Office of Pesticide Programs' Comparison of Allender, RDFgen, and MaxLIP Decomposition Procedures

February 1, 2000

Presented to FIFRA Scientific Advisory Meeting
March 1, 2000

List of Abbreviations

BDL	Below Detection Limit
CDF	Cumulative Distribution Function
CLT	Central Limit Theorem
CV	Coefficient of Variation
DEEM	Dietary Exposure Evaluation Model
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
IQR	Inter-quartile Range
LOD	Limit of Detection
LOQ	Limit of Quantitation
MaxLIP	Maximum Likelihood Imputation Procedure
MLE	Maximum Likelihood Estimation
PDP	Pesticide Data Program
RDFgen	Residue Data File Generator
SAP	Scientific Advisory Panel
SD	Standard Deviation

Table of Contents

EXECUTIVE SUMMARY

On May 26, 1999, the Environmental Protection Agency's (EPA) Office of Pesticide Programs (OPP) issued a draft document entitled "Use of the Pesticide Data Program (PDP) in Acute Risk Assessment" which identified a statistical methodology for applying existing information from the U.S. Department of Agriculture's (USDA) Pesticide Data Program (PDP) report to risk assessments of the acute exposure to pesticide residues in food (see 64 FR 28485-28487; also see EPA's web page--www.epa.gov/pesticides). This methodology (hereafter referred to as the "Allender method") provided a statistical procedure for estimating single-serving pesticide residue distributions from distributions of composite samples of fruits and vegetables. OPP has used the Allender method to generate and then incorporate single-serving data into a probabilistic exposure estimation model, such as the Monte Carlo method, in order to estimate acute dietary exposure to pesticide residues in foods.

In May, 1999, OPP also presented the Allender "decompositing" method to the SAP for its scientific review and recommendations. As part of its presentation, OPP acknowledged a number of limitations and inadequacies in this methodology. Although the SAP stated, "EPA has identified a reliable statistical methodology for applying existing information from the U.S. Department of Agriculture's (USDA) Pesticide Data Program (PDP)," the Panel nevertheless recommended that OPP actively explore the feasibility of using other methods (specifically, maximum likelihood methods with censored data) which might better deal with the issues associated with acute dietary exposure estimation. Several public commenters in response to the above-cited May 26,1999 FR Notice also made this recommendation. Accordingly, OPP has investigated available alternative methods and compared them to OPP's current method.

The purpose of this document is to provide an update on the issues being addressed by OPP concerning the decompositing procedure and the progress, to date, on its investigations into alternate methods. This paper briefly describes the three decomposition methods under consideration and provides details on OPP's comparison of these procedures: (a) the "Allender method," (b) "MaxLIP" by JSC Sielken/Novartis and "RDFgen" by Novigen Sciences [the authors of the MaxLIP and RDFgen methodologies are providing additional background information in their separate presentations to the SAP]. OPP invites the SAP to compare the characteristics and behavior of these three methods and to make recommendations, if desired, as to which method might be most suitable for decompositing field trial, market basket, PDP, or other food residue data for use in probabilistic acute dietary (food) exposure estimation.

OPP's comparison of the different methods was done by using both (a) theoretical data designed to reflect differences in such characteristics as skewness, censoring, number of samples, and number of distributions and (b) empirical (real world) pesticide data collected by USDA's Pesticide Data Program (PDP) and others. Based on this analysis, OPP makes the following observations with respect to the three decomposition methods and their performance characteristics when applied to both theoretical and empirical (actual) data:

(1) Based on the analysis using both hypothetical and empirical data sets, estimates of the high percentiles of daily exposure calculated using residues measured in composite samples are much lower than estimated exposures using "decomposited" residue values. Composite residue values tend to underestimate daily exposure by 30% - 50% at the upper percentiles.

(2) All methods appeared comparable and seemed to do reasonably well at predicting single-item *residues* at up to approximately the 90th percentile, regardless of the data set which was used. This was true of both the theoretical and empirical datasets. As the number of distributions increased, moderate censoring was imposed, or number of data points decreased, the ability of the methods to predict the upper percentile residue values appeared to deteriorate to varying degrees.

(3) The presence of multiple distributions and censoring appear to have the most effect on each methods ability to adequately deconvolute residue values while skewness of the distribution and number of composite residue values seemed to have the least.

(4) In many cases, the RDFgen and Allender procedures appeared to predict too large a "spread" in the data, particularly in the lower percentiles. Nevertheless, this did not appear to affect the exposures (as predicted by DEEM) in the region of regulatory interest (e.g., >95th percentile)

(5) Despite the findings in (2) and (4) above, the most accurate decompositing method rarely overestimated or underestimated the exposure of the 99.9th percentile by more than 15%, compared to the calculation using the parent data set, when using hypothetical data. The differences between the estimates obtained using the best method and the parent data set were even smaller at lower percentiles.

(6) All methods seemed be able to predict the 99.9th percentile *exposure* (as determined by DEEM) reasonably well and no method appeared to have a significant bias toward over- or under-prediction. At the 99th percentile exposure and below, the methods appeared to be essentially equivalent, with each method predicting the same exposure as the original (parent data).

In addition to seeking the SAP's comments and recommendations, OPP will shortly make its analysis comparing the above three decompositing methods available for public comment along with additional information on MaxLIP and RDFgen. After consideration of the SAP's comments and all public comments received, OPP will issue a revised policy.

I. INTRODUCTION

In March, 1998, OPP's Health Effects Division presented a draft Monte Carlo guidance document to the FIFRA Scientific Advisory Panel. As part of this guidance document, OPP stated that it:

> ... will not allow use of monitoring data as a distribution of residues for most raw commodities, because data from composite samples do not adequately represent the range of residues in a single serving size sample, and the relationship between the residues measured in a composite sample and the range of residues in the individual samples that make up the composite is not established for most chemical/commodity combinations. From limited data that are available, OPP has observed that residues in single serving samples can be higher by an order of magnitude or more than residues in the corresponding composite sample. While field trial data are also based on composite samples, they are generally measured at the "farm gate." Because residues may decline during shipping, handling, and/or processing before food consumption at the "dinner plate," field trial data are considered sufficiently conservative for use in an acute dietary risk assessment.

In short, OPP was concerned that unit-to-unit variation in pesticide residues within a composite sample could be significant and could potentially result in significant underestimates of acute risk if such calculations relied on residue data based on residue levels in *composite* sample. OPP believed that this issue deserved further attention since individual residues in a composite sample can at times significantly exceed the average residues which would be measured in a composite sample.

The March 1998 SAP agreed with OPP stating:

> ...the use of monitoring data derived from composite samples seems inappropriate for direct use in acute dietary unchanged exposures for those reasons described. It is clear that if the Agency is protecting against single-day exposures, it would be inappropriate to utilize composite samples for evaluating acute risks... It would be incorrect to use these data from composite samples without adjustment.

In May, 1999, OPP returned to the SAP with this issue, proposing a methodology (Allender method) for "decompositing" the residues obtained from composite samples. That is, the methodology, when applied, would permit the use of monitoring data from PDP, FDA or other monitoring programs in which residue data are collected on *composite* samples to be "statistically adjusted" such that the residues on *individual items* which comprise that composite can be simulated. As stated in OPP's background document for the SAP session,

> The challenge to OPP has been to extrapolate from PDP composite data to provide single unit values for use in acute risk assessments. In statistical terms, given composite samples collected by the USDA, OPP is faced with the challenge of estimating the parameters that describe the original population of residue concentrations in servings of fruits or vegetables. Specifically, the problem is to estimate the population mean μ and the population variance (σ^2) from a set of composite samples where only the composite sample mean **(**x_bar**)**, the composite sample variance (s^2), and the number of units in each composite is known. With the estimation of the population parameters (μ and σ^2) and assuming that the distribution of residues in fruits and vegetables follows a lognormal distribution (as established in previous goodness-of-fit studies), the function that describes chemical residues on fruits/vegetables is adequately established and ready for application into one of the components of the Monte-Carlo model for the acute risk assessment.

As part of its presentation, OPP acknowledged a number of limitations and inadequacies in the OPP-proposed methodology. For example, OPP stated that the methodology assumes that the individual items which comprise the composite sample are selected at random and are independent. In reality, the sampling program is designed such that the individual units in any given composite likely share the same treatment history and thus the individual units within a composite are not randomly selected. In addition, the procedure assumes that there is no correlation between individual item residues in a composite sample; in reality, there is likely to be at least some correlation because of the shared treatments. Finally, the proposed procedure generally requires a minimum of 30 composites which have residues which exceed the limit of detection in order to assure that there is enough representation in the sample such that there is adequate simulation of the entire range of potential single serving residues.

Although the SAP stated that "EPA has identified a reliable statistical methodology for applying existing information from the U.S. Department of Agriculture's (USDA) Pesticide Data Program (PDP)," the Panel nevertheless recommended that OPP actively explore the feasibility of using other methods (specifically, maximum likelihood methods with censored data) which better deal with the issues associated with the exposure estimation issues that were the focus of this session. Specifically, the Panel indicated that it was encouraged by the methodology and data introduced by a public commenter, and, although the Panel were unable to critically review this information, it encouraged OPP to conduct further exploration and either adopt this methodology or incorporate the concepts presented into the methodology which is ultimately selected.

As a follow-up to the SAP's recommendations, OPP has been investigating methods which could be used to better estimate single-item residue distributions from distributions of composite residue measurements. *The purpose of this current background document is to provide an update on the issues being addressed by OPP concerning the decompositing procedure and the progress, to date, on its investigations into alternate methods.* This paper

briefly describes three decomposition methods under consideration and provides details on OPP's testing of these procedures. OPP invites the SAP to compare the characteristics and behavior of these three methods and to make recommendations, if desired, as to which method might be most suitable for decompositing field trial, market basket, PDP, or other food residue data for use in probabilistic acute dietary (food) exposure estimation.

This document is divided in six sections. Section I is this introduction which discusses past SAP reviews and comments on the issue of composite sampling and acute risk assessment. Section II of this document provides general background information on the decomposition procedures under consideration and being presented to the SAP. Specific, and more detailed, background documents for each of the three decomposition methods have been provided to the panel and are also available in the Public Docket. Section III of this document describes some of the standardized sample data sets and procedures (but not results) used to investigate how well the model-predicted single-item values correspond to actual single-item values from whence the composite samples came. This is done for theoretical data sets (which were specifically generated to simulate as many real-world data anomalies as possible) in which "back-prediction" of original (theoretically-based) single-item values by the proposed decomposition routines from simulated composites as well as with actual (empirical) data sets generated by USDA's PDP and others. The generated data (i.e., predicted single-serving residues) resulting from application of each of the three decomposition methodologies is presented in Section IV and compared to the original data which the decomposition methodologies were designed to reproduce. The next section (Section V) summarizes overall OPP observations and conclusions regarding these methods and their abilities to predict a single-serving distribution from a distribution of residues in a composite sample. References are presented in Section VI. Finally, the questions which OPP would like to have addressed by the SAP are presented in the last section (Section VII).

II. DECOMPOSITION PROCEDURES UNDER CONSIDERATION

Three decomposition procedures are being presented to the SAP for their review and consideration. One procedure (**Allender Method**) was formally presented to a previous SAP during May 1999 and was reviewed by the SAP at that time. Also introduced to the SAP at that time by public commenters were two additional procedures which the Panel was asked to consider: the **MaxLIP** (for Maximum Likelihood Imputation Procedure) decomposition method which relied on maximum likelihood estimation techniques for censored data and **RDFgen**, a method introduced by Novigen Sciences. A brief introduction to each of the three procedures is provided below. Additional information concerning the details of each method is being provided as supplementary material to the SAP and is being introduced into the relevant OPP docket.

A. MaxLIP Method

One method OPP is investigating was a method introduced at the May, 1999 SAP by Dr. Robert L. Sielken, Jr. of JSC Sielken and Dr. Leslie Bray of Novartis Crop Protection. This method (called MaxLIP for Maximum Likelihood Imputation Procedure) is a three-step procedure which uses maximum likelihood estimation techniques and Monte Carlo simulation to estimate the distribution of single serving residue concentrations from a database of composite residue concentrations. The final estimate provided by MaxLIP is a distribution of single-item residue concentrations which is applicable for use in a probabilistic dietary risk assessment for food items that are generally consumed individually or in small numbers.

Briefly, the procedure consists of three major steps:

> STEP 1: This step consists of using a maximum likelihood estimation (MLE) technique to determine the parameters of a lognormal distribution of single-item residues that would, via Monte Carlo techniques, generate the best approximation to the observed sample distribution of composite residue concentration values. In essence, the MLE procedure is used to generate a lognormal distribution of the single-item residue concentrations (or a mixture of up to five lognormal distributions if the measured composite residues are assumed to have come from a series of distinct lognormal distributions) that is "most likely" to have generated the observed sample of measured composite residue concentrations. The result of this step is an estimated mean and variance of a lognormal distribution (or a series of means and variances of lognormal distributions if more than one source distribution is assumed) that maximize the likelihood function. Since there are no closed-form solutions for the mean and standard deviation that maximize the likelihood function, MaxLIP uses computer-intensive numerical search procedures. Although this step of the method assumes that the distribution of residues in treated *single-servings* from whence the composite residue concentrations came is lognormal, the method makes no implicit assumption concerning the distribution of composite residue concentrations (i.e, the distribution of composite residues is not assumed to be lognormal or any other specific distribution).[1] This step can incorporate such factors as percent crop treated, treated composites containing non-detectable (<LOD) residues, composites comprised of items whose residues are positively correlated, composites comprised of a variable number of items, and composites comprised of both treated and non-treated items. Significantly, the method permits the single-serving residues to be derived from up to five *independent* lognormal distributions, and uses imputation procedures for "below

[1] As stated by the SAP in its report on the May 1999 presentation, even if the distribution of underlying single-sample residues were perfectly lognormal, the residue found in each composite is effectively a series of weighted arithmetic means of some number of single samples. A distribution of weighted arithmetic means of samples collected from a lognormal population is not expected to be lognormal itself.

limit of detection" (BDL) observations, rather than assigning these to be ½ LOD or aggregating these as one single-valued (½ LOD) part of the distribution.

STEP 2: Given the lognormal distribution of single-item residues generated in Step 1 (or the series of up to five lognormal distributions), Step 2 of the MaxLIP procedure consists of repeatedly sampling the single-item residue distribution to generate samples of single servings which, when composited, match the composite residue concentrations present in the composite sample database. This step uses the (lognormal) approximating, single-item distribution to generate a set of single-item residue values which more closely match the specific, observed composite residue values present in the database. This step provides a distribution of single-item residue concentrations which is more robust than the distribution produced in STEP 1 in that it better matches the observed composite residue concentrations and is less dependent on the theoretical characteristics of the parent lognormal distribution, or series of distributions, of single-item residues generated in STEP 1.

STEP 3:. Step 3 provides the option to perform Steps 1 and 2 not only for a single lognormal distribution but also for a mixture of two, three, four, or five lognormal distributions. The MaxLIP user can run the MaxLIP software specifying that 1, 2, 3, 4, or 5 lognormal distributions be considered. Each such run would deconvolute the composite distribution into a mixture of the specified number of singe-item residue distributions. The result would include a separate mean, standard deviation, and proportion for each separate distribution as well as the joint log-likelihood value. The user can compare these log-likelihood values using a likelihood ratio test to determine the statistical significance of the different log-likelihood values for different mixtures.[2] It is the responsibility of the user to select the appropriate number of distributions into which the input composite distribution is deconvoluted (between one and five) based on what is expected from

[2] For example, the user may choose to deconvolute the distribution of composite values into between one and five separate distributions, each of which has its own separate and associated mean, standard deviation, proportion and loglikelihood. If the user chose to deconvolute the composite input distribution into two separate and distinct distributions, one result of the deconvolution might be one lognormal distribution with a mean of 2.34 ppm and a standard deviation of 2.06 ppm comprising 62% of the sample, with a second distribution with a mean of 1.32 ppm and a standard deviation of 0.78 ppm comprising 38% of the sample. Associated with this deconvolution into two separate distributions would be a loglikelihood value which would be used to test if two distributions were significantly better than one distribution or significantly worse than three distributions. It is the responsibility of the user to select the appropriate number of distributions into which the input distribution is deconvoluted (between one and five) based on the program's output of the loglikelihood value and the results of the user-initiated and conducted likelihood ratio test.

agricultural practice and/or based on the program's output of log-likelihood values and the results of the user-conducted likelihood ratio test[3]

B. RDFgen Method

A second method of decompositing composite residues into their constituent individual components was presented by Novigen Sciences, Inc. In the methodology used under this scheme, each observed composite sample is assumed to be comprised of individual, single-items whose (unmeasured) residues are obtained from a lognormal distribution with a mean residue equal to the measured residue in the composite sample from which the single-item residues are derived. In addition, the standard deviation associated with the individual residue values in each composite is a function of the standard deviation of the measured composite samples (specifically, the assumed standard deviation is equal to the product of the standard deviation calculated from the composite samples and the square root of a user-estimated average or typical number of single items within a composite). This latter calculation is derived from the Central Limit Theorem[4]. For each measured composite sample (comprised of individual items whose defined mean is equal to the measured composite residue value), individual residue values are selected from a simulated lognormal distribution with the defined mean and standard deviation via a Latin Hypercube simulation. Sampling is continued until such time as the mean of the individual single-sample residues selected is within 5% of the measured composite concentration. For treated composite samples with BDL observations, the RDFgen methodology assumes that residues are present at ½ LOD (and thus the composite residue will be decomposited into a series

---

[3] Testing of the MaxLIP program by OPP revealed some inconsistencies in the MaxLIP-generated loglikelihood values. This was communicated to Dr. Sielken who indicated that this was likely a result of the iteration procedure locating a *local* maximum as opposed to a *global* maximum and that this would be corrected in subsequent versions of MaxLIP. OPP believes that the anomolous decreases in the loglikelihood values as the number of specified distributions is increased and the occasional generation of unrealistic means and standard deviations are symptomatic of this situation and will be eliminated in a subsequent version of MaxLIP.

[4]Briefly, the distribution of composite sample residues represents a distribution of sample means since the measured value in a composite sample is simply the mean of the residues in the single-item units making up the composite. The Central Limit Theorem states that the distribution of these means (i.e., the distribution of the measured composite sample residues) will be nearly normal provided the sample size is sufficiently large; *this is true regardless of the nature of the specific distribution of the single-item residues*. The Central Limit Theorem goes on to state that the standard deviation of the means (i.e., the standard deviation of the composite sample measurements) is $\sigma/(N^{1/2})$ where σ is the standard deviation of the parent population and N is the size of each sample (here, the number of items in a composite). Thus, it is possible to estimate the standard deviation (and variance) of the parent population of residues by taking the product of the standard deviation of the composites and the square root of the number of items in each composite.

of single-serving residues whose mean value is ½ LOD). Such a process is repeated for each composite value until a distribution of single-item residue concentrations is generated which corresponds to (and is consistent with) each sampled composite. The result is a series of synthesized single-item pesticide residue concentrations which can be used in a probabilistic exposure assessment[5].

C. Allender Method

The Allender procedure was originally presented to the SAP in May, 1999 and the initial background material made available to the SAP at that time has been made available again to the current SAP. The method is applied only to composite samples with detectable residues, with a separate single-valued distribution at ½ LOD assumed for the treated non-detects. Briefly, the method relies on three pieces of information to estimate a distribution of single-item concentrations: (i) the mean of each composite sample collected; (ii) the number of samples collected (n); and (iii) an estimated average number of units within each sample (N). The method proceeds in two steps:

(1) Calculation of the mean and variance of the sample composite values[6]; and

(2) Adjustment of the composite sample variance estimated in (1) above to correct for the number of single-item units comprising the composite.

Specifically, step (2) above consists of multiplying the variance obtained in (1) by $\sqrt{N}$, in accordance with the Central Limit Theorem (CLT). That is, the Allender method assumes that there is one (lognormal) distribution or residues in single-items from which the composite sample is obtained, that the assignment of each item to a composite is random and independent (and that

[5] The background document prepared by Novigen Sciences on the DEEM software program presents a recent modification of the RDFgen program which OPP has not had the opportunity to use or test in its comparison protocol. The modification proposed by Novigen incorporates the concept of a coefficient of variation in refining the estimate of the sample standard deviation and is expected to produce an estimate single-serving distribution which would be considerably "narrower" than the one currently produced by the version of RDFgen currently available to OPP for testing.

[6] The background paper presented to the SAP in May, 1999 indicated that the mean and standard deviation of the composite samples were to be calculated on a log-basis (i.e., the composite residue values were to be log transformed, with the mean and standard deviation of the log-transformed values calculated. These log-transformed would then to be converted back to the real scale by using back-transformation formulae). The Allender calculation done in the present paper does not perform this transformation and the means and standard deviations are calculated directly from the data with no intermediate log-transformation.

no correlations between residues in a composite are present), and the CLT applies.

The information developed in steps (1) and (2) (namely an estimated mean and standard deviation of the distribution of single-item residues) is then used to create the individual values that comprise this (lognormally-distributed) population and are used as individual item residue inputs in OPP's acute dietary risk assessments.

Overall, then, two of the three methods (Allender and RDFgen) use the standard deviation of the composite samples to calculate the standard deviation of a parent distribution (by way of the CLT and $N^{1/2}$ factor). The Allender method assumes that the composites are derived from a single (common) lognormal parent distribution from which all composites samples are obtained, while the RDFgen method assumes that each composite is derived from its own unique lognormal distribution with a mean equal to the mean of the specific composite sample and a standard deviation equal to $N^{1/2}$ * SD of all composites. Each of these methods assigns a value of ½ LOD to treated samples with BDL residues and uses an average or typical number of items in a composite to estimate the standard deviation of the single-serving population. In contrast, the third method under consideration (MaxLIP) uses a maximum likelihood estimation procedure to estimate the "most likely" (or series of up to five "most likely") distributions from which the observed distribution of composite residues might be seen. Residue values associated with treated BDL samples (i.e., censored observations) are imputed using MLE techniques, and the number of items in each composite (as opposed to an average of typical number over all composites) is explicitly considered. In a second step, the MaxLIP method produces from this distribution a series of single-serving values which, when composited, would (nearly) exactly match the "observed" composite distribution.

Given these similarities and differences in the basic approach, OPP believes it to be useful to subject a series of standardized data sets to each of the three decomposition methodologies being presented to the SAP. The design of this comparison is discussed in Section III, while the results of the comparison are discussed in Section IV.

III. COMPARISON OF PROTOCOLS AND METHODS

In order to understand better the three decomposition methodologies being presented to the SAP, OPP initiated a protocol to evaluate the distribution of *model-predicted* single-item residue concentrations developed from composite data. The protocol used both actual measured (i.e., empirical) and simulated (theoretical) single-item residue concentrations. OPP believes that it is appropriate to use both empirical and theoretical data in comparing the three decomposition schemes (Allender method, Novigen's RDFgen, and JSC Sielken's MaxLIP) because no single set of data, considered separately, permits a wide range of conditions and situations to be fully considered and evaluated. Given the nature of the residue values and data collection protocol, OPP believes that it is important that each method be evaluated and compared through as comprehensive a set of data and conditions as possible.

This section provides details of the protocol used to compare each of the three proposed decomposition methods. The protocols cover both the testing of theoretically-generated data and empirical (real-world) data. OPP considers the use of *theoretical* distributions and their artificially-generated (simulated) composites to be useful since it allows for controlled investigation of the impact of important real-world factors (such as skewness of the distribution, level of censoring, presence of multiple distributions, and the number of available composite values) on the simulated computer-generated single-serving residues. Although limited in quantity and availability, the use of *empirical* data from specially conducted composite/single-serving studies is similarly considered to be valuable since it permits comparisons between decomposited data based on actual analytical measurements of residues in composited samples and the corresponding residues measured on a single-item basis.

The comparison was conducted in two parts in accordance with the above rationale. The first part involved comparing the predictive capabilities of the three decomposition protocols based on *theoretical* data which was artificially generated to meet testing requirements (e.g, varying skewness, degree of censoring, number of component distributions, etc). The second part involved comparing the prediction capabilities of the three decomposition protocols based on *empirical* data which was produced by PDP (including a discussion of PDP data on aldicarb on potatoes and analysis of a second recent special study conducted by PDP which has not yet been formally released) and industry (diazinon on peaches) as part of a series of special studies designed to investigate the composite-single serving relationship. These are described in further detail below.

The comparison protocol used by OPP did not incorporate a number of factors which can be considered and incorporated by one or more of the decomposition procedures being presented to the SAP. For example, the theoretical analysis done here did not incorporate percent crop treated factor, assumed a uniform number of items in a composite, considered that each composite was comprised of only treated items, and assumed random assignment of single items to a composite (no correlation of residues within a composite). Although the degree of censoring was evaluated, censoring was limited to 33%. In reality, censoring of PDP or other monitoring data may be substantially higher than this. Nevertheless, OPP believes that the analysis conducted may serve as a useful adjunct to the Panel in assessing the ability of each decomposition method bing presented to adequately predict residues in single-serving items.

A. Comparison Using Theoretical Distributions

The objective of this first part of the comparison scheme was to evaluate each of the three methodologies described above in reliably “back-predicting” a given log-normal parent distribution of single-item residues, with each decomposition procedure’s “back-prediction” ability based on a set of simulated composites (10 or 30 composites of 15 “items” each)

developed (or chosen) from the parent distribution[7].

The accuracy of the predicted distribution was assessed by evaluating how well the model-generated single serving data "match" or otherwise compare to the original data. Such a comparison was done by comparing how well the predicted mean, standard deviation, interquartile range, absolute range or spread, and upper-percentile residue values generated (or implied) by each methodology match with the corresponding statistics associated with the original, or "test", parent distribution. This comparison of numerical values such as the predicted vs. original mean, predicted vs. original standard deviation, etc. was extended (and more effectively conveyed) by also examining *graphical plots* of the distributions and performing one or more statistical tests which compare the equality of distributions. A wide variety of graphical methods have been developed for this manner of exploration including frequency histograms, stem and leaf plots, dot plots, line plots for discrete distributions, box and whisker plots, and one-way scatter plots. These graphical methods are all intended to permit visual inspection of the density function corresponding to the distribution of the data and can assist in examining the data for skewness, behavior in the tails, rounding biases, presence of multi-modal behavior, and data outliers. Some of the specific graphical and other analyses which were performed in this decomposition method comparison are described below:

> One-way scatter and box-and-whisker plots can be a very effective graphic display for summarizing the distribution of a data set. They permit not only a graphical point-by-point comparison of two or more distributions, but also effectively display various significant percentiles (e.g., 25^{th}, 50^{th}, 75^{th} etc) and outlying (or extreme) data points. Box plots provide easily explained and comprehended visual summaries of:
>
> - the center of the data (median - the center line of the box)
>
> - the spread in the data (inter-quartile range - the box length)

---

[7] Specifically, a composite value is simply the average of the items comprising that composite. Thus, as an example, a total of 450 single items were randomly selected from a defined parent distribution (e.g., a lognormal distribution with an arithmetic mean of 0.1 and a standard deviation of 0.1) using Crystal Ball software. A total of 30 composite values were simulated by averaging the generated values from the parent distribution in contiguous blocks consisting of 15 values each. Thus, the first 15 values were averaged to produce the estimate of the first composite, the second contiguous block of 15 were averaged to produce the second composite value etc. until the last contiguous block of 15 is averaged to produce the 30^{th} composite. In this way, a total of simulated 30 composite values are generated which can be entered into the decomposition software in an attempt to "re-create" the original parent distribution. Each decomposition routine was assessed by comparing how closely the generated values (or the distribution of generated values) matched those of the original lognormal parent distribution with defined mean and standard deviation.

• the skewness (quartile skew - the relative size of the box halves)

• the range (whiskers - lines from the ends of the box to some other elected endpoint, e.g., the last data point within a distance of 1.5x the inter-quartile range (IQR) from the mean.)

In the box plots used in this paper here (for an example, see Figure 1), the whiskers extend to the last data point within one step beyond either end of the box, where a step is defined as 1.5 times the inter-quartile range. Data points beyond 1.5 steps of either end of the box are plotted as individual points. The values at the extreme left and right ends of the box plots represent the natural log of the concentrations for the overall minimum and maximum values, respectively (considering values from all methods together). The dots below each box plot are a one-way scatter plot with each point representing one original or predicted residue value. The density of points (or degree of dithering) in the one-way scatter plot represents the probability density of the values. When constructed in this manner, the box plot provides a rapid visual impression of the prominent features of the data. The median (or central line within the box) shows the location of the center of the data. The spread of the central 50% of the data is represented by the length of the box. The length of the whiskers (relative to the box) show how stretched the tails of the distribution are and provide and indication of the skewness of the data. Individual points which extend beyond the whiskers are outside values which may be further investigated and provide clues as to the distributional form. If the distribution is symmetric (e.g., as with a normal distribution), the box will be divided into two equal halves by the median, the upper- and lower- end whiskers will be the same length, and the number of extreme data points will be distributed equally on either end of the plot.

Frequency histograms are a second method by which two or more distributions can be compared and is a convenient and readily-understood way of communicating distributional information (see Figure 1 for an example). A frequency histogram is a graphical estimate of the empirical probability density function. Frequency histograms can be plotted on both linear and logarithmic scales. In the histograms used in this document, the distribution is plotted on a logarithmic scale, so a “normal” curve here represents lognormal data.

Empirical Q-Q (quantile-quantile) plots can be particularly effective in making detailed comparisons of distributions of two sets of data and are constructed by plotting in a two-dimensional scatter plot the quantiles of the first distribution against the quantiles of the second. If two distributions are identical, then all the plotted points will lie on a 45 degree straight line (y=x). The location and magnitude of departures from this “y=x”line emphasize differences between the two distributions. Q-Q plots tend to emphasize differences in the tails of a distribution (as opposed to P-P plots in which differences in the central portion of the distribution are emphasized). In the empirical Q-Q plots shown in this document, the residues are log-transformed.

Kolmogorov-Smirnov two-sample test statistics are not visual plots *per se*, but rather a calculated statistic which evaluates the null hypothesis of two identical populations. Specifically, the Kolmogorov-Smirnov Test is a non-parametric test based on the maximum absolute difference between the theoretical and sample cumulative distribution functions (CDFs). Large values of this statistic indicate a poor fit, while small values indicate a good fit. The Kolmogorov-Smirnov test is most sensitive around the median and less sensitive in the tails and is best at detecting shifts in the empirical CDF relative to the known CDF. Because it is specific to continuous data, it is often considered more appropriate than the chi-square test which is intended for use with categorical data. Although generally considered to be less proficient at detecting differences in spread among distributions, the Kolmogorov-Smirnov two sample test is considered to be more powerful than the chi-square test.

In addition, the comparison procedure initiated by OPP evaluated the "robustness" of each of the three proposed decomposition methodologies to various perturbations or characteristics of the data. The residue data typically available to OPP are rarely "perfect" from a statistical standpoint. That is, samples are frequently small (e.g., <30), are often considerably skewed (with a long right tail), are frequently moderately to heavily censored at the limit of detection or quantitation, and are likely to be comprised of a "mixture" of distributions with varying (and unknown) proportions. Thus, OPP believes that any proposed decompositing methodology must be sufficiently "rugged" to adequately deal with these situations. Therefore, OPP tested each proposed methodology by varying the following parameters in an attempt to better simulate real-world data which each method would be expected to adequately deal with:

N, the number of composites which are collected. As N decreases from 30 composite residue values to 10, the ability of each proposed methodology to accurately predict the original parent distribution is expected to diminish. The ability of each method to predict the parent distribution based on decreasing numbers of composite samples was evaluated. OPP considers it important that any method used be able to adequately decomposite samples with as few as ten or so samples with detections.

CV, the coefficient of variation of the parent distribution. As the skewness (or "tailedness") of the original parent distribution increases from a CV of 0.5 (low skew), to a CV of 1.0 (moderate skew), to a CV of 2.0 (high skew), the ability of each method to accurately predict the original parent distribution was assessed[8]. It was expected that as the skewness increases from low to high skew, the ability of each method to predict the original parent distribution would decrease. It is

[8] The results of the analyses conducted with parent populations with low skew (CV=0.5) are not shown in this paper. Overall, all three proposed decomposition methods performed reasonably well under these circumstances.

important that any method selected be able to effectively decomposite samples from moderately- to highly-skewed parent populations.

Level of censoring. As the composite values become increasingly censored [increasing from no censoring (0%) to moderate censoring (33%)], the ability of each methodology to adequately predict the parent distribution was assessed. The decomposition methodology selected should appropriately consider censored data and be able to adequately predict the parent distribution when moderate to heavy censoring is present.

Number of separate parent distributions. As the number of separate parent distributions is increased from one to two, the ability of each method to adequately predict the original distribution was assessed. As composite samples will frequently be derived from multiple distributions, it is important that any decomposition methodology be able to adequately handle composites derived from multiple distributions.

The results of these analyses are presented in Section IV.

B. Comparison Using Empirical Distributions

In an effort to investigate the decomposition methodologies' behavior with actual measured residue data, a second part of the comparison procedure was performed by decompositing a variety of actual data sets on which both composite and single-item residue analyses have been performed. These tests spanned a variety of data sets to include specially conducted field trials in which all of numerous samples were obtained from a single field and where all analyzed samples were subjected to the same (identical) pesticide treatment at the same time. In addition, the decompositing procedure was applied to data sets from USDA Pesticide Data Program (PDP) special studies specifically conducted to evaluate the variability in single-item residue concentrations in composite samples and assess its significance. The three sample data sets which were evaluated or reviewed are presented below:

(1) Industry-conducted single serving and composite analyses of diazinon on peaches

(2) USDA PDP single serving and composite analyses of aldicarb on potatoes

(3) USDA PDP single-serving and composite analyses of a widely-used

agricultural pesticide on a common single-serving fruit item[9].

A brief description of each data set is presented below, along with a number of important caveats that apply when using the above "real-world" composite analyses. The results of these comparisons are presented in Section IV of this document.

1. Novartis Single-Serving Peach Study

The first set of data which were decomposited by each decomposition method were residue data from a trial conducted on peaches in Georgia in 1998. The pesticide was sprayed on a peach orchard, with peaches harvested shortly after spraying[10]. A total of twenty composites were collected with each composite comprising 10 peaches. Individual peaches were homogenized and the homogenate split into two portions with one portion going to the composite sample for ten peaches and the other bing used for the single-serving pesticide residue analysis for that peach. This resulted in a total of 20 composite analyses and 200 associated individual-item analyses (10 per composite).

OPP has used the 20 composite samples as input values for the three decomposition routines to compare the predicted distribution with the measured residues in the 200 individually-analyzed samples. Results of these analyses are presented in Section IV.

2. PDP Single Serving Potato Study

In 1997, PDP conducted a special survey of aldicarb concentrations in single-item potatoes in an effort to investigate the relationship between measured composite residues and residues present in single potatoes that comprise that composite. This survey was initiated in response to a request from EPA and was designed to re-evaluate the tolerance for use of aldicarb

[9] This data has not yet been officially released by USDA and has not been subjected to their full QA/QC or reconciliation procedures. OPP's normal practice is to avoid the use of USDA PDP data which is not available for public review and has not been publically released. Nevertheless, USDA and OPP have agreed to make a one-time exception to their usual policy and permit the limited release of this data provided that the name of the crop and pesticide were not disclosed and the disclaimer made that QA/QC and reconciliation procedures have not been completed. It is anticipated that formal release of this data and crop/pesticide identity will occur in late February, 2000.

[10] For this trial, peaches were harvested *before* the label-specified pre-harvest interval in an effort to maximize residue values and thus do not reflect residue concentrations which would normally be encountered. All unsampled peaches from the test orchard were destroyed in accordance with the experimental nature of this trial.

on potatoes . The survey targeted potato samples which originated in four states with registered uses for aldicarb on potatoes (FL, ID, OR, and WA), and samples were collected for 13 months from December 1, 1996 through December 31, 1997 by eight participating states.

A total of 342 composite potato samples (consisting of 10 potatoes each) were analyzed for aldicarb and its metabolites. However, matched composite and single-serving analyses, were performed on only a subset of the 342 collected composites. Specifically, for each of the collected composite samples (consisting of 10 potatoes each), potatoes were washed and cut in half lengthwise, with one-half of each potato separately labeled and frozen for possible later single serving analysis if certain pre-established trigger criteria were met. The remaining ten half-portions were composited together for analysis. The analysis of these half portions was conducted in batches (i.e., analytical sets) of 12-15 samples per batch[11]. The limit of detection for these analyses was either 0.004 ppm or 0.005 ppm depending upon the metabolite determined (the sulfoxide or sulfone).

Subsequent single serving analyses were performed by the laboratory on only those stored, frozen single potatoes which corresponded to the highest composite measurements in each batch or analytical set. That is, of the 28 analytical sets analyzed, only 16 sets contained one or more samples with detectable levels of aldicarb or its metabolites and only the 10 single serving halves associated with the highest composite in each of those 16 sets were analyzed. Thus, for the purposes of the single-serving study, a total of 16 composite analyses and their 160 single-serving counterparts (10 per composite) were available. The study demonstrated the wide variation in occurrence of residues within the single-serving samples when compared to the reported composite value. Aldicarb sulfoxide individual serving results varied in magnitude by up to 7.4 times the reported composite value while aldicarb sulfone in individual single-serving results ranged up to 6.1 times the measured composite value.

This study was NOT purposefully conducted to elucidate a general model for the composite-single serving relationship since it selected only a very specific set of samples (namely those with among the highest residues) to analyze on a matched composite/single serving basis. For example, the 16 composite samples and their single serving counterparts selected for analysis represented the highest reported value in a set which itself comprised 12-15 composite samples and thus were not a random sample of potatoes. This study, instead, was designed to be simply a "first look" at how different concentrations in individual potatoes could be from the composite sample from which they came and if concentrations in single potatoes could exceed the tolerance. Thus, while the study supported the belief that concentrations in individual items in a composite could be substantially higher than the composite value, the study is not particularly well-suited to comparing the three decomposition methods under consideration since only composites with detected or quantified residues above a certain trigger value were subsequently analyzed on a

[11] That is, 12-15 samples were extracted and analyzed at the same time along with the required QA/QC spikes, blanks, standards, and controls for that analytical set according to PDP standard operating procedures (SOPs)

single-serving basis. Thus, the PDP potato samples which were analyzed were not really a random sample and this non-random sampling in the composite selection also led to a nonrandom grouping of single servings within a composite with a bias toward high values.

Given the limitations described above, OPP had elected not to attempt to use the aldicarb data collected by PDP in its comparison procedures.

3. PDP Single-Serving Special Study

Finally, PDP is currently conducting a single-serving study as an added component of its national sampling plan. Sampling for this single-serving study was begun in January 1999 at half the normal monthly sampling rate and is expected to continue through to December 1999. As stated in III.B.3., the data have not been subjected to full QA/QC or reconciliation procedures and have been released to the Agency with the caveat that the identity of the specific crop and pesticide not be released until such time that formal public release of the data by USDA occurs. Samples were collected according to established PDP sampling procedures. Composite samples were formed by combining the halves of each of ten fruits selected for analysis. An individual fruit from the sample was selected for single-serving analysis[12]. A total of 334 composite samples were analyzed with one fruit from each composite removed and analyzed separately as a single serving. This single-serving analysis was done regardless of the concentration detected in the composite (i.e., there was no "trigger concentration" (as there was in the PDP potato single serving study) which was required to be present in the composite before the single fruit was analyzed).

The analysis of this data, however, had to proceed somewhat differently than the analysis of the theoretical data due to complications arising from use of "real-world" residues with the significant intrinsic limitations detailed below:

1) true zeroes (from untreated commodities) could not be differentiated from "below detection level" determinations (here <0.006 ppm). That is, it was not known if a non-detect was due to a sample not being treated or due to it being treated but having concentrations less than the detection limit of 0.006 ppm

2) all detectable levels which were between the detection limit of 0.006 ppm and the quantitation limit (i.e., three times this level or ca. 0.02 ppm) were censored and instead reported as one-half the quantitation limit. That is, if a residue was measured as present between 0.006 ppm and 0.02 ppm, it was reported by the laboratory as 0.01 ppm.

3) Items (1) and (2) above resulting in a significant portion of the data being non-detect

[12]Beginning in May 1999, the PDP protocol was changed to require, at all times, analysis of one of the remaining halves of the ten fruits comprising the composite. Therefore, results for a single-serving sample are available for each corresponding composite for the entire year.

(i.e., below detectable levels due to either absence of treatment or residue levels below the limit of detection) or being reported as one-half the LOQ (i.e., 0.01 ppm). OPP recognizes that a) many of the non-detects are truly not treated and therefore should not be decomposited; and b) the 0.02 ppm LOQ is an artificial constraint and the "true" single-item distribution of residues extends well below the 0.02 ppm censoring level- - *that is, there is no true "spike" or agglomeration of residues at 0.01 ppm and this is only an artifact of the analytical method and reporting protocol.*

Given these limitations in the single-serving analyses and the requirement that these single-serving analyses serve as the reference population (or "gold-standard" or benchmark) for comparing the three decomposition methods being presented to the SAP, the following assumptions were made in an attempt to "correct" the data to allow the use of the full distribution as a standard reference population for comparison purposes:

(1) If no detectable residue (i.e., residue was reported by PDP as <0.006 ppm) was found in analyses of <u>either</u> the single-item sample <u>*or*</u> the associated composite sample from which the single-item was obtained, OPP made the assumption that the sample was not treated and this was therefore removed from the data set and not included as part of the distribution of residues in treated commodities. This was true of 31% of the paired (i.e., composite and associated single serving samples) samples, and thus a detectable residue was found in the composite sample and/or the associated single serving sample in approximately 69% of the cases. This compares favorably to estimates of the percent of the crop treated with the pesticide in question.[13] Thus it seems reasonable to assume for purposes of this comparison protocol that if the pesticide of interest was not found in either the composite or the associated single-serving analysis, then the sample was not treated and should not be decomposited or considered as part of the distribution of treated commodities. This assumption removes 31% of the paired samples from the data set.

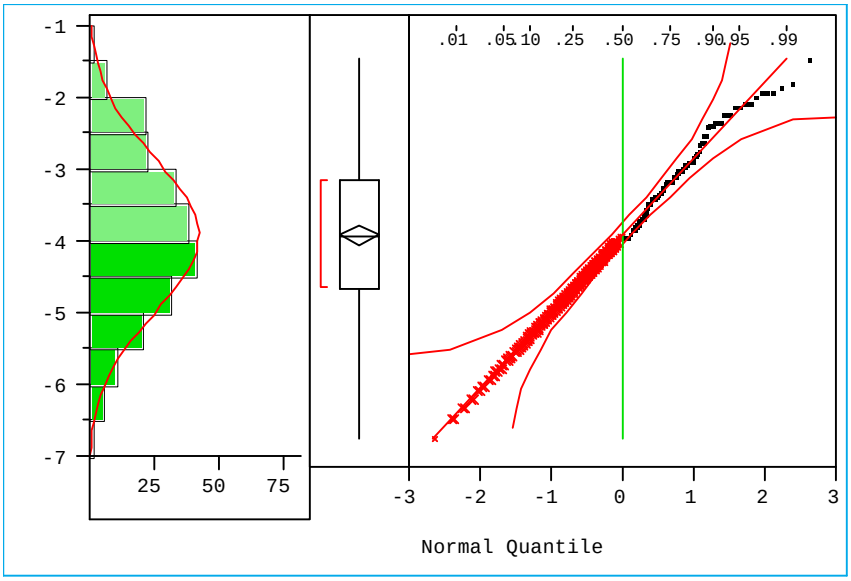
(2) It was assumed that the residues in the remaining single-serving samples (i.e., 69% of the dataset) were lognormally distributed and that the true values which were associated with the approximately half of the data which were censored at 0.02 ppm could be adequately imputed by means of Helsel's Robust Method (Helsel 1990, ILSI, 1998) using Maximum Likelihood Estimation techniques. This was done with the result shown in the normal probability plot on the following page (with residues plotted as their natural logarithms)[14]. As is apparent from the histogram and normal probability plot shown

---

[13] OPP's Biological and Economic Analysis Division estimated that approximately 64%, 76%, and 80% of the crop in question was treated with the pesticide of interest in 1996, 1997, and 1998, respectively.

[14] A better characterization of the observations between the LOD (0.006) and the LOQ (0.02) would be to treat them as censored values between 0.006 and 0.02 rather than just less than 0.02. This could be incorporated into the MaxLIP procedure and substantially impact its

below, the lower half of the distribution was imputed (using Helsel's method and MLE techniques) from the observed and measured residue values in the upper half of the distribution[15]. Given the resulting plot, OPP believes that this is a reasonable means of "filling-out" the remainder of the distribution. In any case, we note that, in our experience, the lower part of the distribution is generally less critical in the exposure and risk assessments performed by OPP and thus believe this to be a reasonable and appropriate attempt to generate an entire distribution of single-serving analyses as a standard to which the proposed decomposition methods can be compared.



With the nature of the reference population established as described above, the composite samples were decomposited using each of the three decomposition methodologies and compared to the distribution of imputed and actual residues obtained from the single-serving analyses (i.e., both the actual residue measurements and the imputed values based on Helsel's Robust method and MLE techniques)[16]. Results of this comparison are shown in Section IV.

IV. RESULTS OF COMPARISON

The results of OPP's comparison of the three decomposition procedures being presented

---

characterization of the distribution below 0.02 ppm.

[15] The imputed portion of the distribution is shown in dark in the histogram and as large X's in the normal probability plot.

[16] Actually, the comparison was not made directly with the entire distribution of decomposited values, but rather with one randomly selected individual item from *each* synthetically generated composite. This was done in order to permit a valid comparison with the PDP single-item analyses since only one item from each composite was routinely selected for a single-serving analysis.

to the SAP are presented in this section. As described earlier, the objective of this first part of the comparison scheme was to assess the adequacy of each of the three methodologies described above in "back-predicting" a given lognormal parent distribution of single-item residue. The accuracy of the predicted distribution was assessed by determining how well the model-generated single serving data "matched" or otherwise compared to the original data by comparing the predicted mean, standard deviation, interquartile range, absolute range or spread, and upper-percentile residue values generated (or implied) by each methodology to the corresponding statistics associated with the original, or "test," parent distribution. The comparison was extended to include examination of key graphical plots of the distributions and performing one or more statistical tests specifically intended to assess the equality of two distributions.

To summarize, results from two sets of comparisons included (a) theoretical analyses conducted with artificial parent/original distributions generated with specific means, standard deviations, skewness, etc. and (b) empirical analyses conducted with PDP or industry-generated data. These analyses are each described and detailed below under "Results from Theoretical Analysis" and "Results from Empirical Analysis," respectively.

A. Results from Theoretical Analysis

The results from each of the theoretical analyses discussed earlier in Section III.A. are discussed in detail in this section. Each section below begins with a description of how the standardized data were generated and used. The statistical tests (including graphical comparisons) are then described, followed by a table which compares various key statistics, and a second table which compares exposures (on a normalized basis) which would be predicted at various percentiles when the individual values are used as residues in the DEEM exposure assessment software. The intent of this latter information is to evaluate whether differences in predicted *pesticide residue* distributions result in significant differences in predicted *pesticide exposures* at the upper percentiles of regulatory interest to OPP.

1. Highly skewed distribution, 30 composites, Uncensored

OPP first investigated the ability of each of the three decomposition methods to "reproduce" the single-item components from a heavily skewed (CV=2) lognormal distribution which was uncensored. Specifically, 450 random "draws" were made from a lognormal distribution with a fixed mean of 0.10, a fixed standard deviation of 0.20 (both on an arithmetic scale), and a coefficient of variation of 2 (by definition). This was done with Crystal Ball software and the resulting distribution of individual values was termed the parent or "original" distribution. A total of 30 "sets" consisting of 15 consecutive values each were created from these 450 values and the arithmetic mean of each of the 30 sets was determined. This procedure, in effect, simulated the compositing process which occurs when 15 single-item units are combined and result in one measured residue concentration being reported which is equal to the weighted-

arithmetic average of the component single-item units[17].

For this analysis, the set of 30 composite values was not artificially censored (i.e., all values were retained). These values were then used as input to each of the three decomposition programs which created (from only the information contained in the 30 averages, or simulated composite values) a set of 450 simulated single item values which was expected to "match" or otherwise reasonably reproduce the parent or original distribution (which, as indicated previously, consisted of 450 random draws from a lognormal distribution with mean equal to 0.10 and standard deviation equal to 0.2)

Detailed graphical and statistical results of these analysis are shown in Attachment 1. Briefly, one method of assessing the accuracy of the predicted distribution is to evaluate how well the model-generated single serving data "match" or otherwise compare to the original data. Such a comparison was done by evaluating how well the predicted mean, standard deviation, interquartile range, absolute range or spread, and various upper-percentile residue values generated (or implied) by each methodology with the corresponding statistics which are characteristic of the original, or "test", parent distribution. In addition, statistical tests specifically designed to compare (on a non-parametric basis) the equality of two distribution were used as were a variety of graphical plots designed to highlight differences and similarities between the original parent data and the synthetic distributions generated by each of the three decomposition procedures.

A tabular summary of various specific statistics characteristic of the parent (and theoretical) and decomposition routine-generated values is shown in Table 1:

[17] This procedure is not precisely equivalent to the compositing procedure used by PDP in that in the procedure used by PDP, units making up the composite are correlated (generally coming from the same field and having been subjected to the same treatment practice), while the procedure used here randomly draws from the entire distribution and the samples are independent and uncorrelated. Nevertheless, there is not normally a correlation coefficient available for PDP composite samples and the assessment methodologies used here to compare the decomposition methods are believed to be adequate to compare the strengths and weaknesses and overall adequacy of the methods.

Table 1. Comparison of Parent and Theoretical Single-Item Distributions With Single-Unit Distributions of Values Generated from Three Decomposition Methods (Highly skewed distribution, 30 composites, Uncensored)

Statistic		Distribution				
		Theoretical[a]	**Parent[a] (original)**	**Allender**	**Novigen**	**MaxLIP-1[d]**
Mean ± SD		0.1000 ± 0.2000	0.10017 ± 0.1898	0.1003 ± 0.1566	0.0985 ± 0.1263	0.1003 ± 0.1543
QUANTILES	min	0.0000	0.0012	0.0021	0.0011	0.0016
	0.50	0.0447	0.0446	0.0529	0.0489	0.0516
	0.75	0.1046	0.1051	0.1133	0.1186	0.1131
	0.90	0.2269	0.2268	0.2246	0.2440	0.2288
	0.95	0.3582	0.3614	0.3400	0.4452	0.3532
	0.99	0.8595	0.8992	0.7650	0.5309	0.7560
	max	∞	2.5308	1.9224	0.7023	1.6195
IQR[b]		--	0.0861	0.0886	0.0996	0.0899
K-S statistic[c] (p value)		--	–	0.0778 (0.115)	0.0467 (0.682)	0.0578 (0.408)

[a] The parent distribution represents the distribution of the 450 generated values, while the theoretical distribution represents the mean, standard deviation (SD), and various selected percentiles of the lognormal distribution with mean = 0.1000 and SD=0.2000
[b] Interquartile Range. This represents the difference between the 75th and 25th percentiles.
[c] Kolmogorov-Smirnov Two-sample Test Statistic with associated p-value.
[d] MaxLIP with 1 lognormal distribution estimated.

A histogram as well as a one-way scatter plot (with associated box and whisker plot) which compares the distribution of the 450 values from each of the three generated distributions with the 450 values from the original test distribution is shown in Figure 1. Additional, more detailed information regarding the distribution of predicted values (included comparative empirical Q-Q plots, Kolmogorov-Smirnov statistics, additional summary descriptive measures, etc.) is presented in Appendix 1.

OPP also evaluated whether potential differences in the distribution of model-predicted single item values (or residues) might lead to significant differences in predicted *exposure* levels at various percentiles as calculated by DEEM (Dietary Exposure Evaluation Model) software. DEEM is the software used by OPP to estimate the distribution of (acute) exposures to the general U.S. population and various population subgroups of interest. DEEM does this by combining reported food consumption figures from USDA's Continuing Survey of Food Intake by Individuals (CSFII) with randomly selected pesticide residue concentrations generally available

from industry-conducted field trials and/or USDA or FDA monitoring data. While it is true in general that differences in the distribution of residues used as one of many inputs to the DEEM model would be expected to produce differences in exposure estimates at various percentiles, it is not necessarily true that these estimated exposure differences at the high end percentiles of particular regulatory interest to OPP are large or even significant from a regulatory perspective. That is, it may be that, despite the differences in "predicted" single-item *residue* concentrations among methods, predicted *exposures* at various regulatory thresholds of interest may not be significantly different.

To test this and evaluate potential differences in estimated pesticide exposures at the upper percentiles resulting from differences among the decomposition methods, OPP has used the parent distribution and model-predicted (decomposited) values as residue concentrations in the DEEM software and estimated exposures at the 99.9th , 99th, and 95th percentiles for two groups (general U.S. population and children 1-6). This analysis was done by each of the three decomposition methods being assessed, with each DEEM predicted "exposure" normalized (for ease of comparison) to the exposure predicted by the original parent distribution at the 99.9th percentile[18]. The analyses were done by assigning original and predicted values as residues to a widely consumed fruit present in the DEEM consumption database: results and/or conclusions might differ if a different commodity were selected or a set of commodities were analyzed simultaneously. For the data set presently being evaluated, normalized exposures as predicted by DEEM for various percentiles are shown in Table 2:

Table 2. Comparison of Normalized Exposures by Decomposition Method (Highly skewed distribution, 30 composites, Uncensored)

Method	**Normalized Exposure[a] (relative to DEEM-predicted exposures from original distribution at 99.9th percentile)**					
	General U.S. Population			**Children 1-6**		
	99.9th	**99th**	**95th**	**99.9th**	**99th**	**95th**
Original	1	0.27	0.08	1	0.27	0.08
Allender	0.89	0.23	0.06	0.88	0.26	0.08
Novigen	0.80	0.25	0.07	0.70	0.28	0.08
MaxLIP	0.89	0.24	0.06	0.88	0.27	0.08

[18]As a hypothetical example, if the DEEM-estimated exposure for the original parent distribution was 10 mg/kg bw/day and was 12, 14, and 16- mg/kg bw/day for the Allender, Novigen, and MaxLIP procedures (all at the 99.9th percentile) and 6, 7, and 8 mg/kgbw/day (all at the 99th percentile), normalized relative exposures for the Allender, Novigen, and MaxLIP procedures would be 1.2, 1.4, and 1.6, respectively for the 99.9th percentile and 0.6, 0.7, and 0.8 at the 99th percentile, respectively.

Composites	0.46	0.19	0.08	0.40	0.18	0.09

[a] Normalized exposures all relate to exposure at the 99.9th percentile (as estimated by DEEM) for the *original* data. These analyses were done by assigning original or decomposition-predicted values as residues to a widely consumed fruit present in the DEEM CSFII (USDA Continuing Survey of Food Intake by Individuals) database. Results would be expected to differ if another commodity was selected or several food commodities were simultaneously analyzed. The residues (and chemical) are hypothetical, and no toxicity value was assigned. Interpretation of the normalized exposures presented in this table should therefore be done with the understanding that the DEEM-estimated exposures in the table are *relative* to an arbitrarily-assigned baseline value, and no health-based implications should be inferred.

2. Moderately skewed distribution, 30 composites, Censored (33%)

OPP next investigated the ability of each of the three decomposition methods to "reproduce" the single-item components from a moderately skewed (CV=1) lognormal distribution which was censored at 33%. Specifically, 450 random "draws" were made from a lognormal distribution with a mean and standard deviation each equal to 0.10 (on an arithmetic scale), and a coefficient of variation of 1 (by definition). This was termed the parent or "original" distribution. As before, a total of 30 "sets" consisting of 15 consecutive values each were created from these 450 values and the arithmetic mean of each of the 30 sets calculated in an attempt to simulate a compositing process. However, the set of 30 composite values was then artificially censored by assigning a default "less than" concentration to all values up to the value representing the 33rd percentile (i.e., the lowest 10 values were to be considered as treated "non-detects" with default values assigned in accordance with the requirements of the protocol to the decomposition method.[19]) These values were then used as input to each of the three decomposition programs which were used to create (from only the information contained in the 30 averages, or simulated composite values) a set of 450 simulated single item values which was expected to "match" or otherwise reasonably reproduce the parent or original distribution consisting of 450 random draws from a lognormal distribution with mean and standard deviation equal to 0.10. This was done in an effort to compare each method's ability to use moderately-

[19] Specifically, a total of 10 of the 30 composite values generated were assigned to be "less than detect" values where the detection limit was determined by the next higher value which was retained. For the Allender method, the 10 "less than detects" were dropped because the Allender method decomposites only *detected* values, with the treated non-detects assigned to a different residue pool; for the Novigen RDFgen procedure, these values were assigned values of ½ the detection limit; for the MaxLIP procedure (which used MLE methods to impute values for these ND residues), the censored values are treated exactly as censored values "less than" the detection limit without having to be replaced with a specific selected value below the detection limit. For consistency in comparison, the histograms and one-way box/scatter plots illustrated in this document for the Allender method do not show these values (which represent treated non-detect residues) which would normally appear as a prominent peak or bulge of values at ½ LOD.

censored data and still obtain reasonable estimates of the distribution of original single-serving values.

Detailed graphical and statistical results of these analysis are shown in Appendix 1. As before, a comparison between the predicted mean, standard deviation, interquartile range, absolute range or spread, and various upper-percentile residue values generated (or implied) by each methodology with the corresponding statistics which are characteristic of the original, or "test", parent distribution and its theoretical counterpart was made to assess how well the model-generated single serving data matched or compared to the original parent data. In addition, statistical tests specifically designed to compare (on a non-parametric basis) the equality of two distributions was used as were a variety of graphical plots designed to highlight differences and similarities between the original parent data and the synthetic distributions generated by each of the three decomposition procedures.

A tabular summary of various specific statistics characteristic of the parent (and theoretical) and decomposition routine-generated values is shown below in Table 3:

Table 3. Comparison of Parent and Theoretical Single-Item Distributions With Single-Unit Distributions of Values Generated from Three Decomposition Methods (Moderately skewed distribution, 30 composites, Censored (33.3%))

Statistic		**Distribution**				
		Theoretical[a]	**Parent**[a] **(original)**	**Allender**	**Novigen**	**MaxLIP1**[d]
Mean ± SD		0.1000 ± 0.1000	0.1000 ± 0.0991	0.1124 ± 0.0793	0.0885 ± 0.1153	0.0982 ± 0.1066
QUANTILES	min	0.0000	0.0064	0.0146	0.00008	0.0031
	0.50	0.0707	0.0705	0.0916	0.0455	0.0640
	0.75	0.1235	0.1241	0.1411	0.1069	0.1228
	0.90	0.2053	0.2045	0.2067	0.2388	0.2168
	0.95	0.2770	0.2848	0.2663	0.3681	0.2937
	0.99	0.4920	0.5068	0.4135	0.5110	0.5383
	max	∞	1.0000	0.6952	0.5975	0.9512
IQR[b]		–	0.0836	0.0812	0.0910	0.0893
K-S statistic[c] (p-value)		--	--	0.1756 (0.000)	0.2422 (0.000)	0.0711 (0.182)

[a] The parent distribution represents the distribution of the 450 generated values, while the theoretical distribution represents the mean, standard deviation (SD), and various selected percentiles of the lognormal distribution with mean = 0.1000 and SD=0.1000.
[b] Interquartile Range. This represents the difference between the 75th and 25th percentiles.
[c] Kolmogorov-Smirnov Two-sample Test Statistic with associated p-value.
[d] MaxLIP with 1 lognormal distribution estimated.

A histogram and one-way scatter plot (with associated box and whisker plot) which compares these distributions is shown in Figure 2. Additional, more detailed information is presented in Appendix 1.

As before, OPP evaluated if potential differences in the model-predicted single serving values might lead to significant differences in predicted exposure levels at various percentiles as determined by DEEM (Dietary Exposure Evaluation Model) software. Also as before, this procedure was followed for each of the three decomposition methods being assessed, with each DEEM predicted "exposure" normalized to the exposure predicted by the original parent distribution at the 99.9th percentile. For the data set presently being evaluated, normalized exposures are shown in Table 4:

Table 4. Comparison of Normalized Exposures by Decomposition Method (Moderately skewed distribution, 30 composites, Censored (33%)

Method	**Normalized Exposure[a] (relative to DEEM-predicted exposures from original distribution at 99.9th percentile)**					
	General U.S. Population			**Children 1-6**		
	99.9th	**99th**	**95th**	**99.9th**	**99th**	**95th**
Original	1	0.33	0.11	1	0.37	0.14
Allender	0.94	0.35	0.12	0.9	0.37	0.16
Novigen	1.09	0.34	0.09	1.03	0.4	0.12
MaxLIP	1.03	0.33	0.10	1.02	0.38	0.14
Composites	0.61	0.26	0.10	0.56	0.26	0.13

[a] Normalized exposures all relate to exposure at the 99.9th percentile (as estimated by DEEM) for the *original* data. These analyses were done by assigning original or decomposition-predicted values as residues to a widely consumed fruit present in the DEEM CSFII (USDA Continuing Survey of Food Intake by Individuals) database. Results would be expected to differ if another commodity was selected or several food commodities were simultaneously analyzed. The residues (and chemical) are hypothetical, and no toxicity value was assigned. Interpretation of the normalized exposures presented in this table should therefore be done with the understanding that the DEEM-estimated exposures in the table are *relative* to an arbitrarily-assigned baseline value, and no health-based implications should be inferred.

3. Moderately skewed distribution, 10 composites, uncensored

OPP next investigated the ability of each of the three decomposition methods to "reproduce" the single-item components from a moderately skewed (CV=1) lognormal distribution with mean and standard deviation equal to 0.1 and which consisted of only 10 values. Specifically, only the initial 150 random draws from the original 450 random draws from the aforementioned lognormal distribution (i.e., mean = standard deviation = 0.1) were used. These 150 values were deemed to be the parent or "original" distribution. As before, a total of 10 consecutive "sets" consisting of 15 values each were created from these 150 values and the arithmetic mean of each of the resulting 10 sets calculated. These values were then used as input to each of the three decomposition programs which were used to create a generated set of 150 simulated single item values which was expected to "match" or otherwise reasonably reproduce the parent or original distribution consisting of the initial 150 random draws. This was done in an effort to compare each method's abilities to use minimal data and still obtain reasonable estimates of the distribution of original single serving values.

Detailed graphical and statistical results of these analysis are shown in Appendix 1. As before, a comparison between the predicted mean, standard deviation, interquartile range, absolute range or spread, and various upper-percentile residue values generated (or implied) by each methodology with the corresponding statistics which are characteristic of the original, or "test," parent distribution was made to assess how well the model-generated single serving data matched or compared to the original parent data. In addition, statistical tests specifically designed to compare (on a non-parametric basis) the equality of two distributions were used as were a variety of graphical plots designed to highlight differences and similarities between the original parent data and the synthetic distributions generated by each of the three decomposition procedures.

A tabular summary of various specific statistics characteristic of the parent and decomposition routine-generated values is shown in Table 5:

Table 5. Comparison of Parent and Theoretical Single-Item Distributions With Single-Unit Distributions of Values Generated from Three Decomposition Methods (Moderately skewed distribution, 10 composites, uncensored)

Statistic	**Distribution**				
	Theoretical[a]	**Parent[a] (original)**	**Allender**	**Novigen**	**MaxLIP1[d]**
Mean + SD	0.1000 ± 0.1000	0.0994 ±0.0938	0.0992 ±0.0992	0.0980 ±0.0936	0.1010 ±0.1133

QUANTILES	min	0.0000	0.0064	0.01	0.0024	0.0046
	0.50	0.0707	0.0705	0.07	0.0660	0.0639
	0.75	0.1235	0.1241	0.12	0.1281	0.1232
	0.90	0.2053	0.2045	0.21	0.2227	0.2254
	0.95	0.2770	0.2848	0.29	0.3116	0.3207
	0.99	0.4920	0.5153	0.55	0.4372	0.5856
	max	∞	0.6099	0.65	0.4563	0.7774
IQR[b]		–	0.0838	0.08	0.0948	0.0909
K-S statistic[c] (p-value)		--	--	0.0667 (0.866)	0.0867 (0.574)	0.0867 (0.574)

[a] The parent distribution represents the distribution of the 150 generated values, while the theoretical distribution represents the mean, standard deviation (SD), and various selected percentiles of the lognormal distribution with mean = 0.1000 and SD=0.1000

[b] Interquartile Range. This represents the difference between the 75th and 25th percentiles.

[c] Kolmogorov-Smirnov Two-sample Test Statistic with associated p-value.

[d] MaxLIP with 1 lognormal distribution estimated.

A histogram and one way scatter plot (with associated box and whisker plot) which compares these distributions is shown in Figure 3. Additional, more detailed information is presented in Appendix 1.

OPP has evaluated if potential differences in the model-predicted single serving values might lead to significant differences in predicted exposure levels at various percentiles as determined by DEEM (Dietary Exposure Evaluation Model) software. For the data set presently being evaluated, normalized exposures are shown below in Table 6:

Table 6. Comparison of Normalized Exposures by Decomposition Method (Moderately skewed distribution, 10 composites, uncensored)

Method	**Normalized Exposure[a] (relative to DEEM-predicted exposures from original distribution at 99.9th percentile)**					
	General U.S. Population			**Children 1-6**		
	99.9th	**99th**	**95th**	**99.9th**	**99th**	**95th**
Original	1	0.34	0.11	1	0.39	0.15
Allender	1.03	0.34	0.11	1.04	0.40	0.15
Novigen	0.99	0.34	0.11	0.96	0.41	0.15

MaxLIP	1.12	0.36	0.11	1.16	0.43	0.15
Composites	0.64	0.28	0.12	0.62	0.29	0.15

[a] Normalized exposures all relate to exposure at the 99.9th percentile (as estimated by DEEM) for the *original* data. These analyses were done by assigning original or decomposition-predicted values as residues to a widely consumed fruit present in the DEEM CSFII (USDA Continuing Survey of Food Intake by Individuals) database. Results would be expected to differ if another commodity was selected or several food commodities were simultaneously analyzed. The residues (and chemical) are hypothetical, and no toxicity value was assigned. Interpretation of the normalized exposures presented in this table should therefore be done with the understanding that the DEEM-estimated exposures in the table are *relative* to an arbitrarily-assigned baseline value, and no health-based implications should be inferred.

4. Moderately skewed distribution, 10 composites, censored (33%)

OPP next investigated the ability of each of the three decomposition methods to "reproduce" the single-item components from a small, moderately skewed (CV=1) lognormal distribution which was also censored at 33%. Specifically, only the initial 150 random draws from the original 450 random draws from the aforementioned lognormal distribution (i.e., mean = standard deviation = 0.1) were used. These 150 values were designated as the parent or "original" distribution. As before, a total of 10 consecutive "sets" consisting of 15 values each were created from these 150 values and the arithmetic mean of each of the resulting 10 sets calculated. However, as before, the set of 10 composite values was artificially censored by assigning a default "less than" concentration to all values up to the value representing the 33rd percentile [i.e., the lowest 3 values of the 10 were assigned treated "non-detects" status with default values assigned in accordance with the requirements of the decomposition method protocol (described in footnote 19)]. These values were then used as input to each of the three decomposition programs which were used to create a generated set of 150 simulated single item values which was expected to "match" or otherwise reasonably reproduce the parent or original distribution consisting of the initial 150 random draws. This was done to investigate how the different methods compared when there were was only minimal data, which was compounded by issues of moderate censoring.

Detailed graphical and statistical results of these analysis are shown in Attachment 1. As before, a comparison between the predicted statistics and the corresponding statistics which are characteristic of the original, or "test," parent distribution was made to assess how well the model-generated single serving data matched or otherwise compared to the original parent data. In addition, statistical tests specifically designed to compare (on a non-parametric basis) the equality of two distributions was used as were a variety of graphical plots designed to highlight differences and similarities between the original parent data and the synthetic distributions generated by each of the three decomposition procedures.

A tabular summary of various specific statistics characteristic of the parent and decomposition routine-generated values is shown below in Table 7:

Table 7.	Comparison of Parent and Theoretical Single-Item Distributions With Single-Unit Distributions of Values Generated from Three Decomposition Methods (Moderately skewed distribution, 10 composites, censored (33%))					
Statistic		**Distribution**				
		Theoretical	**Parent (original)**	**Allender**	**Novigen**	**MaxLIP1[d]**
Mean ± SD		0.1000 ± 0.1000	0.0994 ± 0.0938	0.1036 ± 0.0726	0.0900 ± 0.1151	0.1011 ± 0.1245
QUANTILES	min	0.000	0.0064	0.0199	0.0008	0.0032
	0.50	0.0707	0.0705	0.0837	0.0468	0.0578
	0.75	0.1235	0.1241	0.1249	0.1135	0.124
	0.90	0.2053	0.2045	0.1857	0.2399	0.2375
	0.95	0.2770	0.2848	0.2472	0.3713	0.334
	0.99	0.4920	0.5153	0.3351	0.4838	0.6323
	max	∞	0.6099	0.5693	0.5740	0.8681
IQR			0.0838	0.0861	0.0984	0.0966
K-S statistic (p-value)		--	--	0.1600 (0.032)	0.2333 (0.000)	0.1333 (0.111)

[a] The parent distribution represents the distribution of the 150 generated values, while the theoretical distribution represents the mean, standard deviation (SD), and various selected percentiles of the lognormal distribution with mean = 0.1000 and SD=0.1000
[b] Interquartile Range. This represents the difference between the 75th and 25th percentiles.
[c] Kolmogorov-Smirnov Two-sample Test Statistic with associated p-value.
[d] MaxLIP with 1 lognormal distribution estimated.

A histogram and one-way scatter plot (with associated box and whisker plot) which compares these distributions is shown in Figure 4. Additional, more detailed information is presented in Appendix 1.

OPP investigated whether potential differences in the model-predicted single serving values might lead to significant differences in predicted exposure levels at various percentiles as determined by DEEM (Dietary Exposure Evaluation Model) software. For the data set presently being evaluated, normalized exposures from DEEM are shown in Table 8:

Table 8.	Comparison of Normalized Exposures by Decomposition Method (Moderately skewed distribution, 10 composites, censored (33.3%))

Method	Normalized Exposure[a] (relative to DEEM-predicted exposures from original distribution at 99.9th percentile)					
	General U.S. Population			Children 1-6		
	99.9th	99th	95th	99.9th	99th	95th
Original	1	0.33	0.10	1	0.37	0.14
Allender	0.94	0.35	0.12	0.90	0.37	0.16
Novigen	1.09	0.33	0.09	1.03	0.4	0.12
MaxLIP	1.04	0.33	0.10	1.02	0.38	0.14
Composites	0.64	0.29	0.12	0.61	0.28	0.15

[a] Normalized exposures all relate to exposure at the 99.9th percentile (as estimated by DEEM) for the *original* data. These analyses were done by assigning original or decomposition-predicted values as residues to a widely consumed fruit present in the DEEM CSFII (USDA Continuing Survey of Food Intake by Individuals) database. Results would be expected to differ if another commodity was selected or several food commodities were simultaneously analyzed. The residues (and chemical) are hypothetical, and no toxicity value was assigned. Interpretation of the normalized exposures presented in this table should therefore be done with the understanding that the DEEM-estimated exposures in the table are *relative* to an arbitrarily-assigned baseline value, and no health-based implications should be inferred.

5. Two Moderately skewed distributions, 25% overlap, 15 Composites per distribution, Censored (33%)

OPP next investigated the ability of each of the three decomposition methods to effectively distinguish and "reproduce" the single-item components from two (25% overlapping) moderately skewed (CV=1) lognormal distributions, each of which were censored at the common 33rd percentile. Specifically, only the initial 225 random draws from the original 450 random draws from the aforementioned lognormal distribution (mean and standard deviation each equal to 0.10) were used. To this were added 225 *additional* random draws from a second lognormal distribution with a mean and standard deviation of 0.31 (structured to have a 25% overlap with the first distribution[20]). These 450 values were designated to be the parent or "original" distribution. A total of 15 "sets" consisting of 15 consecutive values each were created from each of these 225 values and two arithmetic means of the resulting 15 sets were calculated. Censoring was performed by designating any value in the bottom third of the ranked *combined* data set as "less than detect." All values (both those above and below the censoring limit) were used with MaxLIP and RDFgen (i.e., BDL values were assigned as either one-half the censoring limit in the

[20] The second distribution was defined such that the 25th percentile of this distribution was equal to the 75th percentile of the first distribution.

case of RDFgen or as "less than" the censoring limit in MaxLIP). Since the Allender method would assign these to a separate, single-valued distribution at ½ LOD, these were not used (and simply dropped) in this method which consequently relied only on the uncensored values. The composite values from each of the two distributions were then used as input to each of the three decomposition programs. This was done to investigate how the different methods compared when the parent distribution actually consisted of two overlapping distributions, and the combined data set was moderately censored.

Detailed graphical and statistical results of these analysis are shown in Appendix 1. As before, a comparison between the predicted statistics and corresponding statistics which are characteristic of the original parent distribution was made to assess how well the model-generated single serving data matched or otherwise compared to the original parent data. In addition, statistical tests specifically designed to compare (on a non-parametric basis) the equality of two distributions were used as were a variety of graphical plots designed to highlight differences and similarities between the original parent data and the synthetic distributions generated by each of the three decomposition procedures.

A tabular summary of various specific statistics characteristic of the parent and decomposition routine-generated values is shown in Table 9:

Table 9. Comparison of Parent Single-Item Distributions With Single-Unit Distributions of Values Generated from Three Decomposition Methods (Two Moderately skewed distributions, 25% overlap, 15 Composites per distribution, Censored (33%))

Statistic		**Distribution**			
		Parent (original)	**Allender**	**Novigen**	**MaxLIP-2[d]**
Mean ± SD		0.2046 ±0.2646	0.2748 ± 0.5237	0.1939 ±0.3284	0.2084 ±0.2106
QUANTILES	min	0.0064	0.0041	0.00003	0.0133
	0.50	0.1211	0.1536	0.0609	0.14
	0.75	0.2476	0.3106	0.2058	0.26
	0.90	0.4501	0.5809	0.5631	0.45
	0.95	0.6375	0.8562	0.8685	0.603
	0.99	1.2726	1.890	1.629	1.05
	max	3.0600	9.082	1.875	1.84
IQR		0.1862	0.2371	0.1946	0.1715
K-S statistic (p-value)		--	0.100 (0.018)	0.3133 (0.000)	0.0978 (0.022)

[a] The parent distribution represents the generation of two lognormal distributions of 225 values each, with means of 0.100 and 0.306 and standard deviations (SD) of 0.100 and 0.306, respectively.
[b] Interquartile Range. This represents the difference between the 75th and 25th percentiles.
[c] Kolmogorov-Smirnov Two-sample Test Statistic with associated p-value.
[d] Likelihood Ratio Test and other ancillary data indicated that MaxLIP (2 distributions) provided the most appropriate fit to the data, with proportions of 0.52 and 0.48 for the two component distributions. $LogLikelihood_{1distr}$= -1.73; $LogLikelihood_{2distr}$ = 7.31 ($p<0.005$); $LogLikelihood_{3distr}$=7.20 (p not calculated); $LogLikelihood_{4distr}$ = 9.75 ($p>0.05$). See Footnote 3 for further explanation of these results.

A histogram and one-way scatter plot (with associated box and whisker plot) which compares these distributions is shown in Figure 5. Additional, more detailed information is presented in Appendix 1.

OPP evaluated if potential differences in the model-predicted single serving values might lead to significant differences in predicted exposure levels at various percentiles as determined by DEEM (Dietary Exposure Evaluation Model) software. For the data set presently being evaluated, normalized exposures are shown below in Table 10:

Table 10. Comparison of Normalized Exposures by Decomposition Method (Two Moderately skewed distribution, 25% overlap, 15 Composites per distribution Censored (33%))

Method	**Normalized Exposure[a] (relative to DEEM-predicted exposures from original distribution at 99.9th percentile)**					
	General U.S. Population			**Children 1-6**		
	99.9th	**99th**	**95th**	**99.9th**	**99th**	**95th**
Original	1	0.30	0.09	1	0.33	0.11
Allender	1.53	0.39	0.11	1.51	0.45	0.14
Novigen	1.25	0.35	0.08	1.18	0.41	0.1
MaxLIP	0.89	0.29	0.09	0.84	0.32	0.12
Composites	0.64	0.26	0.09	0.58	0.26	0.12

[a] Normalized exposures all relate to exposure at the 99.9th percentile (as estimated by DEEM) for the *original* data. These analyses were done by assigning original or decomposition-predicted values as residues to a widely consumed fruit present in the DEEM CSFII (USDA Continuing Survey of Food Intake by Individuals) database. Results would be expected to differ if another commodity was selected or several food commodities were simultaneously analyzed. The residues (and chemical) are hypothetical, and no toxicity value was assigned. Interpretation of the normalized exposures presented in this table should therefore be done with the understanding that the DEEM-estimated exposures in the table are *relative* to an arbitrarily-assigned baseline value, and no health-based implications should be inferred.

6. Two Moderately skewed distributions, 10% overlap, 15 composites per distribution, Censored (33%)

Finally, OPP investigated the ability of each of the three decomposition methods to effectively distinguish and "reproduce" the single-item components from two (10% overlapping) moderately skewed (CV=1) lognormal distributions, each of which were censored at the common 33rd percentile. As before, only the initial 225 draws from the original 450 random draws from the aforementioned (mean = SD = 0.1) lognormal distribution were used. To this were added 225 additional random draws from a second lognormal distribution with a mean and standard deviation of 0.842.[21] These 450 values were deemed to be the parent or "original" distribution. A total of 15 "sets" consisting of 15 consecutive values each were created from each of these 225 values and two arithmetic means of the resulting 15 sets were calculated. Again, censoring was performed by designating any value in the bottom third of the ranked *combined* data set as "less than detect" and treating these values as per each decomposition methods protocol (see IV.A.5)

Detailed graphical and statistical results of these analysis are shown in Attachment 1. As before, a comparison between the predicted mean, standard deviation, interquartile range, absolute range or spread, and various upper-percentile residue values generated (or implied) by each methodology with the corresponding statistics which are characteristic of the original, or "test," parent distribution was made to compare how well the model-generated single serving data matched or otherwise compared to the original parent data. In addition, statistical tests specifically designed to compare (on a non-parametric basis) the equality of two distributions was used as were a variety of graphical plots designed to highlight differences and similarities between the original parent data and the synthetic distributions generated by each of the three decomposition procedures.

A tabular summary of various specific statistics characteristic of the parent and decomposition routine-generated values is shown below in Table 11:

Table 11. Comparison of Parent and Theoretical Single-Item Distributions With Single-Unit Distributions of Values Generated from Three Decomposition Methods (Two Moderately skewed distribution, 10% overlap, 15 composites per distribution, Censored (33%))

Statistic	**Distribution**			
	Parent (original)[a]	**Allender**	**Novigen**	**MaxLIP-2[d]**
Mean ± SD	0.4797 ±0.7463	0.6501 ±1.1214	0.0464 ±0.9708	0.4859 ±0.7390

[21] This produced a second distribution which overlapped the first by 10%

QUANTILES	min	0.0064	0.0075	4.69e-6	0.0103
	0.50	0.2064	0.2820	0.0755	0.199
	0.75	0.6096	0.6936	0.4804	0.581
	0.90	1.1940	1.5226	1.2613	1.23
	0.95	1.6318	2.5007	2.0995	1.84
	0.99	3.5892	5.9084	5.1883	3.71
	max	8.4239	9.7446	6.2807	6.86
IQR[b]		0.5387	0.4769	0.5768	0.4967
K-S statistic[c] (p-value)		--	0.1311 (0.001)	0.3644 (0.000)	0.0511 (0.567)

[a] The parent distribution represents the generation of two lognormal distributions of 225 values each, with means of 0.1000 and 0.8424 and standard deviations (SD) of 0.1000 and 0.8424, respectively
[b] Interquartile Range. This represents the difference between the 75th and 25th percentiles.
[c] Kolmogorov-Smirnov Two-sample Test Statistic with associated p-value.
[d] Likelihood Ratio Test and ancillary information indicated that MaxLIP (2 distributions) provided the most appropriate fit to the data, with predicted proportions of 0.51 and 0.49 for the two component distributions. $LogLikelihood_{1distr}$= -33.14; $LogLikelihood_{2distr}$ = -13.26 ($p<0.005$); $LogLikelihood_{3distr}$= -9.68 ($p<0.05$); $LogLikelihood_{4distr}$ = -11.17 ($p>0.05$) See Footnote 3 for further explanation of these results.

A histogram and one-way scatter plot (with associated box and whisker plot) which compares these distributions is shown in Figure 6. Additional, more detailed information is presented in Appendix 1.

OPP evaluated if potential differences in the model-predicted single serving values might lead to significant differences in predicted exposure levels at various percentiles as determined by DEEM (Dietary Exposure Evaluation Model) software. For the data set presently being evaluated, normalized exposures are shown below in Table 12:

Table 12. Comparison of Normalized Exposures by Decomposition Method (Two Moderately skewed distribution, 10% overlap, 15 composites per distribution, Censored (33.3%))

Method	**Normalized Exposure[a] (relative to DEEM-predicted exposures from original distribution at 99.9th percentile)**					
	General U.S. Population			**Children 1-6**		
	99.9th	**99th**	**95th**	**99.9th**	**99th**	**95th**
Original	1	0.28	0.07	1	0.32	0.10

Allender	1.50	0.39	0.09	1.57	0.44	0.13
Novigen	1.32	0.32	0.06	1.32	0.38	0.09
MaxLIP	1.02	0.29	0.07	1.02	0.34	0.10
Composites	0.67	0.26	0.09	0.61	0.27	0.12

[a] Normalized exposures all relate to exposure at the 99.9th percentile (as estimated by DEEM) for the *original* data. These analyses were done by assigning original or decomposition-predicted values as residues to a widely consumed fruit present in the DEEM CSFII (USDA Continuing Survey of Food Intake by Individuals) database. Results would be expected to differ if another commodity was selected or several food commodities were simultaneously analyzed. The residues (and chemical) are hypothetical, and no toxicity value was assigned. Interpretation of the normalized exposures presented in this table should therefore be done with the understanding that the DEEM-estimated exposures in the table are *relative* to an arbitrarily-assigned baseline value, and no health-based implications should be inferred.

B. Results from Empirical Analysis

The results of the empirical analysis discussed in Section III.B. are discussed in detail in this section. Each section below begins with a brief description of the specific study under consideration. Statistical comparisons are then described, followed, as before, by a table which compares various key statistics, and a second table which compares exposures (on a normalized basis) which would be predicted at various percentiles were the individual values to be placed in to the DEEM exposure assessment software.

1. Novartis Georgia Peach Study

The first set of data which were decomposited by each decomposition method were residue data from an experimental field trial conducted by Novartis Crop Protection in Georgia in 1998. The pesticide was sprayed on a peach orchard, with peaches harvested shortly after spraying. A total of twenty composites were collected with each composite comprising 10 peaches. Both the composite sample and each individual peach within the composite were analyzed for the pesticide resulting in a total of 20 composite analyses and 200 individual single serving analysis (10 per composite). OPP has used the 20 composite samples as input values for the three decomposition routines.

Detailed graphical and statistical results of these analysis are shown in Appendix 1. As before, a comparison between the predicted mean, standard deviation, interquartile range, absolute range or spread, and various upper-percentile residue values generated by each methodology with the corresponding statistics which are characteristic of the actual single serving data was made to compare the model-generated single serving data to the original parent data. In addition, statistical tests specifically designed to compare (on a non-parametric basis) the equality of two distributions was used as were a variety of graphical plots designed to highlight differences and similarities between the original parent data and the synthetic distributions

generated by each of the three decomposition procedures.

A tabular summary of various specific statistics characteristic of the parent and decomposition routine-generated values is shown below in Table 13:

Table 13. Comparison of Novartis Peach Field Trial Study Single-Item Residue Distribution With Single-Unit Distributions of Values Generated from Three Decomposition Methods

Statistic		**Distribution**			
		Field trial data[a]	**Allender**	**Novigen**	**MaxLIP-1[d]**
Mean ± SD		0.1446 ± 0.2047	0.1470 ± 0.1921	0.1432 ± 0.1754	0.1439 ± 0.2174
QUANTILES	min	0.001	0.000	0.0004	0.0033
	0.50	0.067	0.08	0.0757	0.0746
	0.75	0.1855	0.17	0.1751	0.1607
	0.90	0.4175	0.36	0.4211	0.3179
	0.95	0.51	0.525	0.5299	0.5110
	0.99	0.973	1.025	0.7604	1.1805
	max	1.499	1.43	0.7718	1.8706
IQR[b]		0.1660	0.13	0.1509	0.1262
K-S statistic[c] (p-value)		--	0.2048 (0.000)	0.0600 (0.837)	0.1600 (0.009)

[a] This distribution represents the standard distribution for comparison purposes
[b] Interquartile Range. This represents the difference between the 75th and 25th percentiles.
[c] Kolmogorov-Smirnov Two-sample Test Statistic with associated p-value.
[d] Likelihood Ratio Test indicated that MaxLIP (1 distribution) provided the most appropriate fit to the data. $LogLikelihood_{1distr}$=28.65; $LogLikelihood_{2distr}$ = 30.43 (p>0.05); $LogLikelihood_{3distr}$= 31.99; $LogLikelihood_{4distr}$ = 34.71.

A histogram and one-way scatter plot (with associated box and whisker plot) which compares these distributions is shown in Figure 7. Additional, more detailed information is presented in Appendix 1.

OPP evaluated if potential differences in the model-predicted single serving values might lead to significant differences in predicted exposure levels at various percentiles as determined by DEEM (Dietary Exposure Evaluation Model) software. For the data set presently being evaluated, normalized exposures are shown below in Table 14:

Table 14. Comparison of Novartis Peach Field Trial DEEM-Predicted Exposures by Decomposition Method

Method	**Normalized Exposure[a] (relative to DEEM-predicted exposures from original distribution at 99.9th percentile)**					
	General U.S. Population			**Children 1-6**		
	99.9th	**99th**	**95th**	**99.9th**	**99th**	**95th**
Original (field trial)	1	0.18	0.01	1	0.27	0.02
Allender	0.95	0.18	0.01	0.97	0.27	0.03
Novigen	0.96	0.19	0.01	0.9	0.28	0.03
MaxLIP1	0.95	0.17	0.03	1.0	0.24	0.03
Composites	0.67	0.19	0.03	0.53	0.23	0.06

[a] Normalized exposures all relate to exposure at the 99.9th percentile (as estimated by DEEM) for the *original* data. These analyses were done by assigning original or decomposition-predicted values as residues to a widely consumed fruit present in the DEEM CSFII (USDA Continuing Survey of Food Intake by Individuals) database. Results would be expected to differ if another commodity was selected or several food commodities were simultaneously analyzed. The residues (and chemical) are hypothetical, and no toxicity value was assigned. Interpretation of the normalized exposures presented in this table should therefore be done with the understanding that the DEEM-estimated exposures in the table are *relative* to an arbitrarily-assigned baseline value, and no health-based implications should be inferred.

2. PDP Potato Study

As indicated earlier, the PDP aldicarb in potatoes study was not conducted in a manner which would make the data amenable to straightforward use in a comparison of decomposition methods. Therefore, OPP has elected not to conduct this analysis.

3 PDP Single Serving Special Study

The final set of data which were decomposited by each procedure being presented to the Panel as a PDP single serving special study conducted on a trial basis as part of its national sampling plan. As described earlier, composite samples were collected, with one individual item from each composite selected for analysis for the pesticide of interest. A total of 334 analytical results from composite samples were available with one item from each composite removed and analyzed separately as a single serving. This single-serving analysis was done regardless of the concentration detected in the composite (i.e., there was no "trigger concentration" (as there was in the PDP potato single serving study) which was required to be present in the composite before the single apple was analyzed).

Detailed graphical and statistical results of these analysis are shown in Appendix 1. As before, a comparison between the predicted mean, standard deviation, interquartile range, absolute range or spread, and various upper-percentile residue values generated by each methodology with the corresponding statistics which are characteristic of the imputed single serving distribution was made to compare the model-generated single serving data to the original data (including imputed values for <LOQ values). In addition, statistical tests specifically designed to compare (on a non-parametric basis) the equality of two distributions was used as were a variety of graphical plots designed to highlight differences and similarities between the original parent data and the synthetic distributions generated by each of the three decomposition procedures.

A tabular summary of various specific statistics characteristic of the parent and decomposition routine-generated values is shown below in Table 15:

Table 15. Comparison of Imputed PDP Single-Item Residue Distribution With Single-Unit Distributions of Values Generated from Three Decomposition Methods

Statistic		**Distribution**			
		Imputed PDP distribution[a]	**Allender**	**Novigen**	**MaxLIP4**
Mean ± SD		0.0346 ± 0.0385	0.0351 ± 0.0119	0.0349 ± 0.0673	0.0306 ± 0.0726
QUANTILES	min	0.0012	0.0142	<0.0001	<0.0002
	0.50	0.0198	0.0325	0.0053	0.0082
	0.75	0.044	0.0422	0.0339	0.0298
	0.90	0.099	0.0504	0.0963	0.0735
	0.95	0.12	0.0579	0.1882	0.131
	0.99	0.16	0.0686	0.3259	0.331
	max	0.24	0.0840	0.4361	0.74
IQR[b]		0.0343	0.0156	0.0332	0.0298
K-S statistic[c] (p-value)		--	0.4808 (0.000)	0.3977 (0.000)	0.4242 (0.000)

[a] This distribution represents the standard distribution for comparison purposes
[b] Interquartile Range. This represents the difference between the 75th and 25th percentiles.
[c] Kolmogorov-Smirnov Two-sample Test Statistic with associated p-value

A histogram and one-way scatter plot (with associated box and whisker plot) which compares these distributions is shown in Figure 8. Additional, more detailed information is presented in Appendix 1.

OPP evaluated if potential differences in the model-predicted single serving values might lead to significant differences in predicted exposure levels at various percentiles as determined by DEEM (Dietary Exposure Evaluation Model) software. For the data set presently being evaluated, normalized exposures are shown below in Table 16:

Table 16. Comparison of PDP Single-Item Special Study DEEM-Predicted Exposures by Decomposition Method

Method	**Normalized Exposure[a] (relative to DEEM-predicted exposures from original distribution at 99.9th percentile)**					
	General U.S. Population			**Children 1-6**		
	99.9th	**99th**	**95th**	**99.9th**	**99th**	**95th**
PDP single-item	1	0.33	0.09	1	0.40	0.13
Allender	0.61	0.26	0.10	0.60	0.28	0.14
Novigen	1.55	0.42	0.08	1.63	0.54	0.12
MaxLIP4	1.51	0.33	0.07	1.76	0.44	0.10
Composites	0.96	0.32	0.10	0.96	0.39	0.14

[a] Normalized exposures all relate to exposure at the 99.9th percentile (as estimated by DEEM) for the *original* data. These analyses were done by assigning original or decomposition-predicted values as residues to a widely consumed fruit present in the DEEM CSFII (USDA Continuing Survey of Food Intake by Individuals) database. Results would be expected to differ if another commodity was selected or several food commodities were simultaneously analyzed. The residues (and chemical) are hypothetical, and no toxicity value was assigned. Interpretation of the normalized exposures presented in this table should therefore be done with the understanding that the DEEM-estimated exposures in the table are *relative* to an arbitrarily-assigned baseline value, and no health-based implications should be inferred.

V. SUMMARY OBSERVATIONS AND CONCLUSIONS

EPA's Office of Pesticide Programs has compared its current decomposition procedure with the those presented to the SAP by JSC Sielken/Novartis (MaxLIP) and Novigen Sciences (RDFgen). This comparison was done by using both (a) theoretical data designed to reflect differences in such characteristics as skewness, censoring, number of samples, and number of distributions and (b) empirical (real world) pesticide data collected by USDA's Pesticide Data Program (PDP) and others. Based on this analysis, OPP makes the following observations with respect to the three decomposition methods and their performance characteristics when applied to both theoretical and empirical (actual) data:

(1) Based on the analysis using both hypothetical and empirical data sets, estimates of the high percentiles of daily exposure calculated using residues measured in composite samples are much lower than estimated exposures using "decomposited" residue values. Composite residue values tend to underestimate daily exposure by 30% - 50% at the upper percentiles.

(2) All methods appeared comparable and seemed to do reasonably well at predicting single-item *residues* at up to approximately the 90th percentile, regardless of the data set which was used. This was true of both the theoretical and empirical datasets. As the number of distributions increased, moderate censoring was imposed, or number of data points decreased, the ability of the methods to predict the upper percentile residue values appeared to deteriorate to varying degrees.

(3) The presence of multiple distributions and censoring appear to have the most effect on each methods ability to adequately deconvolute residue values while skewness of the distribution and number of composite residue values seemed to have the least.

(4) In many cases, the RDFgen and Allender procedures appeared to predict too large a "spread" in the data, particularly in the lower percentiles. Nevertheless, this did not appear to affect the exposures (as predicted by DEEM) in the region of regulatory interest (e.g., >95th percentile)

(5) Despite the findings in (2) and (4) above, the most accurate decompositing method rarely overestimated or underestimated the exposure of the 99.9th percentile by more than 15%, compared to the calculation using the parent data set, when using hypothetical data. The differences between the estimates obtained using the best method and the parent data set were even smaller at lower percentiles.

(6) All methods seemed be able to predict the 99.9th percentile *exposure* (as determined by DEEM) reasonably well and no method appeared to have a significant bias toward over- or under-prediction. At the 99th percentile exposure and below, the methods appeared to be essentially equivalent, with each method predicting the same exposure as the original (parent data).

VI. REFERENCES

Helsel, DR 1990. Less than obvious: statistical treatment of data below the detection limit. Environ. Sci. Technol. 24(12): 1766-1774.

ILSI (International Life Sciences Institute). 1998. Aggregate Exposure Assessment. ILSI Risk Science Institute Workshop Report. Washington, D.C.

Office of Pesticide Programs (OPP). Health Effects Division. 1999. "A Background

Document for the Session: *Statistical Methods for Use of Composite Data in Acute Dietary Exposure Assessment* of the May 26, 1999 Meeting of the FIFRA Scientific Advisory Panel". May 11.

VII. QUESTIONS FOR PANEL

Question 1

The current PDP program collects residue data on approximately 5 lb. composite samples, whereas the residue values of interest in acute risk assessment are associated with residue concentrations in single items of produce. In order to make better and fuller use of the current PDP data, OPP is currently using its own decomposition method in an effort to convert residues from a "composite" basis to a "single-item" basis which was presented to the SAP in May, 1999.

> **Two additional methods for decompositing pesticide residue data have been presented to the SAP (RDFgen and MaxLIP). What are the overall strengths and weaknesses of each method with respect to their ability to adequately represent pesticide residues in single unit items ?**

Question 2

The OPP comparison attempted to gauge each decomposition method's performance against several standard sets of data which reflected differences in number of samples, degree of skewness, amount of censoring, and number of distributions.

> **Each method may be sensitive to various "imperfections" , limitations, or characteristics of real-world data. For example, often data from many fewer than 30 composite samples are available for decomposition. Frequently, the data are censored and/or are heavily left-skewed. Many times, the composite samples may have been collected from a multitude of separate and distinct pesticide residue distributions. How sensitive are the two methods being presented to the SAP for consideration to these different factors? Does each method being presented tot the SAP have an adequately robust statistical underpinning?**

Question 3

Despite an adequate statistical underpinning and overall robustness, there may be specific situations in which characteristics of available data may make it unreasonable to expect a method to adequately deconvolute a dataset comprised of composite samples and decomposition should be avoided as it may produce invalid or questionable output data.

> **What limitations does the Panel see in the decomposition methodologies being presented to the SAP (e.g., minimum number of samples, degree of censoring, etc.)? In what specific kinds of situations might each presented methodologies fail or be likely to fail?**

Question 4

In contrast to OPP's original decomposition method which was presented to the SAP in May 1999, the MaxLIP and RDFgen methods being presented to the current Panel do not assume that PDP residue measurements are derived from one overall lognormal distribution of residues. MaxLIP permits up to five distinct residue distributions, while RDFgen permits any number of residue distributions and assumes that each composite measurement is derived from its own distribution.

> **The MaxLIP method is able to account for only up to five separate distributions of residues and the user must use the Likelihood Ratio Test to determine if an adequate number of distributions is modeled. Does the Panel have any comments on this aspect of the program and how might this affect the adequacy of the decompositions which are performed? In contrast, RDFgen assumes that each that each composite is derived from a separate and distinct distribution and decompositing is performed by using the standard deviation of composite value measurements and assuming (once adjusted) that this applies to each composite. Does the Panel have any comments on these differences in approach and assumptions?**

Question 5

Although limited in scope, OPP's comparison of each method's ability to accurately predict individual item residue levels based only on information in residue levels in composite samples did not appear to provide any clear evidence of systematic over- or under-estimation of residues in decomposited samples. All three methods did not necessarily perform equally well (particularly at the upper and lower tails of the distribution) under all circumstances in predicting single-item *residue levels*, but differences in predicted *exposure levels* (and therefore risk levels) appeared to differ to a much lesser extent. This situation is not unexpected: it is often not the extreme upper tail of a *residue* distribution which are responsible for driving the 99.9^{th} or 99^{th} percentile *exposure* levels, but rather a combination of reasonable (but high end) consumption and reasonable (but high end) residue levels of one or two frequently consumed agricultural commodities. That is, it is not necessarily true that significant differences in predicted residue levels in the upper tail (e.g., $>95^{th}$ percentile) of the residue distribution will as a matter of course result in significant differences in predicted exposure levels at the upper tails of the exposure distribution since it is a combination of both consumption and residue levels over a wide variety of commodities which determine high-end exposure levels.

> **Does the Panel have any thoughts, insights, or concerns about the potential for underestimation or overestimation (or other biases) of residue levels by each of the two decomposition procedures being presented for consideration? Does any concern regarding over/under estimation extend to concern about over/under**

estimation of exposures (and therefore risks)? Can any characteristic statements be made about over- or under-estimation at various percentile levels (e.g., median, 75th, 90th, 99th 99.9th percentiles)?

FIGURES

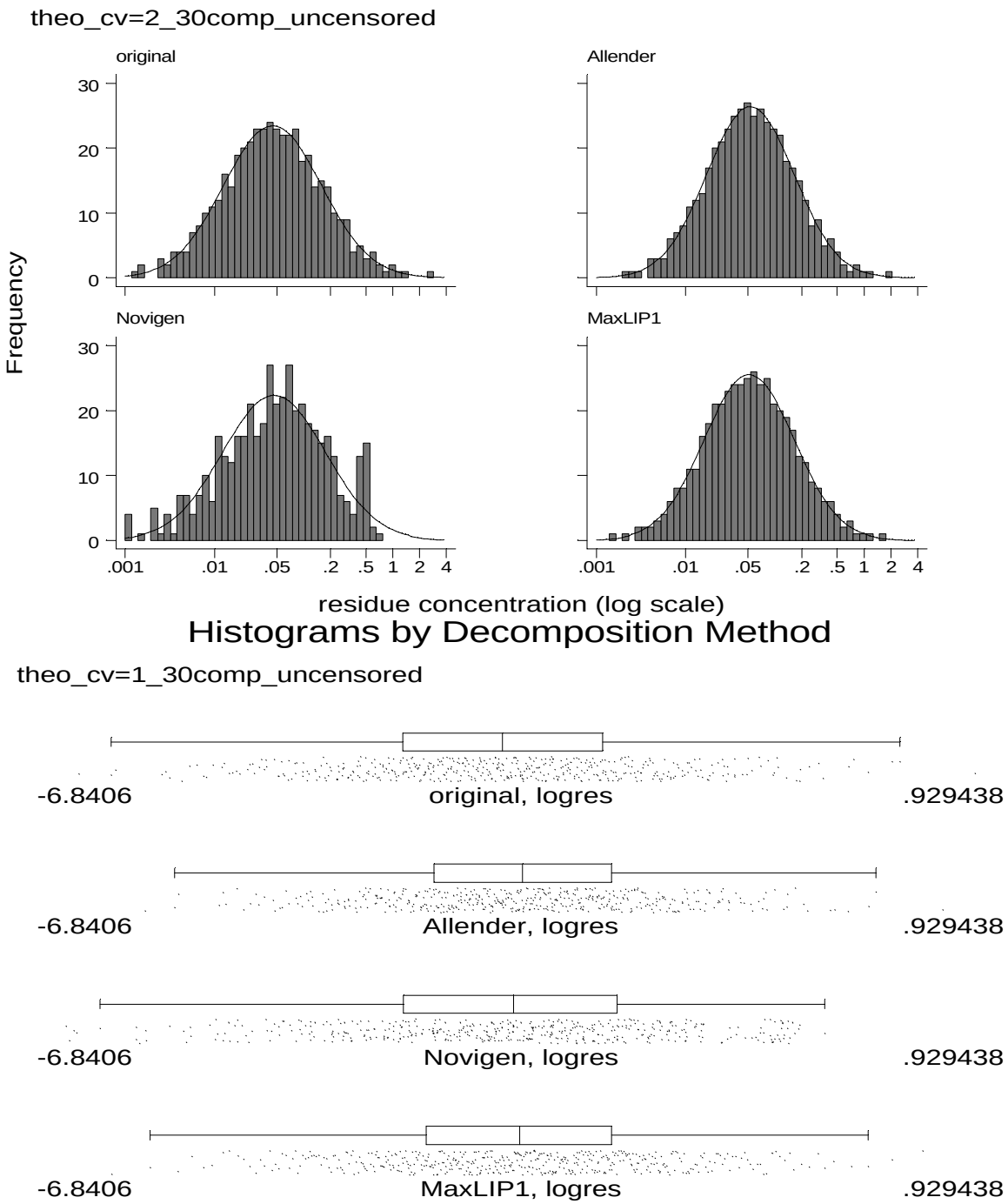

theo_cv=2_30comp_uncensored
original
Allender
Novigen
MaxLIP1
Frequency
0
10
20
30
.001
.01
.05
.2
.5
1
2
4
residue concentration (log scale)
Histograms by Decomposition Method
theo_cv=1_30comp_uncensored
-6.8406
original, logres
.929438
Allender, logres
Novigen, logres
MaxLIP1, logres
ln(residue concentration)
One-way Box diagrams by Decomposition Method

Figure 1

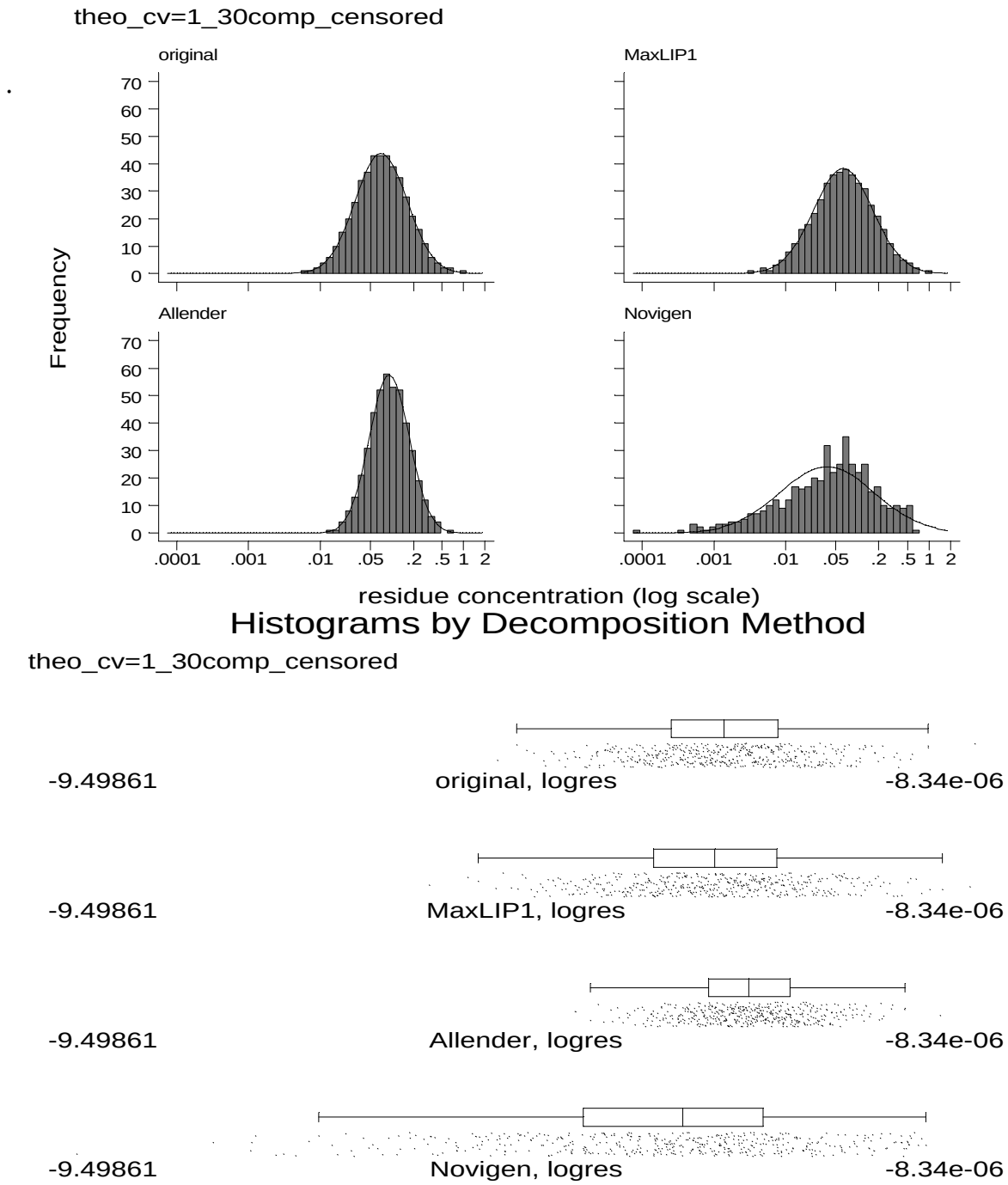
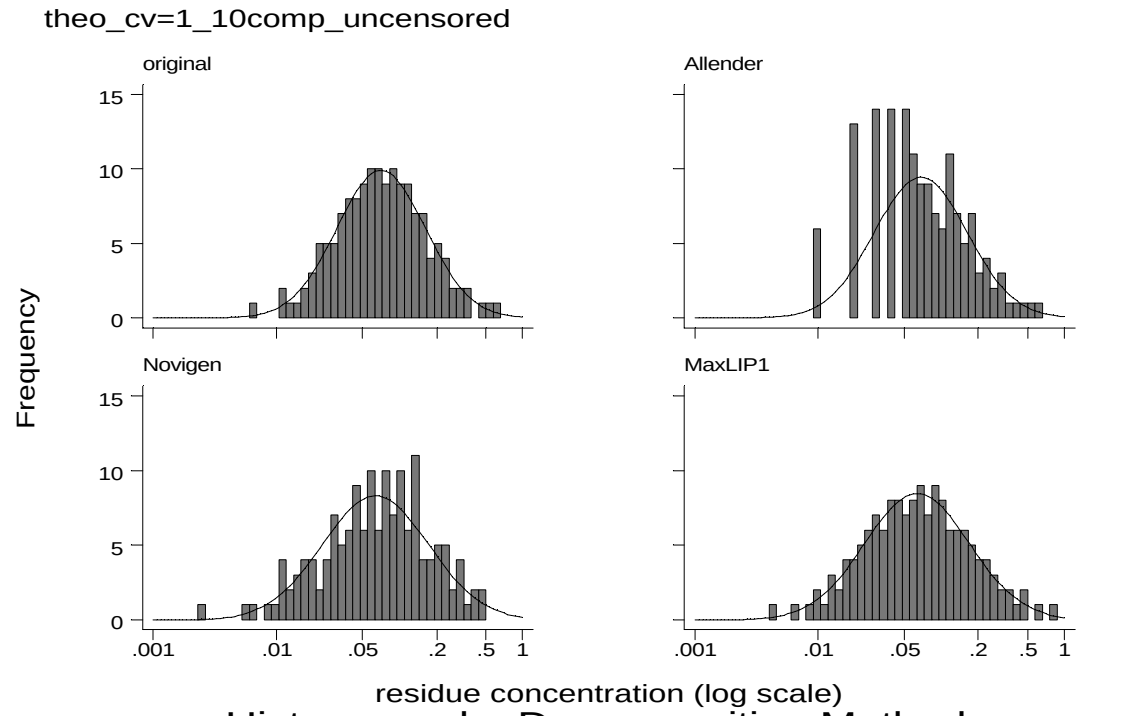

theo_cv=1_30comp_censored
original
MaxLIP1
Allender
Novigen
Frequency
0
10
20
30
40
50
60
70
.0001
.001
.01
.05
.2
.5
1
2
residue concentration (log scale)
Histograms by Decomposition Method
theo_cv=1_30comp_censored
-9.49861
original, logres
-8.34e-06
-9.49861
MaxLIP1, logres
-8.34e-06
-9.49861
Allender, logres
-8.34e-06
-9.49861
Novigen, logres
-8.34e-06
ln(residue concentration)
One-way Box diagrams by Decomposition Method

Figure 2


theo_cv=1_10comp_uncensored
original
Allender
Novigen
MaxLIP1
Frequency
0
5
10
15
.001
.01
.05
.2
.5
1
residue concentration (log scale)
Histograms by Decomposition Method

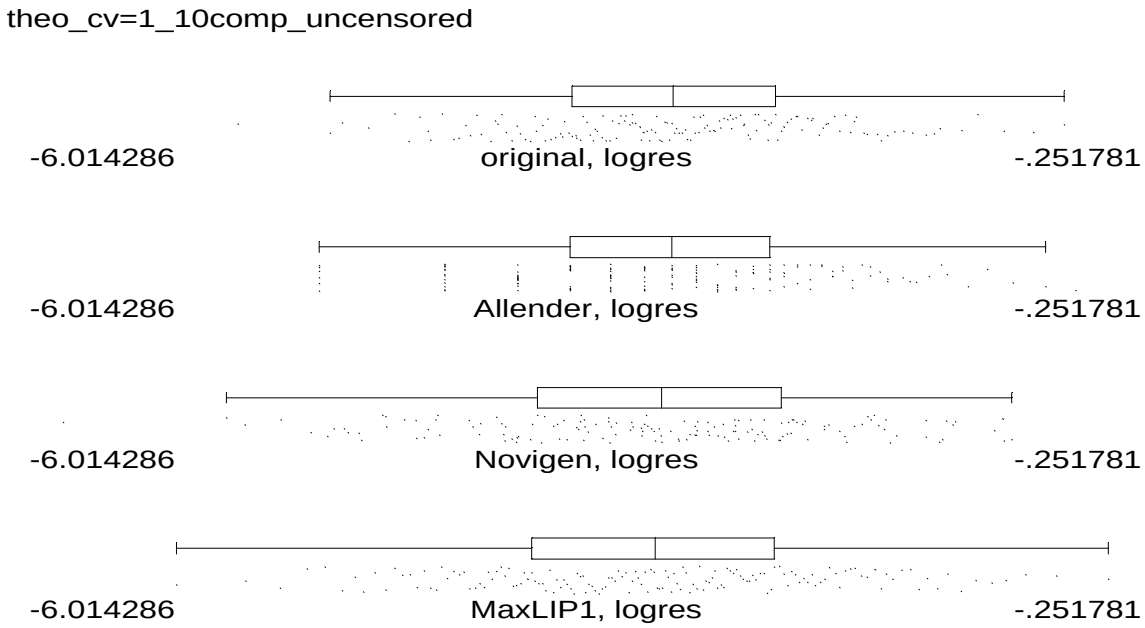

theo_cv=1_10comp_uncensored
-6.014286
original, logres
-.251781
-6.014286
Allender, logres
-.251781
-6.014286
Novigen, logres
-.251781
-6.014286
MaxLIP1, logres
-.251781
ln(residue concentration)
One-way Box diagrams by Decomposition Method

Figure3

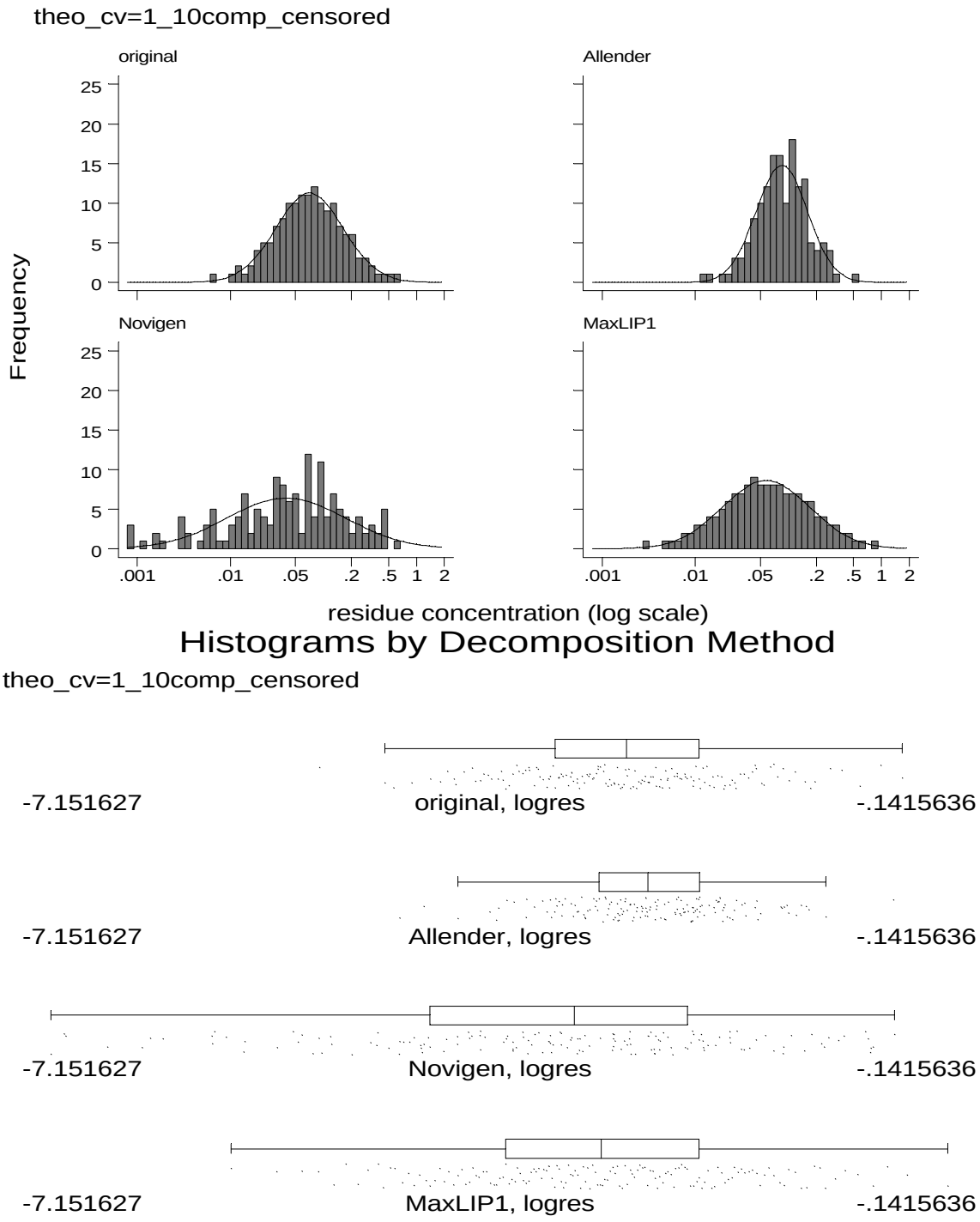

theo_cv=1_10comp_censored
original
Allender
Novigen
MaxLIP1
Frequency
0
5
10
15
20
25
.001
.01
.05
.2
.5
1
2
residue concentration (log scale)
Histograms by Decomposition Method
theo_cv=1_10comp_censored
-7.151627
original, logres
-.1415636
-7.151627
Allender, logres
-.1415636
-7.151627
Novigen, logres
-.1415636
-7.151627
MaxLIP1, logres
-.1415636
ln(residue concentration)
One-way Box diagrams by Decomposition Method

Figure 4

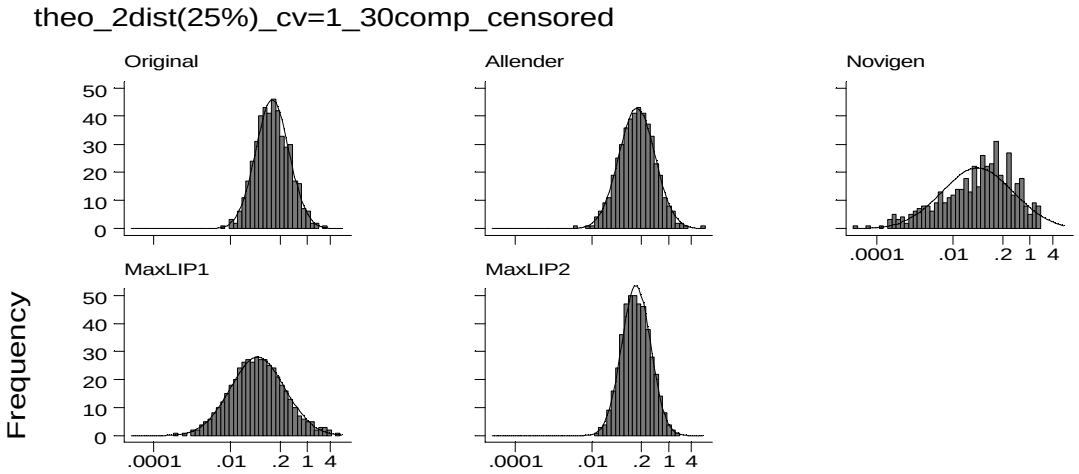

theo_2dist(25%)_cv=1_30comp_censored
Original
Allender
Novigen
MaxLIP1
MaxLIP2
Frequency
0
10
20
30
40
50
.0001
.01
.2
1
4
residue concentration (log scale)
Histograms by Decomposition Method
theo_2dist(25%)_cv=1_30comp_censored
-10.58859
Original, logres
2.206326
Allender, logres
Novigen, logres
MaxLIP1, logres
MaxLIP2, logres
ln(residue concentration)
One-way Box diagrams by Decomposition Method

Figure 5

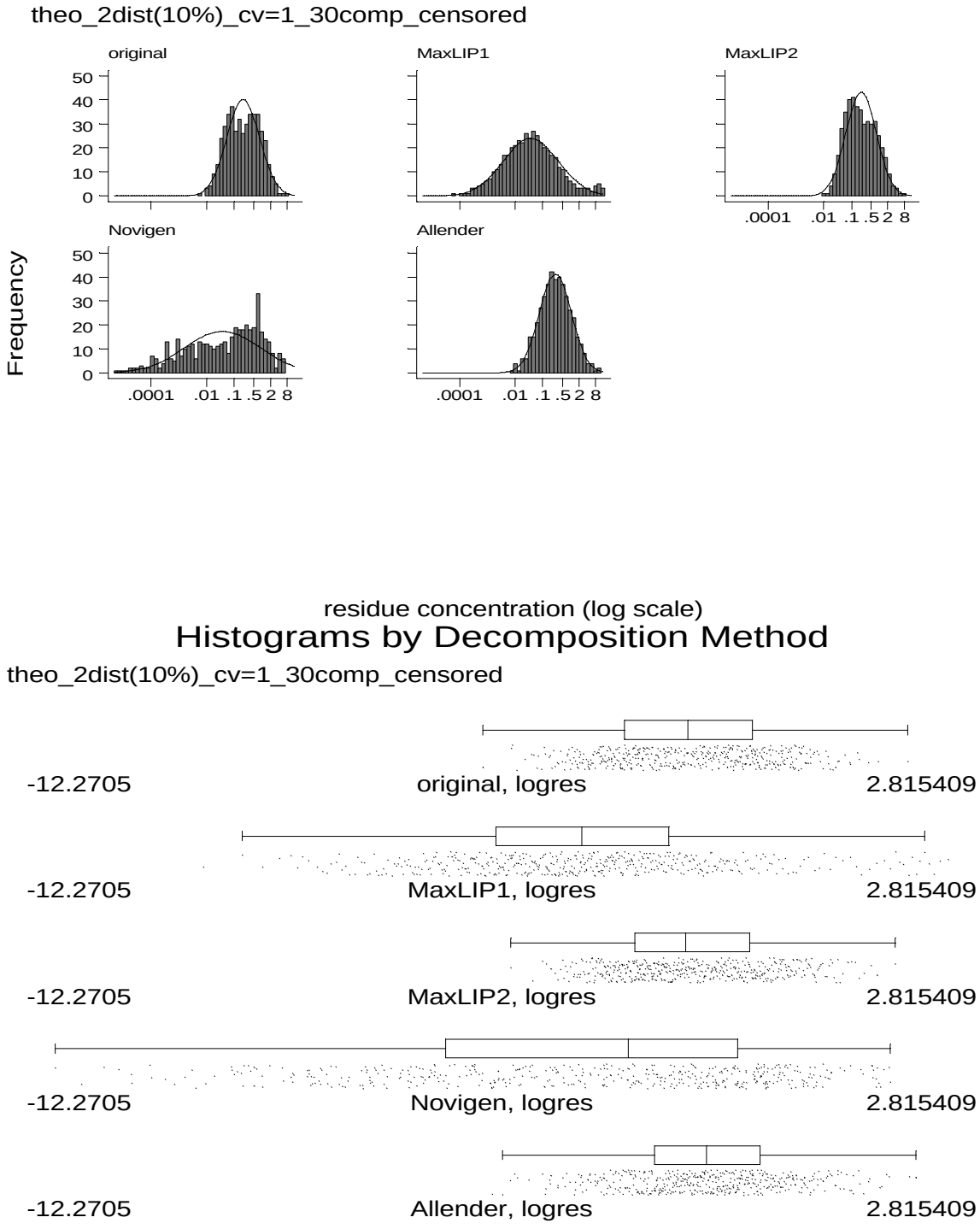
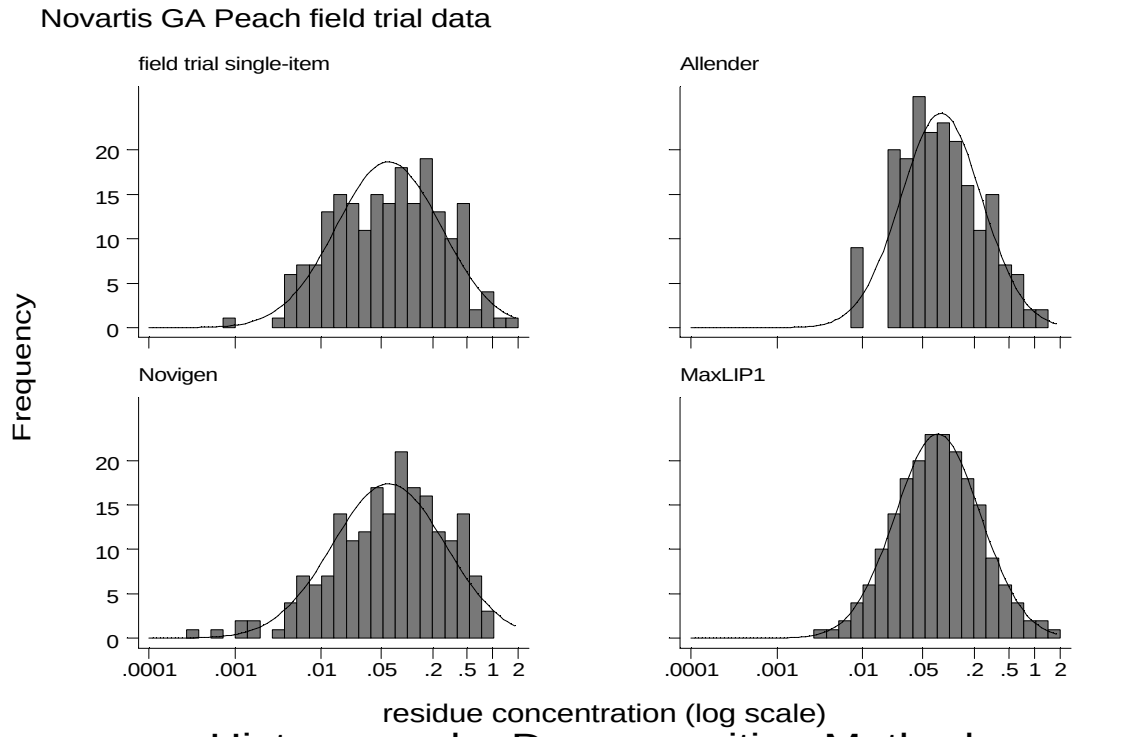

theo_2dist(10%)_cv=1_30comp_censored
original
MaxLIP1
MaxLIP2
Novigen
Allender
Frequency
0
10
20
30
40
50
.0001
.01
.1
.5
2
8
residue concentration (log scale)
Histograms by Decomposition Method
theo_2dist(10%)_cv=1_30comp_censored
-12.2705
original, logres
2.815409
MaxLIP1, logres
MaxLIP2, logres
Novigen, logres
Allender, logres
ln(residue concentration)
One-way Box diagrams by Decomposition Method

Figure 6


Novartis GA Peach field trial data
field trial single-item
Allender
Novigen
MaxLIP1
Frequency
0
5
10
15
20
.0001
.001
.01
.05
.2
.5
1
2
residue concentration (log scale)
Histograms by Decomposition Method

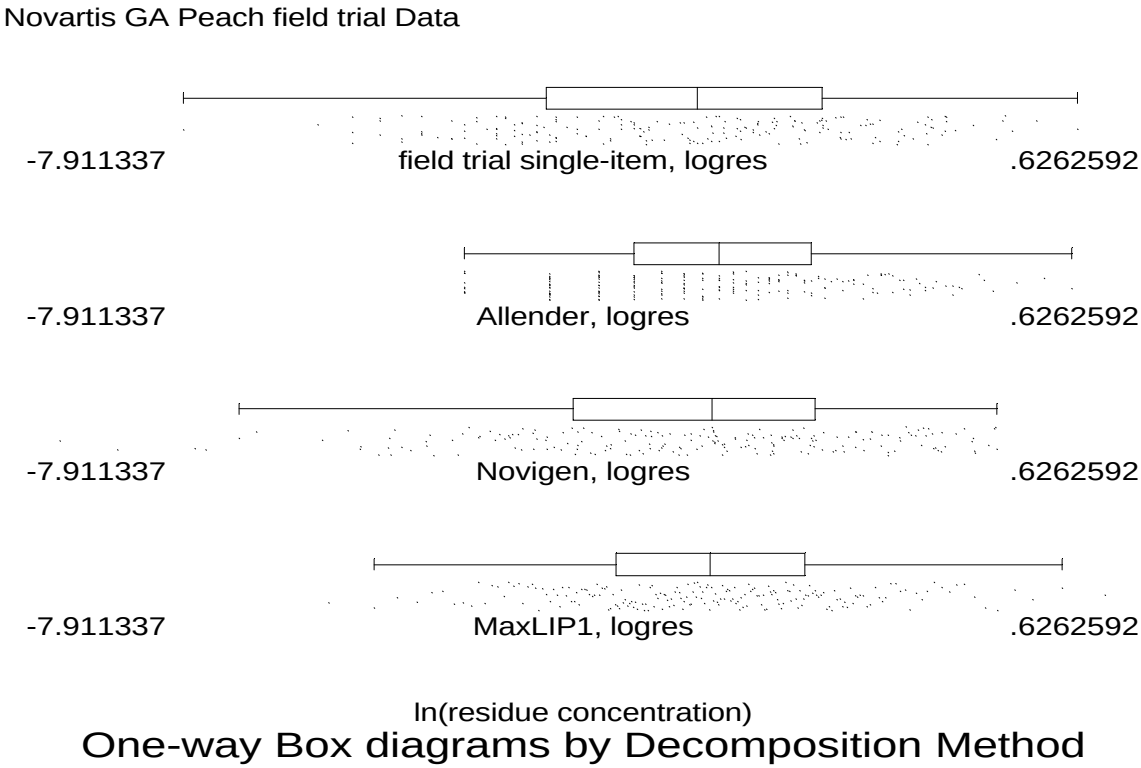
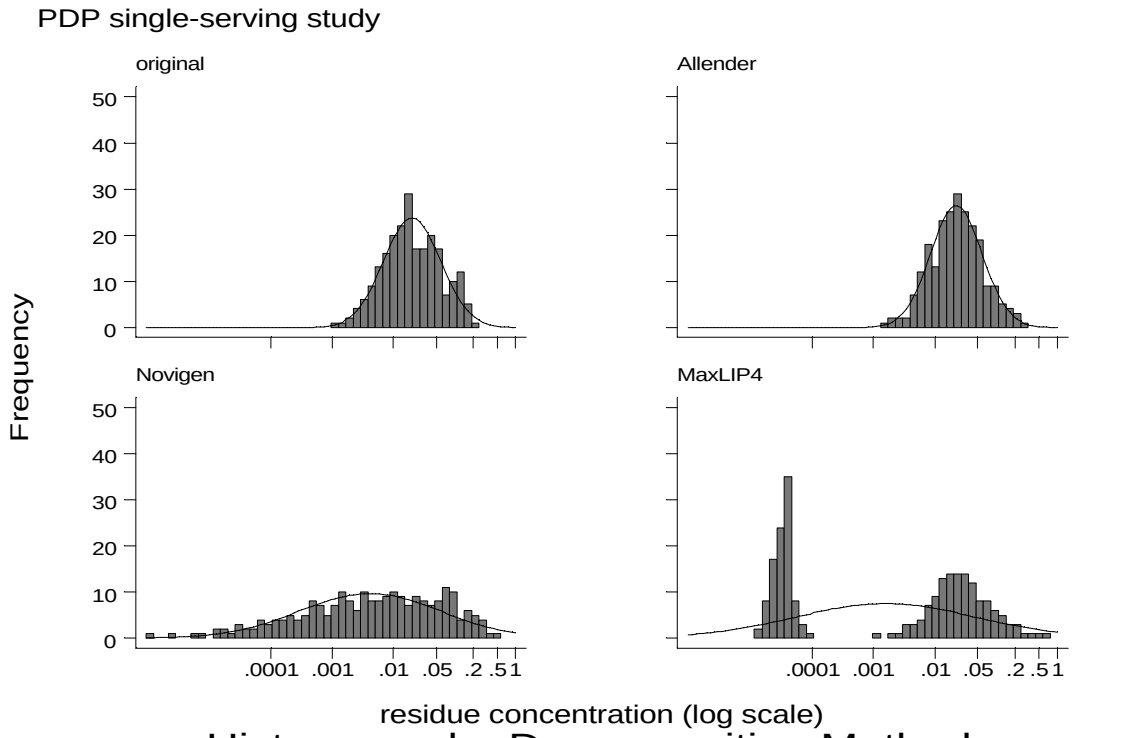

Novartis GA Peach field trial Data
-7.911337
field trial single-item, logres
.6262592
-7.911337
Allender, logres
.6262592
-7.911337
Novigen, logres
.6262592
-7.911337
MaxLIP1, logres
.6262592
ln(residue concentration)
One-way Box diagrams by Decomposition Method

Figure 7


PDP single-serving study
original
Allender
Novigen
MaxLIP4
Frequency
0
10
20
30
40
50
.0001 .001 .01 .05 .2 .5 1
residue concentration (log scale)
Histograms by Decomposition Method

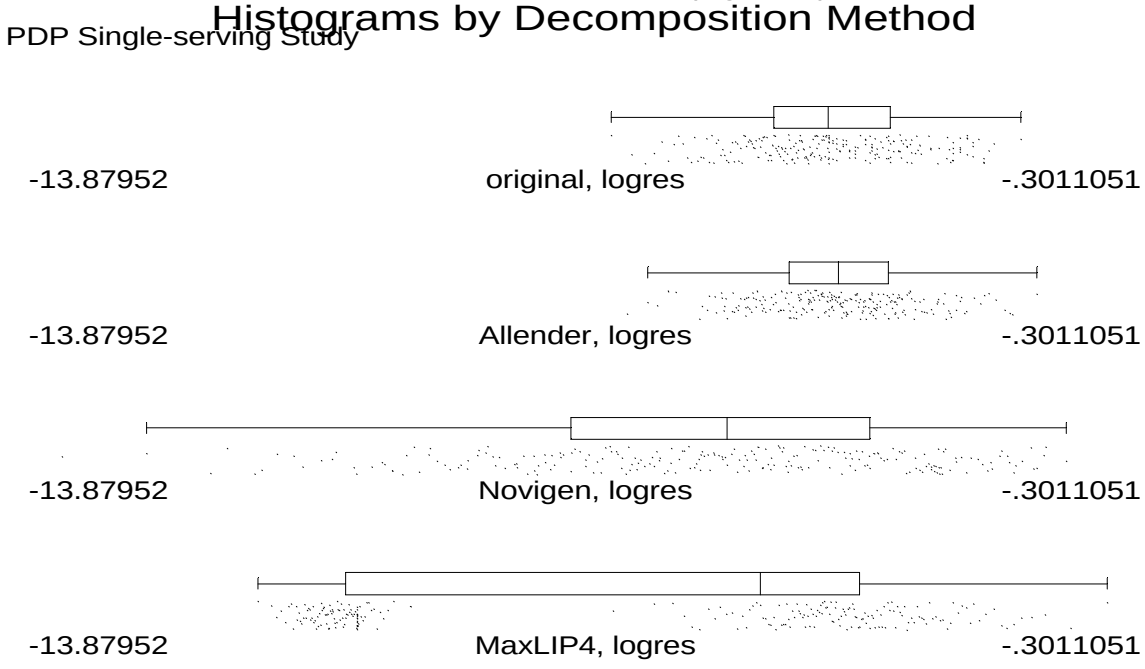

PDP Single-serving Study
-13.87952
original, logres
-.3011051
-13.87952
Allender, logres
-.3011051
-13.87952
Novigen, logres
-.3011051
-13.87952
MaxLIP4, logres
-.3011051
ln(residue concentration)
One-way Box diagrams by Decomposition Method

Figure 8

Appendix 1