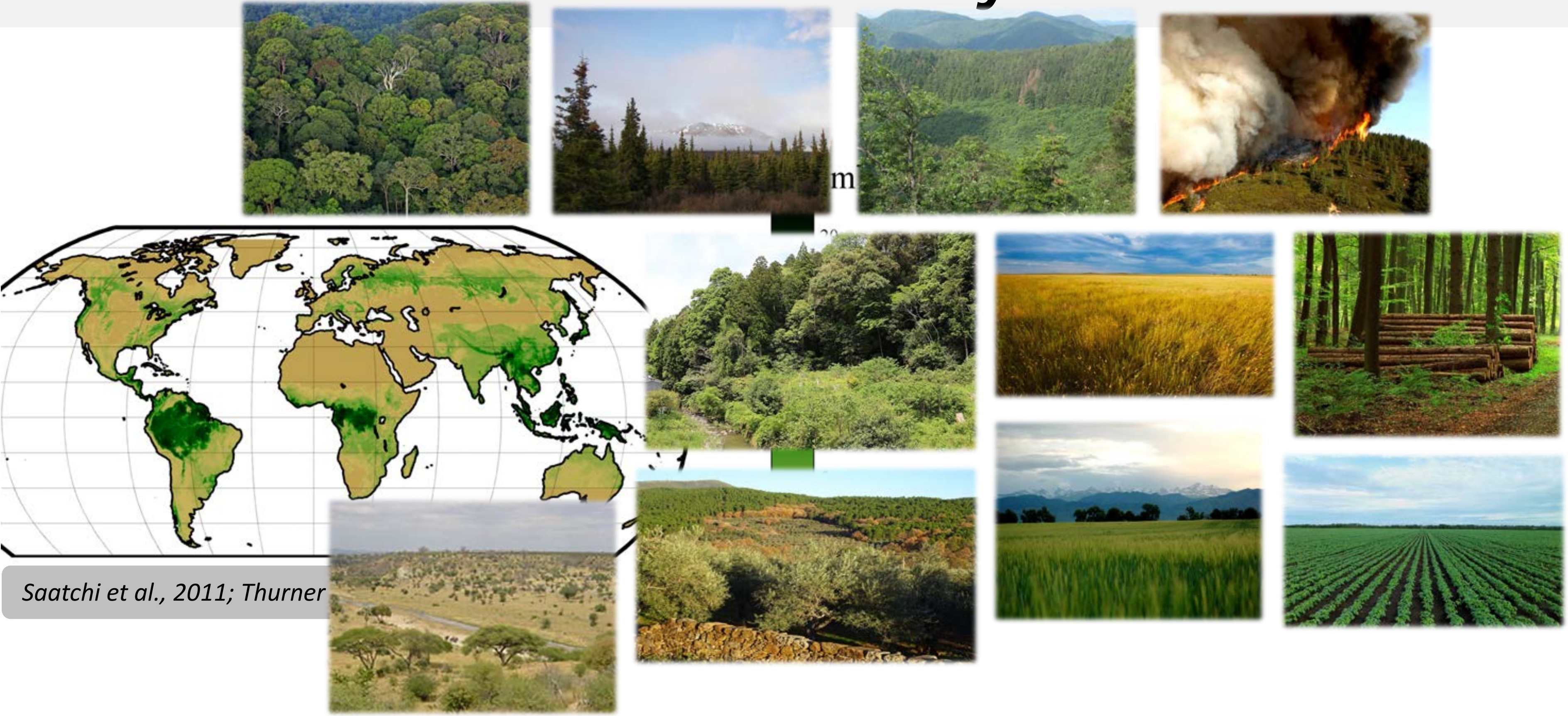
[e.g. De Kauwe et al., 2015]

$$\frac{\partial C}{\partial t} = f(X, \theta)$$

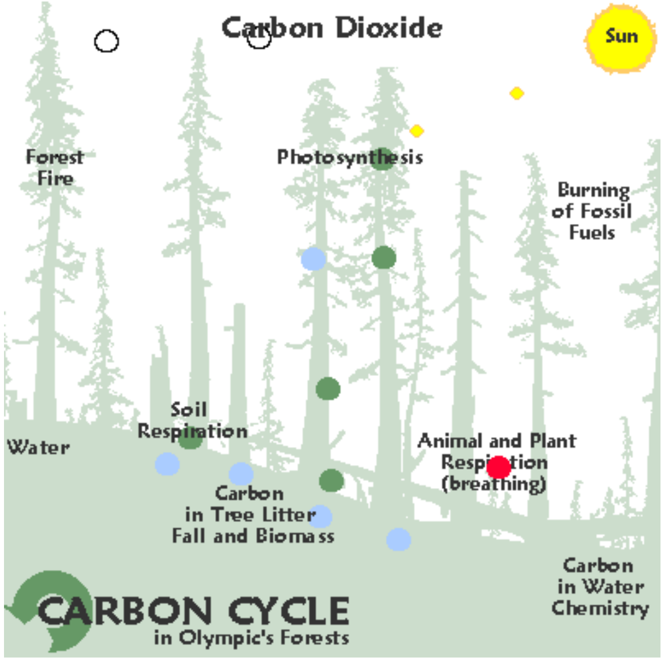


[ESA]

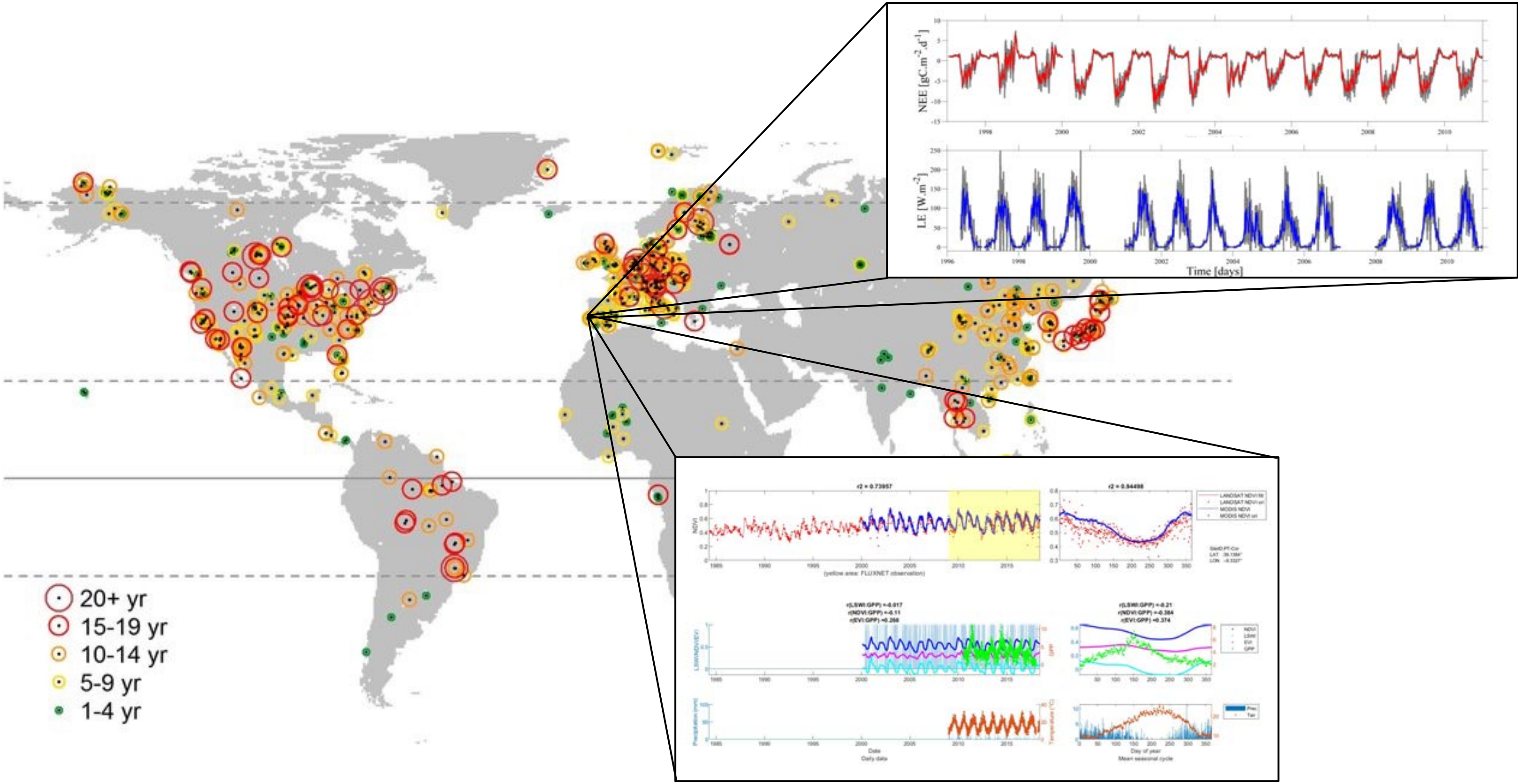
Land carbon cycle



Saatchi et al., 2011; Thurner



Carbon Dioxide
Sun
Forest Fire
Photosynthesis
Burning of Fossil Fuels
Soil Respiration
Water
Animal and Plant Respiration (breathing)
Carbon in Tree Litter Fall and Biomass
Carbon in Water Chemistry
CARBON CYCLE
in Olympic's Forests



[Chu et al., 2017]

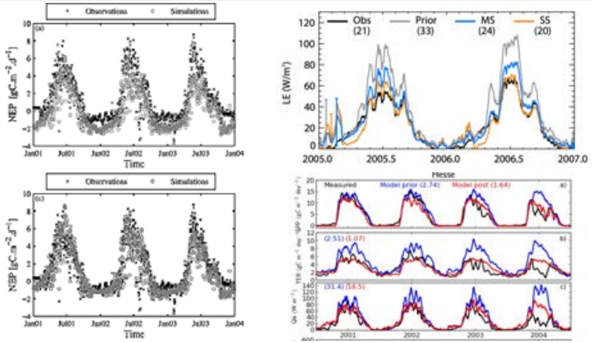
Modelling terrestrial ecosystem fluxes

In situ

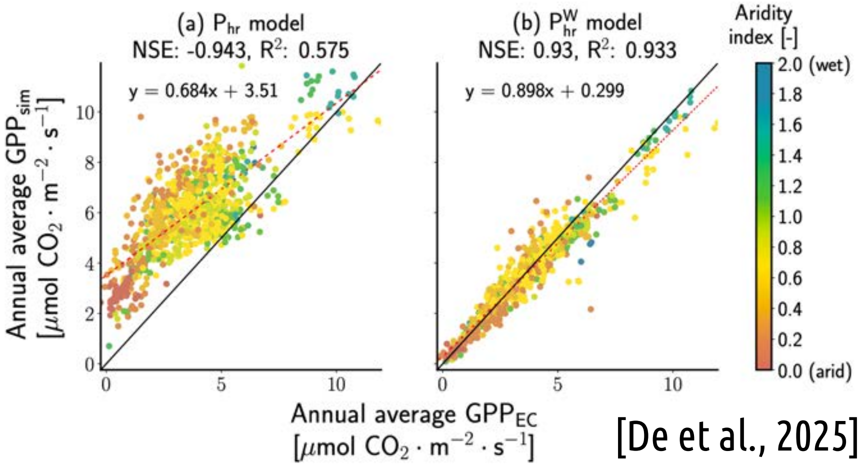
Parameter inversion

$$\hat{\theta}_i = \arg\min \Omega_\mathrm{i}(y_i, X_i, \theta)\,;\theta \in \mathcal{D}$$

$$\Omega_\mathrm{i} = \sum_{v=1}^{N}\sum_{t=1}^{T}\left[\frac{y_{v,i,t} - M_v(X_{i,t}, \theta)}{\sigma_{v,i,t}}\right]^2$$

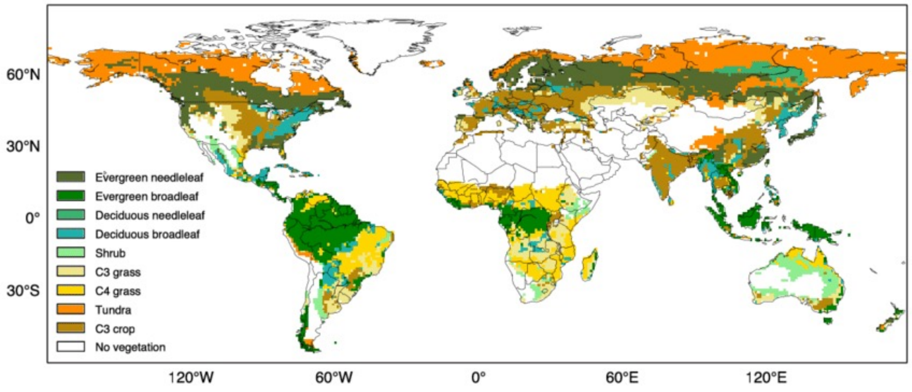


[Thum et al., 2017; Kuppel et al., 2015; Carvalhais et al., 2010]

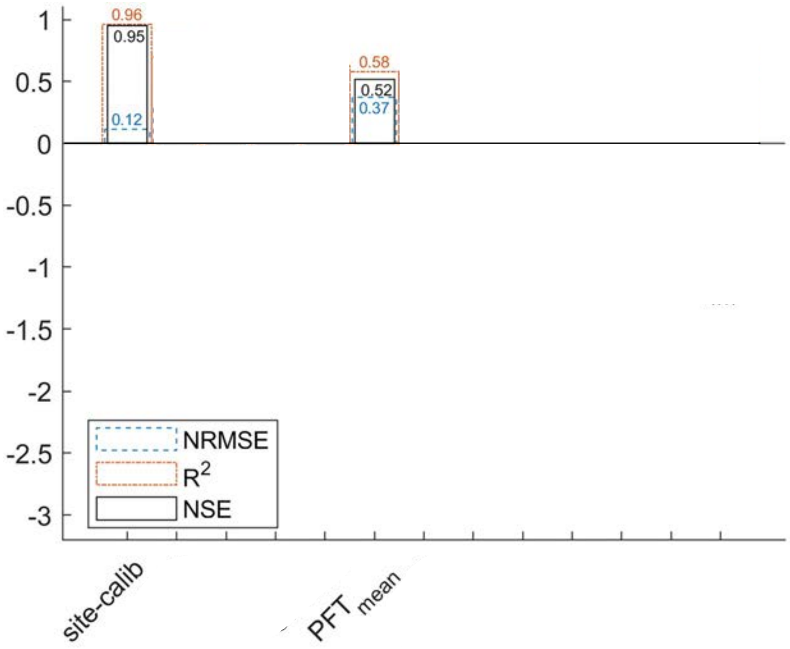


[De et al., 2025]

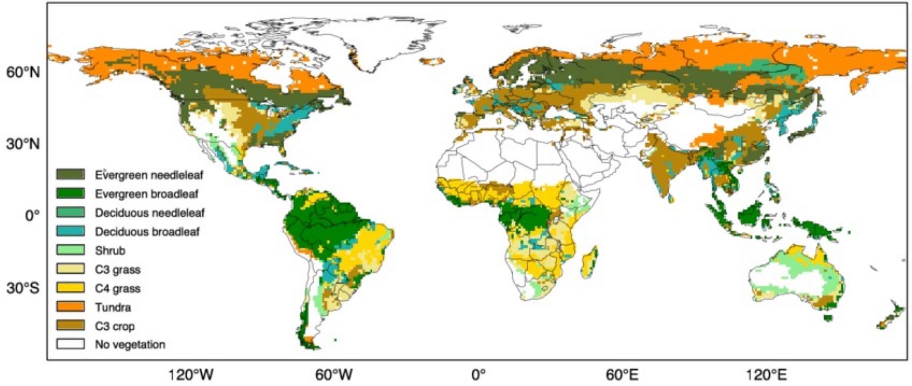
Challenges in generalizing parameterizations



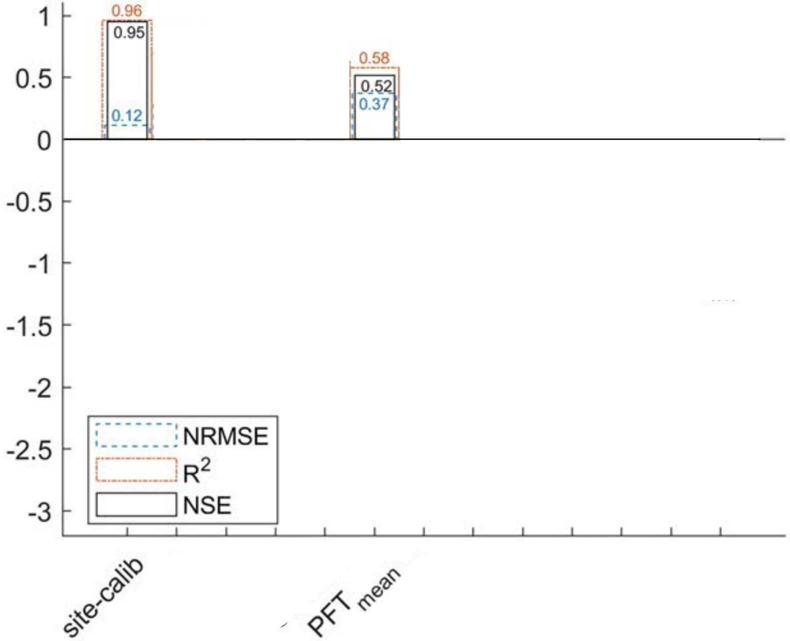
$$\hat{\theta}_{PFT} = \bar{\theta}_{i \in PFT}$$



Testing PFT-based solutions

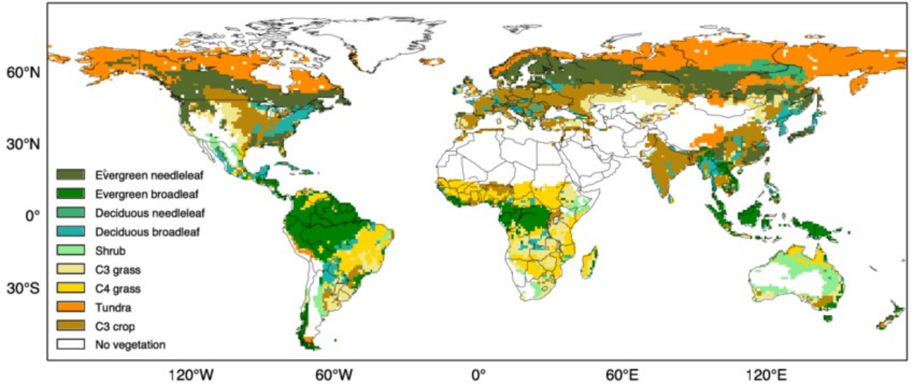


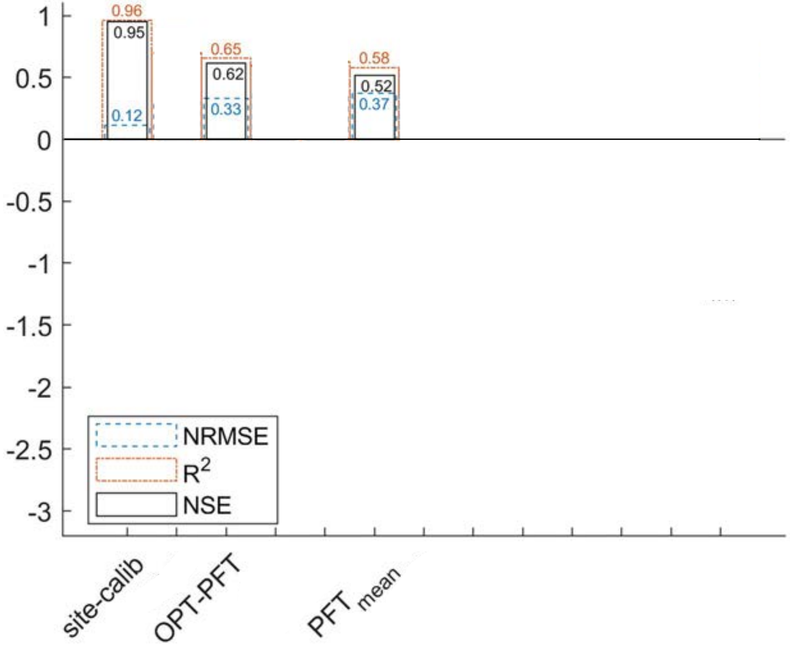
$$\hat{\theta}_{PFT} = \arg\min \sum \Omega_i(y_i, X_i, \theta)\,;\ \theta \in \mathcal{D}; i \in PFT$$



[Kuppel et al., 2012]

Testing PFT-based solutions

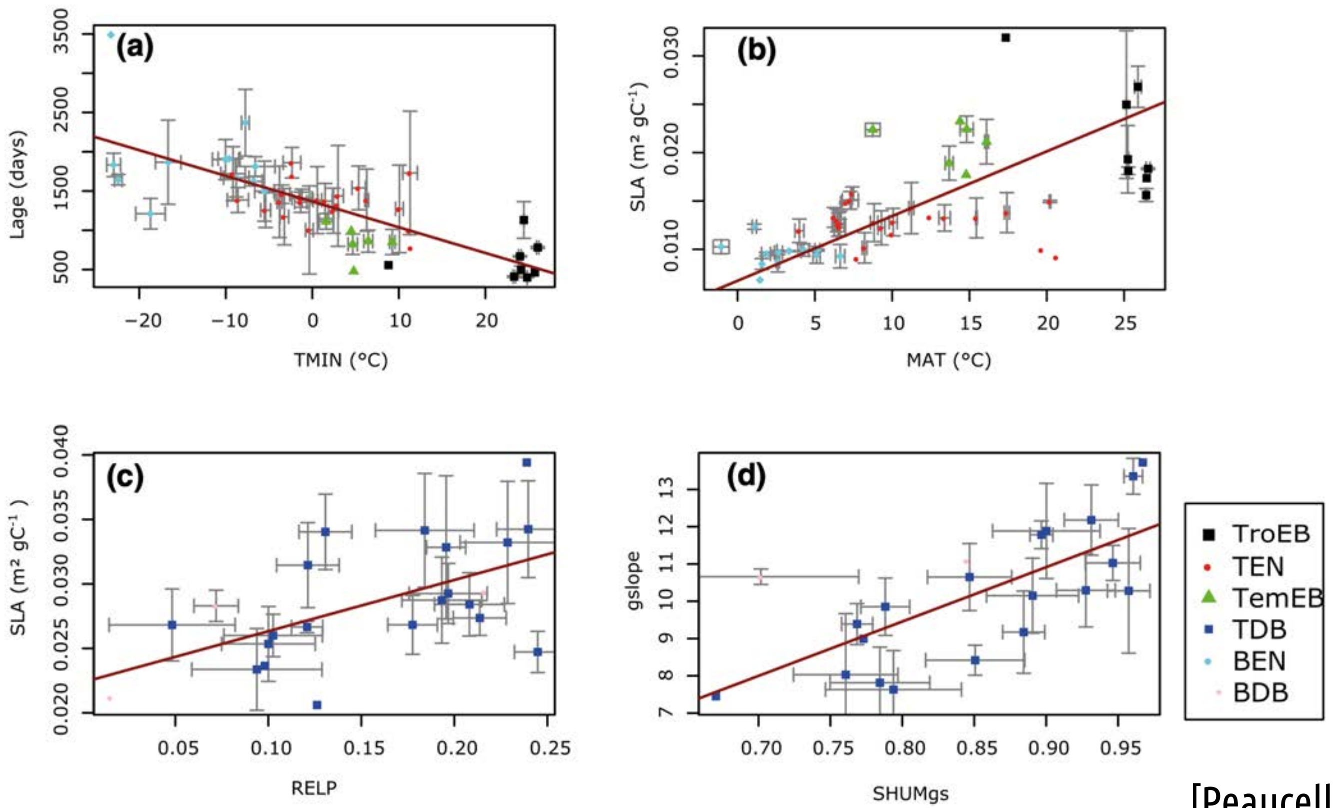




$$\hat{\theta}_{PFT} = \arg\min \sum \Omega_{\mathrm{i}}(y_i, X_i, \theta)\,;\ \theta \in \mathcal{D}; i \in PFT$$

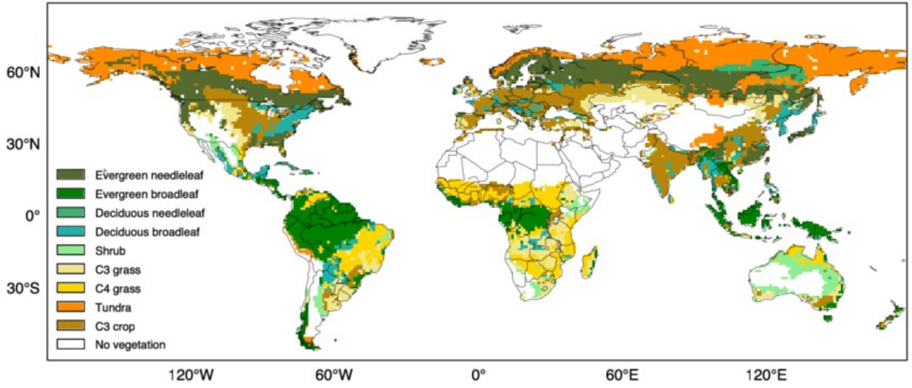
[Kuppel et al., 2012]

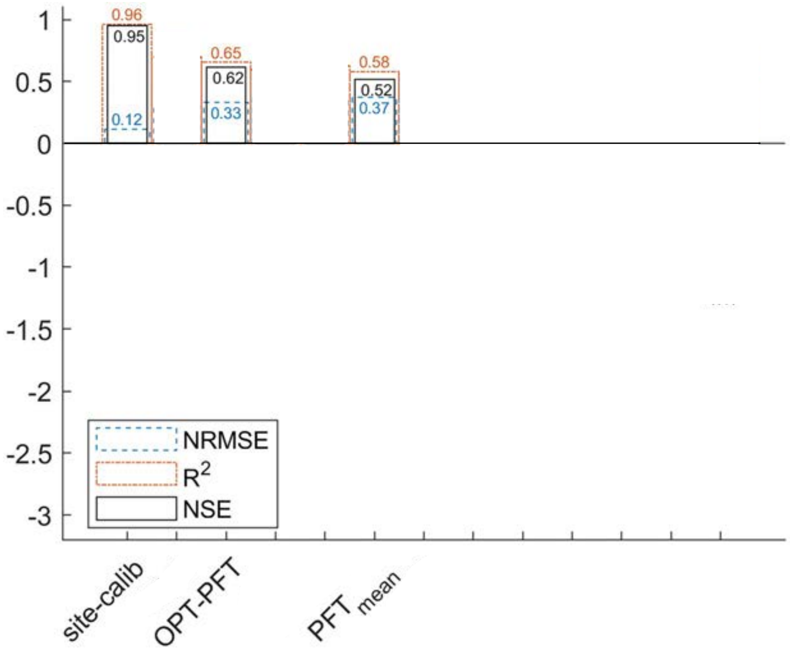
Towards continuous parameterizations



[Peaucelle et al., 2012]

Learning inverted parameters

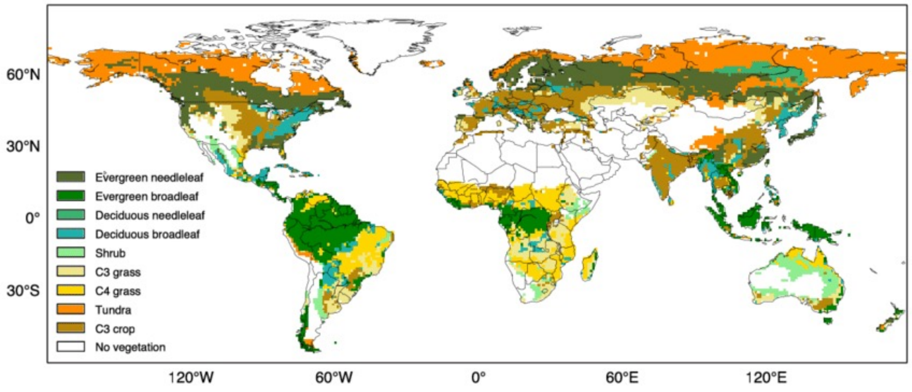


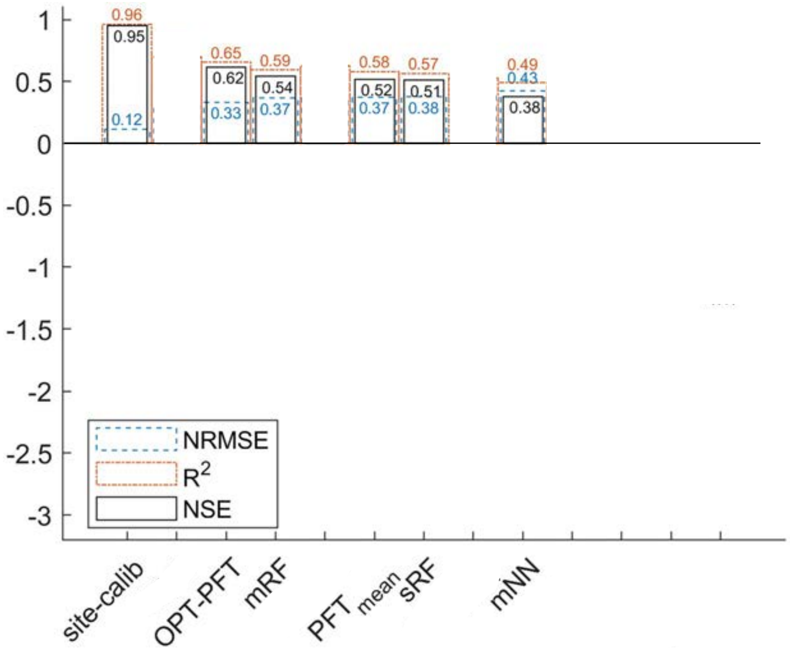


1. $\hat{\theta}_i = \arg\min \Omega_i(y_i, X_i, \theta)\ ;\ \theta \in \mathcal{D}$
2. $\hat{\theta} = h(\bar{X}_i, PFT, \ldots)$

[Bao et al., 2024]

Learning inverted parameters





1. $\hat{\theta}_i = \arg\min \Omega_i(y_i, X_i, \theta)\,;\ \theta \in \mathcal{D}$
2. $\hat{\theta} = h(\bar{X}_i, PFT, \ldots)$

[Bao et al., 2024]

Learning parameter models end-to-end

- Learn g

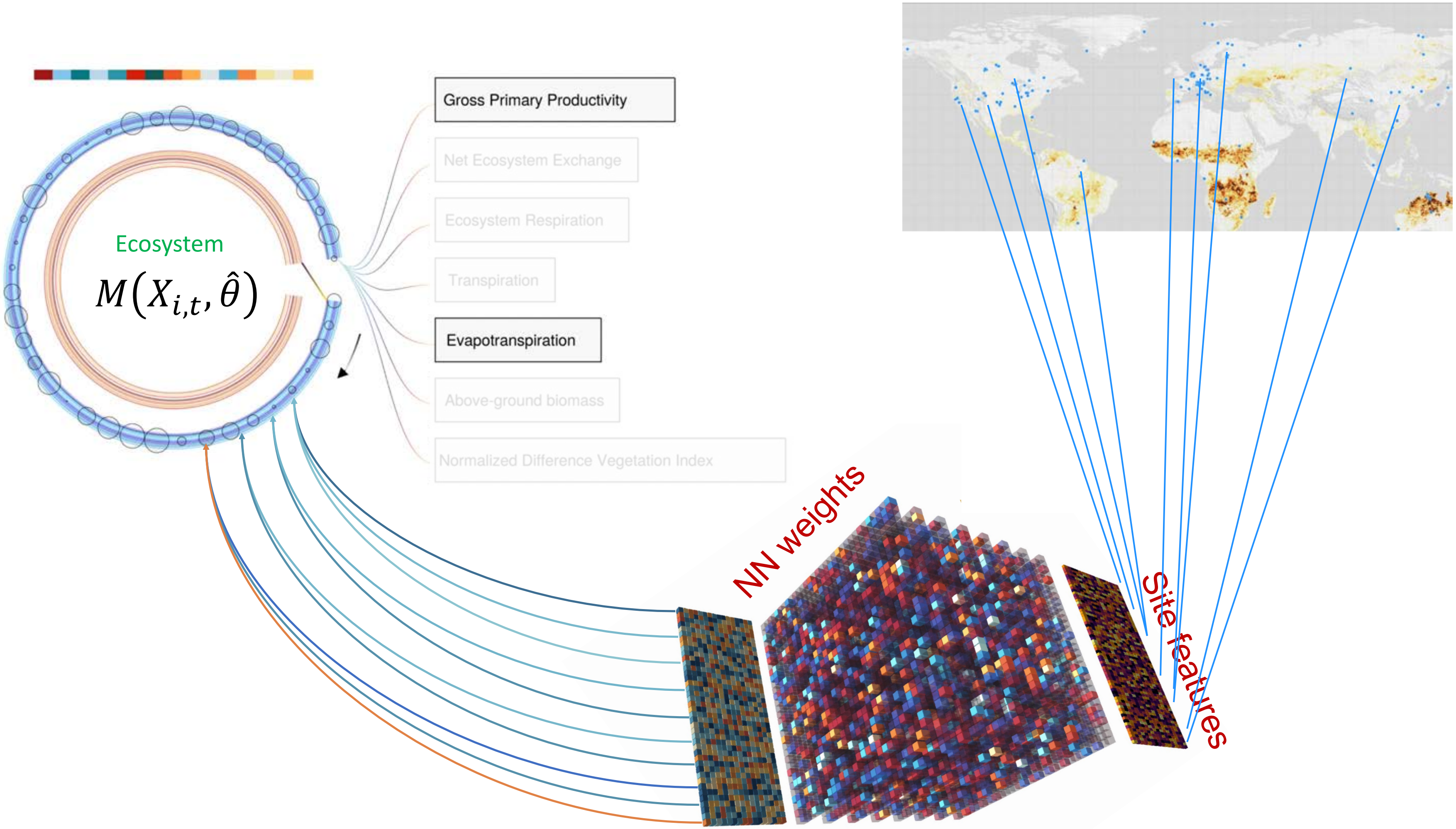
$$\hat{\theta} = g(\bar{X}_i, PFT, \dots)$$

- Such that

$$\hat{y}_{i,t} = M\left(X_{i,t}, \hat{\theta}\right)$$

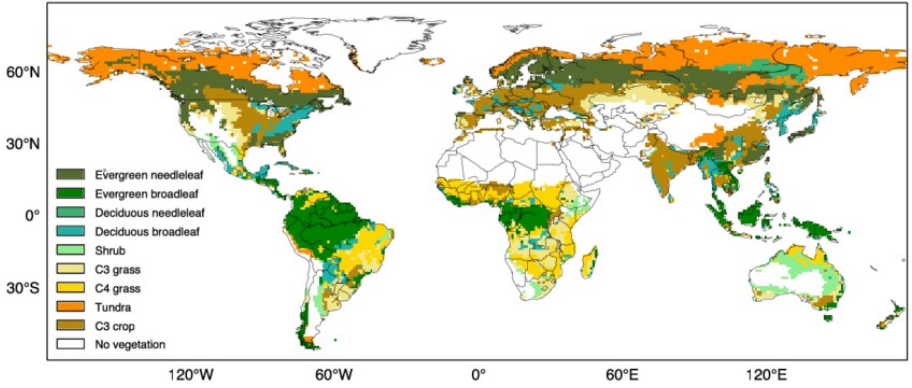
- Minimizing

$$\Omega_{\mathrm{i}} = \sum_{v=1}^{N} \sum_{t=1}^{T} \left[\frac{y_{v,i,t} - M_v\left(X_{i,t}, \theta\right)}{\sigma_{v,i,t}}\right]^2$$

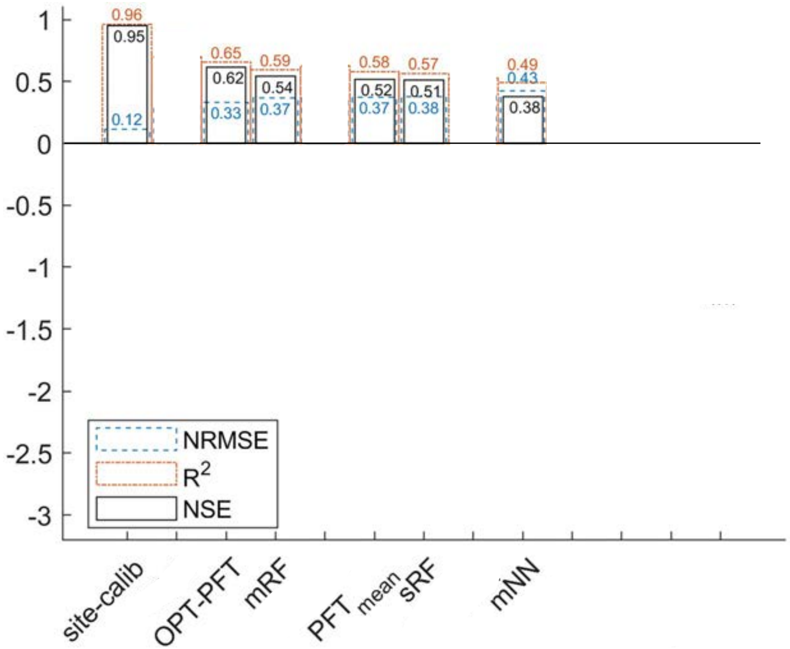


[Bao et al., JAMES, 2024; Alonso et al., in prep.]

Learning inverted parameters

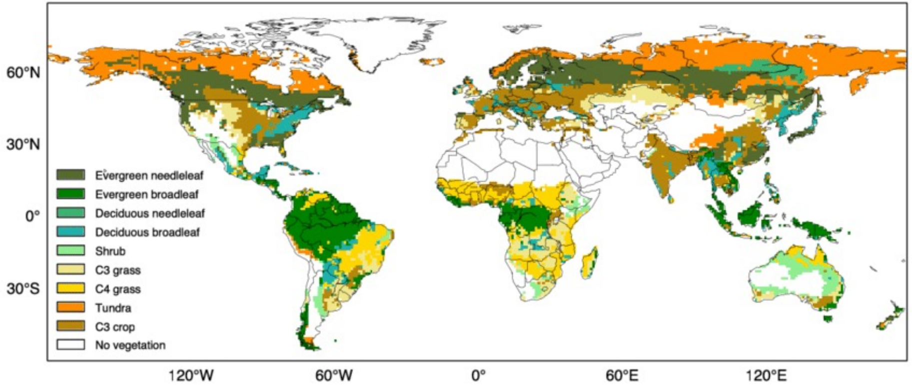


1. $\hat{\theta}_i = \arg\min \Omega_i(y_i, X_i, \theta)\,;\ \theta \in \mathcal{D}$
2. $\hat{\theta} = h(\bar{X}_i, PFT, \ldots)$

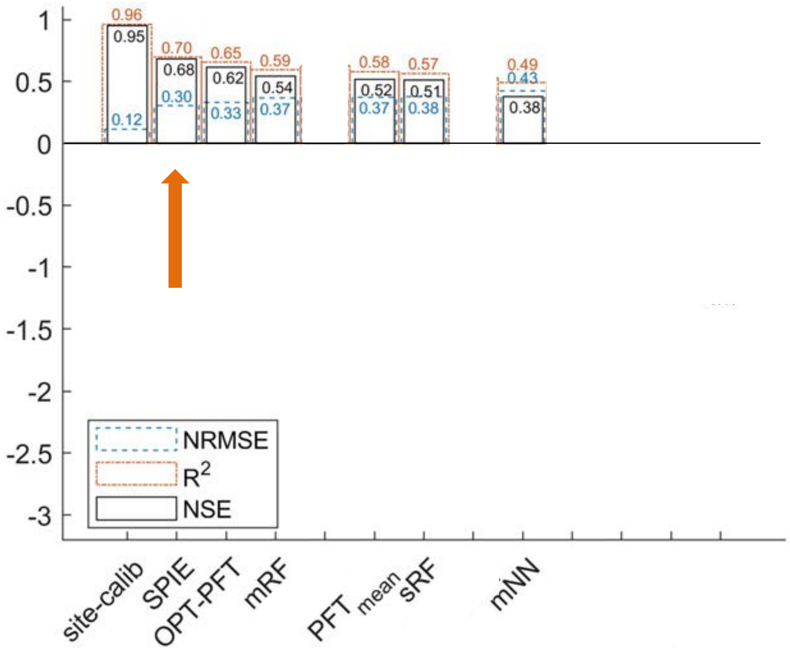


[Bao et al., 2024]

Learning inverted parameters

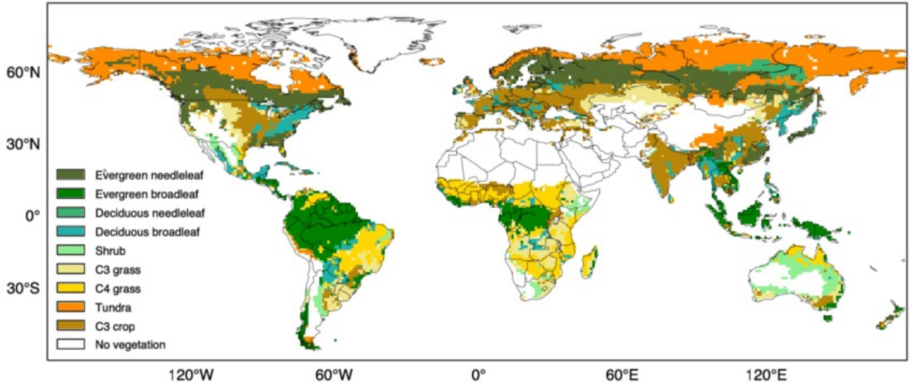


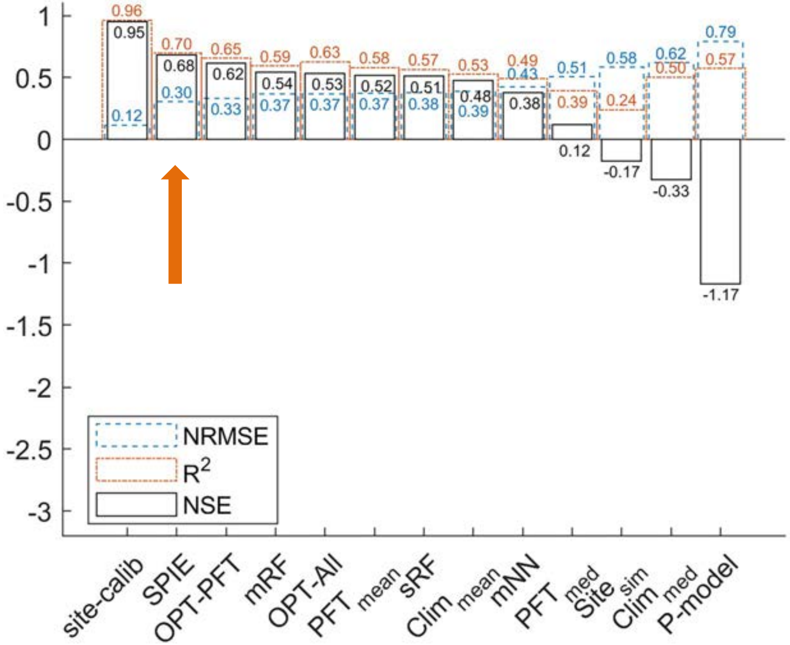
1. $\hat{\theta}_i = \arg\min \Omega_i(y_i, X_i, \theta)\,;\ \theta \in \mathcal{D}$
2. $\hat{\theta} = h(\bar{X}_i, PFT, \ldots)$



[Bao et al., 2024]

Comparing experiments



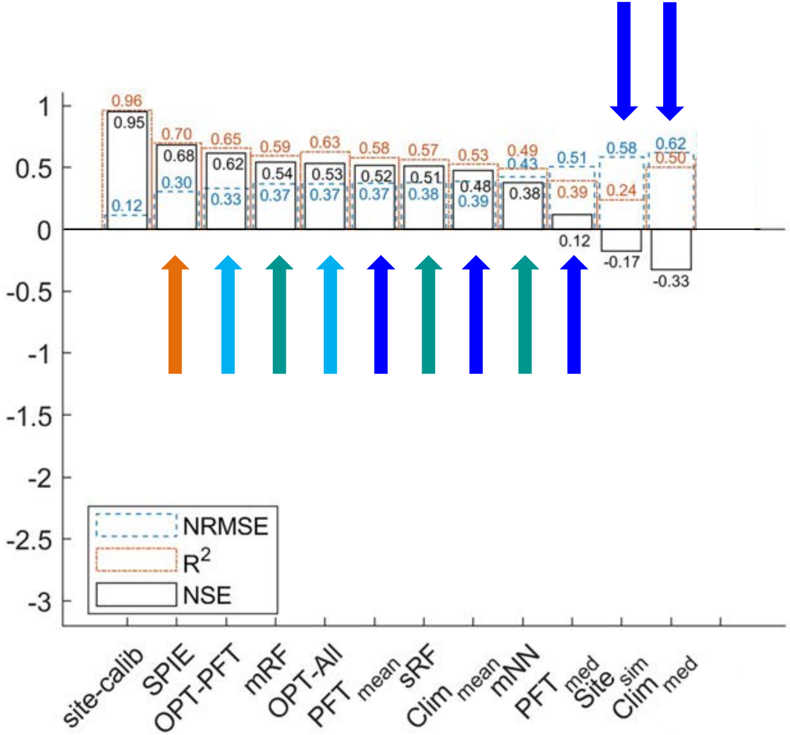


1. $\hat{\theta}_i = \arg\min \Omega_i(y_i, X_i, \theta)\,;\ \theta \in \mathcal{D}$
2. $\hat{\theta} = h(\bar{X}_i, PFT, \dots)$

[Bao et al., 2024]

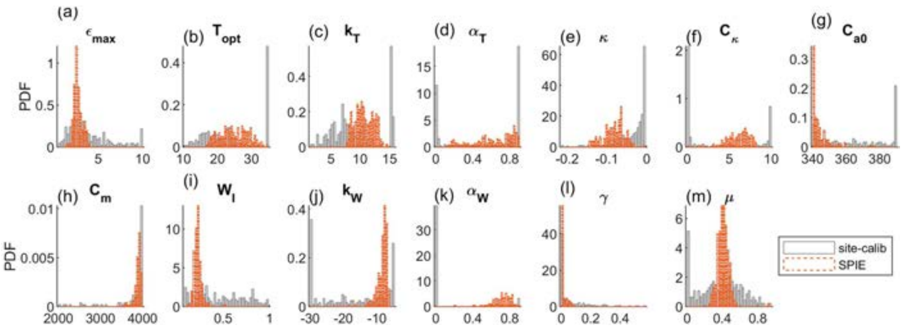
Comparing experiments

- Baseline (site inversion)
- Clustering approaches (PFT, climate, ...)
- PFT-/global-based optim.
- Learning site optimezed parameters
- End-to-end parameter learning

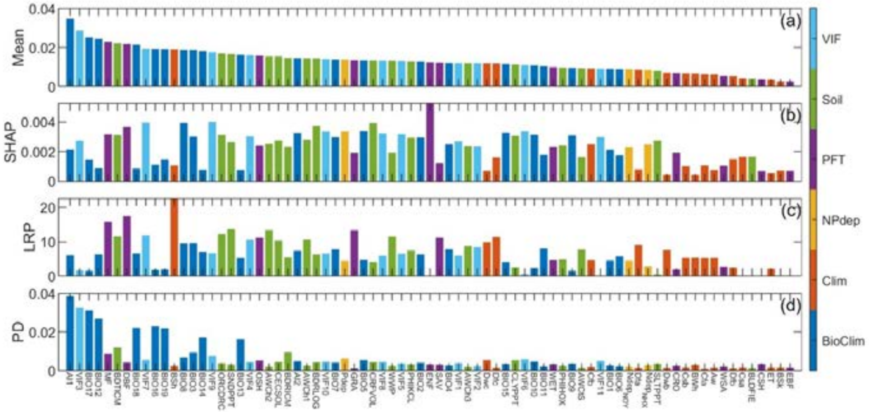


[Bao et al., 2024]

Constraints and relevant features



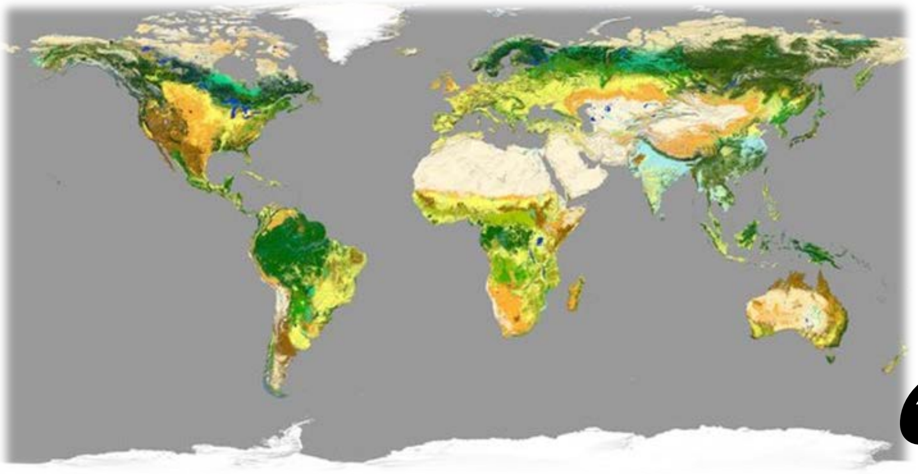
Tighter parameter variability across sites.

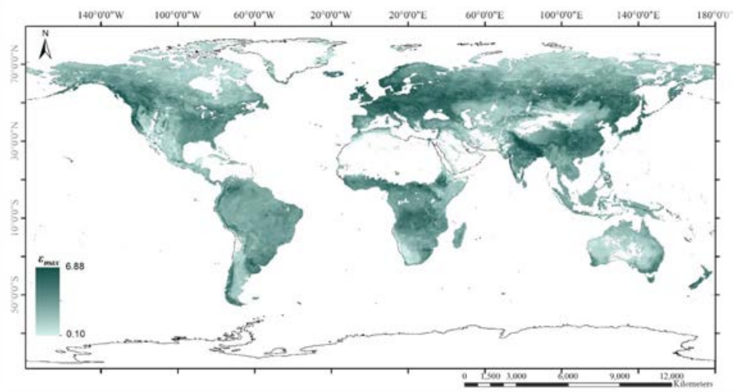


Large number of relevant features

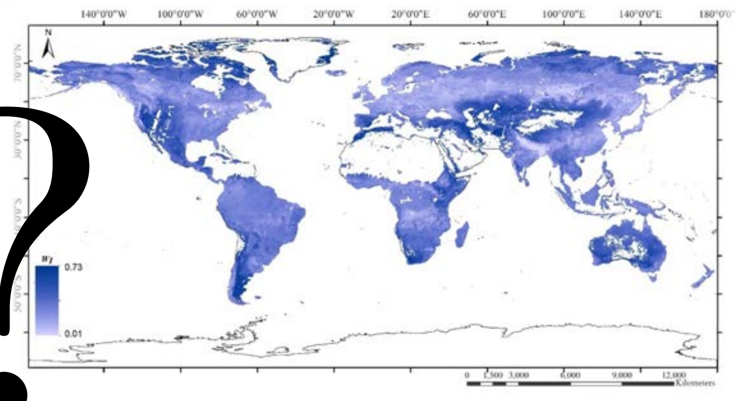
PFTs needed but not dominant

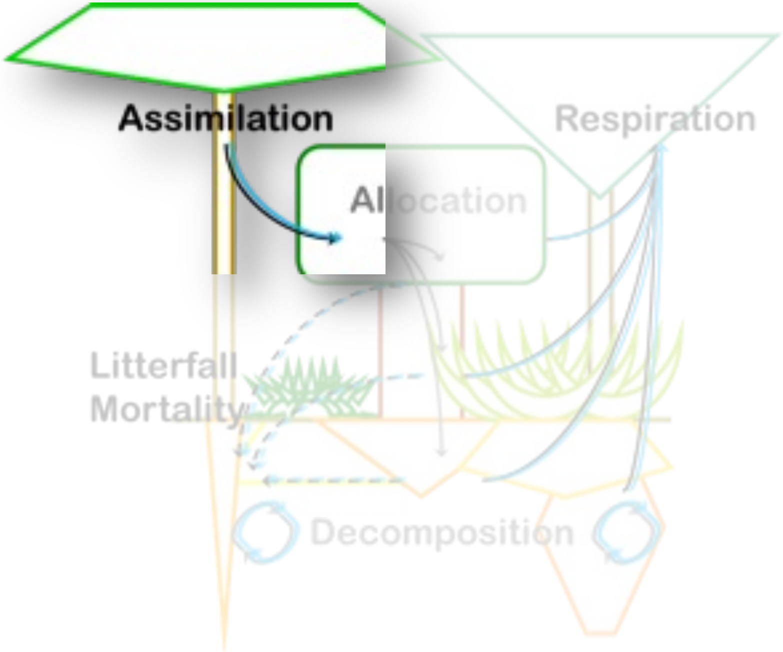
From classes to continuous (natural?) patterns





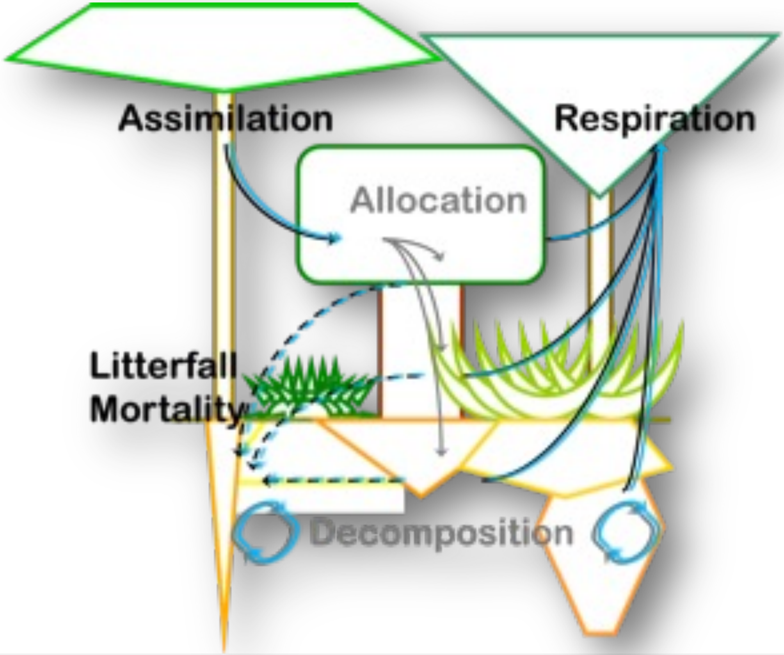
?





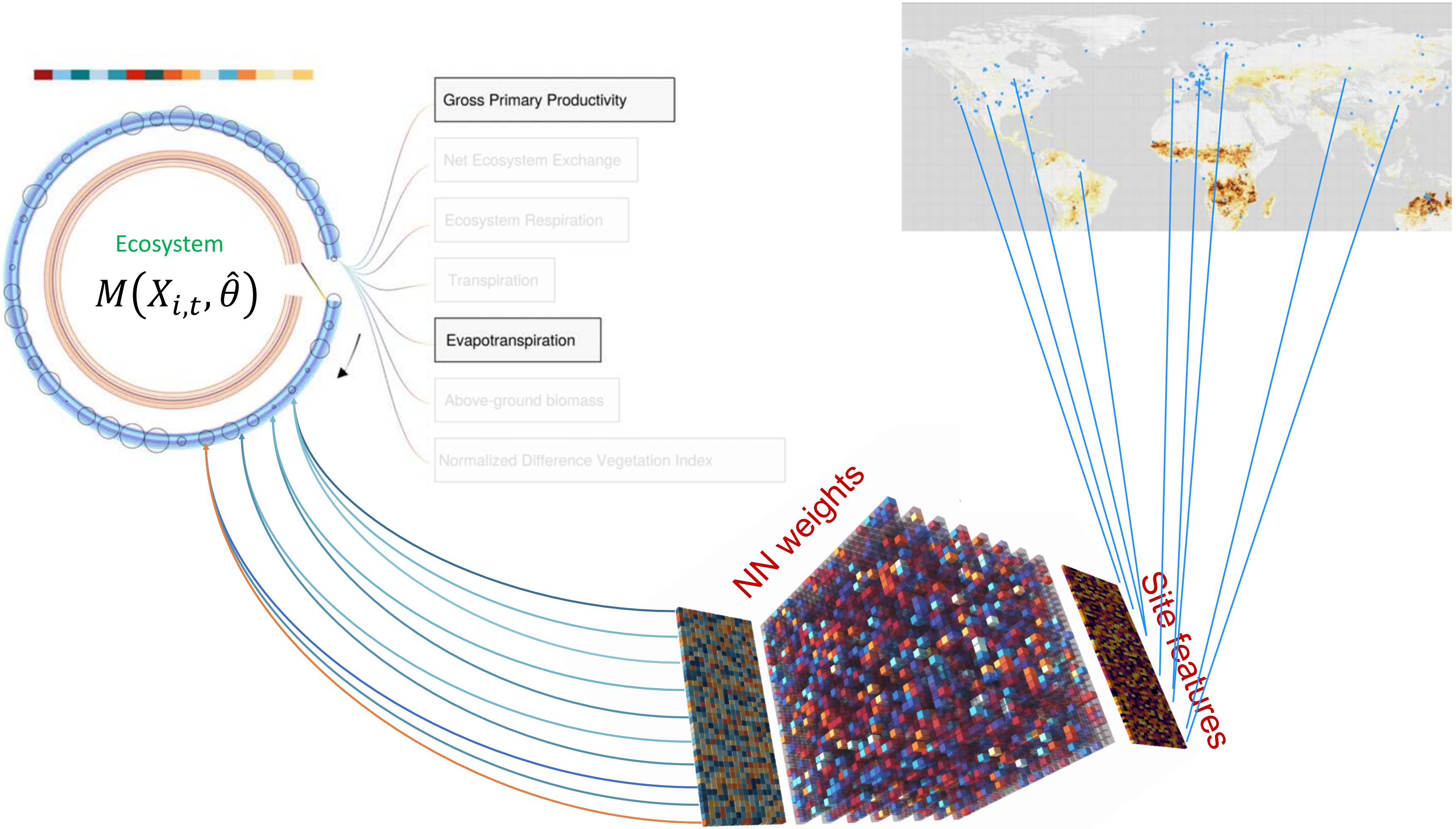
[Bao et al., JAMES, 2024]

PARAMETERIZATIONS

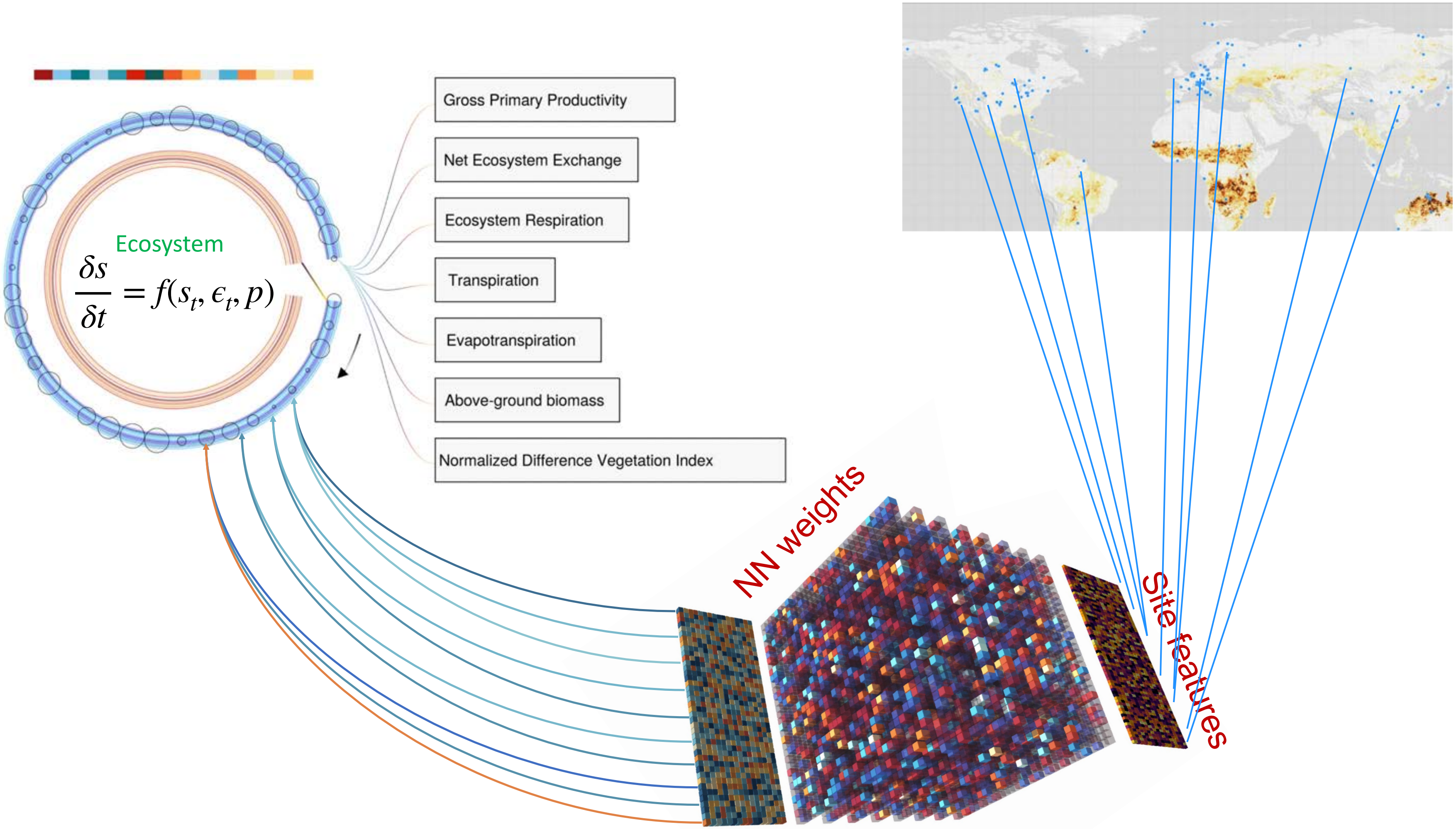


[Alonso et al., in prep.]

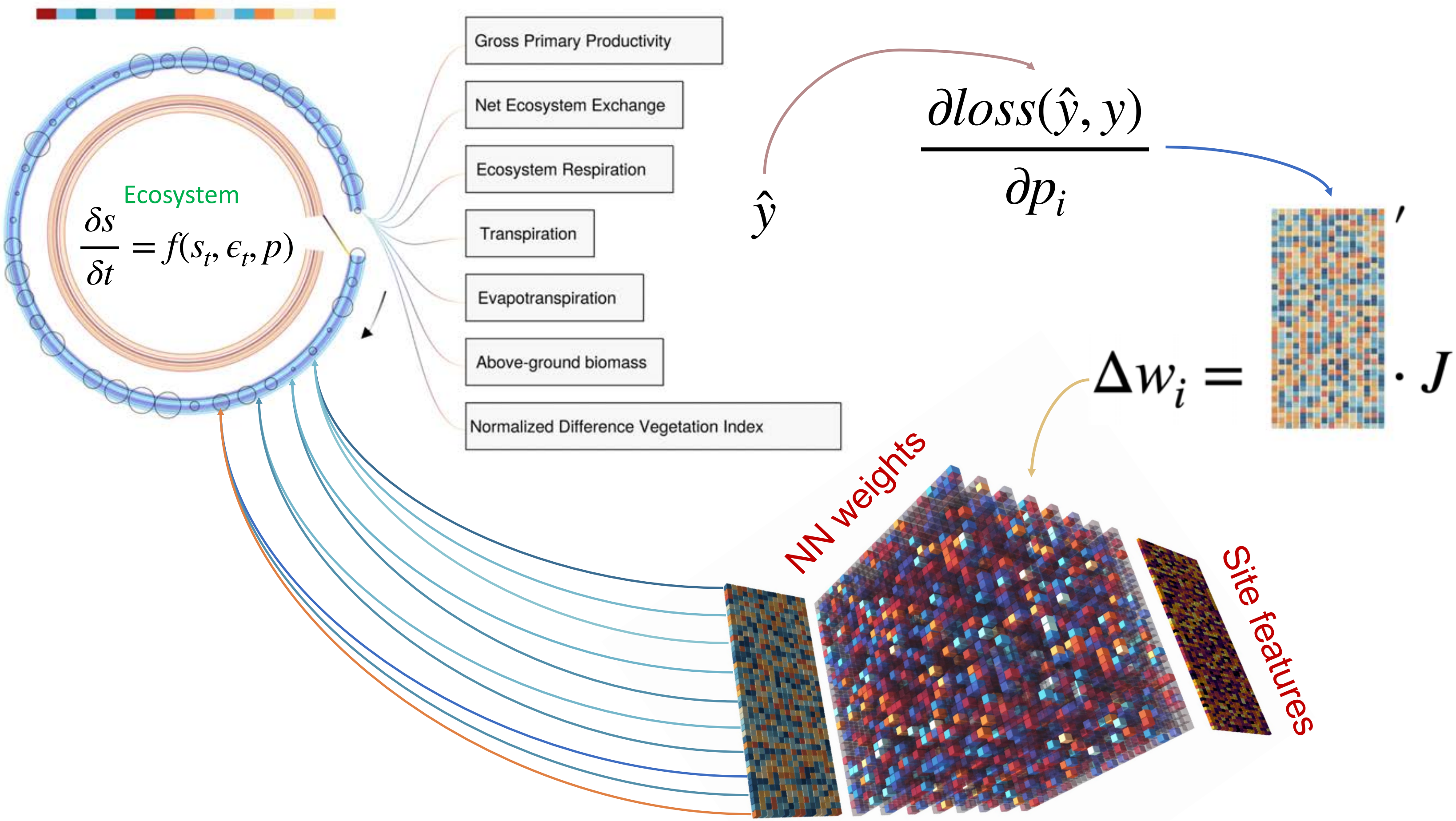
PARAMETERIZATIONS



[Bao et al., JAMES, 2024; Alonso et al., in prep.]

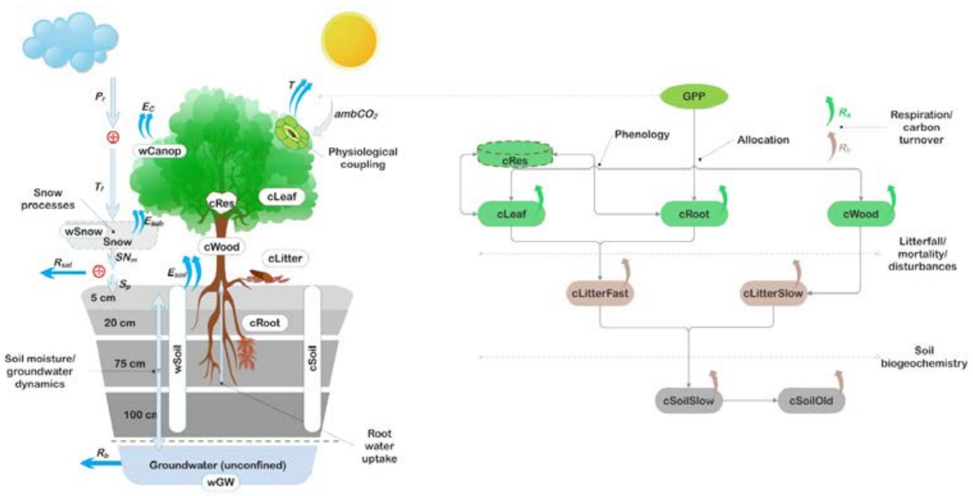


[Alonso et al., in prep.]



[Alonso et al., in prep.]

Terrestrial ecosystem model



[Koirala et al., in prep.]

Prognostic C/H2O cycle model: light use efficiency model, incl. physiological coupling between C/H2O fluxes; root water uptake conditioned on below ground plant carbon stocks; C allocation based on the co-limiting resources availability; decomposition dependent on temperature and moisture dynamics.

Constraints: GPP, NEE, Reco, evapotranspiration, transpiration, above ground biomass, phenology (NDVI)

Experiment

In situ

Parameter inversion

$\hat{y} = M(\theta, u)$

$\hat{\theta} = \arg\min_{\theta} \chi^2$

χ^2: multiple constrain cost/loss

Optimizer: CMAES

Hybrid modelling approach

Learning a model

$\hat{y} = M\left(\hat{\theta}, u\right)$

$\hat{\theta} = NN\left(\vec{X}, w, b\right)$

$\vec{X}$: features including

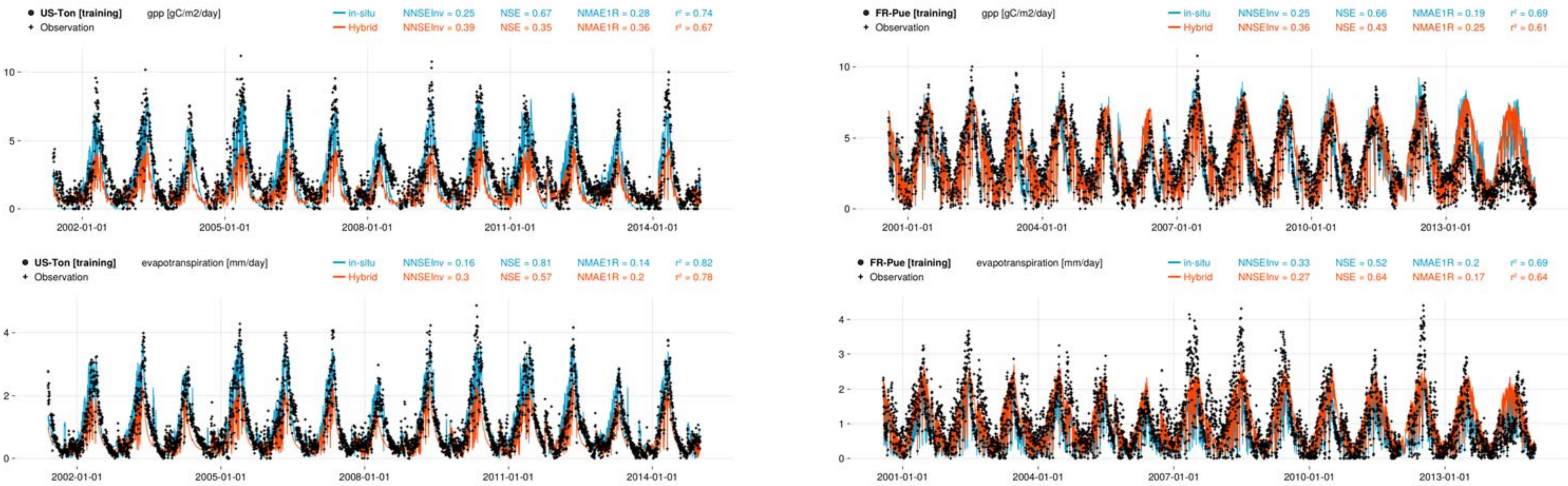
a) bio/climate, soils, PFT

b) PFT

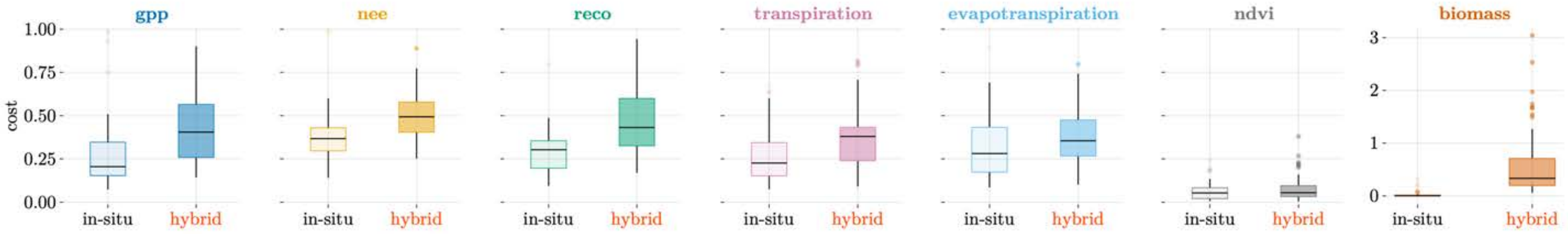
χ^2: multiple constrain cost/loss

Results: in-situ & hybrid

Results: in-situ & hybrid

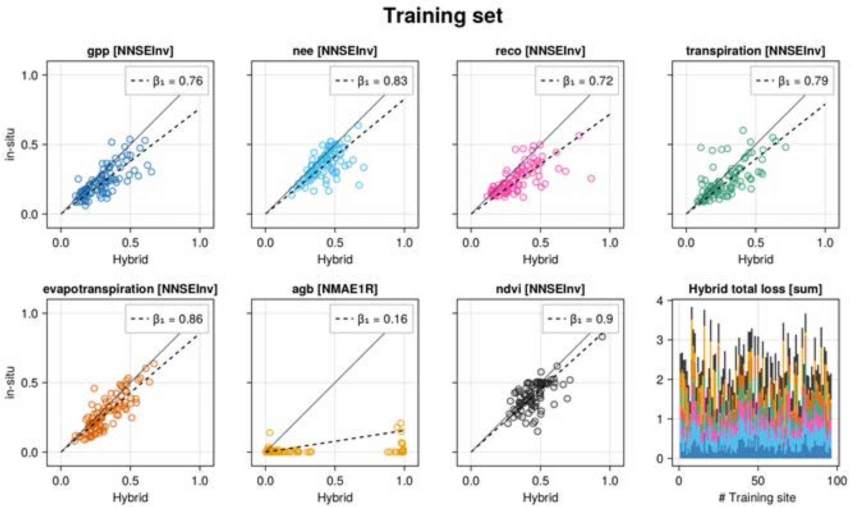


Results: in-situ & hybrid



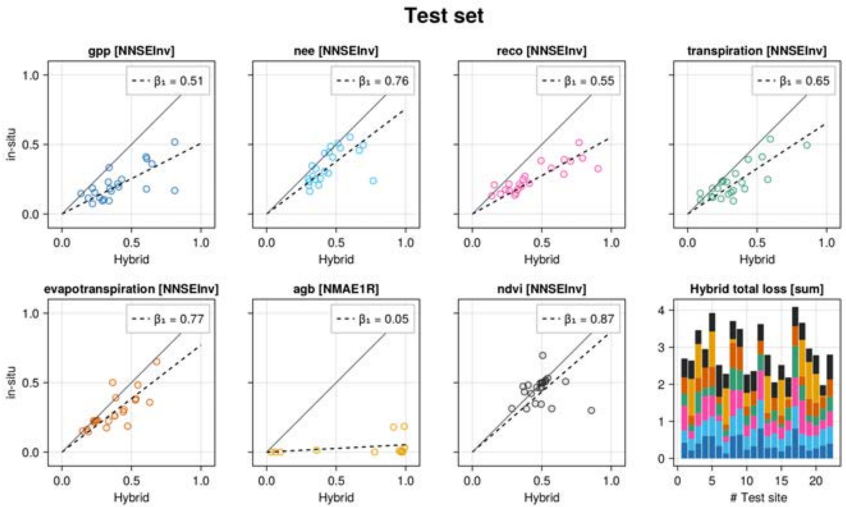
Results: in-situ & hybrid

in-situ model performance



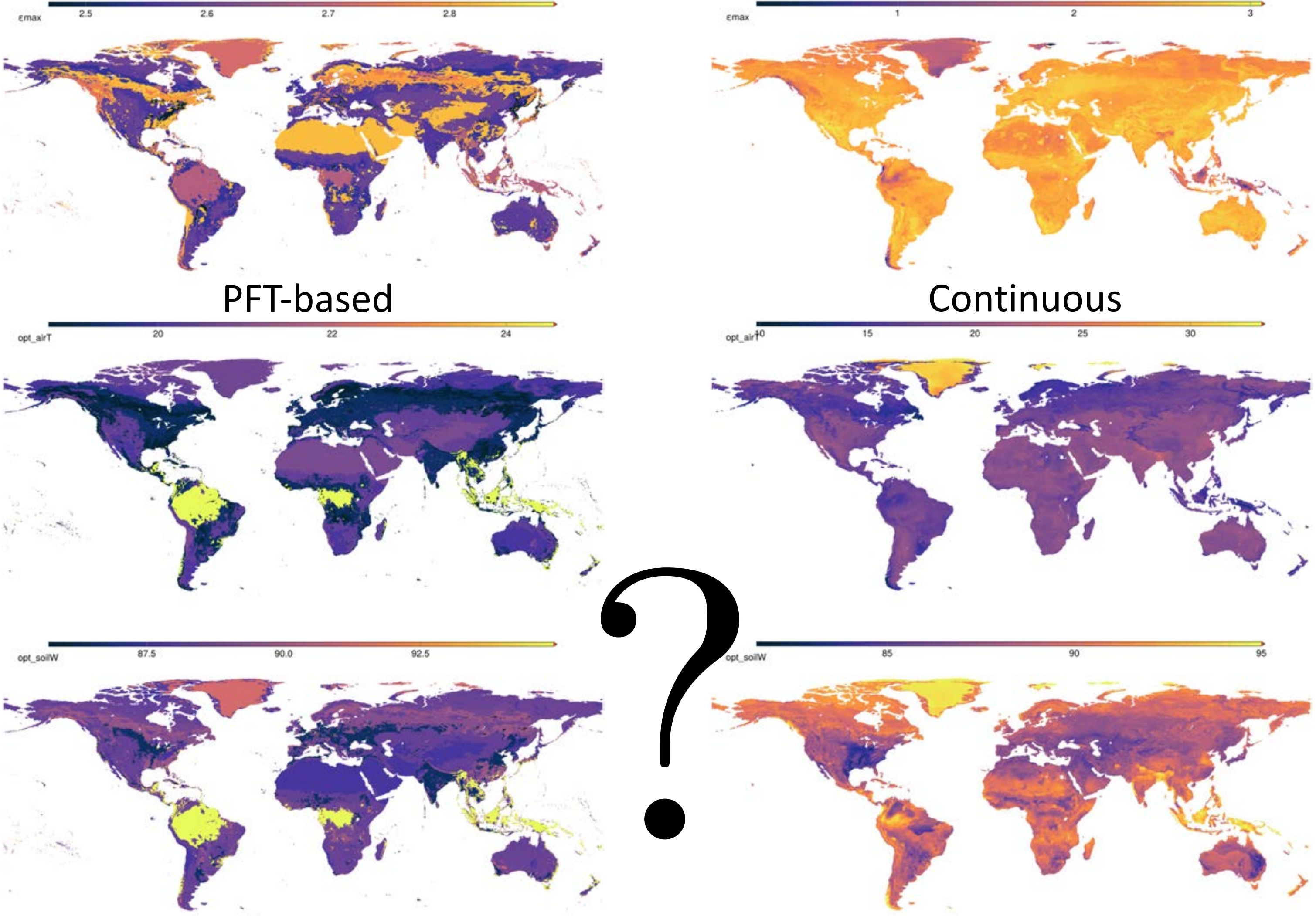
hybrid model performance

in-situ model performance



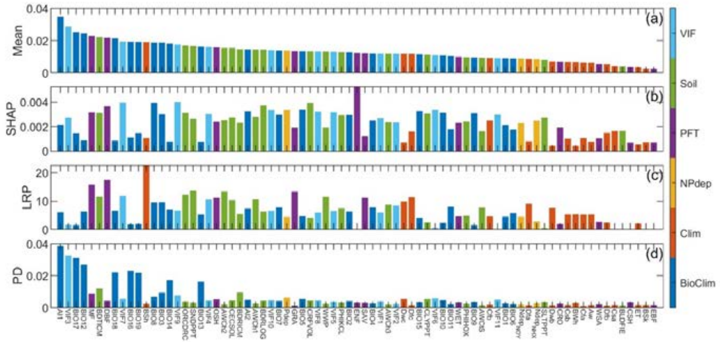
hybrid model performance

Upscaled parameters

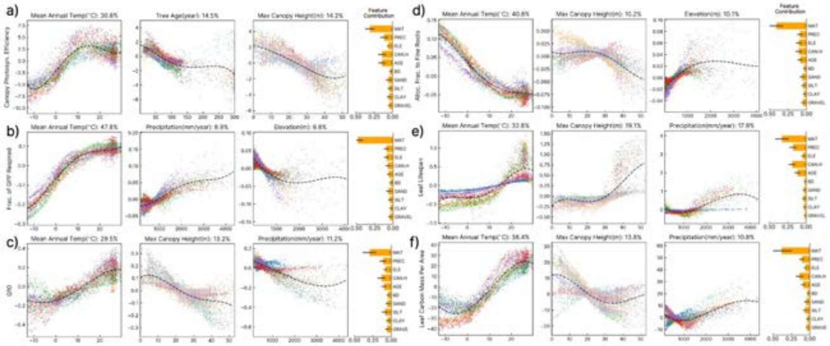


Early findings

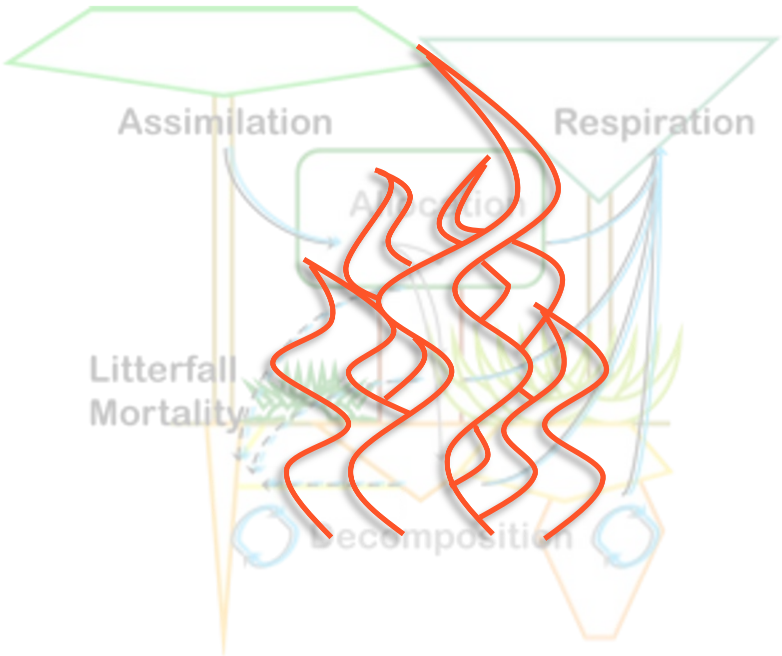
- PFT still a determining factor of parameter variability
  - Plant form/structure
- Global parameter variations do not follow PTF patterns
- Individual feature relevance contingent on approach
  - Sensitivity dependent on ensemble member



[Bao et al., 2024]



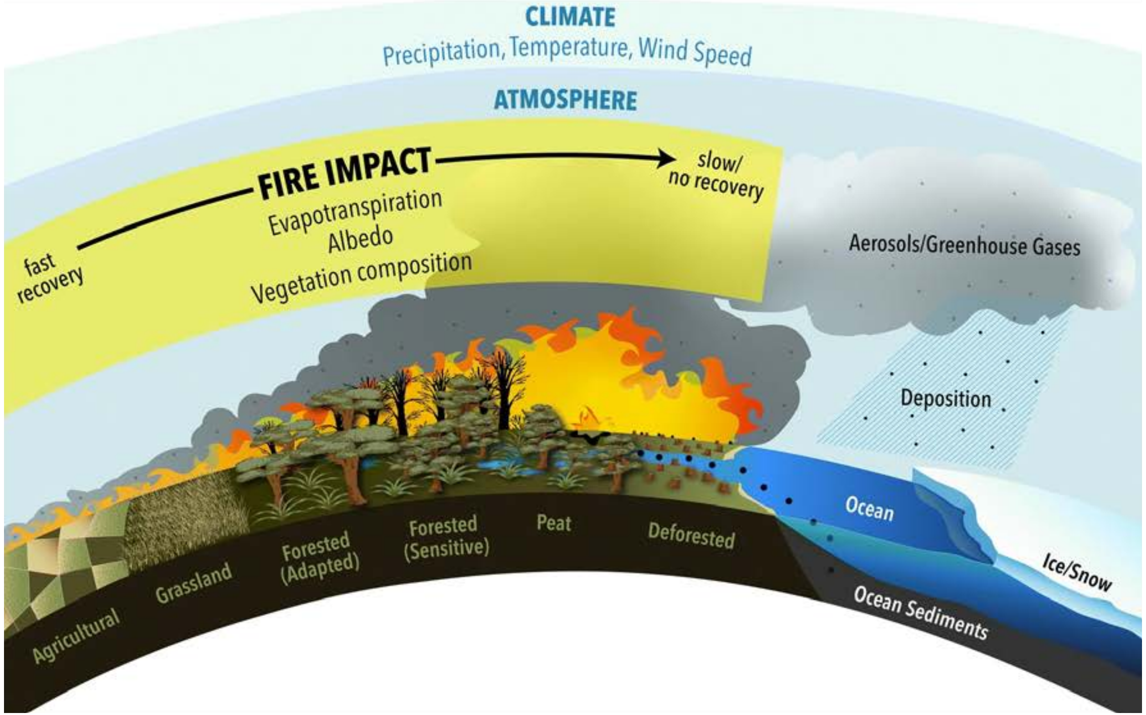
[Fang and Gentine, 2025]



[Son et a., JAMES, 2024]

LEARNING MODEL STRUCTURES

Fire impacts on Earth system & modelling challenge



[Lasslop et al., 2019]

ML/DL applications in fire science

Burned area prediction

- Global [Zhang et al., 2022]
- Canada [Cheng et al., 2008]
- Australia [Bergado et al., 2021]
- Portugal [Li et al., 2021]

Fire occurrence prediction

- Fire incidence [Dutta et al., 2013]
- Lightning ignition [Coughlan et al., 2021]
- Fire susceptibility [Zhang et al., 2021]

Fire spread prediction

- Front spread [Hodges et al., 2019]
- Spread dynamics [Subramanian et al., 2018]

Fire weather prediction

- Extreme condition [Langerquist et al, 2017]
- FWI [Son et al., 2022]

Fuel availability prediction

- Fuel load [D'ESTE et al., 2021]

[Jain et al., 2020]

Process Abstraction

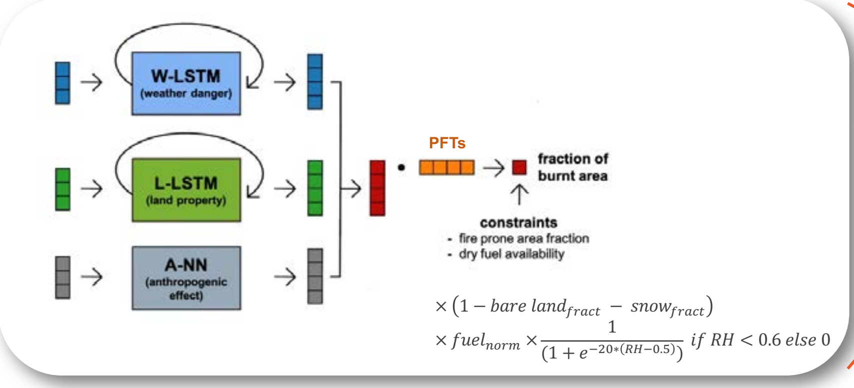
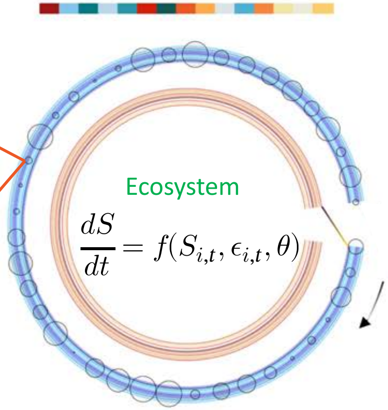


Table 2. Model input dataset

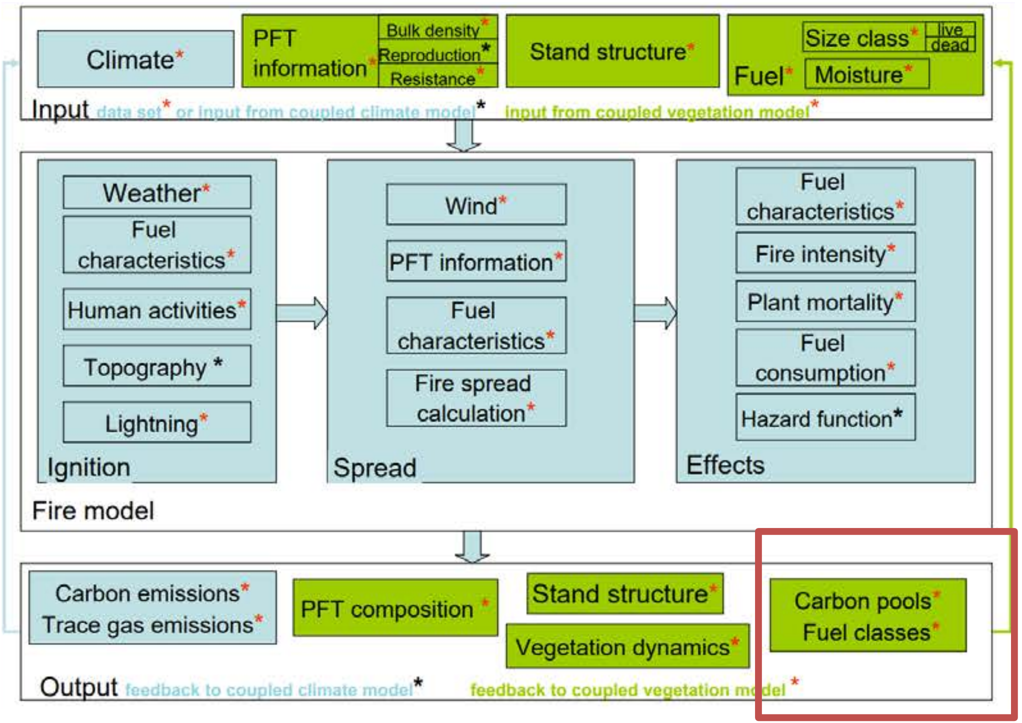
Weather driven fire danger (W-LSTM)	temperature temperature anomaly	ERA5 (Hersbach et al., 2020)
	specific/relative humidity specific/relative humidity anomaly	
	wind speed	
	precipitation	
	lightning climatology	LIS/OTD (Cecil et al., 2014)
Land properties (L-LSTM)	volume of water in soil 4 layers lv1: 0-7cm, lv2: 7-28cm, lv3: 28-100cm, lv4: 100-289cm volume of water anomaly (4 levels)	ERA5
	LAI LAI anomaly	MODIS (MCD15A3H)
	Topography : elevation, slope, roughness	(Amatulli et al., 2018)
	fuel (above ground plant litter)	JSBACH4
	fraction of 9 plant functional types (PFTs) - snow - tree broadleaf evergreen / deciduous - tree needleleaf evergreen / deciduous - shrub evergreen / broadleaf deciduous - grass - bare land	
Anthropogenic effect (A-NN)	population density	HYDE3.2 (Klein Goldewijk et al., 2017)
	gross domestic product (GDP)	(Kummu et al., 2018)
	human development index (HDI)	
	total road density	GRIP4 (Meijer et al., 2018)
	land use (14) states	LUH2 (Hurtt et al., 2020)

target : GFED4 (fraction of BA)
resolution : daily, 0.25x0.25
data split : 2004-10 (train), 2011-15 (test)

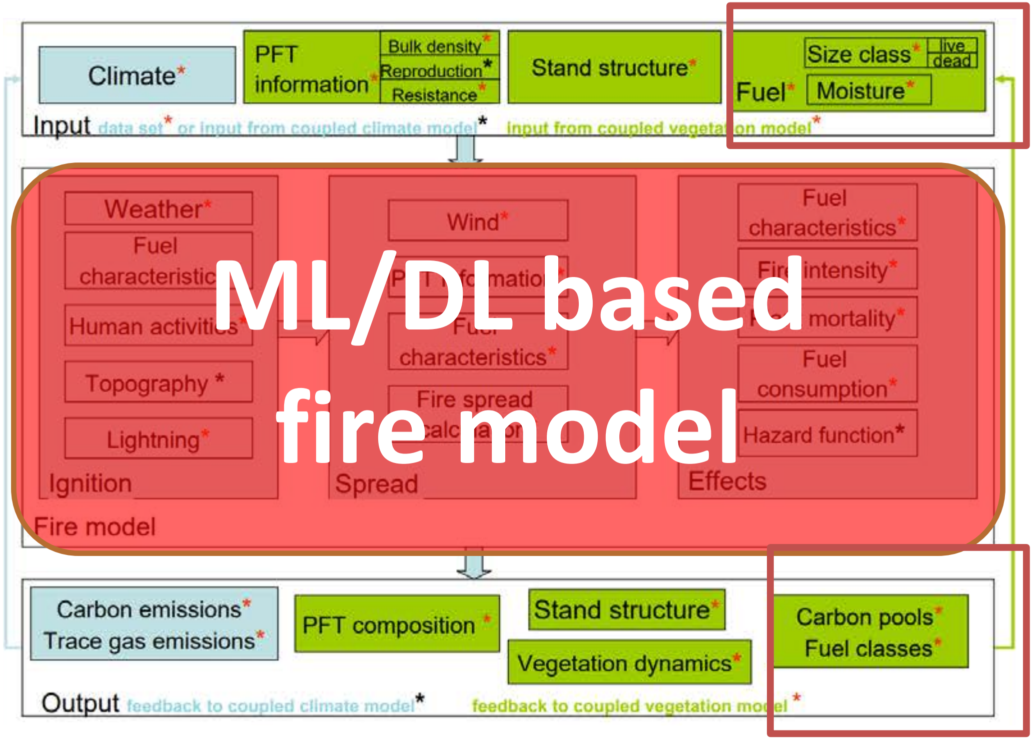
GFED4 14 regions



[Son et al., JAMES, 2024]



[Thonicke et al., 2010]


Climate*
PFT information*
Bulk density*
Reproduction*
Resistance*
Stand structure*
Fuel*
Size class*
live
dead
Moisture*
Input data set* or input from coupled climate model*
input from coupled vegetation model*
ML/DL based fire model
Weather*
Human activities*
Topography *
Lightning*
Ignition
Wind*
Spread
Fuel characteristics*
Hazard function*
Effects
Fire model
Carbon emissions*
Trace gas emissions*
PFT composition *
Stand structure*
Vegetation dynamics*
Carbon pools*
Fuel classes*
Output feedback to coupled climate model*
feedback to coupled vegetation model *

Data

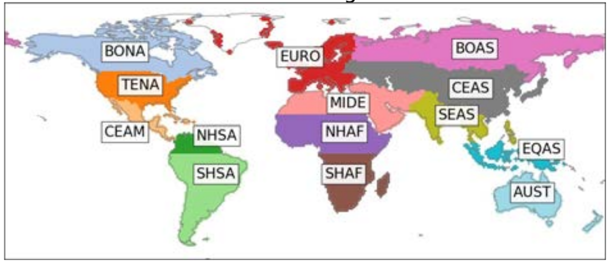
Table 2. Model input dataset

Weather driven fire danger (W-LSTM)	temperature temperature anomaly	ERA5 (Hersbach et al., 2020)
	specific/relative humidity specific/relative humidity anomaly	
	wind speed	
	precipitation	
	lightning climatology	LIS/OTD (Cecil et al., 2014)
Land properties (L-LSTM)	volume of water in soil 4 layers lv1: 0-7cm, lv2: 7-28cm, lv3: 28-100cm, lv4: 100-289cm volume of water anomaly (4 levels)	ERA5
	LAI LAI anomaly	MODIS (MCD15A3H)
	Topography : elevation, slope, roughness	(Amatulli et al., 2018)
	fuel (above ground plant litter)	JSBACH4
	fraction of 9 plant functional types (PFTs) - snow - tree broadleaf evergreen / deciduous - tree needleleaf evergreen / deciduous - shrub evergreen / broadleaf deciduous - grass - bare land	

Anthropogenic effect (A-NN)	population density	HYDE3.2 (Klein Goldewijk et al., 2017)
	gross domestic product (GDP)	(Kummu et al., 2018)
	human development index (HDI)	
	total road density	GRIP4 (Meijer et al., 2018)
	land use (14) states	LUH2 (Hurtt et al., 2020)

target : GFED4 (fraction of BA)
resolution : daily, 0.25x0.25
data split : 2004-10 (train), 2011-15 (test)

GFED4 14 regions



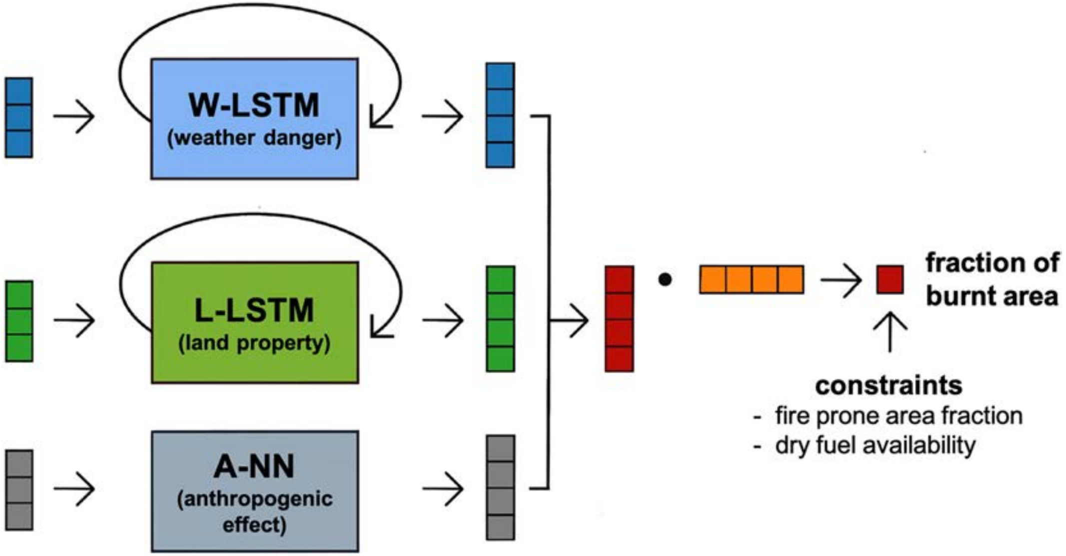
Modelling

Land surface scheme of ICON

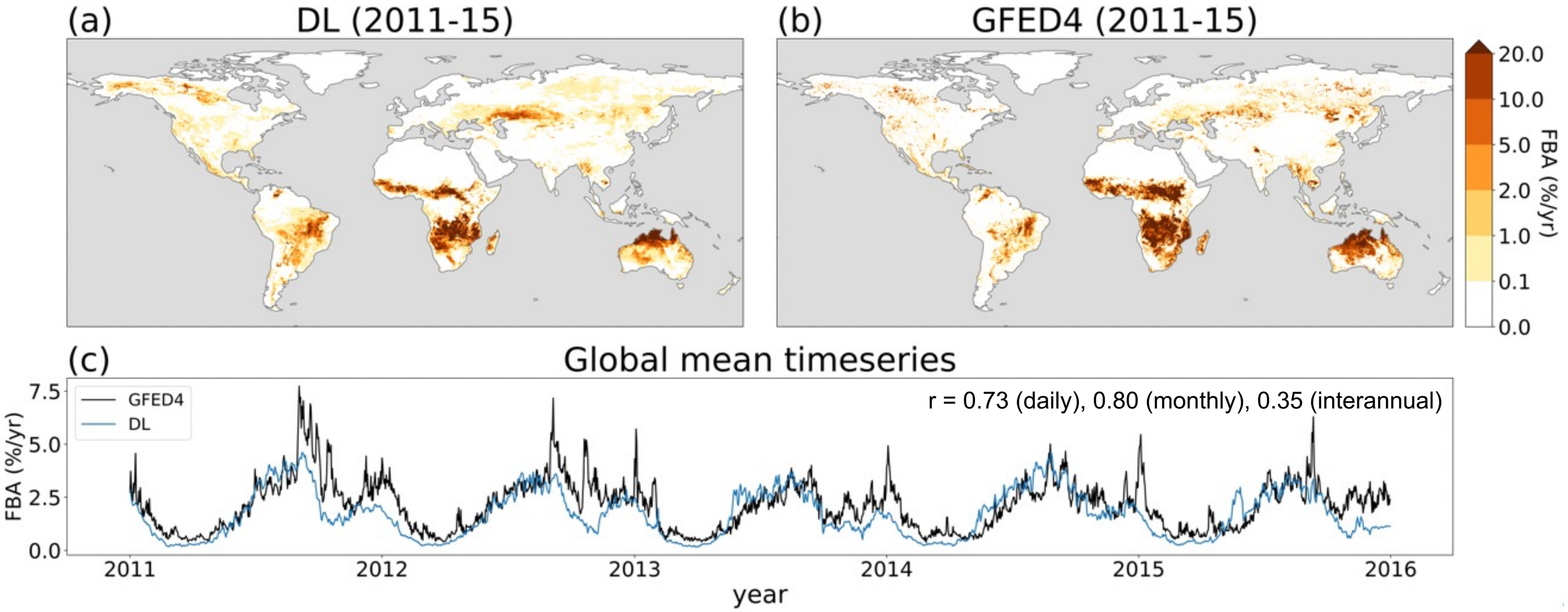
- JSBACH4

DL-fire model

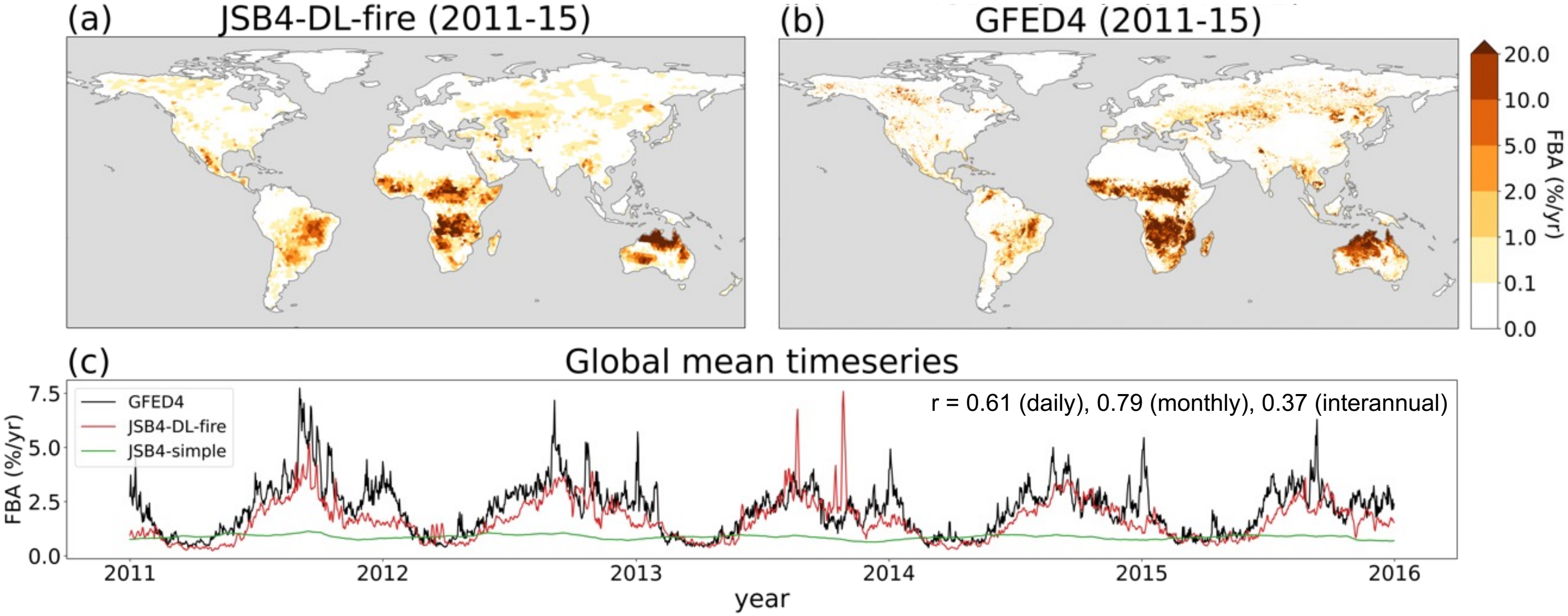
- Son et al 2024



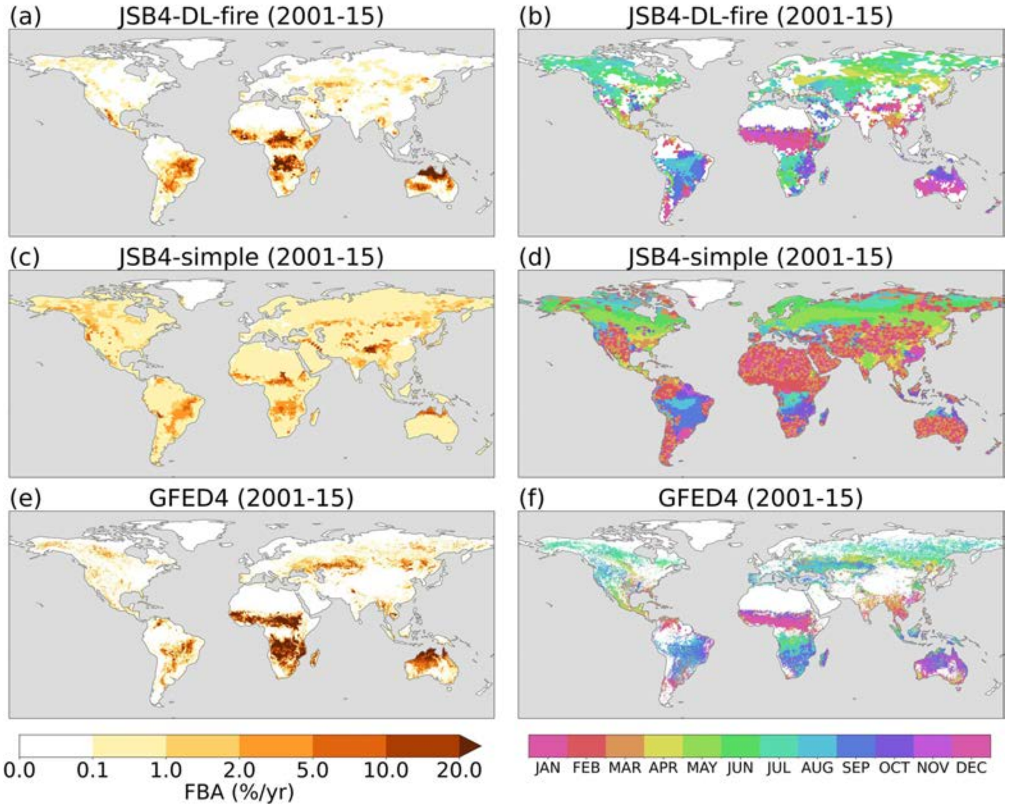
DL-fire model evaluation (offline inference)



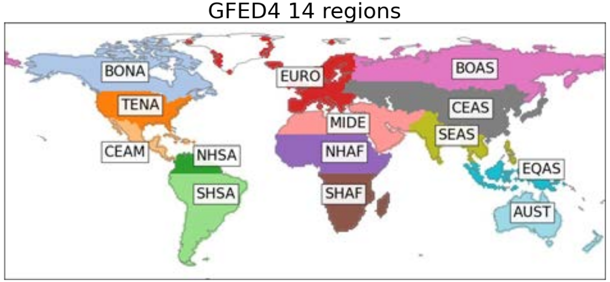
DL-fire model evaluation (online inference)

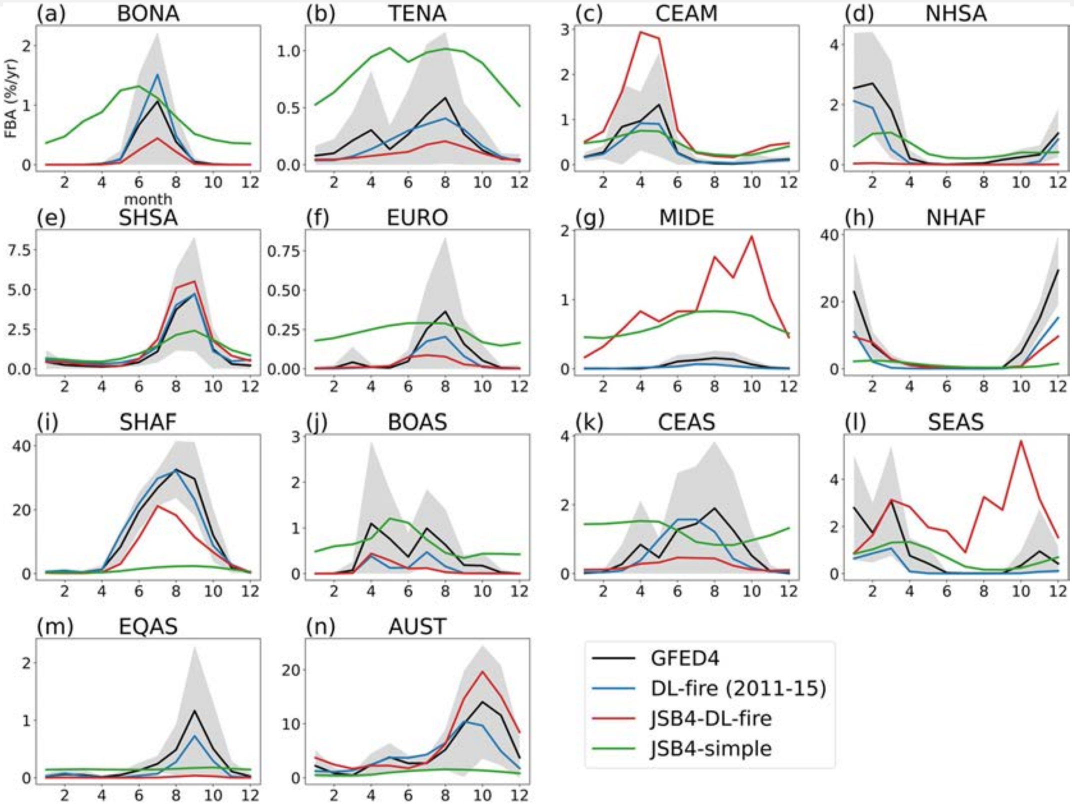


Peak fire season



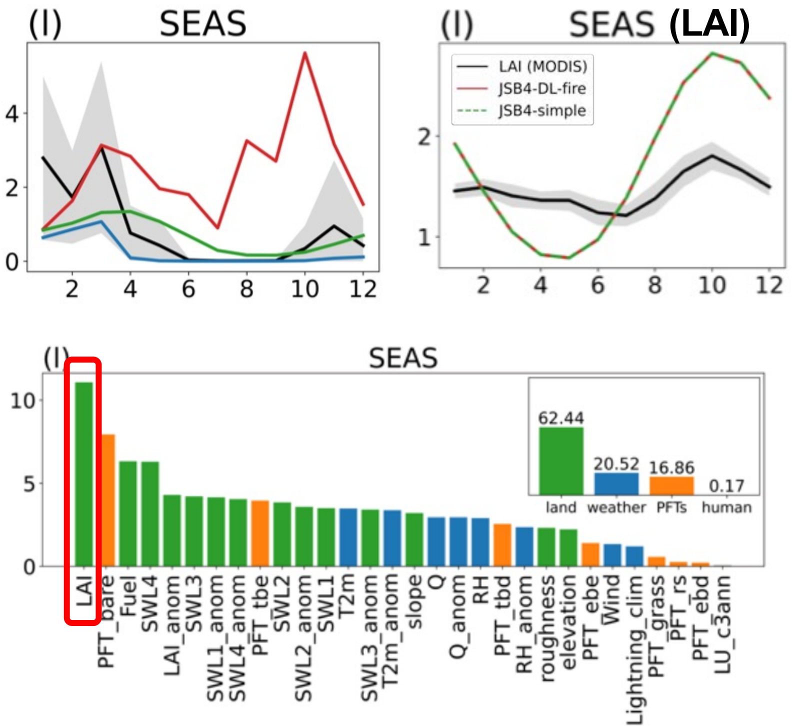
Spatial variability





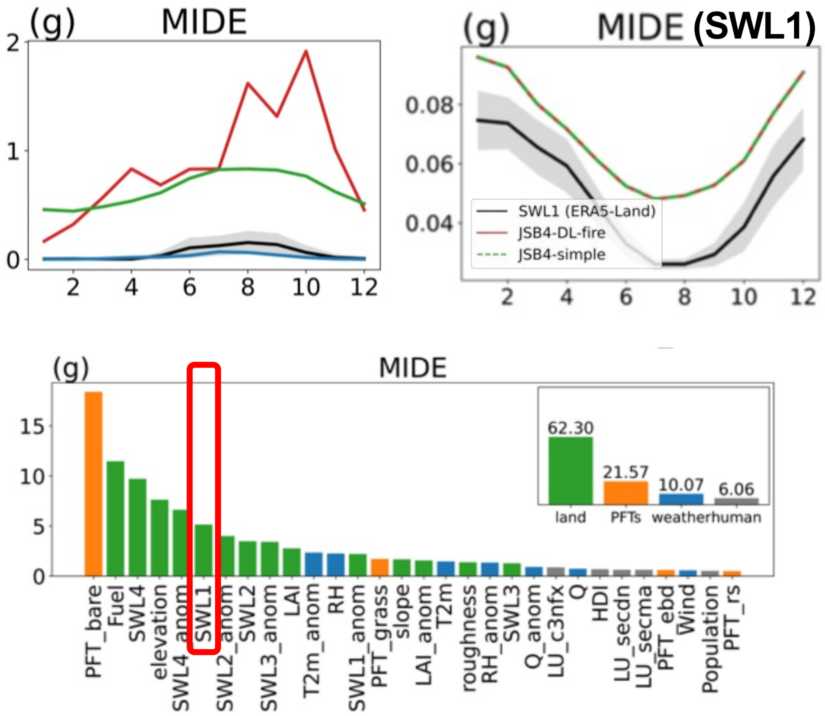
UNDER THE HOOD

Internal bias limitations in JSBACH-DL-Fire

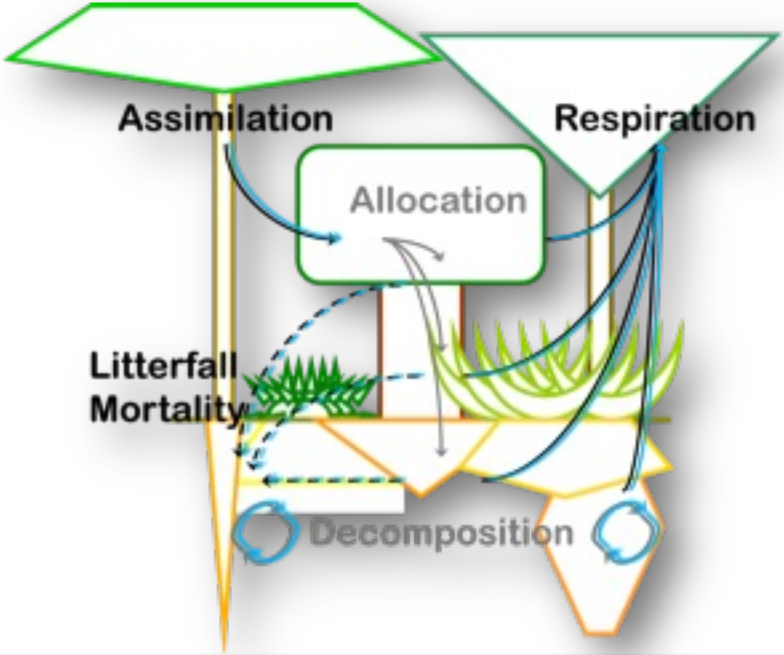


1. LAI	LAI
2. PFT_bare	Bare land fraction
3. Fuel	Fuel
4. SWL4	Volume of water in soil in the 4th layer
5. LAI_anom	LAI anomaly
6. SWL3	Volume of water in the soil 3rd layer
7. SWL1_anom	Anomaly of volume of water in the soil 1st layer
8. SWL4_anom	Anomaly of volume of water in the soil 4th layer
9. PFT_tbe	Tree broadleaf evergreen fraction
10. SWL2	Volume of water in the soil 2nd layer
11. SWL2_anom	Anomaly of volume of water in the soil 2nd layer
12. SWL1	Volume of water in the soil 1st layer
13. T2m	Temperature
14. SWL3_anom	Anomaly of volume of water in the soil 3rd layer
15. T2m_anom	Anomaly of temperature
16. slope	Slope
17. Q	Specific humidity
18. Q_anom	Anomaly of specific humidity
19. RH	Relative humidity
20. PFT_tbd	Tree broad deciduous fraction

Internal bias limitations in JSBACH-DL-Fire

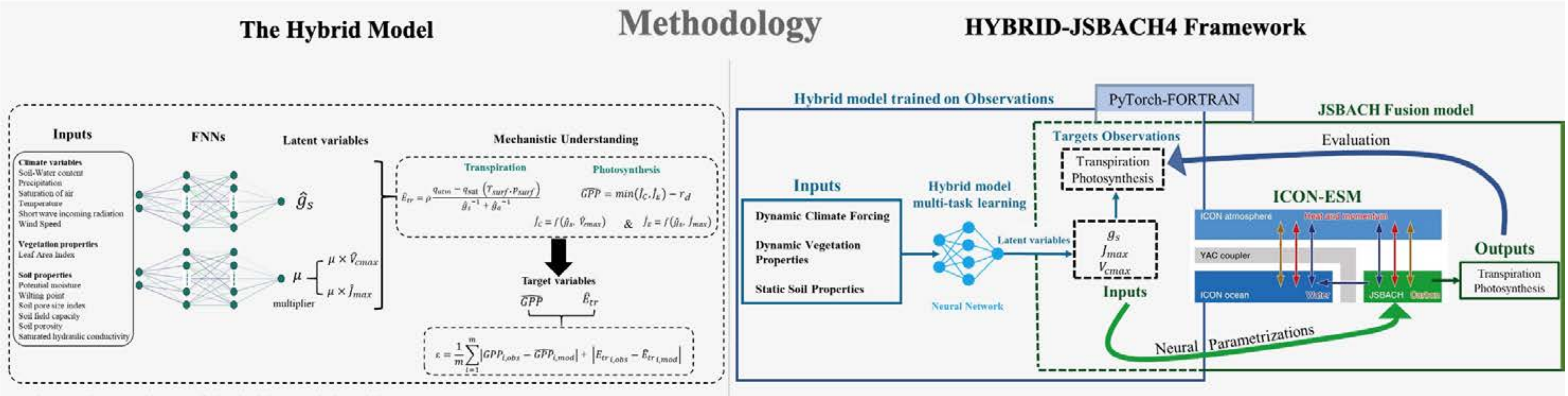


1. LAI	LAI
2. PFT_bare	Bare land fraction
3. Fuel	Fuel
4. SWL4	Volume of water in soil in the 4th layer
5. LAI_anom	LAI anomaly
6. SWL3	Volume of water in the soil 3rd layer
7. SWL1_anom	Anomaly of volume of water in the soil 1st layer
8. SWL4_anom	Anomaly of volume of water in the soil 4th layer
9. PFT_tbe	Tree broadleaf evergreen fraction
10. SWL2	Volume of water in the soil 2nd layer
11. SWL2_anom	Anomaly of volume of water in the soil 2nd layer
12. SWL1	Volume of water in the soil 1st layer
13. T2m	Temperature
14. SWL3_anom	Anomaly of volume of water in the soil 3rd layer
15. T2m_anom	Anomaly of temperature
16. slope	Slope
17. Q	Specific humidity
18. Q_anom	Anomaly of specific humidity
19. RH	Relative humidity
20. PFT_tbd	Tree broad deciduous fraction

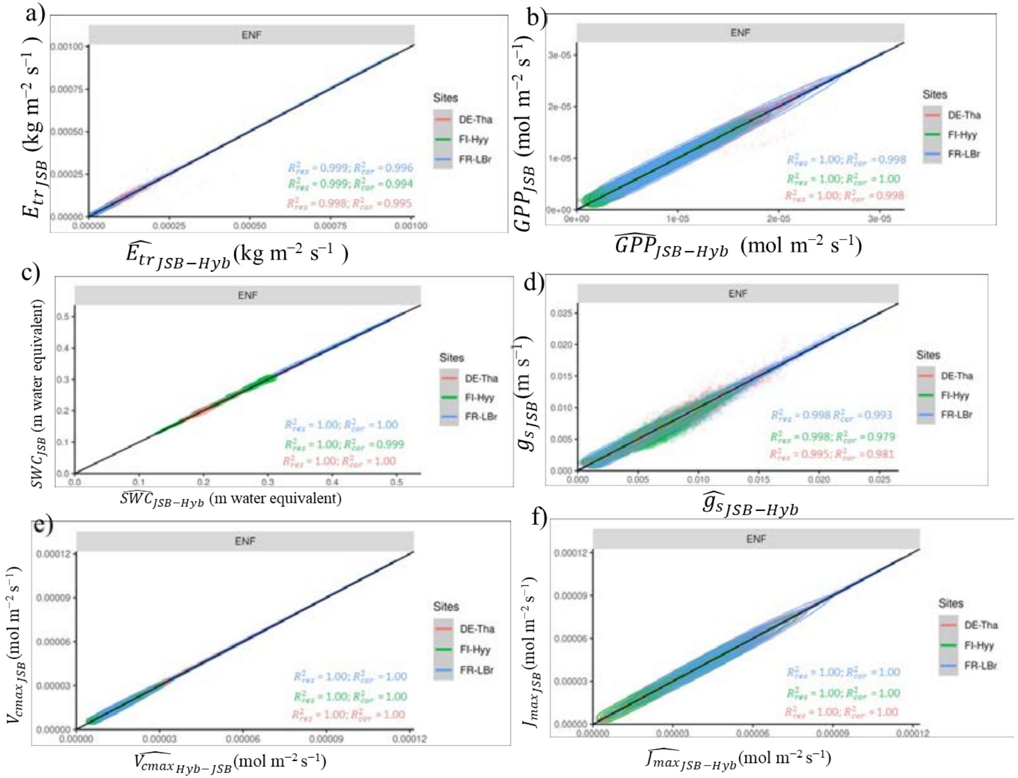


[El Ghawi, 2025; in prep.]

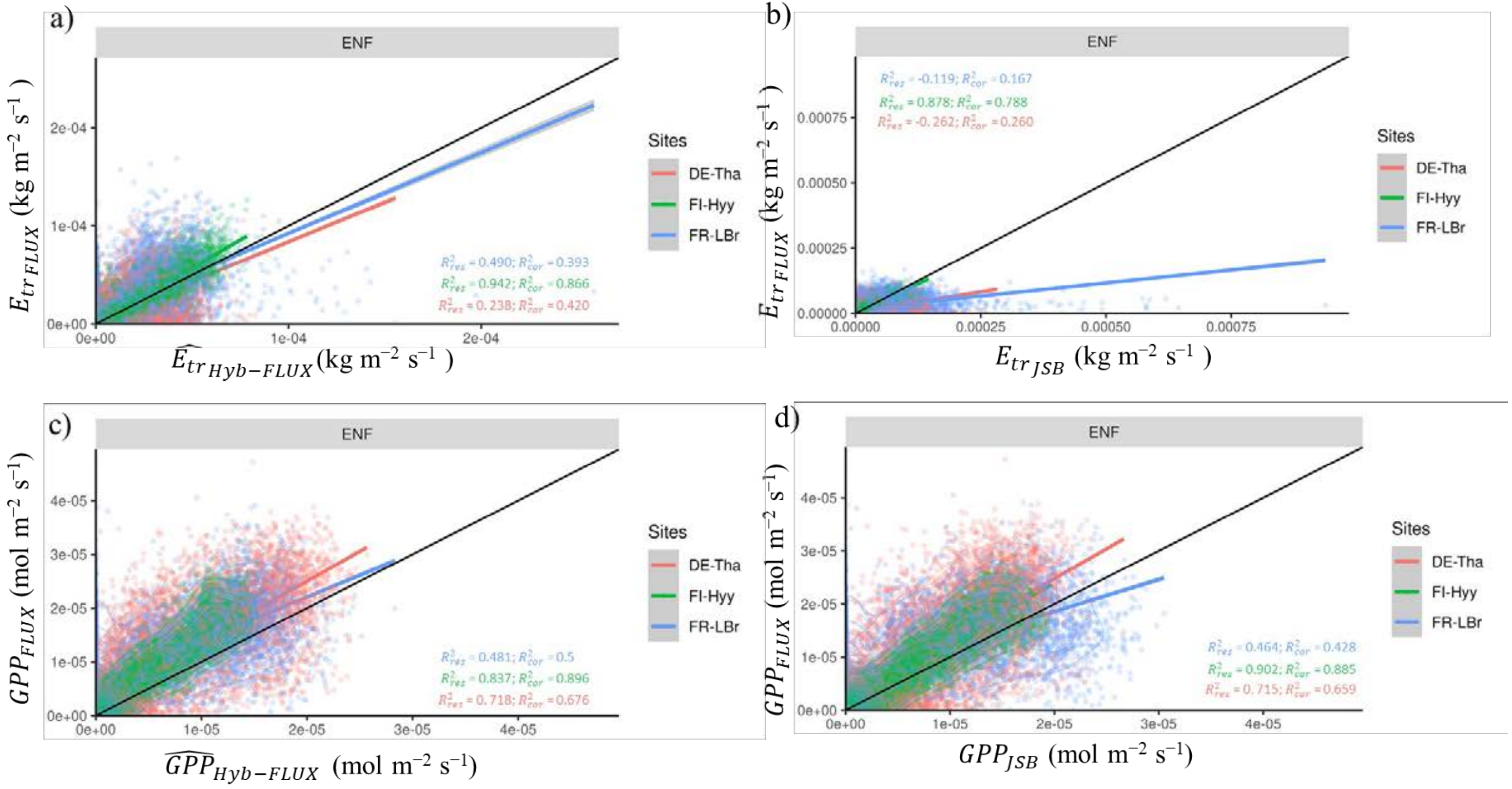
PHOTOSYNTHESIS / CONDUCTANCE


Methodology
The Hybrid Model
Inputs
FNNs
Latent variables
Mechanistic Understanding
Climate variables
Soil-Water content
Precipitation
Saturation of air
Temperature
Short wave incoming radiation
Wind Speed
Vegetation properties
Leaf Area Index
Soil properties
Potential moisture
Wilting point
Soil pore size index
Soil field capacity
Soil porosity
Saturated hydraulic conductivity
multiplier
Transpiration
Photosynthesis
Target variables
HYBRID-JSBACH4 Framework
Hybrid model trained on Observations
PyTorch-FORTRAN
JSBACH Fusion model
Targets Observations
Transpiration
Photosynthesis
Evaluation
Inputs
Dynamic Climate Forcing
Dynamic Vegetation Properties
Static Soil Properties
Hybrid model multi-task learning
Neural Network
Latent variables
Inputs
ICON-ESM
ICON atmosphere
YAC coupler
ICON ocean
Water
JSBACH
Carbon
Outputs
Transpiration
Photosynthesis
Neural Parametrizations

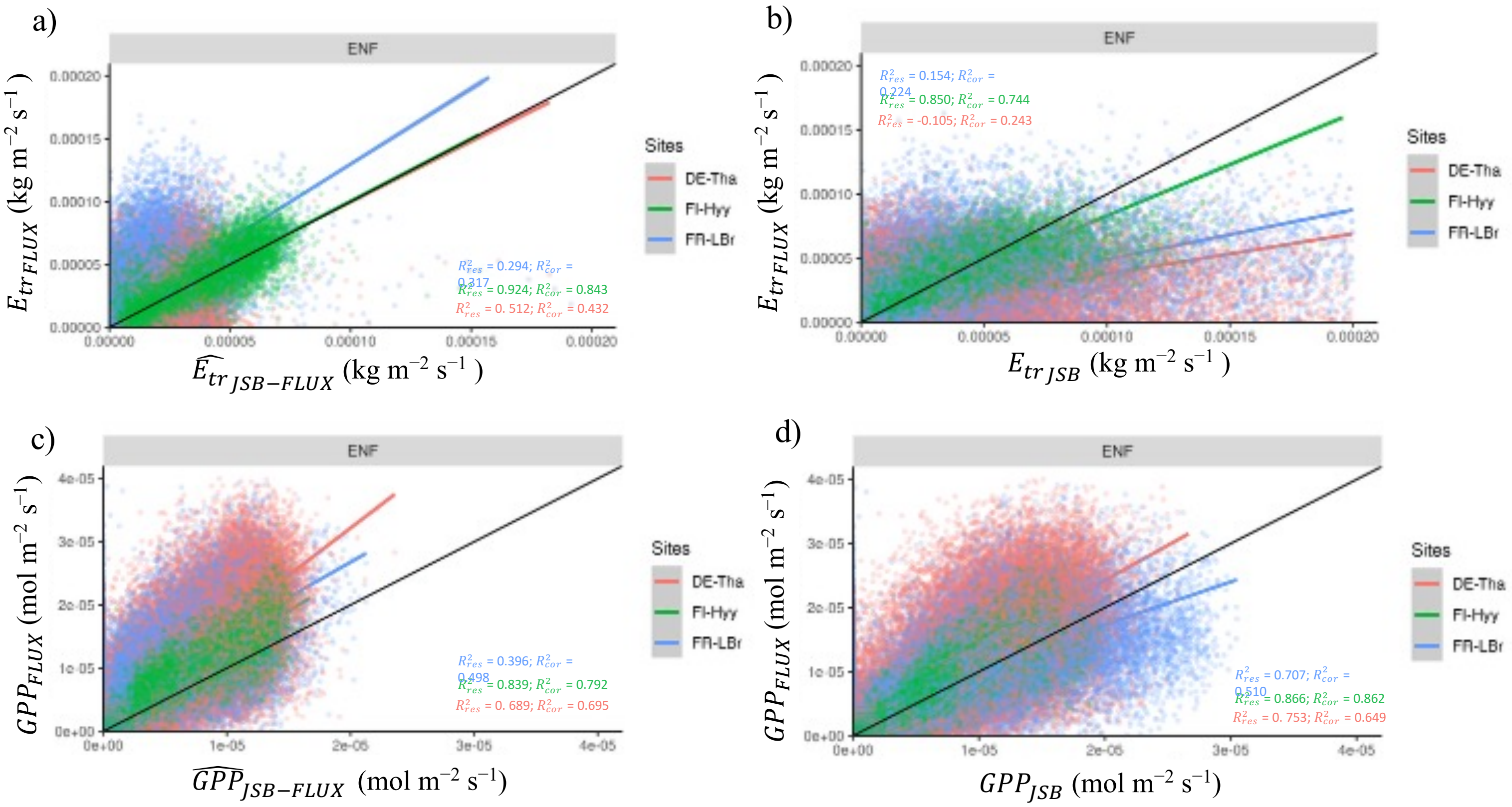
Larning JSBACH parametrizations



Training hybrid on FLUXNET



Integrating hybrid in JSBACH



- Process abstraction

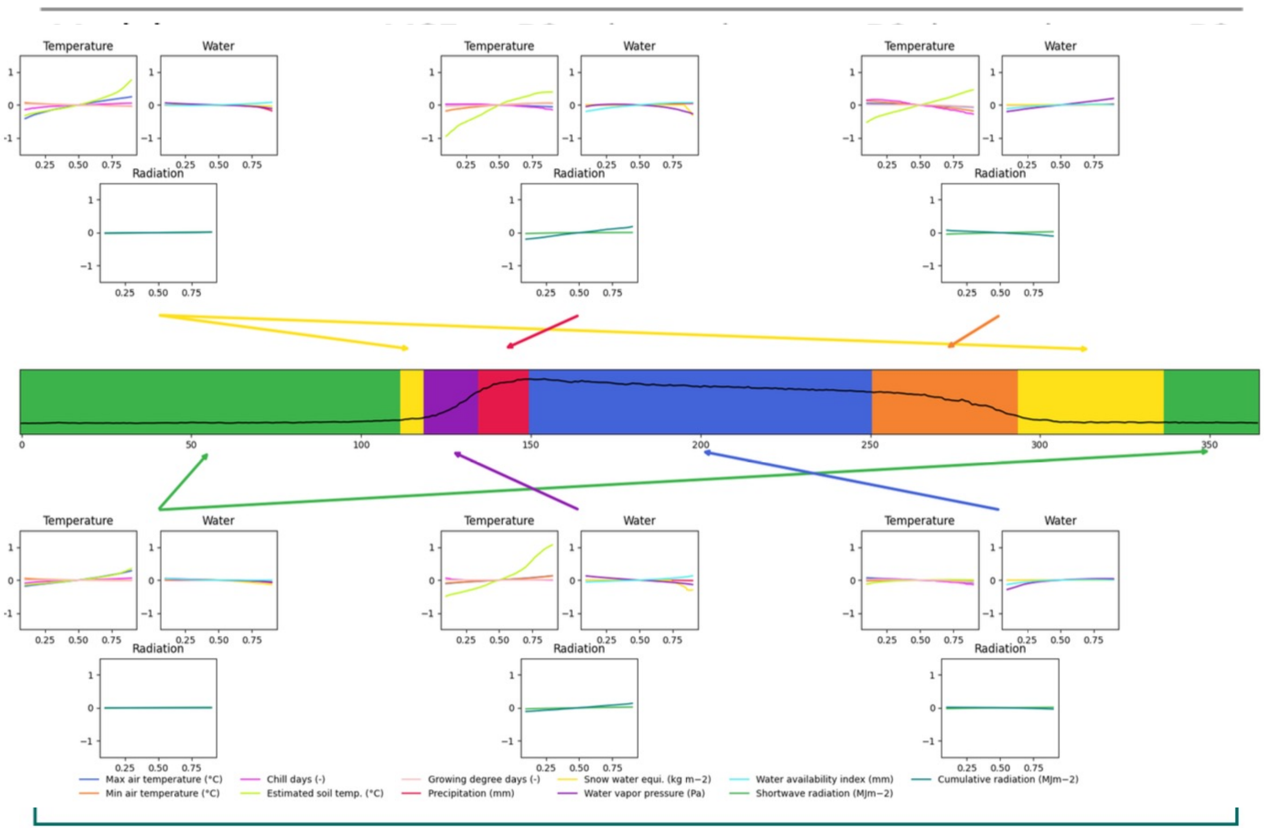
  $\Longrightarrow$ Leading to uncertainty reduction

- Off-line training + online inference

  $\Longrightarrow$ Error inflation

[Reimers et al., in prep.]

PHENOLOGY



Seasonal changes in sensitivities to drivers

OVERALL…

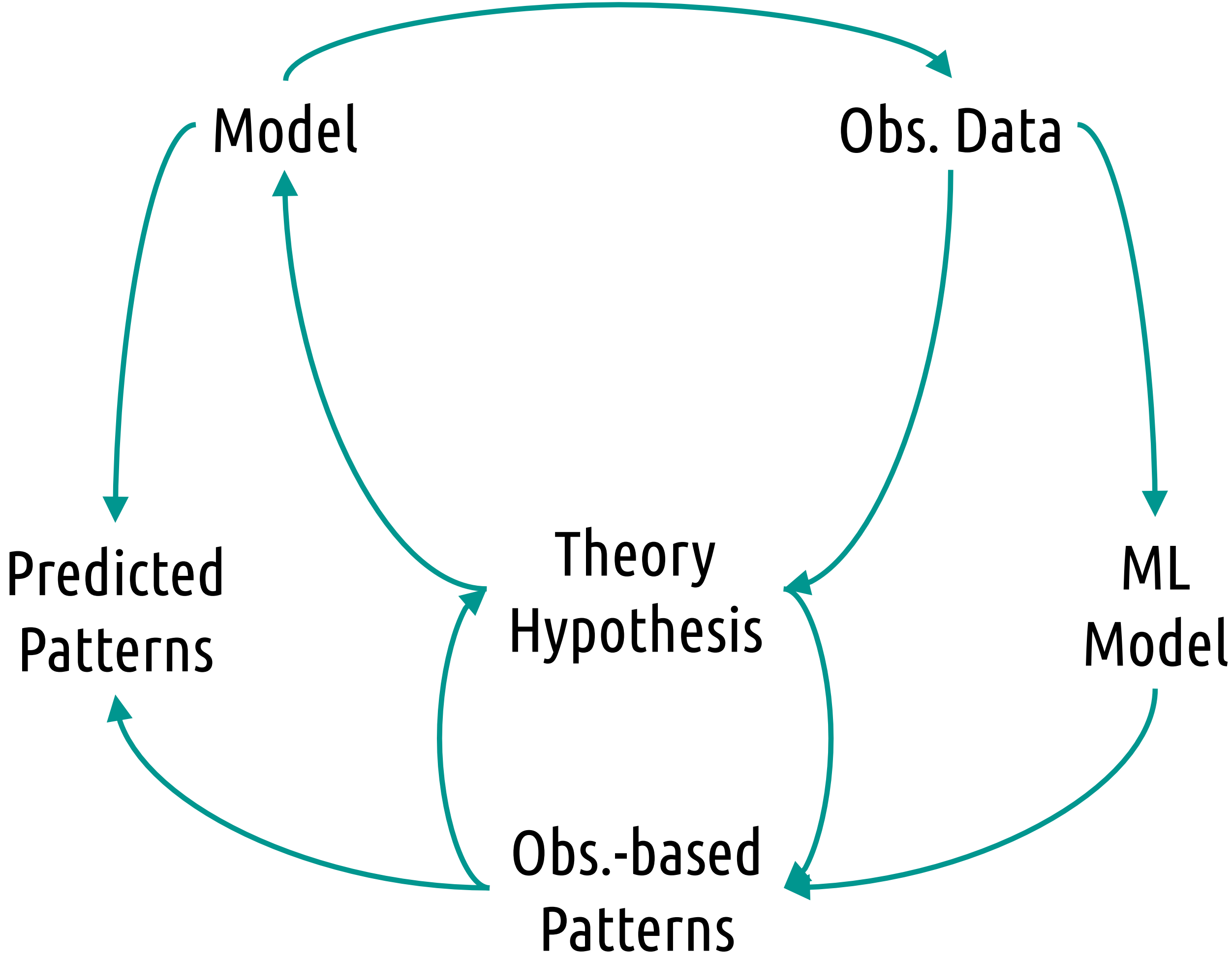
Overall

- ML support for
  $\Longrightarrow$ improved parameterizations (c.f. history matching / SBI)
  $\Longrightarrow$ process abstraction / representation

- Expands information content from observations to models
  $\Longrightarrow$ improved fits / reduction of uncertainties
  $\Longrightarrow$ off-line learning online inference $\Longrightarrow$ **differentiability**

- Generates “new” features / patterns
  $\Longrightarrow$ explainability ambiguities
  $\Longrightarrow$? hypothesis $\Longrightarrow$? learning (for us...)

HYPOTHESIS-DRIVEN / DATA-DRIVEN SCIENCE



[adapted from Reichstein et al., 2019]

Shanning Bao, Rackhun Son, Markus Reichstein, Lazaro Alonso, Siyuan Wang, Johannes Gensheimer, Ranit De, Tobias Stacke, Veronika Gayler, Julia Nabel, Reiner Schnur, Christian Requena-Mesa, Alexander J. Winkler, Stijn Hantson, Sönke Zaehle, Ulrich Weber, Bernhard Ahrens, Basil Kraft, Fabian Gans, Pierre Gentine, Gustau Camps-Valls, Sujan Koirala, Alexander Brenning

Torbern Tagesson, Michael Liddell, Andreas Ibrom, Sebastian Wolf, Ladislav Šigut, Lukas Hörtnagl, William Woodgate, J. Rose Cleverly, Mika Korkiakoski, Lutz Merbold, T. Andrew Black, Marilyn Roland, Anne Klosterhalfen, Peter D. Blanken, Sara Knox, Simone Sabbatini, Bert Gielen, Donatella Zona, Leonardo Montagnani, Rasmus Fensholt, Georg Wohlfahrt, Ankur R. Desai, Eugénie Paul-Limoges, Marta Galvagno, Albin Hammerle, Georg Jocher, Borja Ruiz Reverter, David Holl, Jiquan Chen, Luca Vitale, Pasi Kolari, M. Altaf Arain,…

ACKNOWLEDGEMENTS

AI4PEX

SINDBAD

STRATEGIES TO INTEGRATE DATA AND BIOGEOCHEMICAL MODELS

A model data integration framework

Start Learn Code Repository

THANK YOU!