https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

The Calibrated Rapid Assimilation and Forecasting Technique (CRAFT) for Earth system and ecological modeling using machine learning and Bayesian estimation

Zachary J. Robbins[1], Lucas Tiede[1], Charlie Koven[2], Ryan Knox[2], Nate McDowell[3], Chonggang Xu[1].

1 Los Alamos National Lab, Los Alamos, NM, USA

2 Lawrence Berkley National Lab, Berkley, CA, USA

3Pacific Northwest National Lab, Richland, WA, USA

Correspondence to: Zachary Robbins (zjrobbins@lanl.gov)

Abstract. Increasing the mechanistic complexity of Earth system and ecological model provides the opportunity for improved understanding with numerical experimentation. However, complexity additionally presents greater difficulty in constraining parameters with data. Determining plausible parameter combinations requires a method by which to incorporate data streams, field observations, and their uncertainty. Bayesian methods of integrating datasets are often limited by the computational limits in running these complex mechanistic models. Machine learning can effectively integrate data and physical models by constructing emulations for the finite simulations needed for parameterization. We present the CRAFT (Calibrated Rapid Assimilation and Forecasting Technique) framework for ecological model parameterization and test it using the mechanism rich ecosystem demographic model, FATES-HYDRO (testing 42 parameters and evaluating it for 6 outputs). This framework uses emulation and parameter reduction to construct more rapidly running emulators and test posterior parametric distribution given observational data. We assess whether this mechanism can emulate the model outputs, the variance across the parameter space, and in future prediction (simulations 2020-2100,) using synthetic model runs. Overall random forest models had an out-of-sample accuracy of 92-99% in reconstructing observational periods and showed no-significant difference with the physical model for change in most parameters (283/288 parameter and output combinations). 95% CI posterior ranges of parameters produced FATES-HYDRO runs that had an RMSE for gross primary productivity (GPP) of 3.748 g C month^{-1}, for evapotranspiration (ET) 1.33 mm H_2O month^{-1}, for soil moisture 0.005 m^{2}m^{-2}, 0.381 MPa for maximum leaf water potential (LWP_{max}), 0.44 MPa for minimum leaf water potential (LWP_{min}), 1.80 m^2 m-1 for runoff (RO) when compared to the synthetic data. Future simulations had a RMSE for GPP of 22.55 gC m^{-2}month^{-1}, ET had a RMSE of 7.82 mm H_2O month^{-1}, RO had an RMSE of 88.80 mm H_2O month^{-1}, monthly leaf water potential had an RMSE of 0.145 MPa, soil water content at 20cm had an RMSE of 0.0138 m^{2}m^{-2} when compared to the synthetic dataset. Overall, we show the CRAFT framework as a rapid and accurate semi-automated method to assimilate data and calculate

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

posterior distributions in complex physical models. This framework could accelerate our scientific discovery through rapid accuracy improvement in process-based modeling and more mathematically robust prediction with constrained uncertainty in model parameters.

1 Introduction

Ecological and Earth systems models are important tools for scientific understanding, natural resource planning and hazards risk assessment (Getz et al., 2018, Schuluter et al., 2019). To make these tools robust for decision making, we must capture the uncertainty in model predictions (Kochenderfer and Mykel 2015). Sources of uncertainty can include model representation of process, external forcings, choice of parameter values, and the initial simulation conditions, all which influence representative and projection uncertainty (Dietze et al., 2017). More so than first principle physical models, ecological models suffer from the dual problem of parameter identifiability (given uncertainty in mathematical representation) and equifinality (differentiating multiple correct conclusions from varied initial conditions and parameter structures). Increasing the number of mechanisms in ecological and land surface models is posited as a method by which models can capture complex system behavior across wide environmental conditions, but this increases the burden on choosing the right parameter values and initial state conditions, making many analytical solutions intractable (Saltelli 2019; Pal and Sharma, 2021; Dietze 2017). Traditional model calibration can constrain models for a finite number of metrics and parameters, but is usually hampered by computational limitations, which prevent conducting a larger number of serial optimizations runs.

Ecological and Earth system parameters can include empirically observed properties and relationships, and often scalers for unquantified or undetermined processes (Fischer et al., 2018). Many modeling exercises require that important parameters be established a priori either from heuristics, experimentation or sensitivity analysis (Dagon et al 2020; Buotte et al.,2020; Koven et al., 2020). Manually tuning parameterization is often implemented to match model performance to observational data, though this is inefficient and ultimately under-represents the uncertainty in the system by minimizing diversity in simulations (assuming equifinity). Thus, large ensemble model simulations with random parameter values drawn from prior ranges are proposed to be constrained by observations (Buotte et al., 2020; Chitra-Tarak et al. 2021; Robbins et al., 2024). However, the criteria for the selection of ensemble members are often arbitrary and normally only a small number of viable ensemble members are used as the entire bound of uncertainty.

Bayesian probabilistic inversion approaches based on Markov-Chain Monte Carlo (MCMC) simulations are the leading approach to due to the lack of analytical solutions (Green and Worden, 2015; Lu et al., 2017). This method provides robust methods for estimating parameter uncertainty through the establishing of a posterior distribution which can then be propagated in future simulation (Lu et al., 2017). However, they normally require millions of sequential iterations to reach convergence in sampling replicates making whole model computational cost impossible and only portions of models can be

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

estimated (Hararuk et al., 2014). Rapidly simulating the outputs of these complex models using machine learning emulation could be a solution to mitigate this computation requirement.

Machine learning is the field of study that enables models to learn and improve from data with minimal formal structure in relationships. Machine learning models can accelerate finite simulation through rapid parameter search in Bayesian and optimization algorithms (Sawada, 2020; Dagon et al., 2020) and understanding primary multivariate and derivative relationships in complex models (Gao et al., 2020, Nonnenmacher et al., 2021) and can be used to reduce dimensionality through recognition of the influential parameters (Lindardatos et al., 2021 but see Strobl et al., 2007). While ML models can identify complex patterns and make accurate predictions within the confines of their training data, they may fail to generalize under non-analog conditions (Luo et al., 2021). However, mechanistic model parameter optimization is generally constrained by the limited temporal and conditional scope of observational data, so the need for a machine learning model to be generalized is much lower than under conditions of prediction, where conditions wholly unseen by the model must be accounted for.

Given the growing need to propagate and represent uncertainty within these models, we present a high-performance computing framework for rapid model emulation and Bayesian parametric uncertainty assessment called CRAFT (Calibrated Rapid Assimilation and Forecasting Technique). We apply this to the FATES-HYDRO model (Xu et al., 2023) an eco-physiology-based plant demographic model consisting of > 40 parameters per plant functional type (PFT) based on the FATES plant demography model (Fisher et al. 2015, Moorcroft et al. 2001). This model takes on average ~ 4 hours to run a relatively short (20-years) single-site parameterization, meaning iterative parameterization can take weeks to months. Here, we provide a method using the strength of Bayesian parameter estimation with a drastically reduced time required to run the model sequentially through machine learning emulation. This provides us with the tools to define the posterior distribution of parameters for a given site in a manner that is rapidly reproducible and methodologically robust.

2. Methods

The CRAFT framework is a model-agnostic Bayesian estimation of posterior parametric distributions using machine learning emulators to provide the computational speed missing in classical numerical models. The CRAFT framework does not aim to emulate the entirety of a numerical model or make it fully generalizable, but rather to emulate individual simulation performance that allows for more rigorous parameterization processes. The process flow is shown in (Fig. 1). First, a prior range of parameters for a physics-based model is established in literature or from heuristics. This is used to construct a Latin hypercube sampler to test the range of the outputs from the physical model. The physical model is run under these sampled parameter values, and the corresponding outputs of interest are used to construct the emulators. Using the Shapley importance (Shapley, 1953) within the emulator, the important parameters are selected for model calibration. This step is important as it can allow us to capture the critical parameter-output relationships to reduce the impact of multi-

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

parametric unidentifiability in MCMC simulations (Massound 2019). Finally, MCMC simulations use the emulators to establish posterior distributions of the parameters to fit the observational data for the site.

FATES is an ecosystem demography model structured by plant size and succession stage (age since last disturbance). For each age-since-last-disturbance 'patch', the model tracks many 'cohorts' consisting of populations of plants of similar sizes belonging to the same PFT. For each cohort, the model provides numerical solutions to biophysical processes, including photosynthesis, respiration, allocation of carbon to different plant organs, water and nutrients, growth and competition, and the likelihood of mortality based on carbon starvation, hydraulic failure, logging, fire, and baseline background disturbances. FATES allocates carbon by photosynthesis to various vegetative pools (leaf, stem, seed, roots, storage). The variation in plant traits that control these processes allows for varied plant carbon–water economic strategies. An optional configuration of FATES, FATES-HYDRO, incorporates plant hydrodynamics to simulate soil-to-atmosphere water flow across different plant organs(absorbing root, transporting root, stem, and leaf) and the corresponding water storage in them for each cohort of trees (Christoffersen et al., 2016; Xu et al., 2023). Explicitly simulating storage in plant organs enables the direct representation of loss of hydraulic conductivity based on tissue water content change, which allows us to estimate the risk of hydraulic failure and the resulting risk of mortality. The water flux is calculated based on water pressure gradients across different plant compartments (rhizosphere, absorbing roots, transporting roots, stem, and leaf; Christoffersen et al., 2016). The water potentials for specific tissues are calculated from relative water content based on three stages of water tissue drainage.

To test the CRAFT method, we used the FATES-HYDRO model at Barro Colorado Island, Panama. We constructed initial parameter ranges from Xu et al., 2023, using published meta-analyses on plant trait parameters required for the FATES-HYDRO model for tropical species. These include species level hydraulic plant traits including resistance to embolism, minimum and maximum water storage, overall tree architecture, stomatal response to drought, leaf photosynthetic and other structural traits (Table S1). Where insufficient measurements were available to construct a statistical distribution, a uniform distribution was used. The sampled parameter values are based on Latin hypercube sampling from the fitted statistical distributions. We ran 2,000 parallel ensemble simulations of the mechanism-based model FATES-HYDRO. These simulations serve as the baseline for the relationships between the input and output variables, the site conditions, climate drivers and the outputs we aim to emulate. These simulations were run using a single PFT. The size structure was initialized from inventory observations and held constant subsequently, so that plant physiology was decoupled from ecosystem structure.

To construct emulators for each output of interest, including gross primary productivity (GPP), evapotranspiration (ET), soil moisture at10cm (H2OSOI), minimum leaf water potential (LWPmin), maximum leaf water potential (LWPmax) and runoff (RO), the model run outputs are structured as data frames with labeled dates (day-of-year and year for LWP, and month and year for all others). These data frames are assigned to the composite input of the 42 parameters and labelled dates as predictors of the output. We perform a split test, where emulators are trained using 70% of the dataset. The trained model is then tested using the rest of the dataset (30%). We tested several emulator types 1) classic random forest, 2) gradient

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

boosted tree regression, 3) artificial neural network, and 4) Gaussian process models, using the scikitlearn python package (Pedrogosa et al., 2011). Each output variable's score was compared to determine the best emulator moving forward. To limit the difficulty in fitting posterior trait distributions, each emulator was tested using Shapley outputs (which ranges between 0-1 based on parameter influence on the final model) to determine the relative contribution of each input variable. We removed variables that scored lower than 0.1 in all output emulators. Then using this reduced set of input variables, models were refit to ensure that the performance was comparable. To understand the emulator's performance in response to the input variables we fit splines to the relationship between each input variable and each output variable for both the process-based FATES-HYDRO and its emulators. These splines were then statistically tested against one another using a Kolmogorov-Smirnov test to determine whether they came from similar distributions. This assesses not just whether the primary relationship is captured by the emulators but whether the Jacobian relationship between changes in parameters are reflected in changes in model outputs of interests. This process was repeated for differences in dry season and wet season values, due to lack of difference in entire year distributions.

The goal of the CRAFT framework is to calculate posterior parameter distributions for models too slow to be run in sequential fashion in Bayesian frameworks. This is done through an adaptive Markov chain Monte Carlo fit using the parameters selected with the emulator fitting. At each step the emulators are fed a set of proposed parameters formatted in a manner matching the training data. The emulators then produce the probable outcomes of the FATES-HYDRO model, which can be used to calculate an estimation for each time step. This can then be used to calculate a log-likelihood based on the proposed and observed data. In the adaptive Markov chain Monte Carlo process, the search step is adjusted following 1,000 moves for every 1,000 proposed moves based on the covariance of the accepted samples captured so far. This optimizes the search motion towards greater acceptance rate for the proposed moves. In this simulation, we used a multivariate normal distribution to determine the initial search. To make sampling and loss function more regularized, parameters were normalized, and outputs were log-transformed and then normalized.

To test whether our method could recover "true" parameters based on the observations, and because these are fundamentally unknowable in real scenarios, we used a synthetic dataset. This synthetic dataset represents a FATES-HYDRO run with known parameters at the same site tested over the same period used in the emulator fitting. However, statistical noise ($\boldsymbol{\sigma}$= 1) was added to the output to provide a more difficult challenge for the emulators. This was used as the "observational" data within the posterior distribution calculations. Samplers were run for 10,000 steps following burn-in. Meta parameters (initial search distance, burn-in, steps to adjust move, percent of covariance) were determined through testing, with a goal of optimizing the acceptance rate (approx. 7-20% acceptance) and reducing the autocorrelation between steps (less than 0.1 following burn-in). FATES-HYDRO simulations using the posterior distributions were then compared to the outputs of the synthetic datasets for the calibration period and a future projection period using the climate projections from CESM SSP2-4.5 Earth system model outputs.

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

3.Results

The final parameters selected for fitting the reduced models were specific leaf area ($m^2\,gC^{-1}$: SLA_{top}), maximum carboxylation rate of Rub. at 25C, canopy top (umol CO^2 $m^{-2}s^{-1}$: Vc_{max25}), Ball-Berry stomatal slope (unitless: BB_{slope}), Xylem taper exponent for sapwood(unitless: p_{taper}), absorbing root radius (mm:r_{s2}) and absorbing root water potential at 50% loss of conductivity(MPa: $p50_{node\_a\_root}$), as well as the relevant time predictors (DOY, month, year: Table S1). The final emulators scored Pearson's correlation greater than 0.9 on all the separated testing data; monthly GPP at 0.994, monthly RO at 0.923, H_2OSOI at 0.959, ET at 0.983, daily LWP_{min} at 0.999, daily LWP_{max} at 0.957 (Table 1, Fig. 2). The predictive variables for the final GPP emulator were SLA_{top}, Vc_{max25}, and BB_{slope} (Fig. 3, Table S2). The most predictive for RO was month, year, and p_{taper} signifying that there was a greater influence from the climate drivers than the plant traits (Table S3). Similarly, H_2OSOI was most influenced by month and year, followed by SLA_{top} (Table S4). LWP_{min}, and LWP_{max} were controlled primarily by plant traits For LWP_{max} the most influential traits were p_{taper}, r_{s2}, BB_{slope}, and $p50_{aroot}$. For LWP_{min}, p_{taper}, Vc_{max25}, BB_{slope}, and SLA_{top} were the most influential (Table S5). ET was primarily influenced by BB_{slope}, Vc_{max25}, month and year (Table S6).

In analyzing the relationship between the predictor variables within the emulators through splines we found that no relationship statistically diverged for the entire year. Splitting the relationship between dry and wet season we found that 5 of 288 parameter and output combinations were statistically distinct. BB_{slope} influence on wet season H_2OSOI, SLA_{top} influence on wet season LWP_{max}, years influence on Dry Season ET, $p50_{aroot}$'s influence on Dry season RO, and r_{s2} 's influence on H_2OSOI, (Fig. 4, Table S7, S8). However, these divergences are not visually discernible, but should be considered as SLA_{top}, which was the third most influential predictor variable for LWP_{max}. The remaining 63 variable pairs all had $P > 0.05$, suggesting no significant difference in parametric sensitivity between the emulator and FATES model outputs.

The posterior distributions from the MCMC emulator process recovered most of the variables in the synthetic distribution (Fig 5). The posterior of the xylem taper (p_{taper}) had a median of 0.179 with a range between 0.09-0.265 compared to the synthetic value of 0.176. The posterior distribution of the Ball-Berry slope was 9.68 with a range between 8.70-10.68, compared to a value of 11.52 in the synthetic set. The posterior distribution of Vc_{Max} had a median of 74.34 with a range of 68.49-79.83 compared to the synthetic value of 77.89. The posterior distribution of SLA_{top} had a median of 0.0233 with a range of 0.0176-0.0290 compared with the synthetic value of 0.0367. The posterior distribution of the absorbing root P_{50} had a median of 1.81 with a range of 0.11-4.81 compared to the synthetic value of 2.41. The posterior distribution of the absorbing root radius had a median of 0.200 with a range of 0.0149-0.616 compared with the synthetic values of 0.313.

When compared to the synthetic data for the parameterization period, the MCMC select 95% CI parameters had an root mean square error (RMSE) for GPP of 3.748 g C $month^{-1}$, for ET 1.33 mm H2O $month^{-1}$, for soil moisture 0.005 m^2m^{-2}, 0.381 MPa for maximum leaf water potential, 0.44 MPa for minimum leaf water potential, 1.80 $m^2\,m^{-1}$ for runoff (Fig. 6). Future simulations of FATES-HYDRO drawn from 95% credible intervals seem to capture most of the trend in future

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

variables when compared to the synthetic dataset (Fig. 7). In comparing the root mean square error between the median of future simulations and the synthetic dataset GPP had an RMSE of 22.55 gC $m^{-2}month^{-1}$, ET had a RMSE of 7.82 mm H2O $month^{-1}$, runoff had an RMSE of 88.80 mm H2O $month^{-1}$, monthly mean leaf water potential had an RMSE of 0.145 MPa, soil water content at 20cm had an RMSE of 0.0138 m^2m^{-2}.

4. Discussion

Our results show that the CRAFT machine learning emulators can be used to capture posterior parameter distributions for error propagation of future simulations of FATES-HYDRO using a Bayesian framework. We also show that these methods can capture not only model outputs but parametric sensitivity responses from the physical model in most cases (283/288 parameter and output combinations). This method was able to capture a full suite of outputs and emulate them accurately (GPP, ET, LWP_{min}, LWP_{max}, RO, SWC) using the full suite of relevant parameters. We further provide a robust Bayesian framework by which to estimate full posterior parameter distributions associated with land surface and ecological model runs. This is an important improvement compared to more arbitrary selection of ensemble members from previous studies (Chitra-Tarak et al. 2021).

Our model was able to pinpoint key parameters identified in other broader sensitivity analyses on this site. For example, the dominating model parameters for GPP and ET are SLA_{top} and Vc_{max25}, which agrees with the finding from Koven et al. (2020), The importance of p_{taper}, r_{s2} and $p50_{node\_a\_root}$ for plant hydrodynamic simulations agree with Xu et al. (2023). The RF models performed worse in prediction of runoff, and soil moisture likely because those are least influenced by PFT-specific water usage parameters. Overall, the posterior parameters (Fig. 5) showed relatively little error in comparison to the observations (Fig. 6) suggesting the selected parameters can, in aggregate, represent the observational data within this site. Despite the relatively small amount of error in selected parameters during the parameterization period, future projections show wide ranges in divergence when compared to the synthetic dataset(Fig. 7). However, the synthetic simulation is generally within the range of future estimation, suggesting the error bounds to be generally correct to the uncertainty in parameterization if without a specificity, particularly in the case of soil water and leaf water potential (Fig. 7).

In contrast to other machine learning emulations, CRAFT achieves parameterization through emulation rather than a full surrogate for the model itself. This is based on our experience needing to simulate single sites or groups of sites based on ground-truthed data. This is particularly necessary to assimilate the larger number of data sources needed for large-scale process modeling utilizing global data and multi-site parameterization for regional simulation (Dukes et al., 2025). This model fitting exercise is common in the estimation of site-specific parameters needed for ecological studies and allows the process-based model to be the one used in future simulations. One advantage of this method is that the emulation allows for the computationally expensive part to be run in parallel (training data development), while still allowing Bayesian data

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

assimilation to be run in serial (owing to the minimal cost in running an emulator: < 0.1 second vs 4 hours). In this fitting exercise, we did not include spatial variation in the inputs, but only temporal ones. This logic is based on our exercise that the surrogate model is only used in the parameterization step under data constraint, and the calibrated process-based model will be used for model prediction under different climate conditions. This more rapid fitting and calibration allows the model to be specific rather than generalized, and thus surrogate models could be created accounting for outputs at each site in the event of multi-site simulations and validation exercises when necessary.

There are many hyperparameters and methods that may optimize this process and will require some amount of testing at each site. Our original selection of Random Forest was based on testing models against several algorithms (gradient boosted regression, simple neural network, random forest), and random forest performed best a majority of the time. This may be due to 1) Random Forest excels under a moderate number of inputs (here, the number of parameters, 42, and the month and year) and a large amount of training data, and 2) the ensemble nature of random forest that allows it to be robust under different parameter inputs. We believe more complex neural networks can perform as well as random forest but did not explore here due to the large number of hyper-parameters associated with these types of models.

The CRAFT framework could build upon other methods for emulators and Bayesian calibration of parameter rich models. Using the land surface model ED2 in a framework utilizing statistical emulators across a parameter space, Fer et al. (2018) found an increase in the parameter search and can be additionally used in a Bayesian sampling context, though did produce generally wider parameter estimations (without considering modeling differences). Dagon et al. (2020) found success with a neural network approach of emulation and principal component analysis, while instead we just worked with normalized variables. As opposed to just removing parameters found to not be influential, principal components (and their transformation) would be a method by which a greater number of parameters can be adjusted in parameter search. Separate from other approaches, we evaluate the Jacobian response of the emulator to change in the parameters of interest, and its correlation between the model and emulator. Effort in testing the number of parameters sets sampled and run using the FATES-HYDRO model required to emulate will be useful in reducing upfront costs. Lu and Ricciuto (2019) showed that a relatively small number of training parameter sets in earth systems modeling (20) can be used to train a model and emulate many sets (1,000) through singular value decomposition, though this is for a single output (GPP). In addition, a process such as CRAFT can use the same emulators to evaluate initial search ranges and find parameter sets where feasible parameters are rare (such as competitive coexistence) such as the search done in Li et al. (2023).

An increase in biases that result from modeling structure may occur during emulation, which is particularly problematic in parameterization exercises covering large areas (Dagon et al. 2020). However, this can be used to identify structural biases in the case where single parameter values will be used to cover large biogeochemical areas (full earth system simulations). Simulations that use truncated temporal dynamics (as in this study) may conceal structural bias. For example, daily cycles might exhibit unrealistic amplitude variations while still producing means that match observational data. This, however, is a possible failing in many of the existing optimization parameterization frameworks and necessitates understanding model outputs at a greater level than just those optimized. While validation of the findings through the

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

structured physical models allow for a reduction in the inductive problems inherent in machine learning (Barbierato and Gatti, 2024), the problem of equifinality in a finite system will still persist in forecasting (i.e. there is only one combination of traits in the real system, and will always be multiple that may reach that systems output given incomplete comparison (Napoletani et al., 2014). This additionally is a problem in numerical modeling with unobservable or uncertain free parameters as well (Reichstein et al., 2019). While we provide outputs for Shapley importances, and capturing the Jacobian relationship between parameters and outputs, further effort in interpreting machine learning emulation could be necessary for diagnostics.

Other land surface process emulators have also been shown to capture comprehensive model behaviors (for some predictive period), using the climate drivers of the complex model at each time step (or multiple prior time steps) to drive the emulator (Baker et al., 2022). While this approach makes more sense for projection, within the context of fixed calibration period, the day of year, month, and year can capture this variation, as they are the same among model runs. Other efforts in this space include incorporating differential equations into the biophysical code base (in this case FATES), to allow for optimization (Aboelyazeed et al., 2025). While this seems very promising, the current computational limits (generally RAM or virtual-RAM) limit this to subsections of a model for optimization. To improve the usability, these differentiable components can be incorporated into these general sensitivity assessments in CRAFT to determine the finite subsection parameter optimization and the larger model superstructure for increased accuracy.

5. Conclusion

Earth system and ecological models are a crucial part of hazards predictions and mitigation but are only as useful as their certainty. As these models trend to increase mechanism, faster and more robust methods of parameter estimation and error uncertainty will be required to determine model efficacy. Toward this end we constructed the CRAFT framework and the associated python workflow to allow for determining key parameters and robustly estimating them. While constructed for the FATES-HYDRO model, the basic framework could be applied to a wide range of models and observations, as the framework itself only assumes parameters and their influence on observations. Future work will incorporate a greater number of emulator options into the CRAFT framework. We have developed a deep-learning version of CRAFT (using Tensorflow; Abadi et al., 2016) that seems to perform on a similar level to the Random Forest version, though with greater effort in hyper-parameter tuning. This includes a more advanced Neural Network testing regime using Tensorflow, which will be released along with the code base of this manuscript and provides a greater range of customization in emulator, while using the same throughflow from model processing to parameter estimation. Neural networks can be greater emulators in the event of uneven probability transitions (such as mortality and/or existence), rather than the ecosystem variables we show here. Additionally, Neural Networks present the option for a single model to predict multiple outputs (for comparison to observations). Additionally, we include additional Bayesian sampling methods through the DREAM algorithm (Lu et al., 2017). CRAFT allows model/parameter customization and can be adapted to various land surface and ecological models.

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

While not evaluated in this study, multiple Markov chains can be run utilizing high performance computing, to handle the uncertainty of convergence in Bayesian estimation. We additionally hope to assess model performance on more complex FATES-HYDRO studies, including dynamic simulations with multiple PFTS across multiple domains, the complexity of which may prove more difficult in finding posteriors. Overall, we show the CRAFT framework as a rapid and accurate semi-automated method to assimilate data and calculate posterior distributions in complex physical models.

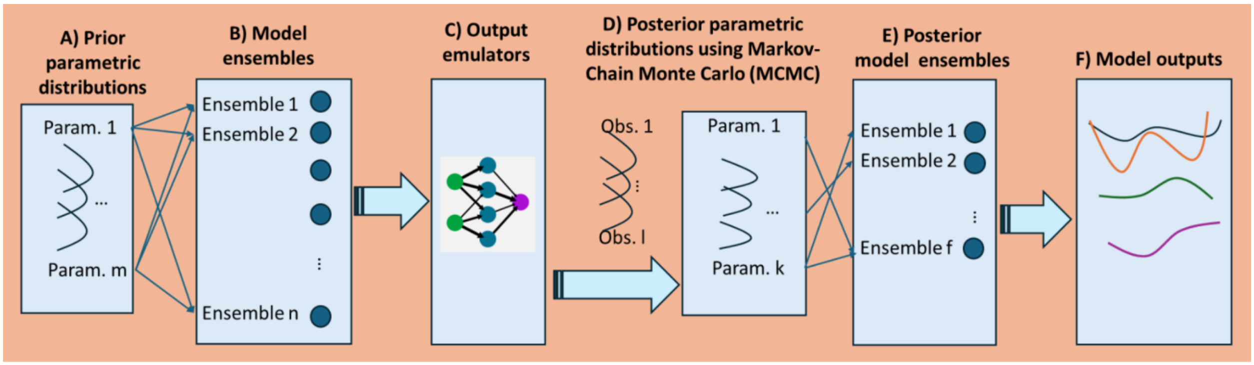
5. Tables and Figures

Table 1: Final training and testing score for Random Forest emulators of the FATES-HYDRO model for Barro Colorado Island, Panama (2000-2018). Models were trained on 1,000 runs of FATES-HYDRO using Latin-Hypercube sampling

Output Variable	Time scale	Training Score	Testing Score
Gross Primary Productivity	Monthly	0.999	0.994
Runoff	Monthly	0.989	0.923
Soil H20	Monthly	0.994	0.959
Evapotranspiration	Monthly	0.997	0.983
Minimum Leaf Water Potential	Daily	1.000	0.999
Maximum Leaf Water Potential	Daily	0.994	0.957

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

Tables and Figures



310 **Figure 1:** An overview of the Calibrated **R**apid Assimilation and Forecasting **T**echnique (CRAFT). This includes A) Establishing a prior range of parameters for a physics-based model, either uniform or where data allows having statistical shape. B) Model ensemble simulations with parameter values drawn using Latin hypercube sampling l. C) Emulators are constructed using machine learning methods that best emulate the time series of the model in relation to time and the physical model parameters. D) Monte Carlo Simulations using emulators across physical model parameters of importance
315 against the observational data for the study site. E) Using the posterior estimation of the parameter groups, propagate uncertainty in future simulations (F). Using the full model with parameter groups in experiments.

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

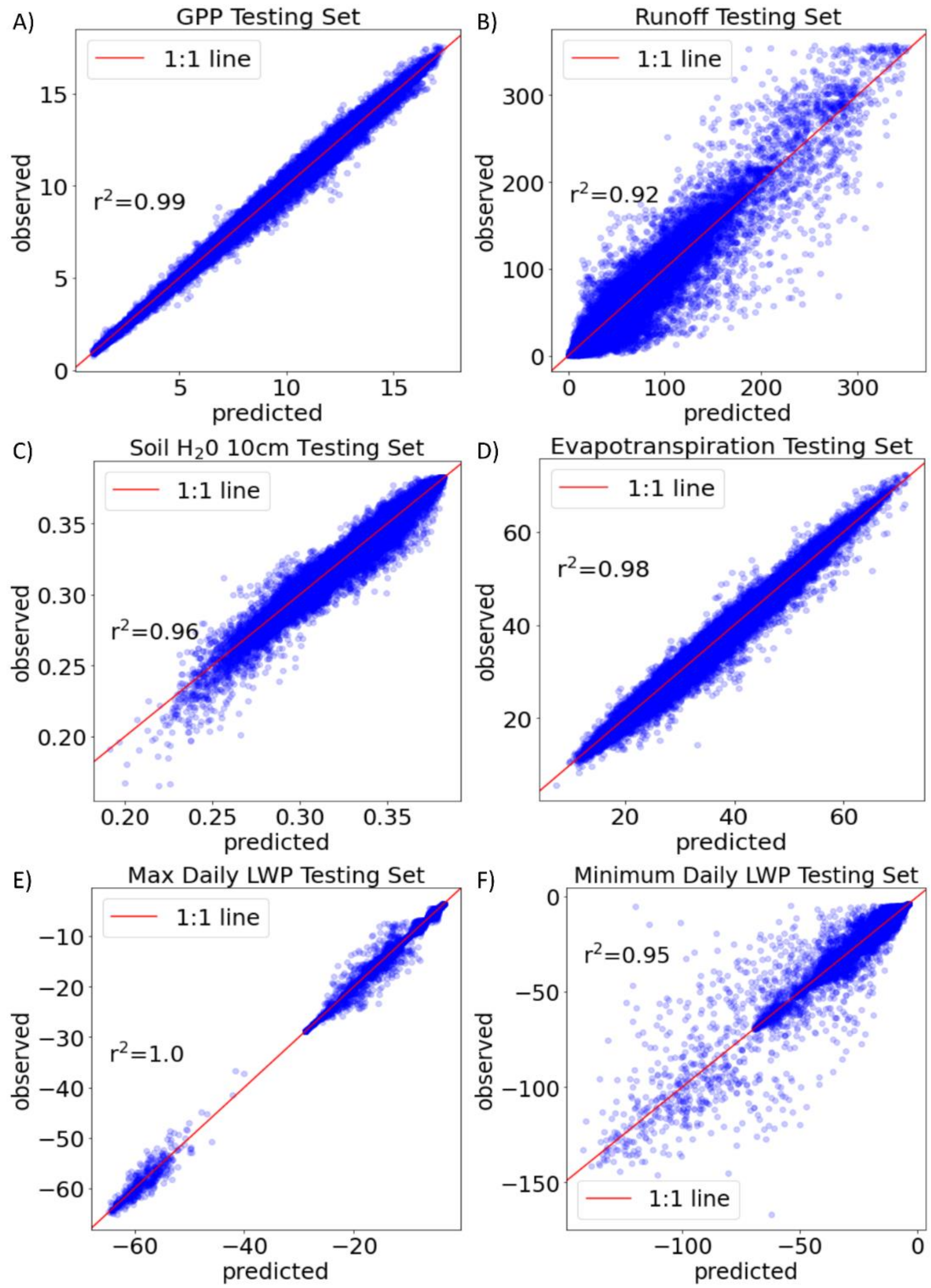


Figure 2: Machine learning predicted values for each output variable and the observed value within the testing data set from
320 FATES-HYDRO

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

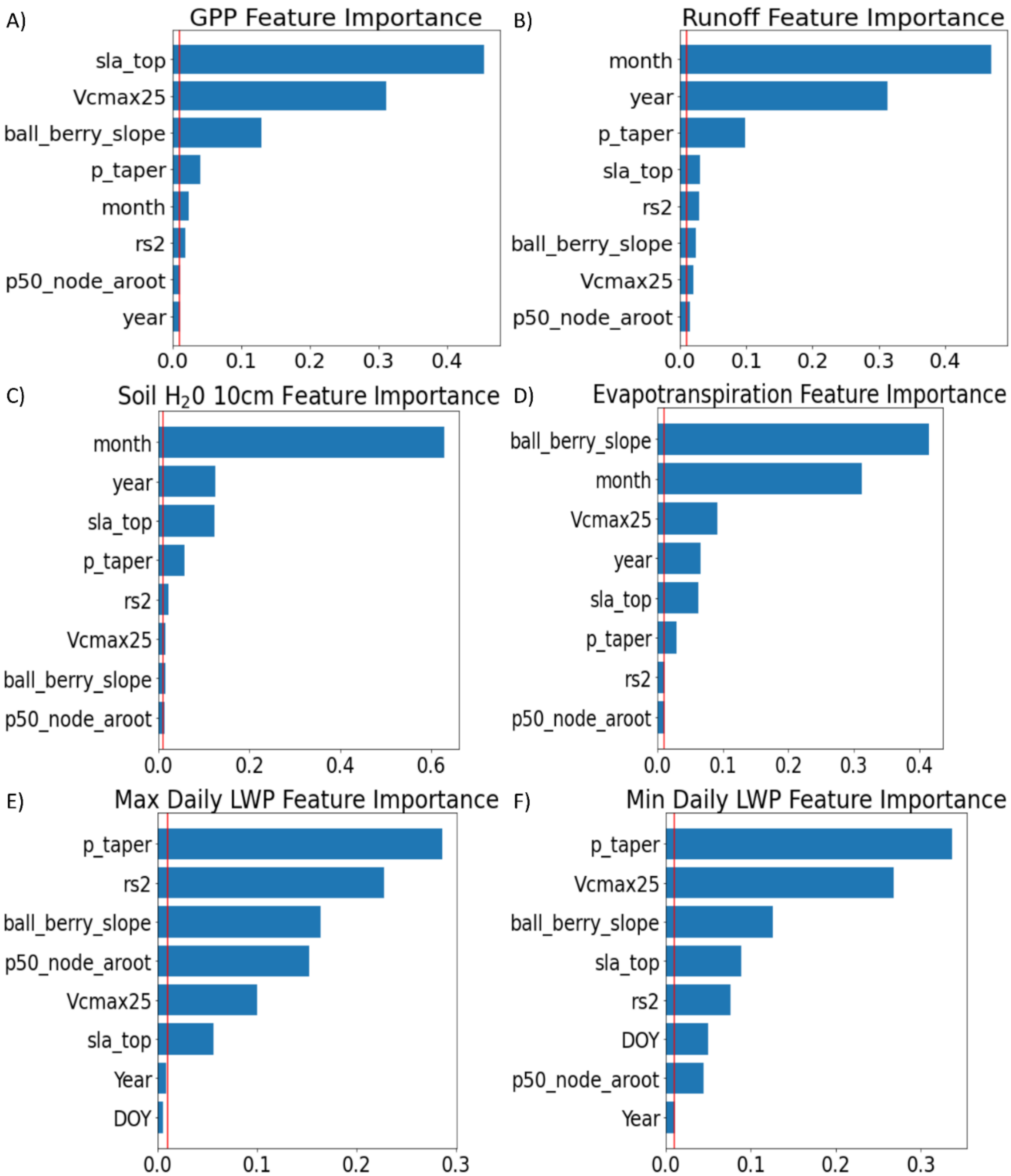
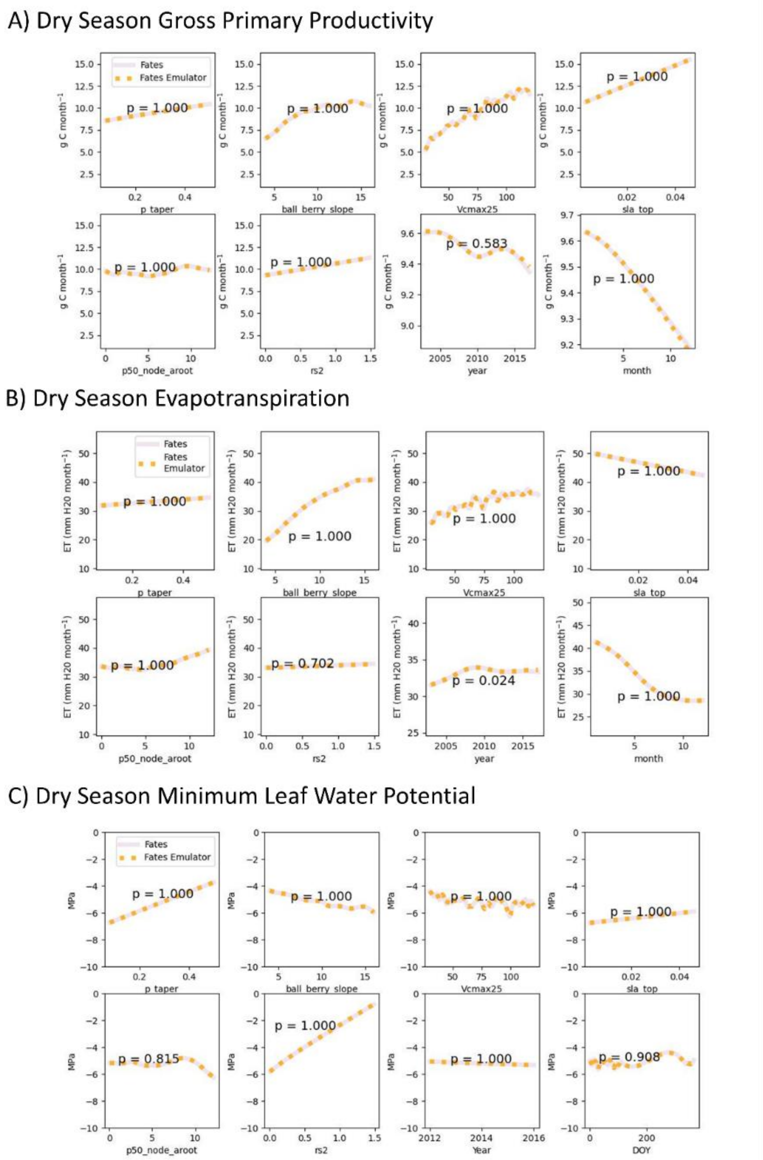


Figure 3: Feature importance in each CRAFT emulator of the FATES-HYDRO model for Barro Colorado Island, Panama. Feature importance calculated from Shapley values.

325

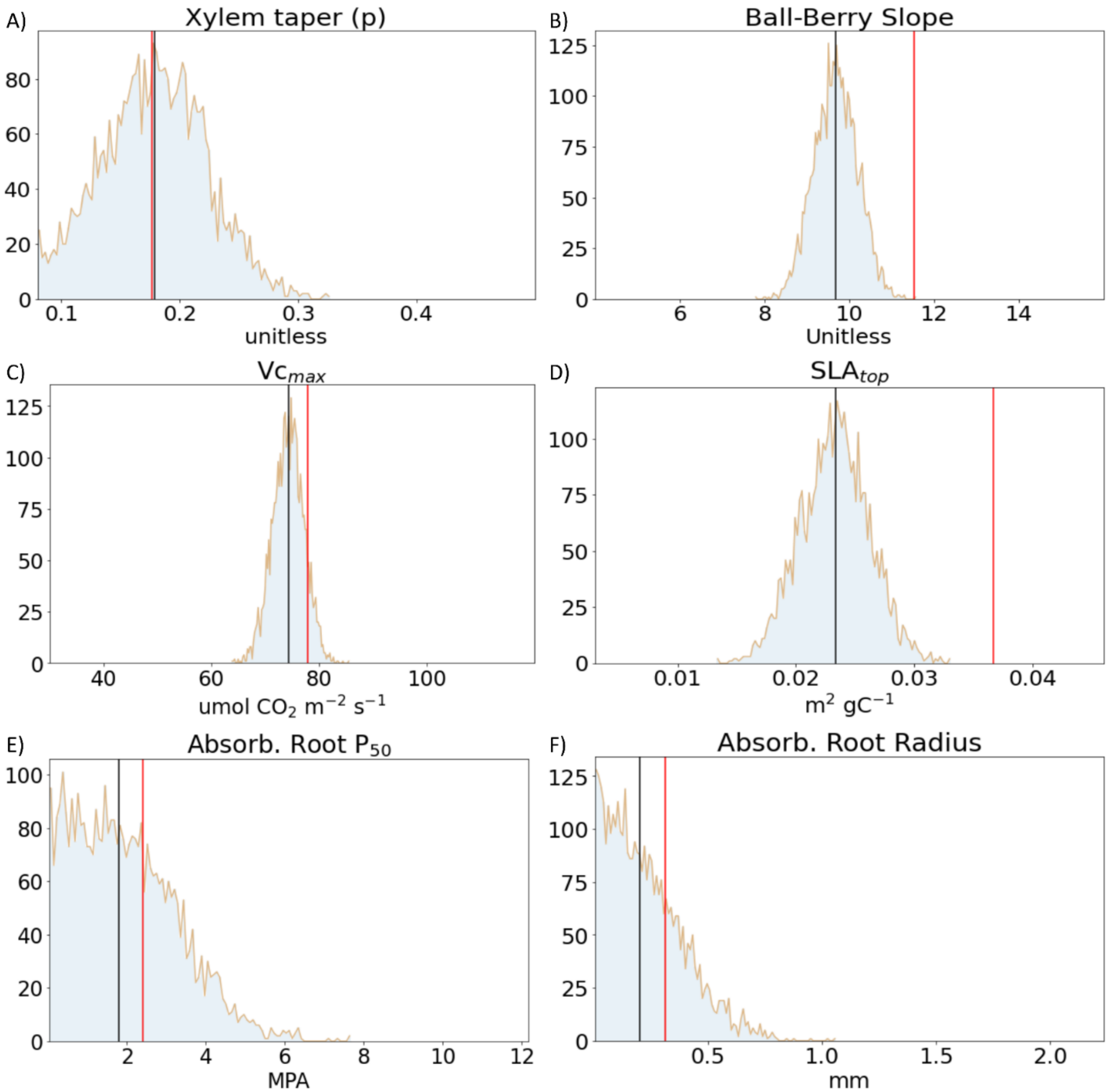
https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026



330

Figure 4: Comparison of the derivative relationship for change in input and out variables of FATES, and an emulator for A) Dry season GPP, B) Dry season minimum leaf water potential, C) Dry season maximum leaf water potential.

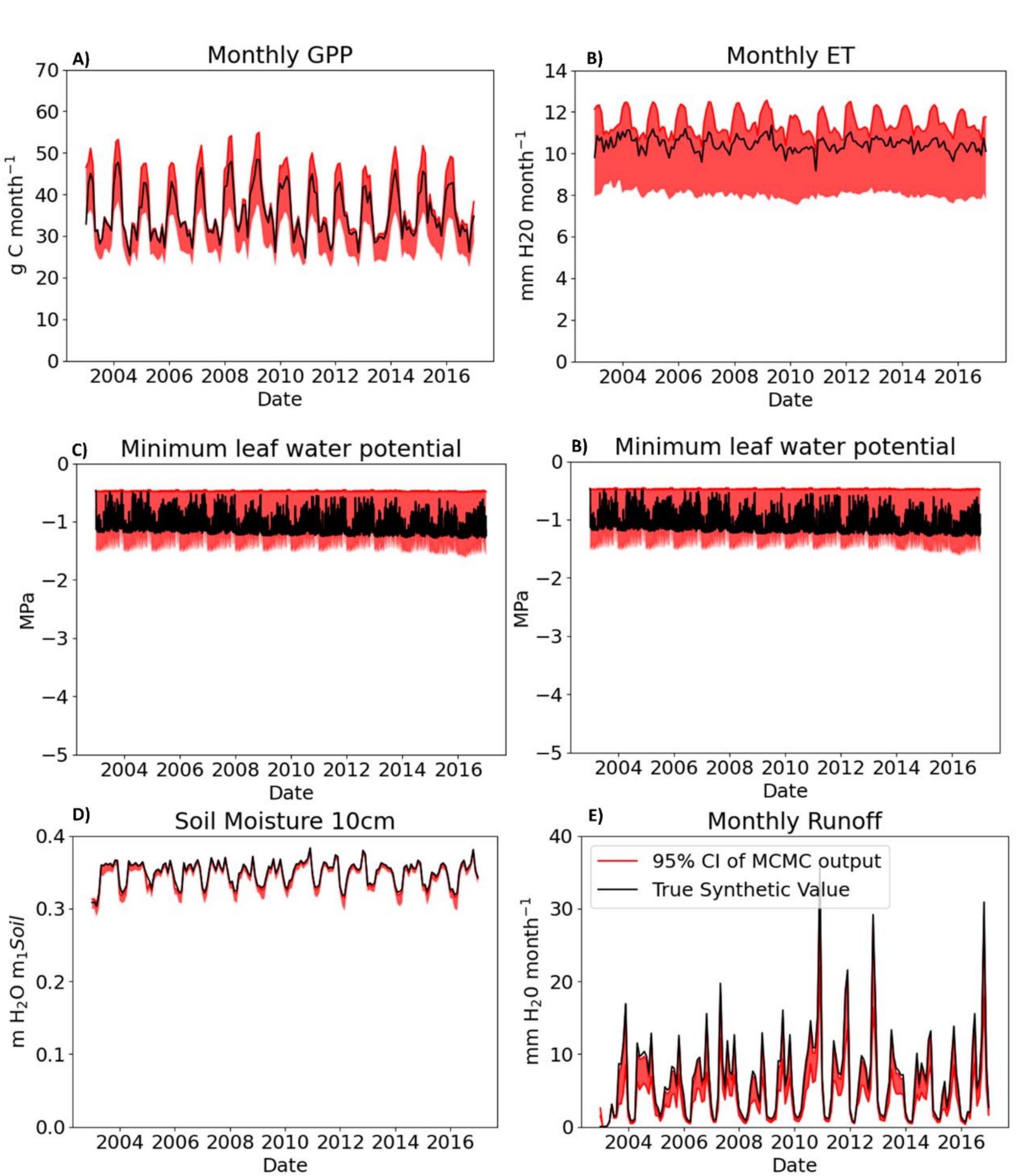
https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026



335

Figure 5: The posterior distribution and median of each tested variable in the CRAFT model fit for the FATES-HYDRO (median in black) at Barro Colorado Island, Panama compared against the synthetic FATES data outputs (red) used in testing parameter recall.

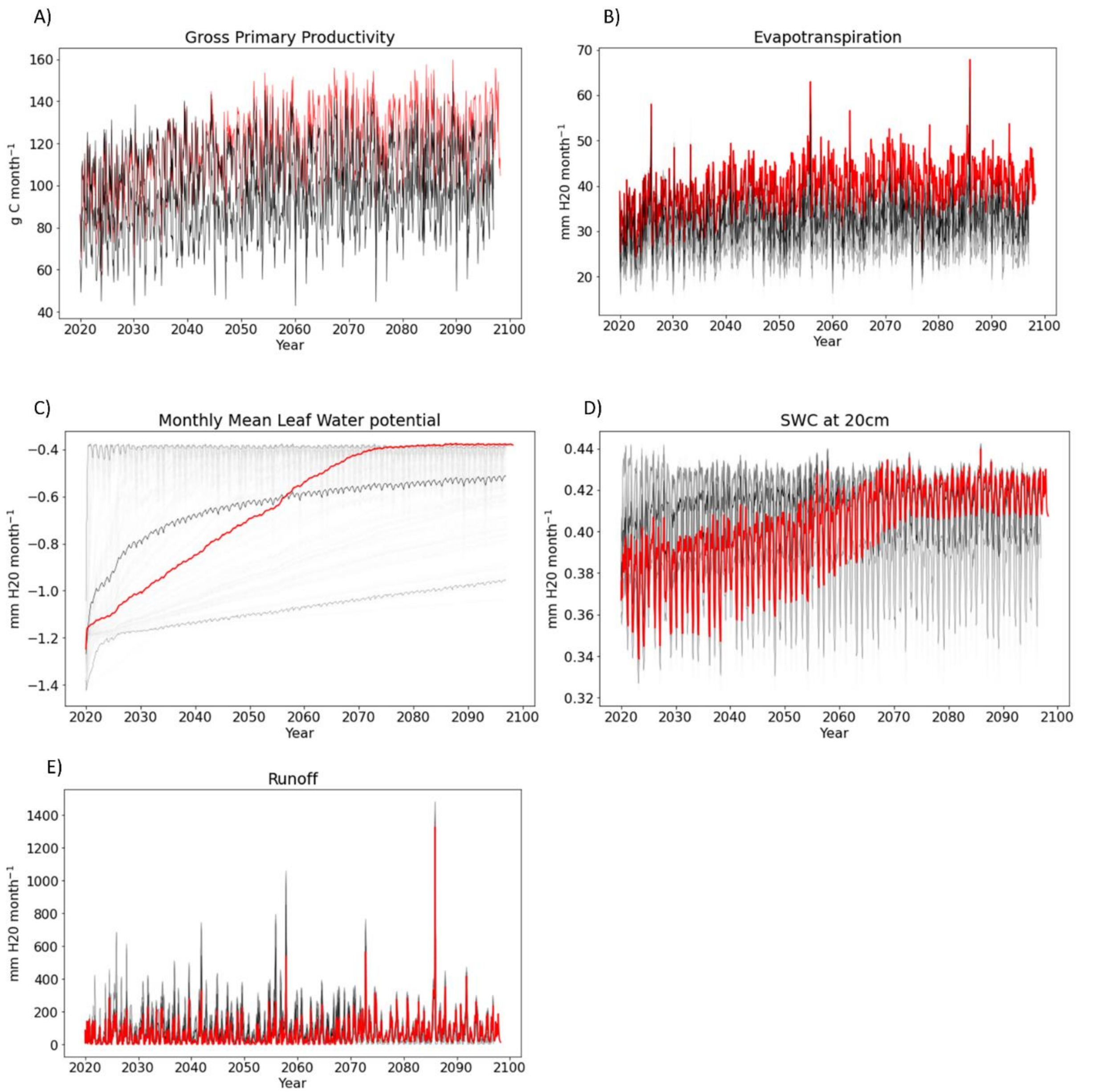
https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026



340

Figure 6: The 95% credible interval of posterior distribution of the CRAFT method used in FATES-HYDRO to simulate Barro Colorado Island, Panama for 2000-2018 and the synthetic run of FATES-HYDRO (black).

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026



345 Figure 7: Future FATES-HYDRO simulations of the Barro Colorado Island, Panama (2020-2100) using the CESM climate model SSP2-4.5. Parameters used in simulation are from the 95% credible interval of posterior distribution of the CRAFT method (outlined in grey with black line as the median, and a synthetic data set fit to in red.

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

Code and data availability

Code is available at 10.5281/zenodo.19921550, continuously updating at https://github.com/lanl/CRAFT

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

ZR, CX- designed the experiments.

ZR, LT- designed the code base and executed experiments.

CK, RK- Provided FATES insight

ZR, LT,CX,CK, NM, RK-planned and contributed to the manuscript.

Competing interests

Authors assert they have no competing interests.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., & Isard, M. (2016). {TensorFlow}: A system for {Large-Scale} machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283.

Aboelyazeed, D., Xu, C., Hoffman, F. M., Liu, J., Jones, A. W., Rackauckas, C., Lawson, K., & Shen, C. (2023). A differentiable, physics-informed ecosystem modeling and learning framework for large-scale inverse problems: Demonstration with photosynthesis simulations. *Biogeosciences*, *20*(13), 2671–2692.

Baker, E., Harper, A. B., Williamson, D., & Challenor, P. (2022). Emulation of high-resolution land surface models using sparse Gaussian processes with application to JULES. *Geoscientific Model Development*, *15*(5), 1913–1929.

Barbierato, E., & Gatti, A. (2024). The challenges of machine learning: A critical review. *Electronics*, *13*(2), 416.

Buotte, P. C., Koven, C. D., Xu, C., Shuman, J. K., Goulden, M. L., Levis, S., Katz, J., Ding, J., Ma, W., & Robbins, Z. (2021). Capturing functional strategies and compositional dynamics in vegetation demographic models. *Biogeosciences*, *18*(14), 4473–4490.

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

Chitra-Tarak, R., Xu, C., Aguilar, S., Anderson-Teixeira, K. J., Chambers, J., Detto, M., Faybishenko, B., Fisher, R. A., Knox, R. G., & Koven, C. D. (2021). Hydraulically-vulnerable trees survive on deep-water access during droughts in a tropical forest. *New Phytologist*, *231*(5), 1798–1813.

Christoffersen, B. O., Gloor, M., Fauset, S., Fyllas, N. M., Galbraith, D. R., Baker, T. R., Kruijt, B., Rowland, L., Fisher, R. A., & Binks, O. J. (2016). Linking hydraulic traits to tropical forest function in a size-structured and trait-driven model (TFS v. 1-Hydro). *Geoscientific Model Development*, *9*(11), 4227–4255.

Dagon, K., Sanderson, B. M., Fisher, R. A., & Lawrence, D. M. (2020). A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5. *Advances in Statistical Climatology, Meteorology and Oceanography*, *6*(2), 223–244.

Dietze, M. C. (2017). Prediction in ecology: A first-principles framework. *Ecological Applications*, *27*(7), 2048–2060.

Dukes, J. S., Xu, C., Liao, C., Novick, K. A., Phillips, R. P., Beverly, D. P., Fang, Y., Jacobs, E. M., McAdam, S. A., & Paudel, I. (2026). Improving the representation of plant water stress and water use in Earth System Models. *New Phytologist*, *249*(1), 39–55.

Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., & Dietze, M. C. (2018). Linking big models to big data: Efficient ecosystem model calibration through Bayesian model emulation. *Biogeosciences*, *15*(19), 5801–5830.

Fisher, R. A., Koven, C. D., Anderegg, W. R., Christoffersen, B. O., Dietze, M. C., Farrior, C. E., Holm, J. A., Hurtt, G. C., Knox, R. G., & Lawrence, P. J. (2018). Vegetation demographics in Earth System Models: A review of progress and priorities. *Global Change Biology*, *24*(1), 35–54.

Fisher, R., Muszala, S., Verteinstein, M., Lawrence, P., Xu, C., McDowell, N., Knox, R., Koven, C., Holm, J., & Rogers, B. (2015). Taking off the training wheels: The properties of a dynamic vegetation model without climate envelopes. *Geoscientific Model Development Discussions (Online)*, *8*(4).

Gao, X., Avramov, A., Saikawa, E., & Schlosser, C. A. (2021). Emulation of Community Land Model version 5 (CLM5) to quantify sensitivity of soil moisture to uncertain parameters. *Journal of Hydrometeorology*, *22*(2), 259–278.

Getz, W. M., Marshall, C. R., Carlson, C. J., Giuggioli, L., Ryan, S. J., Romañach, S. S., Boettiger, C., Chamberlain, S. D., Larsen, L., & D'Odorico, P. (2018). Making ecological models adequate. *Ecology Letters*, *21*(2), 153–166.

Green, P. L., & Worden, K. (2015). Bayesian and Markov chain Monte Carlo methods for identifying nonlinear systems in the presence of uncertainty. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *373*(2051), 20140405. https://doi.org/10.1098/rsta.2014.0405

Hararuk, O., Xia, J., & Luo, Y. (2014). Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 403–417.

Kochenderfer, M. J. (2015). *Decision making under uncertainty: Theory and application*. MIT press. https://books.google.com/books?hl=en&lr=&id=hwxZCgAAQBAJ&oi=fnd&pg=PR7&dq=Kochenderfer,+Mykel+J.+Decision+making+under+uncertainty:+theory+and+application.+MIT+press,+2015.&ots=s3gruWv6rC&sig=7ADHKx9XNzsQ1iTm1ocymkjWpkc

Koven, C. D., Knox, R. G., Fisher, R. A., Chambers, J. Q., Christoffersen, B. O., Davies, S. J., Detto, M., Dietze, M. C., Faybishenko, B., & Holm, J. (2020). Benchmarking and parameter sensitivity of physiological and

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

vegetation dynamics using the Functionally Assembled Terrestrial Ecosystem Simulator (FATES) at Barro Colorado Island, Panama. *Biogeosciences*, *17*(11), 3017–3044.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, *23*(1), 18.

Lu, D., & Ricciuto, D. (2019). Efficient surrogate modeling methods for large-scale Earth system models based on machine-learning techniques. *Geoscientific Model Development*, *12*(5), 1791–1807.

Lu, D., Ricciuto, D., Walker, A., Safta, C., & Munger, W. (2017). Bayesian calibration of terrestrial ecosystem models: A study of advanced Markov chain Monte Carlo methods. *Biogeosciences*, *14*(18), 4295–4314.

Massoud, E. C. (2019). Emulation of environmental models using polynomial chaos expansion. *Environmental Modelling & Software*, *111*, 421–431.

Moorcroft, P. R., Hurtt, G. C., & Pacala, S. W. (2001). A method for scaling vegetation dynamics: The ecosystem demography model (ED). *Ecological Monographs*, *71*(4), 557–586.

Napoletani, D., Panza, M., & Struppa, D. C. (2014). Is big data enough? A reflection on the changing role of mathematics in applications. *Notices of the American Mathematical Society*, *61*(5), 485–490.

Nonnenmacher, M., & Greenberg, D. S. (2021). Deep emulators for differentiation, forecasting, and parametrization in Earth science simulators. *Journal of Advances in Modeling Earth Systems*, *13*(7), e2021MS002554.

Pal, S., & Sharma, P. (2021). A review of machine learning applications in land surface modeling. *Earth*, *2*(1), 174–190.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Robbins, Z., Chambers, J., Chitra-Tarak, R., Christoffersen, B., Dickman, L. T., Fisher, R., Jonko, A., Knox, R., Koven, C., & Kueppers, L. (2024). Future climate doubles the risk of hydraulic failure in a wet tropical forest. *New Phytologist*, *244*(6), 2239–2250.

Robbins, Z., & Xu, C. (2026). Calibrated Rapid Assimilation and Forecasting Technique (CRAFT) (Publication). Zenodo. https://doi.org/10.5281/zenodo.19921551

Saltelli, A. (2019). A short comment on statistical versus mathematical modelling. *Nature Communications*, *10*(1), 3870.

Sawada, Y. (2020). Machine learning accelerates parameter optimization and uncertainty assessment of a land surface model. *Journal of Geophysical Research: Atmospheres*, *125*(20), e2020JD032688.

Schlüter, M., Haider, L. J., Lade, S. J., Lindkvist, E., Martin, R., Orach, K., Wijermans, N., & Folke, C. (2019). Capturing emergent phenomena in social-ecological systems. *Ecology and Society*, *24*(3).

Shapley, L. S. (1953). *A value for n-person games*.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 25.

Xu, C., Christoffersen, B., Robbins, Z., Knox, R., Fisher, R. A., Chitra-Tarak, R., Slot, M., Solander, K., Kueppers, L., & Koven, C. (2023). Quantification of hydraulic trait control on plant hydrodynamics and risk of

https://doi.org/10.5194/egusphere-2026-2433
Preprint. Discussion started: 5 June 2026

hydraulic failure within a demographic structured vegetation model in a tropical forest (FATES–HYDRO V1. 0). *Geoscientific Model Development*, *16*(21), 6267–6283.