Sensitivity of simulated soil water content, evapotranspiration, gross primary production and biomass to climate change factors in Euro-Mediterranean grasslands

Gianni Bellocchi, Z. Barcza, R. Hollós, M. Acutis, E. Bottyán, L. Doro, D. Hidy, E. Lellei-Kovács, S. Ma, J. Minet, et al.

Sensitivity of simulated soil water content, evapotranspiration, gross primary production and biomass to climate change factors in Euro-Mediterranean grasslands

G. Bellocchi [a], Z. Barcza [b,c,*], R. Hollós [b,d], M. Acutis [e], E. Bottyán [b], L. Doro [f,g], D. Hidy [h], E. Lellei-Kovács [i], S. Ma [j], J. Minet [k], V. Pacskó [l,m], A. Perego [e], F. Ruget [n,1], G. Seddaiu [f], L. Wu [o], R. Sándor [a,d]

[a] UCA, INRAE, VetAgro Sup, Unité Mixte de Recherche sur l'Écosystème Prairial (UREP), Clermont-Ferrand, France
[b] ELTE Eötvös Loránd University, Department of Meteorology, Budapest, Hungary
[c] Czech University of Life Sciences Prague, Faculty of Forestry and Wood Sciences, Prague, Czech Republic
[d] HUN-REN Centre for Agricultural Research, Agricultural Institute, Martonvásár, Hungary
[e] University of Milan, Department of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy, Milan, Italy
[f] University of Sassari, Desertification Research Centre, Sassari, Italy
[g] Texas A&M AgriLife Research, Blackland Research Center, Temple (TX), United States of America
[h] Hungarian University of Agriculture and Life Sciences, MTA-MATE Agroecological Research Group, Gödöllő, Hungary
[i] HUN-REN Centre for Ecological Research, Institute of Ecology and Botany, Vácrátót, Hungary
[j] Tottori University, International Platform for Dryland Research and Education, Tottori, Japan
[k] University of Liège, Arlon Environment Campus, Arlon, Belgium
[l] Lechner Knowledge Center, Space Remote Sensing Unit, Hungary
[m] ELTE Eötvös Loránd University, Doctoral School of Earth Sciences, Budapest, Hungary
[n] French National Institute for Agricultural Research, Modelling Agricultural and Hydrological Systems in the Mediterranean Environment, Avignon, France
[o] Rothamsted Research, North Wyke, Okehampton, United Kingdom

ARTICLE INFO

ABSTRACT

Grassland models often yield more uncertain outputs than arable crop models due to more complex interactions and the largely undocumented sensitivity of grassland models to environmental factors. The aim of the present study was to assess the impact of single-factor changes in temperature, precipitation, and atmospheric [CO_2] on simulated soil water content (SWC), actual evapotranspiration (ET), gross primary production (GPP) and yield biomass, and also to link the sensitivity analysis with experimental results. We employed an unprecedented multi-model framework consisting of seven grassland models at nine sites with different environmental characteristics in Europe and Israel, with two management options at three sites. For warming/cooling and wetting/drying, models showed general consistency in the direction of SWC and ET changes, but less agreement regarding GPP and biomass changes. The simulated responses consistently revealed an overall positive effect of CO_2 enrichment on GPP and biomass, while the direction of change differed for SWC and ET. Comparing with single-factor experimental manipulations, SWC simulations slightly underestimated the observed effect of warming, while the overall mean model sensitivity for biomass (+7.5%) closely matched the mean response observed with 1–2 °C warming (+6.6%). The models exhibited lower sensitivity of SWC to wetting or drying compared to the experiments. The overall mean sensitivity of biomass to drying was -4.3%, contrasting with the mean experimental effect size of -9.6%, which proved to be more realistic than the mean wetting effect (+3.2%, against +38.9% in the field trials). The simulated sensitivity of SWC to CO_2 enrichment was markedly underestimated, while the biomass response (+12.0%) closely matched the observations (+17.5%). Although the multi-model averaging did not manifestly improve the realism of the simulations, it ensured a realistic response in the direction of change to varying conditions. The results suggest a paradigm shift in grassland modelling meaning that the usual practice of model optimisation/validation needs to be complemented by a sensitivity analysis following

the approach presented. The results also highlight the importance of model improvements, especially in terms of soil hydrology representation, a key environmental driver of grassland functioning.

1. Introduction

Biophysical and biogeochemical agro-ecosystem models represent key processes such as plant photosynthesis, ecosystem respiration and evapotranspiration related to environmental drivers and management practices (e.g. grazing, cutting, fertilisation). As such, they can quantify the energy and matter exchanges in the ecosystem (e.g. Brilli et al., 2017) and are widely used to quantify ecosystem services (e.g. Pokovai et al., 2020). In this way, the models provide an integrated system perspective in which vegetation, soil, weather and management factors interact dynamically to simulate system feedbacks and revise management alternatives (Antle et al., 2017; Tonitto et al., 2018). These models are widely used to estimate nutrient and water exchanges between the atmosphere and the biosphere, as well as to project the potential impacts of climate changes while examining adaptation options by local stakeholders and policy makers (White et al., 2011). However, a number of issues have been raised regarding the uncertainties (i.e. any deviation from the unattainable ideal of fully deterministic knowledge of a system) of using ecosystem models (Tubiello et al., 2007; Soussana et al., 2010). This is especially true in simulating climate change impacts. Climatic changes continue to challenge the scientific and policy-making communities. Therefore, efforts to model the inherently unknown future behaviour of complex and inter-related systems have led to a focus on the uncertainties associated with framing possible future outcomes (e.g. Schwanitz, 2013; Butler et al., 2014).

The sensitivity of models to changing global environmental conditions (notably, temperature, precipitation and atmospheric carbon dioxide concentration [CO_2]) is well documented for annual crops (e.g. Asseng et al., 2013 for wheat; Bassu et al., 2014 for maize; Li et al., 2015 for rice), but such information is largely missing for grasslands. Grassland models simulate the herbage mass and take into account sward components, including representations of grazing processes. Several aspects of grassland modelling pose unique challenges, mostly stemming from the need to represent several interacting species, and the persistence of plants over several years, forcing models to consider residual effects over time (Kipling et al., 2016). While most biophysical processes in grassland models are similar to those used in arable crop models (e.g., photosynthesis, growth, water and nutrient uptake from the soil), there are additional factors specific to grasslands (Jones et al., 2017). These factors include grazing or mowing practices, interactions within plant species, and the influence of biodiversity on ecosystem dynamics. Grassland models also account for the seasonal variability of forage availability and quality, as well as the impacts of management practices on vegetation dynamics. Incorporating these factors, grassland models provide a comprehensive understanding of the unique characteristics and dynamics of grassland ecosystems.

Here, the focus is on permanent grasslands, which are semi-natural or have been sown since at least five years (EC, 2008), usually composed of several plant species. Grassland models can represent different environments with adequate accuracy, particularly, if the input information is sufficiently detailed (e.g. Graux et al., 2013). However, simulations of carbon (C), water and energy fluxes are inherently uncertain because grasslands are highly complex ecosystems and their behaviour is affected by multifaceted interactions of climate drivers with water availability, nutrients, soil, vegetation and management conditions (Soussana et al., 2013). A thorough assessment of the variation in the response of different grassland models to climate change factors is thus essential to assess the provision and the continuity of grassland services in the future. Before projecting what the future achievements might be under a changing climate using coupled climate and impact models, it is critical to determine the extent to which grassland simulations may vary depending on the model used.

To enlighten communities on the need to improve and evaluate grassland models, detailed grassland data were collected and synthesised from alternative sources as part of the FACCE-MACSUR project ("Modelling European Agriculture with Climate Change for Food Security"; https://www.faccejpi.net/en/faccejpi/actions/core-theme-1/knowledge-hub-macsur.htm). In the first part of this effort (Sándor et al., 2017), three grassland-specific models as well as three grassland-adapted crop models and three land biosphere models (describing vegetation by plant functional types including grasslands), were evaluated with respect to yielded biomass, soil water and temperature dynamics at a variety of sites across Europe and Israel, both without and with calibration. Comparing model simulations with observations and contrasting model results helped to identify biophysical processes that are needed to be improved to better simulate grassland functioning. Being process-based, these models are often employed to simulate the effects of projected climate changes on grassland vegetation and soil cycles. However, these models have not been directly (or, in the case of some models, only to a limited extent) assessed to determine their sensitivity to environmental factors. Sándor et al. (2017) suggested to "progress in grassland modelling by investigating the sensitivity of models to climate and management drivers". There is a challenge in ensuring robust modelling approaches under changing climate conditions, as the implicit assumption that well-designed and calibrated models under current conditions will remain valid under future climate realisations (with, coincidentally, a shift in the pattern of extreme climate events being reported; Field et al., 2012) may be unrealistic.

This study builds upon the work by Sándor et al. (2017), who documented model calibration in the same study sites. In this study, the primary objective was to explore how grassland models differ in simulating the response to climate change factors like temperature, precipitation and atmospheric CO_2 concentration. Analysing the sensitivity of model simulations to gradients of these factors can provide insights into model structure and system behaviour not evident from modelling exercises performed under current climate conditions. In fact, the aim of this one-factor-at-a-time (single-factor) sensitivity analysis used in this study was to reveal the relationship between the varied factors and outputs, as the other factors had their baseline values. By showing these relationships, sensitivity analysis can help to understand model mechanisms (e.g. ten Broeke et al., 2016), as unexpected sensitivity (or insensitivity) can lead to questioning mechanistic understanding and the relative roles of different processes.

In order to analyse the single-factor sensitivity of simulated dry matter, water and C fluxes to climate change factors, modelling scenarios were designed using a unique set of European and Israeli grassland sites, where models were run under altered weather conditions created by changing temperature, precipitation and atmospheric [CO_2] inputs. To do this, we varied each climate change factor within a predefined range and evaluated the resulting output. Variations of the drivers in the sensitivity runs were prescribed primarily to be large enough to identify and isolate any major differences in model responses (interacting with actual management), while remaining within the range of multi-year variability to which European and Israeli grassland ecosystems may be exposed. By focusing on three climate change factors and varying only one at a time, we deliberately narrowed the scope of the analysis not only to simplify the modelling work, but also, and more importantly, to compare the sensitivity results with experimental evidence from published trials in which the same factors were varied (ecosystem manipulation experiments). In this way, this study links model sensitivity analysis with experimental results, providing evidence of the relevance of model responses to varying factors.

By executing virtual (*in silico*) single-factor ecosystem manipulations based on an exceptionally large number of target sites and grassland models, and focusing on four output variables, this study addresses the following main question: (1) Is there consistency amongst model results in terms of the direction and magnitude of response to climate change factors? (2) To what extent does model sensitivity vary across different sites and how does it relate to environmental conditions? (3) Are model results consistent with those of published single-factor climate manipulation experiments?

The study aims to diagnose the grassland models and provide an alternative method to benchmark them, and most importantly to provide a solid reference for future studies. Comparing the structure of the models and relating the differences to the findings is out of scope of the present study as it goes well beyond the context of the present work.

2. Materials and methods

2.1. Core logic of the study

In this study, we employ a multi-model framework with distinct characteristics which sets it apart from studies that primarily focus on single-model applications (Asseng et al., 2013; Bassu et al., 2014; Li et al., 2015; Martre et al., 2015; Sándor et al., 2017; Wallach et al., 2018). The key feature of this approach lies in the extensive inclusion of multiple models, diverse sites and varying conditions, where individual models are not discussed. Instead, the focus is on analysing model results together, emphasising the ensemble behaviour of the models. To facilitate this analysis, the models are treated as "black boxes" within such studies. They serve as tools that generate results influenced by various modelling practices, including overall parameterisation, consistency checks and validation. By treating the models in this manner, we can better understand and interpret their outputs in the context of the study's objectives.

To quantify the sensitivity of seven grassland models to various climate change factors within a multi-model framework, we adopt a large ensemble approach to comprehensively explore how these models simulate and respond to changes in environmental conditions. Furthermore, a significant aspect of this study is the thorough comparison of model sensitivity with observation data collected from field experiments. By incorporating observational data, we enhance the robustness of our analysis and provide a basis for validating and refining the models' representation of real-world conditions.

2.2. Study sites

Nine long-term grassland sites across Europe and Israel were selected for this study, spanning data from five to 31 years (Fig. 1, Table 1).

These sites exhibit a wide range of climatic conditions (Table S1 in Supplementary material) and soil properties (Table S2 in Supplementary material), making them ideal candidates for evaluating the models' behaviour. Four of them – Laqueuille (LAQ), Monte Bondone (MBO), Grillenburg (GRI), Oensingen (OEN) – are long-established semi-natural grasslands with vegetation types representative of the zone, equipped with flux-tower measuring systems where the volumetric soil water content (SWC), evapotranspiration (ET) and gross primary production (GPP) were recorded with half-hourly or hourly resolution amongst others.

Five yield-specific experimental sites are focused on the harvested aboveground biomass production. At some sites (LAQ, KEM, ROT), data from two management options were available. The French flux-tower site of Laqueuille was grazed with two intensities of management using two levels of nitrogen (N) fertilisation (LAQ1: intensive; LAQ2: extensive). At the German site of Kempten, grassland biomass was harvested at two cutting intensities (KEM1: two cuts per year; KEM2: four cuts per year). In Rothamsted (United Kingdom), the same amount of N was applied in two forms (ROT1: NO_3-N; ROT2: NH_4-N). Hereafter,
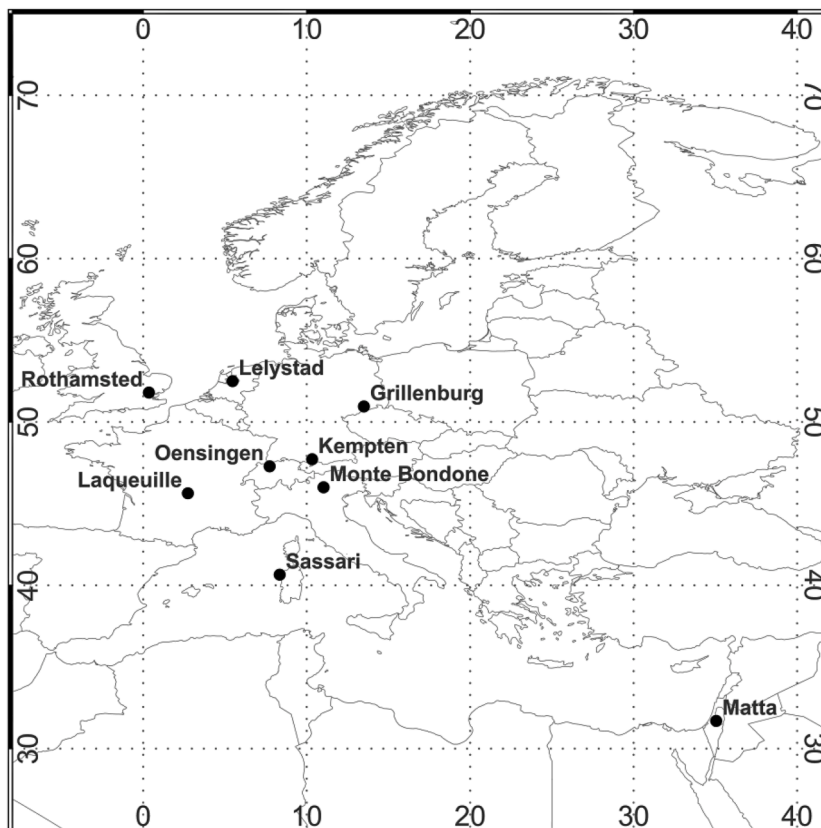

Fig. 1. Geographical location of the grassland sites used in the study.

Table 1
Grassland sites used in this study.

Site (Country) / Characteristics	Flux-tower (FT) sites				Non-flux-tower (NFT) sites				
	Laqueuille (France)	Monte Bondone (Italy)	Grillenburg (Germany)	Oensingen (Switzerland)	Kempten (Germany)	Lelystad (The Netherlands)	Matta (Israel)	Rothamsted (United Kingdom)	Sassari (Italy)
Acronym	LAQ	MBO	GRI	OEN	KEM	LEL	MAT	ROT	SAS
Latitude	45°38′ N	46°00′ N	50°57′ N	47°17′ N	47°43′ N	52°30′ N	31°42′ N	51°48′ N	40°39′ N
Longitude	02°44′ E	11°02′ E	13°30′ E	07°44′ E	10°20′ E	05°28′ E	35°03′ E	00°21′ E	08°21′ E
Elevation (m a.s.l.)	1040	1500	380	450	730	-4	620	128	68
Grassland utilisation	Grazing	Mowing	Mowing	Mowing	Mowing[1]	Mowing	Mowing	Mowing	Mowing
N fertilisation (kg N ha^{-1} yr^{-1})	210 (LAQ1), 0 (LAQ2)	–	–	214	–	–	–	48[2]	90
Years of simulation	2004–2010	2003–2010	2004–2008	2002–2008	2004–2009	1994–1998	2007–2011	1981–2001	1993–1998
Reference	Klumpp et al. (2011)	Wohlfahrt et al. (2008)	Prescher et al. (2010)	Ammann et al. (2007)	Schröpel and Diepolder (2003)	Schils and Snijders (2004)	Golodets et al. (2013)	Silvertown et al. (2006)	Cavallero et al. (1992)

[1] Two treatments: intensive (KEM1) and extensive (KEM2) cutting. [2] Two treatments: NH$_4$-N (ROT1) and NO$_3$-N (ROT2) fertilisers.

the term "site" identifies either a physical location or a location-management combination.

The range of climatic conditions represented by each site is given by indicators of aridity and heat wave frequency (Supplementary material, Fig. S1).

2.3. Grassland models

Sensitivity simulations were performed at each location using seven models that had undergone a full assessment in a previous study (Sándor et al., 2017). At the four flux-tower sites, the models were assessed against detailed soil temperature and water flux datasets, while the evaluation was limited to considering harvested biomass data at the other sites. A brief description of the models follows, while details of the processes and parameter settings that distinguish each of them are provided in Supplementary material of Sándor et al. (2017). The identities of models were kept anonymous by using model codes from 1 to 7 (as in Sándor et al., 2017). Two models – PaSim and SPACSYS – were originally developed to simulate grassland systems. PaSim (Ma et al., 2015) simulates water, C and N cycles in grassland plots at sub-daily time step via climate, soil biology and physics, vegetation and management (including grazing animals) modules. SPACSYS (Wu et al., 2007) is a multi-dimensional, field-scale, daily time step model of C and N cycles between plants, soils and microbes, with a fine representation of the root system. The following three models – ARMOSA, EPIC and STICS – are generic crop models with grassland options. ARMOSA (Perego et al., 2013) estimates N dynamics in the soil-crop-atmosphere continuum and evaluates the impact of management on shallow and groundwater quality via modules for energy, water, C and N balances, and plant development and growth. EPIC, originally developed to estimate soil productivity affected by erosion (Williams et al., 2008), is designed to allow the simulation of a large variety of crops and grasses with unique parameter values. STICS (Brisson et al., 2003) is a generic, daily-step, patch-scale model that covers many crops and climate, soil and management conditions, and is configured to simulate either sown or established mown grasslands. Two vegetation models – Biome-BGCMuSo and CARAIB – include grasslands as a biome type. Biome-BGCMuSo (Hidy et al., 2016) implements a multilayer soil module, improved grassland phenology and management routines in Biome-BGC, originally developed to simulate undisturbed ecosystems, with allometric relationships used to initialize C and N pools. CARAIB (Warnant et al., 1994), a vegetation model based on C assimilation processes in the biosphere, uses a range of plant functional types, including C3 and C4 grasses.

2.4. Simulation study design

The above models were chosen for their suitability in exploring the responses to climate change factors and their relevance to grassland ecosystems. We performed simulations for each site, varying three global change factors: temperature (T), precipitation (PREC) and atmospheric CO_2 concentration ([CO_2]). The conceptual framework of the approach is presented in the Supplementary Material.

Standardised manipulation scenarios were applied to each factor to ensure comparability of the simulation results for the following output variables: soil water content, evapotranspiration, GPP and harvested aboveground biomass (HAB; referred to here as yield biomass or simply biomass and used as a generic proxy for aboveground biomass). To understand the responses and behaviour of the models, a single-factor (T, PREC or [CO_2]) approach was designed.

At each site, the baseline weather series (including [CO_2]=380 ppmv, that is the molar fraction of CO_2 in dry air expressed in parts per million volume or μmol mol^{-1}) were modified by changing (after Sándor et al., 2016): 1) temperatures by -25%, -10%, -5%, +5%, +10%, +25% of the observed standard deviation; 2) precipitations by -25%, -10%, -5%, +5%, +10%, +25% of the observed standard deviation (calculations made on rainy hours or days); atmospheric [CO_2] by +5%, +10%, +15%, +25%, +50%, +100%. A single standard deviation (*SD*) value was calculated for each variable at each site for the whole period of available temperature and precipitation observations (Table 2), based on the hourly weather reports available at the flux sites and the daily data available at the other sites.

The construction of thermal and pluviometric gradients calculated from percentage values of *SD* is meant to take into account local variability, reflecting a larger gradient for sites with higher *SD*. In this way, the individual sensitivity runs were not intended to be modelling scenario studies, which can be defined as using synoptic and physically consistent changes or variations in all drivers. We did not attempt to link changes in a certain driver (e.g. atmospheric [CO_2]) to changes in a potentially related driver (e.g. surface air temperature). Results obtained by using climate scenarios or purely climatological configurations are instead reported in companion global studies (as in the frame of AgMIP, http://www.agmip.org; e.g. Franke et al., 2021). For this study, variations in single drivers were prescribed primarily to be large enough (Table 2) to identify and isolate any major differences in model responses (interacting with actual management), while remaining within the range of multi-year variability to which European and Israeli grassland ecosystems may be exposed.

The model simulations were performed by each modelling team (as described in Sándor et al., 2017). The simulation results were collected

Table 2

[CO$_2$], temperature (T) and precipitation (PREC) scenarios for sensitivity analysis at each site (*SD*=standard deviation).

Site	Atmospheric [CO$_2$] (ppmv)						
	baseline	+5%	+10%	+15%	+25%	+50%	+100%
	380	399	418	437	475	570	760
	Mean annual air temperature (°C)						
	-25%·*SD*	-10%·*SD*	-5%·*SD*	baseline	+5%·*SD*	+10%·*SD*	+25%·*SD*
LAQ (*SD*=7.29)	5.90	6.62	7.35	7.71	8.08	8.81	9.54
MBO (*SD*=7.59)	3.33	4.09	4.85	5.23	5.61	6.37	7.13
GRI (*SD*=8.21)	6.50	7.32	8.14	8.55	8.96	9.78	10.61
OEN (*SD*=8.34)	7.22	8.05	8.89	9.30	9.72	10.56	11.39
KEM (*SD*=7.94)	6.36	6.93	7.39	7.64	7.88	8.33	8.79
LEL (*SD*=6.49)	8.08	8.73	9.38	9.71	10.03	10.68	11.33
MAT (*SD*=8.10)	21.40	22.21	23.04	23.42	23.93	24.65	25.45
ROT (*SD*=5.52)	8.53	9.12	9.55	9.87	10.13	10.72	11.31
SAS (*SD*=5.98)	14.82	15.42	16.02	16.31	16.62	17.21	17.81
	Mean annual precipitation total (mm)						
	-25%·*SD*	-10%·*SD*	-5%·*SD*	baseline	+5%·*SD*	+10%·*SD*	+25%·*SD*
LAQ (*SD*=1.6)	734	850	998	1085	1166	1334	1502
MBO (*SD*=2.2)	705	796	916	1003	1092	1269	1445
GRI (*SD*=1.5)	635	730	860	946	1033	1207	1382
OEN (*SD*=7.4)	843	954	1102	1197	1293	1485	1676
KEM (*SD*=7.5)	832	922	1027	1089	1161	1302	1443
LEL (*SD*=5.4)	614	682	761	813	876	999	1123
MAT (*SD*=11.6)	334	366	405	432	472	549	627
ROT (*SD*=5.0)	534	593	664	710	760	858	956
SAS (*SD*=7.6)	408	453	507	538	571	637	703

and pre-processed centrally before evaluation. The quality control included consistency check of the daily model results (see Supplementary material for more details).

The sensitivity of the individual models at each site, for the three global change factors and for the four output variables, was calculated. Our definition of sensitivity is the percent change (*S*) in an output variable (*O*) that results from a change in a given factor (*R*). We calculated *S* based on the mean annual output variables ($\overline{O}$) relative to the reference (i.e. the calibrated run from Sándor et al., 2017) at each site ($\overline{O_b}$):

$$S = 100 \cdot \frac{\overline{O} - \overline{O_b}}{\overline{O_b}} \quad (1)$$

This definition allowed us to directly compare the simulation results with experimental evidence from a variety of sites worldwide. Experimenters typically report the results of single-factor climate change experiments as effect size (Leuzinger et al., 2011; Andresen et al., 2018), which is in fact defined in the same way as *S* in our case. In this study, the sensitivity of the models to global change factors is also expressed as the slope of the linear function fitted to the individual *S* values (see below) as function of the relative change of the factor studied. In this sense, the slope provides generalised information on the sensitivity of the model where the linear relationship approximates *S* as a function of the unit change of the given factor. To provide unambiguous information about sensitivity measures, we use the term 'effect size' when discussing individual responses. In other occasions, the text defines the meaning of the sensitivity measure.

SWC was typically simulated at all sites by the models and has proven straightforward to use. SWC refers to the topsoil (0–10 cm depth) for all models (Sándor et al., 2017). For ET and GPP, some modelling groups simulated those variables only at the eddy covariance sites, but some groups also provided them for the biomass sites. We used all simulations of ET and GPP when available.

Yield biomass was simulated as HAB at sites with mowing (all sites except Laqueuille; Table 1) and was extracted from simulated data at the day of cutting. At Laqueuille, only grazing is present, but HAB observations made in specific days with destructive sampling were available for model calibration/evaluation (see Sándor et al., 2017). For our analysis, HAB was set equal to the modelled aboveground biomass data

at sampling days in each year, for both intensive and extensive sites. The constructed yield dataset represents the overall productivity (and economic value of the grasslands) and is the production indicator at each site (although it is not identical to net primary production).

In order to compare the response of the different models for the four output variables at the 12 study sites, a linear approximation was first used (Cure and Acock, 1986). A linear regression was performed (using R's inbuild *lm*() function; R Core Team, 2023) to approximate the slope of the effect size as a function of change in the given environmental variable (T, PREC or [CO$_2$] in physical units). In order to determine if the slope of the regression was significantly different from zero (the null hypothesis posited a slope of zero, indicating no significant effect of the environmental variable), a t-statistic test was conducted using the *lm*() function. In some cases, the sensitivity (effect size) showed a non-linear response to the climate change factors. In these cases, the slope of the linear regression still holds information on the sensitivity, but logically the shape of the response function is also a property of the given model, mentioned in the Results section. Note that for T and PREC, non-linearity did not typically mean curvature in the results within the range of the simulations, but rather that it was represented by some noise with an overall linear response. In the case of [CO$_2$], for some model-site combinations, there was a clear curvature (asymptotic behaviour indicating saturation) in sensitivity. In these cases, the slope of the linear regression still captures the overall response. In this study, the slopes are reported in the form of heat maps to check the consistency of the results per site and per model.

According to the core logic of the study (Section 2.1) the multi-model mean was also calculated for each site separately. For the temperature and precipitation manipulations, most of the results indicated a linear response (at least in the ranges studied), but it was associated with asymmetry in some cases (i.e. a different response for warming and cooling, as well as a different response for wetting and drying). To deal with asymmetry, two separate lines were fitted to the data (i.e. a piecewise linear regression was performed) such that the linear function was forced to pass through the origin, [0,0], corresponding to reference climate. One line was fitted to the sensitivity belonging to the positive T/ PREC scenarios, and another to the negative scenarios. As both lines were forced through the origin (zero T/ PREC change), the two lines are

still connected and represent the cooling/warming effect (in the case of T), and wetting/drying (in the case of PREC). We report the slopes as the result of the sensitivity of the multi-model ensemble for each particular site. For $[CO_2]$, the response (S plotted against increasing $[CO_2]$) was non-linear is some cases, so here the ensemble sensitivity was expressed as the effect size belonging to the 760 ppmv scenario (i.e. doubled $[CO_2]$ concentration; see Table 2). Note that the linear regression forced through the reference gave similar results. In the case of $[CO_2]$, the change is unidirectional (only $[CO_2]$ increase is simulated), so the sensitivity is expressed as a single number instead of two as in case of T and PREC.

We note that the linear approximation is only valid within the ranges defined by the extent of the manipulations (Table 2), thus extrapolation outside the bounds of the ranges is not proposed. Any deviation from the linear approximation may be due to missing data in some cases, which are difficult to separate from the modelled (causal) responses. In this sense, although some multi-model responses show a small curvature also for T and PREC, it does not affect the robustness of the results. The slope of the regression lines should be interpreted as an approximation to the actual sensitivity function (effect size versus the extent of the manipulation).

2.5. Data collection on experimental data

To assess the realism of the individual simulations, as well as the realism of the multi-model ensemble mean, we conducted a literature review in which we collected published data from single-factor ecosystem manipulation experiments focusing on the four studied output variables (SWC, ET, GPP and HAB). Google Scholar was used primarily for literature searches. We also used some of the data collected in the study by Piseddu et al. (2021).

Due to the diversity of climatic conditions, experimental designs and technologies applied for climate manipulation, some decisions were made prior to data processing. For better comparability, sites with extremely cold (tundra) or hot (deserts) conditions were excluded from the analysis. Experiments performed under waterlogged conditions (e.g. bogs) were also excluded. In field experiments, several technologies are used to achieve the required change in environmental conditions. For each type of experiment, we have highlighted the method most comparable to the single-factor modelling exercise we performed.

In single-factor warming experiments open top chambers (OTC; Marion et al., 1997), infrared heaters (IR; Kimball, 2005; Aronson and McNulty, 2009), soil heating (by cables), closed chambers and soil transplantation techniques are typically used in the field (see Wang et al., 2019 for a summary). In many cases, the warming was not continuous, and in some cases the warming was asymmetric between day and night. In addition, open-top chambers are known to affect not only temperature, but also wind speed and radiation (and probably other factors; see Dabros et al., 2010), so the use of experimental data from OTCs is subject to uncertainties. We excluded sites where soil heating (via wires) was done instead of overall ecosystem heating (our design handled atmospheric heating through virtually manipulated air temperature). Soil transplant experiments and results obtained in closed chambers were also excluded. A comparison of experimental data and simulated sensitivities was done for the OTC and IR heating experiments. As OTC-based experiments cannot be considered as single-factor climate manipulations, we used IR heating results as another benchmarking option to evaluate model results, as IR heating is considered the closest method to single-factor ecosystem warming conditions (Aronson and McNulty, 2009).

Precipitation experiments manipulate either the amount or distribution of rainfall, or both. They may apply rain addition or rain exclusion to different extents. All types of experiments were included in the present study. For rain exclusion experiments, those that only manipulated the extent of rainfall but not its distribution were handled separately. In this type of experiments, a shelter with a fixed location

throughout the year or growing season, with a roof made of pitched bands of different widths, is typically used (Yahdjian and Sala, 2002). In this way, a specific precipitation rate can be excluded, while the rainfall distribution remains unchanged. This type of design corresponds to the model simulations we have applied in the present work. The other set of experiments changed both the amount and distribution of precipitation by excluding rainfall during a specific period of the growing season, while there was no exclusion during the other periods. For these experiments, fixed roofs with close to 100% exclusion (Köchy and Wilson, 2004), or rain sensor-activated curtains were used to cover the plots when it rained and removed when it stopped raining (Beier et al., 2004). Rainfall addition experiments also used different amounts and distributions, typically with water collected during natural rainfall events. In this study, experiments with an unchanged precipitation distribution are distinguished, as they are considered more comparable with our simulation design.

In compiling the data from elevated CO_2 trials, only the results from free-air CO_2 enrichment (FACE) experiments and open-top chambers were considered, as they represent the responses of plants and ecosystems to elevated atmospheric $[CO_2]$ in a natural environment, mostly avoiding the influence of other conditions. Experiments in closed chambers or greenhouses were not used because of possible uncertain effects on other environmental factors, plant size or growth issues caused by constraints imposed by the enclosures. Due to the known issues with OTCs, the FACE technology was considered in this study as the most comparable to the single-factor modelling exercise we performed.

After the literature review, the collected data were checked and critically filtered. All experimental data with missing information on the magnitude/characteristics of the manipulation or climatic conditions were removed. Experience showed that a relatively large number of publications used data from the same experimental sites with a wide temporal coverage. Particular attention was paid to exclude duplicate sites and to use the longest available datasets. Only annually or multiyearly interpretable data were retained (i.e. results from short-term experiments were ignored). In many cases, data collection consisted of scanning and digitising the graphs presented in journal articles, using customised software written in IDL 6.3 (NV5 Geospatial Solutions, Inc.; formerly distributed by Research Systems, Inc.).

Considering observed SWC in the field trials we extracted data that is comparable to the simulations that refer to the 0–10 cm depth (observed SWC was typically representative to the topsoil 0–5 cm, 0–10 cm, 0–15 cm, 5–10 cm, 0–20 cm depths, or it was measured at 5 or 10 cm depth).

In all cases, the mean effect size was calculated over multiple years. Site properties (mean annual temperature and precipitation sum, latitude and longitude) were retrieved from the published studies and used in the analysis presented. Supplementary material contains the list of publications used in the study.

2.6. Core logic for the presentation of the results

Given the complexity and high dimensionality of the study, which includes nine target sites, seven grassland models, four output variables and three *in silico* global change experiments with seven scenarios (including the reference run; Table 2), the presentation of the results had to be carefully designed. Our intention was to be informative and to avoid redundancy and repetition. The presented results are organized according to the type of the manipulation, which means that warming, wetting/drying and $[CO_2]$ manipulation are presented separately. The sensitivity of the four output variables from seven calibrated models is illustrated with respect to six alternative scenarios at different sites, using yearly averaged (SWC) or aggregated (ET, GPP, biomass) simulation results.

First, the consistency and robustness of the model results are analysed, with a focus on the overall direction and magnitude of the responses. The consistency check is visualized in the form of easily interpretable heatmaps. In these heatmaps, a positive response (e.g. an

increase in the simulated variable due to heating/wetting and a decrease due to cooling/drying) is indicated by red shades, while a negative response is visualised by blue shades.

Second, the overall multi-model mean was used to represent a robust estimate of responses. In our previous study (Sándor et al., 2017), the multi-model median (MMM) was used as a metric of the multi-model ensemble performance. The arithmetic mean (per climate change variable and per magnitude of change) was used here for practical reasons, due to the high degree of imbalance in the data used, as some modelling teams did not provide data for some sites and/or output variables. Calculating the mean rather than the median is likely to be less biased, as selecting the median value could lead to discontinuities in the responses, whereas using the mean of the responses reduces the risk of abrupt changes along the gradients of factor changes.

Third, the model responses are compared with those obtained from the field experiments. The individual model responses and the multi-model mean are also presented and evaluated against field evidence. Model grouping is also employed for a more comprehensive analysis. Model grouping, in this context, means that the seven models are categorized into three groups according to the model type (see Section 2.3). Consequently, the two grassland models, the three crop models and the two generic vegetation models are grouped together, and the resulting three groups are analysed as mini-ensembles. To construct these mini-ensembles, a simple averaging is used based on the available model responses at the site level.

3. Results

3.1. Sensitivity to air temperature

3.1.1. Overall sensitivity of SWC, ET, GPP and biomass to temperature
Fig. 2 shows the sensitivity of the models to temperature changes at each site, aggregated by output variable. According to the plots, SWC and ET are somewhat less sensitive to temperature change than GPP and biomass. In some cases, non-linearity is present but is not discussed in detail here. As mentioned in Section 2.4, non-linear response indicates variability (noise) but not curvature in the data.

The overall mean sensitivity (multi-model and multi-site; i.e. averaged for all models across all sites) of SWC to T is -0.7% $°C^{-1}$, indicating drying with increasing temperature. There is remarkable consistency amongst the models in the direction of SWC change, indicated by the predominance of blue colours in Fig. 2a, with only one model at one site that shows a positive response to increasing temperature (model 2 at MBO). The strongest negative response is associated with model 4 at site LAQ1. Averaged across all sites, model 4 has the strongest negative response, with a large site-to-site variability (SD is 1.1% $°C^{-1}$, but results are missing for some sites), while the smallest responsiveness is associated with model 6. The standard deviation of the model-specific results is fairly proportional to the magnitude of the response. Looking at the model results for the different sites, the spread of sensitivity (expressed here as the SD of the slopes) is highest at ROT1 and lowest at MAT (again, data are not available for some sites).

Considering ET, the overall sensitivity is 3.5% $°C^{-1}$. Fig. 2b shows that, in general, the models agree on the direction of change (indicated
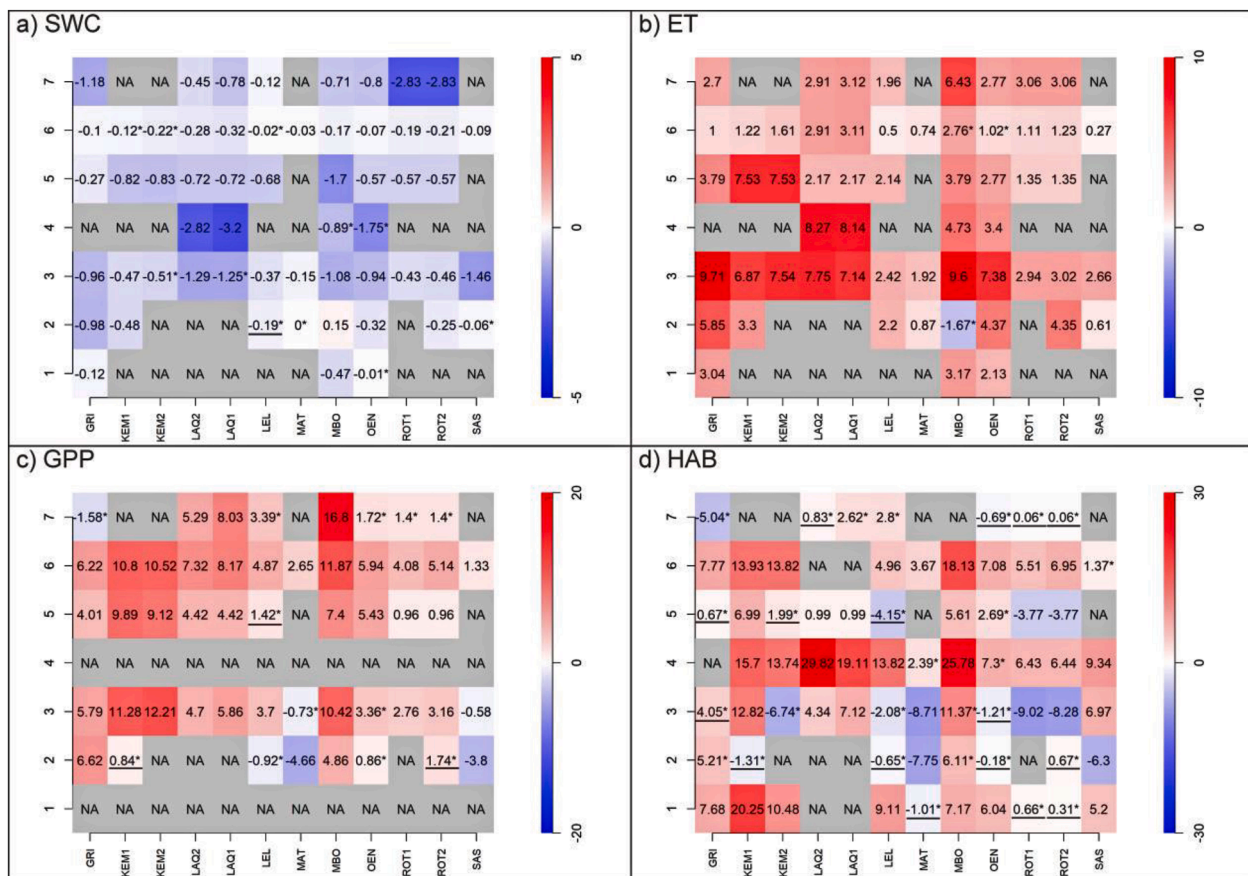

Fig. 2. Sensitivity of simulated SWC (a), ET (b), GPP (c) and yield biomass (d) to temperature change for the seven models (vertically arranged and numbered 1 to 7) at all sites (horizontally arranged). Sensitivity is expressed as the slope of the linear regression describing the dependence of the effect size on temperature change and has the unit of% $°C^{-1}$. The asterisk indicates that the response is slightly non-linear. NA indicates that the model result was not available. The slopes that are not significantly different from zero ($p>0.05$) are marked with underlines.

by the red shading) with one exception (again, model 2 at MBO). When averaged across all sites, model 4 is the most sensitive to temperature change, while model 6 is the least sensitive (the slopes are 6.1%°C^{-1} and 1.5%°C^{-1}, respectively). Focusing on the site-specific results, MBO has the highest variability (SD=3.5%°C^{-1}) while the spread of slopes is the smallest at MAT. Again, the missing simulations are a source of uncertainty.

The simulated sensitivity of GPP to temperature change is somewhat less consistent across sites and models (Fig. 2c). Of the 50 calculated slopes, six are negative (indicating a decrease in GPP with increasing T) while 44 are positive. The overall mean sensitivity is 4.6%°C^{-1} (SD is 4.3%°C^{-1}). When averaged across all available sites, model 6 shows the highest positive responsiveness, while model 2 shows only low sensitivity (models 1 and 4 did not provide usable data for GPP). The variability of the model results across sites is not related to the model-specific means. Focusing on the site-specific sensitivity, the results are consistent at eight sites in terms of the direction of change (i.e. most of the colours in the red gradient in Fig. 2c), but for others the results are inconsistent (GRI, LEL, MAT, SAS). The largest uncertainty (SD of the slopes) is associated with KEM1 (4.9%°C^{-1}), while the smallest was estimated for LAQ2.

Focusing on the most relevant output variable from the economic point of view (biomass), the result of the sensitivity analysis is the least consistent. According to Fig. 2d, 17 model-site combinations had a negative sensitivity to warming, while 52 slopes were positive. The overall mean sensitivity is 4.6%°C^{-1}, while the SD is 7.8%°C^{-1}. Model 4 is associated with the highest sensitivity (13.8%°C^{-1}, with no missing data), while model 2 has an overall slope of -0.5%°C^{-1} for the sites (some sites are missing). Model 7 shows negligible sensitivity. Focusing on the individual sites, a lack of consistency was observed in some cases. At sites LAQ1, LAQ2 and MBO, the models agreed on the positive sensitivity of biomass to T, but missing data could interfere with this conclusion. The spread of slopes is highest at LAQ2 (SD is 14.0% C^{-1}; slopes range from 0.8 to 29.8%°C^{-1}; again, some modellers did not provide data) and lowest at OEN (SD is 3.8%°C^{-1}).

3.1.2. Multi-model sensitivity

Using sensitivity measures (i.e. effect size as in Eq. (1)) quantified based on annually aggregated data, multi-model mean responses were calculated for each altered T scenarios for each site separately, using the available model results. As in the results presented above, in some cases simulation outputs were missing, which inevitably affected the results.

The multi-model sensitivity was calculated using a piecewise linear regression. Fig. 3 shows illustrative results for biomass based on the T scenarios using data from the coolest (Monte Bondone) and the warmest

(Matta) site. At Matta, the overall effect is negative, meaning that warming leads to a decline in biomass (and cooling to an increase). Some models suggest the opposite effect, as shown in Fig. 2. The magnitude of the sensitivity to cooling (-1.7%°C^{-1}) is smaller than that of warming (-2.9%°C^{-1}). At Monte Bondone, the models agree on the direction of change and the overall sensitivity is slightly larger (13.7%°C^{-1}) to cooling than to warming (11.0%°C^{-1}). Supplementary material contains the multi-model sensitivity for all sites and the four output variables studied (Figs. S3-S6).

Table 3 summarises the multi-model mean sensitivity measures expressed as the slope of the piecewise linear regression models.

The slopes distinctly vary by site, and asymmetry is observed for some of the variable-site combinations (see also Fig. 3 and Supplementary material). As cooling is not a plausible scenario under global warming, we focus here on the warming outcome. In the case of SWC, the robust multi-model responses agree on the direction (negative effect) with a mean sensitivity of -0.6%°C^{-1}. For ET, the sign is the same for all sites with an overall multi-site mean sensitivity of 3.2%°C^{-1}. For GPP, the effect of warming is negative at the two hottest and driest sites (Matta and Sassari) but positive at the other sites. The multi-site mean is 3.8% °C^{-1}. For biomass, the multi-model sensitivity is negative at three sites (the two Rothamsted sites and Matta) and positive at the others. The highest sensitivity is associated with Kempten1 and Monte Bondone. The multi-site mean sensitivity is 4% °C^{-1}.

Focusing on biomass as the most relevant output variable, site-specific, multi-model sensitivity was correlated with basic climatic conditions (Tm: mean annual temperature; Pm: mean annual precipitation) and site elevation. The strongest (non-linear) relationship was found between the multi-model mean sensitivity and Tm (sensitivity = -8.72 · ln(Tm) + 23.59, R^2=0.54; Fig. S7 in Supplementary material). If the sites with two treatments were merged and a mean sensitivity was calculated for both treatments (Kempten, Laqueuille and Rothamsted), the relationship became even stronger (R^2=0.67, not shown here). For the other two output variables (GPP and ET), the relationship was approximately linear (R^2=0.44 for Pm and R^2=0.53 for elevation).

3.1.3. Comparison with experimental data

Single-factor climate manipulation experiments (in this context, experimental warming) provide an opportunity to evaluate the realism of the presented model-based results. Based on an extensive data collection effort on the response of SWC, ET, GPP and biomass to experimental warming worldwide, a comparison was made with simulated sensitivities. The climate manipulation experiments are rarely associated with an artificial warming of less than 1 °C. It is thus not straightforward to evaluate the results of this study, which is simulated
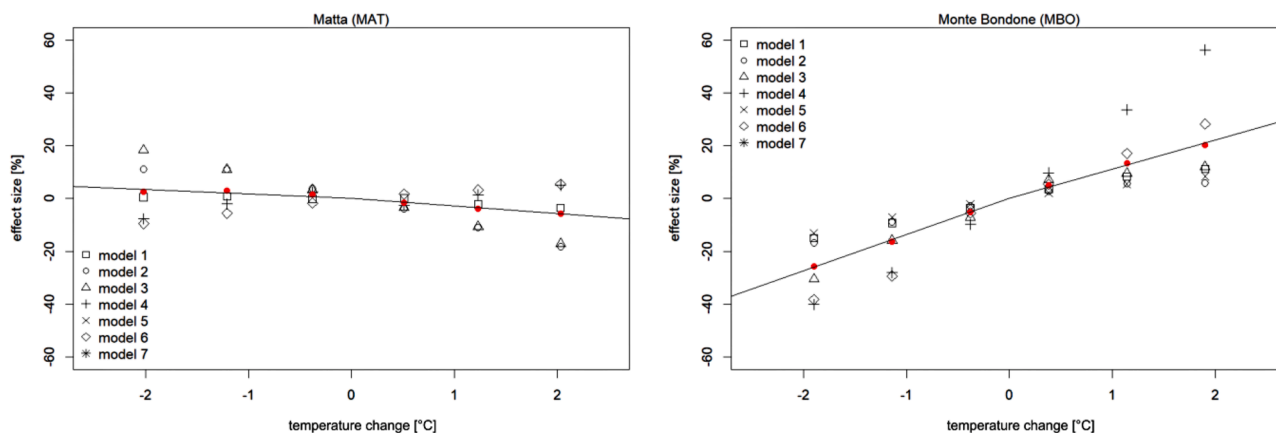

Fig. 3. Mean multi-model sensitivity of simulated biomass to temperature change at the warmest (Matta) and the coldest (Monte Bondone) sites. The black symbols represent the individual models, while the red dots show the mean multi-model effect size calculated for each T manipulation. The slopes of the piecewise linear regression fits are taken as the mean multi-model sensitivity across all sites. Note that the difference between the two slopes is small for both sites, which mean that the piecewise nature of the fit is hardly visible.

Table 3
Mean multi-model sensitivity for the studied output variables at all sites to temperature, expressed as the slopes of the piecewise linear regression models for both directions (warming and cooling) relative to the reference climate. The unit of slopes is% °C^{-1} in all cases. The slopes are significantly different from zero ($p<0.05$) in most of the cases (exceptions are marked with underlines).

| Site	SWC		ET		GPP		biomass	
	cooling	warming	cooling	warming	cooling	warming	cooling	warming
Grillenburg	-0.54	-0.51	4.6	4.1	5.1	3.3	6.8	3.5
Kempten1	-0.47	-0.48	4.5	5.1	8.1	8.3	11.9	10.8
Kempten2	-0.46	-0.52	4.4	4.4	7.8	7.3	6.4	2.8
Laqueuille1	-1.37	-1.15	4.9	4.6	7.1	6.2	7.8	7.1
Laqueuille2	-1.23	-0.99	4.9	4.8	6.1	4.8	8.4	9.6
Lelystad	-0.26	-0.29	1.8	1.9	3.2	1.8	4.7	2.1
Matta	-0.06	-0.06	1.2	1.1	<u>-0.3</u>	-1.5	-1.7	-2.9
Monte Bondone	-0.68	-0.71	4.4	3.9	10.6	9.9	13.7	11.0
Oensingen	-0.78	-0.50	3.4	3.4	4.1	2.8	4.1	1.8
Rothamsted1	-1.03	-0.98	2.3	2.0	2.9	1.7	0.4	<u>-0.4</u>
Rothamsted2	-0.91	-0.86	2.8	2.4	3.1	2.0	1.1	<u>-0.3</u>
Sassari	-0.57	-0.51	1.2	1.1	-0.9	-1.2	4.0	2.6

with a lower warming.

The climatic conditions of the experimental sites mostly differ from the European and Israeli grassland sites included in the present study (see Supplementary material, Fig. S8). Nevertheless, the magnitude and direction of the observed effects are indicative and provide real-world information on grassland response (without explicitly taking into account essential information such as species composition, which is beyond the scope of this study).

For SWC, the simulations tend to underestimate the magnitude of the observed effect (Fig. 4). Between 1 and 2 °C warming, S of the individual models (averaged across sites) vary between -2.8% and -0.2%. The mean S of all models is -1.0% (SD is 1.2%), while it is almost the same for the mean multi-model S (-0.98%; SD is 0.5%). Considering the mini-ensembles, the mean S is -2% for the crop models (with a SD of 1.4%), -0.8% for the grassland models (SD=0.47%) and -0.6% for the generic vegetation models (SD=0.4%) (refer to Supplementary material, Fig. S9). In contrast, the mean S from the observations is -5.8% (SD is -4.7%; $n = 10$). Focusing only on IR heaters, the mean S is -6.5% (SD is -5.4%; $n = 7$). With few observations available for ET and GPP (Figs. S10 and S11), the overall impression is that models tend to underestimate the magnitude of S.

Focusing on biomass (Fig. 5), between 1 and 2 °C of warming, the simulated individual S-values vary between about -14 and 60%, while

the multi-model means vary between -3.8% and 22.3%. For the same temperature range, the experiment-based effect sizes vary between -12 and 39.5%. If we only use the data from the IR-heated sites, the effect sizes vary between -3.6 and 38.8%. Averaging all available data between 1 and 2 °C, S of the individual models vary between -2.3% and 24.0% (the latter is associated with model 4). The mean S from all individual models and sites is 6.6% (SD=13.2%), while it is 7.1% (SD=7.1%) based on the multi-model means. Considering the model-type-based grouping, the mean S is 13.7% for the crop models (with SD of 1.2%), 0.47 for the grassland models (SD=8.2%), and 4.3% for the generic vegetation models (SD=5.5%) (Supplementary material Fig. S12). Using the experimental data between 1 and 2 °C of warming, the overall mean is 7.5% (SD=16%; $n = 15$). If only sites with IR heating are considered, the mean S is 8.8% (SD=14%; $n = 9$). Consistency between simulated and observed data is considerable, especially for the multi-model means, although the climate and site characteristics of the study sites (*in silico* climate manipulations) and the real sites differ.

Fig. S13 (Supplementary material) shows the relationship between Tm and sensitivity slopes calculated from field experiments for sites using IR heating (assuming a linear response with warming magnitude). Comparing Figs. S7 and S13, it is difficult to find consistency between the patterns. Sensitivity (expressed here as% °C^{-1}) can reach negative values for both field trials and model results, but the overall decrease of
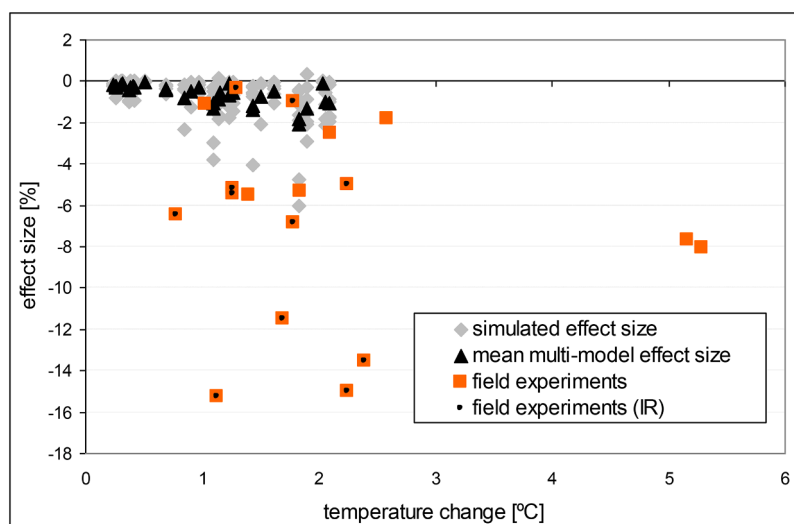
Fig. 4. Comparison of the simulated and observed sensitivities of SWC (expressed here as S, that is identical to effect size;%) to temperature manipulation. grey symbols: S values as function of the magnitude of warming (°C) calculated from all models for each site and for each temperature scenario. Black triangles: mean multi-model effect size for all sites and scenarios. Red squares: S values based on the literature review using all collected data. The dotted red squares indicate data from experimental sites using infrared heating technology, which is considered the closest method to the true single-factor experiments.
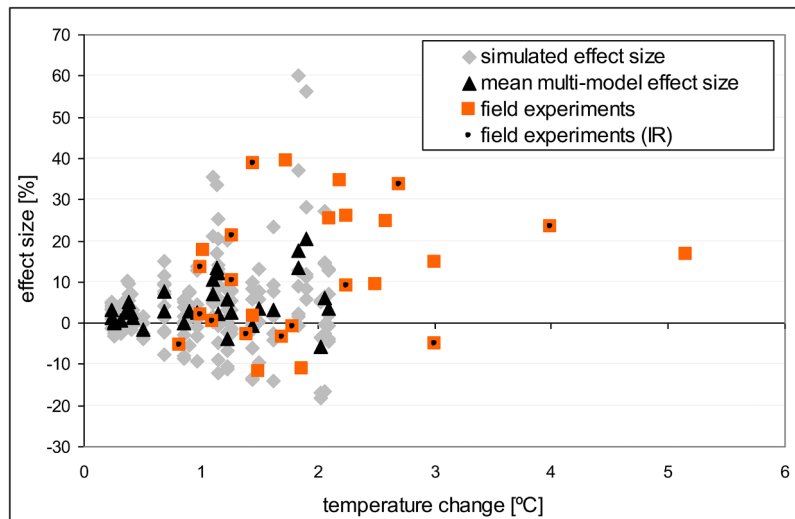
Fig. 5. Same as Fig. 4 but for biomass.

sensitivity with increasing Tm is not detectable in the experiments. At cooler sites associated with Tm between 5 and 10 °C the responses are broadly similar. As mentioned above, the climate of the warming experiments and the MACSUR sites is different and, due to the scarcity of data, it is difficult to draw solid conclusions.

3.2. Sensitivity to precipitation

3.2.1. Overall sensitivity of SWC, ET, GPP and biomass to precipitation

Fig. 6 shows the sensitivity of the models to PREC changes. According to the plots, SWC and GPP are somewhat less sensitive to PREC change than ET and HAB. In some cases, non-linearity is present for some model-site combinations but is not discussed here. In a few cases a
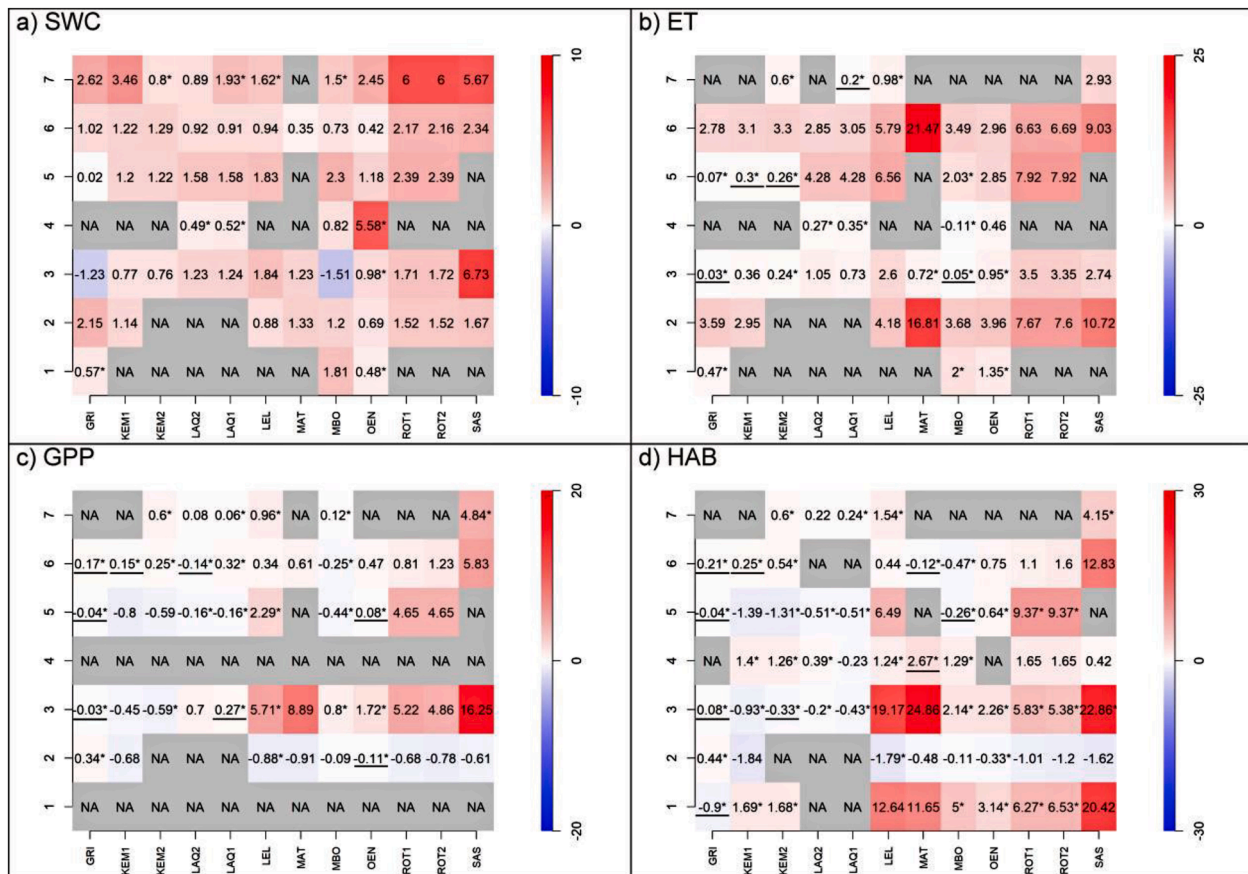

Fig. 6. Sensitivity of the simulated SWC (a), ET (b), GPP (c) and yield biomass (d) to precipitation change for the seven models (arranged vertically and numbered 1–7) across all sites (arranged horizontally). Sensitivity is expressed as the slope of the linear regression describing the dependence of effect size on precipitation change and has the unit of% 100 mm^{-1}. The asterisk indicates that the response is slightly non-linear. NA indicates that the model result was not available. The slopes that are not significantly different from zero ($p>0.05$) are marked with underlines.

strongly non-linear response (with curvature) was depicted by some models (e.g. model 3 GPP at GRI, model 5 GPP at LAQ2). These cases do not affect the overall results, but provide an interesting insight into the behaviour of the given model.

According to Fig. 6a, for SWC there is a general consistency amongst the models regarding the direction of change by PREC (the exception is model 3 at two sites). The overall sensitivity is 1.7% 100 mm^{-1}, while the strongest mean response is associated with model 7 (with a slope of 3.0% 100 mm^{-1}). Considering the overall aggregate response by site, the largest variability is associated with SAS (*SD*=2.5% 100 mm^{-1}, but here the results of only 4 models were available).

Considering ET, the consistency of the models is remarkable (the only exception is model 4 at MBO; Fig. 6b). The overall sensitivity of ET to PREC change is 3.6% 100 mm^{-1}. The strongest positive response is associated with models 3 and 5, but it is interesting to note that this above-average response is mostly caused by the large responses at MAT and SAS (the driest sites) for both models. At MAT, model 6 predicts 21.5% 100 mm^{-1} sensitivity. This behaviour is also clear from the aggregate sensitivity per site, where MAT and SAS have the largest responsiveness with the largest spread (*SD*=10.9% 100 mm^{-1} at MAT and 4.1% at SAS 100 mm^{-1}, but data are missing).

Fig. 6c indicates that the results are inconsistent for GPP. Out of the 50 individual slopes, 20 are negative (although typically small, -0.4% 100 mm^{-1} on average), while the others are positive. The overall sensitivity is 1.3% 100 mm^{-1}. Model 3 is associated with the largest responses, up to 16.3% 100 mm^{-1} at MAT. Considering the individual models, the mean slope across the sites is positive for all but model 2 (but some models did not provide any GPP data). The mean slope is the largest for model 3 (3.6% 100 mm^{-1}). At the individual site level, the results are not consistent in terms of the direction of change (there is no site where the direction of change is uniform). MAT and SAS are associated with the largest uncertainty (*SD*) of slopes at site-specific aggregation level.

Similarly to the T scenarios, the results are not consistent in terms of the sensitivity of the simulated biomass to PREC change (Fig. 6d). Out of the 66 simulated slopes, one third are negative, while the others are positive. The overall mean sensitivity is 3.0% 100 mm^{-1} (*SD* is 5.9% 100 mm^{-1}). Models 1 and 3 provided the largest sensitivities (6.8 and 6.7 100 mm^{-1} for model 1 and model 3, respectively). Model 2 depicts a negative sensitivity for all but one site, with an overall slope of -0.9% 100 mm^1. At the site-level aggregation, there is no consistency in the simulated direction of the change. The uncertainty of the slopes is the largest at MAT and SAS (*SD* is larger than 10% 100 mm^{-1}), but is also quite large at LEL (7.7% 100 mm^{-1}). The uncertainty is the smallest at LAQ1.

3.2.2. Multi-model sensitivity

Fig. 7 shows excerpts from the full multi-model sensitivity analysis of the biomass simulations based on the PREC scenarios using data from the driest (MAT) and the wettest (OEN) site.

The overall effect is positive at both sites, which means a decline in biomass due to the decrease and the increase due to additional PREC. At Matta the response is asymmetric and strong, while at Oensingen the PREC addition leads to a smaller response. Supplementary material contains the multi-model sensitivity for all sites and for all four output variables (Figs. S14-S17).

Table 4 presents the multi-model sensitivity slopes (separately for wetting and drying), which can be interpreted as the best estimate for sensitivity to increasing (wetting) and decreasing (drying) PREC at the sites. In some cases, curvature can be detected in the multi-model mean responses, but in-depth analysis and presentation of the analytical function of the possible non-linear fit was beyond the scope of this study. The curvature can be real, but in some cases missing data or extreme model responses can cause this non-linear pattern. We point out here that, within the ranges defined by the PREC scenarios, we use a linear model that approximates the real response with some error.

For SWC, sensitivity slopes are positive at all sites. The overall mean response is 1.7% 100 mm^{-1}. For ET, a consistency in slopes is observed with a relatively large response at Matta. The overall slope is 4.1% 100 mm^{-1}. In the case of GPP, mean slopes are slightly below zero at some sites, while they are positive at the other sites. The two driest sites (MAT and SAS) are associated with the highest sensitivity. The grand ensemble response is 1.6% 100 mm^{-1}. For biomass, the results are similar to those calculated for GPP. The negative slopes are small while the positive slopes are somewhat larger. MAT and SAS are the most responsive sites. The mean multi-model slope is 3.2% 100 mm^{-1}. An interesting feature of Table 4 is the fact that for drying the responses tend to be larger than for wetting.

Considering the site-specific sensitivity slopes for biomass, site conditions were used as predictors both for drying and wetting separately. For drying, the strongest linear relationship was found between the multi-model mean sensitivity and Pm (sensitivity = -0.02 · Pm + 21.47, R^2=0.85; Fig. S18 in Supplementary material). In the case of Tm, the relationship was non-linear with R^2=0.78. Pm was the best predictor of wetting (sensitivity = -0.01· Pm + 11.23, R^2=0.78; Fig. S19 in Supplementary material), but the predictive value of Tm is also considerable (variance explained was 67%, non-linear in shape). Site elevation was not a satisfactory predictor in any of the cases.

3.2.3. Comparison with experimental data

The published results from the precipitation manipulation experiments were used to benchmark the sensitivity of the presented models in
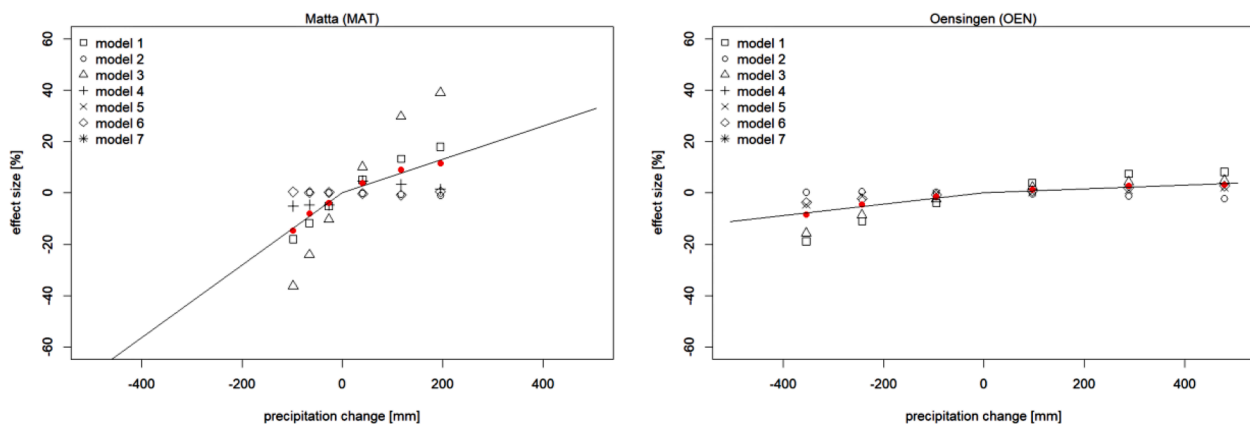

Fig. 7. Mean multi-model sensitivity of simulated biomass to precipitation change at the driest (Matta) and the wettest (Oensingen) sites. The black symbols represent the individual models, while the red dots show the multi-model mean effect size calculated for each PREC manipulation. The slopes of the piecewise linear regression fits are considered as the mean multi-model sensitivity in the two sites.
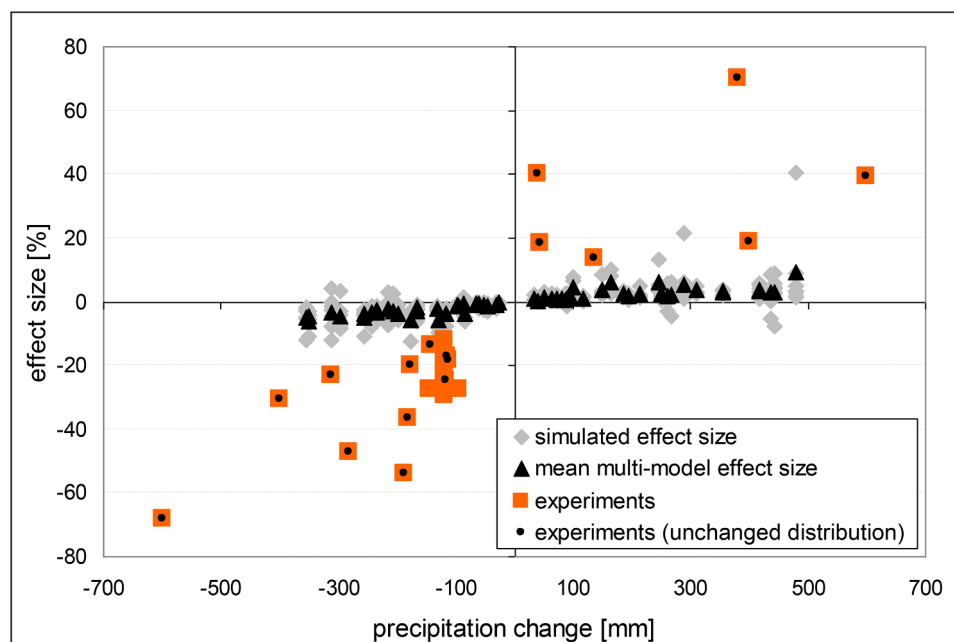
Fig. 8. Comparison of the simulated and observed sensitivities of SWC (expressed here as *S*, that is identical to effect size;%) to precipitation manipulation. grey symbols: *S* values as function of the magnitude of precipitation change (mm) calculated from all models for each site and for each temperature scenario. Black triangles: mean multi-model effect size for all sites and scenarios. Red squares: *S* values based on the literature review using all collected data. The dotted red squares represent experiments using precipitation manipulation without altering the distribution of precipitation (comparable to the modelling approach presented in the study).

Table 4
Mean multi-model sensitivity for the studied output variables at all sites, expressed as the slopes of the piecewise linear regression models for both directions (drying and wetting) relative to the reference climate. The unit of slopes is% 100 mm^{-1} in all cases. The slopes are significantly different from zero ($p<0.05$) in most of the cases (exceptions are marked with underlines).

Site	SWC		ET		GPP		biomass	
	drying	wetting	drying	wetting	drying	wetting	drying	wetting
Grillenburg	1.07	0.68	1.58	0.87	0.49	-0.10	-0.16	-0.83
Kempten1	1.87	1.02	2.18	1.40	-0.43	-0.45	0.23	-0.34
Kempten2	1.32	0.89	2.03	1.17	-0.02	-0.27	0.45	0.01
Laqueuille1	1.63	0.93	2.20	1.36	0.31	-0.02	-0.32	-0.17
Laqueuille2	1.26	0.84	2.69	1.67	0.32	-0.03	-0.07	0.01
Lelystad	1.85	1.23	4.70	3.72	2.87	1.16	7.61	4.82
Matta	0.96	0.94	11.47	13.01	5.24	2.63	14.09	6.52
Monte Bondone	1.43	0.75	2.50	1.53	0.15	-0.03	1.73	1.03
Oensingen	1.37	1.87	2.68	1.73	1.02	0.26	2.19	0.75
Rothamsted1	3.17	2.53	7.86	5.64	3.43	1.98	5.94	2.72
Rothamsted2	3.19	2.52	7.83	5.60	3.47	1.95	5.84	2.82
Sassari	4.40	3.91	7.41	5.66	8.60	5.24	13.19	7.64

terms of wetting (i.e. water addition similar to irrigation) and drying (precipitation exclusion). We highlighted the experimental results associated with the manipulation that does not affect the distribution (i. e. timing) of precipitation events, as they are likely to be more comparable to the simulation results presented.

The climate of the field experiments and that of the studied sites are presented together in Supplementary material (Fig. S20). As in the case of the warming trials, the climatic conditions are rather diverse and the thermo-pluviometric conditions differ between the sites studied and the experimental sites.

The sensitivity of SWC to wetting or drying is lower in the models than in real-world experiments, both for individual models and for the mean multi-model response (Fig. 8; see also Fig. S21 in the Supplementary material, where the precipitation change is expressed in terms of percentage of the mean annual precipitation). This statement holds valid even if we only consider experiments with the precipitation distribution unaltered (dotted red squares).

For drying, the *S* values of the individual models (averaged across sites) vary between -5.4% and -2.0%. The mean *S* value of the individual models is -2.8% (*SD*=2.7%), which is very close to the multi-model mean *S* values (-2.7% on average; *SD*=1.8%). The mean *S* value is -4.4% for the mini-ensemble of crop models (*SD*=3.3%), -2.1% for the grassland models (*SD*=1.6%) and -2.4% for the vegetation models (*SD*=1.6%) (Fig. S22). In contrast, in the field trials, the overall mean is -24.3% (*SD*=9%; one value with 600 mm manipulation was removed from the observations). If we consider only observations with an unaltered precipitation distribution, the mean *S* value is even larger (-28.6%; *SD*=13.5%). For wetting, the situation is similar, with model-specific *S* values (averaged across sites) varying between 1.5% and 5.3%, with overall mean *S* values of 2.8% (*SD*=4%) for the individual models and 2.7% (*SD*=2%) for the mean multi-model responses. Interestingly, these values are exactly the same magnitude as for drying. Model grouping leads to mean *S* values of 4.4% (*SD*=4.2%), 1.9% (*SD*=1.6%) and 2.3% (*SD*=1.5%) for the crop, grassland and vegetation models, respectively

(Fig. S22). For the observations, the overall *S* value is 29.6% (*SD*=21.9%). For the other output related to the water cycle (ET), only limited observational evidence is available (Fig. S23), but the models seem to underestimate the sensitivity.

Considering the biomass simulations with the drying scenarios, Fig. 9 shows that some of the individual model-site combinations provided *S* values that are consistent with the observed effect size values (see also Fig. S24 in the Supplementary material). The *S* values of the individual models vary between -1.3% and -0.1%. The overall mean of all models, sites and scenarios is -4.3% (minimum and maximum are -51.2% and 11.1%, respectively; *SD*=8.9%). The mean of the mean multi-model *S*-values is -4.1% (*SD*=5.2%). Focusing on the mini-ensembles, the mean *S* is -5.2% for the crop models (*SD*=5.7%), -4.6% for the grassland models (*SD*=8.5%) and -2.6% for the vegetation models (*SD*=5.3%) (Fig. S25). The climate manipulation experiments indicate a somewhat higher responsiveness, with a mean *S*-value of -9.6% (*SD*=18.2%; *n* = 34; one extreme PREC manipulation was removed). If we only use the data from the rainfall exclusion experiments with an unaltered precipitation distribution, the mean sensitivity is -10.6% (*SD*=13.9%; *n* = 11).

Focusing on the water addition experiments, the model-specific *S* values vary between -1.6% and 8.4%, while the mean *S* of all simulations is 3.2% (*SD*=7.9%). The mean of the multi-model effect sizes is 3.0% (*SD*=4.3%). Model-type specific multi-model *S* is 3.6% (*SD*=4.7%) for the crop models, 3.1% for the grassland models (*SD*=6.1%), and 2.1% for the vegetation models (*SD*=4.9%) (Fig. S25). In contrast, the mean *S* value of the field trials is 38.9% (*SD*=62.6%, *n* = 13; one data point with 967 mm water addition was excluded). If we further exclude two reported *S*-values above 100% (two data points from a semi-arid site in Mongolia), the mean *S* value is 14.9% (*SD*=15.7%).

Using a linear approximation, the effect size values from the field trials were converted into sensitivity slopes (Section 2.4). Fig. S26 in Supplementary material shows the relationship between Pm from the field experiments and the derived sensitivity slopes for drying. Comparing Fig. S26 with Fig. S18 makes it clear that the strong covariation between Pm and the simulated slopes are not detectable from the real-world data. It is remarkable that the low Pm values are not represented by the *in silico* results, so it is not possible to draw strong conclusions for the dry sites. There is a better agreement between the simulated and observed Pm-sensitivity slope relationship for wetting (Figs. S19 and S27 in Supplementary material). For Pm values above

400 mm, the sensitivity slopes (effect size per 100 mm manipulation) are similar (with one outlier). Again, dry sites associated with Pm below 400 mm are not represented by the simulations, which means that there are no comparable results.

3.3. Sensitivity to atmospheric [CO2]

3.3.1. Overall sensitivity of SWC, ET, GPP and biomass to [CO2]

Fig. 10 shows the sensitivity of the models to changes in [CO2] expressed simply as the effect size with a doubled [CO2] level. According to the plots, the responsiveness of the studied output variables to CO2 fertilisation is rather diverse.

Fig. 10a shows that the models differ in terms of the direction of SWC change. While some models show consistently positive responses (indicated by red shading), others simulated a negative effect. The overall sensitivity is 0.2% for a doubled [CO2] level (*SD* is 1.0%). Excluding model 7, which only simulated one site, model 6 shows the largest negative response (-0.3% for doubled [CO2] level). The largest positive response is associated with model 3. If individual sites are considered, SAS shows the highest uncertainty (*SD*=2.7%). There is no consensus on the direction of change at the individual site level.

According to Fig. 10b, ET is also associated with lack of consistency. Some models show a strong negative response to a doubling of the [CO2] level (mostly models 2, 3 and 4), while the rest of the models depicts a small positive change. There is a remarkable similarity in this pattern with the SWC effect. The overall sensitivity of ET to the doubling of the [CO2] level is -3.0% (*SD*=6.0%). When models 1–4 are averaged across all sites, negative responses are simulated (the largest in absolute terms is model 3, with an effect size of -9.9%), while models 5–6 depict a positive response. However, missing data interfere with the results. If we scrutinise the results at the site level, the sensitivities are not consistent. The largest spread of sensitivities is found at MAT (*SD*=10.0%), while the smallest at MBO (*SD*=4.3%).

In the case of GPP (Fig. 10c), the picture is surprisingly concordant and the consistency is remarkable. The overall mean effect size is 20.5% for the double [CO2] level (*SD*=16.0%). The most responsive model is clearly model 3 (38.7% mean sensitivity across all sites), while the least responsive is model 2 (4.7%). Using site-level aggregation, the spread of sensitivity is the smallest at GRI and the largest at SAS.

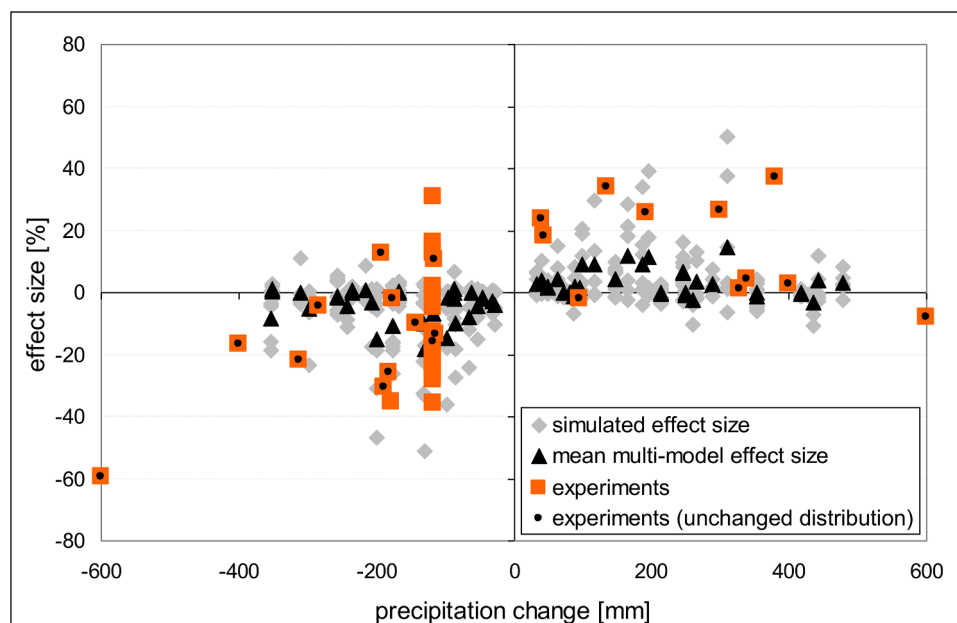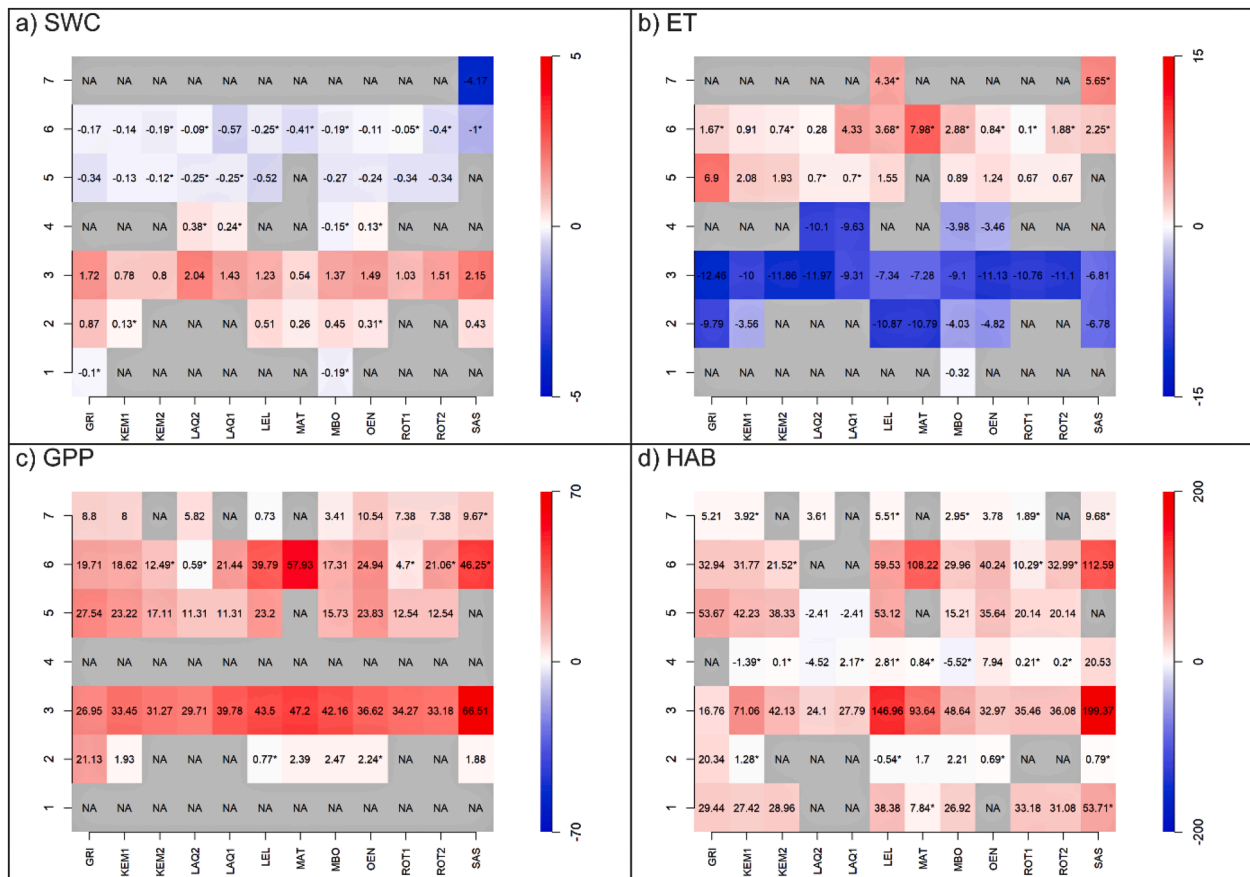The biomass response to the doubled [CO2] level is considerable in

Fig. 9. Same as Fig. 8 but for biomass.

Fig. 10. Sensitivity of simulated SWC (a), ET (b), GPP (c) and yield biomass (d) to changes in $[CO_2]$ for the seven models (arranged vertically and numbered 1 to 7) at all sites (arranged horizontally). Sensitivity is expressed as the effect size for the doubled $[CO_2]$ level (760 ppmv in this study). The asterisk indicates that the response is slightly non-linear. NA indicates that the model result was not available.

some cases and relatively consistent, revealing an overall positive effect (Fig. 10d). Out of the 68 individually simulated effect sizes, only six depict a small negative response. Some models suggest a fairly large response, up to 200% (model 3 at SAS). Four effect sizes above 100% and 10 above 50% were simulated. The overall mean sensitivity to the doubled $[CO_2]$ level is 27.8% (SD=36.3%). Model 3 is the most responsive (64.6% on average). Model 4 is rather insensitive to the doubled $[CO_2]$ level (2.1% on average for the sites). At the individual site level, the variability of sensitivity is large. The largest spread of the effect size is at SAS (SD=76.9%), while the smallest is at LAQ2

(SD=13.1%).

3.3.2. Multi-model sensitivity

Due to the typical nonlinear response to elevated $[CO_2]$, the mean multi-model sensitivity is quantified by the effect size pertaining to the 720 ppmv manipulation (doubled $[CO_2]$ level). Fig. 11 shows the multi-model sensitivity for biomass at two sites with a contrasting response.

At LAQ2, the multi-model response is close to linear and relatively small. However, the result is not protected from bias, as models 1 and 6, which are typically associated with larger responses, did not provide
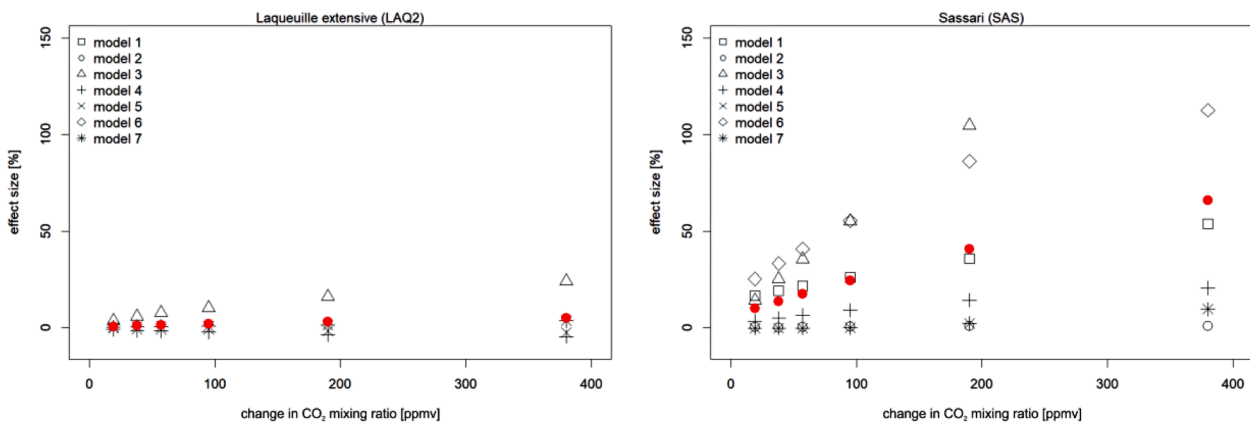
Fig. 11. Mean multi-model sensitivity of simulated biomass to $[CO_2]$ at two sites with contrasting response (LAQ2 and SAS). The black symbols represent the individual models, while the red dots show the multi-model mean effect size calculated for each $[CO_2]$ manipulation. The mean effect size for the 760 ppmv manipulation represents the mean multi-model sensitivity.

data for this site (see Fig. 10d). At SAS, the effect size is much larger, which is reasonable as some of the individual models proposed an effect size larger than 100%. Supplementary material contains the multi-model effect sizes for all sites and for the four output variables (Figs. S28-S31).

Table 5 summarises the effect sizes of 720 ppmv for the four output variables by site. SWC is associated with positive responsiveness, with the exception of SAS, where the effect of model 7 (which only provided data for SAS; Fig. 10a) is clearly dominant. The overall mean effect size is 0.2%. For ET, there is considerable consistency in the direction of change (decrease in ET) with a mean responsiveness of -3.0%. The results for GPP are also consistent and range between 11.9 and 35.8% (mean 20.6%). The results for biomass are similar, but we again mention that missing data from the most responsive models may bias some results, at least for LAQ1 and LAQ2. The maximum effect size is attributed to SAS. The grand ensemble is 26.2%, which can be converted to a linear response of about 7% 100 ppmv^{-1} (but the response is typically non-linear with saturation).

Considering the site-specific, ensemble results for biomass, a linear relationship was found between the multi-model mean effect size and Pm, but a non-linear relationship was detected between the mean effect size and Tm. The explained variance was 50% in both cases and the relationship did not improve if the two LAQ sites were removed from the analysis (Fig. S32 and Fig. S33 in Supplementary material, which also contain the fitted equations). For site elevation, the relationship was nonlinear with low variance explained.

3.3.3. Comparison with experimental data

The published results of the [CO_2] manipulation experiments using OTC and FACE technology were used to benchmark the sensitivity of the presented models in terms response to elevated [CO_2]. Considering the climatic conditions of the experimental sites, they differed mostly in the case of SWC, while for biomass the climate of some sites was closer to that of the sites included in the present study (Supplementary material, Fig. S34).

The sensitivity of SWC to CO_2 manipulation is obviously markedly lower for models than for the real-world experiments, both for the individual site-model combinations and for the mean multi-model response (Fig. 12). The S values of the individual models (averaged across sites for all [CO_2] manipulations) vary between -1.0% and 0.5%. The mean S value of the individual models is 0.0% (SD is 0.5%), which is equal to the mean multi-model S values (SD is 0.1%). Considering the mini-ensembles constructed for the three model types, mean S is -0.2% for the crop models (SD=0.8%), 0.4% for the grassland models (SD=0.4%) and -0.2% for the vegetation models (SD=0.2%) (Fig. S35). In contrast, for the field trials the overall mean is 8.6% (SD=8.8%). If we consider only observations made within FACE experiments only (neglecting open top chambers), the mean S is 9.5% (SD=10.6%).

Considering biomass, model-specific S values vary between -5.5%

and 199.0%. Individual modelspecific S values (averaged across all sites) vary between 0.7% and 25.8%, while the mean S of all simulations is 12.0% (SD=20.4%; Fig. 13). The average of the mean multi-model effect sizes is 11.5% (SD=11.8%). Considering the different groups of models, mean S is 4.9% (SD=5.1%) for the crop models, 15.8% (SD=16.7%) for the grassland models and 18.9% (SD=23.6%) for the vegetation models (Fig. S37). Focusing on the mean S from all considered [CO_2] manipulations the mean is 17.5% (SD=15.1%, n = 21). If we neglect the OTC-based CO_2 enrichment studies the mean effect size is 14.9% (SD=14.9%, n = 15).

Using a linear approximation, the effect size values from the field trials were converted into sensitivity values (here effect size) referring to the doubled CO_2 concentration. Fig. S38 and Fig. S39 in Supplementary material show the relationship between Tm and Pm from the field experiments and the derived effect size values. If we compare Fig. S32 with Fig. S38, they are broadly consistent, meaning that at warmer temperatures the effect size for doubled [CO_2] is usually larger. Similar conclusions can be drawn based on the comparison of Fig. S33 and Fig. S39, meaning that the effect size decreases with increasing precipitation.
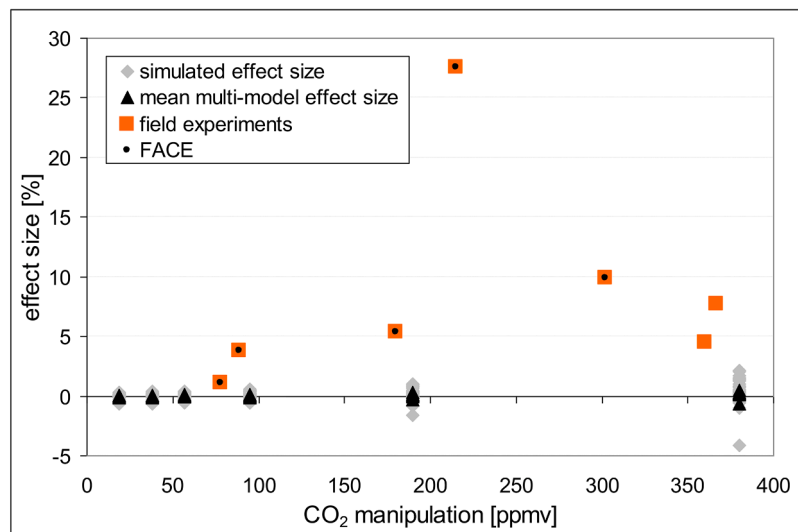
4. Discussion

This section first discusses the simulation results by grouping the water balance outputs (SWC and ET in Section 4.1) and the production outputs (GPP and HAB in Section 4.2). The remainder discusses the comparison of the simulations with the results of manipulation experiments in grassland ecosystems worldwide (Section 4.3) and the limitations of the study (Section 4.4).

### 4.1. Soil water content and evapotranspiration

In process-based models, SWC affects stomatal conductance (which in turn affects photosynthesis and biomass accumulation), ET, plant stress and, eventually, senescence. Simulation of SWC is performed using a variety of methods within the models (e.g. Table B1 in Sándor et al., 2017). Soil water redistribution is commonly accounted for in models using the Richards (1931) equation, which describes the transient flow of water from initial conditions. This equation plays a crucial role in modelling soil hydrological processes within the rhizosphere for multilayer soil hydrology models. Simpler approaches such as the tipping-bucket method (Ritchie, 1981) are also used in vegetation models. ET is calculated by diverse methods like the well-established Penman-Monteith equation or, with less data requirements, the Priestley-Taylor equation (Sándor et al., 2017). The methodologies used by the seven models in this study can be considered representative of state-of-the-art models. However, these different representations of SWC and ET may limit the actual comparison of the models, as this variable is defined according to each model's own scheme. Construction of the model-type specific mini-ensembles shed some light on the consequences of the diverging model structures, though the number of models was low for more robust findings.

Based on Figs. 2, 6 and 10, and also Figs. S3, S4, S14, S15, S28 and S29 in Supplementary material, the temperature, precipitation and [CO_2] scenarios had a varying effect on the simulated SWC and ET changes (in line with the conceptual scheme of Fig. S2 in the Supplement). In general, as air temperature increases, SWC decreases in parallel with the increase in ET. The response of the models to precipitation changes also matched what was expected (i.e. wetting causes an increase in SWC and ET, while drying has the opposite effect). The models were mostly consistent in terms of these changes for temperature and water supply changes. In contrast, for increasing [CO_2] the SWC and ET responses were inconsistent indicating possible issues with the simulated effect of [CO_2] on stomatal aperture (e.g. increasing [CO_2] usually reduces stomatal conductance to water vapour). The results revealed that the models responded differently to changes in environmental conditions due to variations in how they calculate the movement of SWC

Table 5
Mean multi-model sensitivity for the output variables studied at all sites, expressed as the mean effect size for a doubling of [CO_2] relative to the reference climate. The unit is%.

Site	SWC	ET	GPP	biomass
Grillenburg	0.29	-3.3	20.8	22.8
Kempten1	0.16	-2.6	17.0	25.2
Kempten2	0.14	-3.1	15.8	22.0
Laqueuille1	0.21	-3.5	24.2	9.2
Laqueuille2	0.52	-5.3	11.9	5.2
Lelystad	0.24	-1.7	21.6	43.7
Matta	0.13	-3.4	35.8	42.4
Monte Bondone	0.17	-2.3	16.2	17.2
Oensingen	0.32	-3.5	19.6	20.2
Rothamsted1	0.21	-3.3	14.7	16.9
Rothamsted2	0.25	-2.9	18.5	24.1
Sassari	-0.65	-1.4	31.1	66.1

Fig. 12. Comparison of the simulated and observed sensitivities of SWC (expressed here as *S*, that is identical to effect size;%) to [CO_2] manipulation. grey symbols: *S* values as function of the change of CO_2 concentration (difference between elevated and reference concentration expressed in ppmv) calculated from all models for each site and for each scenario. Black triangles: mean multi-model effect size for all sites and scenarios. Red squares: *S* values based on the literature review using all collected data. Dotted red squares indicate data from the experimental sites using FACE technology (that is more comparable to the modelling approach presented in the study).
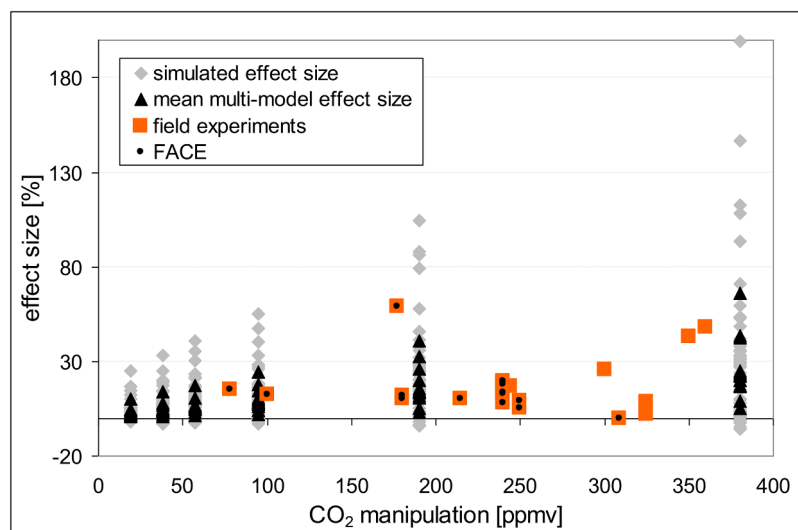
Fig. 13. Same as Fig. 12 but for biomass.

along the soil profile and the ET calculation algorithm.

In addition to affecting grassland functioning, SWC is also known to be closely linked to weather patterns. For example, it was shown that during heatwaves, SWC depletion can increase the severity and duration of drought episodes (e.g. Seneviratne et al., 2006, 2012), leading to a decrease in ET and a parallel increase in sensible heat flux that enhances land-atmosphere coupling with consequences for plant functioning and C balance (Ciais et al., 2005). Teuling et al. (2010) showed that the response of grasslands to drought is different from that of forests, mainly due to their shallower rooting depth, with grasslands having a stronger impact on the soil temperature increase during heatwaves than forests in the prolonged phase of the event.

Due to the well-recognised importance of SWC for atmospheric processes, climate research currently focuses on estimating soil moisture variability (Seneviratne et al., 2006; Diodato et al., 2014; Wu et al., 2015; Fu et al., 2022; Li et al., 2022). Simulation of SWC with acceptable quality is thus a prerequisite for the application of state-of-the-art Earth

System Models (ESMs; Ciais et al., 2013). Analogous to numerical weather and climate prediction models, ESMs estimate SWC using approaches originally developed for ecosystem models (e.g. Balsamo et al., 2009; Chen and Dudhia, 2001a, b). In particular, the functioning of grassland systems cannot be represented without an explicit simulation of the soil water balance due to the role of water limitations (or, in some cases, excess SWC leading to root stress or asphyxia) in several biological processes in plants (Nagy et al., 2010; Hidy et al., 2012). Evaluating the quality of SWC simulation results under different environmental and management conditions is thus crucial to clarify their applicability, as the results may point to the need for further developments and adjustments. As SWC directly affects ET, improving the representation of soil hydrology could improve the quality of ET simulations, reducing the issue of obtaining satisfactory simulation results for wrong reasons (Martre et al., 2015).

4.2. Gross primary production and yield biomass

Across all sites and simulations, GPP (Fig. S30) and yield biomass were observed to respond positively to increasing atmospheric [CO$_2$] (Fig. 13), with substantial consistency amongst models (Fig. 10c and d, respectively). Although the different models were observed to be differently sensitive to changes in T (Fig. 2c, d) and PREC (Fig. 6c, d), the mean values (Figs. 3 and 7, respectively) are consistent with the expected dynamics. This suggests that the construction of multi-model means is useful to ensure a realistic response in terms of direction of change to varying environmental conditions.

Differences in mechanistic representations of gross photosynthesis and biomass allocation could explain the varying and sometimes divergent responses amongst the models. For instance, models based on the approach of Farquhar et al. (1980) rely on the dependence of photosynthesis on the maximum rate of carboxylation of ribulose bisphosphate carboxylase and the maximum rate of photosynthetic electron transport (Wullschleger, 1993), which in turn depend on the response of assimilation rate to sub-stomatal CO$_2$ concentration and further modulated by temperature. In contrast, models based on light use efficiency (LUE) have a different logic and functional representation (e.g. in the case of STICS), and may be difficult to parameterise due to uncertainties in the LUE term, which is typically estimated using look-up tables of maximum LUE for a given vegetation type and then adjusting these values downwards, depending on meteorological variables and environmental stressors (e.g. Sims et al., 2006).

Some observed mechanisms that are relevant in real-world grasslands may not be present in models. For example, under warming conditions a higher C accumulation in the root system is expected, reflecting the physiology of root respiration. Increased root weight (i.e. decreased root respiration rate and/or increased C requirements as a fraction of C fixed in photosynthesis) may be due to more rapid depletion of nutrients in the root environment as a consequence of increased growth at higher temperatures, but these effects are likely to be smaller in systems where nutrient supply is higher (e.g. intensively-managed systems), and to increase with decreasing soil fertility (e.g. Lambers et al., 1996). When the production of new structures is the process most affected by nutrient shortages, a decrease in the shoot:root ratio is to be expected, according to the concept of C-N equilibrium (e.g. Thornley, 1972). Mediated by plant biomass inputs through their effects on available C in the rooting zone, warming has the effect of increasing the mineralisation of soil organic matter (e.g. Carrillo et al., 2010): more nutrients are released and, with more roots, plants can take up this excess nutrient. In addition, moderate drought can lead to a shortage of water supply, which favours C partitioning to roots (Pilon et al., 2013). Under these conditions, low management intensity can alleviate N limitation by stimulating N mineralisation and uptake (Dangal et al., 2016). Water stress is likely to reduce root respiration rates per unit root mass, but increase the fraction of total assimilates required for root respiration, due to a higher allocation of biomass to roots. Rising temperatures and low soil moisture affect both photosynthesis and plant respiration. Moreover, as changes in seasonal water availability have pronounced effects on individual species, adaptation and functional biodiversity would buffer the effect of climate extremes on C cycling, but the basis for biological adaptation and the role of biodiversity remain unclear (Reichstein et al., 2013). The resolution of this study did not allow us to examine the limitations of different modelling solutions to deal with these complex patterns. It would require further study, as they may lead to an over- or underestimation of biomass with cascading effects related to the nature of multi-species ecosystems that remain on the same soil for a long time.

Unlike annual crops (requiring replanting every year), permanent grasslands are characterised by production from the same vegetation cover, generally as mixtures of plant species with a high dissimilarity of traits. Source-sink relationships in grasslands are highly dynamic and vary according to plant growth stage, fertiliser application, grazing and mowing intensity and timing. Under these conditions, it is difficult to define some of the plant parameters or management-related settings in the models. For example, the relationship between aboveground production and standing biomass after cutting or grazing can only be approximated. This means that plant regrowth after cutting cannot be readily estimated from the standing biomass, which inevitably affects the quality of the simulations. In addition, the relationship between aboveground biomass and plant residues is driven by the senescence process of the various plant organs, which are strongly influenced by the stage of development at which they are removed. Plant residues added to the soil during plant growth can trigger high N release (e.g. Oelmann et al., 2011), while stimulating soil microbial decomposition processes (with feedback on biomass growth). In most cases, modellers have little or no control over senescence processes and litterfall, and departure of estimated yields from observations can be ascribed to suboptimal estimation of related parameters.

4.3. Comparison with manipulation experiments

Ecosystem manipulation experiments provide an opportunity to understand the response of the plant-soil system to changing environmental conditions (Norby and Luo, 2004; Ainsworth and Long, 2005; Soussana and Lüscher, 2007; De Boeck et al., 2008; Norby and Zak, 2011; Song et al., 2019; Ainsworth and Long, 2020; Pan et al., 2022). However, despite major advances in the understanding of processes and the availability of a wide array of experimental data, uncertainties remain large. Focusing specifically on grasslands, species distribution, disturbance regimes, climatic conditions including extremes, soil conditions and management all contribute to uncertainty and a wide diversity of responses to changing environmental conditions (Song et al., 2019). Additional uncertainty arises from the diversity of technologies used for manipulation (Wang et al., 2019; Section 2.5). Since a single-factor change in temperature, precipitation or [CO$_2$] already triggers a cascade of events in the ecosystem, the usability of multi-factor experiments for understanding processes and supporting model development is limited. For this reason, and also because the literature is limited for such modelling exercises, we focused here exclusively on single-factor climate manipulation experiments.

This study was not intended to provide a comprehensive review of the results of single-factor manipulation experiments. It means that the data collection based on the literature search was not done in a systematic ways that is used e.g. in meta-analyses (Rustad et al., 2001; Lin et al., 2010; Wu et al., 2011; Dieleman et al., 2012; Wang et al., 2019; Lu et al., 2013; Xu et al., 2013; Fu et al., 2015; Unger and Jongen, 2015; Wilcox et al., 2017; Chen et al., 2020; Piseddu et al., 2021; Shao et al., 2022). Instead, we used strict data selection criteria to include experiments that could be considered comparable to the modelling exercise. This limited the availability of the data. Nevertheless, careful checking of data quality, removal of duplicates and rejection of problematic experimental results ensured a solid basis for comparison.

Studies focusing on the spatial variability of the observed responses are scarce (e.g. Wilcox et al., 2017). In this study, we used site conditions to identify patterns in biomass responses (using a simple linear approach) as an additional benchmark for the sensitivity analysis. The results indicated that for some manipulations (drying and [CO$_2$] enrichment), the overall relationship between site conditions and model sensitivity was consistent with observations, but for other manipulations (warming and wetting), there was a lack of consistency. Given our expectation to capture spatial and temporal variability of grassland functioning through modelling, the results highlight potential issues that need to be addressed through model development and repeated benchmarking. At larger spatial scales, model results should be interpreted with caution, as responses (reflecting different site and climatic conditions) could be problematic, especially in the case of warming and drying.

The study highlighted that the response of SWC and ET to global change factors is not fully realistic within the spread of simulated

responses obtained with the models studied (although limited observed data were available for ET). It is generally accepted that major issues exist regarding the CO_2 fertilisation effect in ecosystem models and ESMs (Smith et al., 2016), but less emphasis is placed on the problems associated with SWC dynamics. According to Li et al. (2022), SWC is highly relevant worldwide, thus improving the representation of soil hydrology is a key step, especially for grasslands, which are more sensitive to water availability than other ecosystems.

Challenges are also posed by the fact that most models cannot account for the long-term acclimation of photosynthetic activity to elevated [CO_2], the so-called down-regulation of photosynthesis, which is related to N availability (Ainsworth and Rogers, 2007). A down-regulation strategy seems necessary to limit the effect of increased [CO_2] on plant growth, also considering that the beneficial effects of higher [CO_2] on C assimilation may be outweighed or downplayed by concomitant changes in other environmental factors (Rosenzweig and Parry, 1994). In contrast, as pointed out by Saban et al. (2019) based on data collected in the proximity of natural CO_2 springs, the response of plants to elevated [CO_2] could remain large and consistent with FACE experiments over longer time periods. In any case, the joint application of experimental data and model simulations are the only way to advance our understanding of grassland response to climate change factors (either single or in combination).

Elevated [CO_2] and warming are known to have fundamentally different effects on C storage and cycling in terrestrial ecosystems (Schimel et al., 2014), meaning that the dominant mechanism might differ depending on climate. Dieleman et al. (2012) proposed to benchmark the models against elevated [CO_2] experiments rather than with warming experiments, given the expected larger effect. Based on our results, we recommend testing the models against single-factor experiments that include warming, precipitation addition/removal and elevated [CO_2] to get a more complete picture of the mechanistic responses of grassland models.

4.4. Limitations of the study

This work extends previous research on grassland modelling (e.g. Sándor et al., 2017) and is intended to be useful as preliminary investigation prior to large-scale studies. This study did not aim to project expected scenario-based impacts of climate change on European grasslands. Rather, the single-factor sensitivity of a set of grassland models to changes in temperature, precipitation and [CO_2] was analysed.

A first limitation of the study is that no combination of changes in temperature, precipitation and [CO_2] was investigated, and the altered scenarios as designed are not equivalent to climate change projections, as they do not fully account for changes in seasonality, the occurrence of extreme events and changes in other climatic variables. In this respect, the changes designed in this study may not cover specific future climate conditions.

Secondly, adaptation of grassland systems to climate change may drastically modify the actual response of grassland production to altered conditions. Adaptation may be due to the inherent ecological capacity of grasslands to evolve to a new state, e.g. through a modification in species composition or through long-term acclimation of existing species (e.g. Way and Yamori, 2014; Noguchi et al., 2015). This could be translated into models by adapting plant parameter values. In this respect, experimental evidence confirms that the response of grasslands to drought varies according to site characteristics (Knapp et al., 2015). In addition, management options can be adapted to altered conditions, e.g. by modifying mowing frequency or grazing pressure to cope with drought (Luna et al., 2023).

Another limitation is that the models were calibrated using only eddy covariance and production data from observational sites, and the calibrated parameters are considered sufficiently representative of European grasslands. Thus, vegetation was modelled as a single group at all grassland sites considered in the study, without accounting for site-to-

site variability of vegetation parameters. In this way, model sensitivity was based on the same loose assumptions that are set down and used to model grassland ecosystem responses in large-scale studies (e.g. Ma et al., 2015). However, a proper calibration at more arid sites would improve model parameterisation and responses to temperature and water constraints typical of these sites. Consideration of biodiversity would be a logical additional step in the future to improve the models (van Oijen et al., 2020).

Another aspect that needs further investigation is the duration of climate manipulation experiments, as during long-term climate manipulation trials the initial larger effect might diminish over time (e.g. Leuzinger et al., 2011; Wang et al., 2012) and uncertainty is large in this aspect due to the scarcity of long-term experiments. In any case, an abrupt change in environmental conditions has different implications from the gradual alteration that characteristics climate change.

The length of the growing season could interfere with the results, which was not explicitly studied. It is necessary to mention that current models have issues with estimating the timing of intensive spring growth and autumn senescence (Piao et al., 2019). This means that changing environmental conditions and the length of the growing season interact in a way that is difficult to quantify. Error compensation or even error amplification could occur within the models due to issues related to phenological patterns.

5. Concluding remarks

This study has shed light on critical aspects of grassland modelling and underscores the paramount importance of rigorous model evaluation. While the existing common practice of model evaluation and optimisation has its merits, our findings call for a paradigm shift, with a proposed benchmarking approach standing out as a more robust, critical and forward-looking alternative. Through a detailed sensitivity analysis focusing on key climate change factors – namely, temperature, precipitation and CO_2 concentration –, we have unveiled the intricate interplay between these factors and grassland model responses. The inherent variability observed in our simulations mirrors similar challenges encountered in crop modelling (e.g. Asseng et al., 2013; Bassu et al., 2014; Wallach et al., 2018). These uncertainties highlight the need for structural enhancements in grassland models. One promising avenue for addressing these uncertainties is the adoption of ensembles, a technique that can mitigate the limitations of the individual models. The implications of this study extend beyond the realm of grassland modelling. It serves as a valuable resource for both model users and the broader modelling community engaged in simulation experiments within this domain. By comparing the results of our selected models with the sensitivities of alternative models, researchers can gain valuable insights that drive innovation and improvement in the field.

In conclusion, our research charts a distinct and actionable course for the refinement of grassland modelling. As we work towards the attainment of accuracy and reliability in agroecosystem modelling, the ongoing journey in this pursuit may yield tangible results, such as informing more sustainable grazing practices. For example, our framework's application may serve to enhance the predictive power of grassland models regarding the impact of changing rainfall patterns on the standing biomass available for livestock and, consequently, on grazing habits.

The collective engagement and expertise within the grassland modelling community are indispensable, as evidenced by the collaborative projects that have supported this research. Moreover, the modeller community is supposed to approach the experimentalist and initiate joint projects. This includes the dissemination of the observed data by the experimentalist so that the modelling community can easily get access and exploit the results. These collaborative endeavours will undoubtedly continue to shape a future where grasslands are a cornerstone of sustainable agricultural practices and ecological balance.

CRediT authorship contribution statement

G. Bellocchi: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Project administration, Supervision. **Z. Barcza:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision. **R. Hollós:** Methodology, Resources, Formal analysis, Writing – original draft, Writing – review & editing. **M. Acutis:** Data curation. **E. Bottyán:** Data curation, Visualization, Investigation, Writing – original draft, Writing – review & editing. **L. Doro:** Software, Data curation. **D. Hidy:** Software, Validation. **E. Lellei-Kovács:** Software, Investigation, Writing – original draft. **S. Ma:** Software, Data curation. **J. Minet:** Software, Data curation. **V. Pacskó:** Data curation, Visualization, Investigation, Writing – original draft, Writing – review & editing. **A. Perego:** Software, Data curation. **F. Ruget:** Software, Data curation. **G. Seddaiu:** Software, Data curation. **L. Wu:** Software, Data curation. **R. Sándor:** Data curation, Visualization, Writing – original draft, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data are available from the corresponding author upon request. In some cases permission is needed from the modellers.

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2023.109778.

References

Ainsworth, E.A., Long, S.P., 2005. What have we learned from 15 years of free-air CO_2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO_2. New Phytol 165, 351–372. https://doi.org/10.1111/j.1469-8137.2004.01224.x.

Ainsworth, E.A., Long, S.P., 2020. 30 years of free-air carbon dioxide enrichment (FACE): what have we learned about future crop productivity and its potential for adaptation? Global Change Biol 27, 27–49. https://doi.org/10.1111/gcb.15375.

Ainsworth, E.A., Rogers, A., 2007. The response of photosynthesis and stomatal conductance to rising $[CO_2]$: mechanisms and environmental interactions. Plant Cell Environ 30, 258–270. https://doi.org/10.1111/j.1365-3040.2007.01641.x.

Ammann, C., Flechard, C.R., Leifeld, J., Neftel, A., Fuhrer, J., 2007. The carbon budget of newly established temperate grassland depends on management intensity. Agric. Ecosyst. Environ. 121, 5–20. https://doi.org/10.1016/j.agee.2006.12.002.

Andresen, L.C., Yuan, N., Seibert, R., Moser, G., Kammann, C.I., Luterbacher, J., Erbs, M., Müller, C., 2018. Biomass responses in a temperate European grassland through 17 years of elevated CO_2. Global Change Biol 24, 3875–3885. https://doi.org/10.1111/gcb.13705.

Antle, J.M., Basso, B., Conant, R.T., Godfray, H.C.J., Jones, J.W., Herrero, M., Howitt, R. E., Keating, B.A., Munoz-Carpena, R., Rosenzweig, C., Tittonell, P., Wheeler, T.R., 2017. Towards a new generation of agricultural system data, models and knowledge products: design and improvement. Agric. Syst. 155, 255–268. https://doi.org/10.1016/j.agsy.2016.10.002.

Aronson, E.L., McNulty, S.G., 2009. Appropriate experimental ecosystem warming methods by ecosystem, objective, and practicality. Agric. For. Meteorol. 149, 1791–1799. https://doi.org/10.1016/j.agrformet.2009.06.007.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A., Boote, K.J., Thorburn, P., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J.E., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Travasso, M., Tao, F., Waha, K., Wallach, D., White, J.W., Wolf, J., 2013. Uncertainties in simulating wheat yields under climate change. Nat. Clim. Change 3, 827–832. https://doi.org/10.1038/nclimate1916.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., Betts, A.K., 2009. A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the integrated forecast system. J. Hydrometeorol. 10, 623–643. https://doi.org/10.1175/2008JHM1068.1.

Bassu, S., Brisson, N., Durand, J.L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S.H., Kumar, N.S., Makowski, D., Müller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses to climate change factors? Global Change Biol 20, 2301–2320. https://doi.org/10.1111/gcb.12520.

Beier, C., Emmett, B., Gundersen, P., Tietema, A., Peñuelas, J., Estiarte, M., Gordon, C., Gorissen, A., Llorens, L., Roda, F., Williams, D., 2004. Novel approaches to study climate change effects on terrestrial ecosystems in the field: drought and passive nighttime warming. Ecosystems 7, 583–597. https://doi.org/10.1007/s10021-004-0178-8.

Brilli, L., Bechini, L., Bindi, M., Carozzi, M., Cavalli, D., Conant, R., Dorich, C.D., Doro, L., Ehrhardt, F., Farina, R., Ferrise, R., Fitton, N., Francaviglia, R., Grace, P., Iocola, I., Klumpp, K., Léonard, J., Martin, R., Massad, R.S., Recous, S., Seddaiu, G., Sharp, J., Smith, P., Smith, W.N., Soussana, J.-F., Bellocchi, G., 2017. Review and analysis of strengths and weaknesses of agro-ecosystem models for simulating C and N fluxes. Sci. Total Environ. 598, 445–470. https://doi.org/10.1016/j.scitotenv.2017.03.208.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussiere, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillere, J.P., Henault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model STICS. Eur. J. Agron. 18, 309–332. https://doi.org/10.1016/S1161-0301(02)00110-7.

Butler, M.P., Reed, P.M., Fisher-Vanden, K., Keller, K., Wagener, T., 2014. Identifying parametric controls and dependencies in integrated assessment models using global sensitivity analysis. Environ. Model. Softw. 59, 10–29. https://doi.org/10.1016/j.envsoft.2014.05.001.

Carrillo, Y., Pendall, E., Dijkstra, F.A., Morgan, J.A., Newcomb, J.M., 2010. Carbon input control over soil organic matter dynamics in a temperate grassland exposed to elevated CO_2 and warming. Biogeosci 7, 1575–1602. https://doi.org/10.5194/bgd-7-1575-2010.

Cavallero, A., Talamucci, P., Grignani, C., Reyneri, A., Ziliotto, U., Scotton, M., Bianchi, A.A., Santilocchi, R., Basso, F., Postiglione, L., Carone, F., Corleto, A., Cazzato, E., Cassaniti, S., Cosentino, S., Litrico, P.G., Leonardi, S., Sarno, R., Stringi, L., Gristina, L., Amato, G., Bullitta, P., Caredda, S., Roggero, P.P., Caporali, F., D'Antuono, L.F., Pardini, A., Zagni, C., Piemontese, S., Pazzi, G., Costa, G., Pascal, G., Acutis, M., 1992. Caratterizzazione della dinamica produttiva di pascoli naturali italiani. Rivista di Agronomia 26 (3), 325–343 nsuppIItalian.

Chen, F., Dudhia, J., 2001a. Coupling an advanced land surface hydrology model with the Penn State NCAR MM5 modeling system. Part I: model implementation and sensitivity. Mon. Weather Rev. 129, 569–585. https://doi.org/10.1175/1520-0493 (2001)129<0569:CAALSH>2.0.CO;2.

Chen, F., Dudhia, J., 2001b. Coupling an advanced land surface hydrology model with the Penn State NCAR MM5 modeling system. Part II: preliminary model validation. Mon. Weather Rev. 129, 587–604. https://doi.org/10.1175/1520-0493(2001) 129<0587:CAALSH>2.0.CO;2.

Chen, Y., Feng, J., Yuan, X., Zhu, B., 2020. Effects of warming on carbon and nitrogen cycling in alpine grassland ecosystems on the Tibetan Plateau: a meta-analysis. Geoderma 370, 114363. https://doi.org/10.1016/j.geoderma.2020.114363.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buhmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A.D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J.M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J.F., Sanz, M.J., Schulze, E.D., Vesala, T., Valentini, R., 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437, 529–533. https://doi.org/10.1038/nature03972.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Le Quéré, C., Myneni, R.B., Piao, S., Thornton, P., 2013. Carbon and other biogeochemical cycles. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp. 465–570.

Cure, J.D., Acock, B., 1986. Crop responses to carbon dioxide doubling: a literature survey. Agric. For. Meteorol. 38, 127–145. https://doi.org/10.1016/0168-1923(86) 90054-7.

Dabros, A., Fyles, J.W., Strachan, I.B., 2010. Effects of open-top chambers on physical properties of air and soil at post-disturbance sites in northwestern Quebec. Plant Soil 333, 203–218. https://doi.org/10.1007/s11104-010-0336-z.

Dangal, S.R.S., Tian, H., Lu, C., Pan, S., Pederson, N., Hessl, A., 2016. Synergistic effects of climate change and grazing on net primary production of Mongolian grasslands. Ecosphere 5, e01274. https://doi.org/10.1002/ecs2.1274.

De Boeck, H.J., Lemmens, C.M.H.M., Gielen, B., Malchair, S., Carnol, M., Merckx, R., Van den Berge, J., Ceulemans, R., Nijs, I., 2008. Biomass production in experimental grasslands of different species richness during three years of climate warming. Biogeosciences 5, 585–594. https://doi.org/10.5194/bg-5-585-2008.

Dieleman, W.I.J., Vicca, S., Dijkstra, F.A., Hagedorn, F., Hovenden, M.J., Larsen, K.S., Morgan, J.A., Volder, A., Beier, C., Dukes, J.S., King, J., Leuzinger, S., Linder, S., Luo, Y., Oren, R., De Angelis, P., Tingey, D., Hoosbeek, M.R., Janssens, I.A., 2012. Simple additive effects are rare: a quantitative review of plant biomass and soil process responses to combined manipulations of CO_2 and temperature. Global Change Biol 18, 2681–2693. https://doi.org/10.1111/j.1365-2486.2012.02745.x.

Diodato, N., Brocca, L., Bellocchi, G., Fiorillo, F., Guadagno, F.M., 2014. Complexity-reduction modelling for assessing the macro-scale patterns of historical soil moisture in the Euro-Mediterranean region. Hydrol. Processes 28, 3752–3760. https://doi.org/10.1002/hyp.9925.

EC (European Commission), 2008. Regulation (EC) No. 1166/2008 of the European Parliament and of the Council of 19 November 2008 On Farm Structure Surveys and the Survey on Agricultural Production Methods and Repealing Council Regulation (EEC).

Farquhar, G.C., Von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species. Planta 149, 78–90. https://doi.org/10.1007/BF00386231.

Field, C.B., Barros, V., Stocker, T.F., Dahe, Q., 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York.

Franke, J.A., Müller, C., Minoli, S., Elliott, J., Folberth, C., Gardner, C., Hank, T., Izaurralde, R.C., Jägermeyr, J., Jones, C.D., Liu, W., Olin, S., Pugh, T.A.M., Ruane, A.C., Stephens, H., Zabel, F., Moyer, E.J., 2021. Agricultural breadbaskets shift poleward given adaptive farmer behavior under climate change. Global Change Biol 28, 167–181. https://doi.org/10.1111/gcb.15868.

Fu, Z., Niu, S., Dukes, J., 2015. What have we learned from global change manipulative experiments in China? A meta-analysis. Sci. Rep. 5 (1), 12344. https://doi.org/10.1038/srep12344.

Fu, Z., Ciais, P., Makowski, D., Bastos, A., Stoy, P.C., Ibrom, A., Knohl, A., Migliavacca, M., Cuntz, M., Šigut, L., Peichl, M., Loustau, D., El-Madany, T.S., Buchmann, N., Gharun, M., Janssens, I., Markwitz, C., Grünwald, T., Rebmann, C., Mölder, M., Varlagin, A., Mammarella, I., Kolari, P., Bernhofer, C., Heliasz, M., Vincke, C., Pitacco, A., Cremonese, E., Foltýnová, L., Wigneron, J.-P., 2022. Uncovering the critical soil moisture thresholds of plant water stress for European ecosystems. Global Change Biol 28, 2111–2123. https://doi.org/10.1111/gcb.16050.

Golodets, C., Sternberg, M., Kigel, J., Boeken, B., Henkin, Z., Seligman, N.G., Ungar, D.E., 2013. From desert to Mediterranean rangelands: will increasing drought and inter-annual rainfall variability affect herbaceous annual primary productivity? Clim. Change 119, 785–798. https://doi.org/10.1007/s10584-013-0758-8.

Graux, A.-I., Bellocchi, G., Lardy, R., Soussana, J.-F., 2013. Ensemble modelling of climate change risks and opportunities for managed grasslands in France. Agric. For. Meteorol. 170, 114–131. https://doi.org/10.1016/j.agrformet.2012.06.010.

Hidy, D., Barcza, Z., Haszpra, L., Churkina, G., Pintér, K., Nagy, Z., 2012. Development of the Biome-BGC model for simulation of managed herbaceous ecosystems. Ecol. Modell. 226, 99–119. https://doi.org/10.1016/j.ecolmodel.2011.11.008.

Hidy, D., Barcza, Z., Marjanović, H., Ostrogović Sever, M.Z., Dobor, L., Gelybó, Gy., Fodor, N., Pintér, K., Churkina, G., Running, S.W., Thornton, P.E., Bellocchi, G., Haszpra, L., Horváth, F., Suyker, A., Nagy, Z., 2016. Terrestrial ecosystem process model Biome-BGCMuSo: summary of improvements and new modeling possibilities. Geosci. Model Dev. 9, 4405–4437. https://doi.org/10.5194/gmd-9-4405-2016.

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.G., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017. Toward a new generation of agricultural system data, models, and knowledge products: state of agricultural systems science. Agric. Syst. 155, 269–288. https://doi.org/10.1016/j.agsy.2016.09.021.

Kimball, B.A., 2005. Theory and performance of an infrared heater for ecosystem warming. Global Change Biol 11, 2041–2056. https://doi.org/10.1111/j.1365-2486.2005.01028.x.

Kipling, R., Virkajärvi, P., Breitsameter, L., Curnel, Y., De Swaef, T., Gustavsson, A.-M., Hennart, S., Högland, M., Järvenranta, K., Minet, J., Nendel, C., Persson, T., Picon-Cochard, C., Rolinski, S., Sandars, D.L., Scollan, N.D., Sebek, L., Seddaiu, G., Bellocchi, G., 2016. Key challenges and priorities for modelling European grasslands under climate change. Sci. Total Environ. 566–567, 851–864. https://doi.org/10.1016/j.scitotenv.2016.05.144.

Klumpp, K., Tallec, T., Guix, N., Soussana, J.-F., 2011. Long-term impacts of agricultural practices and climatic variability on carbon storage in a permanent pasture. Global Change Biol 17, 3534–3545. https://doi.org/10.1111/j.1365-2486.2011.02490.x.

Knapp, A.K., Hoover, D.L., Wilcox, K.R., Avolio, M.L., Koerner, S.E., La Pierre, K.J., Loik, M.E., Luo, Y., Sala, O.E., Smith, M.D., 2015. Characterizing differences in precipitation regimes of extreme wet and dry years: implications for climate change experiments. Global Change Biol 21, 2624–2633. https://doi.org/10.1111/gcb.12888.

Köchy, M., Wilson, S.D., 2004. Semiarid grassland responses to short-term variation in water availability. Plant Ecol 174, 197–203. https://doi.org/10.1023/B:VEGE.0000049098.74147.57.

Lambers, H., Stulen, I., van der Werf, A., 1996. Carbon use in root respiration as affected by elevated atmospheric CO_2. Plant Soil 187, 251–263. https://doi.org/10.1007/BF00017091.

Leuzinger, S., Luo, Y., Beier, C., Dieleman, W., Vicca, S., Körner, C., 2011. Do global change experiments overestimate impacts on terrestrial ecosystems? Trends Ecol. Evol. 26, 236–241. https://doi.org/10.1016/j.tree.2011.02.011.

Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida III, M., Nakagawa, H., Oriol, P., Ruane, A.C., Ruget, F., Singh, B., Singh, U., Tang, L., Tao, F., Wilkens, P., Yoshida, H., Zhang, Z., Bouman, B., 2015. Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Global Change Biol 21, 1328–1341. https://doi.org/10.1111/gcb.12758.

Li, W., Migliavacca, M., Forkel, M., Denissen, J.M.C., Reichstein, M., Yang, H., Duveiller, G., Weber, U., Orth, R., 2022. Widespread increasing vegetation sensitivity to soil moisture. Nat. Commun. 13 (1), 3959. https://doi.org/10.1038/s41467-022-31667-9.

Lin, D., Xia, J., Wan, S., 2010. Climate warming and biomass accumulation of terrestrial plants: a meta-analysis. New Phytol 188, 187–198. https://doi.org/10.1111/j.1469-8137.2010.03347.x.

Lu, M., Zhou, X., Yang, Q., Li, H., Luo, Y., Fang, C., Chen, J., Yang, X., Li, B., 2013. Responses of ecosystem carbon cycle to experimental warming: a meta-analysis. Ecol 94, 726–738. https://doi.org/10.1890/12-0279.1.

Luna, D.A., Pottier, J., Picon-Cochard, C., 2023. Variability and drivers of grassland sensitivity to drought at different timescales using satellite image time series. Agric. For. Meteorol. 331, 109325 https://doi.org/10.1016/j.agrformet.2023.109325.

Ma, S., Lardy, R., Graux, A.-I., Ben Touhami, H., Klumpp, K., Martin, R., Bellocchi, G., 2015. Regional-scale analysis of carbon and water cycles on managed grassland

systems. Environ. Model. Softw. 72, 356–371. https://doi.org/10.1016/j. envsoft.2015.03.007.

Marion, G.M., Henry, G.H.R., Freckman, D.W., Johnstone, J., Jones, G., Jones, M.H., Lévesque, E., Molau, U., Mølgaard, P., Parsons, A.N., Svoboda, J., Virginia, R.A., 1997. Open-top designs for manipulating field temperature in high-latitude ecosystems. Global Change Biol 3, 20–32. https://doi.org/10.1111/j.1365-2486.1997.gcb136.x.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rotter, R.P., Boote, K.J., Ruane, A.C., Thorburn, P.J., Cammarano, D., Hatfield, J.L., Rosenzweig, C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S.N., Nendel, C., O'leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherback, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2015. Multi-model ensembles of wheat growth: many models are better than one. Global Change Biol 21, 911–925. https://doi.org/10.1111/gcb.12768.

Nagy, Z., Barcza, Z., Horváth, L., Balogh, J., Hagyó, A., Káposztás, N., Grosz, B., Machon, A., Pintér, K., 2010. Measurements and estimations of biosphere-atmosphere exchange of greenhouse gases – grasslands. In: Haszpra, L. (Ed.), Atmospheric Greenhouse gases: the Hungarian perspective. Springer, Dordrecht, pp. 91–119.

Noguchi, K., Yamori, W., Hikosaka, K., Terashima, I., 2015. Homeostasis of the temperature sensitivity of respiration over a range of growth temperatures indicated by a modified Arrhenius model. New Phytol 207, 34–42. https://doi.org/10.1111/nph.13339.

Norby, R.J., Luo, Y., 2004. Evaluating ecosystem responses to rising atmospheric CO_2 and global warming in a multi-factor world. New Phytol 162, 281–293. https://doi.org/10.1111/j.1469-8137.2004.01047.x.

Norby, R.J., Zak, D.R., 2011. Ecological lessons from free-air CO_2 Enrichment (FACE) Experiments. Annu. Rev. Ecol. Evol. Syst. 42 (1), 181–203. https://doi.org/10.1146/annurev-ecolsys-102209-144647.

Oelmann, Y., Buchmann, N., Gleixner, G., Habekost, M., Roscher, C., Rosenkranz, S., Schulze, E.-D., Steinbeiss, S., Temperton, V.M., Weigelt, A., Weisser, W., Wilcke, W., 2011. Plant diversity effects on aboveground and belowground N pools in temperate grassland ecosystems: development in the first 5 years after establishment. Global Biogeochem. Cycles 25, GB2014. https://doi.org/10.1029/2010GB003869.

et al. Pan, Y., Jackson, R.B., Hollinger, D.Y., Phillips, O.L., Nowak, R.S., Norby, R.J., Oren, R., Reich, P.B., Lüscher, A., Mueller, K.E., Owensby, C., Birdsey, R., Hom, J., Luo, Y., 2022. Contrasting responses of woody and grassland ecosystems to increased CO_2 as water supply varies Nat. Ecol. Evol. 6, 315–323. https://doi.org/10.1038/s41559-021-01642-6.

Perego, A., Giussani, A., Sanna, M., Fumagalli, M., Carozzi, M., Alfieri, L., Brenna, S., Acutis, M., 2013. The ARMOSA simulation crop model: overall features, calibration and validation results. Ital. J. Agrometeorol. 3, 23–38.

Piao, S., Liu, Q., Chen, A., Janssens, I.A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., Zhu, X., 2019. Plant phenology and global climate change: current progresses and challenges. Global Change Biol 25, 1922–1940. https://doi.org/10.1111/gcb.14619.

Pilon, R., Picon-Cochard, C., Bloor, J.M.G., Revaillot, S., Kuhn, E., Falcimagne, R., Balandier, P., Soussana, J.-F., 2013. Grassland root demography responses to multiple climate change drivers depend on root morphology. Plant Soil 364, 395–408. https://doi.org/10.1007/s11104-012-1371-8.

Piseddu, F., Bellocchi, G., Picon-Cochard, C., 2021. Mowing and warming effects on grassland species richness and harvested biomass: meta-analyses. Agron. Sustain. Dev. 41, 74. https://doi.org/10.1007/s13593-021-00722-y.

Pokovai, K., Hollós, R., Bottyán, E., Kis, A., Marton, T., Pongrácz, R., Pásztor, L., Hidy, D., Barcza, Z., Fodor, N., 2020. Estimation of agro-ecosystem services using biogeochemical models. Időjárás 124, 209–225. https://doi.org/10.28974/idojaras.2020.2.4.

Prescher, A.-K., Grünwald, T., Bernhofer, C., 2010. Land use regulates carbon budgets in eastern Germany: from NEE to NBP. Agric. For. Meteorol. 150, 1016–1025. https://doi.org/10.1016/j.agrformet.2010.03.008.

R Core Team, 2023. R: A language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M.D., Seneviratne, S.I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D.C., Papale, D., Rammig, A., Smith, P., Thonicke, K., van der Velde, M., Vicca, S., Walz, A., Wattenbach, M., 2013. Climate extremes and the carbon cycle. Nature 500, 287–295. https://doi.org/10.1038/nature12350.

Richards, L.A., 1931. Capillary conduction of fluid through porous mediums. Phys 1, 318–333. https://doi.org/10.1063/1.1745010.

Ritchie, J.T., 1981. Water dynamics in the soil-plant-atmosphere system. Plant Soil 58, 81–96. https://doi.org/10.1007/BF02180050.

Rosenzweig, C., Parry, M.L., 1994. Potential impact of climate change on world food supply. Nature 367, 133–138. https://doi.org/10.1038/367133a0.

Rustad, L.E., Campbell, J.L., Marion, G.M., Norby, R.J., Mitchell, M.J., Hartley, A.E., Cornelissen, J.H.C., Gurevitch, J., News, G.T.C.E., 2001. A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming. Oecologia 126, 543–562. https://doi.org/10.1007/s004420000544.

Saban, J.M., Chapman, M.A., Taylor, G., 2019. FACE facts hold for multiple generations; Evidence from natural CO_2 springs. Global Change Biol 25, 1–11. https://doi.org/10.1111/gcb.14437.

Sándor, R., Acutis, M., Barcza, Z., Doro, L., Hidy, D., Köchy, M., Minet, J., Lellei-Kovács, E., Ma, S., Perego, A., Rolinksi, S., Ruget, F., Sanna, M., Seddaiu, G., Wu, L.,

Bellocchi, G., 2017. Multi-model simulation of soil temperature, soil water content and biomass in Euro-Mediterranean grasslands: uncertainties and ensemble performance. Eur. J. Agron. 88, 22–40. https://doi.org/10.1016/j.eja.2016.06.006.

Sándor, R., Barcza, Z., Hidy, D., Lellei-Kovács, E., Ma, S., Bellocchi, G., 2016. Modelling of grassland fluxes in Europe: evaluation of two biogeochemical models. Agric. Ecosyst. Environ. 215, 1–19. https://doi.org/10.1016/j.agee.2015.09.001.

Schils, R., Snijders, P., 2004. The combined effect of fertiliser nitrogen and phosphorus on herbage yield and changes in soil nutrients of a grass/clover and grass-only sward. Nutr. Cycl. Agroecosystems 68, 165–179. https://doi.org/10.1023/B:FRES.0000019045.90791.a4.

Schimel, D., Stephens, B.R., Fisher, J.B., 2014. Effect of increasing CO_2 on the terrestrial carbon cycle. Proc. Natl. Acad. Sci. U.S.A. 112, 436–441. https://doi.org/10.1073/pnas.140730211.

Schröpel, R., Diepolder, M., 2003. Auswirkungen Der Grünlandextensivierung auf Einer Weidelgras-Weißklee-Weide im Allgäuer Alpenvorland. Schuleund Beratung, Heft 11/2003, Seite III–13 Bis III–15. Bayerisches Staatsministerium für Landwirtschaft und Forsten, Munich.. German.

Schwanitz, V.J., 2013. Evaluating integrated assessment models of global climate change. Environ. Model. Softw. 50, 120–131. https://doi.org/10.1016/j.envsoft.2013.09.005.

Seneviratne, S.I., Lüthi, D., Litschi, M., Schär, C., 2006. Land-atmosphere coupling and climate change in Europe. Nature 443, 205–209. https://doi.org/10.1038/nature05095.

Seneviratne, S.I., Nicholls, N., Easterling, D., Goodess, C.M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., Zhang, X., 2012. Changes in climate extremes and their impacts on the natural physical environment. In: Field, C.B., Barros, V., Stocker, T.F., Qin, D., Dokken, D.J., Ebi, K.L., Mastrandrea, M.D., Mach, K.J., Plattner, G.K, Allen, S.K., Tignor, M., Midgley, P.M. (Eds.), Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. Cambridge University Press, Cambridge, pp. 109–230.

Shao, J., Zhou, X., van Groenigen, K.J., Zhou, G., Zhou, H., Zhou, L., Lu, M., Xia, J., Jiang, L., Hungate, B.A., Luo, Y., He, F., Thakur, M.P., 2022. Warming effects on grassland productivity depend on plant diversity. Global Ecol. Biogeogr. 31, 588–598. https://doi.org/10.1111/geb.13441.

Silvertown, J., Poulton, P., Johnston, E., Edwards, G., Heard, M., Biss, P.M., 2006. The Park Grass Experiment 1856–2006: its contribution to ecology. J. Ecol. 94, 801–814. https://doi.org/10.1111/j.1365-2745.2006.01145.x.

Sims, D.A., Rahman, A.F., Cordova, V.D., El-Masri, B.Z., Baldocchi, D.D., Flanagan, L.B., Goldstein, A.H., Hollinger, D.Y., Misson, L., Monson, R.K., Oechel, W.C., Schmid, H. P., Wofsy, S.C., Xu, L., 2006. On the use of MODIS EVI to assess gross primary productivity of North American ecosystems. J. Geophys. Res. G: Biogeosci. 111, G04015. https://doi.org/10.1029/2006JG000162.

Smith, W., Reed, S., Cleveland, C., Ballantyne, A., Anderegg, W., Wieder, W., Liu, Y., Running, S., 2016. Large divergence of satellite and Earth system model estimates of global terrestrial CO_2 fertilization. Nat. Clim. Change 6, 306–310. https://doi.org/10.1038/nclimate2879.

Song, J., Wan, S., Piao, S., Knapp, A.K., Classen, A.T., Vicca, S., Ciais, P., Hovenden, M.J., Leuzinger, S., Beier, C., Kardol, P., Xia, J., Liu, Q., Ru, J., Zhou, Z., Luo, Y., Guo, D., Adam Langley, J., Zscheischler, J., Dukes, J.S., Tang, J., Chen, J., Hofmockel, K.S., Kueppers, L.M., Rustad, L., Liu, L., Smith, M.D., Templer, P.H., Quinn Thomas, R., Norby, R.J., Phillips, R.P., Niu, S., Fatichi, S., Wang, Y., Shao, P., Han, H., Wang, D., Lei, L., Wang, J., Li, X., Zhang, Q., Li, X., Su, F., Liu, B., Yang, F., Ma, G., Li, G., Liu, Y., Liu, Y., Yang, Z., Zhang, K., Miao, Y., Hu, M., Yan, C., Zhang, A., Zhong, M., Hui, Y., Li, Y., Zheng, M., 2019. A meta-analysis of 1,119 manipulative experiments on terrestrial carbon-cycling responses to global change. Nat. Ecol. Evol. 3 (9), 1309–1320. https://doi.org/10.1038/s41559-019-0958-3.

Soussana, J.-F., Lüscher, A., 2007. Temperate grasslands and global atmospheric change: a review. Grass Forage Sci 62, 127–134. https://doi.org/10.1111/j.1365-2494.2007.00577.x.

Soussana, J.-F., Graux, A.I., Tubiello, F.N., 2010. Improving the use of modelling for projections of climate change impacts on crops and pastures. J. Exp. Bot. 61 (8), 2217–2228. https://doi.org/10.1093/jxb/erq100.

Soussana, J.-F., Barioni, L.G., Ben Ari, T., Conant, R., Gerber, P., Havlik, P., Ickowicz, A., Howden, M., 2013. Managing grassland systems in a changing climate: the search for practical solutions. In: Michalk, D.L., Millar, G.D., Badgery, W.B., Broadfoot, K.M. (Eds.), Proceedings of the 22nd International Grassland Congress, 15–19 September, Sydney, pp. 10–27.

ten Broeke, G., van Voorn, G., Ligtenberg, A., 2016. Which sensitivity analysis method should I use for my agent-based model? J. Artif. Soc. S. 19, 5. https://doi.org/10.18564/jasss.2857.

Teuling, A.J., Seneviratne, S.I., Stöckli, R., Reichstein, M., Moors, E., Ciais, P., Luyssaert, S., van den Hurk, B., Ammann, C., Bernhofer, C., Dellwik, E., Gianelle, D., Gielen, B., Grünwald, T., Klumpp, K., Montagnani, L., Moureaux, C., Sottocornola, M., Wohlfahrt, G., 2010. Contrasting response of European forest and grassland energy exchange to heatwaves. Nat. Geosci. 3, 722–727. https://doi.org/10.1038/ngeo950.

Thornley, J.H.M., 1972. A balanced quantitative model for root:shoot ratios in vegetative plants. Ann. Bot. 36, 431–441. https://doi.org/10.1093/oxfordjournals.aob.a084602.

Tonitto, C., Woodbury, P.B., McLellan, E.L., 2018. Defining a best practice methodology for modeling the environmental performance of agriculture. Environ. Sci. Policy 87, 64–73. https://doi.org/10.1016/j.envsci.2018.04.009.

Tubiello, F.N., Soussana, J.-F., Howden, S.M., 2007. Crop and pasture response to climate change. Proc. Natl. Acad. Sci. U.S.A. 104, 19686–19690. https://doi.org/10.1073/pnas.0701728104.

Unger, S., Jongen, M., 2015. Consequences of changing precipitation patterns for ecosystem functioning in grasslands: a review. Progress in Botany. In: Lüttge, U., Beyschlag, W. (Eds.), Progress in Botany. Springer, Cham. https://doi.org/10.1007/978-3-319-08807-5_14, 76.

van Oijen, M., Barcza, Z., Confalonieri, R., Korhonen, P., Kröel-Dulay, G., Lellei-Kovács, E., Louarn, G., Louault, F., Martin, R., Moulin, T., Movedi, E., Picon-Cochard, C., Rolinski, S., Viovy, N., Wirth, S.B., Bellocchi, G., 2020. Incorporating biodiversity into biogeochemistry models to improve prediction of ecosystem services in temperate grasslands. Rev. Roadmap. Agron. 10, 259. https://doi.org/10.3390/agronomy10020259.

Wallach, D., Martre, P., Liu, B., Asseng, S., Ewert, F., Thonburn, P.J., van Ittersum, M., Aggarwal, P.K., Ahmed, M., Basso, B., Biernath, C., Cammarano, D., Challinor, A.J., De Sanctis, G., Dumont, B., Rezaei, E.E., Fereres, E., Fitzgerald, G.J., Gao, Y., Garcia-Vila, M., Gayler, S., Girousse, C., Hoogenboom, G., Horan, H., Izaurralde, R.C., Jones, C.D., Kassie, B.T., Kersebaum, K.C., Klein, C., Koehler, A.-K., Maiorano, A., Minoli, S., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G.J., Palosuo, T., Priesack, E., Ripoche, D., Rötten, R.P., Semenov, M.A., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Fao, F., Wolf, J., Zhang, Z., 2018. Multi-model ensembles improve predictions of crop-environment-management interactions. Global Change Biol 24, 5072–5083. https://doi.org/10.1111/gcb.14411.

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, 1–13. https://doi.org/10.1007/s00442-011-2172-0.

Wang, N., Quesada, B., Xia, L., Butterbach-Bahl, K., Goodale, C.L., Kiese, R., 2019. Effects of climate warming on carbon fluxes in grasslands – A global meta-analysis. Glob. Change Biol. 25, 1839–1851. https://doi.org/10.1111/gcb.14603.

Warnant, P., François, L., Strivay, D., Gérard, J.-C., 1994. CARAIB: a global model of terrestrial biological productivity. Global Biogeochem. Cycles 8, 255–270. https://doi.org/10.1029/94GB00850.

Way, D.A., Yamori, W., 2014. Thermal acclimation of photosynthesis: on the importance of adjusting our definitions and accounting for thermal acclimation of respiration. Photosyn. Res. 119 (1–2), 89–100. https://doi.org/10.1007/s11120-013-9873-7.

White, J.W., Hoogenboom, G., Kimball, B.A., Wall, G.W., 2011. Methodologies for simulating impacts of climate change on crop production. Field Crop. Res. 124, 357–368. https://doi.org/10.1016/j.fcr.2011.07.001.

Wilcox, K.R., Shi, Z., Gherardi, L.A., Lemoine, N.P., Koerner, S.E., Hoover, D.L., Bork, E., Byrne, K.M., Cahill, J., Collins, S.L., Evans, S., Gilgen, A.K., Holub, P., Jiang, L., Knapp, A.K., LeCain, D., Liang, J., Garcia-Palacios, P., Peñuelas, J., Pockman, W.T., Smith, M.D., Sun, S., White, S.R., Yahdjian, L., Zhu, K., Luo, Y., 2017. Asymmetric responses of primary productivity to precipitation extremes: a synthesis of grassland precipitation manipulation experiments. Global Change Biol 23, 4376–4385. https://doi.org/10.1111/gcb.13706.

Williams, J.R., Arnold, J.G., Kiniry, J.R., Gassman, P.W., Green, C.H., 2008. History of model development at Temple. Texas. Hydrol. Sci. J. 53, 948–960. https://doi.org/10.1623/hysj.53.5.948.

Wohlfahrt, G., Anderson-Dunn, M., Bahn, M., Balzarolo, M., Berninger, F., Campbell, C., Carrara, A., Cescatti, A., Christensen, T., Dore, S., Eugster, W., Friborg, T., Furger, M., Gianelle, D., Gimeno, C., Hargreaves, K., Hari, P., Haslwanter, A., Johansson, T., Marcolla, B., Milford, C., Nagy, Z., Nemitz, E., Rogiers, N., Sanz, M.J., Siegwolf, R.T.W., Susiluoto, S., Sutton, M., Tuba, Z., Ugolini, F., Valentini, R., Zorer, R., Cernusca, A., 2008. Biotic, abiotic, and management controls on the net ecosystem CO_2 exchange of European mountain grassland ecosystems. Ecosystems 11, 1338–1351. https://doi.org/10.1007/s10021-008-9196-2.

Wu, Z., Dijkstra, P., Koch, G.W., Penuelas, J., Hungate, B.A., 2011. Responses of terrestrial ecosystems to temperature and precipitation change: a meta-analysis of experimental manipulation. Global Change Biol 17, 927–942. https://doi.org/10.1111/j.1365-2486.2010.02302.x.

Wu, W., Lan, C.-W., Lo, M.-H., Reager, J.T., Famiglietti, J.S., 2015. Increases in the annual range of soil water storage at northern middle and high latitudes under global warming. Geophys. Res. Lett. 42, 3903–3910. https://doi.org/10.1002/2015GL064110.

Wu, L., McGechan, M.B., McRoberts, N., Baddeley, J.A., Watson, C.A., 2007. SPACSYS: integration of a 3D root architecture component to carbon, nitrogen and water cycling – model description. Ecol. Modell. 3-4 343–359. https://doi.org/10.1016/j.ecolmodel.2006.08.010.

Wullschleger, S.D., 1993. Biochemical limitations to carbon assimilation in C3 plants – a retrospective analysis of the A/Ci curves from 109 species. J. Exp. Bot. 44, 907–920. https://doi.org/10.1093/jxb/44.5.907.

Xu, W., Yuan, W., Dong, W., Xia, J., Liu, D., Chen, Y., 2013. A meta-analysis of the response of soil moisture to experimental warming. Environ. Res. Lett. 8, 044027 https://doi.org/10.1088/1748-9326/8/4/044027.

Yahdjian, L., Sala, O.E., 2002. A rainout shelter design for intercepting different amounts of rainfall. Oecologia 133, 95–101. https://doi.org/10.1007/s00442-002-1024-3.