SUMMARY OF PUBLIC CONSULTATION

VT0014 Estimating organic carbon stocks using digital soil mapping, v1.0

A draft of *VT0014 Estimating organic carbon stocks using digital soil mapping, v1.0* was open for public consultation between February 20, 2025 and April 4, 2025. This document includes a list of all comments received and the developer's response.

The structure of the developer's responses emphasizes that many comments raised the same question or set of questions. In addition to detailed responses, references are also included to the "GENERAL RESPONSES FROM DEVELOPER" section at the bottom of this document. These references are listed in the following format: "Response: See general response 'Alignment with existing VCS methodologies.'".

KEY QUESTIONS

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	**Organization**	**Comment**	**Developer's Response**
1	The Nature Conservancy	It is clear that the tool is intended to detail how the tool can be used within the Quantification Approaches allowed in Verra's ALM methodologies. It is not clear whether the tool can be used in other Verra methodologies where soil carbon is an included GHG pool (i.e VM0047).	The tool can be used with any VCS methodology that requires an estimate of SOC as a percentage by mass, SOC stock, bulk density, or how these quantities are changing over time. The tool is not specifically being developed to be forward compliant with revisions to existing methodologies, but where possible, the teams revising existing methodologies have contributed to the tool and we have attempted to avoid inconsistencies

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
			between the tool and future methodologies.
2	Ambipar	Certainly. The description of the proposed tool is aligned with VM0042 and VMD0053.	Thank you for your comment.
3	Boomitra	Yes	Thank you for your comment.
4	Anonymous #4	"This could be improved. Integration with existing VCS methods (VM0032, VM0042, VMD0053) will facilitate uptake, as the tool will provide another avenue for accessibility. This is particularly the case for extensive regions, such as rangelands, where costs of extensive soil carbon sampling can be fiNAncially prohibitive. Opportunities for improvement: 1.I am unsure why CN0137 is called a "tool". I interpret this document as another "method" designed to segue with existing VCS methods (see following comment). A "tool" would be a software, algorithm, model or similar allowing application. 2.It would be more appropriate to phrase CN0137 a "method". Note that a "methodology" is not the same as a "method". A "methodology" is a body of methods, rules, and postulates employed by	Within the VCS Program, the terms methodology, module, tool, verification and validation have specific and formal meanings. Under these definitions, CN0137 is classified as a "tool" because it is designed to support multiple methodologies rather than function as a standalone quantification framework. Validation refers to the third-party assessment of whether a project design meets Verra's rules and eligibility criteria, typically conducted prior to registration. Note that in CN0137 validation is distinct from model validation, as described in the draft of CN0137. Verification refers to the confirmation, by a VVB, that the project's reported outcomes, such as SOC stock changes, have occurred and are quantified correctly. The terms "stratum" and "quantification unit" are defined in VM0042. A stratum, as defined in VM0042 is a "subset of each quantification unit within which the value of a variable, and the processes leading to change in that variable, are relatively homogenous." A quantification unit, as defined in VM0042 is a "defined area within the project for which GHG emission reductions and carbon dioxide removals (reductions and removals) are estimated using the selected quantification approach." CN0137 uses definitions that are consistent with, but not identical to, VM0042 because, as a tool, it has been designed to support multiple methodologies. -- See general response "Alignment with existing VCS methodologies."

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
		a discipline. It is the study of methods. In contrast, a "method" is a particular procedure for accomplishing or approaching something, especially a systematic or established one. Many Verra "methods" suffer from such mislabeling: they are ultimately "methods" rather than "methodologies". 3.Mismatch/inconsistency between statistical evaluation criteria in VMD0053 and CN0137. Users of the DSM must align with VMD0053. Does this mean that proponents need to evaluate model fits across statistical evaluation criteria in both VMD0053 (MPE, MB, PMU, histograms of model residuals etc) as well as those in CN0137 (R2, 90% prediction intervals, lack of bias)? There is some duplication that should be avoided. Further clarity in CN0137 section 2.1.2 would be useful. 4.Section 3. Definitions of "verification" and distinction from "validation" would be useful. The method developer uses both such terms in CN0137, particularly in Section 5 in use case 2."	
5	Viresco Solutions Inc	The tool covers DSM for spatial and temporal SOC quantification and therefore aligns well with Verra methodologies that require such quantification. However, the tool is not well described for those not	Appendix 3 of the CN0137 draft made available for public comment contains a bibliography that serves as a primer on DSM for SOC quantification.

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
		already familiar with use of DSM applications for SOC quantification in both space and time. General reference information on DSM for SOC quantification focuses primarily on spatial analysis at a single point in time. Therefore, it may be difficult for many project developers to appreciate how the tool aligns with existing methodologies. An appendix acting as a primer on DSM for quantification of spatial and temporal SOC (with temporal aspects being emphasized including discussion of the most relevant literature examples) would help readers better understand how the tool may be incorporated into existing methodologies.	
6	Growindigo private limited	Yes, it is clear from the summary description on how the proposed tool aligns with existing Verra methodologies.	Thank you for your comment.
7	Anonymous #2	"Section 2 on: In general, the summary document aligns with existing Verra methods, in particular the requirements for process model usage e.g. the use of the Independent Modelling Expert and the Model Validation Report, the requirement for sampling from the project area for model validation and the intention of scientific rigour. However, there are instances where the tools do not align with existing Verra methods or where there are features of DSM that warrant further attention. As examples:	As noted in the draft of CN0137 made available for public comment, the requirement is consistent with VM0042, which states that sampling must be conducted within the same season. We are assuming that there are 4 seasons per year, which is why we selected a 12 week window (about 3 months) for the sampling window. We note that there are regions with < 4 seasons, such as parts of the wet tropics in Brazil more correctly characterized as having wet and dry seasons. In response to these comments we have removed the explicit time window and now use the phrase "same season," which is consistent with VM0042.

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
		Section 3. EAPE requires at least one peer reviewed publication which appears inconsistent with the requirements for a body of peer-review evidence (or other supporting data) in VM0042 and VM0053 e.g. "Shown in peer-reviewed scientific studies to successfully predict changes in SOC resulting from the changes in ALM practices". Section 5.2. 2. "Soil samples collected prior to the project start date may be used for model calibration, but validation samples must be collected within six weeks of the target date used for validation." This is not consistent with VM0042 and perhaps this should have the same wording i.e. VM0042 states that "… sampling must be conducted during the same season". The rationale is that season lengths vary with geography and other factors therefore some degree of flexibility is needed. Section 3. Prediction support unit. This term is confusing and warrants further explanation of why this term is required for DSM and how it differs from existing VERRA requirements for soil data e.g. soil samples from individual locations (e.g. cores) or from composites from multiple locations. It would help to have examples of how different PSU's area determined (based on land area and volume of soil, core or composite etc). It would also be helpful to align with VM0042 and	

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
		VDM0053 definitions for soil sample, sample point, quantification unit, etc, with PSU then a distinct factor that is composed of these. 2.1.1 "When credits have been issued in the absence of model validation, a cumulative carbon stock change adjustment ensures the accuracy of credits over the project's lifetime by reconciling previously generated credits against validated measurements." This seems at odds to the Verra Standard which states that it will not issue carbon credits (Verified Carbon Units or VCUs) based solely on a model before the project has been validated and verified by a third-party auditor. It is not clear how this corresponds to VM0042 or VDM0053 which do not discuss cumulative carbon stock change adjustments per se. 2.1. Use Case 2 (DSM to predict stock) has an implication that baseline control sites will be required (as per VM0042 Measure and Remeasure). There is some mention of these but not much detail. This warrants more explanation. If baseline control sites are to be the norm then will DSM be used for the interstitial years, to allow carbon credits to be issued? This is the implication of the diagram (Figure 1 top part) which shows a smaller number of samples being collected in between the 5-year points. But it is not clear from the text. Ultimately, it would appear that the validation requirements for DSM are setting different, potentially weaker, thresholds for model predictions compared to	

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
		process modelling – as outlined in more detail in the following sections.	
8	Hyphen Global AG	Yes	Thank you for your comment.
9	Anonymous #3	Yes	Thank you for your comment.
10	South Pole	Yes, it's clearly stated.	Thank you for your comment.
11	Anonymous #5	Additional requirements on calibration and validation of the DSM prior to every verification, is not necessarily aligned, as recalibration of the model might require resampling more frequently than every 5 years.	Recalibration of the DSM is not required prior to every verification. Under CN0137, recalibration is at the discretion of the project proponent, based on whether new data or other considerations warrant an update to model parameters. What is required, however, is that the model's predictive performance is validated using physical soil samples at least once every 5 years in accordance with VCS program rules. This ensures ongoing scientific rigor without mandating recalibration. This approach offers flexibility for proponents who choose to "freeze" model parameters, while still maintaining a robust standard for credit issuance based on empirical validation.
12	Anonymous #6	"In section 2.1. Key concepts and rationale, there are two main uses of the DSM tool: Use Case 1: DSM to initialize and/or true-up a BGCM. Use Case 2: DSM as a primary measurement approach. We would like to understand if the 'Measure and model' approach as it is described in	The tool provides new implementation pathways under existing quantification approaches, but does not replace current quantification approaches.

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
		VM0042 v2.0 and 2.1 will remain as an option. Or will this DSM tool replace it?	
13	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Please see our submission under the section 'General Feedback' below.	Thank you for your comment.
14	Bayer Crop Science	"Yes, the summary description does a decent job of outlining the alignment, but it could be strengthened. The question refers to DSM as a "tool." However, a more appropriate wording would be to refer to it as an "approach." -Alignment is Mentioned: The ""SUMMARY DESCRIPTION"" (Page 4) explicitly states that the tool is designed to support the use of DSM when ""applying existing VCS methodologies."" It also mentions specific methodologies like VM0032 and VM0042. -Specific Applications: The summary clarifies how the tool supports different quantification approaches within VM0042 (Measure and Model, Measure and Re-Measure). -Areas for Improvement: The summary could benefit from a more direct statement of how this tool changes or augments existing methodologies. For example, explicitly stating that it provides a standardized approach to DSM calibration, validation, and uncertainty estimation, which may be lacking in current	Within the VCS Program, the terms methodology, module, tool, verification and validation have specific and formal meanings. Under these definitions, CN0137 is classified as a "tool" because it is designed to support multiple methodologies rather than function as a standalone quantification framework. Validation refers to the third-party assessment of whether a project design meets Verra's rules and eligibility criteria, typically conducted prior to registration. Note that in CN0137 validation is distinct from model validation, as described in the draft of CN0137. Verification refers to the confirmation, by a VVB, that the project's reported outcomes, such as SOC stock changes, have occurred and are quantified correctly. The terms "stratum" and "quantification unit" are defined in VM0042. A stratum, as defined in VM0042 is a "subset of each quantification unit within which the value of a variable, and the processes leading to change in that variable, are relatively homogenous." A quantification unit, as defined in VM0042 is a "defined area within the project for which GHG emission reductions and carbon dioxide removals (reductions and removals) are estimated using the selected quantification approach." CN0137 uses definitions that are consistent with, but not identical to, VM0042 because, as a tool, it has been designed to support multiple methodologies. -- The tool defined DSM as "the use of spatially explicit computer models to predict soil properties using gridded ancillary variables." We agree and have

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
		methodologies, would be helpful. Also, explicitly stating that the tool does not change existing rules for baselines, uncertainty deductions, etc. (as mentioned on Page 7) would provide further clarity. Additionally, it would be helpful to outline how this tool explicitly interacts with VM0053. -DSM, as defined in the summary, "is the use of spatially explicit computer models to predict soil properties using gridded ancillary variables." Examples emphasize the use of airborne and satellite remote sensing data as "gridded ancillary variables". Other forms of data, commonly collected "on-the-go"" in transects using ground-based sensors, have historically been used as DSM variables (eg. EMI, ECa, GPR, NIR, passive gamma). It's recommended to explicitly state if these types of data sources are included as DSM variables within the context of this tool.	clarified the tool in response to this comment.
15	NMI-agro	Yes this is clearly described in the documentation.	Thank you for your comment.
16	Seqana GmbH	"Yes. The current tool presents a big step forward in allowing for flexible models to derive DSM-based estimates for both use cases, QA1 - measure and model (as initial "measurement") and QA2 - measure and	See general response "Issues beyond the scope of CN0137."

Q1: Is it clear from the summary description how the proposed tool aligns with existing Verra methodologies?			
#	Organization	Comment	Developer's Response
		remeasure (as "measurements"). Until now there was a clear analogy between these VM42 defined QAs and the dichotomy of model-based and design-based inference to arrive at project-level estimates for SOC stock changes. The proposed tool breaks this analogy as it allows for model-based inference under QA2 - measure and remeasure. While we welcome the integration of innovative DSM approaches, we urge the authors to explicitly allow for design-based estimators to be permitted in parallel to the model-based estimation. In particular model-assisted design-unbiased estimators should be explicitly allowed, that make use of both ground-measured samples and a DSM model."	

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
17	The Nature Conservancy	The rigor of the revalidation requirements does not appear to be the same as the	See general response "Alignment with existing VCS methodologies" and general response "DSM under

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		rigor for revalidation requirements of biogeochemical models in VMD0053. Equations 1 and 2 suggest that DSM models are validated based on their predictions of SOC stocks at a single point in time. This is different than the requirements for biogeochemical models in VMD0053 which are validated based on their ability to predict changes in SOC stocks over time. It is critical to standardize this unit of validation across tools. DSM models that meet the proposed model validation requirements for SOC stocks may still calculate a highly inaccurate estimate of SOC stock changes since SOC stock changes are very small relative to background SOC stocks. All models should therefore be validated based on their ability to quantify changes in SOC stocks over time so that they can be used to accurately quantify and propagate uncertainty in project SOC stock changes to VCU calculations. The 5-year frequency for model revalidation is reasonable and similar to VM0042 requirements for the validation of biogeochemical models. However it is not clear whether the requirement to collect additional soil samples at every verification event applies to both Use Cases or only Use Case 2.	VM0042 version 2.1."

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
18	Ambipar	Yes, model recalibration and revalidation are in line with the requirements of VM0042 ensuring rigor. I believe that the same revalidation, i.e. at each verification, should also be applied to Quantification Approach 1 of VM0042 methodology.	See general response "DSM under VM0042 version 2.1."
19	Boomitra	The tool requires validation sampling to happen precisely within 6 weeks of the target date of SOC stock prediction for the verification period. This requirement makes large scale projects very difficult to implement under this tool. Validation sampling is required to be representative of the different strata present in a project, but often, in large scale projects, it may be difficult to define the relevant strata so quickly after the start of the project. Occasionally, the strata that are defined for the project may even keep changing until the completion of the validation audit. Furthermore, for projects with significant geographical spread, different farmers have different start dates and cropping calendars, sometimes making the project-level target date not necessarily the ideal time for sampling certain farms. For example, in Boomitra's India project (VCS 2889), which has 100,000+ farmers enrolled, farmers in South India experience different growing seasons than their counterparts in North India, leading to different dates of practice implementation with a given year, and different ideal dates	As noted in the draft of CN0137 made available for public comment, the requirement is consistent with VM0042, which states that sampling must be conducted within the same season. We are assuming that there are 4 seasons per year, which is why we selected a 12 week window (about 3 months) for the sampling window. We note that there are regions with < 4 seasons, such as parts of the wet tropics in Brazil more correctly characterized as having wet and dry seasons. In response to these comments we have removed the explicit time window and now use the phrase "same season," which is consistent with VM0042.

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		for soil sampling. The huge geographical spread and number of farmers also prolongs the amount of time required to collect the necessary geospatial information from the ground (eg. precise farm polygons with all ineligible areas excluded) to conduct a thorough stratification. In our own experience in VCS 2889, it takes more than 6 months to collect all the necessary information for a proper stratification when a large batch of farmers (eg. 100,000+) are onboarded into the project in a given year. Thus, 6 week is an insufficient time even for stratification in these cases. After stratification, it can take 2-3 months for all the sampling to be carried out by a trained field team, spanning remote areas of the country. If the vision of the tool is to encourage the creation of these types of large-scale projects, it must be feasible to implement in projects like VCS 2889. Furthermore, it may not be scientifically necessary to impose such a strict timing requirement for generating credits in a conservative manner. Increments in soil carbon stocks occur very slowly, and statistically significant SOC change may not be observable from physical soil sampling for multiple years. Thus, a soil sample taken a year from the target date is not likely to be different from one taken exactly on the target date. Additionally, if the SOC DSM model is cross-validated	

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		with the validation soil samples, and if we assume that SOC levels normally increase with the adoption of improved practices, then collecting the validation soil samples at some point after the start date of the practices would only direct the initial SOC predictions upwards and thus be conservative in the calculated SOC change and credits issued. Our suggested modification to this requirement is to have validation sampling carried out for a given cohort of farms by the time of validation of the cohort, as opposed to a stringent t0+/-6 weeks	
20	Anonymous #4	Suggested improvements: 1. Quality of data used for calibration. While digital soil carbon datasets abound, their quality, depth of measurement, and measurement approaches vary. Quality controls will be required to ensure that calibration data are not derived from digitized data but are physically sampled data. If this control were not imposed, it could result in the scenario where proponents use outputs from another model (e.g. digital soil data from a government website) to "calibrate" their DSM. 2."True-up" tends to be a US-centric term (section 2.1). "Re-calibration" may be more intuitive and accessible to an international audience. I note that in Figure 1 on page 8	Because the term is used within VCS methodologies, we use the term "true-up" in CN0137. We also use the phrase, "cumulative carbon stock change adjustment", which is formally defined in the CN0137 glossary. We note that "true-up" is not synonymous with "recalibration" or with "cumulative carbon stock change adjustment." A cumulative carbon stock change adjustment refers to a crediting process that corrects for systematic errors based on cumulative performance over time. This can involve adjustments to previously issued credits, and may or may not involve recalibrating the underlying model. A "true-up" is an update to the model prediction error. True-up can occur in the absence of cumulative carbon stock change adjustment and re-calibration. A related question is what would occur if project activities result in greater losses in SOC under project activities than under baseline conditions. CN0137 is designed to ensure that crediting is based on net positive change in soil organic carbon stocks, verified through field

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		in the second last column (lower panel), "re-validation" is mentioned. I don't see why a proponent would not be required to re-calibrate their approach (either DSM or biogeochemical model (BCGM)) when new data arise, because more measured data would be conducive to better predictive performance and more rigorous credit issuances. 3.Use case 1 (using a DSM to initialise a BCGM) could do with some further details. Does this mean that a DSM is used to initialise the BCGM at each re-calibration step? Or is the DSM only used once, at project outset? 4.Section 2.1.2 (model statistical evaluation criteria): I suggest that minimum measured data requirements be added; coverage and goodness of fit will be skewed by the quantum of data used for evaluation. The minimum sample size should be proportional to the number of independent variables and desired level of statistical power. A general guideline is to have at least 10 observations per variable. 5.Section 2.1.2 unclear how to proceed if one or more of these criteria are unsuccessful. This could be up to the discretion of the VVB and IME, as presumably greater uncertainty deductions will apply.	measurements and subject to conservative accounting. If a project demonstrates that SOC stock has decreased under project conditions relative to the baseline, no credits would be issued for that monitoring period. If the decrease leads to an over-crediting situation from earlier rounds (due to prior estimated increases), this discrepancy would be corrected via the cumulative carbon stock change adjustment when allowed by the VCS program and the applied methodology. -- One comment suggested that sample sizes for model validation should be related to the number of variables in the model (e.g., 10 samples per variable). We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137 does not require a fixed number of samples for model calibration, recalibration or model validation. Project-specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly, CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		6.Adding a requirement to validate the DSM for above-ground biomass would be expected to improve rigour of predicted SOC stocks. If a DSM performs poorly in prediction of a variable that is visible from satellite imagery (ie above-ground biomass), one might also infer that the DSM would similarly perform poorly in predicting a variable that is invisible to satellite imagery (SOC stocks). 7.It is unclear how over- or under- credit issuances in earlier crediting rounds will be treated when the model is re-calibrated/re-validated. Predicted SOC stock change will often differ with re-calibration; if on previous rounds credits were over/under issued, it should be clear how this will be accounted for in subsequent crediting issuances. The Australian Government 2015 Method previously used a regression approach to avoid excessive SOC crediting for changes in climate (rather than management) but this was subsequently removed from the 2018 and 2021 Methods to reduce complexity, noting that comparing two points in time was perceived a more accurate reflection of abatement (CER 2021). I later note footnote 14 on page 23. This is essential information and should be part of the main text of the method, not relegated to footnotes (see comment relating to	performance as new samples are added to the calibration. In all cases, the sampling plan must align with the target methodology. -- Several comments questioned what would happen if one of the three validation tests is unsuccessful. As described in the draft available for public and in the revised version of CN0137, credit generation under the tool requires all three tests to be successfully passed. A failure to pass compels the project proponent to recalibrate the model or cease activities under the project. This standard ensures integrity within the VCS carbon program. One comment asked why the bias test is not required between model validation events. We have clarified that results from all three tests must be reported in the DSM-MVR between model validation events. These summaries are not reviewed by the VVB until model validation, which must occur at least once every five years. -- The DSM models described in CN0137 are used to predict SOC stocks or SOC concentration directly, not aboveground biomass. Some DSM models use vegetation indices (e.g., NDVI) that can be proxies for aboveground biomass, but aboveground biomass is never predicted by a DSM in CN0137. -- Some comments noted that the term "true-up" may be more common in the United States. To improve clarity, we have used both the term "true-up" (which appears in VM0042) and the more descriptive phrase "cumulative carbon stock change adjustment", which is formally defined in the CN0137 glossary. We note that "true-up" is not synonymous with "recalibration." A cumulative carbon stock change adjustment refers to a crediting process that corrects for systematic errors based on cumulative performance over time. This can involve adjustments to previously issued credits, and may or

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		footnotes). 8.It needs to be clear what happens if the PCs cause greater losses in carbon than baseline conditions (between sampling events and overall).	may not involve recalibrating the underlying model. The key purpose of a true-up is to ensure credit accuracy over the life of a project, not to refine the model itself. This approach aligns with the VCS crediting framework, which requires that verified emissions reductions or removals reflect net cumulative performance over time. A related question is what would occur if project activities result in greater losses in SOC under project activities than under baseline conditions. CN0137 is designed to ensure that crediting is based on net positive change in soil organic carbon stocks, verified through field measurements and subject to conservative accounting. If a project demonstrates that SOC stock has decreased under project conditions relative to the baseline, no credits would be issued for that monitoring period. If the decrease leads to an over-crediting situation from earlier rounds (due to prior estimated increases), this discrepancy would be corrected via the cumulative carbon stock change adjustment ("true-up") mechanism. This ensures that only verifiable net gains in SOC over time result in credit issuance. -- See general response "Outcome-based validation" and general response "DSM under VM0042 version 2.1."
21	Viresco Solutions Inc	For many situations, including projects on smallholder farms, data availability will cause issues as there will be minimal training data on SOC stock changes to develop the DSM model. The first application of the model will focus on estimating SOC at a single point in time, meaning that the model will not yet test its ability to predict SOC stock changes. Therefore, it is worth recommending that	Under the tool, DSM is being used to predict SOC stock at multiple points in time. Changes in SOC stock are calculated using the empirical difference in these estimates of SOC stock and their associated variances. The variance in SOC stock at each point in time is estimated using geostatistical methods, and the variance in the change in SOC stock explicitly accounts for the covariance in the error of SOC stock over time. This ensures rigor in the estimates of SOC stock and stock changes and their associated uncertainties. -- One

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		project developers consider the first verification, shown in the Figure 1 example, as more than just a refinement of the DSM model. This step may reveal that the model does not inadequately predict SOC stock changes based on the established validation criteria. Changing the covariates including more time-varying factors may be necessary. Therefore, the recommendations should suggest that the first verification measurements may need to be more intensive than subsequent verification exercises. Also, it may reveal that some effort will be needed to collect and derive useful time-varying covariates. Under Procedures, Model Development (point 4), the prediction intervals are an important aspect of the model validation. The requirements are minimal, only requiring the method be used in at least one peer-reviewed publication, and is properly justified. However, discrepancies in the nomenclature between the publication and project documentation can create ambiguity regarding the method's application. To ensure clarity, the methodology must be explicitly documented using the project-specific nomenclature for data, making its application transparent. The treatment of uncertainty within the tool is different than VM0042 because it	comment noted that the nomenclature used by project proponents should match the nomenclature used in peer-reviewed studies that justify the use of a particular method. We agree and have clarified this in the text of the tool. Another comment noted the statement in the draft tool, "All software, computer code, data and other dependencies must be documented, archived, version-controlled and available on request." This comment asked to whom these would be made available. All software, computer code, data and other dependencies must be archived, version-controlled and available upon request by the IME, VVB, or VCS. -- One comment suggested that soil sampling variation should be incorporated into the tool. This comment is based on a misunderstanding. Soil sampling uncertainty is incorporated into CN0137. The measurement error component of soil sampling uncertainty is assimilated into model training and validation, and is represented in estimates of model prediction error. The spatial component of soil sampling uncertainty is handled through geostatistical methods. -- See general comment "Outcome-based validation" and general comment "DSM in smallholder farms."

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		explicitly includes the impact of spatial covariance on increasing the uncertainty of estimates of SOC stocks and their changes. The spatial covariance is ignored in VM0042, risking those uncertainties to be underestimated. In this regard, the tool provides a more conservative estimate of uncertainty of SOC stocks and stock changes than VM0042.	
22	Growindigo private limited	Yes, timeline for recalibration and revalidation are reasonable and scientifically justifiable, except for the six week window period required for validation.	As noted in the draft of CN0137 made available for public comment, the requirement is consistent with VM0042, which states that sampling must be conducted within the same season. We are assuming that there are 4 seasons per year, which is why we selected a 12 week window (about 3 months) for the sampling window. We note that there are regions with < 4 seasons, such as parts of the wet tropics in Brazil more correctly characterized as having wet and dry seasons. In response to these comments we have removed the explicit time window and now use the phrase "same season," which is consistent with VM0042.
23	Anonymous #2	5.1. The text would benefit from some rewording to make the processes clearer. We found it hard to follow the calibration and recalibration processes set out in section 5.1. with different views across our team around what is proposed. Over-fitting the DSM with recalibration looks like an almost inevitable consequence of recalibration (at least with the same validation set) and it is not clear how this can be avoided. The text states that	We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		recalibration is allowed [5.1.6)] if the coverage test fails. The risk of recalibrating until the coverage test passes, is that the model just gets fitted to the validation data (i.e. overfitting). The document acknowledges this: "… but users must avoid fitting the model to the validation set after the properties with the validation set are known." But it is not clear how this can be avoided under these conditions. 5.2.3. The number of samples required for recalibration is estimated to be at least 10% of the numbers used in the previous validation. The text would benefit from justification for 10% or, preferably flexibility for a project to determine (and justify) how much recalibration data is necessary, perhaps with 10% as a minimum. The guideline allows data from outside the project area to be used for calibration (Section 5.1, step 2). There needs to be assurances that these data are representative and not biased. If calibration data is biased (e.g. from certain soil type or land use) there is a risk of model could be overfitting on unrepresented areas. Regarding frequency, soil organic carbon changes occur slowly, often requiring 3–5 years to detect meaningful differences due to management practice changes. A 4-year revalidation with 2 -year recalibration cycle may therefore strike a balance in scientific rigor and feasibility.	does not require a fixed number of samples for model calibration, recalibration or model validation. Project-specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly, CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in performance as new samples are added to the calibration. In all cases, the sampling plan must align with the target methodology. -- See general response "Outcome-based validation."

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
24	Anonymous #3	Yes	Thank you for your comment.
25	South Pole	The frequency and requirements of calibration/validation can be clarified better, in particular for the timing of soil sampling requirements at validation and the case of project joint validation and verification. It's not unusual that in the Use Case 2 (DSM as a primary measurement approach) there could be the need to back-estimate the SOC stock and other variables for t=0. This is actually possible under the Quantification Approach 1 (Measure-and-Model) in VM0042. However, this scenario is not taken into consideration in the tool, so it should be explicitly stated if this scenario is contemplated. Clarifications should also be included regarding soil sampling timing requirements at validation, e.g. soil samples for model validation must be carried out before the project request to start the validation process to the VVB/VCS Standard (instead of mentioning soil sampling should be carried out at t=0). Moreover, in the Tool it's not considered the case when project validation and verification are taken out at the same time ("Joint Validation and Verification"). What are the sampling requirements in this situation? This could be clearly stated in	See general response "Frequency of model calibration, model validation, and verification."

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		the Tool.	
26	Anonymous #5	A re-calibration of the DSM will require that new literature shall be used for data collection,, which might be increasingly hard to find on a, for example, annual basis. On the other hand, collection of soil measurements on a more frequent basis that every 5 years will result in non-scalable approach, or result in low-quality data. Small annual increases in the soil due to a practice change might be too small to be detected on an annual basis due to within field SOC stock variation. There is no mention of back-modelling soil samples in this approach, must be collected at PAI's/strata's date/month of project entry (very restrictive).	See general response "Frequency of model calibration, model validation, and verification"
27	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Please see our submission under the section 'General Feedback' below.	Thank you for your comment.
28	Bayer Crop Science	The frequency and requirements are a balance between rigor and feasibility, but some aspects could be improved. Strengths: -The requirement for validation at the project's first verification and at least every five years thereafter (Page 5) provides a good baseline for ensuring long-term model accuracy. -The flexibility to recalibrate at each	We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		verification event (Use Case 2) offers project proponents options to update their models more frequently. -The cumulative carbon stock change adjustment (Pages 5-6) acts as a safeguard to reconcile previously generated credits. Potential Concerns and Suggestions: -Recalibration Requirements: The requirement that at least 10% of the number of samples used in the previous model validation must be used for recalibration (Section 5.2.3) seems reasonable. However, the tool should clarify the criteria for situations where the initial number of samples is very high. Perhaps a diminishing returns clause could be added (e.g., "…up to a maximum of X samples for recalibration"). -Model Validation Locations (Section 5.2.3): The requirement to resample the same locations as the prior model validation is a good idea in theory, but in practice, it might be difficult to achieve due to logistical constraints or changes in land access. The tool should provide guidance on how to handle situations where resampling all previous validation locations is not possible, beyond the current statement that the subset selected must be identified at random. For example, could a statistically representative subset be selected based on a documented	does not require a fixed number of samples for model calibration, recalibration or model validation. Project-specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly, CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in performance as new samples are added to the calibration. In all cases, the sampling plan must align with the target methodology. -- Several comments questioned what would happen if one of the three validation tests is unsuccessful. As described in the draft available for public and in the revised version of CN0137, credit generation under the tool requires all three tests to be successfully passed. A failure to pass compels the project proponent to recalibrate the model or cease activities under the project. This standard ensures integrity within the VCS carbon program. One comment asked why the bias test is not required between model validation events. We have clarified that results from all three tests must be reported in the DSM-MVR between model validation events. These summaries are not reviewed by the VVB until model validation, which must occur at least once every five years. -- See general response "Issues beyond the scope of CN0137."

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		rationale? -Recalibration Validation Tests: The tool requires the project-specific recalibration data to pass the model validation tests of coverage and R2 (see Section 5.1(10)), but the model validation bias test is not required between model validation events (Section 5.1.6). This seems inconsistent. If coverage and R2 are important, bias is also important. -Cost Implications: The sampling requirements for both validation and recalibration need to be carefully considered in terms of cost, especially for smallholder projects. While the tool aims to reduce costs through DSM, the sampling requirements could still be a barrier. Further exploration of cost-effective sampling strategies might be beneficial.	
29	NMI-agro	Five-year based intervals with a possibility for intermediate recalibration within that period suffices.	Thank you for your comment.
30	Seqana GmbH	Yes and no. Overall we commend the approach to validate the models on data from the project site itself. This solves the representation problem inherent to current model validation requirements under VMD53. We also appreciate the BIAS=0 and R2 > 0 requirements. However, from the current draft it seems that model validation is only ever required	See general response "Outcome based validation."

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		on a temporal snapshot of SOC stocks at single points in time and not directly on the SOC stock changes over time. The estimation bias should be evaluated at every verification/credit issuance event. This is important to avoid over- and under-crediting as much as possible. The critical estimates that must be unbiased are not mean model predictions at temporal snapshots, but ultimately must be the difference in SOC stock changes over time between project and baseline control site. While, this in theory can be satisfied by ensuring the unbiasedness of individual model-based mean estimates as proposed with t-tests, the suggested t-tests must be designed differently. Foremost, the relevant error in this scenario is they type II error of falsely accepting the null hypothesis with the associated beta probability. If multiple t-tests are to be carried out or if a validated model from such a t-test is to be re-used multiple times, the beta value must be appropriately small to avoid an overall high false negative probability of assuming that a 0 bias, when in reality it is not. As currently designed the probability for an overall Bias on a SOC stock change estimate can be up to 59%, even though	

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		the model validation criteria as laid out are all passed.	
31	ChrysaLabs	Suggestion #1: Section 5.1.6 : Recalibrations requirements and Section 5.2.3 : Sampling requirements for recalibration We recommend that Verra clarify the assumptions under which a 10% remeasurement rate is statistically sufficient to ensure low uncertainty. As shown in Potash et al. (2025), such low remeasurement intensity is only effective under specific conditions: large field counts (N > 10,000), randomized treatment assignment, and two-stage cluster sampling. Without these design features, a 10% sample could yield biased or high-variance estimates. Verra should specify minimum remeasurement thresholds based on project size, sampling variance, and desired confidence levels. The Verra tool should clearly specify under which conditions a 10% remeasurement rate is statistically valid and how it impacts uncertainty deduction : -Include a minimum sampling density per stratum when applicable. -Provide design sensitivity analyses, particularly when relying on remote sensing. -Clarify the types of landscapes or project	We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137 does not require a fixed number of samples for model calibration, recalibration or model validation. Project-specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly, CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in performance as new samples are added to the calibration. In all cases, the sampling plan must align with the target methodology. -- See general response "Issues beyond the scope of CN0137."

Q2; Are the frequency and requirements for model recalibration and revalidation reasonable, ensuring rigor with feasibility? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		sizes where 10% remeasurement is appropriate. Suggestion #2: Appendix 6: Uncertainty Calculation Example To support improved uncertainty modelling, we recommend including comparative scenarios showing how uncertainty is propagated in DSM models using remote sensing alone versus those that integrate in-situ spectroscopy or traditional soil sampling. Appendix 6 provides a helpful starting point for calculating the uncertainty of SOC stock estimates. However, the current examples do not differentiate between uncertainty profiles associated with various measurement approaches, such as: - Remote sensing-derived covariates, - In-situ spectroscopy-based models, - Traditional sampling methods (e.g., core-based lab analysis).	

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Quantification Approach 1 of VM0042? If not, what changes would you suggest?

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Quantification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
32	The Nature Conservancy	No. Is Equation 6 meant to be the point of departure from the tool to VM0042? $CO2soil_{t,\Delta t}$ is similar but not identical to $\Delta CO2soil_{t}$ in VM0042 Equations 44 and 45. We would recommend clarifying and revising the equations in the tool so that the link to VM0042 is clear. additionally, it is unclear how the baseline scenario would be quantified over time when digital soil mapping is used under Quantification Approach 1 – Measure and Model? Would projects be required to establish physical baseline control sites to manage, sample and model over time? Or would digital soil mapping only be used to recalibrate biogeochemical model simulations of the project scenario? We would recommend the latter in the same way that VM0042 currently requires biogeochemical models to be "trued-up" with measurements from the project scenario every 5 years. Finally, the approach for incorporating the uncertainty in DSM predictions does not align with the framework for uncertainty in VM0042 v2.1. VM0042 is very clear about	See general response "DSM under VM0042" and general response "Alignment with existing VCS methodologies."

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Quantification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		which sources of uncertainty must be accounted for depending on the quantification approach used (i.e. QA 1 must account for Model prediction error, Sampling error, and Measurement error, QA2 must account for Sampling error and Measurement error). It should be clarified where and how the outputs of the equations in Section 5.1 of the tool plug into Section 8.6 of VM0042. This lack of clarity on how to integrate the tool with VM0042 will discourage project developers from adopting this tool.	
33	Ambipar	Yes, the proposed tool provides clear guidance for the overlay of the results obtained in the DSM and BGCM. The concern of both VERRA scientists and VVBs should focus on ensuring that the calibration and validation of both models is carried out following rigorously the criteria for stratification of the project area, soil sampling, SOC and BD analysis, and calculation of soil C stocks and uncertainties. As well as, ensuring QA/QC of the information generated and presented for BL and Project emissions.	Thank you for your comment.
34	Boomitra	We note that the tool's draft, as published for public comment, omits details on how certain aspects of the tool would translate to implementation under Quantification Approach 1 (QA1). Under QA1, DSM model predictions of SOC are used as an	See general response "DSM under VM0042 version 2.1."

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Quantification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		input to a process based model. In this scenario, the error propagation associated with the DSM must include all aspects of spatial aggregation before and after usage of the process based model, as there are non-zero covariances between different predictions and different sample units. If the process-based model is used with pixel-level predictions from the DSM as a direct input, then the uncertainty propagation must include spatial covariances in the aggregation of the process-based model outputs to the project level. If the process-based model is instead used at the sample unit level, there is an added spatial covariance to be accounted for to get from the pixel-level to the sample unit level. The tool draft provides some guidance in Appendix 5, but does not provide the remaining guidance to go from process based model output to the project level. We would	
35	Anonymous #4	1.Application of digital soil mapping (DSM) to a level with suitable rigour to meet the standards of the VCS – particularly VMD0053 – will require specialist expertise. This requirement may reduce accessibility. I note, however, that existing methods require similar levels of expertise to be able to participate. For example, proponents using Quantification Approach 1 require specialist process-based skills and knowledge to be able to satisfactorily	The DSM models described in CN0137 are used to predict SOC stocks or SOC concentration directly, not aboveground biomass. Some DSM models use vegetation indices (e.g., NDVI) that can be proxies for aboveground biomass, but aboveground biomass is never predicted by a DSM in CN0137. -- One comment asserted that DSM models are less accurate than process-based soil biogeochemical models because DSM is a "black box" and process-based biogeochemical models are based on biological and soil processes. There is no evidence that process-based biogeochemical

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Qua ntification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		meet statistical evaluation requirements of VMD0053. 2.DSMs often use satellite imagery, among other covariates, for training and validation. Such imagery is often optical and based on reflectance indices (such as NDVI). This means that there are two key sources of error: (1) in the prediction of above-ground biomass, (2) in relationships used to convert above-ground biomass (and other covariates) into soil organic carbon (SOC) stocks. This contrasts with predictions from biogeochemical model (BCGM), as most BCGMs are driven by climate data; one could argue that climate data are generally more robust than remote sensing data, as the latter are measured closer to the project site, and are more likely to be contiguous (cf. satellite imagery which is often obscured by clouds). 3.Related to the previous point, DSMs have the advantage of capturing spatial variability, but they are less accurate than BCGM for predicting temporal change (because the latter incorporate biological and physical plant and soil processes, whereas the former are primarily based on black-box statistics). Incorporation of intuitive plant or soil processes within DSMs can significantly improve the prediction of SOC (Yang et al. 2025). 4.The requirement that all validation data must come from within the project area (5.1.2b, page 14) could be prohibitive for	models are more accurate than DSM, and it is debatable whether "black box" models are more accurate than process-based ones. Research in the last few years has demonstrated that data-driven methods have surpassed the predictive accuracy of process-based knowledge in numerous fields, including weather forecasting, skin cancer detection and protein folding. Please see the general comment "Outcome-based validation." -- See general response "Outcome-based validation" and general comment "Alignment with existing VCS methodologies."

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Qua ntification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		adoption, particularly if this data can only be SOC physical measurements. Many project proponents have difficulty sourcing measured SOC datasets, particularly at project outset. The uniqueness of each declared crop functional group and PC combination will further constrain SOC validation data. Provided SOC validation data are peer-reviewed, for the same soil type, IPCC climate zone, crop functional group and PC, I see no reason why validation data cannot be sourced from outside the project area.	
36	Viresco Solutions Inc	For clarity, a figure should be presented that demonstrates how the tool substitutes for conventional soil sampling under quantification approach 1 as currently described in VM0042. Although currently described in words, a figure would make the relationship more obvious. Alternatively, Verra could consider providing detailed examples of integrating DSM output with BGCMs while also clarifying the roles and responsibilities of IMEs and the project developers within an appendix. VM0042 states that "Reporting of SOC stock changes from direct measurements under Quantification Approaches 1 and 2 must occur on an equivalent soil mass (ESM) basis". This represents sound science as changes in soil bulk density	See general response "DSM under VM0042 version 2.1."

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Qua ntification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		over time will produce apparent changes in SOC stocks that confound the real change in SOC stocks. Expressing SOC stocks on an ESM basis greatly reduces the artifact of SOC stock change from changing bulk density. However, the proposed tool does not meet the VM0042 requirement to express SOC on an ESM basis. Under Procedures, 1.4 Calibration Data, point 5, the soil cores may be depth-aligned to target depths using a method described in at least one peer-reviewed publication. Such manipulation could make it impractical to reliably report SOC stock changes on an equivalent mass basis. Therefore, a section should be added which describes the details of how the data from the tool is converted to an equivalent mass basis to report SOC stock changes. This needs to be integrated with any "depth-alignment" method used. The method used for converting to equivalent soil mass needs to be described in at least one peer-reviewed publication and justified and presented using the nomenclature of the project documentation. The application of the tool for quantification approach 2 does not deal with BGCM but equivalent soil mass also affects SOC stock change under quantification approach 2. In this case, there can also be artifacts of bulk density change between the control sites and project sites due to	

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Quantification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		different land management, for example is soil disturbance changes in the project versus the control sites. So again, it is important that the tool be designed to provide SOC stock mass on a ESM basis.	
37	Growindigo private limited	To some extent yes, we have added some questions in general feedback.	Thank you for your comment.
38	University of Illinois Urbana-Champaign	No, I think two of the three model validation criteria are flawed and that there is insufficient validation of the geostatistical modelling component. I make specific comments on each of these below.	Thank you for your comment.
39	Anonymous #2	2.1. on While the document is clear about the use of DSM in BGCMs, it has provided no evidence that their use is appropriate for this purpose and that it would result in accurate predictions. As per VDM0053, there should be a requirement for a body of scientific evidence to demonstrate that use of DSM in BGCMs can predict soil C change reliably - either from peer review publications or relevant approved case-studies. A body of evidence will help to assure stakeholders that their combined use is fit for purpose. Whilst we appreciate the potential opportunities from linking these two modelling approaches, there is on-going (rapidly evolving) scientific debate about their joint predictive capacity. e.g. Xia et al 2025 Coupling Remote Sensing With a Process Model for the Simulation of	See general response "DSM under VM0042 version 2.1" and general response "Outcome-based validation."

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Qua ntification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		Rangeland Carbon Dynamic's - Xia - 2025 - Journal of Advances in Modeling Earth Systems - Wiley Online Library; Broeg, et al. (2024) Spatiotemporal Monitoring of Cropland Soil Organic Carbon Changes From Space. Global Change Biology https://doi.org/10.1111/gcb.17608. To support with assurance of linked models' performance – we would like to see clear emphasis on the use of independent testing (as opposed to validation). For example, DSM for soil carbon tends to produce under-dispersed predictions, which are typically reflected in low R2 such as those in Zhou et al. (2022), who reported a maximum R2 value of 0.44. This under-dispersion in model predictions can be problematic as input in further modelling processes and produce misleading results that would not be detected without the use of independent direct measurement. There is no mention of how to propagate uncertainty from the DSM into the BGCM. The predictions of the BGCM should also incorporate the uncertainty of the DSM modelled values. This is a key missing factor that needs incorporating. Appendix 5 of this document refers to VM0042 (Appendix 6) for sampling error propagation, employing a Monte Carlo method to estimate total error (sampling and model). The total uncertainty is decomposed into sampling and modelling	

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Qua ntification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		uncertainties using standard variance decomposition where sampling is assumed to be stratified sampling. This method's applicability to DSM is questionable and necessitates thorough elaboration in the DSM-specific document CN0137.	
40	Anonymous #3	Yes	Thank you for your comment.
41	Anonymous #5	Yes.	Thank you for your comment.
42	Anonymous #6	Yes, no suggestions to add.	Thank you for your comment.
43	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Please see our submission under the section 'General Feedback' below.	Thank you for your comment.
44	Bayer Crop Science	The guidance on integrating DSM with BGCMs is a good starting point but could be more detailed. Strengths: -The tool clearly outlines the two use cases: DSM to initialize/true-up a BGCM and DSM as a primary measurement approach (Page 5). -Appendix 5 provides a procedure for Monte Carlo error propagation when using DSM to initialize/true-up a BGCM. Areas for Improvement: -Uncertainty Propagation: While Appendix 5 outlines Monte Carlo error propagation, it would be helpful to provide more specific guidance on how to propagate the spatial	See general response "DSM under VM0042 version 2.1," general response "Issues beyond the scope of CN0137," and general response "Frequency of model calibration, model validation, and verification."

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Quantification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
		uncertainty of DSM predictions through the BGCM. The current guidance focuses on propagating the variance of the mean SOC stock, but it doesn't explicitly address how to account for the spatial covariance of model prediction errors within the BGCM. This is a crucial aspect of DSM and should be addressed more thoroughly. -Model Compatibility: The tool should provide guidance on selecting BGCMs that are compatible with DSM outputs. Not all BGCMs may be suitable for integration with DSM data. -True-up Frequency: The tool should clarify the relationship between the DSM model validation frequency (at least once every 5 years) and the BGCM true-up frequency. Should the BGCM be trued-up every time the DSM model is validated, or can the true-up occur more frequently?	
45	NMI-agro	How the robust amount of ground-truthed soil samples is determined (Figure 1, step 1 collection initial validation soil samples, example is 1000) is not sufficiently clear. See feedback on equation 8 in general feedback.	We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137 does not require a fixed number of samples for model calibration, recalibration or model validation. Project-

Q3: Does the tool provide clear and sufficient guidance on how DSM outputs should be integrated with BGCMs, such as under Qua ntification Approach 1 of VM0042? If not, what changes would you suggest?			
#	Organization	Comment	Developer's Response
			specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly, CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in performance as new samples are added to the calibration. In all cases, the sampling plan must align with the target methodology. -- See general response "Outcome-based validation."
46	Seqana GmbH	The guidance is likely sufficient. Two open questions remain: 1. Should a model validation in accordance with VMD53 in this case pass the appropriate VMD53 criteria with an initialization from a DSM as it shall be used in the project context? 2. If a DSM model considerably underestimates the sampling variance in terms of its point predictions, can it still be considered a valid source for initialization of a BGCM?	See general response "Alignment with existing VCS methodologies" and general comment "Model-validation criteria."

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
47	The Nature Conservancy	Yes, the concept in Section 2.1.1 is clearly articulated, and it is good that validation is required every 5 years in the same way it is required for biogeochemical models. However, we find the Use Case concept to be confusing alongside the current logic for Quantification Approaches in VM0042. Digital soil mapping is conceptually similar to biogeochemical modelling. In both cases a model is calibrated and periodically validated using measured SOC stocks and used to predict SOC stocks and stock changes in the carbon project. The main difference is simply the Nature of the model (remote-sensing-derived or process-based). We therefore think it's confusing and misleading to allow for digital soil mapping to be used under a 'Quantification Approach 2 – Measure and Remeasure'. Rather, it would make more sense to only allow digital soil mapping under Quantification Approach 1 – Measure and Model and expand that approach to clarify that both Digital Soil Mapping and Biogeochemical Modeling are allowable tools. We think it's misleading to say that DSM is a primary measurement approach in Section 2.1 when physical soil samples are required to be collected every time a	See general response "Alignment with existing VCS methodologies" and general response "DSM under VM0042 version 2.1."

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		project would verify credits.	
48	Ambipar	Yes, the concept of Validation-Oriented Approach with Safeguards is clearly articulated. I would like to reiterate that the concern of VERRA and VVB scientists should focus on ensuring that the calibration and validation of both models are carried out strictly following the criteria for stratification of the project area, soil sampling, SOC and BD analysis, and calculation of soil C stocks and uncertainties. As well as ensuring QA/QC of the information generated and presented for BL and Project emissions.	Thank you for your comment.
49	Boomitra	Yes it would.	Thank you for your comment.
50	Anonymous #4	1.Similar to process-based models, transparency pertaining to the DSM input assumptions will be critical. This is because models can be manipulated to give us any answer the user desires. As such, it is critical that input assumptions are defensible; preferably hinged upon peer-reviewed scientific literature. Transparency of, and rationale underpinning, model input assumptions are arguably even more important than model outputs or evaluation of the quality of simulations in relation to observed data. 2.Use case 1 (using a DSM to initialise a BCGM) could do with some further details.	We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137 does not require a fixed number of samples for model calibration, recalibration or model validation. Project-specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly,

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		Does this mean that a DSM is used to initialise the BCGM at each re-calibration step? Or is the DSM only used once, at project outset? 3.Section 2.1.2 (model statistical evaluation criteria): I suggest that minimum measured data requirements be added; coverage and goodness of fit will be skewed by the quantum of data used for evaluation. The minimum sample size should be proportional to the number of independent variables and desired level of statistical power. A general guideline is to have at least 10 observations per variable. 4.Section 2.1.2 unclear how to proceed if one or more of these criteria are unsuccessful. This could be up to the discretion of the VVB and IME, as presumably greater uncertainty deductions will apply. 5.More worked examples (perhaps added as appendices) would be valuable. This particularly relates to uncertainty of model predictions, uncertainty deductions and spatial covariance in model prediction errors. 6.Adding a requirement to validate the DSM for above-ground biomass would be expected to improve rigour of predicted SOC stocks. If a DSM performs poorly in prediction of a variable that is visible from satellite imagery (ie above-ground biomass), one might also infer that the DSM would similarly perform poorly in	CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in performance as new samples are added to the calibration. In all cases, the sampling plan must align with the target methodology. -- Several comments questioned what would happen if one of the three validation tests is unsuccessful. As described in the draft available for public and in the revised version of CN0137, credit generation under the tool requires all three tests to be successfully passed. A failure to pass compels the project proponent to recalibrate the model or cease activities under the project. This standard ensures integrity within the VCS carbon program. One comment asked why the bias test is not required between model validation events. We have clarified that results from all three tests must be reported in the DSM-MVR between model validation events. These summaries are not reviewed by the VVB until model validation, which must occur at least once every five years. -- Some comments noted that the term "true-up" may be more common in the United States. Because the term is used within VCS methodologies, we use the term "true-up" in CN0137. We also use the phrase, "cumulative carbon stock change adjustment", which is formally defined in the CN0137 glossary. We note that "true-up" is not synonymous with "recalibration" or with "cumulative carbon stock change adjustment." A cumulative carbon stock change adjustment refers to a crediting process that corrects for systematic errors

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		predicting a variable that is invisible to satellite imagery (SOC stocks). 7. It is unclear how over- or under- credit issuances in earlier crediting rounds will be treated when the model is re-calibrated/re-validated. Predicted SOC stock change will often differ with re-calibration; if on previous rounds credits were over/under issued, it should be clear how this will be accounted for in subsequent crediting issuances. The Australian Government 2015 Method previously used a regression approach to avoid excessive SOC crediting for changes in climate (rather than management) but this was subsequently removed from the 2018 and 2021 Methods to reduce complexity, noting that comparing two points in time was perceived a more accurate reflection of abatement (CER 2021). I later note footnote 14 on page 23. This is essential information and should be part of the main text of the method, not relegated to footnotes (see comment relating to footnotes). 8.It needs to be made clear what happens if the PCs cause greater losses in carbon than baseline conditions (between sampling events and overall).	based on cumulative performance over time. This can involve adjustments to previously issued credits, and may or may not involve recalibrating the underlying model. A "true-up" is an update to the model prediction error. True-up can occur in the absence of cumulative carbon stock change adjustment and re-calibration. A related question is what would occur if project activities result in greater losses in SOC under project activities than under baseline conditions. CN0137 is designed to ensure that crediting is based on net positive change in soil organic carbon stocks, verified through field measurements and subject to conservative accounting. If a project demonstrates that SOC stock has decreased under project conditions relative to the baseline, no credits would be issued for that monitoring period. If the decrease leads to an over-crediting situation from earlier rounds (due to prior estimated increases), this discrepancy would be corrected via the cumulative carbon stock change adjustment when allowed by the VCS program and the applied methodology. -- See general response "Outcome-based validation," general response "DSM under VM0042 version 2.1," general response "Issues beyond the scope of CN0137," and general response "Alignment with existing VCS methodologies."
51	Viresco Solutions Inc	Once the equivalent soil mass is dealt with (see above response to question 3), the approach is well described and provides good assurance of model validity for	See general response "Issues beyond the scope of CN0137."

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		application to VM0042. One potential concern is weak guardrails against applying new innovations – which may ultimately pose risks if not properly managed.	
52	Growindigo private limited	To some extent yes, we have added some questions in general feedback.	Thank you for your comment.
53	Anonymous #2	2.1.1. We do not see sufficient guidance in this document to demonstrate that a DSM or DSM-BGCM model will be good enough at predicting either stock or stock change since the text is confusing around the independent testing of DSM modelling performance. In essence, there appears to be confusion between establishing the validity of a model prediction versus the quality of a model prediction. We recognise that this document serves to address the requirements in VMD0053 v2.0 that a validated model demonstrates "satisfactory performance in terms of goodness of fit and characterization of model prediction error." Therefore, it is vital that DSM models are tested using independent data to clearly demonstrate that models are capable of accurate predictions beyond the boundaries of their calibration and validation data. This requirement is not clear in the text. 5.1. We find that an independent dataset for validation and model verification is vital to obtain a true measure of model performance. Although	See general response "Outcome-based validation."

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		cross-validation (LOO or K-fold) are viable methods for validating model performance, we find them to be insufficient for the case of mapping soil organic carbon stocks. Based on studies conducted both in-house and through third-party model providers, we have evidence to suggest that sophisticated spatial models for carbon stock mapping rely on spatial autocorrelation, not covariate information, to make predictions in the project area. When compared to ordinary kriging (OK), sophisticated models display almost identical predictive performance to the autocorrelation-only model. As soil is highly variable and the dependence is non-stationary, we believe that the performance of these models is overestimated by non-spatial cross-validation techniques, i.e., k-fold cross validation, which could have a deleterious effect on the estimation of carbon stocks and, ultimately, carbon stock changes. The Model Validation Requirements in Section 2.1.2 are insufficient to prove the model is "accurate, precise, and unbiased" across the project area. The requirements mentioned in the document, i.e., coverage, goodness-of-fit and unbiased residuals, are a minimum check of the validity of the model. For metrics to inform the accuracy of the model, a value would have to be placed on error metrics to describe the quality of the approximation that the predictions	

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		represent. In the case of precision, there must be guidance as to the width of the confidence interval as well as well as the expectation of the R2, given that this, too, will inform the quality of the predictions. Guidance should make it clearer that these metrics can only speak to the validity of the model predictions, not to the accuracy and precision, given that models can still be valid under the proposed criteria but still produce poor predictions. A model could meet all 3 model validation requirements, but the model predictions could still be far from the truth (i.e. accurately predicting ground-truthed data). For example, a model could be imprecise even if unbiased. Section 2.1.2 would benefit from details on how the DSM will be calibrated, validated and tested with reference to change in SOC stock over time as opposed to SOC stock, as per point 1 in the summary description. In Section 5.1 Model Development, point 12 provides the formulation for the calculation of the mean and variance of the change in SOC between t0 and t1. These equations provide the mean and variance of the difference between two random variables, it neglects to include the uncertainty attached to predictions at t0 and t1. Without the propagation of uncertainty from modelled observations at t0 and t1, the change between the two points can be misleading. Guidance should provide a	

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		protocol for the propagation of uncertainty either analytically or through Monte Carlo methods 2.1. on There is no mention of the importance of maintaining the appropriate scale during model validation and model prediction. Tan et al. (2023) shows how important scale is in the determination of covariate effects. If the model is validated in at a specific scale, then the model predictions must remain at the same scale and cannot be used for downscaling or upscaling the Section 4. "A calibration and validation dataset is used to calibrate and validate a model of SOC stock." There is some inconsistency between this statement and Section 3 where independent sample data are used for model validation. Section 3: "Independent sample data… Independent sample data are used for model validation". It would be helpful to establish whether the validation datasets is an independent hold-out set of data that have not been used in model calibration / validation.	
54	Anonymous #3	Yes	Thank you for your comment.
55	South Pole	It should be better presented what is meant for t0, e.g. it's the starting date of the project? it can be a date before project validation process start?	t0 is the first date of DSM measurement, whether or not that coincides with the start of the project.
56	Anonymous #5	Yes.	Thank you for your comment.

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
57	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Please see our submission under the section 'General Feedback' below.	Thank you for your comment.
58	Bayer Crop Science	The Validation-Driven Approach with Safeguards is generally well-articulated, but some aspects need further clarification. Strengths: -The emphasis on time-specific and project-specific model validation (Page 5) is a strong point. -The three validation tests (coverage, goodness of fit, lack of bias) provide a clear and objective framework for assessing model performance (Page 6). -The cumulative carbon stock change adjustment (Page 5) offers a crucial safeguard against over-crediting. Areas for Improvement: -Independent Modeling Expert (IME): The role of the IME is critical to the credibility of the tool. Appendix 1 outlines the IME's responsibilities, but it could be strengthened by specifying the IME's access to project data and models. The tool should explicitly state that the IME has the right to request any data, code, or documentation necessary to assess the validity of the DSM model. -Transparency: While the tool mentions that DSM-MVRs and IME assessment reports will be made public (Page 41), it should also emphasize the importance of transparency in all aspects of the DSM	See general response "Outcome-based validation" and general response "Issues beyond the scope of CN0137."

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		modelling process. This includes making the calibration and validation datasets, covariate data, and model code publicly available (subject to reasonable confidentiality constraints).	
59	NMI-agro	We offer the perspective having supported carbon offsetting projects from a scientific angle. In the general feedback we added safeguards for more rigor to increase assurances.	Thank you for your comment.
60	Seqana GmbH	Yes, it is clearly articulated. However, there is currently insufficient assurance as to the BIAS=0 test criteria as explained above. This criteria should ultimately apply to the SOC stock change quantification. With the current design of the t-tests for assuring BIAS=0, the probability for an overall BIAS can be up to 59%, while still passing the t-test criteria. In addition, it is unclear how the 90% coverage probability criterion provides safeguards for project level uncertainty estimates such as a confidence interval on a project-level estimate. Literature or a mathematical derivation supporting this assertion would be much appreciated.	See general response "Model-validation criteria."
61	Treefera	We value the outcome-based and validation-driven approach and the flexible choice of models it allows. The three model validation requirements are clearly	See general response "Model-validation criteria."

Q4: Is the Validation-Driven Approach with Safeguards concept clearly articulated, and would it provide the necessary assurances to you and your stakeholders?			
#	Organization	Comment	Developer's Response
		articulated and cover important aspects of model validity. However, $R^2 > 0$ as a minimum threshold is unexpectedly low. A model with an R^2 that is only slightly above 0 would not satisfy model quality standards. The reasoning for the approach not requiring a higher R^2 for models was explained in the workshop on the 6th of March (uncertainty deduction providing a robust mechanism for conservatism), but is missing in the document. Furthermore, we would welcome the provision of guidance on suitable R^2 scores, perhaps in the form of a recommended target range for model fit, to encourage best practices and to catch models with unsuitable performance early on in the process.	

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for smallholder farms?

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for sm allholder farms?			
#	Organization	Comment	Developer's Response
62	The Nature Conservancy	Yes, pending consideration of our feedback to the questions above.	Thank you for your comment.

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for smallholder farms?			
#	Organization	Comment	Developer's Response
63	Ambipar	Yes. I believe that the method described in the proposed tool can be applied worldwide. However, I am not sure about its application on smallholder farms, due to the fragmentation of agricultural management practices and land use history compared to the resolution of the spatial information needed to generate the DSM model. In this sense, the recommendation is not to use the proposed tool and opt for Quantification Approach 2 of VM0042.	See general response "DSM in small-holder farms."
64	Boomitra	Refer to our response to Question 2 above. The tool is generally written in an encompassing manner that enables applicability globally, with the exception of the t0+/-6 weeks sampling requirement.	Thank you for your comment.
65	Anonymous #4	1.Application of digital soil mapping (DSM) to a level with suitable rigour to meet the standards of the VCS – particularly VMD0053 – will require specialist expertise. This requirement may reduce accessibility. I note, however, that existing methods require similar levels of expertise to be able to participate. For example, proponents using Quantification Approach 1 require specialist process-based skills and knowledge to be able to satisfactorily meet statistical evaluation requirements of VMD0053. 2. I see this method as a subset - and thus another option with - existing methods	Thank you for your comment.

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for smallholder farms?			
#	Organization	Comment	Developer's Response
		(particularly VM0042). This will likely improve accessibility, particularly for farms in extensive zones with large area.	
66	Viresco Solutions Inc	The tool can use a spatially dense set of covariates, allowing it to capture the impact of vegetation, weather, and land management variations among smallholder farms, offering advantages over conventional soil sampling. However, as discussed in the question 2 response, the model may not always include the most relevant time-varying covariates for such heterogenous land use of small holder farms. This limitation is only revealed when actual SOC change data from verification and/or validation measurements are available. Nevertheless, DSM remains to be a more cost-effective approach when compared to soil sampling alone to quantify SOC change in this context.	Thank you for your comment.
67	Growindigo private limited	We think that the tool is a bit complex for implementation in developing countries with smallholder farms, particularly with high soil type/textural diversities and multi-crop, dynamic baseline, scenarios. We have presented some questions in the general feedback.	See general response "DSM in smallholder farms."
68	Anonymous #2	Theoretically the methods are applicable globally across different farming systems. There are many published examples of DSM being applied in diverse farming	See general response "Outcome-based validation" and general comment "DSM in smallholder farms."

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for smallholder farms?			
#	Organization	Comment	Developer's Response
		systems around the world. However, the challenge remains in demonstrating that the model predictions are accurate / reliable irrespective of location e.g. Potash et al. 2023 showed that the relationship between covariates and response (SOC) differs across sites. Smallholder farms can be mapped just so long as the model is fitted locally using data (covariates and response) at the same scale, and under the assumption of weak stationarity. Assuming the term smallholder farms means farmers in low- income countries in LATAM, Africa and Asia (predominately) then the challenge will also be obtaining sufficient contemporary spatial and temporal data, in particular field data, for model calibration, validation and independent testing (c.f. Cerratelli et al, 2021).	
69	Anonymous #3	Yes it will be applicable to smallholder farmers and use should be allowed worldwide.	Thank you for your comment.
70	South Pole	No significant comments regarding the worldwide applicability of the tool, but further comments regarding the applicability conditions in specific geographies and ecosystems (see the following feedback).	Thank you for your comment.
71	Anonymous #5	Potentially, with the right expertise.	Thank you for your comment.

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for smallholder farms?			
#	Organization	Comment	Developer's Response
72	Anonymous #6	In Section 5.2.1, the tool mandates the use of Expected Average Project Effect (EAPE) to determine the sample size for model validation, relying on peer-reviewed literature or proprietary data. However, this poses a challenge for regions like Latin America, where such data may be scarce or non-existent. An alternative approach could be to utilize regional expert knowledge, historical data from similar projects, or adapted values from analogous contexts, ensuring that the EAPE estimation is informed by the best available regional insights, even in the absence of peer-reviewed or proprietary data.	See general response "Outcome-based validation."
73	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Please see our submission under the section 'General Feedback' below.	Thank you for your comment.
74	Bayer Crop Science	The tool has the potential to be applicable worldwide, including for smallholder farms, but some adjustments may be needed. Strengths: -The tool's flexibility in model development (Page 6) allows for adaptation to different environmental conditions and data availability. -The allowance for composite samples (Page 24) is particularly relevant for smallholder farms, where individual field sizes may be small. Potential Concerns and Suggestions:	See general response "Issues beyond the scope of CN0137."

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for smallholder farms?			
#	Organization	Comment	Developer's Response
		-Data Availability: The tool relies on the availability of remote sensing data, geospatial data, and soil data. In some regions, particularly in developing countries, these data may be scarce or of poor quality. The tool should provide guidance on how to address data limitations, such as using alternative data sources, employing data imputation techniques, or adjusting the sampling design. -Technical Expertise: Implementing DSM requires a certain level of technical expertise, which may not be readily available in all regions or among all smallholder farmers. The tool should consider capacity-building initiatives to train local experts in DSM techniques. -Cost: As mentioned earlier, the sampling requirements could be a barrier for smallholder farms. The tool should explore cost-effective sampling strategies that are appropriate for small-scale operations. -Smallholder Heterogeneity: The tool must acknowledge that smallholder farms are often highly heterogeneous in terms of management practices, soil types, and environmental conditions. This heterogeneity needs to be adequately captured in the sampling design and model calibration.	
75	NMI-agro	Yes, as long as smallholder farms are aggregated and their areas are known. We do note that especially in the context of	See general response "Issues beyond the scope of CN0137."

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for smallholder farms?			
#	Organization	Comment	Developer's Response
		scope 3 in geographies in the global south (this tool is for offsetting, not insetting of course) in many cases the smallholder farm shapefiles are not exactly known. Clear delineation of management and control sites to prove additionality requires thorough guidelines, in particular when projects include groups of smallholder farmers, and the number of farmers might change over time. How is a change or update in farm area size dealt with? This is likely to occur, especially in projects in the global south.	
76	Seqana GmbH	Yes. The critical question with respect to smallholder farms is whether on a small spatial scale the micro- to meso-scale variability of SOC stock and SOC stock changes or model prediction error variances can be sufficiently captured with a reasonable sample size in order to allow for appropriate quantification. However, this issue applies to the wider context of SOC stock and SOC stock change quantification and extends beyond this specific tool.	Thank you for your comment.
77	ChrysaLabs	The DSM Tool appears calibrated to large, contiguous agricultural zones (e.g., U.S. Midwest). In fragmented agricultural systems, with small and diverse fields, the assumptions around spatial continuity, covariate coherence, and remeasurement intensity may not hold. We recommend Verra include adaptations or guidance for	The tool does not make assumptions about spatial continuity, covariate coherence, or remeasurement intensity, and these issues have no bearing on the effectiveness of DSM, which can be applied to large, contiguous areas or to small isolated fragments.

Q5: Do you believe the methods described in the tool are applicable worldwide within the defined conditions, including for sm allholder farms?			
#	Organization	Comment	Developer’s Response
		implementing the DSM Tool in such contexts, including increased minimum sampling rates, adjusted uncertainty thresholds, and allowance for in-situ spectroscopy to supplement or replace sparse spatial covariates.	

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer’s Response
78	The Nature Conservancy	We recommend that the tool include the concept of the validation domain that is included in VMD0053. This would clarify what geographies and soil types an approved DSM can be used in. Regions with low-resolution satellite data may struggle to meet certain criteria in this tool. We would recommend clarifying if/how data resolution is considered in uncertainty quantification and propagation.	See general response “Alignment with existing VCS methodologies” and general comment “Outcome-based validation,” and general response “DSM in smallholder farms.”
79	Ambipar	I believe that the proposed tool would not be applicable in regions where there is low resolution in geospatial information,	Thank you for your comment.

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer's Response
		regions with high interference for information acquisition, such as excess clouds and soil cover. Also, regions where there is great variability in soil types, depth and textures. In this sense, the recommendation is not to use the proposed tool and opt for Quantification Approach 2 of VM0042.	
80	Boomitra	Given the tree cover requirements, one would presume that this tool is not permitted to be used for orchards. It may be useful to explicitly clarify this. There are many existing VM42 projects with orchards. Boomitra has also developed a Social Carbon project (SC9) that has smallholder orchard enrolled and uses remote sensing of SOC. We would refer you to review that and consider the inclusion of orchards as well in this tool.	Numerous comments questioned the domain of application of DSM. Some questions the appropriateness of the < 30% tree cover threshold included in the draft version of the tool. Some requested justification for the threshold, while others argued that it was unnecessary or overly restrictive. In response to these concerns, we have removed the threshold from the tool. Project proponents will now be responsible for justifying alignment between the tool and the applied methodology at the time of project registration. This change does not signal an intention to expand the tool's scope to include agroforestry systems. Rather, the revision acknowledges that a fixed threshold is unnecessary to determine project eligibility. For example, an orchard with 50% tree cover may be consistent with the tool's design if the spatial arrangement of trees allows for ground-level observation from remote sensing between trees. Other comments asked for clarification about the types of soils that are amenable to DSM. DSM is applicable to all soils, subject to the constraints in CN0137.

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer's Response
81	Anonymous #4	Section 5. Use of multiple localized DSMs (e.g. for subset project area) would each require independent calibration and validation of that localized DSM. This is because each DSM is effectively an independent model. Having multiple DSMs will improve SOC prediction within a certain region, but each DSM must be validated for the crop functional group and PC in question.	See general response "Outcome-based validation."
82	Viresco Solutions Inc	In geographies with variable topographies (i.e. hummocky or hilly fields), soil erosion, which can cause both soil loss and deposition within the landscape (e.g. tillage erosion) will be difficult to include within DSM. The historical effects influence both SOC stocks and SOC stock changes. These effects are further compounded by variations in recent land use history, making high-quality training data difficult to obtain. More intensive soil sampling initially and during first verification and first validation exercises would be recommended to provide greater amount of SOC data to develop and test the model using cross-validation methods for these conditions. Incorporating data from erosion models and Digital Elevation Models (DEMs) could improve DSM's ability to account for erosion-driven SOC dynamics. It's important to note that it is likely still more cost-effective to account for soil erosion with DSM, compared to	Thank you for your comment.

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer's Response
		conventional soil sampling.	
83	Anonymous #2	The document does not state whether DSM is applicable to all soils. Peats and peaty (organo-mineral) soils can have unpredictable relationships with conventional covariates, making them particularly challenging to model. These soils are fairly common in cool temperate arable fields and grasslands. Guidance would be helpful for these soils e.g. peat soils are excluded (as per VM0042) but ideally guidance on how to address pockets of these soils within fields / PSUs. Land / soil cover is problematic for remote sensing (e.g. cover crops to permanent cover). Guidance would be useful for this. Agricultural land in "extreme" conditions e.g. steep slopes, extended flooding etc will be challenging and there may need to be exclude these features in the project area, or for the modelling to handle these as distinct areas.	Numerous comments questioned the domain of application of DSM. Some questions the appropriateness of the < 30% tree cover threshold included in the draft version of the tool. Some requested justification for the threshold, while others argued that it was unnecessary or overly restrictive. In response to these concerns, we have removed the threshold from the tool. Project proponents will now be responsible for justifying alignment between the tool and the applied methodology at the time of project registration. This change does not signal an intention to expand the tool's scope to include agroforestry systems. Rather, the revision acknowledges that a fixed threshold is unnecessary to determine project eligibility. For example, an orchard with 50% tree cover may be consistent with the tool's design if the spatial arrangement of trees allows for ground-level observation from remote sensing between trees. Other comments asked for clarification about the types of soils that are amenable to DSM. DSM is applicable to all soils, subject to the constraints in CN0137.
84	Anonymous #3	4.3 We question the 30% canopy threshold in the applicability conditions as there are researchers that have been able to successfully quantify and detect SOC changes where canopy cover is higher.	Numerous comments questioned the domain of application of DSM. Some questions the appropriateness of the < 30% tree cover threshold included in the draft version of the tool. Some

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer's Response
		This condition should be removed. If high tree canopy cover results in poor model fit it will be mitigated with confidence deductions.	requested justification for the threshold, while others argued that it was unnecessary or overly restrictive. In response to these concerns, we have removed the threshold from the tool. Project proponents will now be responsible for justifying alignment between the tool and the applied methodology at the time of project registration. This change does not signal an intention to expand the tool's scope to include agroforestry systems. Rather, the revision acknowledges that a fixed threshold is unnecessary to determine project eligibility. For example, an orchard with 50% tree cover may be consistent with the tool's design if the spatial arrangement of trees allows for ground-level observation from remote sensing between trees. Other comments asked for clarification about the types of soils that are amenable to DSM. DSM is applicable to all soils, subject to the constraints in CN0137.
85	South Pole	For rangelands and agroforestry concepts the methods proposed seems to be not well suitable. Rangeland areas with significant tree presence seem not to be well suited for the Tool. Tree cover can be less than 30% but mapped area definition could set out a large portion of project area, where tree presence is quite common. What are the requirements / guidelines to measure tree canopy cover and guarantee applicability in rangelands/agroforestry projects for instance?	Numerous comments questioned the domain of application of DSM. Some questions the appropriateness of the < 30% tree cover threshold included in the draft version of the tool. Some requested justification for the threshold, while others argued that it was unnecessary or overly restrictive. In response to these concerns, we have removed the threshold from the tool. Project proponents will now be responsible for justifying alignment between the tool and the applied methodology at the time of project registration. This change does not signal an intention to expand the

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer's Response
		Moreover, the covariates pre-processing procedure (5.1.3.1) doesn't clarify if there is a threshold for tree identification and if bushes should be also excluded from the mapping area. In the specific, what are the requirements/guidance conditions to define a "pixel containing tree", are remote sensing data products allowed and what defines a pixel containing a tree - especially for projects with agroforestry? It seems challenging to implement the tool for agroforestry contexts, since tree presence can be significant and there it's not clear how the mapped area and the covariates pre-processing procedures need to be considered.	tool's scope to include agroforestry systems. Rather, the revision acknowledges that a fixed threshold is unnecessary to determine project eligibility. For example, an orchard with 50% tree cover may be consistent with the tool's design if the spatial arrangement of trees allows for ground-level observation from remote sensing between trees. Other comments asked for clarification about the types of soils that are amenable to DSM. DSM is applicable to all soils, subject to the constraints in CN0137.
86	Anonymous #6	Regions facing data scarcity may lack comprehensive information on Soil Organic Carbon (SOC) data, particularly with regards to sampling dates. To address this limitation, a viable solution is to construct models utilizing aggregated data, potentially on a yearly basis, thereby enabling the development of reliable models despite data gaps.	See general response "DSM in smallholder farms."
87	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Please see our submission under the section 'General Feedback' below.	Thank you for your comment.
88	Bayer Crop Science	Conditions/Geographies: -Areas with Limited Data: As mentioned above, the tool may not be suitable in areas with limited remote sensing data, geospatial data, or soil data.	See general response "Issues beyond the scope of CN0137" and general response "DSM in smallholder farms."

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer's Response
		-Areas with Complex Topography: In areas with highly variable topography, DSM may be challenging due to the difficulty of accurately modelling soil properties in complex terrain. -Areas with High Cloud Cover: Persistent cloud cover can limit the availability of optical remote sensing data, which is a key input for DSM. -Areas with Permanent Flooding: The tool explicitly excludes project areas with permanent flooding (Page 13), which limits its applicability in some wetland environments. Suggested Changes: -Alternative Data Sources: The tool should provide guidance on using alternative data sources in data-scarce regions, such as citizen science data, participatory mapping, or historical records. -Model Complexity: In areas with limited data or complex environmental conditions, simpler DSM models may be more appropriate than complex machine learning models. The tool should provide guidance on selecting models that are appropriate for the available data and the complexity of the environment. -Integration with Local Knowledge: The tool should encourage the integration of local knowledge and traditional ecological knowledge into the DSM process. Local farmers often have valuable insights into soil properties and management practices	

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer's Response
		that can improve the accuracy of DSM models. Smallholders -Noise: The tool must acknowledge that smallholder farms are often highly heterogeneous in terms of management practices, soil types, and environmental conditions. This heterogeneity needs to be adequately captured in the sampling design and model calibration. -Project Scope: Defining project scope, practice, soil, and environment – ability to do that will be more important than in large conventional farming.	
89	NMI-agro	We've worked with projects globally, mostly in South America, Asia, Africa, North America and Europe. We do note that available remote sensing data quality and covariate abundance is not the same everywhere. The way the module is set up is modular so it should be able to work globally.	Thank you for your comment.
90	Seqana GmbH	What happens if a DSM cannot be validated? We believer there should be a clear path for project proponents to switch between DSM model-based estimates and design-based estimates involving ground measured data (e.g. Horvitz-Thompson estimator or Hansen-Hurwitz estimator (Cochran, 1977)). Such design-based estimates can be based purely on ground	Several comments questioned what would happen if one of the three validation tests is unsuccessful. As described in the draft available for public and in the revised version of CN0137, credit generation under the tool requires all three tests to be successfully passed. A failure to pass compels the project proponent to recalibrate the model or cease activities under the project. This standard ensures integrity within the VCS carbon program. One comment asked why the bias test is not required between model validation events. We

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?

#	Organization	Comment	Developer's Response
		measurements by way of a probability sample as described in the tool for the purpose of model validation. Alternatively, the design-based estimation can be model-assisted where the estimator remains design-unbiased by virtue of the probability sample and in addition the variance of the estimate (uncertainty) can be reduced through the model predictions (Brus, 2000; Dumelle et al., 2022; Viscarra Rossel et al., 2016). In such a case both the samples and the model could still be used for estimation, even if model validation fails. One strict requirement for the use of design-based estimators should be design-unbiasedness, which can be achieved by the criteria for sampling designs already contained in VM42 (probability sample) and expanded on in the present tool. In terms of uncertainty requirements such a design-based estimator should follow the same approach as laid out in equations 70/71 of VM42 v2.1 (December 2024) or as laid out in equation 5 of the tool CN0137 (the difference being that the variance components are not variances of the mean of predicted targets but variances of the mean estimates of the targets (i.e. SOC stock and SOC stock change)). Brus, D.J., 2000. Using regression models in design-based estimation of spatial means of soil properties. Eur. J. Soil Sci. 51, 159–172.	have clarified that results from all three tests must be reported in the DSM-MVR between model validation events. These summaries are not reviewed by the VVB until model validation, which must occur at least once every five years.

Q6: What conditions or geographies are there where the proposed methods would not be suitable? What changes would you suggest?			
#	Organization	Comment	Developer's Response
		https://doi.org/10.1046/j.1365-2389.2000.00277.x Cochran, W.G., 1977. Sampling techniques, 3d ed. ed, Wiley series in probability and mathematical statistics. Wiley, New York. Dumelle, M., Higham, M., Ver Hoef, J.M., Olsen, A.R., Madsen, L., 2022. A comparison of design-based and model-based approaches for finite population spatial sampling and inference. Methods Ecol. Evol. 13, 2018–2029. https://doi.org/10.1111/2041-210X.13919 Viscarra Rossel, R.A., Brus, D.J., Lobsey, C., Shi, Z., McLachlan, G., 2016. Baseline estimates of soil organic carbon by proximal sensing: Comparing design-based, model-assisted and model-based inference. Geoderma 265, 152–163. https://doi.org/10.1016/j.geoderma.2015.11.016	

GENERAL FEEDBACK

Section 2 Summary Description

Section 2 Summary Description			
#	Organization	Comment	Developer's Response
91	Boomitra	Boomitra thanks Verra for the opportunity to submit a public comment on the CN0137 tool draft. Boomitra is the first and only project developer as of today to get a project registered (ID 2887) under the VCS Standard that utilizes remote sensing to measure soil organic carbon stocks, which is one of the techniques that this proposed Verra tool provides guidance on. Additionally, before the time this tool reaches final approval, Boomitra will likely have 4 other projects registered under the VCS Standard that also utilize remote sensing to measure soil organic carbon stocks. Thus, Boomitra's public comment to this tool focuses on our learnings during the development, validation, registration and verification of real projects that leverage the systems that the tool provides guidance for. Firstly, Boomitra strongly supports the development and concept of the tool – it is a necessary endeavour to make all agricultural land management projects economic viable at large scale and at today's carbon market prices. Furthermore, it gives the necessary foundation for building long-term reversal monitoring systems for agricultural projects beyond	Thank you for your comment.

Section 2 Summary Description			
#	Organization	Comment	Developer's Response
		the project longevity period. Additionally, the tool is aligned with the structuring of existing Boomitra projects, notably including the validation sampling metrics and tests (eg. R2, RMSE, bias), and incorporation of spatial correlations in error propagation.	
92	Anonymous #2	We see great potential for the application of DSM in soil C projects, and aim to be constructive in our feedback to help see the methods progress to practical use. If it is helpful we are available for further discussions. Digital soil mapping ....predict soil properties using gridded ancillary variables ". DSM does not exclusively use grided ancillary variables and the text should not exclude the use of other data sources.	Several comments noted that DSM does not exclusively require the use of gridded input features or predictions on a grid. Although we anticipate that most users of the tool will use gridded input features, and will wish to generate predictions on a grid, the comments are correct, so we have modified the text in response to this comment. Examples of non-gridded inputs could be data from proximal sensors, crop type, and other ancillary data.
93	Anonymous #2	2.1.1. How will a buffer be determined from cumulative carbon stock change adjustments when there is no confidence / evidence that a non-validated model is reliably predicting change in soil C stocks? Calculation of SOC stock values. The document would benefit from stating the equations required to calculate soil C stock, perhaps by referencing VM0042? A related minor detail but it would also help to be clear that bulk density is fine fraction bulk density as per IPCC guidelines.	See general response "Issues beyond the scope of CN0137."
94	Anonymous #2	"The tool specifically addresses …collection of physical soil samples,	Some comments requested specific guidance on the appropriateness of laboratory methods for quantifying

Section 2 Summary Description			
#	Organization	Comment	Developer's Response
		including guidelines on … and the selection of laboratory methods." It is not clear what the tool is addressing here since we cannot find further information in the document about laboratory methods.	SOC content. The gold standard for quantifying SOC content at the present time is dry combustion. CN0137 does not make specific recommendations because access to laboratory methods for quantifying SOC content varies regionally, and must therefore be context dependent, and because new methods may become available over time.
95	Anonymous #4	1. Application of digital soil mapping (DSM) to a level with suitable rigour to meet the standards of the VCS – particularly VMD0053 – will require specialist expertise. This requirement may reduce accessibility. I note, however, that existing methods require similar levels of expertise to be able to participate. For example, proponents using Quantification Approach 1 require specialist process-based skills and knowledge to be able to satisfactorily meet statistical evaluation requirements of VMD0053. 2. Data quality for calibration and validation. While digital soil carbon datasets abound, their quality, depth of measurement and measurement approaches vary widely. Quality controls will be required to ensure that calibration data are not derived from digitized data but are physically sampled data. If this control were not imposed, it could result in scenarios where proponents use outputs from another model (e.g. digital soil data from a government website) to "calibrate" their DSM. 3. DSMs often use satellite imagery,	The DSM models described in CN0137 are used to predict SOC stocks or SOC concentration directly, not aboveground biomass. Some DSM models use vegetation indices (e.g., NDVI) that can be proxies for aboveground biomass, but aboveground biomass is never predicted by a DSM in CN0137. -- One comment asserted that DSM models are less accurate than process-based soil biogeochemical models because DSM is a "black box" and process-based biogeochemical models are based on biological and soil processes. There is no evidence that process-based biogeochemical models are more accurate than DSM, and it is debatable whether "black box" models are more accurate than process-based ones. Research in the last few years has demonstrated that data-driven methods have surpassed the predictive accuracy of process-based knowledge in numerous fields, including weather forecasting, skin cancer detection and protein folding. Please see the general comment "Outcome-based validation." -- Numerous comments questioned why CN0037 is called a "tool" and not a "method" or some other term, or criticized the use of the terms "verification," "validation," "stratum," and "quantification unit." Within the VCS Program, the terms methodology, module, tool, verification and validation have specific and formal meanings. Under these definitions, CN0137 is classified as a "tool" because it is designed to support multiple

Section 2 Summary Description			
#	Organization	Comment	Developer's Response
		among other covariates, for training and validation. Such imagery is often optical and based on reflectance indices (such as NDVI). This means that there are two key sources of error: (1) in the prediction of above-ground biomass, (2) in relationships used to convert above-ground biomass (and other covariates) into soil organic carbon (SOC) stocks. This contrasts with predictions from biogeochemical model (BCGM), as most BCGMs are driven by climate data; one could argue that climate data are generally more robust than remote sensing data, as the latter are measured closer to the project site, and are more likely to be contiguous (cf. satellite imagery which is often obscured by clouds). 4. Related to the previous point, DSMs have the advantage of capturing spatial variability, but they are less accurate than BCGM for predicting temporal change (because the latter incorporate biological and physical plant and soil processes, whereas the former are primarily based on black-box statistics). Incorporation of intuitive plant or soil processes within DSMs can significantly improve the prediction of SOC (Yang et al. 2025). 5. The requirement that all validation data must come from within the project area (5.1.2b, page 14) could be prohibitive for adoption, particularly if this data can only be SOC physical measurements. Many project proponents have difficulty sourcing measured SOC datasets, particularly at	methodologies rather than function as a standalone quantification framework. Validation refers to the third-party assessment of whether a project design meets Verra's rules and eligibility criteria, typically conducted prior to registration. Note that in CN0137 validation is distinct from model validation, as described in the draft of CN0137. Verification refers to the confirmation, by a VVB, that the project's reported outcomes, such as SOC stock changes, have occurred and are quantified correctly. The terms "stratum" and "quantification unit" are defined in VM0042. A stratum, as defined in VM0042 is a "subset of each quantification unit within which the value of a variable, and the processes leading to change in that variable, are relatively homogenous." A quantification unit, as defined in VM0042 is a "defined area within the project for which GHG emission reductions and carbon dioxide removals (reductions and removals) are estimated using the selected quantification approach." CN0137 uses definitions that are consistent with, but not identical to, VM0042 because, as a tool, it has been designed to support multiple methodologies. -- See general response "Alignment with existing VCS methodologies" and general response "Outcome-based validation."

Section 2 Summary Description			
#	Organization	Comment	Developer's Response
		project outset. The uniqueness of each declared crop functional group and PC combination will further constrain SOC validation data. Provided SOC validation data are peer-reviewed, for the same soil type, IPCC climate zone, crop functional group and PC, I see no reason why validation data cannot be sourced from outside the project area. 6.I am unsure why CN0137 is called a "tool". I interpret this document as another "method" designed to segue with existing VCS methods (see following comment). A "tool" would be a software, algorithm, model or similar allowing application. 7.It would be more appropriate to phrase CN0137 a "method". Note that a "methodology" is not the same as a "method". A "methodology" is a body of methods, rules, and postulates employed by a discipline. It is the study of methods. In contrast, a "method" is a particular procedure for accomplishing or approaching something, especially a systematic or established one. Many Verra "methods" suffer from such mislabelling: they are ultimately "methods" rather than "methodologies". 8.Similar to process-based models, transparency pertaining to the DSM input assumptions will be critical. This is because models can be manipulated to give us any answer the user desires. As such, it is critical that input assumptions are defensible; preferably hinged upon	

Section 2 Summary Description			
#	Organization	Comment	Developer's Response
		peer-reviewed scientific literature. Transparency of, and rationale underpinning, model input assumptions are arguably even more important than model outputs or evaluation of the quality of simulations in relation to observed data.	
96	Growindigo private limited	1.Justification for Validating Both BCGM and DSM in Use Case 1 Could you elaborate on how Use Case 1 benefits project proponents in terms of feasibility, implementation efficiency, and cost savings? In this case, project proponents are required to validate both the BCGM and DSM tools through IME. However, under the standard VM0042 Quantification Approach 1, only the BCGM typically requires validation. What is the rationale for this additional requirement, and how does it enhance the overall credibility or applicability of the approach? 2.DSM Validation Frequency in Use Case 2 In Use Case 2, DSM is used for direct quantification of soil organic carbon (SOC) stocks and stock changes, requiring validation at the start of the project and at least once every five years. Given that the measure-remeasure approach is already incorporated into VM0042, how does this additional validation requirement benefit the project proponent?	See general response "DSM under VM0042."
97	NMI-agro	This feedback was prepared by the entire NMI Wageningen Team: Sven Verweij, Yuki Fujita, Gerard Ros and Tessa van der	Some comments noted that the term "true-up" may be more common in the United States. Because the term is used within VCS methodologies, we use the term "true-

Section 2 Summary Description			
#	Organization	Comment	Developer's Response
		Voort. We are available for follow-up. This tool is already in excellent shape, the recommendations we have are for improved rigor, clarity and to be able to assure stakeholders and future credit buyers. Our experience is based on scientific (geostatistical) expertise, experience with remote sensing-based modelling and working with project developers as IMEs. Section 2.1. Use Case 1: DSM to initialize and/or true-up a BGCM "In Use Case 1, methods described in the tool provide SOC stock values for initializing and/or truing-up a BGCM. In this case, the DSM model is validated (and optionally recalibrated) at project start and at least once every 5 years. " Please clarify what you mean by a true-up. We recommend to include validation (ground-truthed) samples. If not, and a true-up is just remote sensing measurement, then you can run an entire carbon project just with remote sensing + biogeochemical modelling, without any ground measurements. This would not be robust and be very bad for assurance and trust in the method. Section 2.1: In VM0042 there is also back-	up" in CN0137. We also use the phrase, "cumulative carbon stock change adjustment", which is formally defined in the CN0137 glossary. We note that "true-up" is not synonymous with "recalibration" or with "cumulative carbon stock change adjustment." A cumulative carbon stock change adjustment refers to a crediting process that corrects for systematic errors based on cumulative performance over time. This can involve adjustments to previously issued credits, and may or may not involve recalibrating the underlying model. A "true-up" is an update to the model prediction error. True-up can occur in the absence of cumulative carbon stock change adjustment and re-calibration. A related question is what would occur if project activities result in greater losses in SOC under project activities than under baseline conditions. CN0137 is designed to ensure that crediting is based on net positive change in soil organic carbon stocks, verified through field measurements and subject to conservative accounting. If a project demonstrates that SOC stock has decreased under project conditions relative to the baseline, no credits would be issued for that monitoring period. If the decrease leads to an over-crediting situation from earlier rounds (due to prior estimated increases), this discrepancy would be corrected via the cumulative carbon stock change adjustment when allowed by the VCS program and the applied methodology. -- R2 was defined in Equation 2 of the draft of CN0137 available for public comment. -- See general response "Issues beyond the scope of CN0137" and general response "Model-validation criteria"

Section 2 Summary Description			
#	Organization	Comment	Developer's Response
		modelling which is allowed. In principle, it is possible to retrieve co-variates from a few years ago. Please consider adding a clause explain how this tool interacts with this possibility of VM0042. Section 2.1.2 Model Validation Requirements. Note that not every model gives you a confidence interval (CI). For instance, XGBoost. Please thus clarify how the CI is defined, and give examples for methods / approaches to assess the uncertainty of the predicted value. Please write the text so that it can cover models the various machine learning or deep learning algorithms such as XGBoost, neural networks, etc.. Section 2.1.2 Model Validation Requirements. Please also clarify in the text that though VM00042 there can be a penalty for a wide uncertainty range: otherwise there is a motivation to just make the uncertainty large so you always get 90% of predictions in 90 CI Section 2.1.2 Model Validation Requirements. Please add definition of r2. Is it the coefficient of determination? Or is it absolute?	

Section 3 Definitions

Section 3 Definitions			
#	Organization	Comment	Developer's Response
98	Anonymous #2	Expected average project effect. This introduces a non-standard (SI) term and adds to confusion when reading the document. It would help to have text which indicates why this term has been used and how it relates to VMD0053 requirements for effect size.	The expected average project effect (EAPE) was defined in the tool as "An estimate from at least one peer-reviewed journal article (or proprietary data where no peer reviewed journal articles are available) of the expected mean change in soil organic carbon (SOC) stocks in the next five years of the project, expressed in units of SOC stock. EAPE is used to guide sample size requirements for model validation." This term was introduced to guide users of the tool when selecting a number of physical soil samples to be used for model validation. This term has been removed from the revised version of the tool in response to public comments and discussion with the VVB. One comment requests that the tool clarify how the concept of EAPE relates to effect size in VMD0053. However, the term "effect size" does not appear in VMD0053 version 2.1. One comment asked whether the EAPE is equivalent to a minimum detectable difference (MDD). No, the EAPE is not equivalent to MDD. The EAPE is an estimate of the average project effect that was used to guide sample collection. It does not specify the sensitivity of the model, and has no bearing on the number of credits generated by the project. One comment asked whether alternative data sources, such as a biogeochemical simulation, could be used to estimate the EAPE. Yes, the EAPE could have been estimated using published or proprietary data of any type. We emphasize that the role of the EAPE was entirely to guide the number of samples collected for model validation. The model must then pass all validation tests independent of the EAPE. That is, if the project proponent selects a number of samples and subsequently fails to pass model validation, the project

Section 3 Definitions			
#	Organization	Comment	Developer's Response
			either ceases to move forward or must recalibrate the model until passage occurs.
99	Anonymous #2	What prevents the project operator from selectively providing a validation dataset that only includes points / PSUs that match the observed values well, and discarding those where the DSM value is very different to observed values? The lack of requirements for a certain size of validation dataset (relative to the variability of the carbon stock across the project) would make this very easy.	Some comments questioned how to prevent fraud within the tool by users who might try to manipulate the validation dataset, or augment validation locations with manure. We agree that ensuring the integrity of the validation dataset is essential to maintaining confidence in the tool. While it is true that all methodologies rely on the good faith of project proponents, CN0137 includes safeguards that help to prevent manipulation: (1) VVB and IME review: All validation datasets are subject to third-party audit by an approved VVB. The VVB and IME must assess whether the validation sampling plan is within the project area, rigorously defined, and properly documented. Any attempt to exclude data or locations without justification would be visible during this audit process. (2) Pre-specified sampling protocols: CN0137 requires that project proponents document the sampling design in advance (e.g., stratification, randomization). Exclusion of data after prediction, and without justification, is a breach of this protocol. (3) Transparency and traceability: All validation points must be geolocated and submitted as part of the project documentation. This creates a record that can be reviewed by VCS or third-party reviewers. (4) Statistical validation requirements: CN0137 does not prescribe a fixed validation sample size, but requires that the model meet quantitative benchmarks on an independent dataset. A small or manipulated sample is unlikely to pass these thresholds. These safeguards help to ensure that validation datasets are credible.
100	Anonymous #4	1. Section 2.1.2 (model statistical evaluation criteria): I suggest that minimum	One comment suggested that sample sizes for model validation should be related to the number of variables in

Section 3 Definitions			
#	Organization	Comment	Developer's Response
		measured data requirements be added; coverage and goodness of fit will be skewed by the quantum of data used for evaluation. The minimum sample size should be proportional to the number of independent variables and desired level of statistical power. A general guideline is to have at least 10 observations per variable. 2. Section 2.1.2 unclear how to proceed if one or more of these criteria are unsuccessful. This could be up to the discretion of the VVB and IME, as presumably greater uncertainty deductions will apply. 3. Mismatch/inconsistency between statistical evaluation criteria in VMD0053 and CN0137. Users of the DSM must align with VMD0053. Does this mean that proponents need to evaluate model fits across statistical evaluation criteria in both VMD0053 (MPE, MB, PMU, histograms of model residuals etc) as well as those in CN0137 (R2, 90% prediction intervals, lack of bias)? There is some duplication that should be avoided. Further clarity in CN0137 section 2.1.2 would be useful. 4. Section 3. Definitions of "verification" and distinction from "validation" would be useful. The method developer uses both such terms in CN0137, particularly in Section 5 in use case 2.	the model (e.g., 10 samples per variable). We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137 does not require a fixed number of samples for model calibration, recalibration or model validation. Project-specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly, CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in performance as new samples are added to the calibration. In all cases, the sampling plan must align with the target methodology. -- Several comments questioned what would happen if one of the three validation tests is unsuccessful. As described in the draft available for public and in the revised version of CN0137, credit generation under the tool requires all three tests to be successfully passed. A failure to pass compels the project proponent to recalibrate the model or cease activities under the project. This standard ensures integrity within the VCS carbon program. One comment asked why the bias test is not required

Section 3 Definitions			
#	Organization	Comment	Developer's Response
			between model validation events. We have clarified that results from all three tests must be reported in the DSM-MVR between model validation events. These summaries are not reviewed by the VVB until model validation, which must occur at least once every five years. -- Numerous comments questioned why CN0037 is called a "tool" and not a "method" or some other term, or criticized the use of the terms "verification," "validation," "stratum," and "quantification unit." Within the VCS Program, the terms methodology, module, tool, verification and validation have specific and formal meanings. Under these definitions, CN0137 is classified as a "tool" because it is designed to support multiple methodologies rather than function as a standalone quantification framework. Validation refers to the third-party assessment of whether a project design meets Verra's rules and eligibility criteria, typically conducted prior to registration. Note that in CN0137 validation is distinct from model validation, as described in the draft of CN0137. Verification refers to the confirmation, by a VVB, that the project's reported outcomes, such as SOC stock changes, have occurred and are quantified correctly. The terms "stratum" and "quantification unit" are defined in VM0042. A stratum, as defined in VM0042 is a "subset of each quantification unit within which the value of a variable, and the processes leading to change in that variable, are relatively homogenous." A quantification unit, as defined in VM0042 is a "defined area within the project for which GHG emission reductions and carbon dioxide removals (reductions and removals) are estimated using the selected quantification approach." CN0137 uses definitions that are consistent with, but not identical to, VM0042 because, as a tool, it has been designed to support multiple methodologies. -- See general response "Alignment with existing VCS

Section 3 Definitions			
#	Organization	Comment	Developer's Response
			methodologies."
101	University of Illinois Urbana-Champaign	It would be helpful to have more details about the prediction support unit, and its relationship to the support unit for the validation measurements. For example in the code appendix the prediction support appears to the a 10 m x 10 m pixel but the validation measurements appear to occur on the same support, which does not make sense and has important consequences (see #2). The tool should provide some specific guidance on choices of support and implications thereof. In the above example, I think a recommendation of taking multiple cores within each pixel for validation would be warranted.	As defined in the draft tool, a PSU is "The land area and volume of soil for which model predictions are calibrated and validated. The PSU refers to individual points or composite samples at specific depth intervals. For example, the PSU could be individual soil cores or composite samples from collections of soil cores over a specified depth range. Where the prediction support unit is the individual soil core, there must be a set of coordinates that define the collection location for the physical soil sample." This term and definition has been changed. We now use the terms "prediction support," and "prediction location." Prediction support is, "The soil volume to a specified depth or depth interval for which a model is calibrated and against which a physical measurement is compared during model validation." Prediction location is, "The location at which a prediction is generated by the model. For example, the model could be used to generate predictions on a 10 by 10 meter grid or at arbitrary locations." The PSU can use composite cores at the discretion of the project proponent. A separate comment requested that we clarify that each PSU can belong to only one stratum. In fact this was stated explicitly in 5.1.2.4 of the draft tool and remains in the current version.
102	South Pole	Not clear the definition. In the PSU definition it's mentioned that it refers to individual points or composite samples at specific depth intervals. So it's referred to the sampled points/plots? And how does it relates to the pixel size(s) of the DSM? It should be better explained to avoid	As defined in the draft tool, a PSU is "The land area and volume of soil for which model predictions are calibrated and validated. The PSU refers to individual points or composite samples at specific depth intervals. For example, the PSU could be individual soil cores or composite samples from collections of soil cores over a specified depth range. Where the prediction support unit

Section 3 Definitions			
#	Organization	Comment	Developer's Response
		creating more confusion with VCS methodology concepts as 'stratum' and 'quantification unit'. --- EAPE: "An estimate from at least one peer-reviewed journal article (or proprietary data where no peer-reviewed journal articles are available) of the expected mean change in soil organic carbon (SOC) stocks" It would be advised to require public documentation for proprietary data too.	is the individual soil core, there must be a set of coordinates that define the collection location for the physical soil sample." This term and definition has been changed. We now use the terms "prediction support," and "prediction location." Prediction support is, "The soil volume to a specified depth or depth interval for which a model is calibrated and against which a physical measurement is compared during model validation." Prediction location is, "The location at which a prediction is generated by the model. For example, the model could be used to generate predictions on a 10 by 10 meter grid or at arbitrary locations." The PSU can use composite cores at the discretion of the project proponent. A separate comment requested that we clarify that each PSU can belong to only one stratum. In fact this was stated explicitly in 5.1.2.4 of the draft tool and remains in the current version. -- Numerous comments questioned why CN0037 is called a "tool" and not a "method" or some other term, or criticized the use of the terms "verification," "validation," "stratum," and "quantification unit." Within the VCS Program, the terms methodology, module, tool, verification and validation have specific and formal meanings. Under these definitions, CN0137 is classified as a "tool" because it is designed to support multiple methodologies rather than function as a standalone quantification framework. Validation refers to the third-party assessment of whether a project design meets Verra's rules and eligibility criteria, typically conducted prior to registration. Note that in CN0137 validation is distinct from model validation, as described in the draft of CN0137. Verification refers to the confirmation, by a VVB, that the project's reported outcomes, such as SOC stock changes, have occurred and are quantified correctly. The terms "stratum" and "quantification unit" are defined in VM0042. A stratum, as

Section 3 Definitions			
#	Organization	Comment	Developer's Response
			defined in VM0042 is a "subset of each quantification unit within which the value of a variable, and the processes leading to change in that variable, are relatively homogenous." A quantification unit, as defined in VM0042 is a "defined area within the project for which GHG emission reductions and carbon dioxide removals (reductions and removals) are estimated using the selected quantification approach." CN0137 uses definitions that are consistent with, but not identical to, VM0042 because, as a tool, it has been designed to support multiple methodologies. -- One comment noted this sentence in the draft tool, "An estimate from at least one peer-reviewed journal article (or proprietary data where no peer-reviewed journal articles are available) of the expected mean change in soil organic carbon (SOC) stocks", and stated that it would be ideal to require public documentation for proprietary data too. However, proprietary data is by-definition not public. The concept being referred to by this comment has been removed from the tool, so this comment is no longer relevant.
103	Anonymous #5	- Central definition in Equation (7) is much more sensible that the equivalent in VM0042. - I would focus on the whole calibration curve (ROC or some such calibration metric) rather than just the 90% interval. - The comment in 5.2.1 1) "*This requirement indicates that rolling enrollments cannot be added to existing strata.*" is not entirely aligned with the VM0042 comment that "*strata should remain stable throughout the project's lifetime*" (or something like that) - I would expect, for a large scale project, that the	This comment noted an inconsistency in notation in Equations 11 and 12, Equations 9, 13 and 14, and in the definition of the variance of the prediction error of the mean SOC stock. We have carefully reviewed these issues and made corrections to the text where necessary.

Section 3 Definitions			
#	Organization	Comment	Developer's Response
		most scientifically valid thing to do at points in time is add new PAIs to existing strata. - The illustrative scenarios in Appendix 2 are quite helpful at clarifying the tacit assumptions to this approach (and VM0042) more generally. Primarily to do with verification frequency and documentation.' - Inconsistent notation for distance metric for the same thing in Equations (11) and (12). - Quantity defined in Equation (12) has two definitions: - "… variance of the mean SOC stock" in the text preceding it, - "Variance of the prediction error of the spatial average" The definition seems to be of the later rather than the former. - "Where measurements of the standardized prediction error" - you can't measure standardized prediction error, but you can take soil measurements and compute it. - Missing r in "Semivariance for the erros" in Equation (13). - Extra "in" for "Semivariance obtained with the pseudo cross variogram for in the errors" in Equation (14). - Inconsistent notation and definition of epsilon in Equations (9), (13), (14).	
104	Seqana GmbH	The term prediction support unit is defined as "the land area and volume of soil…" and "...refers to individual points or composite	Suggested that we avoid the abbreviation PSU for prediction support unit, because it can be confused with primary sampling unit. We agree and have implemented

Section 3 Definitions			
#	Organization	Comment	Developer's Response
		samples at specific depth intervals." (p.12). It is unclear whether this definition should really entail individual points consisting of a coordinate pair, as such points cannot have an area or volume > 0. Please, clarify that the prediction support unit refers to a pixel, block, or field with an area and volume soil > 0, that can be linked to an individual soil core point location by spatial relationships (e.g. contains, intersects). While the exact definition of this prediction support unit can be left to individual project proponents/MRV providers, there should be strict requirement for them to transparently define this as an areal unit with area > 0. This is relevant for uncertainty quantification, as the MC integration uses the block supports or area covered by the prediction support units for sampling. Additionally, we would suggest to refrain from the abbreviation PSU in this case, as this can be used in the related context of a multi-stage sampling design to refer to the primary sampling unit.	this change. This comment also suggested that we clarify that the prediction support unit is always a volume of soil linked to covariate data, and cannot be a point, since a point has no area. We have clarified this issue in the tool.
105	Treefera	Throughout the draft, the term "calibration" is appropriately used to describe the process of fitting the DSM model to known data, with the corresponding data subset referred to as "calibration data". In the machine learning and AI community, the terms "training" and "training data" are commonly used to refer to the same process and dataset. To support clarity—	We have clarified that calibration data is commonly called training data in the machine-learning community.

Section 3 Definitions			
#	Organization	Comment	Developer's Response
		particularly for interdisciplinary stakeholders—we suggest including a note in the Definitions section clarifying that "calibration (data)" in this document is equivalent to what is commonly known as "training (data)" in the machine learning community. Additionally, we note that in Section 5.1 Model Development, the term "training" is used once in place of "calibration". For consistency and to avoid confusion, we recommend using "calibration (data)" exclusively throughout the document.	

Section 4 Applicability Conditions

Section 4 Applicability Conditions			
#	Organization	Comment	Developer's Response
106	The Nature Conservancy	Can Verra clarify whether digital soil mapping as described in this tool includes proximal sensing?	Several comments noted that DSM does not exclusively require the use of gridded input features or predictions on a grid. Although we anticipate that most users of the tool will use gridded input features, and will wish to generate predictions on a grid, the comments are correct, so we have modified the text in response to this comment. Examples of non-gridded inputs could be data from proximal sensors, crop type, and other ancillary data.

Section 4 Applicability Conditions			
#	Organization	Comment	Developer's Response
107	Anonymous #4	1. Applicability conditions section 4: there are two number '1s' 2. Applicability conditions section 4: need to clarify if woody biomass (e.g. agroforestry or silvopastoral systems), application of organic amendments etc is out of scope. Here, it may be more consistent to refer to VM0042. 3. Treatment of baseline conditions/control sites could be clarified (5.1.13 implies that control sites may be used). Baseline control sites – variously known as "dynamic baselines" – where baseline activities are run in parallel to areas subject to management intervention – is strongly recommended. Use of a static baseline period (e.g. fixed period prior to project initiation) can mean that SOC measurement occurs during a period subject to weather anomaly (e.g. low SOC due to preceding drought). If by chance the PC coincided with the end of the drought (for example), SOC crediting may partly occur due to increased rainfall, rather than PC. Use of a paired baseline - where SOC is monitored on an adjacent parcel of land and compared with SOC on the project area at the same time - obviates the impact of climate on SOC. It is worth noting however that some proponents will not be able to run control sites, as this requires dedicated management over the long term, as well as significant land resources. Use of the second term in brackets in 5.1.13 (equation 6) disadvantages those	Noted this sentence in the draft of CN0137 available for public comment: "All samples must be matched with covariate features that coincide in space and time with the location and sample date." and requested clarification of its meaning. This statement means that covariates must be from the same time and location as a given physical soil sample being used for model calibration. For example, if a physical soil sample is collected on a given field on a given date, the covariates paired with that sample must coincide with the location and sample date. This would mean, for example, that it must be possible to extract the pixel value within which the sample is located. It would also mean that remote time-varying data must be coincident in time with physical soil sampling. If the sample had been collected today, but time-varying remote sensing data were acquired two years earlier, this criterion would not be satisfied. -- See general response "Issues beyond the scope of CN0137."

Section 4 Applicability Conditions			
#	Organization	Comment	Developer's Response
		proponents who do not use baseline control sites, because the second term is said to be set to zero in such cases.	
108	Growindigo private limited	12.Use of External Data for Calibration The document allows data from outside the project area to be used for calibration, while requiring all validation data to be sourced exclusively from within the project area. Could you clarify the conditions under which external data can be used for calibration? What safeguards are in place to ensure that external calibration data do not introduce biases that could affect the model's accuracy?	See general response "Outcome-based validation."
109	NMI-agro	Section 4: method is permissible for <30% tree canopy cover. Please add description how this is determined. Can examples be provided if this would cover or not cover e.g. cacao or coffee plantations? This is very important for implementation of this method for agro-forestry. These type of systems are prevalent in the southern hemisphere. By adding this context you can help project developers.	Numerous comments questioned the domain of application of DSM. Some questions the appropriateness of the < 30% tree cover threshold included in the draft version of the tool. Some requested justification for the threshold, while others argued that it was unnecessary or overly restrictive. In response to these concerns, we have removed the threshold from the tool. Project proponents will now be responsible for justifying alignment between the tool and the applied methodology at the time of project registration. This change does not signal an intention to expand the tool's scope to include agroforestry systems. Rather, the revision acknowledges that a fixed threshold is unnecessary to determine project eligibility. For example, an orchard with 50% tree cover may be consistent with the tool's design if the spatial arrangement of trees allows for ground-level observation from remote sensing between trees. Other comments asked for clarification about the types of soils that are amenable to DSM. DSM is applicable to all soils,

Section 4 Applicability Conditions			
#	Organization	Comment	Developer's Response
			subject to the constraints in CN0137.
110	Seqana GmbH	Update numbering: currently 1) 1) 2) 3) 4)	The numbering error in section 4 has been corrected.

Section 5 Procedures - Overall Section Feedback

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
111	The Nature Conservancy	Our overarching concern relates to our response to Question 2. Without updating the tool to validate DSM models against changes in SOC stocks over time, we are concerned that projects using the tool would have a lower bar to clear than projects using VMD0053 and therefore be open to legitimate scrutiny over the rigor of their claimed climate impact.	See general response "Alignment with existing VCS methodologies"
112	Boomitra	The tool requires potential IMEs to submit a Qualification Form and be approved by Verra before they may be contracted by the VVB in an audit. We feel that this procedure, while well intentioned, results in an unnecessary time lag between when the tool is published and when many projects can get validated under the tool. (Projects will have to wait until IMEs are approved	See general response "Issues beyond the scope of CN0137."

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		and available). Furthermore, it can cause significant audit delays for new projects if all Verra-approved IMEs are occupied with existing projects. We are starting to see a similar situation already, with a project backlog faced by IMEs under VMD53. The backlog is projected to worsen with the recent increase in scope for VMD53 to all soil methodologies. We believe it would be prudent to avoid creating the same situation for this tool. Our suggested alternative is that the VVB must independently select a DSM IME for the audit team assigned to validate/verify a given project, similar to how VVBs already arrange teams of experts today for audits across various methodologies.	
113	Anonymous #2	"The property being validated is always the prediction of SOC stock, even where separate components of SOC stock (SOC content and BD) are independently predicted. The variance of the estimate of the mean change in SOC stock over the duration of the project relative to baseline conditions is used to compute uncertainty and the associated uncertainty deduction under t applied methodology (see Sections 5.3 and 5.4." This section is not clear and may benefit from being split into two sections.	Thank you for your comment.
114	Anonymous #2	"All software, computer code, data and other dependencies must be documented, archived, version-controlled and available	Comment noted that the nomenclature used by project proponents should match the nomenclature used in peer-reviewed studies that justify the use of a particular

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		on request." To whom will these be made available?	method. We agree and have clarified this in the text of the tool. Another comment noted the statement in the draft tool, "All software, computer code, data and other dependencies must be documented, archived, version-controlled and available on request." This comment asked to whom these would be made available. All software, computer code, data and other dependencies must be archived, version-controlled and available upon request by the IME, VVB, or VCS.
115	Anonymous #4	1.Quality of data used for calibration. While digital soil carbon datasets abound, their quality, depth of measurement, and measurement approaches vary. Quality controls will be required to ensure that calibration data are not derived from digitized data but are physically sampled data. If this control were not imposed, it could result in the scenario where proponents use outputs from another model (e.g. digital soil data from a government website) to "calibrate" their DSM. 2."True-up" tends to be a US-centric term (section 2.1). "Re-calibration" may be more intuitive and accessible to an international audience. I note that in Figure 1 on page 8 in the second last column (lower panel), "re-validation" is mentioned. I don't see why a proponent would not be required to re-calibrate their approach (either DSM or biogeochemical model (BCGM)) when new data arise, because more measured data would be conducive to better predictive performance and more rigorous credit	See general response "Outcome-based validation." -- Some comments noted that the term "true-up" may be more common in the United States. Because the term is used within VCS methodologies, we use the term "true-up" in CN0137. We also use the phrase, "cumulative carbon stock change adjustment", which is formally defined in the CN0137 glossary. We note that "true-up" is not synonymous with "recalibration" or with "cumulative carbon stock change adjustment." A cumulative carbon stock change adjustment refers to a crediting process that corrects for systematic errors based on cumulative performance over time. This can involve adjustments to previously issued credits, and may or may not involve recalibrating the underlying model. A "true-up" is an update to the model prediction error. True-up can occur in the absence of cumulative carbon stock change adjustment and re-calibration. A related question is what would occur if project activities result in greater losses in SOC under project activities than under baseline conditions. CN0137 is designed to ensure that crediting is based on net positive change in soil organic carbon stocks, verified through field measurements and subject to conservative accounting. If a project demonstrates that SOC stock has decreased under project conditions relative to the baseline, no credits would be issued for

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		issuances. 3.Use case 1 (using a DSM to initialise a BCGM) could do with some further details. Does this mean that a DSM is used to initialise the BCGM at each re-calibration step? Or is the DSM only used once, at project outset?	that monitoring period. If the decrease leads to an over-crediting situation from earlier rounds (due to prior estimated increases), this discrepancy would be corrected via the cumulative carbon stock change adjustment when allowed by the VCS program and the applied methodology. -- See response to general comment "DSM under VM0042 version 2.1."
116	Growindigo private limited	3.Representative Soil Sampling and Correlation with Environmental Covariates The document does not provide a clear soil sampling plan, despite the critical role of representative sampling in Quantification Approach 2. Additionally, given that the tool relies on gridded ancillary variables to predict soil properties, how will these be effectively correlated with environmental and remote sensing covariates if soil sampling follows a stratified random approach, as outlined in VM0042, rather than a gridded method? Should gridded sampling be implemented to ensure consistency and accuracy, and if so, what standardized protocols should be followed? 4.Clarity on DSM Monitoring, Verification, and Reporting (MVR) Steps The document does not provide clearly defined DSM MVR steps, instead leaving implementation to the discretion of project proponents. This lack of specificity introduces ambiguity. Additionally, the tool outlines three validation requirements—coverage, R-square, and model prediction error—but does not provide sufficient supporting data or visual representation.	One comment suggested that sample sizes for model validation should be related to the number of variables in the model (e.g., 10 samples per variable). We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137 does not require a fixed number of samples for model calibration, recalibration or model validation. Project-specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly, CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in performance as new samples are added to the calibration. In all cases, the sampling plan must align

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		While Figure 2 is included, it does not adequately illustrate these aspects. Could these validation requirements be further elaborated with supporting data and visual examples, similar to the approach used in VMD0053? 5.Clarification on Statistical Tests for Model Prediction Error (Page 16) There appear to be two potentially overlapping statements regarding the assessment of model prediction error: •Point 8 states that a one-sample t-test with α = 0.05 should be used to determine whether the mean model prediction error is significantly different from zero. •Point 10(c) states that the model prediction error should not be significantly different from zero under a two-tailed, one-sample t-test. Could you clarify whether these statements convey the same meaning, or if there is a distinct difference that needs to be considered in order to meet all three validation tests of the model? 6.Clarification on Treatment of depth in the model (Page 23): The requirement of model calibration for each depth is clear, however, if the chosen BGCM model does not enable the input of SOC stock values of >10 cm, then what measures should the project proponent undertake for further comparisons and calculations?	with the target methodology. -- The draft tool referred to a t test on the model prediction error in two ways. One was a two-tailed, one sample t test and the other was a one sample t test. In fact these meant the same thing. We have clarified that the test is two tailed in both cases. -- One comment stated: "The requirement of model calibration for each depth is clear, however, if the chosen BGCM model does not enable the input of SOC stock values of >10 cm, then what measures should the project proponent undertake for further comparisons and calculations?" The DSM model must generate predicted values of SOC stock at the target depth or depth range required by the applied methodology. -- One comment noted the requirement in the tool that the year in which a prediction support unit was included in the sampling design must be the top-level stratum. This comment further noted the statement in the tool that rolling enrollments cannot be added to existing strata, and suggested that this is "a very heavy implication for project proponents," because "it is not always possible that all the enrollments happen" in a given stratum at once. However, this requirement simply means that enrollments in different years must be within different strata. This does not impose any special burden on project proponents, and ensures the integrity of strata over time. -- A worked example that includes variogram calculation is in Appendix 6. This example includes computer code illustrating the calculation of three distinct variogram models. Project proponents should verify that the locations selected will provide coverage of point distances in the 0 to 500 m range. If the sample does not provide adequate sampling within the 0 to 500 meter range necessary to fit the variogram, the project proponent should modify the sampling criteria using a clustered or multistage sampling design. We have

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		7.Clarification on Sampling design (Page 24): "The year in which a prediction support unit was included in the sampling design must be the top-level stratum. This requirement indicates that rolling enrollments cannot be added to existing strata". This is a very heavy implication for project proponents. It is not always possible that all the enrollments happen in a strata at once. Further clarification is required on this condition. 8.Clarification on Spatial Covariance Requirements (Page 25) The document states that when a variogram is required to account for spatial covariance in the variance of mean SOC stock, project proponents must ensure that a variety of inter-point distances (ranging from 0–500 m) are adequately represented among the sample locations. If the initial probability sample does not achieve this, additional sampling locations should be selected. Could you provide further clarification or guidance on how to effectively implement this requirement? A more detailed explanation with practical examples would be helpful. 9.Efficiency of Stratified Random Sampling in Large Project Areas (Page 26) The document states that in cases where a project area is very large or contains strong gradients in environmental variables or ALM conditions, applying stratified random	updated the text of CN0137 with this clarification. -- The draft of CN0137 stated that in very large projects, stratified random sampling is likely to be inefficient. Some comments emphasized the importance of random samples. Other comments questioned whether random samples are feasible. We believe that stratified random sampling (as defined in VM0042) can sometimes be inefficient because it requires project proponents to travel very large distances without collecting large numbers of samples. In our experience, sampling is a critical bottleneck. There is a limited time window when sampling is possible (e.g., after snow melts and before planting, or after harvest and before the ground freezes, with other constraints elsewhere in the world). Poor weather, difficulty gaining access to required areas, or disruptions to field crews can prevent sampling plans from succeeding. When random samples are desired across a very large area, a crew might travel a long distance to visit a single field where only one or two samples are collected in accord with the random sampling plan, then travel to the next location. This results in a slow accumulation of samples over time, and raises the risk that the sampling program will fail to collect a majority of the samples desired. An alternative is cluster sampling, described in detail in Cochran 1977, which enhances operational flexibility while maintaining statistical rigor. -- Several comments questioned the requirement that validation data must be collected within 6 weeks of the target date used for validation. Some comments incorrectly described this requirement as a "six week window" (it is a 12 week window), and others stated that it is inconsistent with VM0042. In fact, as noted in the draft of CN0137 made available for public comment, the requirement is consistent with VM0042, which states that sampling must be conducted within the

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		sampling to the entire area is likely to be inefficient. Could you clarify why stratified random sampling is considered inefficient in such cases? What alternative sampling approaches would be recommended to improve efficiency while maintaining statistical robustness? 10.Six-Week Sampling Window Reference (Page 26) Reference 17 states that the six-week time window is consistent with VM0042, which requires sampling and re-sampling campaigns to be conducted during the same season over time. However, VM0042 does not explicitly mention a six-week requirement for sampling and re-sampling. Could you clarify the basis for this six-week requirement and how it aligns with VM0042? A six week window is a quite short time-frame for large projects with large enrolment areas. 11.Sampling guidelines for variogram calculations (Page 30) The requirement of 100 points for variogram calculations is rather high. For projects with small holder farmers in developing countries, it is unlikely to obtain such large amounts of samples in one and in small area/district (where 'soil type' diversity is prevalent for example). Furthermore, considering the costs involved in soil sampling this requirement is heavy on project proponents. Could	same season. We are assuming that there are 4 seasons per year, which is why we selected a 12 week window (about 3 months) for the sampling window. We note that there are regions with < 4 seasons, such as parts of the wet tropics in Brazil more correctly characterized as having wet and dry seasons. In response to these comments we have removed the explicit time window and now use the phrase "same season," which is consistent with VM0042. -- A worked example that includes variogram calculation is in Appendix 6. This example includes computer code illustrating the calculation of three distinct variogram models. Project proponents should verify that the locations selected will provide coverage of point distances in the 0 to 500 m range. If the sample does not provide adequate sampling within the 0 to 500 meter range necessary to fit the variogram, the project proponent should modify the sampling criteria using a clustered or multistage sampling design. We have updated the text of CN0137 with this clarification. -- See general response "Outcome-based validation" and "DSM in smallholder farms."

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		further clarification or alternative plan be provided for this requirement?	
117	Anonymous #3	5.1.2 The goodness of fit requirement R>0 may not be a sufficient threshold to meet the definition of materiality set out in the VCS standard (4.1.10 - 4) 5.1.4.5 We question allowing the use of soil from any depth for calibration purposes 5.1.2. a Suggestion, define the minimum percentage split between validation and calibration data sets 5.1.4 b. Allowing a procedure with only one peer reviewed publication to estimate prediction intervals is too lenient 5.1.2.11 is very lenient and offers little guidance which will complicate matters for developer and VVB 5.1.4 Why is aligning soil cores with depth optional?	See general response "Model-validation criteria" and general comment "Outcome-based validation."
118	South Pole	"Multiple localized DSM models may be used for different subsets of the project area. Where multiple localized models are used, the model validation procedure must be conducted in aggregate across the outputs of the localized models for the entirety of the project area." It is unclear if the localized models do have to be separately validated before aggregation, as well as validation of aggregated outputs.	See general response "Outcome-based validation."

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
119	Anonymous #6	Under section 5. Procedures, page 13, it is mentioned that: 'The property being validated is always the prediction of SOC stock, even where separate components of SOC stock (SOC content and BD) are independently predicted'. This raises an important point regarding how SOC stocks are estimated or determined. Depending on whether SOC stocks are directly (measure SOC% and BD, then calculate SOC stocks) or indirectly (separate prediction of SOC% and BD, then calculate SOC stock) estimated, results can differ (see Miller et al. 2015). Additionally, in VM0042 v2.1, the equivalent soil mass (ESM) approach is recommended for measuring SOC stocks, which requires dry soil mass rather than BD. If a project developer is currently using this methodology, then ESM approach will be required in year 5 for resampling. It is not clear how this will be accomplished with the proposed DSM tool. Under section 5.3. Computing the Variance of the Average SOC stock, it is mentioned that: 'Where map errors are spatially structured, estimates of the variability in mean estimates such as the standard error will significantly underestimate the precision of the mean. Geostatistical methods address this issue by quantifying and accounting for spatial dependence in the errors.'	Comment requested clarification about compatibility with the equivalent soil mass (ESM) approach required under VM0042 v2.1. CN0137 allows SOC stocks to be predicted either directly or as the product of independently predicted SOC concentration and bulk density; however, the property being validated is always the predicted SOC stock, not its components. CN0137 is compatible with the SOC stock on an ESM basis. In cases where ESM is required, project proponents must ensure that SOC stock measurements used for model validation are derived on an ESM basis. -- Geostatistical methods must be used under CN0137 no matter where predictions are generated. Sampling requirements are not greater for the variogram calculation than for model validation. The requirement in CN0137 is for a wide range of inter-point distances in the 0 - 500 m range. Note that the number of inter-point distances scales with the square of the number of points. For example, a sample of 100 point locations results in 10,000 interpoint distances.

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		It is not clear how this geostatistical approach will be used, specifically in terms of sampling requirements. For example, if the project proponents create a model and then make the predictions for specific sampling points instead of a map, is this approach still needed? Regarding the sampling requirements, the tool specifies that the spatial covariance of model prediction errors must be quantified with a valid method such as fitting a variogram. It is noted that if a variogram is fit, it is required that: "a variety of inter-point distances in the 0–500 m range are adequately represented among the sample locations". This suggests that the sampling density needed to fit a variogram to account for the spatial covariance would most likely be higher than the current soil sampling density used under the Measure and Model approach. Is this assumption correct?	
120	NMI-agro	Section 5.1: Model Development. Step 6: We recommend to add safeguards to the guidelines for setting sampling locations for validation. At present, there are some unwelcome incentives. We've evaluated many projects and know that loopholes can be misused by projects, so would be better to prevent this. First, how do we make sure that validation samples cover the entire variability of parameters that occur? In the present set-up you are rewarded for	The draft of CN0137 stated that in very large projects, stratified random sampling is likely to be inefficient. Some comments emphasized the importance of random samples. Other comments questioned whether random samples are feasible. We believe that stratified random sampling (as defined in VM0042) can sometimes be inefficient because it requires project proponents to travel very large distances without collecting large numbers of samples. In our experience, sampling is a critical bottleneck. There is a limited time window when sampling is possible (e.g., after snow melts and before

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		picking similar spots, this way you will get your 90 in 90 quickly and need few samples. You can already identify similar spots based on remote sensing. However, then you overlook dissimilar areas which are crucial to understand the full picture of soil carbon stocks. Alternatively, based on experience, we know samples are just taken proximal to the road/access point. This also leads to unrepresentative validation samples. So: How do you make sure validation sample selection is representative of the area? We recommend that sampling for validation should be either random or constraint (e.g. Minasny &McBratney, 2006 worked on conditioned Latin hypercube sampling). The take-away from this recommendation is that in the present set-up there is an adverse incentive to take few & unrepresentative samples, and we recommend to include some suggestions to avoid this. Section 5.2.1. Please clarify how it will be demonstrated that the samplings are representative? Please add benchmarks to constrain this. For instance, one could ask to show the distribution and range of the co-variates. Section 5.2.1. step 2: please specify when variogram is made. Refer to appendix or Appendix 6: Example uncertainty calculation or include reference to R/python packages/functions to make it easier for	planting, or after harvest and before the ground freezes, with other constraints elsewhere in the world). Poor weather, difficulty gaining access to required areas, or disruptions to field crews can prevent sampling plans from succeeding. When random samples are desired across a very large area, a crew might travel a long distance to visit a single field where only one or two samples are collected in accord with the random sampling plan, then travel to the next location. This results in a slow accumulation of samples over time, and raises the risk that the sampling program will fail to collect a majority of the samples desired. An alternative is cluster sampling, described in detail in Cochran 1977, which enhances operational flexibility while maintaining statistical rigor. -- A worked example that includes variogram calculation is in Appendix 6. This example includes computer code illustrating the calculation of three distinct variogram models. Project proponents should verify that the locations selected will provide coverage of point distances in the 0 to 500 m range. If the sample does not provide adequate sampling within the 0 to 500 meter range necessary to fit the variogram, the project proponent should modify the sampling criteria using a clustered or multistage sampling design. We have updated the text of CN0137 with this clarification. -- Equation 8 has been removed from the current draft of the tool. -- Some comments questioned how to prevent fraud within the tool by users who might try to manipulate the validation dataset, or augment validation locations with manure. We agree that ensuring the integrity of the validation dataset is essential to maintaining confidence in the tool. While it is true that all methodologies rely on the good faith of project proponents, CN0137 includes safeguards that help to prevent manipulation: (1) VVB and IME review: All validation datasets are subject to

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		the user to implement this. Section 5.2.1. equation 8: Please clarify in which step and how each component of this equation is determined. E.g. sigma: when is this value determined? These are the residuals between model prediction and actual values. Are the actual validation values from the validation sampling? But this equation is supposed to give the # of validation samples. Is sigma set by user? Is it set to be maximum allowable vale for residual bias? Please clarify. With equation, to run it, you either need to have a value for sigma or n samples to start. Please clarify. We also note that if sigma = 0 you need zero validation samples: this is not an outcome that makes sense. We recommend to add a safeguard on this as well. EAPE: can it also be based on validated VMD0053 model outcomes? Please consider. Section 5.2.3 We recommend adding a safeguard here as well. Consider recommending that re-sampling after the 5-year period must take place on	third-party audit by an approved VVB. The VVB and IME must assess whether the validation sampling plan is within the project area, rigorously defined, and properly documented. Any attempt to exclude data or locations without justification would be visible during this audit process. (2) Pre-specified sampling protocols: CN0137 requires that project proponents document the sampling design in advance (e.g., stratification, randomization). Exclusion of data after prediction, and without justification, is a breach of this protocol. (3) Transparency and traceability: All validation points must be geolocated and submitted as part of the project documentation. This creates a record that can be reviewed by VCS or third-party reviewers. (4) Statistical validation requirements: CN0137 does not prescribe a fixed validation sample size, but requires that the model meet quantitative benchmarks on an independent dataset. A small or manipulated sample is unlikely to pass these thresholds. These safeguards help to ensure that validation datasets are credible. -- One comment suggested that sample sizes for model validation should be related to the number of variables in the model (e.g., 10 samples per variable). We note that the number of validation samples required to evaluate model predictions is not directly related to the number of independent variables in the model. This is because CN0137 does not aim to validate the underlying model structure or its coefficients; rather, it validates the predictions generated by the model through comparison with independently collected soil sample data. As such, the focus is on ensuring that the validation sample set is large enough. Statistical rigor of the model validation is in the sampling design, not in maintaining a fixed observation-to-variable ratio. CN0137 does not require a fixed number of samples for model calibration,

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		different locations to avoid unwelcome incentives. If you know the future sampling location, there is an incentive to add manure or carbon on that target location. Thus, if sampling locations are the same, the measurement can be tricked. Or the party taking samples should not be the same as the one determining the sample locations (switch roles). Section 5.2.3 Step 2, continued safeguard: Consider adding a minimum about of samples. 10% of minimally 50-100 samples? Otherwise you can end up with an n=1 out of 10, that is not going to be enough for a statistically robust recalibration.	recalibration or model validation. Project-specific factors will dictate the number of samples used, which must be evaluated by the IME. Accordingly, CN0137 allows flexibility: some projects will demand larger sample sizes than others. For example, a project in a heterogeneous region with a relatively poorly performing model will demand a larger validation sample size than a project in a homogeneous region with a well-performing model. A model that has been calibrated with 250,000 physical soil samples is likely to experience diminishing returns in performance as new samples are added to the calibration. In all cases, the sampling plan must align with the target methodology. -- See general response "Alignment with existing VCS methodologies" and general response "Outcome-based validation."
121	Seqana GmbH	Allow for model-assisted design-unbiased estimators and provide for an accessible way to switch between the model-based, model-assisted, and design-based estimators: The current tool presents a big step forward in allowing for flexible models to derive DSM-based estimates for both use cases, QA1 - measure and model (as initial "measurement") and QA2 - measure and remeasure (as "measurements"). Until now there was a clear analogy between these VM42 defined QAs and the dichotomy of model-based and design-based inference to arrive at project-level	See general response "Issues beyond the scope of CN0137."

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		estimates for SOC stock changes. The proposed tool breaks this analogy as it allows for model-based inference under QA2 - measure and remeasure. While we welcome the integration of innovative DSM approaches, we urge the authors to explicitly allow for design-based estimators to be permitted in parallel to the model-based estimation. In particular model-assisted design-unbiased estimators should be explicitly allowed, that make use of both ground-measured samples and a DSM model. This is especially important, as the model-based estimates can be biased and lead to over- or under-estimation of the ground-observed SOC stock changes. To avoid leaving a project proponent in jeopardy when model validation fails because of a bias, there should be a clear path for project proponents to switch between DSM model-based estimates and design-based estimates involving ground measured data (e.g. Horvitz-Thompson estimator or Hansen-Hurwitz estimator (Cochran, 1977)). Such design-based estimates can be based purely on ground measurements by way of a probability sample as described in the tool for the purpose of model validation. Alternatively, the design-based estimation can be model-assisted where the estimator remains design-unbiased by virtue of the probability sample and in addition the variance of the estimate	

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		(uncertainty) can be reduced through the model predictions (Brus, 2000; Dumelle et al., 2022; Viscarra Rossel et al., 2016). In such a case both the samples and the model could still be used for estimation, even if model validation fails. One strict requirement for the use of design-based estimators should be design-unbiasedness, which can be achieved by the criteria for sampling designs already contained in VM42 (probability sample) and expanded on in the present tool. In terms of uncertainty requirements such a design-based estimator should follow the same approach as laid out in equations 70/71 of VM42 v2.1 (December 2024) or as laid out in equation 5 of the tool CN0137 (the difference being that the variance components are not variances of the mean of predicted targets but variances of the mean estimates of the targets (i.e. SOC stock and SOC stock change)). In addition to the scientific concerns there are practical reasons for allowing model-assisted estimation in parallel to model-based estimators: Model-assisted SOC estimations offer a cost-effective alternative to model-based approaches, especially for smaller projects that face financial or capacity constraints. Unlike model-based methods, they do not require validation by Independent Modelling Experts (IMEs) and VVBs,	

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		avoiding the high costs and delays associated with model validation. This is because model-assisted approaches rely on sample data rather than predictions, making them unbiased by design and posing no risk to carbon credit integrity. Considering the challenge of assembling a new list of IMEs for DSM model validation, this could offer a path for already allowing DSM in practice while there is not enough IME capacity yet. The only impact of model-assisted estimation is on uncertainty deductions, which change predictably based on the model's R^2: if R^2 is positive, uncertainty is reduced; if zero, it remains the same; and if negative, it increases—prompting developers to revert to purely sampling design-based estimates. Since this approach cannot overstate credits, there's no need for additional validation steps like checking model coverage or prediction intervals. Model-assisted estimations thus offer a practical, credible path forward for projects that might otherwise be excluded from model-based benefits. For the uncertainty quantification and deduction for such model-assisted and design-based estimators the existing equations 62/70/71 in VM42 v2.1 (December 2024) shall be applicable. The flexibility to allow for and switch between model-based, model-assisted,	

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		and design-based estimates should apply to both quantification approaches measure and model and measure and remeasure. For references see answer to question 6 above.	
122	Indigo Ag	Concern 2: If using DSMs for Quantification Approach 2, it is unclear from the proposal whether one would be required to use the DSM for both the project and baseline scenario. If project developers were allowed to use the DSM for just the baseline scenario but not the project (or vice versa), it could lead to severe bias. For example, suppose a DSM were applied in the "control fields," but direct measurement via soil sampling were used in the credited fields. (Note that this approach would be very tempting to do, because it is difficult to get permission to collect soil samples in fields outside the project). Then the DSM could inaccurately predict zero stock change because it doesn't have many features that are sensitive to SOC changes over time, while the other method (e.g., direct soil sampling with design-based inference) could be used in credited fields. If SOC stocks were increasing in the control sites, then this approach would be biased toward over-crediting. We think it's implied that one would need to use DSM in both the project and baseline quantification, due to	See general response "Alignment with existing VCS methodologies," general response "Outcome-based validation," and general response "Issues beyond the scope of CN0137."

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		language in VM0042 v2.1, such as "Where more than one quantification approach is allowable for a given gas and source, more than one approach may be used provided that the same approach is used for a given quantification unit in both the project and baseline scenarios" (page 19), but we feel it is important for the authors to specify this directly in the proposal. Similarly, we feel it is important that the authors clearly specify that the same model is applied in each year of the project to estimate SOC stocks (and trained with the same set of training data). The scenario that concerns us, and the reason we think this should be explicitly stated in the text, is the following: Say you have a model trained at time 0 using one set of training observations, and a model trained at time 1 with a different set of training observations (on a different set of fields). If the training observations at time 1 have higher SOC content than the training observations at time 0, you will produce consistently higher estimates of SOC at time 1, and it will appear that significant stock changes have occurred (when in reality, it was just caused by sampling different sets of locations at the two points in time). Based on the comment about cumulative accounting in footnote 14 on page 23, we believe the authors intended for the same model to be used for both the initial stock and the stock at a later date, but we think	

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		this point could be further emphasized in the proposal. Concern 3: The proposal presents an approach for monitoring SOC stock change that does not require management data in order to build models. While management data is not needed to train and predict SOC stock change, we feel it is important for the proposal to be explicit that project developers are still required to collect management data for all project fields, in order to demonstrate that practice changes have occurred, which is required to generate carbon credits under VM0042. Management data is also important to collect in order to quantify other greenhouse gas emissions such as N2O, CH4, and CO2 from fossil fuel use, which are important to account for to ensure increasing SOC does not come at the expense of equivalent increase in greenhouse gas emissions. In addition, management data would need to be collected on control fields in order to establish a counterfactual scenario to compare against the project scenario. As we move to more remote sensing approaches, and therefore visit fewer farms, it is critical that we are still able to tie changes in carbon stocks back to	

Section 5 Procedures - Overall Section Feedback			
#	Organization	Comment	Developer's Response
		practice changes made by farmers in the program.	

Section 5.1 Model Development

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
123	Anonymous #2	5.1.4 Calibration Data. "The target variable is … but the tool permits use of augmented or synthetic calibration data provided that all validation data follow the guidance in Section 5.2." It needs to be stipulated that the use of augmented or synthetic calibration data must include and propagate the uncertainties associated with these data.	See general response "Alignment with existing VCS methodologies."
124	Anonymous #2	5.1.4 "All samples must be matched with covariate features that coincide in space and time with the location and sample date." It is not clear what is required from this statement, please clarify.	One comment noted this sentence in the draft of CN0137 available for public comment: "All samples must be matched with covariate features that coincide in space and time with the location and sample date." and requested clarification of its meaning. This statement means that covariates must be from the same time and location as a given physical soil sample being used for model calibration. For example, if a physical soil sample is collected on a given field on a given date, the covariates paired with that sample must coincide with the location and sample date. This would mean, for example,

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
			that it must be possible to extract the pixel value within which the sample is located. It would also mean that remote time-varying data must be coincident in time with physical soil sampling. If the sample had been collected today, but time-varying remote sensing data were acquired two years earlier, this criterion would not be satisfied.
125	Anonymous #4	1.Treatment of baseline conditions/control sites could be clarified (5.1.13 implies that control sites may be used). Baseline control sites – variously known as "dynamic baselines" – where baseline activities are run in parallel to areas subject to management intervention – is strongly recommended. Use of a static baseline period (e.g. fixed period prior to project initiation) can mean that SOC measurement occurs during a period subject to weather anomaly (e.g. low SOC due to preceding drought). If by chance the PC coincided with the end of the drought (for example), SOC crediting may partly occur due to increased rainfall, rather than PC. Use of a paired baseline - where SOC is monitored on an adjacent parcel of land and compared with SOC on the project area at the same time - obviates the impact of climate on SOC. It is worth noting however that some proponents will not be able to run control sites, as this requires dedicated management over the long term, as well as significant land resources. Use of the second term in brackets in 5.1.13 (equation 6) disadvantages those	The minimum value for k is 1, when cross-validation is not used, as described in the text of CN0137. The maximum value for k is the number of samples used to calibrate the model, implying a leave-one-out cross-validation procedure. Between these extremes, project proponents have discretion to use other kinds of cross-validation, such as k-fold or geographically-based cross validation. k-fold uses a number of "folds," each of which is a random partition of the dataset (e.g., a dataset with 1,000 observations might be divided into 10 random partitions). Geographically-based cross validation uses non-random partitions based on geography. For example, all measurements within a given field might constitute a single fold, with the number of folds equal to the number of sampled fields in the dataset. -- One comment asked whether model prediction error should be aggregated over space and time, or only over space, (with reference to 5.1.8 in the draft of CN0137 available for public comment). Aggregation should occur at a given time, t, over space. It should not occur over time. This has been clarified in the revised tool. -- The comment correctly notes an ambiguity between the definition of the prediction support unit and the description in Equation 3. The definition notes that the prediction support unit can be a point, but Equation 3 assumes it is an area. We have clarified the ambiguity. The "prediction support" is the soil volume to a specified depth or depth interval for

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		proponents who do not use baseline control sites, because the second term is said to be set to zero in such cases. 2.Section 5. Use of multiple localized DSMs (e.g. for subset project area) would each require independent calibration and validation of that localized DSM. This is because each DSM is effectively an independent model. Having multiple DSMs will improve SOC prediction within a certain region, but each DSM must be validated for the crop functional group and PC in question. 3.Section 5.1.2 and 5.1.5a: splitting of data into the number of calibration and validation datasets (k) could be better explained. Does k have minimum/maximum thresholds? 4.Section 5.1.4: conflict here with VMD0053. CN0137 suggests that any procedure in the peer-reviewed literature is acceptable; 90% prediction intervals are more prescriptive in VMD0053. 5.Section 5.1.6b: it needs to be explicitly stated that no further calibration, hyperparameter tuning (or other form of parameter adjustment) is permissible after calibration is finalised. 6.Section 5.1.8: it needs to be clarified if this is for each prediction support unit (which can be as fine as an individual core and depth increment at a particular time point) or whether model error should be aggregated across all t and prediction support units.	which a model is calibrated and against which a physical measurement is compared during model validation. For example, if a model had been calibrated using individual soil cores to a depth of 30 cm, the prediction support unit is the 30 cm soil core. -- Equation 6 is the mean carbon dioxide removal estimate in soil. A comment suggested that the equation contains a logical error because it is expressed in terms of removal and as a difference. However, we have carefully reviewed Equation 6 and confirm that there is no logical error. The commenter noted that removal can be positive (sequestration) or negative (emission). Say that the first term (the project area) is a net removal (positive), but the second term (baseline) is negative (emission). This would mean that the difference being computed by equation 6 is a positive number minus a negative number. Subtraction of a negative is addition, so the baseline emission would be added to the net removal of the project. This is correct. For example, imagine that the project increases from 1 to 2 units, and the baseline decreases from 4 to 3 units. This would result in (2 - 1) - (3 - 4) = (1) - (-1) = (1) + (1) = 2 units of net carbon removal. This is the correct estimate because the credit claim is based on the net change in the project relative to the baseline. Such a scenario implies that a positive claim could be made even when the project is losing SOC over time. This could occur when the project is losing less than the baseline. For example, imagine that the project decreases from 2 to 1 units, and the baseline decreases from 4 to 2 units. This would result in (1 - 2) - (2 - 4) = (-1) - (-2) = (-1) + (2) = 1 unit of net carbon removal. Whether such a claim is desirable or not under VCS rules is beyond the scope of CN0137. We simply note here that the structure of Equation 6 is correct. -- The tool clarifies that extraneous features must be excluded

Section 5.1 Model Development

#	Organization	Comment	Developer's Response
		7.Section 5.1.10. Ai is defined as the area of the prediction support unit, but earlier where the PSU is defined, the developer notes that a PSU can be a "volume of soil" or specific depth with defined latitude and longitude. Equation 3 would then be undefined. 8.Section 5.1.12. Equation 5 speaks of delta SOC across the project area but I think subscripts should be listed for each CFG and PC (in which case SOC variances will differ). 9.Use of the second term in brackets in 5.1.13 (equation 6) disadvantages those proponents who do not use baseline control sites, because the second term is said to be set to zero in such cases. 10.Suggest that some sense-checking of Equation 13 be conducted. Carbon dioxide removals will not always be a positive number. Drought may cause losses of SOC between sampling events. If is negative, then this term would be added to the first in Equation 6 for example. Losses in SOC between events at least need to be mentioned, and their implications for variance (which cannot be negative) should be stated. 11.Use of foot notes could be minimised. Often these contain useful/essential information (e.g. footnote 9). I suggest that most of the footnote information be incorporated into the text. This should help improve understanding of the method. 12.Dependence/independence of	from the mapped area. -- See general response "Issues beyond the scope of CN0137," "Outcome-based validation," "Alignment with existing VCS methodologies," and "Model-validation criteria."

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		calibration/validation data using k-fold validation could be mentioned within Figure 2. 13.Section 5.1.1. Fencelines should be excluded from the mapped area, as should woody vegetation. 14.Section 5.1.3.1 could be mentioned earlier in the method (when covariates are first mentioned on page 5 and 9) or at least refer to the guidelines in this section and Appendix 3 when covariates are first mentioned. 15.Section 5.1.4. Ambiguity in whether proponents are validating for SOC concentration (%), SOC stock (tonnes SOC/ha) or both. I suggest SOC stocks, because bulk densities (g/cm3) can change between sampling intervals e.g. due to tillage or livestock trampling. 16.Section 5.1.4.1 refers to validation samples, but the section per se (5.1.4) is referring to calibration. 17.Section 5.1.4.4 also root organic matter and inorganic carbon (carbonates etc) 18.Section 5.1.5.2: use of separate models for differing depths would require one validation per depth increment.	
126	University of Illinois Urbana-Champaign	1. The model validation requirement about uncertainty interval coverage has several complications in practice: (i) Assuming independent errors, a perfectly calibrated DSM has a 50% probability of failing the coverage test as	See general response "Model-validation criteria."

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		the empirical coverage follows a binomial distribution centered at 0.9. With reasonable sample size the coverage will be close to 90% but equally likely to be just under or over. (ii) But the independence assumption in (i) is not expected to hold here (hence the geostatistical modelling). With dependent errors it is quite possible for a calibrated DSM to have substantially less than 90% empirical coverage. (iii) As described above, the tool is vague on the role of sampling/measurement error within each prediction support. This error could play a role here as well, for example if the DSM errors are calibrated to one measurement per pixel but multiple measurements per pixel are used for validation. 2. The use of null hypothesis significance testing (NHST) to assess unbiasedness of model predictions has a couple of issues leading to a very weak test of unbiasedness. First I document these issues and then I propose an alternative. (i) it is implied that if the mean error is not significantly different from zero then we can conclude that the model "lacks bias" or is "unbiased", i.e. accept the null hypothesis, but that is not appropriate.	

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		(ii) Why is the 5 year EAPE used to determine the threshold for unbiasedness? (iii) Overall, with the parameters provided (5% significance and 80% power) the sample size calculated by formula 8 leads to a weak test of unbiasedness. iv) The test perversely discourages larger validation sample sizes as they are more likely to end up rejecting the unbiased null hypothesis For example suppose we have an EAPE after 5 years of 1.5 Mg ha-1. Then the standard error for the bias (independent of the prediction standard deviation) using the sample size in equation 8 will be approximately 0.5 Mg ha-1. Thus a 95% interval will have a width of 2 Mg ha-1. This is very large. So long as the point estimate for the bias is between -1 and 1 Mg ha-1, the tool concludes the DSM is unbiased. For example a typical 95% CI for the bias might be -0.5 to 1.5 Mgha-1 and the tool would consider such a DSM to be unbiased. To me this test is very weak. Instead of moving forward under an assumption of unbiasedness it would be more rigorous to propagate the mean error and uncertainty about it through to the final estimate. In many cases this bias will cancel itself out. But in projects with separate project and baseline sites and separate geostatistical models for each,	

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		the biases will not cancel and their uncertainties will compound. For example, if both the project and baseline models have mean errors with standard errors of 0.5 Mg ha-1, then this adds 0.7 Mg/ha to the standard error of the project effect estimate. Had we used the tool to take the above test as evidence of no bias, this uncertainty would have been incorrectly dropped.	
127	South Pole	Image Processing (Section 5.1.3.1, Page 22): The document states that optical images must undergo atmospheric correction before use. Could you specify which methods are recommended or acceptable? Moreover, is there a minimum cloud cover mask percentage to be applied to deem the covariate processing accurate? --- “ Data from outside the project area may be used for calibration, but all data in the validation set must come from within the project area” . Is there any requirement on the provenance of the calibration dataset? --- Data Sources (Section 5.1.3, Page 21): Is there a minimum set of recommended satellite sensors (e.g., Sentinel-2, Landsat-8), or can commercial sensor data be used? --- Spatial and Temporal Resolution (Section 5.1.3.3, Page 22): Is there a minimum	Several comments questioned what would happen if one of the three validation tests is unsuccessful. As described in the draft available for public and in the revised version of CN0137, credit generation under the tool requires all three tests to be successfully passed. A failure to pass compels the project proponent to recalibrate the model or cease activities under the project. This standard ensures integrity within the VCS carbon program. One comment asked why the bias test is not required between model validation events. We have clarified that results from all three tests must be reported in the DSM-MVR between model validation events. These summaries are not reviewed by the VVB until model validation, which must occur at least once every five years. -- Numerous comments questioned the domain of application of DSM. Some questions the appropriateness of the < 30% tree cover threshold included in the draft version of the tool. Some requested justification for the threshold, while others argued that it was unnecessary or overly restrictive. In response to these concerns, we have removed the threshold from the tool. Project proponents will now be responsible for justifying alignment between the tool and the applied methodology at the time of project registration. This

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		threshold for spatial and temporal resolution to ensure data quality? --- Use of Field Data (Section 5.2, Page 24): The document requires validation data to come exclusively from the project area but allows external calibration data. How is the representativeness of these external data ensured in highly variable spatial contexts? Moreover, which is the definition of "representative sample" and what are the thresholds that should be considered in defining its representativeness? --- Validation Tests (Section 5.1.6, Page 23): Three validation tests are required (coverage, goodness of fit, and bias). What happens if a model passes two of these tests but fails one? Are adjustments allowed to improve accuracy without fully recalibrating the model? --- Mapped Area: the applicability condition of the tool refers to max 30% forest cover in the mapped area. Then, why do pixels with trees need to be excluded from the mapped area? Is there a tree cover threshold to exclude pixels with trees, e.g. what happens if a pixel contains one/two trees? Is the tree cover limited to trees or also bush cover must be considered? --- There must be a known month and year in which the sample was collected. Should the same month (season) be kept for the	change does not signal an intention to expand the tool's scope to include agroforestry systems. Rather, the revision acknowledges that a fixed threshold is unnecessary to determine project eligibility. For example, an orchard with 50% tree cover may be consistent with the tool's design if the spatial arrangement of trees allows for ground-level observation from remote sensing between trees. Other comments asked for clarification about the types of soils that are amenable to DSM. DSM is applicable to all soils, subject to the constraints in CN0137. -- Several comments questioned the requirement that validation data must be collected within 6 weeks of the target date used for validation. Some comments incorrectly described this requirement as a "six week window" (it is a 12 week window), and others stated that it is inconsistent with VM0042. In fact, as noted in the draft of CN0137 made available for public comment, the requirement is consistent with VM0042, which states that sampling must be conducted within the same season. We are assuming that there are 4 seasons per year, which is why we selected a 12 week window (about 3 months) for the sampling window. We note that there are regions with < 4 seasons, such as parts of the wet tropics in Brazil more correctly characterized as having wet and dry seasons. In response to these comments we have removed the explicit time window and now use the phrase "same season," which is consistent with VM0042. -- See general response "Issues beyond the scope of CN0137" and general response "Outcome-based validation."

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		future calibrations (rainy/dry season may affect the process?) --- Moreover, is there a minimum spatial resolution (pixel size) in the definition of the mapped area? How should it be estimated? Is there a minimum spatial data quality requirement (e.g. spatial resolution) to comply with? In the case of grouped projects, each area has to be considered independently for the modelling purposes? --- Samples may be collected prior to the start date of the project. Unsure if it is mandatory or not prior to the project start date? If the project activity implies soil disturbance soil bulk density may be affected, impacting the modelling process. --- The target variable is typically SOC content, BD, or SOC stock measured using individual soil cores or composite samples, but the tool permits use of augmented or synthetic calibration data Is model calibration based on purely remote-sensing data products allowed? and purely synthetic (simulated) data?	
128	Seqana GmbH	In equation 5 on page 17: The definition of rho (ρ) seems to be faulty: "Correlation between the standard deviations of SOC stock at times t and t+delta" → it is the correlation between the predicted (estimated) SOC stock means at time t and t+delta. It should be made clear	We added a subscript, t, to rho in response to this comment. Rho is correctly defined as the correlation of errors of the prediction of SOC stock at times t and t + delta. The comment correctly notes that the correlation depends on t and t + delta, meaning that a separate value for rho needs to be estimated for all combinations of t and t + delta. We added a subscript to the equation

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		in the notation that, in general, rho depends on t and delta, meaning: a separate rho needs to be estimated for every t and delta (under a stationarity assumption, rho would be independent of t but still dependent on delta). Using the symbol rho without subscripts (t and/or delta) misleadingly suggests that it is sufficient to estimate a single rho value for all values of t and delta. It would be beneficial to clearly differentiate between the rho in equation 5 and the rho from the variogram expressed in equation 10. Maybe a different subscript or symbol could be chosen.	to clarify this.
129	Treefera	The current guidance in Section 5.1.3.3 states that time-varying covariates must maintain the same temporal structure between calibration and prediction. However, this remains somewhat ambiguous in practice—particularly for remote sensing covariates—and would benefit from further clarification. Specifically, we recommend that the guidance address the following questions regarding optical satellite data (including primary imagery and derivative indices such as NDVI): 1) Should these variables be calculated as an average over a period before or around the soil sampling date (e.g., an average of the four weeks prior to sampling, or ±2 weeks around the date), or would a single	A curated list of papers, including examples with standard practices for commonly used remote sensing inputs is provided in Appendix 3. -- See general response "Issues beyond the scope of CN0137" and general response "Outcome-based validation."

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		image closely matching the sampling date be considered representative? 2) Is there a recommended or required temporal window for using optical satellite-based covariates? For instance, how long before or after the sampling date would an individual satellite image still be considered appropriate for pairing with the soil sample. How should seasonality (environmental and regarding agricultural practices) be taken into account appropriately, and would covariates have to maintain seasonal coherence during calibration and prediction phases similar to time interval and temporal relationship? 3) Is it acceptable to use a composite—such as the median reflectance from a cloud-free mosaic over the sampling month—as a covariate? Including examples or outlining standard practices for these commonly used remote sensing inputs would significantly enhance clarity and support consistent implementation.	
130	Indigo Ag	Other concerns: Guidance is not provided in the module for users who want to build a DSM of carbon	Comment requested guidance for project proponents wishing to develop DSM models that directly predict SOC stock change. Such models are not allowed under CN0137. This was an intentional decision to simplify

Section 5.1 Model Development			
#	Organization	Comment	Developer's Response
		stock change over time (rather than a DSM of stocks at one point in time). In order to use this module, users are encouraged to build static carbon stock models at multiple time-steps that are less sensitive to stock change. Therefore, this module guides users away from building more robust stock change models. We feel that guidance should be included for users who aim to develop DSMs of SOC stock change directly. The validation criterion 10.c on page 16 needs another parameter to be specified: a threshold below which we consider bias to be tolerably small. Without that specified, project developers would either prefer to have very few data points (which is not what we want to encourage) or choose a less rigorous threshold (also not what we want to encourage). [Note that the same issue occurs in VM0042's criteria for matching on SOC concentration: "The percentage must not be significantly different from the mean percentage SOC of the linked quantification unit at a 90% confidence level" (page 23 of VM0042 v2.1).] The suggested formula for sample size (equation 8, on page 25) implies that the threshold for significance may be the expected effect size. That amounts to saying that the bias can be so large that all the credits should not have been created. That target does not seem rigorous enough.	calibration and validation. -- See general response "Model-validation criteria" and general response "Outcome-based validation."

Section 5.2 Soil Sampling

Section 5.2 Soil Sampling			
#	Organization	Comment	Developer's Response
131	Anonymous #2	5.2.1 6. Sample size calculation The formula looks similar to MDD in VM0042. However, the standard deviation is not that of the soil carbon stock, but of the model prediction error. It is unclear how the two approaches compare, and therefore what the implication might be for sample sizes. It is not clear how the Model Prediction Error can be used to calculated how many samples to collect since the model requires data from the project area to generate the model prediction error. Is it possible that MPE will be determined from data from other sources? If so: then this is not clear from the text. 5.1. point 12. Calculation of change in soil carbon stock The document would benefit from clearer descriptions around the determinations of mean and variance of soil C change using DSM. For example, given the previous comment around PSUs, we are not clear how the mean predicted change is generated from the DSM surface.	The comment notes that Equation 8 in the version CN0137 available for public comment resembles Equation 3 in VM0042 v2.1 but substitutes the expected average project effect (EAPE) and the model-prediction error (MPE) for the "minimum detectable difference" and the "standard deviation of ΔSOC" used in VM0042. Equation 8 has been removed from CN0137. The comment correctly notes that the standard deviation in the numerator of CN0137 is unknown in the absence of validation samples, which is one reason why Equation 8 has been removed. The purpose of Equation 8 in CN0137 was guidance only - its purpose was to "estimate the approximate sample size." Regardless of the sample size, the model must subsequently pass the bias, coverage and R^2 tests. In place of Equation 8, the updated draft of CN0137 clarifies the relationship between sample size and model calibration, recalibration and model validation, noting the importance of IME assessment to "determine whether the sample number is sufficient for the model to meet the three validation criteria and aligned with requirements in the applied methodology." -- See general response "Outcome-based validation" and general response "Alignment with existing VCS methodologies."
132	Anonymous #4	1.Section 5.2.1. Treatment of missing (or minimum requirements for) validation data samples should be explicit. 2.Section 5.2.1.4: the fact the each PSU	One comment asked how the tool can ensure that unsampled units treated as "missing at random" do not introduce systematic errors into predictions. By definition, unsampled units that are "missing at random"

Section 5.2 Soil Sampling			
#	Organization	Comment	Developer's Response
		must below to only one stratum should also be listed where the PSU is first defined. 3.Section 5.2.1.5: requirement for baseline control sites should be mentioned where Use Case 2 is first introduced. 4.Section 5.2.1.6.: Need to be clear how this standard deviation is computed. 5.Section 5.2.2 and 5.2.3. It would be more consistent with VM0042 to use "Quantification Approach" rather than "Use Case" terminology. 6.Section 5.2.2 replace "applied methodology" with "VMD0053". 7.Section 5.3.2. Reference to the diagram in Figure 2 would be useful here. 8.Section 5.3. Suggested number of Monte Carlo samples (L) would be worthwhile.	will not introduce systematic errors. It is when samples are missing nonrandomly that the likelihood of systematic errors arises. For example, if unsampled locations happened to be from areas where model performance was better, or worse, on average, then missing locations would be non-random, and systematic error could result. This is why the draft tool specifies that validation samples should be a probability sample of the project area or justified as missing at random. A related comment noted that if "SOC variation is not random but tied to management practices, soil types or precipitation, this assumption [missing at random] could distort results." There are three points we would like to make in response to this comment. First, encouraging the use of a probability sample will prevent spatial variation in unsampled units from introducing systematic error into model validation. This is the purpose of using a probability sample. Second, it is not non-random variation in SOC that is the issue with missingness, but non-randomness in the accuracy of model output. Either way, a probability sample of the project area would ensure that systematic variation in the project does not undermine model validation. We note that these same principles underpin validation in all VCS methodologies that require samples of a project area. The same comment asked whether there should be "mandatory sensitivity analyses" to assess whether missing data could bias SOC estimates. However, this comment seems to misunderstand what is being referred to as "missing data". Through the use of this term, the tool is referring to unsampled locations. These locations are missing because they were never visited, not because of some error or a desire to exclude them. An example can help to illustrate this concept. Imagine a poll designed to forecast the outcome of an election. If 1,000 randomly

Section 5.2 Soil Sampling			
#	**Organization**	**Comment**	**Developer's Response**
			selected people were asked how they would likely vote out of a voting population of 100,000,000, the 99,999,000 people not consulted by the pool are “missing at random”. There is no way to perform a sensitivity analysis to determine what these other people, not consulted by the poll, might have said. This is why it is important that the 1,000 people who were consulted in the poll were selected at random. If they were selected at random, then the assumption that unsampled individuals are missing at random is robust, and a prediction from the poll is valid. However, when the poll is non-random, for example, by focusing on certain geographic locations or demographic groups, missingness is non-random. Under such a scenario a sensitivity analysis would be warranted. Such an analysis is beyond the scope of CN0137, because CN0137 encourages the use of a probability sample and justification for “missing at random.” -- One comment requested clarification about the definition of the prediction support unit (PSU), and suggested that multiple soil cores should be collected within each PSU. As defined in the draft tool, a PSU is “The land area and volume of soil for which model predictions are calibrated and validated. The PSU refers to individual points or composite samples at specific depth intervals. For example, the PSU could be individual soil cores or composite samples from collections of soil cores over a specified depth range. Where the prediction support unit is the individual soil core, there must be a set of coordinates that define the collection location for the physical soil sample.” This term and definition has been changed. We now use the terms “prediction support,” and “prediction location.” Prediction support is, “The soil volume to a specified depth or depth interval for which a model is calibrated and against which a physical

Section 5.2 Soil Sampling			
#	Organization	Comment	Developer's Response
			measurement is compared during model validation.“ Prediction location is, “The location at which a prediction is generated by the model. For example, the model could be used to generate predictions on a 10 by 10 meter grid or at arbitrary locations.” The PSU can use composite cores at the discretion of the project proponent. A separate comment requested that we clarify that each PSU can belong to only one stratum. In fact this was stated explicitly in 5.1.2.4 of the draft tool and remains in the current version. -- Some comments questioned our use of the term “Use Case” and suggested that instead we use the term “Quantification Approach” to align with VM0042. There are two reasons why we selected and retained the term “Use Case.” First, not all VCS methodologies use the term Quantification Approach, so the use of the term would be inappropriate in the tool, which must be applicable under multiple VCS methodologies. Second, “Use Case” is not redundant with “Quantification Approach” in VM0042. For example, there are two use cases under Quantification Approach 1 (point-based, and area-based). The Use Cases in CN0137 describe specific implementations of Quantification Approaches 1 and 2 in VM0042 in addition to other implementations in other VCS methodologies. -- One comment requested a recommendation on the number of Monte Carlo samples. The tool provides an outcome-based recommendation based on the precision of estimate. The number of Monte Carlo samples required will vary depending on context. Users of the tool should consult Wadoux and Heuvelink (2023), which explicitly discusses the number of Monte Carlo samples necessary to produce a precise estimate of the variance of the mean using the method described in the tool. -- See general response “Alignment with existing VCS methodologies” and general response “Issues beyond

Section 5.2 Soil Sampling			
#	Organization	Comment	Developer's Response
			the scope of CN0137."
133	University of Illinois Urbana-Champaign	The representativeness of the validation set is important and mentioned in the tool but not defined as far as I can tell. In order for the validation to truly validate the model with respect to the entire project area, this representativeness should be defined in terms of the covariate space of the model and not just spatially. The distribution of these covariates in (i) the validation set and (ii) the entire project area should be compared to ensure that there is significant overlap. Note that for models with a large number of covariates spatial representativeness alone does not guarantee covariate representativeness.	See general response "Outcome-based validation."
134	South Pole	Where the project proponent does not resample all previous validation locations, the subset selected from the original locations for resampling must be identified at random. In the case of grouped projects, should the random resampling be done exclusively within the same area subset?	See general response "Issues beyond the scope of CN0137."
135	Seqana GmbH	We appreciate the overall unbiasedness validation criterion and the sample size guidance with equation 8. However, there are several issued with the BIAS t-test requirement and equation 8. 1. From a project operational perspective, we believe the bias criterion is especially	See general response "Model-validation criteria."

Section 5.2 Soil Sampling			
#	Organization	Comment	Developer's Response
		important at any verification event when credits are issued. So we disagree with the sentence in section 5.1.6: "The model validation bias test is not required between model validation events." (p.24). In footnote 14 (p.23) you provide an example where an intermediate verification event under-estimates the overall sequestration rate. However, it is equally possible that such an event could over-estimate the overall sequestration rate. Especially thinking about large-scale projects, even a slight over-estimation of the mean sequestration rate could lead to a depletion of buffer/insurance credit pools. At the moment it seems unclear how project proponents would deal with this. Could you clarify? 2. The BIAS test must be adapted when applied on a temporal snapshot rather than on the change basis: For any SOC stock change estimation a DSM model under this tool will have to make at least four predictions: dSOC = (SOCt+dt_proj - SOCt_proj) - (SOCt+dt_bsl - SOCt_bsl). Ultimately, BIAS(dSOC)=0 should be guaranteed. The probability that this is not the case, although the required t-tests pass could be quite large. The error rate to control in this case must be type II error, false negative error rate,	

Section 5.2 Soil Sampling			
#	Organization	Comment	Developer's Response
		beta: the error of accepting the null hypothesis (BIAS=0) although it is false. In equation 8 beta is set at 0.8. We believe this is an honest mistake and the intent was for power (1-beta) to be 0.8 and accordingly beta=0.2. With that said, the probability of wrongly accepting BIAS=0 would still be rather high. One example: Let's assume that the bias test passes with β=0.2 and that valid model(s) is(are) applied individually on all 4 prediction problems. What is the maximum probability that at least one of the estimates is biased even though the tests are passed (assuming independent tests/predictions)? (3a) P(at least one Type II error) = 1 − (1 − β)^4 (3b) P(at least one Type II error) = 1 − (1−0.20)^4 = 1 − (0.80)^4 = 1 − 0.4096 = 0.5904 -> 59% probability of at least one bias. What would be the right quantity for beta to keep the probability of an overall bias at 5%? (5a) P(no Type II error across 4 predictions) = 0.95 (5b) (1−β)^4 ≥ 0.95 (5c) 1−β ≥ 0.95^(1/4) (5d) 0.95^(1/4) ≈ 0.9876 (5e) β = 1 − 0.9876 = 0.0124	

Section 5.2 Soil Sampling			
#	Organization	Comment	Developer's Response
		So when individual predictions are made and the type two error should be controlled by a t-test, the associated beta should be 0.0124 or 1.24%. Alternatively, the target variable for the t-test could be altered from a single snapshot individually for project and baseline to the difference between project and baseline either at a temporal snapshot (requiring 2 predictions/t-tests) or as the change over time between the differences between project and baseline estimates (requiring 1 prediction/t-test). On the flip side the alpha probability is only concerned with falsely rejecting the null-hypothesis when it is actually true, which is of interest to the project proponent/MRV provider but not relevant from a regulatory perspective. So the choice of the alpha level can be left completely to the project proponent/MRV provider. It is their business risk to detect a bias, when there actually is none. 3. Further, we would like to ask for some guidance for how to estimate the parameter σ (Standard deviation of the model prediction error) ex-ante before data are sampled from the project site itself.	

Section 5.3 Computing the Variance of the Average SOC Stock

Section 5.3 Computing the Variance of the Average SOC Stock			
#	Organization	Comment	Developer's Response
136	South Pole	- Variance Estimation (Section 5.3, Page 27): The document mentions the use of geospatial methods to estimate uncertainty in predictions. Could you provide clear guidelines on variogram selection and adjustment?	A worked example that includes variogram calculation is in Appendix 6. This example includes computer code illustrating the calculation of three distinct variogram models. Project proponents should verify that the locations selected will provide coverage of point distances in the 0 to 500 m range. If the sample does not provide adequate sampling within the 0 to 500 meter range necessary to fit the variogram, the project proponent should modify the sampling criteria using a clustered or multistage sampling design. We have updated the text of CN0137 with this clarification.
137	Seqana GmbH	For clarity and ease of understanding we would suggest to restructure sections 5.3 and 5.4 to first cover standardized errors and variogram fitting (including the cross variogram procedure) and then the Monte Carlo integration procedure explicitly, as laid out by Wadoux et al (Wadoux and Heuvelink, 2023). Wadoux, A.M.J. -C., Heuvelink, G.B.M., 2023. Uncertainty of spatial averages and totals of natural resource maps. Methods Ecol. Evol. 14, 1320–1332. https://doi.org/10.1111/2041-210X.14106	This comment appears to be based on a misunderstanding. CN0137 already describes the variogram fitting procedure of Wadoux and Heuvelink (2023) with notation that matches that publication.

Section 5.5 Guidance on Variogram Selection and Fitting

Section 5.5 Guidance on Variogram Selection and Fitting			
#	Organization	Comment	Developer's Response
138	University of Illinois Urbana-Champaign	In realistic examples variogram models may not fit the empirical data well and there may be substantial uncertainty about the best variogram model and its parameters. Note that the example in appendix 6 potentially understates this issue for several reasons: the errors are assumed to come from a known variogram model; the two assumptions identified in #2; the relatively large sample size of 1000, which is not needed to meet the tool's bias test or variogram bin size requirements. I modified the appendix code (excellent that this was shared) to take the above into account and found myself regularly generating variograms with poor fit even though the procedure met the tool's requirements. I recommend two changes to the tool to accommodate variogram uncertainty: (i) Geostatistical model uncertainty quantification The traditional geostatistical methods suggested in the tool, while widely used, have limitations for their major purpose here of uncertainty quantification. Specifically, the tool is written with respect to the traditional empirical approach to variogram estimation as opposed to the	One comment suggested the use of maximum likelihood or Bayesian methods for variogram uncertainty analysis, recommended cross-validation of the fitted variogram, and suggested that CN0137 propagate variogram uncertainty through downstream calculations. We highlight that (1) (restricted) maximum likelihood suggested by the comment is not a method for estimating parameter uncertainty, but simply an alternative approach to obtaining point estimates, and that (2) (restricted) maximum likelihood estimation can be used in the tool for parameter estimation, as can the methods of moments. We agree that Bayesian methods estimate the uncertainty of model parameters. But in geostatistical models, parameter uncertainty is usually quantified with Taylor series approximation and the Fisher information matrix. There are examples in Zhu and Stein (2006) and Kitanidis (1987), with examples in the soil science literature in Marchant and Lark (2007) and Wadoux et al., (2019). Although the quantification of variogram parameter uncertainty is possible, it is uncommon. A literature review in response to this comment yielded no examples documenting the impact of parameter uncertainty on estimates of the variance of the spatial mean (the quantity of interest to CN0137), including within the field of DSM specifically. Because VCS methods require documentation in peer-reviewed literature before they can be adopted, the integration of variogram uncertainty into CN0137 is not currently possible. A second comment suggested validation of the variogram function using k-folds cross-validation or through simulation. However, the steps proposed in CN0137 and in Appendix 6 follow best-practices in

Section 5.5 Guidance on Variogram Selection and Fitting			
#	Organization	Comment	Developer's Response
		more modern likelihood approach [1]. One limitation of the traditional approach is that it does not quantify uncertainty in the variogram. More modern approaches such as (restricted) maximum likelihood and Bayesian methods quantify variogram uncertainty and should be preferred in this application. One possible benefit of the traditional approach is that it does not make distributional assumptions, though these are made eventually by the tool to form an uncertainty interval. The tool would be more rigorous if it required estimation of variogram uncertainty and that this uncertainty was propagated through to the final estimate of project mean SOC stock change/effect. [1] Christianson, Ryan B., Ryan M. Pollyea, and Robert B. Gramacy. "Traditional kriging versus modern Gaussian processes for large-scale mining data." Statistical Analysis and Data Mining: The ASA Data Science Journal 16.5 (2023): 488-506. (ii) Geostatistical validation The tool focuses on validating the point predictions of the DSM but lacks validation of the geostatistical component. Two ways to do this are through k-fold cross validation and through fake data simulation. Geostatistical modelling is often non-trivial. The tool and code appendix	geostatistics. The steps necessary to choose and validate a variogram have been discussed in several references. The steps are: (1) Plot the experiment variogram; (2) Use at least the three models described in CN0137 to fit the experimental variogram (spherical, exponential, Gaussian). These models can be fitted to the empirical variogram using the method of moments (e.g. weighted least squares), restricted maximum likelihood, Bayesian methods, or any alternative that has been demonstrated in peer-review. (3) Plot and compare the fitted variograms to the experimental variogram, as illustrated in Appendix 6. (4) Choose the model with the smallest residual sum of squares or using a model selection criterion, such as AIC or BIC. This procedure is well-documented and does not require cross-validation, as requested by the comment. Please consult Oliver and Webster (2014, Section 3.1), or Webster and Oliver (2007, Section 5.6), or McBratney and Webster (1986) for examples of best practices consistent with the steps outlined here.

Section 5.5 Guidance on Variogram Selection and Fitting			
#	Organization	Comment	Developer's Response
		presents a simplified picture of variogram estimation that hides this and the tool would be stronger if it required validation of this important step.	

Section 6.1 Data and Parameters Available at Validation

Section 6.1 Data and Parameters Available at Validation			
#	Organization	Comment	Developer's Response
139	NMI-agro	Section 6.1 Data parameters: please keep units consistent MgC/ha. Do not work with acres. This is a recipe for conversion errors.	All units in the tool are in SI units (no references to acres).

Appendix 1: Assessment by Independent Modeling Expert (IME)

Appendix 1: Assessment by Independent Modeling Expert (IME)			
#	Organization	Comment	Developer's Response
140	Seqana GmbH	When DSM is used for model-based estimates, IMEs should not only be required to review model validation, but	See general response "Issues beyond the scope of CN0137."

Appendix 1: Assessment by Independent Modeling Expert (IME)			
#	Organization	Comment	Developer's Response
		also the uncertainty estimation with the DSM model. We agree that model validation is important. However, from an operational project perspective, arguably the uncertainty estimation is more important. Therefore, the IMEs should not only be required to confirm the validity of the model on the basis of the model validation criteria, but should also be required to confirm the appropriate processes and calculation methods are in place to correctly estimate uncertainties.	

Appendix 2: Illustrative Scenarios for Project Quantification Lifecycle

Appendix 2: Illustrative Scenarios for Project Quantification Lifecycle			
#	Organization	Comment	Developer's Response
141	The Nature Conservancy	The scenarios in Figure 1 and Appendix 2 are very helpful for visualizing the application of the tool in the project lifecycle. We would love to see more examples like this in future methodologies/tools. Related, it would be helpful to clarify in these figures which of the two use cases is being followed.	Thank you for your comment.

Appendix 6: Example Uncertainty Calculation

Appendix 6: Example Uncertainty Calculation			
#	Organization	Comment	Developer's Response
142	University of Illinois Urbana-Champaign	The code appendix makes two simplifying assumptions that I believe should be corrected for a more realistic example: (i) As alluded to in my comment on the prediction support unit, there needs to be sampling/measurement variability in the simulated samples. In my experience in the US Midwest, the variability (standard deviation) within a 10m x 10m pixel for SOC stock is about 8 Mg ha-1. This means that sampling a single core per pixel incurs an error in validation of about 8 Mg ha-1. Even with 10 random cores per pixel, this is only reduced to 2.5 Mg ha-1 (while increasing the number of cores prohibitively). This error will make variogram estimation noisier. (ii) It appears the code assumes that the true prediction error standard deviation at each prediction support unit is known during validation (as the denominator in the standardized prediction errors) In a more realistic scenario where these are known imprecisely, the estimated variogram will be noisier (see below).	One comment made two recommendations for changes to Appendix 6. First, assess sampling variation within individual pixels, and second, relax the assumption that the model prediction error is known. We agree that increasing the sample size within pixels will reduce uncertainty, and as stated in the text of the tool, composite samples are permitted. We also agree that in the real-world, the model prediction error is known less precisely than is the case under the simulation in Appendix 6. However, the purpose of Appendix 6 is not to create a simulation that exactly matches all conditions of the real world, but rather to demonstrate the basic functions of the tool with computer code so that the process is independently reproducible.

Overall General Feedback

Overall General Feedback			
#	Organization	Comment	Developer's Response
143	The Nature Conservancy	We thank Verra for the development of this tool and the opportunity to comment on it. There has been substantial interest in digital soil mapping as a tool to scale agricultural and rangeland carbon projects for years, and it is exciting to now have draft guidance from standards for stakeholders to react to.	Thank you for your comment.
144	Ambipar	The proposed tool (CN0137) represents a promising advance for estimating SOC stocks using DSM in projects for the generation of carbon credits in agriculture, applying VM0042. We emphasize our concern with the use of both DSM and BGCMs, regarding the calibration and validation of the models and these application scales. In this sense, we reinforce that their effectiveness depends on the quality of the input data – especially the analytical results of SOC and BD – and the rigor in the application of sampling, analysis and QA/QC procedures. This alignment must be confirmed in practice by project proponents (real world), through the careful implementation of area stratification, representative sampling and reliable laboratory analyses. The concept of Validation-Oriented Approach with Safeguards is well articulated and offers a robust framework, in which calibration, validation and QA/QC	See general response "Alignment with existing VCS methodologies" and general comment "DSM under VM0042 version 2.1."

Overall General Feedback			
#	Organization	Comment	Developer's Response
		provide the necessary guarantees for stakeholders. Nevertheless, the responsibility lies with VERRA scientists and VVBs to ensure that these procedures are applied with the utmost rigor, ensuring that hyperparameters, critical variables and covariates, such as variability in soil types, depth and textures, and climate conditions, as well as annual variations in management practices, land cover and diverse land use history, are duly considered. Although the tool is applicable on a global scale, specific challenges arise due to limitations in the resolution of geospatial information or where data capture is hampered by adverse conditions for generating an accurate DSM model. In summary, the CN0137 tool presents a robust proposal that is aligned with the requirements of VM0042. However, its implementation requires extreme rigor in calibration, validation and obtaining high-quality analytical data.	
145	Anonymous #1	- We hope this tool is being developed to align with proposals for VM0032 v2.0 and VM0042 v3.0, so the tool is future proofed and still relevant for these new methodologies. - Can the tool be applied in VM0051? In the context of requiring SOC measurement and monitoring in Quantification Approach 1 of VM0051.	One comment asked whether the tool can be used under other VM0047. Another asked about VM0051. The tool can be used with any VCS methodology that requires an estimate of SOC as a percentage by mass, SOC stock, bulk density, or how these quantities are changing over time. The tool is not specifically being developed to be forward compliant with revisions to existing methodologies, but where possible, the teams revising existing methodologies have contributed to the tool and we have attempted to avoid inconsistencies between the

Overall General Feedback			
#	Organization	Comment	Developer's Response
		- In Scenario 2, there is a non-grouped project where the model is not validated until verification. We therefore assume that it is permissible to therefore have an unvalidated model at validation for all methodologies where the DSM tool is applicable (as long as the GHG reduction credits are issued basis a validated model at verification). Can this be confirmed?	tool and future methodologies.
146	Regrow Ag	General Comments: Regrow Ag is generally in favor of the creation of a DSM tool to reduce soil sampling costs. Outstanding questions and comments related to the proposed CN0137 tool are as follows: Measure/Re-measure approach: unclear whether baseline control sites are required for measure-re-measure approach, and how model true-up would be accomplished at year 5 without baseline control sites. Recommend clarifying to include baseline control sites in order to support model true-up at year five. Soil sampling uncertainty: recommend that soil sampling uncertainty be incorporated into the methodology. Also clarify whether surveyed soil data rather than direct soil samples are allowed, and to capture the uncertainty related to those values if so.	One comment suggested that soil sampling variation should be incorporated into the tool. This comment is based on a misunderstanding. Soil sampling uncertainty is incorporated into CN0137. The measurement error component of soil sampling uncertainty is assimilated into model training and validation, and is represented in estimates of model prediction error. The spatial component of soil sampling uncertainty is handled through geostatistical methods. -- The tool specifically requires soil samples at all model validation events. -- See general response "Issues beyond the scope of CN0137" and general response "Alignment with existing VCS methodologies."

Overall General Feedback			
#	Organization	Comment	Developer's Response
		Soil samples: recommend increased clarity on how this aligns with the existing VM0042 guidance that requires a minimum number of soil samples per strata and.or if the use of the DSM tool has a different requirement for minimum samples not restricted to a definition of discrete strata. Model validation: Recommend that DSM tool have similar requirements as VMD53 for the validation data attributes (e.g. section 5.2.3) to harmonize across methods. Overall Regrow recommends that this tool be used to supplement but not replace direct soil samples at its current version, given the density of soil samples needed to conduct rigorous quantification within reasonable uncertainty.	
147	Hyphen Global AG	The Digital Soil Mapping (DSM) techniques described in the draft VCS tool (CN0137) could be complemented with continuous atmospheric-based flux measurements. Atmospheric-based instrumentation and methods quantify greenhouse gas (GHG) emissions and removals into and out of the atmosphere (fluxes) from all underlying processes (see Burba, 2022 and Pavelka et al, 2018). Flux measurements from on-site instrumentation enable continuous quantification of GHG fluxes for specific geographic areas.	See general response "Issues beyond the scope of CN0137."

Overall General Feedback			
#	Organization	Comment	Developer's Response
		Through atmospheric-based monitoring, emissions reductions – of CO_2, CH_4 and N_2O – are reflected as decreases in the flux quantities entering the atmosphere derived from on-site instrumental measurements. A decreasing flux provides observational evidence of sequestration, either in the soil or vegetation. Vegetation biomass carbon and nitrogen stock can be quantified on the basis of samples and laboratory testing and/or by the use of standard crop-specific biomass stock coefficients. Data on biomass stocks, in combination with atmospheric flux data obtained, for example, with eddy covariance and/or soil flux chamber instrumentation, provides further evidence constraining the whereabouts, in the carbon-pool sense, of any remaining unaccounted-for carbon or nitrogen not entering the atmosphere. Such observational evidence can be further combined with activity-based calculations using standard emission factors to estimate any off-site vegetation-, livestock-, transport- or production-related emissions associated with biomass harvest and off-take. By process of elimination, this combination of observations and calculated emissions confines any remaining carbon and nitrogen not found in the atmosphere (as CO_2 or N_2O) or vegetation biomass, or	

Overall General Feedback			
#	Organization	Comment	Developer's Response
		identified through calculations of any associated off-site emissions, to having been sequestered in the soil. Such ancillary information therefore provides both an independent verification of DSM outputs as well as a source of data for DSM model calibration. Burba, G. (2022). Eddy Covariance Method for Scientific, Regulatory and Commercial Applications. Lincoln, Nebraska: LI-COR Biosciences. https://www.licor.com/env/products/eddy-covariance/ec-book/ Pavelka, M., Acosta, M., Kiese, R., Altimir, N., Brümmer, C., Crill, P., Darenova, E., Fuß, R., Gielen, B., Graf, A., Klemedtsson, L., Lohila, A., Longdoz, B., Lindroth, A., Nilsson, M., Marañón Jiménez, S., Merbold, L., Montagnani, L., Peichl, M., Pihlatie, M., Pumpanen, J., Serrano Ortiz, P., Silvennoinen, H., Skiba, U., Vestin, P., Weslien, P., Janous, D., and Kutsch, W. (2018). Standardisation of chamber technique for CO_2, N_2O and CH_4 fluxes measurements from terrestrial ecosystems, Int. Agrophys., 32, 569-587. https://doi.org/10.1515/intag-2017-0045	
148	Bayer Crop Science	-Clarity and Accessibility: While the tool is generally well-written, some sections could be made more accessible to a broader audience. For example, the sections on uncertainty estimation could be simplified	See general response "Issues beyond the scope of CN0137."

Overall General Feedback			
#	Organization	Comment	Developer's Response
		and made more intuitive. -Advantages to this Tool/Approach: At the beginning of the document, it would be great to outline the advantages of this tool over a traditional measure + model or measure + re-measure approach. For example, if soil sampling is required with this tool regardless of approach, why should a project developer choose this tool/approach over others? -Cost-Effectiveness: The tool should provide more guidance on how to balance the cost of DSM with the benefits of improved SOC quantification. This is particularly important for smallholder projects, where cost is a major constraint. -Real-World Examples: The tool would benefit from more real-world examples of how DSM has been successfully implemented in different agricultural systems and regions.	
149	NMI-agro	Generic feedback to facilitate IME process: please add boxes (similar to VMD0053) with an overview of the requirements as set by this tool. This will facilitate project developers and speed up the IME evaluation process.	Thank you for your comment.
150	Indigo Ag	Thank you for providing us the opportunity to comment on your proposal to quantify soil organic carbon (SOC) using digital soil mapping (DSM). We are eager and excited to see DSMs incorporated into VM0042, and we greatly appreciate you bringing us	This comment asserts that the use of DSM weakens standards in VM0042. We disagree with this assessment, and suggest that the comment ignores ways in which the tool adds to the rigor of existing VCS methodologies and modules. The approach described in CN0137 is at least as rigorous as the current version of

Overall General Feedback			
#	Organization	Comment	Developer's Response
		closer to that reality. However, we have some concerns with how the proposal is currently written that we feel are critical to address before the proposed tool is adopted. In brief, as written currently: - The proposed approach weakens the rigorous standards of VM0042 and VMD0053 by removing the guardrail that predictions of stock change over time (relative to a counterfactual) are accurate and have well-calibrated uncertainty. - The proposed approach may allow incongruous methods to be used between project and baseline accounting, which would open up potentially large systematic biases. - The proposed approach does not make it explicit that project developers are still required to collect management data in order to demonstrate that practices changes have occurred on project fields and to quantify other sources of emissions. We outline these concerns below in greater detail Concern 1: While we support the use of DSMs to initialize biogeochemical models (Quantification Approach 1), we feel that	quantification approach 2 in VM0042, and arguably more rigorous, as noted by public comments, because CN0137 explicitly deals with geostatical sources of error that are ignored in VM0042 version 2.1, requires a sample within the project area for model validation, and never extrapolates in space or time. We very specifically considered rigor when developing CN0137 so that the scenario described by this comment cannot occur. The comment asserts that project proponents could "build many models with many different predictors, and select the one that maximizes the credits they could generate." However, we emphasize that the tool requires model validation against independent measurements at least once every five years within the project area, consistent with standards in existing VCS methodologies. Thus, we interpret CN0137 as meeting the requirements of Quantification Approach 2 in VM0042 version 2.1 with the addition of geostatistical handling of the error. -- Some comments questioned how to prevent fraud within the tool by users who might try to manipulate the validation dataset, or augment validation locations with manure. We agree that ensuring the integrity of the validation dataset is essential to maintaining confidence in the tool. While it is true that all methodologies rely on the good faith of project proponents, CN0137 includes safeguards that help to prevent manipulation: (1) VVB and IME review: All validation datasets are subject to third-party audit by an approved VVB. The VVB and IME must assess whether the validation sampling plan is within the project area, rigorously defined, and properly documented. Any attempt to exclude data or locations without justification would be visible during this audit process. (2) Pre-specified sampling protocols: CN0137 requires that project proponents document the sampling design in advance (e.g., stratification, randomization).

Overall General Feedback			
#	Organization	Comment	Developer's Response
		the use of DSMs for measure/re-measure (Quantification Approach 2) weakens the rigorous standards of VM0042 and VMD0053. As written, the proposal does not require developers to prove that their models can accurately measure change in SOC stocks. Developers are only required to prove that their models can accurately measure SOC stocks themselves. While DSMs are good at describing how SOC stocks vary across the landscape, there is limited evidence that DSMs are able to accurately measure change in SOC stocks over time. In order to use DSMs for Quantification Approach 2, we believe that developers should prove that their models are sensitive to SOC stock change by validating against repeated soil samples (so they have measurements of carbon stock change to compare their models against). Without this added guardrail, VM0042 could allow models with very little sensitivity to stock changes, or sensitivity in the wrong direction (e.g., if baseline fields have a bias toward stock declines). Making the validation criteria be based on stock change would greatly reduce the risk of allowing inaccurate models to be used for crediting. Here is an example of how inaccurate models could be used for crediting under this new tool: A developer could build a strong DSM model that can explain 90% of the variance in SOC stock across the	Exclusion of data after prediction, and without justification, is a breach of this protocol. (3) Transparency and traceability: All validation points must be geolocated and submitted as part of the project documentation. This creates a record that can be reviewed by VCS or third-party reviewers. (4) Statistical validation requirements: CN0137 does not prescribe a fixed validation sample size, but requires that the model meet quantitative benchmarks on an independent dataset. A small or manipulated sample is unlikely to pass these thresholds. These safeguards help to ensure that validation datasets are credible. -- See the general comment "Outcome-based validation."

Overall General Feedback			
#	Organization	Comment	Developer's Response
		landscape. However, static variables (like soil type and historical climate) might explain 88% of the variance, while variables that do vary from one year to the next (like spectral data from remote sensing) might only explain 2% of the variance. This would appear to be a very strong model, but it is not well suited to model change in SOC through time, because the most important variables are static (they are the same in every year of the model). Because project developers are not required to validate that the spectral data are picking up on real SOC change, they could build many models with many different predictors, and select the one that maximizes the credits they could generate (e.g., the model that over-predicts SOC stock change in the project fields, and under-predicts SOC stock change in the baseline fields). Because they are only required to validate SOC stocks, their chosen model would still appear very accurate (since the static variables are doing the heavy lifting), but they have now gamed the module to maximize their credits in ways that bias toward over-crediting. Therefore, without added guardrails, there is no motivation for the developer to build a model that accurately captures stock change; they can simply select a model that predicts stocks well and maximizes the credits they can generate. The precedent in VMD0053 is to check for	

Overall General Feedback			
#	Organization	Comment	Developer's Response
		bias and confidence interval coverage for impact (namely, SOC stock change in a treated field minus SOC stock change in a control plot), not for SOC stocks. That gives assurance that the method works well for the variable of interest, carbon credits. The proposed use of DSMs in Quantification Approach 2 lowers this standard by merely requiring efficacy in predicting SOC stocks.	
151	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Project area and mapped area 1. We would like to ask Verra to align the definitions of mapped area and project area: According to the VCS Project Definitions v4.5 (p. 19), the project area is "The geographic area in which the project activities are implemented." When using this definition in conjunction with, e.g., the non-eligible areas cited in VM32: "Non-eligible areas (e.g., if a project activity relates to improved grazing management – grassland areas where there is a change in practice due to the project –, the KML fi le should only be for the participating grasslands, as defi ned in this methodology, and should exclude any surrounding land, such as croplands, that may be part of the property" The resulting hectarage/acreage is the project area. This also aligns with our experience with VVBs. However, the DSM tool provides the following definition: "pixels that partially or	One comment requested clarification on the definition of "mapped area" and "project area". The comment stated that "In our experience, these are typically one and the same." These concepts are not the same in CN0137, as described in section 5.1.1 "Definition of the mapped area," which explicitly contrasts the mapped area with the project area. -- Numerous comments questioned the domain of application of DSM. Some questions the appropriateness of the < 30% tree cover threshold included in the draft version of the tool. Some requested justification for the threshold, while others argued that it was unnecessary or overly restrictive. In response to these concerns, we have removed the threshold from the tool. Project proponents will now be responsible for justifying alignment between the tool and the applied methodology at the time of project registration. This change does not signal an intention to expand the tool's scope to include agroforestry systems. Rather, the revision acknowledges that a fixed threshold is unnecessary to determine project eligibility. For example, an orchard with 50% tree cover may be consistent with the tool's design if the spatial arrangement of trees allows for ground-level observation from remote sensing between trees. Other comments asked for clarification

Overall General Feedback			
#	Organization	Comment	Developer's Response
		completely contain buildings, trees, waterways, or roads must be excluded from the mapped area. These exclusions do not change the project area." The DSM Tool appears to be using two terms—'project area' and 'mapped area.' In our experience, these are typically one and the same. We would welcome clarification and alignment from Verra on these definitions. 2. Similarly, CN0137 requires that "the mapped area contains less than 30% tree canopy cover," yet VM0032, for instance, defines grasslands as "lands with more than 250 mm mean annual precipitation covered by natural and managed herbaceous cover that lack trees over 5m in height with greater than 50% canopy cover." In order to utilize the tool in grassland settings and for VM0032, these criteria need to be aligned across ALM methodologies.	about the types of soils that are amenable to DSM. DSM is applicable to all soils, subject to the constraints in CN0137.
152	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Sample size, uncertainty equations and sampling 3. Would alternative data sources, including biogeochemical models and proprietary data, be allowed to estimate the Expected Average Project Effect (EAPE) if they provide scientifically robust estimates, even if journal articles are available? 4. Should EAPE not be simply called the 'Minimum Detectable Difference'? If not, what is the difference between MDD and EAPE?	The expected average project effect (EAPE) was defined in the tool as "An estimate from at least one peer-reviewed journal article (or proprietary data where no peer reviewed journal articles are available) of the expected mean change in soil organic carbon (SOC) stocks in the next five years of the project, expressed in units of SOC stock. EAPE is used to guide sample size requirements for model validation." This term was introduced to guide users of the tool when selecting a number of physical soil samples to be used for model validation. This term has been removed from the revised version of the tool in response to public comments and

Overall General Feedback			
#	Organization	Comment	Developer's Response
		5. Can you clarify if the DSM tool would supersede VMD53 for sampling uncertainties?	discussion with the VVB. One comment requests that the tool clarify how the concept of EAPE relates to effect size in VMD0053. However, the term "effect size" does not appear in VMD0053 version 2.1. One comment asked whether the EAPE is equivalent to a minimum detectable difference (MDD). No, the EAPE is not equivalent to MDD. The EAPE is an estimate of the average project effect that was used to guide sample collection. It does not specify the sensitivity of the model, and has no bearing on the number of credits generated by the project. One comment asked whether alternative data sources, such as a biogeochemical simulation, could be used to estimate the EAPE. Yes, the EAPE could have been estimated using published or proprietary data of any type. We emphasize that the role of the EAPE was entirely to guide the number of samples collected for model validation. The model must then pass all validation tests independent of the EAPE. That is, if the project proponent selects a number of samples and subsequently fails to pass model validation, the project either ceases to move forward or must recalibrate the model until passage occurs. -- See general response "Alignment with existing VCS methodologies."
153	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Model Validation and Performance Criteria 6. Is the requirement that R^2 must be > 0 sufficient to ensure a model's predictive strength as a minimum criteria? If so, are poor model performance and prediction uncertainty accounted for via deductions and could you make this explicit in the DSM documentation? If not, would a higher threshold (e.g., $R^2 > 0.5$) provide more confidence in model performance? 7. How can the validation process ensure	One comment asked how the tool can ensure that unsampled units treated as "missing at random" do not introduce systematic errors into predictions. By definition, unsampled units that are "missing at random" will not introduce systematic errors. It is when samples are missing nonrandomly that the likelihood of systematic errors arises. For example, if unsampled locations happened to be from areas where model performance was better, or worse, on average, then missing locations would be non-random, and systematic error could result. This is why the draft tool specifies that

Overall General Feedback			
#	Organization	Comment	Developer's Response
		that "missing at random" assumptions do not introduce systematic errors in SOC predictions? 8. If SOC variation is not random but tied to management practices, soil types or precipitation, this assumption could distort results, e.g. leading to over-reliance on indirect indicators rather than direct soil sampling. Could you please discuss? 9. Should there be mandatory sensitivity analysis to check whether missing data could bias SOC estimates?	validation samples should be a probability sample of the project area or justified as missing at random. A related comment noted that if "SOC variation is not random but tied to management practices, soil types or precipitation, this assumption [missing at random] could distort results." There are three points we would like to make in response to this comment. First, encouraging the use of a probability sample will prevent spatial variation in unsampled units from introducing systematic error into model validation. This is the purpose of using a probability sample. Second, it is not non-random variation in SOC that is the issue with missingness, but non-randomness in the accuracy of model output. Either way, a probability sample of the project area would ensure that systematic variation in the project does not undermine model validation. We note that these same principles underpin validation in all VCS methodologies that require samples of a project area. The same comment asked whether there should be "mandatory sensitivity analyses" to assess whether missing data could bias SOC estimates. However, this comment seems to misunderstand what is being referred to as "missing data". Through the use of this term, the tool is referring to unsampled locations. These locations are missing because they were never visited, not because of some error or a desire to exclude them. An example can help to illustrate this concept. Imagine a poll designed to forecast the outcome of an election. If 1,000 randomly selected people were asked how they would likely vote out of a voting population of 100,000,000, the 99,999,000 people not consulted by the pool are "missing at random". There is no way to perform a sensitivity analysis to determine what these other people, not consulted by the poll, might have said. This is why it is important that the 1,000 people who were

Overall General Feedback			
#	Organization	Comment	Developer's Response
			consulted in the poll were selected at random. If they were selected at random, then the assumption that unsampled individuals are missing at random is robust, and a prediction from the poll is valid. However, when the poll is non-random, for example, by focusing on certain geographic locations or demographic groups, missingness is non-random. Under such a scenario a sensitivity analysis would be warranted. Such an analysis is beyond the scope of CN0137, because CN0137 encourages the use of a probability sample and justification for "missing at random." -- See general response "Model-validation criteria."
154	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Reliance on DSM alone for SOC quantification, global applicability and control plots 10. Can you elaborate on how DSM-based SOC predictions can ensure proper uncertainty estimations/deductions given that uncertainty will vary with the availability of training data, especially in heterogeneous landscapes with limited soil data? 11. The document does not discuss potential biases in using DSM alone without sufficient ground-truthing (e.g. under VM32 v1, sample size is determined with the legacy methodology "CDM, A/R methodological tool: Calculation of the number of sample plots for measurements within A/R CDM project activities"). Can you please discuss such potential biases? Alternatively, if VM32 v2 is in the process of updating the sample size calculations, would it be appropriate to postpone	One comment asked for an explanation of the benefit of the tool in understudied regions or geographies. We see two clear benefits. First, biogeochemical models are not able to be applied worldwide because models have not been developed for all geographies and conditions.CN0137 overcomes this problem using DSM. DSM models are portable, meaning that they can be applied anywhere in the world where calibration and validation data exist. Because DSM models under CN0137 are always validated within the project area, DSM models do not depend on validations from peer-reviewed literature, a contrast with VMD0053. Second, DSM extends the value of physical soil samples. Because the size of SOC stock and how it is changing over time is computed by DSM, the number of physical soil samples required is significantly less than the number that would be required under a measure-remeasure approach, such as Quantification Approach 2 of VM0042. Both of these advantages benefit DSM worldwide, but especially in understudied regions or geographies. -- See general response "Outcome-based validation" and general response "Model-validation

Overall General Feedback			
#	Organization	Comment	Developer's Response
		applying DSM until v2 is fully approved or does the tool provide safeguards that supersede VM32 v1 requirements in this regard? 12. What is the potential applicability and benefit of the tool in understudied regions/geographies? 13. Control plots are essential for utilizing a measure re-measure approach under VM0042 V2.1 (though not VM0032 v1.0). While DSMs are a promising tool to help solve for control plots, particularly in rangeland settings, they are not in scope of this tool. We encourage Verra to include the use of DSM within control plots that are statistically defi ned using matching criteria in the future. Could you please comment on this likelihood?	criteria" and general response "Alignment with existing VCS methodologies."
155	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Carbon Credit Integrity & Recalibration Risks (Use Case 2) 14. In Use Case 2, how can early credit issuance be managed to avoid the risk of over-crediting prior to full model re-validation? Should initial credits only be issued following successful model re-validation? Could Figure 1/Appendix 2 delineate whether Use Case 1 or 2 is being utilized in order to better demonstrate where they differ?	We have clarified that Figure 1 corresponds with Use Case 2.
156	On behalf of Cultivo Land PBC and Kateri Environmental Corp	Transparency, Scrutiny and Open Data 15. Will the biogeochemical models within the DSM tool be subject to the same rules and requirements as those outlined in	See general response "DSM under VM0042 version 2.1" and general response "Outcome-based validation."

Overall General Feedback			
#	**Organization**	**Comment**	**Developer's Response**
		VM32 and VM42 (e.g. public availability)? 16. How can transparency in model assumptions and calibration data be improved?	
157	On behalf of Cultivo Land PBC and Kateri Environmental Corp	DSM tool providers and IMEs 17. Roughly, how many IMEs is Verra expecting to approve in the next 6-12-24 months? 18. Would prospective IMEs require both under-the-hood access and appropriate training from tool providers to become accredited? What incentives might encourage more scientists outside the carbon markets to take on the role of IMEs? 19. What is the process from which a statistical or machine learning method published in Web of Science journals is subsequently verified for specific DSM usage in accordance with CN1037's protocols and outcome-based performance? We're trying to better understand how this ensures consistency in DSM model acceptance and dissuades the use of biased or untested models.	See general response "Alignment with existing VCS methodologies."

GENERAL RESPONSES FROM DEVELOPER

1.1 Overview

Public comments to CN0137 were positive and constructive, with few requests for minor changes to the draft tool, and no requests for fundamental changes. We are grateful to Verra and members of the public who submitted feedback in response to the public draft of CN0137.

Below we provide general responses to seven themes that arose repeatedly in public comments. These responses address: key differences between digital soil mapping and current approaches under VCS programs, including outcome-based validation; issues of alignment between CN0137 and existing VCS methodologies; specific alignment with VM0042; issues beyond the scope of CN0137 to consider; the cadence of calibration, validation and verification under CN0137; applications of DSM to smallholder farms; and model validation criteria. These general comments and our responses clear-up some misunderstandings about DSM and its relationship with existing quantification approaches, including VMD0053. In all cases, changes have been made to the revised version of CN0137 that strengthen its clarity and interpretation, and we are grateful to the commenters.

1.2 Outcome-based validation

CN0137 is founded on outcome-based validation. Unlike VM0042 Quantification Approach 1, where a biogeochemical model is pre-validated on external datasets and credit integrity is protected by stringent input controls and true-up assessment of the model prediction error, CN0137 judges a DSM model solely on its ability to produce precise, unbiased estimates of SOC stock inside the project area. Innovation in model structure and data sources is encouraged; the decisive test is whether the model passes validation on an independent sample collected within the project boundary. This is consistent with the rigor of Quantification Approach 1 in VM0042, with the distinction that initial model validation is based on physical soil samples within the project area.

The tool requires validated point-in-time SOC stocks at t0 and subsequent time points. Change in SOC stock is then derived by difference. Validating each snapshot provides a strict test of model performance, and permits explicit propagation of spatial and temporal covariance into the variance of the mean stock-change estimate.

1.2.1 Calibration and validation data

Some comments questioned why the tool permits calibration data from outside the project area, but requires validation data exclusively within the project area. Other comments questioned why the tool does not impose stringent requirements on calibration data, such as a specific list of acceptable inputs.

Under an outcome-based validation framework, the test of model acceptability is empirical performance on an independent sample within the project area. Credit integrity is upheld at the output stage, not by controlling inputs. Accordingly, the tool is agnostic to calibration inputs. Field measurements, legacy trials, simulated values, or third-party datasets, even those collected far from the project site, may be used for model calibration. And model validation is project-specific. The model must satisfy the validation criteria using a hold-out validation set collected exclusively within the project area.

1.2.2 Safeguards

CN0137 contains numerous safeguards that uphold the quality of the carbon claim. Because CN0137 is outcome based, requiring model validation using physical soil samples within the project area and IME assessment, project proponents cannot “cherry-pick” easy locations that result in favorable apparent performance. The VVB and IME audit sample locations to ensure compliance with the tool. A DSM model that generates under dispersed predictions, biased predictions, or that lacks predictive power will not meet validation criteria specified in the tool. If a model that passes validation criteria is insufficiently precise to detect a change in SOC stock it will be penalized by the uncertainty deduction in the applied methodology. This discourages users from selecting a validation sample that is too small. Even if a model passes validation, it must still generate a precise estimate of the change to avoid a large uncertainty deduction.

1.3 Alignment with existing VCS methodologies

Many comments questioned how CN0137 aligns with existing VCS methodologies and modules. The table below summarizes key differences between CN0137 and VMD0053 version 2.1.

Theme	VMD0053 (process models)	CN0137 (DSM tool)
Validation domain	Relies on a domain-based approach to validation from peer-reviewed data sets; allows validation outside the project area.	Validation must use a sample collected inside the project boundary; validation is not based on practice categories or crop functional groups.
What is validated	Ability to reproduce ΔSOC (stock change) over multiyear intervals.	Ability to reproduce point-in-time SOC stocks at ≥ 2 dates; ΔSOC is obtained by differencing, with propagation of spatial–temporal covariance.
Bias & coverage tests	Required (bias; 90% PI covers at least 90% of observations).	Same two tests, plus minimum $R^2 > 0$
Calibration-data constraints	Strict limits on calibration data, update of model prediction error must occur at least once every five years; data within the project boundary is not compulsory.	Agnostic to calibration data; strict limits on validation data; validation and update of model prediction error must occur at least once every five years; data within the project area is compulsory.
Length of validation	A model may be re-used indefinitely unless its domain or parameter set changes. The model prediction error must be updated at least once every five years.	The model must be validated at least once every five years. The model prediction error must be updated at least once every five years.
Credit-issuance safeguard	If 90 % coverage fails, proponents may request a waiver; model prediction error informs the uncertainty calculation.	Failure of any test stops crediting; model prediction error informs the uncertainty calculation..

1.3.1 Four key differences between CN0137 and VMD0053

There are three key differences between CN0137 and VMD0053.

First, CN0137 is outcome-based, not input controlled. CN0137 does not prescribe a fixed catalogue of calibration inputs. Instead, it demands that the model prove, on an independent sample within the project area, (1) that its predictions are unbiased, (1) that its prediction intervals provide nominal coverage, and (3) that it explains at least some of the variation in independent validation data.

Second, VMD0053 assumes spatial transferability, but CN0137 avoids it. Under VMD0053 a model validated in one location can be applied at a distant site if the project domain is the same. CN0137 forbids this extrapolation: every DSM must demonstrate its performance within the project area where credits are sought. VMD0053 permits, but does not require measurements within the project boundary (see page 18 of VMD0053 version 2.1).

Third, CN0137 aligns the temporal scale of model validation with crediting, but VMD0053 does not. Process models are sometimes benchmarked against decades-long trials (e.g. Falloon and Smith 2000; Zhang et al. 2015). In contrast, CN0137 requires every DSM to pass validation at the five-year (or shorter) interval over which credits are issued. This avoids the assumption that long-term average sequestration is predictive of short-term changes.

1.3.2 Error sources in VM0042 version 2.1

One comment questioned whether CN0137 is out of alignment with VM0042, which requires accounting for model prediction error, sampling error, and measurement error. CN0137 accounts for all three of these error sources, but does so using methods that more closely represent the Monte Carlo approach to error propagation in VM0042, as opposed to the analytical decomposition of error described in VM0042 (e.g., Section 8.6.1.1 of VM0042 version 2.1). Model prediction error in CN0137 is the aggregated estimate of the variance of the mean SOC stock, or the variance of the mean change in SOC stock. This error includes sampling error (from the geostatistical aggregation described in CN0137, based on the methods of Wadoux and Heuvelink 2023), and measurement error (which is represented by the nugget of the variogram used to aggregate uncertainty the spatial mean). Appendix 6 illustrates exactly how error sources in CN0137 result in a credit claim and associated uncertainty deduction under VM0042 version 2.1 using a data simulation and computer code. Appendix 6 also illustrates the relationship between components of the uncertainty deduction in VMD0032 and CN0137.

1.4 DSM under VM0042 version 2.1

Because DSM produces an estimate of SOC stock, it can be appropriately applied under Quantification Approach 2 (measure-remeasure). One comment suggested that because DSM is model-based, it should be placed under Quantification Approach 1 (measure and model). This comment asserted that DSM is "conceptually similar to biogeochemical modeling." However, the output of DSM under CN0137 is an estimate of SOC stock at a single point in time, whereas biogeochemical models under Quantification Approach 1 produce a simulation of the change in SOC over time that may result from project activities. When this estimate of SOC stock from DSM is used to initialize a process-based biogeochemical model, the project proponent is implementing Quantification Approach 1 as described in VM0042 version 2.1. CN0137 does not change Quantification Approach 1 in VM0042. The only distinction is the source of SOC data used to initialize the biogeochemical model (physical soil samples or spectroscopy under the current implementation, and DSM in CN0137). In both cases the physical interpretation of the initialization data remains the same, and all other implementation and validation procedures remain identical.

Quantification Approach 2 uses measurements of SOC stock at two points in time to infer the change in SOC stock. This is identical to the use of DSM in CN0137, with the exception that CN0137 includes error propagation criteria based on geostatistics.

1.5 Issues beyond the scope of CN0137

A number of comments raised questions that are beyond the scope of CN0137 to address. These comments fall into two categories: general comments or questions of implementation that are at the discretion of the project proponent and likely to be context dependent, and questions or comments related to existing VCS methodologies. These include requests for financial analyses of cost-effective sampling strategies; direct comparisons between CN0137 and alternative implementations of Quantification Approach 2 in VM0042; guidance in the selection of process-based biogeochemical models under Quantification Approach 1 in VM0042; additional worked examples of the tool beyond the code repository contained in Appendix 6; the addition of guardrails against new innovations; a request for discussion of heterogeneity in smallholder farms; guidance on the collection of citizen-science or historical data; guidance on the selection of an appropriate DSM model architecture; a request to recommend the inclusion of traditional ecological knowledge; discussion of baseline-control scenarios (these are handled by the applied methodology in all cases); a request for discussion of alternative carbon pools and their constituents; a request for recommendations of specific use cases, sensors, and image processing routines; requests for an explanation of the algorithms within specific machine-learning model architectures (e.g., how confidence intervals are defined in gradient-boosted regression trees); requests for design-based estimation strategies (the tool does not operate under the design-based paradigm); and requests for changes to the IME or VVB evaluation process. We emphasize that answers to many of these questions can be found in the set of representative publications in Appendix 3 of the draft tool.

1.6 Frequency of model calibration, model validation, and verification

Three comments addressed the frequency of model calibration, model validation, and verification. One comment asked whether CN0137 permits back-estimation of t0 under Quantification Approach 2 in VM0042. In all cases, CN0137 follows the applied methodology.

One comment mentioned that small annual increases in SOC may not be detectable annually. Because the tool is outcome-based, if changes are not detectable, or they are detectable but uncertainty is large, application of CN0137 will result in a large uncertainty deduction, including the possibility of no economic value. This is a component of the "validation with safeguards" approach in CN0137. Only those claims that are rigorous and verifiable are supported by the tool.

Finally, one comment requested clarification about the distinction between DSM model validation and true-up of a process-based biogeochemical model. It is important to emphasize that DSM under Quantification Approach 1 is ordinary Quantification Approach 1, with the sole distinction that estimates of SOC stock for model initialization are from DSM. The DSM model must be validated at the time point represented by the initialization of the biogeochemical model. All other model validation, verification and true-up requirements are governed by the applied methodology.

1.7 DSM in smallholder farms

One of the most important benefits of the outcome-based approach described in CN0137 is that it is readily applicable to small-holder farms. This is because (1) DSM requires fewer physical soil samples than are necessary under existing Quantification Approach 2, and (2) DSM model validation does not require an extensive peer-reviewed history of the type required for process-based soil biogeochemical models under VMD0053.

One comment mentioned the "fragmentation of agricultural management practices and land use history" in comparison to the resolution of the data needed to calibrate a DSM model as a potential concern. But fragmentation of management practices imposes no special challenge to DSM. This comment likely reflects the assumption that DSM models are validated in the same way that process-based biogeochemical models are validated under VMD0053. Under the domain-based validation of VMD0053, fragmentation of land management practices can be challenging, because a different parameter set may be required from field to field. However, under CN0137, model validation is not based on crop functional groups or practice categories. Validation occurs using physical soil samples within the project area.

One comment noted "data scarcity" as a challenge to the use of CN0137 in small-holder farms. However, CN0137 was explicitly developed to take advantage of globally available data products. There are numerous remote sensing, topographic, climate, and weather-related variables that are available as gridded data products worldwide. CN0137 always requires local measurements from physical soil samples for model validation. This is consistent with other VCS methodologies, which require physical soil samples and / or farm practice data.

One comment noted that the sampling design must capture heterogeneity in small holder farms. We agree, and note that CN0137 requires a validation sample within the project area.

1.8 Model-validation criteria

One comment suggested that a statistical power requirement should be adopted to guide the number of samples necessary for model validation. This comment expressed the concern that in the absence of a formal power requirement, the sample size guidance in Equation 8 of the draft tool could lead to “a weak test of unbiasedness.” The working group responsible for developing CN0137 discussed this question and concluded that a power test is unnecessary for the following reasons.

First, DSM models are typically highly accurate when evaluated against point-in-time measurements. Correct estimation of the variance of the mean change in SOC stock is therefore likely to be more important than the bias test.

Second, power-requirements reflect a tradeoff between Type II error and cost. Statistical power is formally defined as the probability that a given statistical model will detect an effect of a given size when the effect is present. By convention, statistical power of 80% is desirable. But even 80% power implies a 20% chance of failing to detect an effect that is present. Why is a 20% failure rate acceptable in the context of a VCS carbon program? The optimal failure rate will be a balance between logistical and financial costs of project administration and the consequences of model bias to the integrity of the credit claim. Neither of these questions has been answered in the context of VCS carbon programs.

Third, determining the appropriate sample size to result in a given power level depends on the magnitude of the effect one wishes to detect. But there is no way to determine the a priori magnitude of the effect with confidence. This is why the sample size recommendation in Equation 8 of the draft tool was distinguished as “the approximate sample size.” The tool now provides an explanation of the role of sample size in model calibration, recalibration, and model validation that is consistent with guidelines in VM0042 version 2.1. The tool notes that “The IME must determine whether the sample number is sufficient for the model to meet the three validation criteria and aligned with requirements in the applied methodology.”