Nitrogen Management Project Protocol
Minimum Data Standard

Public Comment Package
December 2013

Table of Contents

Minimum Data Standard for Consideration in Quantification Methodology Development for the Nitrogen Management Project Protocol[1]

D.1 Introduction

As noted throughout the Nitrogen Management Project Protocol (NMPP), the Reserve plans to expand the list of project activities under this protocol as new data and quantification methodologies become available. The lack of field data on N_2O emissions for different regions, crops, and nitrogen management practices has been a significant limitation in the development of further quantification approaches, particularly a lack of data from "pairwise" or "side-by-side" comparisons (e.g. comparisons of baseline and project treatments on the same field in a given year). As such, this Minimum Data Standard provides general guidelines for establishing field experiments, and identifying existing data sets, that could be considered by the Reserve to support standardized quantification methods. The data are critical to methodology development, although many other elements are also required to develop and adopt a full methodology. Specifically, these field experiments could be used to develop, calibrate, and validate quantification methods.[2] These guidelines are referred to throughout the protocol as "minimum data standards."

D.1.1 Methodologies and Priorities for Future Protocol Expansion

The Reserve encourages field experiments and the development of reference data sets to support a variety of quantification approaches. Though the NMPP includes a Tier 2 quantification methodology (e.g. using standardized region-specific emission factors to quantify emission reductions from the project activity[3]), future expansions of the NMPP may incorporate other kinds of approaches and methodologies.. The Reserve has not made a determination of preference between Tier 2 and Tier 3 methods (e.g. higher order quantification methods, such as validated biogeochemical models or comprehensive field sampling). Robust yet simple regional Tier 2 emission factors may be better suited for cropping systems that cover large areas, have management practices that are fairly homogenous, and that are grown in relatively simple rotations. Examples of such cropping systems are rain fed corn systems (included in Version 1.0 of the NMPP), irrigated corn systems, or wheat cropping systems. Tier 3 approaches, including validated biogeochemical models, may be preferred for specialty crops, such as vegetables or fruit, for which the management often varies and that are grown in more complex rotations.

Comment [T1]: PUBLIC COMMENT: On an ongoing basis, the Reserve plans to revisit the relative strength of Tier 2 and Tier 3 methods, as well as hybrid approaches such as the USDA N_2O method. Are there recently published studies and/or methodologies which you believe the Reserve should consider? Are you aware of studies underway and/or methodologies under-development that the Reserve should be tracking?

While the main purpose of the minimum data standard is to provide guidance for field-level experiments resulting in reference data sets, reference data sets will also be reviewed by the Reserve to determine whether they are appropriate for developing a Tier 2 methodology, for validating a Tier 2 methodology, or for calibrating and validating a Tier 3 methodology (e.g. DNDC). In addition to the data sets themselves, stakeholders are encouraged to develop and

[1] The Minimum Data Standard originally appeared as Appendix D of the Nitrogen Management Project Protocol, both versions 1.0 and 1.1.
[2] For a discussion of calibration and validation of models, please see the "C-AGG White Paper Uncertainty in Models and Agricultural Offset Protocols," available at: http://c-agg.org/cm_vault/files/docs/temp_file_C-AGG_Uncertainty_White_Paper_7-5-121.pdf
[3] As defined by the Volume 4 of the 2006 IPCC Guidelines for National Greenhouse Gas Inventories.

submit new Tier 2 or Tier 3 quantification methodologies, developed from reference data sets meeting this standard.[4]

D.1.2 Process for Future Protocol Expansion

The Minimum Data Standards, presented originally as an appendix and now available on the NMPP website, will serve as a framework for the Reserve to determine whether reference data are sufficiently robust. Stakeholders are encouraged to submit new reference data sets and quantification methodologies to the Reserve at any time, and the Reserve will review new data submittals on an ongoing basis. Information on this submittal process is available on the NMPP webpage. Stakeholders should complete an NMPP New Data Submittal Form, which will collect the necessary information for the Reserve to assess whether the data set meets the minimum data standards. Later in the process, the Reserve will need access to the reference data set itself.

Comment [T2]: PUBLIC COMMENT: Please note: the New Data Submittal Form and a document describing the Data Submittal Process are both included in this Minimum Data Submittal packet currently out for public comment.

At minimum, reference data sets submitted to the Reserve must meet these requirements, in terms of study design, temporal frequency of sampling, and other such criteria. Meeting these requirements ensures that data adhere to a minimum quality threshold and were derived using the type of experimental design that is critical for data to be useful. However, meeting minimum requirements does ***not*** guarantee the Reserve will be able to use a data set. The Minimum Data Standard also includes a discussion of "best practice," in an effort to encourage researchers to develop more robust data sets. Implementation of best practices is not required but is strongly encouraged. Following best practices may significantly increase the chance that that the Reserve can adopt a methodology using the collected data, because they can enhance statistical power and better address confounding factors that could influence results (e.g., unusual weather events, insufficient sampling to miss hot spots of N2O emissions in space or time). Best practices should be viewed as guidelines for designing experiments and collecting and analyzing data that will improve the chances of generating a robust data set for inclusion in the NMPP, based on the educated judgment of highly qualified research scientists and Reserve staff.

When the Reserve reviews data sets for consideration of inclusion in the NMPP, the Reserve will be evaluating the quality of that dataset and the ability to develop a quantification methodology from it. As needed, the Reserve will consult with scientists who have previously been involved in the NMPP Science Advisory Committee and other scientists who are active in agriculture, GHG emissions, and ecosystem science, who will assist the Reserve's in evaluating submitted data sets.

It should be noted that the Reserve will consider other criteria, in addition to the relative strength of a submitted data set, before a management practice can be included as a project activity in the protocol. In particular, the Reserve must conduct a thorough assessment of:

- Whether sufficient data exists to develop a performance standard and preliminary assessments show a project activity is likely to be additional
- The economic and technical feasibility, as well as the mitigation potential, of the management practice that reduces N_2O emission under consideration
- The existence of baseline N_2O emission measurements for the practice, region, and/or cropping system considered; and

[4] Other than the fact that the underlying data for new Tier 2 and 3 methodologies should meet the requirements of the Minimum Data Standard, it is outside the scope of this document to address minimum standards for quantification methodologies.

- The total acreage and intensity of use of nitrogen fertilizer for the cropping system in question;

If a practice is determined to be non-additional, or there is very little information on trends of practice adoption, the Reserve is unlikely to move forward with quantification methodology development, regardless of how strong a quantification data set is. The Reserve does not require that data submittals include information on the other factors that will be part of the Reserve's comprehensive review of a candidate practice. However, researchers are encouraged to consider such criteria prior to developing an experiment, and whenever possible, stakeholders are encouraged to submit such complementary data along with their New Data Submittal Form, as doing so will allow the Reserve to expedite consideration of whether a new quantification methodology belongs in the NMPP.

Once the Reserve identifies reference data sets and associated methodologies appropriate for use in expanding the protocol, the Reserve may decide to contract for additional expertise and/or reconvene a stakeholder workgroup to support the protocol revision. As with any new project type, once the new project type has been developed and included in the protocol, the protocol will be released for a 30-day public comment period before the revision is considered for adoption by the Reserve Board.

D.2 Minimum Data Standards for Field Experiments

The minimum data standards apply to the reference data collected in field experiments and used to develop, validate or calibrate new or existing N_2O quantification approaches, including both statistical models and process-based biogeochemical models. Reference data is defined as, "field-level experimental data generated from new or previously completed experiments that comply with the minimum data standards, the results of which have been published[5]". A citation for the original publication of the experiment in which the data was generated must be provided for existing data.

Comment [T3]: PUBLIC COMMENT: The Reserve is seeking stakeholder feedback on this revised definition of reference data, which now allows for submittal of data that has not been published in peer-reviewed journals, as long as other criteria in the Minimum Data Standard are met. Should the Reserve accept datasets published in both peer-reviewed and non-peer reviewed research? Do you think allowing non-peer reviewed studies maintains an appropriately high bar for high quality data? Other than the criteria in the minimum data standard, are there other additional criteria that the Reserve should consider a minimum filter for judging data quality in a case like this?"

D.2.1 Method of Data Collection

Requirement:

1) Reference data should be collected using either chamber-based or tower-based (micrometeorological) methods.[6]
2) Reference data should be collected using sample collection methods consistent with methods commonly cited in the peer-reviewed literature or methods described in standard soil methods manuals available at the start of the experiment(s) yielding the reference data. More specifically, the Reserve requires that data be collected in a manner consistent with one or more of the following sets of guidelines: the USDA Agricultural Research Service (ARS) GRACEnet Chamber-based Trace Gas Flux

[5] While the Reserve requires that reference data sets have been published, the Reserve will evaluate research and reference data sets that have been published both in scientific peer-reviewed journals, as well as those that have been published in other non-peer reviewed papers.

[6] Though both methods are acceptable, chamber-based methods are currently the least expensive option for measuring N_2O emissions from agricultural fields, as the materials required for building the chambers are very affordable, and analytical tools used for N_2O concentration measurements, such as gas chromatography, have become omnipresent in analytical laboratories. Tower-based methods (micrometeorological techniques) to measure N_2O emissions have been developed and have the advantage of being non-intrusive while providing continuous time series. Nevertheless, high investment costs make their use in replicated experiments somewhat less attractive at present.

Measurement Protocol[7], the Soil Science Society of America Methods of Soil Analysis[8], the Canadian Society of Soil Science Soil Sampling and Methods of Analysis[9], the Long-Term Ecological Research Standard Soil Methods[10], or Methods in Ecosystem Science[11]. If an alternative protocol or combination of protocols is used, the data set submission will need to provide a copy of the sampling protocol and a comparison of how it differs from the required protocols.

Comment [T4]: PUBLIC COMMENT: Previously, the Reserve required that data collection follow the GRACEnet Chamber-based Trace Gas Flux Measurement Protocol, which is primarily used by ARS scientists at present. However, the Reserve is proposing broadening this requirement to allow for sample collection methods that are consistent with methods published in the books listed at left. The Reserve is seeking comment on whether the protocols now included are sufficiently robust for inclusion. Are there any additional protocols that are equally robust that the Reserve should consider for inclusion?

D.2.2 Intensity of Data Collection

Due to the high spatial and temporal variability of N_2O emissions, accurate N_2O quantification necessitates a minimum temporal and spatial intensity of data collection, as outlined below.

Guidelines on how to handle flux estimates classified as outliers are included in Section D.2.3.

D.2.2.1 Spatial Frequency and Scale of Data Collection

N_2O emissions are subject to high spatial variability that occurs at multiple geographic scales, including within a field, across fields within a landscape, and across landscapes (e.g. a Land Resource Region or a Major Land Resource Area). Large variability among field measurements contributes to uncertainty around predicted emission reductions for any quantification approach. Independent reference data should be gathered from a sufficient number of sites so that reference data can be divided into separate calibration and validation data sets. If calibration data are collected primarily from one area within a larger region (such as a Land Resource Region), an extensive validation data set, including data from other areas within the region, might allow validation of the model for a much larger geographic area.

Comment [T5]: PUBLIC COMMENT: At present, the term "region" is not defined. Is it important for the Reserve to define region within the MDS? Or is a more flexible interpretation preferable (allowing regions to be defined as methodologies are developed)? One proposed definition of region would allow the data submitter to propose the region, based on how far they believe a reference data set could be applied. Would this be an appropriate definition? Or should the Reserve consider USDA regional definitions, Land Resource Regions or some other well recognized definition?

In this section, guidelines are provided to ensure that the reference data account for spatial variability at different scales. The terminology for "field," as defined in Section 2.2.1 of the NMPP, is different from the terminology used in the design of agricultural experiments, in which a field represents a random variable and may encompass multiple plots with different treatments, which are defined as the suite of management practices implemented on a field. In these guidelines, the Reserve uses "replicate plot" to refer to the smallest experimental unit and "field" or "study site" to designate a greater unit with multiple replicate plots. Within each plot there are many functional locations, which the Reserve defines as, "areas of a field that are biologically or physically distinct from other areas in ways that likely relate to different N_2O emissions, and that lend themselves to separation for purposes of trace gas measurements. Examples of functional locations include ridges or hollows, areas where fertilizer has been applied in bands and areas where fertilizer has not been directly applied, areas immediately adjacent to irrigation lines and areas that are generally less well-irrigated, tractor furrows, and tree rows."

[7] Parkin, T.B., & Venterea, R.T., 2010. Available at www.ars.usda.gov/research/GRACEnet.
[8] Methods of Soil Analysis. Part 1 and Part 2. Microbiological and Biochemical Properties (Soil Science Society of America Book, No 5) (Soil Science Society of America Book Series)
[9] Rochette, P. and N. Bertrand. 2008. "Soil Surface Gas Emissions", in Soil Sampling and Methods of Analysis, 2nd ed. Eds: Carter, M.R. and E.G. Gregorich. pp 851-853.
[10] Holland, E.A., G.P. Robertson, J. Greenberg, P.M. Groffman, R.D. Boone, J.R. Gosz. "Soil CO2, N2O, and CH4 Exchange", in Standard Soil Methods for Long-Term Ecological Research. Eds: G. P. Robertson, D.C. Coleman, C.S. Bledsoe, P. Sollins. pp 185-201.
[11] Matson, P. and A Goldstein 2000. Biogenic trace gas exchanges. Chapter 15, pp. 235-248 in Sala, O., R.B. Jackson, H.A. Mooney and R.W. Howarth, Eds. Methods in ecosystem science, Springer-Verlag.

The spatial frequency and scale of data collection must adhere to the following guidelines:

Requirement:

1) **The dimensions of the flux chambers**: The surface area covered by the flux chamber should be large enough to capture small-scale variability in N_2O fluxes (e.g. distribution of fertilizer granules in the soil, presence of decomposing crop residues, pattern of macropores or earthworm casts in the soil, etc.). Surface area for an individual chamber typically covers between ~300 and ~3000 cm^2.
2) **The number of flux chambers within a replicate plot**: At minimum, at least two flux chambers are required within each replicate plot. Where possible, flux chambers should be placed such that all major functional locations are represented (See best practice guidelines for additional information on stratified sampling).
3) **The number of replicate plots per field / study site**: The reference data should cover a minimum of 3 replicate plots per treatment (i.e. management practice) and per field. For a side-by-side ("pairwise") comparison, there should be at least two treatments, with one treatment representing the baseline scenario and one treatment representing the project scenario. The reference (or "baseline") scenario should represent typical or conventional management practices (e.g. business as usual) and is distinct from a scenario that attempts to capture background emissions under minimal management.
4) **The number of fields / study sites, replicated at different geographic locations**: The field(s) should be representative of the conditions within the area in which the reference data sets will be used, and over which the data submitter proposes the reference data set be extrapolated. Therefore, multiple fields are to be used that are located at different sites and locations (e.g. different counties, different states). Ideally, the fields are also chosen to represent some of the most commonly occurring soil types in a region. However, having multiple sites may be challenging. If a data submittal fails the minimum field requirement, submitters will be required to submit additional information to assist the Reserve in assessing the potential to extrapolate the data set to other areas. This information shall include:
    a. A list of comparable experiments in similar climates and soil types that submitters believe can be utilized for a data comparison. Data and analysis products from these sites need to be robust and also adhere to the minimum data standard.
    b. A list of locations submitters would consider suitable for comparative field studies to augment their data set if resources were available.[12] This can be submitted in conjunction with Requirement 1 in section D.3.

Best Practice:

1) **Number of flux chambers per functional location and number of functional locations within a replicate plot**: In many cropland systems, multiple functional locations can be identified within a plot. It is recommended that sampling design be stratified, placing flux chambers at multiple functional locations within each plot so as to reduce estimates of error around N_2O emission estimates by taking advantage of understanding variation across the field in factors such as soil moisture, temperature and N concentration, all of which influence N_2O emissions. Under a stratified sampling approach, the Reserve encourages at least one chamber per functional location on

Comment [T6]: PUBLIC COMMENT: Currently, there is no strict requirement on minimum or maximum chamber size/surface area. If this range was made a requirement, would this be a reasonable range of acceptable chamber sizes? What are the sizes of a typical chamber used for these experiments?

Comment [T7]: PUBLIC COMMENT: Currently the Minimum Data Standard requires at least 3 replicate plots and 2 chambers per functional location, but the Reserve is considering revising these minimum standards, as proposed at left.
1) Would you recommend that the Reserve revise these standards as proposed?
2) Would you recommend that the Reserve require a minimum of 2 chambers per replicate plot or 2 chambers per functional location (i.e. furrow, side dress, beam) within each replicate plot, which would then potentially require many more chambers per replicate plot, field or treatment?
3) Would you recommend that the Reserve adjust current or recommended requirements based on the size of the chambers being used within the experiment, with different requirements for studies using large (~3,000 cm2) vs. smaller (~300 cm2) chambers? What would be the advantages and disadvantages of establishing minimum acceptable cumulative surface area under chambers (within a plot)? If so, what would be an appropriate value?

Comment [T8]: PUBLIC COMMENT: How much additional confidence is gained as each additional site location (i.e. a different geographic region not an additional field at an existing study location) is added in an experimental design? Would you recommend that the Reserve consider a minimum number of replicate study sites at different locations? If the Reserve were to allow for data submittals that did not include multiple study sites, what is the potential for the experiment to overestimate emission reductions?

[12] Please note, this information is for reference only and for further defining the "region" over which a methodology, based on the reference data, might be applied. The Reserve is not in the position to provide funds for such research.

major functional locations within a replicate plot. In calculating total fluxes across a replicate plot, the proportional area that each functional location covers across the field needs to be taken into account.

2) **Number of treatments per study:** At minimum, a side-by-side ("pairwise") comparison, includes two treatments, with one treatment representing the baseline scenario and one treatment representing the project scenario. However, implementing and monitoring more than one potential project treatment is encouraged, so as to collect data on a wider variety of project activities. Any number of potential project activities could be implemented together as the "project treatment" on a given field (e.g. add nitrogen inhibitors, add a cover crop, trial of different N rates, or N rate reduction with the addition of cover crops).
3) **The number of replicate plots per field:** Although the Reserve will consider data sets based on 3 replicate plots per treatment per field, at least 5 replicate plots per treatment per field is encouraged.
4) **Preliminary data and power analysis.** Ideally, prior to commencing a study, preliminary data should be used to estimate variability among functional locations and among randomly-placed chambers, to inform the number of plots needed in advance of full-scale design and implementation. The Reserve recommends that submitters conduct a power analysis to estimate how many plots and chambers would be likely to enable them to detect differences in N_2O emissions of varying magnitude. The Reserve suggests doing so by using data from analogous studies from other sites, drawing on submitters' own previous research or by collaborating or contracting with other research groups who have such data and/or analysis capacity.

D.2.2.2 Temporal Frequency and Scale of Data Collection

In addition to high spatial variability, N_2O emissions are also subject to variability over time.

Requirement:

1) **Flux measurement frequency:** Flux measurements should, at minimum, be collected at least once per week (every seven days). Following fertilization and first irrigation, the Reserve requires that flux measurements be collected twice a week for 30 days to capture potential variability in N_2O emissions.
2) **Time of day for flux measurements:** Measurements also should represent the daily variations in N_2O fluxes. Multiple flux measurements could be made during one day. However, one flux measurement collected per day is acceptable, so long as the research team has sufficient background data for the site to characterize emissions over daily cycles and therefore extrapolate daily fluxes from single measurements per day. If measurements are limited to one per day, they should be collected at a time that avoids minimum or maximum daily temperatures; mid-morning or early evening is recommended.
3) **Duration of flux measurement collection:** Flux measurements must be collected at a minimum over the complete duration of a growing season and extend over at least two consecutive growing seasons. In regions where there is a freeze/thaw cycle and/or winter precipitation, one complete year (12 months) of flux measurements is required to account for N_2O emissions around freeze/thaw cycles and winter precipitation. Where reference data sets do not meet this full year requirement, a list of comparable experiments in the region, with similar climates and soil types that did collect year-round flux measurements may be acceptable for supplementing non-growing season data. If no year round data are available from other experiments, the Reserve will consider using

Comment [T9]: PUBLIC COMMENT: There are a number of possible approaches, and the Reserve would appreciate consideration of whether the proposed (or other) approaches are more appropriate as Best Practice. An alternative approach would be to use 1 chamber per functional area and use more plots. Another would be to use 3 chambers per functional area in some plots, and 1 per functional location in other plots. Please comment on the relative advantages and disadvantages of these approaches. Further, please comment on how the size of the plot and/or the size of a chamber might be taken into account when determining the size of a plot.

Comment [T10]: PUBLIC COMMENT: Would you recommend that the Reserve retain this newly proposed best practice in the MDS? Would a more explicit recommendation be helpful? For example, the Reserve might "encourage investigators to collect preliminary data from at least 4 chambers per functional location in a test plot and from at least 10 chambers deployed at random across the field at one point in time." Aside from providing funds, what could the Reserve do to make it easier for submitters to follow this recommendation?

Comment [T11]: PUBLIC COMMENT: Do you think this is a reasonable minimum standard, recognizing that the Reserve will also encourage more frequent sampling as best practice? Is this too stringent, such that it will leave quality data sets out of the running?

these data sets for validation purposes on a case by case basis, but they will not be used to develop new quantification methodologies.

4) **Addressing gaps in data:** Due to unforeseeable weather conditions, issues with measurement devices, and other challenges, some gaps in the data set may be unavoidable. If gaps in a data set do arise, gaps in data collection must be clearly identified, in both text and any graphs included in the New Data Submittal. Graphs should make very clear where gaps in data collection occurred. Submitters should also explain how they interpolated estimates during gaps (e.g., linear interpolation between adjacent two measurement times, some kind of model).

Comment [T12]: PUBLIC COMMENT: Many field experiments measuring GHG fluxes in agriculture focus on growing season emissions only; however the Reserve is concerned that in regions with freeze/thaw cycles or winter precipitation, significant N_2O emissions resulting from these events may not be captured with those sampling designs. Is it reasonable to require year-round measurement in these regions? Is one full year of "typical" weather conditions sufficient to get a sense of the non-growing season flux? If so, how should "typical" be defined (e.g. average temperatures +/- a certain percent?) For studies that did not collect year-round data, can comparable experiments within the region (with similar climates and soil types) that did collect non-growing season data be used to supplement the growing season data of the primary study of interest?

Best Practice:

1) **Flux measurement frequency:** It is strongly advised to increase the measurement frequency to daily flux measurements following any agronomic or environmental event known to be associated with major N_2O fluxes (e.g., fertilization (discussed above), tillage, irrigation, rain, or harvest) including any events that may result in crop damage (e.g., major insect damage). Daily flux measurements after such events should continue until N_2O emissions return to pre-event levels (i.e. up to 60 days after the event).
2) **Year-round flux measurement collection:** Year-round flux data are preferable in all regions (as described above) and two full years of data are encouraged.
3) **Anomalous weather during a study:** Flux measurements over additional growing seasons will be necessary if the two consecutive growing seasons for which measurements were collected exhibited anomalous weather conditions, with respect to that region, or if the two years had very similar weather patterns, making extrapolation to other common weather conditions at that site challenging.

D.2.3 Outliers and Analysis

When experimental data are collected, some samples may have values that are very large or small compared to most other samples from that treatment. Such samples are often referred to as outliers, and can occur at different points in the data collection process. Removal of outliers at different points within the experimental process can have significant impacts on annual estimates of N_2O fluxes from the field to the atmosphere.

The Reserve distinguishes between outliers within three stages of the experimental process:

1) Data from individual gas samples collected from chambers over a time course used to estimate the rate of N_2O emissions from soil in the chamber[13]. These outliers are generally handled according to standard protocols within each research laboratory.
2) N_2O flux rate data aggregated at each chamber at each sampling date and time as a result of a set of individual vials from the specific date and time. Methods for retaining or excluding these outliers vary considerably, and how these values are treated is of interest to the Reserve.
3) N_2O flux rate data averaged across multiple chambers in a plot or treatment on a sampling date and time. It is unusual for values to be excluded at this scale of analysis but it is of interest to the Reserve when it occurs.

[13] For a given chamber, gas samples are collected in vials over approximately 20 minutes, and the concentration of N_2O in the chamber – and hence in the vials -- is expected to increase linearly over that time. Sometimes, values from 3 or 4 vials will appear linear and one will appear to bear little or no relation to the other vials. Such outliers can be caused by inadequate closure of flux chambers, leaky sampling vials, and errors in sample collection or analysis. Laboratory protocols often call for removal of these outliers in a routine and standardized fashion.

Deciding how to handle these second and third types of outliers is more challenging. A rate of N_2O emission from the soil is calculated for each chamber, and these flux estimates are sometimes much higher than other chambers established in a similar treatment or plot. In some cases, laboratory protocols call for excluding any value that is more than, say, 3 standard deviations from the mean, but it is difficult to justify that approach. N_2O emissions from soil are known to be highly variable, and "outliers" may simply reflect this variability either in space or in time. As such, outliers may not be removed arbitrarily during this stage of experimentation, but rather must be justified with a rationale based on physical factors (such as having reason to believe a chamber was placed in a very unusual area of the field and hence does not represent the area it was intended to represent, or due to equipment malfunction).

In order to understand the cumulative effect of removal of data the Reserve makes the following requirements.

Requirements:

1) Data submittals must include a statement clearly explaining the protocol they utilized and how researchers analyzed their data, including a description of how they treated outliers at each of the previously mentioned stages of data collection and analysis. All types of outliers should be included in this statement including both positive and negative outliers.
2) Stakeholders submitting data that successfully pass through the first round of Climate Action Reserve data review will be required to submit additional analysis on the relative effect of the removal of outliers on their data. Specifically, stakeholders will also need provide estimates of annual N_2O fluxes calculated with and without the removal of stage 2 and stage 3 outliers from their data, and they must discuss the pattern of outliers and justification for removing those points. Discussion must include the % of total chamber-based flux estimates removed, whether outliers occurred always at the same chamber location or were concentrated on particular sampling dates, and whether the researchers observed features of the landscape or conditions that likely explain these extreme values.

The Reserve will not dismiss a data submittal for removing outliers if removals are justified and explained, and/or the removals do not result in significant changes (+/- 10%) to the annual N_2O flux.

Comment [T13]: PUBLIC COMMENT: Do you recommend any changes in this general approach to responding to submittals from which outliers have been removed? How large a change in N_2O flux estimates, resulting from removal of outliers, would you consider significant? Is +/- 10% appropriate?

D.3 Applying Understanding from Field Experiments to Regions

In order to assess the extent to which the submitted data set is able to be extrapolated within and to other regions the Reserve seeks additional information about the relevance of the experiment to the region.

Comment [T14]: PUBLIC COMMENT: The Reserve is seeking comment on which attributes (that can be easily estimated across large landscapes) have the most influence on N_2O emission fluxes. What type of analyses might the Reserve recommend to extrapolate from a single site (or very few sites) to the larger region? What level of error is introduced as one extrapolates data further away from a core research site? Do Tier 3 models offer a sound method that the Reserve could use to evaluate the distance across which a data set could be extrapolated with confidence?

Requirements:

1) Stakeholders will be asked to propose and justify a geographic "application region" over which a data set (or the subsequently developed quantification methodology) could be extrapolated. The Reserve recommends that the justification include a comparison of weather and climate, soil characteristics, and management practices between the study sites and the application region. This can be submitted in conjunction with Recommendation 4B in Section D.2.2.2.

2) Summaries of growing season and experimental conditions during the field trials should be included along with a discussion of whether representative conditions (e.g. temperature, precipitation, etc.) were “typical” or “average” for that region. For example, a comparison of the experimental growing season(s)’s mean annual temperature and precipitation data to data collected over the preceding ten year period could indicate whether N_2O emissions measured for the period are representative of a “typical” year, or rather a cold, hot, wet or dry year.
3) Summaries of soil properties including soil type, soil texture, soil water holding capacity, soil organic carbon levels, for the experiment should be documented and described. The relationship of these site attributes to a region should be considered when selecting replicate plots and fields for inclusion in an experiment. Sites should be chosen so that results can be applied to the widest possible area within the region. Choosing sites that include fields with multiple soil types, for example, could significantly increase the number of fields to which results could be reasonably extrapolated. Likewise, the management practices executed on the field trials should be selected to represent the largest possible area within the region; if conservation tillage, drip irrigation, and manure-spreading are all widely used, then including all these practices, as seen on working farm fields in the region, as baseline cases in the experiments would be helpful.

In addition to these criteria relating to the applicability of the field experiment to a broader region new data submittal forms will also require information related to the study location, cropping systems, management practices, yields and other data.

Comment [T15]: PUBLIC COMMENT: New data submissions must show that growing season climate and experimental conditions at the study site are representative of “typical conditions” in the region. The data submittal form notes that key attributes to be considered, at minimum, include temperature and precipitation. What other attributes should be considered to ensure that environmental conditions during the study represent generally-occurring conditions in the region?

Comment [T16]: PUBLIC COMMENT: New data submissions must show that soil properties are representative of “typical conditions” in the region, by discussing attributes such as soil type, soil texture, soil water holding capacity, soil organic carbon levels, etc.… Are there any other attributes that should be considered to ensure that soils at the study site represent generally-occurring soils in the region?

New Dataset Submittal Form, Nitrogen Management Project Protocol

A. Information on Parties Submitting Data

1. Name:
2. Title:
3. Organization:
4. Phone:
5. Email:
6. Other involved parties:
7. If you are submitting a data set that you were otherwise not involved in developing, do you have consent from the researchers to submit their data to the Reserve?

B: Basic Information on Dataset

1. Reference data sets must be published. Please provide complete citations for any publications in which the dataset being submitted has been published. If you are not listed as one of the contributing authors for the study, please explain your relationship to and/or interest in submitting this dataset.
2. Where did the study take place? Please provide information on the location of each study site (MLRR or LRR, state, county, city, etc.)
3. Please list the years in which data were collected, noting in which years data were collected year-round and in which years data were collected during the growing season only.
4. Please describe the crop type and/or rotations, including *all* crops in the rotation, cover crops, and fallow years
5. Briefly describe the proposed activity or practice change that the dataset demonstrates will consistently result in N2O emission reductions. Does this activity potentially impact other GHG emission sources, sinks or reservoirs? In which sources, sinks, or reservoirs would GHG reductions occur?
6. Are you submitting this data set for purposes of developing a new Tier 2 methodology? Validating an existing Tier 2 or 3 methodology?

C: Management descriptors (Please describe the following for each treatment evaluated)

1. Tillage practices (conventional till, reduced till, strip till, conservation till, no till, etc.)
2. Rate (amount) of nitrogen applied
3. Types of nitrogen fertilizer applied (please include all, including both organic and inorganic sources)
4. How nitrogen was placed and applied (banded, broadcast, injected, etc.)
5. Timing and frequency of nitrogen application
6. Irrigation type, application frequency and amount
7. Residue handling (left on field, removed from field, including percent remaining incorporated into field, etc.)

8. Other inputs used (i.e. starter fertilizer, phosphorus, etc.)
9. Impact on yields (please describe what, if any, impact the project activity has had on yields. Was any other agronomic event potentially responsible for external impacts on yields over the course of the study?)

D. Study Design Spatial and Temporal Sampling

1. Please provide the following information on the spatial design for each study site:
    i. Number of replicate plots
    ii. Size of replicate plots
    iii. Number of chambers per plot
    iv. Are chambers randomly distributed? Or do they follow a method of stratified sampling? If sampling is stratified, which and how many functional locations were examined?
    v. Is flux data collected via chamber-based methods, tower-based (micrometeorological) methods, or other methods?
    vi. Does the dataset use a sample collection method consistent with one of the following soil methods?
        - USDA ARS GRACEnet Chamber-based Trace Gas Flux Measurement Protocol; the Soil Science Society of America Methods of Soil Analysis
        - Canadian Society of Soil Science Soil Sampling and Methods of Analysis
        - Long-Term Ecological Research Standard Soil Methods
        - Methods in Ecosystem Science.
        - No, the dataset is not consistent with one of the methods above. (Please provide a copy of the sampling protocol implemented and a comparison of how it differs from the required protocols.)
    vii. Please provide a brief description of methods of data collection, including chamber design, sample collection and handling, and gas analysis.
    viii. (Optional) Please provide a visual representation or diagram of the chamber and experimental design.
2. Please provide the following information on the temporal frequency of measurements taken at each study site:
    i. How many times per week were flux measurements typically taken during the experiment?
        a. Were measurements taken with increased frequency after fertilization? If so, what was the increased frequency and for how long were measurements taken at this increased frequency?
        b. Were measurements taken with increased frequency after first irrigation? If so, what was the increased frequency and for how long were measurements taken at this increased frequency?
    ii. At what time of day were flux measurements taken? Were multiple flux measurements taken daily?

E. Location descriptors

1. Please describe the soil characteristics of the study site, including soil texture or series, soil organic matter content.
    i. Is this soil typical of the region?
    ii. If not, what soil is typical of the region?
2. Please describe the weather of each study site:

  i. What is the precipitation amount and distribution for the study site (or closest weather station) in the years of the study?
  ii. What is the average annual precipitation for the study site (or closest weather station)?
  iii. Does weather during the study years represent weather "typical" of the region? If not, please discuss why.
  iv. Is the study site located in a region that experiences freeze/thaw cycles?
    a. If yes, does the dataset include at least one full year of flux measurement data?
    b. If no, please provide a list of comparable experiments in the region that did collect year-round flux measurements.
  v. Is the study site located in a region that experiences non-growing season precipitation (and lacks freeze/thaw cycles)?
    a. If yes, does the dataset include at least one full year of flux measurement data?
    b. If no, please provide a list of comparable experiments in the region that did collect year-round flux measurements.
3. Describe any notable environmental and/or agronomic events that occurred during the experiment and indicate if they have influenced GHG emissions or resulted in crop damage

F. Treatment of Outliers

1. In the Minimum Data Standard, the Reserve distinguishes between outliers at three distinct stages of the experimental process. Please describe how outliers were treated at each of stage of data collection and analysis in the study (as discussed in the minimum data standard). Please address all types of outliers, including both positive and negative outliers.

Technical Questions to Inform the Minimum Data Standard for the Nitrogen Management Project Protocol

The Reserve is seeking feedback on the following questions to inform the draft Minimum Data Standard out for public comment. The Minimum Data Standard formerly appeared as Appendix D in the Nitrogen Management Project Protocol. In the future the Minimum Data Standard will live on the NMPP webpage.

2. On an ongoing basis, the Reserve plans to revisit the relative strength of Tier 2 and Tier 3 methods, as well as hybrid approaches such as the USDA N2O method. Are there recently published studies, calibration/validation efforts and/or methodologies which you believe the Reserve should consider? Are you aware of any such studies underway and/or methodologies under-development that the Reserve should be tracking?

3. The Reserve is seeking stakeholder feedback on whether to allow for submittal of data that has not been published in peer-reviewed journals, as long as other criteria included in this data standard are met. In particular, this is meant to allow for submittal of high quality reference data industry groups, sectoral experts, and government agencies (i.e. USDA, studies commissioned by the California Air Resources Board). Should these additional datasets be accepted as equal to a peer-reviewed study? Do you think this revision maintains an appropriately bar for high quality data? Are there other criteria that are more appropriate and useful as a minimum filter for judging data quality in a case like this?

4. Previously, the Reserve required that data collection follow the GRACEnet Chamber-based Trace Gas Flux Measurement Protocol, which is primarily used by ARS scientists currently. However, the Reserve is considering broadening this requirement to include protocols that use sample collection methods that are consistent with methods published in the following books, and the Reserve is seeking comment on whether these protocols are sufficiently robust for inclusion herein: Methods of Soil Analysis (Ch 47, by Rolston, in Volume 1), published by the Soil Science Society of America; Soil Sampling and Methods of Analysis (Rochette and Bertrand in Carter and Gregorich 2008), published by the Canadian Society of Soil Science; Standard Soil Methods for Long Term Ecological Research (Holland et al. 1999, ch 10 in Robertson et al. 1999); or Methods in Ecosystem Science (Sala et al. 2000). Are there any additional protocols that are equally robust that the Reserve should consider for inclusion? Noting that one of the major longterm GRACEnet projects is the creation of a national database with data on greenhouse gas fluxes, soil carbon stocks, biomass yield, etc, how might the Reserve simultaneously be more flexible in recognizing other flux measurement protocols, while also encouraging more scientists to use GRACEnet protocols (allowing these data to be uploaded into GRACEnet at a later date)?

5. The Reserve has been debating what the minimum temporal frequency of sampling should be in the Minimum Data Standard, both in terms of a bare minimum requirement as well as a "best practice" to encourage studies to aspire to. The Reserve is proposing a minimum requirement of weekly flux measurements, and increased frequency (at least 2x week) for a

minimum of 30 days following fertilization and first irrigation. Do you think this is a reasonable minimum standard, recognizing that the Reserve will also encourage more frequent sampling as best practice? Is this too stringent, such that it will leave quality data sets out of the running? At present the Minimum Data Standard encourages as best practice an increased frequency to DAILY flux measurements for up to 60 days following ANY agronomic event. What other guidance could the Reserve include as part of "best practice"?

6. Many field experiments measuring GHG fluxes in agriculture focus on growing season emissions only; however the Reserve is concerned that in regions with freeze/thaw cycles or winter precipitation, significant N2O emissions resulting from these events may not be captured with growing season-only sampling designs. Is it reasonable to require year-round measurement in these regions? Is one full year of "typical" weather conditions sufficient to get a sense of the non-growing season flux? If so, how should "typical" be defined (e.g. average temperatures +/- a certain percent?). For studies that did not take year-round data, can comparable experiments within the region (with similar climates and soil types) that did take non-growing season data be used to supplement the growing season data of the other study?

7. The Minimum Data Standard includes a section that discusses applying understanding from field experiments to a larger region, and the data submitter is asked to propose and justify a region over which data set could be extrapolated. However, the Minimum Data Standard currently does not define region, and the Reserve has not yet established guidelines on how far data might be extrapolated within a region. The Reserve is eliciting feedback to understand the potential for extrapolating data from a single (or minimal) number of sites to larger areas. As such, the Reserve is seeking comment on which attributes (that can be easily estimated across large landscapes) have the most influence on N2O emission fluxes. What type of analyses might the Reserve recommend to extrapolate from experiments conducted at a single site (or very few sites) to the larger region? In at least a semi-quantitative sense, what level of error is introduced as one extrapolates data further away from a core research site? Do Tier 3 models offer a sound method that the Reserve could use to evaluate the distance across which a data set could be extrapolated with confidence? Is it important for the Reserve to define region within the Minimum Data Standard in a more concrete way than "how far a data set can be extrapolated"? and if so, should the Reserve consider USDA regional definitions, Land Resource Regions, or some other well recognized definition?

8. The Minimum Data Standard states that the study must be replicated on multiple fields or study sites at different locations across a given region. The distance between locations would depend on the size of the region, over which the reference data might eventually be extrapolated. Study sites may cover an area the size of a small county (a few hundred square miles up to a few thousand square miles), a state (perhaps 100,000 square miles), or a USDA Land Resource Region (perhaps 100,000 - 500,000 square miles). Due to inherent variability across agro-ecosystems, the Reserve is seeking comment on what might be an appropriate minimum number of replicate study sites at different locations within a region. What minimum number would you recommend? How would that answer differ depending upon how a region is defined? How much additional confidence is gained as each additional site location (i.e. a

different geographic location not simply an additional replicate plot) is added in an experimental design?

11. The revised Minimum Data Standard recommends that submitters conduct preliminary sampling at one point in time, prior to commencing the study, to estimate variability across the field and among functional locations (defined as biophysical features like furrows or fertilizer bands thought to influence N2O emissions), but the Reserve does not make explicit recommendations for designing this preliminary sampling exercise. Should the Reserve provide more explicit guidance? The Reserve could recommend, for example, 4 chambers per functional location in a test plot and 10 chambers deployed at random across the field. Would you support a recommendation like this? How would you modify it? Would you recommend more or fewer chambers per functional location and/or chambers deployed at random? What additional guidance about collecting and using preliminary data would you recommend the Reserve offer?

12. The Minimum Data Standard recommends that submitters conduct a power analysis using data from analogous studies from other sites, by drawing on their own previous research or by collaborating or contracting with other research groups who have such data and/or analysis capacity. Would you recommend that the Reserve retain this best practice in the MDS? Aside from providing funds, which is not the role the Reserve usually plays, what could the Reserve do to make it easier for submitters to follow this recommendation?

14. Please see attached for a memo outlining questions and considerations of spatial sampling design for N2O emissions from agricultural fields on which the Reserve is eliciting feedback.

15. The Reserve's draft Minimum Data Standard proposes that the Reserve will not dismiss a data submittal on the basis of removing outliers if this removal is justified, explained, and/or does not result in significant changes to the annual N2O flux. Do you recommend any changes in this general approach to responding to submittals from which outliers have been removed? How large a change in N2O flux estimates, resulting from removal of outliers, would you consider large enough to warrant possibly removing the data set from consideration by the Reserve (i.e., to be "significant")? Would you recommend the Reserve further increase/decrease the proposed definition of "significant" from +/- 10% of the annual N2O flux?

17. The Reserve requires that new data submissions include a discussion of whether growing season climate and experimental weather conditions at the study site are representative of typical conditions in the region. The data submittal form notes that key attributes to be considered are: precipitation, temperature, and whether the site experiences freeze/thaw cycles and/or winter precipitation. What other attributes should be considered to ensure that environmental conditions during the study represent generally-occurring conditions in the region?

18. The Reserve requires that new data submissions include a discussion of whether soil properties at the study site are representative of typical conditions in the region. The data submittal form notes that key attributes to be considered are: soil type, soil texture, soil water holding capacity, and soil organic carbon levels. Are there any other attributes that should be

considered to ensure that soils at the study site represent generally-occurring soils in the region?

Considerations of Spatial Sampling Design for N_2O Emissions from Agricultural Fields

This memo (Question 14 in the above list of technical questions) concerns studies whose objective is to ascertain whether implementing a management practice (or set of practices) over an entire field decreases N_2O emissions from the field, and by how much, relative to a field using different (likely more common) practices. The Climate Action Reserve is seeking input about the minimum requirements for number and distribution of chambers used to determine whether N_2O emissions differ between paired fields to which a different management practice (or set of practices) has been applied. The Reserve is considering both what criteria to use as a *minimum requirement* and how to describe *best practice* for distributing chambers in these fields.

The following questions are intended to help elicit the wisdom and perspectives of scientists such as yourself who have experience working with trace gas fluxes in agriculture, or closely related investigations, in order to help the Reserve develop these requirements and best practice recommendations. The requirements and recommendations are part of the Reserve's Minimum Data Standard for the Nitrogen Management Project Protocol, which aims to drive high quality research that can underpin development of methodologies that quantify N_2O emissions. While direct answers to the questions below will be helpful to the Reserve, thoughts they spur about what minimum data requirements for the spatial distribution of samples should look like, along with the reasoning behind those thoughts, would be equally welcome.

Question 1: Emissions of N_2O are known to be highly variable in both space and time. Please consider any existing data sets that you have collected or analyzed.

a) How many plots per field were used to collect these data?
b) How many chambers per plot were used to collect these data?
c) Approximately what were the dimensions of the chambers?
d) If you used more than 2 chambers per plot and 3 plots per field, could you have detected differences using only 2 chambers per plot and 3 plots per field?
e) Were the chambers within each plot placed at random, or were their locations stratified by biophysical field properties like furrows or fertilizer bands?
f) Assuming no stratification of chamber placement based on biophysical attributes like furrows or fertilizer bands, how many plots and chambers per plot would you consider best practice in your study system to detect and estimate differences in N_2O fluxes between fields to which different management practices were applied?
g) If possible, please briefly describe the study system to which your other answers refer (e.g., dryland wheat in Nebraska)
h) Would you expect your answer to (d) to change in other, widespread agricultural ecosystems with which you are familiar? Please very briefly describe an example to illustrate your perspective.

Question 2: A particular point of debate is whether to sample at random within plots or with to stratify sampling according to biophysical field features like furrows or fertilizer bands that are believed, or known, to exhibit clear differences in N_2O emissions) and possibly in the response of those emissions to management. We term these features functional locations. **Please comment on the following proposed sampling design and recommend best practices for sampling; feel free to support or modify this proposal, or offer a completely different approach.**

There are two key points to keep in mind:

(1) Rather than measuring the entire emissions from a field, we measure emissions from a small number of areas in the field, called plots. We take an average of measurements from several plots and also use the measurements from those plots to calculate an uncertainty, or error, around that average value. Using both the average value and the estimate of uncertainty allows us to estimate the chance that the estimates from the two fields really are different from each other. To reiterate: We use measurements from the scale of the plot to estimate emissions from the field; to do this we need to arrive at a single estimate from each plot.

(2) We know that there is biophysical variation across the field, and we would like to use this variation to increase the accuracy of our estimates of field-wide emissions, and to decrease the uncertainty around these estimates.

Putting these points together, it becomes apparent that attempts to account for heterogeneity across the field (e.g., inside and outside of fertilizer bands, on berms vs. in furrows) need to occur within each plot (stratification), or else we need to have enough plots, randomly distributed across the field, that we assume they capture these different functional zones in proportion to their occurrence on the landscape. If we use stratified sampling, then we need to address two points:

The estimated emissions from each plot will need to be averaged with those of other plots in that field to yield an estimate of the emissions from the entire field, along with an estimate of error. Any accounting for the relative contribution of different functional locations needs to happen within the plot. If, for example, fertilizer bands account for 15% of the area of a field, and if emissions from fertilizer bands differ from emissions from other functional locations, you would want the plot-level emission estimates to weight the contributions of these different functional locations accordingly.

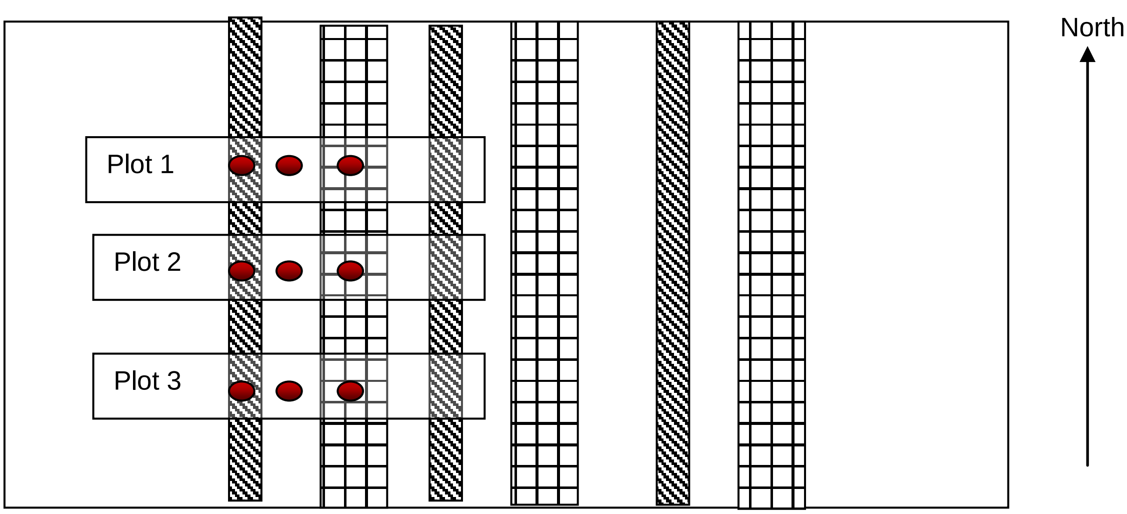

Figure 1. Plots in a field with 3 functional locations (furrows, berms, fertilizer bands). Red ovals are chambers. Hatched, checkered, and blank areas are functional locations (e.g., flat areas, furrows, and bands where fertilizer is applied).

We need a method for deciding whether functional locations really do yield different emissions in a field; if they do not, then it is hard to justify weighting emissions from each according to its area. Suppose a field has 3 functional locations. As depicted in Figure 1, one could use the following approach:

There are 3 chambers per plot, and within each plot one chamber has been placed in each functional location. Clearly the functional locations occupy different areas of the field. The first question is whether emissions from the different functional locations are actually different, on a per-area basis. It is not possible to answer this question definitively within each plot because we have only one chamber per functional area within the plot, and determining whether two values are actually different requires an estimate of error around each value. There are three options:

(1) Calculate the mean value of the three chambers in each functional location, across plots, and determine an estimate of error. Using the means and estimates of error, compare emissions among functional zones. If these values are NOT different from each other, proceed according to (a) below. If they ARE different, proceed according to (b) below.

   a. Weighting by area is NOT justified. Simply take the mean value of the 3 chambers in each plot. Flux for plot 1 = 1/3*{(flux from chamber in flat area)+(flux from chamber in fertilizer band)+(flux from chamber in furrow)}.

   b. Weighting by area IS justified. Determine the area of each functional location across the field. Within each plot, weight the contribution of the flux from each chamber according to those areas. For example, if functional location A represents 85% of the field area, functional location B represents 5%, and functional location C represents 15%, then: Flux for plot 1 = (flux from chamber in flat area *0.85)+(flux from chamber in furrow *0.05)+(flux from chamber in fertilizer band *0.15). The actual areas of the functional locations within the plot are not relevant.

   A weakness of this approach may be that it effectively uses each chamber twice: once to establish whether fluxes actually differ among functional locations, and a second time to estimate mean fluxes for each plot.

(2) Include at least 3 chambers per functional location per plot. This approach increases sampling effort by a factor of 3, but it enables a determination of whether the fluxes differ among functional locations within each plot. As with approach (1), a finding of different fluxes among functional locations leads to a calculation of fluxes for each plot that weights functional locations by area over the entire field, whereas a finding of no difference leads to a calculation of plot-level emissions that does not involve weighting. An open question is how to handle situations where chambers in one plot indicate differences among functional locations and chambers in another plot do not. (In approach 1, this determination cannot be made on an individual plot basis, so the question does not arise.)

(3) A determination of whether emissions differ among functional location may be made on the basis of prior observations in the literature, for similar situations. In this case, a determination of different fluxes would point to weighting, as in (1b).

Question 3: If you are feeling ambitious, please consider helping the Reserve to assess the pros and cons of using a stratified sampling design by examining an actual data set or simulated data that reflect the kinds of variation and distributions commonly observed in N_2O emissions from agricultural fields. Using these data, one could quantify, for example, how much the ability to detect differences in N_2O emissions from two fields increases by using 3 functional areas per plot with 1 chamber each, compared to 2 chambers per plot, placed at random. Any quantitative information about how additional fields and plots beyond the current minimum requirement increases the ability to detect differences would be helpful. If you have access to such data, are mathematically inclined, and are facile with statistical software, the Reserve would be grateful for the results of investigations such as these.

While we believe your answers to these specific questions will be very useful to the Reserve as it develops the minimum requirements and best practice guidelines, the Reserve is equally interested in your expert judgment about what these requirements and practices should look like, beyond or complementary to these specific questions.