https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

Improving Terrestrial Carbon Flux Simulations With Machine Learning and Global Earth Observations

Christian Seiler[1]

[1]School of Environmental Studies, Queen's University, BioSciences Complex, 116 Barrie St., Kingston, Ontario, Canada, K7L 3J9

Correspondence: Christian Seiler (christian.seiler@queensu.ca)

Abstract. The land carbon cycle can act as both a negative and positive climate feedback. Currently, it serves as a negative feedback, absorbing about one-third of anthropogenic CO_2 emissions. However, multi-model studies project a weakening of this sink, with the potential for a future shift to a carbon source. Significant inter-model differences persist, limiting confidence in these projections. Some of these discrepancies may arise from parameter uncertainty. Advances in artificial intelligence, computing, and Earth observations now offer new opportunities to better constrain key model parameters. While previous studies have shown that parameter optimization can substantially improve model performance, they have not explored its impact on the future carbon balance. To address this gap, I use a machine learning algorithm to optimize 28 model parameters based on 13 global Earth observation datasets. The resulting parameter set is then applied in carbon cycle simulations under historical conditions and a high-emissions future scenario. Results show that optimization significantly improves model performance, particularly for gross primary productivity (GPP), leaf area index, and sensible heat flux. Globally, optimized net biome productivity is lower than in the default simulation (33% lower from 1960 to 2022 and 43% lower from 2015 to 2100) due to reduced GPP and increased autotrophic respiration. Regionally, optimization tends to weaken both carbon sinks and sources, reducing the contrast between them. In conclusion, parameter tuning can substantially alter historical and future carbon fluxes, with effects comparable to adding new processes. To reduce inter-model spread, modeling groups should integrate advanced parameter optimization frameworks into their model development cycle.

1 Introduction

The global carbon cycle has the potential to both mitigate and amplify climate change. The Earth's terrestrial biosphere currently absorbs more CO_2 than it releases, making it a net carbon sink (Friedlingstein et al., 2024). This sink sequesters about one-third of all anthropogenic CO_2 emissions, acting as a negative feedback mechanism that mitigates climate change. Paleoclimate data on the other hand shows that an increase in atmospheric CO_2 concentration raises global mean surface temperatures, which in turn drives further increases in atmospheric CO_2 (van Nes et al., 2015). Here, the carbon cycle is acting as a positive feedback mechanism, amplifying a warming trend. The empirical evidence for a positive carbon cycle feedback raises the question whether the terrestrial carbon sink will eventually transition from a net carbon sink to a net carbon source as global mean surface temperatures continue to rise.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

Empirical data from the global carbon flux monitoring network indicate that over the past decade, plants have reached the thermal maximum for photosynthesis during the three warmest months each year (Duffy et al., 2021). The same study suggests that as temperatures rise, respiration rates are expected to continue increasing while photosynthesis declines, potentially reducing the land sink strength by nearly half over the next two decades under a business-as-usual emission scenario.

Climate models can help separate the effects of changes in atmospheric CO_2 concentration (carbon-concentration feedback) and climate (carbon-climate feedback) on land-atmosphere carbon fluxes. Results from the latest generation of climate models, known as Earth System Models (ESMs), indicate that the carbon-concentration feedback is negative, meaning the land acts as a net carbon sink. In contrast, the carbon-climate feedback is positive, suggesting the land functions as a net carbon source (Arora et al., 2020). When both feedback mechanisms act in concert, ESMs project that the land will continue to act as a carbon sink by the end of the 21st century, but the strength of this sink is subject to substantial uncertainties, ranging from about 2 to 7 PgC per year for a fossil-fuel-intensive scenario (Canadell et al., 2021). Longer-term simulations with a set of five ESMs and one ESM of intermediate complexity project that the terrestrial biosphere will switch from a carbon sink to either a neutral state or a carbon source by the year 2300 (Koven et al., 2022). Here as well, the inter-model differences remain substantial, ranging between -7 to 0 PgC per year.

The annual publication of the global carbon budget (GCB) provides regular updates on historical carbon fluxes and the strength of the terrestrial carbon sink (Friedlingstein et al., 2024). This carbon sink is estimated using an ensemble of terrestrial biosphere models referred to as the trends in the land carbon cycle project (TRENDY) ensemble. In the context of TRENDY, the models are driven offline with quasi-observed meteorological data. However, many of the TRENDY models also serve as the land surface components of fully coupled ESMs used for the future projections discussed above. Comparing the TRENDY ensemble against a wide range of global Earth observations shows that while model performance is generally reasonable, substantial room for improvement remains (Seiler et al., 2022). Notably, there is a large inter-model spread in net biome productivity (NBP) and gross primary productivity (GPP), and most models tend to overestimate the leaf area index across all latitudes when compared to multiple remote sensing datasets (their Figure 2).

To summarize, empirical data and numerical modelling both indicate that the strength of the land carbon sink will decline as temperatures rise. However, the rate of this decline and a potential shift from a carbon sink to a carbon source remain highly uncertain. The current generation of models that simulate terrestrial carbon fluxes is still subject to considerable biases and large inter-model differences. Improving model performance for the historical period may yield more reliable future projections. One source of uncertainty arises from parameters that are not fundamental constants, but whose values depend on environmental conditions. The true values of these free parameters are often unknown and must be selected within physically consistent uncertainty ranges. While traditional approaches to parameter tuning rely on expert judgment, recent advancements in machine learning, coupled with increasing computational power and the availability of global Earth observations, have opened up new opportunities for more systematic and efficient parameter tuning.

A common approach for parameter optimization is referred to as data assimilation. Data assimilation has a long history in weather forecasting, where it uses global Earth observations to estimate the initial conditions of state variables such as temperature, humidity, and wind speed. However, the same technique can be used to optimize parameters, essentially by treating

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

model parameters as initial conditions. The technique can use a variety of data streams, ranging from local in-situ observations to globally gridded datasets at daily, monthly, and annual time scales (MacBean et al., 2022). For instance, Kuppel et al. (2012) optimized 21 model parameters using daily carbon and energy fluxes from 12 eddy covariance measurement sites, reducing the model's root-mean-square error by 22%. MacBean et al. (2018) optimized 17 parameters using monthly satellite-derived measures of sun-induced chlorophyll fluorescence (SIF), reducing global mean annual GPP biases from 46% to 25%. Bacour et al. (2023) optimized 15 parameters using eddy covariance data, satellite-derived normalized difference vegetation index, and atmospheric CO_2 concentration data measured at stations, resulting in a stronger global carbon sink that better aligns with estimates from the GCB. Their study also found that parameter optimization is more effective when multiple data streams are used simultaneously during the optimization process.

The examples above apply to the Organising Carbon and Hydrology In Dynamic Ecosystems (ORCHIDEE) model, which forms the land surface component of the Institute Pierre Simon Laplace (IPSL) ESM (IPSL-CM5). The approach uses a Bayesian framework that iteratively minimizes a cost function using a gradient-based method referred to as the Limited-memory Broyden-Fletcher-Goldfarb-Shanno Bound (L-BFGS-B) algorithm (Tarantola, 2005; Byrd et al., 1995). This widely used algorithm has also been applied to other models, such as the Joint UK Land Environment Simulator (JULES) (Raoult et al., 2016) and the Jena Scheme for Biosphere-Atmosphere Coupling in Hamburg (JSBACH) (Schürmann et al., 2016). Other commonly used optimization algorithms include the adaptive Metropolis - Hastings - based Markov Chain Monte Carlo algorithm, which has been employed for parameter estimation in the Lund–Potsdam–Jena General Ecosystem Simulator (LPJ-GUESS) (Kallingal et al., 2024), and the Ensemble Adjustment Kalman Filter, which has been applied to the Community Land Model (CLM) (Fox et al., 2018).

Previous carbon cycle parameter optimization studies have focused on improving model performance against historical observations, without assessing their impact on future projections. As a result, it remains unclear how parameter optimization influences the simulated strength of the future land carbon sink. This study addresses that gap by presenting the first explicit assessment of how parameter optimization affects projections of future carbon fluxes. The approach is implemented using the Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC) (Melton et al., 2020), which serves as the land surface component of the Canadian Earth System Model (CanESM) (Swart et al., 2019) and contributes to the TRENDY ensemble (Friedlingstein et al., 2024). Although recent efforts have advanced CLASSIC parameter optimization (Gauthier et al., 2024), a general-purpose optimization framework, capable of targeting any subset of CLASSIC parameters and integrating diverse observational data streams, has yet to be developed. Whereas previous studies typically constrain model parameters using observations over relatively short time periods (Peylin et al., 2016; Schürmann et al., 2016; MacBean et al., 2018), the optimization in this study is performed over a much longer time span (1701–2020), allowing sufficient time for parameter effects to fully manifest. In addition, the optimization simultaneously targets multiple variables using multiple observational data streams per variable. This approach helps account for observational uncertainty and reduces the risk of overfitting to any single dataset.

The following sections detail the methods used in this study, including a description of the model, simulation protocol, model parameters, machine learning algorithm, and evaluation framework. The results section begins by evaluating the relia-

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

bility of the approach using synthetic data, followed by an analysis of how optimization affects parameter values and model performance. I then assess the impact of optimization on net biome productivity (NBP) under both recent and future climate conditions. Furthermore, I examine the robustness of the results by testing the effects of different optimization settings. The discussion section highlights key findings, addressing both the potential and limitations of my approach. Finally, I discuss the importance of improving estimates of future carbon fluxes for climate change mitigation policies and emphasize the need for climate modelling groups to integrate advanced parameter optimization methods into their model development cycle.

2 Methods

2.1 Land Surface Model

The Canadian Land Surface Scheme Including Biogeochemical Cycles (CLASSIC) forms the land surface component of the Canadian Earth System Model (CanESM) (Melton et al., 2020; Swart et al., 2019). CLASSIC's physical component, known as CLASS, partitions the land surface into four distinct subareas: bare ground, snow-covered ground, canopy-covered ground, and snow-covered canopy. The model solves the energy balance equations for the canopy layer and the underlying ground separately. The canopy comprises a single layer that encompasses four plant functional types (PFTs): needleleaf trees, broadleaf trees, grasses, and crops. The offline model configuration used here has 20 soil layers ranging from 10 layers of 0.1 m thickness, gradually increasing to a 30 m thick layer for a total ground depth of over 61 m. Water fluxes are computed within the permeable soil depth of the ground column, excluding underlying bedrock layers. Temperature calculations encompass both soil and bedrock layers. Furthermore, the model computes the temperature, mass, albedo, and density of the single-layer snowpack, as well as the temperature, interception, storage of rain and snow on the canopy, and the temperature and depth of ponded water on the ground surface.

CLASSIC's biogeochemical component, referred to as CTEM, is a dynamic vegetation model coupled to CLASS. CLASS provides essential physical land surface data, including soil moisture, soil temperature, and net radiation, which are used by CTEM to simulate photosynthesis in response to [CO_2]. CTEM accounts for three living vegetation components (leaves, stem, and roots) and two dead carbon pools (litter and soil). These five carbon pools are monitored for nine PFTs, mirroring the four PFTs from CLASS. For instance, needleleaf trees are subdivided into deciduous and evergreen types, while broadleaf trees are categorized as cold and drought-deciduous and evergreen types. Additionally, crops and grasses are differentiated based on their photosynthetic pathways, resulting in C3 and C4 versions. The names and corresponding acronyms of the PFTs used in this study are needleleaf evergreen tree (NDL.EVG), needleleaf deciduous tree (NDL.DCD), broadleaf evergreen tree (BDL.EVG), broadleaf deciduous cold tree (BDL.DCD.CLD), broadleaf deciduous dry tree (BDL.DCD.DRY), C3-photosynthesis grass (GRASS.C3), C4-photosynthesis grass (GRASS.C4), C3-photosynthesis crop (CROP.C3), and C4-photosynthesis grass (CROP.C4). The spatial allocation of PFTs is predefined.

Model inputs that vary in time include seven meteorological variables (downwelling SW radiation, downwelling LW radiation, surface precipitation rate, surface air pressure, specific humidity, air temperature, and wind speed), [CO_2], land cover, and population density. Another input variable is lightning density, which is based on climatological monthly values. The main

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

processes simulated by the biogeochemical component of CLASSIC and that are used in my study include photosynthesis, canopy conductance, tissue turnover, allocation of carbon, phenology and crop harvesting (Arora and Boer, 2005b), dynamic root distribution (Arora and Boer, 2003), maintenance, growth and heterotrophic respiration (Melton et al., 2015), wildfires (Arora and Boer, 2005a; Arora and Melton, 2018), and land use change (Arora and Boer, 2010). Note that the model's nitrogen cycle (Asaadi and Arora, 2021; Kou-Giesbrecht and Arora, 2022) has been turned off.

2.2 Parameter Uncertainty Ranges

There are about 100 CLASSIC model parameters, depending on the exact model configuration. To tune the model efficiently, my optimization focuses on the most influential parameters. S.N. et al. (2025) conducted a Global Sensitivity Analysis for CLASSIC, identifying the most influential parameters for carbon and turbulent heat fluxes for seven major biomes. While 20 parameters were consistently influential, variations across sites increased the total to 28 parameters. Many parameters have multiple values to account for different PFTs. Parameters with 12 values differentiate among nine PFTs, while those with four values distinguish between needleleaf trees, broadleaf trees, grasses, and crops. Parameters with two values separate trees from grasses or crops. As a result, the total number of parameter values tuned increases to 165. The present study optimizes these 165 parameters within their respective uncertainty ranges (Table 1). The two parameters that were frequently most influential were the maximum carboxylation rate (v_{max}) and the canopy extinction coefficient (k_n). Given the importance of both parameters, I have refined their respective uncertainty ranges using empirical data provided by the TRY database (Kattge et al., 2020) and literature.

The TRY data base provides v_{max} values for different plant species. Classifying these species into PFTs, I obtained PFT-specific values for needleleaf evergreen trees ($n = 998$), needleleaf deciduous trees ($n = 9$), broadleaf evergreen trees ($n = 2456$), broadleaf deciduous trees ($n = 3813$), C3 metabolism herbs/grasses ($n = 965$), and C4 metabolism herbs/grasses ($n = 11$). The PFT classification does not distinguish between cold versus dry broadleaf deciduous trees or grasses versus crops. Therefore, I applied the uncertainty ranges from broadleaf deciduous trees to both cold and dry broadleaf deciduous trees. Likewise, I applied the C3 metabolism herbs/grasses to both C3 grasses and C3 crops, and similarly for C4 metabolism herbs/grasses. For most PFTs, I defined the uncertainty range as the interquartile range of the observations. However, in the case of needleleaf deciduous trees, C4 grasses and C4 crops, I used the total observed range as the uncertainty range due to the small sample size. For C4 grasses and crops, I set the lower bound of the uncertainty range to the model's default value, which was lower compared to the smallest observed value.

The k_n uncertainty ranges are based on a meta-analysis presented by Zhang et al. (2014). Their study provides means and standard deviations of k_n during the whole growth season for sites located in needleleaf forest ($n = 15$), broadleaf forest ($n = 9$), grassland ($n = 17$) and croplands ($n = 35$). These values are utilized as uncertainty ranges for the PFTs. For C3 crops, I selected the default model parameter as the lower bound, which was lower compared to the value provided by the aforementioned meta-analysis. All other parameter uncertainty ranges were obtained from S.N. et al. (2025), who determined these ranges from literature and expert judgement.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

2.3 Model Parameter Optimization

Model parameter optimization is an iterative process in which a machine learning algorithm identifies parameter values that improve model performance over time (Figure 1a). The algorithm initially proposes multiple sets of randomly selected parameter values within their prescribed uncertainty ranges. These values, combined with historical climate data, are used to conduct simulations with CLASSIC for a selection of grid cells. The resulting model outputs are then compared against global Earth observations to assess how well each parameter set performs. Based on this evaluation, the machine learning algorithm proposes a new set of parameter values aimed at improving model performance. This iterative process continues until no further improvement is observed. Once the optimal parameter values are determined, I run the model globally using the optimized parameter values during the spinup, transient, and projected future simulations. The next few paragraphs describe how the machine learning algorithm works.

Model optimization can be performed using a variety of machine learning algorithms, each with its own strengths and limitations. This application employs a Genetic Algorithm (GA), implemented in the open-source programming language *R* by Scrucca (2013). A key advantage of GAs is their ability to explore a broad solution space, reducing the risk of getting trapped in local optima (Eiben and Smith, 2015). Additionally, they do not require the objective function to be continuous and can be efficiently parallelized on high-performance computing systems.

The defining feature of GAs is their ability to optimize parameter values by mimicking biological evolution. In this process, the set of parameters to be optimized is analogous to a chromosome, with each parameter representing a gene (Goldberg, 1989). The main steps of a GA are (i) selection of a population, (ii) crossover, and (iii) mutation. As a first step, the user specifies model parameter uncertainty ranges for all parameters that participate in the optimization process. In this adoption of a GA, I normalize parameter values such that they become dimensionless and vary between zero and one:

$$\hat{x_i} = \frac{x_i - \min}{\max - \min}, \tag{1}$$

where $\hat{x_i}$ is the normalized value of the parameter i, x_i is the actual parameter value, and $\min$ and $\max$ are the parameter's lower and upper boundary values, respectively. Early versions of GAs represented parameter values in binary form (Goldberg, 1989). However, I use the more common real-value representation, working directly with normalized parameter values. Each parameter is represented as an element in the vector $X = (x_1, x_2, ..., x_n)$, where each parameter x_i falls within its uncertainty range, spanning from zero to one. This vector represents the chromosome, with individual parameter values serving as genes.

The GA starts the optimization by producing a population of n chromosomes, where each chromosome consists of genes with values that are randomly chosen within the uncertainty range of 0 to 1. The population forms the first generation in the optimization process. The fitness function runs CLASSIC n-times using each of the n chromosomes as an input. The GA then calculates the relative fitness of each chromosome, as detailed in Section 2.4. This fitness value determines the selection of parent chromosomes, with well-performing chromosomes having a higher probability of being chosen. Selection can be performed in various ways. In this application, I use a method known as fitness-proportional selection with fitness linear scaling (Scrucca, 2013), where the selection probability P_i of a chromosome is given as:

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

$$P_i = \frac{f_i'}{\sum_{j=1}^{N} f_j'}, \tag{2}$$

where f_i' is the linearly scaled fitness of an individual chromosome i and N is the population size. The linearly scaled fitness is given by:

$$f_i' = a * f_i + b, \tag{3}$$

where f_i is the fitness of chromosome i and a and b are scaling coefficients that are computed as follows:

$$a = \bar{f}/\delta, \tag{4}$$

$$b = -f_{min} * \bar{f}/\delta, \tag{5}$$

$$\delta = \bar{f} - f_{min}, \tag{6}$$

where $\bar{f}$ and f_{min} are the mean and minimum fitness of all chromosomes, respectively. Once each chromosome has an assigned selection probability, the population is then randomly sampled (with replacement) where chromosomes with a higher selection probability are more likely to be selected. This is illustrated in Figure 1b, which shows the chromosomes being selected along with their corresponding fitness values. As a result, the best-performing strings get more copies, the average-performing strings stay even, and the worst-performing strings die off (Goldberg, 1989).

Selection alone does not improve solutions. The exploitation of current solutions is achieved through crossover, which can be performed in various ways. The method I have adopted is local arithmetic crossover, where crossover is applied to all genes, but the extent to which the offspring gene is influenced by Parent 1 or Parent 2 varies from gene to gene (Scrucca, 2013):

$$O_1 = \alpha P_1 + (1-\alpha) P_2, \tag{7}$$

$$O_2 = \alpha P_2 + (1-\alpha) P_1, \tag{8}$$

where α is a random value taken from a uniform distribution spanning between zero and one. Each parameter receives a different α value. Figure 1c shows how the values of Parent 1 and Parent 2 differ from Offspring 1 and Offspring 2. Here, the

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

values of each gene, represented as dots, are divided by the corresponding values from Parent 1. The figure demonstrates that the offspring values are bounded by the corresponding values from the parents. However, the exact location varies, as α differs among genes.

Next, the GA performs mutations that randomly alter genes. The purpose of mutation is to explore the whole parameter space, preventing the optimization from being trapped in a local maximum. The mutation function I used is the uniform random mutation, which randomly modifies one gene per chromosome (Figure 1d). After selection, crossover, and mutation, the new generation of chromosomes is used as inputs for the fitness function, and the process starts from the beginning.

The building block hypothesis offers an interpretation of why the seemingly random process of selection, crossover and mutation leads to parameter optimization. At the core of this hypothesis lies a concept referred to as schema. A schema is a parameter vector that restricts some parameters to subranges that are more narrow than the uncertainty ranges (e.g. $0.3 < x_i < 0.5$ rather than $0 < x_i < 1$). Schemata can differ in length, order, and fitness. A short schema constrains a small but meaningful subset of parameters. A low-order schema constrains parameters moderately, avoiding both excessive precision and excessive freedom. A high-fitness schema contains solutions that perform above average. The building block hypothesis states that short, low-order, and high-fitness schemata are more likely to propagate over generations (Eiben and Smith, 2015). These short, low-order, and high-fitness schemata form building blocks, which GAs assemble to create increasingly optimal solutions over time. Selection increases the proportion of high-fitness building blocks over generations. Crossover recombines information from two parent building blocks to create new offspring. Mutation introduces random changes to parameters, helping the algorithm explore new areas of the search space. The repeated selection, recombination, and mutation refine and assemble building blocks, leading to increasingly optimized solutions.

Parameter optimization is computationally expensive. To make the process feasible, I optimize parameters for 160 randomly selected model grid cells (Figure 2). These grid cells are distributed across the world's major biomes, with the number of grid cells per biome proportional to the biome's area. The total number of grid cells was chosen to be a multiple of the number of cores per compute node available on the high-performance computing (HPC) system, which provides 40 cores per node. Since each grid cell runs on a separate core, selecting a multiple of 40 ensures that the computation is carried out efficiently, minimizing idle cores and optimizing runtime. The optimization takes approximately two weeks to complete using 20 compute nodes on a Digital Research Alliance of Canada's HPC system (Beluga).

2.4 Model Evaluation

This section describes how I quantify model fitness. The approach is used as the cost function in the optimization process, as well as for measuring model performance when assessing the impact of parameter optimization on global runs. In both cases, performance is quantified using the Automated Model Benchmarking R package developed by Seiler et al. (2022), which quantifies model performance using a skill score system that is based on the International Land Model Benchmarking (ILAMB) framework (Collier et al., 2018). The method employs five scores that assess the model's annual mean bias (S_{bias}), monthly centralized root-mean-square-error (S_{rmse}), the timing of the seasonal peak (S_{phase}), inter-annual variability (S_{iav}), and spatial distribution (S_{dist}). The main steps for computing a score usually include (i) computing a dimensionless statistical

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

metric, (ii) scaling this metric onto a unit interval, and (iii) computing a spatial mean. All scores are dimensionless and range from zero to one, where increasing values imply better performance. These properties allow me to average skill scores across different statistical metrics in order to obtain an overall score for each variable ($S_{overall}$). A full description of AMBER is provided by Seiler et al. (2021).

The optimization is performed for multiple data streams simultaneously using six standard land surface variables: surface albedo (ALBS), gross primary productivity (GPP), latent heat flux (HFLS), sensible heat flux (HFSS), leaf area index (LAI), and land surface temperature (LST) (Table 2). To reduce the risk of overfitting, I rely on multiple observation-based reference datasets for each variable where available. Specifically, I use three products for ALBS (CERES, GEWEXSRB, and MODIS), two for GPP (FLUXCOM and GOSIF), and two each for HFLS and HFSS (CLASSr and FLUXCOM). For LAI, I include three datasets (AVHRR, Copernicus, and MODIS), while for LST, I use a single product (MODIS). These datasets are described in Seiler et al. (2022), with full references provided in Table 2.

To evaluate global simulations, I also compare my results against global NBP, which represents the net carbon flux between the terrestrial biosphere and the atmosphere. Although NBP is difficult to estimate directly, it can be derived as the difference between fossil fuel emissions (E_{FOS}) and the atmospheric growth rate (G_{ATM}), the ocean sink (S_{OCEAN}), and the cement carbonation sink (S_{CEMENT}), all of which are available from 1960 onwards:

$$NBP = E_{FOS} - G_{ATM} - S_{OCEAN} - S_{CEMENT}. \tag{9}$$

Friedlingstein et al. (2024) provides not only the annual values of each term but also the corresponding uncertainty ranges, expressed in terms of standard deviation (σ), for all terms except S_{CEMENT}. Using uncertainty propagation, I estimate the uncertainty of global NBP as follows:

$$\sigma_{NBP} = \left(\sigma^2_{E_{FOS}} + \sigma^2_{G_{ATM}} + \sigma^2_{S_{OCEAN}}\right)^{0.5}. \tag{10}$$

2.5 Robustness to Genetic Algorithm Settings

The parameter optimization results presented in this study are based on the default settings of the GA algorithm. To assess the robustness of these results, I examine how alternative GA configurations influence the optimization process. To keep computational costs manageable, this analysis is conducted using a reduced experimental setup, with a smaller subset of grid cells (40), a shorter optimization period (2001–2010), and a reduced number of generations (19).

The GA optimization algorithm offers six selection functions, five crossover functions, and four mutation functions, yielding 120 possible combinations. Exhaustively evaluating all of them is not feasible given the computational costs associated with running CLASSIC. A more efficient strategy is a step-wise approach where I vary one component at a time while keeping the others fixed. Specifically, I first test all six selection functions using the default crossover and mutation functions. Next, I test all five crossover functions while keeping the selection and mutation functions fixed at their defaults, followed by testing all four mutation functions with default selection and crossover settings. Results from these 13 experiments indicate that the

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

mutation function has the largest impact on model performance, followed by the selection function. Based on these findings, I fix the mutation function to the best-performing option and then test all selection functions, keeping the crossover function at its default (5 experiments). After identifying the best-performing selection function, I fix it as well and proceed to test the different crossover functions (4 experiments), this time using the previously identified optimal selection and mutation settings. I then use this ensemble of 22 experiments to evaluate how different settings influence parameter optimization.

3 Results

3.1 Optimization

As a first step, I assessed the effectiveness of the optimization through a synthetic data test. Rather than comparing the model output to global Earth observations, I compared it to model output generated with a known set of parameter values. If the optimization successfully recovers these known values, the process can be considered effective. The test was conducted using the nine default parameter values for the maximum carboxylation rate. Figure 3 demonstrates that the optimization accurately reproduces these parameter values, even when the true parameter values coincide with the minimum of the uncertainty range, as observed for C4 grasses and C4 crops.

Next, I performed the optimization using all 28 parameters and 13 observation-based datasets. To enhance the efficiency of the optimization, the first chromosome of the initial generation contains the default parameter values. Figure 4 illustrates how the median AMBER score improves from generation to generation, surpassing the default score that results from using default parameter values after approximately ten generations (denoted by the red line) and levelling off after about 20 generations. The score spread among population members (grey-shaded area) decreases over generations until the interquartile range of the population almost matches that of the best-performing chromosome (stippled line), indicating that minimal improvement can be expected beyond 25 generations.

Comparing default against optimized parameter values shows that the optimization affected many parameter values, with differences of up to 60% (Figure 4a). For the maximum carboxylation rate, optimization increased the parameter values for some PFTs (needleleaf evergreen trees, broadleaf evergreen trees, broadleaf dry deciduous trees, and C3 crops), left them unchanged for C4 grasses, and decreased them for the remaining PFTs (needleleaf deciduous trees, broadleaf cold deciduous trees, C4 crops, and C3 grass).

Figure 4b shows the normalized parameter values for the best-performing chromosome in each generation. The initial value on the left corresponds to the default parameter settings, while the final value on the right represents the optimized values. Over the course of 25 generations, the parameter values of the best-performing chromosome changed six times, with each adjustment improving model performance.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

3.2 Global Model Performance

Using the optimized parameter values obtained above, I conducted a global simulation in which both the spinup and transient runs were based on the optimized parameter values. The results were compared against global Earth observations, and the resulting scores were evaluated against those from a simulation using the default parameter values. Figure 6 shows that, in most instances, model optimization improved the scores. This is particularly evident for leaf area index and sensible heat flux, where improvements were observed across multiple statistical metrics (bias, centralized root-mean-square error, seasonality, interannual variability, and spatial correlation). Gross primary productivity also showed improved scores for bias and centralized root-mean-square error. While some variables did not improve, the score reductions were not statistically significant at the 5% level and were less than 0.01. Plotting the score differences as a histogram shows that, in most cases, the scores improved (up to 0.04), while any deterioration in performance remained minimal ($<$0.01) (Figure 7).

Figure 8 illustrates the difference between the optimized and default absolute biases, where negative values indicate a reduction in absolute bias due to optimization. For leaf area index, GPP, and sensible heat flux, optimization reduced biases across all biomes, with improvements exceeding 1 m^2 m^{-2}, 0.5 gC m^{-2} day^{-1}, and 5 W m^{-2}, respectively. In the case of latent heat flux, the optimization effect is smaller compared to that for sensible heat flux (up to 2.5 W m^{-2}) and is concentrated in fewer regions (not shown).

3.3 Net Biome Productivity

In this section, I evaluate how optimization affects global NBP for the recent past and projected future. A comparison of the default and optimized NBP values reveals that neither simulation perfectly replicates the peaks and troughs of the reference NBP provided by the GCB 2024 (GCB2024 NBP, hereafter) Friedlingstein et al. (2024) (Figure 9a). However, the interannual variability produced by both simulations is similar to observed patterns. Analyzing the 20-year running means shows that both simulations are, for the most part, within the uncertainty range of observations (Figure 9b). Nonetheless, both simulations exhibit a weaker trend compared to observations, and the optimized simulation consistently produces lower NBP values than the default run.

Focusing on the NBP accumulated between 1960 and 2022 reveals that the optimization reduces cumulative NBP by 33% (Figure 9c). The default run exceeds, while the optimized run falls below, the reference NBP estimate of 55 PgC (68 PgC and 45 PgC, respectively). Comparing the time series using a Taylor plot shows that the performances of the default and optimized runs are similar (Figure 9d). The optimized run has a lower standard deviation, bringing it closer to the reference NBP (blue arcs). Additionally, the optimized run has a slightly lower root-mean-square error (grey arcs) but also a slightly lower correlation coefficient (stippled lines). The distance between the optimized run (red dot) and the reference NBP (grey dot) is somewhat smaller than the distance between the default run (black dot) and the reference NBP, indicating that optimization results in a modest overall improvement in performance. However, the differences are too minor to be considered meaningful. Examining future projections under the fossil fuel-intensive SSP5-8.5 scenario shows that the cumulative NBP for the 2015-2100 period

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

is 43% lower in the optimized compared to the default run (Figure 9e). The absolute difference is 89 PgC, equivalent to almost nine years of current emissions.

The optimization reduces global NBP because of a reduction in GPP and an increase in autotrophic respiration. Focusing on the CRUJRA-driven simulation, optimization reduces GPP by 3% and increases autotrophic respiration by 3% compared to the default simulation (Table 3). This effect is only partially offset by reduced heterotrophic respiration (-6%) and lower wildfire emissions (-33%). Examining carbon stocks reveals that optimization decreases the vegetation carbon pool (-16%), which aligns with the reduced GPP and increased autotrophic respiration. The decline in GPP also leads to a reduction in the soil organic carbon pool (-12%), explaining the observed decrease in heterotrophic respiration. Similar patterns emerge in the CanESM5-driven simulations (Table 3).

Carbon fluxes and pools are not independent. For instance, an increase of GPP increases biomass and autotrophic respiration. The increase in biomass leads to more litter, which increases soil organic carbon and heterotrophic respiration. Also, an increase in biomass enhances emissions from wildfires. Let us now identify which process dominates the impact of optimization on NBP. Table 4 presents a correlation matrix for the differences between the optimized and default variables. For example, ΔNBP represents the difference between the optimized and default cumulative NBP. Focusing on the CRUJRAv2-driven simulations reveals that ΔNBP is most strongly correlated with differences in GPP (ΔGPP) (R = 0.37) (Table 4). The correlation coefficient between ΔNBP and heterotrophic respiration (ΔRH) is also relatively high (R = 0.33); however, this does not necessarily imply that ΔRH drives ΔNBP. First, the correlation coefficient is positive rather than negative, and second, ΔRH is strongly correlated with ΔGPP (R = 0.88), suggesting that it primarily responds to changes in GPP.

The correlations between ΔNBP and differences in autotrophic respiration (ΔRA) (R = 0.05) and wildfire emissions (ΔfFire) (R = -0.08) are minor. In both cases, the differences in these fluxes are more strongly correlated with ΔGPP (R = 0.47). This indicates that the impact of optimization on NBP is largely driven by differences in GPP, with variations in the other fluxes arising mainly as a response to GPP. Similar patterns are observed in the CanESM5-driven runs (Table 4), although here the influence of wildfire emissions is more pronounced (R = -0.24 for the correlation between ΔNBP and ΔfFire).

After evaluating the impact of optimization on global NBP, we now shift the focus to regional effects. During the historical period simulated by CLASSIC with CRUJRAv2 forcing, the strongest carbon sinks are concentrated in tropical rainforests, with additional contributions from the Eastern U.S. (Figure 10a). Meanwhile, the strongest carbon sources are located in parts of Southern Brazil, Western Africa, Indonesia, and the Midwest U.S. This spatial pattern is consistent across both the default and optimized simulations (Figure 10a and c). However, optimization generally reduces NBP in regions where ecosystems act as net sinks and increases it in areas that function as sources (Figure 10e). For example, optimized NBP is lower in the Amazon and higher (i.e., less negative) in Indonesia. This adjustment effectively narrows the contrast between carbon sinks and sources (Figure 10g).

Simulations under future climate conditions (SSP5-8.5) indicate that while tropical forests in Central Africa and Southeast Asia become stronger carbon sinks, the Amazon rainforest shifts from a carbon sink to a carbon source (Figure 10b). The same simulation also projects more pronounced carbon sinks at higher latitudes. These patterns are consistent across both the default

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

and optimized simulations. However, as before, optimization generally reduces sink strength and weakens sources, thereby narrowing the contrast between carbon sinks and sources (Figure 10h).

3.4 Robustness to Genetic Algorithm Settings

The results presented above are based on the default GA configuration, which employs *fitness-proportional selection with fitness linear scaling* as a fitness function, *local arithmetic crossover* as a crossover function, and *uniform random mutation* as a mutation function. However, alternative selection, crossover, and mutation strategies could also be applied. This section examines the impact of 22 different GA settings on optimization performance within a reduced experimental setup. Figure 11a compares AMBER scores across configurations, with each line representing a different GA setting. The colour of each line indicates the correlation between the normalized parameter values of that configuration and those of the default setting, based on the best-performing chromosome in the final generation. The figure illustrates that optimization success is highly sensitive to the choice of GA configuration. While the default GA setting performs well, two other configurations yield even better results. The correlation coefficients range from –0.09 to +0.67, indicating that the resulting parameter values can differ substantially from those obtained using the default setting. Notably, the highest correlation is observed for the best-performing configuration, suggesting strong similarity between its parameter values and those of the default. The performance of the default GA setting aligns closely with that of the majority of configurations (12 out of 22), suggesting that it offers a good balance between effectiveness and general applicability (Figure 11b).

Interestingly, configurations that yield performance similar to the default GA setting are not necessarily strongly correlated with the default parameter values. This may indicate that comparable performance can be achieved through different combinations of parameters, or that only a subset of parameters strongly influences model performance. To explore this further, I calculated the interquartile range (IQR) for each of the 165 parameters, providing insight into how their values vary across GA configurations. The IQRs vary widely, from less than 0.1 to 0.8 (Figure 11c), suggesting that some parameters are tightly constrained by the optimization process, while others are not. Focusing on the ten parameters with the smallest IQRs, I found that in seven cases, the default parameter values fall within these narrow interquartile ranges (Figure 11d). Assuming that these tightly constrained parameters are key drivers of model performance, this result suggests that the default configuration captures the essential dynamics of the system. Re-running the full optimization with a slightly better-performing GA configuration would likely lead to a different, but still suboptimal solution, without materially affecting the conclusions. Together, these findings underscore the robustness of the optimization results and support the use of the default GA settings in the main analysis.

4 Discussion

This study demonstrates the effectiveness of optimizing poorly constrained land surface model parameters using machine learning and global Earth observations. The results show that optimization enhances model performance considerably, particularly for GPP, leaf area index, and sensible heat fluxes. Using a prolonged optimization period (1701–2022), this improvement is

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

achieved without the need to switch parameter values between the spinup and transient runs or the use of global scaling factors (Schürmann et al., 2016; MacBean et al., 2018). As a result, the same parameter values can be consistently applied in future climate change projections. Evaluating the impacts on the terrestrial carbon sink shows that optimization weakens the cumulative NBP by 33% from 1960 to 2022 under observed climate conditions and by 43% from 2015 to 2100 under the SSP5-8.5 scenario. This reduction arises from the combined effects of decreased GPP and increased autotrophic respiration. Regionally, the optimization reduces the contrasts between carbon sinks and sources by lowering GPP and enhancing autotrophic respiration in strong sinks, while having the opposite effect in regions that act as strong carbon sources. The following paragraphs explore these results in greater detail, discuss the study's limitations, and highlight opportunities for future research. Furthermore, I will argue that advanced methods of regular parameter tuning should become an integral part of the model development cycle. Finally, I will examine the broader implications of this research for climate change policy.

One of the key questions is whether the optimized NBP is more reliable than the uncalibrated NBP. On the one hand, optimization improves model performance, suggesting that the optimized fluxes are more reliable. On the other hand, optimization neither improves nor deteriorates global NBP estimates when compared against GCB2024 NBP (Friedlingstein et al., 2024). More importantly, neither the default nor the optimized simulations reproduce the GCB2024 NBP trend. One way to improve simulated global NBP directly would be to optimize CLASSIC parameters using the GCB2024 NBP estimates as a target. However, this would require running the model for all grid cells during the optimization process, which is computationally prohibitive. A feasible alternative would be to conduct global optimization using an emulator, a statistical representation of the model (Holden et al., 2015). Emulators require minimal computational resources, making it possible to optimize parameters for global applications (McNeall et al., 2024). Additionally, the reduced computational costs would allow for extended simulation times, including the model spin-up phase. The emulator could also be used to assess how the optimization process is influenced by all possible combinations of selection, crossover, and mutation functions, as well as their respective hyperparameters.

The impacts of optimization vary across different variables. The most significant improvements are observed in GPP, leaf area index, and sensible heat fluxes. In contrast, no noticeable changes in performance are detected for albedo and land surface temperature. This outcome aligns with the selection of parameters being tuned, which were chosen based on a global sensitivity analysis that identified the most influential parameters for carbon and turbulent heat fluxes (S.N. et al., 2025). Including additional variables in the optimization helps ensure that the enhancements in carbon and heat fluxes do not compromise model performance elsewhere. However, improving albedo and land surface temperature would require a broader or different selection of parameters.

Optimization leads to a reduction in carbon stocks, raising the question of whether the optimized values remain consistent with observations. Global datasets synthesizing empirical measurements vary considerably, with reported ranges spanning from 264.6 to 482.5 PgC for vegetation carbon and 1143.4 to 2708.0 PgC for soil organic carbon (Seiler et al., 2022). In the CRUJRAv2-driven run, the optimized vegetation carbon is 395 PgC, which falls within this range. However, the optimized soil organic carbon (1026 PgC) is lower than the reported range. This study did not include carbon stocks as target datasets in the optimization process due to the large uncertainties in these products. Moreover, comparing modeled carbon stocks with reference data requires additional processing, including extracting the top 1.5 m soil layers and distinguishing above-ground

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

from below-ground biomass to align model outputs with observations. This adds to the already substantial computational burden, as this processing must be repeated for each of the 2500 simulations. However, future optimization efforts could assess how incorporating carbon stocks as constraints affects the optimization outcomes.

The CLASSIC configuration used in the present study prescribes the maximum carboxylation rate. Alternatively, the model can calculate the maximum carboxylation rate based on the dynamic interaction between the carbon and nitrogen cycles. The nitrogen cycle is particularly important for accounting for the modulating effect of nitrogen limitation on the CO_2 fertilization effect (Thornton et al., 2007). In the case of CLASSIC, the nitrogen cycle leads to a weaker carbon sink that absorbs about 50% less carbon between 2015 and the end of this century under SSP5-8.5 (Kou-Giesbrecht and Arora, 2022). Past research has shown that model performance deteriorates considerably when the model's nitrogen cycle is activated (Arora et al., 2023). The large uncertainty of nitrogen cycle parameters presents an opportunity to refine them using the parameter optimization system introduced here.

This research evaluates climate change projections using meteorological forcing from the CanESM5 parent model. Consequently, I do not fully explore the potential impacts of climate change across the range of changes produced by different ESMs. The projected NBP changes should therefore be understood as the response specific to CanESM5. For example, CanESM5 projects a future decline in precipitation in the Amazon basin, resulting in a strong regional carbon source, while projecting an increase in precipitation in Central Africa, leading to a regional carbon sink (Seiler et al., 2024). These projections will differ depending on the choice of forcing data, given the large inter-model differences in future precipitation projected for the tropics (Eyring et al., 2021). An analysis of future NBP projections made by a multimodel ensemble is provided by Canadell et al. (2021).

Furthermore, I evaluate the impact of only a single, fossil fuel-intensive emission scenario (SSP5-8.5). The SSP5-8.5 scenario falls under climate category eight (C8), characterized by warming exceeding 4°C compared to pre-industrial levels (Riahi et al., 2022). In contrast, business-as-usual scenarios, also referred to as reference scenarios, such as *CurPol*, explore the consequences of continuing along the path of climate policies implemented in 2020, with only gradual strengthening thereafter. These scenarios fall under climate category seven (C7), which limits warming to 4°C compared to pre-industrial levels throughout the 21st century. It may therefore be argued that SSP5-8.5 is less likely to occur given current trends in climate change mitigation policies (Hausfather and Peters, 2020). Nonetheless, the advantage of using the SSP5-8.5 scenario in this study is that it provides a stronger climate signal and enables the exploration of impacts in a worst-case scenario.

The results demonstrate that parameter optimization has a strong influence on net carbon fluxes, aligning with previous studies that have reported even more pronounced changes in terms of net ecosystem exchange or NBP (Peylin et al., 2016; Schürmann et al., 2016; Bacour et al., 2023). Under future climate conditions, the optimized carbon fluxes lead to a land carbon sink that absorbs 43% less carbon between 2015 and the end of this century. The magnitude of this impact is comparable to adding a new process to the model. For instance, comparing simulations with and without dynamic carbon-nitrogen interaction shows that the nitrogen cycle leads to a weaker carbon sink that absorbs about 50% less carbon between 2015 and the end of this century under SSP5-8.5 (Kou-Giesbrecht and Arora, 2023).

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

The large impact of parameter optimization on net carbon fluxes presents a compelling argument for modelling groups to incorporate advanced optimization methods into their model development cycle. However, it must be noted that model tuning can be a double-edged sword. On one hand, optimization ensures that the model performs as well as possible for the given parameter uncertainty ranges. This reduces the likelihood that mismatches between model output and observations are due to parameter uncertainty, making it easier to identify issues such as coding errors, poor process representation, or missing processes. On the other hand, optimization can mask deficiencies by compensating for model limitations, potentially making the model appear more accurate than it actually is and complicating efforts to diagnose problems. Moreover, interactions between parameters give rise to equifinality, where different combinations of parameter values can produce similar outcomes (Beven and Freer, 2001). However, these challenges are not unique to automated optimization, they also apply to traditional manual tuning. When a model developer incorporates a new process, they typically adjust parameter values to align the model output with reference data, which can similarly obscure underlying deficiencies. Moreover, as models evolve, parameter values that were previously selected may no longer be the most appropriate. For these reasons, adopting more advanced parameter optimization techniques in the model development cycle is a logical and necessary step toward improving model accuracy and reliability.

The model configuration used here is an offline simulation with prescribed [CO_2] and meteorological forcing. Therefore, the impacts of optimization on NBP do not alter atmospheric [CO_2] or the climate. In a fully coupled simulation, lower NBP would imply a faster increase in [CO_2] and temperature, affecting NBP in return. This feedback could be evaluated in an emissions-driven simulation where CLASSIC runs within CanESM and the carbon cycle is fully coupled. Such simulations would not only be scientifically relevant but also important for global climate change mitigation policy, as explained next.

The goal of the 2015 Paris Agreement is to limit the warming of global mean surface temperature to well below 2°C and preferably 1.5°C compared to pre-industrial values (United Nations Framework Convention on Climate Change (UNFCCC), 2015). The remaining carbon budget that is consistent with the 1.5°C global warming target is estimated at 200 $PgCO_2$ (55 PgC) for the year 2024 (>50% chance) (Forster et al., 2024). Current CO_2 emissions are about 40 $PgCO_2$ per year. If emissions remain at present levels, we will likely exhaust the remaining carbon budget consistent with the 1.5°C warming target by the end of this decade.

The remaining carbon budget is calculated using the Model for the Assessment of Greenhouse Gas Induced Climate Change (MAGICC) emulator (Meinshausen et al., 2011). MAGICC is a simple carbon cycle-climate model with a hemispherically averaged upwelling-diffusion ocean coupled to an atmospheric layer and a globally averaged carbon cycle model. Version 7.0 is calibrated against ESMs, such as CanESM5, which are part of the Coupled Climate-Carbon Cycle Model Intercomparison Project (C4MIP) (Jones et al., 2016; Meinshausen et al., 2020). Any advancements in CanESM could therefore improve future estimates of the remaining carbon budget and, consequently, influence climate change mitigation policy. This further strengthens the case for adopting more advanced methods of parameter optimization.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

5 Conclusions

In conclusion, this study demonstrates the effectiveness of optimizing poorly constrained land surface model parameters using machine learning and global Earth observations. The optimization improves model performance considerably, especially for GPP, leaf area index, and sensible heat fluxes. The optimization also weakens the terrestrial carbon sink, reducing the amount of carbon absorbed between 2015 and 2100 by 43%. This substantial impact underscores the need to address parameter uncertainty to minimize inter-model differences. To enhance model accuracy and reliability, modelling groups should integrate advanced parameter optimization methods into their development cycle. Doing so will not only improve model performance but also support more robust climate change mitigation policies.

Code and data availability. The DAISY R package can be accessed at https://github.com/cseilerQueens/daisy.git, and the parameter uncertainty ranges, scripts for running DAISY, and scripts for analyzing outputs are available at https://github.com/cseilerQueens/data-assimilation. A repository containing all scripts, optimization outputs, figures, and global Earth observations used in the optimization process is available at https://doi.org/10.5281/zenodo.15009417. The Genetic Algorithm package is available at https://cran.r-project.org/web/packages/GA/index.html. The AMBER code can be accessed at https://gitlab.com/cseiler/AMBER. A repository containing all global Earth observations used for evaluating global model performance is available at https://doi.org/10.5281/zenodo.7799563. The CLASSIC code can be accessed at https://doi.org/10.5281/zenodo.3522407. The CLASSIC input data can be downloaded from https://crudata.uea.ac.uk/cru/data/hrg/ (CRU-JRAv2) and https://data.isimip.org (bias-adjusted CanESM5).

Author contributions. C.S. conceived the study, performed the analysis, and wrote the manuscript.

Competing interests. The author declares no competing interests.

Acknowledgements. I thank Luca Scrucca for insightful discussions on the Genetic Algorithm he developed and that I adopted. I am also grateful to Andrew Adams for his technical support in parallelizing the code on a Digital Research Alliance of Canada High-Performance Computer. I appreciate my discussions with Natasha McBean on data assimilation and thank Raj S.N. for providing the list of influential parameters and their associated uncertainty ranges. I acknowledge the support of the Canadian federal government scientists who are part of the CLASSIC team. This research was partially enabled by the Digital Research Alliance of Canada (alliancecan.ca) and was funded by a Research Initiation Grant from Queen's University.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

References

Arora, V. K. and Boer, G. J.: A Representation of Variable Root Distribution in Dynamic Vegetation Models, Earth Interact., 7, 1–19, 2003.

Arora, V. K. and Boer, G. J.: Fire as an interactive component of dynamic vegetation models, Journal of Geophysical Research: Biogeosciences, 110, 2005a.

Arora, V. K. and Boer, G. J.: A parameterization of leaf phenology for the terrestrial ecosystem component of climate models, Glob. Chang. Biol., 2005b.

Arora, V. K. and Boer, G. J.: Uncertainties in the 20th century carbon budget associated with land use change, Glob. Chang. Biol., 16, 3327–3348, 2010.

Arora, V. K. and Melton, J. R.: Reduction in global area burned and wildfire emissions since 1930s enhances carbon uptake by land, Nat. Commun., 9, 1326, 2018.

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., et al.: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models, Biogeosciences, 17, 4173–4222, 2020.

Arora, V. K., Seiler, C., Wang, L., and Kou-Giesbrecht, S.: Towards an ensemble-based evaluation of land surface models in light of uncertain forcings and observations, Biogeosciences, 20, 1313–1355, 2023.

Asaadi, A. and Arora, V. K.: Implementation of nitrogen cycle in the CLASSIC land model, Biogeosciences, 18, 669–706, 2021.

Bacour, C., MacBean, N., Chevallier, F., Léonard, S., Koffi, E. N., and Peylin, P.: Assimilation of multiple datasets results in large differences in regional- to global-scale NEE and GPP budgets simulated by a terrestrial biosphere model, Biogeosciences, 20, 1089–1111, 2023.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol. (Amst.), 249, 11–29, 2001.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C.: A Limited Memory Algorithm for Bound Constrained Optimization, SIAM J. Sci. Comput., 16, 1190–1208, 1995.

Canadell, J., Monteiro, P., Costa, M., Cotrim da Cunha, L., Cox, P., Eliseev, A., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P., Piao, S., Rogelj, J., Syampungani, S., Zaehle, S., and Zickfeld, K.: Global Carbon and other Biogeochemical Cycles and Feedbacks. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, pp. 673–816, 2021.

Claverie, M., Matthews, J. L., Vermote, E. F., and Justice, C. O.: A 30+ Year AVHRR LAI and FAPAR Climate Data Record: Algorithm Description and Validation, Remote Sensing, 8, 263, 2016.

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., and Randerson, J. T.: The International Land Model Benchmarking (ILAMB) system: design, theory, and implementation, Journal of Advances in Modeling Earth Systems, 10, 2731–2754, 2018.

Duffy, K. A., Schwalm, C. R., Arcus, V. L., Koch, G. W., Liang, L. L., and Schipper, L. A.: How close are we to the temperature tipping point of the terrestrial biosphere?, Sci Adv, 7, 2021.

Eiben, A. E. and Smith, J. E.: Introduction to evolutionary computing, Springer, 2015.

Eyring, V., Gillett, N., Achuta Rao, K., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P., Kosaka, Y., McGregor, S., Min, S., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System. In Climate Change 2021: The Physical Science

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, pp. 423–552, 2021.

Forster, P. M., Smith, C., Walsh, T., Lamb, W. F., Lamboll, R., Hall, B., Hauser, M., Ribes, A., Rosen, D., Gillett, N. P., Palmer, M. D., Rogelj, J., von Schuckmann, K., Trewin, B., Allen, M., Andrew, R., Betts, R. A., Borger, A., Boyer, T., Broersma, J. A., Buontempo, C., Burgess, S., Cagnazzo, C., Cheng, L., Friedlingstein, P., Gettelman, A., Gütschow, J., Ishii, M., Jenkins, S., Lan, X., Morice, C., Mühle, J., Kadow, C., Kennedy, J., Killick, R. E., Krummel, P. B., Minx, J. C., Myhre, G., Naik, V., Peters, G. P., Pirani, A., Pongratz, J., Schleussner, C.-F., Seneviratne, S. I., Szopa, S., Thorne, P., Kovilakam, M. V. M., Majamäki, E., Jalkanen, J.-P., van Marle, M., Hoesly, R. M., Rohde, R., Schumacher, D., van der Werf, G., Vose, R., Zickfeld, K., Zhang, X., Masson-Delmotte, V., and Zhai, P.: Indicators of Global Climate Change 2023: annual update of key indicators of the state of the climate system and human influence, Earth Syst. Sci. Data, 16, 2625–2658, 2024.

Fox, A. M., Hoar, T. J., Anderson, J. L., Arellano, A. F., Smith, W. K., Litvak, M. E., MacBean, N., Schimel, D. S., and Moore, D. J. P.: Evaluation of a data assimilation system for land surface models using CLM4.5, J. Adv. Model. Earth Syst., 10, 2471–2494, 2018.

Friedlingstein, P., O#39;Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., and Zeng: Global carbon budget 2024, Earth System Science Data Discussions, 2024, 1–133, 2024.

Gauthier, C. B., Melton, J. R., Meyer, G., SN, R. D., and Sonnentag, O.: Parameter optimization for global soil carbon simulations: Not a simple problem, Authorea Preprints, 2024.

Goldberg, D.: Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley Longman Publishing Co., Inc, 1989.

Hausfather, Z. and Peters, G. P.: Emissions–the 'business as usual' story is misleading, Nature, 577, 618–620, 2020.

Hobeichi, S., Abramowitz, G., and Evans, J.: Conserving Land-Atmosphere Synthesis Suite (CLASS), J. Clim., 2019.

Holden, P. B., Edwards, N. R., Garthwaite, P. H., and Wilkinson, R. D.: Emulation and interpretation of high-dimensional climate model outputs, Journal of Applied Statistics, 42, 2038–2055, 2015.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., and Zaehle, S.: C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6, Geosci. Model Dev., 9, 2853–2880, 2016.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Sci Data, 6, 74, 2019.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., et al.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17, 1343–1365, 2020.

Kallingal, J. T., Lindström, J., Miller, P. A., Rinne, J., Raivonen, M., and Scholze, M.: Optimising CH_4 simulations from the LPJ-GUESS model v4.1 using an adaptive Markov chain Monte Carlo algorithm, Geosci. Model Dev., 17, 2299–2324, 2024.

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., Yu, L., and Weller, R. A.: Surface Irradiances Consistent with CERES-Derived Top-of-Atmosphere Shortwave and Longwave Irradiances, J. Clim., 26, 2719–2740, 2013.

Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Tautenhahn, S., Werner, G. D. A., Aakala, T., Abedi, M., Acosta, A. T. R., Adamidis, G. C., and et al.: TRY plant trait database - enhanced coverage and open access, Glob. Chang. Biol., 26, 119–188, 2020.

Kou-Giesbrecht, S. and Arora, V.: Compensatory effects between CO2, nitrogen deposition, and temperature in terrestrial biosphere models without nitrogen compromise projections of the future terrestrial carbon sink, Geophysical Research Letters, 50, e2022GL102 618, 2023.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

Kou-Giesbrecht, S. and Arora, V. K.: Representing the dynamic response of vegetation to nitrogen limitation via biological nitrogen fixation in the CLASSIC land model, Global Biogeochem. Cycles, 36, 2022.

Koven, C. D., Arora, V. K., Cadule, P., Fisher, R. A., Jones, C. D., Lawrence, D. M., Lewis, J., Lindsay, K., Mathesius, S., Meinshausen, M., Mills, M., Nicholls, Z., Sanderson, B. M., Séférian, R., Swart, N. C., Wieder, W. R., and Zickfeld, K.: Multi-century dynamics of the climate and carbon cycle under both high and net negative emissions scenarios, Earth Syst. Dyn., 13, 885–909, 2022.

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A. D.: Constraining a global ecosystem model with multi-site eddy-covariance data, Biogeosci. Discuss., 9, 3317–3380, 2012.

Li, X. and Xiao, J.: Mapping Photosynthesis Solely from Solar-Induced Chlorophyll Fluorescence: A Global, Fine-Resolution Dataset of Gross Primary Production Derived from OCO-2, Remote Sensing, 11, 2563, 2019.

MacBean, N., Maignan, F., Bacour, C., Lewis, P., Peylin, P., Guanter, L., Köhler, P., Gómez-Dans, J., and Disney, M.: Strong constraint on modelled global carbon uptake using solar-induced chlorophyll fluorescence data, Sci. Rep., 8, 1973, 2018.

MacBean, N., Bacour, C., Raoult, N., Bastrikov, V., Koffi, E. N., Kuppel, S., Maignan, F., Ottlé, C., Peaucelle, M., Santaren, D., and Peylin, P.: Quantifying and reducing uncertainty in global carbon cycle predictions: Lessons and perspectives from 15 years of data assimilation studies with the ORCHIDEE terrestrial biosphere model, Global Biogeochem. Cycles, 36, 2022.

McNeall, D., Robertson, E., and Wiltshire, A.: Constraining the carbon cycle in JULES-ES-1.0, Geosci. Model Dev., 17, 1059–1089, 2024.

Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration, Atmos. Chem. Phys., 11, 1417–1456, 2011.

Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P. J., Reimann, S., Smith, S. J., van den Berg, M., Velders, G. J. M., Vollmer, M. K., and Wang, R. H. J.: The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500, Geosci. Model Dev., 13, 3571–3605, 2020.

Melton, J. R., Shrestha, R. K., and Arora, V. K.: The influence of soils on heterotrophic respiration exerts a strong control on net ecosystem productivity in seasonally dry Amazonian forests, Biogeosciences, 12, 1151–1168, 2015.

Melton, J. R., Arora, V. K., Wisernig-Cojoc, E., Seiler, C., Fortier, M., Chan, E., and Teckentrup, L.: CLASSIC v1.0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – Global sensitivity and site-level performance, Geoscientific Model Development, 13, 2825–2850, https://doi.org/10.5194/gmd-13-2825-2020, 2020.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, Remote Sens. Environ., 83, 214–231, 2002.

Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maignan, F., Chevallier, F., Ciais, P., and Prunet, P.: A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle, Geosci. Model Dev., 9, 3321–3346, 2016.

Raoult, N. M., Jupp, T. E., Cox, P. M., and Luke, C. M.: Land-surface parameter optimisation using data assimilation techniques: the adJULES system V1.0, Geosci. Model Dev., 9, 2833–2852, 2016.

Riahi, K., Schaeffer, R., Arango, J., Calvin, K., Guivarch, C., T., H., Jiang, K., E., K., R., M., Peters, G., A., R., Robertson, S., A.M., S., J., S., M., T., and van Vuuren, D.: Mitigation pathways compatible with long-term goals. In IPCC, 2022: Climate Change 2022: Mitigation of

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2022.

Schürmann, G. J., Kaminski, T., Köstler, C., Carvalhais, N., Voßbeck, M., Kattge, J., Giering, R., Rödenbeck, C., Heimann, M., and Zaehle, S.: Constraining a land-surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System V1.0, Geosci. Model Dev., 9, 2999–3026, 2016.

Scrucca, L.: GA: A Package for Genetic Algorithms in R, Journal of Statistical Software, 53, 1–37, https://doi.org/10.18637/jss.v053.i04, 2013.

Seiler, C., Melton, J. R., Arora, V. K., and Wang, L.: CLASSIC v1.0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – Part 2: Global benchmarking, Geoscientific Model Development, 14, 2371–2417, 2021.

Seiler, C., Melton, J. R., Arora, V. K., Sitch, S., Friedlingstein, P., Anthoni, P., Goll, D., Jain, A. K., Joetzjer, E., Lienert, S., et al.: Are terrestrial biosphere models fit for simulating the global land carbon sink?, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 946, 2022.

Seiler, C., Kou-Giesbrecht, S., Arora, V. K., and Melton, J. R.: The Impact of Climate Forcing Biases and the Nitrogen Cycle on Land Carbon Balance Projections, Journal of Advances in Modeling Earth Systems, 16, e2023MS003 749, 2024.

S.N., R. D., Seiler, C., and Monahan, A. H.: Global Sensitivity Analysis of the Historical Carbon Sink Across Biomes, Submitted to Atmosphere-Ocean, 2025.

Stackhouse, Jr, P. W., Gupta, S. K., Cox, S. J., Zhang, T., Mikovitz, J. C., and Hinkelman, L. M.: The NASA/GEWEX surface radiation budget release 3.0: 24.5-year dataset, Gewex news, 21, 10–12, 2011.

Strahler, A. H., Muller, J., Lucht, W., Schaaf, C., and others: MODIS BRDF/albedo product: algorithm theoretical basis document version 5.0, MODIS, 1999.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., Salzen, K. v., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geoscientific Model Development, 12, 4823–4873, 2019.

Tarantola, A.: Inverse problem theory and methods for model parameter estimation, SIAM, 2005.

Thornton, P. E., Lamarque, J.-F., Rosenbloom, N. A., and Mahowald, N. M.: Influence of carbon-nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability, Global Biogeochem. Cycles, 21, 2007.

United Nations Framework Convention on Climate Change (UNFCCC): Paris Agreement, https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement, adopted at the 21st Conference of the Parties to the UNFCCC (COP21), Paris, France, 2015.

van Nes, E. H., Scheffer, M., Brovkin, V., Lenton, T. M., Ye, H., Deyle, E., and Sugihara, G.: Causal feedbacks in climate change, Nat. Clim. Chang., 5, 445–448, 2015.

Verger, A., Baret, F., and Weiss, M.: Near real-time vegetation monitoring at global scale, IEEE Journal of Selected Topics in, 2014.

Wan, Z., Hook, S., and Hulley, G.: MODIS/Terra Land Surface Temperature/Emissivity Monthly L3 Global 0.05Deg CMG V061, https://doi.org/10.5067/modis/mod11c3.061, 2021.

Zhang, L., Hu, Z., Fan, J., Zhou, D., and Tang, F.: A meta-analysis of the canopy light extinction coefficient in terrestrial ecosystems, Front. Earth Sci., 8, 599–609, 2014.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

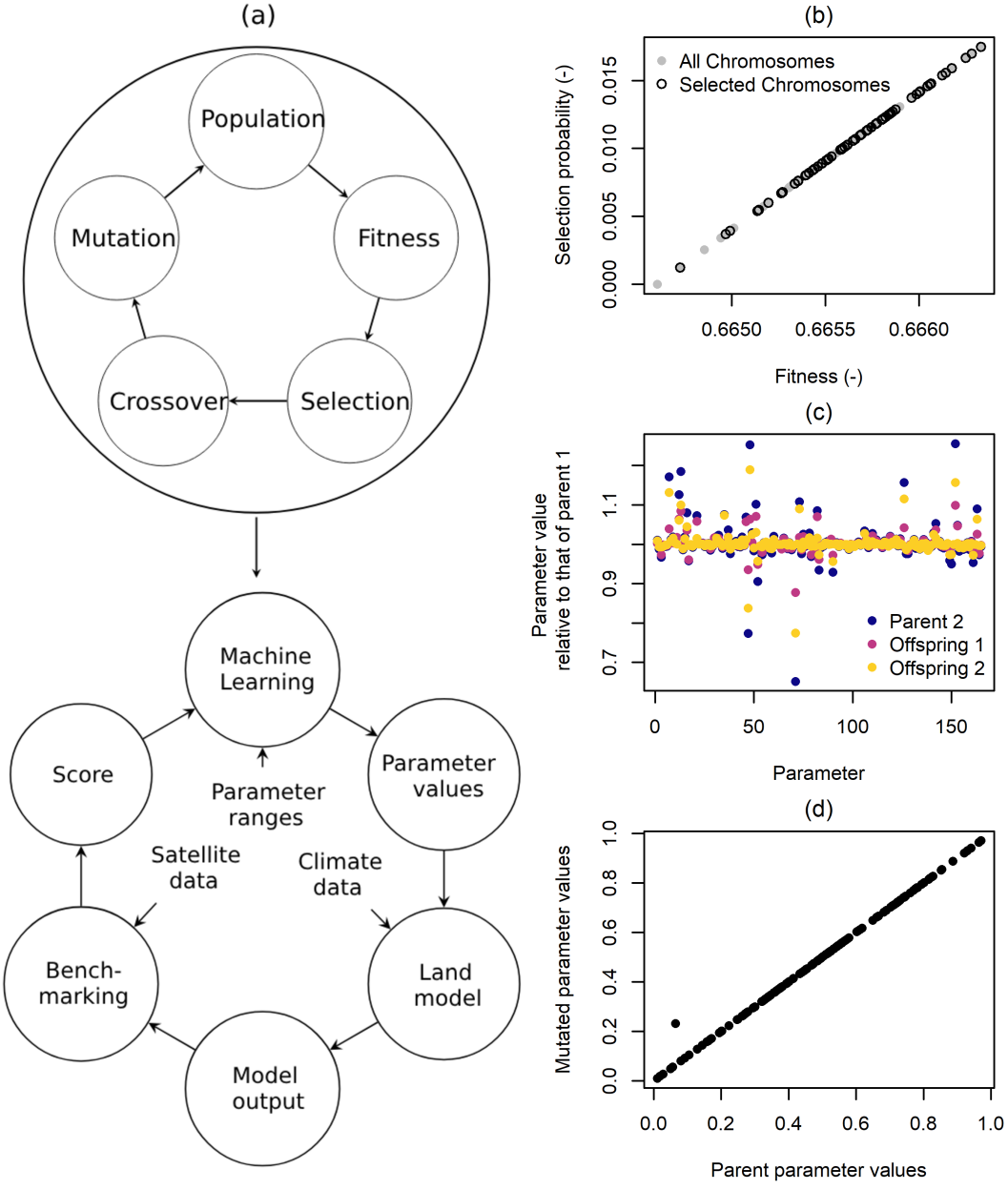

Figure 1. (a) Parameter optimization flowchart, (b) selection, (c) crossover, and (d) mutation.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

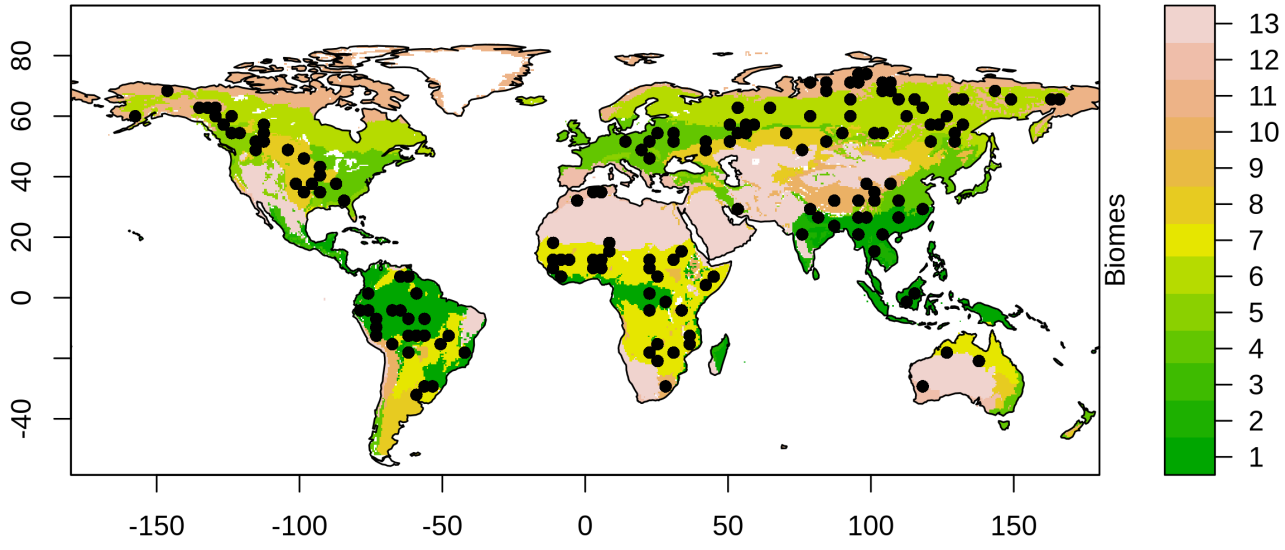

Figure 2. Locations of 160 randomly selected model grid cells used for parameter optimization. The number of selected grid cells per biome is weighted by the biome's area, including (1) Tropical and Subtropical Moist Broadleaf Forests, (2) Tropical and Subtropical Dry Broadleaf Forests, (3) Tropical and Subtropical Coniferous Forests, (4) Temperate Broadleaf and Mixed Forests, (5) Temperate Conifer Forests, (6) Boreal Forests/Taiga, (7) Tropical and Subtropical Grasslands, Savannas and Shrublands, (8) Temperate Grasslands, Savannas and Shrublands, (9), Flooded Grasslands and Savannas, (10) Montane Grasslands and Shrublands, (11) Tundra, (12) Mediterranean Forests, Woodlands and Scrub, and (13) Deserts and Xeric Shrublands.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

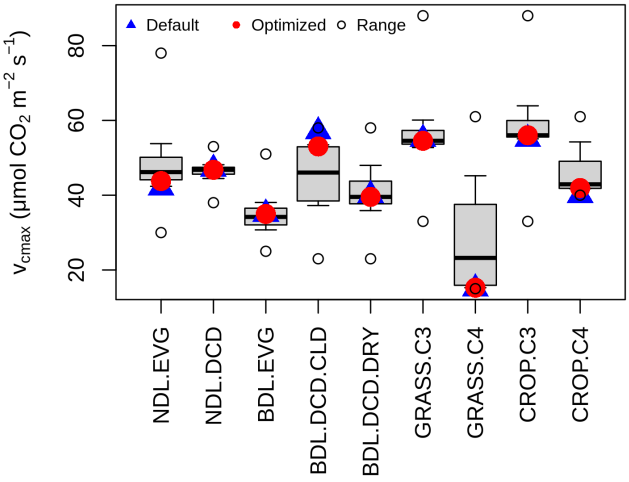

Figure 3. Synthetic data test when optimizing maximum carboxylation rate against modelled GPP for nine Plant Functional Types, namely needleleaf evergreen tree (NDL.EVG), needleleaf deciduous tree (NDL.DCD), broadleaf evergreen tree (BDL.EVG), broadleaf deciduous cold tree (BDL.DCD.CLD), broadleaf deciduous dry tree (BDL.DCD.DRY), C3-photosynthesis grass (GRASS.C3), C4-photosynthesis grass (GRASS.C4), C3-photosynthesis crop (CROP.C3), and C4-photosynthesis grass (CROP.C4). The boxplot displays the interquartile range, median, and 95th percentiles of the sampled parameter values, while the circles represent the full parameter uncertainty range.

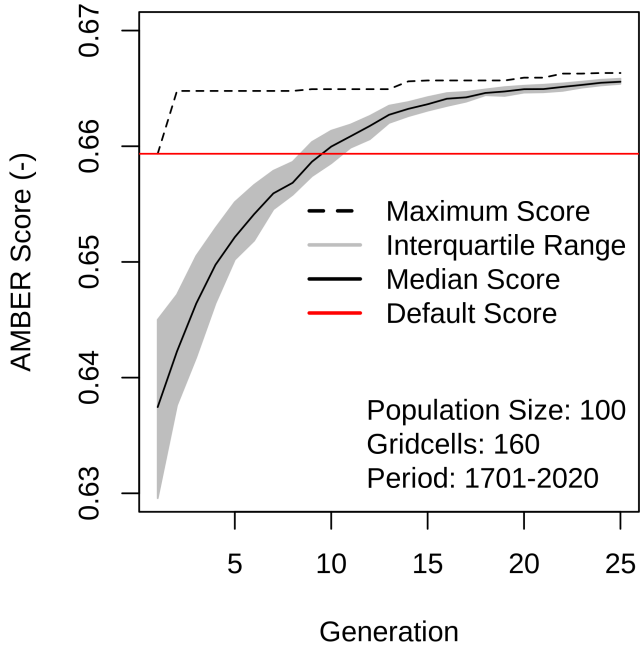

Figure 4. AMBER scores per generation, where each generation consists of 100 chromosomes.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

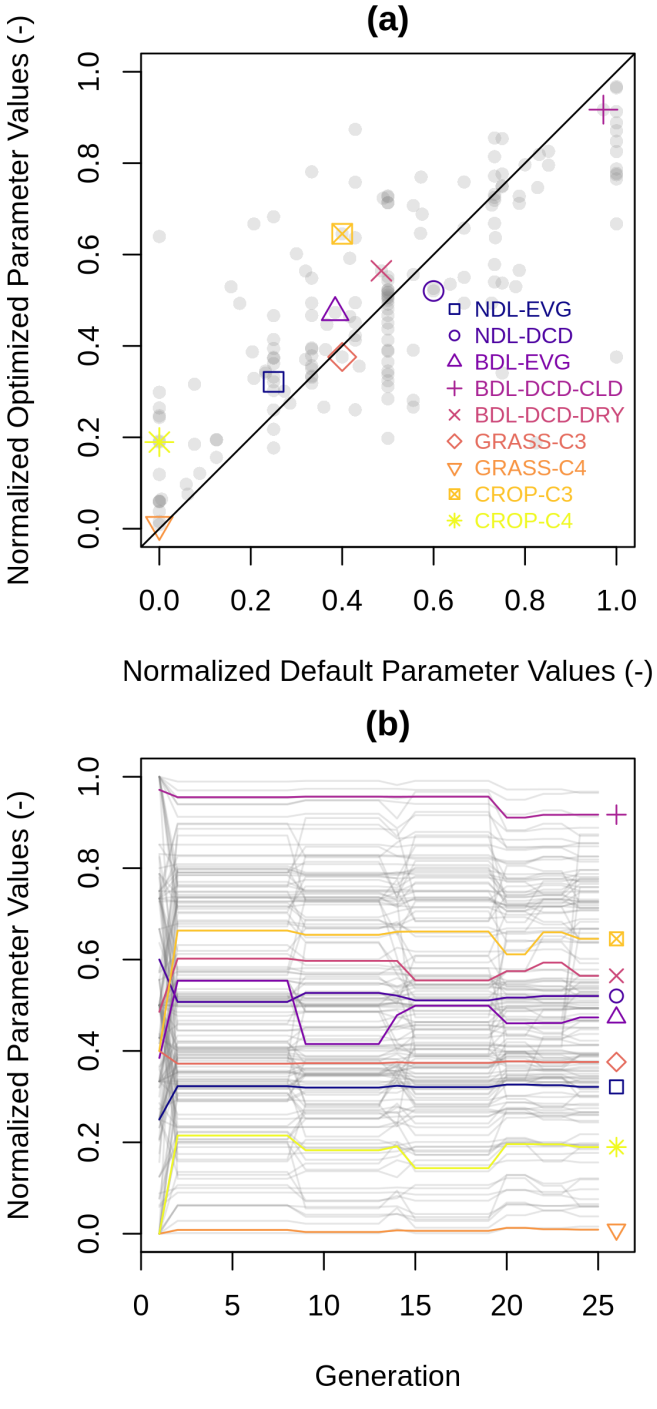

Figure 5. (a) Default versus optimized parameter values and (b) optimized parameter values of the chromosome with the highest fitness for a given generation. The coloured symbols and lines refer to the maximum carboxylation rate of nine PFTs.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

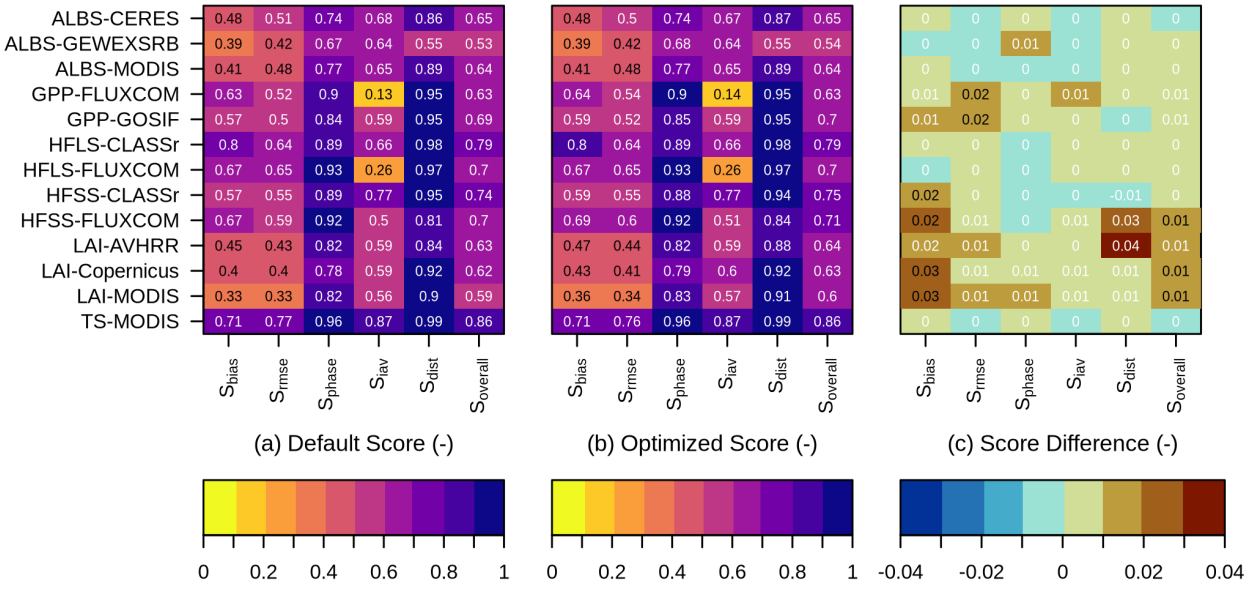

Figure 6. AMBER scores when running the model globally with (a) default and (b) optimized parameter values, with (c) presenting their differences. In (c), black numbers indicate statistically significant differences at the 5% level.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

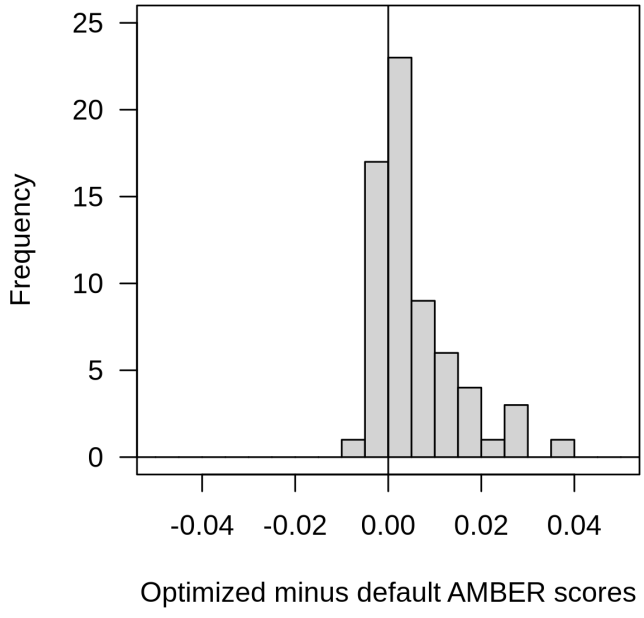

Figure 7. Optimized minus default AMBER score values of S_{bias}, S_{rmse}, S_{phase}, S_{iav}, and S_{dist}, where positive values imply model improvement.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

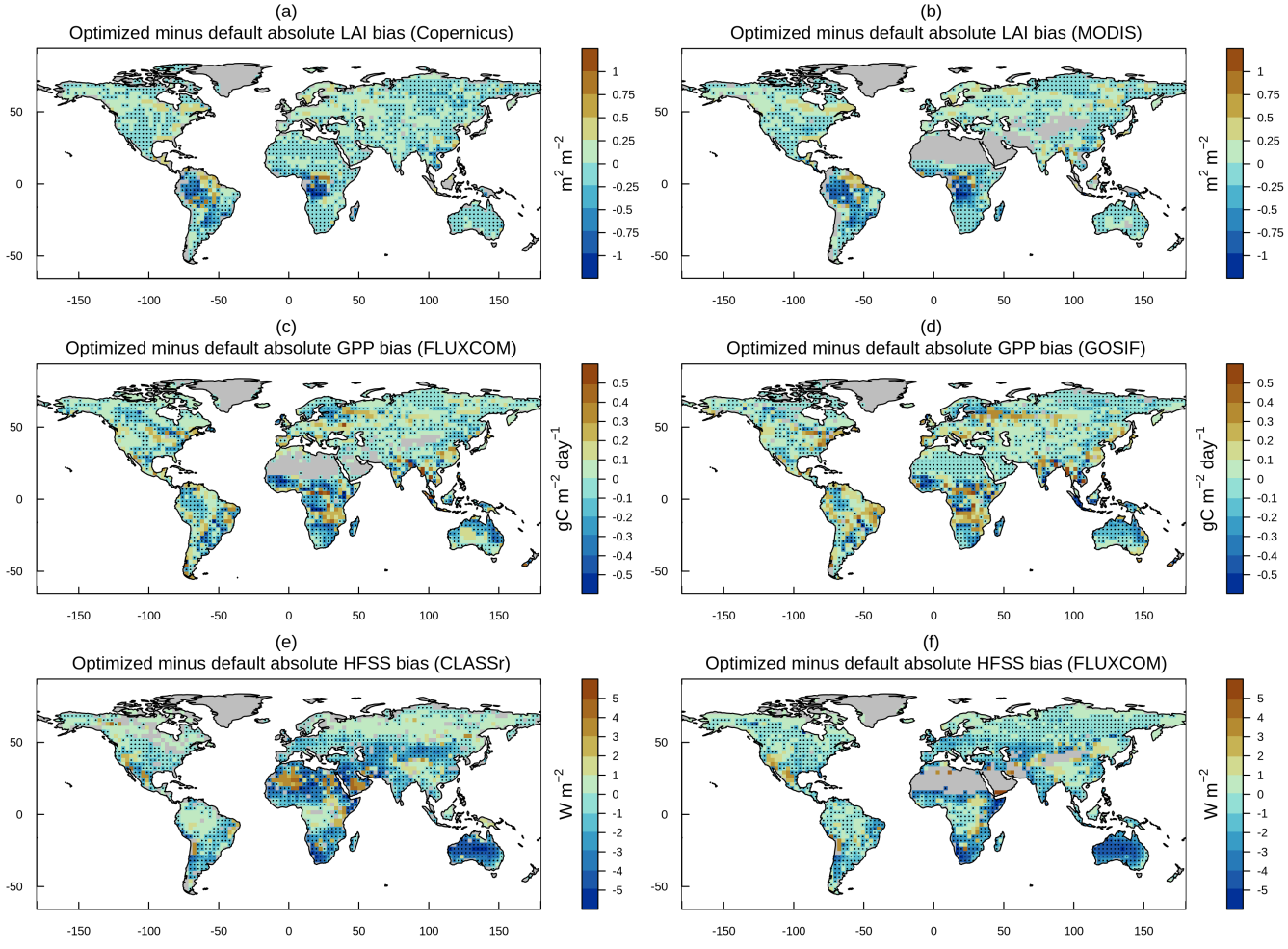

Figure 8. Optimized minus default absolute bias for leaf area index from (a) Copernicus and (b) MODIS, GPP from (c) FLUXCOM and (d) GOSIF, and sensible heat flux from (e) CLASSr and (f) FLUXCOM, where negative values imply bias reduction as denoted with black dots.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

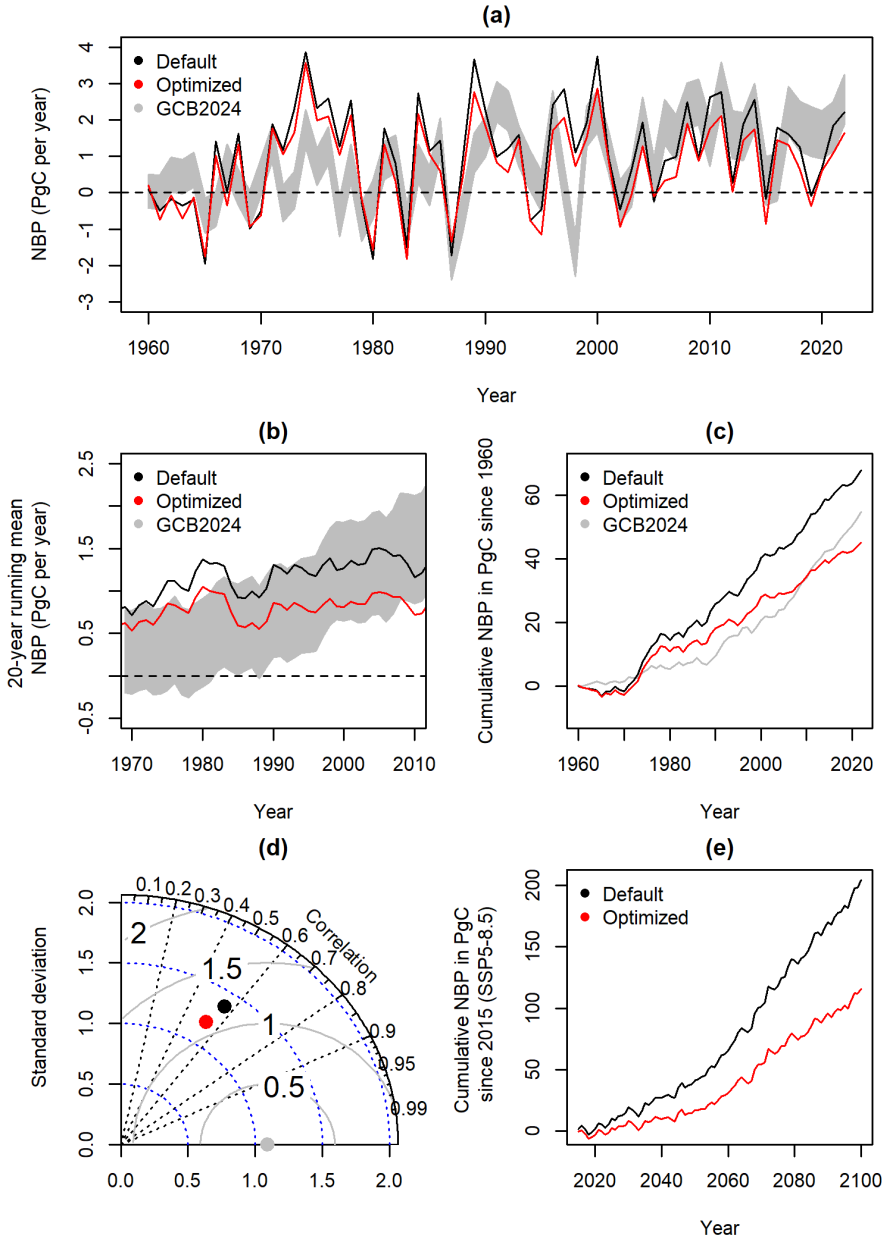

Figure 9. Global net biome productivity (NBP) simulated by CLASSIC with default (black) and optimized (grey) parameter values and reference NBP provided by GCB2024 (grey), shown as: (a) annual means, (b) 20-year running means, (c) annual cumulative values, (d) Taylor diagram of annual means, and (e) annual cumulative values under SSP5-8.5. In panels (a)-(d), default and optimized NBP are based on CRUJRAv2 forcing, while panel (e) uses bias-adjusted CanESM5 forcing.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

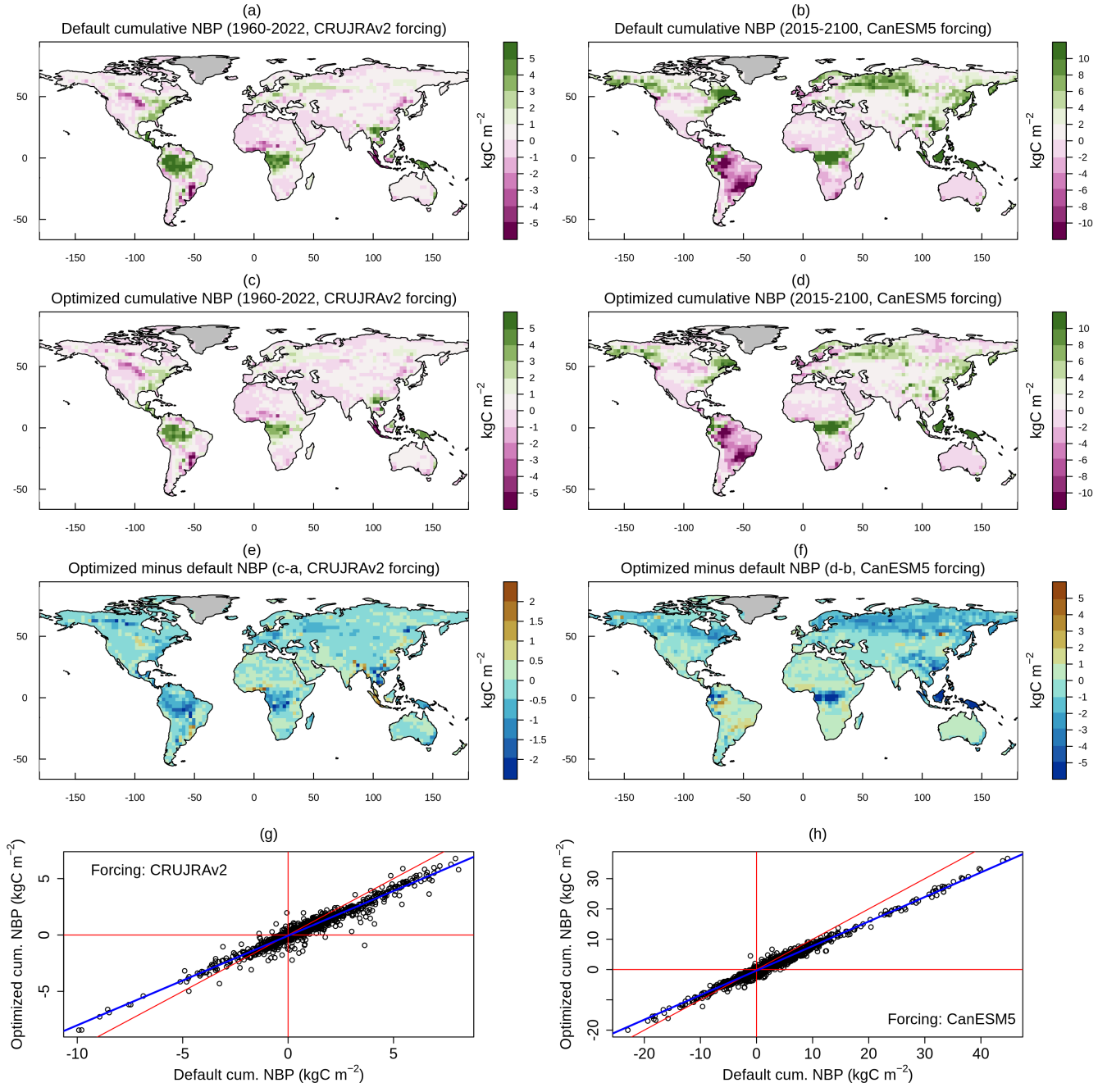

Figure 10. Cumulative NBP under CRUJRAv2 forcing using (a) default parameter values and (c) optimized parameter values, with (e) showing the corresponding differences. Similarly, cumulative NBP under CanESM5 forcing is shown using (b) default parameter values and (d) optimized parameter values, with (f) illustrating the corresponding differences.

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

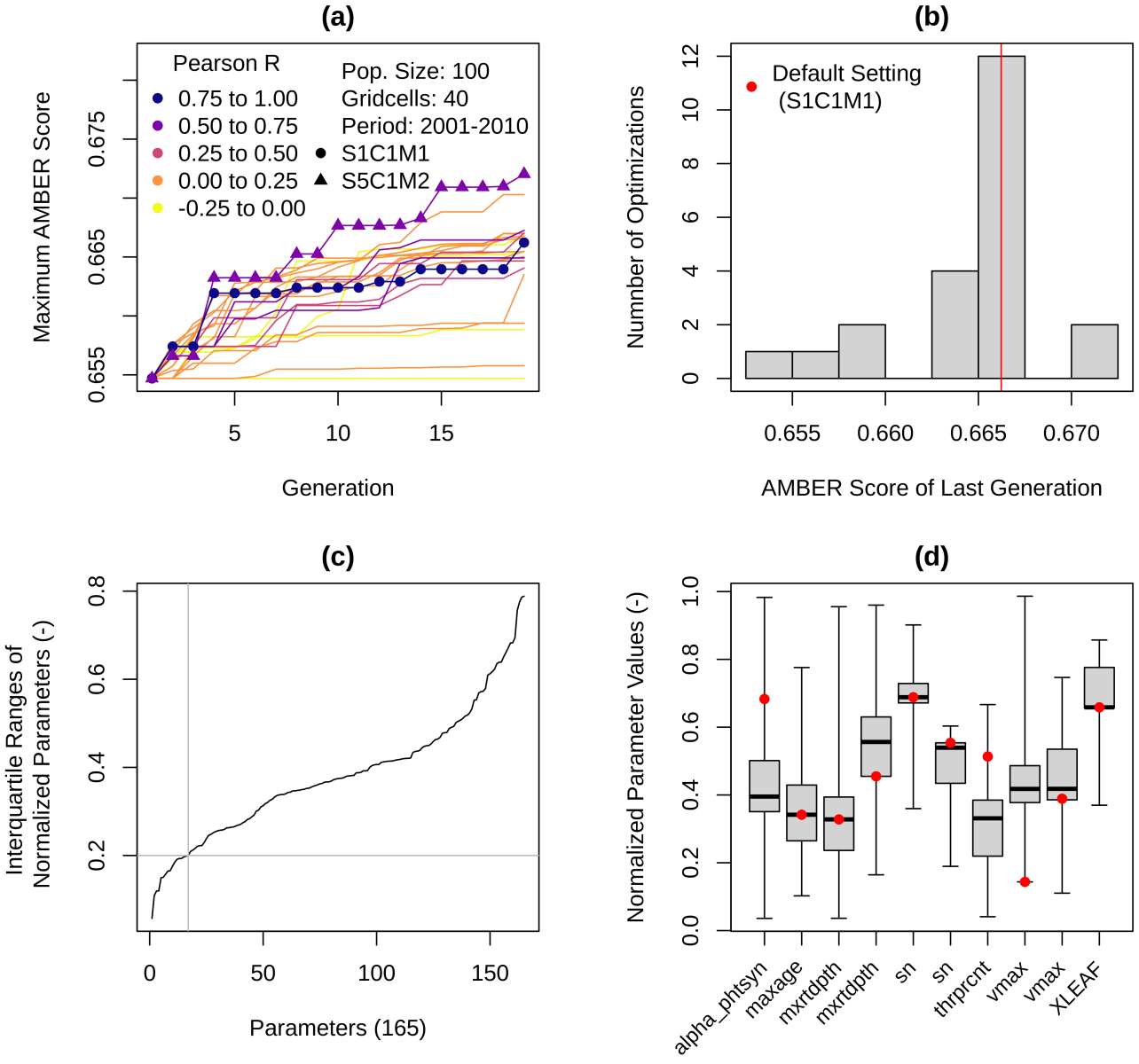

Figure 11. Impact of 22 different Genetic Algorithm (GA) settings on a subset of 40 grid cells over a shorter optimization period (2001–2010), where (a) shows the maximum AMBER score of the population as it evolves across generations under different GA settings, (b) is the frequency distribution of maximum AMBER scores obtained across all optimizations, (c) gives the interquartile ranges of all parameter values aggregated over the 22 GA settings, and (d) shows the interquartile ranges (grey boxes) and total ranges (whiskers) of the ten parameters exhibiting the smallest interquartile ranges across all GA settings. When the same parameters appear in (d), it is because they correspond to different plant functional types (PFTs).

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

Table 1. Parameters that participate in the optimization, where *Process* refers to (1) photosynthesis, (2) autotrophic respiration, (3) allocation (4) phenology, (5) rooting, (6) mortality (7) physical, and *Values* denotes the number of values per parameter.

	Parameter	Description	Process	Values
1	alpha_phtsyn	quantum efficiencies	1	9
2	beta2	photosynthesis coupling or curvature coefficients	1	1
3	gamma_w	photosynthesis down regulation parameter	1	1
4	kn	canopy extinction coefficient	1	9
5	sn	exponent for soil moisture stress	1	9
6	vmax	maximum carboxylation rate	1	9
7	vpd0	stomatal conductance parameter	1	9
8	grescoef	growth respiration coefficient	2	9
9	minlvfr	minimum live wood fraction	2	1
10	omega	allocation parameter	3	9
11	coldlmt	number of days with temperature threshold	4	2
12	coldthrs	initiating harvest mode threshold	4	2
13	kappa	minimum stem and root biomass parameter	4	9
14	lfespany	leaf life span	4	9
15	minslai	minimum storage LAI	4	9
16	roothrsh	root temperature threshold for initiating leaf onset	4	1
17	thrprcnt	percentage of max. LAI that can be supported	4	9
18	abar	root profile parameter	5	9
19	alpha	root growth parameter	5	9
20	avertmas	average root biomass	5	9
21	mxrtdpth	maximum rooting depth	5	9
22	maxage	maximum plant age	6	9
23	albnir	near-infrared albedo	7	9
24	TCCLAY	thermal conductivity of clay particles	7	1
25	TCSAND	thermal conductivity of sand particles	7	1
26	XLEAF	leaf dimension factor used for calculating the leaf boundary resistance	7	9
27	ZOLNG	natural logarithm of soil roughness length	7	1
28	ZOLNS	natural logarithm of snow roughness length	7	1

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

Table 2. Global Earth observations used for model evaluation, where ALBS is surface shortwave albedo, GPP is gross primary productivity, HFLS is latent heat flux, HFSS is sensible heat flux, LAI is leaf area index, and LST is land surface temperature.

Variables	Source	Approach	Period	Reference
ALBS	CERES	radiative transfer model	2000-2012	Kato et al. (2013)
ALBS	GEWEXSRB	radiative transfer model	1984-2007	Stackhouse et al. (2011)
ALBS	MODIS	bidirectional reflectance distribution function	2000-2014	Strahler et al. (1999)
GPP	FluxCom	machine learning	1980-2013	Jung et al. (2020)
GPP	GOSIF	statistical model	2000-2017	Li and Xiao (2019)
HFLS, HFSS	CLASSr	blended product	2003-2009	Hobeichi et al. (2019)
HFLS, HFSS	FluxCom	machine learning	2001-2013	Jung et al. (2019)
LAI	AVHRR	artificial neural network	1982-2010	Claverie et al. (2016)
LAI	Copernicus	artificial neural network	1999-2019	Verger et al. (2014)
LAI	MODIS	radiative transfer model	2000-2017	Myneni et al. (2002)
LST	MODIS	MODIS LST algorithm	2000-2020	Wan et al. (2021)

Table 3. Mean carbon fluxes (net biome productivity, gross primary productivity, autotrophic respiration, heterotrophic respiration, and emissions from fires) and stocks (vegetation carbon and soil organic carbon) for CLASSIC simulations driven with CRUJRAv2 and CanESM5 data using default (def) and optimized (opt) parameter values. The carbon fluxes and pools correspond to the last 20 years of the simulations, i.e. 2003-2022 and 2081-2100 for the CRUJRAv2 and CanESM5 runs, respectively. Also given are the relative differences between simulations that use the optimized versus the default parameter values.

Simulations and their differences	NBP (PgC yr^{-1})	GPP (PgC yr^{-1})	RA (PgC yr^{-1})	RH (PgC yr^{-1})	fFire (PgC yr^{-1})	cVeg (PgC)	cSoil (PgC)
(a) CLASSIC-CRUJRAv2-def	1.34	133.07	59.05	69.91	2.14	469.40	1163.89
(b) CLASSIC-CRUJRAv2-opt	0.86	129.50	60.93	65.77	1.43	394.56	1026.27
Optimization impact: b - a	-0.48	-3.57	1.88	-4.14	-0.71	-74.84	-137.62
(b - a) / a × 100 (%)	-36	-3	3	-6	-33	-16	-12
(c) CLASSIC-CanESM5-def	4.30	235.44	107.44	120.25	3.34	781.98	1349.98
(d) CLASSIC-CanESM5-opt	2.64	224.17	108.69	110.34	2.42	638.73	1169.43
Optimization impact: d - c	-1.66	-11.27	1.25	-9.91	-0.92	-143.25	-180.55
(d - c) / c × 100 (%)	-39	-5	1	-8	-28	-18	-13

https://doi.org/10.5194/egusphere-2025-2517
Preprint. Discussion started: 16 June 2025

Table 4. Pearson correlation coefficients for differences between optimized and default NBP, GPP, autotrophic respiration (RA), heterotrophic respiration (RH) and emissions from fires (fFire). Coefficients denoted with an asterisk are not statistically significant at the 5%-level.

	CRUJRAv2 (1960-2022)					CanESM5, SSP5-8.5 (2015-2100)				
	ΔNBP	ΔGPP	ΔRA	ΔRH	ΔfFire	ΔNBP	ΔGPP	ΔRA	ΔRH	ΔfFire
ΔNBP	1.00	0.37	0.05	0.33	-0.08	1.00	0.32	-0.20	0.31	-0.24
ΔGPP	0.37	1.00	0.47	0.88	0.47	0.32	1.00	0.51	0.89	0.22
ΔRA	0.05	0.47	1.00	0.03*	0.09	-0.20	0.51	1.00	0.12	0.06
ΔRH	0.33	0.88	0.03*	1.00	0.33	0.31	0.89	0.12	1.00	0.12
ΔfFire	-0.08	0.47	0.09	0.33	1.00	-0.24	0.22	0.06	0.12	1.00