Advancing Evaporation and Runoff Simulation: Incorporating CO2 and Environmental Variables into Stomatal Conductance Utilizing Mixed Generalized Additive Models

By

Nastaran Chitsaz
Doctor of Philosophy (PhD)

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy (PhD)
College of Science and engineering
Nov 2024

Table of contents

Abstract

Catastrophic floods and runoff events are increasingly prevalent due to the influence of anthropogenic climate change and variability. Evapotranspiration (ET), the process that governs the exchange of water and energy between the atmosphere, land surface, and groundwater, plays a crucial role in the simulation of runoff within hydrological models. However, accurately estimating ET remains a significant challenge for these models. Presently, many hydrological models rely heavily on potential evapotranspiration (PET) models, as observed ET data is often limited. PET refers to the maximum possible water loss from the soil and vegetation to the atmosphere when water is not limited. PET estimation is challenging due to the complexity of the processes involved and the various sources of uncertainty.

The main source of uncertainty in PET simulation is neglecting the impact of CO_2 on plant water use, which leads to inaccurate runoff simulation. In response to the rising CO_2 concentration, plants close their stomata and decrease stomatal conductance (g_s), which can reduce the amount of water loss through transpiration. A decrease in plant transpiration and an increase in water use efficiency can result in greater antecedent soil moisture and, therefore, increased runoff. Hence, runoff simulations need to consider the relative role of climate change in ecosystems through the PET equation. However, the response of plants to CO_2 varies significantly between different biomes and plant species around the world. In addition, the effects of CO_2 on plant physiology and morphology have complex interactions with other environmental variables such as air temperature (TA), radiation (R), vapour pressure deficit (VPD), and soil water content (SWC). Therefore, the response of plants to CO_2 is characterised by high uncertainty with significant knowledge gaps.

In the first and second chapters of this thesis, the mixed generalised additive model (MGAM) as a nonlinear machine learning technique investigates the plants' response to CO_2 and environmental variables. MGAM analyses the direct and interactive effects of CO_2 and environmental variables on g_s with appropriate sets of statistical covariates between variables. Using eddy covariance flux tower datasets for different vegetation types including crop, deciduous broad-leaf forest, evergreen needle-leaf forest, and grass, shows that MGAM improved g_s simulation by up to 50% increase in Nash-Sutcliffe Efficiency (NSE) compared with conventional g_s simulation models. The MGAM model highlighted the interactive effects of CO_2, VPD, and SWC for crops and grasses. The interactive effects of CO_2, VPD, and TA were identified as important for trees and grasses. In the third chapter of this thesis, the simulated g_s by MGAM was added to the Penman-Monteith PET equation to incorporate vegetation response to environmental variables as a part of the PET equation. The modified PET improved runoff simulation up to a 13% increase in NSE, especially in wet conditions when the role of PET is more significant in runoff fluctuation. The results of this study show that conventional PET models need modification by considering the vegetation response to interactive effects of environmental variables through g_s simulation. This modification leads to a more accurate estimation of water balance elements especially under wet climatic conditions.

Declaration

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university

2. and the research within will not be submitted for any other future degree or diploma without the permission of Flinders University; and

3. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Nastaran Chitsaz

July 2024

Acknowledgement

I would like to express my appreciation to my supervisors, Prof Okke Batelaan, Associate Prof Huade Guan and Dr Margaret Shanafield, for their support throughout my PhD journey. Their guidance and expertise have enriched my academic experience.

I would like to acknowledge the invaluable guidance and insights provided by my advisors, Prof Lu Zhang from CSIRO and Dr Wendy Sharples from BOM. Their expertise and perspective have offered significant contributions to the development and application of my research.

I am grateful to the staff of Flinders University, particularly the Institute of National Centre for Groundwater Research and Training (NCGRT) within the School of Science and Engineering, for their continuous support and invaluable assistance. Their resources, facilities, and expertise have played a pivotal role in the successful completion of my research work.

Acknowledgment is also due to the Australian Government Research Training Program Scholarship for the financial support provided, which enabled me to focus fully on my doctoral studies.

To my family, I owe an immense debt of gratitude for their unconditional love, encouragement, and understanding throughout this challenging academic pursuit. Their steadfast support has been a constant source of motivation and inspiration.

Lastly, I extend my sincere thanks to my friends who have stood by my side during this journey. Your friendship, companionship, and shared experiences have brought joy, laughter, and balance to my life, making this academic journey even more memorable.

Without the support and contributions of all these individuals and institutions, the completion of this PhD thesis would not have been possible.

Chapter 1: Introduction

1.1 Introduction and overview

For millions of years, plants have pulled carbon out of the atmosphere through photosynthesis. Since the start of the Industrial Revolution in 1750, the burning of coal and oil has released a growing percentage of previously buried carbon back into the atmosphere. In the 1960s annual carbon dioxide (CO_2) emission was estimated at 11 billion tons per year, while it increased to 36.6 billion tons in 2022 (Friedlingstein et al., 2022). Based on the published evidence in the Intergovernmental Panel on Climate Change (IPCC), rising atmospheric CO_2 concentrations trap long-wave solar radiation, inducing warming of the earth's surface (Change, 2013). During the late 19th and early 20th centuries (1880-1950 in Fig. 1) both global temperature and atmospheric CO_2 have increased slowly; temperature increased by an average of 0.04° C per decade and atmospheric CO_2 levels rose by around 20 ppm (Lindsey, 2023). There was a rapid change in temperature and atmospheric CO_2 from the late 1950s to 2020; the CO_2 climbed nearly 100 ppm (5 times as fast) and the rate of warming averaged 0.14° C per decade (Lindsey, 2023). The increase in the air and ocean temperature at a global scale resulted in an extensive reduction in ice cover and snow and rising sea levels. The long-lasting changes in temperature and CO_2 are defined as global climate change, which in turn causes extreme weather events such as floods, droughts, and bushfires (Abbass et al., 2022).

The increase in surface temperature caused by global warming increases atmospheric moisture holding capacity and accelerates precipitation (Kim et al., 2023). There has been an increase in the frequency and magnitude of extreme precipitation events in the early twenty-first century, faster than previously anticipated (Kim et al., 2023). Consequently, catastrophic floods and stormwater events have resulted from these increases in precipitation. As a result of anthropogenic climate change and variability, biodiversity, ecosystem functioning, and human well-being are threatened (Abbass et al., 2022;

Abrahms et al., 2023). There is an estimated global socio-economic cost of US$143 billion per year as a result of extreme events associated with climate change in the last twenty years. Human life accounts for 63% of this cost (Newman & Noy, 2023).

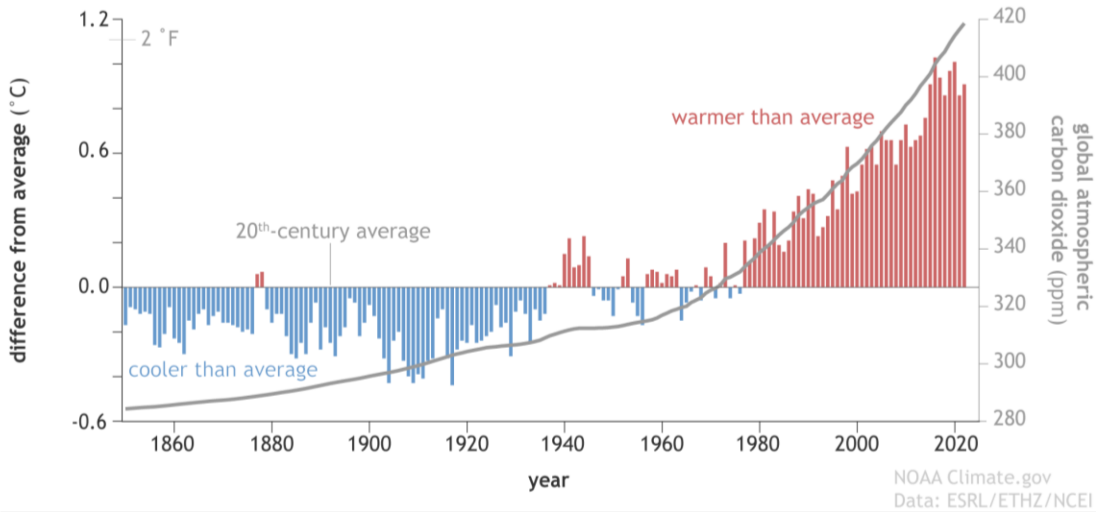


Figure 1 Global average temperature and atmospheric carbon dioxide. Yearly temperature compared to the twentieth-century average (red bars mean warmer than average, blue bars mean colder than average) from 1850–2022 and atmospheric carbon dioxide amounts (grey line). Adapted from reference (Lindsey, 2023).

The increase in annual precipitation caused by rising atmospheric CO_2 leads to higher runoff and frequent flooding (Cui et al., 2020). However, in some regions, runoff changes are not solely influenced by atmospheric processes. In response to the rising atmospheric CO_2 concentration, plants close their stomata, which can reduce the amount of water loss through transpiration. A decrease in plant transpiration and an increase in water use efficiency can result in greater antecedent soil moisture and therefore increased runoff even in the absence of precipitation changes (Fowler et al., 2019). Hence, the runoff simulations for flood management need to consider the relative role of climate change in ecosystems through both soil evaporation and plant physiological responses through transpiration, otherwise known as evapotranspiration (ET). The process of water loss from soil and plants encapsulates the complexity of hydrological cycles under changing climatic conditions. The uncertainty associated with ET estimation is one of the main limitations of accurate runoff simulation (Zhao et al., 2019).

Most hydrological models use potential ET (PET) as a basis for running runoff simulations (Pimentel et al., 2023). PET refers to the amount of water that would be evaporated and transpired by vegetation if sufficient water were available (Peiris & Döll, 2023). A literature review identifies approximately 50 different PET estimation methods, which are divided into three categories as follows: 1) ET as a function of air temperature (TA) only (such as Hargreaves-Samani), 2) ET as a function of TA and radiation (R) (such as Priestly & Taylor), and 3) combinations that are affected by TA, R, wind speed (U), and humidity (hs) (such as Penman-Monteith) (H. Hargreaves & A. Samani, 1985; Monteith, 1965; Oudin et al., 2005; Priestley & Taylor, 1972). Among many PET methods, the Penman-Monteith PET method offers accurate yet simple approximations to the more complex climate model systems (McMahon et al., 2013; Milly & Dunne, 2016). Despite this, plant responses to CO_2 and climate variables are neglected in all PET simulation models, resulting in inaccurate estimation of runoff in hydrological models (Peiris & Döll, 2023; Zhou et al., 2023). Therefore, understanding vegetation response to CO_2 environmental variables in the PET equation is crucial for improving runoff simulation. This has been discussed in more detail in Chapter 4 of the thesis.

1.2 Plant response to CO_2

The stomata, the small pores on leaf surfaces, are responsible for the exchange of gases, mainly water vapour and CO_2, between the leaf and the atmosphere (Fig. 2) (Hetherington & Woodward, 2003). Despite occupying only 5% of the leaf surface, stomata exert major influences on the water and carbon cycle of the planet (Hetherington & Woodward, 2003). At the global scale, there is approximately 110,000 km^3 yr^{-1} of precipitation, while evaporation and transpiration (ET) are approximately 70,000 km^3. The highest rates of transpiration occur in tropical areas with uniform and warm forests with 32,000 $km^3\ yr^{-1}$ of water vapour passing through stomata, which is double the amount of water vapour in the atmosphere (15,000 $km^3\ yr^{-1}$) (Hetherington & Woodward, 2003).

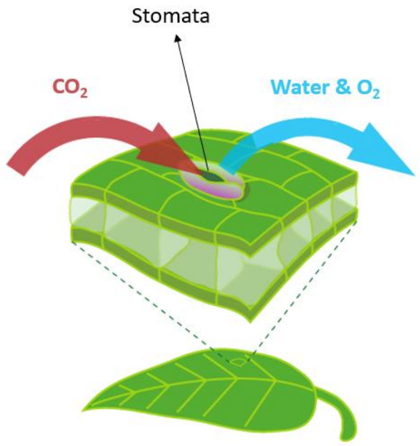


Figure 2 Stomata on the surfaces of leaves control the gas exchange water vapour and CO_2 between the interior of the leaf and the atmosphere (Evolution., 2017).

Elevated CO_2 has two distinct physiological effects on plants as demonstrated in Figure 3. Firstly, plants can reduce their stomatal conductance (g_s) by closing their stomata in response to rising CO_2. Since atmospheric CO_2 levels are increasing, a lower conductance is required to maintain the carbon flux necessary for sustaining photosynthesis. Therefore, reduced conductance results in less water loss through transpiration, which in turn increases soil water content (SWC) and global runoff (Fig. 3a). Secondly, rising CO_2 increases vegetation biomass and vegetation cover or canopy leaf area (LAI), which reduces the impact of stomatal closure on transpiration while increasing transpiration. Therefore, increased transpiration reduces the SWC and runoff (Fig 3b). In regions with high levels of vegetation cover, the first effect of CO_2 on plants is dominant. Therefore, increased CO_2 primarily causes stomatal closure and decreased transpiration rather than increased LAI and vegetation biomass (Zhou et al., 2023). In addition, the changes in LAI due to elevated CO_2 is limited by many other factors such as water and nitrogen deficiency or heat damage (Ågren, 1983; Liu et al., 2023; Warren et al., 2011; Zhou et al., 2023). Therefore, it is generally agreed that the decreasing effects of CO_2 on transpiration are not offset by an increase in LAI (Tor-ngern et al., 2015).

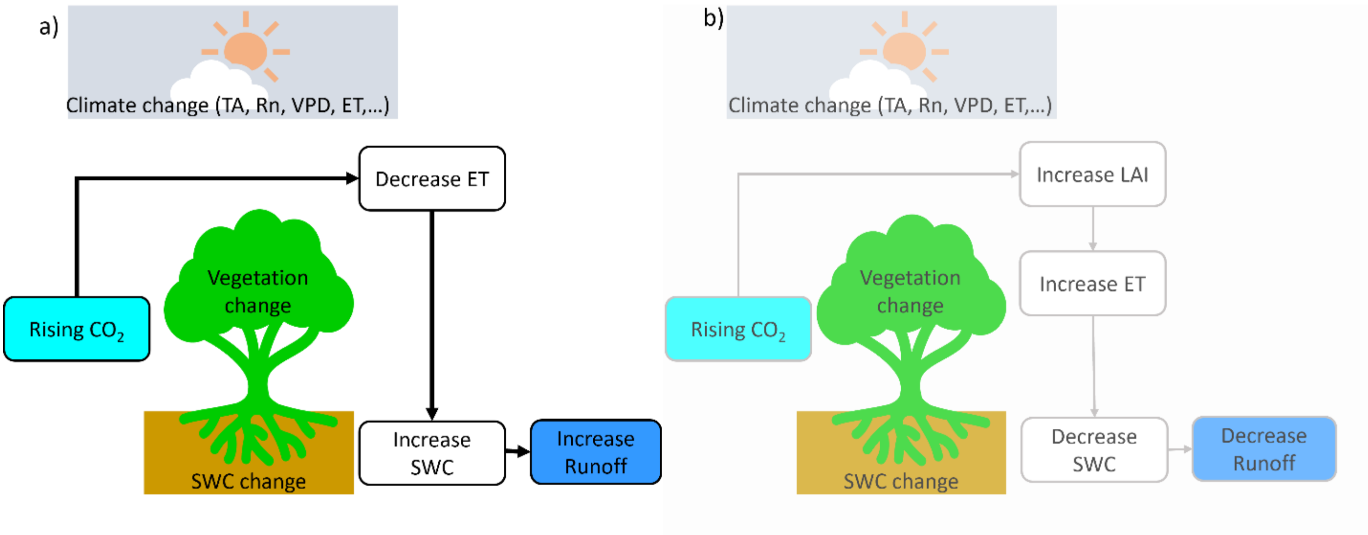


Figure 3 Elevated CO_2 can cause two different physiological effects on plants: a) Rising CO_2 reduces the ET, and consequently increases soil water content (SWC) and runoff, b) Rising CO_2 increases leaf area index (LAI) and evapotranspiration (ET), and consequently decreases SWC and runoff. The first effect of CO_2 is more dominant than the second one.

1.3 Uncertainty in plant response to CO_2

Plants adjust their physiological performance in response to any changes in climate and environmental conditions to improve their growth or survival under extreme conditions (Gimeno et al., 2016). Although stomata typically close in elevated CO_2, the magnitude of this response can vary. Stomatal optimisation theory suggests that stomatal opening to allow CO_2 uptake inevitably comes at the expense of H_2O loss (Cowan & Farquhar, 1977). Thus, stomata should maximise photosynthetic uptake minus the carbon cost of water used in transpiration. This optimisation theory has been applied many times to successfully predict stomatal response to environmental conditions (Arneth et al., 2002; Buckley & Schymanski, 2014; Katul et al., 2010; Vico et al., 2013). Nevertheless, optimisation theory has failed to predict the correct response of the plant to CO_2 under all conditions (Lloyd & Farquhar, 1994; Manzoni et al., 2011; Manzoni et al., 2013; Medlyn et al., 2011). The response of plants to CO_2 varies significantly between different biomes and plant species around the world. Therefore, the response of plants to CO_2 is characterised by high uncertainty with significant knowledge gaps. It has been shown that there is a significant difference between g_s response to CO_2 between shrubs, herbs, and trees (Ainsworth & Long, 2005; Ainsworth & Rogers, 2007). Experimental studies have shown no sensitivity of g_s to elevated CO_2 in some tropical biomes (Wesolowski et al., 2020). The average percentage reduction in g_s by elevated CO_2 varies by vegetation type from 50% in dense meadows, to 15% in broadleaved forests, and to less than 10% in coniferous forests (Körner et al., 2007). There is some evidence that mature forests in mid and high latitudes exhibit a much smaller response of g_s to CO_2 than young trees (Gimeno et al., 2018; Körner et al., 2005). In contrast, some studies have claimed that younger trees' assimilation and transpiration rates increase rapidly to a maximum rate and then stay constant or decline as the vegetation matures, and this process is independent of CO_2 concentrations (Donohue et al., 2017). The response of g_s to CO_2 for different vegetation types has been discussed in Chapter 3 of the thesis.

1.4 Uncertainty in plant response to interactive effects between CO_2 and environmental variables

The effects of CO_2 on plant physiology and morphology overlap with those of other environmental variables (Xu et al., 2013). The global TA is expected to increase in the future, while relative humidity (hs) is expected to decrease (Arias et al., 2021). There is a strong relationship between vapour pressure deficit (VPD) and TA and both are expected to increase in the future (Park Williams et al., 2013). Despite large spatial variability, precipitation is expected to increase on average (Arias et al., 2021). The joint influence of changes in CO_2 and other variables on g_s must be considered due to the interactions between environmental variables (Vicente-Serrano et al., 2022). Precipitation impacts on g_s are reflected by SWC and VPD (Kimm et al., 2020). In response to VPD, plants close their stomata to prevent excessive water loss, which reduces g_s, but this reduction is alleviated by high levels of CO_2 (De Kauwe et al., 2021; Yuan et al., 2019). Moreover, plant response to VPD is highly dependent on plant species, leaf characteristics, and plant height (Lansu et al., 2020). The response of plants g_s to TA is also complex. Through the control of transpiration and cooling effects, stomata play a crucial role in preventing leaf surfaces from reaching excessive TA (Damour et al., 2010). There is evidence that high TA causes an increase in g_s to provide evaporative cooling to the leaf when there is enough available water (Urban et al., 2017). Therefore, plants can endure very high TA by dissipating heat through conduction, convection, and evaporative cooling (Marchin et al., 2022). When TA is high and VPD is increased, there is a severe drought, resulting in the hydraulic failure of the plant water transport system (Adams et al., 2017). As a result, plants come closer to the critical temperature threshold, causing the tree's crown to become thinner, which leads to tree death (Marchin et al., 2022). Thus, an increase in TA causes an increase in g_s up to the threshold TA; TA in exceedance of the threshold value may degrade reserved soil water during long heat episodes and cause leaf cell mortality and a decrease in g_s (Urban et al., 2017). This response of plant g_s to TA is more severe at higher VPD (Purcell et al., 2018; Urban et al., 2017). Drought integrates atmospheric and soil drying but is often identified in terms of soil water availability (Schwalm et al., 2012). VPD and SWC are often correlated and both affect the fluctuation in g_s of plants (Sulman et al., 2016). A reduction in SWC intensifies the effects of VPD on decreasing g_s (Kimm et al., 2020; Sulman et al., 2016).

There are several empirical models to simulate g_s as a function of SWC, VPD and other environmental variables. In these models, multiple linear regression is used to quantify the trends in the attribution of g_s variance to key environmental variables (Kimm et al., 2020; Sulman et al., 2016). Various other empirical and semi-empirical models for g_s simulation are widely used in land surface models (LSMs), described in section 1.5. The interactive effects of environmental variables on g_s have been discussed in Chapter 3 of the thesis.

1.5 Land surface models (LSMs) in g_s simulation

In the face of a world that is rapidly changing both physically and economically, LSMs serve as a useful tool for informing policy about land use and water use management. The LSMs links climate, soil, water, and vegetation to describe energy and water exchanges. Modelling vegetation physiology and soil biogeochemistry is a part of the LSMs; the physical structure of vegetation and the process of photosynthesis affect the exchange of momentum, energy, water, and CO_2 at the land-atmosphere boundary. For LSMs to simulate vegetation response to climate, understanding how photosynthesis, transpiration, and g_s interact through stomata is crucial (Blyth et al., 2021). The g_s simulation in most of the LSMs can be categorized into semi-empirical and empirical approaches. Most g_s models are semi-empirical approaches, combining physiological hypotheses and empirical functions (Damour et al., 2010). All semi-empirical LSMs incorporate net photosynthesis rate (A_n) and g_s relationships as a constraint to couple carbon and water processes. Consequently, due to the coupling between A_n and g_s, these models are expected to have high g_s simulation accuracy. Nevertheless, measurements of A_n are complicated and require extensive experiments or instruments that measure photosynthetic gas exchange and chlorophyll *a* fluorescence over the same area in plants sample (Luo et al., 2016).

The empirical g_s simulation models have different structures that are independent of A_n as a variable, simplifying the input data processing. However, empirical models are dependent on the stress functions of each environmental factor such as R, TA, VPD, SWC, and CO_2. These models assume that environmental factors, are independent without any synergistic interactions (Jarvis et al., 1976). The stress functions of environmental variables have different structures with heavy parameterization (Qi et al., 2023). Initially, empirical models hypothesized that the change in CO_2 concentration was very small and could be ignored (Li et al., 2014). However, CO_2 levels have increased as a result of global climate change, which both directly and indirectly affects agriculture and hydrology by lowering g_s (Stocker, 2014). Consequently, more researchers are engaging in the analysis of how g_s responds to CO_2 by utilising piecewise linear functions (Morison, 1998). Moreover, studies have demonstrated that as CO_2 levels increase, the rate of decreasing g_s gradually lessens. Based on the analysis of physiological and biochemical mechanisms of stomatal activity associated with changing CO_2, the hyperbolic model for the response of g_s to CO_2 was developed, which reflects a more realistic simulation of the g_s-CO_2 interrelationship (Li et al., 2019).

There are several limitations to conventional (semi-empirical and empirical) g_s simulation models, including the need for previous experimental work focussing on the effect of different environmental variables on plants (Table 1). This process involves several calibrated parameters, which are difficult to quantify as they are determined through regression analysis and a complex calibration process with existing datasets. These models, therefore, may not fully capture stomatal response under notable changes in climate conditions compared to the datasets on which these parameters are calibrated (Powell et al., 2013; Saunders et al., 2021). As a result, conventional models require re-parameterisation to be suitable for any changes in vegetation phenology or physiology caused by the variation in climate and growing season (Oliver et al., 2022; Trugman et al., 2018). The structure of empirical and semi-empirical LSMs in g_s simulation has been discussed in Chapter 2 of the thesis.

Table 1 The empirical and semi-empirical models in g_s simulation in LSMs.

	LSMs	**Limitations of the equation**
Empirical	Community Land Model (CLM) Joint UK Land Environment Simulator (JULES) Lund-Potsdam-Jena managed Land (LPJmL)	• Dependency to A_n variable that requires extensive experiments and measurement. • Ignoring variations in vegetation responses to environmental variables.
Semi-empirical	JSBACH Noah	• Heavy parameterization and calibration. • Assumption that environmental variables are independent without any synergistic interactions.

1.6 Machine learning (ML) in hydrology and evapotranspiration

Machine learning (ML) models have become popular in earth sciences in recent years, enabling the discrete classification and estimation of important dynamic variables such as carbon fluxes, precipitation, and river discharge; but also, geospatial variables that are hard to map, such as forest cover, and soils (Koppa et al., 2022). ML models can make full use of the available data, learn complex patterns and relationships between variables and maintain greater consistency with the input data (Zhao et al., 2019). The backbone of all ML models is statistical models which, provide the methodologies and principles of ML models and allow us to interpret the results of these models (Bzdok et al., 2018; Reichstein et al., 2019).

One of the initial branches of ML is known as artificial neural networks (ANNs) inspired by the human brain's neural structure (Sakunthala et al., 2017). The mathematical operations are performed through series of neurons (nodes) that are organized into different layers, such as input and output layers, and hidden layers, which connect the input and output layers (Kalu et al., 2022). Despite numerous advantages, ANNs have several limitations such as a slow learning process which, leads to time-consuming training, and complex structure, which makes it difficult to define the necessary number of neurons and layers (Sakunthala et al., 2017). One of the newest ML approaches is known as random forests (RF) or random decision trees. RF is useful for classification, regression tasks, and prediction with various variables. Although RF accuracy and robust results are the advantages of this model, a large number of variables might make RF useless and sluggish (Nguyen, 2015; Ziegler & König, 2014). In addition, in the presence of outliers and skewed distributions, the RF performance may not be accurate (Latif & Ahmed, 2023; Nguyen, 2015). Another ML model that has been used widely in hydrological studies is the support vector machine (SVM). SVM is relatively memory efficient for classification and regression tasks for a high number of variables (Latif & Ahmed, 2023). SVM is not suitable for datasets with missing values, and the complex training process makes it an inefficient model for large size of datasets (Patle & Chouhan, 2013). Addressing the limitation of each ML model requires the improvement of the underlying statistics. An important statistical development in the last decade is the introduction of generalized linear models (GLM) and multiple linear regression (MLR), which provides progression in the application in hydrology and environmental research (Ravindra et al., 2019). These models are applied to predict an outcome (dependent variable) as a function of one or more predictors (independent variables), which are correlated with the outcome (Ravindra et al., 2019). The main restrictions of GLMs and

MLRs are that these models cannot handle multiple outcome variables simultaneously; and if the observations are not independent of each other, they will give biased standard error in their estimations and consequently misleading statistical inference (Wu & Little, 2011). A generalized additive model (GAM) is a statistical modelling technique that extends the GLM concept and addresses the limitations of GLMs (Ravindra et al., 2019). GAMs can capture nonlinear and complex relationships that GLMs cannot. GAM provides a structure for generalizing GLMs by allowing the additivity of nonlinear functions of the variables (Wood et al., 2016). GAM offers an open-ended solution in case of considerable noise in the predictor variables. It also exhibits the best fit in the case of nonlinear relationships between the predictor and the independent variable (Wood, 2016; Wood, 2017).

In recent years, there has been an increase in the application of ML theories in hydrological and climatic analysis (Kalu et al., 2022). The data-driven stochastic techniques that integrate ML present advantages over physically-based techniques in hydrological interactions. The advantages of ML are easily determined in the operation of data-driven techniques towards parameter estimations, calibration procedures, and its efficiency in handling different sources of uncertainties better than their physically-based counterparts (Kalu et al., 2022). ML models promise to show notable progress in monitoring the multi-scale climatic influences on sub-regional and continental hydrology, simulation of rainfall-runoff, agriculture, groundwater, and other multi-physical trends (Tikhamarine et al., 2020).

In order to evaluate the literature and identify the application of ML and statistical analysis in hydrology and evapotranspiration, keyword co-occurrence analysis was employed. The keyword analysis was performed using VOSViewer™ software, version 1.6.20 (Leiden University, The Netherlands). The following keywords were used in the analysis for searching journal articles: hydrology, machine learning, statistical analysis, linear model, nonlinear model, artificial neural networks, support vector machine, and random forests. Based on the Scopus database, 1597 journal articles published between 2014 to 2024 (last ten years) were selected. The results of the co-occurrence and total links are presented in Figure 4. Different colours represent different clusters that have more links together, while the size of each circle is proportional to the occurrence of the keyword. The analysis was conducted first on the application of ML in hydrology (Fig. 4a), and then the result was narrowed down to the application of ML in evapotranspiration (Fig. 4b). The analysis result showed SVM and RF as the most common ML models in hydrology, ANNs along with linear and multiple linear have been used in many studies while the nonlinear regression analysis and GAM were the least evaluated models for the hydrology studies. The analysis also revealed that most of the ML models have already been used to estimate evapotranspiration (ET), but there is a gap in applying nonlinear regression analysis and GAM in ET studies. ET estimation requires multiple interacting hydroclimatic variables that affect different aspects of plant physiology in a highly nonlinear manner at multiple timescales (Koppa et al., 2022). Therefore, the capability of GAM in addressing the nonlinear functions between various variables raises the assumption that this model could be the best fit for the nonlinear nature of the ET estimation.

a)

b)

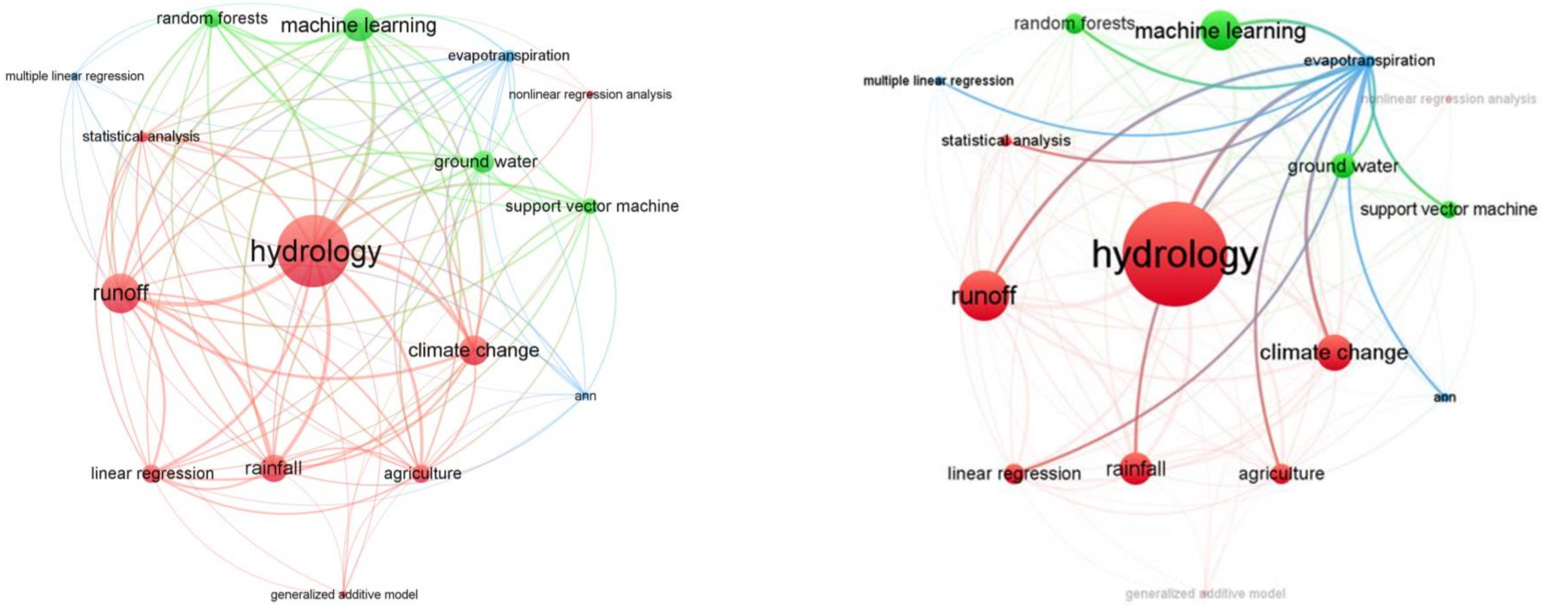


Figure 4 The results of keywords co-occurrence analysis in hydrology fields, among 1597 journal articles published between 2014-2024 in the Scopus database, using VOSViewer™ software. Different colours show different clusters that have more links together, and the size of each circle is proportional to the occurrence of the keyword, a) links to show the applied ML in hydrology studies, b) links to show the applied ML in evapotranspiration (ET) studies.

1.7 Knowledge gap and objectives

There are several approaches for simulating and predicting ET and g_s since they are key elements of the global water cycle. The physically-based approaches derive ET and g_s empirically using climate variables; these models are easy to interpret, but they use fixed environmental variables for simulation for all vegetation types and disregard variations in vegetation responses to environmental variables (Dombrowski et al., 2022). As a result, conventional models must be re-parameterised to accommodate changes in vegetation phenology or physiology caused by climatic and growing season variations (Oliver et al., 2022). Another limitation of conventional models is that they do not optimally extract information from data, and their heavy characterisation and parameterisation are limited by fitting to the existing dataset, which inhibits their generalisation (Liu & Mishra, 2017; Zhao et al., 2019). Although ML algorithms cover the limitations in conventional models, their inability to be interpreted hinders their understanding (Koppa et al., 2022). Therefore, we need to combine conventional models with ML algorithms as 'Hybrid' models to preserve the advantages of conventional models, such as physical consistency and interpretability, as well as those of ML algorithms, such as more realistic data-driven formulations of processes that are not fully understood. Furthermore, considering differences in vegetation response to environmental variables, the ML model should include appropriate sets of statistical covariates with high nonlinear interaction between environmental variables in ET and g_s simulation.

Another knowledge gap discussed in this thesis is the neglect of the vegetation's response to environmental variables in runoff simulations by hydrological models. The moisture transport away from the evaporating surface is presented by PET equations in hydrological models. The PET equations link moisture transport to meteorological data, including TA, R, and VPD (Pimentel et al., 2023). However, the vegetation response to CO_2 and climate variables, which are presented as g_s, is overlooked in PET equations (Ballarin et al., 2023; Yang & Roderick, 2019). The modification of PET by adding g_s as a function of CO_2 has been investigated by several studies, which have shown improvements in runoff simulations at a global scale (Ballarin et al., 2023; Yang et al., 2019; Zhang et al., 2023). However, the g_s variable in modified PET is assumed to be a linear function of CO_2. This does not align with the nonlinear CO_2-g_s function in the real environment (Li et al., 2019). Additionally, the interactive effects of environmental variables such TA, R, VPD, and SWC on g_s are not included in the PET equation, despite their significant influence on plants g_s (Greve et al., 2017; Milly & Dunne, 2016; Zhou et al., 2023). Thus, more work is required to firstly understand whether vegetation response could be more accurately simulated in a PET equation that includes these variables; and secondly if this modification in PET equation can improve runoff simulation accuracy.

This research work had two aims. First, it aimed at simulating g_s using a nonlinear ML model along with physical constraints in order to achieve realistic results. The combined model preserves the advantages of both physical models (physical consistency and interpretability) and ML models (data adaptability and more realistic data-driven formulation). The MGAM as a nonlinear ML model is used for this aim as it may be capable of g_s simulation through optimal extraction of information from observed data. In addition, it can analyse the realistic, nonlinear interaction between environmental variables with appropriate sets of statistical covariates in g_s simulation. The second overall aim of this study is to investigate the role of g_s in a modified PET equation, which incorporates the vegetation response to environmental variables. As PET is an important term in hydrological models for runoff simulation, the modified PET is assumed to improve the runoff simulation accuracy. In this aim, the vegetation response to environmental variables is simulated through g_s and then incorporated into the PET equation for runoff simulation improvement.

The following research objectives were determined to achieve the research aims.

1. Improving the g_s simulation by Mixed Generalised Additive Model (MGAM) as a nonlinear ML.
    a. Objective 1: Comparison between conventional models and MGAM for g_s simulation (addressed in Chapter 2)
    b. Objective 2: Global sensitivity analysis of different g_s simulation models (addressed in Chapter 2)
    c. Objective 3: Generalization of MGAM in g_s simulation for different vegetation types (addressed in Chapter 3)
2. Enhancing the runoff simulation by including g_s in the PET equation.
    a. Objective 1: Comparison between conventional PET simulation model and modified PET model by adding simulated g_s by MGAM (addressed in Chapter 4)
    b. Objective 2: Comparison between runoff simulation by conventional PET and modified PET (addressed in Chapter 4)
    c. Objective 3: The role of CO_2 and environmental variables on runoff simulation for different climate conditions (wet and dry conditions) (addressed in Chapter 4)

To address the first aim of this research, the MGAM model was trained to learn the relationship between VPD, CO_2, R, TA, and SWC for g_s simulation (addressed in Chapter 2). The direct and interactive effects of environmental variables on g_s were measured by smooth and tensor functions in MGAM. Then the results of the g_s simulation by the MGAM model were tested against observed g_s data and compared with the results of the conventional models (semi-empirical and empirical models). The global sensitivity analysis presented the sensitivity of g_s to direct and interactive effects of key environmental variables in MGAM and conventional models. Moreover, the difference in vegetation response to environmental variables, which is overlooked in conventional models, was highlighted by the different structures of MGAM model for each vegetation type, including crop (CRO), deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), and grass (GRA) (addressed in Chapter 3). Visualisation methods in ML models were applied to MGAM to show the contribution of each environmental variable in g_s simulation. The interactive effects between key environmental variables in g_s fluctuation, which is neglected in conventional models, were presented in MGAM visualisation.

To address the second aim, the simulated g_s by MGAM has been added to the Penman-Monteith PET equation (PET_{PM}), to incorporate vegetation response to environmental variables as a part of the PET equation (addressed in Chapter 4). The modified PET (PET_{MGAM}) is expected to improve runoff simulation, especially in wet conditions when the role of PET is more significant in runoff fluctuation. Therefore, the GR4J rainfall-runoff model was used to compare runoff simulation by PET_{MGAM} and PET_{PM} with observed runoff values in different climate conditions of dry, wet, and extreme wet. Then the sensitivity analysis of PET determined the contribution of key environmental variables in PET simulation for different climate conditions to interpret the performance of PET models in runoff simulations.

1.8 Thesis outline

Table showing each chapter's title and type of publication.

Chapter title	Chapter	Type
Introduction and literature review	1	research overview /literature review
Paper 1: Evaluating CO_2 effects on semi-empirical and empirical stomatal conductance simulation in land surface models	2	research output (published)
Paper 2: The Impact of Environmental Variables on Canopy Conductance: Advancing Simulation with Nonlinear Machine Learning Model	3	research output (published)
Paper 3: Enhanced runoff simulation: improved evapotranspiration via vegetation response to climate conditions utilizing machine learning	4	research output (to be submitted to a peer-reviewed journal)
Conclusion and outlook	5	conclusion

1.9 References

Abbass, K., Qasim, M. Z., Song, H., Murshed, M., Mahmood, H., & Younis, I. (2022). A review of the global climate change impacts, adaptation, and sustainable mitigation measures. *Environmental Science and Pollution Research*, *29*(28), 42539-42559. https://doi.org/10.1007/s11356-022-19718-6

Abrahms, B., Carter, N. H., Clark-Wolf, T. J., Gaynor, K. M., Johansson, E., McInturff, A., Nisi, A. C., Rafiq, K., & West, L. (2023). Climate change as a global amplifier of human–wildlife conflict. *Nature Climate Change*, *13*(3), 224-234. https://doi.org/10.1038/s41558-023-01608-5

Adams, H. D., Zeppel, M. J. B., Anderegg, W. R. L., Hartmann, H., Landhäusser, S. M., Tissue, D. T., Huxman, T. E., Hudson, P. J., Franz, T. E., Allen, C. D., Anderegg, L. D. L., Barron-Gafford, G. A., Beerling, D. J., Breshears, D. D., Brodribb, T. J., Bugmann, H., Cobb, R. C., Collins, A. D., Dickman, L. T., . . . McDowell, N. G. (2017). A multi-species synthesis of physiological mechanisms in drought-induced tree mortality. *Nature Ecology & Evolution*, *1*(9), 1285-1291. https://doi.org/10.1038/s41559-017-0248-x

Ågren, G. I. (1983). Nitrogen productivity of some conifers. *Canadian Journal of Forest Research*, *13*(3), 494-500. https://doi.org/10.1139/x83-073

Ainsworth, E. A., & Long, S. P. (2005). What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO2. *New Phytologist*, *165*(2), 351-372. https://doi.org/https://doi.org/10.1111/j.1469-8137.2004.01224.x

Ainsworth, E. A., & Rogers, A. (2007). The response of photosynthesis and stomatal conductance to rising [CO2]: mechanisms and environmental interactions. *Plant Cell Environ*, *30*(3), 258-270. https://doi.org/10.1111/j.1365-3040.2007.01641.x

Arias, P., Bellouin, N., Coppola, E., Jones, R., Krinner, G., Marotzke, J., Naik, V., Palmer, M., Plattner, G.-K., & Rogelj, J. (2021). Climate Change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; technical summary.

Arneth, A., Lloyd, J., Santruckova, H., Bird, M., Grigoryev, S., Kalaschnikov, Y. N., Gleixner, G., & Schulze, E.-D. (2002). Response of central Siberian Scots pine to soil water deficit and long-term trends in atmospheric CO2 concentration. *Global Biogeochemical Cycles*, *16*(1), 5-1. https://doi.org/https://doi.org/10.1029/2000GB001374

Ballarin, A. S., Sousa Mota Uchôa, J. G., dos Santos, M. S., Almagro, A., Miranda, I. P., da Silva, P. G. C., da Silva, G. J., Gomes Júnior, M. N., Wendland, E., & Oliveira, P. T. S. (2023). Brazilian Water Security Threatened by Climate Change and Human Behavior. *Water Resources Research*, *59*(7), e2023WR034914. https://doi.org/https://doi.org/10.1029/2023WR034914

Blyth, E. M., Arora, V. K., Clark, D. B., Dadson, S. J., De Kauwe, M. G., Lawrence, D. M., Melton, J. R., Pongratz, J., Turton, R. H., Yoshimura, K., & Yuan, H. (2021). Advances in Land Surface Modelling. *Current Climate Change Reports*, *7*(2), 45-71. https://doi.org/https://doi.org/10.1007/s40641-021-00171-5

Buckley, T. N., & Schymanski, S. J. (2014). Stomatal optimisation in relation to atmospheric CO2. *New Phytologist*, *201*(2), 372-377. https://doi.org/https://doi.org/10.1111/nph.12552

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, *15*(4), 233-234. https://doi.org/10.1038/nmeth.4642

Change, I. C. (2013). The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge, 2013).

Cowan, I. R., & Farquhar, G. D. (1977). Stomatal function in relation to leaf metabolism and environment. *Symp Soc Exp Biol*, *31*, 471-505.

Cui, J., Piao, S., Huntingford, C., Wang, X., Lian, X., Chevuturi, A., Turner, A. G., & Kooperman, G. J. (2020). Vegetation forcing modulates global land monsoon and water resources in a CO2-enriched climate. *Nature Communications*, *11*(1), 5184. https://doi.org/10.1038/s41467-020-18992-7

Damour, G., Simonneau, T., Cochard, H., & Urban, L. (2010). An overview of models of stomatal conductance at the leaf level. *Plant, cell & environment*, *33*(9), 1419-1438. https://doi.org/https://doi.org/10.1111/j.1365-3040.2010.02181.x

De Kauwe, M. G., Medlyn, B. E., & Tissue, D. T. (2021). To what extent can rising [CO2] ameliorate plant drought stress? *New Phytologist.*, *231*(6), 2118-2124. https://doi.org/https://doi.org/10.1111/nph.17540

Dombrowski, O., Brogi, C., Hendricks Franssen, H. J., Zanotelli, D., & Bogena, H. (2022). CLM5-FruitTree: a new sub-model for deciduous fruit trees in the Community Land Model (CLM5). *Geosci. Model Dev.*, *15*(13), 5167-5193. https://doi.org/https://10.5194/gmd-15-5167-2022

Donohue, R. J., Roderick, M. L., McVicar, T. R., & Yang, Y. (2017). A simple hypothesis of how leaf and canopy-level transpiration and assimilation respond to elevated CO2 reveals distinct response patterns between disturbed and undisturbed vegetation. *Journal of Geophysical Research: Biogeosciences*, *122*(1), 168-184. https://doi.org/https://doi.org/10.1002/2016JG003505

Evolution., U. o. C. M. o. P. s. U. (2017). *Stomata – an ancient insight into a modern problem*. https://evolution.berkeley.edu/ancient-fossils-and-modern-climate-change/leaves-with-microscopic-mouths/

Fowler, M. D., Kooperman, G. J., Randerson, J. T., & Pritchard, M. S. (2019). The effect of plant physiological responses to rising CO2 on global streamflow. *Nature Climate Change*, *9*(11), 873-879. https://doi.org/10.1038/s41558-019-0602-x

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., . . . Zheng, B. (2022). Global Carbon Budget 2022. *Earth Syst. Sci. Data*, *14*(11), 4811-4900. https://doi.org/10.5194/essd-14-4811-2022

Gimeno, T. E., Crous, K. Y., Cooke, J., O'Grady, A. P., Ósvaldsson, A., Medlyn, B. E., & Ellsworth, D. S. (2016). Conserved stomatal behaviour under elevated CO2 and varying water availability in a mature woodland. *Functional Ecology*, *30*(5), 700-709. https://doi.org/https://doi.org/10.1111/1365-2435.12532

Gimeno, T. E., McVicar, T. R., O'Grady, A. P., Tissue, D. T., & Ellsworth, D. S. (2018). Elevated CO2 did not affect the hydrological balance of a mature native Eucalyptus woodland. *24*(7), 3010-3024. https://doi.org/https://doi.org/10.1111/gcb.14139

Greve, P., Roderick, M. L., & Seneviratne, S. I. (2017). Simulated changes in aridity from the last glacial maximum to 4xCO2. *Environmental Research Letters*, *12*(11), 114021. https://doi.org/https://doi.org/10.1088/1748-9326/aa89a3

H. Hargreaves, G., & A. Samani, Z. (1985). Reference Crop Evapotranspiration from Temperature. *Applied Engineering in Agriculture*, *1*(2), 96-99. https://doi.org/https://doi.org/10.13031/2013.26773

Hetherington, A. M., & Woodward, F. I. (2003). The role of stomata in sensing and driving environmental change. *Nature*, *424*(6951), 901-908. https://doi.org/10.1038/nature01843

Jarvis, P. G., Monteith, J. L., & Weatherley, P. E. (1976). The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Biological Sciences*, *273*(927), 593-610. https://doi.org/https://doi.org/10.1098/rstb.1976.0035

Kalu, I., Ndehedehe, C. E., Okwuashi, O., Eyoh, A. E., & Ferreira, V. G. (2022). An assimilated deep learning approach to identify the influence of global climate on hydrological fluxes. *Journal of Hydrology*, *614*, 128498. https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.128498

Katul, G., Manzoni, S., Palmroth, S., & Oren, R. (2010). A stomatal optimization theory to describe the effects of atmospheric CO2 on leaf photosynthesis and transpiration. *Annals of Botany*, *105*(3), 431-442. https://doi.org/10.1093/aob/mcp292

Kim, J., Porter, J., & Kearns, E. J. (2023). Exposure of the US population to extreme precipitation risk has increased due to climate change. *Scientific Reports*, *13*(1), 21782. https://doi.org/10.1038/s41598-023-48969-7

Kimm, H., Guan, K., Gentine, P., Wu, J., Bernacchi, C. J., Sulman, B. N., Griffis, T. J., & Lin, C. (2020). Redefining droughts for the U.S. Corn Belt: The dominant role of atmospheric vapor pressure deficit over soil moisture in regulating stomatal behavior of Maize and Soybean.

Agricultural and Forest Meteorology, *287*, 107930. https://doi.org/https://doi.org/10.1016/j.agrformet.2020.107930

Koppa, A., Rains, D., Hulsman, P., Poyatos, R., & Miralles, D. G. (2022). A deep learning-based hybrid model of global terrestrial evaporation. *Nature Communications*, *13*(1), 1912. https://doi.org/https://doi.org/10.1038/s41467-022-29543-7

Körner, C., Asshoff, R., Bignucolo, O., Hättenschwiler, S., Keel, S. G., Peláez-Riedl, S., Pepin, S., Siegwolf, R. T., & Zotz, G. (2005). Carbon flux and growth in mature deciduous forest trees exposed to elevated CO2. *Science*, *309*(5739), 1360-1362. https://doi.org/10.1126/science.1113977

Körner, C., Morgan, J., & Norby, R. (2007). CO2 fertilization: when, where, how much? In *Terrestrial ecosystems in a changing world* (pp. 9-21). Springer. https://link.springer.com/chapter/10.1007/978-3-540-32730-1_2

Lansu, E. M., van Heerwaarden, C. C., Stegehuis, A. I., & Teuling, A. J. (2020). Atmospheric Aridity and Apparent Soil Moisture Drought in European Forest During Heat Waves. *Geophysical Research Letters*, *47*(6), e2020GL087091. https://doi.org/https://doi.org/10.1029/2020GL087091

Latif, S. D., & Ahmed, A. N. (2023). A review of deep learning and machine learning techniques for hydrological inflow forecasting. *Environment, Development and Sustainability*, *25*(11), 12189-12216. https://doi.org/10.1007/s10668-023-03131-1

Li, S., Hao, X., Du, T., Tong, L., Zhang, J., & Kang, S. (2014). A coupled surface resistance model to estimate crop evapotranspiration in arid region of northwest China. *Hydrological Processes*, *28*(4), 2312-2323. https://doi.org/https://doi.org/10.1002/hyp.9768

Li, X., Kang, S., Niu, J., Huo, Z., & Liu, J. (2019). Improving the representation of stomatal responses to CO2 within the Penman–Monteith model to better estimate evapotranspiration responses to climate change. *Journal of Hydrology*, *572*, 692-705. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.03.029

Lindsey, R. (2023). If carbon dioxide hits a new high every year, why isn't every year hotter than the last? *NOAA Climate.gov*. https://www.climate.gov/news-features/climate-qa/if-carbon-dioxide-hits-new-high-every-year-why-isn%E2%80%99t-every-year-hotter-last

Liu, D., & Mishra, A. K. (2017). Performance of AMSR_E soil moisture data assimilation in CLM4.5 model for monitoring hydrologic fluxes at global scale. *Journal of Hydrology*, *547*, 67-79. https://doi.org/https://doi.org/10.1016/j.jhydrol.2017.01.036

Liu, M., Sun, A. Y., Lin, K., Luo, W., Tu, X., & Chen, X. (2023). Estimating Dynamic Non-Water-Limited Canopy Resistance Over the Globe: Changes, Contributors, and Implications. *Water Resources Research*, *59*(9), e2022WR034209. https://doi.org/https://doi.org/10.1029/2022WR034209

Lloyd, J., & Farquhar, G. D. (1994). (13)C discrimination during CO(2) assimilation by the terrestrial biosphere. *Oecologia*, *99*(3-4), 201-215. https://doi.org/10.1007/bf00627732

Luo, H.-h., Merope, T.-m., Zhang, Y.-l., & Zhang, W.-f. (2016). Combining gas exchange and chlorophyll a fluorescence measurements to analyze the photosynthetic activity of drip-irrigated cotton under different soil water deficits. *Journal of Integrative Agriculture*, *15*(6), 1256-1266. https://doi.org/https://doi.org/10.1016/S2095-3119(15)61270-9

Manzoni, S., Vico, G., Katul, G., Fay, P. A., Polley, W., Palmroth, S., & Porporato, A. (2011). Optimizing stomatal conductance for maximum carbon gain under water stress: a meta-analysis across plant functional types and climates. *Functional Ecology*, *25*(3), 456-467. https://doi.org/https://doi.org/10.1111/j.1365-2435.2010.01822.x

Manzoni, S., Vico, G., Palmroth, S., Porporato, A., & Katul, G. (2013). Optimization of stomatal conductance for maximum carbon gain under dynamic soil moisture. *Advances in Water Resources*, *62*, 90-105. https://doi.org/https://doi.org/10.1016/j.advwatres.2013.09.020

Marchin, R. M., Backes, D., Ossola, A., Leishman, M. R., Tjoelker, M. G., & Ellsworth, D. S. (2022). Extreme heat increases stomatal conductance and drought-induced mortality risk in vulnerable plant species. *Global Change Biology*, *28*(3), 1133-1146. https://doi.org/https://doi.org/10.1111/gcb.15976

McMahon, T. A., Peel, M. C., Lowe, L., Srikanthan, R., & McVicar, T. R. (2013). Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic

synthesis. *Hydrol. Earth Syst. Sci.*, *17*(4), 1331-1363. https://doi.org/10.5194/hess-17-1331-2013
Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V. M., Crous, K. Y., De Angelis, P., Freeman, M., & Wingate, L. (2011). Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology*, *17*(6), 2134-2144. https://doi.org/https://doi.org/10.1111/j.1365-2486.2010.02375.x
Milly, P. C. D., & Dunne, K. A. (2016). Potential evapotranspiration and continental drying. *Nature Climate Change*, *6*(10), 946-949. https://doi.org/https://doi.org/10.1038/nclimate3046
Monteith, J. L. (1965). Evaporation and environment. *Symp Soc Exp Biol*, *19*, 205-234.
Morison, J. I. L. (1998). Stomatal response to increased CO2 concentration. *Journal of Experimental Botany*, *49*(Special_Issue), 443-452. https://doi.org/10.1093/jxb/49.Special_Issue.443
Newman, R., & Noy, I. (2023). The global costs of extreme weather that are attributable to climate change. *Nature Communications*, *14*(1), 6103. https://doi.org/10.1038/s41467-023-41888-1
Nguyen, T. T. (2015, 8-10 Oct. 2015). An L1-Regression Random Forests Method for Forecasting of Hoa Binh Reservoir's Incoming Flow. 2015 Seventh International Conference on Knowledge and Systems Engineering (KSE),
Oliver, R. J., Mercado, L. M., Clark, D. B., Huntingford, C., Taylor, C. M., Vidale, P. L., McGuire, P. C., Todt, M., Folwell, S., Shamsudheen Semeena, V., & Medlyn, B. E. (2022). Improved representation of plant physiology in the JULES-vn5.6 land surface model: photosynthesis, stomatal conductance and thermal acclimation. *Geosci. Model Dev.*, *15*(14), 5567-5592. https://doi.org/10.5194/gmd-15-5567-2022
Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, *303*(1), 290-306. https://doi.org/https://doi.org/10.1016/j.jhydrol.2004.08.026
Park Williams, A., Allen, C. D., Macalady, A. K., Griffin, D., Woodhouse, C. A., Meko, D. M., Swetnam, T. W., Rauscher, S. A., Seager, R., Grissino-Mayer, H. D., Dean, J. S., Cook, E. R., Gangodagamage, C., Cai, M., & McDowell, N. G. (2013). Temperature as a potent driver of regional forest drought stress and tree mortality. *Nature Climate Change*, *3*(3), 292-297. https://doi.org/https://doi.org/10.1038/nclimate1693
Patle, A., & Chouhan, D. S. (2013, 23-25 Jan. 2013). SVM kernel functions for classification. 2013 International Conference on Advances in Technology and Engineering (ICATE),
Peiris, T. A., & Döll, P. (2023). Improving the quantification of climate change hazards by hydrological models: a simple ensemble approach for considering the uncertain effect of vegetation response to climate change on potential evapotranspiration. *Hydrol. Earth Syst. Sci.*, *27*(20), 3663-3686. https://doi.org/https://doi.org/10.5194/hess-27-3663-2023
Pimentel, R., Arheimer, B., Crochemore, L., Andersson, J. C. M., Pechlivanidis, I. G., & Gustafsson, D. (2023). Which Potential Evapotranspiration Formula to Use in Hydrological Modeling World-Wide? *Water Resources Research*, *59*(5), e2022WR033447. https://doi.org/https://doi.org/10.1029/2022WR033447
Powell, T. L., Galbraith, D. R., Christoffersen, B. O., Harper, A., Imbuzeiro, H. M. A., Rowland, L., Almeida, S., Brando, P. M., da Costa, A. C. L., Costa, M. H., Levine, N. M., Malhi, Y., Saleska, S. R., Sotta, E., Williams, M., Meir, P., & Moorcroft, P. R. (2013). Confronting model predictions of carbon fluxes with measurements of Amazon forests subjected to experimental drought. *New Phytologist*, *200*(2), 350-365. https://doi.org/https://doi.org/10.1111/nph.12390
Priestley, C. H. B., & Taylor, R. J. F. (1972). On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters. *Monthly Weather Review*, *100*, 81-92.
Purcell, C., Batke, S. P., Yiotis, C., Caballero, R., Soh, W. K., Murray, M., & McElwain, J. C. (2018). Increasing stomatal conductance in response to rising atmospheric CO2. *Annals of Botany*, *121*(6), 1137-1149. https://doi.org/https://doi.org/10.1093/aob/mcy023
Qi, Y., Zhang, Q., Hu, S., Wang, R., Wang, H., Zhang, K., Zhao, H., Zhao, F., Chen, F., Yang, Y., Tang, G., & Hu, Y. (2023). Applicability of stomatal conductance models comparison for persistent water stress processes of spring maize in water resources limited environmental

zone. *Agricultural Water Management*, *277*, 108090. https://doi.org/https://doi.org/10.1016/j.agwat.2022.108090

Ravindra, K., Rattan, P., Mor, S., & Aggarwal, A. N. (2019). Generalized additive models: Building evidence of air pollution, climate change and human health. *Environment International*, *132*, 104987. https://doi.org/https://doi.org/10.1016/j.envint.2019.104987

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195-204. https://doi.org/https://doi.org/10.1038/s41586-019-0912-1

Sakunthala, S., Kiranmayi, R., & Mandadi, P. N. (2017, 17-19 Aug. 2017). A review on artificial intelligence techniques in electrical drives: Neural networks, fuzzy logic, and genetic algorithm. 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon),

Saunders, A., Drew, D. M., & Brink, W. (2021). Machine learning models perform better than traditional empirical models for stomatal conductance when applied to multiple tree species across different forest biomes. *Trees, Forests and People*, *6*, 100139. https://doi.org/https://doi.org/10.1016/j.tfp.2021.100139

Schwalm, C. R., Williams, C. A., Schaefer, K., Baldocchi, D., Black, T. A., Goldstein, A. H., Law, B. E., Oechel, W. C., Paw U, K. T., & Scott, R. L. (2012). Reduction in carbon uptake during turn of the century drought in western North America. *Nature Geoscience*, *5*(8), 551-556. https://doi.org/10.1038/ngeo1529

Stocker, T. (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.

Sulman, B. N., Roman, D. T., Yi, K., Wang, L., Phillips, R. P., & Novick, K. A. (2016). High atmospheric demand for water can limit forest carbon uptake and transpiration as severely as dry soil. *Geophysical Research Letters*, *43*(18), 9686-9695. https://doi.org/https://doi.org/10.1002/2016GL069416

Tikhamarine, Y., Souag-Gamane, D., Ahmed, A. N., Sammen, S. S., Kisi, O., Huang, Y. F., & El-Shafie, A. (2020). Rainfall-runoff modelling using improved machine learning methods: Harris hawks optimizer vs. particle swarm optimization. *Journal of Hydrology*, *589*, 125133. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125133

Tor-ngern, P., Oren, R., Ward, E. J., Palmroth, S., McCarthy, H. R., & Domec, J.-C. (2015). Increases in atmospheric CO2 have little influence on transpiration of a temperate forest canopy. *205*(2), 518-525. https://doi.org/https://doi.org/10.1111/nph.13148

Trugman, A. T., Medvigy, D., Mankin, J. S., & Anderegg, W. R. L. (2018). Soil Moisture Stress as a Major Driver of Carbon Cycle Uncertainty. *Geophysical Research Letters*, *45*(13), 6495-6503. https://doi.org/https://doi.org/10.1029/2018GL078131

Urban, J., Ingwers, M., McGuire, M. A., & Teskey, R. O. (2017). Stomatal conductance increases with rising temperature. *Plant signaling & behavior*, *12*(8), e1356534-e1356534. https://doi.org/https://doi.org/10.1080/15592324.2017.1356534

Vicente-Serrano, S. M., Miralles, D. G., McDowell, N., Brodribb, T., Domínguez-Castro, F., Leung, R., & Koppa, A. (2022). The uncertain role of rising atmospheric CO2 on global plant transpiration. *Earth-Science Reviews*, *230*, 104055. https://doi.org/https://doi.org/10.1016/j.earscirev.2022.104055

Vico, G., Manzoni, S., Palmroth, S., Weih, M., & Katul, G. (2013). A perspective on optimal leaf stomatal conductance under CO2 and light co-limitations. *Agricultural and Forest Meteorology*, *182-183*, 191-199. https://doi.org/https://doi.org/10.1016/j.agrformet.2013.07.005

Warren, J. M., Norby, R. J., & Wullschleger, S. D. (2011). Elevated CO2 enhances leaf senescence during extreme drought in a temperate forest. *Tree Physiology*, *31*(2), 117-130. https://doi.org/10.1093/treephys/tpr002

Wesolowski, A., Blackman, C. J., Smith, R. A., Tissue, D. T., & Pfautsch, S. (2020). Elevated CO2 did not stimulate stem growth in 11 provenances of a globally important hardwood plantation species. *Frontiers in Forests and Global Change*, *3*(66). https://doi.org/https://doi.org/10.3389/ffgc.2020.00066

Wood, S. N. (2016). Just Another Gibbs Additive Modeler: Interfacing JAGS and mgcv. *Journal of Statistical Software*, *75*(7), 1 - 15. https://doi.org/10.18637/jss.v075.i07

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. CRC press.

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548-1563. https://doi.org/https://doi.org/10.1080/01621459.2016.1180986

Wu, W., & Little, T. D. (2011). Quantitative Research Methods. In B. B. Brown & M. J. Prinstein (Eds.), *Encyclopedia of Adolescence* (pp. 287-297). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-373951-3.00034-X

Xu, Z., Shimizu, H., Yagasaki, Y., Ito, S., Zheng, Y., & Zhou, G. (2013). Interactive Effects of Elevated CO2, Drought, and Warming on Plants. *Journal of Plant Growth Regulation*, *32*(4), 692-707. https://doi.org/10.1007/s00344-013-9337-5

Yang, Y., & Roderick, M. L. (2019). Radiation, surface temperature and evaporation over wet surfaces. *Quarterly Journal of the Royal Meteorological Society*, *145*(720), 1118-1129. https://doi.org/https://doi.org/10.1002/qj.3481

Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R., & Donohue, R. J. (2019). Hydrologic implications of vegetation response to elevated CO2 in climate projections. *Nature Climate Change*, *9*(1), 44-48. https://doi.org/10.1038/s41558-018-0361-0

Yuan, W., Zheng, Y., Piao, S., Ciais, P., Lombardozzi, D., Wang, Y., Ryu, Y., Chen, G., Dong, W., Hu, Z., Jain, A. K., Jiang, C., Kato, E., Li, S., Lienert, S., Liu, S., Nabel, J. E. M. S., Qin, Z., Quine, T., . . . Yang, S. (2019). Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Science Advances*, *5*(8), eaax1396. https://doi.org/https://doi.org/10.1126/sciadv.aax1396

Zhang, Y., Zheng, H., Zhang, X., Leung, L. R., Liu, C., Zheng, C., Guo, Y., Chiew, F. H. S., Post, D., Kong, D., Beck, H. E., Li, C., & Blöschl, G. (2023). Future global streamflow declines are probably more severe than previously estimated. *Nature Water*, *1*(3), 261-271. https://doi.org/10.1038/s44221-023-00030-7

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., & Qiu, G. Y. (2019). Physics-Constrained Machine Learning of Evapotranspiration. *Geophysical Research Letters*, *46*(24), 14496-14507. https://doi.org/https://doi.org/10.1029/2019GL085291

Zhou, S., Yu, B., Lintner, B. R., Findell, K. L., & Zhang, Y. (2023). Projected increase in global runoff dominated by land surface changes. *Nature Climate Change*, *13*(5), 442-449. https://doi.org/https://doi.org/10.1038/s41558-023-01659-8

Ziegler, A., & König, I. R. (2014). Mining data with random forests: current options for real-world applications. *WIREs Data Mining and Knowledge Discovery*, *4*(1), 55-63. https://doi.org/https://doi.org/10.1002/widm.1114

Chapter 2: Publication 1

Evaluating CO_2 effects on semi-empirical and empirical stomatal conductance simulation in land surface models

The manuscript was published in the journal of Hydrology.

PUBLICATION 1

This section is to be completed by the student and co-authors. If there are more than four co-authors (student plus 3 others), only the three co-authors with the most significant contributions are required to sign below.

Please note: A copy of this page will be provided to the Examiners.

Full Publication Details	Evaluating CO_2 effects on semi-empirical and empirical stomatal conductance simulation in land surface models
Section of thesis where publication is referred to	Chapter 2

Student's contribution to the publication	85	%	Research design
	100	%	Data collection and analysis
	95	%	Writing and editing

Outline your (the student's) contribution to the publication:

Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

APPROVALS

By signing the section below, you confirm that the details above are an accurate record of the students contribution to the work.

Name of Co-Author 1	Huade Guan	Signed		Date	14/06/2024
Name of Co-Author 2	Margaret Shanafield	Signed		Date	13/6/024
Name of Co-Author 3	Okke Batelaan	Signed		Date	13/6/2024

Nomenclature	
A_n	Net photosynthesis rate (μmol/m^2s)
C_s	Atmospheric carbon dioxide concentration (μmol CO_2 mol^{-1})
C_i	Intercellular carbon dioxide concentration (μmol CO_2 mol^{-1})
g_s	Stomatal conductance (mol/m^2s)
h_s	Relative humidity
iWUE	Intrinsic water use efficiency (μmol CO_2/mol H_2O)
L*	CO_2 compensatory point (μmol CO_2 mol^{-1})
LAI	Leaf area index (m^2/m^2)
R	Net radiation (W/m^2)
S	Soil water content (m^3/m^3)
T_{air}	Air temperature (°C)
T_m, T_0	Maximum and minimum air temperature (°C)
VPD	Vapour pressure deficit (kPa)
a, a_1-a_4, D_0, g_{min}, g_0, g_1, L, m	Calibration parameters
CLM	Community Land Model
JULES	Joint UK Land Environment Simulator
LPJmL	Lund-Potsdam-Jena managed Land
MCMC	Markov-Chain Monte Carlo
NSE	Nash-Sutcliffe efficiency coefficient
RMSE	Root Mean Square Error
MAE	Mean Absolute Error

2.1 Abstract

Ongoing changes in climate and carbon dioxide (C_s) in the atmosphere have profound effects on plant transpiration and, consequently, on the water balance. Land surface models (LSMs) reflect plant response to these changes by simulation of stomatal conductance (g_s). However, the plant response is not well understood and varies with climate. In this study, the simulation of g_s within different LSMs is reviewed and a new approach, a Mixed Generalized Additive Model (MGAM) for g_s simulation, is developed. The alternative g_s estimation is proposed as a solution for the high parameterisation uncertainty in semi-empirical g_s simulation models, and high dependency on mathematical functions for environmental stress factors in empirical g_s simulation models. MGAM has high Pearson and Spearman correlations (87% and 85%, respectively) and efficiency coefficients (71%), with low error values (0.07 mol/m^2s) in g_s simulation. The global sensitivity analysis of the MGAM approach shows the necessity of considering the interaction between C_s and other key climate variables in g_s simulation. The

high accuracy and low uncertainty to the first-order key climate factors in g_s simulation highlight the MGAM model's importance in future studies.

2.2 Introduction

The increase in atmospheric carbon dioxide concentration (C_s) has global direct and indirect effects on the earth system (Yiqi et al., 1999). The C_s has increased from 275 to 415 ppm since the industrial revolution in 1760. The change in the C_s has profoundly affected the climate system and earth's primary productivity (Friedlingstein et al., 2020). High levels of C_s may inhibit photorespiration and increase the net photosynthesis rate (A_n) hence providing better growth and yield productivity (Ahmed et al., 2019). Additionally, C_s significantly stimulates light-saturated photosynthesis and reduces transpiration, which collectively leads to higher light-use efficiency and water use efficiency (WUE) in plants, consequently changing ecosystem water balance (Reinecke et al., 2021; Yiqi et al., 1999). Therefore, an accurate simulation model to clearly reveal the effects of C_s on plants is highly required.

In plants, stomata consist of microscopic pores formed by a pair of guard cells that control the exchange of water, energy and C_s between leaf and atmosphere via regulating stomatal conductance (g_s) (Wu et al., 2021). The elevated C_s effects on g_s is highly variable between climates (Li et al., 2019; Yang et al., 2021; Yang et al., 2019; Zhang et al., 2021) and vegetation (Donohue et al., 2017; Zhu et al., 2021), which is not well understood. The effect of C_s on g_s has been shown contradictory in different studies leading to uncertainty in g_s simulation. For example, a reduction in g_s or transpiration rate has been claimed to be caused by increasing C_s (Faralli et al., 2019; Gimeno et al., 2016; Leuzinger, 2007; Lin et al., 2001; Ward et al., 2012). In contrast, others have argued that a C_s increase does not significantly reduce the g_s (Uddling et al., 2009; Walker et al., 2019). Moreover, increasing leaf area index (LAI) by elevated C_s may offsets the reduction in g_s in unmature plants (Duursma et al., 2016; Norby & Zak, 2011; Purcell et al., 2018). The effect of C_s on g_s is more complex due to interactions of C_s and other climate variables such as atmospheric water vapour, net radiation (R), or air temperature (T_{air}) (Arora et al., 2020; Ceppi & Gregory, 2017; De Kauwe et al., 2021; Haworth et al., 2013; Liao et al., 2021; Medlyn et al., 2001). Several studies have claimed that high vapour pressure deficit (VPD) outweighs the decreasing effect of C_s on g_s (Flexas et al., 2004; Morgan et al., 2004; Xu et al., 2016). Moreover, in some locations with dry conditions and high temperature, C_s has an increasing effect on g_s (Purcell et al., 2018).

The dynamic vegetation response to climate variables (e.g., humidity, VPD, temperature and Cs) has been introduced into various land surface models (LSMs) since the early 2000s (Blyth et al., 2021; Damour et al., 2010; Lei et al., 2014; Liu & Mishra, 2017; Reinecke et al., 2021). The g_s simulation in LSMs can be categorised into semi-empirical and empirical approaches. The semi-empirical approaches integrate biological, physical, and biochemical processes in plants (Lawrence et al., 2020; Lawrence et al., 2019). The semi-empirical g_s simulations are constrained due to large uncertainties caused by model parameterisations (Jiménez et al., 2011; Seneviratne et al., 2010). In addition, g_s simulations with this approach do not account for important stress processes related to plant hydraulics such as water, humidity, or the C_s effect (Green et al., 2019; Wang et al., 2009). Vegetation responds differently to climate variables and C_s changes (Xu et al., 2016). However, the sensitivity of g_s to C_s and environmental conditions is not well-established in semi-empirical g_s simulation approaches (Franks et al., 2017; Jarvis et al., 1976; Konings et al., 2017). Alternatively, empirical g_s simulations use statistical correlations between g_s, environmental factors, and transpiration (Li et al., 2019; Pan et al., 2015). The empirical g_s simulation includes simplifying assumptions and does not consider the net photosynthesis rate (A_n), while the A_n is generally part of the semi-empirical approaches (Damour et al., 2010). Empirical g_s simulation models usually do consider the stress functions of plants; however, the definition of appropriate stress functions is challenging in these models. For example, Noah LSM includes a Jarvis empirical g_s simulation model (Jarvis et al., 1976), in which multiplicative combinations of stress functions (e.g., R, T_{air}, soil moisture, and VPD) are applied to scale down stomatal conductance from the optimal condition (Kumar et al., 2011). The equations of these stress functions vary among different studies (Granier & Loustau, 1994; Wang et al., 2020; Wang et al., 2016; Zeppel et al., 2008). The C_s effect on leaf stomatal conductance can also be added into a Jarvis type model. However, the addition of C_s effects on g_s simulation as a simple linear or hyperbolic function is another challenging issue in these models (Li et al., 2019).

Despite extensive research on LSMs, large uncertainties still exist in quantifying the magnitude of environmental variables on g_s (Friedlingstein et al., 2020). The large discrepancy among independent studies can be attributed to deficiencies in model structures, lack of sufficient measurements, ill-calibrated model parameters, and uncertainty in forcing data (Pan et al., 2020). The uncertainty sources not only stem from meteorological conditions and soil moisture but are also intensified by the physiology and structure of vegetation (Best et al., 2015; Damour et al., 2010). Knowledge of the uncertainties in the g_s and transpiration estimated from different sources is a prerequisite for future water balance prediction (Blyth et al., 2021).

The objective of this study is to evaluate and improve leaf stomatal conductance g_s simulation in LSMs. The data from the Free-Air Carbon dioxide Enrichment (FACE) experiment with leaf level measurement of g_s by (Duursma et al., 2016) and (Gimeno et al., 2016) is used in this study. We first review g_s simulation in semi-empirical and empirical approaches in different LSMs to identify the sources of uncertainty and sensitivity to C_s changes and climate variables. Then we propose a new Mixed Generalized Additive Model (MGAM) for the g_s simulation with lower uncertainty and capable of accounting for the interactions of various environmental influences on g_s. The aim of this research is to provide an alternative solution for future LSM model development through the study of vegetation responses to key climate interactions. The results of this analysis will allow us to predict the effects of anthropogenic climate change on water balance.

2.3 Methodology and data sources

2.3.1 Overview of approaches to gs estimation

The g_s is a key variable in hydrological modelling of plant water use (Wu et al., 2021). The stomatal regulation of g_s and transpiration in plants are quantified by LSMs (Blyth et al., 2021). The LSMs incorporate C_s effects on g_s by semi-empirical and empirical modelling approaches, as presented in Table 1.

Table 1 The g_s simulation equations in LSMs and MGAM with calibrated parameters

Eq. #		g_s approaches in different LSMs	Equation	Calibration parameters	Calibration results
1	**Semi-empirical g_s approaches**	CLM4.5-g_s-Eq.1 based on (Collatz et al., 1991; Oleson et al., 2013)	$g_s = m\frac{1.6 \times A_n}{C_S}h_s + bB$ $B = \begin{cases} 1 & S \geq S_{crit} \\ \frac{S - S_{wilt}}{S_{crit} - S_{wilt}} & S_{wilt} < S < S_{crit} \\ 0 & S \leq S_{wilt} \end{cases}$	m b (mol/m²s)	9 0.01
2		CLM5-g_s-Eq.2 based on (Ball, 1987; Medlyn et al., 2011)	$g_s = g_0 + g_1\frac{1.6 \times A_n \times h_s}{C_S}$	g_0 (mol/m²s) g_1	0.0009 8.01
3		CLM5-g_s-Eq.3 based on (Brooks & Farquhar, 1985; Leuning, 1990; Leuning, 1995)	$g_s = g_0 + g_1\frac{1.6 \times A_n}{(C_S - L^*)(1 + \frac{VPD}{D_0})}$ $L^* = 42.7 + 1.68 \times (T_{air} - 25) + 0.012 \times (T_{air} - 25)^2$	g_0 (mol/m²s) g_1 D_0 (kPa)	-0.06 6.7 3.87
4		CLM5-g_s-Eq.4 based on (Arneth et al., 2002)	$g_s \approx g_0 + (1 + \frac{g_1}{\sqrt{VPD}})\frac{1.6 \times A_n}{C_S}$	g_0 (mol/m²s) g_1(kPa$^{0.5}$)	-0.029 4.29
5		JULES-g_s-Eq.5 based on (Best et al., 2011; Cox et al., 1998)	$g_s = \frac{1.6 \times A_n}{C_S - C_i}$	-	-
6		JULES-g_s-Eq.6 based on (Best et al., 2011; Cox et al., 1999)	$g_s = \frac{1.6 \times A_n}{C_S - C_i}$, $C_i = x \times C_S$	x	0.78

Eq. #		g_s approaches in different LSMs	Equation	Calibration parameters	Calibration results
7		LPJmL-g_s-Eq.7 based on (Haxeltine & Prentice, 1996; Sitch et al., 2003)	$g_s = g_{min} + \frac{1.6 \times A_n}{C_S(1-L)}$	g_{min} (mol/m²s) L	-0.09 0.83
8	Empirical g_s approaches	JSBACH-g_s-Eq.8 based on (Knauer et al., 2015)	$g_s = \beta \times \frac{1.6 \times A_{n_{max}}}{C_S - C_i}$, $Ci = x \times C_S$ $\beta = 1 - a_1 \times exp(a_2 \times \frac{S_{max} - S}{S_{max} - S_{min}})$	x a_1 a_2	0.84 0.41 0.61
9		Noah-g_s-Eq.9 based on (Jarvis et al., 1997; Jarvis et al., 1976; Kumar et al., 2011; Li et al., 2019)	$g_s = g_{s_{max}} \times f(VPD) \times f(S) \times f(C_S)$ $g_s = g_{s_{max}} \times \exp(-a_1 \times VPD) \times 1 - a_2 \times \exp(a_3 \times \frac{S_{max} - S}{S_{max} - S_{min}}) \times \frac{1}{1 + a_4 \times \left(\frac{C_S}{a_5} - 1\right)}$	a_1 a_2 a_3 a_4 a_5	0.38 0.052 2.12 0.38 397
10		MGAM-g_s-Eq.10	$g_s = s(C_s) + s(VPD) + s(S) + s(T_m) + s(h_s) + t_i(VPD, T_m, C_s)$	-	-

2.3.2 Semi-empirical g_s simulation approaches

A common assumption in the semi-empirical g_s simulation approach is the "big leaf" theory, a representation of the leaf-level photosynthesis that treats the canopy like one big leaf (Farquhar, 1989). The equation CLM5-g_s-Eq.2 (Table 1) is a well-known semi-empirical g_s simulation model developed by Ball et al. (1987) and is used in the LSM "Community Land Model" (CLM) (Ball, 1987). In this method, g_s responds to A_n, relative humidity (h_s), and C_s. The CLM5-g_s-Eq.2 method has been criticised for its use of inaccurately simulated A_n values (Damour et al., 2010). Therefore, Leuning (1990; 1995) modified the CLM5-g_s-Eq.2 method by adding the CO_2 compensatory point ($L^{\times}$) and replacing the h_s with VPD to form the CLM5-g_s-Eq.3 method. This method was further improved to CLM5-g_s-Eq.4 by modifying the incorrect g_s simulation when C_s is equal to $L^{\times}$ (Arneth et al., 2002).

The Joint UK Land Environment Simulator (JULES) uses a big leaf approach in JULES-g_s-Eq.5, where g_s is connected to leaf-air CO_2 exchange, so the intercellular carbon dioxide (C_i) variable is added to the equation (Cox et al., 1999) (Table 1). In other studies, the C_i is replaced by a function of C_s; $C_i = x \times C_s$, where x is a calibrated parameter (Knauer et al., 2015; Knorr, 2000). This replacement was implemented in JULES-g_s-Eq.6 and used in cases where the C_i is not available. The Lund-Potsdam-Jena managed Land (LPJmL) model uses a similar approach to JULES. However, the g_{min} has been added as a vegetation specific minimum stomatal conductance (Table 1), which should be calibrated for different vegetation types (Sitch et al., 2003).

In all semi-empirical LSMs, an A_n and g_s relationship is included as a constraint to couple carbon and water processes. However, LSMs have different representations of the A_n and g_s relationship, particularly in the capability for simulating the CO_2 concentration effect. Moreover, the coupling or decoupling of g_s and A_n is still debated among different studies (Ameye et al., 2012; Collatz GJ et al., 1992; Drake et al., 2018; Krich et al., 2022; Schulze et al., 1973; Tuzet et al., 2003; J. Urban et al., 2017; von Caemmerer & Evans, 2015; Yun et al., 2020). Therefore, for comparing semi-empirical LSMs in g_s simulation, we have assumed g_s is unknown and all other parameters, including A_n and climate variables, are known as input in the g_s simulation equations in Table 1.

2.3.3 Empirical gs simulation approaches

In empirical models, various interconnections between plant components and environmental conditions are defined via empirical mathematical concepts for the g_s simulation. These models estimate g_s independent of the A_n variable. The result of empirical g_s models strongly depends on the quality of the observed input data (Jaiswal et al., 2020). The effects of environmental conditions on plants are computed as stress functions. One of the empirical g_s simulation models is the Jarvis equation integrated in the Noah LSM, which estimates g_s directly by reducing a maximum g_s for the optimal environmental condition using stress functions of actual environmental conditions (Liu et al., 2019). In this research, different forms of stress functions were reviewed from various studies (Table S1). Most versions of the Jarvis equation treat C_s effects on g_s as a simple linear process, except the model used by Li et al. (Li et al., 2019). Another empirical g_s simulation model is JSBACH-g_s-Eq.8 (Table 1). This model uses a maximum value of A_n (without stress) instead of variable A_n and has a soil moisture stress function (Knauer et al., 2015). The soil moisture stress function in JSBACH has been modified in this study based on different soil moisture stress functions in Table S1.

2.3.4 MGAM for gs simulation

MGAM is a powerful modelling technique to simulate complex nonlinear relationships between variables and responses (Wood et al., 2016). This approach is used when at least one parameter or variable appears to be nonlinear. The MGAM simulation process is based on developing multiple nonlinear functions to predict the outcome of the dependent or independent variables and parameters with the help of the degree of relationship among them. MGAM uses flexible regression functions (smoother function), which model the relationships between covariates and outcomes where the shape of the function itself varies between different groups of datasets (Hastie et al., 2009). Apart from the regular smooth function (S) to reflect the nonlinearity of variables, a tensor function (t_i) can be used when the interaction between variables is statistically significant. The generic form of the MGAM model is

$$f(x) = \sum_{k=1}^{K} \beta_k b_k(x) \tag{11}$$

where, $f(x)$ is a smoother function, b_k are basis functions, β_k are corresponding coefficients, and K is referred to as basis size or basis complexity. The coefficients of the basis functions are optimised to ensure the appropriate complexity of the models. The large basis size could lead to overfitting, but it is counteracted by a penalty term, to maximise the penalised log-likelihood as in Eq. 12.

$$L_\rho = L - \lambda\beta^T S\beta \tag{12}$$

where, L_ρ is penalised log-likelihood, L is the model likelihood, S is the penalty matrix, $\beta^T S\beta$ is the penalty term for vector β, and λ controls the trade-off between log-likelihood and penalty term (Wood, 2016).

At a low value of $\lambda=0$, the penalty has no effect, and the model is too complex with high wiggliness, but at high values of $\lambda\rightarrow\infty$, the penalty is high resulting in a simple linear model (Hastie et al., 2009; Wahba, 1990). The 'nls' and 'mgcv' packages in R are used for MGAM-g_s simulation in this study (Baty et al., 2015; Wood et al., 2016).

The structure of g_s simulation in MGAM can be described as Eq. 13.

$$g_s = \sum_{m=1}^{M} f(x_m) \tag{13}$$

where, M are the effective variables on g_s (e.g., climate variables, C_s, and soil moisture). Each of the effective variables has the smoother function $f(x)$ (Eq. 11), which contains basis functions with relevant coefficients.

2.4 Description of the dataset and the case study

The input data of this study were collected from the Western Sydney University website (Duursma, 2015; Duursma et al., 2016). The data is the result of the Eucalyptus FACE experiment (EucFACE) with Eucalyptus-dominated mature woodland in western Sydney (Australia, 33°37´S, 150°44´E, 30 m a.s.l.) from October 2012 to November 2013 (Duursma et al., 2016). The case study is characterised as a humid temperate-subtropical transitional climate (Duursma et al., 2016; Gimeno et al., 2016). The mean annual precipitation is 800 mm and the mean annual temperature is 17 °C from 1881 to 2014; the estimated potential evapotranspiration (PET) is 1350 mm from 1950 to 2000 (Duursma et al., 2016; Zomer et al., 2008). The soil at this site is loamy sand with more than 75% sand content in the top 50 cm, and sandy clay loam with more than 30% silt and clay from 50 to 300 cm depth (Crous et al., 2015; Duursma et al., 2016).

The EucFACE consisted of six 25 m diameter circular plots (rings), each ring having 39 ± 3 canopy trees, with approximately 17 dominant and co-dominant canopy-forming trees. Seven campaigns of leaf gas exchange and water potential measurements were performed. The leaf-level CO_2 and H_2O exchange measurements were performed with four open-flow portable photosynthesis systems (Li-6400, Li-Cor, Inc., Lincoln, NE, USA). The A_n and g_s were measured under 1800 µmol m^{-2} s^{-1} photon flux density (provided by the in-built Li-6400 red-blue LED lamp). The C_s level was increased gradually from ambient level (390 µmol CO_2 mol^{-1}) to elevated level (540 µmol CO_2 mol^{-1}), starting from September 2012, and reached to full operation model in February 2013. Three rings were exposed to elevated C_s, while the three ambient rings were used as control plots. The elevated C_s does not affect LAI of the mature trees in the case study (Duursma et al., 2016). This study focuses on 160 observed data points over 11 days, which had all the necessary variables for both the semi-empirical and empirical g_s simulation models (the rest of the data does not contain all variables). The main objective of this study is finding the effects of CO_2 changes on g_s, and available data had the full coverage for the CO_2 range from ambient to elevated level and covered all seasons 2012-2013.

2.5 Calibration and validation processes of g_s simulation models

2.5.1 MCMC-Bayesian calibration

Bayesian inference is an important approach for calibration, especially in complex environmental and ecological models (Speich et al., 2021). The Markov-Chain Monte Carlo

(MCMC) algorithm is the methodological backbone of the Bayesian approach (Speich et al., 2021). The evaluation of parameterisation in Bayesian theory is based on the likelihood as the goodness of fit. The likelihood p(D|θ) is defined as a probability of observation data (D) occurring given the model parameterisation with θ. The term θ represents different parameterisations of the model. The best choice for θ is the value with the highest likelihood (maximum likelihood estimation) defined in Eq. 14 (Hartig et al., 2012).

$$p(D|\theta) \propto e^{\frac{-|M(\theta)-D|^2}{2\times\sigma^2}} \tag{14}$$

where, $M(\theta)$, are the model prediction results from the parameterisation θ, D is the observed data, and σ is the standard deviation of the error.

Moreover, additional independent information related to parameters should be investigated in the parameterisation. We have used Bayes theory to merge independent information into the likelihood function, as shown in Eq. 15.

$$p(\theta|D) = \frac{p(D|\theta) \times p(\theta)}{p(D)} \tag{15}$$

where, $p(\theta|D)$ is posterior density (or probability density) that summarises the information for probable values of θ. The posterior density ($p(\theta|D)$) depends on the likelihood ($p(D|\theta)$), and a new term of $p(\theta)$ that is called the prior. In each iteration, this new information will be merged with the existing information by using the posterior distribution from the old data as the prior for the new data (Hartig et al., 2012).

The posterior density calculation in Bayesian inference is computationally demanding due to its high dimensionality. Therefore, MCMC was used to generate a sample of data from the posterior distribution to solve this problem. The MCMC performs a random walk in parameter space by the stochastic Markov process. The Markov process was chosen such that the probability of each parameter combination is proportional to its posterior density. There are different algorithms for the Markov process, such as Metropolis-Hastings, Gibb's sampling, Sequential Monte Carlo, or Differential Evolution (DEzs) (Speich et al., 2021). The DEzs process was chosen in this study as the MCMC algorithm as it is more efficient than other methods. In DEzs, different datasets run in parallel; therefore, choosing an appropriate scale and orientation of the distribution is more efficient than other Markov algorithms (ter Braak & Vrugt, 2008). The 'BayesianTools' and 'mcmc' packages in R are used for the calibration process in this study (Geyer & Johnson, 2020; Hartig et al., 2019).

2.5.2 Cross-validation technique

160 observed data points were used in this study for the g_s simulation. 80% of the data was used to train the models and for the calibration processes, while 20% was used to test the models. A 10^{th} fold cross-validation was used with ten iterations. The training data were randomly split into 10 folds, and the model was trained by 9 folds, then it was validated by the remaining 10^{th} fold. The 'caret' package in R was used for the cross-validation (Kuhn, 2021).

2.6 Intrinsic Water Use Efficiency response from g_s models

The intrinsic Water Use Efficiency (iWUE) is defined as the ratio of carbon assimilation (μmol/m^2s), A_n, over g_s (Eq. 16), which is used to measure the adaptability of plants to changes in environmental conditions (Zhang et al., 2019).

$$\text{iWUE} = \frac{A_n}{g_s} \tag{16}$$

The iWUE has received considerable attention due to the recent increase in iWUE in many ecosystems. Several observational (Keenan et al., 2013; Mastrotheodoros et al., 2017) and theoretical (Knauer et al., 2017) studies attributed this phenomenon to rising C_s (Zhang et al., 2019). In LSMs, iWUE is a new index that reflects plants' adaptability to changing environmental conditions (Blyth et al., 2021; Zhang et al., 2019).

In this study, the observed A_n and simulated g_s for the various referenced models were used to estimate iWUE. The response of iWUE to the C_s scenarios is calculated from different g_s simulation models.

2.7 Sensitivity analysis

The Sobol sensitivity analysis is used to determine how much of the variability or the uncertainty of the output model depends on each of the input indices (variables and parameters). Also, it determines if these indices act singularly or if there are interactions between different indices. The Sobol method is a variance-based uncertainty and sensitivity analysis that represents the first, second, and total order of variance-based estimators to understand how output variance is attributed to individual indices or the interaction between indices (Puy, 2021).

It is a common approach to measure local sensitivity (or one-at-a-time analysis) to define the model output changes in terms of one-index variation when all other indices are maintained at a fixed value (Saltelli et al., 2019). This approach does not sufficiently identify the interactions between indices. The Sobol method (as a global sensitivity analysis) fills this gap by studying the interactions of uncertain parameters on the output of the simulation model, even for nonlinear systems (Saltelli et al., 2008). The Sobol method perturbs input indices based on their ranges and then defines the model output uncertainty using variance as in Eq. 17 (Puy, 2021; Saltelli et al., 2008).

$$\mathrm{V(y)} = \mathrm{V_{xi}[E_{x\sim i}(y|x_i)] + E_{xi}[V_{x\sim i}(y|x_i)]} \tag{17}$$

where, $V_{xi}[E_{x\sim i}(y|x_i)]$ and $E_{xi}[V_{x\sim i}(y|x_i)]$ are the first-order effects of the x_i and residual, respectively, E(.) and V(.) are the mean and variance operators, $y = f(x)$ is a scalar output and $x = x_1, x_2, \ldots, x_k$ are uncertain inputs parameters, $x{\sim}i$ denotes all parameters except x_i. $V(y)$

can be decomposed to all partial variances up to the k[th] order as Eqs. 18-19 (Saltelli et al., 2008).

$$\mathrm{V(y)} = \sum_{\mathrm{i=1}} \mathrm{V_i} + \sum_{\mathrm{i}} \sum_{\mathrm{i<j}} \mathrm{V_{ij}} + \cdots + \mathrm{V_{1,2,\ldots,k}} \tag{18}$$

where,

$$\begin{aligned} \mathrm{V_i} &= \mathrm{V_{xi}}[\mathrm{E_{x\sim i}}(\mathrm{y}|\mathrm{x_i})]\ , \\ \mathrm{V_{ij}} &= \mathrm{V_{xi,xj}}\left[\mathrm{E_{x\sim i,j}}\left(\mathrm{y}\middle|\mathrm{x_i},\mathrm{x_j}\right)\right] - \mathrm{V_{xi}}[\mathrm{E_{x\sim i}}(\mathrm{y}|\mathrm{x_i})] - \mathrm{V_{xj}}\left[\mathrm{E_{x\sim j}}\left(\mathrm{y}|\mathrm{x_j}\right)\right] \end{aligned} \tag{19}$$

The Sobol indices are then calculated as Eq. 20 (Saltelli et al., 2008).

$$\mathrm{S_i} = \frac{\mathrm{V_i}}{\mathrm{V(y)}}\ , \quad \mathrm{S_{ij}} = \frac{\mathrm{V_{ij}}}{\mathrm{V(y)}} \tag{20}$$

where, S_i is the first-order effects of x_i, and S_{ij} is the second-order effect of (x_i, x_j).

Total order index T_i, which is the first-order effects of x_i and its interactions with all other parameters can be measured by Eq. 21 (Saltelli et al., 2008).

$$\mathrm{T_i} = 1 - \frac{\mathrm{V_{x\sim i}}[\mathrm{E_{xi}}(\mathrm{y}|\mathrm{x_{\sim i}})]}{\mathrm{V(y)}} = \frac{\mathrm{E_{x\sim i}}[\mathrm{V_{xi}}(\mathrm{y}|\mathrm{x_{\sim i}})]}{\mathrm{V(y)}} \tag{21}$$

As an example, for a three-dimensional model, the total-order index of x_1 is the sum of the first, second, and third-order effects of x_1 as in Eq. 22.

$$\mathrm{T_1} = \mathrm{S_1} + \mathrm{S_{1,2}} + \mathrm{S_{1,3}} + \mathrm{S_{1,2,3}} \tag{22}$$

The methodology used in this study has been summarised in Fig. 1. The input data was divided to train and test data for calibration and test of g_s simulation approaches in different LSMs. The input data, including A_n, climate variables, and calibrated parameters, were defined for each g_s simulation approach in LSMs. MCMC-Bayesian calibration process was performed for g_s simulation in semi-empirical and empirical models. The calibration process requires a calibration range for each fitted parameter in semi-empirical and empirical models (Fig. S1). The 10th fold cross validation process was performed for all simulation models. MGAM is independent of A_n and MCMC-Bayesian calibration. The differences between the various g_s simulation approaches in LSMs and MGAM are evaluated through g_s simulation, iWUE estimation, and global sensitivity analysis.

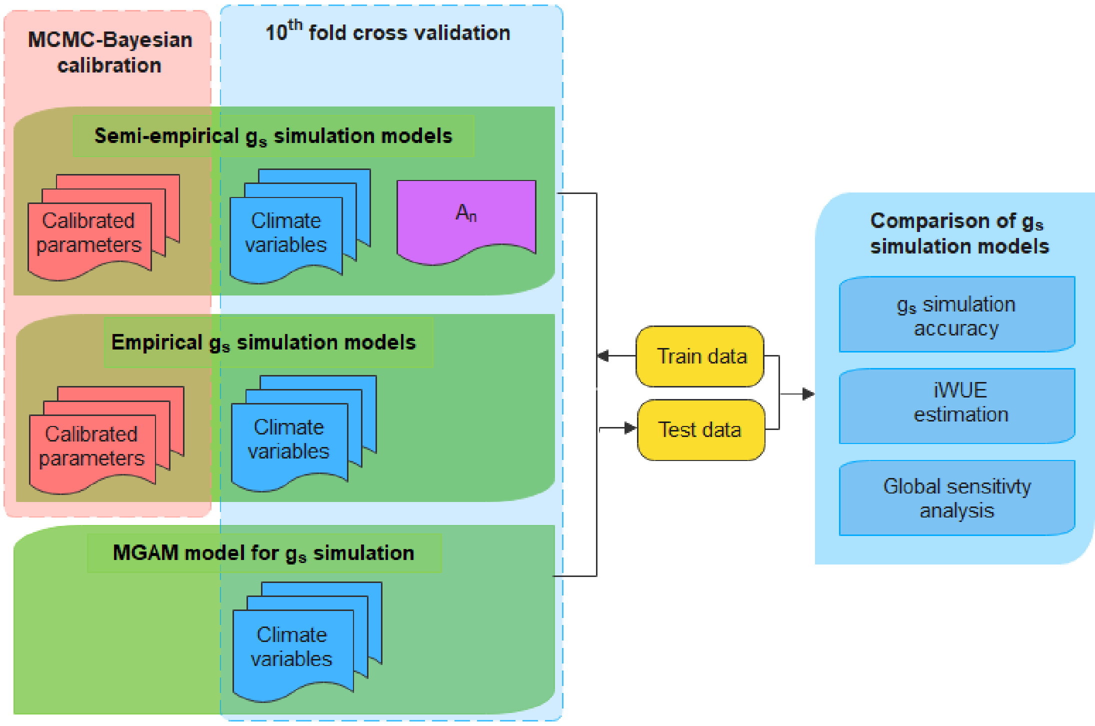


Figure 1 Conceptual diagram of the methodology. The g_s simulation in semi-empirical and empirical models depend on the A_n variable and calibration process. All g_s simulation models were validated by 10th fold cross validation, and compared by g_s simulation, iWUE estimation, and global sensitivity analysis.

2.8 Results

2.8.1 g_s simulation results

The calibration process for semi-empirical and empirical g_s simulation models has been performed with the Bayesian and MCMC method (Table 1). The results of the g_s simulation for all models were compared with the observed g_s values for both the training and testing data (Fig. 2). Simulation performance was measured by assessment criteria such as correlation coefficients (Pearson, Spearman, and Kendall), Nash-Sutcliffe efficiency coefficient (NSE), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) (Fig. 3). The semi-empirical g_s simulation approaches performed well based on all the assessment criteria for testing and training. The coefficient values for Pearson, Spearman, Kendall, and NSE for test data were 88-97%, 91-97%, 75-89%, and 73-94%, respectively. The error values for semi-empirical g_s simulation models were 0.03-0.07 and 0.02-0.05 (mol/m^{2}s) for RMSE and MAE, respectively. The results of the empirical g_s simulation models showed a lower accuracy compared to the semi-empirical g_s simulation models. The JSBACH-g_s-Eq.8 showed 82%, 76%, 59%, and 66% for Pearson, Spearman, Kendall, and NSE coefficient, respectively, while these values for the Noah-g_s-Eq.9 were 73%, 71%, 53%, and 52%, respectively. The JSBACH-g_s-Eq.8 RMSE and MAE values were respectively 0.08 and 0.06 (mol/m^{2}s), while Noah-g_s-Eq.8 error values were respectively 0.11 and 0.09 (mol/m^{2}s).

The results of the MGAM-g_s simulation model in correlation and efficiency coefficients were 87%, 85%, 65%, and 71% for respectively Pearson, Spearman, Kendall, and NSE coefficients in test data. The error values in the MGAM-g_s model were 0.07 and 0.06 (mol/m^{2}s) for respectively RMSE and MAE. This demonstrates that the MGAM-g_s simulation model has better results than the empirical g_s simulation approaches for the test data. It is worth noting that all models shown in Fig. 3 have been cross-validated, so there is no overfitting or underfitting between the training and test results. The slight improvement in test data for the MGAM model may result from model performance on unseen and randomly chosen test datasets or a low number of test data samples.

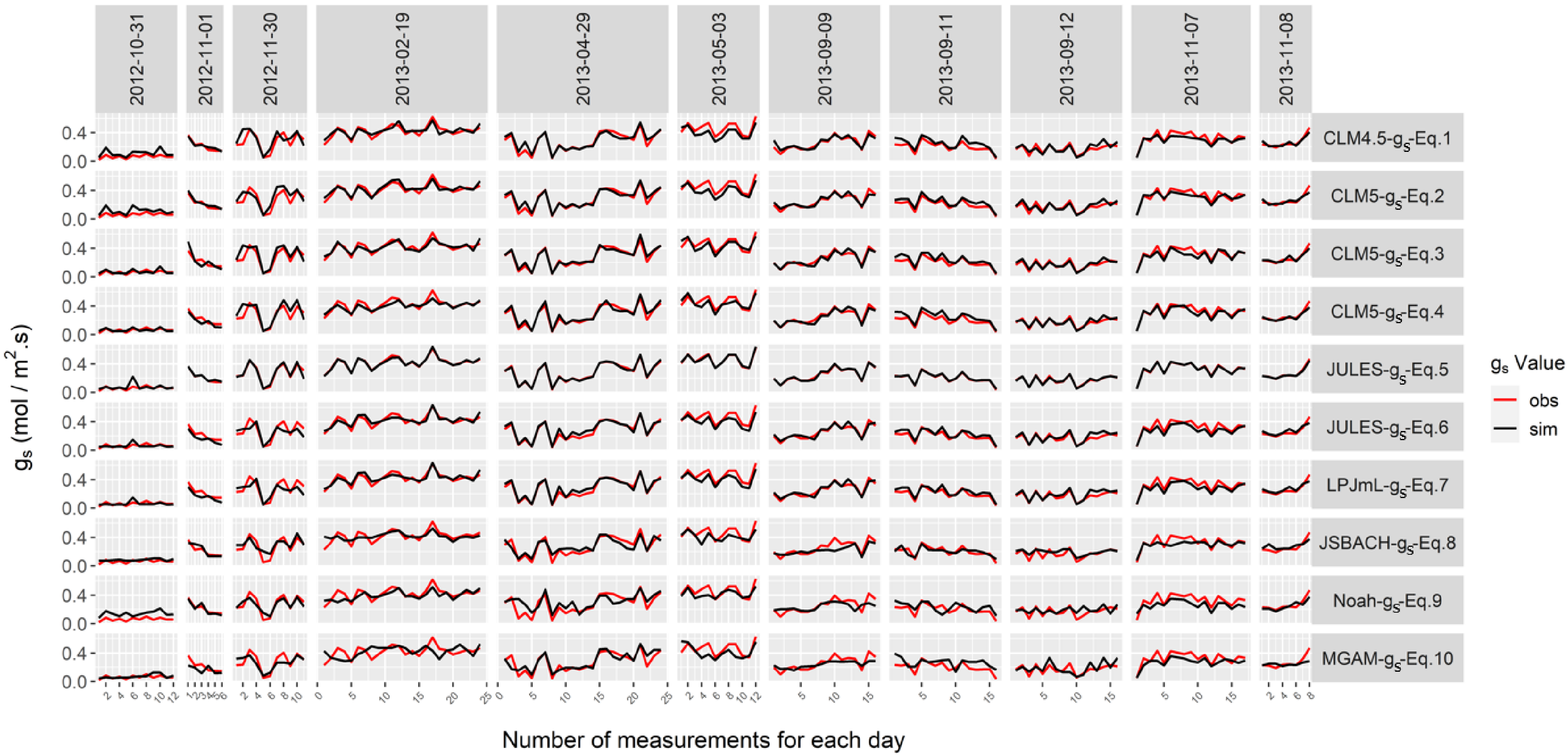


Figure 2 Simulated (black line) and observed (red line) g_s for 11 days (columns) and ten different g_s simulation approaches (rows).

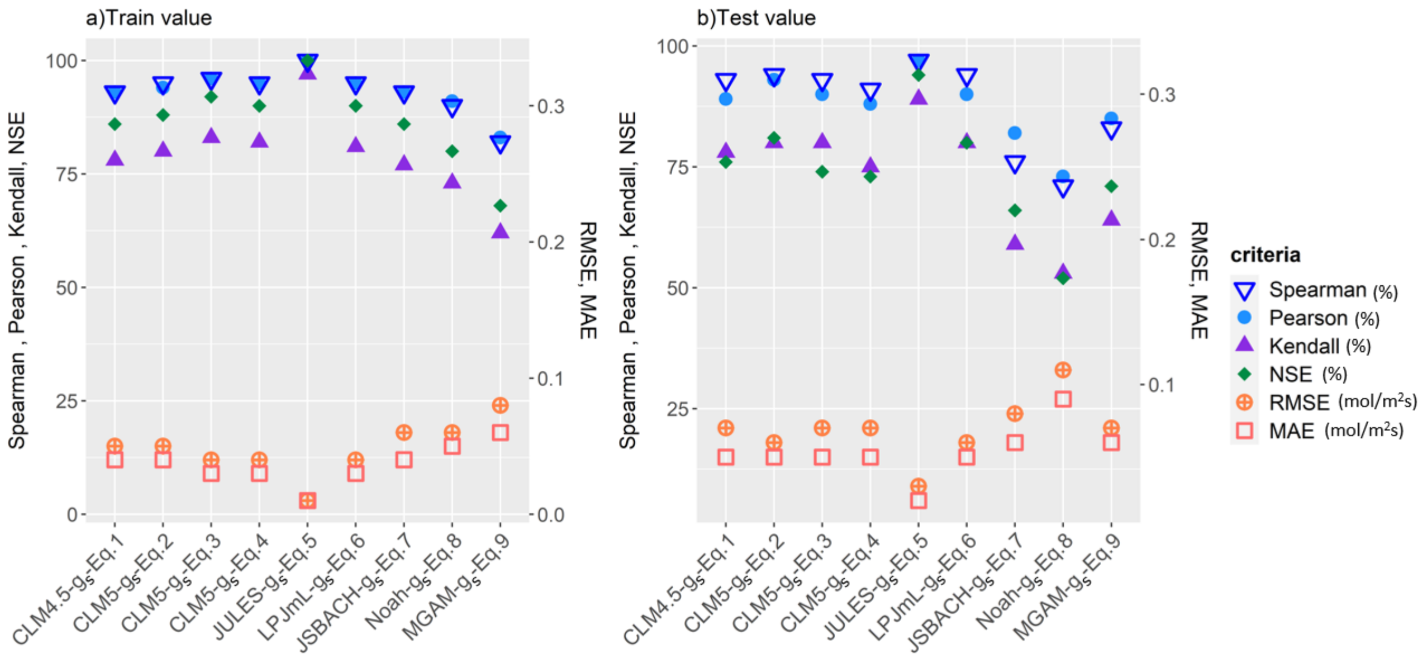


Figure 3 Performance of gs simulation approaches as measured by six criteria for a) training; and b) testing.

2.8.2 The impact of g_s on iWUE

The scatter plot in Fig. 4 confirms that elevated C_s will increase the iWUE as expected. The iWUE for all g_s simulation models increased by 18 to 25% for the C_s gradual rise from 390 to 540 μmol CO_2 mol^{-1}. The results for the semi-empirical g_s simulation models of iWUE for both scenarios show a better fit. However, for the g_s values less than 0.1 mol/m^2s, the simulated iWUE was lower than the observed iWUE, especially for CLM4.5-g_s-Eq.1 and CLM5-g_s-Eq.2. A similar result was observed for the g_s less than 0.1 mol/m^2s in JSBACH-g_s-Eq.8, Noah-g_s-Eq.9, and MGAM-g_s-Eq.10. However, the dispersion was mostly located in the lower part of the iWUE-g_s curve, while the upper part of the curve was more concentrated for both simulated and observed iWUE, expect few points for elevated C_s in Noah-g_s-Eq.9.

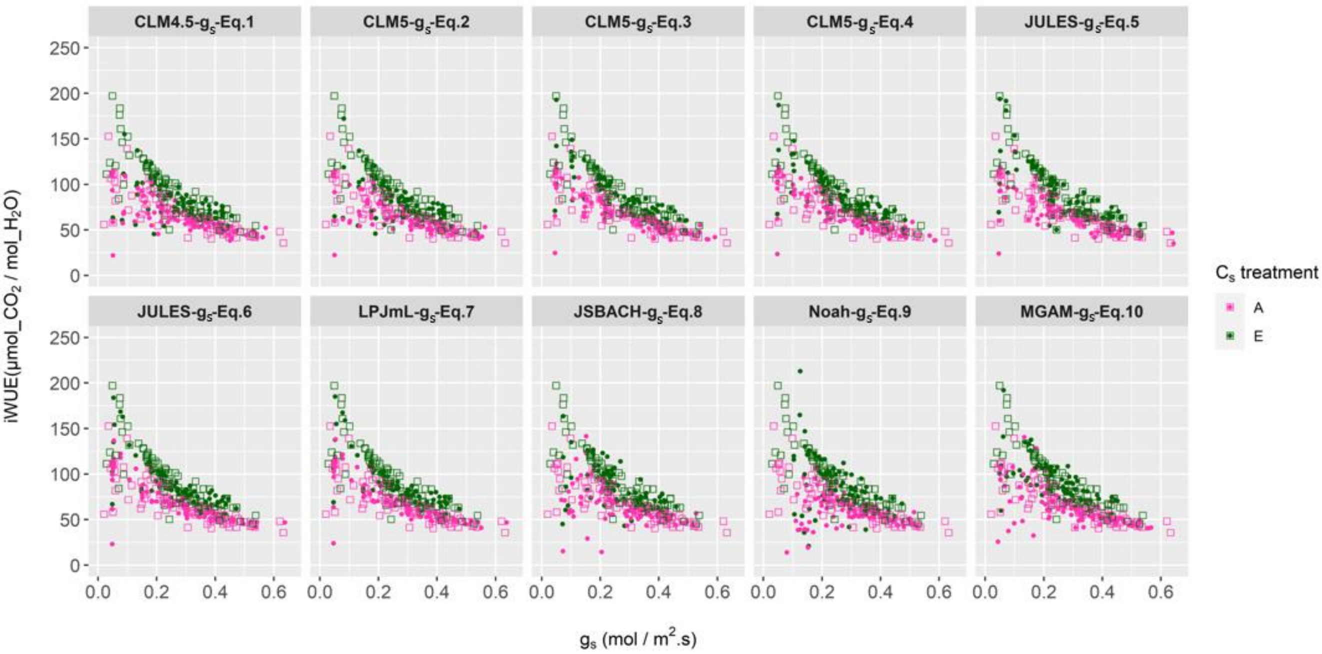


Figure 4 The iWUE related to the observed and simulated g_s for both ambient and elevated C_s scenarios. The pink and green squares are observed values for respectively ambient and elevated C_s scenarios. The pink and green dots are simulated values for respectively ambient and elevated C_s scenarios.

2.8.3 Sensitivity analysis of g_s to control indices

The variance-based sensitivity analysis of g_s to control indices (all parameters and variables) in each g_s simulation approach was performed (Fig. 5 and Table 2). The first-order effects of indices (S_i), which define the dominating indices in the g_s uncertainty, are shown in red bars in Fig. 5. The second (S_{ij}) and total order effects (T_i), which define the effects of the interaction of two and all indices, are shown in green, and purple bars, respectively (Eqs. 18-19). In this part, all input indices have been perturbed (Fig. S1), and then the most effective indices, which contribute to the g_s variability have been ranked among different variables (e.g., A_n, h_s, S, VPD, C_i, C_s, T_m) and calibrated parameters (e.g., a, a_1-a_4, D_0, g_{min}, g_0, g_1, L, m) for each g_s simulation equation. Sensitivity values lower than 0.05 were eliminated to distinguish the dominating indices better from the unimportant ones (Zhang et al., 2015). Hence, the T_i values in some indices were higher than the sum of S_i and S_{ij}.

All semi-empirical g_s simulation approaches show high sensitivities to the indices (Table 2). The first-order effects of indices are high, 75% to 94%. The high value of S_i in semi-empirical g_s simulation models was attributed to calibrated parameters such as b and m in CLM4.5-g_s-Eq.1, g_0 in all CLM5-g_s-Eq.2-3-4, a in JULES-g_s-Eq.6, and g_{min} and L in LPJmL-g_s-Eq.7. A_n also had high effects on g_s variance in all semi-empirical g_s simulation models (except LPJmL-g_s-Eq.7). The g_s sensitivity in semi-empirical g_s simulation models was significantly affected by calibrated parameters, which shows the high g_s uncertainty to calibrated parameters. There was no effective interaction (second-order sensitivity) between variables in these approaches.

The g_s sensitivity in JSBACH-g_s-Eq.8 was highly affected by the calibrated parameter (a_1). The S_i value was 0.86. However, there was no effective interaction between indices for g_s variation. The g_s sensitivity in Noah-g_s-Eq.9 was different because g_s was less sensitive to indices (S_i = 0.33). The key control indices in Noah-g_s-Eq.9 were the calibrated parameter (a_2) and the S variable. There was little interaction between VPD-a_2 and VPD-S, which was negligible.

The MGAM-g_s-Eq.10 model was expected to have a different sensitivity of g_s to input variables. We have obtained the basis functions in MGAM by testing all possible combinations of environmental variables, and the combination with the highest simulation accuracy has been suggested in Table 1. The basis functions, which illustrate the influence of interactive environmental effects on g_s, vary based on the type of vegetation since, as reported previously, each type of vegetation is sensitive to the specific interactive environmental effects (Kimm et al., 2020; Yang et al., 2022). As a result, sensitivity analyses for different basis functions are not necessary for this type of vegetation. The MGAM had lower first-order sensitivity (S_i=0.33) to key climate variables, such as VPD, T_m, C_s, h_s, and S. The g_s variation was more sensitive to VPD and T_m, while other variables indirectly affected g_s due to their interactions. The MGAM also has considerable interaction between VPD-C_s and VPD-T_m as second-order sensitivity.

Table 2 Sum of S_i and S_{ij} values for different g_s simulation models. The S_i defines the direct effects of dominating indices in g_s sensitivity, and the interactions of indices are shown by S_{ij}.

	CLM4.5-g_s-Eq.1	**CLM5-g_s-Eq.2**	**CLM5-g_s-Eq.3**	**CLM5-g_s-Eq.4**	**JULES-g_s-Eq.5**	**JULES-g_s-Eq.6**	**LPJml-g_s-Eq.7**	**JSBACH-g_s-Eq.8**	**Noah-g_s-Eq.9**	**MGAM-g_s-Eq.10**
Sum of S_i	0.75	0.8	0.88	0.85	0.8	0.85	0.94	0.86	0.33	0.33
Sum of S_{ij}	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.19	0.33

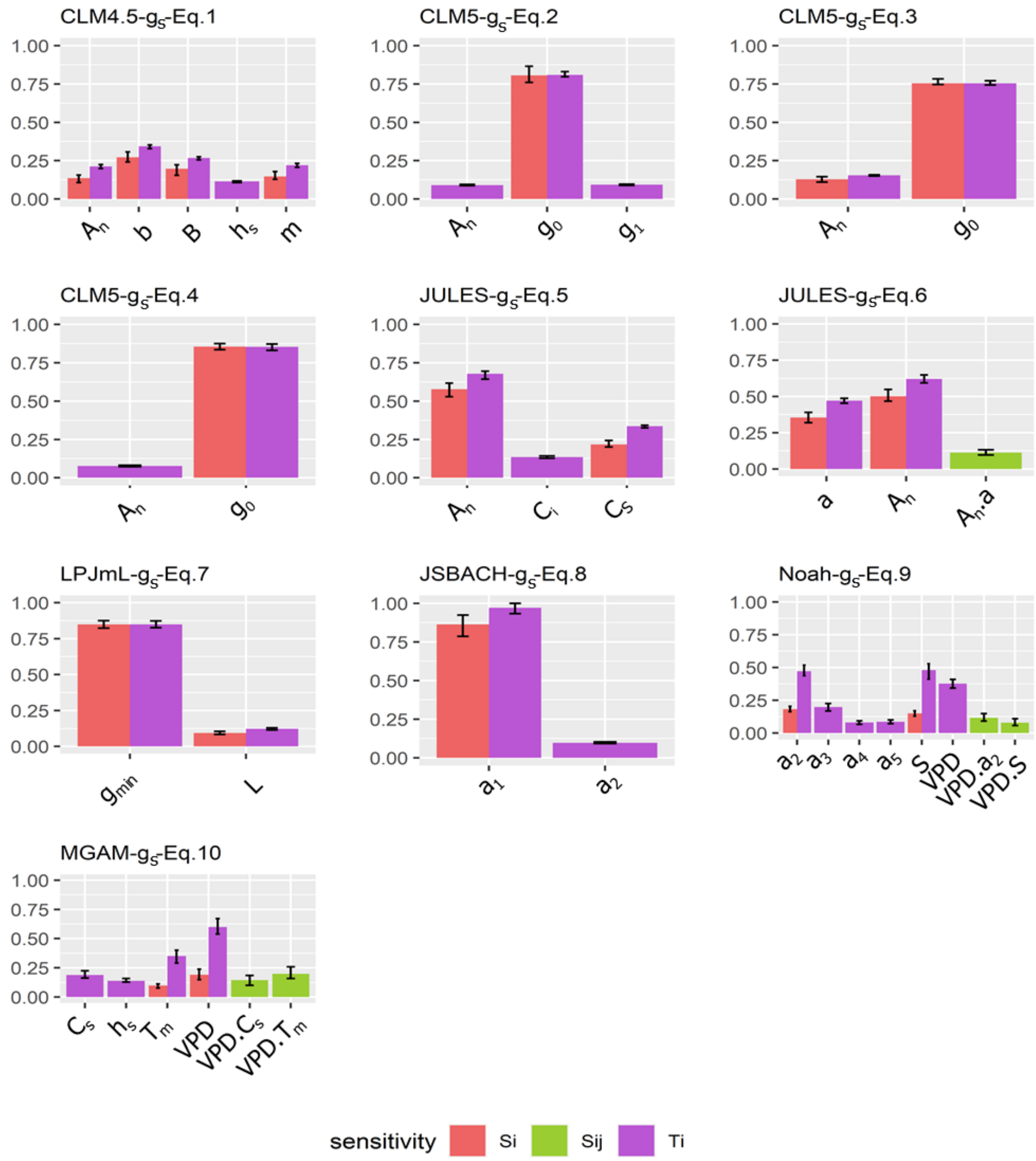


Figure 5 Sensitivity analysis (S_i, S_{ij}, and T_i) of g_s to different indices. The zero and one along the y-axis indicate insensitivity and sensitivity of g_s to indices, respectively. Sensitivities of less than 0.05 have been eliminated. The S_i shows the direct effects of indices in the g_s uncertainty, S_{ij} shows interactions of two indices and T_i is the sum of all direct and interactions effects of indices.

2.9 Discussion

2.9.1 Semi-empirical and empirical g_s simulation approaches

In this study, the semi-empirical g_s simulation models appear to have high accuracy in g_s simulation performance. This result is not surprising because these g_s equations contain the A_n

variable, which is highly correlated with g_s and directs stomatal behaviour (Radin et al., 1988). However, the large uncertainties to parameterisation and the A_n variable, simplification of tree canopy as a big-leaf in semi-empirical g_s simulation models constrains semi-empirical g_s simulation models in global scale (Blyth et al., 2021; Zhang et al., 2020). In contrast, the empirical g_s simulation models, such as Noah-g_s-Eq.9 and JSBACH-g_s-Eq.8, have reduced the uncertainties and limitations of semi-empirical g_s simulation models. However, finding the appropriate mathematical functions regarding environmental factors and complex calibration processes affects the efficiency of empirical models in g_s simulation (Knauer et al., 2015; Lhomme et al., 1998). Moreover, applying the proper value for g_{smax} and A_{nmax} in Noah-g_s-Eq.9 and JSBACH-g_s-Eq.8, respectively are other limitations of these models.

As mentioned above, the limitations of semi-empirical and empirical g_s simulation approaches call for a new method for g_s simulation. The MGAM-g_s simulation model does not require the A_n variable and calculates g_s directly from environmental variables, bypassing the need for ranges of unknown fitted parameters in the MCMC-Bayesian calibration process, g_{smax}, and A_{nmax} values. Climate variables (e.g., C_s, h_s, VPD, T_m) and soil water content (S) were used to simulate g_s in this new approach. The MGAM-g_s model has the following advantages: it relaxes the assumptions and limitations of semi-empirical g_s simulation models; it simulates g_s independent of the A_n variable; it can simulate g_s accurately. As a result, there is less first-order uncertainty regarding key climate variables compared to semi-empirical models, and more complex mathematical concepts regarding stress functions are removed from empirical models.

2.9.2 The impact of g_s on iWUE

The g_s impact on iWUE gives another viewpoint to the g_s simulation model's performance. In all models, the iWUE was enhanced when C_s increased. However, there were some differences depending on the g_s value (higher or lower than 0.1 mol/m^{2}s). Similar results have been reported (Li et al., 2017; Mathias & Thomas, 2021; Zhang et al., 2019), suggesting iWUE improvements in C_s enrichments scenarios. However, there was no identification of the different impact of g_s on iWUE regarding g_s values. To the best of our knowledge, it is the first time that the impact of g_s on iWUE has been evaluated based on the different values of g_s. The results show that most of the LSMs underestimate simulated iWUE, for g_s value less than 0.1 mol/ m^2s. However, they performed better at higher g_s values. The underestimation of the simulated iWUE for low g_s values (g_s lower than 0.1 mol/m^2s) was greater using JSBACH-g_s-Eq.8, Noah-g_s-Eq.9, and MGAM-g_s-Eq.10. Noah-g_s-Eq.9 has shown both overestimation and underestimation for the simulated iWUE at low g_s values, especially for the elevated C_s scenario.

2.9.3 The importance of sensitivity analysis of g_s simulation approaches

The g_s sensitivity analysis results, presented in section 3.3, and the key controlling indices that affect g_s variance were identified for each g_s simulation approach. The semi-empirical g_s simulation models showed high first-order sensitivity to calibrated parameters and the A_n variable. The sensitivity of g_s to calibrated parameters, which change by vegetation type, makes these models computationally demanding. However, these models did not show any interaction between variables as second-order sensitivity (Fig. 5). Jiménez et al. (2011) applied an intercomparison of LSMs output and highlighted the difficulties in using LSMs models and the necessity of improve formulations to cope with model uncertainties (Jiménez et al., 2011). Blyth et al. (2021) have suggested that LSMs need improvements to represent important processes in the real world such as interactions between climate variables and vegetations (Blyth et al., 2021).

The new MGAM-g_s simulation approach can define interaction between key climate variables' effects on g_s which has consistency to the real world. The sensitivity analysis results in Fig. 5 and Fig. S1 for MGAM-g_s-Eq.10 show that VPD and T_m are key climate variables, which affect g_s variation, in addition to the interaction between VPD-T_m and VPD-C_s. The increase in VPD causes a reduction in g_s values (Fig. S1 for MGAM-g_s-Eq.10), which confirms previous studies (Creese et al., 2014; Inoue et al., 2021; Jiao et al., 2019). In higher VPD conditions, the guard cells (two cells that surround a stoma) are vulnerable to turgor loss and close the stomata to decrease the conductance of gas diffusion and water loss via stomata (Inoue et al., 2021). The T_m variable increases g_s for a small margin (Fig. S1). The increase in g_s by T_m can be explained by mesophyll conductance increase, which supplies more water for evaporation and increases guard cell turgor and stomata aperture (Josef Urban et al., 2017). When g_s is increased (by a high temperature), the trees increase their rate of evaporative cooling to survive in hot and dry conditions (Josef Urban et al., 2017). Since the increase in temperature enhances the VPD, the interaction between VPD-T_m on g_s is important (von Caemmerer & Evans, 2015). The global sensitivity analysis of MGAM-g_s-Eq.10 shows the interaction between VPD-T_m (Fig. 5). The interaction between VPD-T_m was shown in several studies; when VPD is high the effect of temperature on g_s is larger than when VPD is low (Purcell et al., 2018; Josef Urban et al., 2017). Another interaction between climate variables that affect g_s simulation in MGAM is VPD-C_s interaction (Fig. 5). Many studies have shown that elevated C_s cause a reduction in g_s. However, for the higher C_s values increased VPD offsets this reduction (Flexas et al., 2004; Xu et al., 2016). The interaction of VPD-T_m and VPD-C_s is justified based on the literature review, as shown above. However, the plant physiological mechanisms are complex and require continuous datasets with a higher quantity and well-controlled environment that is hard to achieve (Josef Urban et al., 2017).

2.9.4 Towards a robust approach in C_s-g_s simulation

The effects of C_s and other climate variables' interactions on g_s and transpiration changes are still debated (Nadal-Sala et al., 2021). The semi-empirical and empirical g_s simulation models have different viewpoints in reflecting C_s effects on g_s. The Jarvis equation uses a linear function to present this relationship (Jarvis et al., 1976). Wang et al. (2005) produced a hyperbolic model to represent g_s response to C_s concentration. They found that the rate of decreasing g_s gradually lessened with C_s increase (Wang et al., 2005). Li et al. (2019) compared versions of the C_s-g_s relationship to find the best physiological and theoretical relationship. They used a combination of linear and hyperbolic equations as a modified-hyperbolic model to improve the accuracy and reliability of C_s-g_s estimation (Li et al., 2019). This study selected the modified-hyperbolic in the Jarvis equation in Noah-g_s-Eq.9 because it had better result than other C_s-g_s simulation approaches (Table 1 and S1).

It is worth noting that in several studies, the understanding of the plant response to C_s changes was through the assumption of keeping other variables at a fixed level (Massmann et al., 2019). This assumption is far from the real-world processes due to the interaction between the different variables. The MGAM-g_s model can reflect the combinational effects of key climate variables on g_s changes. Although this new approach highlights the interaction of VPD-C_s and VPD-T_m in g_s variation, more details should be linked to vegetation growth stages. However, due to the absence of a comprehensive and continuous dataset for a whole year, justification of the climate interactions by C_s through the whole growth period of a plant was not possible. Therefore, more studies on the MGAM approach are suggested for different climates and vegetation types.

2.10 Conclusion

The intercomparison of g_s models and their global sensitivity analysis showed the high sensitivity and dependency of semi-empirical g_s simulation models to parameterisation and A_n. This makes it difficult to extend these models to a global scale. To improve g_s simulation, the complex climate-vegetation interactions should be understood. The g_s simulation in empirical models considered climate variables effects on vegetation. However, their calibration process and complex plants' stress functions make it challenging to use them in new locations with different climates.

The introduced approach of MGAM-g_s captures important processes in real-world soil-atmosphere-vegetation interactions, while maintaining an appropriate level of parsimony to permit global-scale simulations without requiring ranges for fitted parameters by the MCMC-Bayesian calibration process. MGAM can represent the interaction of different key climate variables in g_s simulation, accomplished by global sensitivity analysis. This achievement improves our understanding of g_s simulation from individual indices level to understanding the g_s variation affected by indices interactions. A robust, nonlinear g_s simulation with MGAM-g_s highlights the effects of VPD-C_s and VPD-T_m interaction on g_s value. This new approach provides an alternative method for land surface modelling of transpiration simulation and water balance prediction. Further MGAM testing with comprehensive data for more vegetation types at the global scale and the full plant growth stage is required.

Notes

The authors declare no competing financial interest.

Acknowledgement

The support from the Australian Government Research Training Program Scholarship and Flinders University is acknowledged.

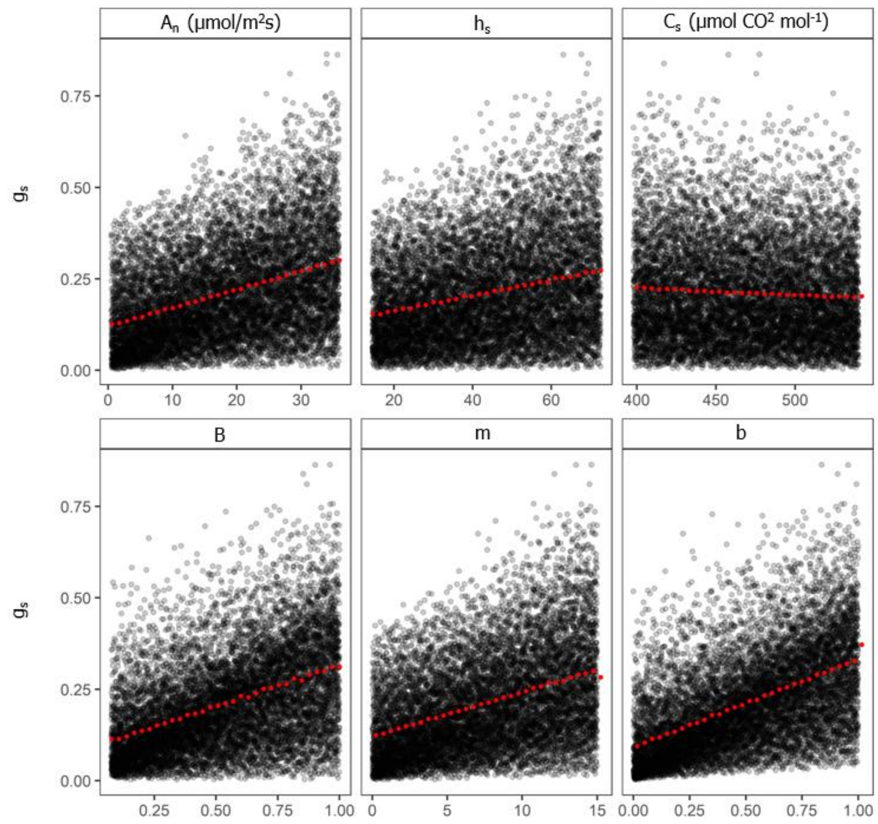
Supplementary Information

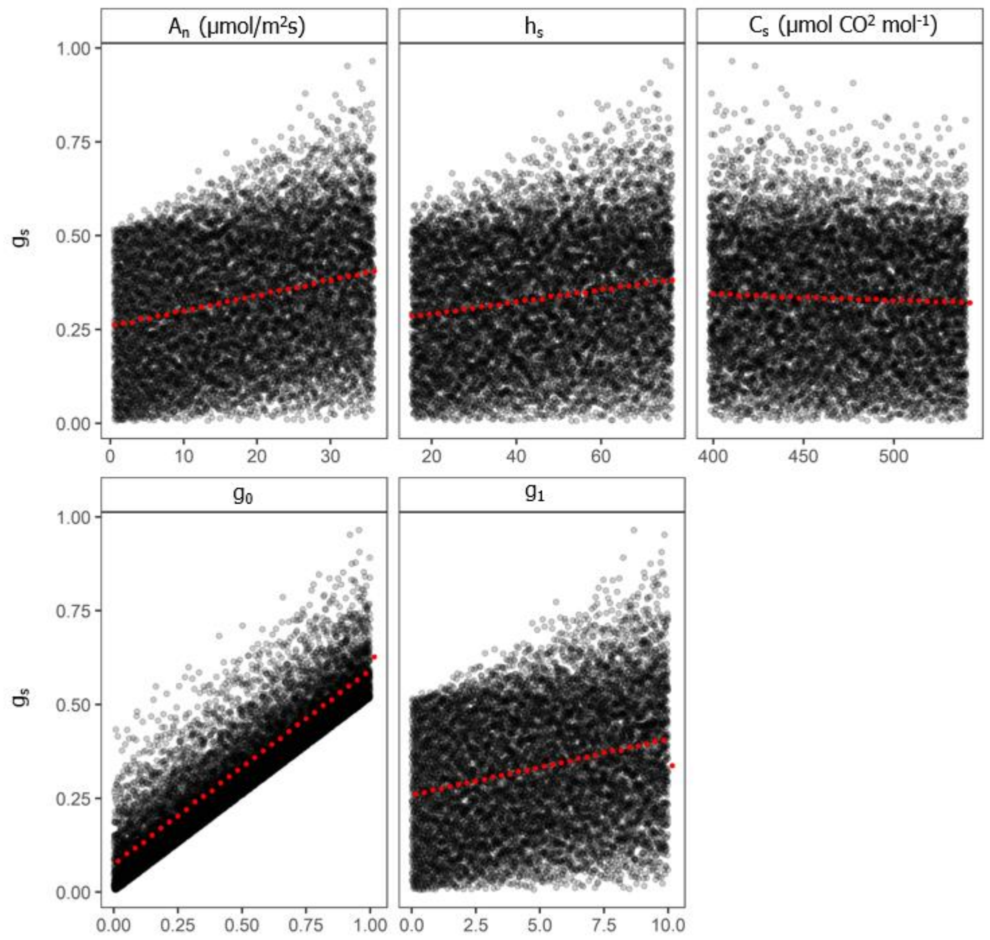
Table S1 Environmental stress functions in the Jarvis g_s model

Eq. #	Reference	g_s equation	Stress functions
S1	(Wang et al., 2014)	$g_s = g_{s_{max}} \times f(R) \times f(VPD) \times f(T_{air}) \times f(S)$	$f(R) = \frac{R}{R + a_1} \times \frac{R_{max} + a_1}{R_{max}}$ $f(VPD) = \exp(-a_2 \times VPD)$ $f(T) = 1 - a_3 \times (T_0 - T_{air})$ $f(S) = \frac{1}{1 + \left({S}/{S_m}\right)^{a_4}}$
S2	(Whitley et al., 2009)	$g_s = g_{s_{max}} \times f(R) \times f(VPD) \times f(S)$	$f(R) = \frac{R}{1000} \times \frac{1000 + a_1}{R + a_1}$ $f(VPD) = a_2 \times VPD \times \exp(-a_3 \times VPD)$ $f(S) = \begin{cases} 1 & S \geq S_{crit} \\ \frac{S - S_{wilt}}{S_{crit} - S_{wilt}} & S_{wilt} < S < S_{crit} \\ 0 & S \leq S_{wilt} \end{cases}$
S3	(Guyot et al., 2017)	$g_s = g_{s_{max}} \times f(R) \times f(VPD) \times f(S)$	$f(R) = \frac{R}{R + a_1}$ $f(VPD) = \exp\left(\frac{-a_2}{VPD + a_3} \times (VPD - VPD_{max})^2\right)$ $f(S) = \frac{1 + \exp(a_4 \times S)}{1 + \exp(-a_5 \times (S - S_{wilt}))}$
S4	(Whitley et al., 2013)	$g_s = g_{s_{max}} \times f(R) \times f(VPD) \times f(S)$	$f(R) = \frac{R}{1000} \times \frac{1000 + a_1}{R + a_1}$ $f(VPD) = \exp\left(\frac{-a_2}{VPD + a_3} \times (VPD - VPD_{max})^2\right)$ $f(S) = min\left\{1, \frac{S - S_{wilt}}{S_{crit} - S_{wilt}}\right\}$
S5	(García-Santos et al., 2009)	$g_s = g_{s_{max}} \times f(R) \times f(VPD)$	$f(R) = \frac{R}{1000} \times \frac{1000 + a_1}{R + a_1}$ $f(VPD) = \exp(-a_2 \times VPD)$
S6	(Harris et al., 2004), (Rodrigues et al., 2016)	$g_s = g_{s_{max}} \times f(R) \times f(VPD) \times f(T_{air}) \times f(S)$	$f(R) = \frac{R}{1000} \times \frac{1000 + a_1}{R + a_1}$ $f(VPD) = \exp(-k_2 \times VPD)$ $f(T) = \left[\frac{(T_{air} - T_0) \times (T_m - T_{air})}{(a_3 - T_0) \times (T_m - a_3)}\right]^{\tau}, \tau = \frac{(T_m - a_3)}{(a_3 - T_0)}$

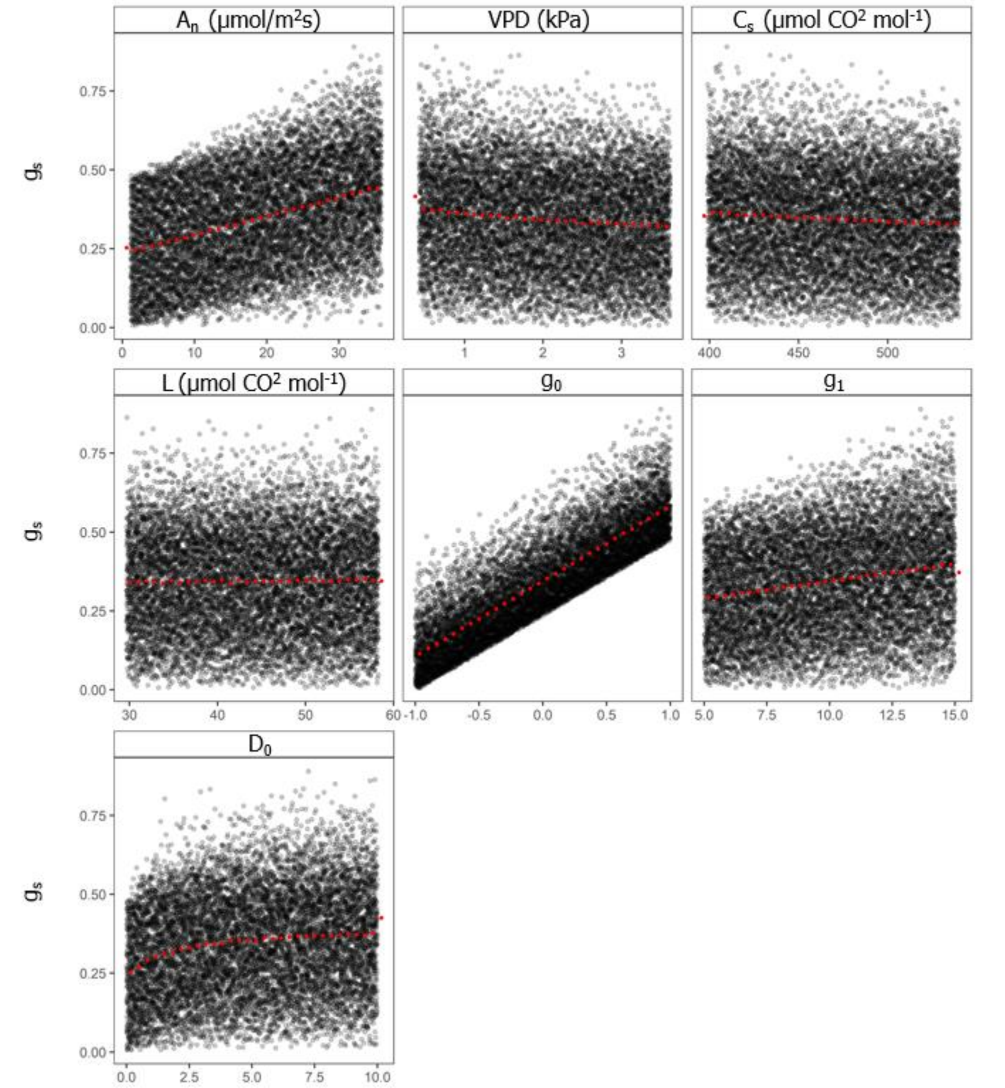
Eq. #	Reference	g_s equation	Stress functions
			$f(S) = \begin{cases} 1 & S \geq S_{crit} \\ \frac{S - S_{wilt}}{S_{crit} - S_{wilt}} & S_{wilt} < S < S_{crit} \\ 0 & S \leq S_{wilt} \end{cases}$
S7	(Granier & Loustau, 1994)	$g_s = g_{s_{max}} \times f(R) \times f(h_s) \times f(S)$	$f(R) = \mathrm{a}_1 \times \frac{R}{R + \mathrm{a}_2}$ $f(h_s) = \frac{1 - (\mathrm{a}_3 \times h_s)}{1 + (\mathrm{a}_4 \times h_s)}$ $f(S) = 1 - \mathrm{a}_5 \times \exp\left(\mathrm{a}_6 \times ((\frac{S_{max} - S}{S_{max} - S_{min}}))\right)$
S8	(Stewart, 1988)	$g_s = g_{s_{max}} \times f(R) \times f(h_s) \times f(T_{air}) \times f(S)$	$f(R) = \frac{R}{1000} \times \frac{1000 + \mathrm{a}_1}{R + \mathrm{a}_1}$ $f(h_s) = 1 - (\mathrm{a}_3 \times h_s)$ $f(T) = \left[\frac{(T_{air} - T_0) \times (T_m - T_{air})}{(\mathrm{a}_3 - T_0) \times (T_m - \mathrm{a}_3)}\right]^{\tau}, \tau = \frac{(T_m - \mathrm{a}_3)}{(\mathrm{a}_3 - T_0)}$ $f(S) = 1 - k_6 \times \mathrm{S}$
S9	(Lhomme et al., 1998)	$g_s = g_{s_{max}} \times f(R) \times f(T_{air}) \times f(VPD)$	$f(R) = \frac{(1 + 0.001 \times \mathrm{a}_1) \times S}{\mathrm{a}_1 + R}$ $f(T) = 1 - \mathrm{a}_2 \times (24.8 - Tair)^2$ $f(VPD) = 1 - \mathrm{a}_3 \times VPD$
S10	(Sommer et al., 2002)	$g_s = g_{s_{max}} \times f(R) \times f(T_{air}) \times f(VPD)$	$f(R) = \frac{R}{1000} \times \frac{1000 + \mathrm{a}_1}{R + \mathrm{a}_1}$ $f(T) = \left[\frac{(T_{air} - T_0) \times (T_m - T_{air})}{(\mathrm{a}_3 - T_0) \times (T_m - \mathrm{a}_3)}\right]^{\tau}, \tau = \frac{(T_m - \mathrm{a}_3)}{(\mathrm{a}_3 - T_0)}$ $f(VPD) = \exp(-\mathrm{a}_2 \times VPD)$
S11	(Li et al., 2019)	$g_s = g_{s_{max}} \times f(R) \times f(T_{air}) \times f(VPD) \times f(S) \times f(C_s)$	$f(R) = \frac{R}{1000} \times \frac{1000 + \mathrm{a}_1}{R + \mathrm{a}_1}$ $f(T) = 1 - \mathrm{a}_2 \times (25 - T_{air})^2$ $f(VPD) = 1 - \mathrm{a}_3 \times VPD$ $f(S) = \begin{cases} 1 & S \geq S_{crit} \\ \frac{S - S_{wilt}}{S_{crit} - S_{wilt}} & S_{wilt} < S < S_{crit} \\ 0 & S \leq S_{wilt} \end{cases}$ $f(C_S) = \frac{1}{1 + \mathrm{a}_4 \times (\frac{C_S}{\mathrm{a}_5} - 1)}$
S12	(Kumar et al., 2011)	$g_s = g_{s_{max}} \times f(R) \times f(VPD) \times f(T_{air}) \times f(S)$	$f(R) = \frac{R}{1000} \times \frac{1000 + \mathrm{a}_1}{R + \mathrm{a}_1}$ $f(T) = 1 - \mathrm{a}_2 \times (25 - T_{air})^2$

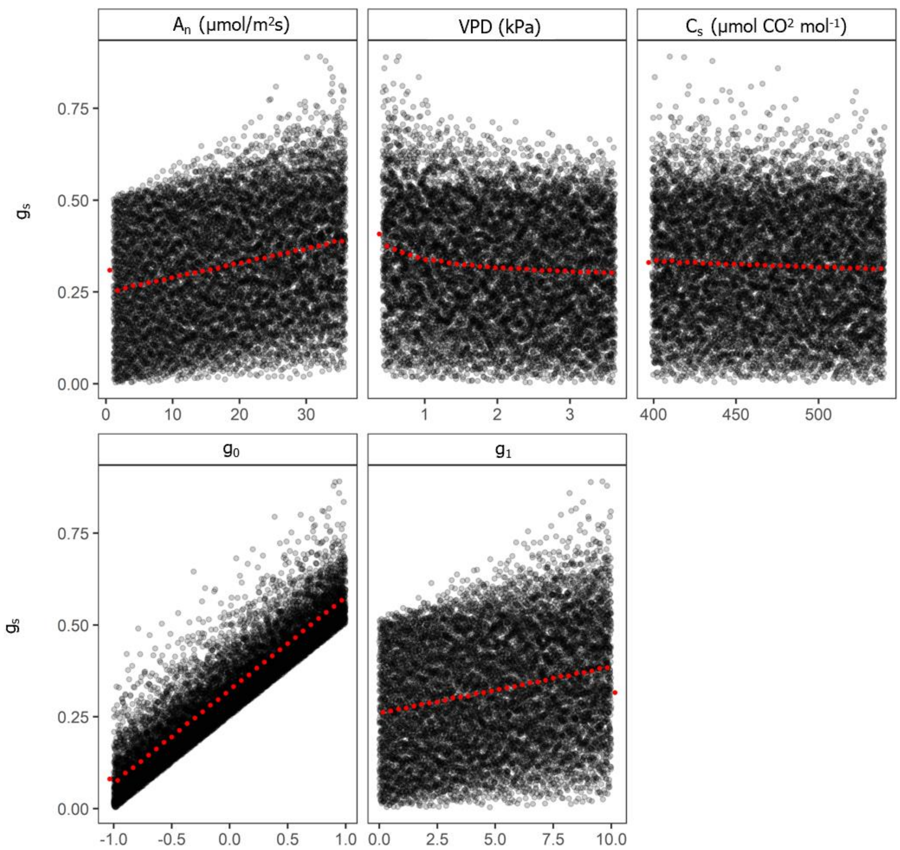
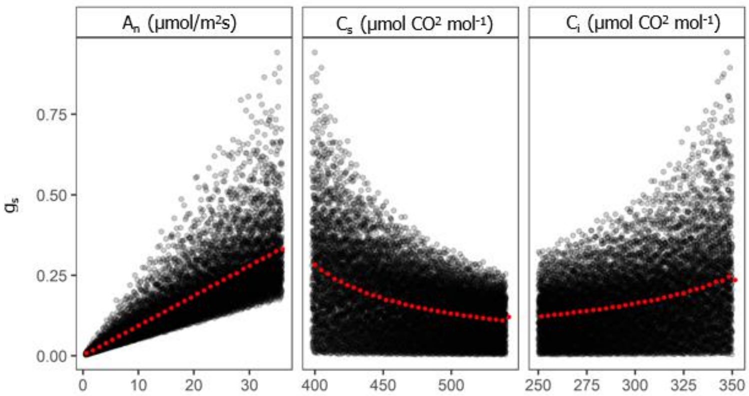
Eq. #	Reference	g_s equation	Stress functions
			$f(VPD) = 1 - \mathrm{a}_3 \times VPD$ $f(S) = \begin{cases} 1 & S \geq S_{crit} \\ \dfrac{S - S_{wilt}}{S_{crit} - S_{wilt}} & S_{wilt} < S < S_{crit} \\ 0 & S \leq S_{wilt} \end{cases}$

CLM4.5-g_s-Eq.1	

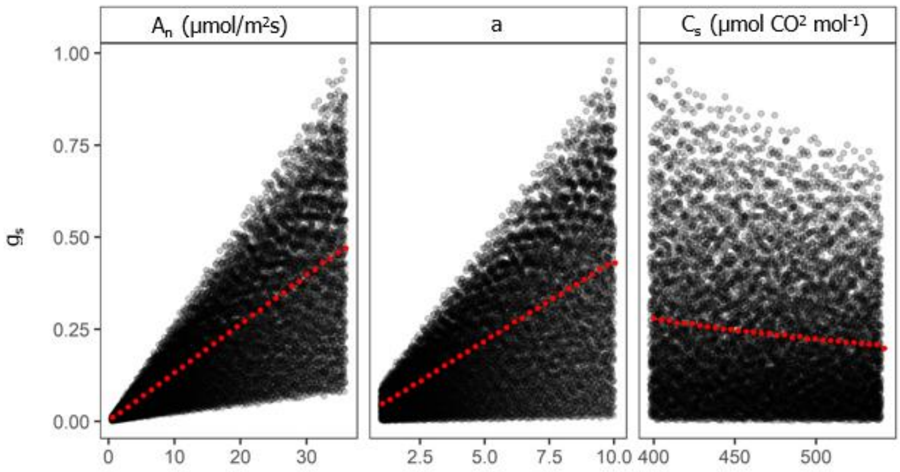
CLM5-g_s-Eq.2	

A_n (μmol/m²s)
h_s
C_s (μmol CO² mol⁻¹)
g_0
g_1
g_s
1.00
0.75
0.50
0.25
0.00
0 10 20 30
20 40 60
400 450 500
0.00 0.25 0.50 0.75 1.00
0.0 2.5 5.0 7.5 10.0

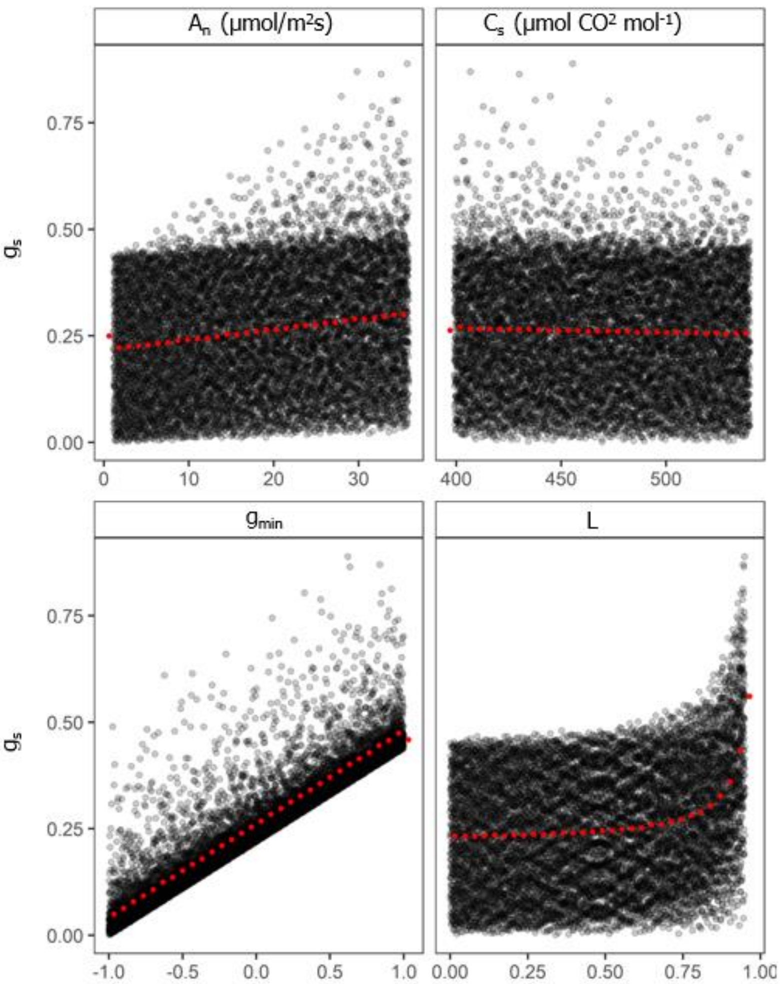
CLM5-g_s-Eq.3	

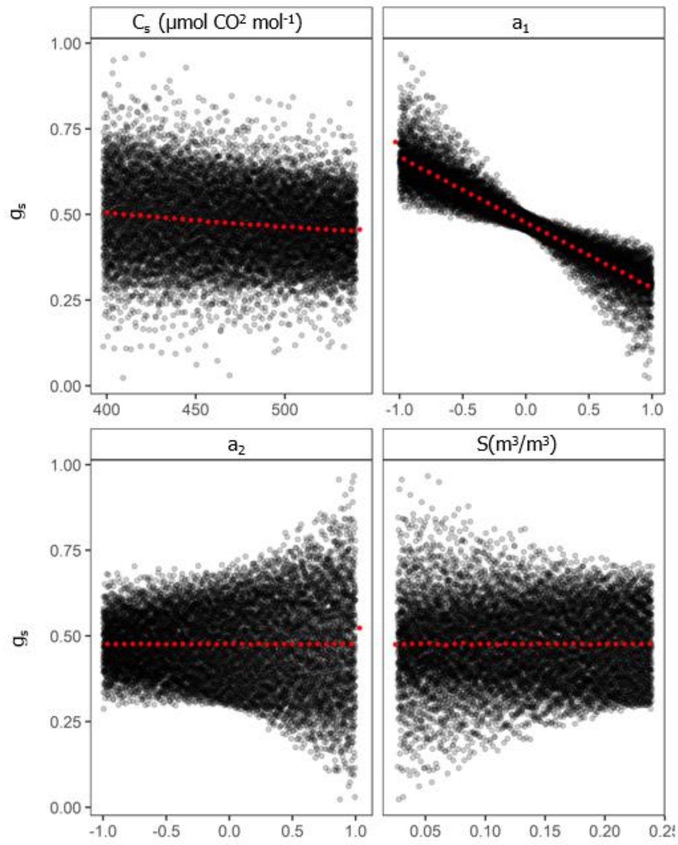
CLM5-g_s-Eq.4	
JULES-g_s-Eq.5	

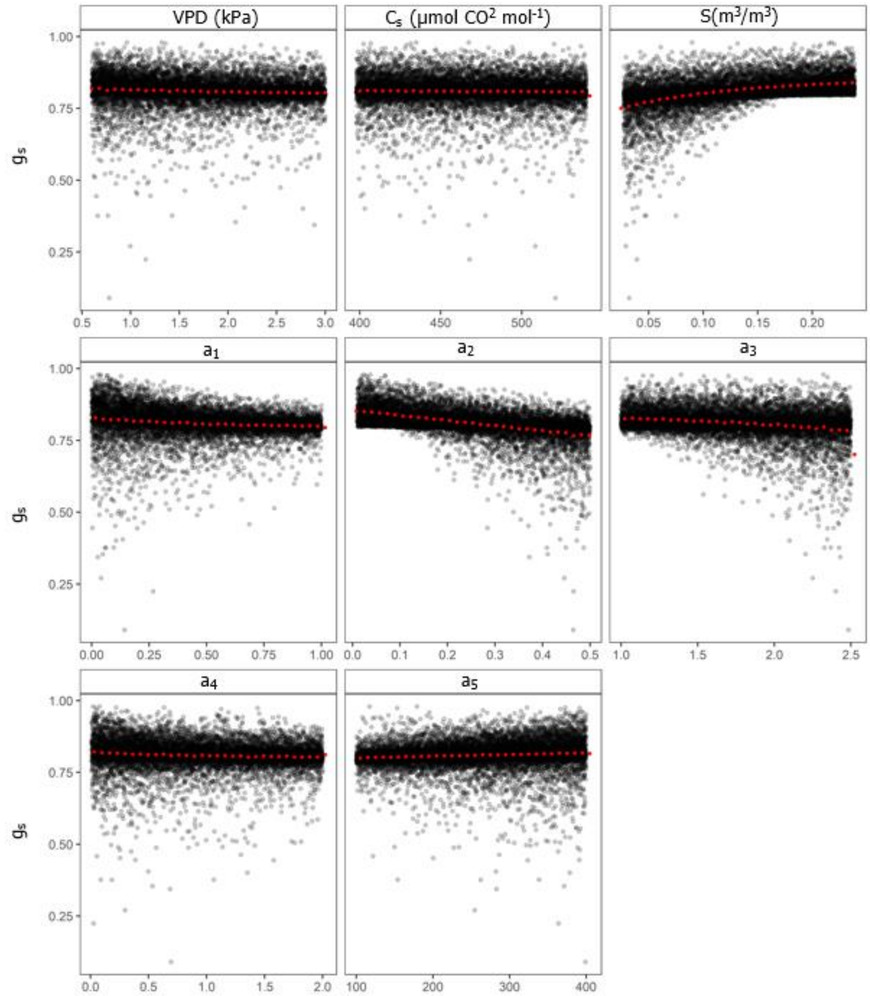
JULES-g_s-Eq.6



LPJml-g_s-Eq.7



JSBACH-g_s-Eq.8	

Noah-g_s-Eq.9	VPD (kPa) C_s (µmol CO^2 mol^{-1}) S(m^3/m^3) a_1 a_2 a_3 a_4 a_5 g_s

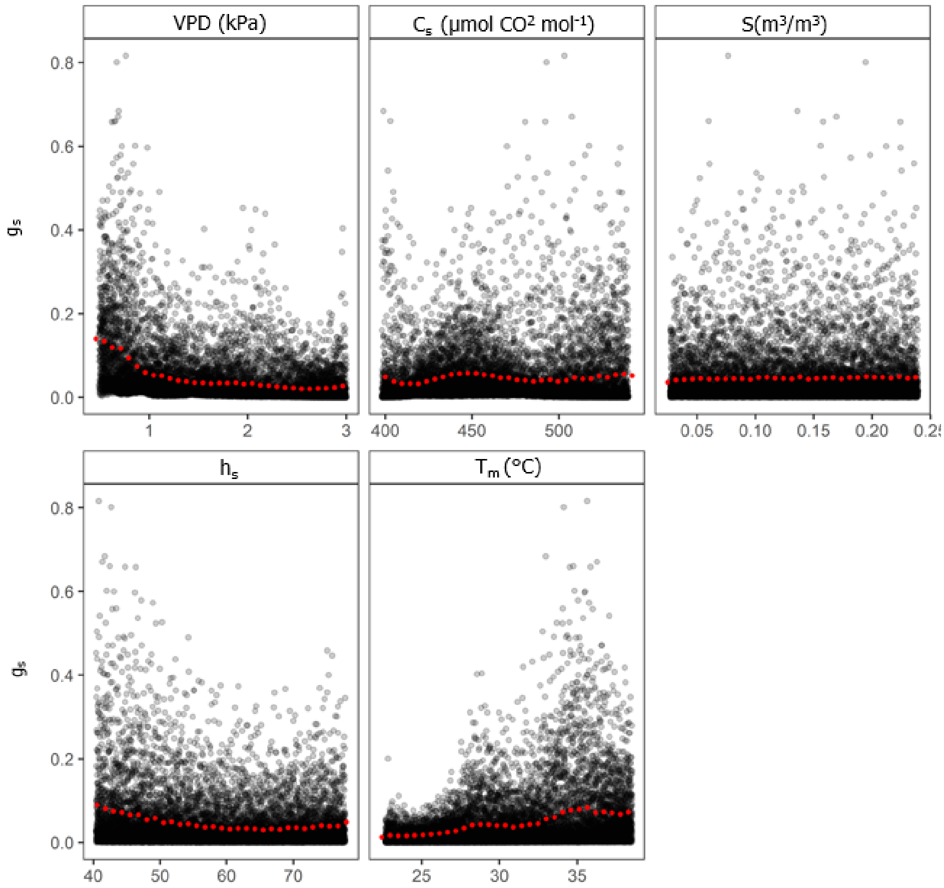
MGAM-g_s-Eq.10	

Figure S1 The variability of normalised g_s for each index range in g_s simulation models. The data points are perturbed each index by global sensitivity analysis and the red dots are the average of g_s.

2.11 References

Ahmed, M., Stöckle, C. O., Nelson, R., Higgins, S., Ahmad, S., & Raza, M. A. (2019). Novel multimodel ensemble approach to evaluate the sole effect of elevated CO_2 on winter wheat productivity. *Scientific Reports*, *9*(1), 7813. https://doi.org/https://doi.org/10.1038/s41598-019-44251-x

Ameye, M., Wertin, T. M., Bauweraerts, I., McGuire, M. A., Teskey, R. O., & Steppe, K. (2012). The effect of induced heat waves on Pinus taeda and Quercus rubra seedlings in ambient and elevated CO2 atmospheres. *New Phytologist*, *196*(2), 448-461. https://doi.org/https://doi.org/10.1111/j.1469-8137.2012.04267.x

Arneth, A., Lloyd, J., Santruckova, H., Bird, M., Grigoryev, S., Kalaschnikov, Y. N., Gleixner, G., & Schulze, E.-D. (2002). Response of central Siberian Scots pine to soil water deficit and long-term trends in atmospheric CO2 concentration. *Global Biogeochemical Cycles*, *16*(1), 5-1. https://doi.org/https://doi.org/10.1029/2000GB001374

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., . . . Ziehn, T. (2020). Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, *17*(16), 4173-4222. https://doi.org/https://doi.org/10.5194/bg-17-4173-2020

Ball, J. T., Woodrow, I. E., Berry, J. A. . (1987). A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. *In Progress in photosynthesis research*, 221-224. https://doi.org/https://doi.org/10.1007/978-94-017-0519-6_48

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.-P., & Delignette-Muller, M.-L. (2015). A Toolbox for Nonlinear Regression in R: The Package nlstools. *Journal of Statistical Software*, *66*(5), 1 - 21. https://doi.org/https://doi.org/10.18637/jss.v066.i05

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J. J., Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. A., Stevens, L., & Vuichard, N. (2015). The Plumbing of Land Surface Models: Benchmarking Model Performance. *Journal of Hydrometeorology*, *16*(3), 1425-1442. https://doi.org/https://doi.org/10.1175/JHM-D-14-0158.1

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R., Ménard, C. B., ... , & Harding, R. J. (2011). The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677-699. https://doi.org/https://doi.org/10.5194/gmd-4-677-2011

Blyth, E. M., Arora, V. K., Clark, D. B., Dadson, S. J., De Kauwe, M. G., Lawrence, D. M., Melton, J. R., Pongratz, J., Turton, R. H., Yoshimura, K., & Yuan, H. (2021). Advances in Land Surface Modelling. *Current Climate Change Reports*, *7*(2), 45-71. https://doi.org/https://doi.org/10.1007/s40641-021-00171-5

Brooks, A., & Farquhar, G. D. (1985). Effect of temperature on the CO2/O2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light. *Planta*, *165*(3), 397-406. https://doi.org/https://doi.org/10.1007/BF00392238

Ceppi, P., & Gregory, J. M. (2017). Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proceedings of the National Academy of Sciences*, *114*(50), 13126-13131. https://doi.org/https://doi.org/10.1073/pnas.1714308114

Collatz GJ, Ribas-Carbo M, & JA, B. (1992). Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants. *Functional Plant Biology*, *19*(5), 519-538.

Collatz, G. J., Ball, J. T., Grivet, C., & Berry, J. A. (1991). Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer. *Agricultural and Forest Meteorology*, *54*(2), 107-136. https://doi.org/https://doi.org/10.1016/0168-1923(91)90002-8

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., & Smith, J. (1999). The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynamics*, *15*(3), 183-203. https://doi.org/http://dx.doi.org/10.1007/s003820050276

Cox, P. M., Huntingford, C., & Harding, R. J. (1998). A canopy conductance and photosynthesis model for use in a GCM land surface scheme. *Journal of Hydrology*, *212-213*, 79-94. https://doi.org/https://doi.org/10.1016/S0022-1694(98)00203-0

Creese, C., Oberbauer, S., Rundel, P., & Sack, L. (2014). Are fern stomatal responses to different stimuli coordinated? Testing responses to light, vapor pressure deficit, and CO2 for diverse species grown under contrasting irradiances. *New Phytol*, *204*(1), 92-104. https://doi.org/ https://doi.org/10.1111/nph.12922

Crous, K. Y., Ósvaldsson, A., & Ellsworth, D. S. (2015). Is phosphorus limiting in a mature Eucalyptus woodland? Phosphorus fertilisation stimulates stem growth. *Plant and Soil*, *391*(1), 293-305. https://doi.org/https://doi.org/10.1007/s11104-015-2426-4

Damour, G., Simonneau, T., Cochard, H., & Urban, L. (2010). An overview of models of stomatal conductance at the leaf level. *Plant, cell & environment*, *33*(9), 1419-1438. https://doi.org/https://doi.org/10.1111/j.1365-3040.2010.02181.x

De Kauwe, M. G., Medlyn, B. E., & Tissue, D. T. (2021). To what extent can rising [CO2] ameliorate plant drought stress? *New Phytologist.*, *231*(6), 2118-2124. https://doi.org/https://doi.org/10.1111/nph.17540

Donohue, R. J., Roderick, M. L., McVicar, T. R., & Yang, Y. (2017). A simple hypothesis of how leaf and canopy-level transpiration and assimilation respond to elevated CO2 reveals distinct response patterns between disturbed and undisturbed vegetation. *Journal of Geophysical Research: Biogeosciences*, *122*(1), 168-184. https://doi.org/https://doi.org/10.1002/2016JG003505

Drake, J. E., Tjoelker, M. G., Vårhammar, A., Medlyn, Belinda E., Reich, P. B., Leigh, A., Pfautsch, S., Blackman, C. J., López, R., Aspinwall, M. J., Crous, K. Y., Duursma, R. A., Kumarathunge, D., De Kauwe, M. G., Jiang, M., Nicotra, A. B., Tissue, D. T., Choat, B., Atkin, O. K., & Barton, C. V. M. (2018). Trees tolerate an extreme heatwave via sustained transpirational cooling and increased leaf thermal tolerance. *Global Change Biology*, *24*(6), 2390-2402. https://doi.org/https://doi.org/10.1111/gcb.14037

Duursma, R. (2015). *Photosynthetically active radiation, canopy gap fraction, and leaf area index at the EucFACE* Western Sydney University. https://doi.org/https://doi.org/10.4225/35/563159f223739

Duursma, R. A., Gimeno, T. E., Boer, M. M., Crous, K. Y., Tjoelker, M. G., & Ellsworth, D. S. (2016). Canopy leaf area of a mature evergreen Eucalyptus woodland does not respond to elevated atmospheric [CO2] but tracks water availability. *Global Change Biology*, *22*(4), 1666-1676. https://doi.org/https://doi.org/10.1111/gcb.13151

Faralli, M., Matthews, J., & Lawson, T. (2019). Exploiting natural variation and genetic manipulation of stomatal conductance for crop improvement. *Current Opinion in Plant Biology*, *49*, 1-7. https://doi.org/https://doi.org/10.1016/j.pbi.2019.01.003

Farquhar, G. D. (1989). "Models of integrated photosynthesis of cells and leaves." Philosophical Transactions of the Royal Society of London. *Biological Sciences*, *323*(1216), 357-367. https://doi.org/https://doi.org/10.1098/rstb.1989.0016

Flexas, J., Bota, J., Loreto, F., Cornic, G., & Sharkey, T. D. (2004). Diffusive and Metabolic Limitations to Photosynthesis under Drought and Salinity in C3 Plants. *Plant Biology*, *6*(3), 269-279. https://doi.org/https://doi.org/10.1055/s-2004-820867

Franks, P. J., Berry, J. A., Lombardozzi, D. L., & Bonan, G. B. (2017). Stomatal Function across Temporal and Spatial Scales: Deep-Time Trends, Land-Atmosphere Coupling and Global Models *Plant physiology*, *174*(2), 583-602. https://doi.org/https://doi.org/10.1104/pp.17.00287

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, A., Arora, V., Bates, N. R., . . . Zaehle, S. (2020). Global Carbon Budget 2020. *Earth System Science Data*, *12*(4), 3269-3340. https://doi.org/https://doi.org/10.5194/essd-12-3269-2020

García-Santos, G., Bruijnzeel, L. A., & Dolman, A. J. (2009). Modelling canopy conductance under wet and dry conditions in a subtropical cloud forest. *Agricultural and Forest Meteorology*, *149*(10), 1565-1572. https://doi.org/https://doi.org/10.1016/j.agrformet.2009.03.008

Geyer, C. J., & Johnson, L. T. (2020). *mcmc: Markov Chain Monte Carlo. R package version 0.9-7*. https://CRAN.R-project.org/package=mcmc

Gimeno, T. E., Crous, K. Y., Cooke, J., O'Grady, A. P., Ósvaldsson, A., Medlyn, B. E., & Ellsworth, D. S. (2016). Conserved stomatal behaviour under elevated CO2 and varying water availability in a mature woodland. *Functional Ecology*, *30*(5), 700-709. https://doi.org/https://doi.org/10.1111/1365-2435.12532

Granier, A., & Loustau, D. (1994). Measuring and modelling the transpiration of a maritime pine canopy from sap-flow data. *Agricultural and Forest Meteorology*, *71*(1), 61-81. https://doi.org/https://doi.org/10.1016/0168-1923(94)90100-7

Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., & Gentine, P. (2019). Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature*, *565*(7740), 476-479. https://doi.org/https://doi.org/10.1038/s41586-018-0848-x

Guyot, A., Fan, J., Oestergaard, K. T., Whitley, R., Gibbes, B., Arsac, M., & Lockington, D. A. (2017). Soil-water content characterisation in a modified Jarvis-Stewart model: A case study of a conifer forest on a shallow unconfined aquifer. *Journal of Hydrology*, *544*, 242-253. https://doi.org/https://doi.org/10.1016/j.jhydrol.2016.11.041

Harris, P. P., Huntingford, C., Cox, P. M., Gash, J. H. C., & Malhi, Y. (2004). Effect of soil moisture on canopy conductance of Amazonian rainforest. *Agricultural and Forest Meteorology*, *122*(3), 215-227. https://doi.org/https://doi.org/10.1016/j.agrformet.2003.09.006

Hartig, F., Dyke, J., Hickler, T., Higgins, S. I., O'Hara, R. B., Scheiter, S., & Huth, A. (2012). Connecting dynamic vegetation models to data — an inverse perspective. *Journal of Biogeography*, *39*(12), 2240-2252. https://doi.org/https://doi.org/10.1111/j.1365-2699.2012.02745.x

Hartig, F., Minunno, F., & Paul, S. (2019). *BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics. R package version 0.1.7.* In https://CRAN.R-project.org/package=BayesianTools

Hastie, T., Robert, T., & JH, F. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). New York: springer.

Haworth, M., Elliott-Kingston, C., & McElwain, J. C. (2013). Co-ordination of physiological and morphological responses of stomata to elevated [CO2] in vascular plants. *Oecologia*, *171*(1), 71-82. https://doi.org/https://doi.org/10.1007/s00442-012-2406-9

Haxeltine, A., & Prentice, I. C. (1996). BIOME3: An equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability, and competition among plant functional types. *Global Biogeochemical Cycles*, *10*(4), 693-709. https://doi.org/https://doi.org/10.1029/96GB02344

Inoue, T., Sunaga, M., Ito, M., Yuchen, Q., Matsushima, Y., Sakoda, K., & Yamori, W. (2021). Minimizing VPD Fluctuations Maintains Higher Stomatal Conductance and Photosynthesis, Resulting in Improvement of Plant Growth in Lettuce [Original Research]. *Frontiers in Plant Science*, *12*(458). https://doi.org/https://doi.org/10.3389/fpls.2021.646144

Jaiswal, R. K., Ali, S., & Bharti, B. (2020). Comparative evaluation of conceptual and physical rainfall–runoff models. *Applied Water Science*, *10*(1), 48. https://doi.org/https://doi.org/10.1007/s13201-019-1122-6

Jarvis, P. G., Massheder, J. M., Hale, S. E., Moncrieff, J. B., Rayment, M., & Scott, S. L. (1997). Seasonal variation of carbon dioxide, water vapor, and energy exchanges of a boreal black spruce forest. *Journal of Geophysical Research*, *102*(D24), 28953-28966. https://doi.org/https://doi.org/10.1029/97JD01176

Jarvis, P. G., Monteith, J. L., & Weatherley, P. E. (1976). The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Biological Sciences*, *273*(927), 593-610. https://doi.org/https://doi.org/10.1098/rstb.1976.0035

Jiao, X.-C., Song, X.-M., Zhang, D.-L., Du, Q.-J., & Li, J.-M. (2019). Coordination between vapor pressure deficit and CO2 on the regulation of photosynthesis and productivity in greenhouse

tomato production. *Scientific Reports*, *9*(1), 8700. https://doi.org/https://doi.org/10.1038/s41598-019-45232-w
Jiménez, C., Prigent, C., Mueller, B., Seneviratne, S. I., McCabe, M. F., Wood, E. F., Rossow, W. B., Balsamo, G., Betts, A. K., Dirmeyer, P. A., Fisher, J. B., Jung, M., Kanamitsu, M., Reichle, R. H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K., & Wang, K. (2011). Global intercomparison of 12 land surface heat flux estimates. *Journal of Geophysical Research*, *116*(D2). https://doi.org/https://doi.org/10.1029/2010JD014545
Keenan, T. F., Hollinger, D. Y., Bohrer, G., Dragoni, D., Munger, J. W., Schmid, H. P., & Richardson, A. D. (2013). Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature*, *499*(7458), 324-327. https://doi.org/https://doi.org/10.1038/nature12291
Kimm, H., Guan, K., Gentine, P., Wu, J., Bernacchi, C. J., Sulman, B. N., Griffis, T. J., & Lin, C. (2020). Redefining droughts for the U.S. Corn Belt: The dominant role of atmospheric vapor pressure deficit over soil moisture in regulating stomatal behavior of Maize and Soybean. *Agricultural and Forest Meteorology*, *287*, 107930. https://doi.org/https://doi.org/10.1016/j.agrformet.2020.107930
Knauer, J., Werner, C., & Zaehle, S. (2015). Evaluating stomatal models and their atmospheric drought response in a land surface scheme: A multibiome analysis. *Journal of Geophysical Research: Biogeosciences*, *120*(10), 1894-1911. https://doi.org/https://doi.org/10.1002/2015JG003114
Knauer, J., Zaehle, S., Reichstein, M., Medlyn, B. E., Forkel, M., Hagemann, S., & Werner, C. (2017). The response of ecosystem water-use efficiency to rising atmospheric CO2 concentrations: sensitivity and large-scale biogeochemical implications. *New Phytologist.*, *213*(4), 1654-1666. https://doi.org/https://doi.org/10.1111/nph.14288
Knorr, W. (2000). Annual and Internannual CO2 Exchanges of the Terrestrial Biosphere: Process-Based Simulations and Uncertainties. *Global Ecology and Biogeography*, *9*(3), 225-252. https://doi.org/https://doi.org/10.1046/j.1365-2699.2000.00159.x
Konings, A. G., Williams, A. P., & Gentine, P. (2017). Sensitivity of grassland productivity to aridity controlled by stomatal and xylem regulation. *Nature Geoscience*, *10*(4), 284-288. https://doi.org/http://dx.doi.org/10.1038/ngeo2903
Krich, C., Mahecha, M. D., Migliavacca, M., De Kauwe, M. G., Griebel, A., Runge, J., & Miralles, D. G. (2022). Decoupling between ecosystem photosynthesis and transpiration: a last resort against overheating. *Environmental Research Letters*, *17*(4), 044013. https://doi.org/10.1088/1748-9326/ac583e
Kuhn, M. (2021). *caret: Classification and Regression Training. R package version 6.0-88*. https://CRAN.R-project.org/package=caret
Kumar, A., Chen, F., Niyogi, D., Alfieri, J. G., Ek, M. B., & Mitchell, K. E. (2011). Evaluation of a Photosynthesis-Based Canopy Resistance Formulation in the Noah Land-Surface Model. *Boundary-Layer Meteorology*, *138*, 263-284. https://doi.org/https://doi.org/10.1007/s10546-010-9559-z
Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., Vertenstein, M., Andre, B. B., G. Ghimire, B. Kampenhout, LV. Kennedy, D. Kluzek, E. Knox, R., & Lawrence, P. L., F. Li, H. Lombardozzi, D. Lu, Y. Perket, J. Riley, W. Sacks, W. Shi, M. Wieder, W. Xu, C. (2020). *CLM5 Documentation*.
Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., . . . Zeng, X. (2019). The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4245-4287. https://doi.org/https://doi.org/10.1029/2018MS001583
Lei, H., Yang, D., & Huang, M. (2014). Impacts of climate change and vegetation dynamics on runoff in the mountainous region of the Haihe River basin in the past five decades. *Journal of Hydrology*, *511*, 786-799. https://doi.org/https://doi.org/10.1016/j.jhydrol.2014.02.029
Leuning, R. (1990). Modelling Stomatal Behaviour and Photosynthesis of Eucalyptus grandis. *Functional Plant Biology*, *17*(2), 159-175. https://doi.org/https://doi.org/10.1071/PP9900159

Leuning, R. (1995). A critical appraisal of a combined stomatal-photosynthesis model for C3 plants. *Plant, cell & environment*, *18*(4), 339-355. https://doi.org/https://doi.org/10.1111/j.1365-3040.1995.tb00370.x

Leuzinger, S., Korner, C. (2007). Water savings in mature deciduous forest trees under elevated CO2. *Global Change Biology*, *13*(12), 2498-2508. https://doi.org/https://doi.org/10.1111/j.1365-2486.2007.01467.x

Lhomme, J.-P., Elguero, E., Chehbouni, A., & Boulet, G. (1998). Stomatal control of transpiration: Examination of Monteith's Formulation of canopy resistance. *Water Resources Research, 34,* 2301. https://doi.org/https://doi.org/10.1029/98WR01339

Li, D., Fang, K., Li, Y., Chen, D., Liu, X., Dong, Z., Zhou, F., Guo, G., Shi, F., Xu, C., & Li, Y. (2017). Climate, intrinsic water-use efficiency and tree growth over the past 150 years in humid subtropical China. *PLoS One*, *12*(2). https://doi.org/https://doi.org/10.1371/journal.pone.0172045

Li, X., Kang, S., Niu, J., Huo, Z., & Liu, J. (2019). Improving the representation of stomatal responses to CO2 within the Penman–Monteith model to better estimate evapotranspiration responses to climate change. *Journal of Hydrology*, *572*, 692-705. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.03.029

Liao, D., Niu, J., Kang, S., Singh, S. K., & Du, T. (2021). Effects of elevated CO2 on the evapotranspiration over the agricultural land in Northwest China. *Journal of Hydrology*, *593*, 125858. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125858

Lin, J., Jach, M. E., & Ceulemans, R. (2001). Stomatal density and needle anatomy of Scots pine (Pinus sylvestris) are affected by elevated CO2. *New Phytologist*, *150*, 665-674. https://doi.org/https://doi.org/10.1046/j.1469-8137.2001.00124.x

Liu, D., & Mishra, A. K. (2017). Performance of AMSR_E soil moisture data assimilation in CLM4.5 model for monitoring hydrologic fluxes at global scale. *Journal of Hydrology*, *547*, 67-79. https://doi.org/https://doi.org/10.1016/j.jhydrol.2017.01.036

Liu, N., Buckley, T. N., He, X., Zhang, X., Zhang, C., Luo, Z., Wang, H., Sterling, N., & Guan, H. (2019). Improvement of a simplified process based model for estimating transpiration under wate' limited conditions. *Hydrological Processes*, *33*, 1670 - 1685. https://doi.org/https://doi.org/10.1002/hyp.13430

Massmann, A., Gentine, P., & Lin, C. (2019). When Does Vapor Pressure Deficit Drive or Reduce Evapotranspiration? *Journal of Advances in Modeling Earth Systems*, *11*(10), 3305-3320. https://doi.org/https://doi.org/10.1029/2019MS001790

Mastrotheodoros, T., Pappas, C., Molnar, P., Burlando, P., Keenan, T. F., Gentine, P., Gough, C. M., & Fatichi, S. (2017). Linking plant functional trait plasticity and the large increase in forest water use efficiency. *Journal of Geophysical Research: Biogeosciences*, *122*(9), 2393-2408. https://doi.org/https://doi.org/10.1002/2017JG003890

Mathias, J. M., & Thomas, R. B. (2021). Global tree intrinsic water use efficiency is enhanced by increased atmospheric CO2 and modulated by climate and plant functional types. *Proceedings of the National Academy of Sciences*, *118*(7). https://doi.org/https://doi.org/10.1073/pnas.2014286118

Medlyn, B. E., Barton, C. V. M., Broadmeadow, M. S. J., Ceulemans, R., De Angelis, P., Forstreuter, M., Freeman, M., Jackson, S. B., Kellomäki, S., Laitat, E., Rey, A., Roberntz, P., Sigurdsson, B. D., Strassemeyer, J., Wang, K., Curtis, P. S., & Jarvis, P. G. (2001). Stomatal conductance of forest species after long-term exposure to elevated CO(2) concentration: a synthesis. *New Phytol*, *149*(2), 247-264. https://doi.org/https://doi.org/10.1046/j.1469-8137.2001.00028.x

Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V. M., Crous, K. Y., De Angelis, P., Freeman, M., & Wingate, L. (2011). Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology*, *17*(6), 2134-2144. https://doi.org/https://doi.org/10.1111/j.1365-2486.2010.02375.x

Morgan, J. A., Pataki, D. E., Körner, C., Clark, H., Del Grosso, S. J., Grünzweig, J. M., Knapp, A. K., Mosier, A. R., Newton, P. C., Niklaus, P. A., Nippert, J. B., Nowak, R. S., Parton, W. J., Polley, H. W., & Shaw, M. R. (2004). Water relations in grassland and desert ecosystems exposed to elevated atmospheric CO2. *Oecologia*, *140*(1), 11-25. https://doi.org/https://doi.org/10.1007/s00442-004-1550-2

Nadal-Sala, D., Medlyn, B. E., Ruehr, N. K., Barton, C. V. M., Ellsworth, D. S., Gracia, C., Tissue, D. T., Tjoelker, M. G., & Sabaté, S. (2021). Increasing aridity will not offset CO2 fertilization in fast-growing eucalypts with access to deep soil water. *Global Change Biology*, *27*(12), 2970-2990. https://doi.org/https://doi.org/10.1111/gcb.15590

Norby, R. J., & Zak, D. R. (2011). Ecological Lessons from Free-Air CO2 Enrichment (FACE) Experiments. *Annual Review of Ecology, Evolution, and Systematics*, *42*(1), 181-203. https://doi.org/10.1146/annurev-ecolsys-102209-144647

Oleson, K., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S., Thornton, P. E., Bozbiyik, A., Fisher, R., Heald, C. L., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., . . . Yang, Z.-L. (2013). *Technical description of version 4.5 of the Community Land Model (CLM)* (National Center for Atmospheric Research (NCAR), Issue.

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Ottlé, C., Poulter, B., Zaehle, S., & Running, S. W. (2020). Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. *Hydrol. Earth Syst. Sci.*, *24*(3), 1485-1509. https://doi.org/https://doi.org/10.5194/hess-24-1485-2020

Pan, S., Tian, H., Dangal, S. R. S., Yang, Q., Yang, J., Lu, C., Tao, B., Ren, W., & Ouyang, Z. (2015). Responses of global terrestrial evapotranspiration to climate change and increasing atmospheric CO2 in the 21st century. *Earth's Future*, *3*(1), 15-35. https://doi.org/https://doi.org/10.1002/2014EF000263

Purcell, C., Batke, S. P., Yiotis, C., Caballero, R., Soh, W. K., Murray, M., & McElwain, J. C. (2018). Increasing stomatal conductance in response to rising atmospheric CO2. *Annals of Botany*, *121*(6), 1137-1149. https://doi.org/https://doi.org/10.1093/aob/mcy023

Puy, A., Piano, S. L., Saltelli, A., & Levin, S. A. . (2021). *sensobol: an R package to compute variance-based sensitivity indices*. http://www.arnaldpuy.com

Radin, J. W., Hartung, W., Kimball, B. A., & Mauney, J. R. (1988). Correlation of Stomatal Conductance with Photosynthetic Capacity of Cotton Only in a CO(2)-Enriched Atmosphere: Mediation by Abscisic Acid? *Plant physiology*, *88*(4), 1058-1062. https://doi.org/https://doi.org/10.1104/pp.88.4.1058

Reinecke, R., Müller Schmied, H., Trautmann, T., Andersen, L. S., Burek, P., Flörke, M., Gosling, S. N., Grillakis, M., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Thiery, W., Wada, Y., Yusuke, S., & Döll, P. (2021). Uncertainty of simulated groundwater recharge at different global warming levels: a global-scale multi-model ensemble study. *Hydrology and Earth System Sciences*, *25*(2), 787-810. https://doi.org/https://doi.org/10.5194/hess-25-787-2021

Rodrigues, T. R., Vourlitis, G. L., Lobo, F. d. A., Santanna, F. B., de Arruda, P. H. Z., & Nogueira, J. d. S. (2016). Modeling canopy conductance under contrasting seasonal conditions for a tropical savanna ecosystem of south central Mato Grosso, Brazil. *Agricultural and Forest Meteorology*, *218*, 218-229. https://doi.org/https://doi.org/10.1016/j.agrformet.2015.12.060

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling & Software*, *114*, 29-39. https://doi.org/https://doi.org/10.1016/j.envsoft.2019.01.012

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

Schulze, E. D., Lange, O. L., Kappen, L., Buschbom, U., & Evenari, M. (1973). Stomatal responses to changes in temperature at increasing water stress. *Planta*, *110*(1), 29-42. https://doi.org/10.1007/bf00386920

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Science Reviews*, *99*(3), 125-161. https://doi.org/https://doi.org/10.1016/j.earscirev.2010.02.004

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M., Thonicke, K., & Venevsky, S. (2003). Evaluation of ecosystem dynamics,

plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *9*(2), 161 - 185. https://doi.org/ https://doi.org/10.1046/j.1365-2486.2003.00569.x
Sommer, R., Sá, T. D. d. A., Vielhauer, K., Araújo, A. C. d., Fölster, H., & Vlek, P. L. G. (2002). Transpiration and canopy conductance of secondary vegetation in the eastern Amazon. *Agricultural and Forest Meteorology*, *112*(2), 103-121. https://doi.org/https://doi.org/10.1016/S0168-1923(02)00044-8
Speich, M., Dormann, C. F., & Hartig, F. (2021). Sequential Monte-Carlo algorithms for Bayesian model calibration – A review and method comparison. *Ecological Modelling*, *455*, 109608. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2021.109608
Stewart, J. B. (1988). Modelling surface conductance of pine forest. *Agricultural and Forest Meteorology*, *43*(1), 19-35. https://doi.org/https://doi.org/10.1016/0168-1923(88)90003-2
ter Braak, C. J. F., & Vrugt, J. A. (2008). Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, *18*(4), 435-446. https://doi.org/https://doi.org/10.1007/s11222-008-9104-9
Tuzet, A., Perrier, A., & Leuning, R. (2003). A coupled model of stomatal conductance, photosynthesis and transpiration. *Plant, cell & environment*, *26*(7), 1097-1116. https://doi.org/https://doi.org/10.1046/j.1365-3040.2003.01035.x
Uddling, J., Teclaw, R. M., Pregitzer, K. S., & Ellsworth, D. S. (2009). Leaf and canopy conductance in aspen and aspen-birch forests under free-air enrichment of carbon dioxide and ozone. *Tree Physiology*, *29*(11), 1367-1380. https://doi.org/https://doi.org/10.1093/treephys/tpp070
Urban, J., Ingwers, M., McGuire, M. A., & Teskey, R. O. (2017). Stomatal conductance increases with rising temperature. *Plant signaling & behavior*, *12*(8), e1356534-e1356534. https://doi.org/https://doi.org/10.1080/15592324.2017.1356534
Urban, J., Ingwers, M. W., McGuire, M. A., & Teskey, R. O. (2017). Increase in leaf temperature opens stomata and decouples net photosynthesis from stomatal conductance in Pinus taeda and Populus deltoides x nigra. *J Exp Bot*, *68*(7), 1757-1767. https://doi.org/10.1093/jxb/erx052
von Caemmerer, S., & Evans, J. R. (2015). Temperature responses of mesophyll conductance differ greatly between species. *Plant Cell Environ*, *38*(4), 629-637. https://doi.org/10.1111/pce.12449
Wahba, G. (1990). *Spline Models for Observational Data*.
Walker, A. P., De Kauwe Martin, G., Medlyn, B. E., Zaehle, S., Iversen, C. M., Asao, S., Guenet, B., Harper, A., Hickler, T., Hungate, B. A., Jain, A. K., Luo, Y., Lu, X., Lu, M., Luus, K., Patrick, M. J., Oren, R., Ryan, E., Shu, S., . . . Norby, R. J. (2019). Decadal biomass increment in early secondary succession woody ecosystems is increased by CO2 enrichment. *Nature Communications*, *10*(1), 1-13. https://doi.org/10.1038/s41467-019-08348-1
Wang, H., Guan, H., Deng, Z., & Simmons, C. T. (2014). Optimization of canopy conductance models from concurrent measurements of sap flow and stem water potential on Drooping Sheoak in South Australia. *50*(7), 6154-6167. https://doi.org/https://doi.org/10.1002/2013WR014818
Wang, H., Guan, H., Liu, N., Soulsby, C., Tetzlaff, D., & Zhang, X. (2020). Improving the Jarvis-type model with modified temperature and radiation functions for sap flow simulations. *Journal of Hydrology*, *587*, 124981. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.124981
Wang, H., Guan, H., & Simmons, C. T. (2016). Modeling the environmental controls on tree water use at different temporal scales. *Agricultural and Forest Meteorology*, *225*, 24-35. https://doi.org/https://doi.org/10.1016/j.agrformet.2016.04.016
Wang, J. L., Yu, G. R., Wang, B. L., Qi, H., & Xu, Z. J. (2005). Response of photosynthetic rate and stomatal conductance of rice to light intensity and CO2 concentration in northern China [(in Chinese with English abstract)]. *Chinese Journal of Plant Ecology*, *29*(1), 16-25. https://doi.org/https://10.17521/cjpe.2005.0003
Wang, S., Yang, Y., Trishchenko, A. P., Barr, A. G., Black, T. A., & McCaughey, H. (2009). Modeling the Response of Canopy Stomatal Conductance to Humidity *Journal of Hydrometeorology*, *10*(2), 521-532. https://doi.org/10.1175/2008jhm1050.1

Ward, E. J., Bell, D. M., Clark, J. S., & Oren, R. (2012). Hydraulic time constants for transpiration of loblolly pine at a free-air carbon dioxide enrichment site. *Tree Physiology*, *33*(2), 123-134. https://doi.org/10.1093/treephys/tps114

Whitley, R., Medlyn, B., Zeppel, M., Macinnis-Ng, C., & Eamus, D. (2009). Comparing the Penman–Monteith equation and a modified Jarvis–Stewart model with an artificial neural network to estimate stand-scale transpiration and canopy conductance. *Journal of Hydrology*, *373*(1), 256-266. https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.04.036

Whitley, R., Taylor, D., Macinnis-Ng, C., Zeppel, M., Yunusa, I., O'Grady, A., Froend, R., Medlyn, B., & Eamus, D. (2013). Developing an empirical model of canopy water flux describing the common response of transpiration to solar radiation and VPD across five contrasting woodlands and forests. *Hydrological Processes*, *27*(8), 1133-1146. https://doi.org/https://doi.org/10.1002/hyp.9280

Wood, S. N. (2016). Just Another Gibbs Additive Modeler: Interfacing JAGS and mgcv. *Journal of Statistical Software*, *75*(7), 1 - 15. https://doi.org/10.18637/jss.v075.i07

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548-1563. https://doi.org/https://doi.org/10.1080/01621459.2016.1180986

Wu, X., Xu, Y., Shi, J., Zuo, Q., Zhang, T., Wang, L., Xue, X., & Ben-Gal, A. (2021). Estimating stomatal conductance and evapotranspiration of winter wheat using a soil-plant water relations-based stress index. *Agricultural and Forest Meteorology*, *303*, 108393. https://doi.org/https://doi.org/10.1016/j.agrformet.2021.108393

Xu, Z., Jiang, Y., Jia, B., & Zhou, G. (2016). Elevated-CO2 Response of Stomata and Its Dependence on Environmental Factors. *Frontiers in Plant Science*, *7*, 657-657. https://doi.org/10.3389/fpls.2016.00657

Yang, L., Feng, Q., Zhu, M., Wang, L., Alizadeh, M. R., Adamowski, J. F., Wen, X., & Yin, Z. (2022). Variation in actual evapotranspiration and its ties to climate change and vegetation dynamics in northwest China. *Journal of Hydrology*, *607*, 127533. https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127533

Yang, Y., McVicar, T. R., Yang, D., Zhang, Y., Piao, S., Peng, S., & Beck, H. E. (2021). Low and contrasting impacts of vegetation CO2 fertilization on global terrestrial runoff over 1982–2010: accounting for aboveground and belowground vegetation–CO2 effects. *Hydrol. Earth Syst. Sci.*, *25*(6), 3411-3427. https://doi.org/10.5194/hess-25-3411-2021

Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R., & Donohue, R. J. (2019). Hydrologic implications of vegetation response to elevated CO2 in climate projections. *Nature Climate Change*, *9*(1), 44-48. https://doi.org/10.1038/s41558-018-0361-0

Yiqi, L., Yiqi Luo, H. A. M., Luo, Y., Mooney, H. A., & Saugier, B. (1999). *Carbon Dioxide and Environmental Stress*. Elsevier Science. https://books.google.com.au/books?id=DRhzoxlQ9zQC

Yun, K., Timlin, D., & Kim, S. H. (2020). Coupled Gas-Exchange Model for C(4) Leaves Comparing Stomatal Conductance Models. *Plants (Basel)*, *9*(10). https://doi.org/10.3390/plants9101358

Zeppel, M. J. B., Macinnis-Ng, C. M. O., Yunusa, I. A. M., Whitley, R. J., & Eamus, D. (2008). Long term trends of stand transpiration in a remnant forest during wet and dry years. *Journal of Hydrology*, *349*(1), 200-213. https://doi.org/https://doi.org/10.1016/j.jhydrol.2007.11.001

Zhang, C., Yang, Y., Yang, D., & Wu, X. (2021). Multidimensional assessment of global dryland changes under future warming in climate projections. *Journal of Hydrology*, *592*, 125618. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125618

Zhang, Q., Ficklin, D. L., Manzoni, S., Wang, L., Way, D., Phillips, R. P., & Novick, K. A. (2019). Response of ecosystem intrinsic water use efficiency and gross primary productivity to rising vapor pressure deficit. *Environmental Research Letters*, *14*(7), 074023. https://doi.org/10.1088/1748-9326/ab2603

Zhang, X. Y., Trame, M. N., Lesko, L. J., & Schmidt, S. (2015). Sobol Sensitivity Analysis: A Tool to Guide the Development and Evaluation of Systems Pharmacology Models. *CPT Pharmacometrics Syst Pharmacol*, *4*(2), 69-79. https://doi.org/10.1002/psp4.6

Zhang, Y., Bastos, A., Maignan, F., Goll, D., Boucher, O., Li, L., Cescatti, A., Vuichard, N., Chen, X., Ammann, C., Arain, M. A., Black, T. A., Chojnicki, B., Kato, T., Mammarella, I.,

Montagnani, L., Roupsard, O., Sanz, M. J., Siebicke, L., . . . Ciais, P. (2020). Modeling the impacts of diffuse light fraction on photosynthesis in ORCHIDEE (v5453) land surface model. *Geoscientific Model Development*, *13*(11), 5401-5423. https://doi.org/10.5194/gmd-13-5401-2020
Zhu, B., Huang, M., Cheng, Y., Xie, X., Liu, Y., Bisht, G., & Chen, X. (2021). Impact of vegetation physiology and phenology on watershed hydrology in a semi-arid watershed in the Pacific Northwest in a changing climate. *Water Resources Management*, *57*(3). https://doi.org/https://doi.org/10.1029/2020WR028394
Zomer, R. J., Trabucco, A., Bossio, D. A., & Verchot, L. V. (2008). Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment*, *126*(1), 67-80. https://doi.org/https://doi.org/10.1016/j.agee.2008.01.014

Chapter 3: Publication 2

The Impact of Environmental Variables on Surface Conductance: Advancing Simulation with a Nonlinear Machine Learning Model

The manuscript was published in the journal of Hydrology.

PUBLICATION 2

This section is to be completed by the student and co-authors. If there are more than four co-authors (student plus 3 others), only the three co-authors with the most significant contributions are required to sign below.

Please note: A copy of this page will be provided to the Examiners.

Full Publication Details	The impact of environmental variables on surface Conductance: Advancing simulation with a nonlinear Machine learning model
Section of thesis where publication is referred to	Chapter 3

Student's contribution to the publication

85	%	Research design
100	%	Data collection and analysis
95	%	Writing and editing

Outline your (the student's) contribution to the publication:

Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

APPROVALS

By signing the section below, you confirm that the details above are an accurate record of the students contribution to the work.

Name of Co-Author 1	Huade Guan	Signed		Date	14/06/2024
Name of Co-Author 2	Margaret Shanafield	Signed		Date	14/6/2024
Name of Co-Author 3	Okke Batelaan	Signed		Date	13/6/2024

3.1 Abstract

Surface conductance (Gs) is a key factor in the Penman-Monteith (PM) equation; the interaction between environmental variables such as CO2 concentration, air temperature (TA), vapor pressure deficit (VPD), soil water content (SWC), and net radiation (R) affects Gs, evapotranspiration and thus impacts the hydrological cycle. These interactions are highly nonlinear and vary among different vegetation types. However, conventional G_s simulation models use fixed interactions between environmental variables in their equations for all vegetation types. Moreover, the characterisation and parameterisation of conventional G_s models is highly uncertain due to the high spatiotemporal variability in key environmental variables and plant parameters, which inhibits their generalisation. This study investigates whether G_s could be estimated more accurately by nonlinear statistical techniques that capture the multiple interactions between the environmental variables that affect G_s for each vegetation type. We compare mixed generalized additive model (MGAM) for G_s simulation with semi-empirical and empirical models at 20 eddy covariance flux tower sites with four different vegetation types at daily and monthly timescales. The results show that the Nash-Sutcliffe Efficiency (NSE) in G_s simulation increased by up to 50% in MGAM model in comparison to the semi-empirical and empirical models. The MGAM model highlighted the interactive effects of CO_2, VPD, and SWC for crops and grasses. The interactive effects of CO_2, VPD, and TA were important for trees and grasses. The results from this study expand our understanding of the ability of G_s simulation models to identify and include the interactive effects of crucial environmental variables on plant transpiration and hydrological processes.

Key words: Surface conductance (G_s); Evaporation and transpiration (ET), Semi-empirical Model; Machine Learning (ML); Penman–Monteith (PM); Eddy Covariance Flux Tower.

3.2 Introduction

Evaporation and transpiration (ET) play a key role in hydrological processes as they return over 60% of global precipitation from the land surface to the atmosphere, and over 95% in arid climates (Koutsoyiannis, 2020; Zhan et al., 2019). ET plays an important role in driving land surface and atmosphere interactions because it links the water, energy, and carbon cycles (Hou et al., 2021; Yang et al., 2022). As the primary form of land-surface and atmosphere vapour exchange and accompanying processes with primary production, ET provides insight into both hydrological and biological processes (Lu et al., 2003; Zhang et al., 2021). Approximately 90% of the water that is absorbed by vegetation is consumed by transpiration through stomata in the leaves (Zang et al., 2012). Consequently, accurate estimations of ET and stomatal conductance (g_s) are important in the planning and implementation of irrigation and water conservation (Hou et al., 2021). While numerous studies have estimated ET and g_s, still there are some limitations in accurately capturing the dynamics of ET and g_s from leaf to ecosystem scale as surface conductance (G_s) due to the complex relationships between climate and vegetation (Page et al., 2018; Zhang et al., 2019). Despite the widespread use of the Penman–Monteith equation (PM) (Monteith, 1965) to simulate ET, the estimation of G_s in this equation remains a challenge (Ershadi et al., 2015; Li et al., 2019; Zhao et al., 2019). For a G_s simulation model to be appropriate, it should incorporate multiple interactions of the environmental variables in a highly nonlinear manner, which is a difficult endeavour and requires complex statistical analysis (Green et al., 2020; Koppa et al., 2022; Liao et al., 2021).

Conventional approaches (e.g., semi-empirical and empirical models) can estimate ET and G_s by flux-based models (which use the residual term in energy balance equation) and physical-based models with empirical equations based on vegetation and climate data (Ershadi et al., 2014; Lei et al., 2014; Polhamus et al., 2013; Zhao et al., 2019). These well-established models are easy to interpret but do not optimally extract information from data (Liu & Mishra, 2017;

Zhao et al., 2019). Furthermore, the high spatiotemporal variability in key plant parameters associated with G_s complicates the characterisation and parameterisation of these models and thus inhibits their generalisation (Abramowitz et al., 2007; Chitsaz et al., 2023; Dou & Yang, 2018; Green et al., 2020; Polhamus et al., 2013). Another limitation of conventional models (semi-empirical G_s simulation models) is the use of fixed environmental variables for G_s simulation for all vegetation types, ignoring differences in vegetation response to environmental variables (Dombrowski et al., 2022). Although conventional models have several calibrated parameters related to specific vegetation types, simulated G_s show large uncertainties to these parameters, which limits robustness in G_s simulation accuracy (Pan et al., 2020). In addition, these calibrated parameters are often determined through fitting on the existing dataset; therefore, they run the risk of not fully capturing vegetation response under the notable changes in climate conditions compared to the datasets on which these parameters are calibrated (Saunders et al., 2021). As a result, conventional models require re-parameterisation to be suitable for any changes in vegetation phenology or physiology caused by the variation in climate and growing season (Oliver et al., 2022).

Several studies have applied conventional G_s simulation models to demonstrate the interactive effects of vapour pressure deficit (VPD) and CO_2 on G_s (De Kauwe et al., 2021; Yuan et al., 2019). However, recent studies have indicated that this interaction is complex since it is also influenced by environmental conditions such as drought and water stress (Birami et al., 2020; Gattmann et al., 2021). Water stress indices, such as soil water content (SWC), are included in some semi-empirical G_s simulation models, but as discrete levels to assume a linear relation between soil moisture levels and G_s (Novick et al., 2016). However, including other climate variables, such as VPD and CO_2, at a continuous level leads to a partial comparison in these models (Kimm et al., 2020). Therefore, the comprehensive comparison of environmental variables individually and interactively is necessary for G_s simulation.

With large amounts of observed data accumulated in recent years, machine learning (ML) models have become increasingly prevalent in G_s simulation (Jung et al., 2019; Koppa et al., 2022; Zhao et al., 2019). ML models can learn complex patterns and relationships between variables and maintain greater consistency with the input data (Reichstein et al., 2019). The ML algorithm is robust in dynamic environments since it adapts to changes in data distribution over time. ML models are trained by measuring dynamic variables such as soil moisture, carbon fluxes, and precipitation in situ to enhance the accuracy and generalisation of estimating G_s and ET (Jung et al., 2019; Koppa et al., 2022). While ML models offer some advantages over conventional models for estimating G_s, combining ML with physical constraints has the potential to yield more promising results than simply replacing conventional models with ML (Reichstein et al., 2019). Combined models preserve the advantages of both physical models (physical consistency and interpretability) and ML models (data adaptability and more realistic data-driven formulation) and accurately estimate G_s (Zhao et al., 2019). However, analysis of the realistic, nonlinear interaction between environmental variables requires appropriate sets of statistical covariates in G_s simulation, considering differences in vegetation response to environmental variables.

This study used a nonlinear statistical model to simulate $G_{s,}$ considering the interactive effects of key environmental variables by quantifying the relationships between their covariates and the predicted G_s. First, we used an inverted PM equation to estimate G_s from observed data of 20 eddy covariance flux tower sites with different vegetation types at both daily and monthly timescales. Then we applied the mixed generalised additive model (MGAM) to optimise the learning of the relationship between VPD, CO_2, net radiation (R), air temperature (TA), and SWC at the continuous level for G_s simulation. We tested the results of the MGAM model against observed G_s data and compared with the results of the semi-empirical and empirical models. The MGAM highlighted the key environmental variables for each vegetation type by meeting the physical constraints. The different models (MGAM with different combinations of environmental variables) found the sensitivity of G_s simulation to direct and interactive effects of key environmental variables. In addition, MGAM visualisation by SHapley Additive exPlanations (SHAP) analysis shows both direct and interactive effects of key environmental variables on G_s fluctuations.

3.3 Data and methodology

3.3.1 Forcing data

This study utilised the Ameriflux sites (20 sites), which measure surface fluxes such as latent heat flux (LE), sensible heat flux (H), and soil heat flux (G), in addition to all required meteorological data for G_s simulation, including CO_2, VPD, TA, SWC, wind speed (U), and vegetation height (h). A variety of representative biomes were presented by eddy covariance flux towers, including crop (CRO), deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), and grass (GRA). Five sites were selected for each vegetation type at different locations in the United States and Canada (Table 1 and Fig. 1). Data were collected from each flux tower site at a daily timescale. Shortwave radiation conditions below 500 Wm^{-2} were also excluded to avoid data for cloudy days, morning dew and evaporation on the plant surfaces, which minimize the effects of soil evaporation (Griebel et al., 2020; Kimm et al., 2020; Nelson et al., 2020; Nie et al., 2021; Zhou et al., 2013). Due to the lack of energy balance closure in flux tower data, the Bowen ratio closure correction technique was used (Ershadi et al., 2014; Wehr & Saleska, 2021).

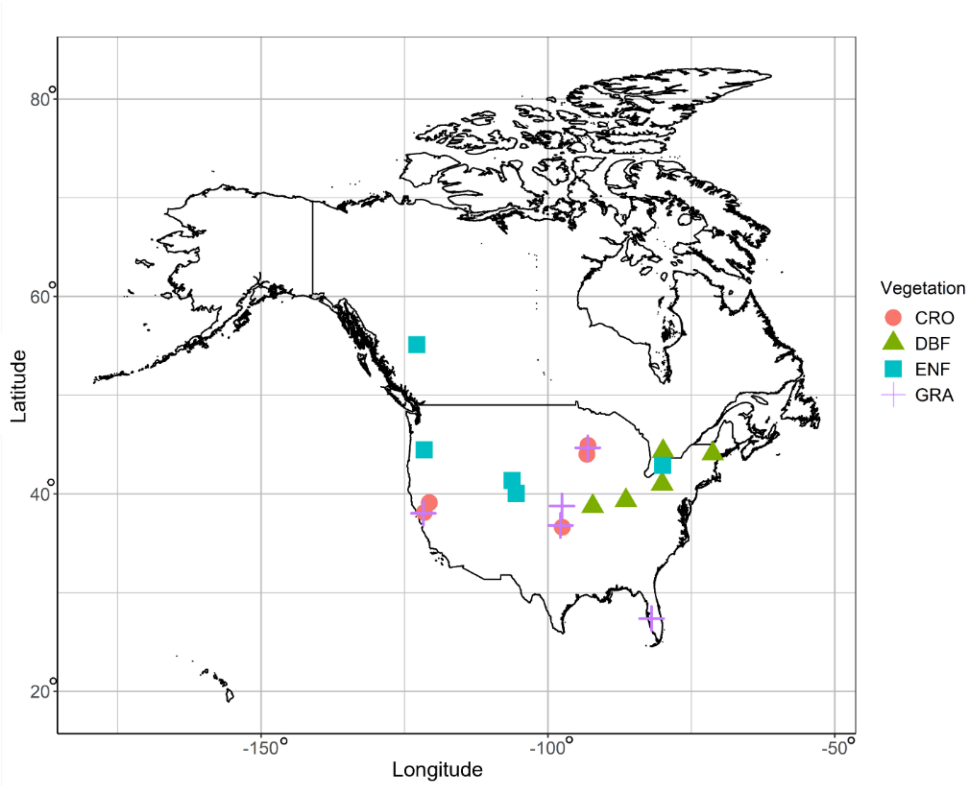


Figure 2 Location of the flux tower sites in the USA and Canada with four groups of vegetation types, including crop (CRO), deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), and grass (GRA).

Table1 Basic information for each Flux tower site for four groups of vegetation types: crop (CRO), deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), and grass (GRA).

	Site name	Years	Latitude	Longitude	Elevation	Vegetation Average age (year)	Average annual PR (mm)	Average annual TA (°C)	Vegetation height (m)	Vegetation	References
CRO	US-ARM	2003-2020	36.60	-97.48	314	-	843	14.7	4.5	Winter wheat, corn, soy, alfalfa	(Raz-Yaseef et al., 2015)
	US-Bi2	2017-2021	38.10	-121.53	-5	-	338	16.0	5.1	Ccorn	(Rey-Sanchez et al., 2021)
	US-Ro5	2017-2020	44.69	-93.05	283	-	879	6.4	3.5	Corn/soybean	(Chu et al., 2021)
	US-Ro6	2017-2021	44.69	-93.05	282	-	879	6.4	2.3	Corn/soybean	(Chu et al., 2021)
	US-Tw2	2012-2013	38.09	-121.64	-5	-	421	15.5	5.2	Twitchell Corn	(Knox et al., 2015)
DBF	CA-Cbo	1994-2020	44.31	-79.93	120	100	876	6.6	44	Red maple, white pine, large-tooth aspen and white ash	(Gu et al., 1999)
	CA-TPD	2012-2017	42.63	-80.55	260	90	1036	8.0	36.6	White oak, red maple, beech, ash, pine	(Beamesderfer et al., 2020)
	US-MMS	1999-2020	39.32	-86.41	275	80-90	1032	10.8	46	Acer saccharum, Liriodendron tulipifera, Quercus spp.	(Roman et al., 2015)

	US-Moz	2004-2019	38.74	-92.20	219	85-180	986	12.1	30	Mixture of abandoned agricultural fields, plantations, second-growth hardwood forests (white oak, sugar maple, hickory).	(Wood et al., 2018)
	US-xBR	2017-2021	44.06	-71.28	232	90-150	1246	5.6	35.68	Red maple, sugar maple, and beech	(Fer & Dietze, 2018)
ENF	CA-LP1	2007-2020	55.11	-122.84	751	80-110	570	2.0	26	Pure lodgepole pine	(Brown et al., 2012)
	CA-TP3	2002-2017	42.70	-80.34	184	45	1036	8.0	16	White pine	(Arain et al., 2022)
	US-GLE	2005-2020	41.36	-106.23	3197	400	1200	0.8	24.4	Abies lasiocarpa, Picea engelmannii_Pinus contorta	(Frank et al., 2014)
	US-Me2	2002-2020	44.45	-121.55	1253	67	523	6.2	32	Ponderosa pine trees and scattered incense cedars	(Kwon et al., 2018)
	US-NR1	1998-2016	40.03	-105.54	3050	97	800	1.5	21.5	Subalpine fir, Englemann spruce, lodgepole pine, aspen, limber pine	(Burns et al., 2015)
GRA	US-A32	2015-2017	36.81	-97.81	335	-	889	33.9	3.77	Grass	(Chu et al., 2021)
	US-KLS	2012-2019	38.77	-97.56	373	-	812	12.0	3	Grass	(Chu et al., 2021)
	US-ONA	2015-2020	27.38	-81.95	25	-	1268	22.3	2.8	Grass	(Silveira, 2021)
	US-Ro4	2014-2021	44.67	-93.07	274	-	879	6.4	2.6	Grass	(Griffis et al., 2011)
	US-Snf	2018-2020	38.04	-121.72	-4	-	381	24.6	3.49	Grass	(Chu et al., 2021)

3.3.2 Deriving G_s by inverting Penman–Monteith equation

The Penman model was originally developed to estimate the potential evaporation from open and saturated land surfaces (Penman, 1948). To describe the effects of partially closed stomata on evaporation under water stressed conditions, the model was generalised by incorporating a surface resistance term in the form of Eq. 1 and Eq. 2 (Monteith, 1965).

$$LE = \frac{\Delta(R - G) + \rho C_p G_a VPD}{\Delta + \gamma(1 + \frac{G_a}{G_s})} \tag{1}$$

$$G_s = \frac{G_a \gamma}{\frac{\Delta(R - G) + \rho C_p G_a VPD}{LE} - (\Delta + \gamma)} \tag{2}$$

where, LE is latent heat flux (Wm^{-2}), Δ is the slope of the saturation vapor pressure-temperature curve (Pa °C^{-1}), R and G are net radiation and soil heat flux (Wm^{-2}), ρ is air density (kg m^{-3}), C_p is specific heat capacity of dry air (J kg^{-1} °C^{-1}), VPD is vapor pressure deficit (Pa), γ is the psychrometric constant (Pa °C^{-1}), G_s and G_a are surface conductance and aerodynamic conductance (m s^{-1}).

The aerodynamic conductance used in the standard PM model is defined in Eq. 3 (Thom, 1972).

$$G_a = \frac{k^2 \times U}{\left[1n\left(\frac{z - d}{z_m}\right) \ln\left(\frac{z - d}{z_h}\right)\right]} \tag{3}$$

where, z is the wind speed measurement height (m), U is wind speed (m s^{-1}), $k = 0.41$ is the von Karman's constant, $d = 0.67 \times h$ is displacement height, h is the canopy height (m), $z_m = 0.123 \times h$, and $z_h = 0.0123 \times h$.

3.3.3 Semi-empirical and empirical G_s simulation models

We used the modified Medlyn model (MM) and Jarvis as semi-empirical and empirical G_s simulation models, respectively. The MM is an optimality-theory model which systematically estimates G_s at ecosystem scale (Lin et al., 2018; Medlyn et al., 2011; Nguyen et al., 2021; Nie et al., 2021). The MM is presented in Eq. 4.

$$G_s = g_0 + g_1 \times \left(\frac{GPP}{VPD^m}\right) \tag{4}$$

where, GPP is gross primary productivity (µmol m^{-2} s^{-1}), VPD (Pa), g_0 (mol m^{-2} s^{-1}), g_1 (Pa^m mol µmol^{-1}), and *m* are fitted parameters. If m=0.5, this approach is equivalent to the Medlyn model, while if m=1, it is equivalent to Leuning's model (Lin et al., 2018). The fitted parameters of the MM model are calibrated by Bayesian Markov-Chain Monte Carlo (MCMC) (Speich et al., 2021). The 'BayesianTools' and 'mcmc' packages in R (Geyer & Johnson, 2020; Hartig et al., 2019) are used for the calibration process in this study.

The Jarvis model is an empirical model to simulate G_s (Bai et al., 2019; Stewart, 1988). In Jarvis model G_s is a function of environmental variables with heavy parameterisation (Jarvis et al., 1976; Qi et al., 2023). This model (Eq. 5) represents the effects of each environmental variable independently, through Eq. 6-10.

$$G_s = G_{s\,max} \times f(R) \times f(TA) \times f(VPD) \times f(SWC) \times f(CO_2) \tag{5}$$

$$f(R) = \frac{R}{1000} \times \frac{1000 + a_1}{R + a_1} \tag{6}$$

$$f(TA) = 1 - a_2 \times (T_{min} - TA)^2 \tag{7}$$

$$f(VPD) = \exp(-a_3 \times VPD) \tag{8}$$

$$f(SWC) = 1 - a_4 \times \exp\left(a_5 \times \frac{SWC_{max} - SWC}{SWC_{max} - SWC_{min}}\right) \tag{9}$$

$$f(CO_2) = \frac{1}{1 + a_6 \times \exp\left(\frac{CO_2}{a_7} - 1\right)} \tag{10}$$

where, G_{smax} is the maximum surface conductance, R is net radiation, T_{min} is the minimum value of air temperature, SWC_{max}, SWC_{min} are the maximum and minimum values of SWC, and a_1 to a_7 are calibrated parameters.

3.3.4 Mixed generalised additive model (MGAM) for G_s simulation

MGAM is a nonlinear statistical technique to simulate complex nonlinear relationships between variables and responses (Hastie et al., 2009; Wood et al., 2016). The MGAM is capable of capturing complex patterns in data, and making accurate predictions on new unseen data; the flexibility and capability of MGAM, which is common in ML models, makes it part of the broader field of ML. The flexible regression functions in MGAM can demonstrate the relationships between covariates and outcomes in the form of Eq. 11.

$$f(x) = \sum_{k=1}^{K} \beta_k b_k(x) \tag{11}$$

where, $f(x)$ is a smoother function, b_k are basis functions, β_k are corresponding coefficients, and K is referred to as basis size or basis complexity. The coefficients of the basis functions are optimised to ensure the appropriate complexity of the models (Wood et al., 2016). The f(x) smoother function should be selected as a smooth function (S) which reflects the nonlinearity of variables directly, or tensor function (t_i) to represent the interaction between variables. The structure of G_s simulation in MGAM can be described as Eq. 12.

$$G_s = \sum_{m=1}^{M} f(x_m) \tag{12}$$

where, m are the effective environmental variables of G_s. Each of the effective variables has a smoother function $f(x)$ (Eq. 11), which contains basis functions with relevant coefficients.

The ‘nls’ and ‘mgcv’ packages in R (Baty et al., 2015; Wood et al., 2016) are used for G_s simulation by MGAM in this study.

3.3.5 SHapley Additive exPlanations (SHAP) analysis

SHAP analysis is based on cooperative game theory to interpret model simulation (Lundberg et al., 2020; Lundberg & Lee, 2017; Mardian et al., 2023). The SHAP value indicates the contribution of each variable or predictor to the model simulation and explains the effect of the high and low values of each variable on the simulated value (Shi et al., 2023). The SHAP value defines the weighted average of marginal contribution of each variable across all coalitions to which the variable belongs (Lee et al., 2023). The SHAP value is calculated by Eq. 13.

$$\varphi_i(f,x) = \sum_{s\subseteq x}\left[\frac{|s|!\,(M-|s|-1)!}{M!}\right]\times[f_x(s)-f_x(s\setminus i)] \tag{13}$$

where, φ is the SHAP value for variable $i=[1,M]$ and M is the number of variables, f is the simulation model, x is sample observation for specific ith variable, s is the subset of possible coalitions of variables. The first bracket of the equation refers to the weighting for each subset of coalitions, and the second bracket refers to the marginal contribution of ith variable, which is the difference between the f model with ($f_x(s)$) and without ($f_x(s\setminus i)$) the ith variable. The higher the SHAP value for each variable, the greater the impact of the variable on the simulation output (Lee et al., 2023). In this study, the SHAP method shows the contribution of VPD, R, Ta, CO_2, and SWC environmental variables in the G_s simulation. The ‘shapviz’ packages in R (Mayer, 2023) is used in this study.

3.3.6 Developing MGAM by environmental variables

We developed several simulation models (Table 2) to examine the direct and interactive effects of environmental variables on G_s in different vegetation types. In model 1 (benchmark model), all direct effects of environmental variables (VPD, CO_2, TA, SWC, and R) were included. Models 2 and 3 eliminated SWC and TA variables, respectively, to determine the sensitivity of G_s to these two variables in different vegetation types. In models 5 - 7, VPD, CO_2, and R were eliminated to illustrate the sensitivity of G_s to these variables. Model 4 illustrated the sensitivity of G_s to the interactive effects of environmental variables. It included functions that add SWC and TA to have interaction between VPD-CO_2 as key environmental variables on G_s fluctuation as identified in the literature review. The structure of model 4 was selected based on the highest NSE in G_s simulation for each vegetation type.

Table 2 Developing MGAM by environmental variables to test the direct and interactive effects of key variables on G_s simulation. S is smooth function which reflects the nonlinearity of variables directly, and ti is tensor function represent the interaction between variables.

Models		**S(VPD)**	**S(CO_2)**	**S(TA)**	**S(SWC)**	**S(R)**	**ti(VPD, CO_2, SWC)**	**ti(VPD, CO_2, TA)**
Model 1 (benchmark)		*	*	*	*	*		
Model 2		*	*	*		*		
Model 3		*	*		*	*		
	CRO	*	*	*	*	*	*	
Model 4	DBF & ENF	*	*	*	*	*		*
	GRA	*	*	*	*	*	*	*
Model 5			*	*	*	*		
Model 6		*		*	*	*		
Model 7		*	*	*	*			

3.4 Results

3.4.1 Validation of MGAM G_s simulation

Two approaches were used to validate the MGAM G_s simulation. In the first approach, the G_s simulation results of MGAM were compared with the MM and Jarvis as semi-empirical and empirical techniques. The results indicated that MGAM performs better in simulating G_s (higher NSE values) than the MM and Jarvis models for all four vegetation types across all flux tower sites (Fig. 2). The MM and Jarvis models required the Bayesian-MCMC calibration technique for fitting parameters. In the MM model, g_0, g_1, and m are fitted parameters (Table S1), and in Jarvis a_1-a_7 are fitted parameters (Table S2).

The second approach for validating MGAM was to train and test the model by inverting the PM equation and determining the G_s value. A 10-fold cross-validation technique was used to validate the results of the MGAM G_s simulation, in which 70% of data was used to train the model and 30% to test it. The training data was divided into ten folds, and the model was trained through nine folds and validated through the tenth fold. This process is repeated to cover all observed values at both training and testing. A cross-validation procedure was conducted using the caret package in R (Kuhn, 2021). Comparing the MGAM performance for the testing data (Fig. 3 for models 1 - 4 and Fig. S2 for models 5 - 7) and the training data (Fig. S3 for models 1 - 4 and Fig. S4 for models 5 - 7) indicated that the models were well trained. Accordingly, the NSE of the models showed good results for test data, which were close to the NSE of the model for the training data.

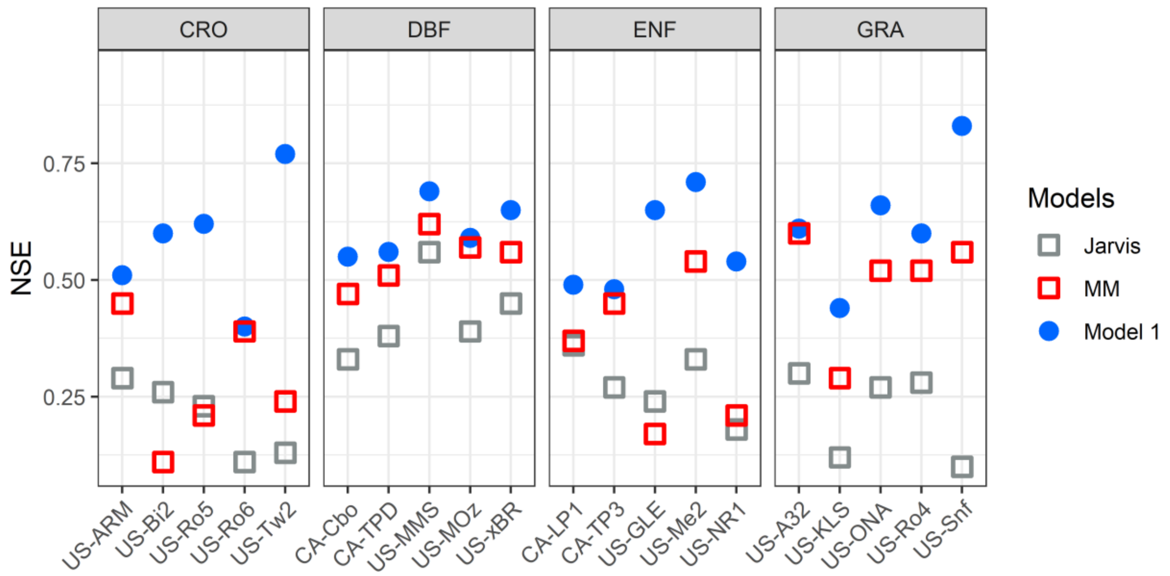


Figure 3 Comparison of MGAM G_s (Model 1 - benchmark) simulation accuracies with MM and Jarvis models in 20 flux tower sites with four vegetation types of crop (CRO), deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), grass (GRA).

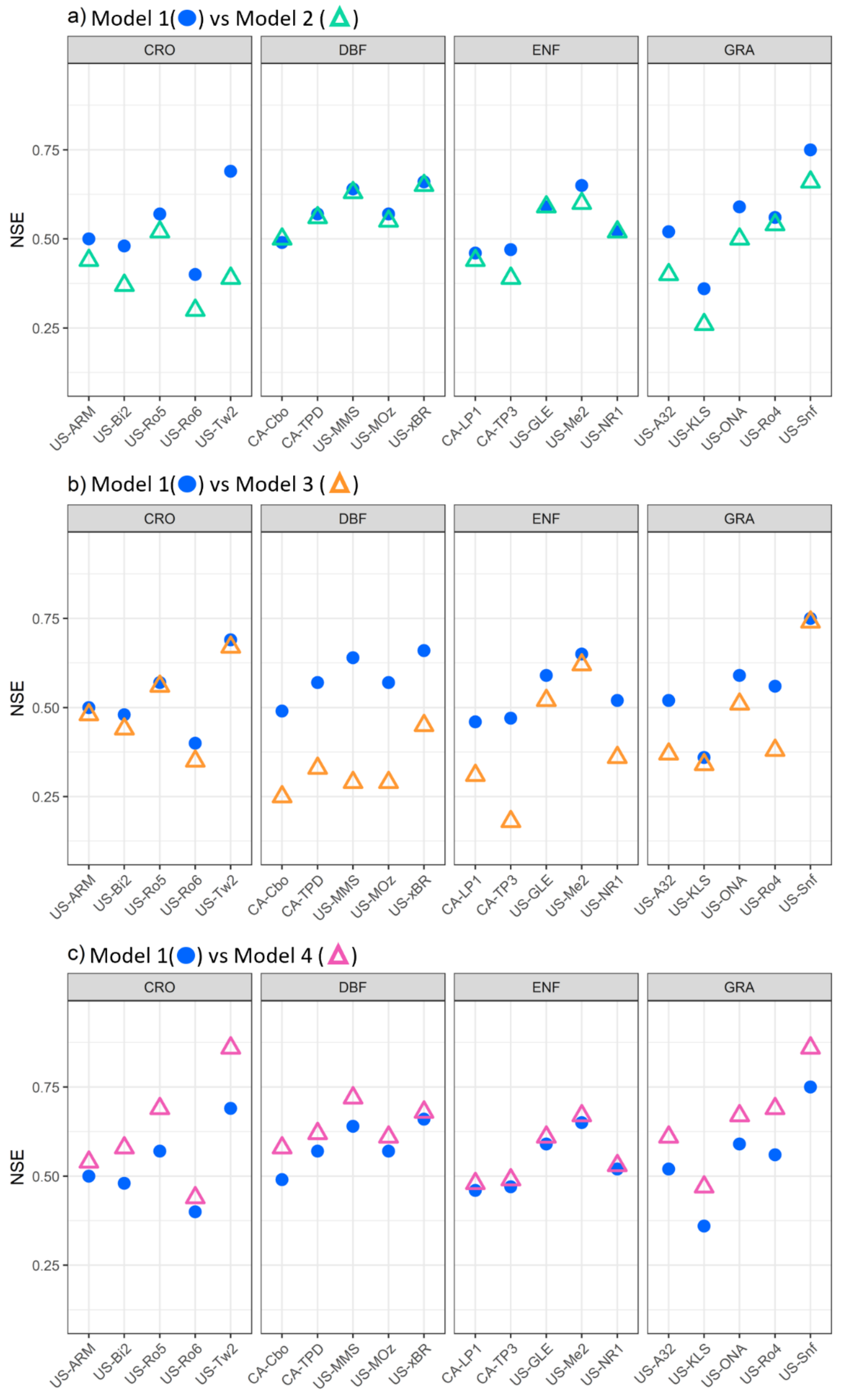


Figure 4 Comparison of G_s simulation accuracies for test data, a) model 1 (benchmark) vs model 2, b) model 1 vs model 3, c) model 1 vs model 4.

3.4.2 The importance of direct and interactive effects of environmental variables on Gs simulation for different vegetation types

The comparison between G_s simulation accuracies for different models (Table 2) demonstrated the sensitivity of G_s to environmental variables in different vegetation types. To demonstrate the role of environmental variables in G_s simulation, model 1 (benchmark, with direct effects of all environmental variables) was compared with other models in terms of their NSE values (Fig. 3 and Fig. S2 for testing data, Fig. S3 and S4 for training data). Comparing the NSE values in model 1 and model 2 (benchmark without SWC), CRO and GRA had a greater difference in NSE, indicating that these plants were sensitive to SWC. DBF and ENF were not sensitive to SWC in models 1 and 2, except for a few sites with on average younger trees such as CA-TP3 and US-Me2 and with mixed vegetation such as US-MOZ (Fig. 3a). DBF and ENF had higher sensitivity to TA in all sites when comparing the NSE value in models 1 and 3 (benchmark without TA). Due to the difference in NSE values between models 1 and 3 (Fig. 3b), GRA was also sensitive to TA in most sites, whereas CRO was not very sensitive to TA in most sites. The interactive effects of environmental variables on G_s simulation are provided in model 4 (Fig. 3c). Due to the greater sensitivity of the CRO to SWC in model 2, the interactive effects of (VPD, CO_2, SWC) demonstrated high improvement in NSE values in model 4. In model 3, the DBF and ENF were more sensitive to TA. Thus, the interactive effects of (VPD, CO_2, and TA) for model 4 demonstrated improvement in NSE values for G_s simulation in these vegetation types. For GRA, both interactive functions of (VPD, CO_2, SWC) and (VPD, CO_2, TA) were applied to model 4. This is because GRA were sensitive to both SWC and TA in accordance with models 2 and 3. We compared differences in NSE values between models 1 and 5 (benchmark without VPD), models 1 and 6 (benchmark without CO_2) and models 1 and 7 (benchmark without R) to determine the sensitivity of vegetation types to VPD, CO_2 and R. When comparing NSE values in model 1 with models 5 - 7, it was revealed that all vegetation types have similar sensitivity to VPD, CO_2 and R. Most flux tower sites had high sensitivity to VPD but low sensitivity to CO_2 (Fig. S2a-c). In all models, the results of the G_s simulation in test data were similar to those in training data (Fig. S3-S4). A schematic path analysis was used for the testing and training data in order to visually demonstrate the effect of each environmental variable on G_s simulation (Fig. 4 and Fig. S5). The NSE value changes in G_s simulation (in %) are presented next to each arrow (path value) and shown as the thickness of arrows for visual comparison. The path values are the average of NSE value changes compared to model 1 (benchmark) for all five flux tower sites in each vegetation type.

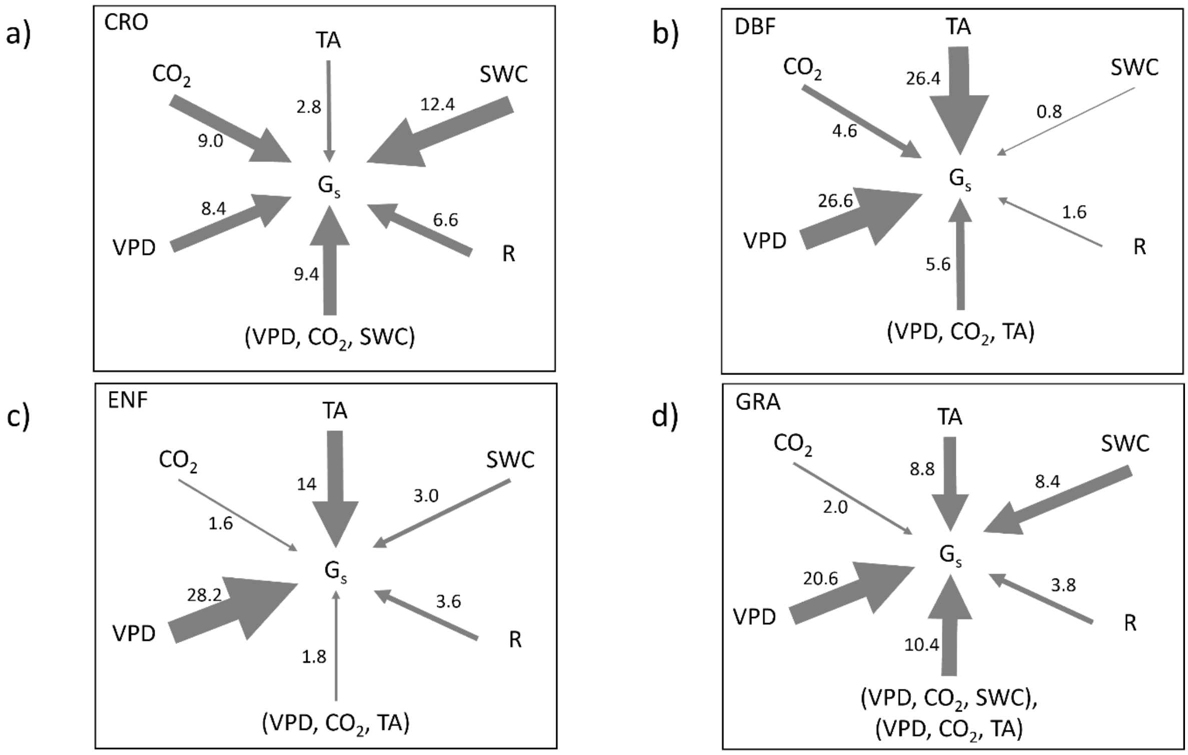


Figure 4 Schematic path analysis for the test data showing the effect of each environmental variable on G_s simulation. The path values and arrow thicknesses are the average NSE value changes compared to model 1 (benchmark) for each environmental variable (%) and for all five flux tower sites in vegetation type: a) crop (CRO), b) deciduous broad-leaf forest (DBF), c) evergreen needle-leaf forest (ENF), and d) grass (GRA).

3.4.3 Visualisation of direct and interactive effects of environmental variables on G_s simulation

The MGAM results were visualised by SHAP analysis to determine the direct effects of each environmental variable on G_s simulation (Fig. 5). All five sites with each vegetation type are considered in the same category. The SHAP value of each observed data shows the changes in weighted average of G_s values forced by each environmental variable (x-axis). The average on SHAP values for all observed data in each variable shows the contribution of variable in G_s simulation (y-axis). The gradient colour (feature value) shows the original value of each variable. The SHAP value for each environmental variables (y-axis) shows that VPD has the highest contribution in G_s simulation for most of the vegetation types. The gradient colour shows that VPD has a decreasing effect on G_s. The SWC has more contribution in G_s simulation for CRO and GRA with an overall increasing effect on G_s for these vegetations. While TA has considerable effects on G_s for DBF, ENF, and GRA, with overall increasing effect on G_s value. The R and CO_2 have increasing and decreasing effects on G_s, respectively, but do not show a high independent contribution in comparison to other key variables.

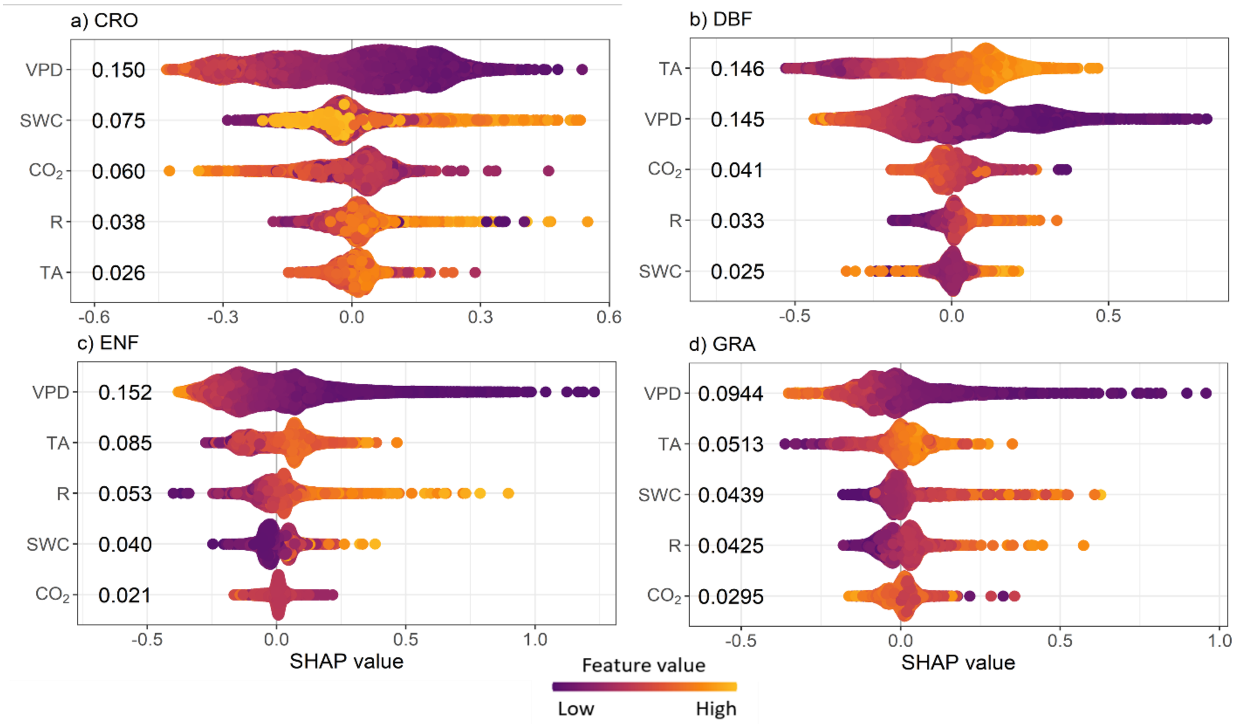


Figure 5 The SHAP value to present the contribution of each environmental variable in G_s simulation. The x-axis shows the SHAP value for all observed data of each environmental variable, y-axis shows the average of SHAP values for each environmental variable, and gradient colour (feature value) shows the original value of each environmental variable.

The visualisation of the interactive effects of environmental variables on G_s was performed for the groups of vegetation with similar responses to the environmental variables (Fig. 6). GRA and CRO are both sensitive to the interactive effects of VPD, CO_2, and SWC; therefore, they are considered in the same category for visualisation (Fig. 6a and 6b). The VPD shows the decreasing effects on G_s in all conditions. At a lower level of CO_2, the decreasing effect of VPD is notable but higher level of CO_2 alleviates the decreasing effects of VPD on G_s values (Fig. 6a). Similar to CO_2, SWC alleviates the decreasing effects of VPD; the increase in SWC increase the G_s values (Fig. 6b). GRA, DBF, and ENF show the sensitivity to TA (Fig 6c-d). TA has increasing effects on G_s, but increases of TA higher than a threshold value degrade the G_s values. The TA effects on G_s are considerable at the higher VPD (Fig. 6c-d). DBF and ENF are grouped in the same category because they both are sensitive to the interactive effects of VPD, CO_2, and TA (Fig. 6d and 6e). Moreover, CO_2 alleviate the decreasing effects of VPD by decreasing the G_s value for DBF and ENF, similar to the CRO and GRA (Fig. 6e and Fig. 6a).

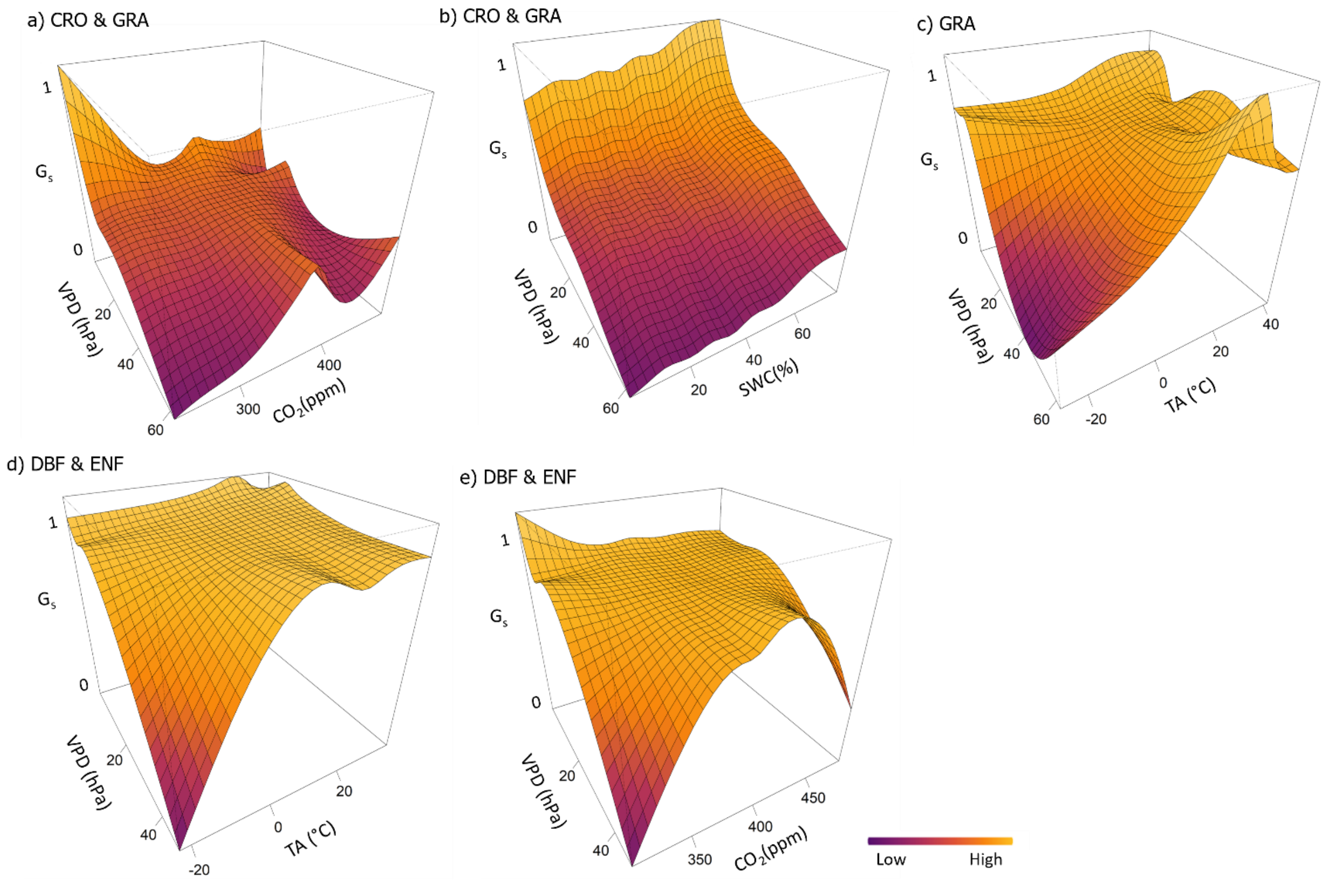


Figure 6 The interactive effects of environmental variables on normalised G_s: a) VPD and CO_2 for CRO and GRA, b) VPD and SWC for CRO and GRA, c) VPD and TA for GRA, d) VPD and TA for DBF and ENF, e) VPD and CO_2 for DBF and ENF.

3.4.4 The role of interactive effects of environmental variables on G_s simulation at monthly timescale

As described in section 3.2, the direct effects of SWC were more important in G_s simulation for CRO, GRA, and younger trees, while G_s simulation in mature trees showed less sensitivity to SWC. In contrast, TA played more important role in improving G_s simulation for trees and GRA. Although CO_2 less directly affected G_s simulation for all vegetation types, its interactive effects with other climate variables showed high improvements in G_s simulation. To further evaluate the role of the interactive effects of environmental variables on G_s, comparisons of the NSE values for models 1 and 4 in the G_s simulation were made at the monthly timescales (Fig. 7). The NSE value at monthly timescale is calculated by the comparison between observed and simulated G_s for each specific months during the whole period of data. At both the beginning (Jan-Mar) and the end (Nov-Dec) of the growth period, interactive effects of (VPD, CO_2, TA) played an important role for most DBF and some ENF sites (CA-LP1 and CA-TP3) (Fig. 7). Furthermore, the interactive effects of (VPD, CO_2, SWC) for CRO, and the interactive effects of (VPD, CO_2, SWC) and (VPD, CO_2, TA) for GRA were crucial throughout the entire growth period irrespective of any specific pattern (Fig. 7).

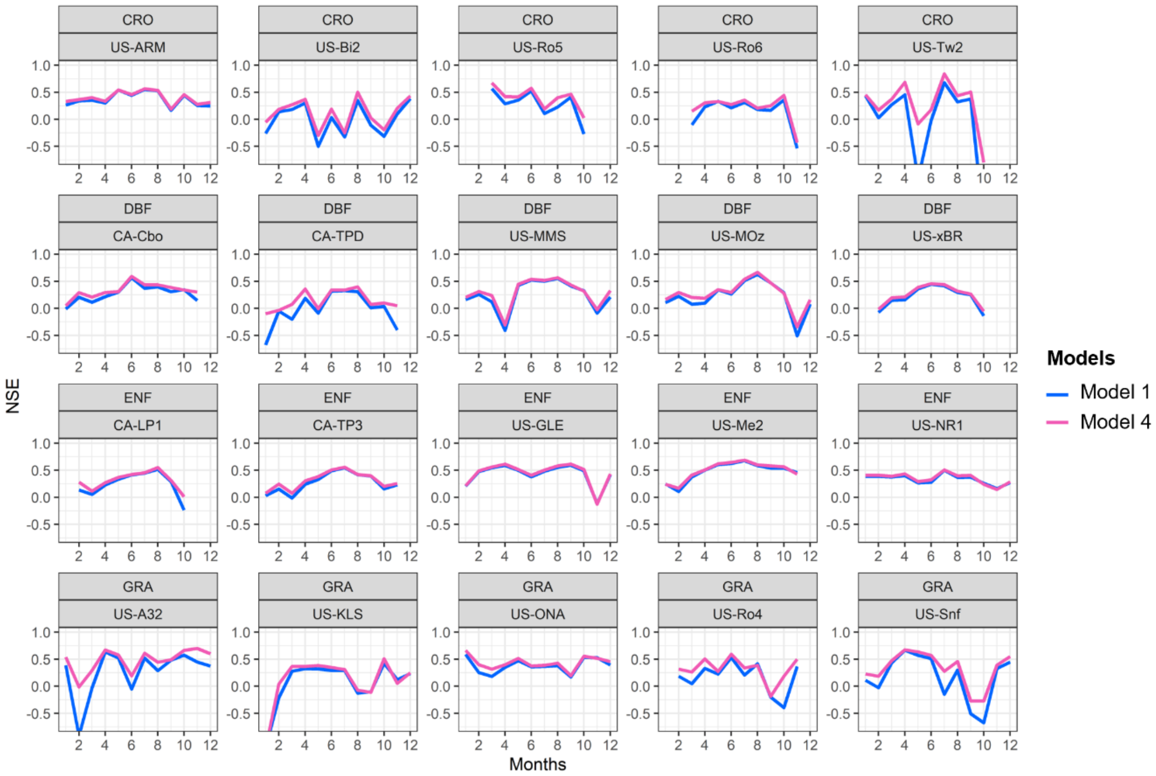


Figure 7 The comparison of monthly G_s simulation accuracy in model 1 (benchmark) containing all direct effects of environmental variables with model 4 by added interactive effects of (VPD, CO_2, SWC) for CRO and GRA, and interactive effects of (VPD, CO_2, TA) for DBF, ENF, and GRA to benchmark model.

NSE values in model 1 and model 2 (Fig. S7) demonstrated the important role played by SWC for CRO and GRA during the entire growth period. However, SWC had only a minor effect on DBF and ENF (excluding US-MOZ, CA-TP3, and US-Me2) at monthly timescales. The sensitivity to SWC results was consistent with those presented in Fig. 3a. NSE values for models 1 and 3 (Fig. S8) revealed that TA played a crucial role in DBF and ENF, particularly at the beginning (Jan-Mar) and the end (Nov-Dec) of the growth period. The GRA also showed a sensitivity to TA over the entire growth period without a specific pattern, whereas the CRO showed a less pronounced response. As presented in Fig. S8, the sensitivities to TA results were consistent with those presented in Fig. 3b.

The role of VPD, CO_2 and R on G_s simulation at a monthly time scale was considered in models 5 - 7. A comparison between NSE values in models 1 and 5 indicated that VPD played an important role throughout the whole growth period for all vegetation types (Fig. S9). The interaction of CO_2 with other environmental variables improved the results of G_s simulation, but its direct effect was less important than that of other environmental variables without any pattern for all vegetation types (Fig. S10). The role of R in G_s simulation at monthly timescale did not show any pattern for different vegetation types (Fig. S11).

3.5 Discussion

3.5.1 Advantages of MGAM over semi-empirical and empirical models

The comparison between MGAM, MM and Jarvis models showed that MGAM outperforms semi-empirical and empirical models in G_s simulation for all vegetation types in the flux tower sites (Fig. 2). The significant calibration process in Jarvis models and the independent combination of environmental variables leads to the challenging analysis of interactive effects of key environmental variables on G_s. According to the MM model, environmental variables such as VPD, SWC, and TA were partially incorporated into photosynthetic rate or GPP as a diffusive flux between the leaf and the atmospheric boundary layer (Ball, 1987; Leuning, 1995; Lin et al., 2018; Medlyn et al., 2011). Although including the photosynthetic rate in G_s models may improve simulation accuracy, analysis of the direct and interactive effect of environmental variables on G_s is challenging when using the photosynthetic rate or GPP (Kimm et al., 2020). The effects of environmental variables on G_s are embedded in photosynthetic rate or GPP and partitioning the effects of each variable is not straightforward. The suggested MGAM approach of this study is capable of measuring the direct effects of each environmental variable by adding or removing each variable. In addition, evaluating the interactive effects of key environmental variables is possible by including the interactive tensor function in MGAM. The MGAM model provided a higher level of accuracy in the G_s simulation than previous models without relying on GPP or photosynthesis variables.

Furthermore, formulating the impact of SWC on G_s is difficult because it is correlated with VPD. Hence, several conventional G_s simulation models incorporated SWC in discrete levels for simplicity. Therefore, it was assumed that each soil moisture level had a linear relationship with G_s. However, other environmental variables are quantified as continuous variables (Novick et al., 2016). As a result, comparisons between these variables are not appropriate (Kimm et al., 2020). In the MGAM model in this study, all variables are treated equally on a continuous scale. Thus, this model was better suited to our goal of examining the direct and interactive effects of key environmental variables.

3.5.2 The direct and interactive effects of environmental variables on MGAM G_s simulation for each vegetation type

In this study, the structure of MGAM was designed to take into account the sensitivity of each vegetation type to environmental variables. An analysis of the sensitivity of vegetation to environmental variables was performed by eliminating or adding each environmental variable and their interactions in the MGAM benchmark model and evaluating the changes of NSE in G_s simulation. Our analysis of the direct effect of each environmental variable on G_s highlighted the notable contribution of VPD, SWC, CO_2, R and TA. VPD had the greatest direct impact on G_s for all flux tower sites without differentiation for the vegetation types (Fig. S2a and Fig. S4a), while SWC had a greater influence on G_s for CRO and GRA and younger trees (Fig. 3a and Fig. S3a), and TA had a greater impact on G_s for GRA and trees (DBF and ENF) (Fig. 3b and Fig. S3b). In comparison with other variables, CO_2 had less direct effect on G_s simulation accuracy. However, its interaction with VPD, SWC, and TA affected G_s simulation accuracy for all vegetation types (Fig. S3c and Fig. 3c).

There is consistency between our results and those found in other studies. As previously reported (Wang et al., 2012; Wertin et al., 2012), TA played an important role in determining the G_s of trees (DBF and ENF). According to Lin et al. (2018), VPD had an important influence on G_s in different vegetation types (Lin et al., 2018). The greatest variation of G_s in the flux tower data for the U.S. Corn Belt was attributed to VPD and SWC (Kimm et al., 2020).

In light of the fact that G_s is highly correlated with photosynthesis, several studies concluded that it is necessary to consider the impacts of both VPD and CO_2 on G_s simulation simultaneously (De Kauwe et al., 2021; Lin et al., 2018; Nadal-Sala et al., 2021). Some studies reported that elevated CO_2 offsets the negative effects of high VPD on G_s (De Kauwe et al., 2021; Yuan et al., 2019). Recent studies, however, suggested that vegetation receives this benefit from elevated CO_2 when they are not exposed to severe droughts or water stresses (Birami et al., 2020; Gattmann et al., 2021). It has been observed that both SWC and VPD affect stomatal conductance, and GPP (Anderegg et al., 2012; Breshears et al., 2013; Sulman et al., 2016). This statement was in accordance with our results, which indicated that CO_2 and SWC alleviate the decreasing effects of VPD on G_s for CRO and GRA (Fig. 6a-b).

VPD played an important role as a driver of carbon and water fluxes, especially during heat waves; since it is highly likely that global temperatures will increase in the future, VPD should also increase (Park Williams et al., 2013). An increase in TA can increase the G_s, but when the increase in TA exceeds a threshold value, it degrades the G_s; the effects of TA on G_s is intensified at higher VPD (Purcell et al., 2018; Urban et al., 2017). In this study, the visualisation of simulated G_s by MGAM clearly showed that higher VPD intensified the effects of TA on G_s for trees (Fig. 6d). However, the elevated CO_2 could mitigate the negative effects of high VPD on G_s for trees (Wang et al., 2012; Wertin et al., 2012), supported by our results (Fig. 6e). Hence, the combination of elevated CO_2 and elevated TA may promote the fixation of carbon and the accumulation of biomass (Morison & Lawlor, 1999; Wang et al., 2012).

Model 4 in this study incorporated the interactive effects of VPD, CO_2, and TA for trees (DBF and ENF) and showed improvement in G_s simulation, which is supported in literature reviews (Mathias & Thomas, 2021; Urban et al., 2017; Wang et al., 2012; Wertin et al., 2012).

3.5.3 Interactive effects of environmental variables on G_s simulation at a monthly timescale

The comparison between interactive effects of environmental variables on G_s simulation revealed that trees were sensitive to the interaction between VPD, CO_2, and TA. NSE values at a monthly timescale for models 1 and 4 showed that this sensitivity was higher during the growth periods of January to March and November to December (Fig. 7). For most of the DBFs and two sites of the ENFs (CA-LP1 and CA-TP3), this specific pattern was evident. A possible explanation for this pattern can be found at the beginning and end of the growth period at the mentioned sites which have lower monthly average TA with higher variation (Fig. S12). Although the US-GLE site also had low average monthly TA values, the lower sensitivity to TA can be justified by its mature trees (average age of 400 years).

A comparison between model 1 and model 3 indicated that trees are highly sensitive to TA both at the beginning and the end of their growth period (Fig. S8). There was also evidence in the literature that TA had a more important effect on G_s at the beginning of the growth period rather than in the middle of it (D'Arrigo et al., 2004; Wertin et al., 2012). For the Loblolly pine, the combined effects of elevated TA and CO_2 on G_s were more important during the cooler months of October as opposed to the warmer months of June and September (Wertin et al., 2012). According to a meta-analysis of plant response to TA, the simultaneous TA and CO_2 treatments had more considerable effects at ambient temperature rather than at elevated temperature (Wang et al., 2012).

3.6 Conclusion

The G_s simulation is one of the most complex parts of the Penman-Monteith equation. As a result of the nonlinear and multiple interactions between environmental variables, the G_s simulation has always been challenging. The MGAM, as an ML model with physical constraints, was capable of simulating G_s through optimal extraction of information from data. MGAM provided higher accuracy than semi-empirical and empirical models in G_s simulation, and its flexibility in applying multiple interactions of key environmental variables makes it an alternative tool for simulating G_s. For forest, crop, and grass ecosystems, the MGAM model developed in this study provided a satisfactory simulation of G_s for both testing and training dataset, suggesting its further use in G_s and ET prediction and generalisation. By employing MGAM to enhance plant ET simulation and prediction, the potential arises for more exploration into comprehending the intricate interplay among ET, precipitation, and streamflow – pivotal elements in water resources and hydrological ecosystems.

Acknowledgement

The support from the Australian Government Research Training Program Scholarship, National Centre for Groundwater Research and Training (NCGRT), and Flinders University is acknowledged.

Supporting Information for

The Impact of Environmental Variables on Surface Conductance: Advancing Simulation with a Nonlinear Machine Learning Model

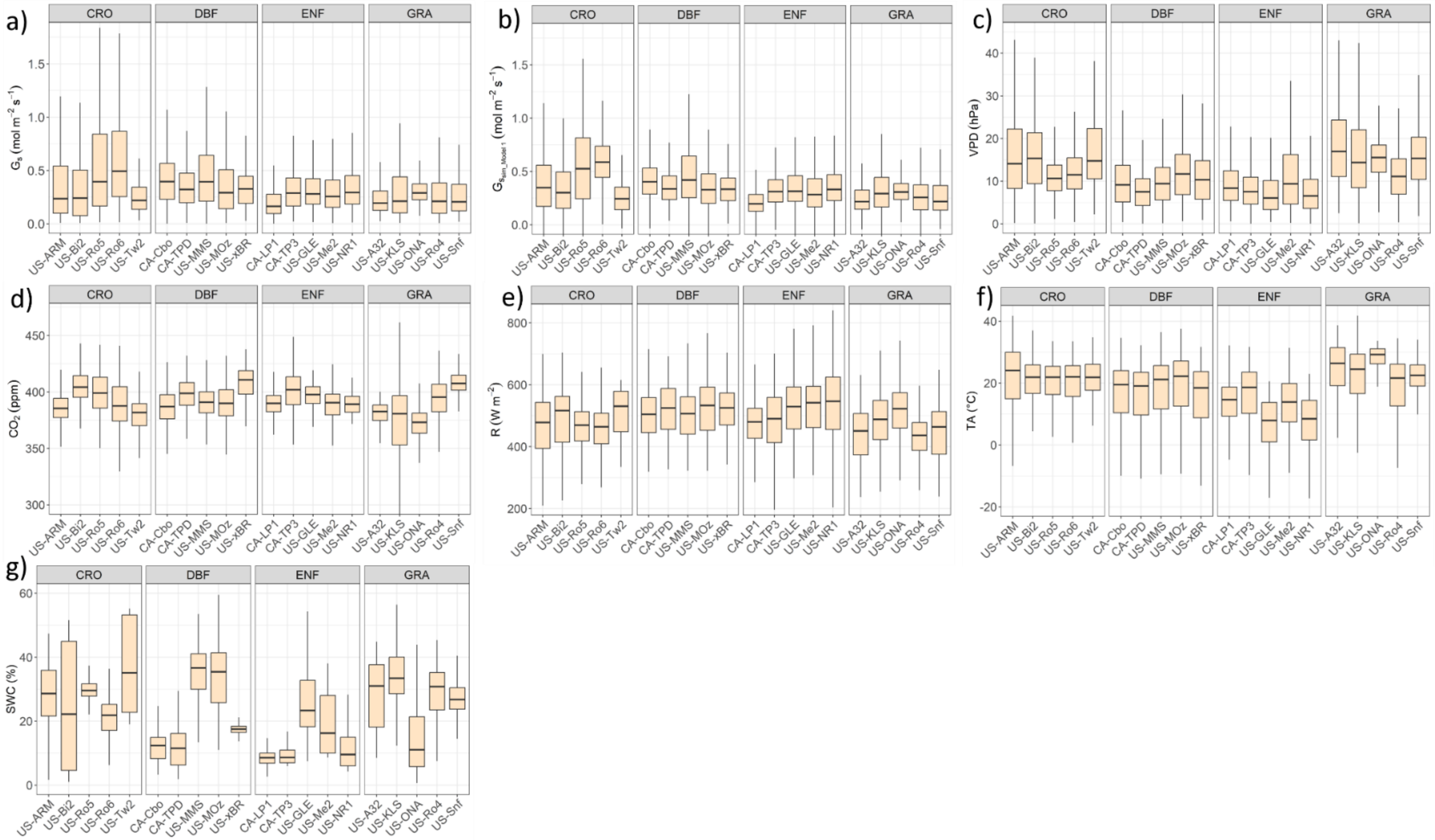


Figure S5 The range of all environmental variables for 20 flux tower sites with four vegetation types of crop (CRO), deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), and grass (GRA), a) observed G_s, b) simulated G_s by model 1 (benchmark), c) VPD, d) CO_2, e) R, f) TA, g) SWC.

Table S1 Bayesian-MCMC calibration results for fitted parameters in Modified Medlyn (MM) model for four groups of vegetation types including crop (CRO), deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), and grass (GRA).

Vegetation types	**Site names**	**g_0 ($\frac{mol}{m^2 s}$)**	**g_1 ($Pa^m \frac{mol}{\mu mol}$)**	**m**
CRO	US-ARM	0.14	30.42	0.66
	US-Bi2	0.28	7.28	0.65
	US-Ro5	0.37	11.47	0.62
	US-Ro6	0.26	26.55	0.63
	US-Tw2	0.18	13.54	1.76
ENF	CA-Cbo	0.24	9.70	0.77
	CA-TPD	0.17	11.81	0.51
	US-MMS	0.18	17.87	0.55
	US-MOz	0.10	21.64	0.59
	US-xBR	0.15	14.86	0.65
DBF	CA-LP1	0.09	18.51	0.97
	CA-TP3	0.11	11.77	0.62
	US-GLE	0.29	5.78	1.45
	US-Me2	0.07	13.67	0.85
	US-NR1	0.25	13.57	0.56
GRA	US-A32	0.05	18.69	0.72
	US-KLS	0.13	20.79	0.63
	US-ONA	-0.05	28.80	0.58
	US-Ro4	0.11	13.86	0.81
	US-Snf	0.04	22.97	0.72

Table S2 Bayesian-MCMC calibration results for fitted parameters in Jarvis model for four groups of vegetation types including crop (CRO), deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), and grass (GRA).

Vegetation types	**Site names**	**a_1**	**a_2**	**a_3**	**a_4**	**a_5**	**a_6**	**a_7**
CRO	US.ARM	167.50	0.02	0.12	1.23	0.00	1.71	132.32
	US.Bi2	100.73	0.00	0.13	0.00	7.67	1.60	399.09
	US.Ro5	104.20	0.40	0.11	1.01	0.02	1.34	109.61
	US.Ro6	251.70	0.05	0.10	1.06	0.00	1.57	112.34
	US.Tw2	236.80	0.55	0.10	1.04	0.00	2.36	110.81
ENF	CA.Cbo	215.14	0.01	0.10	1.24	0.01	2.22	101.28
	CA.TPD	203.04	0.03	0.10	1.02	0.01	2.54	299.44
	US.MMS	398.80	0.15	0.10	1.01	0.00	2.17	282.96
	US.MOz	104.56	0.06	0.08	1.01	0.00	2.28	272.51
	US.xBR	112.62	0.01	0.08	1.30	0.02	2.60	215.38
DBF	CA.LP1	150.69	0.02	0.18	1.05	0.00	2.55	319.39
	CA.TP3	101.35	0.02	0.14	1.14	0.00	2.36	107.98
	US.GLE	397.07	0.75	0.34	1.00	0.00	1.92	398.37
	US.Me2	336.64	0.77	0.28	1.03	0.00	2.14	115.44
	US.NR1	396.54	0.17	0.18	1.02	0.00	2.00	186.88
GRA	US.A32	379.17	0.60	0.08	1.01	0.00	2.74	192.59
	US.KLS	132.90	0.78	0.10	1.00	0.00	1.87	114.54
	US.ONA	150.57	0.11	0.10	1.23	0.00	2.70	106.93
	US.Ro4	349.69	0.47	0.09	1.00	0.00	2.34	244.63
	US.Snf	342.04	0.98	0.31	1.46	0.00	2.09	100.93

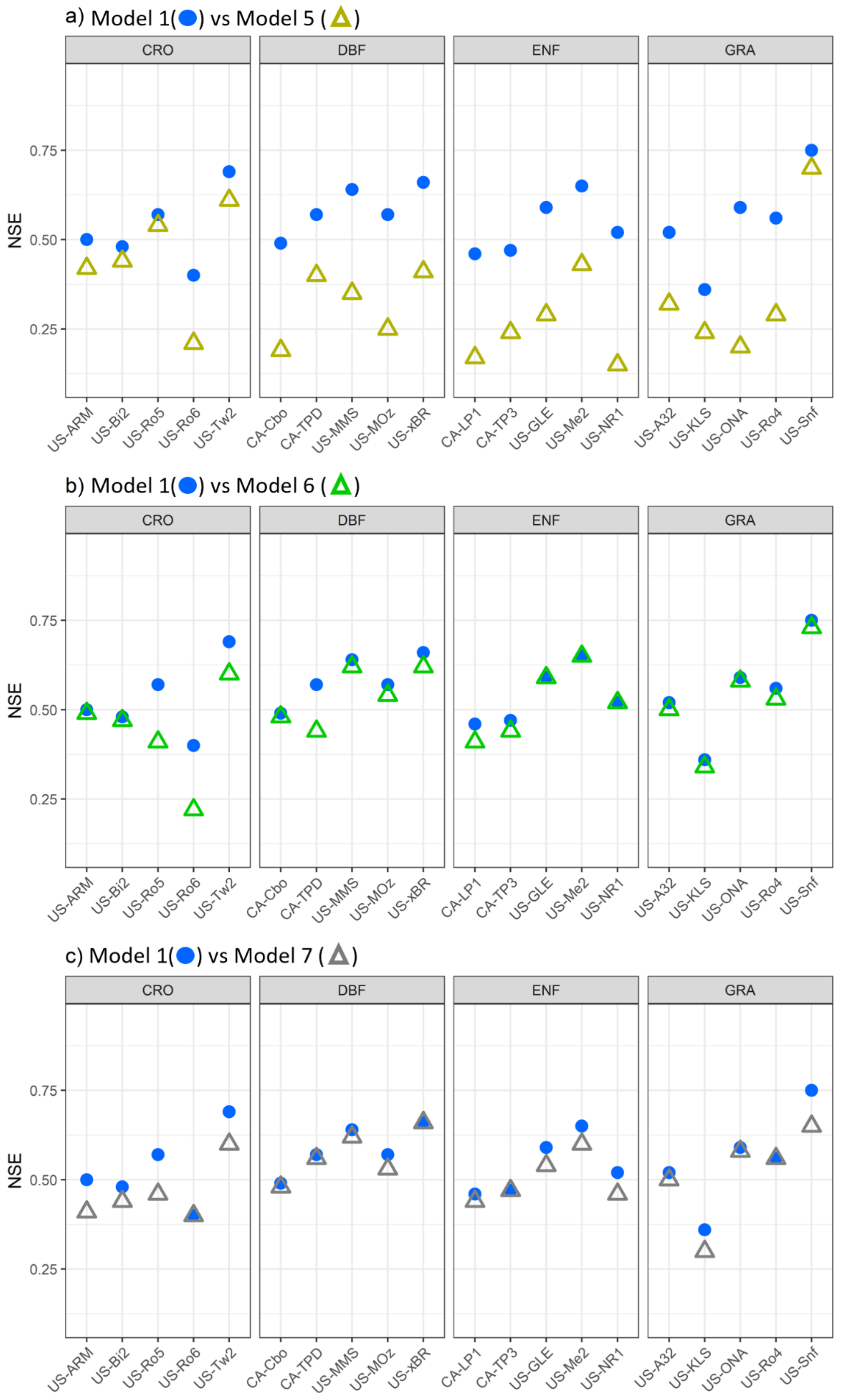


Figure S2 Comparison of G_s simulation accuracies for test data, a) model 1 (benchmark) vs model 5, b) model 1 vs model 6, c) model 1 vs model 7.

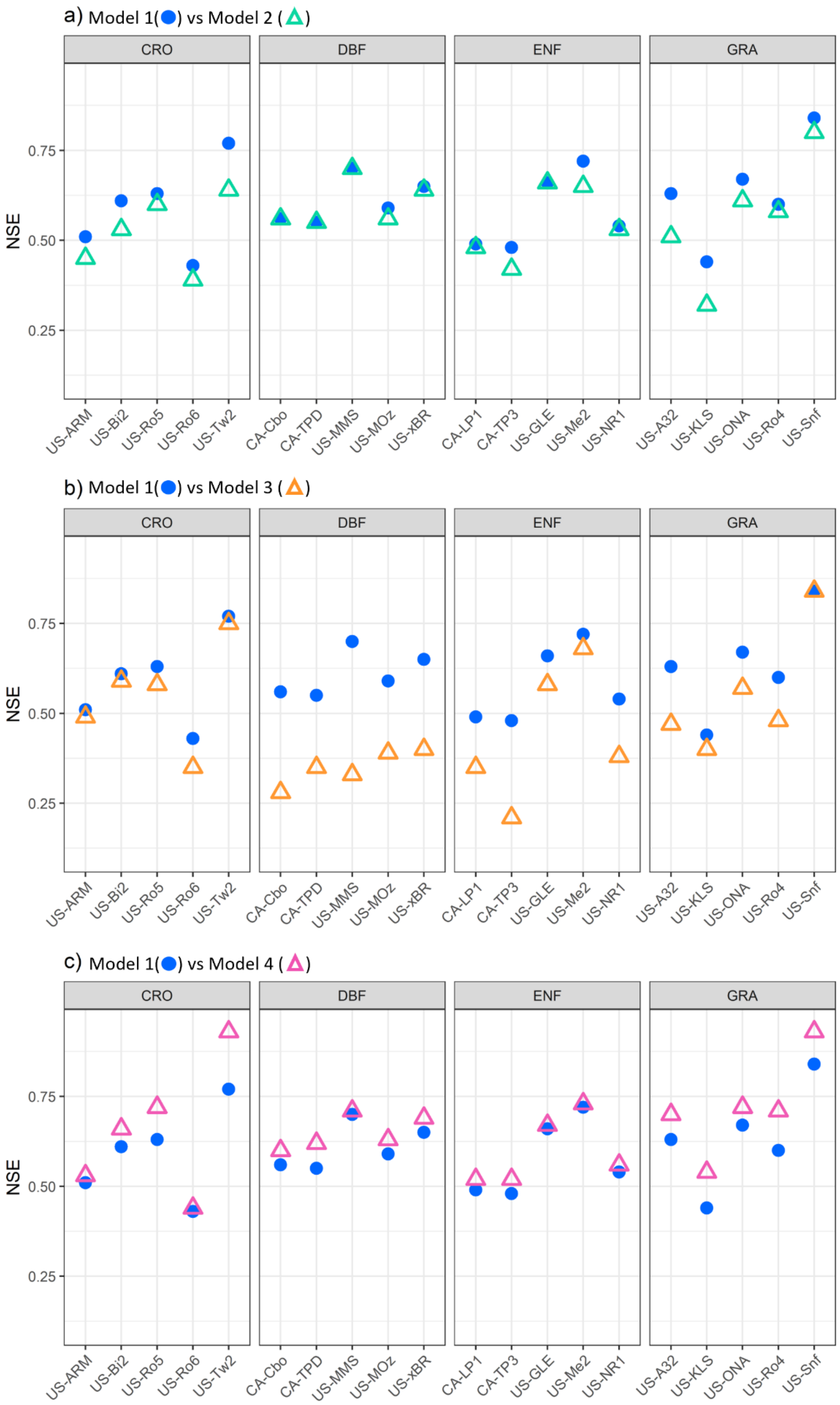


Figure S3 The comparison of G_s simulation accuracies for training data, a) model 1 (benchmark) vs model 2, b) model 1 vs model 3, c) model 1 vs model 4.

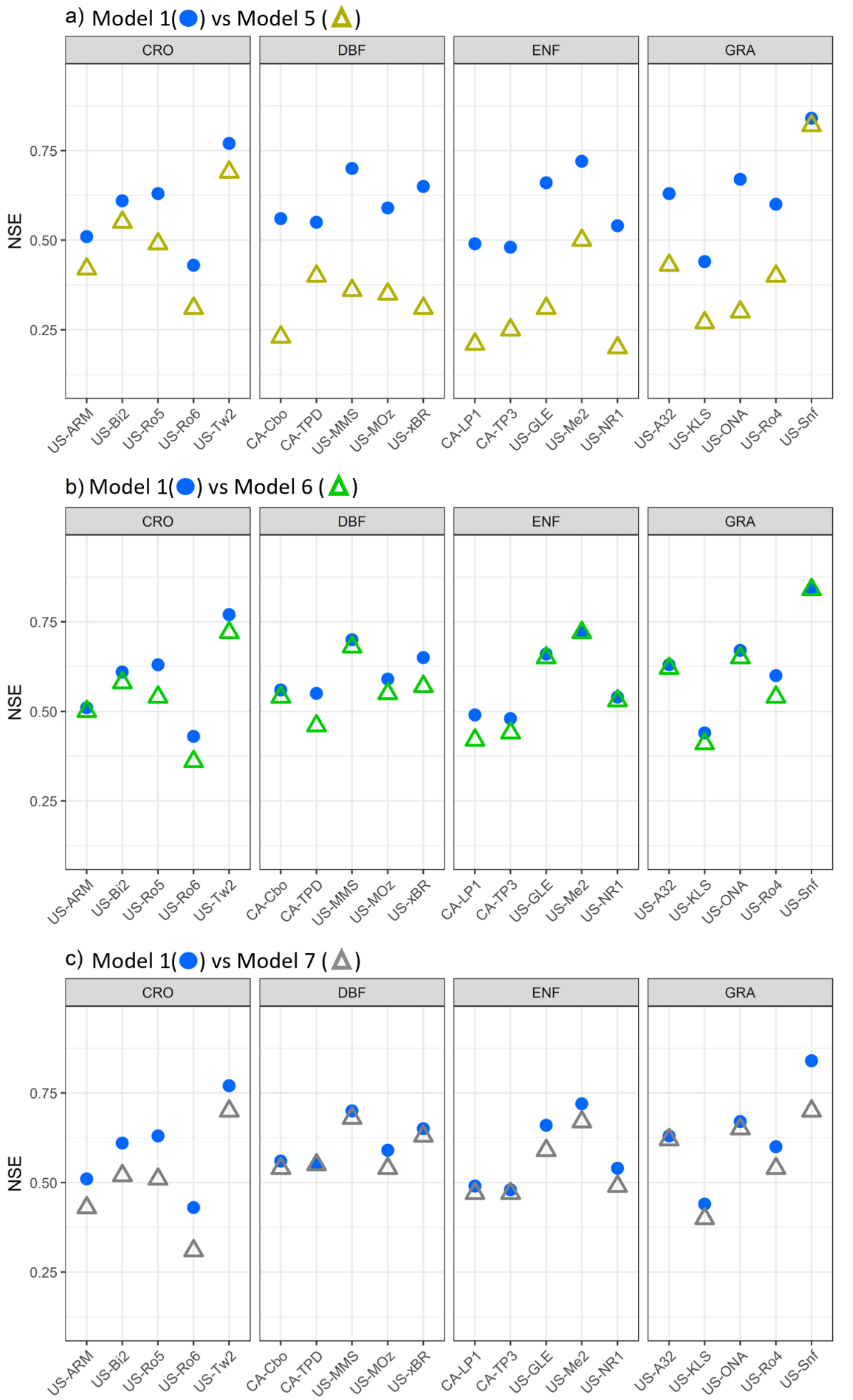


Figure S4 The comparison of G_s simulation accuracies for training data, a) model 1 (benchmark) vs model 5, b) model 1 vs model 6, c) model 1 vs model 7.

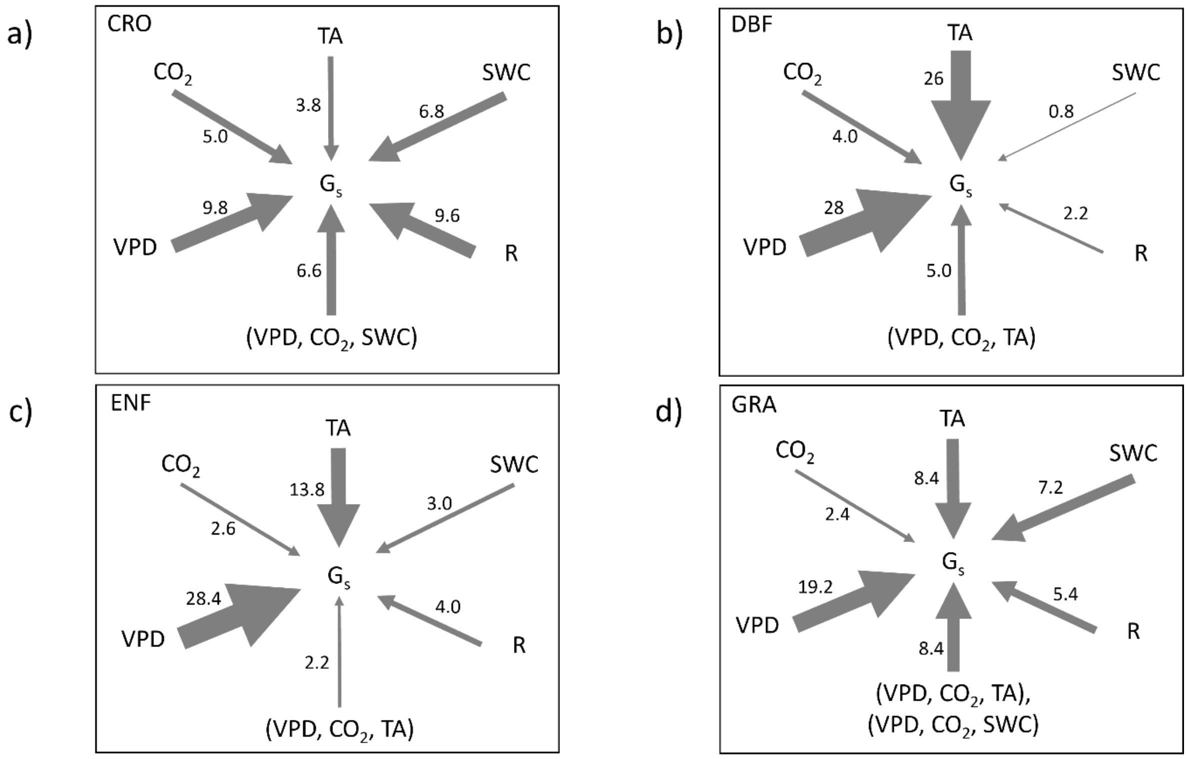


Figure S5 Schematic path analysis for the training data showing the effect of each environmental variable in G_s simulation. The path values and arrow thicknesses are the average NSE value changes compared to model 1 (benchmark) for each environmental variable (%) and for all five flux net sites in vegetation type: a) crop (CRO), b) deciduous broad-leaf forest (DBF), c) evergreen needle-leaf forest (ENF), and d) grass (GRA).

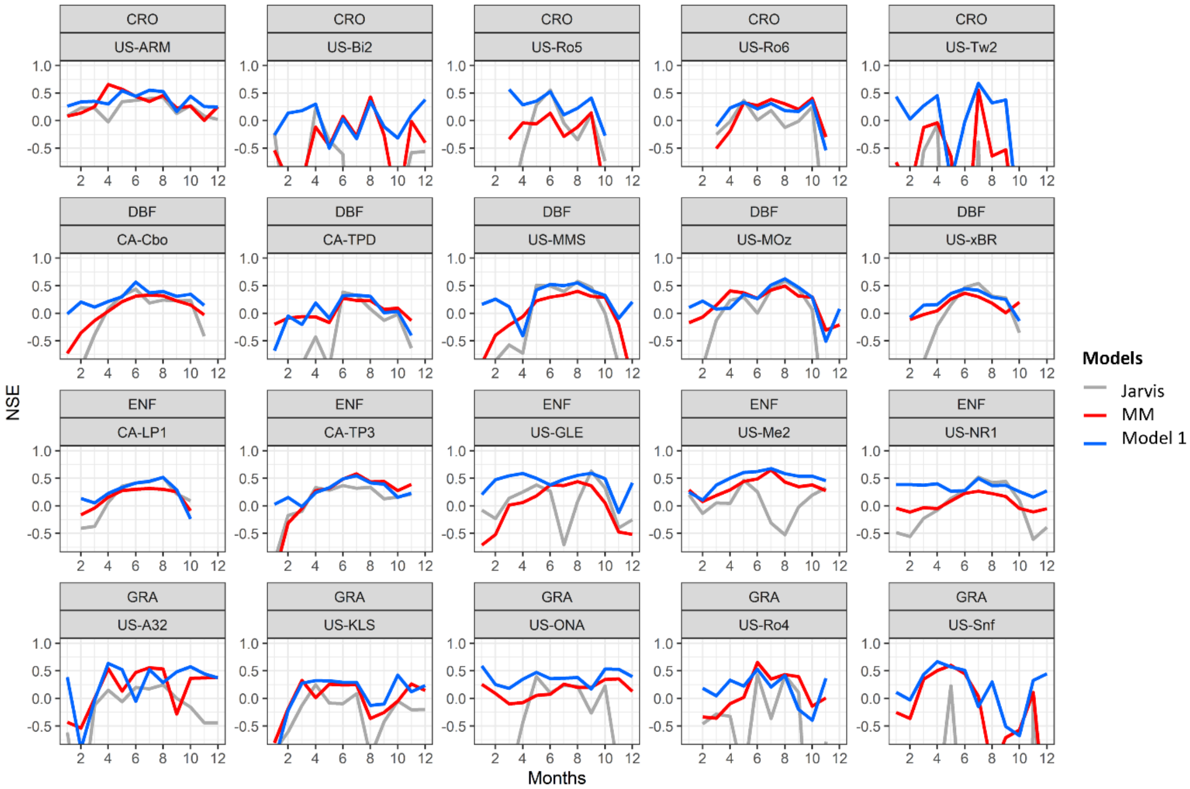


Figure S6 The comparison of monthly G_s simulation accuracies in Model 1 (benchmark) contains all direct effects of environmental variables with Jarvis and MM models.

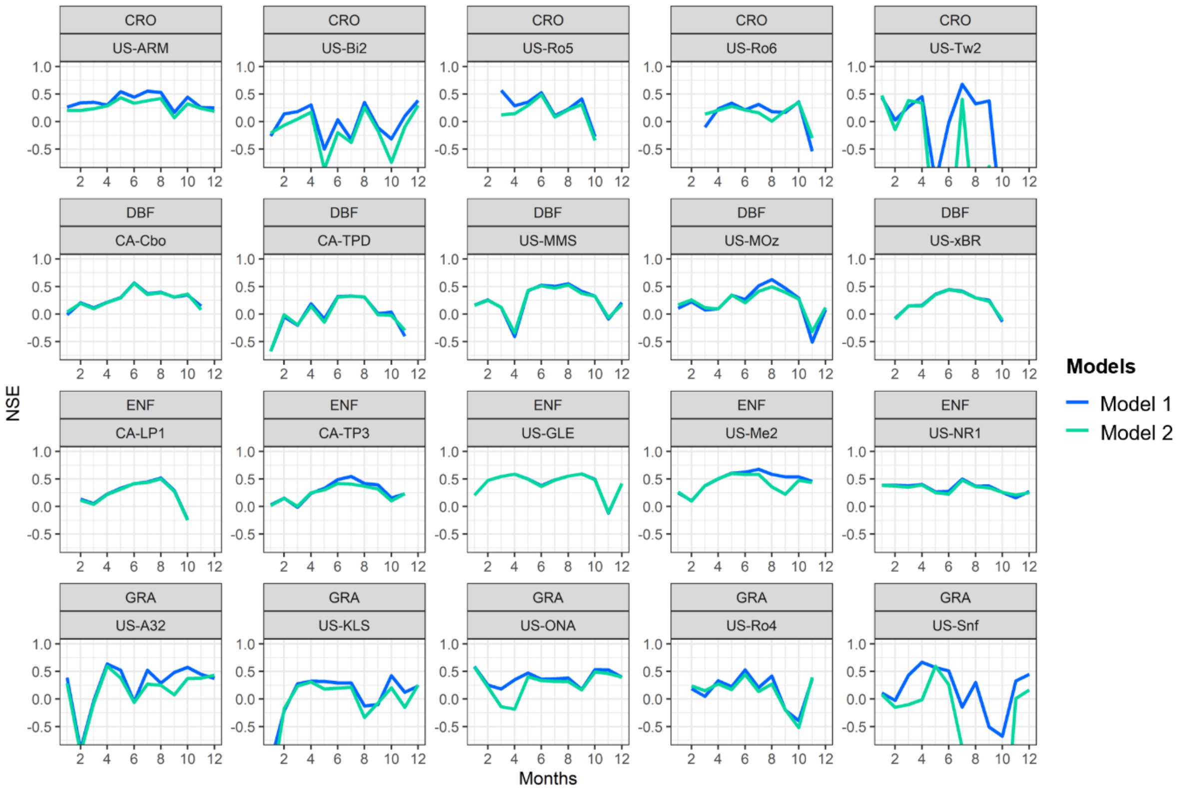


Figure S7 The comparison of monthly G_s simulation accuracies in Model 1 (benchmark) contains all direct effects of environmental variables with Model 2 (benchmark without SWC).

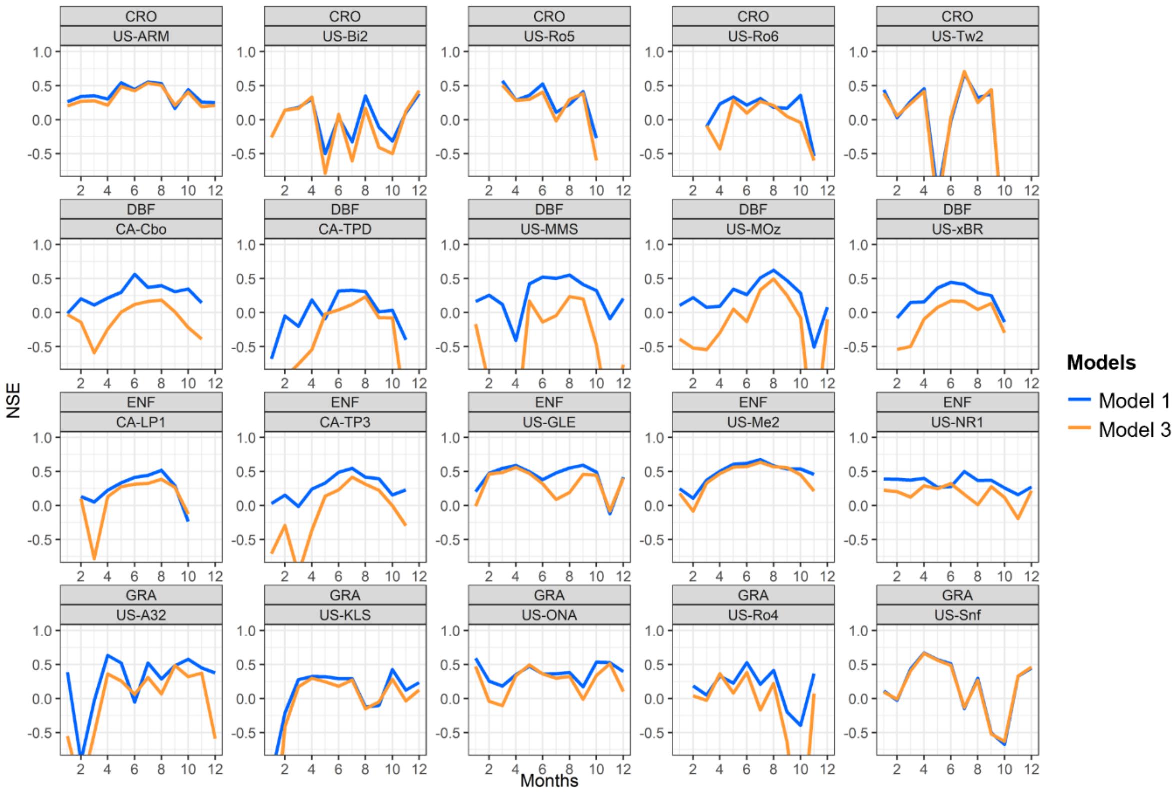


Figure S8 The comparison of monthly G_s simulation accuracies in Model 1 (benchmark) contains all direct effects of environmental variables with Model 3 (benchmark without TA).

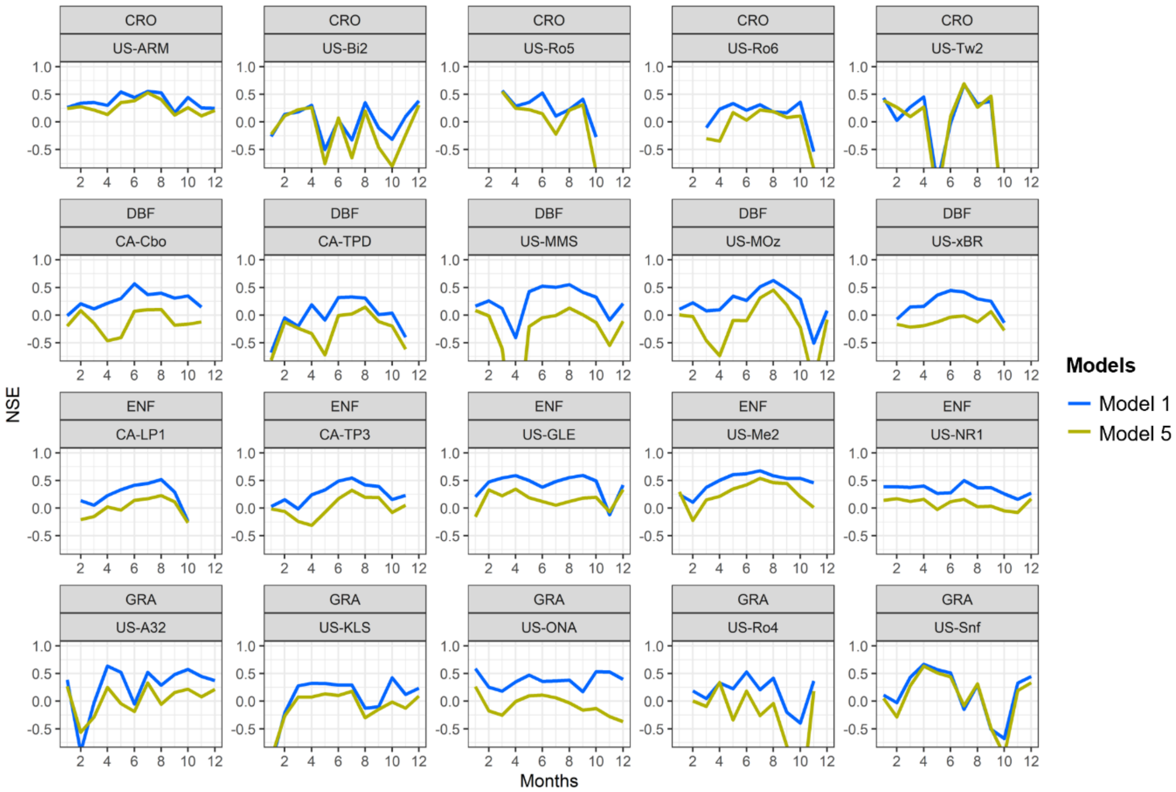


Figure S9 The comparison of monthly G_s simulation accuracies in Model 1 (benchmark) contains all direct effects of environmental variables with Model 5 (benchmark without VPD).

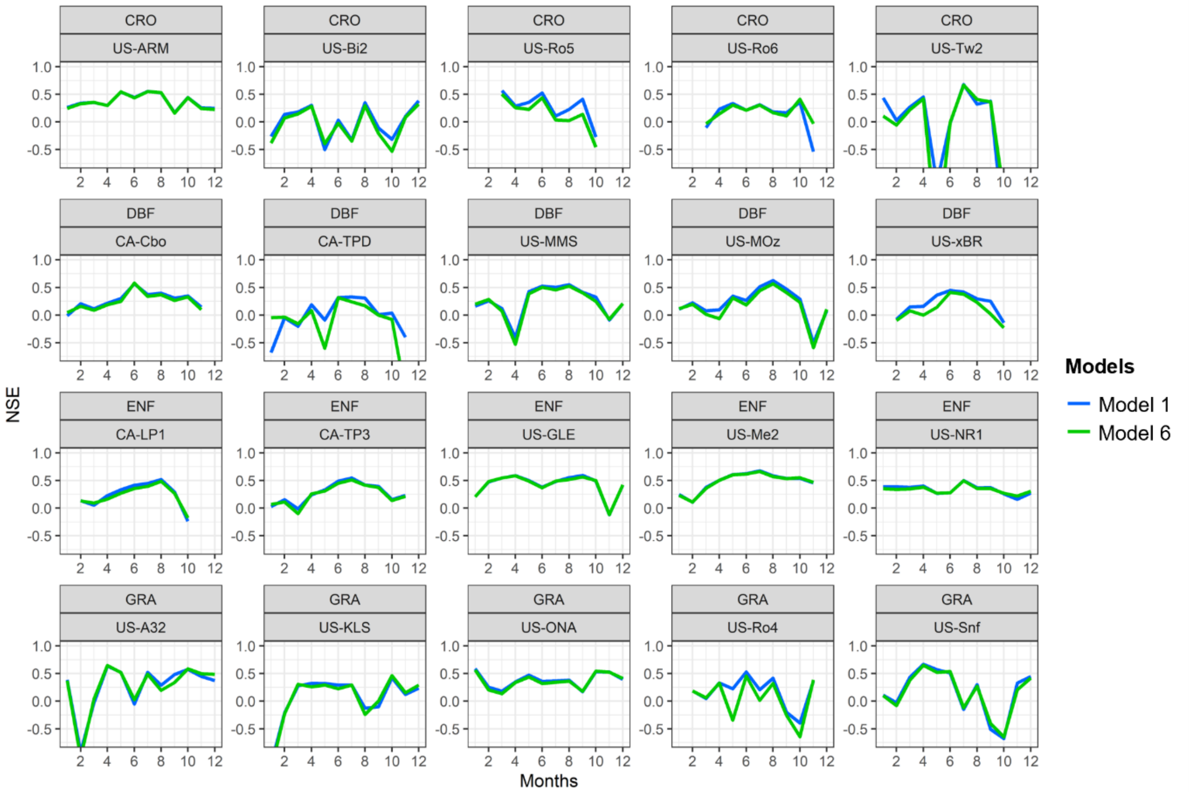


Figure S10 The comparison of monthly G_s simulation accuracies in Model 1 (benchmark) contains all direct effects of environmental variables with Model 6 (benchmark without CO_2).

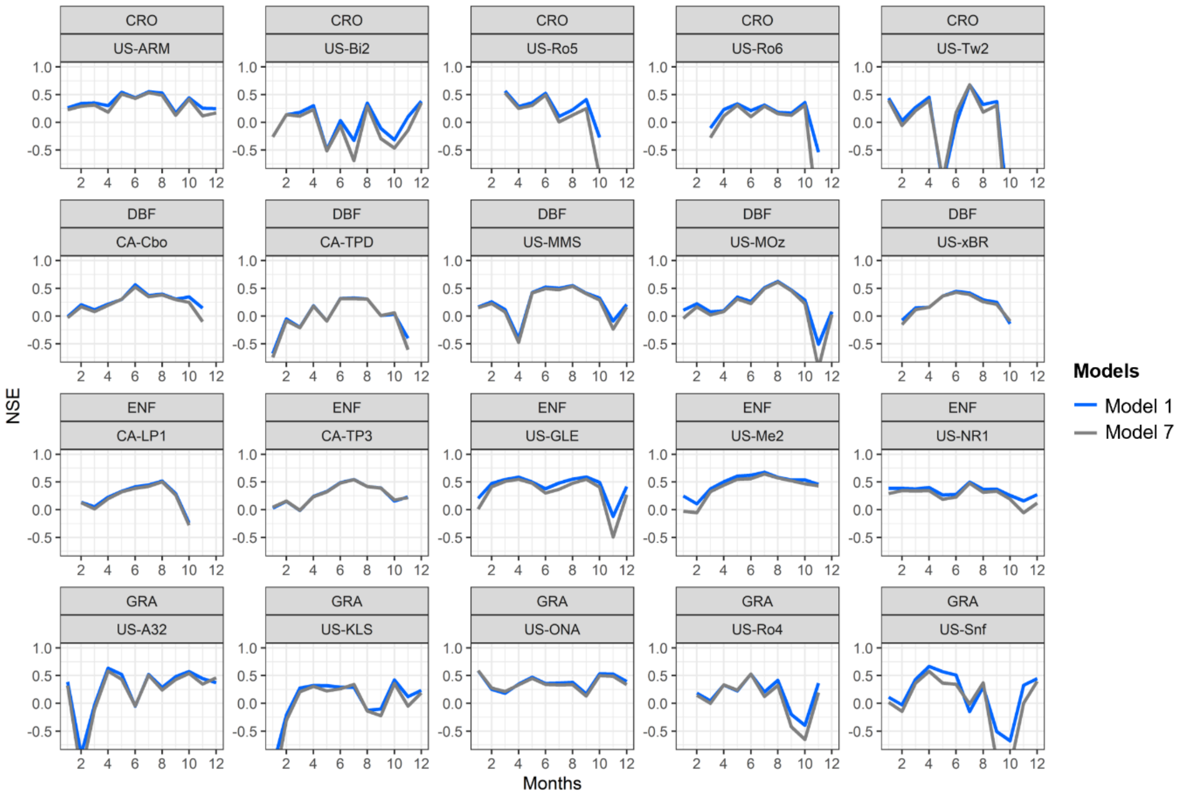


Figure S11 The comparison of monthly G_s simulation accuracies in Model 1 (benchmark) contains all direct effects of environmental variables with Model 7 (benchmark without R).

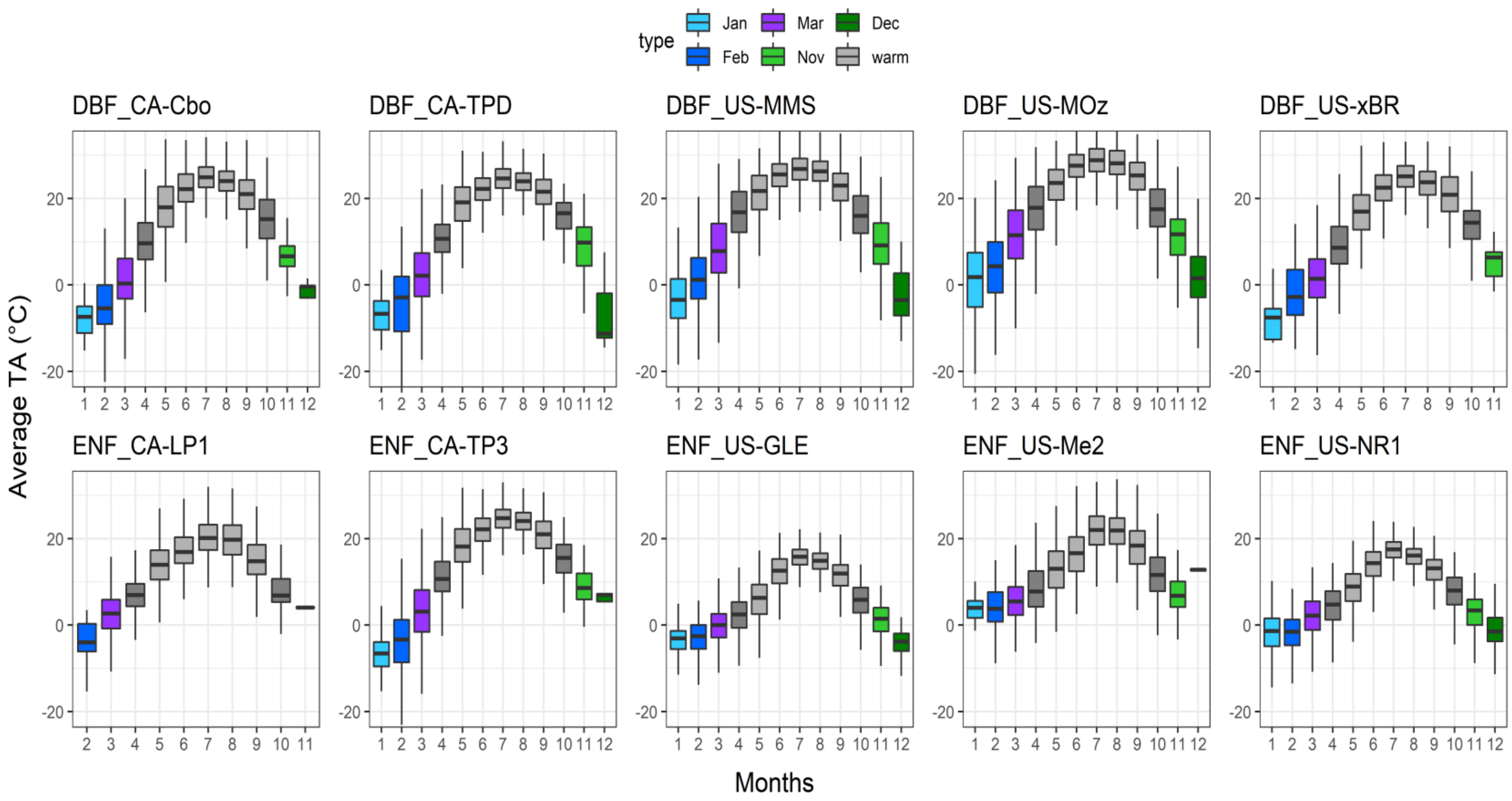


Figure S12 Monthly average temperature for DBF and ENF with more sensitivity to TA at the beginning (Jan-Mar) and the end (Nov-Dec) of the growth period.

3.7 Reference

Abramowitz, G., Pitman, A., Gupta, H., Kowalczyk, E., & Wang, Y. (2007). Systematic Bias in Land Surface Models. *Journal of Hydrometeorology*, *8*(5), 989-1001. https://doi.org/https://doi.org/10.1175/jhm628.1

Anderegg, W. R. L., Berry, J. A., Smith, D. D., Sperry, J. S., Anderegg, L. D. L., & Field, C. B. (2012). The roles of hydraulic and carbon stress in a widespread climate-induced forest die-off. *Proceedings of the National Academy of Sciences*, *109*(1), 233-237. https://doi.org/https://doi.org/10.1073/pnas.1107891109

Arain, M. A., Xu, B., Brodeur, J. J., Khomik, M., Peichl, M., Beamesderfer, E., Restrepo-Couple, N., & Thorne, R. (2022). Heat and drought impact on carbon exchange in an age-sequence of temperate pine forests. *Ecological Processes*, *11*(1), 7. https://doi.org/https://doi.org/10.1186/s13717-021-00349-7

Bai, Y., Li, X., Zhou, S., Yang, X., Yu, K., Wang, M., Liu, S., Wang, P., Wu, X., Wang, X., Zhang, C., Shi, F., Wang, Y., & Wu, Y. (2019). Quantifying plant transpiration and canopy conductance using eddy flux data: An underlying water use efficiency method. *Agricultural and Forest Meteorology*, *271*, 375-384. https://doi.org/https://doi.org/10.1016/j.agrformet.2019.02.035

Ball, J. T., Woodrow, I. E., Berry, J. A. . (1987). A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. *In Progress in photosynthesis research*, 221-224. https://doi.org/https://doi.org/10.1007/978-94-017-0519-6_48

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.-P., & Delignette-Muller, M.-L. (2015). A Toolbox for Nonlinear Regression in R: The Package nlstools. *Journal of Statistical Software*, *66*(5), 1 - 21. https://doi.org/https://doi.org/10.18637/jss.v066.i05

Beamesderfer, E. R., Arain, M. A., Khomik, M., & Brodeur, J. J. (2020). The Impact of Seasonal and Annual Climate Variations on the Carbon Uptake Capacity of a Deciduous Forest Within the Great Lakes Region of Canada. *Journal of Geophysical Research: Biogeosciences*, *125*(9), e2019JG005389. https://doi.org/https://doi.org/10.1029/2019JG005389

Birami, B., Nägele, T., Gattmann, M., Preisler, Y., Gast, A., Arneth, A., & Ruehr, N. K. (2020). Hot drought reduces the effects of elevated CO2 on tree water-use efficiency and carbon metabolism. *New Phytologist*, *226*(6), 1607-1621. https://doi.org/https://doi.org/10.1111/nph.16471

Breshears, D., Adams, H., Eamus, D., McDowell, N., Law, D., Will, R., Williams, A., & Zou, C. (2013). The critical amplifying role of increasing atmospheric moisture demand on tree mortality and associated regional die-off [Opinion]. *Frontiers in Plant Science*, *4*. https://doi.org/https://doi.org/10.3389/fpls.2013.00266

Brown, M. G., Black, T. A., Nesic, Z., Fredeen, A. L., Foord, V. N., Spittlehouse, D. L., Bowler, R., Burton, P. J., Trofymow, J. A., Grant, N. J., & Lessard, D. (2012). The carbon balance of two lodgepole pine stands recovering from mountain pine beetle attack in British Columbia. *Agricultural and Forest Meteorology*, *153*, 82-93. https://doi.org/https://doi.org/10.1016/j.agrformet.2011.07.010

Burns, S. P., Blanken, P. D., Turnipseed, A. A., Hu, J., & Monson, R. K. (2015). The influence of warm-season precipitation on the diel cycle of the surface energy balance and carbon dioxide at a Colorado subalpine forest site. *Biogeosciences*, *12*(23), 7349-7377. https://doi.org/https://doi.org/10.5194/bg-12-7349-2015

Chitsaz, N., Guan, H., Shanafield, M., & Batelaan, O. (2023). Evaluating CO2 effects on semi-empirical and empirical stomatal conductance simulation in land surface models. *Journal of Hydrology*, *620*, 129385. https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.129385

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., Arain, M. A., Arkebauer, T. J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Bracho, R., Brown, S., . . . Zona, D. (2021). Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites.

Agricultural and Forest Meteorology, *301-302*, 108350. https://doi.org/https://doi.org/10.1016/j.agrformet.2021.108350

D'Arrigo, R. D., Kaufmann, R. K., Davi, N., Jacoby, G. C., Laskowski, C., Myneni, R. B., & Cherubini, P. (2004). Thresholds for warming-induced growth decline at elevational tree line in the Yukon Territory, Canada. *Global Biogeochemical Cycles*, *18*(3). https://doi.org/https://doi.org/10.1029/2004GB002249

De Kauwe, M. G., Medlyn, B. E., & Tissue, D. T. (2021). To what extent can rising [CO2] ameliorate plant drought stress? *New Phytologist.*, *231*(6), 2118-2124. https://doi.org/https://doi.org/10.1111/nph.17540

Dombrowski, O., Brogi, C., Hendricks Franssen, H. J., Zanotelli, D., & Bogena, H. (2022). CLM5-FruitTree: a new sub-model for deciduous fruit trees in the Community Land Model (CLM5). *Geosci. Model Dev.*, *15*(13), 5167-5193. https://doi.org/https://10.5194/gmd-15-5167-2022

Dou, X., & Yang, Y. (2018). Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Computers and Electronics in Agriculture*, *148*, 95-106. https://doi.org/https://doi.org/10.1016/j.compag.2018.03.010

Ershadi, A., McCabe, M. F., Evans, J. P., Chaney, N. W., & Wood, E. F. (2014). Multi-site evaluation of terrestrial evaporation models using FLUXNET data. *Agricultural and Forest Meteorology*, *187*, 46-61. https://doi.org/https://doi.org/10.1016/j.agrformet.2013.11.008

Ershadi, A., McCabe, M. F., Evans, J. P., & Wood, E. F. (2015). Impact of model structure and parameterization on Penman–Monteith type evaporation models. *Journal of Hydrology*, *525*, 521-535. https://doi.org/https://doi.org/10.1016/j.jhydrol.2015.04.008

Fer, I., & Dietze, M. (2018). *Bartlett Forest (BART)*. https://sites.nicholas.duke.edu/clarklab/projects/forecasting-community-dynamics-the-mast-system/bartlett-forest-bart/#:~:text=Forest%20stands%20are%2090%2D150,Betula%20alleghaniensis%20(yellow%20birch. https://sites.nicholas.duke.edu/clarklab/projects/forecasting-community-dynamics-the-mast-system/bartlett-forest-bart/#:~:text=Forest%20stands%20are%2090%2D150,Betula%20alleghaniensis%20(yellow%20birch).

Frank, J. M., Massman, W. J., Ewers, B. E., Huckaby, L. S., & Negrón, J. F. (2014). Ecosystem CO2/H2O fluxes are explained by hydraulically limited gas exchange during tree mortality from spruce bark beetles. *Journal of Geophysical Research: Biogeosciences*, *119*(6), 1195-1215. https://doi.org/https://doi.org/10.1002/2013JG002597

Gattmann, M., Birami, B., Nadal Sala, D., & Ruehr, N. K. (2021). Dying by drying: Timing of physiological stress thresholds related to tree death is not significantly altered by highly elevated CO2. *Plant, cell & environment*, *44*(2), 356-370. https://doi.org/https://doi.org/10.1111/pce.13937

Geyer, C. J., & Johnson, L. T. (2020). *mcmc: Markov Chain Monte Carlo. R package version 0.9-7*. https://CRAN.R-project.org/package=mcmc

Green, J. K., Berry, J., Ciais, P., Zhang, Y., & Gentine, P. (2020). Amazon rainforest photosynthesis increases in response to atmospheric dryness. *Science Advances*, *6*(47), eabb7232. https://doi.org/https://doi.org/10.1126/sciadv.abb7232

Griebel, A., Bennett, L. T., Metzen, D., Pendall, E., Lane, P. N. J., & Arndt, S. K. (2020). Trading Water for Carbon: Maintaining Photosynthesis at the Cost of Increased Water Loss During High Temperatures in a Temperate Forest. *Journal of Geophysical Research: Biogeosciences*, *125*(1), e2019JG005239. https://doi.org/https://doi.org/10.1029/2019JG005239

Griffis, T. J., Lee, X., Baker, J. M., Billmark, K., Schultz, N., Erickson, M., Zhang, X., Fassbinder, J., Xiao, W., & Hu, N. (2011). Oxygen isotope composition of evapotranspiration and its relation to C4 photosynthetic discrimination. *Journal of Geophysical Research: Biogeosciences*, *116*(G1). https://doi.org/https://doi.org/10.1029/2010JG001514

Gu, L., Fuentes, J. D., Shugart, H. H., Staebler, R. M., & Black, T. A. (1999). Responses of net ecosystem exchanges of carbon dioxide to changes in cloudiness: Results from two North American deciduous forests. *Journal of Geophysical Research: Atmospheres*, *104*(D24), 31421-31434. https://doi.org/https://doi.org/10.1029/1999JD901068

Hartig, F., Minunno, F., & Paul, S. (2019). *BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics. R package version 0.1.7.* In https://CRAN.R-project.org/package=BayesianTools
Hastie, T., Robert, T., & JH, F. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). New York: springer.
Hou, M., Tian, F., Ortega-Farias, S., Riveros-Burgos, C., Zhang, T., & Lin, A. (2021). Estimation of crop transpiration and its scale effect based on ground and UAV thermal infrared remote sensing images. *European Journal of Agronomy*, *131*, 126389. https://doi.org/https://doi.org/10.1016/j.eja.2021.126389
Jarvis, P. G., Monteith, J. L., & Weatherley, P. E. (1976). The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Biological Sciences*, *273*(927), 593-610. https://doi.org/https://doi.org/10.1098/rstb.1976.0035
Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., & Reichstein, M. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific Data*, *6*(1), 74. https://doi.org/https://doi.org/10.1038/s41597-019-0076-8
Kimm, H., Guan, K., Gentine, P., Wu, J., Bernacchi, C. J., Sulman, B. N., Griffis, T. J., & Lin, C. (2020). Redefining droughts for the U.S. Corn Belt: The dominant role of atmospheric vapor pressure deficit over soil moisture in regulating stomatal behavior of Maize and Soybean. *Agricultural and Forest Meteorology*, *287*, 107930. https://doi.org/https://doi.org/10.1016/j.agrformet.2020.107930
Knox, S. H., Sturtevant, C., Matthes, J. H., Koteen, L., Verfaillie, J., & Baldocchi, D. (2015). Agricultural peatland restoration: effects of land-use change on greenhouse gas (CO2 and CH4) fluxes in the Sacramento-San Joaquin Delta. *Global Change Biology*, *21*(2), 750-765. https://doi.org/https://doi.org/10.1111/gcb.12745
Koppa, A., Rains, D., Hulsman, P., Poyatos, R., & Miralles, D. G. (2022). A deep learning-based hybrid model of global terrestrial evaporation. *Nature Communications*, *13*(1), 1912. https://doi.org/https://doi.org/10.1038/s41467-022-29543-7
Koutsoyiannis, D. (2020). Revisiting the global hydrological cycle: is it intensifying? *Hydrol. Earth Syst. Sci.*, *24*(8), 3899-3932. https://doi.org/https://10.5194/hess-24-3899-2020
Kuhn, M. (2021). *caret: Classification and Regression Training. R package version 6.0-88.* https://CRAN.R-project.org/package=caret
Kwon, H., Law, B. E., Thomas, C. K., & Johnson, B. G. (2018). The influence of hydrological variability on inherent water use efficiency in forests of contrasting composition, age, and precipitation regimes in the Pacific Northwest. *Agricultural and Forest Meteorology*, *249*, 488-500. https://doi.org/https://doi.org/10.1016/j.agrformet.2017.08.006
Lee, Y.-G., Oh, J.-Y., Kim, D., & Kim, G. (2023). SHAP Value-Based Feature Importance Analysis for Short-Term Load Forecasting. *Journal of Electrical Engineering & Technology*, *18*(1), 579-588. https://doi.org/10.1007/s42835-022-01161-9
Lei, H., Yang, D., & Huang, M. (2014). Impacts of climate change and vegetation dynamics on runoff in the mountainous region of the Haihe River basin in the past five decades. *Journal of Hydrology*, *511*, 786-799. https://doi.org/https://doi.org/10.1016/j.jhydrol.2014.02.029
Leuning, R. (1995). A critical appraisal of a combined stomatal-photosynthesis model for C3 plants. *Plant, cell & environment*, *18*(4), 339-355. https://doi.org/https://doi.org/10.1111/j.1365-3040.1995.tb00370.x
Li, X., Kang, S., Niu, J., Huo, Z., & Liu, J. (2019). Improving the representation of stomatal responses to CO2 within the Penman–Monteith model to better estimate evapotranspiration responses to climate change. *Journal of Hydrology*, *572*, 692-705. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.03.029
Liao, D., Niu, J., Kang, S., Singh, S. K., & Du, T. (2021). Effects of elevated CO2 on the evapotranspiration over the agricultural land in Northwest China. *Journal of Hydrology*, *593*, 125858. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125858
Lin, C., Gentine, P., Huang, Y., Guan, K., Kimm, H., & Zhou, S. (2018). Diel ecosystem conductance response to vapor pressure deficit is suboptimal and independent of soil moisture.

Agricultural and Forest Meteorology, *250-251*, 24-34. https://doi.org/https://doi.org/10.1016/j.agrformet.2017.12.078

Liu, D., & Mishra, A. K. (2017). Performance of AMSR_E soil moisture data assimilation in CLM4.5 model for monitoring hydrologic fluxes at global scale. *Journal of Hydrology*, *547*, 67-79. https://doi.org/https://doi.org/10.1016/j.jhydrol.2017.01.036

Lu, J., Sun, G., McNulty, S. G., & Amatya, D. M. (2003). Modeling actual evapotranspiration from forested watersheds across the southeastern United States1. *Journal of the American Water Resources Association*, *39*(4), 887-896. https://doi.org/https://doi.org/10.1111/j.1752-1688.2003.tb04413.x

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin Jordan, M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56-67. https://doi.org/https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions* Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA.

Mardian, J., Champagne, C., Bonsal, B., & Berg, A. (2023). A Machine Learning Framework for Predicting and Understanding the Canadian Drought Monitor. *Water Resources Research*, *59*(8), e2022WR033847. https://doi.org/https://doi.org/10.1029/2022WR033847

Mathias, J. M., & Thomas, R. B. (2021). Global tree intrinsic water use efficiency is enhanced by increased atmospheric CO2 and modulated by climate and plant functional types. *Proceedings of the National Academy of Sciences*, *118*(7). https://doi.org/https://doi.org/10.1073/pnas.2014286118

Mayer, M. (2023). *shapviz: SHAP Visualizations. R package version 0.9.1*. https://CRAN.R-project.org/package=shapviz

Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V. M., Crous, K. Y., De Angelis, P., Freeman, M., & Wingate, L. (2011). Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology*, *17*(6), 2134-2144. https://doi.org/https://doi.org/10.1111/j.1365-2486.2010.02375.x

Monteith, J. L. (1965). Evaporation and environment. *Symp Soc Exp Biol*, *19*, 205-234.

Morison, J. I. L., & Lawlor, D. W. (1999). Interactions between increasing CO2 concentration and temperature on plant growth. *Plant, cell & environment*, *22*(6), 659-682. https://doi.org/https://doi.org/10.1046/j.1365-3040.1999.00443.x

Nadal-Sala, D., Medlyn, B. E., Ruehr, N. K., Barton, C. V. M., Ellsworth, D. S., Gracia, C., Tissue, D. T., Tjoelker, M. G., & Sabaté, S. (2021). Increasing aridity will not offset CO2 fertilization in fast-growing eucalypts with access to deep soil water. *Global Change Biology*, *27*(12), 2970-2990. https://doi.org/https://doi.org/10.1111/gcb.15590

Nelson, J. A., Pérez-Priego, O., Zhou, S., Poyatos, R., Zhang, Y., Blanken, P. D., Gimeno, T. E., Wohlfahrt, G., Desai, A. R., Gioli, B., Limousin, J.-M., Bonal, D., Paul-Limoges, E., Scott, R. L., Varlagin, A., Fuchs, K., Montagnani, L., Wolf, S., Delpierre, N., . . . Jung, M. (2020). Ecosystem transpiration and evaporation: Insights from three water flux partitioning methods across FLUXNET sites. *Global Change Biology*, *26*(12), 6916-6930. https://doi.org/https://doi.org/10.1111/gcb.15314

Nguyen, M. N., Hao, Y., Baik, J., & Choi, M. (2021). Partitioning evapotranspiration based on the total ecosystem conductance fractions of soil, interception, and canopy in different biomes. *Journal of Hydrology*, *603*, 126970. https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126970

Nie, C., Huang, Y., Zhang, S., Yang, Y., Zhou, S., Lin, C., & Wang, G. (2021). Effects of soil water content on forest ecosystem water use efficiency through changes in transpiration/evapotranspiration ratio. *Agricultural and Forest Meteorology*, *308-309*, 108605. https://doi.org/https://doi.org/10.1016/j.agrformet.2021.108605

Novick, K. A., Ficklin, D. L., Stoy, P. C., Williams, C. A., Bohrer, G., Oishi, A. C., Papuga, S. A., Blanken, P. D., Noormets, A., Sulman, B. N., Scott, R. L., Wang, L., & Phillips, R. P. (2016). The increasing importance of atmospheric demand for ecosystem water and carbon fluxes.

Nature Climate Change, *6*(11), 1023-1027. https://doi.org/https://doi.org/10.1038/nclimate3114

Oliver, R. J., Mercado, L. M., Clark, D. B., Huntingford, C., Taylor, C. M., Vidale, P. L., McGuire, P. C., Todt, M., Folwell, S., Shamsudheen Semeena, V., & Medlyn, B. E. (2022). Improved representation of plant physiology in the JULES-vn5.6 land surface model: photosynthesis, stomatal conductance and thermal acclimation. *Geosci. Model Dev.*, *15*(14), 5567-5592. https://doi.org/10.5194/gmd-15-5567-2022

Page, G. F. M., Liénard, J. F., Pruett, M. J., & Moffett, K. B. (2018). Spatiotemporal dynamics of leaf transpiration quantified with time-series thermal imaging. *Agricultural and Forest Meteorology*, *256-257*, 304-314. https://doi.org/https://doi.org/10.1016/j.agrformet.2018.02.023

Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Ottlé, C., Poulter, B., Zaehle, S., & Running, S. W. (2020). Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling. *Hydrol. Earth Syst. Sci.*, *24*(3), 1485-1509. https://doi.org/https://doi.org/10.5194/hess-24-1485-2020

Park Williams, A., Allen, C. D., Macalady, A. K., Griffin, D., Woodhouse, C. A., Meko, D. M., Swetnam, T. W., Rauscher, S. A., Seager, R., Grissino-Mayer, H. D., Dean, J. S., Cook, E. R., Gangodagamage, C., Cai, M., & McDowell, N. G. (2013). Temperature as a potent driver of regional forest drought stress and tree mortality. *Nature Climate Change*, *3*(3), 292-297. https://doi.org/https://doi.org/10.1038/nclimate1693

Penman, H. L. (1948). Natural evaporation from open water, hare soil and grass. *Proc R Soc Lond A Math Phys Sci*, *193*(1032), 120-145. https://doi.org/https://doi.org/10.1098/rspa.1948.0037

Polhamus, A., Fisher, J. B., & Tu, K. P. (2013). What controls the error structure in evapotranspiration models? *Agricultural and Forest Meteorology*, *169*, 12-24. https://doi.org/https://doi.org/10.1016/j.agrformet.2012.10.002

Purcell, C., Batke, S. P., Yiotis, C., Caballero, R., Soh, W. K., Murray, M., & McElwain, J. C. (2018). Increasing stomatal conductance in response to rising atmospheric CO2. *Annals of Botany*, *121*(6), 1137-1149. https://doi.org/https://doi.org/10.1093/aob/mcy023

Qi, Y., Zhang, Q., Hu, S., Wang, R., Wang, H., Zhang, K., Zhao, H., Zhao, F., Chen, F., Yang, Y., Tang, G., & Hu, Y. (2023). Applicability of stomatal conductance models comparison for persistent water stress processes of spring maize in water resources limited environmental zone. *Agricultural Water Management*, *277*, 108090. https://doi.org/https://doi.org/10.1016/j.agwat.2022.108090

Raz-Yaseef, N., Billesbach, D. P., Fischer, M. L., Biraud, S. C., Gunter, S. A., Bradford, J. A., & Torn, M. S. (2015). Vulnerability of crops and native grasses to summer drying in the U.S. Southern Great Plains. *Agriculture, Ecosystems & Environment*, *213*, 209-218. https://doi.org/https://doi.org/10.1016/j.agee.2015.07.021

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195-204. https://doi.org/https://doi.org/10.1038/s41586-019-0912-1

Rey-Sanchez, C., Wharton, S., Vilà-Guerau de Arellano, J., Paw U, K. T., Hemes, K. S., Fuentes, J. D., Osuna, J., Szutu, D., Ribeiro, J. V., Verfaillie, J., & Baldocchi, D. (2021). Evaluation of Atmospheric Boundary Layer Height From Wind Profiling Radar and Slab Models and Its Responses to Seasonality of Land Cover, Subsidence, and Advection. *Journal of Geophysical Research: Atmospheres*, *126*(7), e2020JD033775. https://doi.org/https://doi.org/10.1029/2020JD033775

Roman, D. T., Novick, K. A., Brzostek, E. R., Dragoni, D., Rahman, F., & Phillips, R. P. (2015). The role of isohydric and anisohydric species in determining ecosystem-scale response to severe drought. *Oecologia*, *179*(3), 641-654. https://doi.org/https://doi.org/10.1007/s00442-015-3380-9

Saunders, A., Drew, D. M., & Brink, W. (2021). Machine learning models perform better than traditional empirical models for stomatal conductance when applied to multiple tree species across different forest biomes. *Trees, Forests and People*, *6*, 100139. https://doi.org/https://doi.org/10.1016/j.tfp.2021.100139

Shi, H., Luo, G., Hellwich, O., He, X., Xie, M., Zhang, W., Ochege, F. U., Ling, Q., Zhang, Y., Gao, R., Kurban, A., De Maeyer, P., & Van de Voorde, T. (2023). Comparing the use of all data or specific subsets for training machine learning models in hydrology: A case study of evapotranspiration prediction. *Journal of Hydrology*, *627*, 130399. https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.130399

Silveira, M. (2021). *AmeriFlux FLUXNET-1F US-ONA Florida pine flatwoods* Version Ver. 3-5). https://doi.org/10.17190/AMF/1832163

Speich, M., Dormann, C. F., & Hartig, F. (2021). Sequential Monte-Carlo algorithms for Bayesian model calibration – A review and method comparison. *Ecological Modelling*, *455*, 109608. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2021.109608

Stewart, J. B. (1988). Modelling surface conductance of pine forest. *Agricultural and Forest Meteorology*, *43*(1), 19-35. https://doi.org/https://doi.org/10.1016/0168-1923(88)90003-2

Sulman, B. N., Roman, D. T., Yi, K., Wang, L., Phillips, R. P., & Novick, K. A. (2016). High atmospheric demand for water can limit forest carbon uptake and transpiration as severely as dry soil. *Geophysical Research Letters*, *43*(18), 9686-9695. https://doi.org/https://doi.org/10.1002/2016GL069416

Thom, A. S. (1972). Momentum, mass and heat exchange of vegetation. *Quarterly Journal of the Royal Meteorological Society*, *98*, 124-134. https://doi.org/https://doi.org/10.1002/qj.49709841510

Urban, J., Ingwers, M., McGuire, M. A., & Teskey, R. O. (2017). Stomatal conductance increases with rising temperature. *Plant signaling & behavior*, *12*(8), e1356534-e1356534. https://doi.org/https://doi.org/10.1080/15592324.2017.1356534

Wang, D., Heckathorn, S. A., Wang, X., & Philpott, S. M. (2012). A meta-analysis of plant physiological and growth responses to temperature and elevated CO(2). *Oecologia*, *169*(1), 1-13. https://doi.org/https://doi.org/10.1007/s00442-011-2172-0

Wehr, R., & Saleska, S. R. (2021). Calculating canopy stomatal conductance from eddy covariance measurements, in light of the energy budget closure problem. *Biogeosciences*, *18*(1), 13-24. https://doi.org/https://10.5194/bg-18-13-2021

Wertin, T. M., McGuire, M. A., & Teskey, R. O. (2012). Effects of predicted future and current atmospheric temperature and [CO2] and high and low soil moisture on gas exchange and growth of Pinus taeda seedlings at cool and warm sites in the species range. *Tree Physiology*, *32*(7), 847-858. https://doi.org/https://doi.org/10.1093/treephys/tps051

Wood, J. D., Knapp, B. O., Muzika, R.-M., Stambaugh, M. C., & Gu, L. (2018). The importance of drought–pathogen interactions in driving oak mortality events in the Ozark Border Region. *Environmental Research Letters*, *13*(1), 015004. https://doi.org/https://doi.org/10.1088/1748-9326/aa94fa

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548-1563. https://doi.org/https://doi.org/10.1080/01621459.2016.1180986

Yang, L., Feng, Q., Zhu, M., Wang, L., Alizadeh, M. R., Adamowski, J. F., Wen, X., & Yin, Z. (2022). Variation in actual evapotranspiration and its ties to climate change and vegetation dynamics in northwest China. *Journal of Hydrology*, *607*, 127533. https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.127533

Yuan, W., Zheng, Y., Piao, S., Ciais, P., Lombardozzi, D., Wang, Y., Ryu, Y., Chen, G., Dong, W., Hu, Z., Jain, A. K., Jiang, C., Kato, E., Li, S., Lienert, S., Liu, S., Nabel, J. E. M. S., Qin, Z., Quine, T., . . . Yang, S. (2019). Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Science Advances*, *5*(8), eaax1396. https://doi.org/https://doi.org/10.1126/sciadv.aax1396

Zang, C. F., Liu, J., van der Velde, M., & Kraxner, F. (2012). Assessment of spatial and temporal patterns of green and blue water flows under natural conditions in inland river basins in Northwest China. *Hydrol. Earth Syst. Sci.*, *16*(8), 2859-2870. https://doi.org/https://doi.org/10.5194/hess-16-2859-2012

Zhan, S., Song, C., Wang, J., Sheng, Y., & Quan, J. (2019). A Global Assessment of Terrestrial Evapotranspiration Increase Due to Surface Water Area Change. *Earths Future*, *7*(3), 266-282. https://doi.org/https://10.1029/2018ef001066

Zhang, C., Yang, Y., Yang, D., & Wu, X. (2021). Multidimensional assessment of global dryland changes under future warming in climate projections. *Journal of Hydrology*, *592*, 125618. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125618

Zhang, Q., Ficklin, D. L., Manzoni, S., Wang, L., Way, D., Phillips, R. P., & Novick, K. A. (2019). Response of ecosystem intrinsic water use efficiency and gross primary productivity to rising vapor pressure deficit. *Environmental Research Letters, 14*(7), 074023. https://doi.org/10.1088/1748-9326/ab2603

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., & Qiu, G. Y. (2019). Physics-Constrained Machine Learning of Evapotranspiration. *Geophysical Research Letters*, *46*(24), 14496-14507. https://doi.org/https://doi.org/10.1029/2019GL085291

Zhou, X.-M., Tang, B.-H., Wu, H., & Li, Z.-L. (2013). Estimating net surface longwave radiation from net surface shortwave radiation for cloudy skies. *International Journal of Remote Sensing, 34*(22), 8104-8117. https://doi.org/https://10.1080/01431161.2013.832002

Chapter 4: Publication 3

Enhanced runoff simulation with improved evapotranspiration accounting for vegetation response to climate variability

The manuscript will be submitted in per-reviewed journal.

PUBLICATION 3

This section is to be completed by the student and co-authors. If there are more than four co-authors (student plus 3 others), only the three co-authors with the most significant contributions are required to sign below.

Please note: A copy of this page will be provided to the Examiners.

Full Publication Details	Enhanced runoff simulation with improved evapotranspiration accounting for vegetation response to CO_2 and climate variability
Section of thesis where publication is referred to	Chapter 4

Student's contribution to the publication

85	%	Research design
100	%	Data collection and analysis
95	%	Writing and editing

Outline your (the student's) contribution to the publication:

Conceptualization, Methodology, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

APPROVALS

By signing the section below, you confirm that the details above are an accurate record of the students contribution to the work.

Name of Co-Author 1	Huade Guan	Signed	Huade Guan	Date	14/06/2024
Name of Co-Author 2	Margaret Shanafield	Signed	Margaret Shanafield	Date	14/6/2024
Name of Co-Author 3	Okke Batelaan	Signed		Date	13/6/2024

4.1 Abstract

Rainfall-runoff simulation plays a crucial role in the prediction of high runoff events. Hydrological models which use potential evapotranspiration (PET) equations have an implicit bias in runoff simulation due to neglecting the role of vegetation responses to environmental variables such as CO_2 concentration, air temperature (TA), vapor pressure deficit (VPD), soil water content (SWC), and net radiation ®. The modification of Penman-Monteith PET (PET_{PM}) by incorporating vegetation response to environmental variables through canopy stomatal conductance (g_s) leads to complexity and uncertainty. In this study, a mixed generalised additive model (MGAM) was used to simulate g_s as a nonlinear function of environmental variables. By including the MGAM g_s into PET_{PM} a PET_{MGAM} was then developed. Using four eddy covariance flux tower sites data with different vegetation types, PET_{MGAM} produced higher Nash-Sutcliffe Efficiency (NSE) and Kling–Gupta efficiency (KGE) values than PET_{PM} for PET estimation and runoff simulation. Results showed that PET_{MGAM} moderated the underestimation of runoff simulated by PET_{PM}, particularly in extreme wet conditions when runoff is more sensitive to PET. Shapley Additive exPlanations (SHAP) analysis revealed that key environmental variables contribute differently to PET estimation in wet and dry climates. Notable changes in SHAP values for different climate conditions were related to CO_2 and soil water content, which are the key environmental variables in PET simulation for wet and dry conditions, respectively. PET_{MGAM}, considers the role of key environmental variables in a modified PET, leading to a more accurate estimate of the water balance elements under extreme wet climate conditions.

4.2 Plain language summary

Potential evapotranspiration (PET) is a key input of many hydrological models for runoff simulation. It is one of the most uncertain hydrological variables, so accurate PET estimation is still challenging. Several studies have shown that neglecting the vegetation response to environmental variables such as CO_2 concentration, air temperature (TA), vapor pressure deficit (VPD), soil water content (SWC), and net radiation ® in current PET equations in hydrological models causes underestimation of runoff, especially in locations with higher precipitation and runoff. We present a method in which the vegetation response to the environmental variables is part of a modified PET equation. This new method improves the PET estimation and consequently, the runoff simulation, especially in extreme wet conditions when precipitation is higher than PET. We show how sensitive the PET is to CO_2 and other environmental variables, allowing for better future simulation and prediction of PET and runoff.

4.3 Introduction

Global warming leads to more extreme precipitation events, intensifying storm runoff (Tabari, 2020; Yin et al., 2018). Runoff from extreme events is a serious societal concern, and it has caused extensive property damage and agricultural losses across the globe (Yin et al., 2018). Global economic losses due to runoff have risen over the past half-century and exceeded $30 billion annually in the past decade (Roxy et al., 2017). Accurate simulation of rainfall-runoff is crucial for analysing and managing extreme rainfall-runoff risks (Wang & Karimi, 2022). Because extreme hydrological events such as high temperatures (TA) and extreme precipitation (Pr) occur more regularly, rainfall-runoff modelling is becoming increasingly important in hydrological forecasting (Pimentel et al., 2023; Yin et al., 2018). In spite of this, hydrological models consistently and significantly underestimate extreme runoff (Ballarin et al., 2023; Milly & Dunne, 2017; Zhou et al., 2023). Hence, it is necessary to identify and address the reasons for this bias in runoff simulation by hydrological models.

Several studies have evaluated the runoff simulation of hydrological models using data obtained from general circulation models (GCMs), (Ballarin et al., 2023; Milly & Dunne, 2017; Yang et al., 2019; Zhou et al., 2023). Hydrological modelling of runoff fed by data from offline climate models was compared with the outputs of the GCM model Coupled Model Intercomparison Project Phase 5-6 (CMIP 5-6) (Ballarin et al., 2023; Liu et al., 2024). CMIP 5-6 showed an increase in runoff over global terrestrial environments by 2100 (Milly & Dunne, 2017; Swann et al., 2016; Zhou et al., 2023). However, offline climate model results underestimated runoff increases, contradicting CMIP5-6 results (Hou et al., 2023; Milly & Dunne, 2016). These contradictory runoff predictions of CMIP5-6 and offline climate model output-driven hydrological models are the results of neglecting vegetation’s stomatal conductance (g_s) response to change in climate and CO_2 in the PET equation of hydrological models (Peiris & Döll, 2023; Vremec et al., 2023; Zhou et al., 2023). CMIP 5-6 uses actual evapotranspiration (AET) to represent the dynamic responses of vegetation to climate variables, whereas PET is the rate of evapotranspiration without water stress, which assumes g_s as a constant with a value of 70 ms^{-1} (Yang et al., 2019; Zhou et al., 2023). Consequently, incorporating g_s as a function of CO_2 and climate variables into PET estimation improves climate model outputs in runoff simulation (Bass et al., 2023; Yang et al., 2019). Despite this, only a few studies have investigated vegetation response to climate variables in the PET equation (Peiris & Döll, 2023; Zhou et al., 2023).

Among many PET methods, the Penman-Monteith PET (PET_{PM}) represents an accurate yet simple approximation to the more complex system embedded in climate models (McMahon et al., 2013; Milly & Dunne, 2016). By representing the vegetation response to atmospheric CO_2 in PET_{PM} through g_s, it is possible to see how PET simulations are improving according to AET data from CMIP 5-6 outputs (Yang et al., 2019; Zhang et al., 2021). However, the g_s variable in PET_{PM} is assumed to be a linear function of CO_2, which does not align with the nonlinear CO_2-g_s behaviour in the real environment (Li et al., 2019). Additionally, the effects of environmental variables such as air temperature (TA), radiation ®, vapour pressure deficit (VPD), and soil water content (SWC) on g_s are not included in the PET_{PM} equation despite their significant influence on plants. Thus, more work is required to understand whether vegetation response could be more accurately simulated in a PET equation that includes these variables.

This study aims to improve runoff simulation by incorporating g_s as a function of CO_2, R, VPD, Ta, and SWC in the PET estimation. We use GR4J as a conceptual rainfall-runoff model to simulate runoff in function of rainfall and PET at the catchment level. First, we simulate g_s as a nonlinear function of environmental variables by extracting data from four eddy covariance flux tower sites with different vegetation types. For this purpose, the mixed generalised additive model (MGAM) is used as a machine learning model to establish a relationship between g_s and environmental variables. Then, the modified PET model including g_s (PET_{MGAM}) is compared with the traditional PET_{PM} model at two levels: by comparing PET_{MGAM} and PET_{PM} by eddy covariance AET from flux towers, and by comparing the accuracy of runoff simulation driven by PET_{MGAM} and PET_{PM}. In addition, to understand the importance of CO_2 and climate variable effects on runoff simulation, the PET_{MGAM} containing the effects of multiple environmental variables on g_s, is compared with $PET_{MGAM\,[CO_2]}$, which only considers the CO_2 effect on g_s. The performance of different PET models is then compared for dry, wet, and extreme wet climate conditions. Finally, the Shapley Additive exPlanations (SHAP) analysis is used to determine the key environmental variables that control PET for these different climate conditions to interpret the performance of PET models in runoff simulations.

4.4 Data and methodology

4.4.1 Forcing data

Surface flux measurements and historical meteorological data were utilised as the input variables to simulate g_s and PET. Eddy covariance flux tower data was used for latent heat flux (LE), soil heat flux (G), sensible heat flux (H), CO_2, R, VPD, TA, and SWC. Four Ameriflux sites in the United States of America were selected with different biomes, including deciduous broad-leaf forest (DBF), evergreen needle-leaf forest (ENF), and crop (CRO) (Table 1 and Figure 1). Data from the gauge stations provided daily precipitation (Pr) and runoff (Q) for each catchment of the Ameriflux sites (Newman et al., 2014).

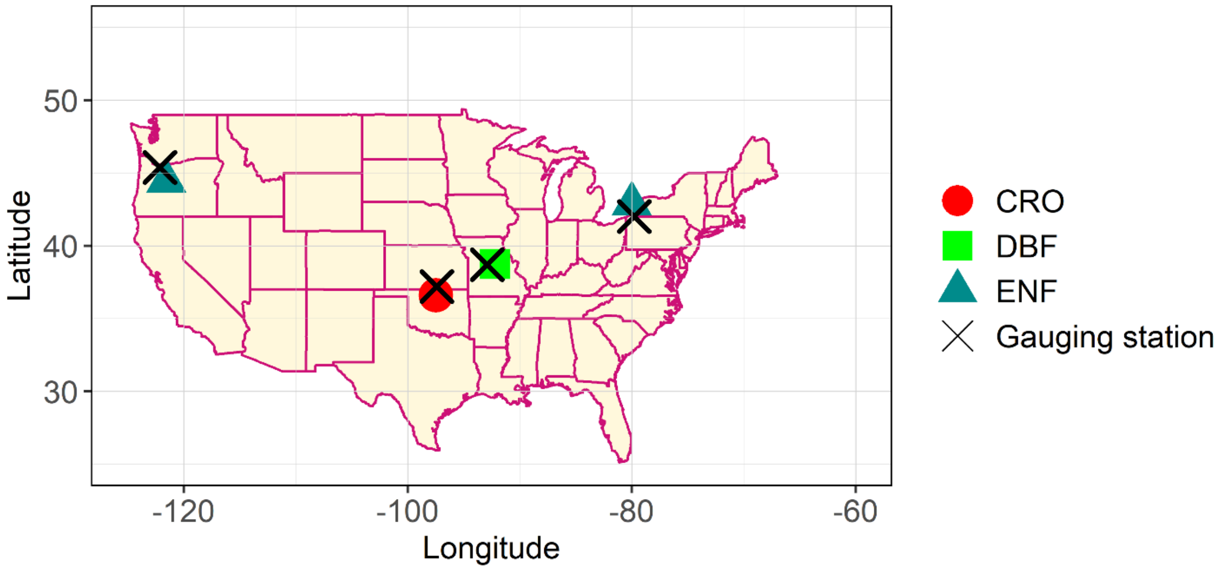


Figure 6. Locations of four flux tower sites and gauging stations with different vegetation types, including crop (CRO), deciduous broad-leaf forest (DBF), and evergreen needle-leaf forest (ENF).

Table 2. Locations and climate conditions of flux tower sites with different vegetation types: crop (CRO), deciduous broad-leaf forest (DBF), and evergreen needle-leaf forest (ENF).

Site name-Vegetation types-Years	Lat	Long	Elevation	Annual TA (°C)	Annual total Pr (mm)	Annual total Q (mm)	Gauging station name-Area (km^2)	Flux tower references
US-ARM (CRO) (2009-2014)	36.60	-97.48	314	14.7	752	95.8	Wellington (398.68)	(Raz-Yaseef et al., 2015)
US-Moz (DBF) (2006-2014)	38.74	-92.20	219	12.1	935	271	Lamine River (1405.76)	(Wood et al., 2018)
US-Me2 (ENF) (2008-2014)	44.45	-121.55	1253	6.2	918	759	Sandy River (683.46)	(Kwon et al., 2018)
CA-TP3 (ENF) (2009-2014)	42.70	-80.34	184	8.0	904	617	French Creek (238.17)	(Arain et al., 2022)

4.4.2 Penman–Monteith PET (PET_{PM})

The PET_{PM} equation for reference crop is as Eq. 1 (Allen et al., 1998).

$$PET_{PM} = \frac{0.408\Delta(R - G) + \gamma \frac{900}{TA + 273} uVPD}{\Delta + \gamma(1 + 0.34u)} \tag{1}$$

where, PET_{PM} is reference evapotranspiration (mm day^{-1}), Δ the slope of the saturation vapor pressure-temperature curve (kPa $°C^{-1}$), R and G are net radiation and soil heat flux (MJ m^{-2} day^{-1}), VPD is vapour pressure deficit (kPa), γ is the psychometric constant (kPa $°C^{-1}$), TA is air temperature (°C), and u is wind speed (m s^{-1}).

In previous studies, PET_{PM} was modified by adding g_s in Eq. 1 using $0.34u = \frac{g_a}{g_s}$; where g_s and g_a are canopy stomatal conductance and aerodynamic conductance (m s^{-1}), respectively. The g_s value can be estimated by latent heat flux (LE) from flux tower data through inversion of the original Penman-monteith model as Eq. 2 to Eq. 4.

$$LE = \frac{\Delta(R - G) + \rho C_p g_a VPD}{\Delta + \gamma(1 + \frac{g_a}{g_s})} \tag{2}$$

$$g_s = \frac{g_a \gamma}{\frac{\Delta(R - G) + \rho C_p g_a VPD}{LE} - (\Delta + \gamma)} \tag{3}$$

where, LE is latent heat flux (Wm^{-2}), ρ is air density (kg m^{-3}), C_p is specific heat capacity of dry air (J kg^{-1} $°C^{-1}$), R and G are in Wm^{-2} g_s and g_a are canopy and aerodynamic conductance (m s^{-1}).

The aerodynamic conductance is defined as (Thom, 1972),

$$g_a = \frac{k^2 \times u}{\left[1n\left(\frac{z-d}{z_m}\right)\ln\left(\frac{z-d}{z_h}\right)\right]} \tag{4}$$

where, z is measurement height (m), u is in m s^{-1}, $k = 0.41$ is von Karman's constant, $d = 0.67 \times h$ is displacement height, h is canopy height (m), $z_m = 0.123 \times h$ is the roughness length for momentum transfer, and $z_h = 0.0123 \times h$ is the roughness length for heat and vapour transfer.

4.4.3 MGAM model for modified PET simulation (PET_{MGAM})

We used the MGAM model to train and simulate g_s using environmental variables from available flux tower data. The training process for MGAM requires g_s values, which were obtained from Eq. 2 to Eq. 4. A nonlinear function of MGAM was then used to demonstrate the relationship between covariates and outcomes (Eq. 5).

$$f(x) = \sum_{k=1}^{K} \beta_k b_k(x) \tag{5}$$

where, f(x) is a smoother function, b_k are basis functions, β_k are corresponding coefficients, and K is referred to as basis size or basis complexity. The coefficients of the basis functions were optimised to ensure the appropriate complexity of the models (Wood et al., 2016). The f(x) smoother function was selected as a smooth function (S) to represent nonlinearity of variables directly, or as a tensor function (t_i) to represent the interaction between variables. The structure of g_s simulation in MGAM can be described as Eq. 6.

$$g_s = \sum_{m=1}^{M} f(x_m) \tag{6}$$

where, M are the effective environmental variables of g_s. Each of the effective variables has a smoother function f(x) (Eq. 5), which contains basis functions with relevant coefficients. Therefore, by replacing the effective environmental variables, the g_s function can be represented as in Eq. 7.

$$g_s = f(\mathrm{VPD}, \mathrm{CO_2}, \mathrm{TA}, \mathrm{SWC}, \mathrm{R}) = \begin{cases} S(\mathrm{VPD}) + S(\mathrm{CO_2}) + S(\mathrm{TA}) + S(\mathrm{SWC}) + S(\mathrm{R}) + ti(\mathrm{VPD}, \mathrm{CO_2}, \mathrm{SWC}) & \text{for} \\ S(\mathrm{VPD}) + S(\mathrm{CO_2}) + S(\mathrm{TA}) + S(\mathrm{SWC}) + S(\mathrm{R}) + ti(\mathrm{VPD}, \mathrm{CO_2}, \mathrm{TA}) & \text{for DBF and} \end{cases} \tag{7}$$

The modified reference crop PET equation was obtained by replacing the g_s as a function of environmental variables into Eq. 1, which can be described as Eq. 8.

$$\mathrm{PET_{MGAM}} = \frac{0.408\Delta(\mathrm{R} - \mathrm{G}) + \gamma \frac{900}{\mathrm{TA} + 273} \mathrm{uVPD}}{\Delta + \gamma\left(1 + \frac{g_a}{g_s = f(\mathrm{VPD}, \mathrm{CO_2}, \mathrm{TA}, \mathrm{SWC}, \mathrm{R})}\right)} \tag{8}$$

$\mathrm{PET_{MGAM\,[CO_2]}}$ follows the same form as Eq. 8, except that g_s is only a function of CO_2.

MGAM was validated by splitting 70% of the data for training the model and the remaining 30% for testing the model, using 10-fold cross-validation technique for training. The ‘nls’, ‘mgcv’ and ‘caret’ packages in R (Baty et al., 2015; Wood et al., 2016) were used for g_s simulation by MGAM and cross-validation in this study.

4.4.4 SHAP analysis

SHAP analysis is based on cooperative game theory to interpret model simulation (Lundberg et al., 2020; Lundberg & Lee, 2017; Mardian et al., 2023). The SHAP value shows the contribution of each variable or predictor to the model simulation and explains the effect of the high and low values of each variable on the simulated value (Shi et al., 2023). The SHAP value defines the average marginal contribution of each variable across all coalitions to which the variable belongs (Lee et al., 2023). The SHAP value is calculated by Eq. 9.

$$\varphi_i(f, x) = \sum_{s \subseteq x} \left[\frac{|s|!\,(M - |s| - 1)!}{M!}\right] \times [f_x(s) - f_x(s \setminus i)] \tag{9}$$

where, φ is the SHAP value for variable $i = [1, M]$ and M is the number of variables, f is the simulation model, x is sample observation for specific ith variable, s is the subset of possible coalitions of variables. The first bracket of the equation refers to the weighting for each subset of coalitions, and the second bracket refers to the marginal contribution of ith variable, which is the difference between the f model with and without ith variable, which is $f_x(s)$ and $f_x(s \setminus i)$, respectively. The higher the SHAP value of each variable, the greater the impact of the variable on the simulation output (Shi et al., 2023). In this study, the SHAP method shows the contribution of each of the environmental variables VPD, R, TA, CO_2, and SWC for PET_{MGAM} simulation at different climate conditions (dry and wet). The SHAP value of each environmental variable enhances the interpretability of the PET and runoff simulation for different climate conditions. The XGBoost, one of the common machine learning models for SHAP analysis, was used through the 'shapviz' and 'xgboost' packages in R (Chen et al., 2023; Mayer, 2023).

4.4.5 Runoff simulation model

We used the daily conceptual rainfall-runoff model GR4J. GR4J is based on a soil moisture store and uses a continuous relationship between moisture level in the soil store and runoff production (De la Fuente et al., 2023; Perrin et al., 2003; Sinha et al., 2022). The GR4J input variables are Pr, PET, and TA. Each day is considered as dry when Pr is less than PET or wet when Pr exceeds PET. Runoff routing is determined using a unit hydrograph (Guo et al., 2020; Perrin et al., 2003). On wet days, the proportion of net Pr (Pr minus PET) is added to the soil moisture store and the remaining effective rainfall contributes to runoff production (Santos et al., 2018). The percolation (or infiltration) leakage and effective rainfall go to the routing store where they are split into two parts routed by two-unit hydrographs. After applying groundwater-surface water exchanging function, the total runoff is simulated by adding these two parts (Wang & Solomatine, 2019). GR4J has four calibration parameters: maximum capacity of the production store (mm), groundwater exchange coefficient (mm), one day ahead maximum capacity of the routing store (mm), and time base of the unit hydrograph (days) (Delaigue et al., 2023; Wang & Solomatine, 2019). For calibration processes, the observed runoff was required as input to the GR4J model. The GR4J model in this study was run with the 'airGR' package in R (Coron et al., 2023; Coron et al., 2017).

The GR4J runoff simulation in this study was used to measure the changes in runoff simulation accuracy when the traditional PET equation was modified by adding the role of g_s. The climate information for four flux tower sites with CRO, ENF, and DBF vegetation was used to train MGAM to simulate g_s as a function of the environmental variables and added g_s into the PET equation (PET_{MGAM}). Then the runoff simulated by GR4J with PET_{MGAM} was compared with the runoff simulated by GR4J with PET_{PM}. In addition, $PET_{MGAM\,[CO_2]}$, which contains g_s only as a function of CO_2 was added to this comparison. Model performance was compared for the different climate conditions based on the ratio of Pr/PET: Pr/PET<1 was considered dry, Pr/PET>1 was considered wet, and the top 5% of the Pr/PET ratio was considered extremely wet. The runoff simulation accuracy with different forms of PET showed the role of g_s as a function of the effective environmental variables in PET estimation at different climate conditions (Fig. 2). Validation of the MGAM and GR4J models was performed with 10-fold cross-validation and time series cross-validation, respectively.

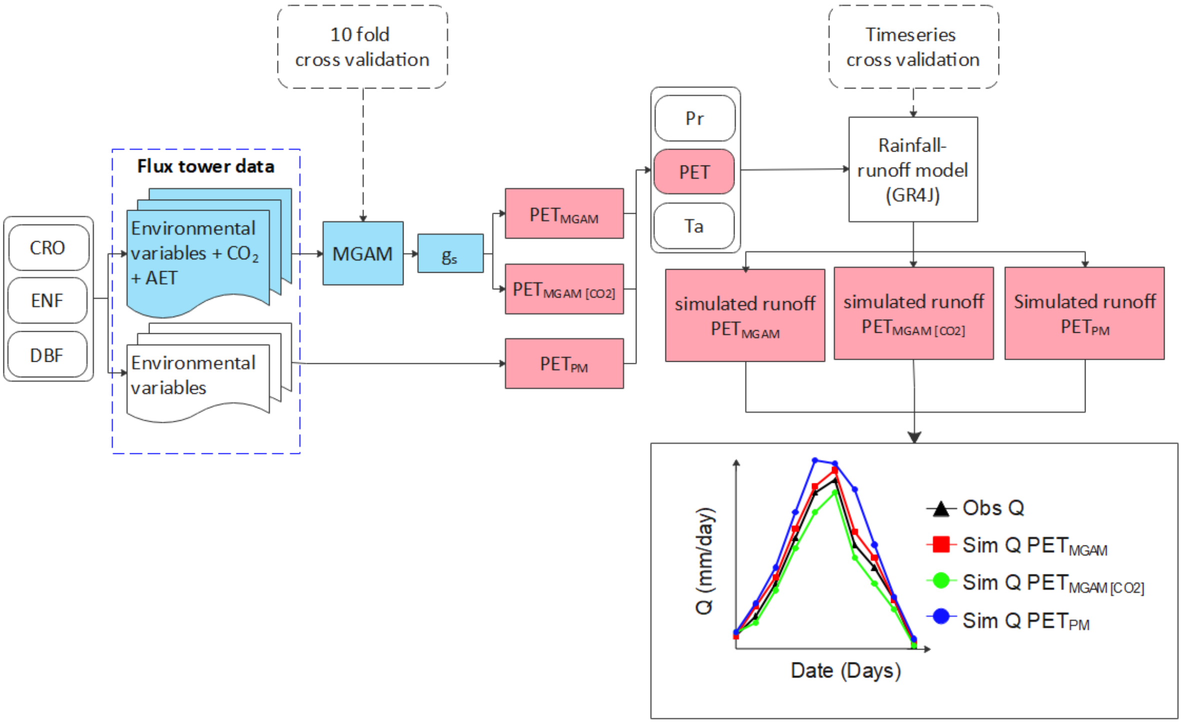


Figure 2 The climate information of crop (CRO), evergreen needleleaf (ENF), and deciduous broadleaf (DBF) vegetation was used to train MGAM to simulate g_s as a function of the environmental variables and added g_s into the potential evapotranspiration (PET) equation (PET_{MGAM}). Then the runoff simulated by GR4J with PET_{MGAM} was compared with the runoff simulated by GR4J with PET_{PM}. In addition, $PET_{MGAM\,[CO2]}$, which contains g_s only as a function of CO_2 was added to this comparison. Model performance was compared for the different climate conditions based on the ratio of Pr/PET: Pr/PET<1 was considered dry, Pr/PET>1 was considered wet, and the top 5% of the Pr/PET ratio was considered extremely wet. The runoff simulation accuracy with different forms of PET showed the role of g_s as a function of the effective environmental variables in PET estimation at different climate conditions. Validation of the MGAM and GR4J models was performed with 10-fold cross-validation and time series cross-validation, respectively.

4.5 Results

MGAM with g_s simulation resulted in a Nash–Sutcliffe efficiency (NSE) value greater than 50% for all flux tower sites (Table S1). The GR4J model performance was evaluated by training and testing processes for different PET inputs as traditional PET (PET_{PM}) and the modified

PET via adding g_s (PET_{MGAM}) (Table S2). The result of the GR4J simulation showed an acceptable performance (NSE higher than 50% for the test and train dataset) of this model. To further investigate PET_{MGAM} and PET_{PM} simulations, the time series results of simulated PET_{MGAM} and PET_{PM} were compared with observed AET from eddy covariance measurements at four flux tower sites (Fig. S1). The time series data for four sites showed higher NSE and Kling–Gupta efficiency (KGE) values for the simulated PET_{MGAM} in comparison to PET_{PM}. Furthermore, simulated PET_{PM} was overestimated compared with simulated PET_{MGAM} and observed AET. As PET plays a greater role in wet conditions (when Pr/PET > 1), the PET_{MGAM} was compared to the PET_{PM} in this situation (Fig. 3a-3d). The results showed that PET_{PM} was substantially overestimated in comparison to PET_{MGAM} and the observed AET. PET_{MGAM} performs better in PET simulation than PET_{PM} under wet conditions (Fig. 3).

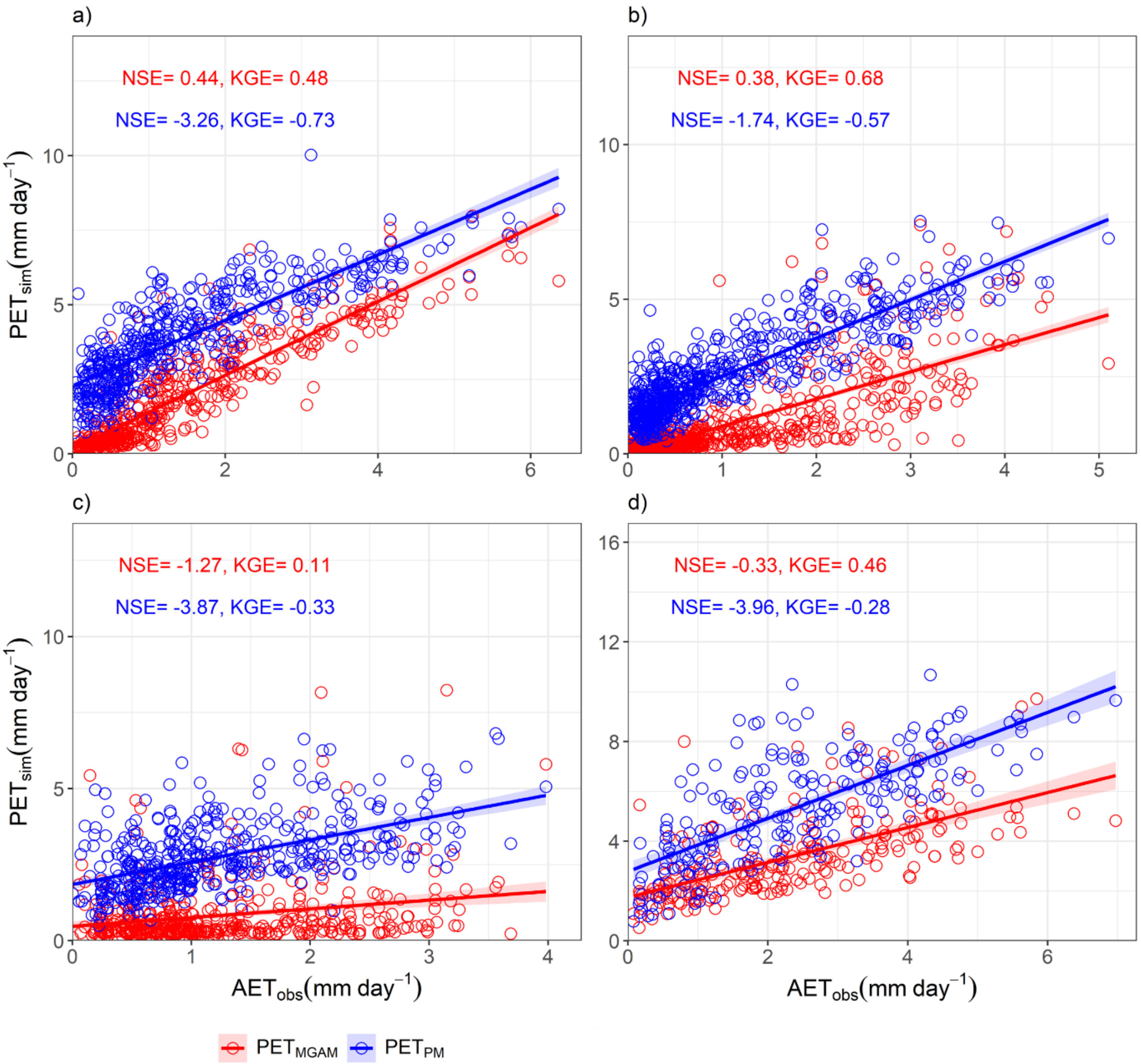


Figure 3 Simulated PET_{PM} and PET_{MGAM} for wet and extreme wet conditions (Pr/PET>1) for four flux tower sites, a) US-MOZ, b) CA-TP3, c) US-Me2, and d) US-ARM.

Based on the results of runoff simulation, it was evident that choosing different PET values resulted in notable differences in runoff simulation accuracy (Fig. 4). The accuracy of runoff simulation was evaluated under dry, wet, and extreme wet conditions separately (Fig. 4a-4d). Across all sites, PET_{MGAM} supported a higher accuracy runoff simulation for extreme wet conditions. However, there were marginal differences between dry and wet runoff simulation accuracy (NSE and KGE values). The comparison of runoff simulation with PET_{MGAM} and PET_{PM} for extreme wet conditions showed that PET_{PM} scenario underestimated the overall runoff across all sites in comparison with the PET_{MGAM} runoff simulation scenario (Fig. 5a-5d).

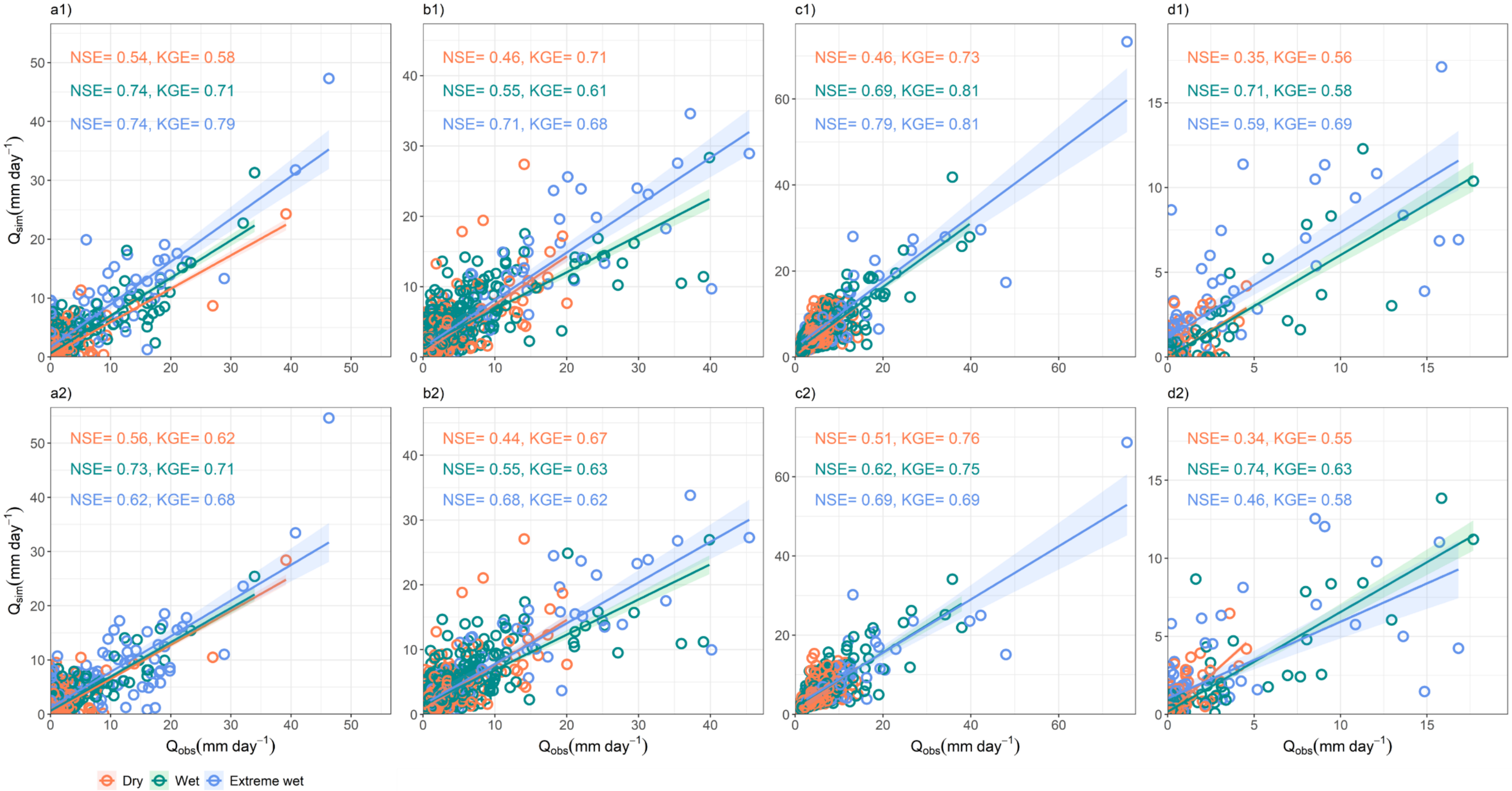


Figure 4 Simulated runoff for dry, wet, and extreme wet conditions by using two PET formulations: (1) PET_{MGAM} and (2) PET_{PM}, for four flux tower sites: a) US-MOZ, b) CA-TP3, c) US-Me2, and d) US-ARM.

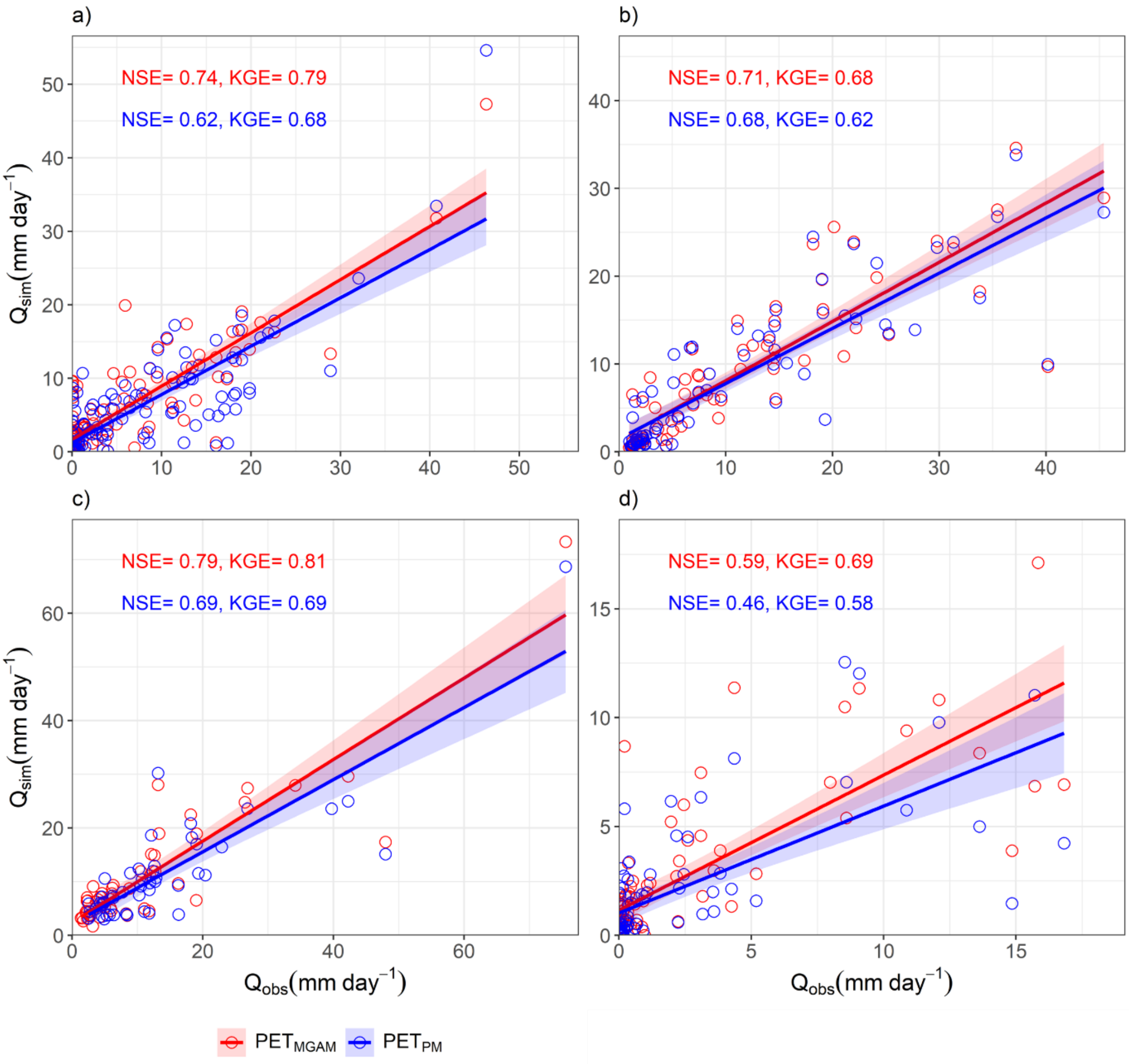


Figure 5 Simulated runoff using PET_{PM} and PET_{MGAM} for extreme wet conditions for four flux tower sites, a) US-MOZ, b) CA-TP3, c) US-Me2, and d) US-ARM.

The results of simulated runoff by adding g_s as a function of only CO_2 into the PET equation ($PET_{MGAM\ [CO2]}$) showed an acceptable performance of this model for wet conditions, but failure in dry conditions (Table S3 and Fig. S2). However, PET_{MGAM} showed better results than $PET_{MGAM\ [CO2]}$ in all climate conditions.

The role of CO_2 and environmental variables on PET_{MGAM} for dry, and wet conditions were investigated in more detail by The SHAP analysis (Fig. 6). The data for extreme wet condition is also included in wet conditions. The SHAP value showed the changes in average PET_{MGAM} values forced by each environmental variable (x-axis). The mean absolute of SHAP values for each variable show the contribution of the variable in the PET_{MGAM} simulation (y-axis). The gradient colour (feature value) shows low and high value of each variable. To focus on the variables that have higher impacts on PET_{MGAM} simulation, all variables are sorted based on the maximum absolute value of their SHAP values. In all four sites, the SHAP value for each environmental variable (y-axis) showed that R contributed most to PET_{MGAM} (Fig. 6a-6d). The gradient colour shows that R had an increasing effect on PET_{MGAM}. However, CO_2 and SWC, which were added indirectly into PET_{MGAM} equation by the g_s value, had completely different contributions for dry and wet conditions. PET simulation showed a stronger role for CO_2 in wet conditions, whereas SWC played a stronger role in dry conditions. Among all four sites, the SWC had the greatest impact on the PET_{MGAM} US-ARM site (with CRO vegetation type) in dry conditions (Fig. 6-d) Dry), while it had the least effect on PET_{MGAM} at the US-MOZ site (covering with mature trees) (Fig. 6-a) Dry).

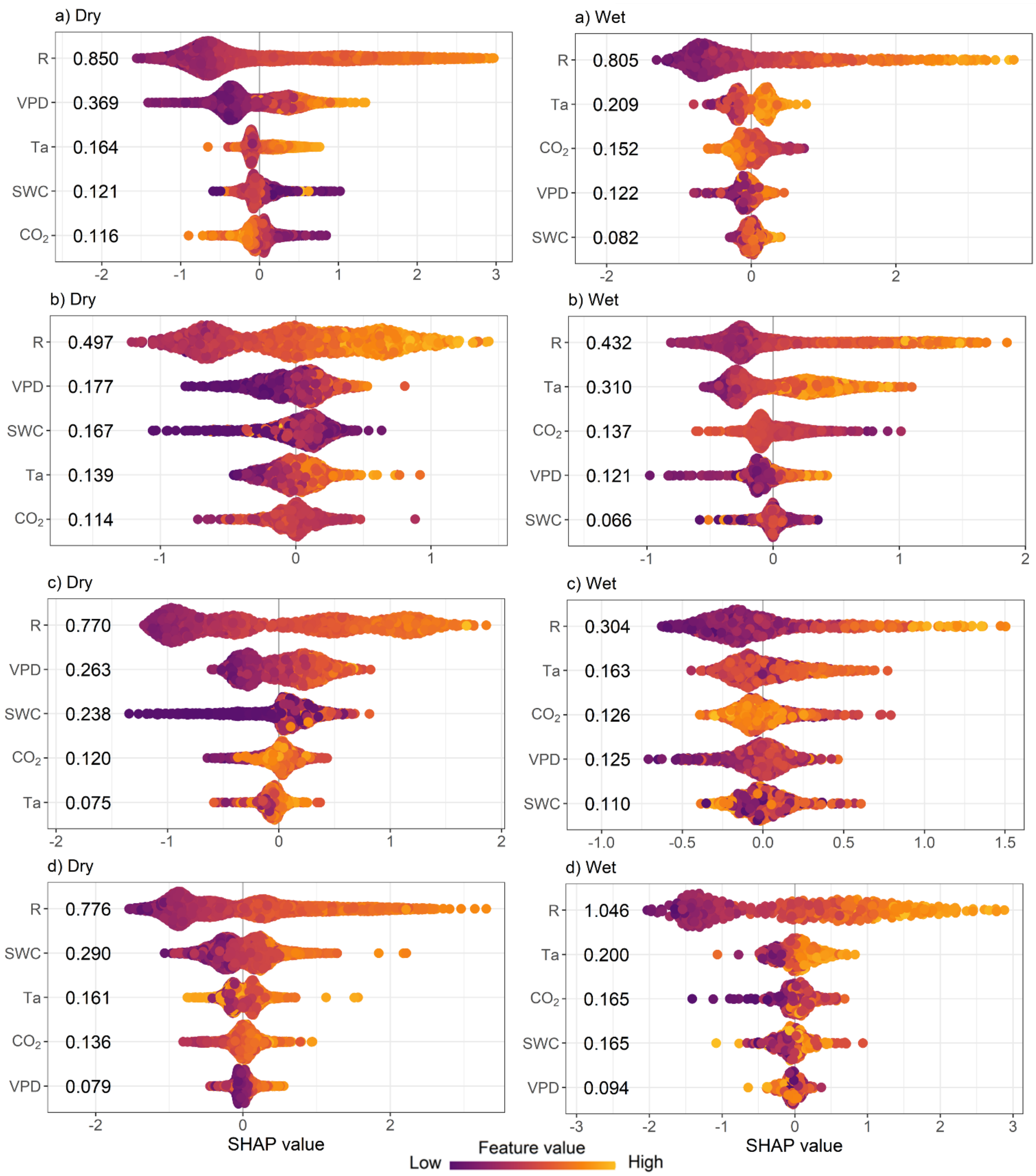


Figure 6 SHAP values to show the role of environmental variables in PET_{MGAM} estimation for wet and dry conditions for four flux tower sites, a) US-MOZ, b) CA-TP3, c) US-Me2, and d) US-ARM. The data for extreme wet condition is also included in wet conditions.

4.6 Discussion

The importance of g_s in PET estimation has been demonstrated in several studies (Bai et al., 2020; Ban & Lettenmaier, 2022; Milly & Dunne, 2017; Vremec et al., 2023). By ignoring g_s in the PET_{PM} equation, PET_{PM} estimation is rendered inaccurate (Liu et al., 2023; Peiris & Döll, 2023; Swann et al., 2016; Zhang et al., 2022). Analysis of climate model outputs from anthropogenic climate change experiments indicates that different versions of the PET equation, such as PET_{PM} for reference crops and PET for open water surface, overestimate PET significantly (Liu et al., 2023; Milly & Dunne, 2017; Zhang et al., 2023). As a result of the PET overestimation, runoff simulation also presented an underestimation of runoff, especially in regions with high precipitation and where the climate models showed increased runoff (Milly & Dunne, 2016; Milly & Dunne, 2017; Zhou et al., 2023). Runoff is significantly underestimated in wet conditions in the above studies, since runoff is more sensitive to PET in these conditions (Chen & Wang, 2022; Roderick et al., 2014; Yang et al., 2019). The results of this study support the overestimation of PET by PET_{PM} at four flux tower sites with different vegetation types. The higher NSE and KGE values between observed AET and simulated PET_{MGAM} present a better estimation of PET_{MGAM} than PET_{PM} (Fig. 3 and Fig. S1). In addition, the simulated PET_{MGAM} in this study shows better results than PET_{PM} in GR4J runoff simulation, especially for extreme wet conditions, when the role of PET is more significant for runoff fluctuation (Fig. 4). Runoff simulation was underestimated when PET_{PM} was used compared to when PET_{MGAM} was used for all four flux tower sites (Fig. 5).

Other studies have also investigated the modification of PET_{PM} (Ballarin et al., 2023; Yang et al., 2019; Zhang et al., 2023). Adding g_s as a function of CO_2 in the PET_{PM} equation has been shown to improve the runoff simulation at a global scale (Zhou et al., 2023). In the above studies, the g_s values in climate model outputs were obtained by inverting the PET_{PM} with climate model outputs of AET. However, they have used a linear function of CO_2-g_s for all vegetation types at global scales, and their findings are based on ensemble climate model CMIP5 outputs. The generalisation of the linear CO_2-g_s equation to real-world data remains an open question for future investigations (Yang et al., 2019). Global climate models' analysis needs careful investigation due to the coarse spatial resolution and imperfect physical parameterisations; even the spatial downscaling of these models may seriously increase inconsistencies in their information content (Milly

& Dunne, 2016; Milly & Dunne, 2017). Therefore, in this study, we extracted the g_s value from AET obtained by flux tower data, which reflects real-world conditions. In addition, we used g_s simulated as a nonlinear function of various environmental variables such as VPD, TA, SWC, and Ra, rather than CO_2, which enables a simulated g_s suitable for all climate conditions.

Despite GR4J fed with PET_{MGAM} showing satisfactory results in runoff simulations, we performed another investigation using $PET_{MGAM\ [CO_2]}$ to analyse the role of CO_2 effects on g_s separately. While the runoff simulation fed by $PET_{MGAM\ [CO_2]}$ provided acceptable results for wet and extremely wet conditions, the results for dry conditions notably degraded. GR4J with $PET_{MGAM\ [CO_2]}$ failed to predict runoff under dry conditions, possibly because other environmental variables play an important role in g_s fluctuation under such conditions. For instance, the decreasing effects of elevated CO_2 on g_s and PET are partially offset by the increasing effects of elevated TA on g_s and PET (Bass et al., 2023; Yang et al., 2019). In addition, lower SWC contributed to the reduction in g_s and PET, especially in warm and dry conditions (Zhou et al., 2023).

The SHAP values provided additional justification for the function of these environmental variables in PET fluctuation at wet and dry conditions separately. The higher SHAP value presents the higher effects of variables on PET fluctuations. The SHAP value for CO_2 and SWC, which are introduced in the PET equation by adding g_s, present notable contributions for dry and wet conditions. The SHAP analysis shows the considerable role of CO_2 rather than SWC in wet conditions. Therefore, considering only CO_2-g_s for runoff simulation results in acceptable outcomes for wet conditions. In contrast, SWC's role in PET simulation is highlighted in dry conditions; hence, neglecting SWC in the $PET_{MGAM\ [CO_2]}$ equation degraded the runoff simulation accuracy in dry conditions. The marginal degradation of NSE and KGE in runoff simulation with $PET_{MGAM\ [CO_2]}$ in comparison to PET_{MGAM} for the US-Moz site (contains mature trees with an average of 130 years old) for dry conditions was justified by the negligible SHAP value of SWC for this site. While the notable decrease of NSE and KGE in runoff simulation with $PET_{MGAM\ [CO_2]}$ for US-ARM was due to the considerable role of SWC for dry conditions at this site. More information about the negligible role of SWC for mature trees against its highlighted role for CRO is discussed in our previous study (Chitsaz et al., 2024).

The GR4J model is a common surface water hydrological model owing to its ease of use, the computational speed that facilitates exploring sensitivity and uncertainty to climate variability as well as reasonable performance in the absence of change (Partington et al., 2022; Razavi et al., 2021; Renard et al., 2010). However, like any other conceptual model, GR4J has some limitations in its ability to capture and represent long-term hydrological changes, such as variation in topography, soil porosities, and geomorphology (Fowler et al., 2020; Peel & McMahon, 2020). The physically-based models, on the other hand can overcome these limitations, but they have a significant cost associated with model development due to field observations, data preparation and extensive parameter calibrations which cause over-parameterisation problems (Camporese et al., 2015). Therefore, physically-based models do not provide a reliable and practical basis for representing environmental changes across a wide range of climate scenarios and locations (Thyer et al., 2024).

The ML algorithms are robust in dynamic environments since they adapt to changes in data distribution over time and have the potential to yield more promising results, however, they are not easy to interpret. Therefore, a combined physically-based model with ML algorithms as 'Hybrid' models to preserve the advantages of physically-based models may allow for capturing complex hydrological processes while leveraging the data-driven capabilities of ML models. This approach can incorporate domain knowledge from existing data and physical constraints, particularly useful when facing hydrological changes.

4.7 Conclusion

A reliable estimate of changes in runoff is essential to mitigate the adverse consequences of hydroclimatic variables such as drought and flooding. However, most hydrological models underestimate extreme runoff due to the inaccurate PET simulation. The result of this study shows that incorporating the vegetation response to environmental variables into the PET equation can improve runoff simulation results. The PET_{MGAM} proposed in this study is a modified PET that incorporates the vegetation response to VPD, TA, Ra, CO_2, and SWC by adding g_s into the traditional PET_{PM}. This new approach presents more reliable results in PET simulations that lead to higher NSE and KGE values in runoff simulation, especially in wet conditions where PET plays a greater role in runoff fluctuations. The sensitivity of PET to each environmental variable highlights the important role of CO_2 and SWC in runoff estimation at wet and dry conditions, respectively.

Supplementary information

Table S3 The NSE values of train and test for MGAM in g_s simulation for four flux tower sites

Site name	Train	Test
US-Moz	0.68	0.66
CA-TP3	0.55	0.50
US-Me2	0.66	0.65
US-ARM	0.52	0.50

Table S2 The NSE (and KGE) values of train and test for GR4J model in runoff simulation for PET_{MGAM} and PET_{PM} as input to the model for four flux tower sites

Site name	PET_{MGAM}		PET_{PM}	
	Train	Test	Train	Test
US-Moz	0.77 (0.79)	0.75 (0.76)	0.76 (0.77)	0.73 (0.75)
CA-TP3	0.66 (0.72)	0.65 (0.70)	0.68 (0.73)	0.65 (0.69)
US-Me2	0.80 (0.87)	0.74 (0.84)	0.76 (0.80)	0.71 (0.78)
US-ARM	0.65 (0.62)	0.64 (0.61)	0.68 (0.60)	0.64 (0.62)

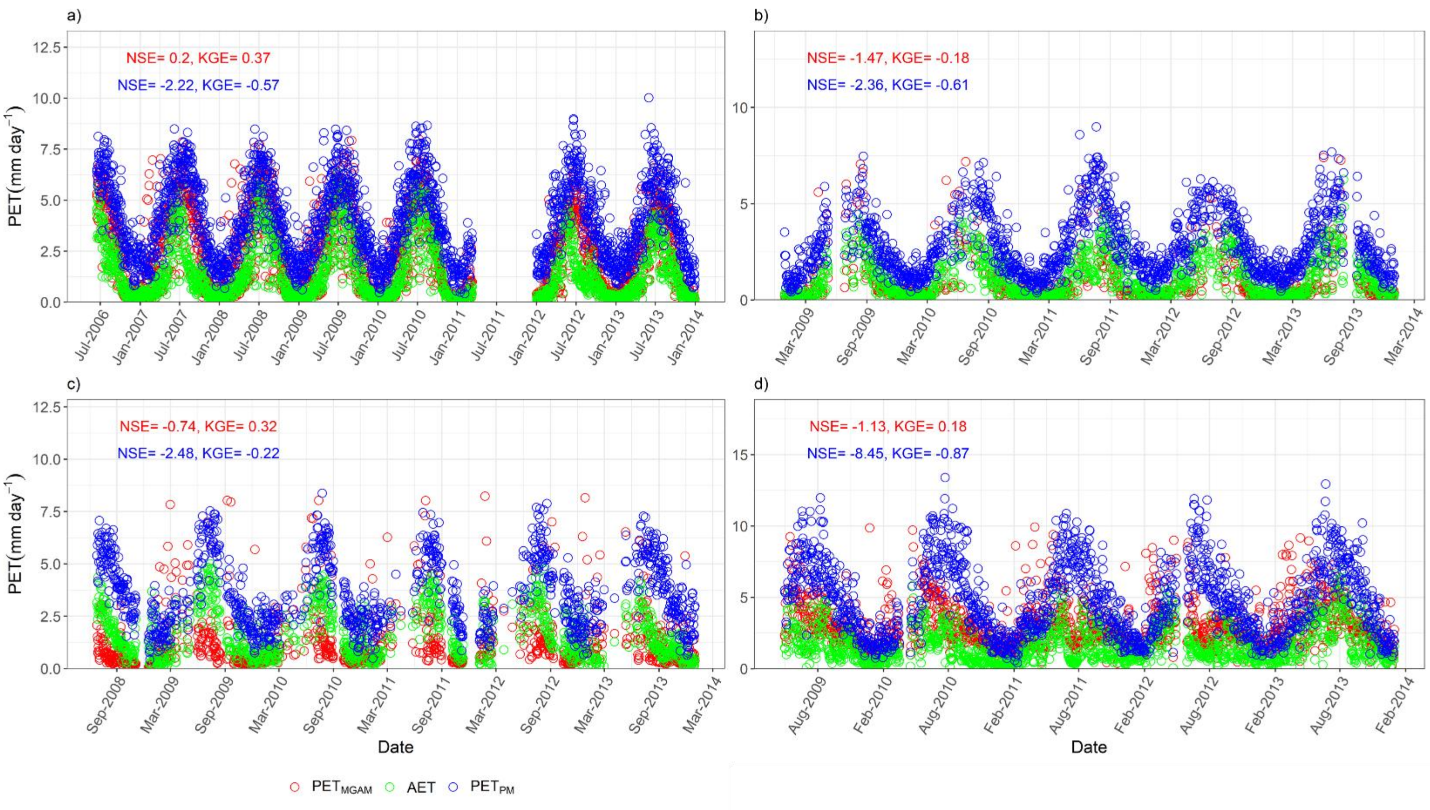


Figure S1 Observed AET and simulated PET$_{PM}$ and PET$_{MGAM}$ for timeseries for four flux tower sites, a) US-MOZ, b) CA-TP3, c) US-Me2, d) US-ARM.

Table S3 The NSE and (KGE) values of train and test for GR4J model in runoff simulation for $PET_{MGAM\ [CO_2]}$ as input to the model for four flux tower sites

Site name	Train	Test
US-Moz	0.79 (0.80)	0.77 (0.78)
CA-TP3	0.63 (0.70)	0.61 (0.67)
US-Me2	0.68 (0.78)	0.64 (0.75)
US-ARM	0.66 (0.62)	0.64 (0.60)

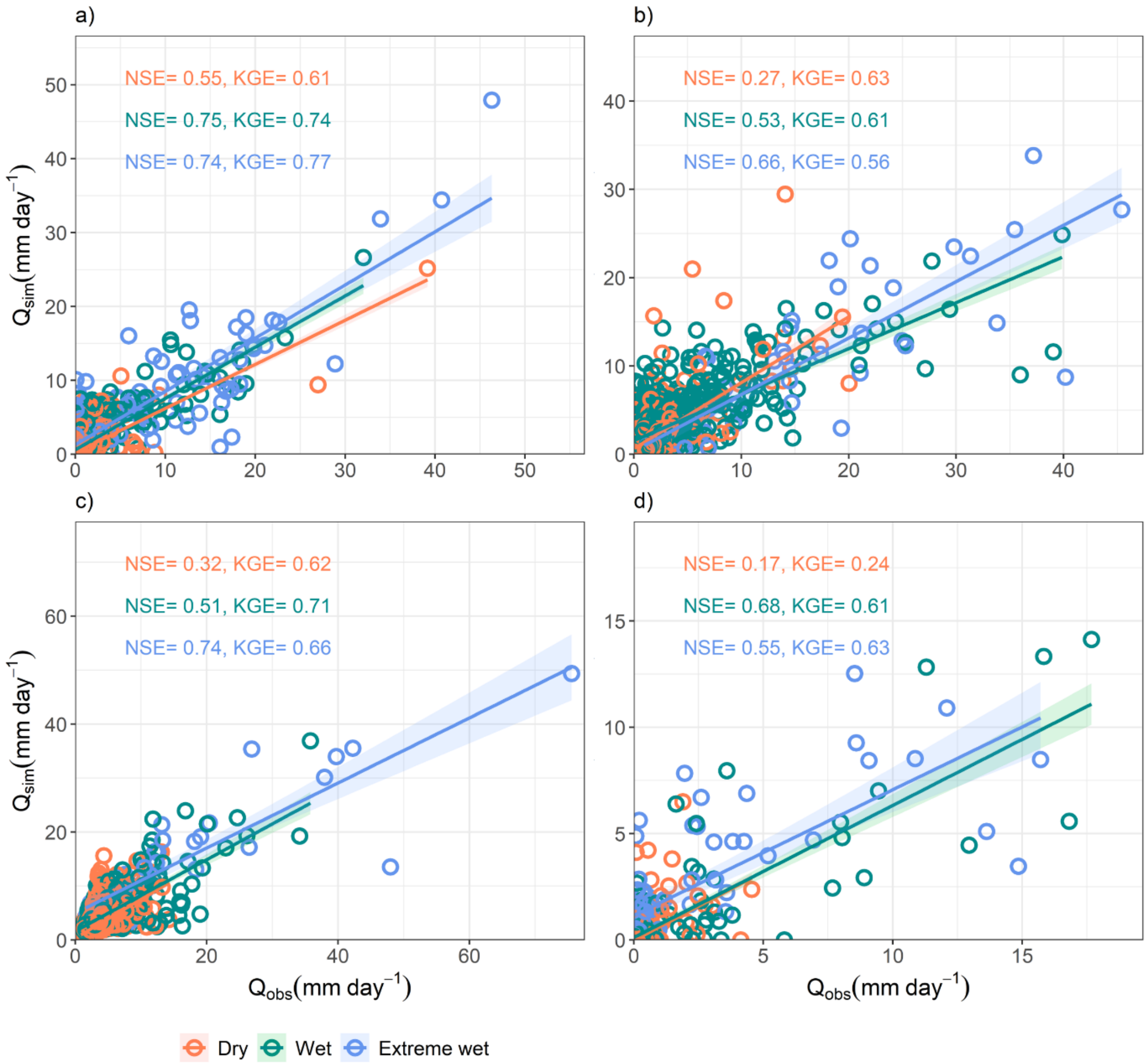


Figure S2 Simulated runoff for dry, wet, and extreme wet conditions by $PET_{MGAM\ [CO2]}$ for four flux tower sites: a) US-MOZ, b) CA-TP3, c) US-Me2, and d) US-ARM.

4.8 References

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome*, *300*(9), D05109.

Arain, M. A., Xu, B., Brodeur, J. J., Khomik, M., Peichl, M., Beamesderfer, E., Restrepo-Couple, N., & Thorne, R. (2022). Heat and drought impact on carbon exchange in an age-sequence of temperate pine forests. *Ecological Processes*, *11*(1), 7. https://doi.org/https://doi.org/10.1186/s13717-021-00349-7

Bai, P., Liu, X., Zhang, Y., & Liu, C. (2020). Assessing the Impacts of Vegetation Greenness Change on Evapotranspiration and Water Yield in China. *Water Resources Research*, *56*(10), e2019WR027019. https://doi.org/https://doi.org/10.1029/2019WR027019

Ballarin, A. S., Sousa Mota Uchôa, J. G., dos Santos, M. S., Almagro, A., Miranda, I. P., da Silva, P. G. C., da Silva, G. J., Gomes Júnior, M. N., Wendland, E., & Oliveira, P. T. S. (2023). Brazilian Water Security Threatened by Climate Change and Human Behavior. *Water Resources Research*, *59*(7), e2023WR034914. https://doi.org/https://doi.org/10.1029/2023WR034914

Ban, Z., & Lettenmaier, D. P. (2022). Asymmetry of Western U.S. River Basin Sensitivity to Seasonally Varying Climate Warming. *Water Resources Research*, *58*(2), e2021WR030367. https://doi.org/https://doi.org/10.1029/2021WR030367

Bass, B., Goldenson, N., Rahimi, S., & Hall, A. (2023). Aridification of Colorado River Basin's Snowpack Regions Has Driven Water Losses Despite Ameliorating Effects of Vegetation. *Water Resources Research*, *59*(7), e2022WR033454. https://doi.org/https://doi.org/10.1029/2022WR033454

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flandrois, J.-P., & Delignette-Muller, M.-L. (2015). A Toolbox for Nonlinear Regression in R: The Package nlstools. *Journal of Statistical Software*, *66*(5), 1 - 21. https://doi.org/https://doi.org/10.18637/jss.v066.i05

Camporese, M., Daly, E., & Paniconi, C. (2015). Catchment-scale Richards equation-based modeling of evapotranspiration via boundary condition switching and root water uptake schemes. *Water Resources Research*, *51*(7), 5756-5771. https://doi.org/https://doi.org/10.1002/2015WR017139

Chen, H., & Wang, S. (2022). Accelerated Transition Between Dry and Wet Periods in a Warming Climate. *Geophysical Research Letters*, *49*(19), e2022GL099766. https://doi.org/https://doi.org/10.1029/2022GL099766

Chen, T., He, T., Khotilovich, M. B. V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., JXie, Lin, M., Geng, Y., Li, Y., & Yuan, J. (2023). *_xgboost: Extreme Gradient Boosting. R package version 1.7.5.1*. https://CRAN.R-project.org/package=xgboost

Chitsaz, N., Guan, H., Shanafield, M., Zhang, L., & Batelaan, O. (2024). The impact of environmental variables on surface Conductance: Advancing simulation with a nonlinear Machine learning model. *Journal of Hydrology*, 131254. https://doi.org/https://doi.org/10.1016/j.jhydrol.2024.131254

Coron, L., Delaigue, O., Thirel, G., Dorchies, D., Perrin, C., & Michel, C. (2023). *airGR: Suite of GR Hydrological Models for Precipitation-Runoff Modelling. R package version 1.7.4*. https://CRAN.R-project.org/package=airGR

Coron, L., Thirel, G., Delaigue, O., Perrin, C., & Andréassian, V. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, *94*, 166-171. https://doi.org/https://doi.org/10.1016/j.envsoft.2017.05.002

De la Fuente, L. A., Gupta, H. V., & Condon, L. E. (2023). Toward a Multi-Representational Approach to Prediction and Understanding, in Support of Discovery in Hydrology. *Water Resources Research*, *59*(1), e2021WR031548. https://doi.org/https://doi.org/10.1029/2021WR031548

Delaigue, O., Brigode, P., Thirel, G., & Coron, L. (2023). airGRteaching: an open-source tool for teaching hydrological modeling with R. *Hydrol. Earth Syst. Sci.*, *27*(17), 3293-3327. https://doi.org/10.5194/hess-27-3293-2023

Fowler, K., Knoben, W., Peel, M., Peterson, T., Ryu, D., Saft, M., Seo, K.-W., & Western, A. (2020). Many Commonly Used Rainfall-Runoff Models Lack Long, Slow Dynamics: Implications for Runoff Projections. *Water Resources Research*, *56*(5), e2019WR025286. https://doi.org/https://doi.org/10.1029/2019WR025286

Guo, D., Zheng, F., Gupta, H., & Maier, H. R. (2020). On the Robustness of Conceptual Rainfall-Runoff Models to Calibration and Evaluation Data Set Splits Selection: A Large Sample Investigation. *Water Resources Research*, *56*(3), e2019WR026752. https://doi.org/https://doi.org/10.1029/2019WR026752

Hou, Y., Guo, H., Yang, Y., & Liu, W. (2023). Global Evaluation of Runoff Simulation From Climate, Hydrological and Land Surface Models. *Water Resources Research*, *59*(1), e2021WR031817. https://doi.org/https://doi.org/10.1029/2021WR031817

Kwon, H., Law, B. E., Thomas, C. K., & Johnson, B. G. (2018). The influence of hydrological variability on inherent water use efficiency in forests of contrasting composition, age, and precipitation regimes in the Pacific Northwest. *Agricultural and Forest Meteorology*, *249*, 488-500. https://doi.org/https://doi.org/10.1016/j.agrformet.2017.08.006

Lee, Y.-G., Oh, J.-Y., Kim, D., & Kim, G. (2023). SHAP Value-Based Feature Importance Analysis for Short-Term Load Forecasting. *Journal of Electrical Engineering & Technology*, *18*(1), 579-588. https://doi.org/10.1007/s42835-022-01161-9

Li, X., Kang, S., Niu, J., Huo, Z., & Liu, J. (2019). Improving the representation of stomatal responses to CO2 within the Penman–Monteith model to better estimate evapotranspiration responses to climate change. *Journal of Hydrology*, *572*, 692-705. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.03.029

Liu, C., Feng, S., Zhang, Q., Hu, J., Ma, N., Ci, H., Kong, D., & Gu, X. (2024). Critical influence of vegetation response to rising CO2 on runoff changes. *Science of the Total Environment*, *906*, 167717. https://doi.org/https://doi.org/10.1016/j.scitotenv.2023.167717

Liu, M., Sun, A. Y., Lin, K., Luo, W., Tu, X., & Chen, X. (2023). Estimating Dynamic Non-Water-Limited Canopy Resistance Over the Globe: Changes, Contributors, and Implications. *Water Resources Research*, *59*(9), e2022WR034209. https://doi.org/https://doi.org/10.1029/2022WR034209

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin Jordan, M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56-67. https://doi.org/https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions* Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA.

Mardian, J., Champagne, C., Bonsal, B., & Berg, A. (2023). A Machine Learning Framework for Predicting and Understanding the Canadian Drought Monitor. *Water Resources Research*, *59*(8), e2022WR033847. https://doi.org/https://doi.org/10.1029/2022WR033847

Mayer, M. (2023). *shapviz: SHAP Visualizations. R package version 0.9.1*. https://CRAN.R-project.org/package=shapviz

McMahon, T. A., Peel, M. C., Lowe, L., Srikanthan, R., & McVicar, T. R. (2013). Estimating actual, potential, reference crop and pan evaporation using standard meteorological data: a pragmatic synthesis. *Hydrol. Earth Syst. Sci.*, *17*(4), 1331-1363. https://doi.org/10.5194/hess-17-1331-2013

Milly, P. C. D., & Dunne, K. A. (2016). Potential evapotranspiration and continental drying. *Nature Climate Change*, *6*(10), 946-949. https://doi.org/https://doi.org/10.1038/nclimate3046

Milly, P. C. D., & Dunne, K. A. (2017). A Hydrologic Drying Bias in Water-Resource Impact Analyses of Anthropogenic Climate Change. *JAWRA Journal of the American Water Resources Association*, *53*(4), 822-838. https://doi.org/https://doi.org/10.1111/1752-1688.12538

Newman, A., Sampson, K., Clark, M. P., Bock, A., Viger, R. J., & Blodgett, D. (2014). *A large-sample watershed-scale hydrometeorological dataset for the contiguous USA. Boulder, CO: UCAR/NCAR*. https://dx.doi.org/10.5065/D6MW2F4D

Partington, D., Thyer, M., Shanafield, M., McInerney, D., Westra, S., Maier, H., Simmons, C., Croke, B., Jakeman, A. J., Gupta, H., & Kavetski, D. (2022). Predicting wildfire induced changes to runoff: A review and synthesis of modeling approaches. *WIREs Water*, *9*(5), e1599. https://doi.org/https://doi.org/10.1002/wat2.1599

Peel, M. C., & McMahon, T. A. (2020). Historical development of rainfall-runoff modeling. *WIREs Water*, *7*(5), e1471. https://doi.org/https://doi.org/10.1002/wat2.1471

Peiris, T. A., & Döll, P. (2023). Improving the quantification of climate change hazards by hydrological models: a simple ensemble approach for considering the uncertain effect of vegetation response to climate change on potential evapotranspiration. *Hydrol. Earth Syst. Sci.*, *27*(20), 3663-3686. https://doi.org/https://doi.org/10.5194/hess-27-3663-2023

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, *279*(1), 275-289. https://doi.org/https://doi.org/10.1016/S0022-1694(03)00225-7

Pimentel, R., Arheimer, B., Crochemore, L., Andersson, J. C. M., Pechlivanidis, I. G., & Gustafsson, D. (2023). Which Potential Evapotranspiration Formula to Use in Hydrological Modeling World-Wide? *Water Resources Research*, *59*(5), e2022WR033447. https://doi.org/https://doi.org/10.1029/2022WR033447

Raz-Yaseef, N., Billesbach, D. P., Fischer, M. L., Biraud, S. C., Gunter, S. A., Bradford, J. A., & Torn, M. S. (2015). Vulnerability of crops and native grasses to summer drying in the U.S. Southern Great Plains. *Agriculture, Ecosystems & Environment*, *213*, 209-218. https://doi.org/https://doi.org/10.1016/j.agee.2015.07.021

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., . . . Maier, H. R. (2021). The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, *137*, 104954. https://doi.org/https://doi.org/10.1016/j.envsoft.2020.104954

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*(5). https://doi.org/https://doi.org/10.1029/2009WR008328

Roderick, M. L., Sun, F., Lim, W. H., & Farquhar, G. D. (2014). A general framework for understanding the response of the water cycle to global warming over land and ocean. *Hydrol. Earth Syst. Sci.*, *18*(5), 1575-1589. https://doi.org/10.5194/hess-18-1575-2014

Roxy, M. K., Ghosh, S., Pathak, A., Athulya, R., Mujumdar, M., Murtugudde, R., Terray, P., & Rajeevan, M. (2017). A threefold rise in widespread extreme rain events over central India. *Nature Communications*, *8*(1), 708. https://doi.org/10.1038/s41467-017-00744-9

Santos, L., Thirel, G., & Perrin, C. (2018). Continuous state-space representation of a bucket-type rainfall-runoff model: a case study with the GR4 model using state-space GR4 (version 1.0). *Geosci. Model Dev.*, *11*(4), 1591-1605. https://doi.org/10.5194/gmd-11-1591-2018

Shi, H., Luo, G., Hellwich, O., He, X., Xie, M., Zhang, W., Ochege, F. U., Ling, Q., Zhang, Y., Gao, R., Kurban, A., De Maeyer, P., & Van de Voorde, T. (2023). Comparing the use of all data or specific subsets for training machine learning models in hydrology: A case study of evapotranspiration prediction. *Journal of Hydrology*, *627*, 130399. https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.130399

Sinha, S., Hammond, A., & Smith, H. (2022). A comprehensive intercomparison study between a lumped and a fully distributed hydrological model across a set of 50 catchments in the

United Kingdom. *Hydrological Processes*, *36*(3), e14544. https://doi.org/https://doi.org/10.1002/hyp.14544

Swann, A. L. S., Hoffman, F. M., Koven, C. D., & Randerson, J. T. (2016). Plant responses to increasing CO₂ reduce estimates of climate impacts on drought severity. *Proceedings of the National Academy of Sciences*, *113*(36), 10019-10024. https://doi.org/doi:10.1073/pnas.1604581113

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, *10*(1), 13768. https://doi.org/10.1038/s41598-020-70816-2

Thom, A. S. (1972). Momentum, mass and heat exchange of vegetation. *Quarterly Journal of the Royal Meteorological Society*, *98*, 124-134. https://doi.org/https://doi.org/10.1002/qj.49709841510

Thyer, M., Gupta, H., Westra, S., McInerney, D., Maier, H. R., Kavetski, D., Jakeman, A., Croke, B., Simmons, C., Partington, D., Shanafield, M., & Tague, C. (2024). Virtual Hydrological Laboratories: Developing the Next Generation of Conceptual Models to Support Decision Making Under Change. *Water Resources Research*, *60*(4), e2022WR034234. https://doi.org/https://doi.org/10.1029/2022WR034234

Vremec, M., Forstner, V., Herndl, M., Collenteur, R., Schaumberger, A., & Birk, S. (2023). Sensitivity of evapotranspiration and seepage to elevated atmospheric CO2 from lysimeter experiments in a montane grassland. *Journal of Hydrology*, *617*, 128875. https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.128875

Wang, A., & Solomatine, D. P. (2019). Practical Experience of Sensitivity Analysis: Comparing Six Methods, on Three Hydrological Models, with Three Performance Criteria. *Water*, *11*(5), 1062. https://www.mdpi.com/2073-4441/11/5/1062

Wang, Y., & Karimi, H. A. (2022). Impact of spatial distribution information of rainfall in runoff simulation using deep learning method. *Hydrol. Earth Syst. Sci.*, *26*(9), 2387-2403. https://doi.org/10.5194/hess-26-2387-2022

Wood, J. D., Knapp, B. O., Muzika, R.-M., Stambaugh, M. C., & Gu, L. (2018). The importance of drought–pathogen interactions in driving oak mortality events in the Ozark Border Region. *Environmental Research Letters*, *13*(1), 015004. https://doi.org/https://doi.org/10.1088/1748-9326/aa94fa

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548-1563. https://doi.org/https://doi.org/10.1080/01621459.2016.1180986

Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R., & Donohue, R. J. (2019). Hydrologic implications of vegetation response to elevated CO2 in climate projections. *Nature Climate Change*, *9*(1), 44-48. https://doi.org/10.1038/s41558-018-0361-0

Yin, J., Gentine, P., Zhou, S., Sullivan, S. C., Wang, R., Zhang, Y., & Guo, S. (2018). Large increase in global storm runoff extremes driven by climate and anthropogenic changes. *Nature Communications*, *9*(1), 4389. https://doi.org/https://doi.org/10.1038/s41467-018-06765-2

Zhang, C., Yang, Y., Yang, D., & Wu, X. (2021). Multidimensional assessment of global dryland changes under future warming in climate projections. *Journal of Hydrology*, *592*, 125618. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125618

Zhang, X.-Y., Jin, J., Zeng, X., Hawkins, C. P., Neto, A. A. M., & Niu, G.-Y. (2022). The Compensatory CO2 Fertilization and Stomatal Closure Effects on Runoff Projection From 2016–2099 in the Western United States. *Water Resources Research*, *58*(1), e2021WR030046. https://doi.org/https://doi.org/10.1029/2021WR030046

Zhang, Y., Zheng, H., Zhang, X., Leung, L. R., Liu, C., Zheng, C., Guo, Y., Chiew, F. H. S., Post, D., Kong, D., Beck, H. E., Li, C., & Blöschl, G. (2023). Future global streamflow declines are probably more severe than previously estimated. *Nature Water*, *1*(3), 261-271. https://doi.org/10.1038/s44221-023-00030-7

Zhou, S., Yu, B., Lintner, B. R., Findell, K. L., & Zhang, Y. (2023). Projected increase in global runoff dominated by land surface changes. *Nature Climate Change*, *13*(5), 442-449. https://doi.org/https://doi.org/10.1038/s41558-023-01659-8

Chapter 5: Conclusion

5.1 Conclusion

This research work had two aims. First, it aimed to improve g_s simulation using a nonlinear ML model to achieve realistic results by investigating the interactive effects of environmental variables on various vegetation types. The second overall aim of this study was to incorporate g_s into modified PET equation to enhance runoff simulation.

For the first aim, we utilised mixed generalized additive models (MGAM) as a nonlinear ML model in g_s simulation to allow the direct and interactive effects between environmental variables. The results of NSE showed that MGAM approach improved g_s simulations compared to conventional models such as empirical and semi-empirical simulation models. Additionally, global sensitivity analysis indicated lower uncertainty in MGAM g_s simulation compared to conventional models (addressed in Chapter 2). Generalising MGAM across different vegetation types highlighted the importance of key environmental variables in g_s simulation. The MGAM g_s simulation results showed that the interactive effects of CO_2, VPD, and SWC were important for crops and grasses, while the interactive effects of CO_2, VPD, and TA were highlighted for trees and grasses (addressed in Chapter 3).

For the second aim, we incorporated MGAM g_s simulation into PET equations; and the modified PET showed higher NSE than the conventional Penman-Monteith PET equation. The modified PET equation improved runoff simulation for different climate and vegetation types. This modification underscored the importance of vegetation's role in hydrological processes (addressed in Chapter 4). MGAM's definition of interactive effects of environmental variables can better elucidate the dominant factors influencing PET and runoff changes, which is of significant value for water resource management and decision-makers.

5.2 Outlook

Our research has focused on enhancing the accuracy of evaporation and runoff simulations by integrating CO_2 and environmental variables into stomatal conductance models using machine learning (ML) and non-linear statistical models. In future, hydrological modelling undergo significant advancements driven by several emerging trends and technological developments (Yang et al., 2021). High-resolution remote sensing data and advanced big data analytics will provide detailed spatial and temporal information, enhancing model precision and enabling real-time monitoring and prediction of hydrological events ("Farmer First:Shifting Paradigms in Agricultural Technology Development," 2011). Additionally, the ability of ML algorithms to handle complex, non-linear relationships and large datasets will make them inevitable tools for simulating hydrological processes under changing environmental conditions (Kalu et al., 2022; Zhu et al., 2022). AI-driven models will offer more accurate predictions of extreme weather events, such as floods and droughts, and better insights into the impacts of climate change on hydrological cycles (Latif & Ahmed, 2023).

The future will also see the widespread adoption of hybrid models that combine the strengths of conceptual, physically-based, and ML models (Zhao et al., 2019). These hybrid approaches will leverage the data-driven capabilities of ML while incorporating physical constraints and domain knowledge from traditional hydrological models (Reichstein et al., 2019). This will enable more comprehensive and reliable simulations of hydrological processes, particularly in diverse and dynamic environments (Koppa et al., 2022). Furthermore, the incorporation of CO_2 and other environmental variables into standard modelling practices will improve the accuracy of evapotranspiration and runoff simulations, accounting for the effects of changing vegetation cover and atmospheric conditions on hydrological cycles (Yang et al., 2019; Zhang et al., 2021).

Our research significantly contributes to these future trends. By incorporating CO_2 and environmental variables into stomatal conductance models, we have improved the accuracy of evaporation and runoff simulations, laying the groundwork for more precise and dynamic hydrological models. The integration of ML and non-linear statistical models in our research demonstrates their potential to handle complex interactions and non-linearities in hydrological processes, providing a robust framework for future modelling efforts. Our proposed hybrid modelling framework combines the strengths of physically-based and ML models, addressing the limitations of each approach and offering a pathway for future models to effectively simulate internal catchment processes and environmental changes across a wide range of scenarios and locations.

The advancements made through this research position hydrological modelling to become more accurate, adaptive, and useful for addressing future environmental challenges. The improved accuracy and reliability of our models will support better decision-making in water resource management, climate change adaptation, and disaster risk reduction. Our work underscores the importance of long-term investigations into how natural and anthropogenic changes affect hydrological cycles, providing valuable insights for policymakers and water resource managers in developing strategies for sustainable water management and climate resilience.

5.3 References

Farmer First:Shifting Paradigms in Agricultural Technology Development. (2011). *Future Agriculture*. https://www.future-agricultures.org/wp-content/uploads/pdf-archive/CAADP_Policy_Brief_04_english.pdf

Kalu, I., Ndehedehe, C. E., Okwuashi, O., Eyoh, A. E., & Ferreira, V. G. (2022). An assimilated deep learning approach to identify the influence of global climate on hydrological fluxes. *Journal of Hydrology*, *614*, 128498. https://doi.org/https://doi.org/10.1016/j.jhydrol.2022.128498

Koppa, A., Rains, D., Hulsman, P., Poyatos, R., & Miralles, D. G. (2022). A deep learning-based hybrid model of global terrestrial evaporation. *Nature Communications*, *13*(1), 1912. https://doi.org/https://doi.org/10.1038/s41467-022-29543-7

Latif, S. D., & Ahmed, A. N. (2023). A review of deep learning and machine learning techniques for hydrological inflow forecasting. *Environment, Development and Sustainability*, *25*(11), 12189-12216. https://doi.org/10.1007/s10668-023-03131-1

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195-204. https://doi.org/https://doi.org/10.1038/s41586-019-0912-1

Yang, D., Yang, Y., & Xia, J. (2021). Hydrological cycle and water resources in a changing world: A review. *Geography and Sustainability*, *2*(2), 115-122. https://doi.org/https://doi.org/10.1016/j.geosus.2021.05.003

Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R., & Donohue, R. J. (2019). Hydrologic implications of vegetation response to elevated CO2 in climate projections. *Nature Climate Change*, *9*(1), 44-48. https://doi.org/10.1038/s41558-018-0361-0

Zhang, C., Yang, Y., Yang, D., & Wu, X. (2021). Multidimensional assessment of global dryland changes under future warming in climate projections. *Journal of Hydrology*, *592*, 125618. https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125618

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., Lin, C., Li, X., & Qiu, G. Y. (2019). Physics-Constrained Machine Learning of Evapotranspiration. *Geophysical Research Letters*, *46*(24), 14496-14507. https://doi.org/https://doi.org/10.1029/2019GL085291

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health*, *1*(2), 107-116. https://doi.org/https://doi.org/10.1016/j.eehl.2022.06.001