Review of the Validation Report for DayCent-CR version DCR 1.0.2 (CAR1459_model_val_DayCentCR_1.0_20210921.pdf)

Proponent: Soil Metrics, LLC, Indigo Ag

Reviewer: Brian McConkey, PhD, Viresco Solutions

Summary

The validation report is satisfactory and meets the requirements of the Requirements and Guidance for Model Calibration, Validation, Uncertainty, and Verification For Soil Enrichment Projects, Version 1.0a and provides the model prediction error required for CAR SEP uncertainty calculations.

Some revisions to the Guidance are warranted for clarity, to provide better guidance to the model validator, and to fill some gaps, and to provide the data and information to help the project verifier assess if modelling application conforms to SEP requirements.

Guidance Section 2

The validation report ("Report") covers all necessary aspects of validation.

Information:

a) Model version was fully specified (version and build) so met that requirement
b) The model calibration procedure was well described so met that requirement.
c) Documentation of parameter set met the requirement that not finer than one LRR without clear documentation was met. The parameter set was general across all LRR and CFG so this requirement was easily met.
d) Justification for splitting data between validation and calibration was met. This was done through k-fold method so the splitting was automated and all data was used for validation.
e) Datasets met the minimum requirement for full citation, experimental locations, specific crops, and practices studies, LRR, IPCC climate zones, soil textures and clay contents, and number of observations. Additional data was provided.
f) Responsible party for validation. The organization, contributors and their specific roles were well identified (the roles are not required)
g) Model version and build was specified so this requirement was met.

The purpose of the Report was to validate a specific version of DayCent-CR model (version DCR1.0 build 1.0) was clear. Background on DayCent Model and this version was provided.

The intent of the validation is for application of the model for Indigo U.S. Project #1 (CAR1459). Therefore, this is a type 1, project specific, validation report.

The calibration procedure is well described as required.

The calibration was constructed with k- folds, with k=5, so met the requirement that data used for calibration was separate from that used for validation for each k fold.

This reviewer found it valuable to have listed in appendix the sites involved in each fold of the cross validation since that is relevant to validation.

The final parameter sets, including those not subjected to calibration, were provided with the validation documentation so that verifier can easily check that the same parameter sets resulting from the validation report were used for baseline and project estimation in a project. Due to the use of the k-fold method, these final parameter sets will not be identical to those used for validation, which have different calibrated parameter sets for each fold. Nevertheless, this reviewer accepts that the requirement is met.

The final parameter set was general across all CFG and PC so conformed completely to the recommendation that the sets be as general as possible.

The Bayesian method was used for model validation so met the Guidance encouragement to use this approach.

Guidance Section 3

Every of 14 additional practices, falling into 4 general practice categories was declared so this requirement was met.

The three crop functional groups, corn, soy, and wheat, were declared with the rationale for categorization declared. Therefore, this requirement was met. See annex B. for comments.

The 13 LRR were declared as well as the IPCC climate zones, so this requirement was met.

The soils were declared in terms of texture and clay content so this requirement was met.

Dataset

General

Dataset were developed of measurements of SOC change for each practice category and CFG. The practice change contained comparisons between at least two levels of practice within each category Therefore this requirement was met.

Further the dataset of met other listed requirements:

- From peer reviewed literature
- Sufficient data to be modeled.
- All studies used were reported
- Studies with stacked (multiple) practices were note sole data used to evaluate a practice category with the exception of soy x organic amendment. A deviation was requested for the

latter combination and the rationale for accepting this deviation is satisfactory for this combination. Therefore, this requirement is met with the accepted deviation request.

- Studies from outside the United States were in IPCC climate zones contained in the project domain
- Only data between two measurements at least 3 years apart.
- The proponents attested that all suitable data sources from the US within the project domain.

With the tabulated data in the report and supplementary data including the spreadsheet Validation_report_supp_dataset_details.xlsx and CAR1459_model_val_DayCentCR1.0.2_dataset.xlsx along with input data for each site, the sites were well documented so that this requirement.

The general procedures for estimating missing data were specified so met this requirement.

Although not a requirement for the Validation, the verifier could require that the transformation to derive the 30 cm SOC mass be provided and this reviewer noticed that information was not always adequately documented for each site in the supplementary data.

Specific dataset requirement

The project domain included 13 LRR so, according, the validation data set cover at least 3 of the declared LRRs so met that specific requirement.

All PCxCFG combinations had at least 3 textures spanning at least 15% clay content so that specific requirement.

Since, one generalized parameter data set was used, there was no need to have separate data sets to validate the parameters specific to subset of LRRs.

The documentation provided is very good including histograms of responses.

Reporting on model validation

This report is type 1, project specific.

There was no need for substitution for a crop that appears in the project.

The values for model prediction error that will be used for SEP Appendix D uncertainty deduction are both justified and stated. The frequentist method proposed using RSME across folds meets the Guidance Requirements that "uncertainty in a model's prediction is determined from comparison to measurements, required in this document to be the same measurements used to validate the model" Some revision to both the Guidance and SEP will be necessary to use the posterior prediction interval for the case of Bayesian calibration.

The calculation of the pooled measurement uncertainty (PMU) just meets the requirement of the guidance. However, there are weaknesses in how the Report address the PMU. The Parties for this Validation report chose to calculate the PMU from reported variance of each treatment in the pair. This is non-standard approach as the generally accepted science is to use the pooled error across all treatments as the best linear unbiased estimate of the uncertainty of the difference between any two treatments in a physical study. In fact, the Validation Guidance specifies using the uncertainty of the difference for the PMU that is most easily extracted from the treatment difference confidence limits.

Since the uncertainty of the difference is more commonly reported than the uncertainty of each treatment in the validation literature, the Report choice limited the number of studies that could be used to estimate the PMU that weakened the validation.

A rationale was provided that the data was not collected for the purposes of validation for SEP. Although it was not a priority of the proponents of the validation, the PMU used in the validation had minimal data support despite being important to assessment of bias as criteria for validation. In fact, although 45 separate studies were used for calibration/validation, only 6 were used for estimation of PMU and, for the 12 PCXCROP combinations validated, one study was used in all 12, one in 8, and one is 6 and 4 only used those same three studies to estimate the PMU. However, given the current lack of guidance around the importance of estimating this value, this reviewer has to accept the that the PMU estimation meets the minimum requirements of the Verification Guidance.

The presentation of the validation meets all the requirements. The presentation was concise while being complete so could serve as a template for other Validation Reports.

The validation met the criterial of mean bias less than the PMU clearly with some exceptions due to data limitations and idiosyncrasies. Consequently, the proponents proposed several important deviations from the Validation Guidance.

The first exception was the problem of validating legume (soybean) for nitrogen fertilization as that is there are insufficient studies to meet validation source needs if there was any N fertilizer applied to soybean in the project. The deviation was that this this combination be allowed in the project if N fertilizer has been validated for another CFG and the legume has been validated for cropping system changes. The proposal rationale is sound to tis reviewer.

The second exception was the model should not be applied to conditions where the model estimates of 5000 g m^{-2}. That magnitude of change SOC change results for organic amendments with large changes over multiple decades that was not well predicted by the model. The proponents provided a proposal that the model can be considered conditionally validated for predicted SOC change <5000 g m^{-2}. This is not an important restriction since the because the magnitude of the SOC change for which there was poor model fit greatly exceed any real SOC change every expected over the maximum 5-yr crediting period for the project.

The third exception was that the model be valid for organic amendment practice change for all project CFG if the model those CFGs have been satisfactorily validated in aggregate. The rationale for this is well presented, and is largely due to limited data, and this reviewer agrees that this lumping across CFG is an acceptable way to assess if the model is validated for organic amendments. In order to make the argument that the model was validated for organic amendments that the model can be considered conditionally validated for SOC change <5000 g m^{-2} for organic amendments for wheat. This rationale was considered acceptable to this reviewer.

Consequently, with acceptance of these exceptions by CAR, the model was deemed to have met the validation guidance requirements for PCxCFGxES for the project domain.

Appendix H contains many helpful suggested changes to Guidance Document.

Annex A Concerns shared with the proponents during review and the current status of their resolution. These concerns refer to prior versions of the validation report.

1) There was question what is the appropriate SOC value for validation as this was not crystal clear since at several places in SEP that mentioned simulating SOC stock themselves. Also, the term "SOC change" could be i) SOC change between baseline and project at one time, or ii) the SOC change over time between the baseline and project (note for first sampling after initial sampling this would be the same as i) but not the same for subsequent samplings) or iii) the SOC change for each treatment individually over time.

Resolution: The Parties for the Report received clarification from the CAR that is SOC difference between control treatment without new practice and treatment with the new practice at one time, this being the analog to the project minus the practice; i.e., meaning i) above.

2) This reviewer believed that the full list of final model parameters, including those not calibrated, should be provided with the validation report to be filed with CAR as part of the validation report. This complete set enables the verifier to assess that the requirement of the validation report that the "parameter sets used when validating the model are the same used when the model is applied to the simulate baselines and project practices" has been met.

For the same reason, a copy of the executable code was also requested and provided. Again, this would be filed with CAR and enable to the Project verifier, if deemed necessary, to check that the model version is the same. Actually, in thinking through the potential verification need, that may be difficult to test if identical as any change to data flow to and from the model for the project that requires code changes will change the executable version. Therefore, it is possible to have the identical version regarding SOC simulation process but the executable version is not bit-by-bit identical to that used for validation. Also, it would unreasonable to expect the verifier to be able to run the executable without all the file structures etc. set up correctly for that specific build. Without being requested, the Parties for the validation report provided the DayCent input files used in the validation sites. These are welcome addition, particularly, the input files as it would enable the verifier, if they deem it necessary, to have the project developer use the project version of the model to run these input files to ensure it reproduces the same modeled values as the version used for validation process. Another benefit of providing the input data is it enables to reviewer of the validation report to check if the description of site management and other data is reasonable according the relevant reference documentation.

In the latest round of review, the reviewer requested the model estimates with the final parameter set. This will enable the verifier to check that the model version is substantively the same as that used for the validation. This is done by having the project developer use the project version to estimate the validation SOC data points using the final parameter data set. These could be added to CAR1459_model_val_DayCentCR1.0.2_dataset.xlsx. This would be superior for verifier than having the executable code.

Resolution: The additional information was provided including the modeled data points.

3) initialization procedures were not sufficiently described.

Resolution: The proponents added a description of the initialization procedures be described including how they relate to those planned for estimating emissions/removals for the baseline and project.

4) The description calibration/validation refers to a fourth step of Monte Carlo simulation (MCS) for estimating SOC difference that, if following the cited reference of Gurung et al. (2020), is fundamentally, not required with the SEP methodology for uncertainty estimation that requires the model structural error, interpreted as model prediction error for validation data. A similar assertation is made under model prediction error section that the MCS of uncertainty propagation is useful. For the Validation Report, the inclusion of the MCS is superfluous because it not used to estimate the model prediction error. If being proposed as a deviation or as method for future consideration, it should go into an appendix.

Resolution: This text has been extensively revised and moved to Appendix H.

5) Requirements for Model Calibration, Validation, Uncertainty, and Verification document did not actually specify clearly that the validation was for SOC stocks change between new practice and a control treatment at one time. Some wording in the SEP itself could be interpreted that the estimate of the SOC stocks was required.

Resolution: The proponents went to CAR and received clarification that it was SOC stock change between with and without a practice at one time.

6) Adequacy of cross validation for model validation. The acceptability of cross validation was specified as acceptable in the Validation Requirements. However, although the utility of cross validation generally is well recognized to assess relative model adequacy when comparing between different alternate models, there was concern that cross validation was not statistically acceptable to evaluate validity of a single model compared with the classical method of validation using a completely independent evaluation (validation) data set from the calibration dataset. The facts that i) the Validation Requirements did not provide any references to support the claim about statistical adequacy, ii) the Validation Requirement authors were now applying those same Requirements for their own purposes, and iii) clear requirement in the Validation Requirements to have separate validation from calibration. After review of literature, this reviewer concluded that cross validation is statistically acceptable "when it is not practical to collect new data to test the model" (Snee, 1977), the situation that exists for validation of GHG emissions and removals for CAR SEP for initial model validation. The cross validation is very powerful technique to use all available data for both calibration and validation and may be required, particularly for non-CO2 emissions.

Resolution: The proponents provided much additional support.

7) The completeness of relevant peer-reviewed studies was not specified. Excluding studies could affect validation results. Without an attestation that all studies with data that meets criteria up to a publication certain date are included, the reviewer is obligated to do their own review to make sure that all eligible studies are included.

Action required: The supplementary CAR1459_model_val_DayCentCR1.0.2_dataset.xlxs lists a second set of sites called "not used" that includes a number of sites in the US in the declared domain of the

project. It is not clear why they are listed, but now the Parties need to explain why the listed US sites were not included.

Action required: Some sites CAR1459_model_val_DayCentCR1.0.2_dataset.xlxs had highlighted cells in yellow but no explanation about what the highlight referred to.

Resolution: The Parties for the Report included the attestation that all appropriate studies meeting specified criteria with publication up to end of 2020 were included.

Resolution: In revised dataset, CAR1459_model_val_DayCentCR1.0_dataset.xlsx, the sites not used were clarified that passed initial criteria for inclusion but did not meet eligibility criteria upon more detailed examination.

Resolution: In revised dataset, CAR1459_model_val_DayCentCR1.0_dataset.xlsx. highlighting was removed.

8) The use of a comparison of bias compared with PMU has issues. This is the first validation report and shows that the proponents found that 76% of data point do not have an uncertainty to estimate the PMU. The 24% that do have uncertainty are definitely not a sampling of all possible observations but are associated with particular studies such that some studies represent a large proportion of the data used for estimating the PMU. If the sites with extractible uncertainty happen to themselves bias the estimate of PMU, the validity of the validation is questionable. This reviewer looked at 5 studies at random that indicated that uncertainty of the difference could not be extracted and found that uncertainty could be extracted from two of those 5 studies (Varvel, 2006; Campbell et al. 2007). This is not to suggest that this is common that uncertainty may be available on many more of the studies but a careful read through references is necessary before claiming that the required uncertainty is unavailable.

Actions required:

a) Table 17 appears to provide the data used for estimating PMU. The N obs and N sites were not defined. To assess the suitability of the data for estimating PMU, the reviewer requires the following data: i) the total number of sites (experiments), the total number of sites for which at least one instance of PMU was extracted, ii) the total number of observations for SOC change in the CFG X PC across sites, iii) the number of observations for which PMU was extracted for with stacked observations used for more than one CFG X PC and iv) the same as iii) but without any stacked observations used for more than one CFG x PC. With this information, the reviewer is able to understand the nature of the data used for estimating the PMU.

b) For the Validation Report need to re-check all the papers without uncertainties for the PMU to confirm that the standard error of the difference is not available. Generally, it can be back calculated, or at least, estimated from the mean separation confidence interval, such as least significant difference. The focus should be on those CFG X PC for which there is <25% of observations having uncertainties.

c) Not yet requested of the proponents, list of the citations used for the PMU for each PCxCFGxES would be useful for the reviewer.

Resolution a): The table (#16 in revision), contains the number of sites used for PMU, the number of stacked sites.

Resolution b): No change in the number of data points for which uncertainty was extracted.

Resolution c): List of citations was provided in table (#16 in the revision).

9) Data documentation. The supplementary CAR1459_model_val_DayCentCR1.0.2_dataset.xlxs has a column for transformation to develop 0-30 cm SOC for studies that do measure that directly. This is useful information and important information for the review. However, in reviewing 5 studies for ability to assess PTDU discussed in section 4.4), two of the studies would have required a transformation (Khan et al. 2007 had measurement to of SOC concentration for 0-30 cm but did not present bulk densities to enable calculation to 30 cm, only sufficient information to quantify SOC stocks for 0-15 or 0-46 cm so a transformation was necessary. Campbell et al. 2007 only presents data to 15 cm so a transformation was necessary). Again, the fact two studies happened not to have sufficient documentation of the transformations do not indicate a problem but having the assumptions documented is important.

Actions required: The Parties to the report need to go through the references again and document how 30 cm was estimated when that data was not provided directly in the data in the reference

Resolution: The documentation was revised and provides better description of transformations used including for the data from the two references with noted problems.

10) Estimation and clarity of the model prediction error.

Action required: Using the model prediction error from the fit of validation data would be fully acceptable under the Guidance. Given the lack of clarity in Verification Guidance, the posterior prediction interval could be used for SEP Appendix D but that would need to be justified by showing it is more conservative in the context of SEP Appendix D, i.e., has a wider 95% interval.

Resolution: Additional detail and explanation of use of posterior prediction error as the model prediction error to be used for SEP Appendix D. Importantly, a reference to the conservativeness of this estimate was provided.

Action required: The verifier needs to see clearly the value of model prediction error that will be used for SEP Appendix D to ensure that value was used during the project until a new validation.

Resolution: New text added that clarifies what will be used for model prediction error for a potential revised SEP Appendix D for the cases of frequentist approach and for Bayesian calibration.

Action required: Clarity that the 90% interval used for validation is derived from the (k-1)-fold Bayesian calibration (i.e., not the interval proposed for SEP Appendix D).

Resolution: The text was modified so it is clear that the 90% intervals are based on the k-1 folds.

Annex B Items that were brought forward to CAR by reviewer from this review (May 18, 2021).

Number	Change	Rationale
1a	Define SOC change clearly as the difference without and with a practice change planned for the project	Clarity
1b	Provide references about the statistical adequacy of k-fold cross validation	Clarity and ensure reviewers are confident about the consistency about the principles of the Guidance and the use of k-fold cross validation.
2a	Specify that all the final model parameters, including those not subject to calibration are filed with CAR with the Validation report	One of requirements and needed by project verifier but not clear in Guidance that it needs to be all parameters.
2b	Specify that the that the Model input files used for validation are filed with the Validation Report.	Useful for Review of Validation Report, enables verifier to determine if the model version/build used for project reporting provides the same results as that used for validation
2c	Provide all modelled SOC data points for observed data points	Useful for the review of Validation Report and also enables the verifier to determine if the model version/build used for project reporting provides the same results as that used for validation
2d	Require that all data of full citation, experimental locations, extracted observed data points, specific crops and practices studied, LRRs and IPCC climate zones, soil textures and clay contents, and number of observations with the validation report. Remove requirement in Guidance document (requirements summary box) to keep additional data.	This with data under 2b and 2c, this would be complete data package and no need for verifiers to require extra data. Therefore, the proponents of model validation do not have to keep data organized for easy retrieval in case reviewers request information in the future.
2e	Specify that extra data required under 2a to 2d needs to be filed with the Validation Report and will only be available to reviewer and verifiers	The publicly available Validation Report needs to high level information to show that the model has been validated. However, there are intellectual property concerns with this detailed data so, since it does need to be publicly available, access should be limited by CAR. That detailed data only needs to be available to the Validation Report reviewer and the Project verifier.
2f	The guidance needs to specify that Information on assumptions and methods used to derive 0-30 cm SOC change for reference sources is required	The transformation to derive the 0-30 cm is important for the reviewer to assess the appropriateness of the use of the observations
2 g	Specify that the when k-fold validation is used properly, the model parameters	Clarify issue for reviewer and verifier.

Number	Change	Rationale
	used for project use do not have to be the same as those used for validation.	
2h	Specify that a description of the initialization procedures used for validation is required including how it that procedure will be similar to that planned for the project application.	Clarity to help reviewer and verifier.
3a	Revise the estimate of measurement variability now called the pooled measurement uncertainty (PMU). Rename the PMU as the pooled treatment difference uncertainty, PTDU, as that is both more accurate and more descriptive name.	The term pooled measurement uncertainty is misleading and confusing since the value required is not just the uncertainty due to measurement error and the actual measurement error is generally not quantified in references reporting on physical experiments. The lack of guidance on estimating the PTDU adds to the problem.
3b	Provide more guidance in the Validation Guidance on what constitutes an acceptable amount of data to derive a valid PTDU. There is no guidance now although does suggest option of using a default value when it is not possible to estimate the value from the validation references.	The PTDU is the essential statistic to assess validity of the model. However, there is no guidance on the criteria for the project developer to decide when there is sufficient extractible data from the validation data to use a PTDU derived from the that data. An appendix to the Validation Guidance that provides general guidance on extracting the PTDU from types of information common in published references reporting on experiments would be useful. Because of the principle of conservativeness, the guidance would include ways to estimate PTDU when the data may be unclear, such as using assumptions that least to conservative estimates, for example, assumptions that lead to a low value for degrees of freedom of the error terms so as not to overestimate the PTDU.
3c	The Validation Guidance provide default PTDU (currently called PMU in Guidance) that correspond to global best practice.	The PTDU is the essential statistic to assess validity of the model. The expectation is that the PTDU should be relatively stable across carefully management experiments so a default should be representative value for different experiments. However, deriving this value from actual validation data can be difficult. If the value is inappropriate because of limited data availability in the project domain, this affects the integrity of the validation. The Guidance

Number	Change	Rationale
		already suggests use of a default value when there is inadequate extractible data but leaves derivation of the default to the project developer. Validating the developer-supplied default value will be a challenge to the reviewer to ensure not gaming. Guidance-supplied defaults would simplify and speed the review process. Further, Guidance- supplied default could be used as part of criteria regarding adequacy of extracted data (see 3b above) – if the PTDU estimated from references is much larger than the default than requirements for extracted data would be stricter to ensure conservativeness is not being undermined (a project derived PTDS < default is more conservative than default so requirement could be less rigorous). Finally, for projects with small domains for which validation data may be limited, the use of default PTDU values will make the general value for PTDU and this would simply the evaluation of PTDU for reviewers, particularly for small-domain projects with limited validation data.
3d	For each PCxCFGxES, include in the Validation Report the a) number of single practice change and the b) number of stacked practice change studies, c) the number of observations used for PTDU estimated from validation data, and d) the citations that contributed values to the PTDU	The PTDU is an essential statistic and the reviewer needs to know the data support for any developer-provided value.
3e	Lower the requirement to validate the N management practice change (NM) for crop functional group of legume crops (LEG)	N management changes is not an important practice change for legumes. Therefore, there is little data for the impact of N management change or legume on GHG emissions or removals. However, N management changes could occur for legumes as part of cropping system changes involving legumes. This then provides assurance that reviewer/verifier will not declare the model implementation unvalidated for such cropping systems when there is the expected situation of insufficient data. This less restrictive validation option will quicken and simplify model validation.

Number	Change	Rationale
3f	Allow stacked PC to be used for each of the stacked practices for the CFG	The stacked practice changes are an important validation as stacking of practices is expected in most practices. There is no reason to assume that multiple use of stacked practices for validation is not conservative. The requirement that there needs to be at least one reference for the PC without stacking for each CFGxPCxES combinations would still be required. Note the special treatment of stacked studies for estimate of the PTDU under 3d.
3g	Lower requirement to validate organic amendments as practice changes. Allow validation if aggregate validation across project crop types is sufficient to validate organic amendments for each crop type.	The amount of data for the affect of organic amendments on emission sources is limited. At the same time the organic amendments change the soil but are not expected to interact with specific crop types compared to interactions with crop types for other practice changes. This problem is expected to occur with all projects involving organic amendments. Allowing this will make project validation more certain, simpler, and quicker.
3h	Deviation requests that the proponents will go to CAR regarding the validation should be specifically included in the Validation Report. The requests document in the Report should provide a thorough rationale based on science and validation data circumstances.	It is difficult for reviewer to complete their assessment on overall validity of model if the deviation requests are not known. With the deviation request specified, the reviewer is well placed to assess the merits of the request due to knowledge of the validation data and project domain. It then allows the reviewer to conditionally accept the validation (assuming the reviewer accepts the merits for the request), pending approval of deviation requests by CAR. Without this conditional approval option, the reviewer would have to reject validity of use of model for the full project domain. CAR should also benefit from having the reviewer assess the merits of the deviation requests as part of the assessment of the validation.
3i	More clearly specify the options for estimating the model prediction error for use in the project	What exactly constitutes model prediction error is sometimes not clear in the literature. In some contexts, it would be making a new prediction for different inputs at the validation sites (exactly the domain defined by the validation sites). In other contexts, it is making predictions for new sites that are

Number	Change	Rationale
		within a specified domain that includes the validation sites. Obviously, the latter definition is what is relevant for the project. Clarifying the accepted methods is therefore necessary. The appropriate model prediction error will be needed for any method of estimating uncertainty of project emissions and removals; this Guidance change thus is needed for any method of estimating project uncertainty specified in the Appendix D of SEP.

Annex C Items that will be brought forward to CAR by reviewer from this review (Oct 7, 2021).

Number	Change	Rationale
3g	Provide more guidance on the values for model prediction error to use in SEP Appendix D. SEP Appendix D may also need revision.	The verifier needs to confirm that the correct value is used and this is determined from the validation.
3h	More guidance and justification for using cross validation	Indigo has developed good draft revisions on this