Received: 6 April 2023 | Accepted: 21 June 2023

DOI: 10.1111/gcb.16896

Global Change Biology WILEY

REVIEW

The challenge of selecting an appropriate soil organic carbon simulation model: A comprehensive global review and validation assessment

Agata Garsia | Antoine Moinet | Carmen Vazquez | Rachel E. Creamer | Gabriel Y. K. Moinet

Soil Biology Group, Wageningen University, Wageningen, The Netherlands

Correspondence
Carmen Vazquez, Soil Biology Group, Wageningen University, Wageningen, The Netherlands.
Email: carmen.vazquezmartin@wur.nl

Abstract

Promotion of soil organic carbon (SOC) sequestration as a potential solution to support climate change mitigation as well as more sustainable farming systems is rising steeply. As a result, voluntary carbon markets are rapidly expanding in which farmers get paid per tons of carbon dioxide sequestered. This market relies on protocols using simulation models to certify that increases in SOC stocks do indeed occur and generate tradable carbon credits. This puts tremendous pressure on SOC simulation models, which are now expected to provide the foundation for a reliable global carbon credit generation system. There exist an incredibly large number SOC simulation models which vary considerably in their applicability and sensitivity. This confronts practitioners and certificate providers with the critical challenge of selecting the models that are appropriate to the specific conditions in which they will be applied. Model validation and the context of said validation define the boundaries of applicability of the model, and are critical therefore to model selection. To date, however, guidelines for model selection are lacking. In this review, we present a comprehensive review of existing SOC models and a classification of their validation contexts. We found that most models are not validated (71%), and out of those validated, validation contexts are overall limited. Validation studies so far largely focus on the global north. Therefore, countries of the global south, the least emitting countries that are already facing the most drastic consequences of climate change, are the most poorly supported. In addition, we found a general lack of clear reporting, numerous flaws in model performance evaluation, and a poor overall coverage of land use types across countries and pedoclimatic conditions. We conclude that, to date, SOC simulation does not represent an adequate tool for globally ensuring effectiveness of SOC sequestration effort and ensuring reliable carbon crediting.

KEYWORDS

carbon credit, carbon market, carbon sequestration, model, simulation, soil organic carbon

1 | INTRODUCTION

Soils are recognized as the second most essential resource to human life on the planet, after water (van Leeuwen et al., 2017). Soils are multifunctional and are crucial in supporting a broad range of ecosystem services, such as food, fibre and energy production, climate regulation, nutrient cycling and biodiversity protection (Kopittke et al., 2022). Regrettably, anthropogenic activities, and particularly agriculture, have degraded one third of soils globally, threatening the welfare of 3.2 billion people (Shukla et al., 2019). Indeed, the green revolution has allowed food production to sharply increase worldwide from the middle of the 20th century, with a more than threefold increase in the last 50 years (www.fao.org/faostat). However, the focus on this single ecosystem service (food production) has come at the expense of others, jeopardizing soil multifunctionality and making agriculture the single largest driver of environmental change (Tilman et al., 2001).

Soil organic carbon (SOC) has long been considered to be the most important indicator of overall soil functioning (Bünemann et al., 2018) and is increasingly argued to be the pivotal element around which all soil functions revolve (Kopittke et al., 2022). Furthermore, soils represent a carbon store of global significance, containing more carbon than global vegetation and atmosphere combined, and representing the equivalent of approximately 160 times the current annual anthropogenic CO_2 emission rate (Friedlingstein et al., 2022). Based on these observations, it is commonly argued that a proportionally small increase in global carbon stocks would not only revert the soil degradation trend by rebuilding the central element of soil multifunctionality, but also contribute to mitigating climate change by removing significant concentrations of CO_2 from the atmosphere (IPCC, 2022; Lal, 2004).

As a result, increasing organic carbon stocks in agricultural soils (SOC sequestration) has become a central strategy to transition agriculture towards more sustainable management and meet the tremendous challenges of ensuring food security, by producing enough food for a growing global population, while mitigating climate change. However, criticism exists around the actual mitigation potential of soils and there are trade-offs between SOC sequestration and food production (Moinet et al., 2023). In addition, the adoption of greenhouse gas removal techniques can lead to deterrence or delay in emission reduction (a risk known as mitigation deterrence; McLaren, 2020). In spite of this, enthusiasm for SOC sequestration has never been higher, with numerous papers (Amelung et al., 2020; Chabbi et al., 2017; Derrien et al., 2023), and international initiatives (The '4p1000 initiative' launched during the COP21 in 2015, the Koronivia workshops held during the COP23 in 2018) and reports (FAO, 2017, 2019) promoting the role of soil carbon sequestration in climate change mitigation.

Matching this societal demand, public and private sectors are seizing the opportunity to promote soil management for SOC sequestration by paying farmers per tons of CO_2 sequestered (or not emitted) in initiatives often referred to as 'carbon farming' (Paul et al., 2023). These initiatives allow farmers to register their fields with commercial certificate providers who certify the increase in SOC stocks. These certificates can then be sold as voluntary emission offsets on carbon markets, with schemes already existing at least in Europe, the United States and Australia (Paul et al., 2023).

Despite the lack of overall regulation and variation in overall methodologies and protocols (Oldfield et al., 2022), current standards for soil carbon certification, such as The GoldStandard Foundation's 'Soil organic carbon framework methodology' (2020), and the 'Methodology for improved agricultural land management' approved by Verra (2020), and the European 'Framework for carbon removals' (European Commission, 2021) agree on a few general principles. The first and perhaps the most important principle is quantification. Accurately quantifying SOC changes over time (e.g. monitoring SOC) after a change in land management practices is indeed a condition without which the achievement of the desired outcome (SOC sequestration) cannot be guaranteed. As summarized by Paul et al. (2023), three options exist for monitoring of SOC stocks, including: (1) direct measurements, (2) remote-sensing and (3) simulation models. Direct measurements are the most reliable method but are costly and time consuming for large areas. Remote-sensing is a potentially cost-effective way to monitor SOC over large areas in the topsoil, but precise estimations require specific conditions (bare soil conditions, low water content, uniform SOC content through the plough layer) and no studies to date have successfully detected SOC changes at the field scale (Paul et al., 2023). While simulation models are the cheapest and the most readily available option, they can result in large uncertainties if not appropriately applied. Yet, models are already in use by a number of certificate providers and in the most publicly available protocols (Oldfield et al., 2022).

Despite large uncertainties due to the complexity of the soil system, models simulating SOC stocks and their temporal changes (thereafter referred to as SOC models) have been described as the only economic SOC monitoring option (Campbell & Paustian, 2015; Lugato et al., 2021). The large number of existing SOC models (87 were reported by Campbell & Paustian, 2015) testifies to both their importance and the complexity of modelling SOC stocks. Different models are suitable for different contexts (i.e. combination of land use and management practices, climate, soil texture, temporal and spatial scales), have different constraints or needs (data availability, objective of the simulation) and may be formulated using different types of processes, each with a specific number of parameters and input data requirements (Campbell & Paustian, 2015; Manzoni & Porporato, 2009; Smith et al., 1997).

Campbell and Paustian (2015) presented a comprehensive overview of SOC models, describing their application domain, scale of use and the type of processes they simulate. This review was critical in bringing SOC model development nearer its scientific and policy applications. More recently, Le Noe et al. (2023) published a review that provides information regarding decomposition kinetics, scope and validation procedures of a selection of SOC models, and shows their evolution across the last 90 years. Despite this effort, the large number of existing models and model versions today, which vary considerably in their applicability and sensitivity, still confronts practitioners and certificate providers with the critical challenge of selecting the right model, appropriate to the specific conditions in which they are working. With respect to this, model validation (the process of assessing the prediction accuracy of a model by comparing simulation results with data obtained by observation and measurement of the real system (Le Noe et al., 2023) and the environmental and land use context in which it has been performed are particularly critical, as they define the boundaries of applicability of the model.

Promoting global SOC sequestration as a carbon removal technology to mitigate climate change poses ethical challenges due to potential trade-offs with other sustainability objectives (Moinet et al., 2023) and mitigation deterrence mechanisms which jeopardize efforts to cut emissions (McLaren, 2020). Notwithstanding, carbon markets for soil certificates are growing and the most recent authoritative assessments argue that climate goals will not be met without CO_2 removals (IPCC, 2022). Accurate model simulations of changes in SOC stocks that are site and management specific are critical now more than ever before. However, clear guidelines are urgently needed to enable educated choices of models and, therefore, fair and realistic allocation of carbon certificates where they are applied.

In this article, we present a comprehensive review of the most updated versions of existing SOC models and a classification of their validation context. Our aim is to facilitate an appropriate choice of models for simulating SOC stocks and their changes over time. The complexity of the task, as described below, prevented the full development of complete selection guidelines, but our work forms a basis for SOC model selection and is a necessary step towards the establishment of a standardized and reliable method for estimating SOC stocks and their changes.

2 | METHODS

To provide a comprehensive list of existing models simulating SOC stocks or SOC changes and their validation context, we developed a search methodology following three main steps, as described in the following subsections. Due to the complexity of model history and the inconsistent reporting of information, the search method evolved over time, which resulted in a complex search methodology requiring a range of search strings used in different search engines and correspondence with authors from relevant papers. The whole process is summarized in Figure 1.

Briefly here, we started by creating a list of named models published between 1933 and March 2022 that had been used to estimate SOC dynamics (Section 2.1). Second, as models often evolve over time in sometimes complex ways (Figure 2), we investigated the history of each model development and identified all relevant model entities for further research (Section 2.2). Third, we evaluated whether these model entities had been validated following a list of criteria for adequate modelling validation, and then searched for a range of information defining the validation context, such as the countries, range of land use types and soil textures in which the model entities had been successfully validated, as well as the availability of the software or code for the model (Section 2.3).

2.1 | Model search and selection

The model search and selection started with the list of models compiled by Campbell and Paustian (2015) which reviewed models published between 1933 and 2009 (and reviewed earlier by Falloon

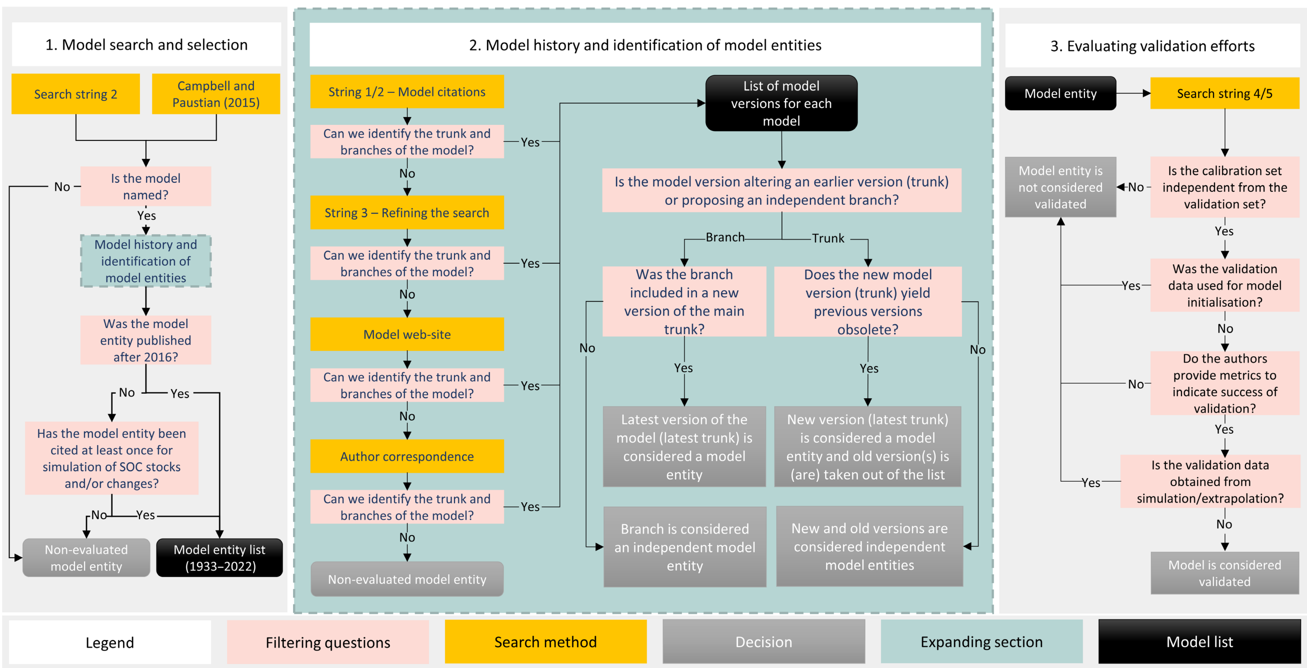

FIGURE 1 Search and identification framework. String searches were mainly input into Web of Science Core Collection.

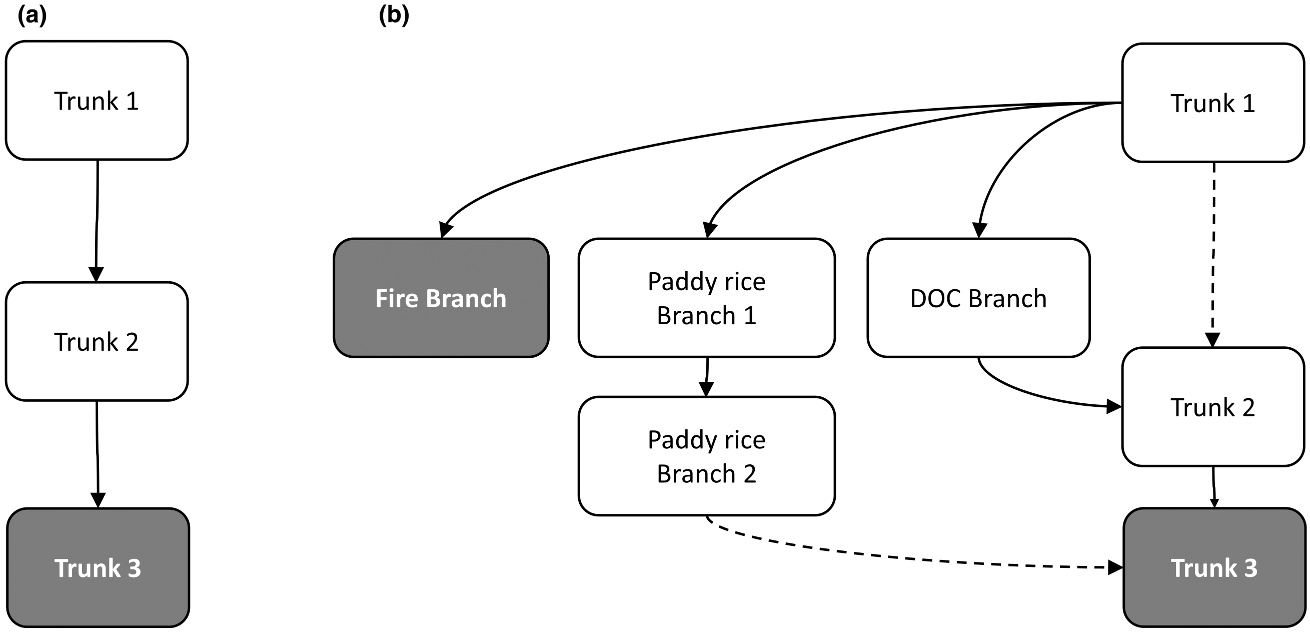

FIGURE 2 Hypothetical examples of (a) linear development history and (b) complex development history of soil organic carbon models. Each box represents one model version. Grey boxes represent model entities. Complex development histories often include several branches, some of which have not been included in updates to the main model or model trunk.

& Smith, 2000; Manzoni & Porporato, 2009; Stockmann et al., 2013). Campbell and Paustian (2015) identified 221 models and narrowed their review list by selecting only 'named' models in which SOC stocks were explicitly included in the model's first version formulation, and whose name had been cited in the title, key word or abstract of a scientific publication.

We drew inspiration from Campbell and Paustian's (2015) approach, but our approach differed in that we also included models whose updated versions (e.g. forms of the same models with a change in the equations and thus in the structure of the model) simulate SOC dynamics, even though the first version formulation does not. For each of them, we verified if there was at least one study, different from the publication paper of the model version indicated by Campbell and Paustian (2015), which adopted it for estimating SOC stocks and/or their changes. We did this by applying the search string 1 (Table S1) on Web of Science Core Collection (WoS-CC). We reviewed titles and abstracts of the resulting papers and kept only the model versions that had been used for SOC dynamic simulation and their publication paper for further analysis.

We also extended the search to include those models that were cited and applied for the simulation of SOC dynamics after the publication of Campbell and Paustian (2015). For this, we identified SOC models published or applied at least once between 2009 and March 2022 using search string 2 (Table S1). This second search string resulted in a comprehensive list of 1180 publications citing models and model versions. This list was further narrowed down through a review of the title and abstract: articles mentioning models that did not simulate SOC dynamics or models without a name were excluded from any further analyses. When the model's name or the model version referred to in the article was not in the list provided by Campbell and Paustian (2015), it was added to the list of models to further analyse, together with its related article(s).

2.2 | Model history and identification of model entities

Model development is rarely linear, with new model versions being developed as updates to the main models (defining the 'trunk' of a model history tree, Figure 2) and 'branching' versions representing modifications of the trunk main or updated versions and may become new models or merge back to the main ones (Figure 2). Therefore, following identification of relevant models or model versions as described in Section 2.1, we set out to identify all relevant model entities, that is, the latest version for each branch and trunk of the model tree (Figure 2). This required that we identified the first model version (trunk 1), new versions of the model and whether these versions had been included or not in any model updates (Figure 2). The publications presenting the first version of each model in our list were either provided by Campbell and Paustian (2015) or were retrieved by an in-text reference search within the selected articles associated with the listed models all the way back to the original publication. For each initial model, we tracked model development by identifying all citations for the paper presenting the model, and searching within them those articles that mention the model's name and SOC as specified in search string 3 (Table S1). The resulting list was explored to identify branches of the model, updates to the model and whether branches had been included or not on those updates (Figure 2). If we were unable to identify the different model versions from the literature, we searched for them on the potential official model website (searching for the model's name together with the word soil carbon in Google Scholar as specified in search string 5 of Table S1). Outdated model versions were excluded from further analyses. Model ensembles, that is, combinations of two or more model versions whose output is aggregated, for example, by averaging, to obtain a better predictive performance than the one of the single models (Kotu & Deshpande, 2015), were also considered as model entities. Finally, we further selected the model entities by looking at the date of publication and at their further use for SOC dynamic simulation. If the model entity was published before 2016, we verified that there was at least one study, different from the model publication paper, which adopted that model entity for estimating SOC stock and/or their changes. To do so we used the option 'refining the search' in WoS-CC and applied string 3 (Table S1) to search within the list of papers citing the publication of each model entity. Moreover, we selected all model entities published after 2016, as the absence of citation in this case could reflect the young age of the publication rather than a lack of interest. Exceptions to this method of model entity identification are reported in Supporting Information, Methods.

2.3 | Identifying whether and in which conditions models were validated

To identify the model validations of each single model entity, we checked within the model entity publication (or reference article) if a validation was performed for SOC stocks and/or their temporal changes. We then used WoS-CC to search for papers citing the model publication, and within those, we used the option 'refining the search' to identify papers that may have evaluated model fit using search string 4 (Table S1). Reference articles of some model entities are not indexed in WoS-CC. Thus, the search of validation papers for such models was performed on Google Scholar using search string 5 (Table S1). Additionally, when evaluating the literature throughout the entire process, we followed up on claims of model validations using in-text citations. Once the validation articles were obtained, we inspected the text to verify the correctness of the validation procedure, and we reported the relevant information related to each validation in terms of both validation context and availability of the code/software and its language. If no validation was found for a certain model entity, we searched for the validation of the immediately preceding model version.

Although we did not systematically list all articles citing models and model versions (including validation papers), we estimate the final list of reviewed papers to near 4000.

In the light of the IPCC guidelines (Shukla et al., 2019), we considered a model to be 'validated' under the following circumstances:

- There was a clear distinction between the calibration and validation data.
- Validation data were not used for model initialization.
- The authors provided metrics indicative of the goodness of fit of the model simulation when compared to the validation set, and the metrics were interpreted to be indicative of a successful validation by the authors. This led to the exclusion of validation efforts that reported only a visual comparison between the observed and simulated values. However, it is important to specify that we considered the simulated datapoints falling within the confidence interval of the measured data a sufficient statistical indicator of goodness of fit, but only if confidence intervals were obtained from measurement of replicates and not if they were calculated with the standard deviation of measurements covering different pedoclimatic conditions. Despite issues with the interpretation (see discussion), we also deemed sufficient the use of linear regression between observed and simulated data, but only if the coefficient of determination (R^2) or the correlation coefficient (r) were given together with a graph of the observed versus simulated data.
- Validation data were directly measured and were not the output from a different model nor the result of an interpolation, thus excluding gridded reconstructed databases, such as SoilGrids (Poggio et al., 2021), the Harmonized World Soil Database, or HWSD (Nachtergaele et al., 2012), the Northern Circumpolar Soil Carbon Database, or NCSCD (Hugelius et al., 2013) and the WISE30sec database (Batjes, 2015). This choice was made due to the uncertainty associated with the methods used to extrapolate information in these databases (Tifafi et al., 2018).

Then, for each model entity, we identified the environmental context of validation and identified the software and code availability.

Specifically, for the validation of each model entity we recorded the conditions regarding:

- Ecosystem type, including cropland (upland herbaceous crops, paddy rice cropland, orchard systems), agroforestry systems, grassland, forest and wetlands. A specific category was used for validations that were done after a land use conversion.
- Location, consisting of nation and possibly (for large nations such as the United States, China and India) sub-national region or province or state.
- Number of soil types or textures in which validation was applied. Due to the diversity of soil type and texture classifications, we report only whether the validation was done on one, two or more than two soil types or textures as identified by the authors.
- Soil depth.
- Temporal scale. The temporal scale of validation can be a single time estimate of SOC stock, or a SOC stock change over a certain period of time (the latter also referred to as 'diachronic validation' by Le Noe et al., 2023). The former occurs when the SOC stock of the model was initialized based on a steady-state assumption (spin-up) or with a value of SOC stock obtained from gridded databases and the simulated SOC stock is compared to one single measured value. The latter occurs when at least two measurements of the SOC stock are made at different times, and at least

the second one is compared with a simulated value. We also report the time step of the model used to perform the validation.
- Spatial scale, including field scale, regional scale (sub-national or cross-national regions), national scale and global scale.

Within the publication for each model entity, we also searched for information about the availability of the code/software and its language. If no clear information was found, we searched for it on the potential official model website (found by searching in Google the model's name together with the word 'soil carbon'). We reported this information when available and clear. If the information was not available, or if the link to the code indicated in the literature did not work, we report 'unclear' in the extended model classification table (Table S2).

3 | RESULTS

A total of 221 model entities (hereinafter also referred to as models) were identified. According to our method, only 64 (29%) of the selected models were validated at least once for single time estimates of SOC stock and/or its changes overtime (Figure 3). The list of validated model entities, together with summarized information related to the model validation status are presented in Table 1. An extended version of the classification with detailed information about model validation contexts and detailed bibliographic information is presented in Table S2. The list of selected model entities that were not validated according to the criteria in Section 2.3 is presented in Table S3. Due to the very large number of citations (942), references for both validated (Table 1; Table S2) and non-validated (Table S3) models are provided as Supporting Information, as well as brief description of the model development history (Table S4).

Those 64 validated models largely cover cropland (44 models), forests (26) and grassland (26) systems (note that the number of models does not sum to 64 as some models are validated for multiple land uses). Other land uses are severely under-represented compared to these three main land use types (Figure 3). Most of the 64 models (43), however, were validated for more than two soil types or textures (Figure 3). In some cases (2) the soil categorization was not clearly defined, which poses questions on the applicability of the models for SOC assessment.

Most strikingly, model validations are not evenly distributed across world regions, with Africa and the middle-east being extremely under-represented (Figure 4). Overall, with the exception of China, model validation is predominantly concentrated in North America, Western Europe and Australia.

Furthermore, we observed that about a third (21) of the models were validated only once, in a single country and for a unique land use type (Figure 5). In contrast, very few models were validated across multiple countries and land uses, and within those, even fewer were validated by multiple studies (Figure 5). For example, MILLENNIALv2 (Abramoff et al., 2022) was validated in 47 countries by a single study using a global dataset covering six land use types. Interestingly, the six land use types are relatively well distributed across countries, with several land use types validated in each country, except for Africa, where most countries have validations for only one land use type (except Senegal, Ivory Coast and Togo which have validations for three land use types; Figure S1). MEMS2 (Zhang et al., 2021) was also validated in a significant number of countries (27), but by only two studies and restricted to grasslands and forests. YASSO07 (Tuomi et al., 2009) is the model whose validations cover the greatest number of countries (49), covering seven land use types across nine studies. It is interesting to note that one of the validations for YASSO07 and the validation for MILLENNIALv2 were done using the same global dataset and together constitute 100% of validations found for many of the African countries. Finally, by far, RothC26.3 (Coleman & Jenkinson, 1996) is the most studied model, with 25 studies validating the model and covering 22 countries.

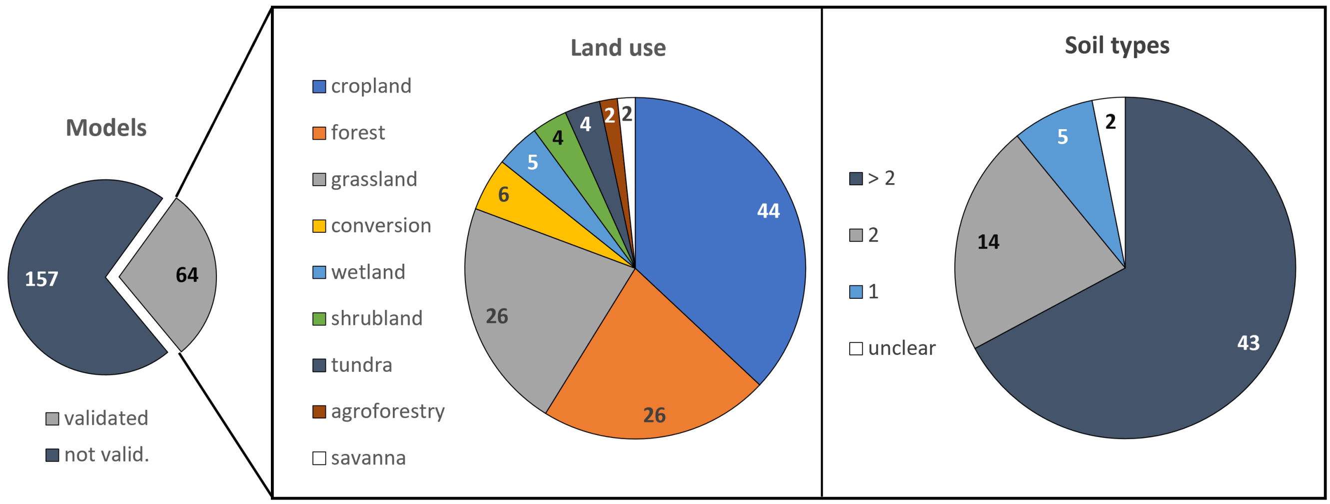

FIGURE 3 Pie charts summarizing some features of the models. From left to right: number of validated and not validated models, number of models validated for each land use (note that the number of models in this chart does not sum to 61 as some models are validated for multiple ecosystems), and number of models validated for one, two and more than two soil types.

TABLE 1 List of model entities validated according to our criteria listed in Section 2.3. For each model, information on the context of validation is provided on land use, countries and continent in which validation data were collected, the number of soil types/texture classes (1, 2 or larger than 2), the depth (cm), the timescale of validation (long: longitudinal; punctual: space for time substitutions with one time point set of measurements) and the number of studies having validated the model. A more complete version of this table including references for the model entity and for its validations is provided in Supporting Information (Table S2).

Model entity	Land use	Countries	Continent	Soils	Depth (cm)	Timescale	Studies
ICBM	Cropland	Sweden	Europe	1	15	Long	1
ICBM/2	Cropland, grassland, conversion	Sweden	Europe	1	20	Long	1
ICBMagroforestry	Agroforestry	Spain, Cameroon, Costa Rica, Sudan	Global	>2	Varying	Long	1
Qmodel(1)	Forest	Sweden	Europe	>2	Unclear	Punctual	1
SOCRATES	Cropland, grassland	United States of America, Canada, Australia, Germany, Sweden, United Kingdom, New Zealand	Global	>2	10	Long	1
ECOSYS	Cropland	Canada	North America	1	15	Long	2
NCSOIL	Cropland, grassland, forest	Germany, United States of America, United Kingdom, Australia, Czech Republic	Global	>2	30	Long	2
CERES-EGC	Cropland	France	Europe	1	30	Long	1
DAISY	Cropland, grassland	Germany, Australia, Czech Republic, United States of America	Global	>2	30	Long	3
DSSAT	Cropland	Canada, Brazil, United States of America, United Kingdom	Global	>2	40	Long	7
RothC26.3	Cropland, grassland, forest	Germany, United States of America, United Kingdom, Australia, Czech Republic, France, Switzerland, Ireland, Spain, Sweden, Canada, Mexico, Thailand, Hungary, Japan, China, India, Argentina, Syria, Iran, Hawaii, Italy	Global	>2	30	Long, punctual	25
Roth-CNP	Cropland, grassland	United Kingdom	Europe	2	23	Long	1
RothC10_N	Cropland, grassland	Italy, Spain, Australia, Syria, United Kingdom	Global	>2	30	Long	3
RothC26.5	Cropland	Italy, Spain, Australia, Syria, United Kingdom	Global	>2	Unclear	Long	1
SOMM	Cropland, grassland, forest	Germany, United States of America, United Kingdom, Australia, Czech Republic	Global	>2	50	Long	2
ROMUL	Forest	Finland	Europe	>2	100	Punctual	2
ROMUL_HUM	Forest, grassland	Russia	Europe	2	50	Long	1
DAYCENT	Cropland	United States of America, Bangladesh, France, UK, Canada, Germany, Brazil	Global	>2	Varying	Long, punctual	7
CANDY	Cropland, grassland, forest	Russia, Ukraine, Belarus, Czech Republic, Germany, United States of America, Australia, United Kingdom	Global	>2	30	Long	5
CNP	Cropland	Germany	Europe	2	30	Unclear	1
CCB	Cropland	Germany	Europe	2	30	Unclear	1
APSIM	Cropland	India, Australia, China	Global	>2	30	Long	4
DNDC9.5	Cropland, grassland	South Korea, India, China, Ireland, United Kingdom, Germany, Czech Republic, Australia	Global	>2	Varying	Long, punctual	6
DNDCv.CAN	Cropland	Canada	North America	2	20	Long	2
DNDCmicrobial	Cropland	United Kingdom	Europe	1	23	Long	1

(Continues)

TABLE 1 (Continued)

Model entity	Land use	Countries	Continent	Soils	Depth (cm)	Timescale	Studies
YASSO07	Forest, cropland, grassland, savanna, conversion, tundra, shrubland	Romania, Switzerland, Georgiou database	Global	>2	Varying	Long, punctual	9
YASSO15	Forest	Germany	Europe	>2	Varying	Long	1
LPJ-GUESS	Cropland	Kenya	Africa	1	150	Long	1
LPX-Bern	Wetland	Canada, United States of America (Alaska)	North America	>2	300	Punctual	1
Century	Conversion, cropland, forest, grassland, shrubland	Sweden, Argentina, United States of America, China, Hungary, United Kingdom, Germany, Australia, Czech Republic, Canada	Global	>2	Varying	Long, punctual	12
CN-SIM	Cropland	Brazil, Switzerland, Australia, United States of America	Global	>2	Unclear	Unclear	1
TEM v5.1	Grassland, forest, tundra	China	Asia	>2	100	Punctual	2
DVM-DOS-TEM	Forest, tundra	United States of America (Alaska)	North America	>2	90	Punctual	1
EPICv1102	Cropland, grassland, conversion	United States of America, Canada, Thailand, China, Italy, Cambodia, Kazakhstan, Argentina, Uruguay	Global	>2	Varying	Long, punctual	14
EPIC	Cropland	Colombia	South America	1	20	Long	1
APEX	Cropland	United States of America	North America	>2	15	Long	1
SPACSYS	Cropland	China	Asia	2	15	Long	2
AMG	Cropland	France	Europe	>2	25	Long	1
orchidee-SOM	Cropland, forest, grassland	France, Republic of Congo, Argentina	Global	>2	Varying	Punctual	2
FBDCAN	Forest	South Korea	Asia	Unclear	Unclear	Punctual	1
FBDC	Forest	South Korea, Turkey	Asia	Unclear	30	Punctual	2
agro-c	Cropland	China, Australia	Global	>2	30	Long	2
teco-R	Grassland	China	Asia	>2	50	Unclear	1
C-TOOL	Cropland	Denmark	Europe	1	25	Long	1
MILLENNIA	Wetland	United Kingdom	Europe	1	100	Punctual	1
ECOSSE	Cropland, forest, grassland, conversion	United Kingdom, Italy, Spain, Finland	Europe	>2	varying	Long, punctual	4
ARMOSA	Cropland	Italy, Kazakhstan, Finland	Global	>2	30	Long	1
carbosoil	Forest, wetland, cropland, agroforestry, shrubland	Italy, Egypt, Spain	Global	>2	75	Punctual	4
SoilGen2	Forest	Belgium, China	Global	>2	100	Punctual	2
Cqestr	Cropland, grassland	Brazil, United States of America, Italy, Canada	Global	>2	Varying	Long	5
MILLENNIAL V2	Forest, grassland, cropland, savanna, shrubland, conversion	Georgiou database	Global	>2	Varying	Punctual	1
MEMS 2	Grassland, forest	United States of America, Portugal, Spain, France, Italy, United Kingdom, Ireland, Belgium, Netherlands, Germany, Austria, Slovakia, Czech Republic, Greece, Poland, Estonia, Macedonia, Hungary, Romania, Bulgaria, Slovenia, Estonia, Latvia, Denmark, Lithuania, Sweden, Finland	Global	>2	Varying	Punctual	2

TABLE 1 (Continued)

Model entity	Land use	Countries	Continent	Soils	Depth (cm)	Timescale	Studies
InTEC	Forest, wetland	Canada	North America	>2	Varying	Punctual	4
N14CP	Forest, cropland, grassland	United Kingdom, Norway, Sweden, Germany, Switzerland, Denmark	Europe	>2	15	Long	1
Quincy v2.0	Forest	Germany	Europe	>2	100	Long	1
JSM	Forest	Germany	Europe	>2	100	Long	1
OC-VGEN	Cropland, grassland	Belgium	Europe	1	120	Punctual	1
Multi-model ensemble	Cropland	France, Russia, United Kingdom, Denmark, Sweden	Europe	>2	20	Long	1
Fun-BioCROP	Cropland	United States of America	North America	1	30	Long, punctual	1
Jules-DOCM	Forest, grassland	Germany, Belgium, China, Canada, Ireland	Global	>2	Varying	Long	1
Jules-Peat	Forest, tundra, grassland, wetland	Sweden, United Kingdom, Belgium, Canada, Ireland, Russia, United States of America, Germany, Norway, Poland, Finland, Greenland, Alaska, Svalbard	Global	>2	Varying	Punctual	1
RZWQM2	Cropland	United States of America	North America	1	30	Long	1
SWAT-C	Grassland	United States of America	North America	>2	150	Punctual	1
CAMPBELL	Cropland	Sweden	Europe	1	15	Long	1

However, these numerous validations for RothC26.3 are restricted to cropland, grassland and forest. Century-v4 (Metherell et al., 1993) and EPICv1102 (Izaurralde et al., 2012) follow RothC26.3 in terms of number of validations with 12 and 14 validation studies respectively. Century-v4 was validated for five land uses across 10 countries. However, reporting of land use type within countries of validation was inconsistent between studies, so it is not possible to know systematically whether each country has one or several land uses validated within them. Generally, inconsistent reporting made it impossible to produce maps such as the one for MILLENNIALv2 (Figure S1) detailing validated land use types within countries.

4 | DISCUSSION

As we introduced at the start of this paper, societal and scientific attention for SOC sequestration has risen steeply in recent years. The most recent IPCC report (IPCC, 2022) substantiated this trend by proposing SOC sequestration in agricultural soils as the third most potent option to support climate change mitigation working towards a 2030 timeline. As a result, voluntary carbon markets are rapidly expanding, and the majority rely on SOC simulation models for verification (Oldfield et al., 2022; Paul et al., 2023). This puts tremendous pressure on simulation models, which are now expected to provide the foundation for a reliable global carbon credit generation system. Our attempt in this paper was to provide a comprehensive list of SOC simulation models, with their validation contexts, aimed at defining guidelines for users on model selection and paving the way to ensuring consistent, transparent, robust and fair carbon crediting schemes. Our work fell short of such guidelines, and revealed a large gap between current state-of-the-art in model validation and the requirements imposed by the urgency and importance of this task.

Our main finding is that, despite a tremendous number of SOC models and hundreds of publications related to the validation and usage of such models, validation contexts are extremely limited. Particularly critical is that the main focus is on the global north (with a few exceptions), with severe under-representation of models suitable for ecosystems in Africa and the middle east, and to a lesser extent central and south America and Asia (except China). While countries of the global south overall hold a much smaller responsibility in past and current cumulated emissions, they are considered to hold an equal share in calculations of the global climate change mitigation potential of SOC sequestration (IPCC, 2022). It is therefore critically important that adequate tools are provided for inclusion of land managers from countries of the global south in carbon farming schemes. The current status of the SOC model validation means that appropriate models for simulation are not available for selection in these countries and we propose that there needs to be significant investment in both appropriate models and validated models for an equitable development of carbon markets globally.

Furthermore, we evidenced that a few models (RothC26.3, Coleman & Jenkinson, 1996; Century-v4, Metherell et al., 1993; and

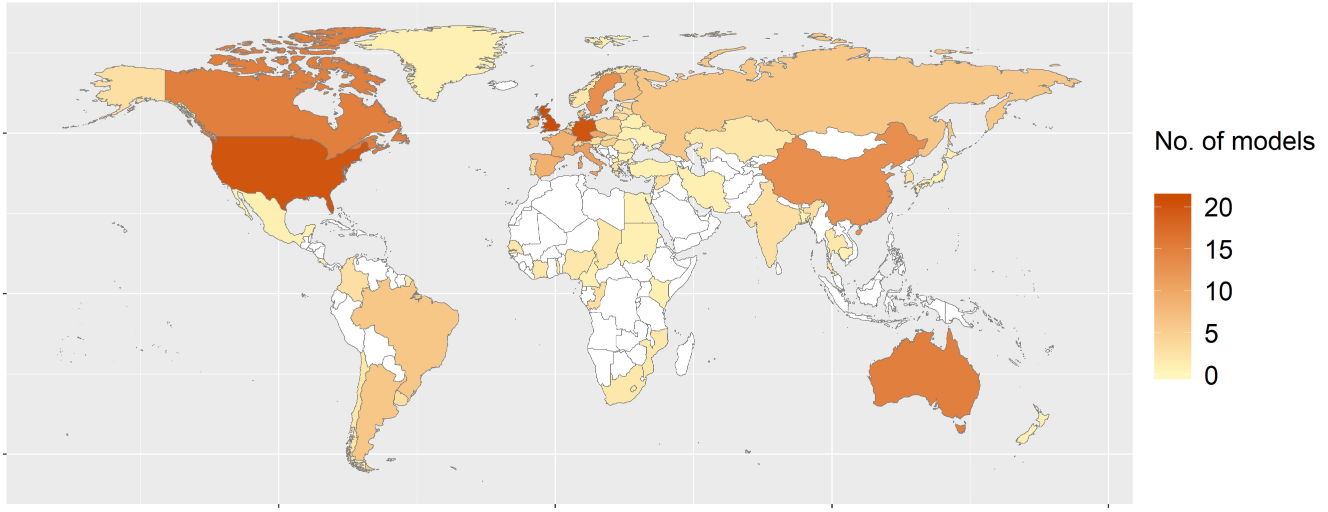

FIGURE 4 Map of model validations per country. The colour intensity represents the number of models that were validated within each country from 0 (white) to 20 (dark red). Multiple validations of a single model in a country are counted as one.

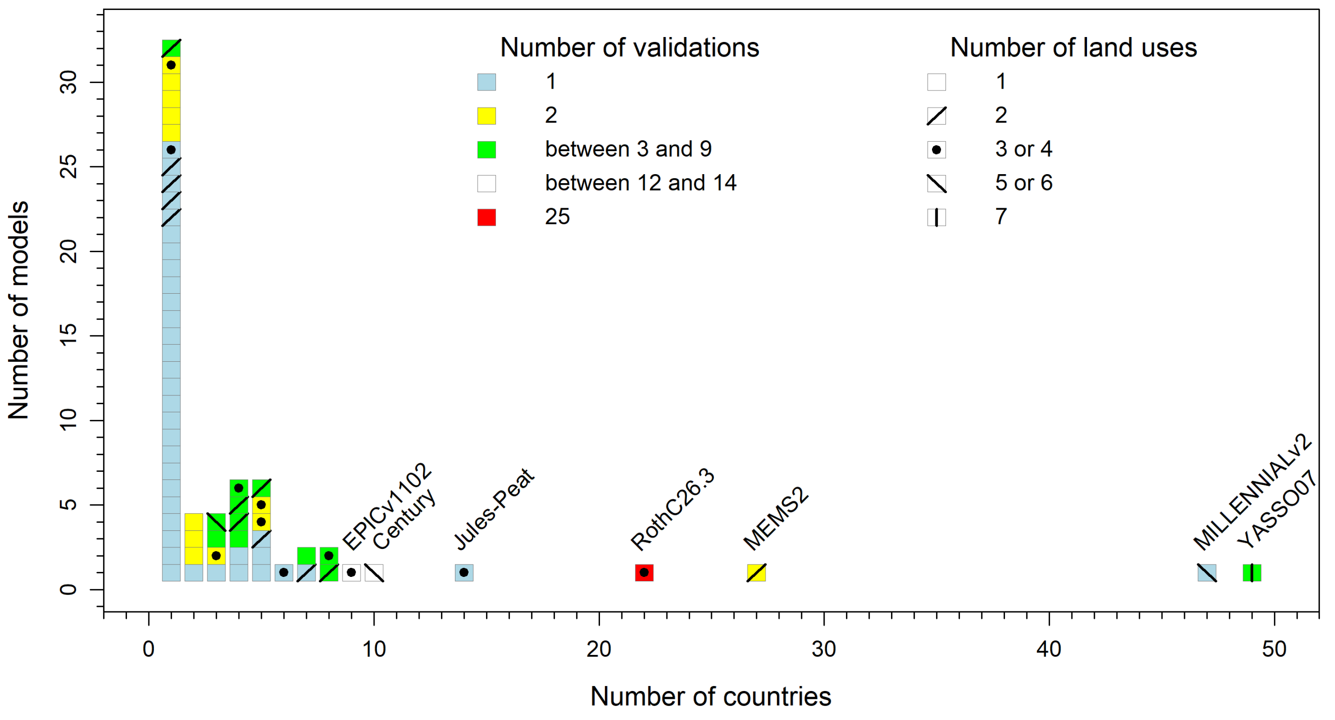

FIGURE 5 Stacked and textured histogram of the number of countries validated by individual models. Each box represents a model. For each model, the number of countries validated by the model is read on the x-axis, the number of publications validating the model is indicating by the box colour, and the number of land use types validated by the model is given by the texture.

EPICv1102, Izaurralde et al., 2012) occupy most of the literature on the topic with numerous validations in many countries, but largely restricted to a few land use types. Here, it is important to stress that models validated multiple times, even across several countries, may have been validated multiple times in similar pedoclimatic conditions and on the same land use. Therefore, although it increases their reliability, the number of validations should not be regarded directly as a proxy for the generality of these models. On the other hand, MILLENNIALv2 (Abramoff et al., 2022 and YASSO07 Tuomi et al., 2009), although validated relatively few times, cover larger geographical areas than other models, largely thanks to the use of a global SOC dataset for validation, and showed good coverage of land use types across countries.

In addition to those obvious limitations in terms of applicability, coverage and generalization of models, we found several further issues regarding the reliability of reported model performance as well as issues concerning the search and identification of relevant model versions (regarding both our own methodology and external factors inherent to inconsistent information reporting in studies). Below, we described these issues and draw a set of preliminary guidelines and a set of warnings to help potential users of our list in Table S2 for adequate model identification.

4.1 | Issues with validation

We found many instances of validation attempts which were lacking a clear distinction between the calibration and validation datasets (and sometimes partial or total overlap between them). We classified the validation attempts in the light of the IPCC guidelines (Shukla et al., 2019). In particular, we made a clear distinction between the term 'validation', also referred to as 'performance evaluation', which is the comparison of model output with data that was not used for calibration (i.e. independent data), and the term 'verification' which refers to the comparison of model output with the data that was used to calibrate the model (also referred to as 'non-independent validation' by Le Noe et al., 2023). As examples of rejected validation attempts, we give Franko et al. (2011) and Riggers et al. (2019) which initialized the SOC pool by giving to it the value that best fitted the SOC dataset on which the model was 'validated' (non-independently), inflating the performance as compared to a SOC simulation in which future SOC values are unknown. Our results are in line with those of Le Noe et al. (2023), who, despite applying a different methodology for their model selection and review found a lack of 'independent validations'.

We also found many instances of lack of clarity about the initialization procedure. It should be made clear whether initial SOC values (total and/or fractionated) are quantified using real measurements or instead evaluated with a spin-up procedure, which relies on the assumption that the SOC content has reached a steady state.

In general, we recommend the use of the IPCC guidelines for model validation (see also Moriasi et al., 2007), and while not the focus of this study, we underline the importance of the calibration procedure to model performance. See for example the efforts by Luo and Schuur (2020) or Tao et al. (2020) to refine parametrization and calibration procedures that lead to better model performance, such as allowing for time and space heterogeneity of parameters, or machine learning techniques to infer parameters. We also call for efforts to try to identify and quantify the different sources of error/uncertainty from the modelling, calibration and initialization assumptions and choices, as well as from the measurements (see some guidelines in Refsgaard, 1997; Refsgaard et al., 2007; Thornton & Rosenbloom, 2005). We also note that, as underlined by Le Noe et al. (2023), 'diachronic validation' which makes use of time-series data of SOC stocks allows for a more precise and trust-worthy evaluation of the model's performance as compared to a validation performed with measurements taken at a single point in time, because it reduces possible errors on initial SOC values and allows to reliably predict temporal trends. Lastly, we want to point to two additional specific issues encountered when assessing model validation performance.

The first issue concerns the use of linear regressions in validation attempts. We often observed the use of a linear regression in model performance estimation, where the simulated (herein called P_i for predicted) values given by the model are taken as an explanatory variable for the measured (herein called O_i for observed) values. This can indeed bring useful information about the overall performance of the model. However, we found numerous examples of poor interpretations of such regressions in the context of model validation. A linear regression allows for a rapid visualization of possible systematic (non-zero intercept) bias and magnitude-dependent (slope different than one) bias. However, the predictions resulting from the linear regression $P_{reg,i} = \text{intercept} + \text{slope} \times P_i$ do not have, unlike P_i, any clear mechanistic/biological interpretation. As a consequence, it is not correct to interpret the coefficient of determination R^2 of the linear regression as a measure of the (mechanistic) model performance, nor is it correct to say that the model explains R^2 per cent of the variation of the data, when in fact R^2 measures how well $P_{reg,i}$ fits the observed data (in other words it evaluates the performance of the linear regression). Ťupek et al. (2016) and Zhang et al. (2020) are good examples of this issue. In both cases, the fitted model did not perform very well, as can be seen by how the slope of the linear regression between the P_i and O_i values deviates from the 1:1 line. The authors, however, incorrectly interpreted the high R^2 from the linear regression (which indicates that the regression line fits the observed data well) as the fraction of variance explained by the mechanistic models. To achieve the aim of evaluating the model performance, one should instead use the Nash–Sutcliffe efficiency, which measures how well the predicted values P_i (i.e. the model) fits the observed data, in the exact same way as R^2 measures how well $P_{reg,i}$ fits the observed data. Note that Ťupek et al. (2016) still complied with the criteria listed in Section 2.3 and was deemed a correct validation. However, in Zhang et al. (2020), the models were calibrated using the validation dataset and the paper therefore did not qualify as a correct validation and was excluded from our classification (Table S2).

The other issue that we deemed worth further explaining relates to cases in which authors compare, not the actual simulated and observed values, but rather mean values obtained from the aggregation of data from contrasting pedoclimatic conditions and mean values of the corresponding simulated values. Indeed, by including data covering a broad range of pedoclimatic conditions, one can increase the variability in the data, and by doing so, increase the chance that the confidence intervals of the observations and the simulations overlap. Moreover, comparing the means in fact comes down to performing a linear regression with only an intercept term forcing the slope to zero. Ostrogović Sever et al. (2021) provide a good example of a validation we discarded due to this issue. As the authors of this article noted themselves, while scatterplots of the observed and predicted values do not point towards a good fit (Ostrogović Sever et al., 2021; Figure 4), aggregating these points to represent the means and standard deviations of particular land use types allow us to make the conclusion that the means of the observations do not differ significantly from the simulations (Ostrogović Sever et al., 2021; Figure 2). Nevertheless, if one has access to the confidence intervals calculated with the standard deviation SD_i of measurement replicates for each datapoint i (as opposed to SD of data with heterogeneous pedoclimatic conditions), we recommend the comparison between RMSE and $RMSE_{95}$ (which is the root mean square error expected if the observations are equal to the simulations plus normally distributed errors with standard deviations SD_i) as explained in Smith et al. (1997).

4.2 | Issues related to the search and identification of model entities

Tracing model histories and identifying all relevant model entities proved very difficult, largely due to inconsistent and sometimes unclear reporting. In addition, tracking model development in the cases in which the model code was open access was also difficult, as model development could, in those cases, take place outside of the control of the authors of the first trunk. It is therefore likely that we missed a few relevant validations for some model entities. As for any attempt of a literature review, one simply cannot browse a complete set of relevant publications, as no database exists that exhaustively reports all publications related to a certain field. In addition, it is not guaranteed that all publications of interest are returned by a given keyword search string.

In addition, the validation search was limited by the complexity of model development histories found in the literature and by our choice to search exclusively for the validations of the identified model entities (last model version of both the trunk and possible branches), and not of their previous versions. Indeed, it is possible that the validations of previous model versions are also relevant for the last version. This occurs when new versions bring modifications that do not impact SOC dynamics. This can be the case when the code is rewritten, when the developers add modules which use the outputs of SOC dynamics without affecting them (i.e. such as a carbon accounting module), or when they add a process which is specific to a certain land use type or crop which does not affect the dynamics of the land use type and crop already simulated in the model. Besides, the model development history includes intricate and non-explicit branching and merging of versions which brings additional difficulty to identifying the model versions whose validations are relevant for the selected model entities. Moreover, modifications to the model that impact the simulated SOC dynamics seem to be more frequent than modifications not affecting SOC dynamics. For this reason, we think that the number of missed model validations should be reasonably low. Overall, we acknowledge a crucial need for systematic tracking of the different versions of the model as well as of the model validations. This would make it easier to update our classification table over time, thus allowing an informed choice of model. Finally, we urge the provision of clear information on the availability of the model code or model software, conditions of use and updated links to the download page, all information that is often not provided in the relevant model literature or model documentation.

4.3 | Model choice and application

Should actors/stakeholders of the carbon farming landscape use the classification provided in this work, we urge them to take into consideration the limits of this work. To be applied in certain conditions, a model should have been validated for the same conditions in regard to all validation aspects (land use type, climate, soil type/texture and depth). However, a few issues arise when trying to precisely identify from our classification the set of experimental contexts for which a model was validated, partly due to the issues described in Section 4.2.

First, we did not report the climatic region or climate information. The reason is those are inconsistently reported in the different papers, with some model validations directly reporting the climate characteristics of the validation sites (annual precipitation and mean temperature) while others mention climatic zones or climate regions based on different climate classifications. It is still possible to identify, within the classification in Table 1, which model entities are applicable to a given climate based on the reported validation location (country and possibly sub-national region/state/province). To this mean, one can use climate maps such as the Köppen–Geiger climate map (Beck et al., 2018) to identify all locations around the world which are characterized by the climate type of interest, and then search within the classification to identify the model entities validated in the corresponding areas. Once this is done, one needs to narrow down the list of possible models based on other pedoclimatic conditions.

Unfortunately, experimental conditions described in the validation papers are often given for the globality of the validated dataset, and it is not always possible to establish a correspondence between experimental conditions and individual samples. As an example, if a paper states that the data cover two soil types S1 and S2, two countries C1 and C2 and two land use types L1 and L2, we do not necessarily know which of the eight pedoclimatic combinations $S \times C \times L$ are effectively represented in the data and thus are considered to be validated. In a number of cases, studies used data that cover broad ranges of pedoclimatic conditions and give statistical indicators of the goodness of fit for the whole dataset. In such cases it is not guaranteed that the model performs well when looking at a subset of the data corresponding to specific experimental conditions (see Plaza et al., 2012 for an illustration of such a situation).

We also underline that the validations reported for a certain model regard the specific combination of model structure and calibration (i.e. the set of parameter values used in the validation run) rather than the model alone. Our classification table reports the validations performed for each model, but no information is given on the calibration procedure. Therefore, when applying a certain model for a simulation, after verifying that it was validated for the specific context of interest, it is necessary to read the original validation paper to determine which (and how) parameters were calibrated and use the same ones.

The result of a model validation (intended as reliability/accuracy of the model output) is a crucial factor that should be also considered when choosing a model. However, there exists a variety of statistical tools that can be applied as quantitative measure of the goodness of fit between simulated and measured data, all coming with their own advantages and drawbacks. The absence of a unique and one-fits-all protocol is reflected in the literature encountered, and validation papers usually do not give a quantitative and systematic measure of the model performance. In addition, the evaluation of the goodness of fit is partly subjective: there is no prescribed threshold above which a statistical indicator guarantees that a model validation was successful (the same holds for the number of samples). Thus, we decided not to try to classify 'good' or 'bad' validations, but rather leave it to the users to decide for each case if the method and accuracy of the validation fits their requirements. We refer back to Section 4.1 as a reminder of a few common flaws found in the reviewed studies in terms of model performance assessment.

These considerations limit our ability to make model recommendations for users willing to estimate the SOC in a specific pedoclimatic context (or estimate the SOC sink potential of specific land management practices). Instead of a ready-to-go model selection from our classification, one needs to check in the table which models have been validated for the desired conditions, and then refer to the corresponding validations and further carefully verify that the paper actually validates the model in the desired context, with the required accuracy (whose level is up to the user's choice), and check which calibration is to be used. If two or more models are selected after this step, we suggest to give a higher level of confidence to the model with the highest variety of validated experimental conditions. Indeed, it is likely that any new experimental context will differ from previously validated conditions regarding detailed and hard to identify parameters such as localized climate variations, slightly different land use/management, difference in microbial communities or in land use histories and levels of ecological succession. A model with a greater variety of validated contexts is then more likely to make accurate prediction in novel experimental conditions.

5 | CONCLUSIONS

The societal pressure and hope placed on SOC simulation models to provide the foundation for fair, consistent, transparent and robust carbon crediting schemes is enormous. We made great efforts to comprehensively report up to date versions of SOC models and their validation contexts, and to offer guidelines for selecting models appropriate to specific pedoclimatic conditions, land use types, temporal and spatial scales. The general lack of clear reporting, numerous flaws in model performance evaluation and the poor overall coverage of land use types across countries and pedoclimatic conditions prevented us from providing such guidelines. To date, SOC simulation does not represent an adequate tool for globally ensuring effectiveness of SOC sequestration efforts and ensuring reliable carbon crediting. Most critically, countries of the global south, particularly in Africa, the least emitting countries that are already facing the most drastic consequences of climate change, are, here again, the most poorly supported. Mitigation deterrence mechanisms and context specific trade-offs between SOC sequestration and other sustainability objectives already pose large challenges to effective use of SOC sequestration as a tool to support sustainability, and even question its relevance at global scale (McLaren, 2020; Moinet et al., 2023). If we are, as society, to promote SOC sequestration based on carbon crediting schemes nonetheless, it is critically urgent that the scientific community and society at large invest massively in supporting its implementation appropriately, everywhere.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Carmen Vazquez https://orcid.org/0000-0002-2384-4658
Gabriel Y. K. Moinet https://orcid.org/0000-0002-5347-8847

REFERENCES

Abramoff, R. Z., Guenet, B., Zhang, H. C., Georgiou, K., Xu, X. F., Viscarra Rossel, R. A., Yuan, W., & Ciais, P. (2022). Improved global-scale predictions of soil carbon stocks with Millennial Version 2. *Soil Biology & Biochemistry*, *164*, 108466. https://doi.org/10.1016/j.soilbio.2021.108466

Amelung, W., Bossio, D., de Vries, W., Kögel-Knabner, I., Lehmann, J., Amundson, R., Bol, R., Collins, C., Lal, R., Leifeld, J., Minasny, B., Pan, G., Paustian, K., Rumpel, C., Sanderman, J., van Groenigen, J., Mooney, S., van Wesemael, B., Wander, M., & Chabbi, A. (2020). Towards a global-scale soil climate mitigation strategy. *Nature Communications*, *11*(1), 5427. https://doi.org/10.1038/s41467-020-18887-7

Batjes, N. H. (2015). *World soil property estimates for broad-scale modelling (WISE30sec). No. 2015/01*. ISRIC-World Soil Information.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Bünemann, E. K., Bongiorno, G., Bai, Z., Creamer, R. E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T. W., Mader, P., Pulleman, M., & Brussaard, L. (2018). Soil quality—A critical review. *Soil Biology and Biochemistry*, *120*, 105–125. https://doi.org/10.1016/j.soilbio.2018.01.030

Campbell, E. E., & Paustian, K. (2015). Current developments in soil organic matter modeling and the expansion of model applications: A review. *Environmental Research Letters*, *10*(12), 123004. https://doi.org/10.1088/1748-9326/10/12/123004

Chabbi, A., Lehmann, J., Ciais, P., Loescher, H. W., Cotrufo, M. F., Don, A., SanClements, M., Schipper, L., Six, J., Smith, P., & Rumpel, C. (2017).

Aligning agriculture and climate policy. *Nature Climate Change*, *7*(5), 307–309. https://doi.org/10.1038/nclimate3286

Coleman, K., & Jenkinson, D. S. (1996). *RothC-26.3—A model for the turnover of carbon in soil*. Paper presented at the evaluation of soil organic matter models, Berlin, Heidelberg.

Derrien, D., Barré, P., Basile-Doelsch, I., Cécillon, L., Chabbi, A., Crème, A., Fontaine, S., Henneron, L., Janot, N., Lashermes, G., Quénéa, K., Rees, F., & Dignac, M.-F. (2023). Current controversies on mechanisms controlling soil carbon storage: Implications for interactions with practitioners and policy-makers. A review. *Agronomy for Sustainable Development*, *43*(1), 21. https://doi.org/10.1007/s13593-023-00876-x

European Commission (EC). (2021). *Sustainable carbon cycles*. Communication from the Commission to the European Parliament and the Council. https://tinyurl.com/4nj3ma4w

Falloon, P. D., & Smith, P. (2000). Modelling refractory soil organic matter. *Biology and Fertility of Soils*, *30*(5), 388–398. https://doi.org/10.1007/s003740050019

Food and Agriculture Organisation of the United Nations (FAO). (2017). *Soil organic carbon: The hidden potential*. https://www.fao.org/3/I6937EN/i6937en.pdf

Food and Agriculture Organisation of the United Nations (FAO). (2019). *Recarbonization of global soils—A dynamic response to offset global emissions*. http://www.fao.org/3/i7235en/I7235EN.pdf

Franko, U., Kolbe, H., Thiel, E., & Liess, E. (2011). Multi-site validation of a soil organic matter model for arable fields based on generally available input data. *Geoderma*, *166*(1), 119–134. https://doi.org/10.1016/j.geoderma.2011.07.019

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

Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., & Swanson, D. K. (2013). The northern circumpolar soil carbon database: Spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth System Science Data*, *5*(1), 3–13. https://doi.org/10.5194/essd-5-3-2013

IPCC. (2022). Climate change 2022: Mitigation of climate change. In P. R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, & J. Malley (Eds.), *Contribution of working group III to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. https://doi.org/10.1017/9781009157926

Izaurralde, R. C., Mcgill, W. B., & Williams, J. R. (2012). *Development and application of the EPIC model for carbon cycle, greenhouse-gas mitigation, and biofuel studies*. https://www.osti.gov/biblio/1055403

Kopittke, P. M., Berhe, A. A., Carrillo, Y., Cavagnaro, T. R., Chen, D., Chen, Q.-L., Román Dobarco, M., Dijkstra, F. A., Field, D. J., Grundy, M. J., He, J. Z., & Minasny, B. (2022). Ensuring planetary survival: The centrality of organic carbon in balancing the multifunctional nature of soils. *Critical Reviews in Environmental Science and Technology*, *52*(23), 4308–4324. https://doi.org/10.1080/10643389.2021.2024484

Kotu, V., & Deshpande, B. (2015). Chapter 2—Data mining process. In V. Kotu & B. Deshpande (Eds.), *Predictive analytics and data mining* (pp. 17–36). Morgan Kaufmann.

Lal, R. (2004). Soil carbon sequestration impacts on global climate change and food security. *Science*, *304*(5677), 1623–1627. https://doi.org/10.1126/science.1097396

Le Noe, J., Manzoni, S., Abramoff, R., Bölscher, T., Bruni, E., Cardinael, R., Ciais, P., Chenu, C., Clivot, H., Derrien, D., Ferchaud, F., Garnier, P., Goll, D., Lashermes, G., Martin, M., Rasse, D., Rees, F., Sainte-Marie, J., Salmon, E., & Guenet, B. (2023). Soil organic carbon models need independent time-series validation for reliable prediction. *Communications Earth & Environment*, *4*, 1–8. https://doi.org/10.1038/s43247-023-00830-5

Lugato, E., Lavallee, J. M., Haddix, M. L., Panagos, P., & Cotrufo, M. F. (2021). Different climate sensitivity of particulate and mineral-associated soil organic matter. *Nature Geoscience*, *14*(5), 295–300. https://doi.org/10.1038/s41561-021-00744-x

Luo, Y., & Schuur, E. A. (2020). Model parameterization to represent processes at unresolved scales and changing properties of evolving systems. *Global Change Biology*, *26*(3), 1109–1117. https://doi.org/10.3389/fdata.2020.00017

Manzoni, S., & Porporato, A. (2009). Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biology & Biochemistry*, *41*(7), 1355–1379. https://doi.org/10.1016/j.soilbio.2009.02.031

McLaren, D. (2020). Quantifying the potential scale of mitigation deterrence from greenhouse gas removal techniques. *Climatic Change*, *162*(4), 2411–2428. https://doi.org/10.1007/s10584-020-02732-3

Metherell, A. K., Harding, L. A., Cole, C. V., & Parton, W. (1993). *CENTURY soil organic matter model environment*. Technical documentation. Agroecosystem version 4.0. Great Plains System Research Unit technical report no. 4.

Moinet, G. Y. K., Hijbeek, R., van Vuuren, D. P., & Giller, K. E. (2023). Carbon for soils, not soils for carbon. *Global Change Biology*, *29*, 2384–2398. https://doi.org/10.1111/gcb.16570

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Nachtergaele, F. O., van Velthuizen, H., Verelst, L., Wiberg, D., Batjes, N. H., Dijkshoorn, J. A., van Engelen, V. W. P., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., & Shi, X. (2012). *Harmonized world soil database (Version 1.2)*. Food and Agriculture Organization of the UN, International Institute for Applied Systems Analysis, ISRIC—World Soil Information, Institute of Soil Science—Chinese Academy of Sciences, Joint Research Centre of the EC. http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HWSD_Documentation.pdf

Oldfield, E. E., Eagle, A. J., Rubin, R. L., Rudek, J., Sanderman, J., & Gordon, D. R. (2022). Crediting agricultural soil carbon sequestration. *Science*, *375*(6586), 1222–1225. https://doi.org/10.1126/science.abl7991

Ostrogović Sever, M. Z., Barcza, Z., Hidy, D., Kern, A., Dimoski, D., Miko, S., Hasan, O., Grahovac, B., & Marjanovic, H. (2021). Evaluation of the terrestrial ecosystem model biome-BGCMuSo for modelling soil organic carbon under different land uses. *Land*, *10*(9), 968. https://doi.org/10.3390/land10090968

Paul, C., Bartkowski, B., Dönmez, C., Don, A., Mayer, S., Steffens, M., Weigl, S., Wiesmeier, M., Wolf, A., & Helming, K. (2023). Carbon farming: Are soil carbon certificates a suitable tool for climate change mitigation? *Journal of Environmental Management*, *330*, 117142. https://doi.org/10.1016/j.jenvman.2022.117142

Plaza, C., Gollany, H. T., Baldoni, G., Polo, A., & Ciavatta, C. (2012). Predicting long-term organic carbon dynamics in organically amended soils using the CQESTR model. *Journal of Soils and Sediments*, *12*(4), 486–493. https://doi.org/10.1007/s11368-012-0477-1

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *The Soil*, *7*(1), 217–240.

Refsgaard, J. C. (1997). Parameterisation, calibration and validation of distributed hydrological models. *Journal of Hydrology*, *198*(1), 69–97. https://doi.org/10.1016/S0022-1694(96)03329-X

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process—A

framework and guidance. *Environmental Modelling & Software*, *22*(11), 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004
Riggers, C., Poeplau, C., Don, A., Bamminger, C., Hoper, H., & Dechow, R. (2019). Multi-model ensemble improved the prediction of trends in soil organic carbon stocks in German croplands. *Geoderma*, *345*, 17–30. https://doi.org/10.1016/j.geoderma.2019.03.014
Shukla, P. R., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H.-O., Roberts, D. C., Zhai, P., Slade, R., Connors, S., van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., … Malley, J. (Eds.). (2019). *IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*.
Smith, P., Smith, J. U., Powlson, D. S., McGill, W. B., Arah, J. R. M., Chertov, O. G., Coleman, K., Franko, U., Frolking, S., Jenkinson, D. S., Jensen, L. S., Kelly, R. H., Klein-Gunnewiek, H., Komarov, A. S., Li, C., Molina, J. A. E., Mueller, T., Parton, W. J., Thornley, J. H. M., & Whitmore, A. P. (1997). A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. *Geoderma*, *81*(1–2), 153–225. https://doi.org/10.1016/s0016-7061(97)00087-6
Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A. B., de Remy de Courcelles, V., Singh, K., Wheeler, I., Abbott, L., Angers, D. A., Baldock, J., Bird, M., Brookes, P. C., Chenu, C., Jastrow, J. D., Lal, R., … Zimmermann, M. (2013). The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agriculture, Ecosystems & Environment*, *164*, 80–99. https://doi.org/10.1016/j.agee.2012.10.001
Tao, F., Zhou, Z., Huang, Y., Li, Q., Lu, X., Ma, S., Huang, X., Liang, Y., Hugelius, G., Jiang, L., Doughty, R., Ren, Z., & Luo, Y. (2020). Deep learning optimizes data-driven representation of soil organic carbon in earth system model over the conterminous United States. *Frontiers in Big Data*, *3*, 17. https://doi.org/10.3389/fdata.2020.00017
The Gold Standard Foundation. (2020). *Soil organic carbon framework methodology. Version 1.0*. https://tinyurl.com/cf36uc6b
Thornton, P. E., & Rosenbloom, N. A. (2005). Ecosystem model spin-up: Estimating steady state conditions in a coupled terrestrial carbon and nitrogen cycle model. *Ecological Modelling*, *189*(1), 25–48. https://doi.org/10.1016/j.ecolmodel.2005.04.008
Tifafi, M., Guenet, B., & Hatté, C. (2018). Large differences in global and regional total soil carbon stock estimates based on SoilGrids, HWSD, and NCSCD: Intercomparison and evaluation based on field data from USA, England, Wales, and France. *Global Biogeochemical Cycles*, *32*(1), 42–56.
Tilman, D., Fargione, J., Wolff, B., D'Antonio, C., Dobson, A., Howarth, R., Schindler, D., Schlesinger, W. H., Simberloff, D., & Swackhamer, D. (2001). Forecasting agriculturally driven global environmental change. *Science*, *292*(5515), 281–284. https://doi.org/10.1126/science.1057544
Tuomi, M., Thum, T., Jarvinen, H., Fronzek, S., Berg, B., Harmon, M., Trofymow, J. A., Sevanto, S., & Liski, J. (2009). Leaf litter decomposition-estimates of global variability based on Yasso07 model. *Ecological Modelling*, *220*(23), 3362–3371. https://doi.org/10.1016/j.ecolmodel.2009.05.016
Ťupek, B., Ortiz, C. A., Hashimoto, S., Stendahl, J., Dahlgren, J., Karltun, E., & Lehtonen, A. (2016). Underestimation of boreal soil carbon stocks by mathematical soil carbon models linked to soil nutrient status. *Biogeosciences*, *13*(15), 4439–4459. https://doi.org/10.5194/bg-13-4439-2016
van Leeuwen, J. P., Saby, N. P. A., Jones, A., Louwagie, G., Micheli, E., Rutgers, M., Schulte, R. P. O., Spiegel, H., Toth, G., & Creamer, R. E. (2017). Gap assessment in current soil monitoring networks across Europe for measuring soil functions. *Environmental Research Letters*, *12*(12), 124007. https://doi.org/10.1088/1748-9326/aa9c5c
Verra: Verified Carbon Standard. (2020). *VM0042 methodology for improved agricultural land management. Version 1.0*. https://tinyurl.com/3emtwyey
Zhang, H. C., Goll, D. S., Wang, Y. P., Ciais, P., Wieder, W. R., Abramoff, R., Huang, Y., Guenet, B., Prescher, A. K., Viscarra Rossel, R. A., Barré, P., Chenu, C., Zhou, G., & Tang, X. L. (2020). Microbial dynamics and soil physicochemical properties explain large-scale variations in soil organic carbon. *Global Change Biology*, *26*(4), 2668–2685. https://doi.org/10.1111/gcb.14994
Zhang, Y., Lavallee, J. M., Robertson, A. D., Even, R., Ogle, S. M., Paustian, K., & Cotrufo, M. F. (2021). Simulating measurable ecosystem carbon and nitrogen dynamics with the mechanistically defined MEMS 2.0 model. *Biogeosciences*, *18*(10), 3147–3171. https://doi.org/10.5194/bg-18-3147-2021

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Garsia, A., Moinet, A., Vazquez, C., Creamer, R. E., & Moinet, G. Y. K. (2023). The challenge of selecting an appropriate soil organic carbon simulation model: A comprehensive global review and validation assessment. *Global Change Biology*, *00*, 1–15. https://doi.org/10.1111/gcb.16896