CLIMATE ACTION RESERVE

SUMMARY OF COMMENTS & RESPONSES
SECOND PUBLIC COMMENT PERIOD
DRAFT SOIL ENRICHMENT PROTOCOL VERSION 1.0

Fourteen sets of comments were received during the second public comment period for the Climate Action Reserve (Reserve) draft Soil Enrichment Protocol Version 1.0. Staff from the Reserve provides responses to the comments below. The second public comment period for the draft protocol was August 11 to August 25, 2020.

The comment letters can be viewed on Reserve's website at http://www.climateactionreserve.org/how/protocols/soil-enrichment/.

COMMENTS RECEIVED BY:

1. Grayson Badgley (**Badgley**)
2. Grayson Badgley, Jeremy Freeman, Eric William Slessarev, Jennifer Pett-Ridge, Jane Zelikova (**Badgley et al.**)
3. Mike Badzmierowski (**Badzmierowski**)
4. California Association of Sanitation Agencies (**CASA**)
5. CarbonPlan (**CarbonPlan**)
6. Environmental Defense Fund (**EDF**)
7. Kaiyu Guan, Zhenong Jin, Evan DeLucia, Bin Peng, Jinyun Tang, Andrew Margenot, Wang Zhou, Ziqi Qin, DoKyoung Lee, and Yuxin Wu (**Guan et al.**)
8. Indigo Ag, Inc. (**Indigo Ag**)
9. Natural Capital Partners (**NCP**)
10. Ken Newcombe, Deane Belfield, Mark Bradford, Karen Haugen-Kozyra, Louisa Kiely, Paul Luu, Sean McMahon, Steve Pacala, Keith Paustian, Sean Penrith, Diego Saez-Gil, Tom Stoddard, Nathan Truitt, Moritz von Unger, Matthew Warnken, Raphael Wood (**Newcombe et al.**)
11. Jonathan Sanderman, Woodwell Climate Research Center (**Sanderman WCRC**)
12. Scott Shibata (**Shibata**)
13. TreePeople (**TreePeople**)
14. Matthew Wiens (**Wiens**)

General Comments

1. I am a little discouraged by the lack of inclusion for preservation of SOC as it is harder to build SOC than it is to preserve it. This is particularly true for the middle of the United States where their topsoil contains much greater levels of SOC than in other parts of the United States. Increasing carbon levels in these sites can be more difficult to achieve but preserving this carbon from being lost via destructive management should be valued greatly. (**Badzmierowski**)

   RESPONSE: Thank you for your comments. The protocol does include preservation of SOC, in several ways, but any crediting for potential preservation activities can only occur during the project crediting period(s). The protocol compares baseline SOC levels with project SOC levels. Projects could then get credited for reducing rates of decline or for increasing SOC, both of which amount to preservation of existing SOC. We do allow for some crediting of avoided SOC emissions under our Grassland Protocol and Forest Protocol (for the avoided conversion project type), as the appropriate modeling of such emissions has been performed for those landscape types, which allowed for their inclusion.

2. The Reserve staff and the working group have done a tremendous job pulling together this groundbreaking offset project protocol. As a project developer, Indigo Ag would not be able to generate and verify high-quality offsets for agricultural projects without the presence of rigorous protocols. The Reserve's protocol development process provides incredible transparency and stakeholder interaction, ultimately raising the quality of the credits we are able to generate and bringing confidence to our farmers, partners, and buyers. We have appreciated the opportunity to support the Reserve and to participate on the SEP working group. The staff should be commended for assembling an excellent group of experts to help craft this final document. We believe there is opportunity for GHG emission reductions and increased soil carbon sequestration on millions of acres across the US, and worldwide, and that this opportunity is now being unlocked through the adoption of this protocol. Indigo Ag fully supports the Reserve Board in approving this protocol for use. (**Indigo Ag**)

   RESPONSE: Thank you for your comments, and continued support in development of this protocol.

3. I am frustrated by what seems like a pattern of avoiding meaningful engagement on substantive issues affecting the efficacy of the SEP. During the initial drafting phase, I received my first copy of the protocol text roughly 72 hours before the first public comment period, which prevented me from providing meaningful comment. When I raised concerns through the public comment process, I was portrayed as having been unengaged. Since then, I have been quick to speak up, only to find my concerns frequently ignored. After what felt like a rushed fourth working group meeting (in which serious topics, like sampling depth, were hardly discussed), I reached out several times, citing literature by Angers et al., 2008, Luo et al., 2010, and Powlson et al., 2014, all of which point to the high likelihood of SOC overestimation if using a fixed sampling depth of 30 cm. (**Badgley**)

   RESPONSE: Thank you for your comments and multiple follow up discussions. With respect to the amount of time given to workgroup members to review the first full draft of the protocol before the first public comment period, we did send the first full draft of the protocol out to all members of the workgroup on March 30th; the protocol was published online alongside the recording of that third workgroup meeting, on April 3rd; we then directly sent you another copy of protocol on April 15th by request; finally, the public comment period commenced on April 17th – 3 working weeks after we first made the full draft protocol available to the workgroup. While at the outset of undertaking this protocol development process we did not anticipate a fourth workgroup meeting, we did find it necessary to add an additional workgroup meeting in advance of the second public comment period, in order to facilitate additional workgroup engagement. That was a three-hour meeting, during which time there was almost constant discussion. Our workgroups are always voluntary, as we recognize that having a very heavy level of engagement for multiple hours can be an onerous task for most workgroup volunteers. However, we do our best to ensure all workgroup members are heard. We regret that you did not feel that your comments were adequately addressed and appreciate that you have continued to follow up outside the workgroup meetings. Our continued verbal and email deliberations with you have resulted in numerous changes to the draft protocol. Some key examples of changes made in response to your feedback include: where you advocated that a minimum sample depth be set at 15-20 cm below historical plow depth, to which we responded with changes that ensure samples are 10-20 cm below historical plow depth; and you advocated we move away from referring simply to categories of historical deep plowing, and instead focus on actual data on historical plow depth, which we adopted. We thank you again for all of your valuable contributions.

4. CASA appreciates CAR's responsiveness to our first round of comments and for reaching out to discuss the issues raised. We specifically support the addition of biosolids in the protocol in recognition of its value as an organic fertilizer/soil amendment that results in the reduction of GHG emissions and sequestration of carbon in the soil below. Biosolids are produced as an inherent part of the wastewater treatment process and can be land applied in agricultural settings helping to mitigate climate change by substituting for the use of fossil fuel intense inorganic fertilizer and by improving long-term sequestration of carbon in soil. Roughly 0.22 gallons of fossil fuel is required to produce every pound of inorganic nitrogen fertilizer, illustrating the tremendous offset gained by using biosolids for land application. Because biosolids are an organic matrix, rich in organic carbon and nitrogen as well as other valuable micro and macro nutrients, biosolids improve soil tilth, reduce the need for irrigation by increasing the soil's water holding capacity, and increase crop production. (**CASA**)

   RESPONSE: We thank you for your feedback during both rounds of public comments and appreciate your participation in this process.

5. As we noted in our first set of comments (May 18th, 2020) selling GHG credits in an agricultural offset program requires a high level of certainty and a robust risk management accounting system, as offset sales result in GHG emissions to the atmosphere elsewhere and many land-use-based credits can be reversible. Unfortunately, in our opinion, the revisions in the second draft of the SEP (August 2020) do not improve the ability of the SEP to meet this requirement. As we stated in our first set of comments, it is our position that achieving a satisfactory level of certainty and risk management capabilities cannot be achieved at the project-scale but rather requires a regionally coordinated landscape-scale program structure. Regional verification will be required to ensure that there are real net changes in carbon stocks and reductions in other GHG emissions. The complexity of the project-level framework proposed in the SEP could lead to perverse outcomes. This potential reduces our confidence that the projects will reflect a net GHG reduction in the

atmosphere when all the actions taken in a region are considered. Furthermore, current models (such as those named in the August 2020 draft, DNDC and COMET-Farm) do not have sufficient resolution and lack sufficient validation to accurately predict SOC and N_2O dynamics at the field level and may only achieve sufficient certainty when applied over large numbers of acres[1]. Using such models for regions and cropping systems not yet calibrated and validated further reduces confidence of real GHG emission reductions and introduces the chance of other unintended consequences. Therefore, models used must provide proof of validity with guidance as to acceptable uncertainty provided within the protocol. (**EDF**)

RESPONSE: Thank you for your comments. We believe projects under this protocol will be employed at the broad landscape scale. Following your earlier feedback, we did include guidance to the effect that larger-scale projects will likely reduce uncertainty. Note that the guidance enshrined in the Model Validation, Calibration and Verification Guidance for Soil Enrichment Projects ensures that models are not used if insufficiently validated. A validation report will need to be made public and will need to be approved by an independent model expert. Furthermore, while models are an important component of quantification, they must be applied in concert with ground-based SOC measurements that are used for ongoing ground-truthing and calibration of quantification results, which in turn provides a safeguard against over-crediting.

6. TreePeople applauds Climate Action Reserve's leadership in developing the Soil Enrichment Protocol 1.0: Reducing emissions and enhancing soil carbon sequestration on agricultural lands. We believe that soil is a nature-based solution for carbon sequestration and climate resilience. In particular, globally, about 95% of our food comes from soils (FAO, 2015); and soils store more carbon than the atmosphere and terrestrial vegetation combined (FAO, 2017). At the same time, we strongly recommend that urban soils should be considered and added in the future protocol development given more than 80% of Americans live in urban areas. (**TreePeople**)

   RESPONSE: Thank you for your comments. The main barriers facing inclusion of urban soils within this protocol are a likely lack of sufficient data to run necessary modelling, and also the likely feasibility challenges associated with including many disparate and relatively much smaller areas. We will consider such inclusion further in future versions of the protocol, or in a future update to the Urban Forest Management Protocol.

7. Although we appreciate that the Reserve has clearly disclosed that its protocol development process is sponsored by Indigo Ag, which intends to earn third-party certified credits under the protocol's final version, we have struggled to get a clear answer from the Reserve about the role Indigo Ag has played in drafting the protocol text and guiding the protocol development process. We ask again because the August draft of the protocol indicates that Indigo Ag has played an even bigger role than the Reserve has so far acknowledged, in apparent contradiction of statements the Reserve made to us earlier this summer. [T]he August draft protocol now recognizes "a financial contribution, research and drafting support" from Indigo Ag and acknowledges eight additional employees of Indigo Ag and two employees from the consultancy TerraCarbon LLC. This disclosure is critically important to the integrity of the protocol development process because the workgroup forms the entire basis of soil science expertise for the protocol, as the Reserve does not itself have any in-house soil science expertise. (**CarbonPlan**)

[1] Chenu et al 2019 https://www.sciencedirect.com/science/article/pii/S0167198718303738?via%3Dihub

RESPONSE: Thank you for your comments. Many offset programs allow proponents who intend to use protocols to themselves develop complete protocols and submit them to the offset program for consideration. Our process is relatively unique in that we do not typically follow such a process, and instead we take on a much more hands-on role in developing protocols ourselves. Nonetheless, as mentioned previously, we draw as heavily as we can on the good work from our peers that has come before us. In this instance, we have increasingly drawn insights from a parallel protocol development effort that Indigo Ag and TerraCarbon LLC are running through Verra/VCS, as well as drawing insights from many other protocols from across the world. We have mentioned this parallel effort during workgroup meetings and have had multiple open discussions with the Verra team about our respective efforts. Given we are now drawing insights/draft text from that specific parallel effort, and given that Indigo Ag has provided drafting support around elements related to soil science, model expertise, farmer behavior, and agronomy, in response to workgroup discussion and public comments, we have updated our attributions page accordingly. These interactions have comprised only a fraction of our outreach to potential contributors, including other workgroup members. Throughout the process the Reserve has had editorial control over the protocol. Our process has always been transparent, and we have answered all concerns and questions put to us transparently. We will continue to work closely with industry partners and will continue to be transparent about such interactions and our broader process, in line with industry best practices.

2.2.1 Defining the Project Activities

8. "Project activities must not decrease carbon stocks in woody perennials on the project area." This has potential to limit management practices aimed at improved grazing and increased productivity of grasslands in areas where encroachment by aspen and other woody species reduces carrying capacity. There is evidence that encroachment of woody species into certain grassland ecosystems (e.g. tall grass prairie) may negatively impact carbon sequestration on pastures. Pinno and Wilson (2011) studied the impact of woody encroachment of grasslands on sandy soil near Regina, Saskatchewan. They found that total ecosystem carbon (above + below ground) was higher in forest ecosystems than in shrubland or grassland; however, soil carbon alone was significantly greater in shrubland or grassland than in forest. Another study, done in the southwestern USA (Jackson et al., 2002) found that dry locations gained ecosystem carbon with woody encroachment, but the wetter tallgrass prairie site lost ecosystem carbon as a result of encroachment. Therefore, allowance should be made in the protocol for control of woody species encroachment into grasslands if it can be shown that total carbon sequestration will increase as a result. (**Wiens**)

   RESPONSE: Thank you for your comments. The intent with the guidance here was to prevent large-scale land-use change, in particular deforestation. Your feedback is apt, and we have updated the guidance further to that effect.

2.4 Non-GHG Impacts of Project Activities

9. The protocol states in section 2.4: "The Reserve requires project developers to demonstrate that their GHG projects will not undermine progress on other environmental issues such as air and water quality, endangered species and natural resource protection, and environmental justice." I understand this is merely "guidance", but I think this does not illustrate the complexity of soils. Increasing SOC in agricultural soils comes with environmental trade-offs, not just positives. For example, just looking at water quality, 1.

Reducing tillage and disturbance of soil should improve soil structure and aggregates, but conversely can lead to increased vertical preferential flow, meaning it will leach elements and contaminants and reduce water quality. 2. Cover crop increases soil organic carbon and may increase soil hydrophobicity in dry summer, which may favor the development of finger flow. 3. The greater amount of organic substances also promotes soil fauna activities, likely producing more macropores and enhancing the abundance of macropore flow. If farmers take up practices to improve SOC but still apply pesticides it could compromise water quality, therefore making the farmer ineligible to receive credits. I think there needs to be clear goals and understand that trade-offs should be expected. Is the goal to increase SOC? Is the goal to increase mineral associated organic matter (the soil carbon fraction thought to last longer in soils (though this once again is not always true)? Should the goal be to preserve the most soil carbon (i.e., preserve soils with high SOC stocks in the central U.S.)? I just feel like this protocol does not include the full picture when it comes to the complexity of soils. Overall, this protocol is much needed to move soil carbon policy forward and I am glad to see this initiative. (**Badzmierowski**)

RESPONSE: Thank you for your comments. For project-level GHG accounting we need to focus primarily on GHG impacts. This general guidance encourages projects to consider these broader issues as best they can, but the focus remains on GHG impacts. With respect to potential water quality impacts from tillage: we are encouraging more efficient nutrient usage, as well as encouraging reduction in exposed soil, which should reduce both nutrient and sediment runoff. With respect to potential impacts of using pesticides, projects must comply with all laws. If laws are in place to address pesticide usage, and if projects contravene such laws, in those circumstances the field in question would not be eligible to generate credits whilst such contravention of laws remain. Again, we focus on changes caused by the project itself, and it is not foreseeable that project activities will themselves cause increases in pesticide usage. With respect to the goals the protocol is intended to address, the goal of the protocol is to realize GHG benefits, through encouraging movement to more sustainable farming systems. The protocol requires the use of complex modelling, which should capture most biogeochemical impacts, in line with best practice. While we do anticipate and encourage positive impacts related to these complementary policy goals, and we hope projects will minimize negative impacts, the protocol must remain primarily focused on promoting positive GHG impacts.

3.3 Project Crediting Period

10. We support the replacement of a 30 year crediting period with 10 year crediting periods which may be extended for two additional 10 year periods. (**CASA**)

    RESPONSE: Thank you for your feedback.

3.4.1 Performance Standard Test (Additionality)

11. I was encouraged to see the adoption of a more rigorous additionality standard that explicitly excludes widely adopted tillage and cropping practices. It is also nice to see Negative List will be up- dated through time. I was less encouraged by section 3.4.1.2, “PST – Project-Specific Means to Demonstrate Additionality.” This section largely invalidates the optimism I had about the new additionality standard. According to this section, “stacking” a second practice change would automatically render all practices, including those excluded on the Negative List, as additional. This raises serious concerns given the expansive definition of “project activities” (Section 2.2.1) adopted by the protocol.

The protocol already heavily relies on modeling. As such, we should use modeling to isolate the contribution of non-additional from additional practices, as opposed to ignoring the additionality standard if more than one project activity is adopted. From our previous working group discussions, I am under the impression that CAR is expecting "stacking" to be a common practice, which further amplifies my concern. If stacking is common, the new additionality standard would effectively be rendered useless.

I have further concerns about section 3.4.1.2, subsection 2 ("Demonstrating new tillage practices are still rotated with conventional tillage"). This section states that if a project can point to **three** fields that still incorporate CT with new tillage practices, the new tillage practice is considered additional. Apart from three being an incredibly small number, the mere existence of Section 3.4.1.2, sub-section 2 raises concerns our ability to evaluate projects as additional. I am familiar with the literature this sub-section is referring to — it is well documented that many farmers, after adopting alternative or no-till, will occasionally revert back to conventional tillage. This pattern raises a simple but tricky question: "what fraction of farmers in region X practice NT agriculture?" As written, 3.4.1.2, sub-section 2, argues that as long as even three fields still alternate between CT and NT, NT should be considered additional. I would argue that if determining additionality is so difficult, do we really have any business issuing credits for these types of projects in the first place? (**Badgley**)

RESPONSE: Thank you for your multiple comments and discussions. The concept of additionality is often addressed through complex approaches taking into account not just uptake rates for a given practice, but also the socio-economic barriers to adoption. We discuss such barriers extensively in Appendix A.

Regarding stacking multiple practices as a means to demonstrate additionality: Each time we develop a protocol, we try to carefully define eligible practices. We then assess additionality for each specific practice, as best we can. We are not aware of any protocol or program that broadly categorizes together multiple eligible practices into an additionality screen for the group of practices. There are very distinct agronomic considerations that a farmer must take into account when considering a move from conventional tillage to any of the very different conservation tillage practices. We have advice from multiple agronomists, and prospective project developers, and our own extensive collective experience working in agricultural climate change policy, that tells us that it can be very hard to convince a farmer to change a single practice, and much harder to convince them to change several practices at once. It is worth noting that there is a paucity of USDA data demonstrating simultaneous adoption of multiple such practices. We have also been asked to consider the suite of potential eligible practices, taken as a whole, and to assess how many farmers are doing all of them together. If we were to do that, then uptake rates of the suite would be virtually non-existent. Again, bundling practices together in that form is not widespread amongst offset protocols or programs.

Regarding the notion that stacking is common practice: To be clear, any practice that is existing at a project start date would be accounted for in the baseline, so stacking would not make any existing practices eligible. We were pushed hard to relax this requirement, but we will not. If a project adopts a practice on the negative list that is new for their field, they must also then adopt a further new practice, in order for either/both to be eligible (or use the exception in Section 3.4.1.2(2), which we discuss in the next paragraph below). We do expect that stacking will be necessary for the development of a feasible project under this protocol, and thus may become common for projects. However, we did not assert that we think stacking is common practice insofar as that term is used in the context of additionality. We have found no evidence to suggest that stacking is currently common practice. Our program assesses a performance standard at the start of a project at a single point in time – as is best practice for many reputable offset programs, and then such assessments can be repeated if a project applies for a second crediting period. If lots of fields within a project start to stack practices, in conditions where stacking was not common practice before the project, that does not make those new practices non-additional. It should be noted however, that each time a *new field* is added to the project, it must be demonstrated that the new field meets the additionality requirements of the protocol at the time the new field started its first eligible activity. Such fields could potentially be screened out as non-additional, if we are able to assess that the practice in question has already become common practice in that given county.

Regarding Section 3.4.1.2(2), the intent there was that projects must provide evidence that demonstrates that the majority of fields in the county in which the subject field is located have either implemented conventional tillage, or a rotation of no-till or reduced-tillage with conventional tillage, during the historical baseline period, as well as identifying 3 specific fields which demonstrate this trend. We have updated the guidance accordingly to make this clearer.

12. [In discussions with CarbonPlan] you indicated you would consider a common practice evaluation to screen out management techniques that are already in regular use. The updated protocol implements a common practice evaluation. Unfortunately, this approach is easily circumvented by the practices you indicated you expect most projects to use going forward. Under the protocol's common practice evaluation, the Reserve will draft a "negative list" of practices that are excluded from eligibility to earn credits (Section 3.4.1.1). If a practice is already employed by at least 50% of the applicable cropland or pasture land in an individual county, then it will be placed on the negative list for that specific county and made ineligible for crediting in that same county — reflecting the logic that commonly implemented practices are not additional. We note that the Reserve provides no justification for its selection of a county-specific 50% threshold and has also not disclosed what the initial list (the "Additionality Tool") would look like.

    We question the sincerity of an additionality screening tool that was apparently not ready to be shared publicly with this latest protocol revision. Although the current draft describes a process for producing the negative list, neither the protocol nor the Reserve website provides the list. Without a list, it was not possible to review the quality of the common practice assessment from primary data within the two-week window provided for public comment. With a list in hand, we would want to evaluate whether there are any practices that, for example, are commonplace in many applications but not quite at the 50% threshold in some locations — a type of evaluation is required to provide confidence in a common practice evaluation, [2] but not included in the protocol's technical appendix.

    The protocol's opacity on common practice may not matter in application because the concept of the "negative list" is unlikely to be used by large projects as a result of a major and explicit loophole. The protocol proposes to render eligible any practice on the negative list when it is combined with any other practice — including another practice on the negative list (Section 3.4.1.2). The protocol also enables projects to point to as few as three other fields in a county to justify an exception for a practice that is otherwise on that county's negative list. These loopholes are so big as to overwhelm the common practice assessment for any project that "stacks" practice changes — as you and Mr. Ebert specifically indicated to us you expect to be common among credited projects under the final protocol.

We do not believe that this additionality standard is credible and therefore we recommend that the Reserve acknowledge instead that the protocol is not testing for additionality. As we have discussed previously, protocols that do not screen for additionality may have important applications in supporting public and private expenditure programs — but they are critically flawed in the context of carbon offsets meant to justify a buyer's greenhouse gas emissions. (**CarbonPlan**)

RESPONSE: Thank you for your comments. Please see our response to comment 4 above, as it addresses the issues you raise in your comment. With respect to SEP Additionality Tool, please note it was an oversight on our part to not release that during the public comment period. That tool was sent out to the workgroup prior to the public comment period, as well as to several members of the public that requested a copy. Please note we have also now uploaded a copy of that tool on the Soil Enrichment Protocol page on our website.

3.4.1.1 PST – Negative List

13. Natural Capital Partners advocates a lower common practice threshold, consistent with the Reserve's other protocols and with norms across offsetting practice. In the Reserve's Discussion Paper, Standardized GHG Accounting for Soil Organic Carbon Accrual on Non-Forest Lands: Challenges and Opportunities, dated September 23, 2019, the Reserve states the following:

    Under a common practice approach, if a practice is undertaken by no more [than] a certain percentage of farmers in a particular region (commonly 5%) it can be treated as additional. A balance typically needs to be made between excluding some early adopters (often seen as unjustly penalizing critical early movers), as they may have implemented the activity without an offset motive, and rewarding laggards (those who "should have" already adopted the GHG-reducing practices, but for some reason have not).

    Carbon finance is ideally a means of encouraging innovative or nascent activities so that they might reach a tipping point, beyond which the support is no longer necessary for widespread adoption. It is difficult to argue that practices already implemented on half of croplands require carbon offsets to spur change. An underpinning of integrity, stringent additionality helps uphold the general legitimacy of offsetting as meaningful climate action. (**NCP**)

    RESPONSE: Thank you for your comments. We settled on the threshold for this performance standard test following the barriers analysis discussed in Appendix A, and following extensive feedback from stakeholders. Please see comment 7 below, which demonstrates significant support for a more flexible approach to setting common practice thresholds, from a broad group of well-respected peers. Please also note that Environment and Climate Change Canada have recently proposed the adoption of a 40% common practice threshold, within their discussion paper 'Considerations for Protocol Development in the Federal Greenhouse Gas Offset System'[2]. This is further indication that there is growing support for a more flexible approach to common practice assessments.

14. Comment to Climate Action Reserve and Verra staff on the assessment of additionality for agricultural land management projects.

    Thank you for taking the time to consider this comment. Both of your organizations have put in important work for the development of innovative methodologies for connecting agricultural land management (ALM) activities with the carbon markets at scale. The National Academy of Sciences 2018 report supports us in the belief that restoration of agricultural soil carbon could simultaneously help resolve three global problems – food security, biodiversity, and climate change. When carbon is restored to agricultural soils by building soil organic matter, it generally makes the soils more fertile, better at retaining water and fertilizer, and more resilient to climate change. Thus, if done correctly, the restoration of soil carbon would result in a quadruple win: increased food production per acre, better livelihoods for farmers and their families, less pressure to convert remaining forests to agriculture, and a substantial carbon sink.

    During these development processes you have both come to the rational conclusion that traditional approaches to assessing additionality for carbon offsets, such as assessment of financial feasibility and practice-based assessments of activity penetration – or common practice – are not appropriate for ALM projects in their current forms. Both registries are proposing novel approaches that break out of the traditional mold applied by the carbon market, yet also maintain a rigorous, investable standard for additionality.

    This letter is intended as a show of support for your new approaches to assessing additionality in agricultural carbon projects.

    Because they are new and different, novel approaches may cause confusion and criticism from carbon market stakeholders, with some voicing strong opposition to these new policies. That is understandable, and public review and comment is crucial for credibility and market confidence. This letter is intended as a demonstration of support for the new approaches being proposed, and to provide an independent perspective on additionality in ALM projects, and why new approaches are needed.

    Why not rely on a financial additionality test?

    Research has demonstrated that increasing long-term financial returns of preferred practices is insufficient to change farmer behavior (Howley, Buckley, O'Donoghue, & Ryan, 2014). Growers do not act purely in pursuit of long-term profit maximization, even if that is how it appears. Growers also consider factors such as maximizing social value, adhering to in-group norms and values, and simplifying the decision-making process through heuristics.

    Barriers faced by farmers that are ignored through simple financial additionality screens include:

[2] As of the date of publication of this summary, this document could be downloaded here: https://www.canada.ca/en/environment-climate-change/services/climate-change/pricing-pollution-how-it-will-work/output-based-pricing-system/carbon-pollution-pricing-considerations-protocol-development.html.

1. Widespread adoption of new ALM practices will require group collaboration to reach a critical mass of market demand for proper equipment, crop inputs, and services. This market demand is needed to signal to farm equipment manufacturers and local equipment providers to commit resources to produce and service new equipment; other input providers to shift to meet demand for inputs unique to new agronomic practices (ranging from physical crop inputs to information and consultation services); local, group, and institutional knowledge to incorporate new information and overcome outdated heuristics; and commodity markets to develop infrastructure to offload commodities with unique, marketable characteristics.
2. Farmers experience long delays between decisions and outcomes, preventing them from receiving immediate feedback to reshape future decisions. Making economically correct decisions requires strong understanding of the short and long-term economic impacts of various decisions. Long delays between decisions and outcomes dilutes the connection between the two for the farmer.
3. Farmers experience significantly more uncontrollable factors, particularly market prices for their products and weather patterns, that impact their overall productivity and profitability. These heightened uncertainties lead growers to prioritize risk mitigation over long-term profit maximization. New practice changes are viewed as inherently risky, mostly due to lack of sufficient information for farmers.
4. Farmers face systemic and structural barriers to widespread adoption of significant practice changes. These barriers include lack of education on specific practice changes (both in universities as well as through generational knowledge transfer), lack of agronomic support through common channels (e.g., seed and equipment dealers, government agencies, university extension), and availability and proximity to markets which value crops produced with more sustainable methods.

These unique factors can slow or halt the adoption of new agricultural practices, ones beneficial to the planet through the reduction of GHG emissions and increases in soil carbon levels. Project financing from the sale of GHG credits is intended to address barriers related to practice change, primarily risk reduction, and should be communicated as such to farmers. Additionally, project financing is often coupled with other values designed to address barriers such as access to information and guidance to not only have success with new practices, but also to have confidence in implementation; this is a vital factor for encouraging behavior change.

Rather than a subjective assessment of simple farm economics, the proposed methodologies rely on demonstration of widespread risk aversion and other barriers to change in the agricultural sector. Appendix A of the Reserve's Soil Enrichment Protocol (SEP) v1.0 includes an assessment of behavior in the agricultural sector that covers the list of barriers above in greater detail.

Why not conduct activity penetration assessments for each individual practice?

The typical approach to assessing common practice for carbon project methodologies is to consider any activity with an estimated or measured penetration level above 5% to be considered "common." This traditional approach has been suggested by at least one public comment to the Reserve's SEP v1.0. Assessing the penetration of individual ALM practices against a threshold of 5% is inappropriate for the following reasons:

1. ALM projects are driving toward multiple practices, so a single practice assessment is not relevant. Over a project lifetime, ALM projects will only have a financially compelling

GHG benefit if the farmer adopts multiple practices. However, it's completely unreasonable to ask a farmer to adopt multiple practices in order to gain entry to the program. Sustainable ALM is a journey, and (as discussed above) farmers are risk averse, so practices are necessarily adopted one at a time. Although some individual practices may be adopted at rates greater than 5% in certain regions, it is exceedingly rare to find such high penetration of farmers adopting multiple practices simultaneously.

2. Practices will not be static at the field level. Sustainable ALM is a journey involving experimentation and continuous assessment of performance that must be tailored to the individual field and farmer. This means the farmer may try practices and abandon them later in favor of an alternative that works better for their farm. Thus, the focus should be on the act of behavior change and GHG performance, which are both incentivized by the proposed methodologies. Successful implementation will involve experimentation and stacking of practices.
3. Whether a practice is "common" must be defined differently for agriculture. A single practice with a 5% penetration rate is essentially a radical experiment in the eyes of the farmers. The 5% threshold for additionality may have been useful or appropriate in the context of renewable energy technologies, or industrial emission sources, but it is neither useful nor appropriate for ALM projects. As discussed above, farmers are often skeptical of practice changes until they are adopted very widely. While it is true that multiple practice adoption is rare, single practice adoption must be eligible to get those farmers in the door. The focus on crediting for performance ensures that the incentives are aligned for farmers to move to multiple practice adoption as quickly as possible to maximize carbon revenues.
4. Different practices would need to be assessed at different scopes. The relevant sphere of influence will be different for different practice changes depending on crop type, political boundaries, access to technology, water availability, soil type, geography, etc. For ALM projects – which involve multiple practice changes, crops, and geographies, grouped together into one project – any assessment conducted on individual practices would be impossible to standardize across the project. Any such assessment should cover the entire project region, focusing on overall practice change, rather than individual practices.
5. Agricultural practice changes are not "one size fits all." The actual nature of the practice change will vary not only between crops and regions, but also between farmers and time periods. The complexity of the changes and the diversity of practices makes it difficult to draw direct comparisons and clear assessments of what is "common" practice.

Conclusion
I, as well as the undersigned stakeholders, strongly support the work of the Climate Action Reserve and Verra to develop new approaches to assessing additionality for ALM projects. The opportunity to positively impact the climate is massive but will only become reality if we are able to take a global approach. We urge you to avoid reverting back to traditional approaches which are neither appropriate nor effective for ALM projects. We urgently need incentives to overcome cultural and economic barriers to change, and climate finance can provide this incentive. This feeds into the need for a sensible, pragmatic approach to additionality for new science-based soil organic carbon methodologies.

Successful ALM methodologies should define eligibility in relation to adoption of practice changes generally, and quantify crediting based on performance (in the form of GHG benefits). The practice changes are needed to get into the program, but the farmers must actually reduce their GHG emissions and/or increase their carbon sequestration in order to benefit from the project. (Newcombe et al.)

RESPONSE: Thank you for your comments. We agree with the general thrust of your comments and believe the current protocol accurately reflects incorporation of your perspectives.

3.4.1.2 Performance Standard Test – Project-Specific Means to Demonstrate Additionality

15. Natural Capital Partners suggests that these project-specific approaches could too easily fall short of ensuring additionality. Three fields don't necessarily compose a statistically valid sample size, and the remote sensing approach only requires identification of "other fields," without specifying a minimum number or acreage. In neither case is it clear that the rotations have to have occurred within the past three to four years. However, the overriding issue is that neither approach is framed in a way that would prove that the county-wide long-term adoption rate is below the common practice threshold (which we still advocate should be lowered). A waiver of the threshold should be based only on evidence that it is wholly inaccurate; the proposed approaches are too open to non-representative examples. (**NCP**)

    RESPONSE: Thank you for your comments. Regarding Section 3.4.1.2(2), the intent there was that projects must provide evidence that demonstrates that the majority of fields in the county in which the subject field is located have either implemented conventional tillage, or a rotation of no-till or reduced-tillage with conventional tillage, during the historical baseline period, as well as identifying 3 specific fields which demonstrate this trend. We have updated the guidance accordingly, to make the guidance clearer.

3.5 Requirements for Permanence

16. The August 2020 SEP draft reduction in length of the Project Implementation Agreement (PIA), dropping the contract period from 30 years to 10 years, and the reasons offered for this change in the *Summary of Comments and Responses*, highlight our concerns over SOC permanence and the accounting options CAR is proposing for managing reversals. The size of anticipated annual SOC enhancement levels are small relative to the existing SOC stock and the literature suggests achieving statistically significant measurable enhancements will require continuous application of conservation practices for a decade or more. The shorter PIA of 10 years in combination with the Tonne Year Accounting option, which will likely be the accounting method chosen by the majority of applicants, does little to encourage long term implementation of the conservation practices expected to result in soil C sequestration. To encourage management which will result in the a more realistic potential for net atmospheric CO2 reduction when SOC enrichment and sold offset emissions are considered, we recommend that Tonne Year Accounting only be allowed as an option after completion of a 30-year PIA. This time frame would be more in line with the anticipated formation of new soil C steady states and C reversals are unlikely to exceed the proportional credit afforded by the Tonne Year Accounting approach. (**EDF**)

    RESPONSE: Thank you for your comments. Please note we reduced the duration of the crediting period, the period over which credits can be generated, from 30 years, to 10 years, though the crediting period may be renewed up to two times for each project or, in the case of aggregated projects, for each field comprising the aggregate.

    We prefer to create flexible options and defer to the market as to which option respective parties wish to employ in their given projects. Projects will have the option to choose a TYA approach, which would result in a commitment to maintain additionally sequestered stocks for less than 100 years, in exchange for proportionately smaller issuance of credits that ensures credits issued reflect the time-value of any sequestered tonnes relative to a 100-year permanence time frame. While the TYA approach does essentially limit crediting relative to actual SOC accumulation according to the permanence commitment made by the project owner at the time of credit issuance, there is an inherent incentive for project owners employing TYA to extend their PIA since any previously sequestered SOC that continues to be maintained—and for which they have not received 100% of the time-value relative to the 100-year permanence timeframe—will be eligible for continued credit issuance at a rate of 1% per year through any extended PIA commitment. Nevertheless, if projects wish to earn more credits at the time SOC is initially sequestered, they can adopt the more traditional tonne-tonne accounting approach, with a 100-year PIA, rather than the alternative TYA approach. In this respect the TYA approach should be viewed as a more conservative crediting approach to avoid the potential for over-crediting.

Whilst the idea of SOC becoming more stable after some time was raised during the workgroup process, we received insufficient information to enable us to enshrine that as a basis for adjusting permanence requirements.

17. The August draft protocol retains the same structure, with confidential PIA terms that are explicitly not required to last for the claimed 100-year duration of credited climate benefits. Worse, the Reserve's response to public comments erroneously and misleadingly suggests that protocol implementation will rely primarily on 100-year PIA contracts (see 37), while the protocol text explicitly retains the option for shorter PIA contract terms (Section 3.5.3). We believe the risk of misleading prospective buyers is extreme if the Reserve purports to offer 100-year permanence protections while only requiring projects to sign project contracts with shorter and potentially confidential durations. The Reserve must clarify its practices here and ensure those statements are consistent with the protocol text. (**CarbonPlan**)

    RESPONSE: Thank you for your comments. Our intent with these requirements is to retain the same 100-year permanence requirement we have adopted consistently since our first Forest protocol was adopted in 2005. Some of our peers may be moving away from this type of requirement, but we are not. We have updated the guidance in this section to make our intent clearer. To clarify things, our intent is that projects could choose to adopt multiple shorter PIAs, provided the total duration of the sum of those PIAs equates to a 100-year commitment (i.e. they could adopt 5 consecutive 20-year commitments). Projects could alternatively seek permission for some other form of compensating mechanism, such as a surety bond, but that would also need to be in place for a total duration of 100 years. Lastly, projects could choose to adopt a tonne-year accounting approach, for which credits are only issued proportional to the time frame of commitment made by a project owner in a PIA to maintain sequestered stocks. Extensive guidance on tonne-year accounting is provided elsewhere. We hope this offers some clarification.

5.3.1 Modeling the Baseline

18. Portions of this section introduce the concepts necessary for selecting the historical baseline on which the modeling will be conducted. Development of the baseline is absolutely crucial for any carbon project and is particularly complex in the case of the SEP, due to the wide range of crops and practices allowed by the Project Definition. In the public comment draft of the SEP, we feel that this section is not quite complete in its guidance around how the

historical baseline is to be modeled during each year of the project. As written, it's not clear whether the averaging of model outputs for one year is meant to drive the inputs of the model for the next year, or whether the averaging is only used for the quantification of the baseline in a given year. It is our opinion that the former is not workable with existing biogeochemical models, but that the latter is a very sensible approach. We recommend making it clear that when modeling multiple historic years separately to determine the baseline for a given year, the outputs from each model run should form the inputs of the model runs for the next year without being averaged together. So if a project has three years in its historical baseline period, it would maintain three separate instances, or threads, of modeling throughout the crediting period. The results of each of these instances would be averaged together in a given year, according to the protocol guidance, to determine the baseline for that year. (**Indigo Ag**)

RESPONSE: Thank you for your comments. Updates to this effect were made and included in the version of the protocol released for the second public comment period.

5.3.1 Contribution to the Buffer Pool

19. [W]e raised concerns about a complex set of calculations that describe what share of a project's credited climate benefits must be set aside for the protocol-level buffer pool, in order to insure against the risk of reversal. We noted that the Reserve's calculations included a series of loopholes that render the vast majority of the calculations irrelevant, including the option for a company to provide a surety bond to a wholly owned subsidiary or parent company — a sign of potential fraud in securities and insurance regulation because closely-held corporate entities can fail together under financial duress. We noted that the choices of parameters for capitalizing the buffer pool contributions were not justified by an evidence, but rather asserted as stated. By taking advantage of the loopholes we identified, a project would only need to set aside 5% of its offset credits for the common buffer pool — a level offered as sufficient to cover the risk of all reversals over the coming 100 years, from floods to fires to bankruptcies. (**CarbonPlan**)

    RESPONSE: Thank you for your comments. Please note the protocol makes simple reference to the potential use of surety bonds, subject to approval by the Reserve. Our own self-interest – in addition to the robust governance and oversight of our Board - necessitates that we not accept illegal or uncertain arrangements.

    Please note, we discussed buffer pool contributions at length during our public workgroup meetings. In the absence of objective standards, we create our own, based on workgroup feedback and our extensive experience with such matters. Please note, we were continually pressed in public workgroup meetings to reduce buffer pool contributions, by a range of workgroup members, but we did not. Please note, as we pointed out in our responses to your previous public comments, the Reserve is committed to stepping in to make whole any reversals the buffer pool is unable to address (please see Section 2.8.1 of the Reserve Offset Program Manual for the means by which the Reserve would make good on any reversals not covered by the buffer pool).

5.5 Emissions from Leakage

20. With [the protocol] being refined to agricultural lands, I still think there is much to question about how focusing on sequestering C will affect yields. The literature is still unclear on this topic and if it decreases yields it may just result in the need more farmland. Is there a way to

account for how acreage in agriculture may shift as a result of adopting these management practices? (**Badzmierowski**)

RESPONSE: Thank you for your comments. Such changes are considered 'leakage', and the 'leakage' mechanisms in Section 5.5 account for the potential shifting of crop production outside of the project area, if yields were to decline within the project area.

Table 6.1 Agricultural Management Data Collection

21. My background is in soil chemistry with a current focus on determining the potential carbon sequestration of land-applied biosolids. I see that biosolids has at least been included as an intervention method based on the first round of comments. I would like to say this is an important first step. Biosolids are produced in significant quantities with three end pathways in the United States: land application, landfill, and incineration. Improving the recycling rate to land is especially important as it can offset fertilizer production and use, while reducing the environmentally deleterious effects of landfilling and incinerating biosolids. I am sure the lack of greater detail in your protocol is a result of lack of quantitative numbers regarding biosolids and resulting soil organic carbon (SOC) stocks. Two other faculty (Dr. Greg Evanylo and Dr. W. Lee Daniels) and I at Virginia Polytechnic Institute and State University are currently working to address this knowledge gap through meta-analysis. We hope that we can partner with you come the new year upon completion of our work. (**Badzmierowski**)

    RESPONSE: Thank you for your comments. The protocol points out activities that are likely to have significant GHG impacts, but eligibility of any given activity will ultimately depend on availability of suitable quantification approaches. The quantification approaches under this protocol include the use of models, and default equations/emission factors. We would be very interested to follow up with you further on your work. Thank you.

22. We support the addition of biosolids to Table 6.1 as an organic soil amendment. (**CASA**)

    RESPONSE: Thank you for your feedback via multiple public comments, and discussions.

6.5 Soil Sampling and Testing Guidance

23. Equivalent soil mass correction is an important issue in SOC accounting. If you are actively building SOC, you are also fluffing up the soil so that there is less soil mass in the upper 30 cm under the improved management. Since SOC stock is OC% × soil mass, you are "losing" soil mass with some SOC in it by always sampling to the same depth. Conversely, if the soil is compacted and there is no change in SOC, fixed depth sampling will result in an apparent sequestration. Both of these scenarios are easily corrected for, especially if you collect an extra depth increment, and there are several good publications (including this recent paper by von Haden and colleagues) suggesting how to do it. The Australian Carbon Farming Initiative methods for SOC require reporting SOC by equivalent mass and describe a simplified approach (https://www.legislation.gov.au/Details/F2018L00089). Models actually implicitly assume an equal soil mass through time because they are initialized to a baseline mass of SOC (contained within the baseline year volume of soil), so this type of field data would be more consistent with model output.

    The revised protocol draft is not clear on these issues. In section 6.5 (p 69), the protocol states "SOC measurement will by necessity include calculation of SOC based on bulk density, as well as the determination of SOC stocks based on either %C by mass, or use of the equivalent soil mass method." But then the minimum standards only specify a single 0-30 cm sample with measurement of bulk density seeming to be optional (p 72 – "If bulk density methods are being used to convert soil carbon concentration to soil carbon stocks, coarse (>2 mm fraction) content corrections to bulk density must be made."). How do you convert from concentration to stock without measurement of bulk density? Good bulk density measurement is critical to SOC accounting and should be spelled out in the protocol. The Reserve claims to have addressed this in response to the public comments (p22 of the summary of comments documents) but this response has not made its way into the new draft. The highlighted statement on p 69 in itself is confusing ("SOC stocks based on either %C by mass, or use of the equivalent soil mass method") – was this sentence meant to read "based on either %C by ***volume***, or…"? While my preference is to require equivalent mass reporting, the protocol should at a minimum provide guidance as to how to do equivalent mass reporting. (**Sanderman WCRC**)

RESPONSE: Thank you for your comments and multiple discussions. Per guidance we have received, including your guidance in this comment, if soil is being fluffed up, then sampling to 30 cm will be conservative. We expect that compaction, as a result of the project, is unlikely to be common, and that compaction during the crediting period would likely be detrimental to soil health and thus could reduce SOC. These two factors should mean that the risk of over-crediting due to significant compaction, caused by the project, should be a non-issue for this protocol. Based on the above, it's much more likely that sampling to 30 cm will be conservative, in the vast majority of cases, and the risk of over-crediting due to compaction is minimal.

We do require measurements of bulk density. We have guidance from multiple soil scientists that estimation of SOC will not be possible without measurements of bulk density, therefore we took that as a given, and did not feel it necessary to spell out bulk density requirements in great detail. We typically don't include lengthy descriptions of appropriate methods for every requirement in our protocol, and so have not included anything more for bulk density measurement here. We will consider including such further guidance in a forthcoming Handbook, which would include practical guidance to help parties implement projects.

With respect to equivalent mass reporting, we have been advised that this method is not widely applied beyond research applications. Based on that guidance, and our long-standing preference to not included detailed guidance on all possible methods employed in our protocols within the body of the protocols themselves, we have chosen not to include such guidance here. Note, we will consider including guidance on the use of this method in a forthcoming Handbook. With respect to the option to assess %C by mass, we understand that the SOC analysis would tell us % C by mass, and then the bulk density analysis allows one to convert the % by mass into a % by volume.

6.5.1 Sample Design and Soil Collection

24. There have been numerous concerns raised over the initial soil sampling protocols in both the public comments and in working group member comments that seem to have been refuted with the publication of Appendix B in the newly released draft protocol. On July 16, 2020, you specifically asked for my opinion on sampling to 40 cm to accommodate for tillage induced redistribution of soil organic carbon (SOC). I responded by adding another reason to sample deeper than 30 cm – in order to make a so-called equivalent soil mass correction.

I appreciate the thought that went into Appendix B as a solution to the sampling depth issue. However, I do not think this negative list test is sufficient. Disk and chisel tillage methods are often conducted at 10” (as noted in Table B2) and even greater depths. Ten inches (25 cm) does not give sufficient soil depth below the plow layer. Additionally, the definition and depth columns in Table B2 are inconsistent (i.e. moldboard say 2-10” in the description but assigns a depth of 30 cm). A much simpler application of this negative list test would be to require the depth of tillage to be reported for the historic period. This should be easy information to gather as it is strictly a function of farm equipment (what size disks or chisels are on your plow?). (**Sanderman WCRC**)

RESPONSE: Thank you for your comments, and multiple discussions. Please note that the Public Comment draft of the protocol included new guidance at both Section 6.5.1, and in Appendix B.1, that directs that actual historical plow depth will be used to determine if historical tillage was typically employed deeper than 20 cm. This should allow us to capture any historical plowing that is of concern. Table B.2 was intended to provide indicative depths, but we’ve now removed it based on your feedback, and the changes noted above.

25. I understand the desire to use a model driven approach so farmers can be credited annually but I also find the “models cannot simulate deeper than 30 cm” argument against mandating soil sampling deeper than 30 cm to be a red herring. Perhaps only models that can simulate deeper than 30 cm should be allowed. Alternatively, if models can only simulate to 30 cm (and some of the most widely used models actually only simulate the top 20 cm), then any gains or losses below model depth can be added or subtracted during the true-up period to compensate for changes happening below the historic plow layer. This would further alleviate concerns about including shifts to no-till practices.

    Sampling to at least 40 cm with a minimum of two increments along with robust measurement of bulk density will eliminate much of the debate around inclusion or exclusion of no-till adoption and allow for the requirement of equivalent mass reporting of SOC stocks. By incorporating information on SOC change in the deeper soil layer during the true-up period, this could eliminate the shortcoming of models only simulating the top 30 cm. (**Sanderman WCRC**)

    RESPONSE: Thank you for your comments, and multiple discussions. We believe the current safeguards built into the protocol regarding any risk of upwards SOC migration are reasonable. With respect to equivalent mass reporting, such a method is already eligible under the protocol. In this light, we do not think it necessary to abandon the hybrid modelling / sampling approach employed in this protocol. We believe this approach to be robust, more flexible, and thus more widely accessible to projects, than an approach favoring only direct sampling.

26. I have three concerns about sampling depth. First, the 30 cm sampling depth is too shallow and would result in over-crediting (Angers et al., 2007; Baker et al., 2007; Govaerts et al., 2009; Luo et al., 2010; Powlson et al., 2014). Though Appendix B raises some interesting points, the peer reviewed literature is unequivocal: deeper sampling is needed. Plowing causes SOC to accumulate just below plow depth. When tillage changes, that deeper SOC begins to dissipate, offset by the accumulation of SOC in the previously plowed region. Failure to sample both the decline in deeper SOC that occurs in tandem with SOC accumulation in the upper layers will lead to over-crediting. Luo et al., 2010 showed that sampling to 40 cm all but eliminated gains in SOC that were detected when sampling at shallower depths. In light of this literature, I had recommended either: a) sampling to 40 cm

(as recommended by Luo et al., 2010, page 229) or b) having a variable sampling depth of 15 to 20 cm below historical plow depth. The protocol adopts neither approach.

Appendix B argues that modeling limitations prevent deeper sampling. If models truly cannot go below 30 cm, then the Protocol should only allow projects that have historically plowed to a depth of 15 cm (10 cm if being conservative). Such a standard would respect both the limits of models and the empirical evidence of how SOC is affected by changes in tillage. Shallow sampling depths will likely result in money for project development flowing to fields that practiced forms of deep tillage that are not explicitly prohibited. This is because the reduction in SOC from changing tillage would be under sampled, yielding more CRTs and higher returns.

My second concern is that the Protocol establishes limits on tillage technologies, as opposed to tillage depth. Such a standard is insufficient. As Table B.2 makes clear, moldboard, ridge till, and chisel tillage all are practiced at depths ranging between 2 to 10 inches (5 to 25 cm) and potentially deeper. Rather than using broad regional averages for specific technologies, the protocol should use specific, field level information. Table 6.1 specifies that projects need to provide information about historical tillage depth. That information could be used to exclude any project that plows below a certain depth (e.g., 15 cm in the case where only 30 cm of soil is required to be sampled). I also believe the last sentence of section 6.5.1 needs revision. It reads: "Fields historically employing deep ripper or moldboard tillage practices may become eligible to be credited for SOC gains if/when they subsequently adopt any tillage practice other than no-till in subsequent reporting periods." SOC takes on the order of tens (20+) of years to equilibrate (Powlson et al., 2014 and references therein). However, this sentence would allow any project that stopped deep tillage prior to the baseline (e.g., 3-5 years) to be project eligible. This would result in over-crediting, as SOC accumulation would be at least partially due to the cessation of a practice specifically prohibited under the protocol.

My third and final concern with sampling is that the fixed 30 cm standard goes against best practices in quantifying National Greenhouse Gas Inventories, as outlined by the IPCC. As we discussed previously, these standards dictate that depths of greater than 30 cm are needed "if it is clear that land-use change and management have a significant impact over the proposed depth increment." Ample evidence exists that changes in tillage practices has a significant impact on SOC at depths that exceed 30 cm, meaning best practices require deeper sampling or exclusion of fields based on historical plow depth. (**Badgley**)

RESPONSE: Thank you for your comments, and multiple discussions. With respect to the potential risk of soil migration, when moving from deep tillage to no-till, the public comment draft of the protocol sets out requirements (at Section 6.5.1 and Appendix B.1) that direct that projects must provide data on historical plow depth, and such data will be used to determine if historical plowing went below 20 cm. These requirements should ensure that samples to 30 cm capture any SOC migration some 10-20 cm below typical historical plow depth (we note that your recommendation was 15-20 cm below plow depth). If a project employed deep tillage historically, and then employs no-till for a very short time, before changing to another tillage practice, then there should be negligible migration of SOC during that short period of no-till. For this reason, we can safely allow projects to again be eligible for generating SOC credits, having switched back to another tillage practice.

We disagree that the SEP SOC guidance violates IPCC best practices. Please note we have feedback from several soil scientists, including on the workgroup and outside of the workgroup, that disagree with your characterization that sampling to 30 cm, in the context of this protocol, is out of line with IPCC best practices. The IPCC generally supports soil sampling to 30 cm, they then advise deeper sampling is good practice 'if it is clear that mixing of cultivated soil occurs at deeper depths, and if land-use change and management have a significant impact over the proposed depth increment' (IPCC 2006)[3]. The IPCC also state that where depth of sampling is set and cultivation mixes soils to a depth deeper than the sample point, then it would be good practice to sample to the depth at which land use and management impact SOC (IPCC 2019)[4]. Given we are excluding historical deep plowing (below 20 cm), deep tillage practices are not typically applied in the US, and we don't expect deep tillage to be common during the project, then we should not expect mixing of soils below 20 cm, let alone 30 cm. Therefore, it is not likely there will be significant impacts at depths below 30 cm. In these circumstances, we believe this protocol follows IPCC best practices. Moreover, we note that 30 cm is set at the absolute minimum and projects are encouraged to go lower with their sampling (with some commenters noting that 30 cm will actually underestimate SOC because positive SOC migration to lower levels will not be captured).

27. We are a group of scientists committed to ensuring that carbon removal efforts are informed by the best available science and data.

    We write with serious concerns about the scientific integrity of the Climate Action Reserve's Soil Enrichment Protocol and its proposed adoption of a 30 cm minimum sampling depth for measuring soil organic carbon (SOC). The peer-reviewed scientific literature clearly documents that this kind of shallow sampling can lead to systematic overestimation of increases in soil organic carbon, jeopardizing the credibility and efficacy of the Protocol. Significantly deeper sampling is required to achieve confidence in the climate benefits created by credited soil management and agricultural practices. If the Reserve fails to address these issues, we are concerned that the Protocol could lead to large-scale over-crediting and harm the development of policies to recognize soil carbon benefits by encouraging the adoption of similarly flawed sampling practices elsewhere.

    Long-term experimental data from the UC-Davis Century Experiment show that increases in soil organic carbon in the upper soil layers can be accompanied by significant decreases across the microbially active soil profile that extends below the plow layer. As a result, Tautges et al. (2019) caution that "focusing only on the surface layer of soil could result in grossly overestimated SOC gains." Paired measurements of soil organic carbon collected from tilled and non-tilled systems suggest a similar pattern, with changes in tillage practices redistributing soil carbon such that apparent gains at the surface are accompanied by losses at depth. These losses can manifest well be- low 30 cm (Bai et al. 2019) and can partly offset or even completely negate gains near the surface (Luo et al., 2010). Such losses would not be observed under the Protocol's proposed sampling requirements. These effects are not limited to changes in tillage regime, similar patterns have also been documented with changes in crop type (Ledo et al. 2020), and carbon isotope studies indicate that a substantial fraction of the soil organic carbon below 30 cm is dynamic on decadal timescales (Baisden and Parfitt 2007, Slessarev et al. 2020). Taken together, these studies paint a consistent picture; measuring soil organic carbon only in the "plow layer" (or upper 30 cm of soil) is not a defensible approach for assessing how land management practices affect total soil organic carbon— yet that is all the Protocol requires.

[3] 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Chapter 2, at 2.38.
[4] 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Chapter 2, at 2.39.

Shallow sampling can also introduce significant measurement bias that is driven by changes in soil bulk density, rather than real accumulation of soil organic carbon. This presents a significant concern with respect to the Protocol because changes in agricultural practices tend to be accompanied by changes in soil density. If sampling depth remains fixed and soil density changes due to new land management practices credited under the Protocol then apparent changes in shallow soil organic carbon could simply reflect differences in the mass of soil sampled at a fixed depth, rather than real changes in soil organic carbon (Lal 209, Wendt and Hauser 2013). Density-driven biases are easily avoided through deeper sampling, since changes in density tend to cancel out with depth, or by sampling at multiple depths and reporting changes to soil organic carbon on mass-equivalent basis. Credible measurements of soil climate benefits require one (or preferably both) of these approaches, but the Protocol treats them as optional, leaving the door open to significant measurement bias.

Together, these shortcomings fundamentally undermine the Protocol's ability to generate real, quantifiable carbon credits. Problems with shallow sampling depths and changes in soil density are well-documented in the peer-reviewed literature and require serious attention in any carbon crediting application. We are concerned that any effort to credit soil carbon benefits that ignores these concerns risks not only over-crediting benefits, but also prejudices the development of an evidence-based soil carbon management paradigm going forward. (**Badgley et al.**)

RESPONSE: We respectfully disagree that this collection of studies supports your contention, that sampling to 30 cm is an indefensible approach for assessing SOC under this protocol. Please see our response to comment 19 above, where we cite IPCC guidance that sampling to 30 cm remains current international best practice, recently reaffirmed by the IPCC in its 2019 update to its 2006 National Inventory Reporting Guidelines.

With respect to your contention that SOC migration from lower to upper layers is not limited to changes in tillage regime, you cite a paper by Ledo et al. (Ledo et al. 2020). We note that this paper pertains to changes from annual to perennial crops, changes from forest to perennial crops, and in both instances in particular changes to perennial grasses, palms and woody plants. With respect to changes from annual to perennial grasses, the report notes 1% increases in SOC at the 0-30 cm depth, as well as even greater 10% gains at the 0-1 m depth, so sampling to 0-30 cm would be conservative for such changes. Please also note that changes from forests to perennials is explicitly excluded (see Section 2.2). Please also note that with respect to palms, it's unlikely such changes could be credited under this protocol, given current model capabilities, and model validation requirements. Thus, we do not believe the findings of this study can reasonably be used to assert that SOC migration is a concern for all possible combinations of crops / practices / regions we expect under this protocol.

With respect to the findings in Tautges et al. (2019), pertaining to SOC migration in the context of winter cover crops, whilst the results do raise valid concerns, we do not feel it's reasonable to extrapolate such concerns across all US farming systems, based on a single study conducted on a few (maize-tomato, wheat/fallow) crop / practice combinations at a single site.

With respect to your contention that shallow sampling can also introduce significant measurement bias that is driven by changes in soil bulk density, please see our response to comment 16. We have received guidance from other soil scientists that sampling to 30 cm introduces bias in the form of being more conservative. As the soil is fluffed up, due to project activities, less SOC build up will be captured in a sample to 30 cm, and thus SOC build up below 30 cm would be conservatively excluded. In addition, we note that Badgley repeatedly advocated for sampling to 15-20 cm below plow depth. The most common plow practices in the US range from 10-20 cm deep. The protocol specification of a minimum sampling depth to 30 cm would therefore be 10-20 cm below plow depth, which other experts, including an IPCC lead author, have indicated is sufficient.

28. We do appreciate the value of the hybrid modeling and soil sampling approach that CAR has included in the SEP. However, with regard to the changes made in the August version in an attempt to address the concerns about the inaccurate assessment of net soil carbon sequestration over the whole soil profile by depending only upon soil samples collected at a 30 cm depth, we think the SEP modification falls short. The exclusion of SOC credits for conversion to no till in fields previously managed with deep tillage techniques as a remedy for 30 cm soil sampling is not supported by recent literature, which finds no net change in carbon over the 1 meter soil profile (see, for example the systematic review of Haddaway et al 2017[5], Chenu et al 2019[6] and the exclusion of SOC enhancement from tillage by Grisom et al 2017[7] and Fargione et.al, 2018[8].) Nayak et al 2019[9], not only points to the need for 1-meter soil samples to adequately measure net soil C sequestration but also points to the evolution of eddy covariance approaches as an upcoming and superior way to measure net soil C and N_2O fluxes. Various efforts are underway to develop high resolution datasets to support testing and validation of soil flux measurements from field level to satellite-based assessments. These techniques may provide simpler and more cost effective methods of measuring net SOC enrichment. However, given the current questions surrounding the validity of using a 30 cm soil sample depth as an accurate measure of SOC enrichment, the CAR protocol should require that a statistically appropriate percentage of soil samples in each stratified collection of samples be taken to 1 meter. While current models may only include predictions to 30 cm, a subsample of 1 meter deep samples would allow the net soil C enhancement over the full soil profile to be compared to the 30 cm sample, allowing for corrections for the 30 cm soil C estimates. Furthermore, the data collected by requiring a subsample of soil samples to 1 meter depth will help stimulate modeling to this depth and work towards resolution of the current debate over the true net soil carbon enrichment. (**EDF**)

    RESPONSE: Thank you for your comments. We have received differing feedback from experts in soil science on this issue. Based on this feedback, and our interpretation of the meta-analysis we cited in the protocol, if we exclude historical deep tillage, and if we don't expect deep tillage to be common in the project, then we should be able to account correctly for SOC migration. Please see our response to comment 19 for further discussion of this issue.

6.6 Modeling Guidance

29. Our May comment letter raised concerns about the kinds of soil carbon models used to calculate projects' carbon credits in between physical soil carbon sampling at five-year intervals. The earlier draft of the protocol from April provided few constraints on the type of

[5] Haddaway et al 2017 https://environmentalevidencejournal.biomedcentral.com/articles/10.1186/s13750-017-0108-9
[6] Chenu et al 2019 https://www.sciencedirect.com/science/article/pii/S0167198718303738?via%3Dihub
[7] Griscom et al 2017 https://www.pnas.org/content/114/44/11645
[8] Fargione et al 2018 https://advances.sciencemag.org/content/4/11/eaat1869
[9] Nayak et al 2019 https://doi.org/10.1016/j.scitotenv.2019.02.125

model or modeling framework that could be used. In contrast, the updated August draft and the accompanying Model Calibration, Validation, and Verification Guidance document contain additional safeguards to ensure that model calibration is based on peer- reviewed evidence and that model calibration reports will be made publicly available. We believe that open access to the basis for crediting methods is essential and sincerely appreciate the Reserve's willingness to rise to this important standard. Thank you. (**CarbonPlan**)

RESPONSE: Thank you for your comments.

7 Verification Guidance

30. We recommend streamlining the verification process as it appears to be overly difficult and would require the expertise of third parties to complete. This may be a disincentive for enrollment in the program. (**CASA**)

    RESPONSE: Thank you for your comment. We have made considerable efforts to make verification of projects registering under the protocol as simple as possible while maintaining standards that provide assurances as to the veracity of the emission reductions being claimed by each project. Given the anticipated complexity of the projects and the quantification of emissions reductions, the verification requirements are relatively straightforward. If you have specific suggestions for modifications that will streamline the verification process while maintaining the integrity of any credits issued, we welcome such suggestions. Regardless, all projects must undergo verification by a third-party verification body. If the suggestion is that verification bodies will be required to hire expert consults to assist in verification, this is already a standard practice of verification bodies in instances where they do not maintain or obtain expertise on staff for all subject matters that are the focus of offset protocols. Nevertheless, verification bodies are required to be trained and accredited for each offset protocol for which they wish to conduct verifications, hiring the services of external consultants in those instances for which they require specific outside expertise to properly complete their review of projects.

8.3.1 Verifying Proper Use of Models

31. [W]e note that the August draft protocol continues to provide that the verification team does not need any expertise in the modeling software a project employs to calculate its carbon credits, so long as the project hires its own expert to do that modeling in the first place. By definition, this is not a reliable means of independently verifying model calculations. It is a system for creating the appearance of third-party verification, when in fact there is none. (**CarbonPlan**)

    RESPONSE: Thank you for your comments. The protocol requires verifiers to be expert in the given model being used, except in very specific circumstances, if safeguards are met. As we have made clear, this type of exception is likely to only be used in very specific circumstances. To reiterate, the DNDC model can only be operated by the world-renowned team at Dagan (an entity that currently hosts the model, holds licensing rights to its use, and whose members contributed to the development of the model). The team at Dagan will provide verifiers with a sensitivity analysis of model inputs, with model outputs, and demonstrate those model outputs have not been altered since Dagan created them. In these specific circumstances, it would be most appropriate for verifiers to focus on the veracity of model inputs, and it would add no real value for verifiers to themselves run the model again. Incidentally, please note we are not aware of a single offset program, or any regulatory program for that matter, that explicitly requires verifiers to obtain raw data and quantification tools, and themselves independently run such tools.

8.3.2 Verification of Soil Samples

32. Our May comment letter raised a number of concerns about the proposed sampling and verification practices in the draft protocol. Chief among them was our concern that the April draft protocol indicated that project verifiers would not conduct independent soil sampling. This is a significant problem because soil carbon cannot be inferred remotely. The only way to reliably calculate soil carbon content is to physically measure it, and we have concerns about the potential for biased or inaccurate measurement if there is no independent verification.

    The biggest change the Reserve made to verification standards in the August draft is to indicate that the question of independent soil sampling is to be left to the determination of the verifier, rather than simply not required. As we noted in our previous comment, placing the burden of choosing rigorous independent sampling on verifiers encourages a “race to the bottom.” With this new optionality, project developers seeking minimal oversight could screen project verifiers by asking whether those verifiers are willing to accept the developers’ soil samples.

    Similarly, there is no actual requirement that verifiers physically visit any project sites. All in-person verification visits can be explicitly waived at the Reserve’s discretion, if replaced by third-party attestations or remote visits instead. The August draft retains the loopholes and special permissions by which the Reserve can waive physical site visits on a case- by-case basis in the future, including for the protocol’s sponsor, Indigo Ag.

    Given the lack of independent soil sampling in the verification process, the ability of participating projects to avoid in-person verification site visits at the discretion of the Reserve, and the lack of third-party expert review of model calculations, we believe the protocol fails to function effectively as a third-party standard. (**CarbonPlan**)

    RESPONSE: Thank you for your comments. With respect to soil sampling requirements, we have included detailed guidance on all mandatory minimum requirements, including extensive guidance for determining sample units, stratification, sample location, resampling, site preparation, and sample depth. All of the approaches taken must be clearly spelt out in a detailed monitoring plan that will be verified. A randomized list of all sample points must also be retained for verification. Verifiers are required to use their professional judgment and expertise, and to study these requirements, the monitoring plan, and the evidence provided to demonstrate the requirements are met. With respect to verification requirements, we took feedback from multiple soil scientists, and verifiers with specific soil science expertise, in crafting these requirements. Verifiers will be equipped with model results, and direct measurement results, and can use their expert judgment to determine if specific verification activities are required in the given context. This is relatively common throughout many regulatory programs, as well as offset programs. With respect to site visit requirements, please note we created a list of every single verification activity we feel will be critical under this protocol, and then assessed whether site visits would add value in the specific context of every single verification item. Following this exercise, we deemed there may be circumstances under which site visits would not add value, over and above alternative sources of information. In the current pandemic conditions, we are all being asked to reassess preconceived notions that physically being present at our workplaces is inherently critical.

8.4.1 Verification Site Visit Requirements

33. The first item of the numbered list of this section requires that the verifier visit "a minimum of one-half the square root of the total number of fields in the project." While this is far more practical than a fixed percentage of fields when verifying an aggregated project with many fields, it still represents an undue burden. For example, a project with 5,000 fields would require 36 individual site visits, or 50 site visits for a project with 10,000 fields. While this sampling intensity would be entirely reasonable for an industrial project, where the conditions at the site (e.g., meters, piping, control equipment, etc.) are likely to remain unchanged throughout the year, and an individual site may represent a significant quantity of GHG reductions, it is overkill for a highly-dispersed agricultural project. Farming conditions and activities vary greatly throughout the year, such that the verifier is unlikely to witness much of the actual project activities through a visit on a single day. In fact, it is likely that actual site activities may be more easily determined through review of periodic remotely-sensed data. In addition, the GHG reductions for any given field are not likely to be material (hence the need for significant aggregation). Moreover, it could be argued that the main source of risk is the project aggregator's practices and procedures around data collection, quantification, and reporting, none of which occur on-farm. (**Indigo Ag**)

    RESPONSE: Thank you for your comment. While we recognize there are limits to the ability of site visits to be fully revelatory under project conditions and activities that are changing throughout the year, we maintain that it is nonetheless a reasonable default standard. We also provide the option for projects to seek approval for alternatives to physical site visits by verifiers, as outlined in Section 8.4.1, such as site visits by an independent third-party with agronomic expertise (e.g., local university extension service staff) or remote site visits. Lastly, we agree that a significant source of risk comes from the practices and procedures of project aggregators and have included verification guidance pointing verification bodies to review project documentation to ensure reported data are reasonable and accurate.

9 Glossary of Terms

34. We support the addition of biosolids to the definition of organic nitrogen fertilizer. (**CASA**)

    RESPONSE: Thank you for your feedback.

Model Calibration, Validation, and Verification Guidance for Soil Enrichment Projects

35. Based on our reasoning in Section A, we identified the following three major parts that the current CAR protocol is required to revise.

    1. Model validation is the only criteria by which a model's merit can be evaluated.

    The requirements for the model described at P73 - 74 in "*Soil Enrichment Protocol"* is far from sufficient. The quoted text provides the criteria of the CAR "*Soil Enrichment Protocol"*:

    "Models used to estimate stock change/emissions may be empirical or process-based, and must meet the following conditions:

1. Publicly available;
2. Shown in at least one peer-reviewed study to successfully simulate changes in soil organic carbon and, where modelling is used for non-reversible emissions impacts, trace gas emissions resulting from changes in agricultural management included in the project description;
3. Able to support repeating the project model simulations. This includes clear versioning of the model use in the project, stable software support of that version, as well as fully reported sources and values for all parameters used with the project version of the model. In the case where multiple sets of parameter values are used in the project, full reporting includes clearly identifying the sources of varying parameter sets as well as how they were applied to estimate stock change/emissions in the project. Acceptable sources include peer-reviewed literature and appropriate expert groups, and must describe the data sets and statistical processes used to set parameter values (i.e., the parameterization or calibration procedure, see guidance described in 5);
4. Incorporate one or more input variables that are monitored ex-post;
5. Validated according to the guidance contained in the external document titled Model Calibration, Validation, and Verification Guidance for Soil Enrichment Projects, using the same parameters or sets of parameters applied to estimate SOC/trace gas emissions in the project."

We think the above conditions could not enable fair and transparent model validation and model uncertainty quantification. We point out that model accuracy, characterized by "model uncertainty", plays the most essential role here, as it directly relates to the final estimated carbon credit from a model. The only way to quantify uncertainty of a model is through model validation. (**Guan et al.**)

RESPONSE: Thank you for your comments. We agree with the reviewers that the only way to quantify uncertainty is through model validation. For this exact reason, the document titled "Model Calibration, Validation, and Verification Guidance for Soil Enrichment Projects" lays out a clear framework for the validation requirements for a model, which must be passed for a model to be used. Model uncertainty quantification was intended to be a component of the model validation process described by the SEP Model Guidance document, but the most recent public comment version of the Model Guidance did not describe requirements clearly or adequately. This concern was addressed in a revision submitted to the Reserve dated 8/26/20 (following a process of external expert review and comments with S. Wood, B. McConkey, K. Paustian, Y. Zhang, and M. Easter), using language such as 'validating a model for performance and uncertainty', and including an end-to-end walkthrough (Figs 3.4.1, 3.4.2, and 3.5.1) of the proposed model validation process, assessing model accuracy as well as the proper setting of model uncertainty bounds (i.e. model prediction error) given observed data. The revised approach penalizes model imprecision and model inaccuracy through larger model prediction error. Model prediction error, in turn, penalizes credit generation by increasing the overall uncertainty of quantified credits. This was not clearly described in the previous version of the Model Guidance document and, we hope the revisions, address the above concerns.

36. We further highlight that model validation is the only judgement criteria of a model's merit. Any model to be used in a carbon credit system should publicly report its "uncertainty" in a reproducible format, which is derived from its validation performance benchmarked with a high-quality ground truth dataset following the standard model validation protocol. No exemption should be permitted for any model, even if it is widely used, peer-reviewed, or has a long history. **(Guan et al.)**

RESPONSE: Thank you for your comments. We agree. The overall objective of the SEP Model Guidance is to achieve this purpose, i.e., of only allowing a model judged as acceptable through a rigorous validation process to be used to issue credits under this protocol. The goal of the requirements and guidance described in the document is to put a framework around the demonstration of model performance in such a standardized approach that it opens up the right components for expert review and scrutiny. Additions were made in the introduction and throughout a revision submitted to the Reserve on 8/26/20 to clarify this intention.

Further, please note that the protocol requirement in Section 6.5 of a model that is "Shown in at least one peer-reviewed study to successfully simulate changes in soil organic carbon and, where modelling is used for non-reversible emissions impacts, trace gas emissions resulting from changes in agricultural management included in the project description;" is not meant to substitute for following the Model Guidance requirements and issuing a Validation Report for any model. Rather, this requirement is meant as an added layer of required scientific scrutiny, i.e., that modeling a project is not the 1st time a model has been used for this type of modeling exercise. All models, no matter their history, will need to issue an accepted Validation Report following SEP Model Guidance requirements to be used in an SEP project.

37. To enable such objective assessment, we strongly recommend developing and compiling an open-source and high-quality dataset through community efforts to make the model validation results transparent and intercomparable. **(Guan et al.)**

    RESPONSE: Thank you for your comments. This is an important concept and one that clearly would be to the benefit in many contexts, including validating a model for use under this protocol. The intention of the Model Guidance, as written, is to allow for use of such a resource- once approved by the Reserve - without requiring its existence. (Section 3.3 "Datasets can be drawn from a benchmark database maintained by a third party, if approved by the Reserve.") We hope that as more projects come in, we are able to contribute towards building such a resource, which can then be used for the benefit of all.

    We recognize the importance of such a resource; our approach was written with the intention to support its development. The Model Guidance document describes requirements on allowable data resources, but also, and perhaps even more importantly, requires that methods for data selection, aggregation, etc. are fully reported. In a revision submitted to the Reserve dated 8/26/20 (following a process of external expert review and comments with S. Wood, B. McConkey, K. Paustian, Y. Zhang, and M. Easter), summary boxes have been added regarding what is required in the Validation Report to make this more clear.

38. **Validation data needs to be the same standard data for all the models, to ensure apple-to-apple comparison.** Requirements for validation dataset described P8 - 10 in "*Model Calibration, Validation, and Verification Guidance for Soil Enrichment Projects"* did not require the same standard data for model validations.

    "Measured datasets must be drawn from peer-reviewed and published experimental datasets with measurements of SOC stock change (and annual measures of N2O and CH4 change if applicable) using control plots to test the practice category. All dataset sources must be reported.

Project developers are expected to use a process for selecting data for model validation that results in the assembly of validation datasets that are representative of the range of peer-reviewed observed results. Project developers must describe the methods, selection process, and data manipulations used to create the dataset applied in the model validation process. This includes describing search terms and databases used to identify available datasets, criteria used to select dataset sources, origin of extracted data (e.g., figures, tables, databases with DOI), original units of data and data uncertainty, and data manipulations used to convert original units into the units described above. The project developer should report the number of validation data measurements of each data type (SOC, N_2O and CH_4) for each project domain combination of practice category and crop functional group, and include a histogram showing the range of validation data values."

As we claim above, "any method should clearly report its uncertainty before its operational use"; the same applies to any proposed process-based models. Most importantly, we emphasize that model validation, a procedure to benchmark model simulation with independent and high-quality observational data, is the only way to quantify model uncertainty. (**Guan et al.**)

RESPONSE: Thank you for your comments. We agree and refer to the response to Comment 30 above for a more complete statement.

39. A reliable protocol for field-level soil carbon sequestration should include the following two aspects:

    (i) Which model variables should be validated?
    (ii) What qualifies as benchmark ground truth data for validation?

    For (i), we believe the soil carbon quantification in this protocol requires quantification of both carbon pools and fluxes of the agroecosystem at a field scale. Table 1 provides a minimum list of the carbon related variables for this purpose, as well as a high-level list.

    For (ii) about model validation benchmark data, to make the model validation results transparent and intercomparable, high-quality observational dataset should be compiled through community efforts. This dataset should ensure site representativeness to include different environmental conditions (e.g., climate, soil properties) and management practices (e.g., different tillage practices, cover crop uses), all at the field level. We should use this standard benchmark data and the same protocol to evaluate different models, and this derived uncertainty metrics should be reported. Thus, instead of debating which model is "better" or "worse", the most objective solution is to validate a model's simulation performance based on the benchmark data. Using this objective way to benchmark different models enables new models to join the available model list, and also motivate them to improve existing models.

    Any model to be used in a carbon credit system should publicly report its "accuracy" in a reproducible format, which is derived from its validation performance benchmarked with a high-quality ground truth dataset following the standard model validation protocol. No exemption is available for any model, even if it is widely used, peer-reviewed or developed by a reputable group or institute.

    It is worth noting that there have been several model intercomparison (MIP) efforts in the research communities for climate models (CMIP) (Eyring et al. 2016) and crop models

(AgMIP) (Rosenzweig et al. 2013), which set guiding examples for agroecosystem or soil biogeochemistry modeling efforts in agricultural carbon sequestration programs. It is also worth noting that the new SMARTFARM program by DOE ARAP-E is developing such a gold-standard and open-source data for benchmarking field-level soil carbon change and GHG emissions (DOE ARPA-E SMARTFARM Program 2020). **(Guan et al.)**

RESPONSE: We agree that (i) and (ii) are critical aspects of assessing a model. However, particularly regarding (i) we would like to draw attention to how 'validation' is defined for the purposes of the Model Guidance document, versus how the authors define validation. Specifically, the authors state:

> "Before we illustrate what a validation system should look like, we should understand the basic principle of validation - the validation data should always be split into two independent parts: in-sample and out-of-sample data (Hastie, Tibshirani, and Friedman 2009). In-sample validation allows one to calibrate and train a model with observations; then the calibrated model should be tested against the "out-of-sample" data. Whereas "in-sample" performance informs where improvements may be possible, we should note that "out-of-sample" model performance is what counts - this basic principle has been frequently violated in the literature, as many work mixed "in-sample" and "out-of-sample" model performance"

We strongly agree with this statement from the authors, particularly with the last sentence. It is for this reason that the Model Guidance document is designed to set requirements for a validation process focused primarily on what the authors are calling 'out-of-sample' validation, and what they describe related to (ii) above. What the authors call 'in-sample' validation the Model Guidance we suggest falls into the 'Model Calibration' section (section 2).

We recognize that modeling experts operationalize terms differently, and for this reason a definition section was added to the version of this guidance submitted to the Reserve on 8/26/20, to be more specific in how terms like 'calibrate', 'validate', 'model prediction error' and other are used in the Model Guidance document. The primary requirement of Model Calibration is that the Calibration datasets are demonstrably independent from those used to validate a model for use in a project under this protocol. An additional requirement for Model Calibration is that the process of calibrating the model is described and reported in the Validation Report (which will be made public after receiving independent expert review).

In the Model Guidance document, Model Calibration is intentionally left less prescriptive than Model Validation, recognizing different models have different structures, and the 'out-of-sample' model performance and model prediction error is of primary importance to verify that a model is appropriate for this protocol. For this reason, as well, Model Guidance focuses on validating model performance simulating changes in SOC and trace gas fluxes with the adoption of changed practices, since it is the accuracy and precision in simulating these changes that will determine credits. Validating the model to simulate SOC stocks is not currently required, since SOC measurement is a required input to use a model under this protocol. Again, this does not prevent the model calibration process from evaluating model performance in these aspects. We are supportive of the idea that a model calibration process (as defined in the Model Guidance document), would be inclusive of the all variables listed in Table 1, and see this as a worthwhile area to develop standards in alignment with the guiding examples. However, the Model Guidance document is currently designed to be inclusive of diverse calibration processes, as long as they are described, the datasets used are independent of Validation, and parameter sets generated via the calibration process are fully declared and properly Validated for use in a project using this protocol.

In the detailed statement for (ii) above, regarding the importance of validation and accuracy reporting please see Responses to Comments 29 and 30 above. Regarding the value of community databases of benchmark data please see Response 3. Representativeness of the datasets we will discuss in our responses to your further comments below.

40. **Model scalability should be addressed through the two-tier validation approach.** Requirements for model validation for the entire Project Domine is described at P10 in "Model Calibration, Validation, and Verification Guidance for Soil Enrichment Projects". The requirements below only show that a model could work at a few sites, which is not enough to show the model's capacity of scalability. The consistent performance with the accepted "accuracy" is required for any randomly selected site.

"Requirement 3: Validating a practice category / crop functional group combination for the entire Project Domain can only be completed if there are measurements of SOC stock and annual N_2O and CH_4 flux change (if applicable) that in total cover:

- At least three declared LRRs for projects within the US (or two IPCC climate zones per each required LRR for projects outside of the US, if and when the Reserve adapts this protocol for use outside of the US)
- At least three declared soil textural classes
- A range in declared clay amount per unit of soil spanning at least 15 percentage points"

We identify a major missing point in this CAR protocol - model scalability. In the current context, a method that works well at one or a few demonstration sites is not enough; the consistent performance with the accepted "uncertainty" is also required when applying to randomly selected sites. The current protocol has no discussion regarding how to ensure the model scalability. Again, instead of based on a model's history or reputation, we should design an appropriate protocol of model validation to address this requirement to test "model scalability". We provide a detailed pathway of how to conduct model validation (Figure 1), including to develop a two-tier validation system, and use community effort to develop open-source data to enable objective model validation, in particular, to test model's performance at random fields, which is the key metrics to determine the extent of model scalability.

Specifically, "model scalability" should not only be demonstrated by model performance at data-rich sites, where parameter calibration is allowed; a true test of "model scalability" should be also demonstrated at many random sites, where only limited measurements are available to truly test the transferability of the model. The latter is what a real-world application entails - we need to quantify soil carbon credit at any given field. Only models that can reproduce the accepted 'accuracy' extent at any random fields can be used as an accepted method for a carbon credit system.

Before we illustrate what a validation system should look like, we should understand the basic principle of validation - the validation data should always be split into two independent parts: in-sample and out-of-sample data (Hastie et al. 2009). In-sample validation allows one to calibrate and train a model with observations; then the calibrated model should be tested against the "out-of-sample" data. Whereas "in-sample" performance informs where improvements may be possible, we should note that "out-of-sample" model performance is what counts - this basic principle has been frequently violated in the literature, as many work mixed “in-sample” and “out-of-sample” model performance (Figure 3).

To achieve the above goal to fully validate the model scalability, we believe a two-tier validation approach is needed (Figure 3). Both tiers of validation data and the usage protocol should be prepared and used by the community, and these results should be reported to the community for fair and transparent comparison.

Tier 1: This tier includes sites that have collected a complete suite of measurements data and can be regarded as gold-standard sites. Example includes ARPA-E SMARTFARM Phase 1 sites (https://arpa-e.energy.gov/?q=arpa-e-programs/smartfarm), USDA LTAR sites (https://ltar.ars.usda.gov/, e.g. Kellogg Biological Station), NEON sites (https://preview.neonscience.org/), and some AmeriFlux sites at cropland and pasture land (https://ameriflux.lbl.gov/). Tier 1 sites enable detailed model calibration and out-of-sample validation. Usually, for the Tier 1 sites, it has the ability to measure whole ecosystem flux (e.g. NEE, GPP), soil carbon flux and stock, plant biomass etc.

Tier 2: This tier includes an extensive number of sites to test the model scalability performance. These sites in general only have limited amounts of ground measurements (for example, most sites may only have reported crop yield and SOC), but they represent the real-world situation for operational use. Validation can be directly made to compare the simulated crop yield, SOC stock and changes with observations. When doing model validation over tier 2 sites, only basic information about site location and management history will be provided, and the modeling team should report their simulation results for independent comparison with observations. (**Guan et al.**)

RESPONSE: Thank you for your comments. Please see the response to comment 33 above for a clarification of Model Guidance use of the term Calibration (what the authors call ‘in-sample’) versus Validation (what the authors call ‘out of sample).

We agree with the importance of model scalability, as well as the conceptual differentiation between Tier 1 and Tier 2 datasets. We are aligned with the authors in the value of Tier 1 gold-standard sites, and see them as particularly important for Model Calibration, as defined in the Model Guidance document.

For validating a model, as defined by the Model Guidance document, measured dataset requirements were intentionally set to be inclusive of Tier 1 and Tier 2 datasets, as long as they are independent from Calibration datasets (barring use of a statistical approach like k folding). We recognize that the amount of information reported by experimental studies varies, and in the revision to model guidance submitted to the Reserve on 8/26/20 (following a process of external expert review and comments with S. Wood, B. McConkey, K. Paustian, Y. Zhang, and M. Easter), requirements were included to report on information needed to run a model versus data provided in published studies (Section 3.3, Requirement 1, bullet 2). However, we did not set any requirements based on these differences. We are not opposed to the concept of Model validation requirements that treat the two types of data tiers separately, but see this as a requirement to consider longer term once the availability of datasets in the context of SEP Model Validation has been more extensively scoped.

Further, The Model Guidance document was written with the intention of ensuring scalability, but while also allowing the validation of models for use in small projects. These include:

- **Model Calibration (Section 2)** requirements for declaration of parameter sets for validation, designed to support the use of model parameters that are broadly generalizable, i.e., with the minimum land unit to vary parameter sets as 1 LRR. In the revision submitted to the Reserve on 8/26/20, guidance was included to clarify these requirements in the Summary Of Requirements for Section 2. The only exceptions may be petitioned for crop growth parameters (e.g., to vary maturity groups within an LRR).
- **Section 3.3** Requirement 2 (validating a model at a field level) and Requirement 3 (validating a model for an entire project domain) from the previous version of Model Guidance were revised to a single scalable **Requirement 2: Specific dataset requirements**. In the revised document, Requirement 2 sets minimum dataset requirements to validate a model that can be adjusted according to the size of the project for which the model will be used.
- **Section 3.3 Requirement 2** The dataset requirements are designed to set minimums for ensuring testing of generalized model performance, but exceeding these minimums is encouraged and incentivized via requirements around validating model prediction error. The revised text aims to expand and clarify these points.

We would like to note to the authors that the intention of the Model Guidance document is to provide a framework of following scientific best practices for Model Validation that allows an external expert to assess whether model simulations and model prediction error is appropriate for use in a SEP project. Where standardized frameworks do not yet exist, for example, benchmark datasets supporting direct model inter-comparisons for SEP projects, the Model Guidance is currently written to require transparency and reporting of methods. We hope that more integrated standards and benchmark datasets can be developed from the learnings gained by modeling groups issuing SEP model Validation Reports.

We thank the authors for the contribution of these comments and look forward to advancing SEP's Model Guidance with these goals in mind.

41. (1) SEP relies heavily on third-party models for reserve tonne estimation. The "peer reviewed journal" standard of acceptance is relatively low. Journals and their reviewers do not necessarily abide by similar publication standards nor have equivalent qualifications or lack of conflicts of interest. An independent third party committee ("Model Committee") be created by the Reserve to develop through a public process more explicit standards for developing experimental data, model development, model evaluation, and model application. Without standardizing this process CRTs and their innate values will invariably have much more risk. The risk of model failure and negative GHG change is non-zero. As such, a third risk term (R_{MF}: Risk - Model Failure) should be considered to be added to the risk pool that covers model risk. R_{MF} can vary depending on out-of-sample model performance metrics determined by the Model Committee. Levels could also be varied based on frequency of on-site monitoring and testing.

    (2) The models referenced in the SEP have been shown in research to produce both under and over estimation. Models inherently are built upon assumptions and data limited in spatial, temporal, and agronomic scope and resolution. Without robust out-of-sample testing and comparison against actual measurements, their accuracy, error, and appropriate application is limited to those assumptions and data from which they were built upon.

(3) These processes and the assumed GHG changes are not as easily modeled or verified via remote sensing as forestry. More frequent verification through on-site testing is necessary to assess the accuracy of the model and its application.

All of the prior noted risks are at least additive to each other and at worst multiplicative. (**Shibata**)

RESPONSE: Thank you for your comments. Through this protocol development process, we have spent extensive time developing Model Validation, Calibration and Verification Guidance for Soil Enrichment Projects, which contains much of the standards you are calling for. Contributors to that document include model experts, and expert soil scientists, both within and external to the workgroup. We will look to continue to refine these requirements through just this type of public expert process you advocate for. With respect to 'risk of model failure', we will take your feedback to those experts, to put that idea through a review process. With respect to testing and comparison of model results against actual measurements, this protocol does require the use of direct soil measurements, which must periodically be compared to model results. With respect to remote sensing, we have made it clear in the protocol that remote sensing will only be used to detect practice change, but will not be used to estimate GHG Impacts of any such practice change.