'Towards a modular, multi-ecosystem Monitoring, Reporting and Verification (MRV) approach for soil organic carbon stock change assessment'

Supplementary information (S1) – Uncertainty assessment

This supplementary information discusses: 1) Uncertainty assessment, 2) Statistical modelling of uncertainty, 3) Sources of uncertainty, 4) Uncertainty propagation, 5) Upscaling uncertainties, and 6) Statistical validation in the context of MRV systems. Main references cited in S3 are listed in Section 7.

1. Uncertainty assessment

Uncertainty quantification plays a pivotal role in MRVs due to their focus on natural systems, which can never be perfectly known. Owing to factors such as spatial variability, complexity of physical, chemical and biological processes, and numerous other error sources, estimations of system states frequently harbour significant uncertainties. Often, these uncertainties can be so substantial that they impede the demonstration of the impact of specific land management practices. To illustrate, uncertainties in estimated SOC stocks for a given field or region might be that large to render reported changes in carbon stocks — between project commencement and conclusion or between baseline and regenerative practice implementation — statistically insignificant. This can result in erroneous conclusions and speculative assertions, potentially leading to unwarranted claims of practice effectiveness (referred to as 'green washing'). Recent articles in prominent outlets like The Guardian[1] and Science[2] have highlighted this concern, caused significant commotion and prompted temporary modifications or withdrawals of certain MRVs. Thus, within MRVs, acknowledging and quantifying uncertainty stands as a crucial obligation. Any claims put forth by projects must invariably be substantiated by statistically rigorous evidence.

The aim of this section is to provide a generic description and short review of uncertainty assessment in the Earth and environmental sciences, whenever possible illustrated with examples from MRV practices.

[1] https://www.theguardian.com/environment/2023/jan/18/revealed-forest-carbon-offsets-biggest-provider-worthless-verra-aoe
[2] https://www.science.org/content/article/farmers-paid-millions-trap-carbon-soils-will-it-actually-help-planet

2. Statistical modelling of uncertainty

Uncertainty finds its most accurate portrayal through probability distributions. Probability distributions of uncertain environmental variables can be complex. It would be naïve to assume that all that is needed is a mean and a variance of a probability distribution, because the uncertainty about some variables is not realistically characterised by a parametric distribution , while cross-correlation, spatial and temporal autocorrelation must also be accounted for [1]. Moreover, there are many sources of uncertainty, which all need a specific approach, as explained in the next subsection.

3. Sources of uncertainty

Many MRVs make use of process-based models during the reporting phase. In such case, there are three main sources of uncertainty that must be considered [e.g., 2] :

1. Uncertainty in the model inputs
2. Uncertainty in the model parameters
3. Uncertainty in the model structure

Uncertainty in model inputs often result from measurement error. Van Leeuwen *et al.* [3], for example, stressed the importance of considering measurement error in wet chemistry data, which is particularly important when soil data are derived from various sources. The same applies for dry combustion data derived using various methods [4]. Sampling and sample preparation errors were found to be of the same order of magnitude as errors caused in the chemical analyses themselves [5-7]. Similar issues arise for soil physical properties [8]. When looking at changes in SOC stocks it is also important to consider whether these were computed on a fixed depth or an equivalent soil mass basis [9,10]. The equivalent soil mass method, for example, is recommended in the GSOC MRV protocol [11] while the IPCC [12] guidelines still consider a fixed depth. Further, Stanley *et al.* [13] indicated that the accuracy of SOC measurements is limited by inherent spatial heterogeneity, variability of laboratory assays, unmet statistical assumptions, and the relatively small magnitude of SOC changes over time, which hampers measuring SOC change.

When models use inputs that are not directly measured but are derived from secondary sources, such as maps (e.g., biophysical variables derived from remote sensing) or expert judgement, additional uncertainty will arise. These could be spatial interpolation errors [14] or regression errors [15, Section 12.3]. Expert judgement uncertainty can be assessed through expert elicitation procedures [16], although this is cumbersome and not entirely free of subjectivity. Complex models represent processes better than simpler models (e.g., Tier 3 approaches are preferred over Tier 1 approaches), but complex models require more inputs. If these inputs are poorly known and have

large uncertainties, then replacing a simple by a complex model might actually deteriorate results [2].

Expert judgement could also be used to quantify model parameter and model structural uncertainty, but this is challenging and can be unreliable. A better approach is to derive these uncertainties from statistical approaches that compare model outputs with independent observations, such as in Bayesian calibration [17] and data assimilation [18,19]. Model structural uncertainty is usually represented by an additive or multiplicative noise term [20,21]. There are also approaches that characterise model structural uncertainty by a multiple modelling approach [22,23]. Quantifying model uncertainty and model validation are still challenging and in practice often lacking [24].

Multiple sources of uncertainty can impact the quality of the model's output. However, not all of these sources require inclusion in an uncertainty analysis. Focusing on the sources with the most significant impact, as determined by uncertainty propagation techniques elaborated below, suffices. While discerning the primary sources of uncertainty beforehand is challenging, employing a Quickscan' [1,25] might aid in this determination process.

Quantification of uncertainty is a challenging and complex task. A related challenge is how to comprehensibly communicate this uncertainty to different stakeholders so that they can correctly interpret the results.

4 Uncertainty propagation

Uncertainty propagation can be traced in various ways, but the most commonly applied methods are the Taylor series method [26-29] and Monte Carlo simulation [1,30,31]. The advantage of the first method is that it is fast and yields an interpretable mathematical equation, while the advantages of the second method are that it is easily implemented, generally applicable and that its approximation errors can be made negligibly small, provided sufficient computing resources are available. Many MRVs listed in Section 3 make use of the Monte Carlo method. Examples of uncertainty propagation analyses in MRV context are given elsewhere [19,32,33].

The end result of an uncertainty propagation analysis is a probability distribution of the model output. Note that in case of Monte Carlo uncertainty propagation this will take an empirical form (i.e., a random sample from the distribution), while the Taylor series method only yields the mean and variance of the distribution, implying that additional assumptions about the shape of the distribution (e.g., normal, lognormal) are needed to fully characterise model output uncertainty. Communication of uncertainty to end users is perhaps best done by presenting the lower and upper limits of a prediction interval (e.g., the 0.05 and 0.95 quantiles of the distribution). For instance, a project might compute and present the lower and upper limits of a 90% prediction interval of the SOC stock difference between the baseline and regenerative practice. Note that if this interval

includes zero (i.e., if the lower limit is negative and the upper limit positive) this implies that the estimated difference is not statistically significant from zero at the 90% confidence level.

5. Upscaling uncertainties

Most variables of interest of MRVs, such as GHG emission and SOC, vary in space and time. Often, the interest is not in the value of these variables at points but in the average or total for an area (e.g., a field, region, entire country or the globe) and/or time period (e.g., a day, month, year or decade). Upscaling the predictions of these variables is easy if predictions are available for the whole area or time period, but quantifying the associated uncertainties is much more difficult. It can only be done if the spatial and temporal correlations of the prediction errors are known and accounted for, using autocorrelation functions and/or semi-variograms [34]. In spite of its importance, this problem seems largely ignored by the scientific community [e.g.,35,36].

Spatial and temporal aggregation lead to a decrease of uncertainty. The uncertainty decrease is largest if errors have a low spatial or temporal correlation. Time series modelling [37] and geostatistics [14] provide methodologies to quantify the spatio-temporal correlations of prediction errors. Szatmári *et al.* [38] used a geostatistical approach (block kriging) to derive the uncertainty of the soil organic carbon stock change over time for Hungary at multiple spatial scales. The study confirmed that uncertainty decreases as the area over which is aggregated increases. At point scale, none of the estimated soil organic carbon changes between 1992 and 2010 were statistically significant, while at the county and country scale they were.

Upscaling to large spatial areas, such as the entire study area, can also be done using a design-based statistical approach [39]. This has the advantage that no model assumptions are needed, but a requirement is that the measurement locations are a probability sample from the area of interest, and that the sample size is sufficiently large for each upscaling area. Some relevant applications of this approach are Singh *et al. [40]* and Karunaratne et al.[41]. See also Section 6 below, where design-based statistical inference is discussed in some more detail from a statistical validation perspective.

6. Statistical validation

The preceding sections tackled the quantification of uncertainty in the outputs of models predicting crucial MRV variables. This is achieved through the application of probability theory and uncertainty propagation analyses. These methods are powerful as they not only assess model output uncertainty but also quantify the contributions of individual uncertainty sources. However, their drawback lies in being 'model-based', entailing the incorporation of various assumptions like stationarity, isotropy, and normality [39,42]. Furthermore, as previously mentioned, estimating the

parameters of probability distributions is often challenging, a difficulty that persists even after adopting simplifying assumptions [43,44].

Contrary to model-based approaches, the 'design-based' approach has the important advantage that it is entirely model-free [39]. It is based on statistical sampling theory and requires random sampling from the population of interest. The simplest example of that is simple random sampling, while more elaborate approaches are stratified random sampling, cluster random sampling, systematic random sampling and model-assisted sampling [39,45]. When probability sampling is conducted both at the start and end of a project, an unbiased estimate of the the SOC stock change over time for an area is obtained with quantified accuracy [45]. Practical implementation of design-based statistical inference requires probability sampling and might entail considerable sample sizes to attain a desired accuracy level (e.g., for attaining adequately narrow confidence intervals ensuring statistically significant estimated SOC stock changes), thereby imposing a substantial resource burden on a project.

7. References

[1] Nol L, Heuvelink GBM, Veldkamp A, et al. Uncertainty propagation analysis of an N_2O emission model at the plot and landscape scale. Geoderma. 2010;159(1-2):9-23.

[2] Heuvelink GBM. Uncertainty analysis in environmental modelling under a change of spatial scale. Nutrient Cycling in Agroecosystems. 1998 1998/03/01;50(1):255-264.

[3] van Leeuwen C, Mulder VL, Batjes NH, et al. Statistical modelling of measurement error in wet chemistry soil data. European Journal of Soil Science. 2022;73:13137.

[4] Grahmann K, Terra JA, Ellerbrock R, et al. Data accuracy and method validation of chemical soil properties in long-term experiments: Standard operating procedures for a non-certified soil laboratory in Latin America. Geoderma Regional. 2022 2022/03/01/;28:e00487.

[5] Wagner G, Desaules A, Muntau H, et al. Harmonisation and quality assurance in pre-analytical steps of soil contamination studies — conclusions and recommendations of the CEEM Soil project. Science of The Total Environment. 2001;264(1–2):103-118.

[6] Fernández-Ugalde O, Jones A, Meuli RG. Comparison of sampling with a spade and gouge auger for topsoil monitoring at the continental scale. European Journal of Soil Science. 2020;71(2):137-150.

[7] Bettingole C, Hanle J, Kane DA, et al. Optimizing Sampling Strategies for Near-Surface Soil Carbon Inventory: One Size Doesn’t Fit All. Soil Systems. 2023;7(1):27.

[8] Holmes KW, Wherrett A, Keating A, et al. Meeting bulk density sampling requirements efficiently to estimate soil carbon stocks. Soil Research. 2011;49(8):680-695.

[9] Ellert BH, Bettany JR. Calculation of organic matter and nutrients stored in soils under contrasting management regimes. Can J Soil Sci. 2002;75:529–538.

[10] Wendt JW, Hauser S. An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. European Journal of Soil Science. 2013;64:58-65.

[11] FAO-GSP. A protocol for measurement, monitoring, reporting and verification of soil organic carbon in agricultural landscapes – GSOC-MRV Protocol. Rome: FAO, ITPS, GSP; 2020. p. 140.

[12] IPCC. 2019 refinement to the 2006 ipcc guidelines for national greenhouse gas inventories: Overview. In: Buendia EC, Guendehou S, Limmeechokchai B, et al., editors.: IPCC; 2019. p. 15.
[13] Stanley P, Spertus J, Chiartas J, et al. Valid inferences about soil carbon in heterogeneous landscapes. Geoderma. 2023 2023/02/01/;430:116323.
[14] Webster R, Oliver A. Geostatistics for Environmental Scientists (2nd ed.). Chichester: Wiley; 2007.
[15] Burt JE, Barber GM, Rigby DL. Elementary Statistics for Geographers (3rd edition). Guilford Press; 2009.
[16] O'Hagan A, Buck CE, Daneshkhah A, et al. Uncertain Judgements: Eliciting Experts' Probabilities. John Wiley & Sons, Ltd; 2006.
[17] Kennedy MC, O'Hagan A. Bayesian calibration of computer models. JR Statist Soc B. 2001;63:425-464.
[18] Xu X, Shi Z, Li D, et al. Soil properties control decomposition of soil organic carbon: Results from data-assimilation analysis. Geoderma. 2016;262:235-242.
[19] Wijmer T, Al Bitar A, Arnaud L, et al. AgriCarbon-EO v1.0.1: large-scale and high-resolution simulation of carbon fluxes by assimilation of Sentinel-2 and Landsat-8 reflectances using a Bayesian approach. Geosci Model Dev. 2024;17(3):997-1021.
[20] Davoudabadi MJ, Pagendam D, Drovandi C, et al. Advanced Bayesian approaches for state-space models with a case study on soil carbon sequestration. Environmental Modelling & Software. 2021 2021/02/01/;136:104919.
[21] Wang H, Prentice IC, Cornwell W, et al. A universal model for carbon dioxide uptake by plants. bioRxiv. 2016 2016-01-01 00:00:00.
[22] Refsgaard JC, van der Sluijs JP, Brown J, et al. A framework for dealing with uncertainty due to model structure error. Advances in Water Resources. 2006 2006/11/01/;29(11):1586-1597.
[23] Liao C, Chen Y, Wang J, et al. Disentangling land model uncertainty via Matrix-based Ensemble Model Inter-comparison Platform (MEMIP). Ecological Processes. 2022 2022/02/08;11(1):14.
[24] Garsia A, Moinet A, Vazquez C, et al. The challenge of selecting an appropriate soil organic carbon simulation model: A comprehensive global review and validation assessment. Glob Chang Biol. 2023 Aug 12.
[25] Janssen PHM, Petersen AC, van der Sluijs JP, et al. A guidance for assessing and communicating uncertainties. Water Science and Technology. 2005;52(6):125-131.
[26] Taylor JR. An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements (2nd Edition). Mill Valley: University Science Books; 1982.
[27] Heuvelink GBM. Error Propagation in Environmental Modelling with GIS. CRC Press; 1998.
[28] Grüneberg E, Ziche D, Wellbrock N. Organic carbon stocks and sequestration rates of forest soils in Germany. Global Change Biology. 2014;20(8):2644-2662.
[29] Magnussen S, Köhl M, Olschofsky K. Error propagation in stock-difference and gain–loss estimates of a forest biomass carbon balance. European Journal of Forest Research. 2014 2014/11/01;133(6):1137-1155.
[30] Lewis PAW, Orav EJ. Simulation Methodology for Statisticians, Operations Analysts, and Engineers (Vol. 1). Pacific Grove, California: Wadsworth and Brooks/Cole; 1989.
[31] Fortin M. Comparison of uncertainty quantification techniques for national greenhouse gas inventories. Mitigation and Adaptation Strategies for Global Change. 2021 2021/03/22;26(2):7.
[32] Jonas M, Bun R, Nahorski Z, et al. Quantifying greenhouse gas emissions. Mitigation and Adaptation Strategies for Global Change. 2019 2019/08/01;24(6):839-852.
[33] Yanai RD, Wayson C, Lee D, et al. Improving uncertainty in forest carbon accounting for REDD+ mitigation efforts. Environmental Research Letters. 2020 2020/11/17;15(12):124002.
[34] Wadoux AMJ-C, Heuvelink GBM. Uncertainty of spatial averages and totals of natural resource maps. Methods in Ecology and Evolution. 2023;14(5):1320-1332.

[35] Plaza C, Zaccone C, Sawicka K, et al. Soil resources and element stocks in drylands to face global issues. Scientific Reports. 2018 2018/09/13;8(1):13788.
[36] Harris NL, Gibbs DA, Baccini A, et al. Global maps of twenty-first century forest carbon fluxes. Nature Climate Change. 2021 2021/03/01;11(3):234-240.
[37] Box GEP, Jenkins GM, Reinsel GC. Time series analysis. John Wiley & Sons; 2008.
[38] Szatmári G, Pásztor L, Heuvelink GBM. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. Geoderma. 2021;403:115356.
[39] De Gruijter JJ, Brus DJ, Bierkens MFP, et al., editors. Sampling for natural resource monitoring. Heidelberg: Springer; 2006.
[40] Singh K, Murphy BW, Marchant BP. Towards cost-effective estimation of soil carbon stocks at the field scale. Soil Research. 2013;50(8):672-684.
[41] Karunaratne SB, Bishop TFA, Odeh IOA, et al. Estimating change in soil organic carbon using legacy data as the baseline: issues, approaches and lessons to learn. Soil Research. 2014;52(4):349-365.
[42] Brus DJ, de Gruijter JJ. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma. 1997 1997/10/01/;80(1):1-44.
[43] Heuvelink GBM. Analysing Uncertainty Propagation in GIS: Why is it not that Simple? Uncertainty in Remote Sensing and GIS 2002. p. 155-165.
[44] Kros J, Heuvelink GBM, Reinds GJ, et al. Uncertainties in model predictions of nitrogen fluxes from agro-ecosystems in Europe. Biogeosciences. 2012;9(11):4573-4588.
[45] Brus J. Spatial sampling with R. New York: Chapman and Hall R/C; 2022.