https://doi.org/10.1038/s44264-025-00067-z

From field to analysis: strengthening reproducibility and confirmation in research for sustainable agriculture

Jeffrey W. White[1] ✉, Kenneth J. Boote[1], Bruce A. Kimball[2], Cheryl Porter[1], Montse Salmeron[3], Vakhtang Shelia[1], Kelly R. Thorp[4] & Gerrit Hoogenboom[1]

Lack of robustness and potential bias are growing concerns for research, including for sustainable agriculture. Research confirmation requires independent duplication of field experiments, modeling and other analyses. Key concepts include "repeatability" (consistency within an experiment), "replicability" (same team, different environments), and "reproducibility" (independent team, different environments). Researchers must improve workflow descriptions, especially regarding crop environments and management. A useful metric is how well research could be reproduced in ten years.

Novel research findings require independent confirmation before attaining acceptance. This process often involves terms such as "repeatability," "replicability," and "reproducibility." The reliability of confirmation has been questioned from multiple perspectives. Low statistical power and researcher bias lead to the reporting of incorrect findings[1]. Pressure to publish and selective publication of results also reduce reproducibility[2]. Post hoc hypothesis formulation and flexibility in analyses are further problems[3]. These and similar studies suggest that the credibility of science is eroding[4,5]. Responding to these concerns, the US National Academies of Sciences, Engineering, and Medicine assessed the status of the confirmation process and recommended measures to improve the rigor and transparency of research, resulting in the report *Reproducibility and Replicability in Science*[6], hereafter referred to as the "NASEM Report."

Over the last four decades, a large body of research has examined how agricultural practices affect the sustainability of production systems, what adaptations might help ensure sustainable production, and what mitigation opportunities exist for issues such as greenhouse gas emissions. These endeavors have generated controversies related to whether reported results are robust or how applicable they are across environments or cropping practices. Examples include tillage effects on soil carbon[7,8], cover crop effects on nitrate leaching[9], and crop responses to climate change, including atmospheric CO_2 ($[CO_2]$)[10–12], temperature[13–15] and relative humidity[16].

Scrutiny of sustainability research will only intensify. As issues such as food security, non-point source water pollution, and climate change gain attention in political and economic arenas, policies are being proposed with large impacts on producers and other stakeholders. Both proponents and opponents of specific policies will demand that supporting research is robust. Examples of concerns with potentially significant economic or legal implications include the reliability of soil carbon credits[17] and the estimated impacts of aerosols on global warming[18].

Additional trends may further erode research integrity and confirmation. "Publish or perish" policies that link publishing in high-impact journals to job advancement may induce researchers to cut corners, including decreasing internal confirmation prior to publication[19] and manipulating data or analyses to enhance apparent statistical significance, often critiqued as "p-hacking"[20]. Funding agencies may support only "novel" or "innovative" research rather than confirmation studies. Journals heavily reliant on manuscript fees may favor lax peer review[21,22]. Papers may be generated from fake data[5,23], while content generated using artificial intelligence may involve plagiarism while reducing creativity and innovation[24]. As field research costs increase[25], researchers may reduce sampled area, replication, or number of trials, further compromising confirmation.

Given the interest in research confirmation and the increasing likelihood of controversies, the confirmation of agricultural research merits examination. The foremost concern is to increase the robustness of findings, but improved confirmation should also reduce the diversion of resources to refuting ill-founded results[26,27].

Focusing on research to support sustainable production, we review the levels of confirmation embodied in the terms "repeatability", "replicability", and "reproducibility", and propose steps toward strengthening reproducibility for both field research and modeling-based studies. Compared to other areas of agricultural research, reproducibility is especially relevant for research on sustainable production. Firstly, sustainable production often involves more complex management than in conventional systems, so research may require more complex treatments and measures of performance, challenging efforts to reproduce research. Secondly, sustainable

[1]Department of Agricultural and Biological Engineering University of Florida, Gainesville, FL, USA. [2]Arid-Land Agricultural Research Center, USDA ARS, Maricopa, AZ, USA. [3]Department of Plant and Soil Sciences, University of Kentucky, Lexington, KY, USA. [4]Grassland Soil and Water Research Laboratory, USDA ARS, Temple, TX, USA. ✉e-mail: jeff.white.az@gmail.com

practices often involve matching inputs to specific, local conditions, making thorough characterization of field environments paramount. Thirdly, the time frames considered for sustainability are longer than for other research, introducing challenges both for documenting management and environmental conditions as well as for independent duplication of field research. Finally, emphasizing on- and off-site impacts implies a direct connection with environmental concerns that invite scrutiny by multiple stakeholders. In this context, numerical modeling and related tools are invaluable for examining how multiple factors may interact in production scenarios and how measured responses may vary spatially and temporally, especially given climate uncertainty.

To frame our discussion of research confirmation, it is helpful to consider a single-season experiment. A specific crop phenotype or environmental property (P_t) at a time (t) is usefully described as a function of the field's initial conditions (time t = 0, $F_{t=0}$), the crop's genetics (G), the environment (E_t), crop management (M_t) and ε_t, representing random error (errors from measurements of P_t, input data, model structure, parameter estimation, emergent properties, and other sources or variation):

$$P_t = f(F_{t=0}, G, E_t, M_t) + \varepsilon_t. \quad (1)$$

Experimental treatments are applied by varying G, E_t, or M_t at one or more locations or time sequences. Experimental units may involve individual plants or experimental plots.

Reproducing a series of P_t values requires conducting confirmatory studies under conditions of G, E_t, and M_t that are relevant to the underlying research problem. In field research, natural variation in E_t precludes perfect duplication of prior results. Researchers attempt to confirm findings via replicated experiments conducted over multiple locations or seasons, usually also seeking to understand how P_t varies with E_t. In numerical modeling, while estimating values of P_t given $F_{t=0}$, G, E_t, and M_t may appear straightforward, efforts to duplicate work often encounter difficulties due to factors including our incomplete understanding of crop processes, uncertainty of model inputs, and software issues, as examined later.

Terminology for confirmation of research findings

The terminology for confirmation of research varies greatly. "Repeatability," "reproducibility," and "replicability" are often interchanged[28]. In many disciplines, their meanings vary with how identical the confirmatory process is to the original experiment or analysis[29]. The NASEM Report defined "reproducibility" as the ability to obtain consistent results using the same input data, computations, methods, code, and conditions of analysis, thus focusing exclusively on computations and rendering the term synonymous with "computational reproducibility." "Replicability" was defined as the ability to obtain consistent results across studies that are directed at the same research question, each study obtaining its own data. Two studies are replicated if they provide consistent results within the expected uncertainty of the study system. "Repeatability" was considered a specialized term from metrology relating to measurements repeated close in time and using the same conditions and equipment.

Use of the terms "repeatability," "replicability," and "reproducibility" in agricultural research shows limited consistency (Table 1) and often diverges from the NASEM Report (Table 2). In agricultural research, "repeatability" usually refers to the ability of a research group to obtain essentially identical results when an analysis or experiment is repeated within a study or under the same conditions as an initial study. Given the inherent variability of E_t in field experiments due to variability of weather, soils and biotic factors, repeatability is difficult to achieve and not fully expected. However, the term appears applicable for individual measurements over short time intervals and for laboratory analyses. For modeling and numerical analyses, repeatability implies that data, scripts, software or processing environments are unchanged from previous work within a research group, and that results are essentially identical.

We consider "replicability" as the ability of a single research group to obtain identical results from a previous study when using the same methods, including numerical analyses (Table 2). The concept includes single-season field experiments repeated over multiple seasons or locations. Replication increases confidence that results hold true, and quantifying the effects of G, E_t and M_t helps indicate how responses might vary across seasons or regions. Computational replicability is seldom discussed in agricultural research but seems synonymous with "repeatability".

"Reproducibility" refers to obtaining comparable results from a study independent of the original. A new field experiment might obtain results that confirm prior research, but involve different cultivars, crops, locations, or management. Such work often seeks both to confirm the original study and to understand whether the results are robust for a broader range of situations. Analogously, reproducibility in modeling involves two situations. The first is when independent researchers use data from the original study and the same or other models to confirm the original results. The second arises when new sets of data are modeled in a manner similar to the original study.

The most notable differences between agricultural research and the NASEM Report (Table 2) are that the Report lacks equivalents for "replicability" in agricultural research and that it thus considers our use of "reproducibility" to be "replicability", while limiting "reproducibility" to computations. A further difference is that we consider confirmation of computations, taken to include modeling and other numerical analyses, as crucial both for original research and independent, external confirmation.

Confirmation of Field Research

Considering $P_t = f(F_{t=0}, G, E_t, M_t)$, the first challenge for strengthening confirmation of field research is to describe $F_{t=0}$, G, E_t, and M_t in sufficient detail that other researchers can understand how the results were obtained and, if desired, reproduce the experiment within the constraints inherent to reproducing the E_t, including soil, weather and biotic conditions. A second challenge is to describe protocols used to obtain values of observed P_t.

For describing F_0, G, E_t, and M_t, the standards first developed by the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) project and subsequently revised by the International Consortium for Agricultural Systems Applications (ICASA) provide a useful vocabulary and data architecture for documenting experiments[30,31]. The standards were used by the Agricultural Model Intercomparison and Improvement Project (AgMIP)[32] and formed the core of the AgMIP data management system[33].

Protocols for measuring P_t present further challenges. Economic yield might be described by the plot area, whether border- or end-rows were excluded, the threshing process, and how moisture contents were determined. Traits such as leaf photosynthesis, canopy reflectance, and soil nutrient concentrations require descriptions of instrument configurations and calibrations, sampling criteria and procedures, conditions during measurements, and data processing, among other metadata. The Prometheus web resource[34] hosts protocols for ecological and environmental plant physiology. The platform protocols.io[35] provides tools for entering, editing, and sharing protocols within a research group, and finalized protocols may be associated with a unique DOI (digital object identifier). However, neither platform is widely used in agricultural research.

Improper data manipulation and misuse of statistical tests can lead to erroneous results. Practices such as modifying analyses to achieve statistical significance ("p-hacking"), formulating hypotheses after observing the results ("HARKing"), and publishing only positive findings ("publication bias") increase the risk of reporting non-existent effects or relationships (false positives or Type I errors), which are seldom reproducible[20,36].

Even for series of well-characterized field experiments, reaching a consensus on implications of results can be challenging. Studies of crop response to elevated atmospheric CO_2 (e[CO_2]) consistently show that growth increases with e[CO_2], but the estimated responses vary[37,38]. Comparisons are difficult due to differences in G, E_t, and M_t among experiments and the methods used to induce e[CO_2], notably the use of enclosed or Open-Topped Chambers (OTCs) vs. Free-Air CO_2 Enrichment (FACE). Kimball et al. [39]. listed twelve ways that the environment inside OTCs can

Table 1 | Examples of how the terms "repeatability," "replicability" and "reproducibility" appear in agricultural research

Topic	Reference	Term	Usage
Characterizing cereal genotypes with imagery	Svensgaard et al. [82]	Repeatability	The "variation associated with the measurement itself".
		Reproducibility	The "impact of different ways of measuring or different measuring conditions".
Quantitative genetics	Visscher et al. [83]	Repeatability	The variation within individuals that is attributable to measurement errors and other random environmental factors.
Agronomic traits in perennial forages	Braz et al. [84]	Repeatability	The degree of correlation among repeated measurements on a single individual.
Field experiments	Lockeretz[85]	Replicability	The ability of responses reported in one experiment to be duplicated either in time or space by the same researchers.
Nitrogen test strips for wheat	Roberts et al. [86]	Replicability	The ability of responses from one on-farm field test to be duplicated in space or time.
Simulation of pesticide leaching	Dubus and Janssen[87]	Replicability	The similarity of responses for simulations from a single model and set of parameter distributions.
On-farm trials	Kool et al. [88]	Reproducibility	The degree to which yield-limiting and yield-reducing factors are described in sufficient detail to allow the experimental conditions to be reproduced.
Crop modeling	Tatsumi, 2016[89]	Reproducibility	The ability of the model to reproduce values reported from field measurements.
Crop modeling	Porter et al. [33]	Reproducibility	The ability to duplicate simulations by fully documenting data, model versions, and other relevant information.
Yield-gap estimation	Van Ittersum et al. [90]	Reproducibility	The ability of a described workflow to be fully recreated by other researchers.

Quoted text for usage indicates exact phrasing from the source publication. Otherwise, the text is our interpretation of usage from the source.

Table 2 | Proposed terminology for confirmation of results on agricultural research on climate change, and the equivalent term from the National Academies of Sciences, Engineering, and Medicine (NASEM) report[6]

Term	Proposed definition for agricultural research	Equivalent in NASEM report
Repeatability	The ability to obtain essentially identical results when an experiment or analysis is repeated under identical conditions using the same instruments, data processing workflows, and modeling or numerical analyses. Confirmation is conducted by the original research group.	Considered only as "measurement repeatability" in relation to when a measurement is repeated by the same operator using the same instrument under constant conditions and close in time. (p. 48)
Replicability	The ability to obtain essentially identical results when an experiment or analysis is repeated under similar conditions as the original study, typically using the same type of instruments, data processing workflows, and modeling or analyses. Confirmation is conducted by the original research group.	Obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data. (p. 46)
Reproducibility	The ability to obtain results from independent studies that are similar enough that they are considered to confirm the findings of previous experiments or analyses. This usually implies somewhat similar growing conditions, treatments, measurements, data processing workflows, and modeling or analyses. In modeling and numerical analyses, this may involve two approaches: using the original dataset but different software or using both different data and software. Conducted by an independent research group.	Obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis. This definition is synonymous with "computational reproducibility," and the terms are used interchangeably in this report. (p. 46)

Page numbers refer to the location of definitions in the NASEM report.

differ from outside. From the literature, they calculated an average increase in growth inside OTCs of 10% at ambient [CO_2]. Such growth increases could be amplified by e[CO_2], leading to greater growth during the exponential phase of crop growth, further enhancing growth responses in OTCs compared to FACE, as first reported by Long and co-workers[10]. In contrast, a 2020 review[12] concluded that fluctuations in e[CO_2], which are prominent in FACE systems, reduce assimilation and growth compared to steady-state e[CO_2]. Hence, responses measured with FACE may be too low, while those from OTCs may be high. Building consensus remains difficult in the absence of studies directly comparing methods for elevating [CO_2].

Confirmation of Crop Simulations

We consider first process-based crop simulation models as these are the most widely used models in sustainable agriculture. Confirmation of models can involve three facets. The simplest concerns the consistency of numerical results: if identical numerical inputs are processed with the original model version and computational environment, outputs should be identical to the originals[40]. The second concerns the confirmation of the mathematical representations of the biophysical processes embodied in each model, whether examined as component processes or complete models, recognizing that model developers differ in their approaches or hypotheses underlying how processes are represented. This is essentially model evaluation, typically comparing simulations of one or more models with process-specific experimental data and conducting sensitivity analyses[41,42]. The third facet is the confirmation of results from model applications under different assumptions, which for sustainability usually involves simulating long-term crop rotations or sequences, potentially varying climatic factors or [CO_2] to mimic climate change. Model confirmation in applications may involve comparisons with field experiments, historical production records, or outputs from other models[43].

Confirming model outputs can involve the three levels outlined in Table 2. In theory, for any level, the process only requires re-running the specified model with the associated datasets. However, obtaining identical results often proves difficult, especially when independent parties attempt to reproduce results, even using the same model. Comparing 455 models of biological processes expressed using Systems Biology Markup Language (SBML), Tiwari and co-workers were unable to reproduce results from half of the models[44].

Factors constraining the confirmation of modeling studies include differences in inputs, parameters, and model versions, the use of stochastic processes such as weather generation, difficulties in interpreting code, and software dependencies (Table 3). Minor differences arise simply from compiling model code under different operating systems, language versions, and compiler settings[40].

Determining how accurately biological, chemical, and physical processes are embodied in crop models is widely discussed as model evaluation[45,46]. Foremost among factors constraining model evaluation are the scarcity of detailed field data and uncertainties in inputs such as genotype-specific traits, soil physical parameters, and initial conditions[47]. Furthermore, field data and simulation outputs are often mismatched due to differences between variables measured in the field vs. what are modeled as state variables[48]. Failure to consider a lack of independence among observed data, typically involving differences among treatments or environments, can inflate apparent model accuracy[49]. Additionally, while field measurements usually describe "realized" crop growth where multiple abiotic and biotic factors limit growth, crop models simulate greater, "attainable" growth constrained by explicitly modeled effects such as water or nutrient deficits or non-optimal temperatures. Thus, simulated growth usually represents an upper boundary for measured values. Finally, crop phenotypes are emergent properties that arise from interacting processes within a complex system. Crop models often include parameters in process equations whose values are estimated because they are difficult to measure directly (e.g., for gene effects). A crop model thus represents an abductive learning framework whereby only a subset of possible solutions is allowed, yet it is impossible to pinpoint a unique solution. This lack of unique solutions, termed equifinality[50], further compromises reproducibility.

AgMIP crop model intercomparisons have partially addressed reproducibility by comparing different models run using identical inputs[15,51]. The intercomparisons, however, have largely focused on how well models described crop growth or environmental effects in specific field trials, rather than on reproducibility among models. An implicit assumption may have been that the field data, including associated G_x, E_t, and M_t, had low uncertainty compared to differences among models. In an analysis of modeling datasets for 426 potato experiments, errors occurred in all elements of the inputs, parameters, and evaluation data[52]. Weather data appeared especially problematic, possibly because weather data are easier to cross-check as compared to soil, management, and crop growth data.

Confirmation of Other Numerical Models

Statistical and geospatial models are also used to investigate issues relating to sustainability, especially for climate change[53,54]. Again, while confirmation of such models seems straightforward, attempts to reproduce analyses from other disciplines have encountered difficulties[55]. In geosciences, Konkol and co-workers recreated analyses and resulting maps or graphs from 41 open-access papers[56]. Analyses from two papers were reproduced without issues and for 33, issues were readily resolved. For two papers, issues were partially resolved, but four papers were considered irreproducible. Similar studies from psychology found that published values frequently could only be reproduced after author consultation[57,58]. Difficulties involved how analytic procedures were reported, and the primary research conclusions were unaffected.

The underlying causes of failure to reproduce numerical analyses parallel those for crop modeling (Table 3). Data may inadvertently be modified over time, such as in the handling of outliers or missing values or by values from databases being updated. Workflows that include different software tools may require manual manipulation of files, increasing the potential for errors. Even when workflows are documented, problems arise. Two analyses of research using the open-source digital notebook Jupyter Notebooks (https://jupyter.org/, verified 2025-02-12) found that results were often unreproducible because the actual computation steps ("cells") differed from the described order[59].

Towards Improved Reproducibility and Confirmation

Strengthening the confirmation of sustainability research requires a substantial shift in research culture, including changes in the attitudes of individuals, teams, and funding agencies[6,60,61]. Researchers must recognize the importance of thoroughly and accurately describing their experiments and analyses, and sharing their data, data collection and analysis protocols, and software in ways that enable reproduction. The planning of field experiments, modeling, and any subsequent numerical analyses should seek to maximize repeatability, replicability, and reproducibility (Table 2). Guided by a project that assessed the reproducibility of computer code[62], we suggest researchers consider whether their work could be reproduced ten

Table 3 | Examples of potential causes of failure to confirm results from simulation modeling or other numerical models

	Potential cause
Data	Outliers or missing data are processed using varying criteria.
	Values are retrieved from a database where records have been altered or are updated continuously.
	Values in parameter files differ due to local edits or versions.
	The user inadvertently accesses different parameter files due to errors in specifying file locations.
	Inputs or parameters have large uncertainties such as due to insufficient data for model calibration.
	Dataset creation involves stochastic effects such as in generating long-term weather series.
Software	Different versions of software are used.
	Executable files of software differ due to differences in compilers, including compiler options and performance with different processors.
	Processing involves stochastic methods such as Monte Carlo integration or genetic algorithms.
	Dependencies such as of packages and libraries that are used by the software evolve and are not fully backward-compatible.
Scripts or code	Different versions of scripts or code are used, especially as coding practices evolve.
	Script or code is written in a manner that is difficult to understand, leading to errors in attempts to duplicate the results.
	Portions of the script or code are inadvertently executed in a different order than in the original study.
	Dependencies such as of packages and libraries that are used by scripts or code evolve and are not fully backward-compatible.
Workflows	Different versions of workflows are used.
	Portions of the workflow are inadvertently executed in a different order than was intended.
	Workflow requires the use of incompatible tools that require manual conversion or transfer of files.
	Workflow is implemented using different programming languages and executed in different processors.

years from now. Our admittedly subjective rationale is that if agricultural research can be reproduced after a decade, it should be reproducible over a longer period, acknowledging the inherent uncertainties in instruments, germplasm, environments, and other factors.

A recurring recommendation, formulated in various manners, is to follow "best" practices for data management and processing that enhance reproducibility[63,64]. For agriculture, key practices for researchers are outlined in Table 4. Publishers might create a certification process that assesses completeness, nomenclature and formatting, including materials, methods, datasets, and software. Researchers in ecology and evolution proposed eight review criteria, including how well a manuscript describes metadata, data processing steps, and sources of secondary data[65]. The journal PLOS Computational Biology implemented a pilot system for peer review of reproducibility[66]. Ideally, compliance would be tested prior to submission, using tools similar to turnitin (https://turnitin.com, verified 2025-02-12) or iThenticate (https://www.ithenticate.com/, verified 2025-02-12).

Researchers may resist change out of concern that it will divert resources from advancing their objectives[7,67,68]. However, changes can enhance research impact, reduce errors, discourage unfounded challenges, and improve compliance with open science directives. A balance must be struck between providing too little data, rendering studies irreproducible, and requiring so much documentation that research suffers.

We suggest resource concerns may be overstated. Valuable information is often available but unreported simply because its importance is unrecognized (e.g., row spacing, sowing depth, fertilizer composition). Similarly, some data are not collected because their perceived value does not justify the cost. For instance, soil samples are often taken before planting but limited to the upper 30 cm. Extending sampling to the maximum rooting depth provides a more complete nutrient balance at minimal extra cost since sampling sites are already established.

Meta-research examining how published research is evolving in terms of replication and reproducibility might identify constraints and needs[69]. A 2011 analysis of methodologies for simulating climate change impacts helped strengthen subsequent studies, although reproducibility was not explicitly addressed[70].

Field research

The first step for field experiments is to improve reporting of F_0, G, E_t, and M_t. Adequately describing the weather, soil profile characteristics, and crop management is essential. Often, data on management exist but are not in an organized digital format. The ICASA standards provide one option for documenting the field environment and crop management[30,31].

Given that sustainability research often concerns quantitative responses of crops or soils to nutrients, temperature, precipitation, $[CO_2]$, and other abiotic factors, the question arises of whether studies can measure responses more accurately, thus strengthening confirmation, especially considering interactions of G, E_t, and M_t. For experiments involving multiple quantitative factors, response surface methods can capture nonlinear responses with a reduced number of plots, while maintaining statistical power[71]. Studies combining $e[CO_2]$ with other factors have predominantly used only two levels per factor, sufficient to detect an interaction with $[CO_2]$ but insufficient to infer the shape of the responses[71].

Confirmation also benefits from coordination among trials to standardize elements of G, E_t, and M_t or measurement protocols. The "China Wheat" study partially standardized wheat (*Triticum aestivum* L.) cultivars, nitrogen and water regimes across five locations from Texas, USA to Alberta, Canada, and used coordinated protocols for growth, spectral reflectance, and canopy temperature measurements[72]. GRACEnet (Greenhouse gas Reduction through Agricultural Carbon Enhancement network)[73] and the Long Term Agroecosystem Research (LTAR) network[74] specifically address sustainability.

Crop modeling and other numerical models

Multiple actions to enhance reproducibility of simulation modeling and other numerical approaches merit consideration. As far as possible, modeling per se and associated analyses should employ peer-reviewed, open-source software. The open-source framework Crop2ML allows interchanging modules among models, which should enhance reproducibility[75,76], although Crop2ML still requires comparisons with external data to identify the most promising approaches. Model inputs, parameters, and control scripts should be placed in public repositories. If the model or analytic software is not open source, then the equations and parameters should be reported in detail. Version control systems such as Git, along with the cloud-based GitHub repository, can assist researchers in tracking model development, and code can also be shared as appendices to journal articles, research web sites, or model repositories such as the CoMSES Net Model Library (https://www.comses.net/)[77].

Crop model intercomparisons have strengthened model-based research for climate change impacts but also highlighted challenges in improving models per se, designing simulation experiments, and analysis of modeling studies. A major constraint remains the scarcity of datasets combining a range of treatments or environmental conditions with adequate information on soils and crop management. Furthermore, data on crop growth and development often constrain accurate model parameterization. There are numerous calls to follow the FAIR Data Principles of datasets being Findable, Accessible, Interoperable, and Reusable[78], and funding sources increasingly require datasets to be released in digital formats. However, datasets in repositories such as the USDA Ag Data Commons (https://agdatacommons.nal.usda.gov/browse; verified 2025-02-12) frequently lack data describing E_t, and M_t.

Coordinated field experiments carefully designed to fill confirmation gaps in model evaluation and application, and that follow adequate protocols for data collection and sharing, are essential to reduce the current bottleneck in experimental data. Design of field trials and protocols would benefit from greater collaboration among experimentalists and crop modelers.

Table 4 | Recommended actions to strengthen the reproducibility of data management and analysis in agricultural research

Action
1. Identify the types and amounts of data to be recorded, including details of crop management, pre-plant soil conditions, and weather using the ICASA or similar standards as a guide for recommended detail.
2. Plan the data management and analysis workflow at the onset of the research, recognizing that workflows can evolve as research progresses.
3. Plan explicitly for data quality control, especially for handling of missing data and outliers and for secondary sources of soil and weather data.
4. Commit to protocols for naming datafiles, variables, scripts, and other digital entities, with emphasis on readability.
5. Manage data using well-documented, open structures such as comma-separated variable format, Open Document Format, or JSON.
6. Conduct analyses within a single open-source programming environment such as R or Python.
7. Ensure the transparency of any model by providing documentation that includes equations, flow charts, and/or pseudocode.
8. Test the workflow frequently, starting from the original data whenever possible.
9. Use tools to encapsulate workflows such as RStudio projects and Jupyter Notebooks.

Model intercomparisons might investigate sources of uncertainty such as model inputs, including initial conditions, model structure, and model parameters[79]. In ecological modeling, researchers cited the need to "break" models, defined as determining "under what conditions the mechanisms represented in a model can no longer explain observed phenomena"[80]. This approach was embodied in the temperature-based sensitivity analyses used to evaluate modeled responses for sorghum (*Sorghum bicolor* (L.) Moench) and dry bean (*Phaseolus vulgaris* L.)[81].

Conclusion

Research related to sustainable agricultural production will face increasing scrutiny and pressure to quantify responses more accurately for factors including soil carbon and nutrient levels, air temperature, precipitation, [CO_2], water deficits, flooding, crop genetics, and biotic factors. Addressing these challenges requires sustained efforts by individual researchers, research groups, funding agencies, and others to strengthen the independent confirmation of scientific results. While the NASEM report increased the visibility of the confirmation process, NASEM terminology seems too narrow for agricultural research. We urge the use of the broader senses of repeatability, replicability, and reproducibility, and emphasize their relevance in field research, crop simulation modeling, and other numerical analyses.

Agricultural research should explicitly plan for reproducibility, recognizing that natural variability in the local environment (E_t) constrains reproducibility of field studies. A useful benchmark is whether the descriptions of experiments and analyses are detailed enough to enable reproduction of the research 10 years later. Further actions include strengthening the digital description of data and protocols, improving experimental designs and statistical analyses, and simulating crops growing under challenging, extreme conditions. These actions are crucial for research to accurately characterize the responses of agricultural systems to multiple challenges, especially in the context of sustainable production and climate uncertainty. Attaining the needed changes may require additional resources, but the benefits of improved confirmation should justify the investment.

Data availability

No datasets were generated or analysed during the current study.

Received: 19 September 2024; Accepted: 10 April 2025;
Published online: 22 May 2025

References

1. Ioannidis, J. P. A. Why most published research findings are false. *PLOS Med.* **2**, e124 (2005).
2. Baker, M. 1500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
3. Munafò, M. R. et al. A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 1–9 (2017).
4. Reinecke, R., Trautmann, T. & Wagener, T. The Critical Need to Foster Reproducibility in Computational Geoscience. https://meetingorganizer.copernicus.org/EGU22/EGU22-2139.html (2022) https://doi.org/10.5194/egusphere-egu22-2139.
5. Brainard, J. Fake scientific papers are alarmingly common. *Science* **380**, 568–569 (2023).
6. Committee on Reproducibility and Replicability in Science et al. Reproducibility and Replicability in Science. 25303 (National Academies Press, Washington, D.C., 2019). https://doi.org/10.17226/25303.
7. Baker, J. M., Ochsner, T. E., Venterea, R. T. & Griffis, T. J. Tillage and soil carbon sequestration—What do we really know? Agriculture. *Ecosyst. Environ.* **118**, 1–5 (2007).
8. Bai, X. et al. Responses of soil carbon sequestration to climate-smart agriculture practices: A meta-analysis. *Glob. Change Biol.* **25**, 2591–2606 (2019).
9. Nouri, A., Lukas, S., Singh, S., Singh, S. & Machado, S. When do cover crops reduce nitrate leaching? A global meta-analysis. *Glob. Change Biol.* **28**, 4736–4749 (2022).
10. Long, S. P., Ainsworth, E. A., Leakey, A. D. B., Nösberger, J. & Ort, D. R. Food for Thought: Lower-Than-Expected Crop Yield Stimulation with Rising CO2 Concentrations. *Science* **312**, 1918–1921 (2006).
11. Tubiello, F. N. et al. Crop response to elevated CO2 and world food supply. *Eur. J. Agron.* **26**, 215–223 (2007).
12. Allen, L. H. et al. Fluctuations of CO2 in Free-Air CO2 Enrichment (FACE) depress plant photosynthesis, growth, and yield. *Agric. For. Meteorol.* **284**, 107899 (2020).
13. Parent, B. & Tardieu, F. Temperature responses of developmental processes have not been affected by breeding in different ecological areas for 17 crop species. *N. Phytologist* **194**, 760–774 (2012).
14. White, J. W., Kimball, B. A., Wall, G. W. & Ottman, M. J. Cardinal temperatures for wheat leaf appearance as assessed from varied sowing dates and infrared warming. *Field Crops Res.* **137**, 213–220 (2012).
15. Kothari, K. et al. Are soybean models ready for climate change food impact assessments? *Eur. J. Agron.* **135**, 126482 (2022).
16. Grossiord, C. et al. Plant responses to rising vapor pressure deficit. *N. Phytologist* **226**, 1550–1566 (2020).
17. Oldfield, E. E. et al. Crediting agricultural soil carbon sequestration. *Science* **375**, 1222–1225 (2022).
18. Schumacher, D. L. et al. Exacerbated summer European warming not captured by climate models neglecting long-term aerosol changes. *Commun. Earth Environ.* **5**, 1–14 (2024).
19. Grimes, D. R., Bauch, C. T. & Ioannidis, J. P. A. Modelling science trustworthiness under publish or perish pressure. *R. Soc. Open Sci.* **5**, 171511 (2018).
20. Greenland, S. et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* **31**, 337–350 (2016).
21. Grudniewicz, A. et al. Predatory journals: no definition, no defence. *Nature* **576**, 210–212 (2019).
22. Ioannidis, J. P. A., Pezzullo, A. M. & Boccia, S. The Rapid Growth of Mega-Journals: Threats and Opportunities. *JAMA* **329**, 1253–1254 (2023).
23. Else, H. & Van Noorden, R. The fight against fake-paper factories that churn out sham science. *Nature* **591**, 516–519 (2021).
24. Lund, B. D. et al. ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *J. Assoc. Inf. Sci. Technol.* **74**, 570–581 (2023).
25. Heisey, P., SunLing, W. & Fuglie, K. Public agriculture research spending and future U.S. agricultural productivity growth: scenarios for 2010-2050. Economic Brief - Economic Research Service, United States Department of Agriculture (2011).
26. Loomis, R. S. Systems approaches for crop and pasture research. in Proc. 3rd Australian Agronomy Conference. Australian Soc. of Agronomy, Parkville, VIC 1–8 (1985).
27. Sheehy, J. E., Sinclair, T. R. & Cassman, K. G. Curiosities, nonsense, non-science and SRI. *Field Crops Res.* **91**, 355–356 (2005).
28. Kenett, R. S. & Shmueli, G. Clarifying the terminology that describes scientific reproducibility. *Nat. Methods* **12**, 699–699 (2015).
29. Plesser, H. E. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. *Front. Neuroinformat* **11**, 1–4 (2018).
30. Hunt, L. A., White, J. W. & Hoogenboom, G. Agronomic data: advances in documentation and protocols for exchange and use. *Agric. Syst.* **70**, 477–492 (2001).
31. White, J. W. et al. Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Computers Electron. Agriculture* **96**, 1–12 (2013).
32. Rosenzweig, C. et al. The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agric. For. Meteorol.* **170**, 166–182 (2013).

33. Porter, C. H. et al. Harmonization and translation of crop modeling data to ensure interoperability. *Environ. Model. Softw.* **62**, 495–508 (2014).
34. Nicotra, A. & McIntosh, E. PrometheusWiki: online protocols gaining momentum. *Funct. Plant Biol.* **38**, iii (2011).
35. Teytelman, L., Stoliartchouk, A., Kindler, L. & Hurwitz, B. L. Protocols.io: Virtual Communities for Protocol Development and Discussion. *PLOS Biol.* **14**, e1002538 (2016).
36. Stefan, A. M. & Schönbrodt, F. D. Big little lies: a compendium and simulation of p-hacking strategies. Royal Society Open. *Science* **10**, 220346 (2023).
37. Kimball, B. A. Crop responses to elevated CO_2 and interactions with H_2O, N, and temperature. *Curr. Opin. Plant Biol.* **31**, 36–43 (2016).
38. Toreti, A. et al. Narrowing uncertainties in the effects of elevated CO_2 on crops. *Nat. Food* **1**, 775–782 (2020).
39. Kimball, B. A. et al. Comparisons of responses of vegetation to elevated carbon dioxide in free-air and open-top chamber facilities. in Advances in Carbon Dioxide Effects Research 113–130 (John Wiley & Sons, Ltd, 1997). https://doi.org/10.2134/asaspecpub61.c5.
40. Thorp, K. R. et al. Methodology to evaluate the performance of simulation models for alternative compiler and operating system configurations. *Computers Electron. Agriculture* **81**, 62–71 (2012).
41. Deligios, P. A., Farci, R., Sulas, L., Hoogenboom, G. & Ledda, L. Predicting growth and yield of winter rapeseed in a Mediterranean environment: Model adaptation at a field scale. *Field Crops Res.* **144**, 100–112 (2013).
42. Thorp, K. R., Marek, G. W., DeJonge, K. C. & Evett, S. R. Comparison of evapotranspiration methods in the DSSAT Cropping System Model: II. Algorithm performance. *Computers Electron. Agriculture* **177**, 105679 (2020).
43. Liu, B. et al. Similar estimates of temperature impacts on global wheat yield by three independent methods. *Nat. Clim. Change* **6**, 1130–1136 (2016).
44. Tiwari, K. et al. Reproducibility in systems biology modelling. *Mol. Syst. Biol.* **17**, e9982 (2021).
45. Wallach, D., Makowski, D., Jones, J. W. & Brun, F. Working with Dynamic Crop Models: Methods, Tools and Examples for Agriculture and Environment. (Academic Press, 2018).
46. Gao, Y. et al. Evaluation of crop model prediction and uncertainty using Bayesian parameter estimation and Bayesian model averaging. *Agric. For. Meteorol.* **311**, 108686 (2021).
47. Luo, Q., Hoogenboom, G. & Yang, H. Uncertainties in assessing climate change impacts and adaptation options with wheat crop models. *Theor. Appl Climatol.* **149**, 805–816 (2022).
48. Kothari, K. et al. Evaluating differences among crop models in simulating soybean in-season growth. *Field Crops Res.* **309**, 109306 (2024).
49. White, J. W., Boote, K. J., Hoogenboom, G. & Jones, P. G. Regression-based evaluation of ecophysiological models. *Agron. J.* **99**, 419–427 (2007).
50. Lamsal, A., Welch, S. M., White, J. W., Thorp, K. R. & Bello, N. M. Estimating parametric phenotypes that determine anthesis date in Zea mays: Challenges in combining ecophysiological models with genetics. *PLOS ONE* **13**, e0195841 (2018).
51. Asseng, S. et al. Uncertainty in simulating wheat yields under climate change. *Nat. Clim. Change* **3**, 827–832 (2013).
52. Ojeda, J. J. et al. Assessing errors during simulation configuration in crop models – A global case study using APSIM-Potato. *Ecol. Model.* **458**, 109703 (2021).
53. Lobell, D. B. & Burke, M. B. On the use of statistical models to predict crop yield responses to climate change. *Agric. For. Meteorol.* **150**, 1443–1452 (2010).
54. Pugh, T. A. M. et al. Climate analogues suggest limited potential for intensification of production on current croplands under climate change. *Nat. Commun.* **7**, 12608 (2016).
55. Peng, R. D. Reproducible Research in Computational Science. *Science* **334**, 1226–1227 (2011).
56. Konkol, M., Kray, C. & Pfeiffer, M. Computational reproducibility in geoscientific papers: Insights from a series of studies with geoscientists and a reproduction study. *Int. J. Geographical Inf. Sci.* **33**, 408–429 (2019).
57. Hardwicke, T. E. et al. Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal Cognition. *R. Soc. Open Sci.* **5**, 180448 (2018).
58. Hardwicke, T. E. et al. Analytic reproducibility in articles receiving open data badges at the journal Psychological Science: an observational study. *R. Soc. Open Sci.* **8**, 201494 (2021).
59. Wang, J., Kuo, T., Li, L. & Zeller, A. Assessing and restoring reproducibility of Jupyter notebooks. in Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering 138–149 (Association for Computing Machinery, New York, NY, USA, 2021). https://doi.org/10.1145/3324884.3416585.
60. Landis, S. C. et al. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* **490**, 187–191 (2012).
61. Munafò, M. R., Chambers, C. D., Collins, A. M., Fortunato, L. & Macleod, M. R. Research culture and reproducibility. *Trends Cogn. Sci.* **24**, 91–93 (2020).
62. Perkel, J. M. Challenge to scientists: does your ten-year-old code still run? *Nature* **584**, 656–658 (2020).
63. Wilson, G. et al. Good enough practices in scientific computing. *PLOS Comput. Biol.* **13**, e1005510 (2017).
64. Leonelli, S., Davey, R. P., Arnaud, E., Parry, G. & Bastow, R. Data management and best practice for plant science. *Nat. Plants* **3**, 1–4 (2017).
65. Jenkins, G. B. et al. Reproducibility in ecology and evolution: Minimum standards for data and code. *Ecol. Evol.* **13**, e9961 (2023).
66. Papin, J. A., Gabhann, F. M., Sauro, H. M., Nickerson, D. & Rampadarath, A. Improving reproducibility in computational biology research. *PLOS Comput. Biol.* **16**, e1007881 (2020).
67. Williams, S. C., Farrell, S. L., Kerby, E. E. & Kocher, M. Agricultural researchers' attitudes toward open access and data sharing. Issues in Science and Technology Librarianship https://doi.org/10.29173/istl4 (2019).
68. Peterson, D. The replication crisis won't be solved with broad brushstrokes. *Nature* **594**, 151–151 (2021).
69. Hardwicke, T. E. et al. Calibrating the scientific ecosystem through meta-research. *Annu. Rev. Stat. Its Application* **7**, 11–37 (2020).
70. White, J. W., Hoogenboom, G., Kimball, B. A. & Wall, G. W. Methodologies for simulating impacts of climate change on crop production. *Field Crops Res.* **124**, 357–368 (2011).
71. Piepho, H.-P., Herndl, M., Pötsch, E. M. & Bahn, M. Designing an experiment with quantitative treatment factors to study the effects of climate change. *J. Agron. Crop Sci.* **203**, 584–592 (2017).
72. Reginato, R. J. et al. Winter wheat response to water and nitrogen in the North American Great Plains. *Agric. For. Meteorol.* **44**, 105–116 (1988).
73. Del Grosso, S. J. et al. Introducing the GRACEnet/REAP Data Contribution, Discovery, and Retrieval System. *J. Environ. Qual.* **42**, 1274–1280 (2013).
74. Goodrich, D. C. et al. Long term agroecosystem research experimental watershed network. *Hydrol. Process.* **36**, e14534 (2022).
75. Midingoyi, C. A. et al. Crop2ML: An open-source multi-language modeling framework for the exchange and reuse of crop model components. *Environ. Model. Softw.* **142**, 105055 (2021).
76. Midingoyi, C. A. et al. Crop modeling frameworks interoperability through bidirectional source code transformation. *Environ. Model. Softw.* **168**, 105790 (2023).
77. Janssen, M. A., Pritchard, C. & Lee, A. On code sharing and model documentation of published individual and agent-based models. *Environ. Model. Softw.* **134**, 104873 (2020).

78. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
79. Tao, F. et al. Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments. *Glob. Change Biol.* **24**, 1291–1307 (2018).
80. Thiele, J. C. & Grimm, V. Replicating and breaking models: good for you and good for ecology. *Oikos* **124**, 691–696 (2015).
81. White, J. W., Hoogenboom, G. & Hunt, L. A. A structured procedure for assessing how crop models respond to temperature. *Agron. J.* **97**, 426–439 (2005).
82. Svensgaard, J. et al. Can reproducible comparisons of cereal genotypes be generated in field experiments based on UAV imagery using RGB cameras? *Eur. J. Agron.* **106**, 49–57 (2019).
83. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era — concepts and misconceptions. *Nat. Rev. Genet* **9**, 255–266 (2008).
84. Braz, T. G. S., Fonseca, D. M., Jank, L., Cruz, C. D. & Martuscello, J. A. Repeatability of agronomic traits in Panicum maximum (Jacq.) hybrids. *Genet. Molecualr Res.* **14**, 19282–19294 (2015).
85. Lockeretz, W. Replicability in agricultural field experiments. *Am. J. Alternative Agriculture* **8**, 50–50 (1993).
86. Roberts, D. C., Brorsen, B. W., Taylor, R. K., Solie, J. B. & Raun, W. R. Replicability of nitrogen recommendations from ramped calibration strips in winter wheat. *Precis. Agric* **12**, 653–665 (2011).
87. Dubus, I. G. & Janssen, P. H. M. Issues of replicability in Monte Carlo modeling: A case study with a pesticide leaching model. *Environ. Toxicol. Chem.* **22**, 3081–3087 (2003).
88. Kool, H., Andersson, J. A. & Giller, K. E. Reproducibility and external validity of on-farm experimental research in Africa. *Exp. Agriculture* **56**, 587–607 (2020).
89. Tatsumi, K. Effects of automatic multi-objective optimization of crop models on corn yield reproducibility in the U.S.A. *Ecol. Model.* **322**, 124–137 (2016).
90. van Ittersum, M. K. et al. Yield gap analysis with local to global relevance—A review. *Field Crops Res.* **143**, 4–17 (2013).

Acknowledgements
The authors thank Carlos D. Messina for input on issues related to simulation modeling. Mention of trade names or commercial products in this publication was solely for the purpose of providing specific information and does not imply recommendation or endorsement by the University of Florida, the University of Kentucky, or USDA. The University of Florida, the University of Kentucky, and USDA are equal opportunity providers and employers.

Author contributions
J.W. and G.H. identified the initial concern and wrote the draft manuscript. K.B., B.K., C.P., M.S., V.S., and K.T. contributed content related to their expertise and provided feedback on organization of the manuscript. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Jeffrey W. White.