GUM measurement error pilot

Jérôme Lavoué

2024-06-19

## Introduction

This document summarizes a short simulation effort aiming at illustrating 2 approaches to deal with measurement error in the statistical evaluation of measurement data.

True exposure levels to airborne chemicals are often assumed to reasonably follow the lognormal distribution. When performing risk assessment, best practive recommend collecting 6-12 samples from the workplace to estimate metrics such as the 95th percentile or exceedance fraction of the OEL. However, the true exposure the measurements are subject to error potentially introduced during the sampling and analysis phases of the measurement process.

Sampling and analysis (S&A) errors are often assumed to be normally distributed with mean zero.

Two landmark documents have studied the impact of measurement error on the statistical evaluation of exposure data. [Grzebyk and sandino](https://www.inrs.fr/media.html?refINRS=ND%202231) in France and [Nicas, Simmons and Spear](https://www.tandfonline.com/doi/abs/10.1080/15298669191365199) in the US used slightly different approximations to demonstrate that S&A error below a CV of 30% has a negligible impact on the final estimate of the exposure metric for typical workplace GSD.

Both teams had to use simplifying assumptions to make their conclusions. In recent decades advances in computing power have let to the rapid spread of Bayesian methods in industrial hygiene and other fields. Bayesian methods are very flexible, and indeed can be easily used to incorporate S&A error in the interpretation of exposure levels.

An Bayesian S&A error model has been developed by our team at University of Montreal during the [Webexpo project](https://www.irsst.qc.ca/en/publications-tools/publication/i/101066/n/webexpo).

In parallel, the [supplement 1 to the ISO Guide to the expression of uncertainty in measurement (GUM)](https://www.iso.org/standard/50462.html) provides a framework to propagate measurement error in the final estimate of a statistical parameter using Monte Carlo simulation.

This document will compare the performance of the Bayesian model to the GUM model for industrial hygiene data interpretation in the context of a simple simulation study.

## Webexpo Bayesian S&A error model

The model is described in details in the scientific reports linked above.

In essence:

true\_concentration is what really happens in the workplace, it is assumed to be lognormal with unknown geometric mean (GM) and geometric standard deviation (GSD).

Mathematically :

* ln(true\_concentration) ~ Normal( ln(GM ) , ln(GSD) )

observed\_concentration is what comes out of the laboratory. For a particular value of true\_concentration, the observed\_concentration is assumed to be equal to the true value plus an error term, which is traditionally assumed to be normally distributed with mean 0 and a known coefficient of variation (CV).

Mathematically :

* observed\_concentration = true\_concentration + error
* error ~ Normal(0,CV)

As an example, for an expanded uncertainty of 50%, the CV would be ~25%.

During the analysis of a sample with this model, one would supply the observed values as well as the assumed CV, and the Bayesian engine would return posterior samples for all metrics of interest. The posterior samples represent the estimated uncertainty distribution of the parameter of interest. It is typical to use the median of the posterior sample as the point estimate, and, e.g., the 70% quantile of the posterior sample as a 70% upper confidence limit.

## GUM approach

This is the approach described by email by Robert Emond in an email from June 6th, 202 as understood (possibly misunderstood) by Jérôme Lavoué

The steps to estimate parameters and or upper limits is as follows:

* sample from a normal distribution around the observed values
* calculate a metric with the new dataset (GM, GSD, 95th percentile, UTL95,70)
* repeat a lot of times - result is a lot of values of the metric
* report the average value of the metric

For the practical implementation of the GUM approach, I used the frequentist estimation methods for the 95th percentile as presented in . To be able to obtain the 70% UCL I used tables in Zwillinger, D.; Kokoska, S. (2000) Standard Probability and Statistics Tables and Formulae. Chapman & Hall / CRC, Boca Raton, FL Pp 176 (section 7.3 Tolerance factors for the normal distribution)

## Example using both approaches

The vector of true concentrations is : 15, 120, 40, 9.2, 200, 23

It was generated from a lognormal distribution with 95th percentile=100 and GSD=2.5

The vector of observed concentration was created by adding normal random noise to the true concentrations. The observed concentrations would be what is accessible to the industrial hygienist. We selected a value for S&E error of 25%, which correspond to an expanded uncertainty of 50%.

The vector of observed concentrations is : 13, 160, 41, 3.9, 220, 29

Table 1 below shows the results of the determination of 4 metrics : gm, gsd, 95th percentile (p95) and the 70% upper confidence limit on the 95th percentile (P95\_ucl). These were calculated using both a Bayesian and Frequentist framework for three conditions :

* Ideal condition : the analysis was performed on the true concentrations, i.e, if there was no measurement error at all
* Naive condition : the analysis was performed on the observed concentrations, but without any provision for measurement error in the estimation process. This represents what is done by most tools today.
* ME condition : the analysis was performed on the observed concentrations, this time with provision for measurement error: The Bayesian measurement error model for the Bayesian analysis and the GUM approach for the Frequentist analysis.

Table 1 : Applying 2 measurement error approaches to one sample

| typeab | gm | gsd | p95 | p95\_ucl |
| --- | --- | --- | --- | --- |
| ideal\_b | 38.1 | 3.2 | 253.7 | 380.2 |
| ideal\_f | 38.2 | 3.3 | 272.8 | 440.8 |
| naive\_b | 35.6 | 4.1 | 355.9 | 574.1 |
| naive\_f | 35.6 | 4.6 | 432.1 | 794.4 |
| me\_b | 37.2 | 3.9 | 343.4 | 549.2 |
| me\_f | 34.5 | 4.7 | 442.7 | 830.8 |
| aIdeal : analysis performed on true concentrations with standard approach, Naive : analysis performed on observed concentrations with standard approach, me : analysis performed on observed concentrations with measurement error approach (Bayesian or GUM) | | | | |
| b\_b : Bayesian method, \_f : frequentist method | | | | |

Observations from T1 :

* GM seems little impacted
* The naive approach clearly shows an overestimation of environmental variability (expected as we observed environmental variability as well as S&E noise )
* For GSD, the frequentist approach considerably increase the estimate compared to naive approach (expected since normal noise is added to the observations in the Monte Carlo procedure)
* For GSD, the Bayesian approach decreases the estimate compared to naive approach (expected since the Bayesian model takes the S&E noise into account when trying to estimate the variability of the true concentrations)
* Overall the Bayesian ME approach seems to get results closer to the ideal results compared to the GUM approach, but not always much closer.

## Limited simulation study

The previous example illustrate the general behavior of both measurement error approaches, but doesn’t really provide useful information as to their performance across many analyses.

TO get a little bit more insight we performed a simulation study for 6 scenarios of sample size and true variability : GSD = 1.5 or 2.5 , n=3 or 6 or 9. We kept the true 95th percentile at 100. For each scenario we generated 1000 datasets of true and observed concentrations using an expanded S&E uncertainty CV of 50%. For each simulated dataset we calculated the same metrics as shown above.

To summarize the results across the 10 000 we calculated, for gm, gsd and P95 the bias, precision and rmse of the estimates. For the 70% UCL on the 95th percentile we calculated the coverage of that UCL i.e. the proportion of time the calculated UCL was indeed greater than the true P95 of 100.

The stability of the simulation was assessed by looking at the variability of the results across the 5 repeats of te entire simulation effort. To limit the amount of information presented, we only show the mean results across the 5 repetitions as the standard deviation across the 5 repetitions were very small (<1%). They are however available in the raw data.

Overall this took approximate 20h of computing time on a modern desktop computer.

### results

#### geometric mean

Table 2a and b below presents the results for the estimation of the geometric mean for GSDs of 2.5 and 1.5. All metrics are relative to the true value of the parameter.

Table 2a : Performance metrics for GM - GSD = 2.5

|  | Biasa | | | Precisiona | | | RMSEa | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | 15.0 | 7.1 | 4.7 | 65.3 | 41.7 | 32.8 | 67.0 | 42.3 | 33.2 |
| ideal\_B | 15.0 | 7.1 | 4.7 | 65.3 | 41.7 | 32.8 | 67.0 | 42.3 | 33.2 |
| naive\_F | 12.2 | 4.0 | 1.4 | 66.7 | 42.5 | 33.3 | 67.8 | 42.7 | 33.3 |
| naive\_B | 12.2 | 4.0 | 1.4 | 66.7 | 42.5 | 33.3 | 67.8 | 42.7 | 33.3 |
| me\_F | 9.6 | 0.9 | -1.8 | 65.1 | 41.2 | 32.2 | 65.8 | 41.2 | 32.3 |
| me\_B | 16.8 | 8.4 | 5.6 | 69.4 | 44.3 | 34.7 | 71.4 | 45.0 | 35.1 |
| aall metrics are relative to the true value of the parameter and expressed in % | | | | | | | | | |

Table 2b : Performance metrics for GM - GSD = 1.5

|  | Biasa | | | Precisiona | | | RMSEa | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | 3.1 | 1.4 | 0.9 | 24.5 | 16.9 | 13.8 | 24.7 | 17.0 | 13.8 |
| ideal\_B | 3.0 | 1.4 | 0.9 | 24.5 | 16.9 | 13.8 | 24.7 | 17.0 | 13.8 |
| naive\_F | 0.6 | -1.6 | -2.3 | 28.8 | 20.0 | 16.3 | 28.8 | 20.0 | 16.4 |
| naive\_B | 0.6 | -1.6 | -2.3 | 28.8 | 20.0 | 16.3 | 28.8 | 20.0 | 16.4 |
| me\_F | -1.8 | -4.5 | -5.4 | 28.2 | 19.4 | 15.7 | 28.2 | 19.9 | 16.6 |
| me\_B | 3.3 | 1.4 | 0.9 | 29.6 | 20.5 | 16.7 | 29.8 | 20.5 | 16.7 |
| aall metrics are relative to the true value of the parameter and expressed in % | | | | | | | | | |

Table 2a and 2.b reveal little influence of the studied methods and parameters on the estimation of GM. The maybe surprising bias values are influence by the skewed distribution of most estimated parameters in this simulation study. The idea is to compare the naive and ME approaches to the ideal approach.

Table 2c and d below presents the same results as above but using different performance metrics : median difference, median absolute deviation, and rmsle.

Table 2c : Alternative performance metrics for GM - GSD = 2.5

|  | Median errora | | | MAD | | | RMSLE | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | 0.04 | -0.08 | -0.03 | 8.80 | 5.89 | 4.76 | 0.50 | 0.36 | 0.29 |
| ideal\_B | 0.04 | -0.08 | -0.03 | 8.80 | 5.89 | 4.76 | 0.50 | 0.36 | 0.29 |
| naive\_F | -0.74 | -0.85 | -0.80 | 8.92 | 5.99 | 4.78 | 0.53 | 0.37 | 0.31 |
| naive\_B | -0.75 | -0.84 | -0.80 | 8.93 | 5.99 | 4.78 | 0.53 | 0.37 | 0.31 |
| me\_F | -1.25 | -1.48 | -1.48 | 8.71 | 5.81 | 4.63 | 0.53 | 0.38 | 0.31 |
| me\_B | 0.11 | 0.06 | 0.08 | 9.28 | 6.24 | 4.98 | 0.53 | 0.37 | 0.31 |
| aTrue value of GM is 22.15 | | | | | | | | | |

Table 2c : Alternative performance metrics for GM - GSD = 1.5

|  | Median errora | | | MAD | | | RMSLE | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | 0.16 | 0.02 | -0.02 | 8.34 | 5.80 | 4.70 | 0.23 | 0.16 | 0.13 |
| ideal\_B | 0.17 | 0.02 | -0.02 | 8.34 | 5.80 | 4.70 | 0.23 | 0.16 | 0.13 |
| naive\_F | -1.43 | -1.72 | -1.75 | 9.75 | 6.81 | 5.51 | 0.28 | 0.20 | 0.17 |
| naive\_B | -1.43 | -1.72 | -1.74 | 9.76 | 6.82 | 5.51 | 0.28 | 0.20 | 0.17 |
| me\_F | -2.62 | -3.20 | -3.33 | 9.53 | 6.61 | 5.34 | 0.29 | 0.21 | 0.18 |
| me\_B | -0.06 | -0.21 | -0.11 | 10.00 | 7.03 | 5.68 | 0.28 | 0.20 | 0.16 |
| aTrue value of GM is 51.3 | | | | | | | | | |

#### geometric standard deviation

Table 3a and b below presents the results for the estimation of the geometric standard deviation for true GSD = 2.5 and 1.5. All metrics are relative to the true value of the parameter.

Table 3a : Performance metrics for GM - GSD =2.5

|  | Biasa | | | Precisiona | | | RMSEa | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | -1.0 | -0.4 | -0.2 | 50.2 | 30.2 | 23.4 | 50.2 | 30.2 | 23.4 |
| ideal\_B | -1.3 | 0.8 | 1.0 | 28.7 | 24.2 | 20.6 | 28.7 | 24.2 | 20.6 |
| naive\_F | 4.0 | 4.2 | 4.3 | 57.3 | 33.9 | 26.2 | 57.5 | 34.2 | 26.6 |
| naive\_B | 1.4 | 4.4 | 4.9 | 31.1 | 26.5 | 22.7 | 31.2 | 26.9 | 23.2 |
| me\_F | 9.1 | 9.0 | 8.9 | 58.1 | 34.0 | 26.3 | 58.8 | 35.2 | 27.7 |
| me\_B | -2.5 | -0.4 | 0.1 | 29.0 | 25.8 | 22.3 | 29.1 | 25.8 | 22.3 |
| aall metrics are relative to the true value of the parameter and expressed in % | | | | | | | | | |

Table 3a : Performance metrics for GM - GSD =1.5

|  | Biasa | | | Precisiona | | | RMSEa | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | -2.6 | -1.2 | -0.7 | 19.5 | 12.6 | 10.1 | 19.6 | 12.6 | 10.1 |
| ideal\_B | 19.7 | 8.8 | 5.7 | 15.2 | 12.5 | 10.2 | 24.9 | 15.3 | 11.7 |
| naive\_F | 6.6 | 8.0 | 8.7 | 39.3 | 21.2 | 17.2 | 39.8 | 22.7 | 19.3 |
| naive\_B | 26.5 | 17.7 | 15.1 | 22.0 | 18.4 | 16.0 | 34.5 | 25.5 | 22.0 |
| me\_F | 15.4 | 16.8 | 17.5 | 37.2 | 20.1 | 16.3 | 40.3 | 26.2 | 23.9 |
| me\_B | 24.7 | 12.7 | 8.5 | 18.2 | 16.1 | 13.1 | 30.6 | 20.5 | 15.6 |
| aall metrics are relative to the true value of the parameter and expressed in % | | | | | | | | | |

Table 3c and d below presents the same results as above but using different performance metrics : median difference, median absolute deviation, and rmsle.

Table 3c : Alternative performance metrics for GM - GSD =2.5

|  | Median errora | | | MAD | | | RMSLE | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | -0.36 | -0.15 | -0.09 | 0.74 | 0.48 | 0.39 | 0.31 | 0.20 | 0.16 |
| ideal\_B | -0.18 | -0.07 | -0.04 | 0.47 | 0.40 | 0.34 | 0.19 | 0.16 | 0.14 |
| naive\_F | -0.29 | -0.06 | 0.01 | 0.81 | 0.53 | 0.42 | 0.32 | 0.21 | 0.17 |
| naive\_B | -0.13 | 0.01 | 0.05 | 0.50 | 0.43 | 0.37 | 0.20 | 0.18 | 0.15 |
| me\_F | -0.18 | 0.05 | 0.12 | 0.83 | 0.53 | 0.42 | 0.31 | 0.21 | 0.17 |
| me\_B | -0.25 | -0.12 | -0.07 | 0.48 | 0.42 | 0.36 | 0.19 | 0.18 | 0.15 |
| aTrue value is 2.5 | | | | | | | | | |

Table 3d : Alternative performance metrics for GM - GSD =1.5

|  | Median errora | | | MAD | | | RMSLE | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | -0.10 | -0.04 | -0.02 | 0.19 | 0.13 | 0.10 | 0.12 | 0.08 | 0.06 |
| ideal\_B | 0.25 | 0.11 | 0.07 | 0.16 | 0.13 | 0.10 | 0.13 | 0.09 | 0.07 |
| naive\_F | 0.00 | 0.08 | 0.10 | 0.26 | 0.17 | 0.14 | 0.15 | 0.11 | 0.10 |
| naive\_B | 0.34 | 0.23 | 0.20 | 0.21 | 0.16 | 0.14 | 0.18 | 0.13 | 0.12 |
| me\_F | 0.12 | 0.20 | 0.23 | 0.25 | 0.16 | 0.13 | 0.16 | 0.13 | 0.12 |
| me\_B | 0.30 | 0.15 | 0.10 | 0.16 | 0.15 | 0.13 | 0.16 | 0.11 | 0.09 |
| aTrue value is 1.5 | | | | | | | | | |

#### 95th percentile

Table 4a and b below presents the results for the estimation of the 95th percentile for true GSD = 2.5 and 1.5. All metrics are relative to the true value of the parameter.

Table 4a : Performance metrics for 95th percentile - GSD =2.5

|  | Biasa | | | Precisiona | | | RMSEa | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | 27.1 | 11.6 | 7.3 | 160.0 | 77.9 | 56.1 | 162.3 | 78.8 | 56.6 |
| ideal\_B | 13.4 | 10.3 | 7.9 | 91.4 | 65.0 | 51.3 | 92.4 | 65.8 | 51.9 |
| naive\_F | 34.7 | 16.8 | 11.9 | 178.6 | 86.3 | 61.7 | 181.9 | 88.0 | 62.9 |
| naive\_B | 15.5 | 13.4 | 11.2 | 96.8 | 69.7 | 55.3 | 98.0 | 71.0 | 56.4 |
| me\_F | 43.1 | 22.0 | 16.4 | 186.0 | 88.0 | 62.6 | 190.9 | 90.7 | 64.8 |
| me\_B | 12.0 | 8.8 | 6.8 | 93.3 | 67.9 | 54.1 | 94.0 | 68.5 | 54.5 |
| aall metrics are relative to the true value of the parameter and expressed in % | | | | | | | | | |

Table 4b : Performance metrics for 95th percentile - GSD =1.5

|  | Biasa | | | Precisiona | | | RMSEa | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | 0.7 | 0.2 | 0.4 | 43.2 | 27.3 | 21.9 | 43.2 | 27.3 | 21.9 |
| ideal\_B | 35.7 | 16.3 | 10.7 | 44.1 | 29.5 | 23.4 | 56.8 | 33.7 | 25.7 |
| naive\_F | 13.8 | 12.5 | 12.7 | 66.2 | 40.8 | 33.4 | 67.6 | 42.7 | 35.8 |
| naive\_B | 44.4 | 27.9 | 22.9 | 54.8 | 38.7 | 32.2 | 70.5 | 47.7 | 39.5 |
| me\_F | 26.5 | 24.0 | 23.8 | 67.1 | 40.9 | 33.3 | 72.2 | 47.4 | 40.9 |
| me\_B | 44.7 | 22.2 | 14.7 | 50.8 | 35.6 | 28.4 | 67.7 | 42.0 | 31.9 |
| aall metrics are relative to the true value of the parameter and expressed in % | | | | | | | | | |

Table 4c and d below presents the same results as above but using different performance metrics : median difference, median absolute deviation, and rmsle.

Table 4c : Alternative performance metrics for 95th percentile - GSD =2.5

|  | Median errora | | | MAD | | | RMSLE | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | -21 | -9 | -5 | 71 | 44 | 34 | 0.88 | 0.59 | 0.48 |
| ideal\_B | -12 | -5 | -3 | 50 | 39 | 32 | 0.68 | 0.53 | 0.44 |
| naive\_F | -18 | -6 | -2 | 78 | 48 | 37 | 0.91 | 0.62 | 0.50 |
| naive\_B | -12 | -4 | -1 | 53 | 41 | 34 | 0.71 | 0.55 | 0.46 |
| me\_F | -12 | -1 | 2 | 81 | 49 | 38 | 0.87 | 0.60 | 0.48 |
| me\_B | -14 | -8 | -5 | 50 | 40 | 34 | 0.70 | 0.56 | 0.47 |
| aTrue value is 100 | | | | | | | | | |

Table 4d : Alternative performance metrics for 95th percentile - GSD = 1.5

|  | Median errora | | | MAD | | | RMSLE | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | -9 | -4 | -2 | 26 | 18 | 14 | 0.39 | 0.26 | 0.21 |
| ideal\_B | 28 | 12 | 8 | 28 | 19 | 15 | 0.40 | 0.27 | 0.22 |
| naive\_F | -0 | 6 | 8 | 35 | 23 | 19 | 0.45 | 0.32 | 0.27 |
| naive\_B | 34 | 22 | 19 | 35 | 25 | 20 | 0.47 | 0.35 | 0.30 |
| me\_F | 12 | 18 | 19 | 35 | 24 | 19 | 0.44 | 0.33 | 0.30 |
| me\_B | 36 | 17 | 11 | 32 | 22 | 18 | 0.45 | 0.32 | 0.26 |
| aTrue value is 100 | | | | | | | | | |

#### 95th percentile UCL

Table 5 below presents the results for the coverage of the calculated 70%UCL : ideally it should be superioror equal to 70%: the 70% UCL should be above the true value at least 70% of the time.

Table 5 : Coverage the 95th percentile 70% UCL (%)

|  | GSD=2.5 | | | GSD=1.5 | | |
| --- | --- | --- | --- | --- | --- | --- |
| Approach | n=3 | n=6 | n=9 | n=3 | n=6 | n=9 |
| ideal\_F | 60 | 64 | 66 | 61 | 64 | 66 |
| ideal\_B | 66 | 67 | 68 | 97 | 87 | 84 |
| naive\_F | 61 | 66 | 68 | 68 | 76 | 79 |
| naive\_B | 66 | 69 | 70 | 96 | 91 | 90 |
| me\_F | 67 | 71 | 72 | 86 | 90 | 93 |
| me\_B | 65 | 66 | 67 | 98 | 91 | 87 |