A machine learning approach targeting parameter estimation for PFT coexistence modeling using ELM-FATES

Lingcheng Li[1] (Lingcheng.li@pnnl.gov)

Yilin Fang[2], Zhonghua Zheng[3], Mingjie Shi[1], Marcos Longo[4], Charles D. Koven[4], Jennifer A. Holm[4], Rosie A. Fisher[5], Nate G. McDowell[1,6], Jeffrey Chambers[4], L. Ruby Leung[1]

1. Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA
2. Earth System and Science Division, Pacific Northwest National Laboratory, Richland, WA, USA
3. Department of Earth and Environmental Sciences, The University of Manchester, Manchester, UK
4. Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
5. CICERO Center for International Climate and Environmental Research, Oslo, Norway
6. School of Biological Sciences, Washington State University, PO Box 644236, Pullman, WA, USA

BERKELEY LAB Lawrence Berkeley National Laboratory

NGEE Tropics

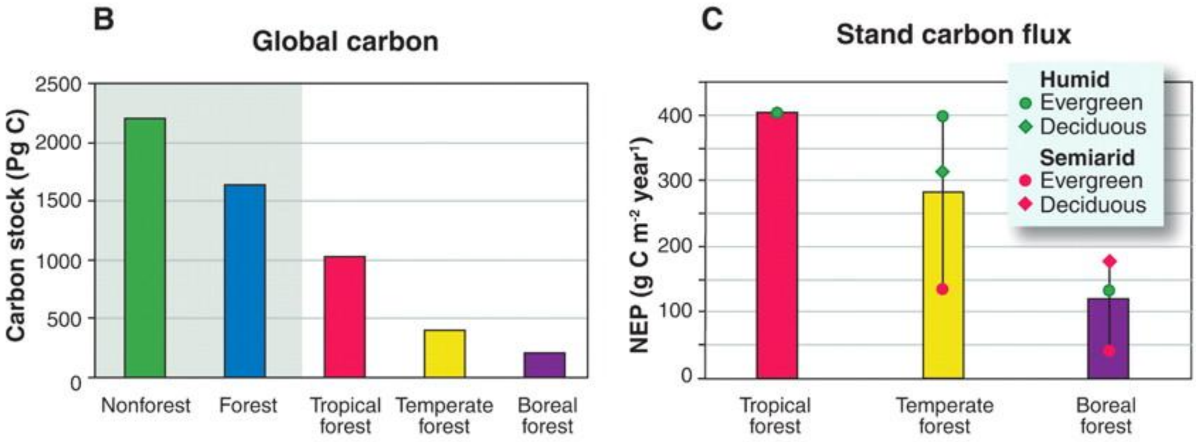
U.S. DEPARTMENT OF ENERGY | Office of Science

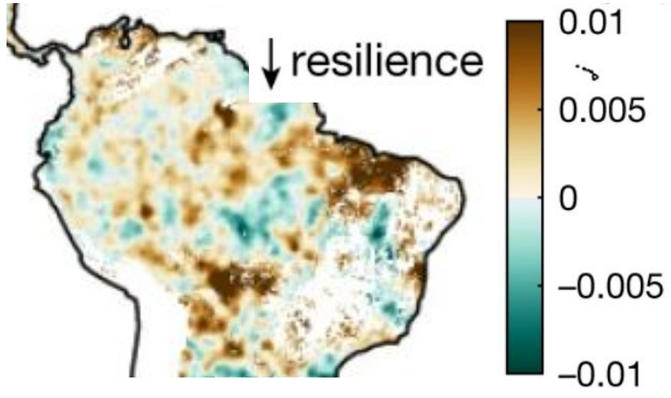
Li, L., Fang, Y., Zheng, Z., Shi, M., Longo, M., Koven, C., Holm, J., Fisher, R., McDowell, N., Chambers, J., and Leung, R.: A machine learning approach targeting parameter estimation for plant functional type coexistence modeling using ELM-FATES (v2.0), EGUsphere [preprint], https://doi.org/10.5194/egusphere-2022-1286, 2023. (under revision)

Tropical forest dynamics are crucial for global carbon cycle

- ~25% of the carbon in terrestrial biosphere
  ~33% of terrestrial net primary production

- Experiencing a significant decline in resilience,
  □increased water limitations and climate variability





- Increasing stress from climate change and deforestation
  e.g., drought, fire, extreme storms

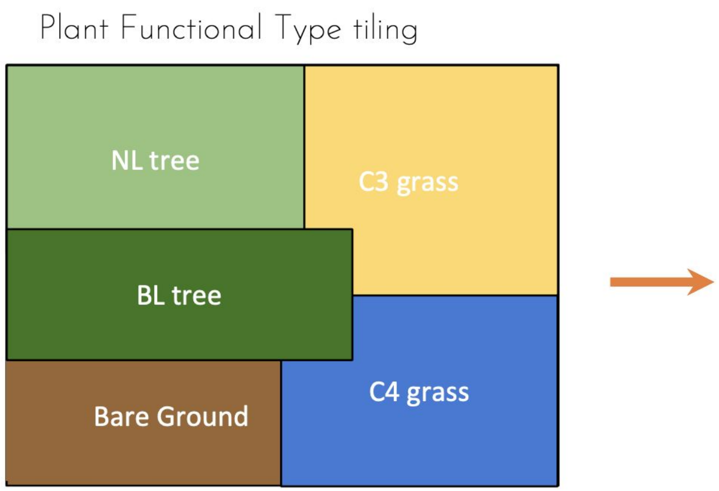
- Better understanding and <u>modeling tropical forest dynamics</u> under climate change

[Bonan et al., 2008; Mitchard et al., 2018; Forzieri et al., 2022]

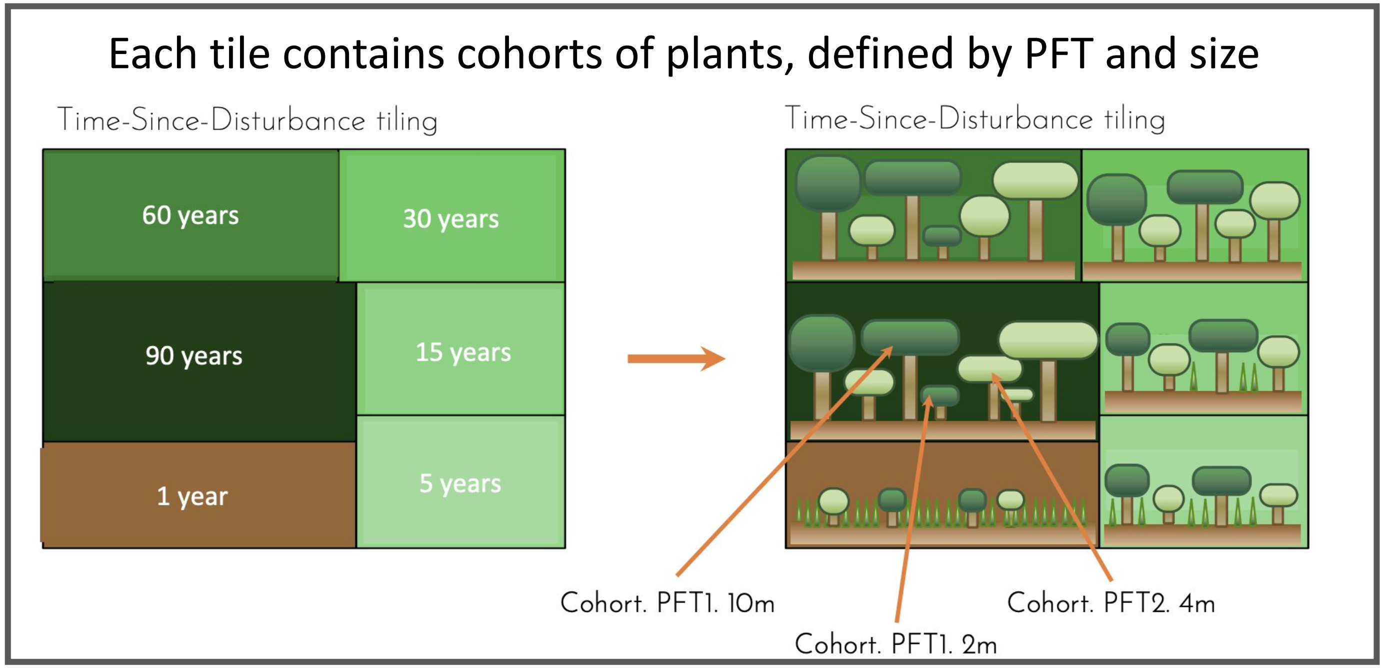
Modeling vegetation dynamics

- Commonly used tools: traditional DGVMs, forest-gap models, and ”cohort-based” models
- Among these tools, "Cohort-based" models have advantages

  Represent sufficient ecosystem dynamics, and maintain relatively high computational efficiency



ELM



ELM-FATES

Represent competition/coexistence between different PFTs

Pacific Northwest NATIONAL LABORATORY

Challenge of coexistence modeling and coexistence theory

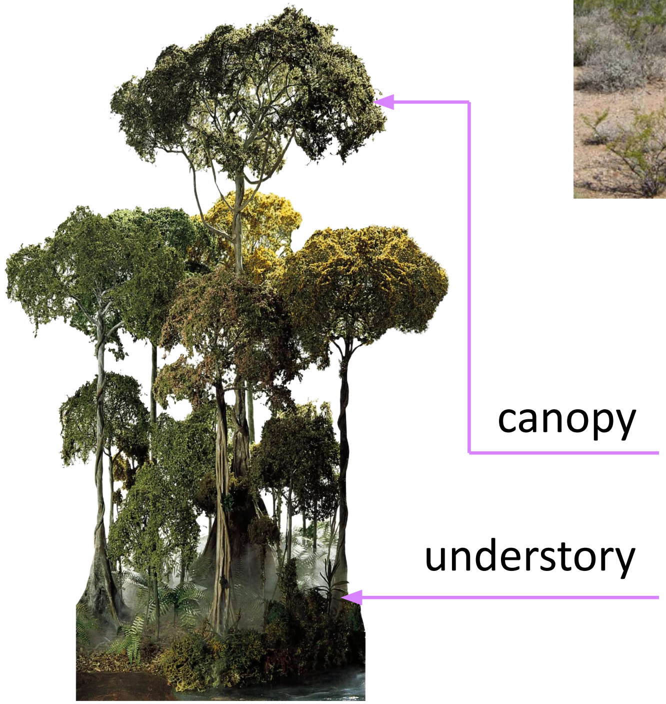
- The challenge in ELM-FATES:
  Reasonably simulate the coexistence of plant functional types (PFTs)

Niche-based coexistence theory

- **Environmental Filtering** → convergence in strategy to adapt to the surrounding environment
- **Niche partitioning** → divergence in strategy to ensure differentiation in resource requirements

Arid Shrub

Humid rainforest



[Kraft et al., 2008; Adler et al., 2013]

Challenge of coexistence modeling and coexistence theory

- The challenge in ELM-FATES:
  Reasonably simulate the coexistence of plant functional types

Niche-based coexistence theory

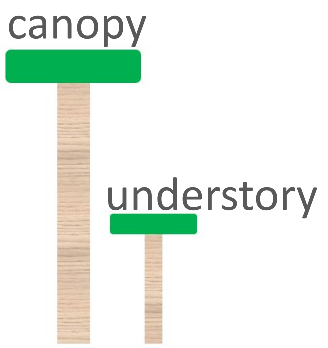
- Environmental Filtering — convergence in strategy
- Niche partitioning — divergence in strategy

ELM-FATES

→ Represent mechanisms/process

→ Reasonable trait parameter values

- C4 vs. C3 PFT



- Vertical canopy structure in ELM-FATES

Limitations in previous studies

Commonly use **filtered ensemble approach** to select parameters

- generate a parameter ensemble
- generate ensemble simulations
- filter out the coexistence runs

Reasonable trait parameter values

- Huang et al. (2020)

  ~1.4%, 70 one-at-a-time experiments before obtaining one reasonable parameter set

- Buotte et al. (2020)

  ~0.3% or 5.5%, two stages of experiments to select optimal parameters

Low efficiency and low percentage of PFTs coexistence experiment !!

Pacific Northwest
NATIONAL LABORATORY

Research goal and testbed

Research goal

Utilize machine learning (ML) to

- alleviate the challenge of modeling PFTs coexistence
- reduce model errors against observations

XGBoost & **SHAP**

Testbed

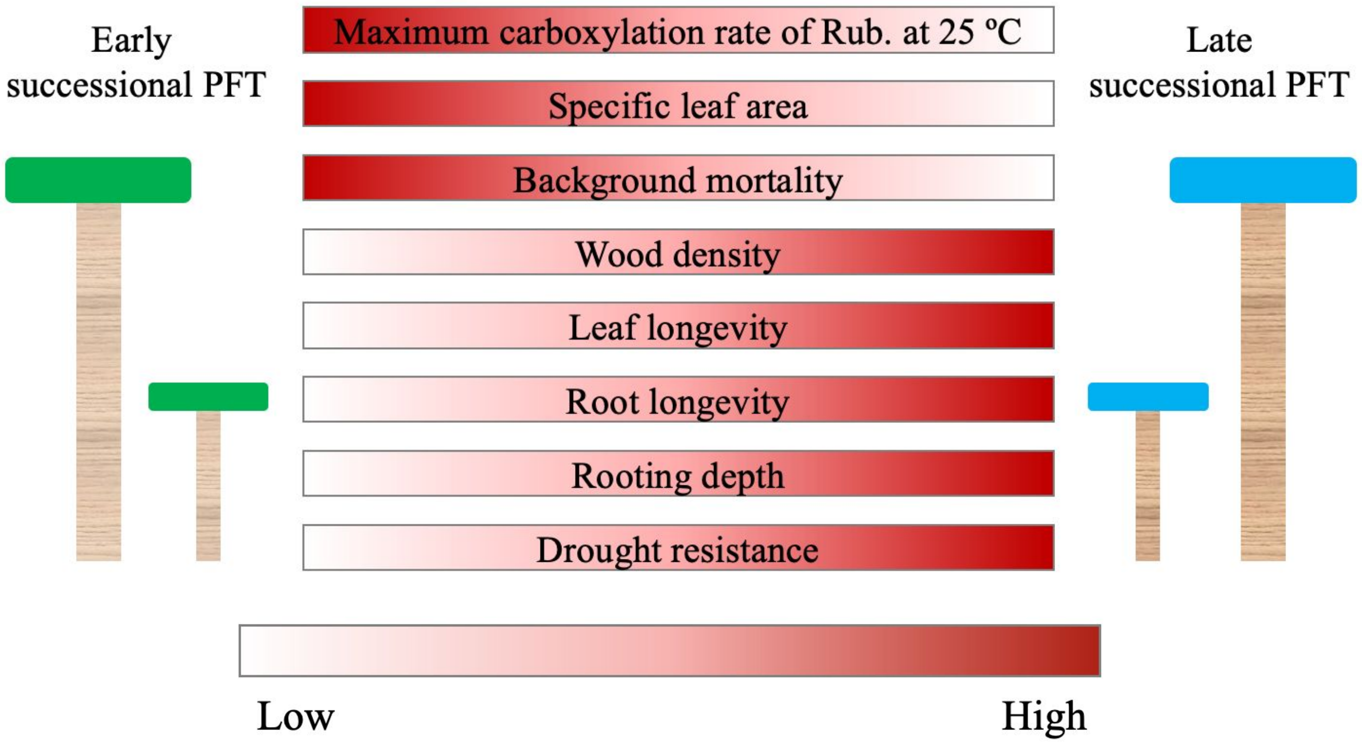
- ELM-FATES
- tropical forest site: Manaus,
- data: meteorological forcing, GPP, ET, SH, AGB

Pacific Northwest NATIONAL LABORATORY

Model configuration

Two PFTs represented in FATES

- Early vs. late successional broadleaf evergreen tropical tree
- represent a primary axis of variability in tropical forests



Pacific Northwest NATIONAL LABORATORY

Model configuration

- 11 trait parameters
- reflect strategic tradeoffs between two PFTs
- trait ranges based on tropical tree measurements

Parameter type	Parameter name	Symbol	Unit	Early PFT	Late PFT	Range
Optimized parameter	Maximum carboxylation rate of Rub. at 25 ºC, canopy top	V_{cmax}	μmol $CO_2/m^2/s$	$V_{cmax,early} > V_{cmax,late}$		40–105
	Specific leaf area, canopy top	SLA	m^2/gC	$SLA_{early} > SLA_{late}$		0.005–0.04
	Background mortality rate	M_{bk}	1/yr	$M_{bk,early} > M_{bk,late}$		0.005–0.05
	Wood density	WD	g/cm^3	$WD_{early} < WD_{late}$		0.2–1.0
	Leaf longevity	L_{leaf}	year	$L_{leaf,early} < L_{leaf,late}$		0.2–3.0
	Maximum size of storage C pool, relative to the maximum size of leaf C pool	CR_{s2l}	—	same		0.8–1.5
Fixed parameter	Root longevity	L_{root}	year	0.9	2.6	—
	Fine rooting distribution profile parameter a	R_a	—	7	7	—
	Fine rooting distribution profile parameter b	R_b	—	2	0.4	—
	BTRAN threshold below which drought mortality begins.	M_{btran}	—	0.4	1.0E-06	—
	Soil water potential at full stomatal closure	$\psi_{closure}$	mm	–113000	–242000	—

Drought resistant:
Late PFT > Early PFT

Pacific Northwest NATIONAL LABORATORY

Overall flowchart and research questions

P1. Parameter sampling

Latin hypercube sampling, and tradeoffs

$V_{cmax,early} > V_{cmax,late}, SLA_{early} > SLA_{late}$

$M_{bk,early} > M_{bk,late}, WD_{early} < WD_{late}$

$L_{leaf,early} < L_{leaf,late}$

P2. Initial FATES experiments

Exp-OBS, consideration of observed trait relationships

Exp-CTR

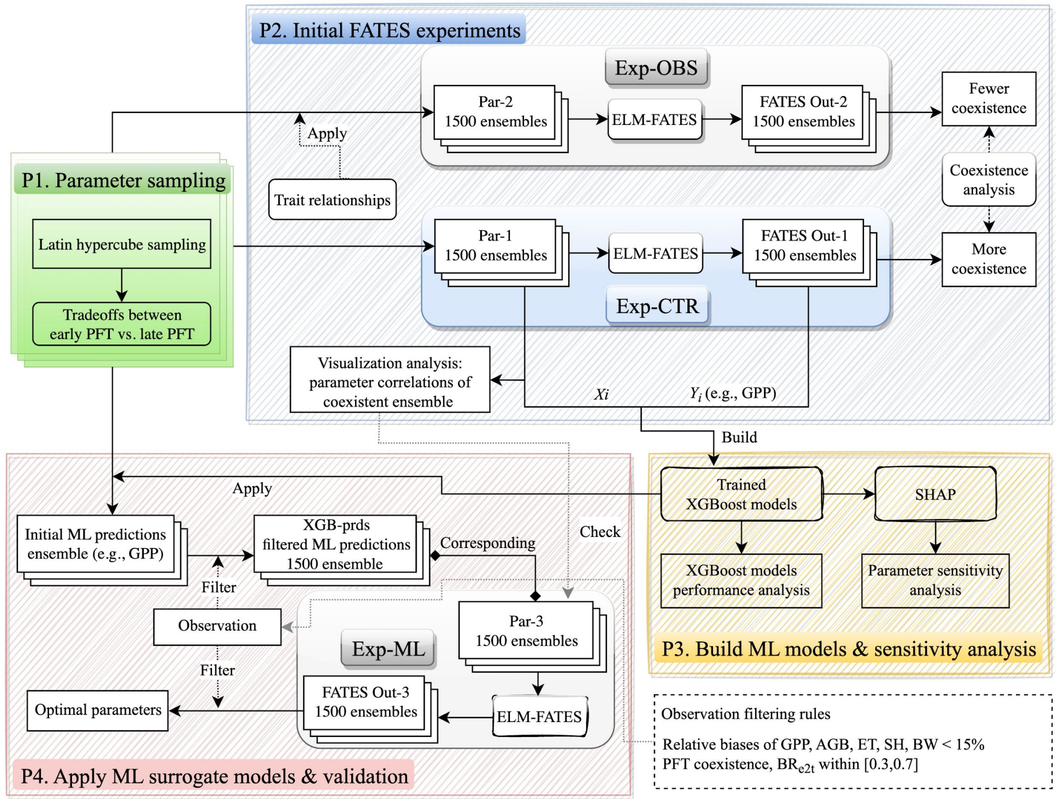
P3. Build ML models and sensitivity analysis

ML models train and test

SHAP importance analysis

P4. Parameter selection and validation

Exp-ML, ELM-FATES simulation using ML selected parameters



Pacific Northwest
NATIONAL LABORATORY

Overall flowchart and research questions

P1. Parameter sampling

Latin hypercube sampling, and tradeoffs

$V_{cmax,early} > V_{cmax,late}$, $SLA_{early} > SLA_{late}$

$M_{bk,early} > M_{bk,late}$, $WD_{early} < WD_{late}$

$L_{leaf,early} < L_{leaf,late}$

P2. Initial FATES experiments

Exp-OBS, consideration of observed trait relationships

<u>Exp-CTR</u>

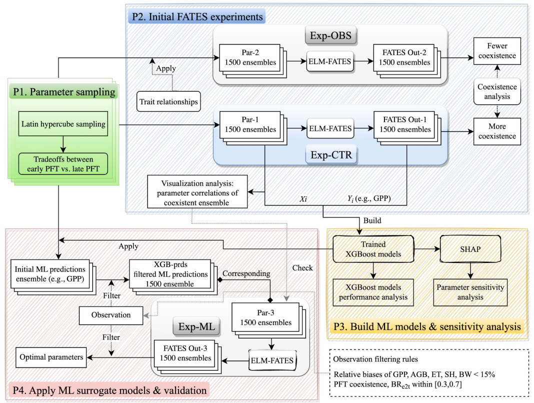
P3. Build ML models and sensitivity analysis

ML models train and test

SHAP importance analysis

P4. Parameter selection and validation

Exp-ML, ELM-FATES simulation using ML selected parameters



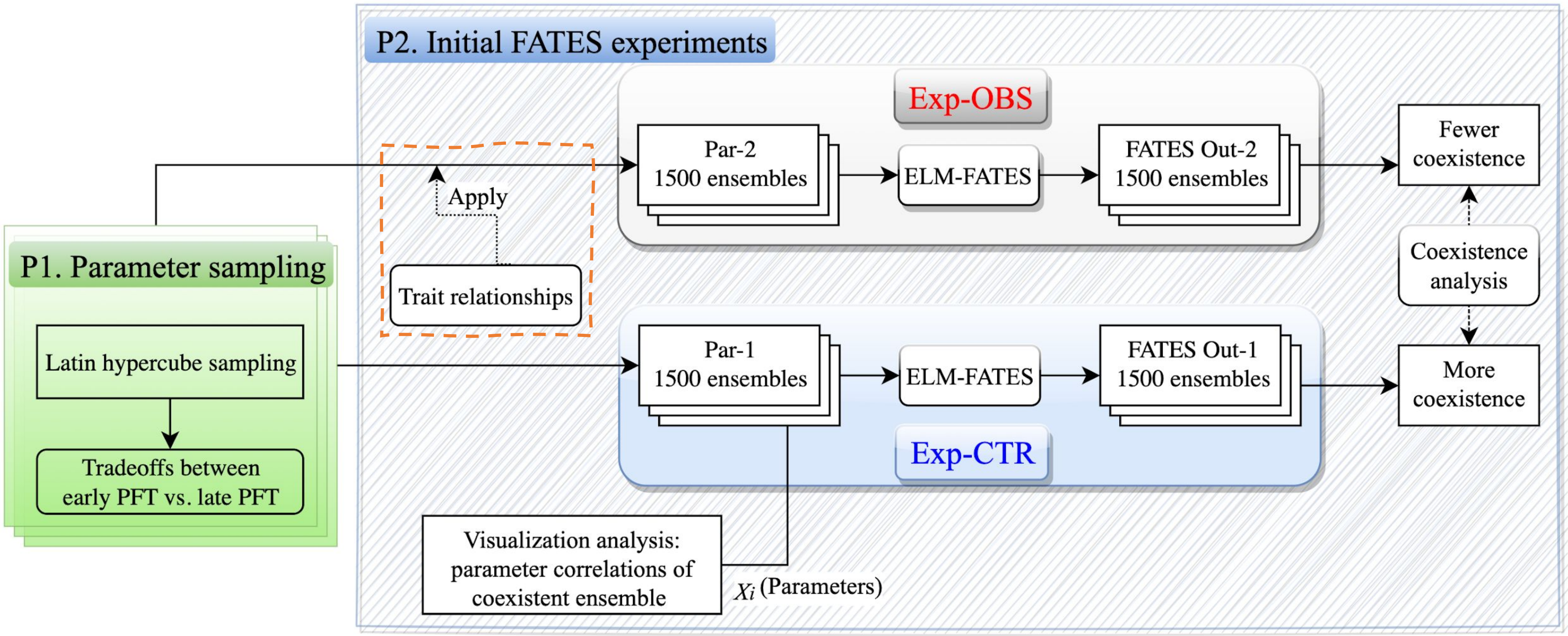
Specific research questions

- Whether observed trait relationships can improve PFTs coexistence?
- Can simple correlations be constructed to improve PFTs coexistence?
- Can ML selected parameter values improve PFTs coexistence

Whether observed trait relationships can improve PFTs coexistence modeling?

Two experiment ensembles

- Exp-CTR, traits tradeoffs
- Exp-OBS, traits tradeoffs + observed trait relationships
- 1500 runs per experiment, 350 years to reach equilibrium state,



Whether observed trait relationships can improve PFTs coexistence modeling?

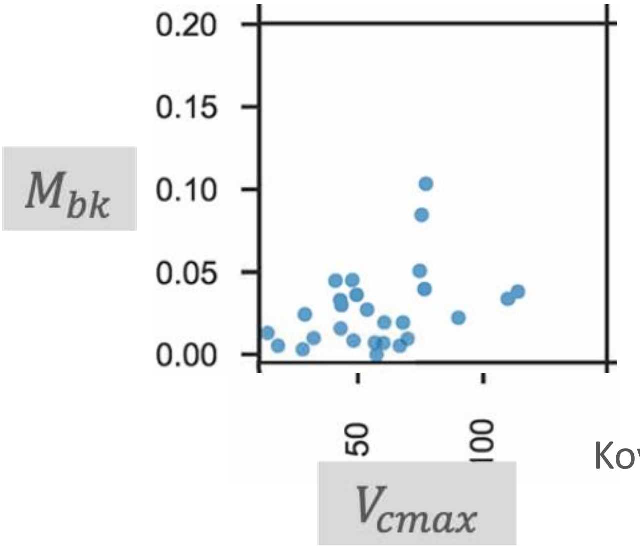
Two experiment ensembles

- Exp-CTR, traits tradeoffs
- Exp-OBS, traits tradeoffs + observed trait relationships
- 1500 runs per experiment, 350 years for each run

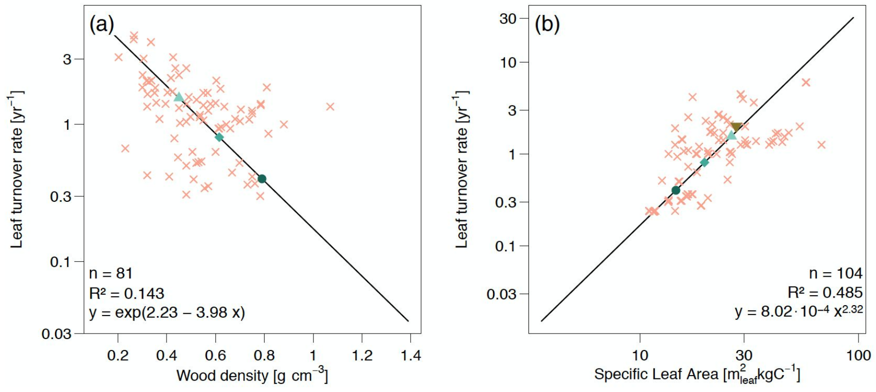
$$M_{bk} = 0.0082 \times e^{(0.0153 \times V_{cmax})} \quad (1)$$

$$L_{leaf} = 0.0001 \times SLA^{(-2.32)} \quad (2)$$

$$WD = -0.583 \times \ln(SLA) - 1.6754 \quad (3)$$



Koven et al., 2020

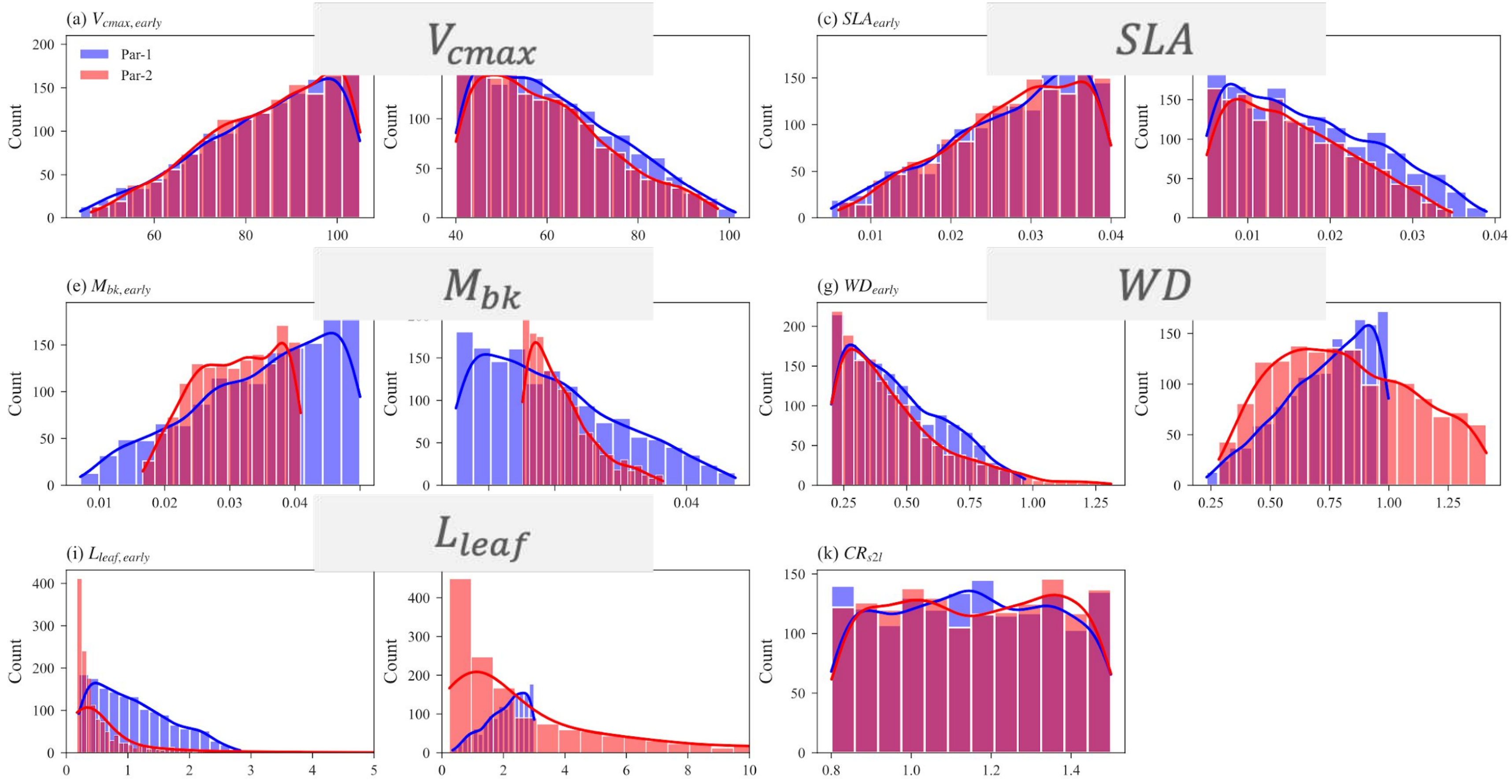


Longo et al., 2020

Whether observed trait relationships can improve PFTs coexistence modeling?

- Exp-CTR, traits tradeoffs
- Exp-OBS, traits tradeoffs + observed trait relationships

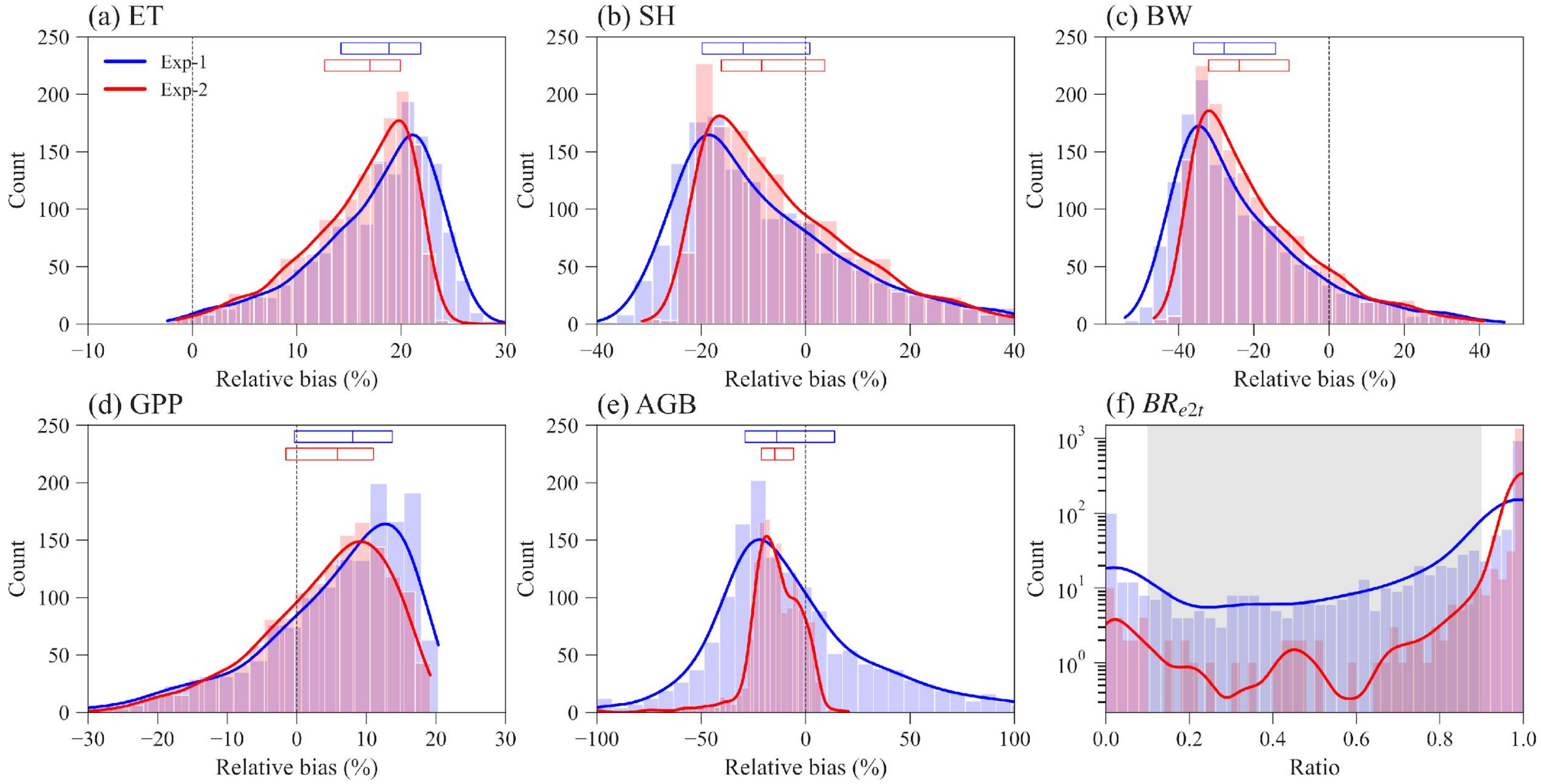
Exp-OBS has narrower M_{bk}, but broader WD and L_{leaf}



Whether observed trait relationships can improve PFTs coexistence modeling?

No

- Exp-CTR, has more PFT coexistence experiments
- Exp-OBS, slight better water carbon and energy simulations, but worse PFT coexistence



PFT coexistence,
Biomass ratio between early PFT and total biomass

$BR_{e2t} \in (0.9, 1.0]$, "early"
$BR_{e2t} \in [0.1, 0.9]$, "coexistence"
$BR_{e2t} \in [0.0, 0.1)$, "late"

Whether observed trait relationships can improve PFTs coexistence modeling?

Why observation constrains do not yield better PFT coexistence ?

1. ELM-FATES limitations

   Implicit representation of trait tradeoff in current ELM-FATES model may not be well balanced, which may differ from the observed trait relationships that lead to coexistence in the real world.

2. Observation data limitation

   Large-scale trait relationships may not reflect the small-scale trait relationships.

3. Simple relationship representation

   The observed trait relationships are based on simplified equations,
   which may not be able to comprehensively reflect tradeoffs between traits.

Koven et al. (2020) and Longo et al. (2020):

$$M_{bk} = 0.0082 \times e^{(0.0153 \times V_{cmax})} \quad (1)$$

$$L_{leaf} = 0.0001 \times SLA^{(-2.32)} \quad (2)$$

$$WD = -0.583 \times \ln(SLA) - 1.6754 \quad (3)$$

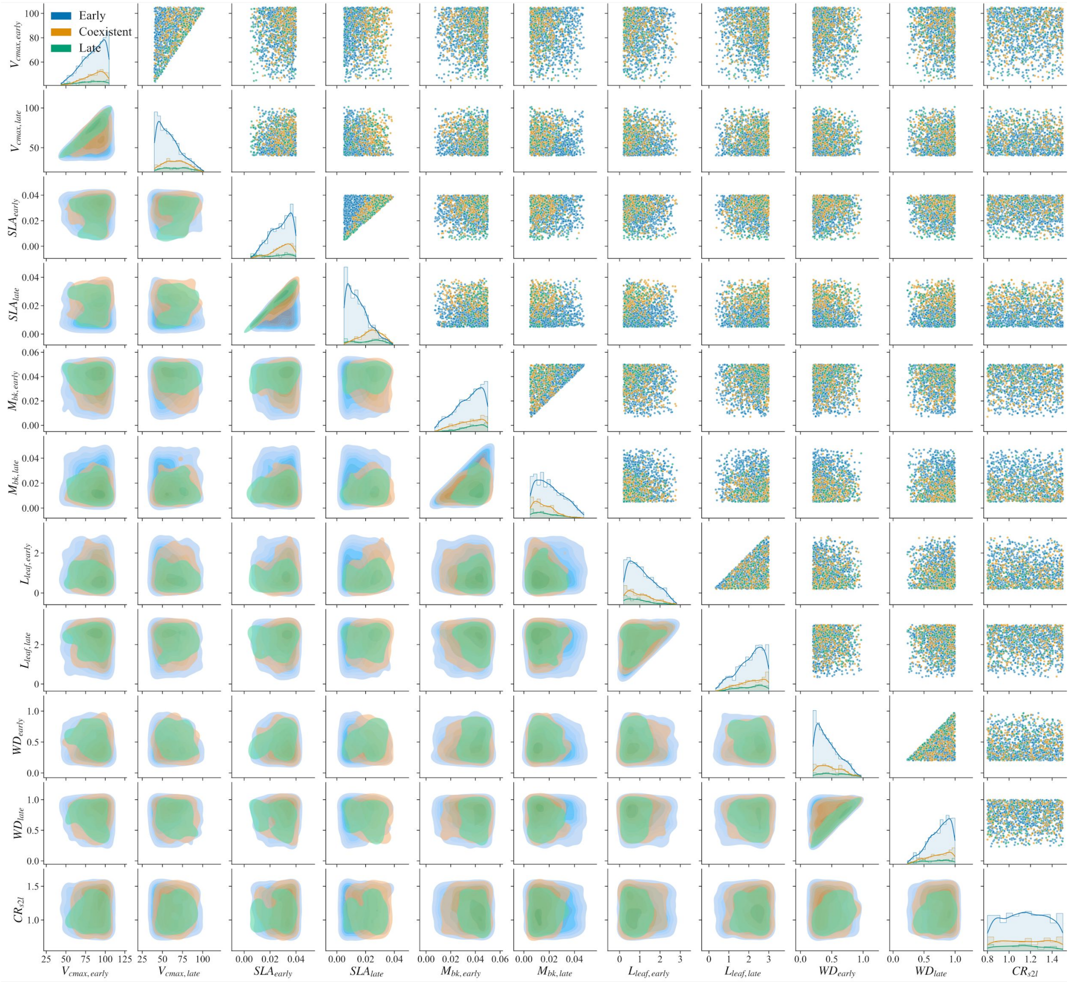
Exp-CTR will be used for the following analysis

Can simple correlations be constructed to guide PFTs coexistence modeling?

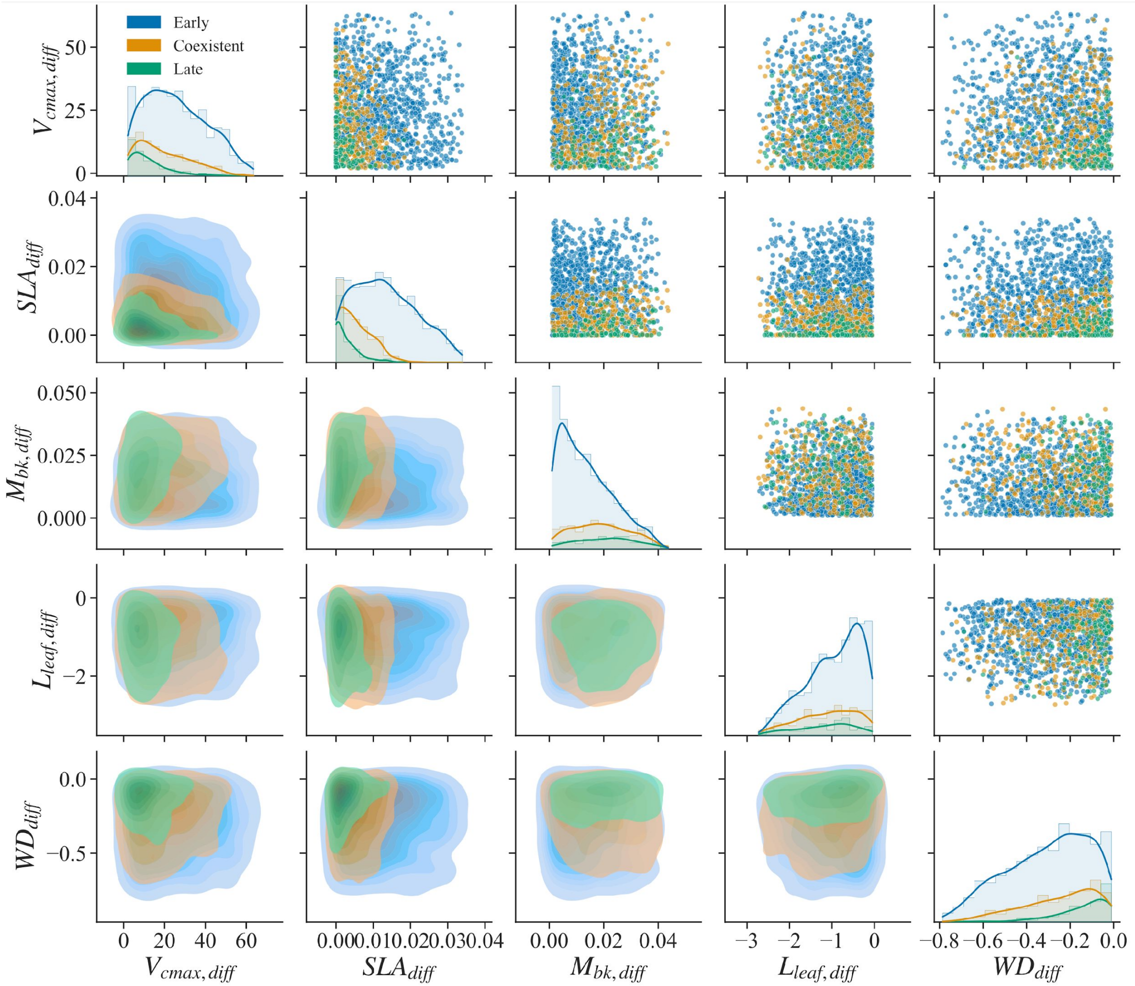
Parameter space of Exp-CTR

● Early
● Coexistent
● Late

Early vs. late parameters



Early–late parameters

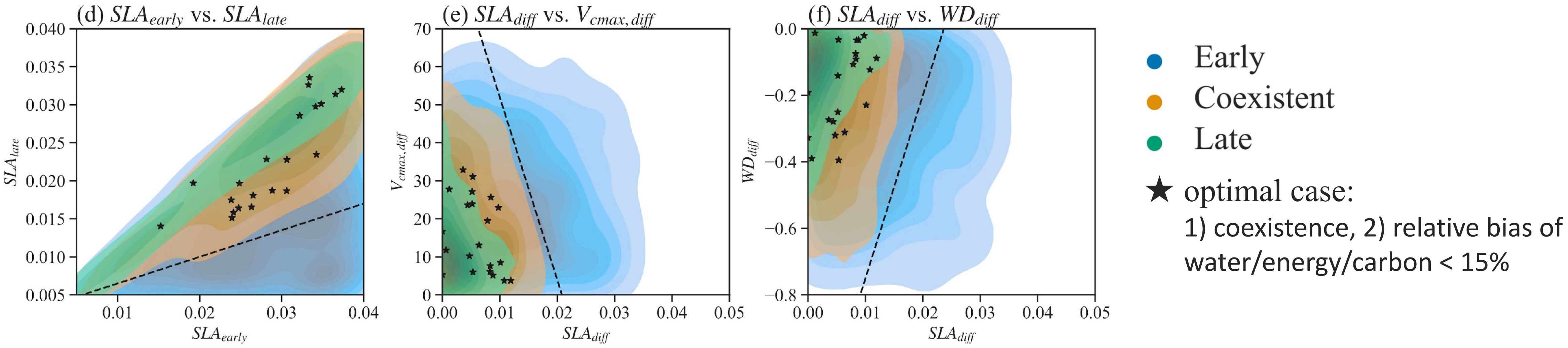


Pacific Northwest NATIONAL LABORATORY

Can simple correlations be constructed to guide PFTs coexistence modeling?

No

Based on Exp-CTR, build empirical simple parameter correlations



- $SLA_{late} > 0.35 \times SLA_{early} + 0.003$
- $V_{cmax,diff} < -4800 \times SLA_{diff} + 100$
- $WD_{diff} > 55 \times SLA_{diff} - 1.3$

→

Within these constrained parameter spaces,

- Coexisting cases increases from 20.6% to 32.6%
- 67.4% is still either early or late
- Optical cases account only about 2.3%

Pacific Northwest NATIONAL LABORATORY

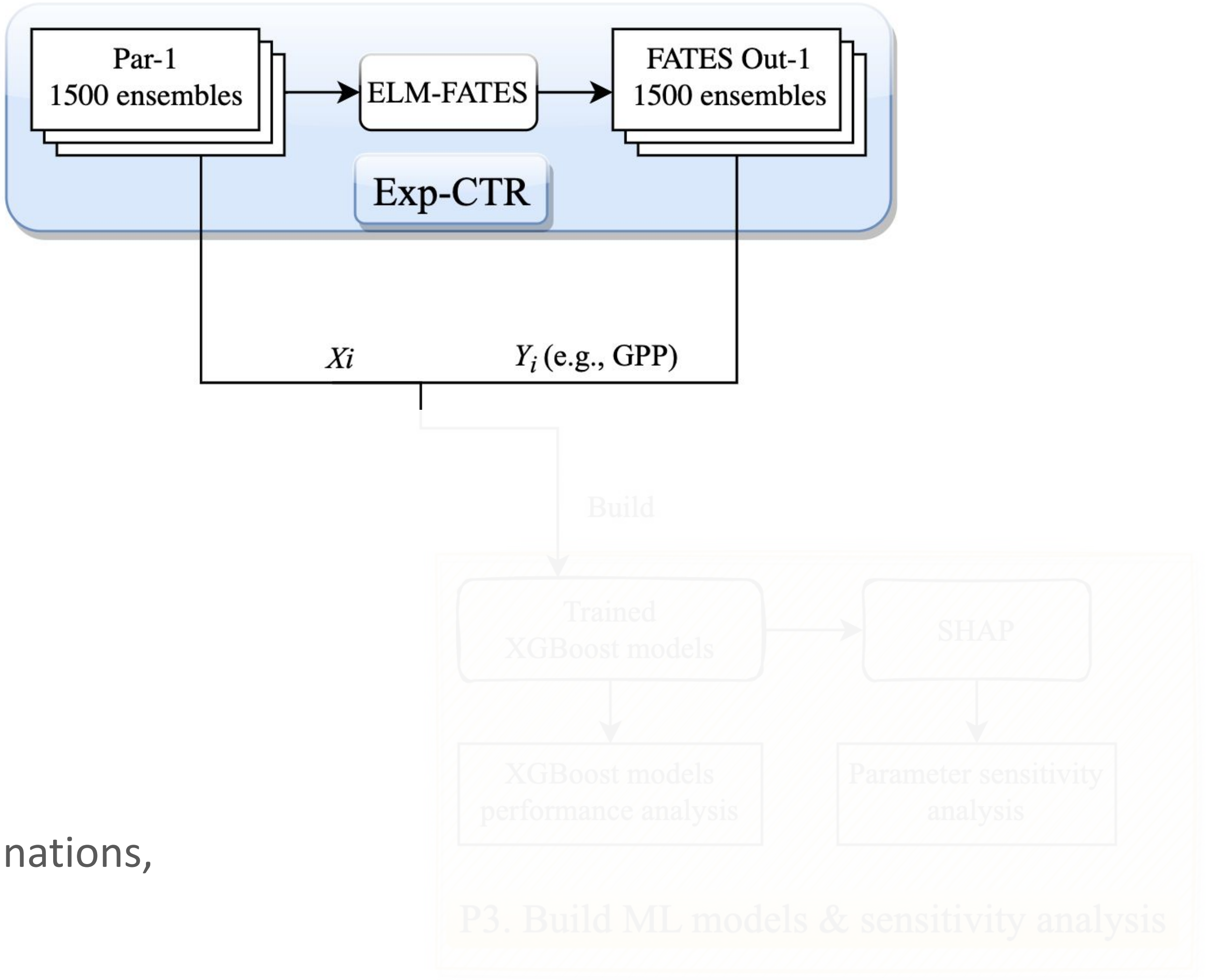
Build ML surrogate models

In Exp-CTR, 1500 samples of

- *Xn*, parameters and their difference

  e.g., $V_{cmax,early}$, SLA_{diff},
- *Yi*, ELM-FATES outputs

  e.g., ET, SH, GPP, AGB, BW



Build emulators

$$Y_i = f_i\,(X_1, X_2, X_3, , \dots)$$

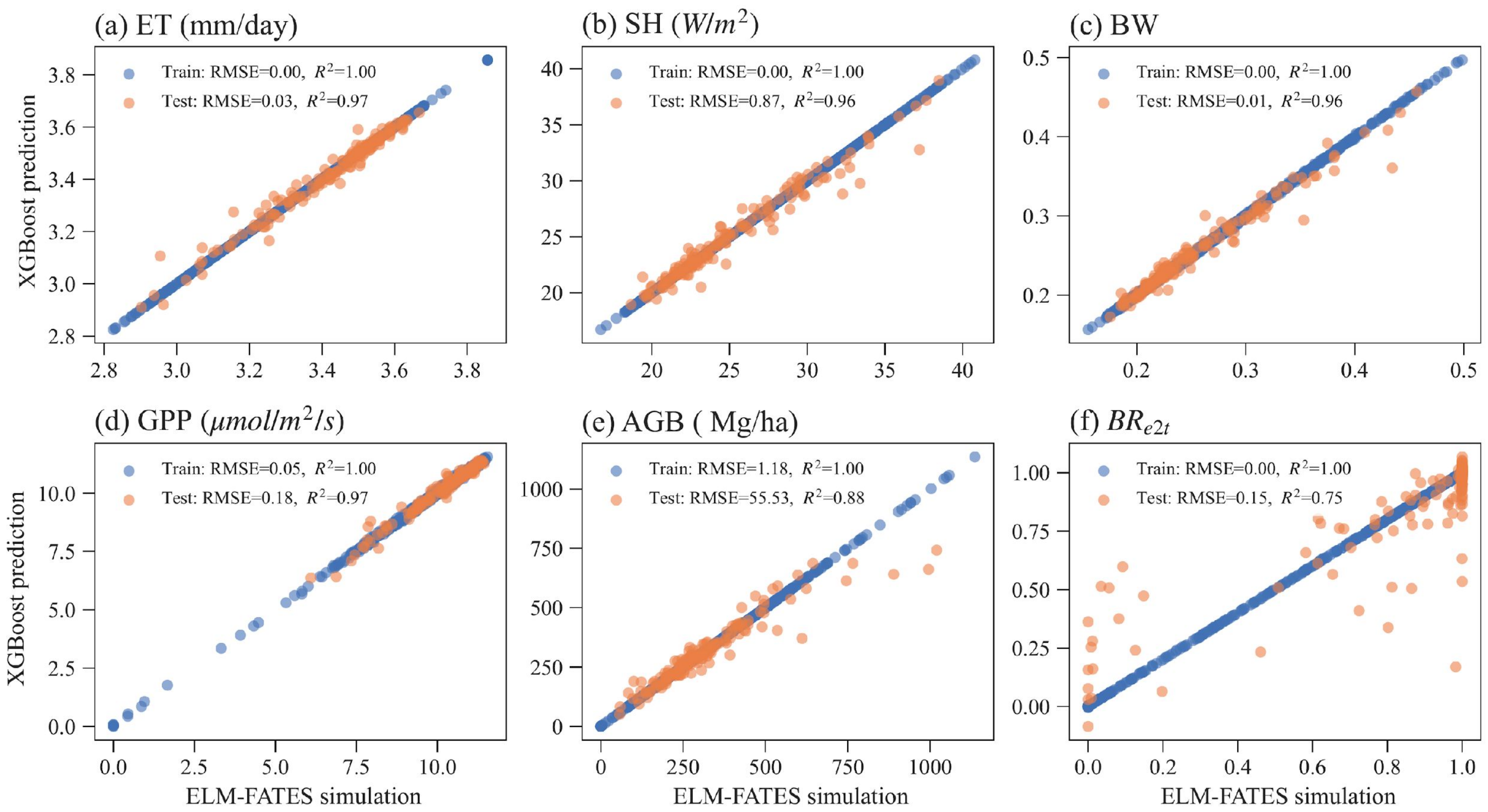
→ Machine learning algorithm
e.g., XGBoost (Chen et al., 2016)

→ SHAP (SHapley Additive exPlanations, Lundberg et al., 2017)

→ Parameters selection

ML surrogate models have good performance

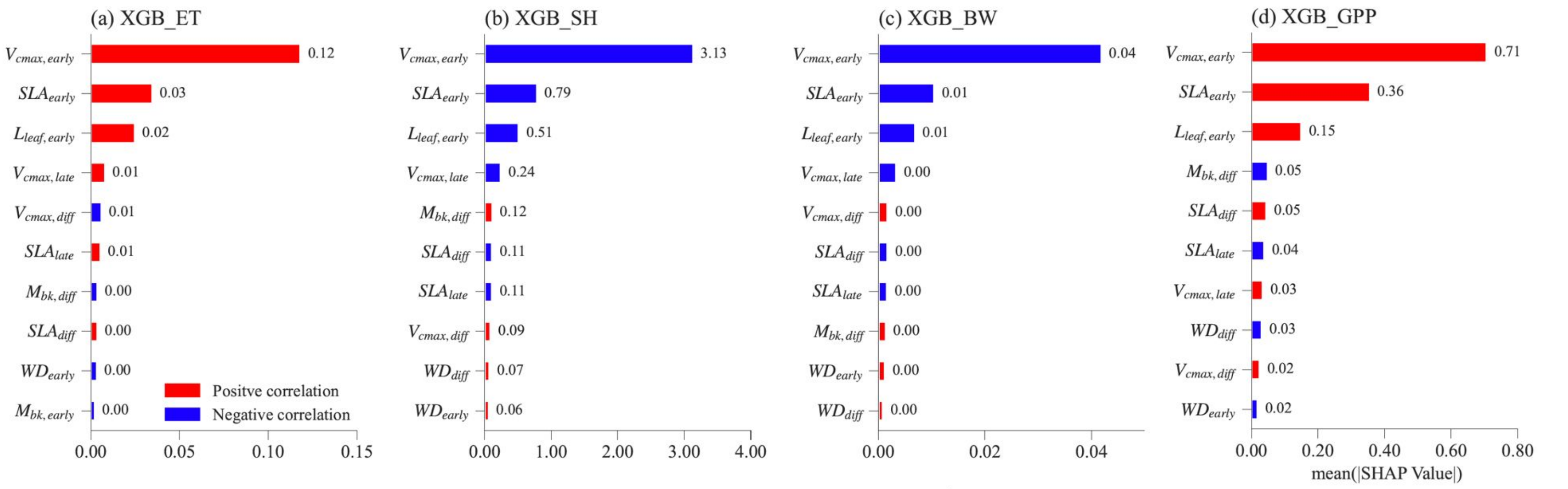
- 6 XGBoost surrogate models: ET, SH, BW, GPP, AGB, and BR_{e2t}
- Overall good performance in training and testing samples
  - AGB and BR_{e2t} are relatively difficult to predict



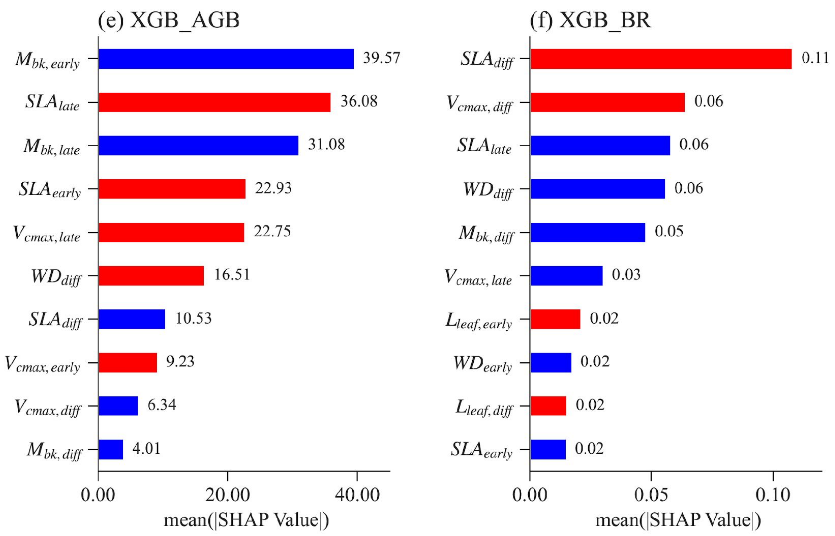
Which parameters are important

- Only 3 features dominate the prediction of ET, SH, BW, and GPP

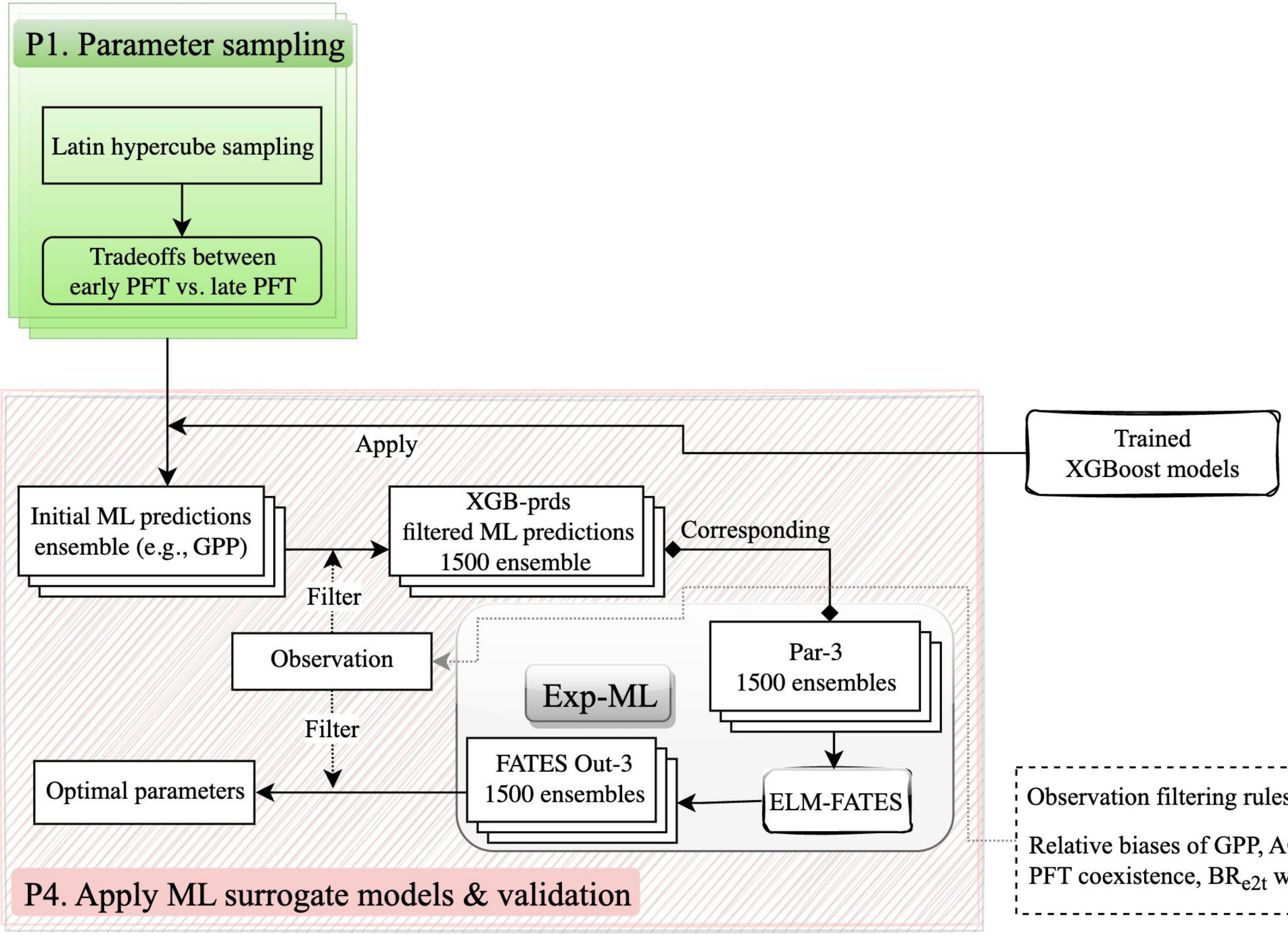
  $V_{cmax,early}$
  SLA_{early}
  $L_{leaf,early}$



- More than 6 features are most important for predicting AGB and BR_{e2t}
- Parameter differences between early and late PFT are very important, e.g, SLA_{diff}, $Vcmax_{diff}$
- Closely related to PFT competition



Parameter selection using ML surrogate models



ET, SH,

BW, GPP,

AGB, BR_{e2t}

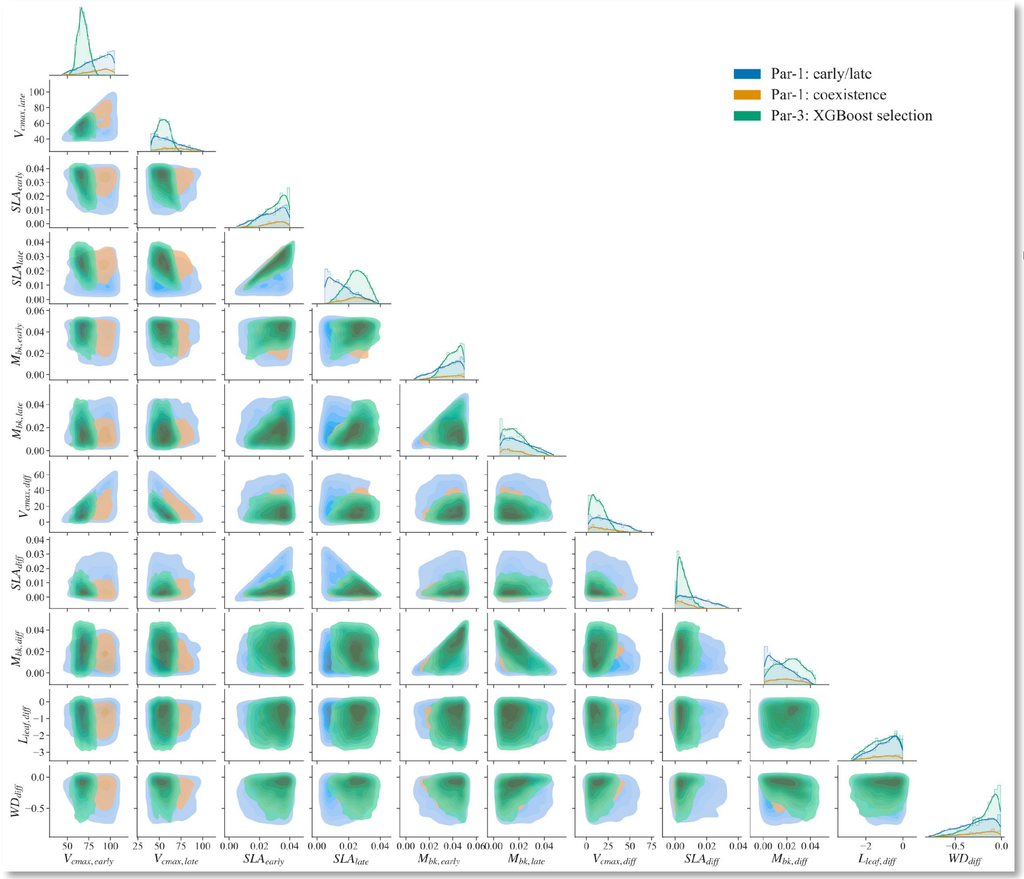
Observation filtering rules

Relative biases of GPP, AGB, ET, SH, BW < 15%

PFT coexistence, BR_{e2t} within [0.3,0.7]

Parameter values selection using ML surrogate models

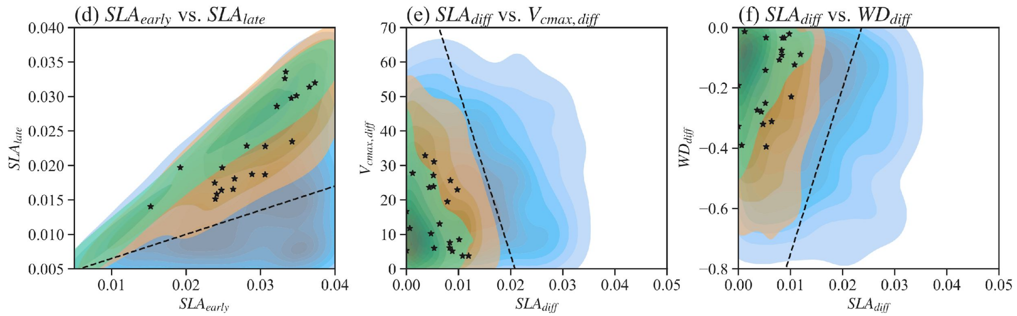
- 99.1% ML selected parameters capture capture the empirical correlations
  - ML surrogate models implicitly learned these simple relationships



ML selected parameters' space

Capture → Empirical correlations of Exp-CTR

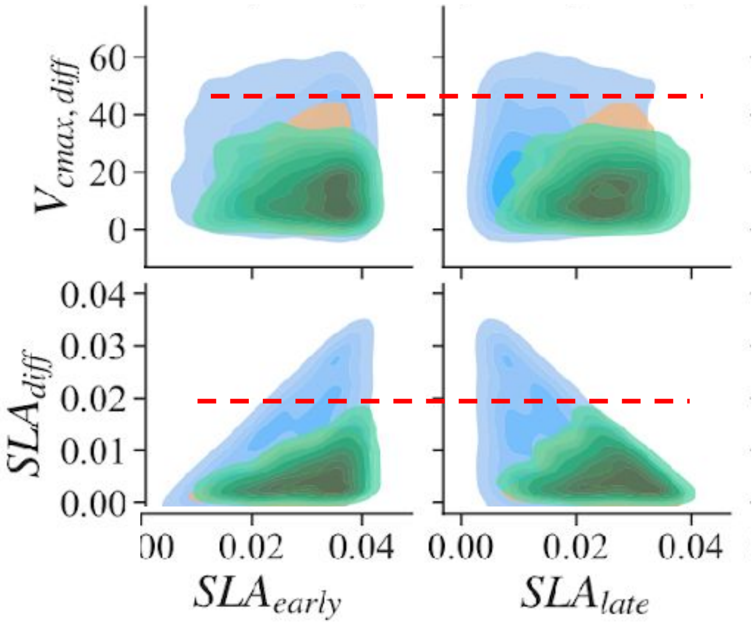
- $SLA_{late} > 0.35 \times SLA_{early} + 0.003$
- $V_{cmax,diff} < -4800 \times SLA_{diff} + 100$
- $WD_{diff} > 55 \times SLA_{diff} - 1.3$



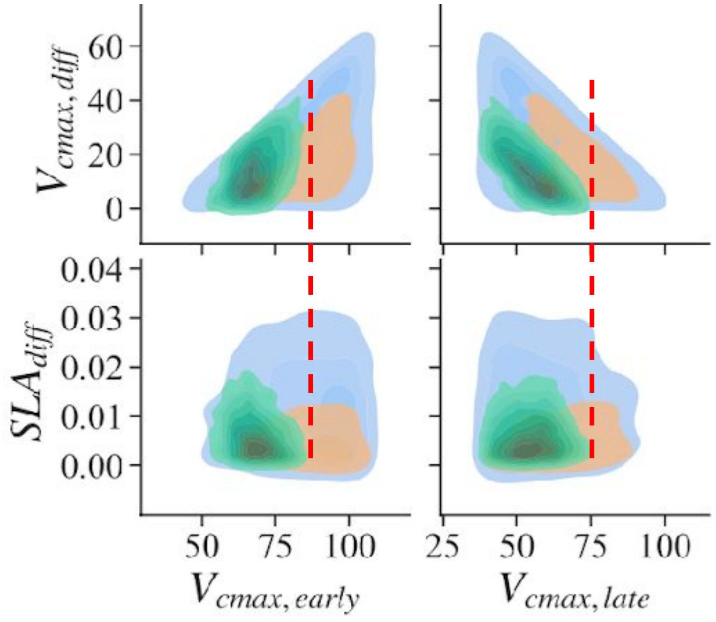
Parameter values selection using ML surrogate models

Comparison between PFT coexistence parameters of Exp-CTR and ML select parameters

- Consistence



- Difference



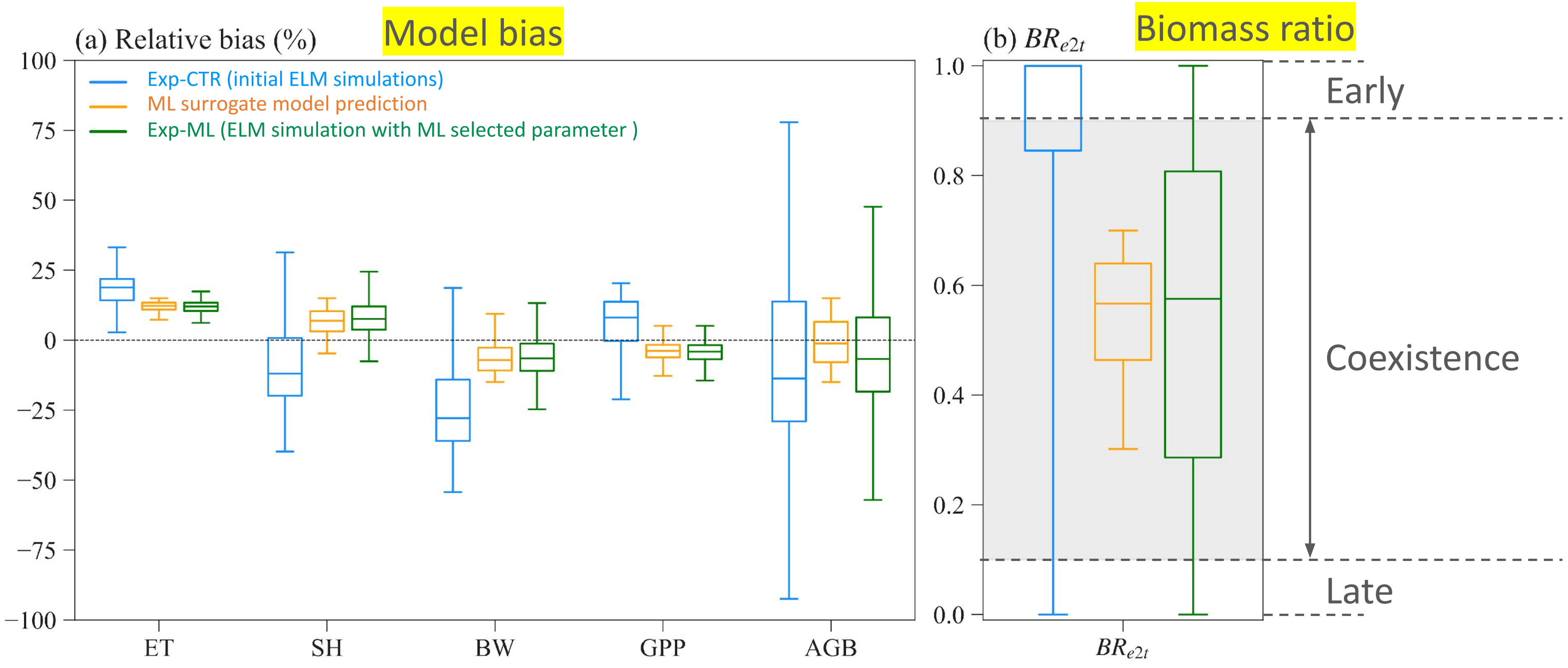
Exp-CTR early/late

Exp-CTR coexistence

ML selection

ML selected parameter values largely improve FATES simulation

- ML selected parameters better capture observations
- ML selected parameters more well-coexistent runs



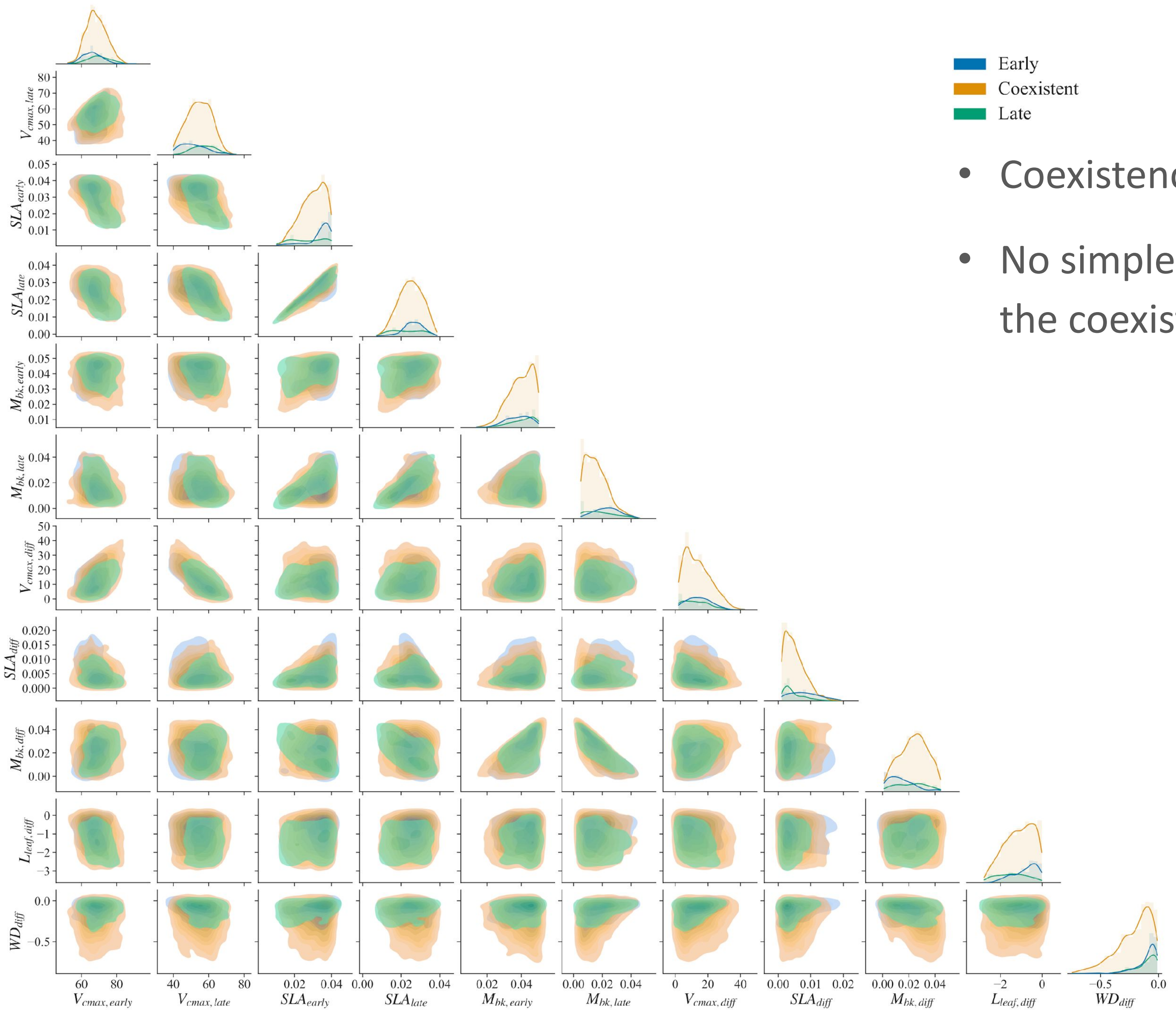
ML selected parameter values largely improve FATES simulation

Compared with Exp-CTR and Exp-ML have

- 3.6 times more coexistence cases, 20% □73%
- 23.6 times more optimal cases, 1.4%□33%, with higher model accuracy

Category	BR_{e2t} ∈[0.1, 0.9	|AGB_bias| < 15%	|GPP_bias| < 15%	|ET_bias| < 15%	|SH_bias| < 15%	|BW_bias| < 15%	Exp-CTR		Exp-ML		Ratio
							count	percent	count	percent	
Late							130	8.7%	174	11.6%	1.3
Coexistence							309	20.6%	1097	73.1%	3.6
Early							1059	70.6%	229	15.3%	0.2
All dead							2	0.1%	0	0.0%	
Total							1500		1500		
Add observation constraints	+						309	20.6%	1097	73.1%	3.6
	+	+					98	6.5%	620	41.3%	6.3
	+	+	+				85	5.7%	618	41.2%	7.3
	+	+	+	+			23	1.5%	572	38.1%	24.9
	+	+	+	+	+		23	1.5%	502	33.5%	21.8
	+	+	+	+	+	+	21	1.4%	495	33.0%	23.6

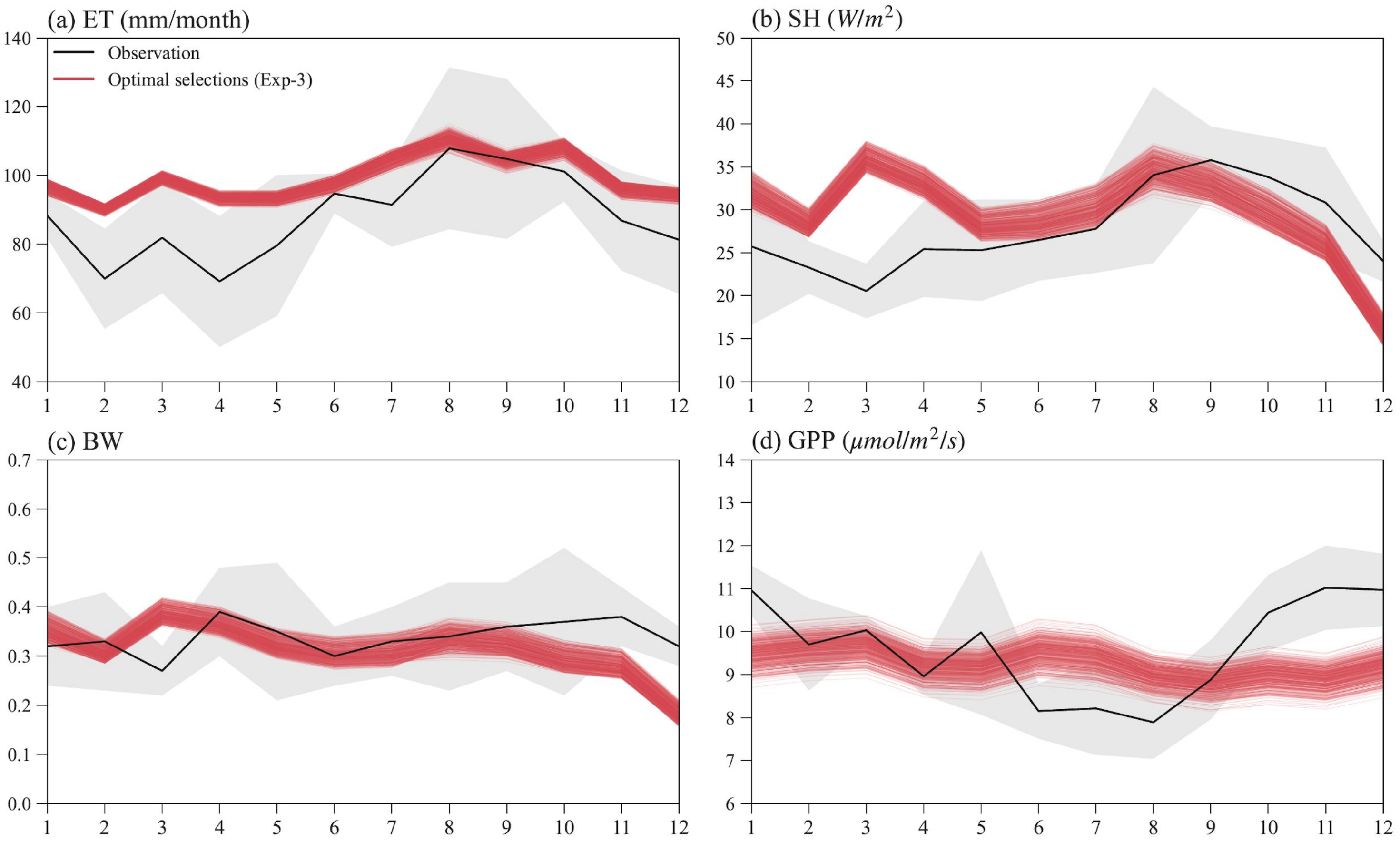
Parameter spaces of Exp-ML



- Coexistence show large overlaps with the early/late
- No simple correlations can be built to distinguish the coexistence from the early and late

ML selected parameter values largely improve FATES simulation

- ML guided optimal simulations reproduces the annual means and seasonal variations of water, energy and carbon fluxes



Pacific Northwest
NATIONAL LABORATORY

Parameter tradeoffs align with niche-based coexistence theory

Environmental Filtering

convergence in strategy

Niche partitioning

divergence in strategy

Niche-based coexistence theory

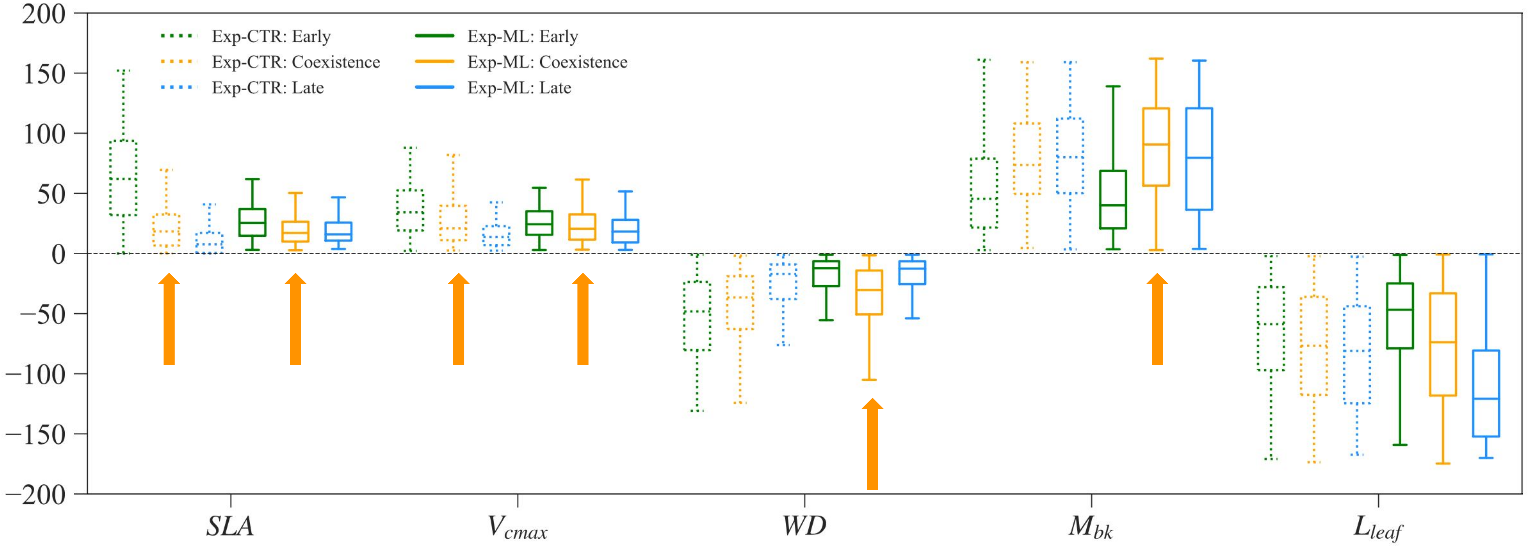
Difference should not be considerable

- Large difference in SLA more likely favors the early PFT

Some degree of differences should exist or balance

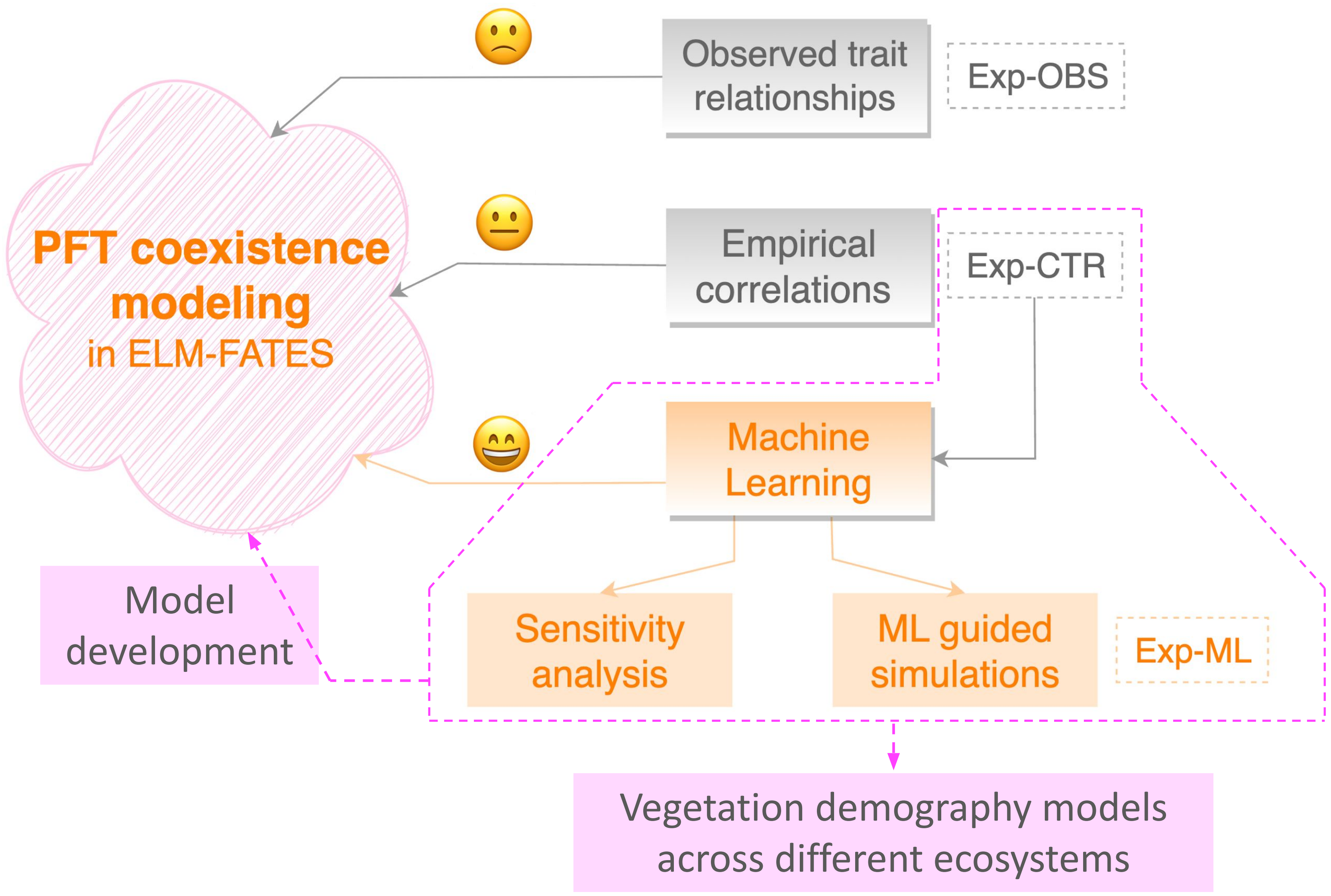
- Small difference in SLA more likely favors the late PFT
- For Exp-CTR, coexistence have intermediate differences in SLA, V_{cmax}, WD, M_{bk} and L_{leaf}
- For Exp-ML, coexistence have intermediate differences in SLA, V_{cmax}, and L_{leaf}

M_bk and WD show large difference but they show tradeoff to make coexistence



Parameter relative difference (%) between early PFT and late PFT

Take Home Message



Li, L., Fang, Y., Zheng, Z., Shi, M., Longo, M., Koven, C., Holm, J., Fisher, R., McDowell, N., Chambers, J., and Leung, R.: A machine learning approach targeting parameter estimation for plant functional type coexistence modeling using ELM-FATES (v2.0), EGUsphere [preprint], https://doi.org/10.5194/egusphere-2022-1286, 2023. (under revision)

Acknowledgement

This research was supported as part of the Next Generation Ecosystem Experiments-Tropics, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research.

NGEE–TROPICS
NEXT-GENERATION ECOSYSTEM EXPERIMENTS–TROPICS

U.S. DEPARTMENT OF ENERGY | Office of Science