WHITE PAPER | SEPTEMBER 2026

The Uncertainty Challenge: Making Soil Carbon Measurement Credible at Scale

Part 2: Metrics and Mathematics of Verra VMD0053 & VT0014

A D-CAT Publication

CONTACT
info@d-cat.co.uk | www.d-cat.co.uk

Digital Content Analysis Technology Ltd

1 Overview

This is the second of our three-part series examining the state of monitoring, reporting and verification (MRV) of soil organic carbon that is intended to provide insight and commentary on current technology, practices and trends as well as remaining challenges.

Part 1 examined the challenges faced when building trust in modelled soil organic carbon (SOC) and how Verra[4] VMD0053[2] and VT0014[3] methodologies address those challenges to allow confidence in the results to be established. In this Part 2 of the series, a technical perspective on the quantification framework is given: digging deeper into the actual mathematics underlying VMD0053 (the model validation module) and VT0014 (the digital soil mapping tool), how uncertainty is propagated, what the various validation metrics are actually testing and how rigorous those tests really are. Part 3 will then consider the challenges and practical realities of implementing SOC MRV services that deliver Verra-compliant results cost-effectively at any scale, to enable widescale adoption.

2 Introduction

Beneath the policy language of Verra's agricultural land management framework sits a genuine mathematical apparatus carefully designed to address a statistical problem of considerable sophistication: how can a limited number of physical observations, combined with models and spatial information, support a defensible estimate of a change in soil carbon stocks across potentially very large and heterogeneous areas?

Verra's VM0042[1], VMD0053 and VT0014 represent increasingly sophisticated attempts to solve this problem:

- VM0042 is the accounting framework. It defines how emission reductions and SOC removals are quantified, including the treatment of uncertainty and the alternative approaches available for measuring SOC change. Verra's current VM0042 v2.2 permits several SOC measurement approaches and includes explicit uncertainty procedures. Essentially, VM0042 asks: "***Does the process-based model adequately represent the system and its uncertainty?***"

- VMD0053 is the model assurance framework. This governs calibration, validation, parameterisation and uncertainty assessment for biogeochemical models used to estimate SOC changes and greenhouse gas fluxes. The current v2.1 has been active since March 2025, while v3.0 is being developed alongside the next major VM0042 revision. Essentially, VMD0053 asks: "***Can sparse soil observations be defensibly transformed into SOC predictions, with quantified uncertainty?***"

- VT0014 is the spatial prediction framework. It provides a formal route for using digital soil mapping, remote sensing, geospatial information and statistical methods to generate spatially explicit SOC estimates, including their uncertainty. It became active at v1.0 in August 2025. Essentially, VT0014 asks: "***How do those measurements and predictions translate into a defensible carbon-accounting outcome?***"

Together, they create a framework that is considerably more rigorous than simply estimating carbon from a model or extrapolating a handful of soil samples across a landscape. With rigour comes complexity, more so in method compliance than mathematics, but complexity that requires specialist expertise to implement, liaise with the Independent Modelling Expert (IME) required by Verra to audit projects for compliance, and ensure correct quantification of change in SOC with due consideration and handling of uncertainties.

The sophistication of the mathematics involved is not in question, as this paper will show, and the standards continue to evolve over time. Each standard is firstly summarised before the key metrics used by each to quantify SOC change over time are set out along with the essential approach to how uncertainties are handled and accounted for. This latter element is key, not only to understanding model characteristics and validity, but to building the trust in the numbers reported for carbon credits – there must be a credible estimate of the uncertainty surrounding the SOC change reported.

KEY MESSAGES

- Validating SOC models is not a simple task due to many factors affecting model uncertainty.
- Verra's framework for addressing the problem is rigorous and based on well-established, credible mathematical metrics and statistical methods, tailored to handle the specific challenges of soil variability and substantial uncertainty.
- Within the VM0042 framework, VMD0053 and VT0014 specify clear methodologies and metrics for establishing model validity, and hence credibility in SOC changes reported.
- VT0014 (for DSM) is more prescriptive and demanding than VMD0053 (for BGCM-based projects), with spatial variability requiring more complex statistical treatment.
- Neither is perfect, but VMD0053 and VT0014 continue to evolve and their current state is sufficient to enable trust and confidence to be built in models, and results from models, that are Verra-compliant.

3 A Deceptively Simple Challenge: Mean SOC Change

At its most basic level, soil carbon accounting appears straightforward: the estimated change over time is simply the difference between the SOC at time *t0* and *t1*. However, in practice, neither value is known for certain – measurements have associated uncertainties and so too do modelled estimates – meaning the actual quantity of interest is not simply a difference between two known numbers, but a statistical estimator of a difference with an associated variance and ultimately an uncertainty interval.

This complexity is fundamental to the mathematical frameworks designed to validate models that are used to estimate mean SOC change. A project can have a large, estimated increase in SOC at the same time as a large uncertainty, which would result in a smaller increase in SOC being submitted for credits. In other words, the carbon stock change and the confidence in that value are inseparable and any validation methodology for carbon accounting must recognise and quantify this dependency.

Further difficulties abound, not the least of which is the fact that soils are spatially correlated, meaning that soil carbon observations are not generally independent and identically distributed observations. Two samples a few metres apart are likely to be more similar than two samples a few kilometres apart, and so the effective amount of information contained in a set of samples may be considerably less than the total number of samples. Consequently, sampling design matters as much as sample count.

Another challenge is the variety of factors affecting the uncertainty in SOC change, including (but not limited to):

- soil sampling error;
- laboratory measurement error;
- bulk-density uncertainty;
- spatial prediction error;
- model parameter uncertainty;
- model input uncertainty.

Covariance must be propagated through uncertainty calculations for errors that are correlated, and the same issue becomes particularly important when combining model outputs from multiple simulations or when scaling pixel-level predictions to field- or project-level.

How then does the Verra framework handle errors, propagate uncertainty and build trust in models and their outputs through a validation methodology, and how does it stand up to the challenges as well as market needs?

4 VM0042 – Accounting Under Uncertainty

VM0042 is the parent methodology, whereas VMD0053 and VT0014 are the technical instruments used to quantify the SOC change reported, whichever of the two modelling approaches is taken. At the methodology level, the main mathematical principle is error calculation with uncertainty propagation.

Uncertainty is treated as compounding: total uncertainty is calculated as the square root of the sum of the squares of each component uncertainty, reflecting standard error-propagation practice for independent error sources feeding into the same final estimate. That combined uncertainty is then evaluated against a confidence interval — commonly a 90% interval — with wider intervals impacting the credits a project can claim. Total uncertainty is calculated as:

$$U_{total} = \sqrt{\sum_{i=1}^{k} U_i^2}$$

Here, U_i is the uncertainty of the i^{th} independent component (e.g., baseline measurement error) and k is the number of components.

This has an important mathematical consequence in that uncertainty doesn’t average out across error sources; instead, it compounds under a square root law. A project with three moderately uncertain inputs (say, baseline SOC measurement error, model residual error, and spatial upscaling error) does not get to simply average those three uncertainty terms — it inherits something close to the largest of them, inflated further by the others. This is precisely why the model validation statistics discussed below matter so much: a poorly characterised model isn't just “somewhat wrong”, it can dominate the entire project’s uncertainty budget.

5 VMD0053 – From Prediction to Model Credibility

VMD0053 governs how a biogeochemical model (BGCM) — commonly RothC, DayCent, or similar models — earns the right to be used for quantification of SOC changes over time. Such models are process-based and don’t simply predict a single number. They model carbon pools and transfers over time using differential or difference equations, driven by environmental conditions and management inputs, and therefore require calibration as well as validation.

VMD0053 therefore provides procedures for model calibration, validation and parameterisation and uses IMEs to audit the process. Model bias (average relative error) is explicitly used to establish a BGCM’s fitness for purpose, with other potential metrics left unspecified, such as for model prediction error, and hence metrics such as Root Mean Square Error (RMSE) could be used to further quantify error. Under VMD0053, a model must demonstrate performance across the relevant range of soils, climates and management conditions, whilst the validation dataset must remain sufficiently independent from the data used to calibrate the model. Otherwise, excellent model statistics can simply reflect overfitting.

Bias

Bias (or error) captures systematic over- or under-prediction:

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)$$

Model bias is a very important metric and is arguably the single most safety-critical statistic in the whole validation stack since a biased model could result in consistent overestimation of SOC gains. So, although a model may have an acceptable RMSE, it could still be systematically biased in one direction.

For a BGCM to be compliant with VMD0053, its bias must be less than or equal to the inherent average background "noise" (measurement error) present in the field validation datasets. This measurement quality threshold is termed the Pooled Measurement Uncertainty (PMU) and ensures that the model's systematic error is no greater than the uncertainty associated with the underlying measurements. The PMU is calculated by pooling the standard errors of individual observations, weighted by their degrees of freedom:

$$PMU = \sigma_{meas} = \sqrt{\frac{\sum_{j=1}^{k} \sigma_j^2 (n_j - 1)}{\sum_{j=1}^{k} (n_j - 1)}}$$

Here, σ_j is the standard error of the j^{th} measured change, n_j is the number of measurements contributing to that observation, and k is the total number of observations across the validation studies.

Root Mean Square Error (RMSE)

Model prediction error could be quantified using RMSE. This focuses on a fundamental: how close are a model's predictions to the reality they are supposed to estimate? RMSE captures this by measuring the average magnitude of the prediction error in the same units as the underlying quantity:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - P_i)^2}$$

Here, O_i is the observed value at point i; P_i is the model's predicted value at point i; n is the number of validation points.

Despite its simplicity and limitations, RMSE is universally understood and gives a sense of model accuracy before further behaviours are assessed. It also brings the benefit of penalising large individual errors more heavily than small ones (because the errors are squared before averaging), which allows it to flag a model that is occasionally wildly wrong even if it is usually close. However, RMSE does not, by itself, tell us whether the uncertainty estimate associated with the prediction is correct.

Benefits and Limitations

Bias alongside RMSE provides confidence from different perspectives: bias shows whether the model is systematically wrong in one direction, while RMSE shows how large the errors are.

These two metrics build confidence but there remains the possibility that the uncertainty reported by the model for its estimates is unrealistic, and so VMD0053 introduces a coverage test: essentially, a test of whether a model is honest about how uncertain its predictions are. The model gives each prediction a range within which the true measured value is expected to fall — VMD0053 requires that at least 90% of the independent validation measurements fall within the model's 90% prediction interval. Importantly, this ensures that the model is tested not only for estimation accuracy, but whether the uncertainty it reports around those estimates is realistic.

6 VT0014 – Spatial Considerations

In recent times, digital soil mapping has emerged as an alternative to BGCMs for estimating SOC change and this has driven the need for a methodology that goes beyond the bounds of VMD0053 and addresses the validation of spatial predictions of SOC over time: VT0014.

VT0014 accommodates the use of physical soil samples, BGCMs and also DSMs (typically machine-learning models trained using satellite imagery, soil and terrain maps, and climate data) to estimate SOC change, and uses the same principles and core metrics as VMD0053 (bias, RMSE, R^2, coverage). However, it necessarily takes these concepts further because modelling is now predicting SOC across an entire landscape, including locations where no soil sample exists. Consequently, VT0014 has to deal with both prediction error and spatial dependence, extending the framework of VMD0053 and adding the mathematics required when those predictions must be continuous across space, including spatial dependence, interpolation, prediction variance and the propagation of uncertainty into an estimate of mean SOC change.

The spatial problem is what VT0014 addresses through rigorous handling of spatial considerations in model uncertainty. Fundamentally, the methodology must handle the fact that SOC predictions at every location have an associated uncertainty, and those uncertainties are themselves influenced by:

- the quality and density of soil measurements;
- the relationship between SOC and the explanatory variables;
- spatial correlation;
- the distance from a prediction location to observations;
- the spatial distribution of the validation samples.

This means that the quantification task is no longer "How wrong is the model on average?" but "How uncertain is the prediction at this particular location, given the available measurements and the spatial variability across the landscape?" Spatial interpolation therefore plays a crucial role in VT0014. Instead of treating every soil measurement as an independent observation, spatial modelling recognises that measurements taken close together are often more similar than measurements far apart.

Unlike VMD0053, VT0014 explicitly defines metrics to quantify the model's errors and uncertainties. Alongside bias and RMSE, the following are stipulated:

The Coefficient of Determination

The coefficient of determination, R^2, describes the proportion of variance in observed values explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}\left(O_i - \overline{O}\right)^2}$$

Here, $\overline{O}$ is the mean of the observed values.

A very modest threshold of $R^2 > 0$ is set as one of the validation tests. R^2 is popular because it's intuitive, but it has a well-known weakness that is relevant here: it measures the strength of the linear relationship, not agreement. This means that a dataset can return an R^2 score of 1 (i.e. a perfect linear fit) even if the dataset is shifted, as long as the total and residual variances remain unchanged. This blindness to systematic scaling and offset errors is exactly why Bias and other metrics are also used.

Variability

Mathematically, VT0014 handles the spatial variability of sample data by means of a semivariogram model (defining the relationship between two data values according to the distance between them):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2$$

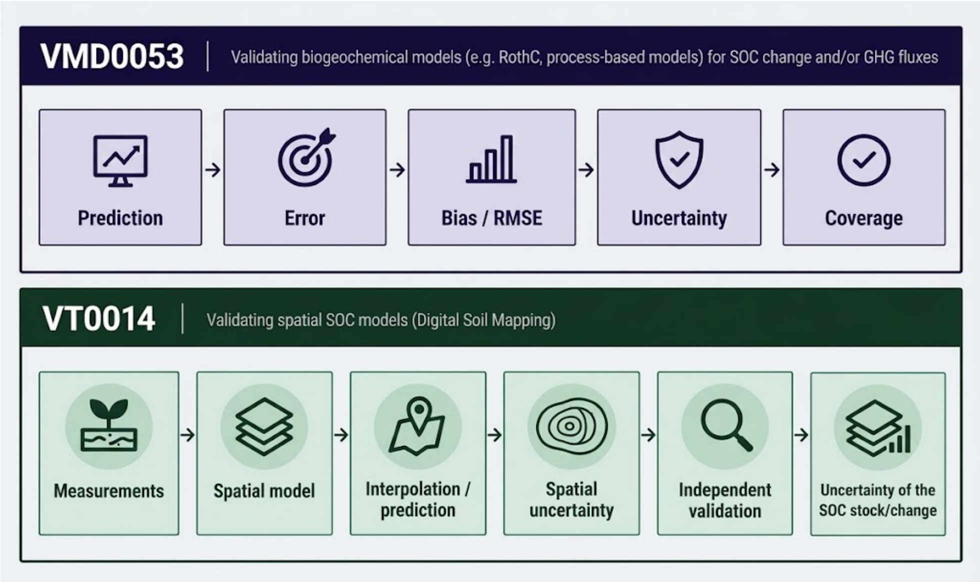
Here, h is the separation (lag) distance between a pair of points; $N(h)$ is the total number of pairs of sample points, $z(x_i)$ and $z(x_i + h)$ are the measured SOC values at location i and at the location with distance h from i.

Determining the spatial relationship accurately is one of the most important steps in VT0014, as soils typically exhibit substantial spatial variation that is not always adequately captured through soil sampling. Knowing how much the SOC measurements become less similar as the distance between them increases allows distinction to be made between spatially correlated variation and unexplained, random variation.

Kriging, a spatial interpolation method for predicting values at unmeasured locations and estimating their spatial uncertainty, is then used to generate estimates across the unsampled locations for comparison against model predictions. VT0014 allows for different kriging approaches to be applied but stipulates the need for transparency of approach and justification, and similar flexibility is shown to the model used for variogram fitting (e.g. spherical, exponential, Gaussian) as long as the model chosen is stated and the reason for the choice is justified.

7 Summary Comparison

The differences between the VMD0053 and VT0014 approaches are readily captured graphically, with a more detailed comparison given in tabular form.



Topic	VMD0053	VT0014
What is being tested	Predictions, behaviour and accuracy of a SOC process model.	Pixel estimates of SOC at a snapshot in time, geographically spaced.
Core metrics	Bias, model prediction error (e.g., RMSE).	Bias, RMSE, R^2, variance.
Primary modelling issues it catches	Model bias.	Spatial extrapolation error.
Independence requirement	Validation data must be mutually exclusive from calibration data.	Validation locations must be spatially held out.
Nature of the underlying claim	A forecast of SOC change, conditioned on correct baselining and assumptions.	An estimate of current SOC stock and its spatial variability.
Statistical sophistication required of reviewers	Time-series and mechanistic-model diagnostics.	Additionally, geospatial statistics.

8 Rigour and Legitimacy

Sound principles of model validation as well as mathematics and statistics therefore sit at the heart of both VMD0053 and VT0014 – these are not novel inventions created specifically for carbon markets, but well-established techniques and metrics carefully constructed to apply rigour and provide confidence. A strong argument can therefore be made, even though there is always room for revision and improvement, that the underlying statistical architecture of the Verra framework is legitimate, encompassing:

- prediction error;
- goodness-of-fit;
- estimation of bias;
- confidence and prediction intervals;
- variance propagation;
- covariance;
- spatial statistics;
- Monte Carlo simulation (used in estimating variance); and
- independent expert review.

Both instruments are legitimate in the sense that they require real statistical testing against real data — this is not “trust us” carbon accounting. However, what will make the MRV process, including Verra accreditation, credible is the integrity of the evidence chain connecting measurement to the final carbon claim.

9 Conclusion

The mathematics underneath VMD0053, VM0042, and VT0014 are more rigorous than perhaps common opinion gives them credit for — real error propagation, exploration of model behaviour, transparency and independent validation. But rigour on paper and rigour in practice are different things. The metrics only do their job if there is sufficient independent expertise to compute and review them correctly, sufficient data density to compute them meaningfully, and trust in all components. With rigorous validation methodologies in place and continually being improved, the ongoing challenge to build confidence in and drive adoption of soil carbon MRV can only be made easier as knowledge and appreciation of how seriously and professionally validation is handled to achieve compliance becomes more widespread.

ABOUT D-CAT

D-CAT provides enterprise-grade observed ESG analytics, combining satellite Earth Observation with cloud-native processing, validated algorithms and AI to produce continuously updated environmental intelligence. Our platform integrates into existing ESG workflows, supporting carbon projects and providing automated SOC modelling and MRV reports for VMD0053 and VT0014 from pixel to farm, regional and national scales.

info@d-cat.co.uk | www.d-cat.co.uk | LinkedIn: Digital Content Analysis Technology

References

1 **VM0042, Improved Agricultural Land Management, v2.2**
https://verra.org/methodologies/vm0042-improved-agricultural-land-management-v2-2/VM0042

2 **VMD0053, Model Calibration, Validation and Uncertainty Guidance for Biogeochemical Modeling for Agricultural Land Management Projects, v2.1**
https://verra.org/methodologies/vmd0053-model-calibration-validation-and-uncertainty-guidance-for-the-methodology-for-improved-agricultural-land-management-v2-1/

3 **VT0014, Estimating Organic Carbon Stocks Using Digital Soil Mapping, v1.0**
https://verra.org/methodologies/vt0014-estimating-organic-carbon-stocks-using-digital-soil-mapping-v1-0/

4 **Verra**
https://www.verra.org