Draft Responses to MARSSIM Revision 2 Charge Questions

Draft Responses to MARSSIM Revision 2 Charge Question 1.1

Charge Question 1.1 on Scan-Only Surveys

1) Are the revisions to MARSSIM concepts and methodologies technically appropriate, useful and clear, and do they provide a practical and implementable approach to performing environmental radiological surveys of surface soil and building surfaces?

1.1 Please identify whether the inclusion and proposed implementation of scan-only surveys (Section 5.3.6.1 and Section 8.5) is appropriate, adequate and clear, especially the discussion on sampling for scan-only measurement method validation or verification.

Charge Question 1.1 Response

The response is provided according to four considerations:

1. *Technical appropriateness of concepts & methodologies*
2. *Practicality of implementing approaches during surveys of surface soil and building surfaces*
3. *Usefulness & clarity of MARSSIM revisions including past misinterpretation of term "scan-only"*
4. *Verification of scan-only surveys with sampling & laboratory analysis*

Technical appropriateness of concepts & methodologies

- Provide additional discussion on instrument response and calibration factors for the specific site or surface being measured
- Discuss site-specific certified reference validation standards for scan-only calibrations with MQO targets equivalent to the 2nd bullet on the "Verification of scan-only surveys ..." slide
- Address overlapping fields of view for scan-only surveys with gamma-ray detectors
- Justify subjecting the minimum detectable concentration for scan-only surveys (Scan MDC) to 50% of the Derived Concentration Guideline Level for wide areas ($DCGL_w$)
- Provide a more complete description of scan-only technology advances since the last revision of MARSSIM
- Develop a stronger, more formal connection between scan-only measurements and laboratory verification results to substantiate quantification requirements for scan-only surveys
- Explain the relationship among MDC recommendations for scan-only surveys, instrument calibration including its rigor for the specific site under investigation, and statistical approaches as well as the extent of conservatism in metrics selected for decision making
- General: Improve the consistency of MDC recommendations relative to guideline levels throughout the document. Explain the rationale for differences.

Technical appropriateness of concepts & methodologies

- Provide additional discussion on instrument response and calibration factors for the specific site or surface being measured
- Discuss site-specific certified reference validation standards for scan-only calibrations with MQO targets equivalent to the 2nd bullet on the "Verification of scan-only surveys ..." slide
- Address overlapping fields of view for scan-only surveys with gamma-ray detectors
- Justify subjecting the minimum detectable concentration for scan-only surveys (Scan MDC) to 50% of the Derived Concentration Guideline Level for wide areas ($DCGL_w$)
- Provide a more complete description of scan-only technology advances since the last revision of MARSSIM
- Develop a stronger, more formal connection between scan-only measurements and laboratory verification results to substantiate quantification requirements for scan-only surveys
- Explain the relationship among MDC recommendations for scan-only surveys, instrument calibration including its rigor for the specific site under investigation, and statistical approaches as well as the extent of conservatism in metrics selected for decision making
- General: Improve the consistency of MDC recommendations relative to guideline levels throughout the document. Explain the rationale for differences.

Practicality of implementing approaches during surveys of surface soil and building surfaces

- Present appropriate quality assurance and quality control definitions for the scanning technology and its recording output
- Update the two-staged technique on scanning with frequent stationary measurements (designed for older equipment without data logging) to accommodate modern systems with continuous data logging of detector location and response
- Incorporate maps of radiation data generated by continuous data-logging scanning systems
- Describe how scan-only equipment relates to data and measurement quality objectives including measurement sensitivity requirements for scan-only surveys
- Strengthen the description of relevant equipment for scan-only surveys
- Provide additional detail on how to perform scan-only surveys
- Illustrate scan-only implementation with validation from sampling and laboratory analysis
- Present implementation examples of replicate measurements over defined scanning areas
- Add detailed insights from a case study

Usefulness & clarity of MARSSIM revisions including past misinterpretation of term "scan-only"

- Revise the term "scan-only" surveys to
  - Generic scanning surveys: Systems with no site-specific calibration that require validation from sampling and laboratory analysis
  - Site-specific scanning surveys: Systems applying site-specific calibrations with reference validation materials that require no additional confirmation from sampling & laboratory analysis
- Clarify expectations associated with the recommended percentages of areas covered by scan-only surveys (e.g., how to address 90% of the area scanned for Class 1 limited by obstacles in comparison to the 100% recommendation)
- Discuss surface versus volumetric contamination in the context of scan-only surveys
- Expand the description of scan-only survey impacts due to different particle emissions
- Incorporate examples with readouts of representative scan-only technology
- Elaborate on the extent to which scan-only surveys should be verified
- Justify why final status surveys based on scan-only measurements would be inadequate
- General: Extend cautionary notes on applying statistical tests to small data sets into other areas of the document. Illustrate the risks from small data sets with examples.

Verification of scan-only surveys with sampling & laboratory analysis

- Maintain that collecting samples for laboratory analysis to validate scan-only measurements is a good quality control practice
- Recommend the measurement quality objective (MQO) for laboratory analysis has a measurement uncertainty of at least 1/3 of that for the scan-only measurements (i.e., laboratory measurement uncertainties are at least 3 times smaller compared to scan-only measurement uncertainties)
- Acknowledge that site-specific instrument calibration with reference validation standards for the contaminants under investigation can remove requirements for subsequent confirmation by physical sampling and laboratory analysis
- Highlight important distinctions between scanning with stationary measurements versus scanning measurements with sampling for laboratory analysis
- Consider a protocol for data-logging scanning systems with (1) site-specific instrument calibration, (2) preselected locations for stationary measurements according to a grid, (3) continuous scanning measurements at a constant speed, and (4) follow-up stationary measurements at locations suspected to have the highest concentrations
- Expand the verification discussion for scan-only surveys to include spatial concentration variability, contamination depth, and anticipated contaminant migration
- Programmatic: Consider a scan-only performance testing program to validate capabilities

Summary comments on Q1.2 Stram, Hamrick, Smith, Wang

Please comment on the inclusion and proposed implementation of Scenario B

Charge questions

There are 4 questions/requests asked in this charge

A. Please comment on the inclusion and proposed implementation of Scenario B (Chapter 4, Section 5.3, and Chapter 8)

B. Is it appropriate to recommend that Scenario B be used only for those situations where Scenario A is not feasible?

C. Are methods for considering background variability in assessing whether the site is indistinguishable from background reasonable and technically accurate?

D. Is the inclusion and proposed implementation of added requirements for retrospective power analysis and the Quantile Test while using Scenario B technically appropriate and discussed adequately and clearly?

Overall comments

1. All panel members support the inclusion of Scenario B
2. Clarity needs to be improved, as well as access to key concepts/references. In some cases concepts are used before they are defined, or defined differently in different places
3. Terminology is complicated
   - Gray region, UBGR, LBGR, relative shift, DL, AL, DCGL all need to be understood in order to do power and sample size calculations for either Scenario A or B
   - This terminology is only standard throughout the multiagency manuals (MARSAME, MARLAP, MARSSIM, etc.) not throughout all of statistics!
   - But once learned the terminology makes overall sense to someone with appropriate statistical training (but see point 2 above)
   - All involved in the MARSSIM process need to also know the terminology and concepts – regulators, planners, stakeholders, contractors, etc.]
4. Question remains whether MARSSIM procedures can be followed by someone w/o good statistical understanding (can it function as a good cookbook)? [everyone needs to be on the same page]

Is it appropriate to recommend that Scenario B be used only for those situations where Scenario A is not feasible?

- The panel agrees that it is reasonable to recommend that Scenario A should be the default scenario except when the DCGL is close to zero.
  - In some situations, where the proposed residual radioactive material criterion is “close to zero”, consideration of Scenario B may be driven by available field instrumentation and laboratory analyses to be performed. Specification of adequate measurement methods is part of the study design.
  - In other situations, consideration of Scenario B may be driven by the fact that the contaminant is also found in nature with high variability in the location to be surveyed. In this case, refinement of instrumentation and analyses will not eliminate the need to rely on Scenario B for ultimate release
  - The panel requested that language indicating that Scenario B was not as desirable as Scenario A be dropped or modified. It should be emphasized that Scenario B has an important and necessary role to play.
    - The presumption that the burden of proof shifts from user to regulator when Scenario B is utilized should be retracted.
    - The burden of proof remains on the MARSSIM user to show (i.e. through retrospective power analysis) that the data in fact strongly supports the null hypothesis when Scenario B fails to reject the null

Are methods for considering background variability in assessing whether the site is indistinguishable from background reasonable and technically accurate?

- Kruskal-Wallis test is appropriate for detection of background variability
- ANOVA gives appropriate estimates of variance of random effect
- Chebyshev inequality for bounds is quite weak (Confidence limits very wide)
  - In Example 9 the suggested LBGR of 3*omega (corresponding to upper ten percent tail using the Chebyshev inequality) seemed rather large by the panel. The text should emphasize that this is an extreme value and that lower values can be considered. Here omega is the square root of the variance of the random effect for background.
- Clarity of presentation could be improved
  - Examples should be stand alone; (e.g., Example 9 depends on equations and Tables from NUREG-1505, e.g. equations 13-3 and 13-13, and Tables 13.1 and 13.5.)
- MARSSIM should acknowledge underlying assumptions of statistical tools and include cautionary notes indicating under what conditions the statistical tests become unreliable

Is the inclusion and proposed implementation of added requirements for retrospective power analysis and the Quantile Test while using Scenario B technically appropriate and discussed adequately and clearly?

- The panel was unanimous about the importance of retrospective power analysis in Scenario B
- The panel found that the Wilcoxon test was technically appropriate, adequately discussed and clear
  - Drawback is that Wilcoxon tests for a shift in medians not means
  - This motivates use of a test for shift in the tails such as the Quantile Test
- However, a better description of the Quantile Test, literature references, discussion of the underlying null and alternative hypotheses, and power, are all needed in the MARSSIM revision
- MARSSIM-2 goes a long way to provide guidance for "what to do" but could do better for "how to do it" with easily understood beginning-to-end detailed step-by-step worked-out stand-alone examples and case studies where all statistical tools are applied and worked out in detail to support the decision making process

Draft Response to MARSSIM Charge Question 1.3

REVISED DRAFT RESPONSES TO CHARGE QUESTION BASED ON DISCUSSION DURING JANUARY PUBLIC SESSION. DO NOT CITE OR QUOTE.

Charge Question 1.3 Has Four Parts

a) **Is the proposed implementation of the of the concept of Measurement Quality Objectives adequately and correctly described, including the concept of measurement uncertainty (Chapter 4 and Appendix D)?**

b) **Is the proposed calculation of measurement uncertainty consistent with the concept of Measurement Quality Objectives?**

c) **Is the method appropriate and practical for both laboratory and field (including scan) measurements?**

d) **Please comment on the concerns of stakeholders that calculating measurement uncertainty for field measurements makes the survey process difficult to implement. In addition, please comment on whether recommendations provided by NIST, ANSI/IEEE and MARLAP for measurement quantifiability should be incorporated further into MARSSIM, Revision 2, or whether the current recommendations should be left as is (e.g., the original MARSSIM requirement that the MDC/MDA should be set at 10-50% of the action level).**

Measurement Quality Objectives and Uncertainty: 1.3 a) - Adequately and Correctly Described in Chapter 4 and Appendix D? (1 of 2)

The material in Chapters 4 and 6 and Appendix D seem to be consistent with relating measurement uncertainty to MQOs.

There does not appear to be a rationale that relates the way to determine if instrument uncertainty meets the MQOs. The basis for selecting instrumentation and methods to adequately classify sites etc., relies on proper instrument selection before measurements are performed.

MQOs should use exact terminology, i.e., standard uncertainty, combined uncertainty or expanded uncertainty, and the coverage factor, k,.

Appendix D –lists measurements for inclusion in uncertainty determination –

- Need to identify Type A/Type B uncertainty components, expected values, sensitivity coefficients and covariance effects
- Specific examples of how to calculate CSU for measurements is needed

Measurement Quality Objectives and Uncertainty: 1.3 a) - Adequately and Correctly Described in Chapter 4 and Appendix D?

(2 of 2)

Recommendations regarding how to set the MDC based on the UBGR are generic and not based on *method* or measurement uncertainty

Relationships between MDC, MQC and UBGR are not well defined

The MQC is mentioned in Chapter 6.4 - how to calculate it should be included in that Chapter or in Chapter 4.

Measurement Quality Objectives and Uncertainty: 1.3 b)Proposed Calculation of Measurement Uncertainty Consistent with the Concept of Measurement Quality Objectives?

Measurement uncertainty and how it should be determined is not adequately described.

- The description of its determination is scattered in the manual and the Appendix
- The references to uncertainty in MARSSIM (more than 300) are valuable, but are information overload for a user of the manual
- An adequate example of a realistic determination of final measurement uncertainty (CSU) is not provided but should be integrated into case studies
- Method uncertainty is not addressed in the selection of instrumentation

Other concepts in MARSSSIM rely on this:

- The MDC/MDA concepts require that measurement/method uncertainty be a well established MQO.
- Scanning and in-place measurement selection require a well established requirement for measurement uncertainty
- Whether or not a $DCGL_{EMC}$ or $DCGL_{W}$ has been exceeded

Measurement Detectability and Quantifiability: 1.3 c)Is the method appropriate and practical for both laboratory and field (including scan) measurements?

Laboratory instrument relationship of MQOs to required *method* uncertainty is described in MARLAP. The MARSSIM method for relating *measurement* uncertainty to MQOs is inconsistent with the guidance provided in MARLAP.

For the experienced MQO and measurement uncertainty practitioner, MARSSIM is a resource for planning and execution of site cleanups. However, the novice is faced with understanding the right questions to ask. From a practical standpoint the manual is lacking details and sufficient examples.

For scan measurements operator speed of scan, distance from the surface and experience all contribute to shared uncertainty. This contribution to measurement uncertainty is not discussed and is not part of the MQO discussion and in general may contribute a greater part of the CSU than just the counting uncertainty.

From the practical approach for a manual user, an example showing an evaluation in interpreting the results and reaching a final decision accounting for possible differences in outcome depending upon approaches to Data Quality Management would be useful. Missing throughout Chapter 4 are appropriate call-ins to details that are presented in Appendix D that would strongly help in using Chapter 4.

Measurement Detectability and Quantifiability: 1.3 c)Is the method appropriate and practical for both laboratory and field (including scan) measurements?

Laboratory instrument relationship of MQOs to required *method* uncertainty is described in MARLAP. The MARSSIM method for relating *measurement* uncertainty to MQOs is inconsistent with the guidance provided in MARLAP.

For the experienced MQO and measurement uncertainty practitioner, MARSSIM is a resource for planning and execution of site cleanups. However, the novice is faced with understanding the right questions to ask. From a practical standpoint the manual is lacking details and sufficient examples.

For scan measurements operator speed of scan, distance from the surface and experience all contribute to shared uncertainty. This contribution to measurement uncertainty is not discussed and is not part of the MQO discussion and in general may contribute a greater part of the CSU than just the counting uncertainty.

From the practical approach for a manual user, an example showing an evaluation in interpreting the results and reaching a final decision accounting for possible differences in outcome depending upon approaches to Data Quality Management would be useful. Missing throughout Chapter 4 are appropriate call-ins to details that are presented in Appendix D that would strongly help in using Chapter 4.

Measurement Quality Measurement Detectability and Quantifiability: 1.3 d) - Calculating Measurement Uncertainty for Field Measurements?

There is no mention of the MDA/MDC set at "10% to 50% of the UBGR" in the text of MARSSIM. The proposed range cited in the charge question, even if it was in the text, is arbitrary and not based on the project specific DQOs which must include method uncertainty for the instruments selected and the other contributors to the measurement uncertainty.

The instrument selection process must ensure that the method uncertainty (includes not only the instruments but the process of the measurement) meets the project DQOs.

Calculation of an MDC or MDA is an *a priori* process. These concepts are not addressed in the NIST, ANSI/IEEE documents and it is unclear what from those documents should be used in MARSSIM based on the text in MARSSIM. The guidance provided in MARLAP could be considered for inclusion in this manual, apart from the material about MQC.

MARSSIM does not cover in sufficient detail Type A and Type B uncertainties especially as it applies to scan measurements. The Type A and B uncertainties defined in the NIST and ISO documents: Type A - determined using statistical processes, Type B determined based on professional or other knowledge. Both types of uncertainty can contain either shared or unshared measurement uncertainties which adds a third dimension to the uncertainty equation. A table or description collecting the most common of all these uncertainties, with a realistic example, would be a significant help to those performing the measurements. If the uncertainty process is too difficult to implement in some limited circumstances, a summary table of various approaches or considerations that might be employed instead to limit effects for data analysis and decision making may be a useful addition. Including scan-only surveys increases the sampling design error, and thus measurement uncertainty becomes more critical.

Recommendations (1)

1. MARSSIM should include detailed step-by-step worked out examples of setting up σ and u_{MR} uncertainty component lists for a few cleanup scenarios with the expected measurement systems to be used.
2. A discussion should be included describing how to select field scanning or in-place measurement instruments.
3. The use of standard terminology and symbols for uncertainty, standard deviation and all statistical parameters should be used throughout the manual.
4. Examples of determining measurement uncertainty for field scanning instruments should be included with details of how the calculations are performed. Included in these examples should be consideration for the following: reliability, robustness, dynamic range, sensitivity, specificity, calibration, specificity, sample collection, heterogeneity, attenuation, temporal effects, operator skill, speed, distance, blank result, surface self-shielding shielding. A table or description collecting the most common of all these uncertainties should go into Chapter 4.
5. MARSSIM-goes a long way to provide guidance for “what to do” but could do better for “how to do it” with easily understood beginning-to-end detailed step-by-step worked-out stand-alone examples and case studies
6. If the uncertainty process is too difficult to implement in some limited circumstances, a summary table of various approaches or considerations that might be employed to limit effects for data analysis and decision making may be a useful addition.
7. Overall Chapter 4 and Appendix D relate measurement uncertainty and MQOs but are short of details on how to relate the MQOs and measurement uncertainty. Additional details to the examples are needed and integrated into case studies.

Recommendations (2)

6. Explicitly define the terms repeatability and reproducibility as they are used in the document.
7. A discussion of where in the gray region the MDC/MDA should lie is needed
8. A discussion of detectability versus measurement result based on the measurement uncertainty is needed.
9. Include a discussion of Type A and Type B uncertainties and how shared and unshared contributions to the measurement uncertainty fit into the Type A or Type B definitions.
10. Formula (6-18) should have a better statement of context. It is an approximation appropriate when the function being evaluated is approximately linear over the region of uncertainty. It does assume the different error terms are uncorrelated and if that's not true, there need to be additional terms reflecting the correlations.
11. In Section 6.4.4 there should at least be some discussion of when this kind of normal-theory confidence interval is appropriate and some mention of the t distribution when the standard deviation is estimated from data.
12. MARRSIM should collect all of the recommendations and put them in a list at the end of each Chapter.
13. An *a posteriori* MARSSIM historical measurement record should be used to validate the MDC practice for MARSSIM cleanup projects, i.e., see the validity and accuracy of the MDC concept.

Recommended Editorial Changes (1)

1. The *u* symbol should be included in the Symbols, Nomenclature, and Notations list. The *u* and *σ* symbols should be used where appropriate, particularly where Δ/u should be used.
2. Appendix D D-53, L 1-3 "The uncertainty of a measurement expressed as combined standard uncertainty includes the counting uncertainty of the measurement instrumentation and the sum of the errors associated with the measurement system." Recommend rewording of the sentence by using the Equation 6-18 terminology that the uncertainties are combined as the root-sum-of-squares.
3. Chapter 6 6-31, Example 8 illustrates *some* of the steps used to estimate an uncertainty, in this case σ_y, for a measurement counting process. Recommend detailing all the steps of this process for estimating measurement uncertainties, and also for the theoretical total standard deviation of the population distribution being sampled, *σ*, that is used for MQO *Δ/σ*.
4. Terminology used throughout MARSSIM should use the guidance of NIST 1297 and ISO GUM.

Recommended Editorial Changes (2)

5. The following abbreviations appear nowhere in the text and only in "Symbols, Nomenclature and Notations"
   5. Y_c
   6. y_D
6. In Figure 4.1 the wording inside the information boxes reference sections of Chapter 4 (i.e., 4.3, 4.4, etc). These do not reflect the title of those sections of Chapter 4 and do not reflect the content of those sections.
7. Page 4-4 L21. Samples are authentic? Perhaps using words like reliable, dependable, trustworthy, valid are more in keeping with terminology used in such investigations?
8. Page 4-5 L6-8. Why not have some of these referenced tables located in Appendix D, or even in Chapter 4? This part of the discussion is central to the entire chapter; why make it difficult for the reader to view the most essential of these tools (such as Tables)?
9. Page 4-5 L16. Measurement performance criteria are not defined here. The term is mentioned in Appendix D.2.2 but it is not defined. Definition is needed in Chapter 4.
10. Page 4-28 Scanning speed, operator training, calibration and tracking of these calibrations needs more details.
11. Page 4-31 L37 "MARSSIM recommends that a realistic or conservative estimate of the MDC be used instead of an optimistic estimate." An example of what might be realistic or conservative would be helpful like 95 % Type I error.

Recommended Editorial Changes (3)

12. Page 4-55. L18. "…alpha of 0.05 and beta of 0.10". Why not use Type-I error alpha and Type-II error beta?
13. Appendix D D-23 L8-11. More information on Scenario B can be found in the NRC draft report NUREG-1505. Why not discuss implications of Type I error alpha for Scenario B? While the citations are thorough, the concept is critical at this point of the discussion, and expansion (brief summary) is reasonable.

Charge Question statement

1.4. Is the discussion of survey requirements for areas of elevated activity technically accurate, appropriate and clear?

In particular, please comment on the decision to maintain the use of the unity rule for multiple areas of elevated activity (Section 5.3.5, Section 8.6 and Appendix O.4).

Are there suggested alternatives to the use of the unity rule?

Charge question 1.4
Technically accurate, appropriate and clear

1. The current discussion in MARSSIM concerning survey requirements for areas of elevated activity is technically accurate and appropriate, but not always clearly written (plain language). Additional details would be helpful, including more examples.

   Examples 7 & 8 could be clearer if the authors expanded the context and more sampling info was included (work the problem from the beginning as a real field situation).

Charge question 1.4 -- Maintain Unity Rule.

2. Maintaining the use of the unity rule was advised with caveats. Most commenters endorsed the limits as stated in the Draft Rev 2 document and coordination with the appropriate regulators since their final agreement is necessary for license termination/site release.
   - If multiple radionuclides are present in elevated areas, the unity rule may be a conservative alternative.
   - Unity rule can serve as the initial screening tool.

Charge question 1.4 summary continued.

3. Additional real-life examples would be useful in demonstrating the concepts. Narrative & equations before examples is inconsistent making examples hard to follow. We suggest adding real-life examples before the narrative and examples.

   In Section 8.4's formulation of unity rule suggests $DCGL_{EMC}$ is a level in addition to, not instead of, the $DCGL_{w}$.

Charge question 1.4: Alternatives available

4. No specific alternatives were presented other than those provided in Section 8 of the proposed MARSSIM revision. A conservative alternative could calculate the total dose (combining external & internal components) at each elevated area with doses from other areas being down-weighted using the inverse-square rule from the main site and the limiting case picked for compliance.

Use of any rule needs to be conservative and not increase the likelihood of inappropriate site release. Don't believe alternatives can exist to *accurately reflect* the potential for dose assessment.

Editorial Comments are available

Detailed editorial comments (Section 5 and 8) are available for the report's editorial section.

CHARGE QUESTION 1.5

Is the discussion of the use of MARSSIM surveys for addressing sites containing discrete radioactive particles technically sound and appropriate, and is the description accurate? In particular, please comment of the rule-of-thumb for determining when use of MARSSIM may not be appropriate for survey units containing discrete radioactive particles (Section 4.12.8 and Appendix O.5).

RESPONSE 1 TO CHARGE QUESTION 1.5

The current discussion in the draft MARSSIM document is incomplete regarding discrete radioactive particles (DRP). The discussion should not be confined solely to Section 4.12.8 or Appendix O.5.

Recommendation.

The subject of DRP should be introduced no later than Chapter 3 because having a known history of DRP will influence future surveys. The description of DRP’s should be supplemented with statements regarding the difficulty in their detection dependent on the suspected radionuclides, the likelihood of multiple DRP’s, and their physical behavior in being highly mobile and a possible contamination source.

RESPONSE 2 TO CHARGE QUESTION 1.5

Dose pathways differ for discrete radioactive particles compared to bulk contamination at the site. As a result, the concept of derived concentration guideline levels (DCGL) as used in MARSSIM is unlikely to be effective for DRP’s.

Recommendation

Because general DRP guidance has not been sufficiently developed, decision aids, perhaps site specific rules of thumb, are needed to address when additional treatment of discrete radioactive particles is warranted. Such decision aids will allow resources to be directed to aspects most important in terms of dose and risk.

RESPONSE 3 TO CHARGE QUESTION 1.5

The caution of not using the MARSSIM survey strategies for $DCGL_{EMC}$ currently contained in the draft is appropriate.

FURTHER RECOMMENDATIONS TO CHARGE QUESTION 1.5

- DRP's should be addressed in those surveys conducted earliest in an attempt to find and remove such particles with follow up verification because of the potential health hazards posed to workers and to avoid having to use a rule of thumb to decide whether to use MARSSIM survey strategies or not.
- The discussion of the DQO and MQO processes in Chapter 4 should address the influence DRP's will exert on method and measurement uncertainties. These processes will affect the detection requirements that will in turn influence instrument selection and use. DRP's are most likely to be discovered by scanning techniques and the DQO and MQO process should identify instrument response patterns that would indicate the presence of a DRP. A figure devoted to summarizing the planning processes for DRP's would be beneficial.

FURTHER RECOMMENDATIONS TO CHARGE QUESTION 1.5

- More thought should be given to developing one or more rules of thumb because the current rule may not have the intended effect. The source detector distance and length of the elevated activity area required for the current rule-of-thumb may not be sufficiently determined.
- Appendix O.5 should be either an example for addressing DRP or an expansion of the basic concepts addressed in the body of the manual.

Charge Question 2.1

"Please comment on whether the description of updated measurement methods and instrumentation information (Chapter 6 and Appendix H) are useful, appropriate, and clear."

General Comment

The description of measurement methods and instrumentation information in Chapter 6 and Appendix H are generally useful and in large part appropriate and clear. However, there are descriptions of concepts and operations that could be made clearer and more useful.

Minimum Detectable Concentration, Detection Limits and similar wording regarding detectability

- **Comment:** Discussion of minimum detectable concentration (MDC) and detectability need to be made clear, and used consistently and concisely. MDC and detectability are NOT the same and therefore are not interchangeable. Further there are cases where the terminology is used without definition, for example in Table 6.9 "LLD" is given units of activity concentration.
- **Recommendation:** Carefully review Chapter 6 and MARSSIM V2 for the correct usage, and thus consistent usage, of L_C, L_D, detectability, LLD, MDA, MDC and similarly related terms.

Complete the modernization of the measurement methods

- **Comment:** Methods described in Chapter 6 have not all been uniformly modernized. For example, $MDCR_{surveyor}$ appears from the context of the document that's value is arbitrary based on what the project team deems acceptable a survey efficiency. What is the applicability of $MDCR_{surveyor}$ for a GPS-enabled scanning system?
- **Recommendation:** Thoroughly review Chapter 6 and remove concepts and terminology that are no longer used or needed. Appendix H was scrubbed of obsolete and outdated instrumentation. A similar scrubbing of instrumentation and methods is warranted for Chapter 6. Were appropriate recent references should be cited.

Quantification of and Description of Uncertainties

- **Comment:** Proper calculation and quantification of the various sources of uncertainty is of paramount importance when it comes to determination of the confidence interval.
- **Recommendation:** Calculation and propagation of uncertainty should be consistent with other common guidance documents, such as NIST/ISO Guide to the Expression of Uncertainty in Measurement

Appendix H

Comment: The updated and expanded Appendix H assure that it will remain the top download for years to come. However, there is room for further improvement.

Recommendation: References to technical information contained in Appendix H needs to be properly cited. An example of where a citation of the literature is required is: Page H-10, lines 23-25. Specific scientific information needs to be properly cited

It is stated in MARSSIM V2 that "*…those interested in purchasing or using the equipment are encouraged to contact vendors* …". Adding references to the scientific literature is a way to give specific information to the reader without becoming a advertising platform or the appearance of an endorsement of a particular manufacturer.

Editorial Comments

Numerous editorial and specific comments are tabulated in the detailed draft comments. Additional editorial comments can be made available.

CHARGE QUESTION 2.2

- Please comment on whether the additional optional methodology for the use of Ranked Set Sampling (Appendix E) for hard-to-detect radionuclides is useful, appropriate and clear.
- The Ranked Set Sampling methodology requires a close, reasonable and provable correlation between an easy-to-measure attribute of the sample (e.g., soil sample size distribution) and the activity level of a hard-to-detect radionuclide. While challenging to implement in practice, the revisions include this optional method to assist sites with designing surveys for hard-to-detect radionuclides, which can be difficult and resource intensive to implement.

WHAT IS RANKED SET SAMPLING (RSS)?

- A method for constructing a random sample that potentially could considerably improve on simple random sampling (SRS) with some small additional effort
- Requires two methods of conducting a measurement, a rough (cheap) method and a precise (expensive) measurement
- The idea is to use the rough measurements to help us construct a better sample for the precise measurements
- It works best when the rough and precise measurements are highly correlated. However, it can still improve on SRS even when they are not.

IDEA BEHIND THE METHOD (example)

- Select 9 sites for sampling, divide them into 3 blocks of 3
- Perform rough measurement at each site
- Within each block of 3, use the rough measurements to classify the sites as “small,” “medium” and “large.”
- Choose the “small” site from the first block, the “medium” site from the second block, the “large” site from the third block.
- Do the precise measurement for these 3 sites. The remaining 6 sites are not used.
- Repeat as many times as desired.
- You could do the same thing based on 4x4 blocks, or 5x5. As far as I can tell, no one in practice goes larger than 5x5.

THEORY

- The estimator from RSS is unbiased to the same population mean as SRS, and has an equal or smaller variance (even if the correlation between the two measurements is low)
- It's possible to calculate a theoretical variance for the RSS estimator if the ranking is essentially perfect (i.e. correctly identifies the smallest, middle and largest values in each block – this doesn't require perfect correlation between the two measurements, but would need something close to that).
- The numbers in Tables E.1-E.4 do assume perfect ranking. Also, it's not stated explicitly but I believe there's an implicit normal distribution assumption in Tables E.1-E.3. Note also the comment about sample sizes being increased by 20% to account for possible missing or unusable data – maybe this is needed, but it counters the claim to greater efficiency compared with SRS.

COMMENTS FROM PANEL MEMBERS

- The description of the method is sound but the style of writing is more suited to a statistician than the general user. Would be good to add more motivation an explanation for the method.
- One panel member felt the method would not provide reasonable direction or assurance that hard to detect high activity concentration areas will be properly selected. Significant information and guidance on how to perform this type of assessment is missing.
- There should be better discussion of practical issues such as when you should use this method at all, what alternatives might be available (e.g. stratified sampling, regression estimator, ratio estimator), how the rough measurements should be determined and what sorts of correlations are needed to make the method useful.
- More detailed worked examples should be provided.

COMMENTS DURING THE DISCUSSION

- Questioning whether the assumptions are satisfied for nuclear survey applications – counts not normal distributions, not clear whether a proxy variable exists that has desired properties
- Need to discuss decision-based outcomes – if this method improves on SRS for calculating confidence intervals, it should improve it for other kinds of statistical outputs as well, but that isn't brought out in the discussion.
- Use calibration methods to use the rough measurements differently (this is similar to the comment about regression or ratio estimators)
- Include cautionary text about when not to use it.
- From one committee member: "Much of the major decision concepts are statistical, and the text is thick with statistician vernacular. It would be merciful if the statistician-ese could be dialed down and in more lay language."

CONCLUSION

- There needs to be a more careful explanation of the method to make this useful to someone without a statistics background
- Need better discussion of the usefulness of the method in the nuclear testing context we are talking about
- More thorough discussion of examples, including whether to use the method at all and how to use the results for other kinds of decision making
- Comparisons with alternative methods, such as regression or calibration
- In addition, numerous technical queries or suggested edits

Draft Responses to MARSSIM Revision 2 Charge Question 2.3

Charge Question 2.3 on Chapter 5 Examples

Please comment on the usefulness and accuracy of new and additional examples provided in Chapter 5.

Charge Question 2.3 Response

General Utility of Examples

Overall, the examples are useful and provide a means for users to see how the manual guidance can be implemented for practical situations. However, each example could use additional detail on calculations and use of the tables cited in the text. It cannot be assumed that the user will be well versed in either sampling theory or use of the statistical models employed in the examples.

First Three Examples are Identical to Rev. 1

1. Scoping Survey Checklist, pages 5-7 - 5-9 (identical to Rev. 1, pages 5-5 - 5-6) Should the checklist refer to the flowchart on page 5-2?

2. Characterization Survey Checklist, pages 5-17 - 5-19 (identical to Rev. 1, pages 5-16 - 5-17)
   The section on characterization surveys provides limited guidance on the scoping of contamination in matrices other than soils (e.g., ground water, surface water and sediments). While such information is important to assessing how site contamination is impacted, MARSSIM of itself does not provide guidance on how to use this additional information to structure site clean up or future site surveys. Should the checklist refer to the flowchart on page 5-3?

3. Remedial Action Support Survey Checklist, page 5-21 (identical to Rev. 1, page 5-20)
   Should the checklist refer to the flowchart on page 5-3?

Next Two Examples deal with WRS

4. Use of WRS Test under Scenario A, page 5-31 (modified from Rev. 1, pages 5-33 - 5-35)
   Example 4 in Rev 2 describes a Scenario A situation, changed from Rev.1 (was soil, now building surface). The description of the process for selecting the number of samples in the reference and test area is clearer in Rev 2 than in Rev 1.

5. Use of WRS Test under Scenario B, page 5-33 (new)
   This new example is similar to Example 4 again using a building surface. The description of the process for selecting the number of samples in the reference and test area is clear and needed as this type of analysis (Scenario B) was lacking in Rev. 1.

   Consider a cautionary note when not to use the WRS. The WRS is addressed in Section 8.4.1 (page 8-27). The reader may not be aware nor seek this guidance before applying effort in Chapter 5.

Example 6 deals with Sign Test

6. Use of Sign Test under Scenario A, page 5-34 (new)

   Provides a new example for a Scenario A decision on the number of samples to be selected. The example clearly identifies the input decisions needed by the project team on Type I/II decision errors, the $DCGL_w$ and the LBGR value. It also clearly shows the calculation involved and the use of the Sign Test table. What would be very helpful to users would be to demonstrate that as the significance of the Type I or II errors increase, from perhaps 0.05 to 0.10, that the measurement uncertainty for the method and instrument selected, must be much smaller (or with less precise measurements increase the number of samples required). This relates directly to the Type I and Type II decision error rates that are based on the project DQO values. This type of discussion then ties the decision made using the data obtained back to the original MQOs and DQOs.

   Same as before: Need a cautionary note when not to use the Sign Test. The Sign Test is addressed in Section 8.3 (page 8-19). The reader may not be aware nor seek this guidance before applying effort in Chapter 5.

Examples 7 and 8 deal with Additional Data Points

7. Determination Whether Additional Data Points are Required, pages 5-39 - 5-40 (modified from Rev. 1, page 5-39)

- Examples 7 and 8 demonstrate how to use the information discussed in 5.3.5 however, there are assumptions made for the initial calculation of the number of samples to be taken that should be stated in the example so it is clear to the reader. For instance, in Example 7, it is not obvious where the value 27 came from. There is a need to show this calculation and the use of Table 5.3 in the example.

8. Determination Whether Additional Data Points are Required, pages 5-40 – 5-42 (modified from Rev. 1, page 5-39)

Example 8 provides good contrast to the scenario depicted in Example 7 and lets the user see how MARSSIM can be an iterative process. A typo needing correction, page 5-41:

> "The grid area encompassed by a triangular sampling pattern of 10 m is approximately ~~86.6~~ 99.1 m^2, as calculated using Equation 5-3:"

Examples 9 and 10 Present Sampling Patterns

9. Random Sampling Pattern, page 5-45 (identical to Rev. 1, page 5-41)
   Example 9 is used to demonstrate a random sampling pattern for a Class 3 area. It incorrectly references Table I.11 as it should be Table I.12. This type of example is beneficial to the user providing guidance on random selection of the sampling locations. However, it is unclear how Table I.12 is used to select the locations depicted on the diagram. The method used should be explained in some detail.

10. Illustration of a Triangular Systematic Pattern in an Outdoor Class 2 Survey Unit, pages 5-46 - 5-47 (identical to Rev. 1, page 5-43)
    Example 10 provides a good means of showing the triangular grid sampling pattern based on the equations used in the chapter. The random start coordinates are stated as a foregone fact without any development. The stated location relates to Table I.12 but it is not obvious how you identify that random start point.

Example 11

11. Example 11 identifies a sample checklist for the FSS. It provides additional details not included in the Rev 1 version. Checklists such as these provide users with both goals to attain during the remediation and final survey status as well as serving as a final quality check that all items have been considered.

Starting with the same section on Final Status Survey, the two document descriptions and examples diverge significantly. Both contain flow charts identifying the flow paths for determination of survey plan, measurement locations and needs assessment. Rev 2 however goes into significant details about selecting the appropriate scenario and an expanded discussion of the gray region and the selection of a lower bound of the gray region.

Conclusions

- Of the 11 examples that appear in Chapter 5, five are unchanged, four are modified (additional information), and two are new.

- The new examples are a significant improvement and are needed. The unchanged examples were mostly reformatted such that they are easier to read and understand. The modified examples included some additional information or steps or changes to data such that they are better suited to inform the reader. A few typos were discovered in these Chapter 5 examples that need correction.

- These types of examples should be integrated into the flow of detailed front-to-back case studies

- In summary, the examples in Chapter 5 are a marked improvement over Rev. 1 examples.

Charge Question 3.1

3.1 Please comment on the revised description of how to set the Lower Bound of the Grey Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).

- The value of site-specific information about the residual level of radioactivity is likely to vary from site to site. The suggestion that the LBGR be set to a value near to the median seen in preliminary data is a good one so long as the preliminary data is reasonably informative. In cases where preliminary data is limited, adherence to some kind of heuristic rule (such as using ½ the $DCGL_W$) probably can't be avoided. Of course the sample size needed can be reevaluated as the data comes in and if the median of a first set of observations is far below the $DCGL_w$ the necessary sample size can be lowered. A development of a good case study (as an example) would probably be more persuasive than the very short description of using various possible preliminary data sources, without giving any specifics. Unless the preliminary data is strong it is natural to fall back on a (however inadequate) "hard and fast" rule.
- Re the above, consider adding a rule of thumb that the LBGR could be based on a gray region that is something like 1.67 times the σ of the measurements, for cases where the available data are very sparse.
- In Ch. 4, indicate clearly that in Scenario A the LBGR is the DL: it is the level below which the measurement is statistically considered to be below the $DCGL_W$.

3.1 Please comment on the revised description of how to set the Lower Bound of the Grey Region (LBGR) and its likely effectiveness in encouraging users to rely on site-specific information for doing so (Chapter 4 and Section 5.3).

- For scenario B, add some material describing use of a series of power curves as a function of the DL for different sample sizes to evaluate whether the DL for a "reasonable" study size is unrealistically large in relation to typical requirements.
- Consider moving the material in Sec. 4.12.3.1 into chapter 5, where it is also covered, with a higher-level summary of sample size determination being presented in chapter 4, and reference (as is already given) to the appropriate sections in chapter 5.
- Add two figures like Fig. 6.1 to Ch 4 to illustrate the relationships in Scenarios A and B, and describe in simple language how the relative shift (width of gray area) depends on the values of σ and the α and β error rates.
- In Section 5.3.3.1 include a short paragraph on "Considerations for Setting the LBGR" prior to the examples, discussing how the site-specific information along with some emphasis on conservatism is applied and why this was chosen as a basis for setting the LBGR.

Charge Question 3.2

Please comment on whether avoiding the use of the term “area factor” improves understandability of the elevated measurement comparison concept (Section 8.6.1).

- Overall there seemed to be agreement that avoiding the use of the rev. 1 ‘area factor’ designation and simply utilizing the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria, does emphasize the need for site-specific modeling or calculations.
- While the understanding of Section 8.6 is improved without the use of the term ‘area factor’. It was suggested that a figure could perhaps be added to enhance overall understanding.
- There are still several sections of the MARSSIM rev. 2 draft that continue to utilize ‘area factor’ and thereby introduce some level of confusion. These sections (i.e., 5.3.5.1, and 5.3.8) should be revised to remove the ‘historical approach discussions’ if appropriate, or to simply refer to the ratio of the $DCGL_{EMC}$ to the $DCGL_W$ rather than ‘area factor’, or delete any reference to ‘area factor’.

Charge Question 3.2

- Section 8.6.3, Example 14 specifically refers to the ‘area factor’, introducing confusion, and should rather refer to the ratio (and thereby also encouraging the importance of site-specific modeling).
- The inclusion of specific items associated with ‘area factor’ in Appendix O, while noted as historical illustrations, becomes rather confusing. No need to include the reasons for the change to not using ‘area factor’ in Rev 2, and these examples could/should be eliminated. Alternatively, include this historical discussion in one part of Appendix O, and remove discussions from elsewhere. Perhaps cite some examples of current State Regulator documents for clarity while clearly noting that MARSSIM rev. 2 encourages utilizing the ratio of the Elevated Measurement Comparison (EMC) release criteria to the wide-area release criteria.
- Glossary includes a definition of area factor but should perhaps not be included or should be made very clear that this is referring to historical definition (from earlier MARSSIM revisions).

Charge Question 3.3

Please comment on the effectiveness of the new organization of Chapter 4 (Considerations for Planning Surveys) to improve the understandability of the Chapter.

- The organization of Chapter 4, MARSSIM, Version 2 is considerably improved over MARSSIM, Version 1, and moving detailed derivations and calculations to Appendix O improved the flow of the information in the Chapter; however additional work is needed:
  - Chapter 2 provides more detail on survey types than Chapter 4, where it seems the material more properly belongs. At a minimum, Figures 2.4-2.8 should be included or referenced in Section 4.3.
  - Section 4.1.1 states Chapter 4 will address all survey types, but in the very next section limits the primary focus to the Final Status Survey (FSS). More discussion is needed on the other survey types, which as noted above could be drawn from material in Chapter 2.
  - Figures 4.1 and 4.2 should appear in reverse order, since 4.2 shows the general case, while 4.1 addresses the special case (i.e., the FSS). In addition, these two figures should not be separated by so much text; they should be discussed in the context of survey types in Section 4.3.
  - Section 4.2 opens with the seven steps of the DQO process, and the subsections following should expressly address those steps in order.
  - Section 4.8.3 should include a rationale for setting the MDC at less than 50% of the UBGR, since that is not discussed in Section 6.3 to which the reader is referred.
  - Consideration should be given to moving Section 4.12 Examples to Appendix O as well as some of the details regarding the calculation of the various types of DCGLs in Section 4.5.3 (i.e., retain a summary, but carry out the math in the appendix) to further improve flow.

Charge Question 3.3

Additional Editorial Remarks

- Section 4.7.1 lists “Basic Terms,” but does not refer the reader to the Glossary. In addition, the glossary does not include all the terms. Missing are: a) sample median, b) parametric tests, c) Student’s t test, and the word “decision” should be included in the term “Type I and Type II errors” to be consistent with the glossary.
- Sections O.5 and O.6 repeat Sections 4.12.8 and 4.12.9, respectively and should be removed.
- There are figures and boxed texts with incorrect references (e.g., in Figure 4.1, the box “Identify Radionuclides” refers to Section 4.3, but Section 4.3 addresses survey types).
- Consider combining Sections 4.4 and 4.5 and adding a precautionary statement to Eq. 4-3 that the dose/risk endpoints must be identical and able to be summed.
- Section 5.3.5 refers to Section 4.2.5, but there is no Section 4.2.5.
- Appendix A.2.1 and A.2.2. do not refer to the appropriate sections in Chapter 4.

Charge Question 3.4

Please comment on the effectiveness of moving derivations from Chapter 5 (Survey Planning and Design) to Appendix O to improve the understandability of the Chapter.

There is general agreement that the change to Chapter 5 by moving derivations to Appendix O is expected to improve understandability. The following were identified in comments as improvements:

- Flowcharts in Figure 5-1 (Scoping Survey), Figure 5-2 (Characterization and Remedial Action Surveys), Figure 5-3 (Final Status Survey), and Figure 5-4 (Integrated Survey Plan for FSS).
- Formatting improvements
- Improved cross-referencing to other chapters and appendices
- Improved section 5.3.1 concerning Scenario selection and LBGR determination

Charge Question 3.4

Suggested improvements for Charge Question 3.4 were primarily editorial in nature.

- Several locations where reference to Appendix O should be added, including Section 5.1 (Introduction), 5.3.3 (WRS), 5.3.4 (Sign Test) and 5.3.5 (EMC).
- Suggest adding to description of Table 5-2 to state that values were increased by 20% to account for missing or unusable data and uncertainty in the calculated value of N.