



10.5281/zenodo.14761278

Preliminary note: Recommendations to the CRCF methodology for mineral soils and agroforestry for strengthening the uncertainty reporting

Sofia Biffi (Aarhus University), Ahmad Al Bitar (INRAE-Cesbio), Eric Ceschia (INRAE-Cesbio), Iris Cronin (Aarhus University), Istem Fer (FMI), Liisa Kulmala (FMI), Edouard Lanckriet (Agrosolutions), Manuel Martin (INRAE), Julius Vira (FMI), Greet Ruyschaert (ILVO)

Introduction

This document summarises the recommendations of the MARVIC consortium regarding minimum requirements and recommendation for the use of level 2 and level 3 models¹ in MRV systems for agricultural carbon removal projects, and related uncertainty estimation throughout the MRV modelling structure.

Key definitions related to modelling

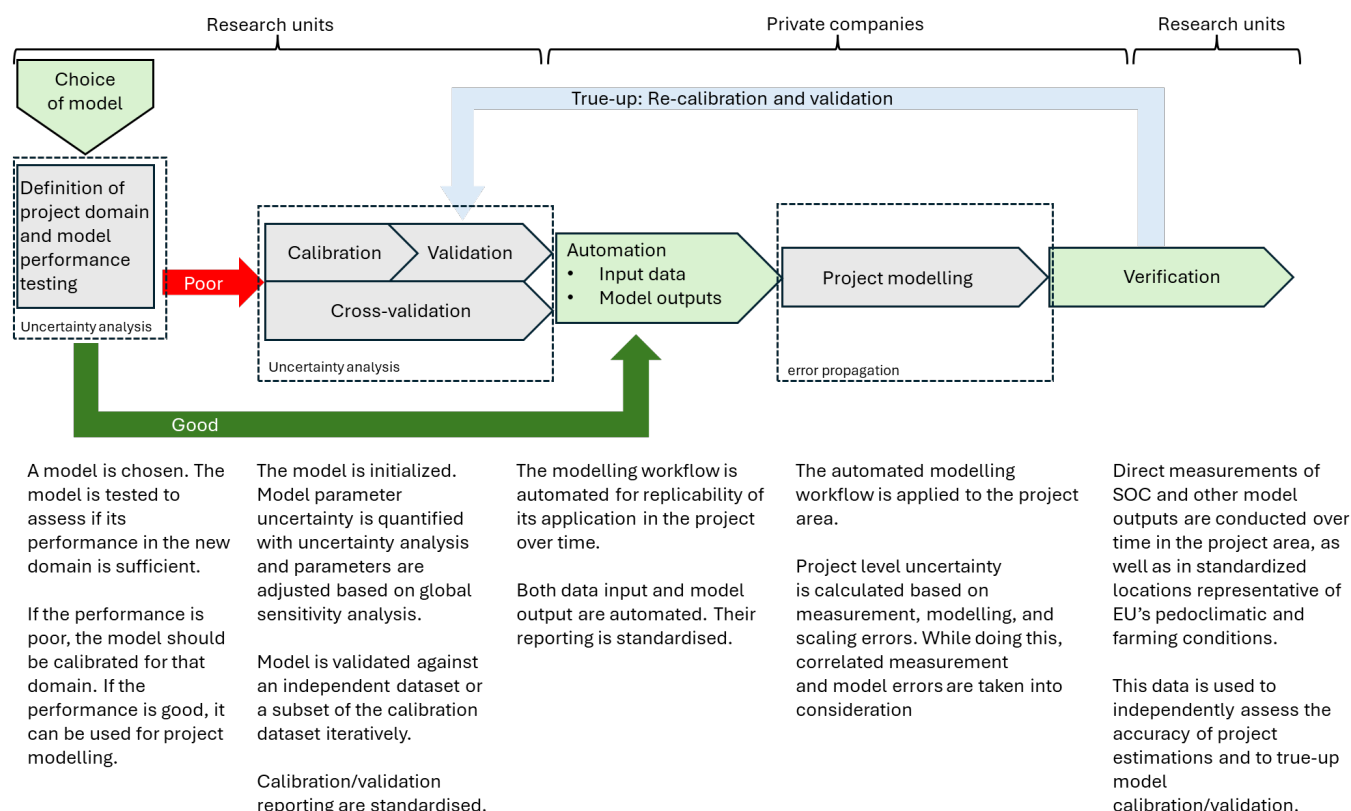
- **Modelling workflow:** all the steps required for preparing and running one or more models to produce model predictions. This includes input data curation, model configuration and parametrization, initialisation, spin-up, and post-processing.
- **Operational processing chain (OPC):** automated workflow that combines a spe-

cific modelling workflow (including the combination of multiple models) with specific data layers input to generate predictions and automate their outputs, ideally in a standardised way.

- **Project modelling:** running an OPC for a Project over a specific Project area, under specific farming conditions, and using relevant data layers (i.e. using an MRV tool across one or multiple farms).
- **MRV modelling structure:** in an MRV system, sequence of actions that are directly linked to running a model (or multiple models) using a modelling workflow, as shown in Figure 1.

¹SOC changes can be assessed with models of varying levels of complexity. Level 1 refers to simple 'empirical' models; Level 2, to 'soil' models (e.g. Roth-C, C-tool, ICBM, Yasso); and Level 3 to 'ecosystem' models (e.g. DAISY, LandscapeDNDC, APSIM), with model complexity and data demand increasing from Levels 1 to 3 (FAO 2019).

Managing the MRV modelling structure in a carbon certification framework such as the CRCF



The models and modelling workflows used for soil carbon quantification must be scientifically robust, which implies that they are calibrated, validated, and verified within their domain of validity. This complex process requires a high degree of scientific expertise to be carried out adequately, and most end-users of carbon certification frameworks (i.e. project developers) do not have such necessary expertise. Thus, we propose to separate the roles and responsi-

bilities of research entities and project developers within the CRCF framework in two stages:

1. Public research bodies (e.g., national research institutes) are responsible for the process of calibrating, validating, and verifying GHG emission reduction and removal models and modelling workflows in their domain of validity (i.e. model domain). This work could be reviewed by an



independent scientific committee. It is important that a process of continuous improvement of the models, based on new data availability and research progress is ensured. It is also recommended that a network of well-monitored benchmark sites is established to independently validate the OPCs used in Project modelling by project developers. To support this process, MARVIC will propose in a future note how networks of benchmark sites could be used for model improvement (i.e. model parametrization, calibration, and validation) and for independent verification of the OPCs.

2. The validated models from stage 1 could be used by public or private project developers in their MRV activities, e.g., by automating them and integrating them in digital tools for use during Project modelling. Private actors are also respon-

sible for part of the verification process through the monitoring over time of their Project Area. The digital end-user tools can be connected to e.g., CAP data, the FMIS (i.e. farm management systems used by the farmers), or lab data. We recommend that the automation process should follow common formatting standards for data input and model output reporting to ensure replicability of GHG emission reduction and removal results and comparability among MRV tools. To support this process, MARVIC will explore the applicability of existing standards to MRV systems.

This process will ensure that end users will utilise carbon MRV tools that are running models parameterized, validated and documented by a trusted (public) third party, while demonstrating that they are using such models within their domain of validity.

The use of models in MRV systems: pitfalls and recommendations for strengthening uncertainties reporting

The use of process-based models/OPCs in an MRV system to estimate SOC stock change and/or GHG emissions change following changes in agricultural practices follows a sequence of actions, which are described in Figure 1. Although these broad steps are generally consistent across different approaches, due to the current lack of standardised

guidelines on the methodology to perform each of these actions, these can each be conducted in different ways. These incompatibilities hinder the replicability and reusability of model outputs, creating deviations within the same project over time, and drastically reducing the comparability of different MRV systems. Importantly, they also



introduce the risk of MRV developers and users making strategic choices to maximise carbon crediting outcomes (i.e. “gaming”). Standardised requirements for how models are used, documented and tested would increase the transparency and replicability of MRV systems and would help towards building public trust in carbon farming markets.

This brief focuses on minimum requirements and recommendations for the use of level 2 and level 3 models² in MRV systems for agricultural carbon removal projects, and related uncertainty estimation throughout the MRV modelling structure. We have reviewed existing standards and make reference to Verra methodologies (Verra VMD0053, 2023 and Verra VM0042, 2024), and to recommendations from Lavallee et al. (2024). For each of the actions in the MRV modelling structure outlined in Figure 1, we briefly summarise existing requirements in Verra certification methodology which can be considered the minimum requirements to adhere to current certification standards (i.e. “Verra requirements”), provide our additional recommendations towards more robust and transparent guidelines (i.e. “Additional recommendations”), and highlight key research aspects that are still needed to inform our additional recommendations (i.e. “Future work”).

Our main recommendations can be summarised in the following four points:

1. Encourage the recognition of publicly available, well documented, and sufficiently calibrated and validated models that can be used by project developers. Models should have a well-defined domain of validity based on standardised definitions of pedoclimatic conditions and of farming practices.
2. The MRV modelling structure should be standardised where possible to increase transparency, comparability, and replicability of MRV systems. This means developing standards in model/OPC selections, model domain of validity definitions, modelling workflows, and input and output reporting.
3. Models and modelling workflows should be continuously updated with new data acquired during the Project verification phase of the MRV modelling structure. This will ensure that over time the model domain of validity will become more specific, covering the majority of pedoclimatic conditions and farming practices in Europe and increasing accuracy in GHG emission reduction and removal predictions.

²SOC changes can be assessed with models of varying levels of complexity. Level 1 refers to simple ‘empirical’ models; Level 2, to ‘soil’ models (e.g. Roth-C, C-tool, ICBM, Yasso); and Level 3 to ‘ecosystem’ models (e.g. DAISY, LandscapeDNDC, APSIM), with model complexity and data demand increasing from Levels 1 to 3 (FAO 2019).

4. Establish, together with Member States, a network of well-monitored benchmark sites for independent verification of the OPCs used by project developers. This

could substitute, to some extent, project by project verification and would result in reduced costs and increased transparency.

1. Choice of model and modelling workflow

Each model introduces different uncertainties to SOC stock estimation based on the model structure and input data requirements. The choice of model and modelling workflow for an MRV system is steered by the availability and resolution of data in the project area, the capacity to model different farming practices, the availability of model expert users, and the required level of accuracy at the scale of application. Currently there is no standardised approach for selecting models based on these considerations.

Any change to the structure, default input data, configuration, or parameters of a model can result in differences in the model outputs. Thus, model versions must be uniquely identified via version control to ensure reproducibility of results. For a given set of input data, configuration and parameters, any copy of the model reporting the same model and modelling workflow version on the same domain must always produce the same model output.

VERRA requirements (VMD0053):

- Provide traceable record of source code, standard parameters, and default input data needed to reproduce a given model output.
- Models must be version controlled and publicly available.

Additional recommendations:

- Model choice should be justified using standardised procedure, such as a decision tree (see forthcoming ORCASA Deliverable 4.2 early 2025).

Future work needed:

- Develop a standardised approach for the choice of model/OPC. MARVIC will propose decision trees as part of T3.1 and T3.3.



2. Model testing and continuous improvement of model domain

Once a model has been chosen, it can be tested in a new area of application to assess if the model can successfully represent local conditions in its simulation, using the data that is available in the project, at the resolution that is available within the project. Model testing covers all the aspects of modelling workflow, as it means running the model using the project input data and assessing its performance via uncertainty analysis. If the performance is poor, the model should be calibrated to better represent the project domain. If the performance is good, the model can be used in the MRV system as it is. Recommendations for how to conduct the modelling workflow of model testing (i.e. initialisation, uncertainty analysis, standardisation of reporting for input and outputs) are covered in the Model Calibration and Validation section below. One key aspect of model testing is the standardisation of the definition of the project domain, which warrants specific recommendations.

Model domain: The combination of environmental and agricultural conditions that can be represented by a model is called domain. In the context of carbon farming, a model domain is defined as:

land use × pedoclimatic conditions × carbon farming practice

Under arable land use, as different crops and crop rotations introduce an additional level of complexity in the model parametrization, the model domain is defined by an additional factor as:

(arable land use) × pedoclimatic conditions × carbon farming practice × crop type

Each MRV project is required by current certification standards to declare the unique combinations of land use, pedoclimatic conditions, and carbon farming practices (and crop types, in an arable setting) that are found within the project area. Evaluating the performance of the model in the local context of the project ensures that the project falls within the model domain (i.e. the model can adequately represent the local project conditions). However, there is currently no standardised definition of the different components of model domain (see Grouping of domain definitions).

Grouping of domain definitions: Because of the high number of combinations of factors denoting a domain, a certain degree of simplification and grouping inevitably must occur when defining a model domain. For example, similar carbon farming practices or similar crop types are usually grouped together (e.g. cereals, root crops etc), as it is unrealistic to aim to a complete rep-

resentation of individual farming practices and their combinations across the benchmarking sites needed to calibrate the models. There is currently no standardised definition of these groups, nor of combination of groups, and this variability of definitions can introduce the possibility of gaming and render different MRV systems incomparable. We recommend that these grouping definitions should be standardised in a workable and context specific way (e.g. set at the Member State level to reflect local farming conditions that can be implemented at the project scale). Continuous improvement of the models with new data for model calibration and validation will ensure that these domains of validity will improve over time, and standardisation of the domain grouping definitions will ensure that different MRV systems will be comparable over time.

VERRA requirements (VMD0053):

- Project must list climate zones (definitions from IPCC, 2019), project soils (texture class from FAO 2015), and crop functional groups.

Additional recommendations:

- Standardise pedoclimate definitions: list climatic groups more specific to Europe (e.g. Environmental Zones of Europe), soil groups including soil type groups (e.g. FAO soil type classification, national

classification).

- Standardise farming practices definitions: list acceptable groupings of carbon farming practices, list acceptable crop type groups.

Future work needed:

- Standardise crop type groups.
- Standardise carbon farming practices groups.
- How to report model and project domain in a standardised way. MARVIC will explore relevant standardisation approaches, such as the Ecological Forecasting Initiative (EFI) standards (Dietze et al., 2023).



3. Model calibration and validation

Although a model can be used successfully in a specific domain, a reliable performance in a different combination of pedoclimate and carbon farming practices is not guaranteed, due to the variability and heterogeneity of soil and crop systems in agricultural landscapes at different scales. If model testing shows poor model performance in the project domain, the model does not adequately represent the project local conditions and should thus be calibrated and validated with additional data that represents such conditions.

Initialisation: Initialization refers to the process of setting up a model in its initial state before it is used. In MRV context this often involves estimating not only the total soil carbon pool size but also its partitioning into different SOC pools with different sensitivity to decomposition. Proper initialization ensures that the model starts in a predictable and valid state and thus it minimizes errors and potential instabilities during simulations.

Calibration: Calibration is the process of constraining model parameters using measured data. Model calibration can be conducted following different methods, and current guidelines do not prescribe a specific methodology, leaving room for innovative approaches. Models typically represent inter-linked processes with different parameters

affecting the interconnected parts of the model. Hence, using only one data stream in calibration may degrade model's overall predictive capacity if some of the interconnected parameters are not well constrained. Using multiple constraints (i.e. multi-data streams on different parameters and processes such as carbon, water, phenology, energy, nitrogen) for calibration enhances the robustness and accuracy of the model by providing diverse and complementary information further reducing uncertainty in parameter estimation. As many model parameters are non-measurable and must be estimated indirectly, adopting probabilistic approaches, such as Bayesian approaches, should be encouraged when informing model parameters and performing model calibration.

Validation: Once a model has been calibrated over a domain, its performance can be assessed against a set of independent data, a process known as validation. Critically, the same exact model version, configuration and model parametrization should be used during the calibration and the validation stage.

Transparency of calibration/validation data: The ability to calibrate and validate a model within a specific project domain is always limited by the availability of field data to validate model predictions within that domain. This type of data typically comes from

long term experiments and/or monitoring networks and is often sparse across pedo-climatic conditions and farming practices, rendering access to data the main barrier to the calibration of models across the EU. There must be full transparency regarding the description of the data used for calibration and validation and existing protocols require a full description of such datasets.

Uncertainty analysis (UA) and sensitivity analysis (SA):

- UA quantifies the uncertainty in model predictions (i.e., the model prediction error) from the distance between observed values and simulated values.
- SA is usually performed following UA, and its aim is to attribute proportions of the uncertainty of the model output to different sources, such as input data, parameters, model structure, and resolution.

To address UA and SA systematically: (1) the sources of uncertainty, which may be specific to the land use and model level considered, must be identified, quantified, and prioritised in a SA; and (2) these results should be properly reported. The method employed to conduct UA and SA can also lead to biased results. Specifically, “local” SA methodologies are commonly recommended for the evaluation of SOC models results (e.g. FAO, 2019), but they are inadequate in the presence of non-linear relationships or interactions between the variables consid-

ered, leading to biased SA results. “Global” SA approaches should be recommended by the CRCF.

VERRA requirements (VMD0053):

Calibration and validation methodology:

- Calibration and validation data must be demonstrably independent either by being two separate datasets with no overlap of research locations, or by a statistical process such as cross-validation.
- Justification for splitting experimental data in calibration and validation dataset.
- Parameter sets used when validating the model must be the same as those used when the model is applied to simulate baselines and project practices.

Calibration:

- Documentation of all internal model parameter sets and justification for any variation within the project area.
- Detailed reporting of datasets used for validation is expected only upon request of the certification body.

Validation:

- Detailed reporting of datasets used for validation.
- Assessment of bias for each practice category using pooled measurement uncertainty (PMU).
- Model prediction error for changes in SOC, N₂O, NH₄ (where relevant) with a minimum coverage of 90% for 90% pre-



diction intervals on independent data (i.e., the 90% prediction intervals should contain the measured value for at least 90% of the validation data).

Additional recommendations

Initialisation:

- Where applicable, model initialisation methodology should be reported and justified, such as duration of initialisation period, calibration of specific parameters, and initial distribution of pools.
- Calibrating slow acting carbon parameters and not just fast ones that are affecting photosynthesis/respiration/phenology.

Calibration and validation methodology:

- Details on the input datasets used should always be reported in a standardised way, including the source and the resolution of all the datasets.
- Input data and model outputs should be versioned with unique identifiers.
- Calibration and validation datasets should be made publicly available.
- Use global sensitivity analysis methods instead of local ones.

Calibration:

- Using multiple constraints (i.e. multi-data streams on different parameters and processes such as carbon, water, phenology, energy, nitrogen) for calibration.
- Detailed justification when calibrating

against derived datasets (e.g. remote sensing based), as the assumptions in derivation and in the model might differ.

- Encourage Bayesian calibration and recommend prior and posterior predictive checks to assess the validity of distributional assumptions before and after calibration.

Validation:

- Encourage independent validation, as cross-validation does not guarantee independence of calibration and validation datasets. However, given the limited availability of calibration and validation data under different pedoclimatic conditions and farming practices across Europe, it is likely that cross-validation will still have to be used.
- Model performance threshold for evaluating model performance should be standardised.
- Define the list of sources of uncertainty for each model level that should be considered in uncertainty analysis so that no relevant sources can be omitted.
- Reporting of model uncertainty from the list of uncertainty sources should be standardised.

Future work

- Standardise the reporting of model calibration and validation procedure, used data streams, parameter sets and outputs. MARVIC will explore relevant

standardisation approaches, such as the EFI standards (Dietze et al., 2023), which cover model outputs but could be expanded to model inputs and parametrization (T3.2 and T3.3).

- Provide recommendations for the list of uncertainty sources for each model level. MARVIC will propose standardised lists (T3.4).
- Definition of good and poor performance: define indicator metrics and

their thresholds of evaluation. MARVIC will propose indicator metric lists and thresholds of evaluation (T3.2 and T3.4)

- Propose guidelines how to move on when no good calibration or validation data sets are available in the region yet for a given model domain.
- Define the minimum dataset size for cross-validation. This is beyond the scope of MARVIC.

4. Automation of input data and model outputs

When running an OPC over a project, automation is important as it facilitates reproducibility and transparency of the modelling workflow in a cost-effective way. Automation can concern both input data and model outputs. The automation of input data concerns any external data needed to run the model, in any format, such as farmer activity data, remote sensing layers, and soil/climate data coming from external data sources. Model outputs reporting is standardized to reduce user uncertainty and ensure provenance tracking. Currently, there are no established guidelines on how to conduct automation of model inputs and outputs.

There are no current requirements for transparent and standardised automation of OPCs.

Additional recommendations:

- Automated workflows should be publicly available, and version controlled.

5. Project modelling and project-level uncertainty estimation

Once a model has been shown to be valid within a certain domain of application and its data input and output have been automated in an OPC, it can be used at the project-level in an MRV digital tool to predict GHG emission reductions and removals following carbon farming practices. Once project locations have been modelled individually (e.g. for each field) to obtain predicted changes from baseline practices, they must be combined (i.e. scaled-up) to obtain the overall GHG mitigation predictions of the project (e.g. for the group fields of a farm or group of farms in the project). Scaling-up introduces additional uncertainty as only a proportion of the total project area can be measured/ modelled (e.g. field SOC stock is estimated from a limited number of soil samples). Scaling error, measurement error, and model prediction error must be combined in a process called ‘uncertainty propagation’ (or error propagation) to determine the precision of GHG emission reduction predictions for the project.

Error propagation through the OPC involves all the components of uncertainty propagation through the modelling workflow (see above), plus the scaling errors. Therefore, it is essential to keep the entire modelling workflow protocol (e.g. model version, input preprocessing, initialization, parameter sets) consistent with the validation proce-

dure to ensure propagated project errors will be compatible with what was demonstrated before.

OPC-level errors have a strong spatial component, and they should be assumed to have spatial dependence unless there is evidence otherwise. Scaling uncertainty should not only concern spatial uncertainty, but also temporal uncertainty, as model prediction errors tend to increase in time (i.e. over the duration of the project and beyond). This is currently an aspect that is not covered by guidelines.

Verra requirements (VM0042, Section 8.6.1):

- Scaling error, measurement error, and model prediction error are assessed separately for each SOC pool/GHG flux and combined via
- Analytical calculation of error propagation: estimated errors are combined to provide an estimate of the total variance of the mean GHG. Assumption that model errors are uncorrelated with the input data and there is no spatial autocorrelation among samples.
- Monte Carlo simulation: uses Bayesian methods to better addresses spatial dependencies in underlying data and skewed error distributions (e.g. Gurning, et al., 2020). Suggest between

500 and 1000 simulations.

Additional recommendations:

- Keep the project workflow consistent, as for validation actions described above.
- Rather than assuming complete independence, investigate spatial autocorrelation of model errors.
- If spatial correlation of errors is diagnosed, correlated errors must be accounted for in calculations.
- Provide plots of error versus time to assess and substantiate the nature of the relationship between model prediction error and time.

- Use a conservative estimate of model prediction error over time.

Future work:

- Investigate potential role of temporal autocorrelation of model errors, instead of assuming that it is constant. MARVIC will address this as part of T3.3.
- Investigate how to aggregate uncertainty estimates from field to project level considering both random and systematic errors to assess whether assuming independent parcel-level errors underestimates project level uncertainty. MARVIC aims to address this as part of T3.3.

6. Verification

The accuracy of the OPC should be evaluated over time in a carbon farming project to ensure accuracy in the estimated SOC stock change, a process known as verification. Verification can be done for every field or every farm within a project through targeted re-measuring. Additionally, it could also be done in a standardised way on a network of well-monitored benchmark sites across Europe, such as lighthouses, on-farm networks, or soil monitoring networks. Public and private project developers could verify the accuracy of their monitoring approach (OPCs) on these benchmark sites, which would facilitate transparency and comparability of their performance.

Verra requirements (VMD0042):

- Monitoring and verification are responsibilities of the project developers. Initial measures of SOC are taken at the project start. SOC is periodically (usually every 5 years) remeasured during the project duration throughout the project area to true-up modelled estimates.

Additional recommendations:

- Monitoring and verification of the project area is responsibility of the project developer. The intensity and frequency of sampling may be reduced from current



Verra requirements (in combination with the recommendation listed below).

- Verification will also be conducted in a standardised way across a network of benchmark sites in Europe. We recommend developing a network of standardised, well-monitored, benchmark sites for the verification of OPCs. These would need to:
 - be representative of pedoclimatic conditions and of farming practices of each Member State.
 - Ideally, present both standard and carbon farming practices so that the GHG emissions reduction and removal prediction uncertainty can be estimated

for both baseline conditions and the activity.

- Have a standardised data sampling protocol to broadly match the needs of models.
- Monitoring for verification purposes must be continuous and for a long period of time because SOC changes slowly and to capture the effects of climate change.

Future work:

MARVIC will propose in a future note the establishment of a network of benchmark sites.

References

- Buendia, E., Tanabe, K., Kranjc, A., Baasansuren, J., Fukuda, M., Ngarize, S., ... & Federici, S. (2019). refinement to the 2006 IPCC guidelines for national greenhouse gas inventories. IPCC, Geneva, Switzerland.
- FAO, 2015. World reference base for soil resources 2014: International soil classification system for naming soils and creating legends for soil maps (IUSS Working Group WRB No. 106). Food and Agriculture Organization of the United Nations (FAO), Rome. Available at: <https://openknowledge.fao.org/server/api/core/bitstreams/bcdecec7-f45f-4dc5-beb1-97022d29fab4/content>
- FAO, 2019. Measuring and modelling soil carbon stocks and stock changes in livestock production systems: guidelines for assessment. Rome, Italy. Available at: <https://openknowledge.fao.org/items/4148cd2c-e08d-4022-ab3f-26dc2efe9312>
- Gurung, R.B., Ogle, S.M., Breidt, F.J., Williams, S.A., Parton, W.J., 2020. Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty. *Geoderma* 376, 114529. <https://doi.org/10.1016/j.geoderma.2020.114529>
- Dietze, M.C., Thomas, R.Q., Peters, J., Boettiger, C., Koren, G., Shiklomanov, A.N., Ashander, J., 2023. A community convention for ecological forecasting: Output files and metadata version 1.0. *Ecosphere* 14, e4686. <https://doi.org/10.1002/ecs2.4686>
- Lavallee, J.M., McConkey, B.G., Kyker-Snowman, E., Dietze, M.C., Harrison, M.T., Liu, K., Smith, W., Grant, B., Karunaratne, S., Marklein, A., 2024. Identifying challenges and advancing guidance for using process-based models in soil emission reduction and removal projects. Available at: <https://library.edf.org/AssetLink/henw61p8uk181u34rh2bk8y-3dwe7lp68.pdf>
- Tonitto, C., Woodbury, P.B., McLellan, E.L., 2018. Defining a best practice methodology for modeling the environmental performance of agriculture. *Environmental Science & Policy* 87, 64–73. <https://doi.org/10.1016/j.envsci.2018.04.009>
- Verra, 2023. VMD0053 Model Calibration, Validation and Uncertainty Guidance for the Methodology for Improved Agricultural Land Management, Version 2.0. Available at: <https://verra.org/methodologies/vmd0053-model-calibration-validation-and-uncertainty-guidance-for-the-methodology-for-improved-agricultural-land-management-v2-0/>
- Verra, 2024. VM0042 Improved Agricultural Land Management, Version 2.1. Available at: <https://verra.org/methodologies/vm0042-improved-agricultural-land-management-v2-1/>