Minimum Data Standard for Consideration in Quantification Methodology Development (Appendix D, NMPP V1.1)

D.1 Introduction

As noted throughout the 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 appendix provides general guidelines for establishing field experiments to develop reference data sets which can be used to develop and/or calibrate and validate standardized quantification methodologies. 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[1]), the NMPP's current Tier 2 approach does not necessarily set precedent for future expansions of the NMPP. 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[1]). 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 for which the management is often varying and that are grown in more complex rotations. Examples of such cropping systems are vegetable or fruit cropping systems.

Reference data sets will be reviewed by the Reserve to determine whether the data is appropriate for developing a Tier 2 methodology, for calibrating and validating a Tier 3 methodology (e.g. DNDC), or for further validating a previously accepted NMPP methodology. In addition to the data sets themselves, stakeholders are encouraged to develop and submit new Tier 2 or Tier 3 quantification methodologies, developed from these reference data sets, including justification of why the selected methodology is most appropriate for that specific crop/state/practice combination.

D.1.2 Process for Future Protocol Expansion

The minimum data standards presented in this appendix will serve as internal guidance for the Reserve in determining whether reference data are sufficiently robust. The Reserve will also maintain a Nitrogen Management Science Advisory Committee (SAC) into the future, and the Reserve will consult the SAC, as needed, when making determinations about the quality of proposed methodologies, their underlying reference datasets, and independent reference datasets.

[1] As defined by the Volume 4 of the 2006 IPCC Guidelines for National Greenhouse Gas Inventories.

Stakeholders are encouraged to submit new reference datasets and quantification methodologies to the Reserve at any time. Information on this submittal process is available on the Nitrogen Management Project Protocol webpage. Stakeholders should complete an NMPP New Data Submittal Form, which will be used to assess whether the dataset meets the minimum data standards included in this appendix. The stakeholder submitting data is also asked to provide recommendations for data sources on adoption rates of a given practice to be used for performance standard development. The Reserve will review new data submittals on an ongoing basis. The Reserve will periodically consult the SAC to determine whether a given data set or proposed quantification methodology should be prioritized for further development and inclusion in the protocol. Criteria to be considered include:

a) The existence of baseline N_2O emission measurements for the practice, region, and/or cropping system considered;
b) The total acreage and intensity of use of nitrogen fertilizer for the cropping system in question;
c) Whether sufficient data exists to develop a performance standard and preliminary assessments show a project activity is likely to be additional; and
d) The economic and technical feasibility, as well as the mitigation potential, of the management practice that reduces N_2O emission under consideration.

Once the Reserve identifies specific protocol expansions, 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 for developing and/or validating new N_2O quantification approaches, and/or validating existing N_2O quantification approaches using independent data.[2] Reference data can be new source data generated during new measuring campaigns or existing data from, *inter alia*, the following sources, so long as the data requirements included in this appendix are met: scientific and technical articles in books, journals and reports; universities and extension services; United States Department of Agriculture; sectoral experts, commodity and stakeholder organizations, and industry groups. A reference to the source of the data must be provided for existing data. For the Reserve to approve reference data for use in a new quantification method, it should comply with the minimum data standards described below.

D.2.1 Method of Data Collection

Reference data should be collected using either chamber-based or tower-based (micrometeorological) methods.[3] Chamber-based methods are currently the least expensive option for measuring N_2O emissions from agricultural fields, as the materials required for

[2] The minimum data standard applies for reference data used for the development of statistical models as well as for the calibration and validation of process-based biogeochemical models proposed for the quantification of N_2O emission reductions.
[3] 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 currently less attractive.

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.

Since methodologies to measure N_2O emissions are continuously improving, specific guidelines for sampling methods are not listed in this protocol. The Reserve will only review datasets for which sample collection methods comply with the most recent peer-reviewed guidelines available for the adopted method at the start of the experiments that yielded the reference data. A brief description of the chamber design, sample collection and handling, gas analysis and data analysis should be provided. For chamber-based measurements, the Reserve recommends following guidelines from the USDA Agricultural Research Service (ARS) GRACEnet Chamber-based Trace Gas Flux Measurement Protocol.[4] Measurements taken through tower-based methods should be consistent with methodologies currently in use in peer-reviewed scientific literature.

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.

D.2.2.1 Temporal Frequency and Scale of Data Collection

Flux measurements should take place at least once per week (every seven days). However, it is strongly advised to increase the measurement frequency following agronomic or environmental events known to be associated with major N_2O fluxes (i.e. tillage, fertilization, irrigation, rain, or harvest). Daily flux measurements after such events should continue until N_2O emissions return to pre-event levels. Note that N_2O responses to such events may not appear until several weeks after the event. This lag effect should be incorporated in the sampling design. It is recognized that due to unforeseeable weather conditions, issues with measurement devices, and other challenges, some gaps in the data set are unavoidable. Guidelines on how to handle outlying values are included in Section D.2.3.

Measurements also should represent the daily variations in N_2O fluxes. Multiple flux measurements could be made during one day. However, one flux measurement taken per day is acceptable, so long as it is taken at a time that corresponds to the daily average temperature (e.g. mid-morning or early evening).

Flux measurements should be taken at a minimum over the complete growing season, but year round flux data is preferable. Reference data should extend over at least two consecutive growing seasons. Flux measurements over additional growing seasons may be necessary if the two consecutive growing seasons for which measurements were taken exhibited anomalous weather conditions, with respect to that region.

D.2.2.2 Spatial Frequency and Scale of Data Collection

N_2O emissions are not only variable over time, but are also subject to high spatial variability. This spatial variability reveals itself at multiple geographic scales, including variability within a field, variability across fields within the same landscape, and across landscapes (e.g. a Land Resource Region or a Major Land Resource Area). In this section, guidelines are provided to ensure that the reference data accounts for spatial variability at those different scales. Note that

---

[4] Parkin, T.B., & Venterea, R.T., 2010. Available at www.ars.usda.gov/research/GRACEnet.

the terminology for “field” in the NMPP, as defined in Section **Error! Reference source not found.**, 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. In these guidelines, the Reserve uses “replicate plot” to refer to the smallest experimental unit and “field” to designate a greater unit with multiple replicate plots. In other words, a replicate plot corresponds to a field as defined in the NMPP.

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

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. due to the number of fertilizer granules present in the chamber, the presence of decomposing crop residues, etc.). Chamber surface areas typically cover between ~300 and ~3000 cm^2.

2. **The number of flux chambers per functional locations within a replicate plot**: In many cropland systems, multiple functional locations with different soil moisture conditions, soil temperature and N concentration can be identified within a replicate plot (For example: middle of the berm, side of the berm, the furrow in annual row crops, tree row versus tractor row in orchards, etc.). It is recommended that flux chambers be strategically placed in multiple functional locations so as to represent the variety within the field appropriately. A minimum of two flux chambers per functional location within a replicate plot is recommended.

3. **The number of replicate plots per field**: The reference data should cover a minimum of 3 replicate plots per treatment (i.e. management practice) and per field. Usually, for a side-by-side (“pairwise”) comparison, there will be at least 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).

4. **The number of fields**: The field(s) should be representative for the conditions within the area in which the reference data sets will be used. Therefore, multiple fields are to be used that are located at different sites and geographic locations (e.g. different counties, different states). Ideally, the fields (and replicate plots within fields) are also chosen to represent some of the most commonly occurring soil types in a region. However, it is recognized that having multiple fields may be challenging.

D.2.3 Outliers

When experimental data are collected, it is very likely that some samples will have values that are considerably larger or smaller compared to replicate samples. Such samples are often referred to as outliers, and can be spatial, temporal or analytical in nature. Analytical outliers can be caused by inadequate closure of flux chambers, leaky sampling vials, errors in sample collection or analysis, etc. and labs can remove analytical outliers in a routine and standardized fashion. However, as N_2O fluxes are known to be very variable in space and time, spatial and temporal outliers are often merely a reflection of the variable nature of the process and should be handled as real data. As such, removal of temporal and spatial outliers is strongly discouraged; the Reserve prefers that submitted reference data include any observed temporal or spatial outliers, with notations as to which outliers were flagged for removal by lab analysis. In some cases, there is a real reason for removing temporally or spatially anomalous data. Examples include local flooding due to a leak in a drip line, enhanced N_2O fluxes due to undesired animal excretions in the flux chamber, etc. Under such situations, temporal and spatial outliers may be removed by the Reserve prior to methodology development, if the outliers were properly identified and a justification is provided with the data set submittal. The extent to which inclusion/exclusion of this value affects the mean should be discussed in this justification.

D.3 Applicability of Field Experiment to a Region

Stakeholders will be asked to propose and justify a geographic applicability region over which a data set (or the subsequently developed quantification methodology) may be extrapolated. It is recommended that the justification includes a comparison of weather and climate, soil characteristics, and management practices between the study sites and the geographic applicability region.

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.

Further, "typical" soil type, soil texture, soil water holding capacity, soil organic carbon levels, etc., for a given region should be considered when selecting replicate plots and fields for inclusion in an experiment. Sites should be chosen for their widest applicability to multiple soil types, etc., within the region. Likewise, the management practices executed on the field trials should be selected so that they represent the overall management within the region.

D.4 Independent Validation and Quantifying Uncertainty

Large uncertainty around field measurements leads to uncertainty around predicted emission reductions for any quantification approach. Therefore, the quantification approach must be robust in situations with high uncertainty. Even though a quantification methodology may ensure that projects meet minimum standards through eligibility and applicability conditions (e.g. conditions for which the model was calibrated), a significant amount of uncertainty may remain, which must be accounted for through an uncertainty deduction mechanism.

According to C-AGG's white paper on uncertainty, analyses of both structural and input uncertainty related to their use must be completed so as to use and apply models appropriately.[5] Input uncertainty for an empirical model is subject to less uncertainty than a biogeochemical model, simply because there are significantly fewer critical inputs. Quantification approaches based on biogeochemical models, and quantification approaches for which the input variables are associated with a significant amount of uncertainty, require a Monte Carlo simulation to assess the effect of uncertainty around input variables on projects' N_2O emissions reduction estimates, as is done in the Reserve's Rice Cultivation Project Protocol (RCPP). In addition, all quantification approaches that are using a biogeochemical process model must include how to parameterize every input parameter to the model. More

[5] C-AGG, Executive Summary: Uncertainty in Models and Agricultural Offset Protocols. Discussion draft, 2012.

specifically, for every input parameter, it must be explained if the parameter has to be set using field measurements, look-up tables, default values, or internal calibration. If internal calibration is used to set certain parameters, the procedures for calibration must be clearly explained, as is done in the RCPP.

Structural uncertainty (termed $\mu_{struct,f}$ in the RCPP and NMPP) represents how well the model performs against measured emissions, regardless of whether that model is an empirical model or a biogeochemical model. To estimate structural uncertainty in the RCPP, for example, independent emissions measurement data (e.g. data that were not used to build the model) for California rice fields were used to "validate" the DNDC model by comparing measured and modeled data.

In the case of this protocol, in which an adaptation of the MSU-EPRI methodology is included (see Section 5), no additional field emissions measurement datasets for N rate trials are currently available for the North Central Region, other than MSU-EPRI's robust data set. This makes it more challenging to validate the methodology and estimate structural uncertainty. However, the developers of the original MSU-EPRI methodology performed a "leave-one-out" cross-validation analysis[6] to approximate the structural uncertainty and found that the uncertainty increased about 2 to 4 percent compared to an uncertainty analysis using non-independent data. The uncertainty quantified using a leave-one-out cross-validation is certainly applicable for areas similar in characteristics to the study sites. However, the uncertainty is likely greater for areas far away from the study sites. As a consequence, the "leave-one-out" approach's 2 to 4 percent increase in uncertainty was considered acceptable by the Reserve for the state of Michigan, where all of the study sites used to develop the MSU-EPRI quantification approach are located. However, an additional 15 percent uncertainty deduction is taken for other states in the NCR to avoid underestimating the structural uncertainty on sites that are far away from the field measurement locations.

When independent data becomes available to validate the model and quantify the structural uncertainty explicitly for the various NCR states, the Reserve plans to adjust the structural uncertainty deduction currently included in the NMPP.[7] This independent reference data should be gathered from a sufficient number of different data points so that the reference data can be divided into separate calibration and validation data sets. If calibration data are taken primarily from one area within a larger region (such as a Land Resource Region), an extensive validation data set, including data points from other areas within the region collected from a number of sources, might allow validation of the model for a much larger geographic area than the model was otherwise developed and calibrated for. It is worth noting that while the MSU-EPRI methodology was adapted and included in the NMPP before independent data was available, this decision is not precedent-setting. The Reserve prefers a full structural uncertainty assessment using validation data that is representative for the geographic applicability region over the leave-one-out approach.

[6] Generally, the goal of a cross-validation analysis is to evaluate the fit of a model to a data set that is independent of the data that were used to train the model. A leave-one-out cross-validation analysis estimates the structural uncertainty by comparing a single observation from the original sample to the outcome predicted by a model that was calibrated using the remaining observations.

[7] The Reserve anticipates that market drivers will try and reduce this uncertainty deduction as soon as possible, hopefully within the next five years.