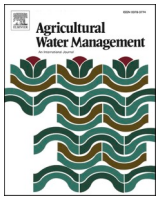
Agricultural Water Management 337 (2026) 110882



Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat



Uncertainties in simulating wheat growth responses to waterlogging

Rogério de S. Nóia-Júnior [a,*], Matthew Tom Harrison [b,*], Taru Palosuo [c], Margarita García-Vila [d], Ke Liu [b], Murilo dos Santos Vianna [e], Tobias KD Weber [f], Jin Zhao [g], Zhentao Zhang [g], Diego Abalos [h], Senthold Asseng [i], Juraj Balkovic [j], Attila Bussay [k], Xianguan Chen [l], Mathieu Delandmeter [m], Benjamin Dumont [m], Liping Feng [n], Roberto Ferrise [o], Christian Folberth [j], Davide Fumagalli [p], Surendra Giri [q], Aram Gorooei [r], Muhammad Habib-Ur-Rahman [s], Yean-Uk Kim [q,t], David Kraus [u], Huan Liu [h], Lioba Martin [u], Julien Morel [v], Claas Nendel [q,w], Gloria Padovan [o], Marlene Palka [q], Johannes Wilhelmus Maria Pullens [h], Reimund P. Rötter [s], Valentina Stocca [i], Maurits van den Berg [v], Heidi Webber [q,x], Jean-Charles Deswarte [y], Jean-Pierre Cohan [z], Alex C. Ruane [aa], Pierre Martre [a]

[a] *LEPSE, Univ Montpellier, INRAE, Institut Agro Montpellier, Montpellier, France*
[b] *Tasmanian Institute of Agriculture, University of Tasmania, Newnham, Launceston, Tasmania, Australia*
[c] *Natural Resources Institute Finland (Luke), Helsinki, Finland*
[d] *Instituto de Agricultura Sostenible, CSIC, Córdoba, Spain*
[e] *Institute of Bio, and Geosciences (IBG-3), Forschungszentrum Jülich, Jülich, Germany*
[f] *Section of Soil Science, University of Kassel, Witzenhausen, Germany*
[g] *College of Resources and Environmental Sciences, China Agricultural University, Beijing, China*
[h] *Department of Agroecology, Aarhus University, Tjele, Denmark*
[i] *Technical University of Munich, School of Life Sciences, Department of Life Science Engineering, Chair of Digital Agriculture, HEF World Agricultural Systems Center, Freising, Germany*
[j] *International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria*
[k] *Arcadia SIT, Milano, Italy*
[l] *College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, China*
[m] *Liege University, Gembloux Agro-Bio Tech, TERRA Teaching and Research Centre, Plant Sciences/Crop Science, Gembloux, Belgium*
[n] *College of Resources and Environmental Sciences, China Agricultural University, Beijing, China*
[o] *Department of Agriculture, Food, Environment, and Forestry, University of Florence, Firenze, Italy*
[p] *Fincons Group, Vimercate, Italy*
[q] *Leibniz Centre for Agricultural Landscape Research (ZALF), Müncheberg, Germany*
[r] *Institute of Crop Science and Resource Conservation (INRES), University of Bonn, Bonn, Germany*
[s] *Department of Crop Sciences, Tropical Plant Production and Agricultural Systems Modelling (TROPAGS), University of Goettingen, Goettingen, Germany*
[t] *Department of Landscape Architecture and Rural Systems Engineering, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea*
[u] *Institute of Meteorology and Climate Research – Atmospheric Environmental Research (IMKIFU), Karlsruhe Institute of Technology (KIT), Garmisch-Partenkirchen, Germany*
[v] *European Comission Joint Research Center (JRC), Ispra, Italy*
[w] *Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany*
[x] *Brandenburg University of Technology, Faculty of Environment and Natural Science, Cottbus, Germany*
[y] *ARVALIS, Villiers-le-Bâcle, France*
[z] *ARVALIS, Loireauxence, France*
[aa] *NASA Goddard Institute for Space Studies (GISS), New York, NY, USA*

ARTICLE INFO

Handling Editor - Xiaomin Mao

ABSTRACT

Waterlogging is an increasingly important stress in wheat growing regions, yet its representation in process based crop models (CMs) remains limited. We used a standardized AgMIP experiment to diagnose how 15 wheat CMs differ in their response to waterlogging timing, duration, and rainfall driven soil saturation. Before fitting

* Corresponding authors.
E-mail addresses: rogeriosouzanoia@gmail.com (R.S. Nóia-Júnior), matthew.harrison@utas.edu.au (M.T. Harrison).

https://doi.org/10.1016/j.agwat.2026.110882
Received 18 July 2026; Received in revised form 14 October 2026; Accepted 18 October 2026
Available online 22 October 2026
0378-3774/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords:
Process-based crop simulation models
Crop stressors
Flooding
Heavy rain
Multi-model analysis

waterlogging specific crop response coefficients to the eight waterlogged Lleida treatments in the current exercise, the ensemble median reproduced these treatments with an rRMSE of 17.0%. We interpret this as a pre fit extrapolative benchmark rather than strict independent validation because CMs retained their existing formulations and parameterizations and one CM reported prior calibration of soil water variables. Fitting waterlogging specific parameters to the same dataset reduced rRMSE to 6.3%, which is calibration goodness of fit rather than an independent test. Large differences among CMs remained when the calibrated parameters were transferred across phenological stages. Stem elongation was the single most frequently identified stage of maximum yield sensitivity, occurring in 4 of 15 CMs, although maxima occurred across all seven stages. Mean inter model variability in relative grain yield, expressed as the coefficient of variation across CMs, increased from 26.7% under 4 day waterlogging to 41.4% under 24 day waterlogging. A family balanced ensemble that first averaged related MONICA and SIMPLACE versions produced nearly identical mean responses ($r = 0.998$; maximum difference 2.8 %age points), indicating that related model versions did not drive the ensemble pattern. Most CMs showed weak yield responses to synthetic rainfall perturbations, including very large events, but these scenarios lack matched observations and therefore diagnose sensitivity rather than model bias. An exploratory structure response analysis found no single directly affected crop process that separated the severe stress response typologies. The strongest result is the large structural disagreement among current wheat CMs under standardized waterlogging perturbations, pointing to the combined importance of hydrological stress generation, crop process representation, parameterization, and development dependent damage and recovery for future model improvement.

1. Introduction

Soil waterlogging, periods of excessive soil moisture leading to full saturation of the root zone, is an increasingly widespread threat to global agricultural production (Liu et al., 2023). In 2023–2024, excess water was the leading cause among reported weather-related crop failures, linked to 40 of 130 reported events (Orla Dwyer, 2025). As the hydrological cycle intensifies under climate change, extreme rainfall is expected to become more frequent and severe in many regions (Fowler et al., 2021; Seneviratne et al., 2021), increasing waterlogging risks across cropping systems. Many reported cases refer to large-scale waterlogging, but small and localized events, often driven by landscape features, can also have substantial cumulative impacts on crop yields (Nóia-Júnior et al., 2025b).

Wheat, a staple crop that supplies about 20% of global calories (Shiferaw et al., 2013), is particularly vulnerable to soil waterlogging (Herzog et al., 2016; Pan et al., 2021). Excessive rainfall and poor drainage can deplete oxygen in the root zone, impair root respiration and nutrient uptake, alter hormone balance, and restrict growth (Daniel and Hartman, 2024; Manghwar et al., 2024). Severe stress can reduce root survival and canopy development while increasing susceptibility to disease and lodging (Nguyen et al., 2016; Pan et al., 2021; Shaw et al., 2013). Yield responses depend strongly on stress timing and duration. Stem elongation and the reproductive period have frequently been identified as sensitive phases in wheat (Marti et al., 2015; de San Celedonio et al., 2014). A recent global meta analysis of 217 studies further showed that flooding effects depend on crop type, growth stage, duration, water depth, and genotype, and highlighted substantial post flood compensatory recovery in wheat (Qiu et al., 2025). Temperature, soil properties, and genotype further modulate these responses (Liu et al., 2021, 2020a, 2020b; Yan et al., 2022).

Recent wheat crop failures, including flooding-induced losses in Brazil (Nóia-Júnior et al., 2025) and China (Jiang Mengnan, 2023) in 2023, and repeated yield declines across north-western Europe in 2016 (Ben-Ari et al., 2018; Nóia-Júnior et al., 2023) and 2024 (NIEMEYER Stefan and VANDENBERG Maurits, 2024), demonstrate that superfluous water can limit yields as much as drought or heat (Li et al., 2019; Nóia-Júnior et al., 2025b). In Europe, spring waterlogging during anthesis and grain filling has become one of the most critical threats to wheat productivity (Nóia-Júnior et al., 2023; Nóia-Júnior et al., 2025b; Zampieri et al., 2017). Wheat production stability is increasingly threatened by yield losses, planting delays, and limited field workability caused by waterlogged soils across major agricultural zones. Depending on terrain and soil properties, rainfall can cause both large-scale flooding and small-scale, patchy waterlogging lasting from days to months (Fu et al., 2023; Kulkarni et al., 2021; Liu et al., 2023; Nóia Júnior et al., 2021; Rötter et al., 2018a; Zampieri et al., 2017). Despite its global relevance, waterlogging remains underrepresented in crop–climate impact assessments using CMs (Kim et al., 2024; Nóia Júnior et al., 2023; Nóia-Júnior et al., 2025a). The vast majority of studies focus on a subset of climatic impact-drivers including drought, heat, or gradual warming, leaving excess water largely unexplored (Kim et al., 2025). While some agronomic studies have quantified waterlogging impacts and mitigation strategies (Chen et al., 2024; Daniel and Hartman, 2024; de San Celedonio et al., 2018; Marti et al., 2015; Zhou et al., 2018), most crop models still do not represent waterlogging adequately (Garcia-Vila et al., 2025). Contemporary waterlogging routines are still in their infancy (Garcia-Vila et al., 2025; Nóia Júnior et al., 2023), with critical deficiencies in simulating capillary rise, crop tolerance and recovery, and impacts on phenology, nitrogen cycling, and yield components (Garcia-Vila et al., 2025). Consequently, uncertainties remain high in simulating crop responses under waterlogged conditions (Nóia-Júnior et al., 2025a).

Crop modeling studies have traditionally emphasized water limited systems (Asseng et al., 2013; Asseng et al., 2015; Jägermeyr et al., 2021; Martre et al., 2015; Pequeno et al., 2021; Webber et al., 2017; Webber et al., 2025). Processes associated with water excess, including oxygen limitation, nitrogen loss under anaerobic conditions, impaired root growth, and post stress recovery, are therefore often simplified or absent. Garcia-Vila et al. (2025) provided a multi model inventory of these processes and identified recurring gaps in capillary rise, crop recovery, waterlogging resistance, soil nitrogen, phenology, and yield components. That work established which processes are represented across CMs. It did not quantify how the same CMs diverge when exposed to identical waterlogging timing, duration, and severity, nor did it test whether waterlogging parameterizations calibrated mainly during stem elongation transfer to other phenological stages. This quantitative diagnosis is the specific gap addressed here.

Two linked sources of uncertainty limit the use of CMs for waterlogging assessment. The first is the representation of crop response to oxygen limited soil conditions. The second is the representation of the hydrological conditions that generate and terminate waterlogging. Most CMs omit landscape position and lateral water flow, which restricts their ability to represent spatially heterogeneous ponding and perched water tables (Garcia-Vila et al., 2025). Recent work has proposed both empirical and mechanistic waterlogging routines (Garcia-Vila et al., 2025; Githui et al., 2022; Kim et al., 2024; Liu et al., 2020b; Rötter et al., 2018a), but their consequences have not been quantified under a common, coordinated perturbation framework. A standardized intercomparison can separate disagreement caused by model formulation from

disagreement caused by differing weather, soil, or management inputs.

We therefore coordinated a global intercomparison within the Agricultural Model Intercomparison and Improvement Project (AgMIP) and the AgMIP Wheat team (Rosenzweig et al., 2013; Asseng et al., 2015, 2013; Garcia-Vila et al., 2025; Martre et al., 2024, 2015). Fifteen wheat CMs were subjected to standardized waterlogging treatments that varied in phenological timing, duration, and rainfall driven soil saturation. Existing evidence for stage dependent sensitivity and root mediated damage comes primarily from experiments rather than coordinated CM intercomparisons. We therefore used these empirical patterns as biological expectations against which CM behavior could be tested, rather than as established properties of model ensembles. We tested three hypotheses. First, based on experimental evidence that waterlogging sensitivity varies with developmental timing, we hypothesized that simulated yield responses would be stage dependent rather than invariant across phenology, with stem elongation and reproductive development representing empirically motivated candidate periods of high sensitivity. Second, within the tested range of 4–24 days, we hypothesized that disagreement among CMs would increase with waterlogging duration. Third, translating the prominence of root hypoxia and root damage in the experimental literature into a testable model structure hypothesis, we expected CMs that link waterlogging directly to root growth or root length density to simulate larger severe stress yield losses than CMs without such a linkage. The objective was to diagnose the conditions and model structures associated with divergence, rather than to rank CMs for universal predictive accuracy.

2. Material and methods

This study was conducted within AgMIP Wheat Phase 4 to diagnose how CM structure, stress timing, stress duration, and hydrological setup influence simulated wheat responses to waterlogging. All teams received common weather, soil, crop, and management inputs and a standardized treatment protocol. Participation was open to wheat modeling groups responding through the AgMIP network and meetings. The design intentionally combines a controlled experimental benchmark with synthetic perturbations so that differences among simulations can be interpreted primarily as differences in CM behavior under the prescribed conditions.

2.1. Simulation protocol

The full protocol distributed to all participating teams is provided in the Supplementary Material. It comprised three sequential steps: (i) calibration of cultivar parameters under non waterlogged conditions followed by calibration of waterlogging specific parameters against the controlled Lleida dataset, (ii) a synthetic sensitivity test in which heavy rainfall generated excess soil water in contrasting soils, and (iii) a transferability test in which the waterlogging parameterization from Step 1 was applied without further crop parameter adjustment across seven phenological stages and six stress durations.

i. In Step 1, CMs were parameterized for the winter wheat cultivar Soissons using four field experiments from Arvalis in France (Nóia-Júnior et al., 2023) and the outdoor controlled experiment of Marti et al. (2015) in Lleida, Spain. Step 1a calibrated phenology, growth, and yield using the Arvalis trials and the non waterlogged Lleida control, without fitting waterlogging specific crop response coefficients to the eight waterlogged Lleida treatments in the current exercise. The resulting parameterization was then used to simulate those eight treatments. We therefore use Step 1a as a pre fit extrapolative benchmark of waterlogging response, rather than strict independent validation, because CMs retained existing formulations and parameterizations and DSSAT Nwheat reported that soil water related variables had been calibrated previously for an earlier waterlogging study. Step 1b additionally fitted waterlogging specific parameters using the full Lleida dataset, which comprised the control and eight waterlogged treatments. These Step 1b comparisons are therefore calibration goodness of fit and are not an independent validation. Nitrogen limitations were avoided, either through fertilization or by disabling nitrogen stress, to isolate the direct crop response to excess soil water.
ii. Step 2 assessed CM sensitivity to rainfall driven soil saturation. Twelve scenarios combined two soils (sandy and clayey) with six rainfall perturbations around anthesis: baseline rainfall, 50 mm added over 3 days, 150 mm over 6 days, 300 mm over 9 days, 500 mm over 12 days, and 1000 mm over 12 days. Added rainfall was distributed uniformly over each period. These scenarios were designed as a non probabilistic stress test, not as estimates of event frequency. The upper perturbations are deliberately far outside the local climatological range. For context, the 1983–2010 normal annual precipitation at the AEMET Lleida station is about 342 mm, so 500 and 1000 mm correspond to approximately 1.5 and 2.9 times mean annual precipitation, respectively (AEMET, 2026). Over a 12 day treatment, these additions are far larger than plausible crop transpiration and must be partitioned among changes in soil storage, evapotranspiration, drainage, and runoff. Harmonized water fate outputs were not available across the ensemble, so we do not quantify that partitioning. The upper treatments are therefore interpreted as joint tests of CM hydrological boundary behavior and crop sensitivity under progressively stronger forcing.
iii. Step 3 tested transferability of the calibrated waterlogging response across crop development. Forty two treatments combined six durations (4, 8, 12, 16, 20, and 24 days) with seven starting stages: seedling emergence, two leaf, three leaf, onset of tillering, onset of stem elongation, anthesis, and onset of grain filling. Crop parameters and waterlogging specific coefficients fitted in Step 1b were held fixed. Soil, weather, and management followed the Lleida setup. The common protocol specified strongly restricted bottom layer drainage and irrigation designed to maintain the prescribed waterlogged condition, while teams could make model specific hydrological adjustments where required to achieve saturation. Because the waterlogging coefficients were fitted mainly to stem elongation treatments, Step 3 is an extrapolation and transferability test across stages rather than independent validation.

For Steps 2 and 3, all simulations used Soissons with the Step 1b waterlogging parameterization. Teams received standardized weather, soil, management, and treatment files and initialized simulations six months before sowing with soil at field capacity. Additional abiotic stresses were minimized, except for the observed field conditions retained in the Arvalis trials. Teams documented model configuration and any hydrological adjustments in standardized ModelSetup templates. These records show that adjustments differed among CMs, for example through bottom layer hydraulic conductivity, impermeable lower boundaries, aeration thresholds, or model specific drainage settings. We use these differences as part of the diagnostic interpretation, but they cannot be separated completely from crop stress routines because common intermediate hydrological outputs were not available for all CMs and scenarios.

2.2. Experimental data for calibration and benchmarking

The principal waterlogging dataset was the outdoor controlled experiment of Marti et al. (2015) at the University of Lleida, Spain (41°37′47″ N, 0°35′47″ E). Winter wheat cv. Soissons was sown on 13 December 2011 and grown outdoors in vertical soil columns 84 mm in diameter and 1.25 m deep. The columns were filled with a loamy sand soil containing 80.8% sand, 15.3% silt, 3.9% clay, 0.25% organic

matter, and a pH of approximately 8.0. The experiment included nine treatments arranged in a randomized complete block design with three replicates and 54 single plant columns per treatment. The main treatments comprised a well drained control and waterlogging durations of 4, 8, 12, 16, 20, and 24 days, imposed at staggered starting dates so that they ended together approximately three days after anthesis. Two complementary treatments imposed 8 or 16 days of waterlogging beginning 10 days after terminal spikelet initiation to test timing within stem elongation. The long durations were deliberately included in the original experiment to quantify the response gradient during stem elongation. The controlled column design isolates waterlogging from drought, nutrient limitation, pest, and disease effects, which makes it useful for process benchmarking, but its single cultivar, single environment, and coarse textured soil restrict inference to field conditions. We therefore use these observations for calibration and benchmarking, not as an independent field validation of the 15 CMs.

Four additional Arvalis field experiments were used in Step 1a to support calibration of Soissons phenology, growth, and yield under non waterlogged or naturally managed field conditions. These trials spanned different years and locations and followed standard agronomic practices. They broaden the cultivar calibration environment, but they do not provide an independent waterlogging validation because waterlogging specific parameters were fitted to the Lleida experiment. All available input and observed data are provided in the Supplementary Material and project archive.

2.3. Process-based crop models and waterlogging response groups

Fifteen CMs participated in the intercomparison (Table 1). They differ in soil water representation, the variable used to diagnose aeration stress, the crop processes modified under excess water, and the persistence of stress after drainage. Table 1 summarizes versions, repositories, direct crop processes affected by waterlogging, and response groups. To make the process basis more explicit, the model specific functional representations and hydrological adjustments reported by participating teams are described below.

Because grain yield output at seedling emergence was unavailable for one CM, the primary clustering used the six phenological stages with complete output from all 15 CMs, from the two leaf stage through onset of grain filling. CMs were grouped using k means clustering of simulated relative grain yield under the 24 day waterlogging treatment, which was treated as a severe stress diagnostic rather than a typical field duration. Cluster number was evaluated for $k = 2–6$ using within cluster sum of squares and mean silhouette width. Robustness of the selected four cluster solution was independently assessed with Ward hierarchical clustering and the adjusted Rand index. Cluster stability was further evaluated by repeating the four cluster analysis at 4, 8, 12, 16, and 20 days and comparing each classification with the 24 day solution. A sensitivity analysis using all seven phenological stages was also conducted for the 14 CMs with complete seven stage grain yield output. We interpret the resulting groups as severe stress response typologies rather than universal CM classes.

2.4. Model specific waterlogging formulations

The participating CMs do not share a single waterlogging production function. We therefore distinguish between the state variable used to diagnose excess water or aeration stress, the functional form or named stress factor, the crop process modified, and the study specific hydrological setup. Supplementary Table S1 summarizes these elements for all 15 CMs and separates model documentation from parameter choices used in this intercomparison. For APSIM Classic, published documentation defines aeration stress from soil water relative to drained upper limit and saturation and applies crop and stage dependent modifiers to root and growth processes (Githui et al., 2022). AquaCrop calibrated an aeration stress parameter acting through transpiration and dry matter

Table 1
Process-based crop models used in the intercomparison study. Table includes CM versions, code repository access links, processes directly affected by waterlogging, and cluster groups based on k-means classification described in the methods.

Process-based crop simulation models	Version	Repository / Website	Processes directly affected by waterlogging	Group
APSIM	v7.9	https://www.apsim.info/download-apsim/downloads/3-may-2017-apsim-7–9-release-notes/	Root growth, radiation use efficiency, biomass production, nitrogen accumulation in organs, grain weight, dry matter partitioning, phenology and grain yield	2
AquaCrop	v7.1	https://www.fao.org/aquacrop/en/	Transpiration and dry matter partitioning	4
Daisy	6.45	https://daisy.ku.dk/download/windows/	Root growth	1
DSSAT-CERES-Wheat	v4.8.2	https://dssat.net/DSSAT-CERES-Wheat/v.4.8.2	Root growth and root length density	1
DSSAT-Nwheat	v4.7	https://github.com/DSSAT	Root growth, biomass production and potential grain size	4
EPIC	v.0810	https://epicapex.tamu.edu/	Biomass production and leaf area index	4
WOFOST	v8.1	https://www.wur.nl/en/Research-Results/Research-Institutes/Environmental-Research/Facilities-Tools/Software-models-and-databases/WOFOST.htm	Transpiration and photosynthesis	3
LandscapeDNDC	1.36	https://ldndc.imk-ifu.kit.edu	Root growth, root length density and photosynthesis	4
MONICA v1	3.6.22	https://github.com/zalf-rpm/monica	Transpiration, dry matter partitioning	4
MONICA v2	3.6.22	https://github.com/zalf-rpm/monica	Transpiration, dry matter partitioning and photosynthesis	2
SIMPLACE v1 (<LINTUL5, Hillflow1D>)	v5.0	https://www.simplace.net	Transpiration	3
SIMPLACE v2 (<LINTUL5, SlimWL>)	v5.0	https://www.simplace.net	Transpiration, root growth and root length density	2
SSM-Wheat	iCrop	http://ssm-crop-models.net/ssm-icrop/	Leaf area index, radiation use efficiency and nitrogen accumulation in organs	2
STICS	v10	https://stics.inrae.fr/	Transpiration, root growth and root length density, leaf area index and radiation use efficiency	2

(continued on next page)

Table 1 (*continued*)

Process-based crop simulation models	Version	Repository / Website	Processes directly affected by waterlogging	Group
WheatSM	v5.1	https://pan.baidu.com/s/1ozi3U9QU-L0UrkZov63WQw?pwd=nzlj	Transpiration, biomass production and photosynthesis	4

partitioning. Daisy calibrated PenWaterFac, a water dependent factor multiplying PenPar1 as a function of relative water content (Theta/-Theta_sat). In DSSAT CERES, saturated hydraulic conductivity constrains layer drainage and excess water affects root length density and root water uptake; SLLL, SDUL, SSAT, SRGF, and SSKS were evaluated during waterlogging calibration. DSSAT Nwheat used lower saturated hydraulic conductivity in the bottom soil layer to generate waterlogging and directly links stress to root growth, biomass production, and potential grain size.

EPIC used a critical aeration factor of 0.85 within its limiting stress framework for biomass production and leaf area, together with a low conductivity bottom layer that permits water accumulation when downward routing is constrained. WOFOST uses a standard oxygen stress formulation in which soil moisture above a critical aeration threshold and the number of consecutive oxygen stress days reduce transpiration; maximum duration effects are reached after four successive stress days. LandscapeDNDC scaled saturated hydraulic conductivity in the bottom layer, while its reported crop pathways include root growth, root length density, and photosynthesis. MONICA v1 calibrated reduced StomataConductance and dry matter partitioning among organs in response to waterlogging, whereas MONICA v2 calibrated CriticalOxygenContent and TimeUnderAnoxiaThreshold. Both MONICA versions used a hydraulic conductivity of 0.003 in the last two soil layers for the prescribed waterlogging treatments.

SIMPLACE v1 couples LINTUL5 with HillFlow1D. HillFlow1D uses a Feddes type pressure head response in which excessively wet soil reduces root water uptake and transpiration between the wetness thresholds HLIM1 and HLIM2; the resulting actual to potential transpiration ratio is propagated to LINTUL growth through the TRANRF stress factor. SIMPLACE v2 calibrated HLIM1 in its SlimWL implementation and used an impermeable bottom layer during waterlogging treatments. SSM Wheat calibrated its transpiration efficiency coefficient and maximum flooding duration, and STICS calibrated sensanox for anoxia sensitivity. WheatSM calculates actual root water uptake as potential crop transpiration multiplied by a water stress factor and directly links waterlogging effects to transpiration, biomass production, and photosynthesis. These formulations and their documentary sources are reported in Supplementary Table S1.

Across the participating CMs, documented stress drivers and setup choices span soil water relative to saturation, air filled porosity, oxygen or anoxia thresholds, pressure head, hydraulic conductivity, and water dependent stress functions, while affected crop processes include root growth, transpiration, photosynthesis, biomass production, radiation use efficiency, partitioning, leaf area index, nitrogen accumulation, phenology, and yield components (Table 1; Supplementary Table S1). Consequently, divergence can arise from the hydrological conditions that generate stress, from the crop process modified after stress is triggered, or from parameter calibration. Harmonized intermediate variables such as soil oxygen status, water table depth, drainage, and runoff were not available across all CMs. The archived daily outputs contained actual evapotranspiration, runoff, and deep percolation for MONICA v2 and actual evapotranspiration for SIMPLACE v2, but not a common water fate diagnostic for the full ensemble. We therefore do not attempt a formal attribution of Step 2 differences to individual hydrological pathways.

2.5. Data analysis

All analyses were conducted in R version 4.3.2 within RStudio (Team, 2020), with additional diagnostic checks reproduced from the archived simulation outputs. In Step 1, relative root mean squared error (rRMSE), mean absolute error (MAE), and mean bias were calculated between the multi model ensemble median and observed grain yield. Because Step 1b observations were used for fitting waterlogging parameters, these statistics quantify calibration goodness of fit rather than independent predictive performance. Inter model spread was treated as structural ensemble variability rather than sampling uncertainty. Fig. 1 therefore displays individual CM simulations together with the ensemble median and interquartile range (IQR). For the subsequent sensitivity analyses, ensemble means and coefficients of variation were retained to characterize average response and structural disagreement across CMs. For a structural dependence sensitivity test, MONICA v1 and v2 were first averaged within a MONICA family and SIMPLACE v1 and v2 within a SIMPLACE family, after which all 13 model families received equal weight.

Maximum sensitivity was defined for each output as the phenological stage containing the lowest simulated value relative to the control across all stage and duration combinations. The denominator was the number of CMs that reported the relevant variable: 15 for grain yield and above ground biomass, 14 for LAI, and 12 for root biomass. A CM that reported a variable but simulated no reduction relative to the control was retained in the denominator but was not assigned a maximum sensitivity stage; exact tied minima across stages were retained. To explore associations between CM structure and severe stress response, we coded only processes explicitly identified in Table 1 as directly affected by waterlogging. For each CM, severe stress response was defined as the mean grain yield loss under 24 day waterlogging across the six phenological stages with complete grain yield output for all 15 CMs. CMs with and without direct effects on root growth or root length density, transpiration, photosynthesis, biomass production or radiation use efficiency, dry matter partitioning, leaf area index, nitrogen accumulation, phenology, and yield components were compared descriptively using medians and interquartile ranges. The number of CMs contributing to each contrast is reported alongside the medians, because several categories contain fewer than five CMs and a median drawn from so few models carries little weight on its own. These contrasts were not interpreted causally because the process categories overlap, some contain few CMs, and MONICA and SIMPLACE include related versions. Root mortality and recovery were not coded as structural attributes because Table 1 does not identify them explicitly for the participating CMs.

3. Results

3.1. Waterlogging response before and after waterlogging specific calibration

Observed grain yield declined approximately linearly as waterlogging duration increased. Before fitting waterlogging specific crop response coefficients to the eight waterlogged Lleida treatments in the current exercise, the ensemble reproduced the overall direction of the response but showed substantial spread and systematic overestimation. For these eight treatments, excluding the well drained control used during Step 1a parameterization, the ensemble median had an rRMSE of 17.0%, MAE of 0.94 t ha−1, and mean bias of + 0.80 t ha−1. After waterlogging specific calibration, the corresponding values were 6.3%, 0.34 t ha−1, and + 0.12 t ha−1. The largest remaining discrepancies occurred in the longest pre anthesis treatments. Under 24 days of waterlogging, observed grain yield was approximately 4.35 t ha−1, compared with ensemble medians of about 5.54 t ha−1 before and 4.87 t ha−1 after waterlogging parameter calibration. The individual CM simulations and IQR show that substantial structural spread remained even after calibration, particularly in the longest treatments.

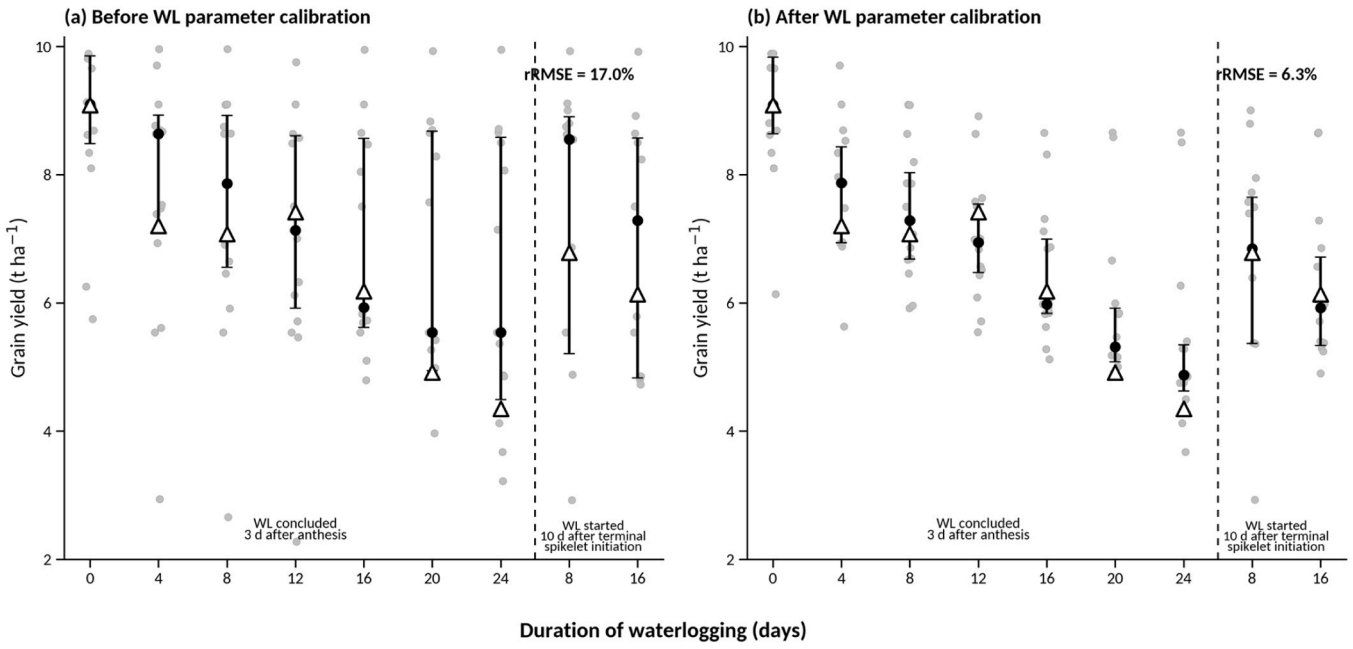


Fig. 1. Simulated impacts of waterlogging on grain yield before and after calibration of waterlogging related parameters. Panels show simulations (a) before and (b) after calibration. Gray points represent individual CM simulations, black circles represent the multi model ensemble median, vertical bars show the interquartile range across CMs, and open triangles represent observed grain yield. rRMSE, MAE, and bias reported in the text were calculated on the eight waterlogged treatments only; the well drained control is plotted for context but excluded because it contributed to Step 1a parameterization. Individual simulations were available from 15 CMs, except for the control treatment after calibration, for which 14 CMs provided output. Observed data are from the outdoor controlled experiment in which the six main waterlogging treatments lasted 4–24 days and were staggered to end approximately three days after anthesis (Supplementary Fig. S1). The two treatments beyond the vertical dashed line involved 8 and 16 days of waterlogging beginning 10 days after terminal spikelet initiation (Marti et al., 2015).

The Step 1a statistics provide a pre fit extrapolative benchmark with respect to fitting the current Lleida waterlogging treatments, whereas the Step 1b statistics quantify calibration goodness of fit. Supplementary Table S2 reports the corresponding metrics using both ensemble mean and median, with and without the control, to make the effect of these choices transparent.

3.2. *Sensitivity in simulating waterlogging impacts*

After calibration with the Lleida waterlogging data, CMs showed markedly different responses when the same parameterizations were transferred across phenological stages and durations (Fig. 2). At 24 days, $k = 4$ produced the highest mean silhouette width among the tested solutions (0.423), compared with 0.405, 0.343, 0.405, and 0.340 for $k = 2$, 3, 5, and 6, respectively. These values lie close together, and a mean silhouette width near 0.42 indicates moderate rather than strong separation, so the internal criterion alone does not identify four groups decisively. Agreement between k means and Ward clustering applied to the same fifteen CMs is likewise a weaker form of corroboration than resampling would provide, although with fifteen CMs a bootstrap would itself be poorly determined. Ward hierarchical clustering reproduced the robust four group solution exactly (adjusted Rand index, ARI = 1.0). Relative to the original Fig. 2 grouping, 14 of 15 CMs retained the same membership; AquaCrop was reassigned from Group 1 to Group 4, giving an ARI of 0.819 between the original and robust partitions. Agreement with the 24 day grouping increased with waterlogging duration, with ARI values of 0.037, 0.318, 0.416, 0.433, and 0.612 for 4, 8, 12, 16, and 20 days, respectively. The sensitivity analysis using all seven stages for the 14 CMs with complete data reproduced the primary grouping exactly (ARI = 1.0; mean silhouette width = 0.423). At 4 days, however, agreement with the 24 day partition was no better than chance, so the four group structure is not recovered at the shorter durations most often encountered in the field. The groups are therefore a typology of severe stress behaviour and should not be read as a general classification of these CMs (Supplementary Fig. S6).

The robust Group 1 contained Daisy and DSSAT CERES and showed little mean yield response through stem elongation under the 24 day treatment, followed by mean losses of about 21% at anthesis and 16% at onset of grain filling. Group 2 (APSIM, MONICA v2, SIMPLACE v2, SSM Wheat, STICS) showed strong early and mid season sensitivity, with mean losses of about 49–78% from the two leaf stage through stem elongation and smaller mean losses at anthesis and grain filling. Group 3 (WOFOST, SIMPLACE v1) had little mean response at emergence but showed its strongest response around tillering and the reproductive period, including mean losses of about 83% at onset of tillering and 74% at anthesis. Group 4 (AquaCrop, DSSAT Nwheat, EPIC, LandscapeDNDC, MONICA v1, WheatSM) showed a more moderate response profile, with the largest mean loss at stem elongation (about 48%) and mean losses of about 35–38% during tillering, anthesis, and grain filling. These profiles summarize group behavior and do not imply that all CMs within a group share the same internal mechanism.

MONICA v1 (Fig. 2i), although classified in Group 4, showed a distinct response with relatively large early losses and limited additional decline across later stages. Similar response diversity occurred for above ground biomass (Supplementary Fig. S3). DSSAT CERES simulated increased biomass under some waterlogging treatments while yield declined, whereas SIMPLACE v1, LandscapeDNDC, and WheatSM showed comparatively smaller biomass responses during yield formation despite substantial yield penalties. These contrasts indicate that similar yield responses can arise from different internal pathways and that cluster membership alone should not be interpreted as evidence of a specific causal mechanism.

The exploratory structure response analysis likewise did not identify a single process representation that separated the severe stress response typologies. CMs with a direct transpiration linkage (n = 8) had a median severe stress yield loss of 49.5% (IQR 39.3–55.6%), compared with 28.2% (IQR 14.4–42.0%) among CMs without a direct transpiration linkage (n = 7). In contrast, CMs with direct effects on root growth or

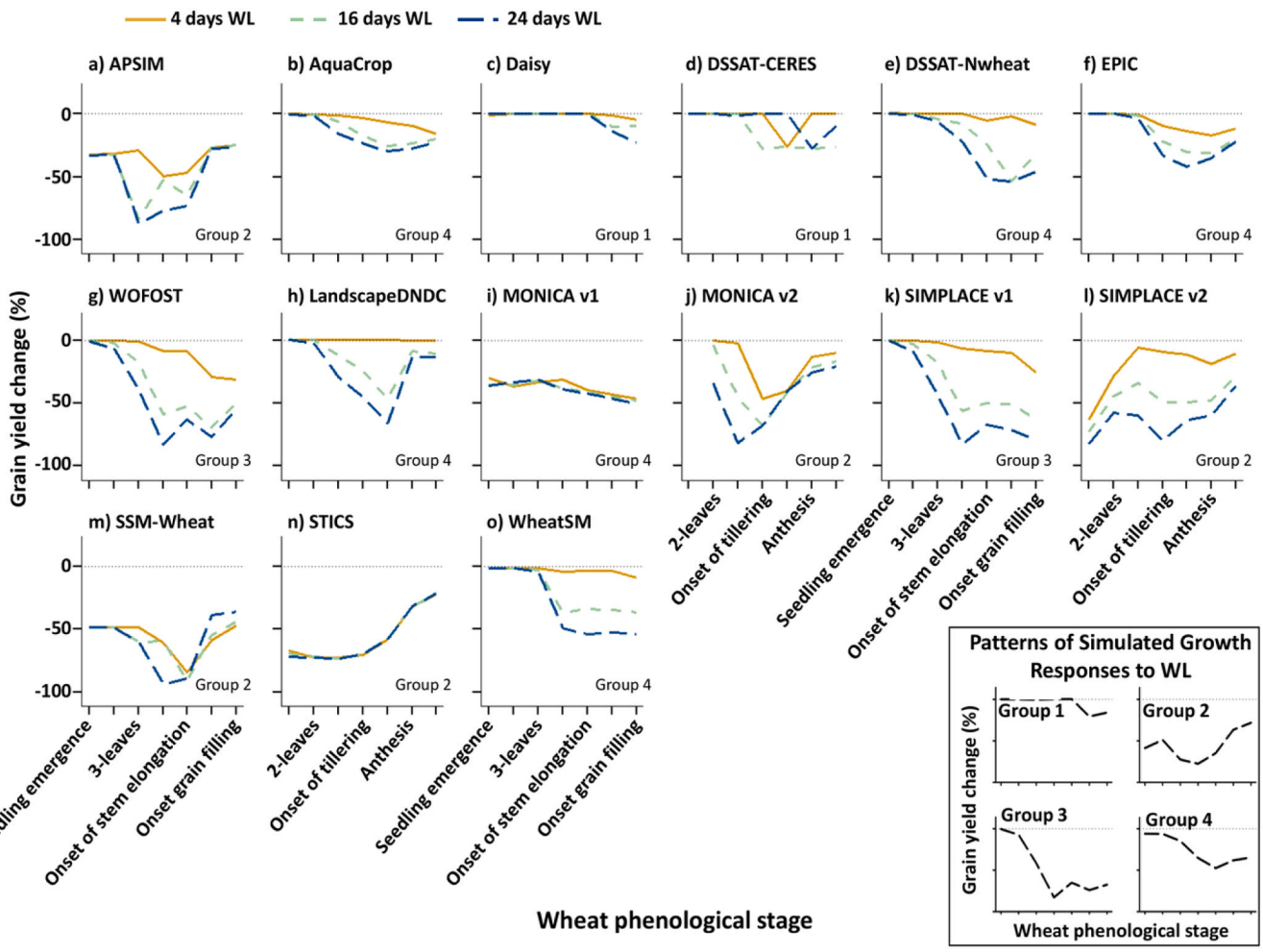


Fig. 2. Sensitivity of CMs to waterlogging across phenological stages and waterlogging duration. Simulated relative yield deviation under varying waterlogging durations (4, 16 and 24 days) applied at seven phenological stages, from seedling emergence to the onset of grain filling. Each line in each panel represents a waterlogging duration across different phenological stages. Horizontal dotted line in each panel indicates the yield of the control treatment (with no waterlogging). Details of CM versions and structural features are provided in Table 1. Group labels correspond to the robust k = 4 classification based on the 24 day relative yield response across the six phenological stages with complete output from all 15 CMs; AquaCrop is assigned to Group 4 in the robust solution. The inset shows arithmetic mean 24 day relative yield responses across all seven displayed stages for CMs within each robust group. At seedling emergence, the Group 2 mean is based on the available CMs because MONICA v2 did not report grain yield for that treatment.

root length density (n = 7) had a median loss of 30.0% (IQR 17.2–54.2%), compared with 42.7% (IQR 32.4–55.3%) among CMs without a direct root linkage (n = 8). Photosynthesis showed almost no median contrast: 40.3% (IQR 33.8–47.2%; n = 4) with a direct linkage versus 40.6% (IQR 21.3–56.7%; n = 11) without. The direction of the root contrast is opposite to the third hypothesis, and the broad overlap in IQRs reinforces that these are exploratory associations rather than evidence for a single causal mechanism. Paired versions also showed modest differences in the CM level mean severe stress loss: MONICA v2 exceeded MONICA v1 by 4.4 %age points (44.9% versus 40.6%), while SIMPLACE v2 exceeded SIMPLACE v1 by 0.5 %age points (59.4% versus 58.9%). Because categories overlap and model versions are not statistically independent, these comparisons are diagnostic associations rather than causal effects (Supplementary Fig. S7).

We identified, for each CM and variable, the phenological stage causing the greatest simulated loss relative to the control across all stage duration combinations. The denominator for each variable was the number of CMs reporting that variable (15 for above ground biomass and grain yield, 14 for LAI, and 12 for root biomass). A CM was assigned a stage of maximum sensitivity only when it simulated an actual reduction relative to the control; CMs that reported a variable but simulated no reduction in any waterlogging treatment remained in the denominator but were not assigned to a stage. When a CM had exactly equal maxima at more than one stage, each tied stage was retained, so percentages can exceed 100% when summed across stages. For above ground biomass, maximum sensitivity occurred at onset of tillering, stem elongation, and onset of grain filling in 20.0% of CMs each, with smaller frequencies at emergence, the two leaf and three leaf stages, and anthesis (Fig. 3a). For grain yield, stem elongation was the single most frequent maximum sensitivity stage (4 of 15 CMs; 26.7%), followed by the three leaf stage and onset of tillering (3 of 15 each; 20.0%), onset of grain filling (2 of 15; 13.3%), and emergence, the two leaf stage, and anthesis (1 of 15 each; 6.7%) (Fig. 3b). No stage represented the maximum yield sensitivity for a majority of CMs.

For LAI, the highest frequency of maximum sensitivity occurred at onset of tillering (7 of 14 CMs; 50.0%), followed by stem elongation (3 of 14; 21.4%). The two leaf, three leaf, and anthesis stages each accounted for 2 of 14 CMs (14.3%), and onset of grain filling accounted for 1 of 14 (7.1%) (Fig. 3c). EPIC reported LAI but simulated no reduction relative to the control across the waterlogging treatments and was therefore retained in the denominator without being assigned a maximum sensitivity stage. DSSAT CERES showed an exact tied minimum across five stages, so these tied stages were retained. For root biomass, onset of tillering was the most frequent maximum (3 of 12 CMs; 25.0%), while

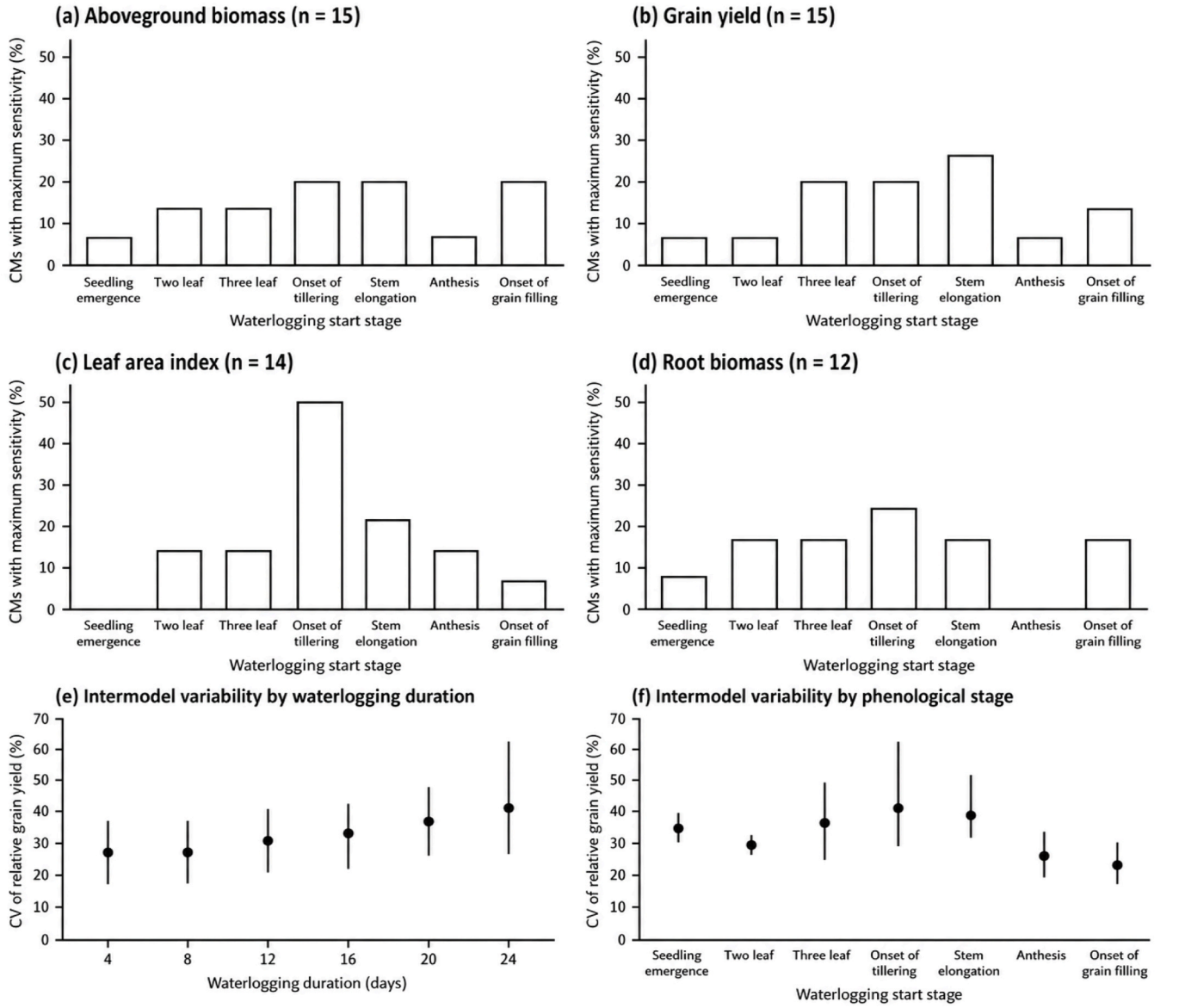


Fig. 3. Phenological stages associated with maximum simulated sensitivity to waterlogging across CMs. Bars show the percentage of CMs exhibiting their maximum simulated reduction relative to the control at each phenological stage for (a) above ground biomass, (b) grain yield, (c) leaf area index, and (d) root biomass. Percentages were calculated using the number of CMs providing each respective variable as the denominator, indicated in each panel. Models that provided a variable but simulated no reduction relative to the control across any waterlogging treatment were retained in the denominator but were not assigned a stage of maximum sensitivity. When a CM showed equal maximum sensitivity at more than one phenological stage, all tied stages were counted, and percentages may therefore sum to more than 100%. Panels (e) and (f) show inter model variability in relative grain yield according to waterlogging duration and phenological stage, respectively. The coefficient of variation was first calculated across CMs for each exact stage by duration treatment; points show the mean treatment level CV within each duration or stage, and vertical lines indicate the minimum to maximum range.

the two leaf, three leaf, stem elongation, and onset of grain filling stages each accounted for 2 of 12 CMs (16.7%) (Fig. 3d). As shown in Table 1, not all CMs report root biomass or LAI, and these variables can also respond indirectly through processes triggered by waterlogging stress.

Inter model divergence increased as stress duration increased. For grain yield, the mean coefficient of variation across stage specific treatments increased from 26.7% under 4 days of waterlogging to 41.4% under 24 days (Fig. 3e). Across phenological stages, mean inter model variability was greatest when waterlogging began at onset of tillering (40.8%) and stem elongation (38.1%), followed by the three leaf stage (35.3%) (Fig. 3f). This pattern supports the hypothesis that structural disagreement becomes more pronounced as the perturbation becomes more severe. The stage of maximum sensitivity itself was not concentrated in a majority of CMs, which further indicates that transfer of a waterlogging parameterization across phenology remains a major source of uncertainty.

3.3. *Multi-model ensemble for simulating waterlogging impacts*

The equal weighted multi model mean showed progressively larger losses in LAI, above ground biomass, root biomass, and grain yield as waterlogging duration increased (Fig. 4). The strongest mean reductions generally occurred when stress began between tillering and stem elongation, while yield remained sensitive into anthesis and grain filling. The ensemble derived slope corresponded to approximately 1.8 %age points of additional yield loss per waterlogged day when stress began at tillering or stem elongation, compared with about 0.5 %age points per day during seedling emergence to the two leaf stage (Supplementary

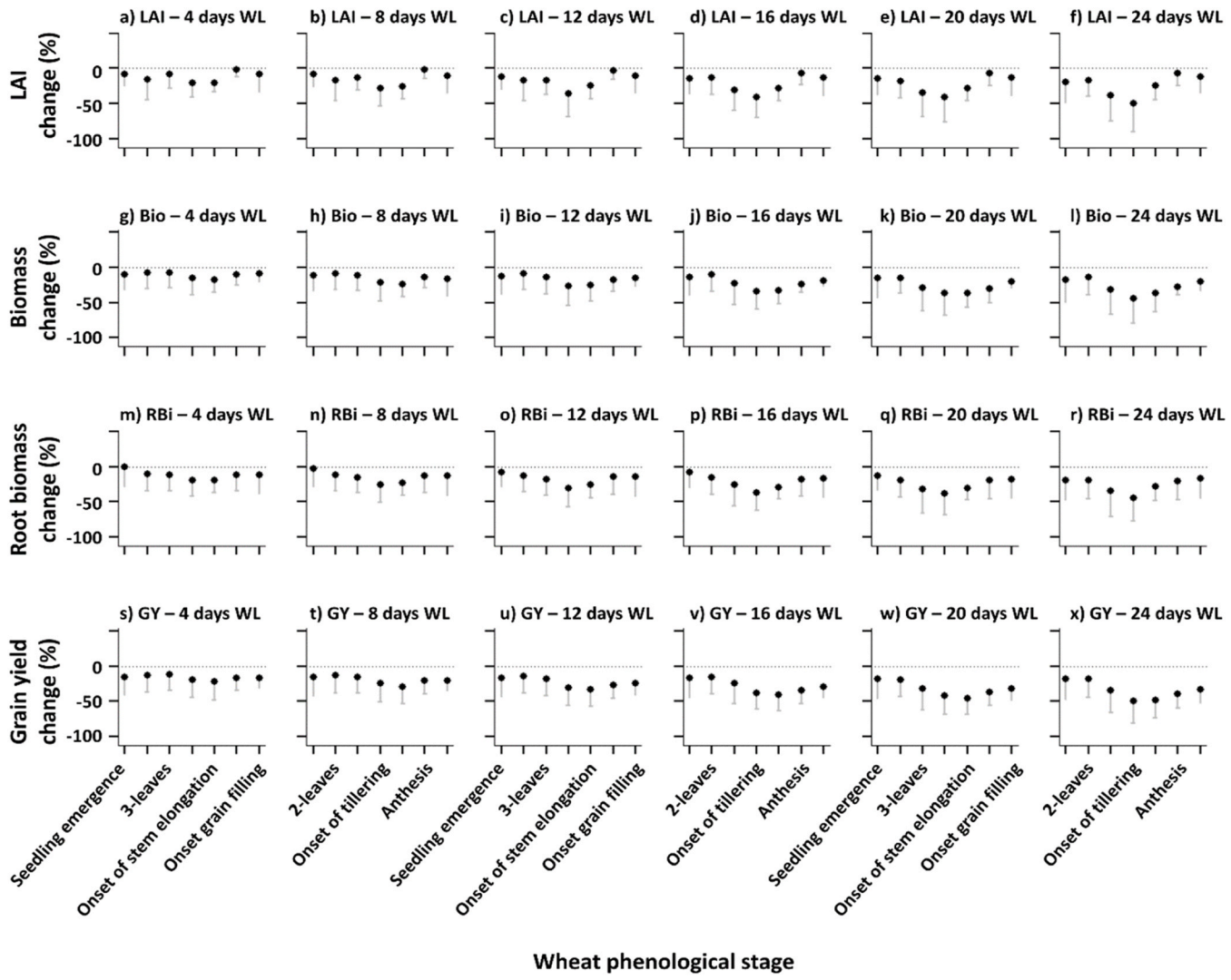


Fig. 4. Sensitivity analysis of the impacts of waterlogging on crop growth simulated by a multi-model mean ensemble. Impacts of simulated waterlogging on (**a-f**) leaf area index (LAI), (**g-l**) total above-ground biomass, (**m-r**) root biomass and (**s-x**) grain yield. The sensitivity analysis was carried out by simulating waterlogging at different phenological stages, from seedling emergence to the onset of grain filling, with six durations varying from 4 to 24 days. The panels indicate, from left to right (e.g. panel **a** to **f**), increasing order of different durations of waterlogging. Dots represent ensemble means based on 15 wheat CMs, with light gray bars indicating one standard deviation across CMs. This standard deviation describes structural spread among CMs and is not an inferential confidence interval; the corresponding median and interquartile range ensemble is given in Supplementary Fig. S5. Sensitivity analysis of the impacts of waterlogging on growth simulated by a multi-model median ensemble is shown in Supplementary Fig. S5.

Fig. S4). These slopes are scenario specific sensitivities of this ensemble, not general biological loss rates. The family balanced ensemble produced nearly the same yield response as the 15 CM mean across the 42 treatments ($r = 0.998$), with a mean absolute difference of 1.2 %age points and a maximum difference of 2.8 %age points. Thus, the main ensemble pattern was not driven by the duplicated MONICA and SIMPLACE model families. For grain yield, the median ensemble in Supplementary Fig. S5 had a similar stage pattern but differed from the mean by as much as 18.9 %age points in individual scenarios, illustrating the influence of extreme CM responses.

At the two stages highlighted most often in empirical work, individual CM responses remained highly heterogeneous. For the 24 day treatment, simulated yield change at stem elongation ranged from a slight increase of 0.6% in Daisy to an 89.3% loss in SSM Wheat. At anthesis, losses ranged from 13.2% in LandscapeDNDC to 76.8% in WOFOST. These ranges should not be interpreted as model bias because Step 3 has no stage specific observations; they quantify the transferability disagreement among CMs after the common calibration step.

3.4. Sensitivity analysis of CMs in simulating heavy rainfall impacts

We evaluated simulated yield sensitivity to synthetic rainfall perturbations from 50 to 1000 mm around anthesis under clayey and sandy soils (Fig. 5). In clay soil, WheatSM and LandscapeDNDC showed the strongest responses, with losses exceeding 30% above 300 mm in LandscapeDNDC and reaching about 70% in WheatSM. The ensemble mean loss was slightly above 10% from 150 mm onward. These values quantify CM sensitivity to the imposed rainfall and soil setup; no observations correspond to these synthetic treatments, so they cannot be used to estimate model bias.

Under sandy soil, ensemble mean losses were about 5% at 150 mm and 10% at 1000 mm (Fig. 5b). Secondary effects of heavy rainfall, including nitrogen leaching and resulting nutrient stress, were intentionally excluded by disabling nitrogen stress or applying additional nitrogen. Disease, lodging, and direct physical rainfall damage were also outside the experimental design. The analysis therefore isolates the modeled pathway from rainfall to soil saturation and direct crop waterlogging stress.

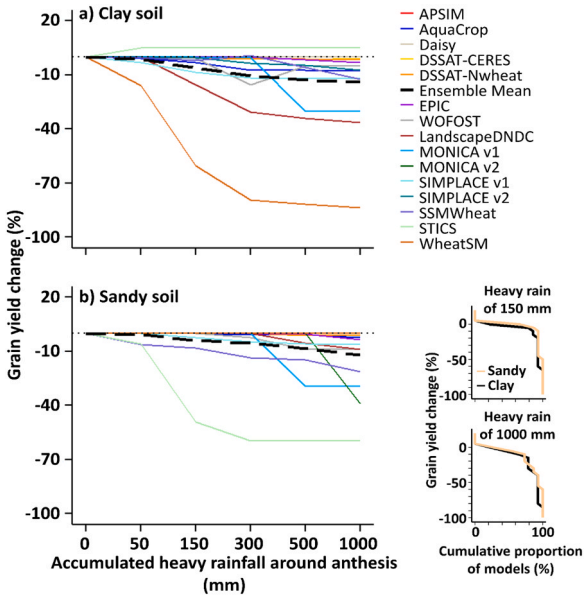


Fig. 5. Simulated yield deviation in response to waterlogging caused by heavy rainfall events around anthesis. Sensitivity analysis was conducted using a multi-model ensemble and individual CMs under varying extreme rainfall events ranging from 0 to 1000 mm (see Supplementary Material for details). Results are shown for clay and sandy soils. Insets display cumulative proportion of CMs simulating a given yield change in response to rainfall events of 150 mm (top inset) or 1000 mm (bottom inset), relative to the control treatment (no additional rainfall).

Across soils, 90% of CMs simulated yield losses below 10% at 150 mm and 80% remained below this threshold even at 1000 mm (insets in Fig. 5). The 500 and 1000 mm treatments are deliberately unrealistic for the Lleida climatology and serve only as upper bound sensitivity tests. The result is therefore that most calibrated CMs respond weakly to rainfall driven saturation under the specified synthetic conditions. Independent field observations with harmonized hydrological measurements would be required to determine whether this weak sensitivity constitutes systematic underestimation. Because the imposed rainfall greatly exceeds plausible evapotranspiration over the 12 day treatment, the added water must be partitioned among soil storage, evapotranspiration, drainage, and runoff. The weak yield response therefore indicates limited transmission of the imposed rainfall forcing to crop stress under many model setups, but the absence of harmonized water fate outputs prevents us from identifying which hydrological pathway dominates across the ensemble. Step 2 should therefore be interpreted as a joint sensitivity of hydrological boundary representation and crop stress formulation.

4. Discussion

The strongest finding of this intercomparison is the large structural disagreement among wheat CMs exposed to identical waterlogging perturbations. Differences occurred in both the magnitude and phenological timing of simulated yield loss, and disagreement increased with stress duration. The controlled Lleida experiment provides a useful calibration benchmark because waterlogging was isolated from other stresses, but it does not establish that the ensemble or any individual CM predicts waterlogging losses accurately across agroecosystems (Asseng et al., 2019). We therefore interpret the analysis as a diagnosis of model behavior and transferability under standardized conditions rather than a ranking of universal predictive performance.

Stage dependence was evident, but no single stage represented the maximum yield sensitivity for most CMs. Stem elongation was the most frequently identified maximum, while other CMs peaked during early vegetative stages, tillering, anthesis, or grain filling. This broad distribution partly reflects that waterlogging coefficients were fitted using treatments concentrated around stem elongation and then transferred unchanged across phenology. It also reflects genuine structural differences in how CMs connect soil saturation to root function, transpiration, photosynthesis, biomass allocation, phenology, and yield components. Empirical syntheses similarly show strong dependence on growth stage and stress duration rather than a universal sensitivity window (Qiu et al., 2025). The experimental literature is itself divided on which stage is most sensitive, with reports variously favouring early establishment, stem elongation, booting and flowering, and grain filling (de San Celedonio et al., 2014; Dickin and Wright, 2008; Marti et al., 2015; Qiu et al., 2025). The dispersion we simulate is therefore not straightforwardly a model deficiency. It also reflects the absence of a consensus empirical benchmark against which any simulated stage response could be judged, which is itself an argument for the multi environment field datasets we call for below.

Large structural uncertainty under severe stress is consistent with multi model studies for other abiotic stresses, where divergence typically increases outside the environmental range used for calibration (Asseng et al., 2015; Wang et al., 2017; Kothari et al., 2022). These comparisons are relevant because they show a recurring modeling problem: alternative formulations can perform similarly under moderate conditions yet diverge strongly under severe perturbation. In the present study, inter model CV increased over the tested range from 4 day to 24 days, while the response groups became progressively more similar to the 24 day partition as duration increased. This supports the second hypothesis only within the tested range. It does not imply that divergence must increase indefinitely: under still more severe stress, a yield floor could instead reduce disagreement as more CMs approach crop failure.

The exploratory structure response analysis reinforces that waterlogging sensitivity cannot be attributed to a single model feature. Direct transpiration effects were associated with a higher median severe stress loss across CMs, whereas direct effects on root growth or root length density were associated with a lower median loss, and photosynthesis showed essentially no median contrast. The inverted direction of the root contrast runs against our third hypothesis. One candidate explanation is the deliberately resource sufficient design of the protocol. Nitrogen limitation was minimized, and the waterlogging treatments were designed to impose excess soil water rather than drought. Under such conditions, penalties routed primarily through root acquisition may propagate less directly to grain yield than penalties applied directly to transpiration or assimilation. This interpretation remains hypothesis generating because root water uptake, nitrogen status, soil oxygen, drainage, and runoff were not harmonized across CMs. It predicts that root linked penalties should become more consequential when post waterlogging water or nitrogen acquisition is limiting, a condition the present protocol was not designed to test. The paired model versions further illustrate the complexity: MONICA v2 had a CM level mean severe stress loss only 4.4 %age points higher than MONICA v1, while SIMPLACE v2 differed from SIMPLACE v1 by 0.5 %age points. These comparisons are descriptive because process categories overlap and versions can differ in parameterization and implementation. Response magnitude therefore emerges from the combination of stress generation, affected crop processes, parameter values, and hydrological setup rather than from the presence of one routine.

The ensemble mean showed the largest average reductions around tillering and stem elongation, broadly consistent with experiments reporting strong wheat yield effects when waterlogging overlaps spike development and the reproductive transition (de San Celedonio et al., 2014; Ghobadi et al., 2017; Marti et al., 2015). Early stage responses were smaller in the ensemble, but this should not be interpreted as evidence of biological tolerance. Field observations can include establishment failure, seedling mortality, pathogen pressure, and later legacy effects that are absent from many CMs. Qiu et al. (2025) further showed

that wheat can express substantial compensatory growth after flooding, indicating that both initial damage and recovery dynamics need to be represented. A CM that omits early mortality but also omits compensatory recovery can reach an apparently plausible final yield through compensating errors.

Most CMs responded weakly to the synthetic rainfall driven saturation experiment (Fig. 5). This is a sensitivity result, not direct evidence of underestimation, because no observations match the 150–1000 mm perturbations and the design intentionally excluded nitrogen stress, disease, lodging, and direct rainfall damage. Real events in southern Brazil in 2023 and France in 2016 combined several of these pathways and therefore cannot serve as direct validation targets for Step 2. Their relevance is conceptual: field yield loss during extremely wet seasons emerges from both waterlogging and associated stresses, whereas the current experiment isolated only the direct modeled waterlogging pathway. Future evaluation should combine field observations of yield with soil water, water table or oxygen status, drainage, nutrient loss, lodging, and disease so that hydrological and crop response errors can be separated.

The increase in CM divergence with stress severity supports a broader need to evaluate model behavior under extremes rather than only under average conditions. Compound stresses add an additional challenge because their effects can be non additive (Ploschuk et al., 2025; Webber et al., 2022). Qiu et al. (2026), for example, found that combined salinity and waterlogging responses shifted from approximately multiplicative effects at moderate stress toward synergistic inhibition beyond stress thresholds, with root aeration, hypoxia adaptation, and energy metabolism playing central roles. This evidence supports interaction and threshold functions grounded in physiology rather than independent stress multipliers when CMs are extended to compound conditions.

Three priorities follow directly from the model defects diagnosed here. First, the wide range of stage specific responses and the absence of a simple relation between direct root routines and yield sensitivity show that CMs need better defined links from soil aeration status to root function, carbon assimilation, and yield formation, together with harmonized hydrological boundary conditions. Second, the transferability test shows that coefficients fitted mainly during stem elongation do not generate a consistent response across other stages. Dynamic damage and recovery functions are therefore needed so that stress history persists appropriately after drainage and can vary with development. Third, the weak direct response to rainfall driven saturation in most CMs, combined with the intentional exclusion of nitrogen loss, disease, and lodging, shows that future evaluation must represent and test interacting pathways that accompany wet extremes. These priorities are supported by the present simulations and are also consistent with physiological and modeling literature (Daniel and Hartman, 2024; Githui et al., 2022; Qiu et al., 2025, 2026).

Several limitations constrain the scope of inference. The waterlogging specific calibration relies on one cultivar, Soissons, and one outdoor soil column experiment in a loamy sand containing 81% sand. The four Arvalis field trials broaden cultivar calibration but do not provide independent waterlogging validation. The Step 2 rainfall experiment uses one season at Lleida and deliberately synthetic forcing. Modeling teams could adjust hydrological parameters or bottom boundary conditions to achieve saturation, and common intermediate outputs such as water table depth, soil oxygen status, drainage, runoff, and complete water balance components were not harmonized across all CMs, preventing a formal separation of hydrological error from crop stress error. The clustering is based on yield responses alone and is recovered most clearly under the severe 24 day treatment, agreement with that partition falling to chance level at 4 days, so it should be interpreted as a severe stress response typology rather than a general model classification. Similarly, the observed increase in inter model variability with duration is demonstrated only over 4–24 days and does not exclude convergence under still more severe stress if yields approach a common floor. Nitrogen related effects, diseases, lodging, direct rainfall damage, genotype differences, and several recovery processes were excluded or incompletely represented. These constraints mean that the study diagnoses structural disagreement under standardized perturbations; it does not validate waterlogging risk prediction across soils, genotypes, or production regions. Multi environment field datasets with complete weather, soil, hydrological, phenological, and yield observations are essential for the next evaluation step.

5. Conclusions

Standardized waterlogging perturbations reveal substantial disagreement among current wheat CMs in both the timing and magnitude of simulated yield loss. Before waterlogging specific crop response coefficients were fitted to the eight waterlogged Lleida treatments in the current exercise, the ensemble median reproduced these treatments with an rRMSE of 17.0%, providing a pre fit extrapolative benchmark rather than strict independent validation; calibration reduced the same metric to 6.3%. Response profiles diverged when the fitted waterlogging parameterizations were transferred across phenological stages, and inter model variability increased with duration over the tested 4–24 day range. Related model families did not drive the ensemble pattern, and the four severe stress response groups were reproduced by hierarchical clustering but were not recovered at the shortest duration. Most CMs also showed weak sensitivity to rainfall driven saturation under the synthetic Step 2 scenarios, a result that requires independent field observations and harmonized hydrological outputs before it can be interpreted as bias. The study therefore provides a quantitative diagnosis of where model uncertainty emerges and identifies three priorities for improvement: explicit and testable links between soil aeration and crop processes, development dependent damage and recovery functions, and evaluation of interacting stresses using field datasets that include hydrological as well as crop observations. These conclusions rest on one cultivar, one outdoor column experiment in a coarse textured soil, and one synthetic rainfall season. They describe how this set of CMs behaves under a common protocol; they do not establish how accurately any of them predicts waterlogging loss across production environments.

CRediT authorship contribution statement

Rogério de S. Nóia-Júnior: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Visualization, Writing – original draft. **Matthew Tom Harrison:** Conceptualization, Project administration, Software, Formal analysis, Data curation, Methodology, Visualization, Supervision, Writing – review & editing. **Taru Palosuo:** Conceptualization, Project administration, Methodology, Visualization, Supervision, Writing – review & editing. **Margarita García-Vila:** Conceptualization, Project administration, Formal analysis, Data curation, Software, Writing – review & editing. **Ke Liu:** Conceptualization, Project administration, Visualization, Software, Formal analysis, Data curation, Writing – review & editing. **Murilo dos Santos Vianna:** Conceptualization, Project administration, Visualization, Writing – review & editing, Formal analysis, Data curation, Software. **Tobias KD Weber:** Conceptualization, Project administration, Visualization, Writing – review & editing. **Jin Zhao:** Conceptualization, Project administration, Visualization, Writing – review & editing. **Zhentao Zhang:** Conceptualization, Project administration, Data curation, Methodology, Visualization, Writing – review & editing. **Diego Abalos:** Software, Formal analysis, Data curation, Writing – review & editing. **Senthold Asseng:** Software, Formal analysis, Data curation, Writing – review & editing. **Juraj Balkovic:** Software, Formal analysis, Data curation, Writing – review & editing. **Attila Bussay:** Software, Formal analysis, Data curation, Writing – review & editing. **Xianguan Chen:** Software, Formal analysis, Data curation, Writing – review & editing. **Mathieu Delandmeter:** Software, Formal analysis, Data

curation, Writing – review & editing. **Benjamin Dumont:** Software, Formal analysis, Data curation, Writing – review & editing. **Liping Feng:** Software, Formal analysis, Data curation, Writing – review & editing. **Roberto Ferrise:** Software, Formal analysis, Data curation, Writing – review & editing. **Christian Folberth:** Software, Formal analysis, Data curation, Writing – review & editing. **Davide Fumagalli:** Software, Formal analysis, Data curation, Writing – review & editing. **Surendra Giri:** Software, Formal analysis, Data curation, Writing – review & editing. **Aram Gorooei:** Software, Formal analysis, Data curation, Writing – review & editing. **Muhammad Habib-Ur-Rahman:** Software, Formal analysis, Data curation, Writing – review & editing. **Yean-Uk Kim:** Software, Formal analysis, Data curation, Writing – review & editing. **David Kraus:** Software, Formal analysis, Data curation, Writing – review & editing. **Huan Liu:** Software, Formal analysis, Data curation, Writing – review & editing. **Lioba Martin:** Software, Formal analysis, Data curation, Writing – review & editing. **Julien Morel:** Software, Formal analysis, Data curation, Writing – review & editing. **Claas Nendel:** Software, Formal analysis, Data curation, Writing – review & editing. **Gloria Padovan:** Software, Formal analysis, Data curation, Writing – review & editing. **Marlene Palka:** Software, Formal analysis, Data curation, Writing – review & editing. **Johannes Wilhelmus Maria Pullens:** Software, Formal analysis, Data curation, Writing – review & editing. **Reimund P. Rötter:** Conceptualization. **Valentina Stocca:** Software, Formal analysis, Data curation, Writing – review & editing. **Maurits van den Berg:** Software, Formal analysis, Data curation, Writing – review & editing. **Heidi Webber:** Software, Formal analysis, Data curation, Writing – review & editing. **Jean-Charles Deswarte:** Resources, Writing – review & editing. **Jean-Pierre Cohan:** Resources, Writing – review & editing. **Alex C. Ruane:** Validation, Writing – review & editing. **Pierre Martre:** Conceptualization, Funding acquisition, Validation, Supervision, Writing – review & editing.

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.

Acknowledgements

P.M. acknowledges support from the metaprogram Agriculture and forestry in the face of climate change: adaptation and mitigation (CLIMAE) and the AgroEcoSystem Division of INRAE. R.S.N.J and P.M. received supports from the European Union's Horizon 2020 research and innovation program under grant agreement N°817970. ACR contributions were made possible through Earth Science Division support for the NASA GISS Impacts Group. The authors also thank Prof. Gustavo Slafer for his valuable contributions to the critical interpretation of the results.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.agwat.2026.110882.

Data availability

Data will be made available on request.

References

Agencia Estatal de Meteorología (AEMET), 2026. Valores climatológicos normales: Lleida, station 9771C, 1983–2010 normals. AEMET, Madrid, Spain (accessed 3 September 2026).

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V. V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P.J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat production. Nat. Clim. Chang. 5, 143–147. https://doi.org/10.1038/nclimate2470.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., 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., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Chang. 3, 827–832. https://doi.org/10.1038/nclimate1916.

Asseng, S., Martre, P., Ewert, F., 2019. Crop simulation model inter-comparison and improvement. In: Boote, K. (Ed.), Advances in Crop Modelling for a Sustainable Agriculture. Burleigh Dodds Science Publishing, Cambridge, UK, pp. 449–476. https://doi.org/10.19103/as.2019.0061.22.

Ben-Ari, T., Boé, J., Ciais, P., Lecerf, R., Van der Velde, M., Makowski, D., 2018. Causes and implications of the unforeseen 2016 extreme yield loss in the breadbasket of France. Nat. Commun. 9, 1627. https://doi.org/10.1038/s41467-018-04087-x.

Chen, J., Zhao, C., Harrison, M.T., Zhou, M., 2024. Nitrogen fertilization alleviates barley (Hordeum vulgare L.) waterlogging. Agronomy 14. https://doi.org/10.3390/agronomy14081712.

Daniel, K., Hartman, S., 2024. How plant roots respond to waterlogging. J. Exp. Bot. 75, 511–525. https://doi.org/10.1093/jxb/erad332.

de San Celedonio, R.P., Abeledo, L.G., Miralles, D.J., 2014. Identifying the critical period for waterlogging on yield and its components in wheat and barley. Plant. Soil. 378, 265–277. https://doi.org/10.1007/s11104-014-2028-6.

de San Celedonio, R.P., Abeledo, L.G., Miralles, D.J., 2018. Physiological traits associated with reductions in grain number in wheat and barley under waterlogging. Plant. Soil. 429, 469–481. https://doi.org/10.1007/s11104-018-3708-4.

Dickin, E., Wright, D., 2008. The effects of winter waterlogging and summer drought on the growth and yield of winter wheat (Triticum aestivum L.). Eur. J. Agron. 28, 234–244. https://doi.org/10.1016/j.eja.2007.07.010.

Orla Dwyer, 2025. Mapped: How extreme weather is destroying crops around the world [WWW Document]. Carbon Brief - Clear on Climate. URL https://interactive.carbonbrief.org/crops-extreme-weather/index.html (accessed 4.21.25).

Fowler, H.J., Lenderink, G., Prein, A.F., Westra, S., Allan, R.P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H.X., Guerreiro, S., Haerter, J.O., Kendon, E.J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., Zhang, X., 2021. Anthropogenic intensification of short-duration rainfall extremes. Nat. Rev. Earth Environ. 2, 107–122. https://doi.org/10.1038/s43017-020-00128-6.

Fu, J., Jian, Y., Wang, X., Li, L., Ciais, P., Zscheischler, J., Wang, Y., Tang, Y., Müller, C., Webber, H., Yang, B., Wu, Y., Wang, Q., Cui, X., Huang, W., Liu, Y., Zhao, P., Piao, S., Zhou, F., 2023. Extreme rainfall reduces one-twelfth of China's rice yield over the last two decades. Nat. Food. https://doi.org/10.1038/s43016-023-00753-6.

Garcia-Vila, M., dos Santos Vianna, M., Harrison, M.T., Liu, K., de S. Nóia-Júnior, R., Weber, T.K.D., Zhao, J., Acutis, M., Archontoulis, S., Asseng, S., Aubry, P., Balkovic, J., Basso, B., Chen, X., Chen, Y., de Jong van Lier, Q., Delandmeter, M., de Wit, A., Dumont, B., Ferrise, R., Folberth, C., Gabbrielli, M., Gaiser, T., Gorooei, A., Hoogenboom, G., Kersebaum, K.C., Kim, Y.-U., Kraus, D., Liu, B., Martin, L., Metselaar, K., Nendel, C., Padovan, G., Perego, A., Seserman, D.M., Scheer, C., Shelia, V., Stocca, V., Tao, F., Wang, E., Webber, H., Zhao, Z., Zhu, Y., Palosuo, T., 2025. Gaps and strategies for accurate simulation of waterlogging impacts on crop productivity. Nat. Food. https://doi.org/10.1038/s43016-025-01179-y.

Ghobadi, M.E., Ghobadi, M., Zebarjadi, A., 2017. Effect of waterlogging at different growth stages on some morphological traits of wheat varieties. Int. J. Biometeorol. 61, 635–645. https://doi.org/10.1007/s00484-016-1240-x.

Githui, F., Beverly, C., Aiad, M., McCaskill, M., Liu, K., Harrison, M.T., 2022. Modelling waterlogging impacts on crop growth: a review of aeration stress definition in crop models and sensitivity analysis of APSIM. Int. J. Plant. Biol. 13, 180–200. https://doi.org/10.3390/ijpb13030017.

Herzog, M., Striker, G.G., Colmer, T.D., Pedersen, O., 2016. Mechanisms of waterlogging tolerance in wheat – a review of root and shoot physiology. Plant. Cell. Environ. 39, 1068–1086. https://doi.org/10.1111/pce.12676.

Jägermeyr, J., Müller, C., Ruane, A.C., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J.A., Fuchs, K., Guarin, J.R., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A.K., Kelly, D., Khabarov, N., Lange, S., Lin, T.-S., Liu, W., Mialyk, O., Minoli, S., Moyer, E.J., Okada, M., Phillips, M., Porter, C., Rabin, S.S., Scheer, C., Schneider, J.M., Schyns, J.F., Skalsky, R., Smerald, A., Stella, T., Stephens, H., Webber, H., Zabel, F., Rosenzweig, C., 2021. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. Nat. Food. https://doi.org/10.1038/s43016-021-00400-y.

Kim, Y.-U., Ruane, A.C., Finger, R., Webber, H., 2025. Robust assessment of climatic risks to crop production. Nat. Food. https://doi.org/10.1038/s43016-025-01168-1.

Kim, Y.-U., Webber, H., Adiku, S.G.K., Nóia Júnior, R. de S., Deswarte, J.-C., Asseng, S., Ewert, F., 2024. Mechanisms and modelling approaches for excessive rainfall stress on cereals: Waterlogging, submergence, lodging, pests and diseases. Agric. For. Meteorol. 344, 109819. https://doi.org/10.1016/j.agrformet.2023.109819.

Kothari, K., Battisti, R., Boote, K.J., Archontoulis, S.V., Confalone, A., Constantin, J., Cuadra, S.V., Debaeke, P., Faye, B., Grant, B., Hoogenboom, G., Jing, Q., van der

Laan, M., Macena da Silva, F.A., Marin, F.R., Nehbandani, A., Nendel, C., Purcell, L. C., Qian, B., Ruane, A.C., Schoving, C., Silva, E.H.F.M., Smith, W., Soltani, A., Srivastava, A., Vieira, N.A., Slone, S., Salmerón, M., 2022. Are soybean models ready for climate change food impact assessments? Eur. J. Agron. 135, 126482. https://doi.org/10.1016/j.eja.2022.126482.

Kulkarni, Anil V., Shirsat, T.S., Kulkarni, Ashutosh, Negi, H.S., Bahuguna, I.M., Thamban, M., 2021. State of Himalayan cryosphere and implications for water security. Water Secur. 14, 100101. https://doi.org/10.1016/j.wasec.2021.100101.

Li, Y., Guan, K., Schnitkey, G.D., DeLucia, E., Peng, B., 2019. Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States. Glob. Chang. Biol. 25, 2325–2337. https://doi.org/10.1111/gcb.14628.

Liu, K., Harrison, M.T., Archontoulis, S.V., Huth, N., Yang, R., Liu, D.L., Yan, H., Meinke, H., Huber, I., Feng, P., Ibrahim, A., Zhang, Y., Tian, X., Zhou, M., 2021. Climate change shifts forward flowering and reduces crop waterlogging stress. Environ. Res. Lett. 16. https://doi.org/10.1088/1748-9326/ac1b5a.

Liu, K., Harrison, M.T., Ibrahim, A., Manik, S.M.N., Johnson, P., Tian, X., Meinke, H., Zhou, M., 2020a. Genetic factors increasing barley grain yields under soil waterlogging. Food Energy Secur. 9, e238. https://doi.org/10.1002/fes3.238.

Liu, K., Harrison, M.T., Shabala, S., Meinke, H., Ahmed, I., Zhang, Y., Tian, X., Zhou, M., 2020b. The state of the art in modeling waterlogging impacts on plants: what do we know and what do we need to know. Earths Future 8. https://doi.org/10.1029/2020EF001801.

Liu, K., Harrison, M.T., Yan, H., Liu, D.L., Meinke, H., Hoogenboom, G., Wang, B., Peng, B., Guan, K., Jaegermeyr, J., Wang, E., Zhang, F., Yin, X., Archontoulis, S., Nie, L., Badea, A., Man, J., Wallach, D., Zhao, J., Benjumea, A.B., Fahad, S., Tian, X., Wang, W., Tao, F., Zhang, Z., Rötter, R., Yuan, Y., Zhu, M., Dai, P., Nie, J., Yang, Y., Zhang, Y., Zhou, M., 2023. Silver lining to a climate crisis in multiple prospects for alleviating crop waterlogging under future climates. Nat. Commun. 14, 765. https://doi.org/10.1038/s41467-023-36129-4.

Manghwar, H., Hussain, A., Alam, I., Khoso, M.A., Ali, Q., Liu, F., 2024. Waterlogging stress in plants: unraveling the mechanisms and impacts on growth, development, and productivity. Environ. Exp. Bot. 224, 105824. https://doi.org/10.1016/j.envexpbot.2024.105824.

Marti, J., Savin, R., Slafer, G.A., 2015. Wheat yield as affected by length of exposure to waterlogging during stem elongation. J. Agron. Crop. Sci. 201, 473–486. https://doi.org/10.1111/jac.12118.

Martre, P., Dueri, S., Guarin, J.R., Ewert, F., Webber, H., Calderini, D., Molero, G., Reynolds, M., Miralles, D., Garcia, G., Brown, H., George, M., Craigie, R., Cohan, J.-P., Deswarte, J.-C., Slafer, G., Giunta, F., Cammarano, D., Ferrise, R., Gaiser, T., Gao, Y., Hochman, Z., Hoogenboom, G., Hunt, L.A., Kersebaum, K.C., Nendel, C., Padovan, G., Ruane, A.C., Srivastava, A.K., Stella, T., Supit, I., Thorburn, P., Wang, E., Wolf, J., Zhao, C., Zhao, Z., Asseng, S., 2024. Global needs for nitrogen fertilizer to improve wheat yield under climate change. Nat. Plants 10, 1081–1090. https://doi.org/10.1038/s41477-024-01739-3.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rötter, 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., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2015. Multimodel ensembles of wheat growth: many models are better than one. Glob. Chang. Biol. 21, 911–925. https://doi.org/10.1111/gcb.12768.

Jiang Mengnan, 2023. Henan wheat harvest hit by heavy rains [WWW Document]. Dialogue Earth. URL https://dialogue.earth/en/digest/henan-wheat-harvest-hit-by-heavy-rains/ (accessed 4.21.25).

Nguyen, T.-N., Son, S., Jordan, M.C., Levin, D.B., Ayele, B.T., 2016. Lignin biosynthesis in wheat (Triticum aestivum L.): its response to waterlogging and association with hormonal levels. BMC Plant. Biol. 16, 28. https://doi.org/10.1186/s12870-016-0717-4.

NIEMEYER Stefan, VANDENBERG Maurits, 2024. Crop monitoring in Europe Severe impacts of hot conditions in the south.

Nóia Júnior, R. de S., Asseng, S., García-Vila, M., Liu, K., Stocca, V., dos Santos Vianna, M., Weber, T.K.D., Zhao, J., Palosuo, T., Harrison, M.T., 2023. A call to action for global research on the implications of waterlogging for wheat growth and yield. Agric. Water Manag. 284. https://doi.org/10.1016/j.agwat.2023.108334.

Nóia Júnior, R. de S., Martre, P., Finger, R., van der Velde, M., Ben-Ari, T., Ewert, F., Webber, H., Ruane, A.C., Asseng, S., 2021. Extreme lows of wheat production in Brazil. Environ. Res. Lett. 16, 104025. https://doi.org/10.1088/1748-9326/ac26f3.

Nóia-Júnior, R. de S., Deswarte, J.-C., Cohan, J.-P., Martre, P., van der Velde, M., Lecerf, R., Webber, H., Ewert, F., Ruane, A.C., Slafer, G.A., Asseng, S., 2023. The extreme 2016 wheat yield failure in France. Glob. Chang. Biol. 29, 3130–3146. https://doi.org/10.1111/gcb.16662.

Nóia-Júnior, Rogério de S., Asseng, S., Müller, C., Deswarte, J.-C., Cohan, J.-P., Martre, P., 2025a. Negative impacts of climate change on crop yields are underestimated. Trends Plant. Sci. https://doi.org/10.1016/j.tplants.2025.05.002.

Nóia-Júnior, Rogério de S., Christo, B.F., Pezzopane, J.E.M., 2025. Extreme weather events in southern Brazil warn of agricultural collapse. Research 2, 100217. https://doi.org/10.1016/j.nexres.2025.100217.

Nóia-Júnior, Rogério de S., Martre, P., Deswarte, J.-C., Cohan, J.-P., Van der Velde, M., Webber, H., Ewert, F., Ruane, A.C., Ben-Ari, T., Asseng, S., 2025b. Past and future wheat yield losses in France's breadbasket. Field Crops Res. 322, 109703. https://doi.org/10.1016/j.fcr.2024.109703.

Pan, J., Sharif, R., Xu, X., Chen, X., 2021. Mechanisms of waterlogging tolerance in plants: research progress and prospects. Front. Plant. Sci.

Pequeno, D.N.L., Hernández-Ochoa, I.M., Reynolds, M., Sonder, K., MoleroMilan, A., Robertson, R.D., Lopes, M.S., Xiong, W., Kropff, M., Asseng, S., 2021. Climate impact and adaptation to heat and drought stress of regional and global wheat production. Environ. Res. Lett. 16, 54070. https://doi.org/10.1088/1748-9326/abd970.

Ploschuk, R.A., Savin, R., Slafer, G.A., 2025. Dual stress, equivalent harm? hypothesizing on the type of interactions between waterlogging and high temperature. Front. Plant. Sci. 15. https://doi.org/10.3389/fpls.2024.1472665.

Qiu, S., Zhang, Y., Dai, J., Dong, H., 2025. Physiological mechanisms and agronomic strategies underlying flood tolerance variability in dryland crops: A global meta-analysis. Field Crops Res. 334, 110146. https://doi.org/10.1016/j.fcr.2025.110146.

Qiu, S., Zhang, Y., Dai, J., Dong, H., 2026. Ecological divergence governs plant resilience to compound salinity-waterlogging stress under global change. Glob. Change Biol. 32, e70875. https://doi.org/10.1111/gcb.70875.

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G., Winter, J.M., 2013. The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agric. For. Meteorol. 170, 166–182. https://doi.org/10.1016/j.agrformet.2012.09.011.

Rötter, R.P., Appiah, M., Fichtler, E., Kersebaum, K.C., Trnka, M., Hoffmann, M.P., 2018a. Linking modelling and experimentation to better capture crop impacts of agroclimatic extremes—A review. Field Crops Res. 221, 142–156. https://doi.org/10.1016/j.fcr.2018.02.023.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Luca, A., Di, Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S.M., Wehner, M., Zhou, B., 2021. The physical science basis. In: Press, C.U. (Ed.), Sixth Assessment Report of the Intergovernmental Panel on Climate Change.

Shaw, R.E., Meyer, W.S., McNeill, A., Tyerman, S.D., 2013. Waterlogging in Australian agricultural landscapes: a review of plant responses and crop models. Crop. Pasture Sci. 64, 549–562.

Shiferaw, B., Smale, M., Braun, H.-J., Duveiller, E., Reynolds, M., Muricho, G., 2013. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Secur. 5, 291–317. https://doi.org/10.1007/s12571-013-0263-y.

Team, Rs., 2020. RStudio: Integrated Development for R.

Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., Rötter, R.P., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Cammarano, D., Challinor, A.J., De Sanctis, G., Doltra, J., Dumont, B., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.-K., Liu, L., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ripoche, D., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wallach, D., Wang, Z., Wolf, J., Zhu, Y., Asseng, S., 2017. The uncertainty of crop yield projections is reduced by improved temperature response functions. Nat. Plants 3, 17102. https://doi.org/10.1038/nplants.2017.102.

Webber, H., Cooke, D., Wang, C., Asseng, S., Martre, P., Ewert, F., Kimball, B., Hoogenboom, G., Evett, S., Chanzy, A., Garrigues, S., Olioso, A., Copeland, K.S., Steiner, J.L., Cammarano, D., Chen, Y., Crépeau, M., Diamantopoulos, E., Ferrise, R., Manceau, L., Gaiser, T., Gao, Y., Gayler, S., Guarin, J.R., Hunt, T., Jégo, G., Padovan, G., Pattey, E., Ripoche, D., Rodríguez, A., Ruiz-Ramos, M., Shelia, V., Srivastava, A.K., Supit, I., Tao, F., Thorp, K., Viswanathan, M., Weber, T., White, J., 2025. Wheat crop models underestimate drought stress in semi-arid and Mediterranean environments. Field Crops Res. 332, 110032. https://doi.org/10.1016/j.fcr.2025.110032.

Webber, H., Martre, P., Asseng, S., Kimball, B., White, J., Ottman, M., Wall, G.W., De Sanctis, G., Doltra, J., Grant, R., Kassie, B., Maiorano, A., Olesen, J.E., Ripoche, D., Rezaei, E.E., Semenov, M.A., Stratonovitch, P., Ewert, F., 2017. Canopy temperature for simulation of heat stress in irrigated wheat in a semi-arid environment: A multi-model comparison. Field Crops Res. 202, 21–35. https://doi.org/10.1016/j.fcr.2015.10.009.

Webber, H., Rezaei, E.E., Ryo, M., Ewert, F., 2022. Framework to guide modeling single and multiple abiotic stresses in arable crops. Agric. Ecosyst. Environ. 340, 108179. https://doi.org/10.1016/j.agee.2022.108179.

Yan, H., Harrison, M.T., Liu, K., Wang, B., Feng, P., Fahad, S., Meinke, H., Yang, R., Liu, D.L., Archontoulis, S., Huber, I., Tian, X., Man, J., Zhang, Y., Zhou, M., 2022. Crop traits enabling yield gains under more frequent extreme climatic events. Sci. Total. Environ. 808, 152170. https://doi.org/10.1016/j.scitotenv.2021.152170.

Zampieri, M., Ceglar, A., Dentener, F., Toreti, A., 2017. Wheat yield loss attributable to heat waves, drought and water excess at the global, national and subnational scales. Environ. Res. Lett. 12, 64008. https://doi.org/10.1088/1748-9326/aa723b.

Zhou, Q., Wu, X., Xin, L., Jiang, H., Wang, X., Cai, J., Jiang, D., 2018. Waterlogging and simulated acid rain after anthesis deteriorate starch quality in wheat grain. Plant. Growth Regul. 85, 257–265. https://doi.org/10.1007/s10725-018-0390-8.