

## **Comments on GIEC's Draft White Paper**

### **Introduction**

1. The paper does not appear to fully address one of the fundamental issues we are facing, the application of statistical methods to data sets which have a relatively small number of samples. Moreover, the paper seems to lack a sufficient level of mathematical sophistication to aid us in our efforts to establish a reliable, easily achievable, and reproducible way to establish a background concentration for a hazardous constituent or regulated substance in soil. The stated purpose of the paper is to provide guidance on how to handle small sample sizes; however, the narrative does not seem to provide a defensible basis for the conclusions given in the included table.

*This paper defines concepts for further discussion of setting background limits, particularly when using small data sets. Previously submitted methods (Reference Athens submittal) more substantial in nature were not accepted by Georgia EPD. This paper focused on the concepts as a starting point for further discussion.*

*The basis for estimating the upper limit of background concentrations is statistical hypothesis testing. The paper suggests that one means of dealing with small sample size is to have a well characterized data set believed to represent typical background concentrations (e.g. Washington, 1992 and 1998) against which one could evaluate the small set of background samples for a specific release or site. Using typical nomenclature, the null hypothesis is:*

*Ho: Background Samples Collected = Expected Background for Constituent*

*With the alternatives:*

*Ha: Background Samples Collected <> Expected Background for Constituent or  
Ha: Background Samples Collected > Expected Background for Constituent.*

*At this point, we believe we have conceptual issues to resolve before proceeding with more substantial statistical problems and justifications.*

2. The paper says there should be a balance between false positive and false negative results in analyses performed on background data sets. However, the levels of confidence and the estimates of acceptable levels of error should also be proposed, as well as methods or criteria that should be used to judge the appropriateness of the balance, level of confidence, or acceptable range of error. In addition, examples of the effectiveness of this approach should be included, which are based on actual field data, to ensure that a statistically significant result that is intended has been achieved.

*We believe the error rates should be reached through consensus discussion once the conceptual framework has been agreed upon. The paper could present a discussion of statistical limits in common usage.*

3. The use of statistical analysis to determine background levels and to compare background levels to site contamination requires that statistical considerations be applied beginning with the development of the sampling plan and throughout the entire data analysis process. Sampling design, sample collection, and sample analysis must all be considered, as well as the impact of planning decisions on the usability of the data to meet the established goals of the project.

*We believe the development of sampling plans is adequately covered in regulatory guidance (e.g. SW-846, Volume II, Chapter 9 Sampling Plan) and is outside the scope of this document.*

### **Types of Background**

4. The definition of background presented in the first sentence differs with the description of background found in the Georgia Rules for Hazardous Site Response. According to the Rules, the background samples must not be affected by a release of a regulated substance. However, the paper restricts the definition to releases from the site. Also, the terms "ambient background," "regional background," and "site background" are used interchangeably in the paper. The terms should be clearly defined, consistently used, and transparent with respect to the examples provided.

*The Georgia Rules for Hazardous Site Response offers no definition of background samples but refers to background samples as being unaffected by releases of regulated materials .  
{ insert text here}.*

*We accept that the definition of background samples precludes releases from the site. We are concerned, however, with the extent of delineation and cleanup required if regulated materials from other sources are present in nominally unaffected soils. The development of estimates for site background above normally expected regional background is consistent with USEPA (19XX) and is also found in Washington ( 1992, 1998X) ( check for other examples in Michigan, Rhode Island, Ontario).*

5. The regulatory requirements under Georgia Hazardous Waste Management Act that address the assembly of a background data set should be discussed. Again, it is important to understand that background samples cannot be affected by the activities which caused the facility or site to be subject to this evaluation.

*We agree with this comment and propose to include the requirements of the Georgia Hazardous Waste Management Act. (Add also a discussion of urban environments where releases may have occurred which are unrelated to the listed site or are in addition to releases from the listed site)*

6. It should be acknowledged that background samples form a distinct population when compared to contamination samples. The comparison of data collected from two

different populations introduces difficulty in analyzing the sample results, specifically, determining from which population the sample comes.

*This comment is consistent with the need to develop estimates for regional background values to better understand the distribution and expected values for regulated materials naturally present in soil. This would be particularly useful information when regulated materials from other processes may affect the distribution of results in otherwise typical background samples.*

7. Background samples, when compared with samples from an area, portion, or variable depths at the site that may be contaminated, should be selected so that the location, depth, lithology and other variables are comparable, so that valid statistical computations can be made using the analytical results of those samples.

*See response to 3.*

8. The use of regional data sets require data assessment to ensure that samples are representative of site conditions and meet the regulatory requirements for background demonstrations.

*It would be incumbent on those who collect regional data to have adequate detail in such a study to include this level of information. We suggest that EPD establish criteria for the acceptability of regional background data and that they oversee the establishment of background a database. The Brunswick data could be put forth as an example of suitable regional data.*

9. More complex settings (e.g., urban areas, industrial parks, ecological habitats, etc.) should be addressed in a separate issue paper. This paper would present criteria for a more detailed sampling plan and evaluation program.

*We believe the issue regarding potential non-site contribution to background is an important part of the discussion and development of guidance. The development of a prescriptive sampling plan, however, should be site specific and is outside the scope of consideration.*

10. The paper moves from "Types of Background" to "Number of Background Samples" without discussing sample locations, sample types and sample patterns. These topics should be addressed in the sampling plan and should reflect the intended statistical applications and confidence intervals. Issues associated with difficulties in determining background concentrations in the vicinity of the site should also be discussed.

*The difficulty introduced by confounding affects will be require addressing the concept of "site background" from non-site related releases (see comment 4 and 9).*

## Number of Background Samples

11. The definition of statistics presented applies only to descriptive statistics and would only be applicable if no inferences, predictions, or comparisons are to be made to the background data set. At a minimum, the definition needs to be clarified with respect to its purpose regarding comparisons of data sets to background concentration data sets.

*Comparison of compliance samples with background samples is the next step after methods for the determination of background have been established.*

12. When the mean and standard deviation are estimated from data, these parameters can be used for a variety of statistical tests beyond just the estimation of the background distribution range for future samples, as stated in the paper.

*Comparison of compliance samples with background samples is a logical step after methods for the determination of background have been established. See 11.*

13. The sentence regarding judgmental sampling appears to indicate that statistical analysis of samples collected with this method is inappropriate because it is based on qualitative information, however, there are times when such analysis may be valid. Please provide a discussion of the basis of the restriction that you think limits its use in this case.

*The paper does not suggest that these methods are inappropriate but that they may be biased high in cases where samples are taken from locations known or suspected to be release areas..*

14. The justification for recommending 4-8 background samples to characterize a small site should be presented. Furthermore, "a small site" should be defined. As you may recall, the search for an appropriate method to determine how to handle, in a statistically and mathematically correct fashion, data sets that have relatively few data points began with GIEC's offer to identify the proper way to address this difficult issue. EPD did then and continued to appreciate GIEC's efforts; however, to date, a consensus has not yet been reached on the statistical evaluation of small data sets.

*Several factors must be considered when decided the appropriate number of background samples to be taken at a small site including statistical power, other sources influencing background, cost, and representative site coverage. We can add additional detail regarding the affect of sample size on evaluating differences from expected background means once we agree on a conceptual framework. We agree that a consensus is not apparent and none was implied by the paper. We hope to present the advantages and*

*disadvantages of collecting few samples and a logical approach for implementing this process.*

15. The justification for reliance upon regional background data sets to determine the acceptability of the site background data set is not fully explained. As mentioned above, the lack of quality control on regional data sets limits their usefulness in site specific applications.

*Regional background data allows a facility to utilize data that has been generated to augment site-specific data and strengthen the statistical power of the description of background. A substantial amount of regional background data has been generated for CERCLA and HSRA sites. This data was collected under the oversight of USEPA or EPD and should be considered available for use at other HSRA sites. See 1 also.*

16. The issue of characterizing sites where ambient background is not well defined should be more thoroughly discussed. The lack of published data to define ambient conditions should not preclude an adequate demonstration of background for a site. Please provide additional clarification on the point you are trying to make on this issue.

*We believe a large background data set is a prerequisite to evaluating site background estimated from small numbers of samples (less than 10). See also 1 and 4*

17. The proposed phased approach to a sampling design should include performance criteria which must be met before the data can be used in a statistical analysis for the site. The approach should also provide checks for meeting those criteria.

*We would be interested in developing such criteria together with EPD.*

18. The paper fails to include the use of background data sets in human health or ecological risk assessments, as provided in the Georgia Rules for Hazardous Waste Management and federal Comprehensive Environmental Response Compensation and Liability Act. For this document to have the greatest use and Utility to both EPD and GIEC, these data sets should also be addressed.

*Please clarify this comment. It appears to be outside the scope of the proposed document.*

### **Balancing Data Quality Needs to the Decisions**

19. The paper states that the detection limit should be below the risk reduction standard. In some instances the detection limit can be the risk reduction standard, as provided for in the Georgia Rules for Hazardous Site Response.

*The site sampling plan should specify maximum detection limits so this problem should not occur too often. When it does the RRS will be suggested as the maximum detection limit.*

20. The presence of greater than 15% non-detects does not require the use of non-parametric Statistical methods. Furthermore, the treatment of data sets that contains multiple detection limits should be addressed. There are published methods for handling censored data sets (data sets that include both the detected and non-detected samples). This issue merits a separate paper. Moreover, EPD recommends that before the first sample in a field investigation is collected, the laboratory selected to conduct the analyses should be advised that multiple detection Limits on analyses for the same hazardous constituent or regulated substance tend to complicate the straightforward analysis of the data that is sought by all involved (See comment # 3).

*This level of discussion can be addressed once we have a consensus on the approach to this issue. It is not an issue for small sample sets, as non-parametric methods and published methods for dealing with censored data sets are not sufficiently robust for this situation.*

*(Editorial notes:*

*The multiple detection limit issue comes from using multiple sources for data, not from changes within a single laboratory (usually). Resampling or using multiple sources may reflect recognition that the background results are more complex than expected.*

*I believe they have misinterpreted the paper. The paper says that for data sets with greater than 85% non-detect values, defaulting to 1/2 the detection limit will have a greater impact on the estimate of the confidence interval than real site data. At > 85 % NDs the data set is no longer normal and should be treated non-parametrically.  
End of editorial notes)*

### **Practical Background Sampling Design and Interpretation**

21. The term "anthropogenic" as presented in the goal for background determination. should be defined.

*" relating to or influenced by the impact of man on nature <~ ecosystems> ",  
Webster's New World Dictionary.*

22. The effect of sample size on the confidence of the interpretation of the results should be quantitatively evaluated and discussed. The methods used to estimate the confidence level for each sample size interval, as provided in the sample size matrix in the table, should be discussed.

*This level of discussion can be addressed once we have a consensus on the approach to this issue.*

23. Using regional values when all the site specific values are "non-detect" is not appropriate. If all the values are non-detect for a specific constituent, then logically the background

value would be the detection limit. Similarly, as the number of non-detects increases in a data set, the background value for that data set should approach the detection limit.

*This assumes the error rates are not of concern (see 2) and that small sample sizes would have acceptable power. A better estimate when one has not rejected the null hypothesis (see 1) is from the larger data set for regional estimates.*

24. The "Methods to Calculate Upper Background Limit" in the table do not include percentiles which can be used for both parametric or non-parametric evaluations of data sets. The decision criteria for using alternatives to the listed methods should also be included. References and examples, should be provided to verify the applicability of the stated methods to the small data sets.

*The table provides two options: 1) default to regional background, or 2) use the highest concentration detected. In the case of all NDs this would be the detection limit (although one should consider the former as more reliable, see 23).*

*Using percentiles with small data sets (less than 10 samples) may not give suitable error rates and are not robust (see also Washington, 1992).*

25. The observation that a method was used by EPD on some sites does not mean it is a preferred method for data analysis at all sites that EPD regulates as inferred in the table.

*This is one of the reasons for a guidance document.*

26. The effect of the low confidence associated with background values in sample sizes of eight or less should be discussed. The low confidence indicates that background values presented by statistical analysis are not reliable and a qualitative evaluation of the on-site versus off-site data may be more appropriate.

*This is the reason for recommending assembling a regional data set. When less well defined background estimates are used, the chance of concluding that a sample is greater than background when it's not, or concluding that it's less than background when it's not, increase the probability of decision error. Therefore, the collection and analysis of reliable background is important to the regulated community as well as the regulators. If this point was not clearly communicated the paper must be revised to bring this point across.*

## Summary Comments and Recommendations

27. The proposed compilation of the existing background data accepted by EPD as a first step in establishing a regional data set will not eliminate the need for site specific background demonstrations as indicated earlier in our comments. Therefore, the benefits and expenditure of resources required to develop a reliable regional data set are not intuitively obvious. Please explain what you think the value of this effort would be to EPD and GIEC, especially with respect to eliminating or minimizing future dispute regarding site specific background concentrations.

*We agree that having a regional data set does not eliminate the need for site specific evaluation. However, when less well characterized background estimates are based on small sample numbers, the chance of concluding that a sample is greater than background when it's not, or concluding that it's less than background when it's not, increase. Therefore, the collection and analysis of reliable background for comparative purposes is important to the regulated community as well as the regulators. The use of regional background data allows a facility to utilize data that has been generated to augment site-specific data and strengthen the statistical power of the background estimate. A substantial amount of regional background data has been generated for CERCLA and HSRA sites. This data was collected under the oversight of USEPA or EPD and should be considered available for use at other HSRA sites*

*Another benefit to the use of regional background data is cost-savings for the regulated community which should be viewed as a benefit to all concerned. Inefficient spending of investigative resources may mean that less money is available for remedial efforts. The alternative is to perform more detailed background measurements for each site.*