dpr

Brian R. Leahy
Director

Department of Pesticide Regulation

M E M O R A N D U M

Edmund G. Brown Jr.
Governor

TO: Randy Segawa
Environmental Program Manager I
Environmental Monitoring Branch

FROM: Bruce Johnson, Ph.D. *Original signed by*
Research Scientist III
Environmental Monitoring
916-324-4106

Frank C. Spurlock, Ph.D. *Original signed by*
Research Scientist III
Environmental Monitoring Branch

DATE: January 18, 2013

SUBJECT: STOCHASTIC EVALUATION OF BACK CALCULATION PROCEDURES FOR ESTIMATING FLUX USING DATA FROM THE LOST HILLS STUDY

Background

The Department of Pesticide Regulation (DPR) has utilized the back-calculation technique (Ross et al. 1996) for estimating fumigant flux. In brief, this technique requires a collection of air monitors surrounding a field where volatilization occurs. These monitors run from periods of generally 6 to 24 hours and measure an average concentration over the monitoring period. Meteorological data collected during the air measurement periods are utilized in conjunction with field and monitor geometry as input to the Industrial Source Complex - Short Term Model (ISCST3).The ISCST3 is then used to estimate air concentrations at the monitors during each monitoring period. In order to run the ISCST3 model, an assumed flux is used (usually 100 ug/m2s or 1 ug/m2s). For each period, a regression is then performed using the ISCST3-modeled values as x values and measured values as y values. The slope of the regression line is used to adjust the artificial flux in order to estimate the actual flux.

The Lost Hills (LHS) Study is being used as a validation exercise for the HYDRUS model. If HYDRUS can be validated, DPR hopes to use the HYDRUS model, at a minimum, to estimate emissions for minor modifications, such as depth of application, in known application methodologies. In order to validate HYDRUS, one needs to compare the HYDRUS-estimated fluxes to the back-calculated fluxes. The error in each flux estimate provides a context for comparing the fluxes.

Majewski (1977) describes an analytical approximation method for estimating the error associated with measuring flux via the aerodynamic flux method. No similar error estimation procedure exists for the back-calculation method. Moreover, the back-calculation method has no generally accepted analysis pathway when the initial regression is not significant. When the initial regression is not significant, DPR sorts the ISCST3 and measured values independently and recalculates the regression based on those sorted values. If a 'significant' regression is still not achieved by sorting, then average measured air concentrations divided by average modeled concentrations for the period is used as a flux estimate. In the LHS study, Sullivan (Ajwa and Sullivan 2011) used a more elaborate method, which involves determination of the significance and size of the intercept, and either redoing a regression with no intercept or when that fails, using the mean measured divided by mean modeled procedure. In no case do Ajwa and Sullivan (2011) use a sorting method.

Because there is no obvious way to estimate variability in the back-calculation method and because of the varying analysis procedures employed when the initial regression is not significant, stochastic approaches were used in an attempt to compare the various methods and estimate the associated variability and bias of the back-calculated flux estimate.

This memorandum is a compendium of some stochastic analyses that we undertook to try to compare the various procedures used the back-calculation methodology. In some sense it is preliminary. We feel, however, that it can be used as a basis for future analysis and discussion. We welcome feedback on the approach that we took in this study and on the results of our analyses.

The purpose of the LHS study was to ascertain how long to hold a "Totally Impermeable Film" type of tarp in order to reduce emissions during tarp cut. The LHS study included four fields and two fumigants: 1,3-dichloropropene (1,3D) and chloropicrin (PIC) applied as the formulated product, PicChlor 60 (Table 1). More details can be found in Ajwa and Sullivan (2011).

Motivation for stochastic examination of various back-calculation analysis methods

A. Comparison of HYDRUS-simulated and ISCST3 Back-calculated Fluxes

The discussion in this section was originally a reason for embarking on stochastic simulations to examine the back-calculation procedure. However, our thinking about this issue has changed over time. Nevertheless, the clear conceptual framework has general applicability to modeling and questions of validation and so we include it.

Ideally, HYDRUS mean flux predictions for each sampling interval *i* ($f_{i,HYDRUS}$) would be compared to *actual* period mean fluxes in the field ($f_{i,actual}$):

$$f_{i,HYDRUS} + \varepsilon_{i,HYDRUS} = f_{i,actual} \qquad (1)$$

The HYDRUS period-wise modeling errors $\varepsilon_{i,HYDRUS}$ would then be compared to the actual flux to assess overall modeling error. However, $f_{i,actual}$ is not measured in field flux studies. It is unknown. Instead, measured air concentrations in conjunction with a model are used to estimate $f_{i,actual}$. Two common modeling approaches used to estimate $f_{i,ctual}$ are vertical profile techniques (e.g. the aerodynamic method, Majewski, 1995), or back-calculation methods using air-dispersion models such as the Environmental Protection Agency's ISCST3 as was done in the LHS study. A comparison between HYDRUS and either type of data amounts to comparing predicted fluxes of two models. For the back-calculation method used in the LHS study, the comparison takes the form:

$$f_{i,HYDRUS} + \varepsilon_{i,HYDRUS} = f_{i,back\_calc} + \varepsilon_{i,back\_calc} \qquad (2)$$

Where $f_{i,back\_calc}$ is the mean back-calculated flux estimate for sampling period i and $\varepsilon_{i,back\_calc}$ is the corresponding period-wise error compare3d to the unknown actual period mean flux. It is apparent from Eq. 2 that any meaningful evaluation of HYDRUS (i.e. $\varepsilon_{i,HYDRUS}$) requires estimates of $\varepsilon_{i,back\_calc}$. Flux estimation methods that do not provide an estimate of error will limit the ability to make comparisons between HYDRUS-estimated and back-calculated estimated flux.

We had hoped to utilize the stochastic simulation results to estimate $\varepsilon_{i,back\_calc}$. However, we now believe that performing this exercise properly requires nearly certain knowledge of the 'true' flux for each period. Otherwise, the stochastic simulations are not being compared to the real thing. The back-calculation procedure relies on the fundamental Gaussian equation which expresses a proportional relationship between the flux and estimated concentration. All other features such as the geometry (that is the positional relationship between the source and the receptor), wind direction and speed and stability classification are fixed for that hour which is being simulated. However, errors of two broad kinds can affect ISCST3 estimates. The first kind of errors are model misspecification errors, where the model does not accurately describe the physical scenario. For example, the way that ISCST3 expresses the impact of stability classes may be incorrect in certain meteorological situations. A transient morning low-level inversion may not be captured by the ISC hourly meteorological summaries.

Another example of misspecification error is the possibility of nonuniform flux across the field. When using ISCST3 for back-calculation, a uniform flux is assumed. But nonuniform, high fluxes in proximity to specific samplers could increase measured concentrations in those samplers and affect the regression results between the modeled and measured concentrations. The nonuniformity of the flux could only be ascertained if flux measurement systems were able to measure flux with a much higher degree of spatial resolution. Off-field air monitors integrate measurements wide areas on the field and cannot distinguish hot-spots.

Another possible misspecification error, analogous to spatial nonuniformity of flux, is temporal nonuniformity of flux. A common pattern after fumigation is higher flux earlier on and then declining flux later. Within a measurement period, flux is assumed to be relatively constant for purposes of back-calculation. If not, then some hourly meteorological conditions are 'weighted' more heavily (during higher flux) over others (lower flux) and, as a consequence, the unweighted simulation will not properly estimate the average concentration. In practice, field sampling schemes often utilize shorter time periods in the periods immediately following fumigation in order to avoid sampling over large flux changes. Later on, the sampling periods are lengthened, reflecting the slower change in flux with the passage of time.

ISCST3 does not realistically represent the downwind movement of source pollutants. The Gaussian equation has been described as a 'lighthouse' model. That is, the predicted downwind air concentrations are modeled as though they occur instantaneously with each changing hour. There is no passage of time reflecting the movement of molecules from the field to downwind locations. Thus ISCST3 does not capture realistic downwind plume movement. Other models have attempted to incorporate a more dynamic plume movement (Scire et al. 2000).

The second kind of error relates to the nature of Gaussian plume prediction. The ISCST3 model is designed to predict the ensemble mean concentration. This predicted ensemble mean will differ somewhat from any individual realization, in particular from the actual realization which occurred during an actual monitoring period. That realization is only one of a theoretically infinite number of realizations making up the ensemble. Within a given hour the same average wind direction and wind speed could be attained by many different time series. Each different realization may produce a slightly different hourly average concentration at a receptor. If it were possible to sample repeatedly using different realizations, but each having the same average wind speed and direction, then in theory the model would predict that average concentration. It is difficult to estimate or measure the variance for this kind of error.

Because of these two types of errors and because we do not have highly certain flux measurements, we cannot use this stochastic analysis to estimate $\varepsilon_{i,back\_calc}$.

B. Flux estimation procedures using the back-calculation method in the Lost Hills Study

After a preliminary step, judging data sufficiency, the back calculation method used in Ajwa and Sullivan (2011) starts with a simple linear regression which yields a slope and an intercept (Figure 1). A series of tests and procedures may lead to a second regression where the regression is forced through the origin (no intercept) or the calculation of the mean measured over mean modeled, in order to obtain a multiplicative factor to adjust the assumed flux in the ISCST3 model. In contrast the DPR method utilizes sorting when the initial regression is not significant. If sorting still does not yield a 'significant' regression, then mean measured over mean modeled is used (Figure 2). The actual spreadsheets in Ajwa and Sullivan (2011) utilize a significance level of 0.054. DPR utilizes a significance level of 0.05.

In both Sullivan's and DPR procedures, the chosen significance level is a screening value used to classify whether the original regression was statistically significant. When a regression is forced through the origin, or when the regression is performed on sorted values, the p value which results is not a true p value in a conventional sense of the term; hence we are describing it as a screening significance level.

For example, after forcing a regression through the origin, the 'significance' level can be misleading (Figure 3). In this example X and Y have no actual functional relationship with the 9 x,y pairs placed at nodes on a square grid. Ordinary least squares regression (OLR) gives completely nonsignificant results ($p>0.9$). However, by forcing through the origin, EXCEL reports a significant regression ($p<.002$) with a high R^2 of 73% (Figure 3). Furthermore, the slope error is 0.18, only 20% of the fitted slope of 0.86 – even though no true relationship between X and Y exists ($r = 0.00$). Given these results the physical interpretation is potentially a problem and a significance level does not reference the same kind of probability as in the OLR with slope and intercept.

The DPR methodology has its own problems. The sorting method will typically lead to a 'significant' regression. That is, $p<0.05$. However, the meaning of this statement is called into question because simply generating random numbers and sorting them will more often than not lead to a 'significant' p value, when no underlying relationship exists. In a 10,000-trial simulation where x values were simulated as uniform random between 0 and 10 and y values were simulated as normal mean 5 and standard deviation 1.0 (no correlation between x and y), 5% of the trials yielded significant OLR regressions. This is exactly what would be expected with completely random data since the significance level is set so that Type I errors (the probability of falsely rejecting the null hypothesis) occur in 5% of the trials. However, after sorting the x and y values and recalculating the regression, 100% of the regressions were statistically significant. Thus in both methods (forcing through the origin or sorting), the significance level is being used as a screening procedure to assess goodness of regression fit, but can be misleading for the reasons described above.

These two contrasting pathway for analyzing field study results in order to estimate flux involve four basic analysis approaches: OLR (simple linear regression with slope and intercept), OLR with no intercept (simple linear regression but forced through the origin), sorting and then regressing, and [meanY/meanX] (divide the mean of measured concentrations by the mean of the ISCST3-modeled concentrations). We propose a stochastic approach for comparing the variability and bias amongst these four basic statistical procedures for estimating the slope.

Methods for stochastic examination of back-calculation

The back-calculation method assumes

$$F_i = L_i T \qquad (3)$$

Where F_i is the actual flux density for period i, T is the assumed flux density used in the ISCST3 model to simulate this period, and L_i is the constant of proportionality between the actual flux density and assumed flux density which is needed to simulate the period concentrations.

Because the Gaussian model assumes a proportional relationship between flux and concentration, there is a proportional relationship between each measured concentration at a particular sampler during a particular period and the ISCST3 estimate for that sampler during that period and the constant of proportionality is the same as in equation (3).

$$M_{ij} = L_i C_{ij} \qquad (4)$$

Where j is the jth sampler during the ith period, M_{ij} is the measured concentration at sampler j during the ith period and C_{ij} is the ISCST3 predicted concentration (based on the assumed flux T) at sampler j for the ith period.

The goal of the back-calculation method is to estimate Li for each period and to estimate the 'true' flux with equation (5). The procedural pathways outlined in Figures 1 and 2 are comprised of four basic methods for estimating Li: OLR, OLR forcing through origin, sorting, and [meanY/meanX]. Sources of error in the back-calculation procedure may include: (1) inhomogeneous spatial or temporal flux during the period, (2) representation of each hour using mean wind direction and wind speed, (3) chemical analytical variability, (4) idiosyncrasies of the particular field such as presence of structures, berms, nearby crops or other physical features which may alter air flow over the field in ways not captured by the meteorological data, (5) transient meteorological events not captured by the meteorological summaries, (6) inadequacy of the stability class or misclassification of stability class in relation to the actual field conditions, and (7) limits to chemical analysis and the representation of nondetects.

To represent the sampling and estimation procedures, the following representational model was adopted:

$$Y = LX \qquad (5)$$

Where Y represents measured values, L is a random variable with a defined arithmetic mean and X is a random variable representing the ISCST3-generated estimates.

Two similar approaches were used for the stochastic simulation analyses. The first approach used 1,3D – based statistics from the 42 periods of field 1 in LHS study and the second used PIC-based statistics from the 42 periods of field 1. The two approaches differed in some details and will be described in separate sections.

1,3D Stochastic Simulations

X is a random variable distributed as described below based on analysis of period 1-42 field 1 1,3D air concentrations, while L is a random variable with a defined arithmetic mean = 100, and a distribution and standard deviation chosen such that the resulting "within sampling interval" Y distribution is similar to the within interval 1,3D period 1-42 ISCST3-modeled air concentrations. For each of the 8 (X,Y) pairs in each of the 1000 simulated monitoring intervals, X and L were independently sampled from their respective sampling distribution to obtain Y (Eq. 5)

The 1000 simulated monitoring intervals of 8 (X,Y) data pairs were used to compare OLR, sorted regression, forced origin regression and [mean Y/mean X] methods for estimating the (known) L. Method comparisons are based on (a) how accurately each method estimated mean L, and (2) the standard deviation of those mean L estimates. The methods were compared using subsets of the 1000 L,Y interval data segregated as to (a) significant OLRs, (b) insignificant OLRs, and (c) OLRs with intercepts significantly different than zero regardless of the significance of the overall regression.

1,3D Selection of Sampling Distributions

In selecting distributions for X and L, the objective was to obtain (X,Y) data that were 'similar' to actual ISCST-modeled (X) and measured (Y) air concentration data obtained in LHS. To assess this we used periods 1-42, field 1, 1,3D data (n=668 measured/modeled air concentration data pairs). By similar we mean:

- The stochastically sampled *X* should have similar within interval μ and σ as the ISCST3-modeled data, and a similar distributional shape based on a cumulative frequency plots of standardized *X* and modeled data. Each standardized datum was calculated as $[(X_{i,j}-\mu_i)/\sigma_i]$ where μ_i is the interval *i* mean X, σ_i is the interval *i* standard deviation of X, and $X_{i,j}$ is the j_{th} datum in the i_{th} interval.

- Stochastically sampled *Y* (each based on the product of a sample from the L distribution and the X distribution, equation 5) should have a similar coefficient of variation (CV = σ/μ) as the measured data, although the mean will differ from measured data since the mean is determined by L's chosen mean of 100. The CV of *Y* was approximately matched to the CV of measured air concentration data by 'tuning' the standard deviation of L. The overall similarity of *Y* and measured air concentration data was evaluated qualitatively using cumulative frequency plots and probability plots of the standardized *Y* and measured air concentration data.

- Distribution of correlations between X and Y should be approximately equal to that for measured and modeled data.

1,3D Results

A. Sampling Distributions

Distribution of X: We obtained a 'nearly best-fit' to the ISCST3-modeled data from the LHS using beta distribution with minimum, maximum, alpha and beta of 0, 35.14, 1.066 and 6.537, respectively. These distribution parameters yielded a mean of the 'within interval' means,standard deviations and CVs for the 1000 synthetic 8 member X data vectors of 4.87, 3.87 and 0.79, respectively. These compare favorably with the mean within interval means, standard deviations and CVs for period 1-42 ISCST3 field 1 data of 4.46, 3.55 and 0.80 (Figure 4).

Distribution of actual 1,3D concentrations, L and the calculated Ys (Eq 3): The Anderson-Darling test statistic for normality of the standard normal deviates of the log-transformed measured air concentrations was large enough to reject the null hypothesis of normality (Figure 5). The more general omnibus D'Agostino-Pearson K2 test (d' Agostino et al. 1990) indicated insufficient evidence to reject the null hypothesis (p=0.105).

Based on the latter normality test and linearity of the plot in Figure 5, we concluded the measured air concentration data were "close" to log normal, and so selected a lognormal sampling distribution for L (thereby obtaining a log normal distribution for Y). The arithmetic mean for that distribution was defined as 100. We found that a standard deviation of 160 provided within interval Y distributions that were generally similar in shape to those of the measured data (Figure 6). The mean within interval CV was 1.29 for the untransformed Y data, comparing favorably to the mean within interval CV of 1.36 for the measured air concentrations.

Using the sampling distributions of X and L described above, 1000 data sets consisting of 8 pairs each of X,Y data were created using Monte Carlo sampling. The resulting X,Y data had similar CVs and correlation structure as the modeled and measured 1,3D air concentrations in LHS Field 1, periods 1-42. The mean within-interval correlation between the Monte Carlo Y and X was 0.53 (N=1000 intervals), while that between the measured and modeled field 1 1,3D air concentrations was 0.55 (N=42 intervals). These X,Y data were the used to assess the accuracy of the 4 methods for estimating the known L.

B. Comparison of Flux Estimation Methods

Significant OLRs: Although the correlations between X and Y were similar to those between measured and modeled, only 33% (334 of 1000) of the simulated intervals yielded significant OLRs as compared to 71 percent among the 42 1,3D field 1 intervals. This is attributable to the larger number of measured versus modeled data pairs (16 pairs for Field 1 periods versus 8 pairs for the simulated X,Y data). Among methods, the [mean Y/mean X] procedure yielded a mean estimate for L of 105 (96-114, 95% CI), that was closest to the true value of 100 (Figure 7); the 95% CI for the remaining 3 methods did not contain 100. In addition, the variance of the mean [mean Y/mean X] was significantly lower than the OLR and sorted OLR methods (e.g., Figures 7 and 8). In a subset of the significant OLRs with $r>0.90$ (N=88; results not shown), the outcome was similar; the [mean Y/mean X] method yielded more accurate and less variable estimates of L than the other methods.

Insignificant OLRs and Significant Intercepts: Two other cases are illustrated here: regressions that were not significant ($p>0.05$), and regressions with significant intercepts regardless of the overall significance of the regression. In both cases, the [mean Y/mean X] procedure outperformed the remaining three procedures with more accurate and less variable L estimates (Figures 9-11).

Other Distributions Examined: A variety of other distributions for X and L in Eq 3 were also used to synthesize additional X,Y interval datasets. These included lognormal, left-censored (at zero) normal and gamma distributions for L. Without exception, the [mean Y/mean X] procedure yielded the most accurate and least variable estimates of L.

C. Slope Estimation and Error Estimates: Simulated data based on a Specific Sampling Interval

We evaluated (a) the slope estimation methods previously discussed, and (b) a method for estimating error using simulated (X, Y) data that were similar to LHS field 1 interval 11 measured/modeled 1,3D air concentrations. Interval 11 was chosen because the CV for the interval 11 measured data (1.44) was similar to the mean within interval CV (1.36) over all 42 intervals; hence interval 11 was considered “representative” from that standpoint.

The procedure for creating the 1000 simulated X,Y data sets was generally similar to that previously described with the following exceptions. The sampling distributions for the random variables X and L in Eq. 6 were the beta distribution and log normal distribution, respectively, as before. However, the parameters of the distributions were chosen such that the resultant means and CVs of X and Y closely matched sampling interval 11 modeled and measured mean and CVs (Table 2). We adjusted the parameters of the X beta distribution and the slope L log normal distribution by trial and error until X,Y means and CVs from samples of those distributions were close to those for interval 11. The 'optimized' parameters were for the beta distribution were minimum = 0, maximum = 13, alpha = 0.5, and beta = 1.2, while those for the L log normal distribution were arithmetic mean = 5.75 and standard deviation = 7.75. One thousand sets of X,Y data were then created using Monte Carlo sampling as before.

Comparison of Slope Estimates by Method: The relative accuracy and variability of the [mean Y/mean X], OLR and sorted OLR methods to predict the true mean of L (slope) was similar to the previous examples: the [mean Y/mean X] yielded the most accurate and least variable estimates regardless of whether a significant correlation existed among the X and Y variables (Figure 12).

The OLR residuals were strongly heteroscadastic, similar to the pattern observed for many of the actual sampling intervals (Figure 13). As a result, meaningful standard errors for the slope cannot be determined using OLR.

PIC Stochastic Simulations

These simulations followed roughly the same procedures as the 1,3-D simulations and used equation 5 as a starting point.

The PIC distribution for X

The 42 PIC periods were stored in an Excel worksheet with indices consisting of period and sampler. Most periods had a full 16 samples for Field 1, though a few had 15 samples. The 'X' column consisted of all of the ISCST3 estimates from Ajwa and Sullivan (2010) for Field 1. This column was fit using Oracle Crystal Ball, Fusion Edition Release 11.1.2.000 (32-bit). The best fitting distribution was a beta distribution with minimum -0.56, maximum 35.14, Alpha 1.06618, Beta 6.53668. As a sidenote, the CB documentation refers to this as the "Beta Distribution". In fact, it is the beta probability density function. If one integrates the beta probability density function, it can be compared to the built-in beta.dist worksheet function in Excel which gives the beta cumulative distribution function. Note that the beta distribution function is only defined on the interval [0,1] so that data will have to be mapped from [min,max] to [0,1] using the obvious linear transform. One can also approximate the CB beta density function by finding the derivative along the beta.dist cumulative distribution function given by Excel.

The CB distribution fitting results gave the lowest Anderson-Darling and Kolmogorov-Smirnov fitting metrics to the beta distribution compared to the other available distributions in CB and hence was chosen as the distribution for representing the ISCST3 PIC modeled values (X) for field 1 from the LHS study.

The PIC distribution for Y (measured PIC concentrations)

Of the 42 periods in Field 1, 22 gave significant ($p<.05$) regressions and 3 showed significant intercepts (Table 3). This amounted to 52% of the period with significant regressions and 7% with significant intercepts.

The mean within period coefficient of variation of the measured PIC concentrations was 1.46. Using equation 5 and a lognormal distribution with a mean of 1 (the 1,3D section used a mean of 100) and simulating 16 X,Y values for each period (the 1,3D section simulated 8 pairs per period), we simulated 1000 periods at each of several standard deviation values of the lognormal ranging from 1.2 to 2.2 (Table 4). For each stochastically simulated period, we calculated whether the initial regression was significant, whether the slope was significant and the coefficient of variation of the Y (‘measured’) values. The fraction of significant regressions declined as the standard deviation increased. The within-period coefficient of variation for the Y values increased and the fraction of significant intercepts was largely unchanged as the standard deviation of the underlying distribution increased.

In order to find a standard deviation which reflected both the fraction of significant regressions and within-period CVs from the LHS study, we separately fit the significant regression fraction as a function of the standard deviation and fit the within-period CV as a function of the standard deviation using Table Curve 2D V5.0, their equation 34: $\ln(y)=a+b*\ln(x)$. In both cases, the r^2 values exceeded 99.9%. The resulting equations (untransformed) were

$$fractsig = e^{(-0.17845)} * SD^{(-0.68456)} \tag{6}$$

$$withinCV = e^{0.28557} * SD^{0.32575} \tag{7}$$

We set up the following equation for optimization which measures the distance between the predicted fraction of significant regressions and the actual fraction and adds to that the distance between the predicted within period CV and the actual within period CV:

$$goal = (e^{(-0.17845)} * SD^{(-0.68456)} - 0.52)^2 + (e^{0.28557} * SD^{0.32575} - 1.46)^2 \tag{8}$$

We used Excel equation solver to find the minimum for *goal*. The minimum was achieved at SD=1.566. At this value for the standard deviation, the mean fraction of significant regressions was calculated to be 0.615 and the mean within period CV was calculated to be 1.54. Approximate 95% confidence intervals for the fraction of significant regressions (0.37 to 0.68) and for the CV (1.32 to 1.60) both encompass the target values of 0.52 and 1.46, respectively. Therefore, the lognormal distribution for L used a mean=1.0 and SD=1.566 for stochastic simulation to compare the back-calculation methods.

The PIC analysis utilized a somewhat different approach to the stochastic simulation than the approach used in the 1,3-D analysis. The latter set up 1000 rows in an Excel spread sheet, each row containing the X,Y values to be regressed. Summaries were calculated based on these 1000 trials. For the PIC analysis, Crystal Ball was used to iteratively sample from the distributions, run the analyses and statistically summarize the results. On each iteration, 16 X,Y values were produced from the specified distributions, OLR was calculated, a regression through the origin was calculated, the mean Y over mean X was calculated, and the two vectors were sorted and regression performed on the sorted vectors. Crystal Ball has a 'forecast' feature which designates a cell to be the collector of statistics on whatever function of the simulated values is desired. For the PIC analysis Forecast cells were used for the means, medians, standard deviations and within period CVs.

In order to separate the two cases where the initial OLR was statistically significant or not, Crystal Ball provides a filter procedure. However, once the filter is defined, statistics on the discarded cases are lost. Therefore, two separate 10,000 run trials were conducted, one to examine the initially significantly OLRs and one to examine the initially non-significant OLRs. This resulted in 6186 cases with significant OLR and 3855 cases where the initial OLR was not significant (Table 5). This fraction of significant OLRs closely reflected the predicted fraction of 0.615 based on the SD of the lognormal distribution that was determined in the optimization of equation 8, and was comparable to the actual observed data (22 significant OLRs/42 periods = 0.524, Table 3). For the synthetic Y values, the average of the within period CVs was 1.54 which was the expected value based on equation 8 and the comparable to the measured value of 1.46. Similarly, for initially nonsignificant OLR, the mean within period CVs for the synthetic Y values was 1.55.

The mean for the [meanY/meanX] estimator was 1.0 regardless of whether the OLR was initially significant or not. The OLR method tended to overestimate the mean by 22% amongst the significant OLR cases and underestimated slopes by 35% for the not significant cases. Sorted OLR overestimated the true value by 58%, regardless of the significance of the initial OLR. The forced OLR overestimated the mean by 10% when the OLR was initially significant, but underestimated by 15% when the initial OLR was not significant. In the DPR procedure if the initial regression is not significant, the next step would be sorting, and thus overestimation may result. In Sullivan's procedure, the next step following an initially nonsignificant OLR is forcing through the origin, which would result in some underestimation, according to the results of these stochastic simulations.

The standard deviations of the slope estimates for each method were also consistent with the 1,3D simulations. The [meanY/meanX] method generally had the lowest standard deviation and coefficient of variation amongst the four methods, regardless of the initial significance of the OLR. For OLR the CV in the initially significant cases was 0.99 (100% as a percentile). The CV was higher than the other methods, though the absolute SD for the OLR was lower than corresponding values for the sorting method. Across the spectrum of initially significant versus initially not significant OLRs, the the [meanY/meanX] estimator performed best.

Effect of additive constant in the PIC simulations

Equation 5 was modified by adding a constant:

$$Y = LX + K \qquad (9)$$

The Gaussian equation embodies a proportional relationship between flux and concentration. Consequently, based on this equation, the expectation would be that K=0. However, the denial of an intercept violates statistical procedures and may lead to grossly incorrect conclusions (Figure 3). Therefore, we decided to examine the performance of the four methods in the presence of a known intercept. The PIC-based stochastic approach was used.

In order to determine how large to make the constant, K, the 10%, 30% and 50% percentile values from the synthetic measured values distributions were determined for both initially significant OLR (five estimates, left hand side of Table 6) and initially not significant OLR eight estimates, right hand side of Table 6). The differences at each percentile were not great and so were averaged together to determine the constant K. We used the 10%, 30%, and 50% percentile values (ie. K=0.05, 0.58, 1.54) as a constant in equation 9. Altogether in this additive constant section, we conducted six stochastic simulations (each with n=1000) to compare the performance of the four methods. There were six simulations when the three levels of K were combined with either significant or nonsignificant OLR (Table 7).

Table 7 provides various statistics on each of the six simulation scenarios. The Trials variable is a count of the number of simulations passing the filter and is the sample size upon which the statistics are based. The filter for the left side was regression significant ($p<0.05$) and the right side was regression not significant ($p>0.05$). The filtered values variable is the number of trials not satisfying the filter criteria. Within each of the six cases the Trials plus Filtered Values = 1000.

The CV of ‘measured’ concentrations refers to the within period coefficient of variations of the Y values generated by simulating equation 9. These synthetic Y values correspond to the measured concentrations in the LHS study. Each CV was calculated based on 16 simulated values. For both significant and nonsignificant OLR, these CVs gradually decreased as K increased. This makes sense since the additive constant would tend to reduce the variance of the resulting Y values.

The Reg. Sign Flag and Int Sign. Flag stand for Regression significance flag and intercept significance flag. These were 0,1-valued functions in the spreadsheet. The Reg.Sign. Flag was 0 when the initial OLR was not significant and 1 when it was. The Int Sign Flag was 0 when the intercept was not significant and 1 when it was. For the Reg. Sign Flag on the initial OLR Significant side, the fraction was always 1.00 and conversely was 0.00 on the nonsignificant side of the table. This was included as a check that the Crystal Ball filtering procedure was being correctly applied.

The Int Sign. Flag mean represents the fraction of cases where the intercept was statistically significant. On both sides of the table, this fraction increased as K increasee. However, the fraction was always larger on the OLR not significant side. For example, with K=1.54, it was 0.28 for the nonsignificant OLR trials compared to 0.19 for the significant OLR trials (Table 7, i.e. 28% compared to 19%). These results mirror the actual data for PIC in Field 1 where 2 of the 20 nonsignificant OLR periods (10%) had significant intercepts compared to only 1 of the 22 significant OLR periods (5%) (Table 3).

As in the case with no additive constant (equation 5), sorting generally seemed to overestimate the true slope value of 1.0, more so than the other methods   In addition standard deviation for the slope from the sorted regression procedure was larger than any of the other methods, regardless of the significance of the OLR (Table 7, Figure 14). When the OLR was significant, the slope from the OLR provided reasonable estimate of the true slope and was stable across increasing the additive constant.

Increasing the additive constant increased the average slope for the force through origin method. When the initial OLR was significant, the forcing strategy increased the average slope estimates, up to a mean of 1.3 when K was at the 50th percentile. When the initial OLR was not significant, forcing through the origin resulted in estimates closest to the true value at the highest K amongst the four estimation procedures.

The differences between the methods in this PIC analysis with the additive constant were all generally within or comparable to the standard deviations (Figure 14).

Conclusions

While this work is preliminary, it presents a compelling perspective on the current procedures for back-calculation to determine flux. The [meanY/meanX] approach gave the best results in terms of accuracy and precision. The [meanY/meanX] method only showed a large bias when the additive intercept was set at the rather extreme value of the 50th percentile of the synthetic measured values. Over the range of conditions we looked at, the [meanY/meanX] method estimated the mean slope with the least bias and exhibited the lowest variance. Sorting generally performed poorest, both in terms of accuracy and precision. The ordinary least squares regression even in the significant regression cases, did not perform as well as the [meanY/meanX] in terms of bias and accuracy.

References

Ajwa, Husein, David Sullivan, Suduan Gao, Randy Segawa, Michael Stanghellini. 2011. Determine Emission Reduction Using Totally Impermeable Film and Waiting Period for Tarp Cutting in a Large Field Fumigation Trial.

Majewski, M.S. 1997. Error evaluation of methyl bromide aerodynamic flux measurements pages 135-153 IN James N. Seiber, James A. Knuteson, James E. Woodrow, N.Lee Wolfe, Marylynn V. Yates and S.R. Yates. Fumigants Environmental Fate, Exposure, and Analysis. ACS Symposium Series 652. American Chemical Society, Washington, D.C.

Ross, L.J., B. Johnson, K.D. Kim, and J. Hsu. 1996. Prediction of methyl bromide flux from area sources using the ISCST model. Journal of Environmental Quality 25(4):885-891.

Scire, Joseph S., David G. Strimaitis and Robert J. Yamartino. 2000. A user's guide for the CALPUFF dispersion model. Earth Tech, Inc. 196 Baker Avenue, Concord, MA 01742. January 2000

Tables

Table 1. Overview of study fields. Each field received an application of PicClor 60 (60% PIC, 40% 1,3D).

Field	Size (Acres)	Tarp Cut (days after application)	Total Monitoring Periods
1	8	16	42
2	2	10	30
3	2	5	20
4	2	5	20

Table 2. Comparison of (X,Y) and interval 11 1,3D (modeled, measured) means and Coefficients of Variation.

	Mean X (N = 1000)	**Modeled air conc (interval 11)**	**Mean Y (N = 1000)**	**Measured air conc (interval 11)**
Mean	3.87	3.93	22.7	21.2
CV	0.94	1.03	1.33	1.44

Table 3. Period breakdown for Field 1 chloropicrin regressions by significance of slope and intercept.

	Slope		
Intercept	Not Significant	Significant	Total
Not Significant	18	21	39
Significant	2	1	3
Total	20	22	42

Table 4. Effect of increasing standard deviation in the lognormal distribution (L) on fraction significant regressions, intercepts and within-period CVs for PIC calculations.

SD for L (lognormal)	Mean Fraction Significant Regressions	Mean Fraction Significant Intercepts	Average Within Period CV
1.2	0.74	0.02	1.41
1.25	0.72	0.02	1.43
1.35	0.68	0.02	1.47
1.5	0.63	0.02	1.52
1.7	0.58	0.02	1.58
1.9	0.54	0.02	1.64
2.2	0.49	0.02	1.72

Table 5. Comparison of back calculation methods based on PIC statistics and equation 5, using 16 X,Y samples per period. Each trial was comprised of 16 pairs of simulated values using equation 5. The average CV for the within period Y values was 1.54 and 1.55 for initially significant and initially non-significant OLR groups, respectively.

Initial OLR significant ($p<0.05$)					
	Statistic	$\frac{\overline{Y}}{\overline{X}}$	**OLR**	**Sorted OLR**	**Forced OLR**
	Trials	6186	6186	6186	6186
	Mean	1.00	1.22	1.58	1.10
	Median	0.88	0.92	1.26	0.90
	SD	0.56	1.22	1.35	0.84
	CV	0.56	0.99	0.85	0.77
Initial OLR not signficant ($p>0.05$)					
	Trials	3855	3855	3855	3855
	Mean	1.01	0.65	1.59	0.85
	Median	0.89	0.47	1.27	0.72
	SD	0.53	0.67	1.19	0.55
	CV	0.52	1.04	0.75	0.65

Table 6. Data used to estimate percentiles for simulated measured PIC concentrations (Y variable in equation 5)

	Initial OLR significant						Initial OLR not significant										
Percentile	m01	m02	m03	m04	m05	avg	m01	m02	m03	m04	m05	m06	m07	m08	avg	grand avg	Percentile
0%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0%
10%	0.03	0.03	0.03	0.03	0.02	0.03	0.07	0.06	0.06	0.06	0.04	0.07	0.06	0.06	0.06	0.04	10%
20%	0.23	0.23	0.24	0.25	0.22	0.23	0.33	0.33	0.30	0.31	0.31	0.30	0.31	0.35	0.32	0.28	20%
30%	0.52	0.52	0.53	0.53	0.49	0.52	0.67	0.65	0.61	0.63	0.61	0.62	0.64	0.69	0.64	0.58	30%
40%	0.90	0.87	0.92	0.92	0.88	0.90	1.11	1.13	1.01	1.02	1.03	1.02	1.09	1.15	1.07	0.98	40%
50%	1.43	1.45	1.47	1.48	1.44	1.45	1.66	1.69	1.54	1.57	1.63	1.59	1.62	1.75	1.63	1.54	50%
60%	2.20	2.18	2.29	2.29	2.23	2.24	2.48	2.43	2.38	2.36	2.38	2.39	2.44	2.68	2.44	2.34	60%
70%	3.55	3.44	3.54	3.57	3.47	3.51	3.68	3.67	3.57	3.38	3.58	3.77	3.46	3.89	3.63	3.57	70%
80%	5.87	5.79	5.79	5.91	5.86	5.84	5.92	5.91	5.67	5.17	5.55	5.94	5.47	6.32	5.74	5.79	80%
90%	10.53	10.48	10.83	10.79	10.87	10.70	10.98	10.79	9.99	9.86	10.23	11.27	10.34	11.52	10.62	10.66	90%
100%	152.5	186.6	157.8	265.8	260.9	204.7	216.2	168.4	354.9	258.8	158.9	316.7	382.8	252.9	263.7	234.2	100%

Table 7. Stochastic simulation results with additive constant, K, (equation 9) at three levels and results for significant initial OLR on the left and non-significant OLR on the right. Trials are the number of cases satisfying the criterion and used to obtain the statistics. Filtered values are cases which were not used. Highlighted cells are the mean slope estimate. The true value for the mean slope was 1.00. See text for further explanation.

	Initial OLR Significant							**Initial OLR Not Significant**						
				Slope estimators							**Slope estimators**			
K=0.05, 10th percentile of measured values	CV of 'measured' conc.	Reg. Sign. Flag	Int Sign. Flag	meany over meanx	force through origin	OLR	slope-sorted	CV of 'measured' conc.	Reg. Sign. Flag	Int Sign. Flag	meany over meanx	force through origin	OLR	slope-sorted
Trials	647	647	647	647	647	647	647	392	392	392	392	392	392	392
Mean	1.49	1.00	0.01	0.99	1.06	1.15	1.48	1.52	0.00	0.03	0.98	0.82	0.61	1.52
Median	1.40	1.00	0.00	0.87	0.88	0.86	1.18	1.41	0.00	0.00	0.89	0.72	0.47	1.26
Standard Deviation	0.44	0.00	0.10	0.49	0.67	0.90	1.05	0.48	0.00	0.17	0.46	0.47	0.55	1.06
Coeff. of Variability	0.29	0.00	9.57	0.49	0.63	0.79	0.71	0.31	---	5.89	0.46	0.57	0.91	0.70
Filtered Values	353	353	353	353	353	353	353	608	608	608	608	608	608	608
K=0.58, 30th percentile of measured values														
Trials	635	635	635	635	635	635	635	384	384	384	384	384	384	384
Mean	1.32	1.00	0.06	1.14	1.16	1.19	1.51	1.32	0.00	0.12	1.13	0.92	0.65	1.63
Median	1.24	1.00	0.00	1.01	0.99	0.93	1.24	1.21	0.00	0.00	1.01	0.80	0.46	1.33
Standard Deviation	0.41	0.00	0.23	0.54	0.72	0.97	1.19	0.47	0.00	0.32	0.56	0.56	0.66	1.33
Coeff. of Variability	0.31	0.00	4.08	0.47	0.63	0.81	0.79	0.36	---	2.75	0.49	0.61	1.02	0.81
Minimum	0.52	1.00	0.00	0.44	0.30	0.17	0.14	0.44	0.00	0.00	0.40	0.22	-0.05	0.28
Maximum	2.96	1.00	1.00	6.04	6.88	8.93	12.60	3.25	0.00	1.00	5.74	4.64	5.55	15.78
Filtered Values	365	365	365	365	365	365	365	616	616	616	616	616	616	616
K=1.54, 50th percentile of measured values														
Trials	619	619	619	619	619	619	619	375	375	375	375	375	375	375
Mean	1.11	1.00	0.19	1.39	1.30	1.22	1.66	1.15	0.00	0.28	1.40	1.08	0.67	1.65
Median	1.04	1.00	0.00	1.27	1.13	0.95	1.33	1.07	0.00	0.00	1.29	0.95	0.47	1.25
Standard Deviation	0.41	0.00	0.39	0.49	0.68	0.96	1.29	0.45	0.00	0.45	0.64	0.77	1.01	1.48
Coeff. of Variability	0.37	0.00	2.07	0.35	0.52	0.79	0.78	0.39	---	1.62	0.46	0.71	1.50	0.90
Minimum	0.41	1.00	0.00	0.64	0.41	0.19	0.28	0.48	0.00	0.00	0.54	0.33	-0.29	0.28
Maximum	3.19	1.00	1.00	5.41	8.07	11.02	13.06	3.49	0.00	1.00	9.33	12.56	16.33	21.49
Filtered Values	381	381	381	381	381	381	381	625	625	625	625	625	625	625

Figures

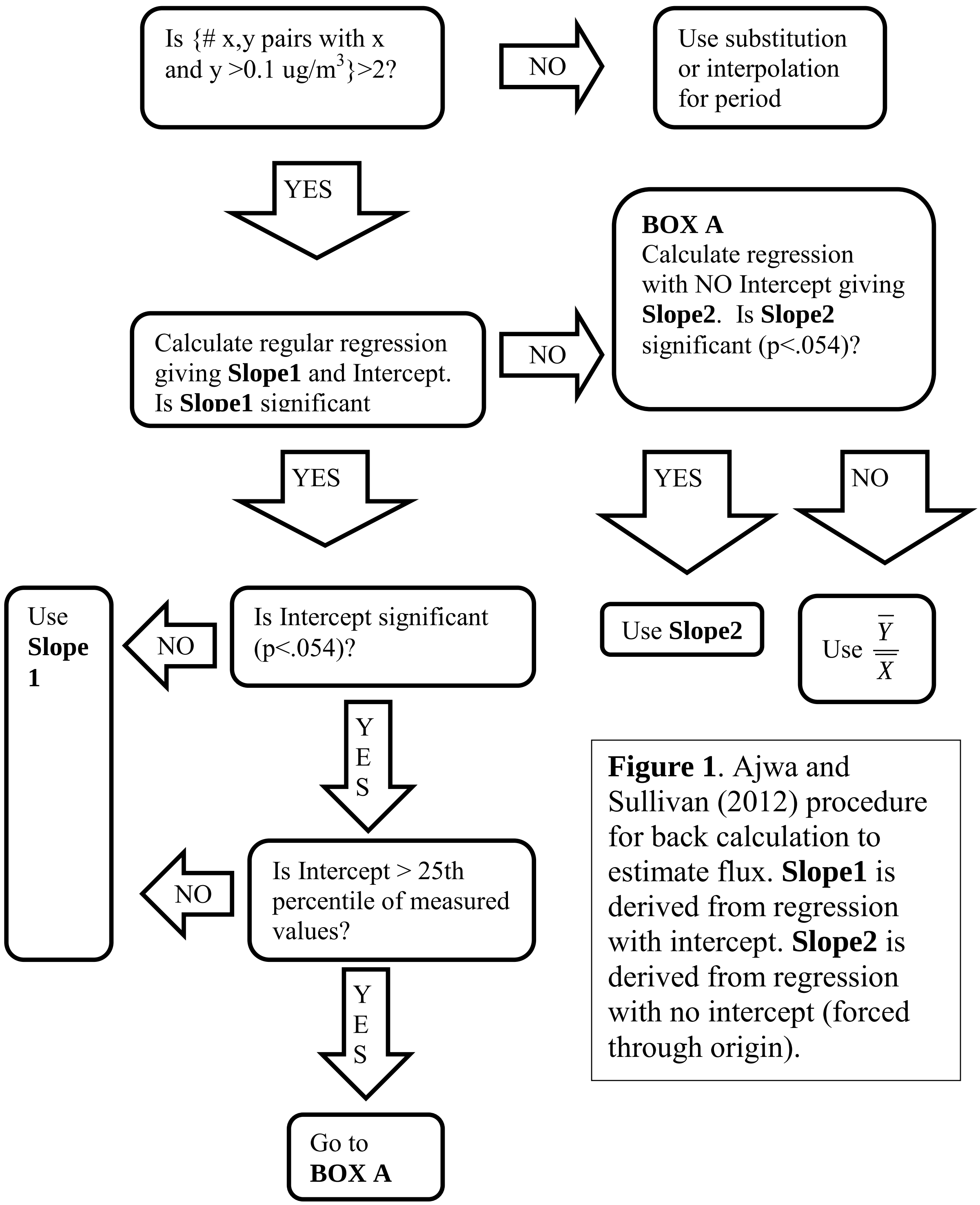


Figure 1. Ajwa and Sullivan (2012) procedure for back calculation to estimate flux. **Slope1** is derived from regression with intercept. **Slope2** is derived from regression with no intercept (forced through origin).

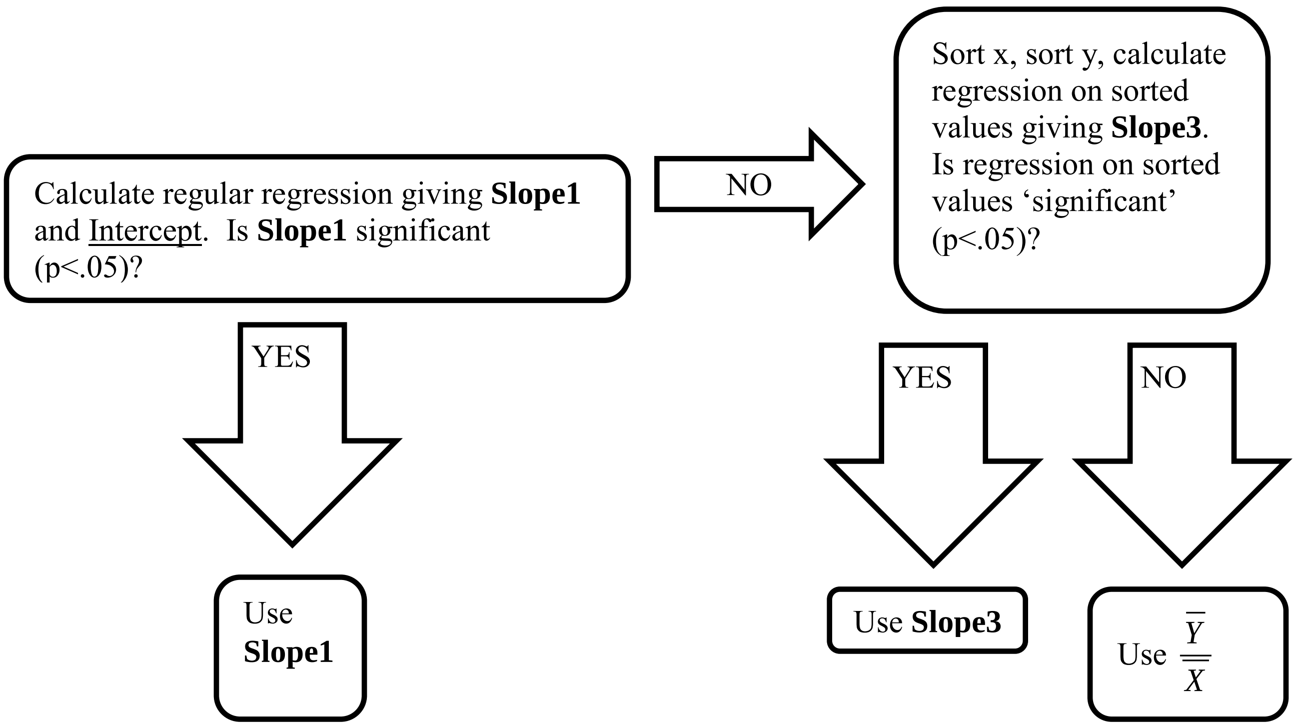
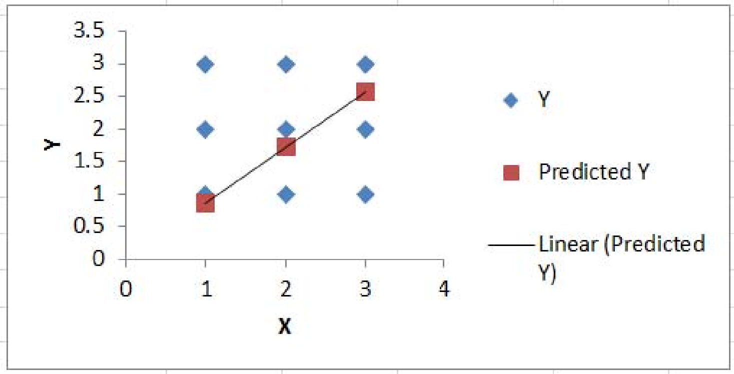


Figure 2. DPR procedure for back calculation to estimate flux. **Slope1** is derived from regression with intercept. **Slope3** is derived from regression of sorted y on sorted x values.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.857142857
R Square	**0.734693878**
Adjusted R Square	0.609693878
Standard Error	1.180193689
Observations	9

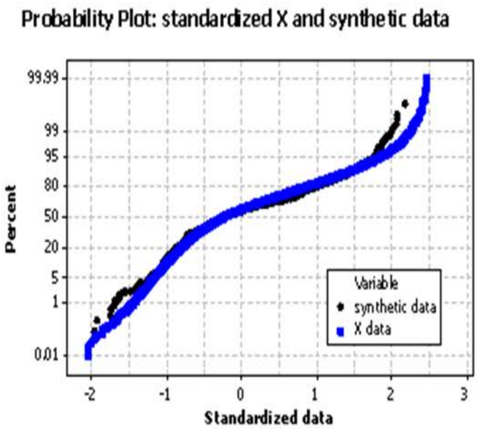


ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	30.85714286	30.85714	22.15385	**0.002191075**
Residual	8	11.14285714	1.392857		
Total	9	42			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A
X Variable 1	0.857142857	**0.18210784**	4.706787	0.001528	0.437201426	1.277084

Figure 3. Example EXCEL 2010 “forced origin regression” output for regression of [1,1; 1,2; 1,3; 2,1; 2,2; 2,3; 3,1; 3,2; 3,3]. OLR yields $R^2 = 0$, significance = 1, slope = 0 for these data.



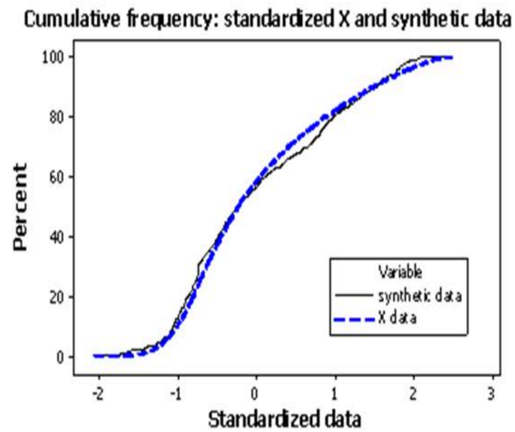


Figure 4. Probability and cumulative frequency plots of standardized "within interval" ISCST3 modeled air concentrations from Lost Hills Study and Crystal Ball generated *X* using a best fit beta distribution for *X*.

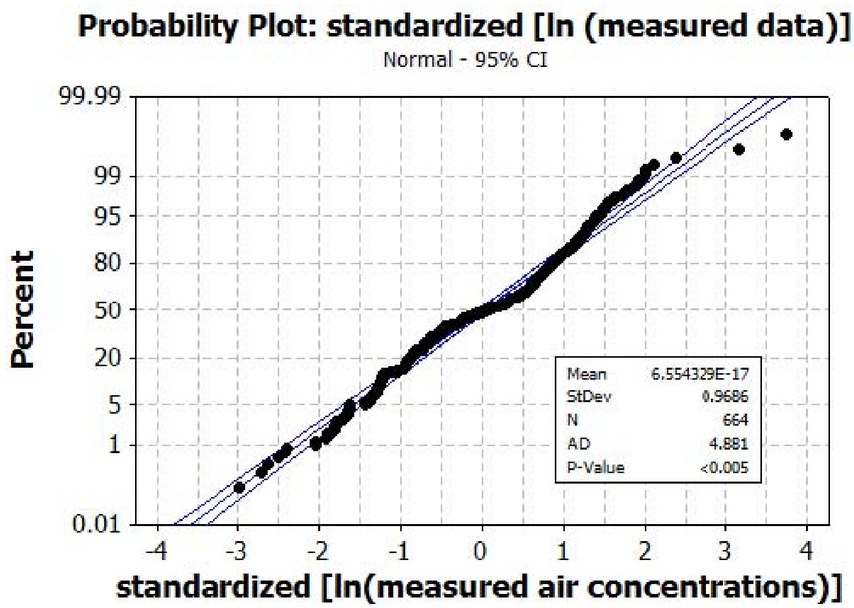


Figure 5. Normal probability plot of 1,3D standardized *ln* (measured data), calculated using individual within interval means (of *ln* measured) and standard deviations. The mean is zero.

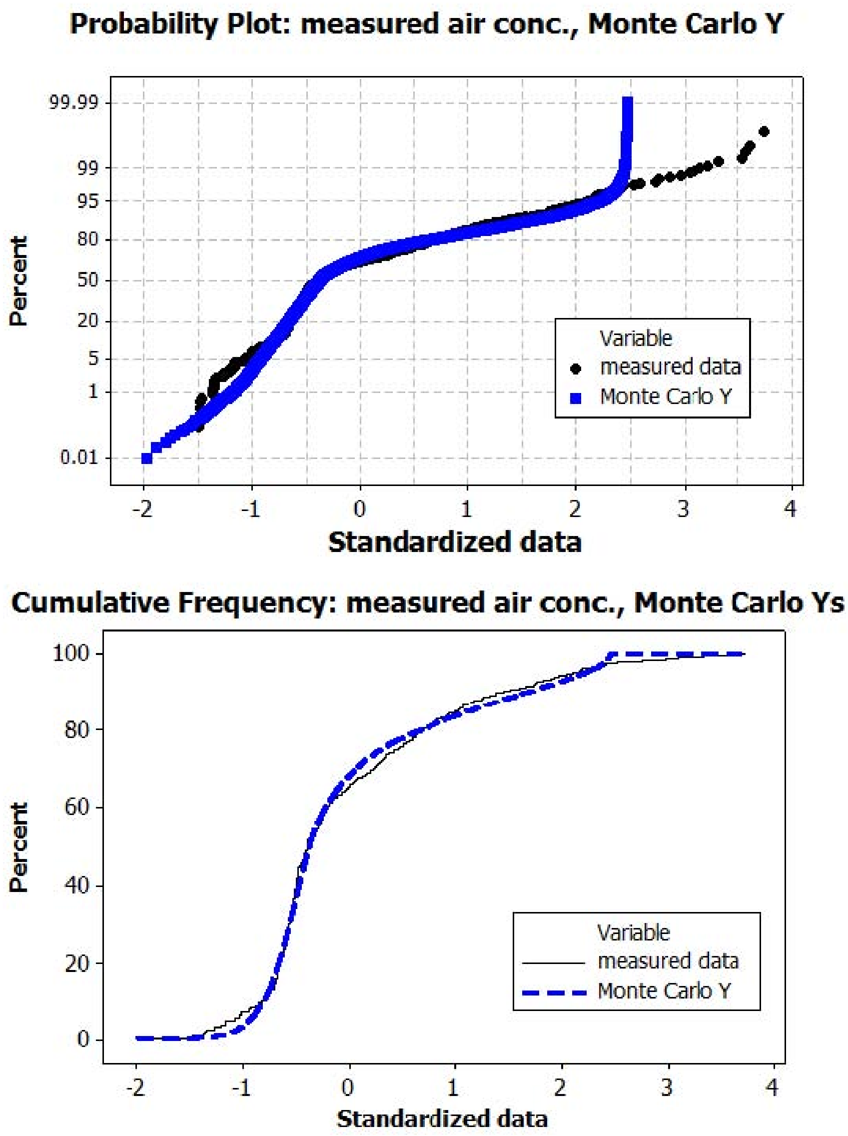
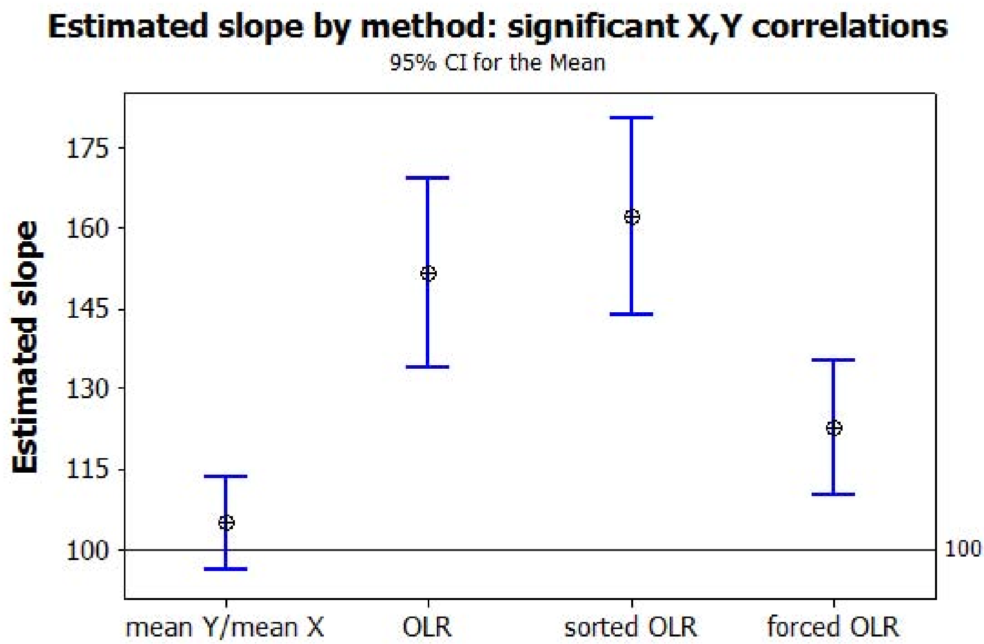


Figure 6. (a) Standardized measured air concentrations and Ys derived from Eq. 6 and (b) cumulative frequency of standardized measured concentrations and Y. Both Y and measured data standardized using within interval means and standard deviations.



Variable	Mean	StDev	CoefVar
mean Y/mean X	104.83	81.23	77.49
OLR	151.72	165.55	109.11
sorted OLR	162.29	168.73	103.96
forced OLR	122.57	116.52	95.07

Figure 7. Comparison of methods for estimating slope L using four methods. Estimates were generated for 334 of 1000 intervals that had significant regressions ($r > 0.707$).

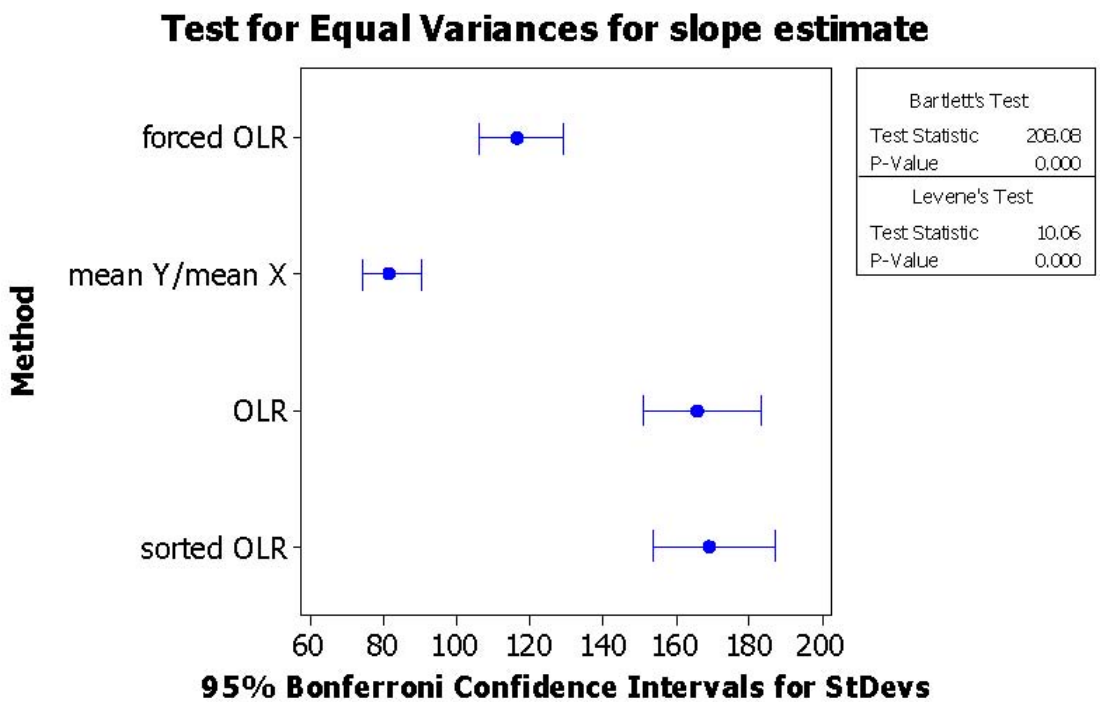
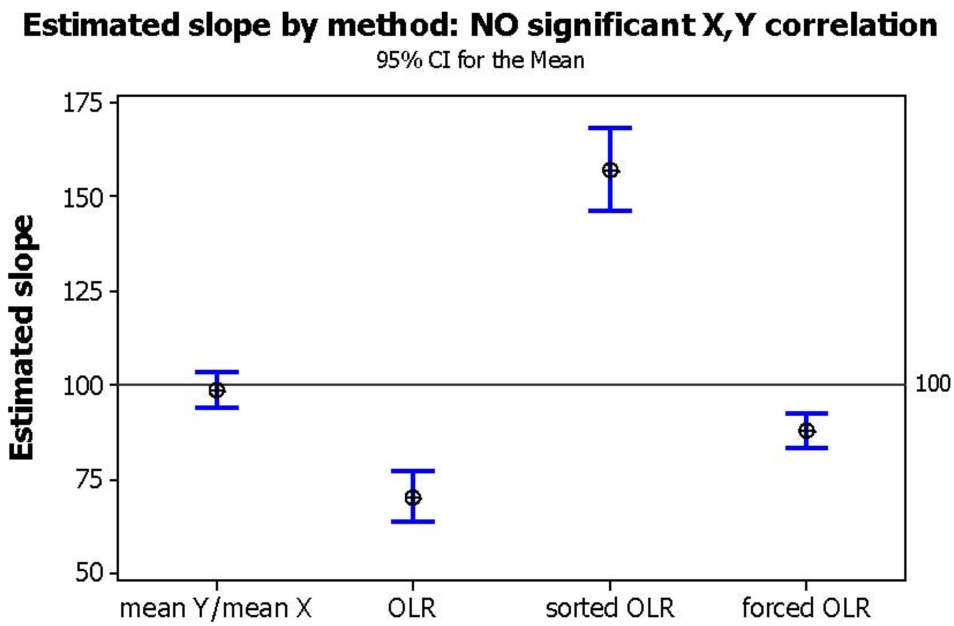


Figure 8. Test of equal variance of L estimates by method. L estimates were generated for 334 of 1000 intervals that had significant regressions (r > 0.707).



Variable	Mean	StDev	CoefVar
mean Y/mean X	98.52	62.17	63.10
OLR	70.01	89.43	127.75
sorted OLR	157.45	144.49	91.77
forced OLR	87.79	64.06	72.97

Figure 9. Comparison of methods for estimating L when OLRs were not significant. Reference line at Y = 100 = "true" mean of L.

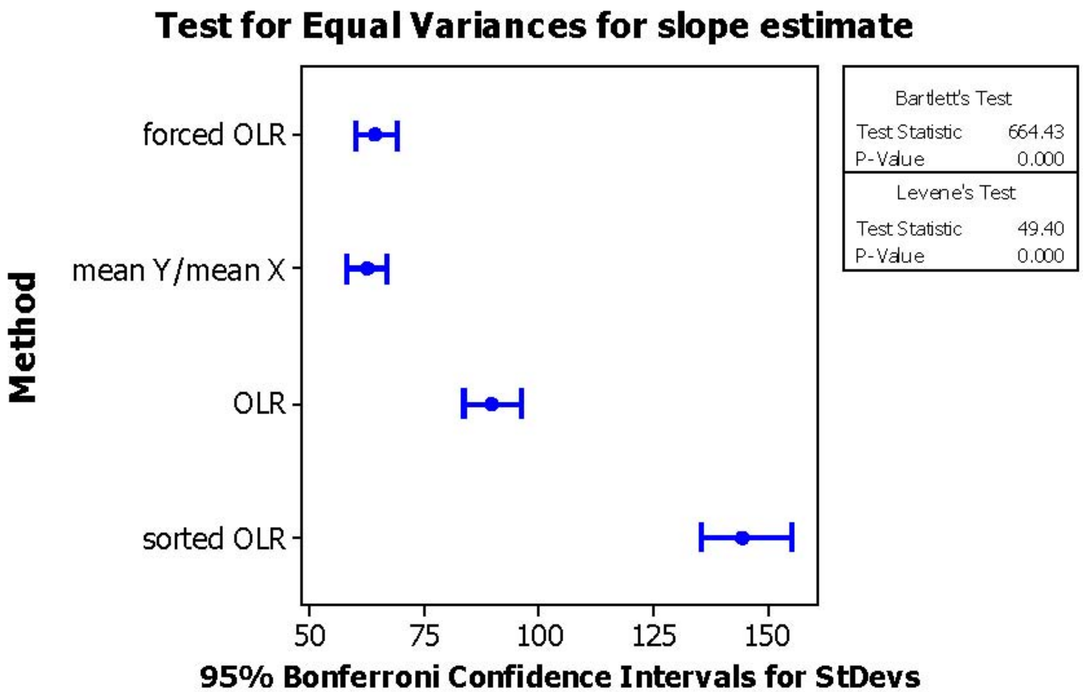
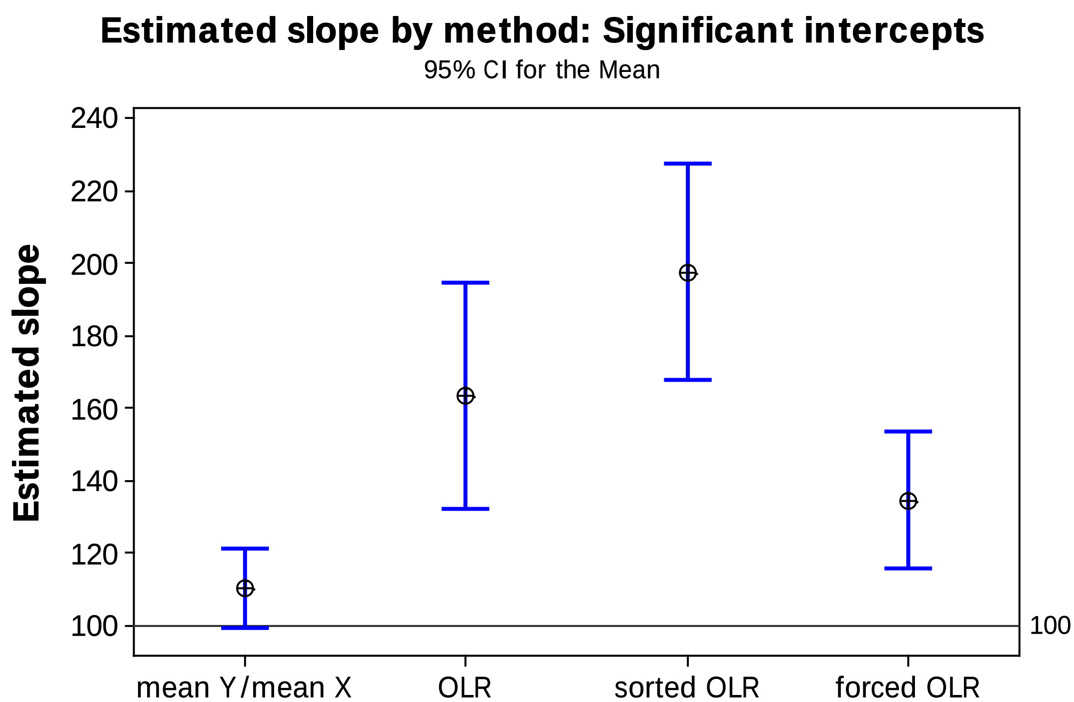
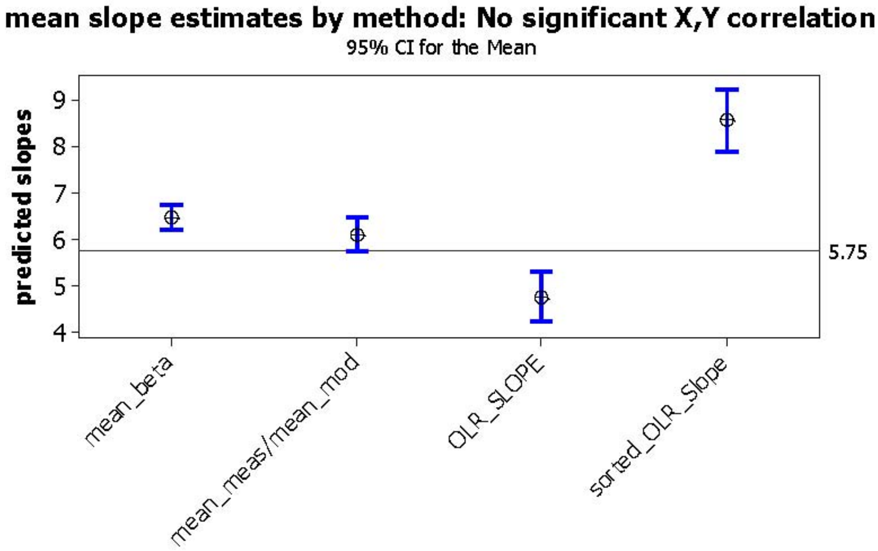


Figure 10. Test of equal variance of L estimates by method. L estimates were generated for 666 of 1000 intervals that had NO significant X,Y correlation (r < 0.707).



Variable	Mean	StDev	CoefVar
mean Y/mean X	110.25	106.89	96.95
OLR	163.2	301.7	184.84
sorted OLR	197.3	286.9	145.37
forced OLR	134.41	182.01	135.41

Figure 11. Comparison of methods for estimating L of those OLRs with significant slopes (includes both significant and insignificant regressions; N = 365). Data were generated using same distributions of X, L as previously described, but were selected from an expanded set of synthetic X,Y data (N = 15000) to obtain increased sample size. Reference line at Y = 100 = “true” mean of L.



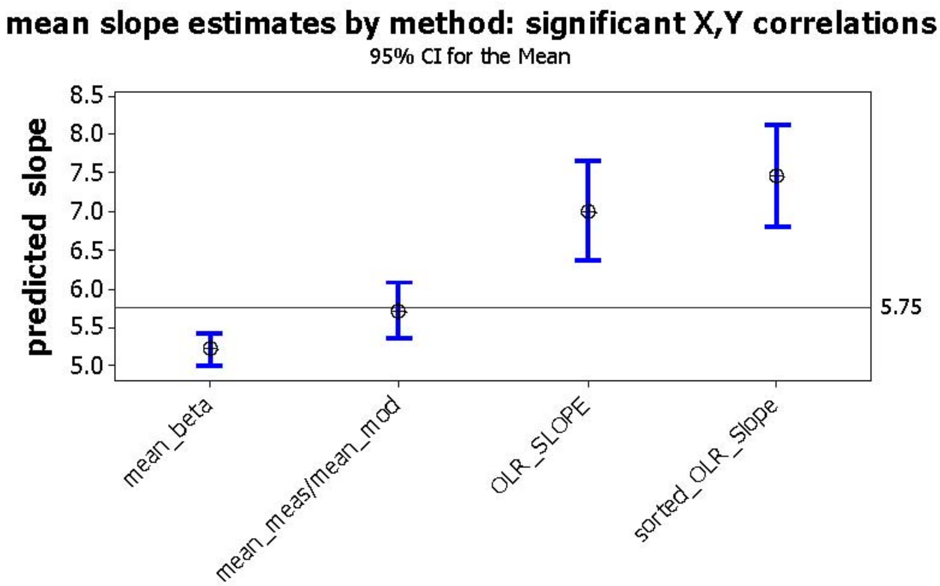


Figure 12. Comparison of slope estimation methods for synthetic data based on specific sampling interval 11 for 1,3D. For this scenario, the 5.75 reference line (y-axis) is the "true" slope (the true population mean of L); error bars are 95% confidence intervals for the mean estimate. Intervals with significant correlations ($r > 0.707$; $N = 465$), intervals with no significant correlation ($N = 535$). "mean_beta" is the mean of the individual trial slopes based on the samples drawn from the lognormal distribution for L in equation 5. The upper plot is the subset of nonsignificant correlations and the lower plot is the subset of significant correlations.

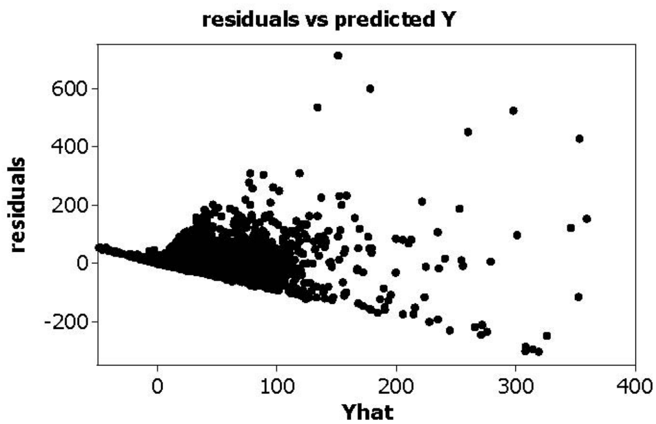


Figure 13. Residuals versus predicted Ys for the 1000 OLRs of the X,Y data used to compare error estimates by method. Data show heteroscadastic regression residuals.

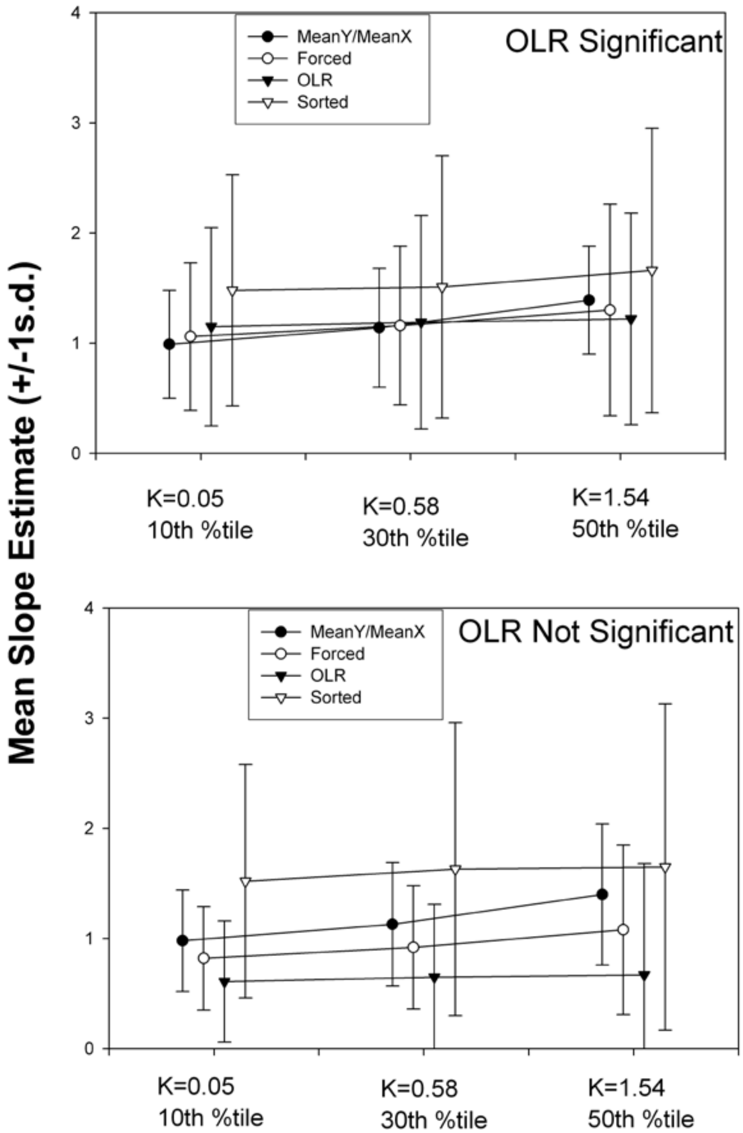


Figure 14. Comparison of [meanY/meanX], Forced through origin, OLR and sorted methods of flux estimation for PIC-based stochastic simulations when there is an additive constant in the basic equation (equation 9). The true slope was 1.0. The additive constant ranges from the 10th to the 50th percentile of the simulated measured values. Note x-axis values offset for clarity.