Methodology for GHG and Co-Benefits in Grazing Systems V1.0

Response to Reviewer 1

Reviewer's Information

Specialization:	Soil ecology, biogeochemistry, rangeland ecology

---

Authors response to reviewer

Please find below a detailed response to each comment. We indicate where in the text and how each comment was addressed.

Reviewer's Comments	Authors Response	Reviewer´s Decision (PASSED/ REQUIRES FURTHER REVIEW with comment)	Author´s final comments
Is the underlying foundation of the methodology clear? It would be helpful up front to explicitly explain whether these methodologies are for generating carbon credits.	This has been clarified at the beginning of the document, Section 1. Methodology Overview, subsection 1.1. Scope. *"(...) to obtain estimates of Soil Organic Carbon (SOC) stocks within a project area, track changes in SOC stock over time to generate Carbon Credits, and measure additional ecological*	passed	*-not required-*

	co-benefits such as animal welfare, ecosystem health, and soil fertility."		
Is the methodology feasible? The methodology seems highly technical, and unless user-friendly tools are developed to support the process I fear that it will be prohibitive for most. E.g., using GIS and remote sensing tools, land cover algorithms, spatial interpolation activities, creation of pedotransfer functions, uncertainty assessment via train-test split approach, accounting for fuel emissions. Who will be using the methodology? It is feasible, yes, but doesn't seem widely scalable in its current form.	We agree with the reviewer about scalability barriers. We are using cutting edge technology and best of science to shape this methodology, which as a drawback constrains its implementation to professionals that have a good GIS and RS background. We are nonetheless doing our best to make it easy to follow without sacrificing the quality. We put a lot of energy into writing this Methodology in a way that is concrete, targeted and avoids any ambiguities. We also generated guidelines and a jupyter notebook to provide an automated version to ease the RS and GIS and statistical analysis part and to lower the time expenditure by the Monitors, to reduce MRV costs. This automation could be further developed to become even more user friendly, although anyone with experience in using GEE or python language should find the current version easy to follow. Our expectation is that given this is an open source methodology, there will be increasing adoption and feedback provided from users that will allow its evolution into better versions with better tools.	Passed - I like the minimum skills section.	*-not required-*

	Past experience with external users: we have handheld two Monitors interested in implementing it, with different levels of RS and GIS technical skills. One has a skilled technical staff and found it relatively easy to implement. They have replicated our results from past vintages and are running new monitoring rounds. The second one has only one GIS and RS technician with a Beginner's level in their staff, and found it more difficult to run the GIS part without our help. For now this version of the Methodology will require the Monitoring staff to have at least someone with an intermediate to high degree of expertise in RS and GIS. We now added the skills required to run it in the beginning of the Methodology doc, in section 1.2.: *1.2. Minimum skills required from the Monitor* *This Methodology requires that the Monitoring Team has the following skills and experience:* ● *soil sampling experience* ● *moderate to strong quantitative spatial analysis and remote sensing skills (e.g. GIS, Google Earth*		

	Engine, geostatistics) ● *some experience in GHG accounting and in environmental monitoring. A professional in the area of agronomics, environmental science, soil science or biology is recommended.*		
Soil Health Co-benefit Requiring participants to generate and use region-specific soil health curves seems unreasonable. Why not just tell them that more is better, less is better, or if there is an optimum and then they can track/assess change over time similar to soil carbon (more is better) in the same way? Alternatively, why not take a scoring approach (needs improvement, fair, good, excellent) for soil carbon as well? Why leave that out? Seems to be a disconnect in approaches for interpretation (change over time for SOC, regional curves for other soil health indicators; comparison with control targets for ecosystem vigor)	The reasoning behind this is that there's no one size fits all recommendation when it comes to nutrients. Different soils under different climates have different nutrient thresholds and optimums. Of course what would be ideal is to have those values already compiled into tables or maps, so that for a given geolocation, there would be an associated list of nutrients with their optimal and minimum values. Because no such a database exists for prescribed grazing management to date that we are aware of, we are recommending users to rely on their local knowledge. Regarding the differences between carbon accountability and soil health, we have done this distinction on purpose, given the changes in SOC and GHG over time are the ones used as input for determining the number of credits issued, so they are assessed quantitatively	Re: region specific soil health curves — I guess my concern is, where are people going to get the data to generate those region-specific curves from in the first place? Re: comparing soil carbon stocks to a database for ideal/optimal levels — have you looked into what (if anything) you could do with something like SoilsRevealed? Partnering w/ folks from ISRIC could help customize it for your purposes.	1) Re: region specific soil health curves: We expanded a bit the text, so now it reads: *Scoring functions should be regionally adapted by the Monitor according to thresholds based on literature or local standards (e.g. scientific papers, local reports, or consultation to soil experts from local universities), which must be cited or attached to the monitoring report.* 2) Re: comparing soil carbon stocks to a database for ideal/optimal levels:

	following a rigorous method. Whereas changes in co benefits are assessed qualitatively and work as a complementary source of information that adds value to the credits. But I think the reviewer has a point in that there´s some disconnect. We have been wondering how to factor in the status of carbon stocks for a given point in time, in comparison to the best-case scenario, for a particular project area. We don't have a database for SOC ideal or optimal levels for the different ecoregions that we can direct our users to, yet. But we are working on some research projects like Regen Lands to get those values for grazing for the different configurations of soil types, climates and topography that exist. We think it will be valuable once we have that information, to factor in and be able to rank the current status of each specific project regarding the optimal status of SOC. It would also help to understand how far is the soil from a potential plateau, and help also generate payments for "maintenance". Today, the credits only pay for recovery of carbon but not so much for maintenance and there are producers that have been doing regenerative management for decades that might be closer to the plateau and thus wont see a significant increment in carbon (which leaves them out of the market) We believe it's not fair that		Yes, we have been following those types of mapping efforts. Unfortunately for instance Soils Revealed uses an empiric estimation of potentials based on the IPCC tier 1 equations, so does not consider regional soils types and climate. We are looking forward to developing or joining efforts to help develop something that's more granular for regenerative management so we can really get a sense of the optimal levels for each area and understand where we are at now and which type of management practices can lead to that optimal scenario. *I will bring this up during the Public comment discussion !*

	those landstewards are left out of finance opportunities. We'd love to discuss if the reviewer has thoughts and ideas around this.		
Are the sampling and measurement protocols robust? Requiring a minimum of four monitoring rounds for a 10-year crediting terms seems like a lot. A minimum of 10 months and maximum of 4 year resampling periods is too frequent and will require a cumbersome number of samples and effort to capture meaningful change in most rangeland environments. I would suggest requiring just three sampling dates. One baseline, one mid, and one end.	This requirement has been changed according to the reviewers suggestion, in section 2.2. "The maximum time between sampling monitoring rounds is five (5) years.". Nonetheless, if project developers want to do more frequent samplings they are allowed.	passed	*-not required-*
When trying to capture change over time, sampling density estimates also require information on the expected effect size (or MDD), which can be affected by climate, management practice, or duration between sampling events. Power analyses for detecting change over time also often use the SD of the difference/change. This does not seem to be incorporated into the sample density calculations, and I'm wondering how the authors account for this.	In section 3.1.1. we added this clarification: "Nonetheless, more comprehensive approaches based on previous sampling and/or GIS analysis are preferred if feasible, in which case a brief description along with a valid scientific reference to the specific approach to estimate the sample size should be cited in the report. "		
Why only require additional metrics be collected on 30% of the samples?	Because soil health co-benefit is less rigorous than SOC stocks accounting, and so it can still be assessed from only 30%	passed	*-not required-*

	of the (random) samples to get a good qualitative result. And by doing so, the laboratory costs for analysis are dramatically reduced, to about ⅓ per sample.		
I am glad to see stratification is not required. Stratification only increases accuracy if done properly, and this step can be highly technical in terms of discerning which variables are important for a given landscape, using appropriate methodologies to define boundaries using those variables, and stratifying accordingly.	n/a required	n/a required	*-not required-*
I expect the methodology will get a lot of critique for having a minimum sampling depth of 10 cm. I personally am not opposed to including shallower depths for soil health purposes, but if these protocols are for carbon credits, I believe that will be seen as a big weakness and recommend the authors consider mandating 30 cm.	This has been addressed and we now mandate 30cm min depth. Please see section 3.1.4., point 2)"The minimum sampling depth is 30 cm. Justification must be provided if sample depth is shallower than 30 cm10-30cm." The other reviewers made the same comment and also we believe this has become the new standard and there´s consensus in the VCMs that 30cm minimum depth is adequate for SOC.	passed	*-not required-*
Why require samples be taken 10 meters away from any tree, shrub, or body of water? If a landscape has trees (e.g., in a oak woodland) that will influence soil carbon and shouldn't that be captured? Or, if a project plants trees or restores their riparian area (actively or passively) as part of their landscape-level approach for managing soil	It would be good to include that into this methodology, but it would require a whole development for AGB carbon accounting or wetlands carbon accounting.We hope that some hybrid methodologies will emerge in our Registry in the near future, although we are not there yet.	Hm, I can see the methodological challenges associated with this. Could you state that this is a soil carbon protocol, which would get around the need to do AGB carbon accounting? It still seems like	

carbon, shouldn't that also be captured?	It would be good to include that into this methodology, but it would require a whole development for AGB carbon accounting. Given how we are coupling RS with samples, in order to be conservative and reduce the error and only account for the additional soil carbon sequestered from the implementation of the eligible practices, we are excluding all pixels with trees or any other artifacts like rocks, roads, man made objects or water that would add error to the satellite calibration. Only the pixels with grassland vegetation cover, where the practices are implemented, are the ones that can be calibrated from regressions or ML using this method. So for the final accountability of SOC stocks, the only surface that's accounted for is the net grasslands area, without these trees and their shadows where we can´t model the pixel values with good accuracy. We hope that some hybrid methodologies will emerge in our Registry in the near future, so that more complete assessments of soil and above ground carbon changes can be run.	one of the bigger limitations of your approach (b/c you will be underestimating soil C stocks for places w/ woody plants present; and missing possibilities for soil C accrual w/ natural revegetation through grazing management. And what if trees establish in points that used to be open grassland?) I just worry it will be met with additional criticism from others.	
It seems funny to me that the sampling depth be 10 cm below the depth of profile alteration. I understand not wanting to analyze, for instance, biochar itself if incorporated into the soil, but	agreed! so we deleted that point! See section 3.1.4.	passed	*-not required-*

wouldn't losses due to tilling etc want to be captured in the area of the soil profile that was directly affected?			
Even default laboratory measurements of simple bulk density require an accurate measurement of volume provided by the participant. This is not something the laboratory can generate, as suggested by the protocol on page 16.	This was corrected and the volume calculation is now more clearly requested to the participant (section 3.2.1.1. BULK DENSITY.)	passed	*-not required-*
I think the protocol will get critique for not requiring rock or gravel content assessment on all bulk density samples.	The main reason for reducing the requirement for gravel content assessment is to reduce sampling efforts to the extent possible. We reviewed the current literature more in depth to adjust the threshold. We concluded, based on our findings, that we should require that rock or gravel content be assessed "if soil has >10 % gravel or the stones are >2 cm …" Some extracts and citations justifying this below: - *Soil organic C concentration can have a greater effect on the SOC stock variance than bulk density and rock fragment volume in non-stony soils (Goidts et al., 2009, Schrumpf et al., 2011). However, Schrumpf et al. (2011) reported that rock fragment volume fraction is more important than SOC concentration to SOC stock variance in soils*	passed	*-not required-*

	with rock fragment concentrations greater than 20%. - In https://www.sciencedirect.com/science/article/abs/pii/S0016706120325921 - *If soil has >10 % gravel or the stones are >2 cm conventional bulk density readings will be inaccurate, as most coarse fragments have bulk densities of 2.2–3.0 g/cm3 (McKenzie et al., 2002).* Coughlan, K., Cresswell, H., & McKenzie, N. (2002). *Soil Physical Measurement and Interpretation for Land Evaluation* . CSIRO PUBLISHING. Retrieved from: https://www.perlego.com/book/1468472/soil-physical-measurement-and-interpretation-for-land-evaluation-pdf (Original work published 2002)-also cited in: https://www.soilquality.org.au/factsheets/bulk-density-measurement We corrected the previous requirement and added the citation in section 3.2.1.2. GRAVEL CONTENT		

I recommend the authors encourage stocks to be calculated using equivalent soil mass rather than a fixed depth approach.	We are now expanding on the option to carry out a ESM approach as an alternative to using the fixed depth approach in section 3.2. We are still unsure that this would always be a better alternative to the FD, given that in this case we are requiring bulk density to be measured always, at each sample and at each point in time. We think that the ESM might be complicated to be implemented correctly, and if samples are required to be split into sublayers, the costs for the sample analysis will be at least double if only split in two (0-15 and 15-30). A couple years ago, we co-authored along with other scientists a report for the ESMC where, after reviewing the literature, we concluded the following: ***Further research is needed in order to define the criteria to choose between either of the methods.Nevertheless, given the simplicity and smaller cost of the FD method, it may be preferred under the following conditions:- Management practices or changes in land use in the area of interest are not expected to create drastic changes in BD;- When sampling to at least 30 cm depths;- If treatments are being compared and BD is consistent between both.*** So, it seems that maybe for 30cm depths and for grazing systems the FD would be adequate, as conservatively we	I think your rationale makes sense and that adding in an option to do ESM is good; you might consider providing some of the rational in text somehow (or supp info) because I think it is one decision point that (like you mention) soil folks will have strong opinions on.	

	are requiring all samples to be measured for BD. might the ESM be better for deeper soil layers? Should we make any recommendations there, to guide our project developers to make the right choice for their projects? We would love to learn from the reviewer´s experience and have a conversation to better understand the advantages and feasibility of implementation. During this review process we also had some conversations with soil experts about this topic, who expressed contradictory opinions.		
It is unclear whether other emissions (livestock, fuel) are accounted for at the beginning and end of the project as well (i.e., the change assessed). It seems like that is not the case, and would that not be a strong limitation to this accounting approach?	GHG emissions are required to be accounted for on a year basis and included in the reports and the calculations for each crediting period. We made it more explicit in the text in section 3.5. and its subsections. E.g.: *GHG emissions must be accounted for on an annual basis and included in the net CO2e reduction calculation (see section 3.6.3.) of every monitoring report for its corresponding period.*	passed	*-not required-*
Are there any alternative or additional steps that should be considered? I would recommend including language requiring participants to use the same analytical/service laboratory over time.	We agree with the reviewer. We now added a note in section 3.2.1. *<u>Note</u>: It is highly recommended that the same analytical procedures and service laboratory are used over time, to reduce additional errors from calibration bias or*	Could you take it a step further and require it? If so, I'd highly recommend it. I'm working on a couple of manuscripts comparing analytical labs w/ colleagues and there are some labs that consistently	I agree with this so much but I'm worried what happens if the lab closes, or changes the lab analysis they perform or something like that? I guess a nice solution is to require it and also allow for exceptions

	changes in the analytical techniques.	produce higher results than others so the easiest way for someone to "gain carbon" would be to send to Lab A in T0 and Lab B in T1. Big risk there.	under a good justification like these.... We now replaced the recommendation with a requirement, that reads as follows: *<u>Note</u>: It is required that the same analytical procedures and service laboratory are used across all sampling events, to reduce additional errors from calibration bias or changes in the analytical techniques. Exceptions are allowed under strong justification: i.e. the laboratory shuts down, or stops providing the service.*
It'd be helpful to provide guidance on the actual methodologies for assessing the additional soil health indicators. E.g., there are different phosphorus extractions that are more-or-less suitable for specific environments.	We find it hard to be too prescriptive in this regard, as our own experience with having some projects running in Australia, others in the US and others in Argentina has proven to us that there are different constraints in different locations that have to do with the local analytical procedures adoption, costs of analysis, access to some reactives that might be due to certain circumstances lacking (imported) or too expensive, etc. We believe that it is important that the best analytical procedures locally available and suitable for the specific environment and soil types assessed are chosen by the local experts rather than the	I agree, makes sense - might be nice to include some of this language in the handbook just to clarify for your readers. Unless I missed it, currently I believe there is no mention of how methods should be chosen in text.	We now added a short paragraph about this in section 3.2.1. What to measure? as follows: *"In terms of which analytical procedures should be followed for each parameter, we will defer to the local experts to choose the best options within the local possibilities. We recommend that standard and well*

	Methodology imposing. But this could be a nice topic for discussion during the public comment period, if the reviewer has a strong opinion here.		*accepted procedures are followed in order to avoid avoidable errors and potential negative criticism about the credits.“*
Provide information on how to handle/treat samples prior to sending to the lab (air dry, etc) – for all analyses.	We now made it clear in the text that they should follow the local lab recommendations for sample collection and handling (considering the Methodology requirements for min. depth, parameters, etc) and they should describe how samples were handled and cite the source of recommendations or standards they followed.	passed	*-not required-*
Additional comments The authors allude to, and put heavy emphasis on, the fact that sample sizes are lower when incorporating remote sensing and spatial interpolation, but do not cite any work on this or provide robust description of why and by how much. I suggest bolstering this.	This is a good observation. Given the innovative nature of this methodology, unfortunately there are no papers yet comparing both sample size requirements that we could reference. We know from previous attempts of the people in Australia to submit their projects to the Australian ERF, that the request to achieve a reasonable error was much higher for their sites (10x they said). There's also a recent publication from Bettigole et al (2023) estimating a required soil sampling intensity of between 0.25 to 2.2 samples /ha to be able to track changes in SOC at 90% confidence within a 5% margin of error, using traditional sampling and analysis coupled with stratification or cLHS to locate the samples and reduce sampling effort. This would mean that for a 1,000 ha	passed	*-not required-*

	rangeland, the number of samples would vary between 250 and 2,200, whereas our calculator estimates between 88-123 samples for the same project size, depending on the site variability. And the difference is even higher for larger properties, as through our approach the number of samples increases non linearly to calibrate satellites or interpolate with a reasonable error. For example, according to Bettigole et al. for a 2,000 ha property between 500 and 4,400 samples should be required (which is extremely unfeasible), whereas our calculator estimates between 102 and 142 samples per sampling round. Nonetheless, we recognize that we have not yet tested the calculator in sufficient sites to be able to scientifically prove that the estimated amount of samples that we are requesting is the optimal to achieve a good performance. We are inferring this from the available literature from interpolation use cases and the use cases from our native projects that have been running our method, and we will continue adjusting as more data comes in. We added some context and relevant references that you will find in a new paragraph at the beginning of section 3.3.1. but those are more related to the GIS approach rather than the sample size comparisons. We appreciate this observation and we think it will be good to		

	publish an analysis in this regard once we have more projects with data to compare performances with the traditional ones.		