Application of the GUM to purity verification

Adriaan M.H. van der Veen[1],
Annarita Baldan[1], Maurice G. Cox[2]
[1]Nederlands Meetinstituut
[2]National Physical Laboratory

Purpose of this lecture

- Proposals for statistically valid approaches to processing purity data
- Demonstration of the flexibility of the GUM

Applications of purity data

- Assessment of purity of 'pure' substances
- Calculation of the 'exact' composition of a calibrant
- Certification of (synthetic) reference materials

Experimental facts

- Purity verification of *o*-xylene

(mg/g)	Experimental	
	x	u
Benzene	***-0.016***	***0.012***
Toluene	***-0.010***	***0.010***
Ethylbenzene	0.021	0.003
p-Xylene	0.338	0.005
m-Xylene	2.801	0.045
Styrene	***-0.027***	***0.020***
n-Nonane	0.414	0.007
Cumene	0.775	0.010

Key requirements

- Model should be able to handle negative estimates of impurities
- Results of the model should be consistent
- Confidence intervals (of e.g. mass fractions) shall be within [0,1]
- Uncertainty budgets shall be ‘GUM-compliant’

Negative estimates

- Assuming the normal distribution, the value of a 'true blank' is expected to lie within

$$-2u_c < x_{obs} < 2u_c$$

with a probability of 95%

Consistency

- Prior to any kind of modelling the list of impurities should be compiled
- Model is as good as the list of impurities
- List of impurities shall be based on knowledge of synthesis/purification of substance of interest

Working with limits

- Observed mass fractions can be outside [0,1]
- Estimated mass fraction cannot be outside [0,1]
- Low levels of impurities hardy measurable $\Rightarrow$ LoD, LoQ
- Sum of all estimated mass fractions shall not exceed unity

Limit of detection

- Usually an arbitrary number, based on
  - Multiple of standard deviation of blank
  - Multiple of repeatability standard deviation
- As a consequence: any approach based on LoD provides an arbitrary estimate and uncertainty!

LoD approach

- Suppose, LoD is equal to *a*

$$x = \frac{1}{2}a$$

$$u(x) = \frac{1}{2\sqrt{3}}a$$

for a 'true blank', assuming the rectangular distribution

Alternative I

- Use of the ‘max’ function

$$x_i = \max\left(0, x_i^{obs}\right)$$

- Model satisfies definition of a pdf
- Drawback

$$p = \int_0^0 \max\left(0, x\right) dx > 0$$

Alternative II

- Use of Bayesian statistics
- Pdf based on product of a ‘vague prior’ and the Gaussian distribution
- Resulting pdf is product of Gaussian and rectangular distribution
- Posterior distribution is pdf of x_i
- Posterior satisfies definition of a pdf

Mean, standard deviation, and CI

- From a pdf, the following parameters can be computed
  - mean (‘best estimate’)
  - standard deviation (‘standard uncertainty’)
  - confidence interval

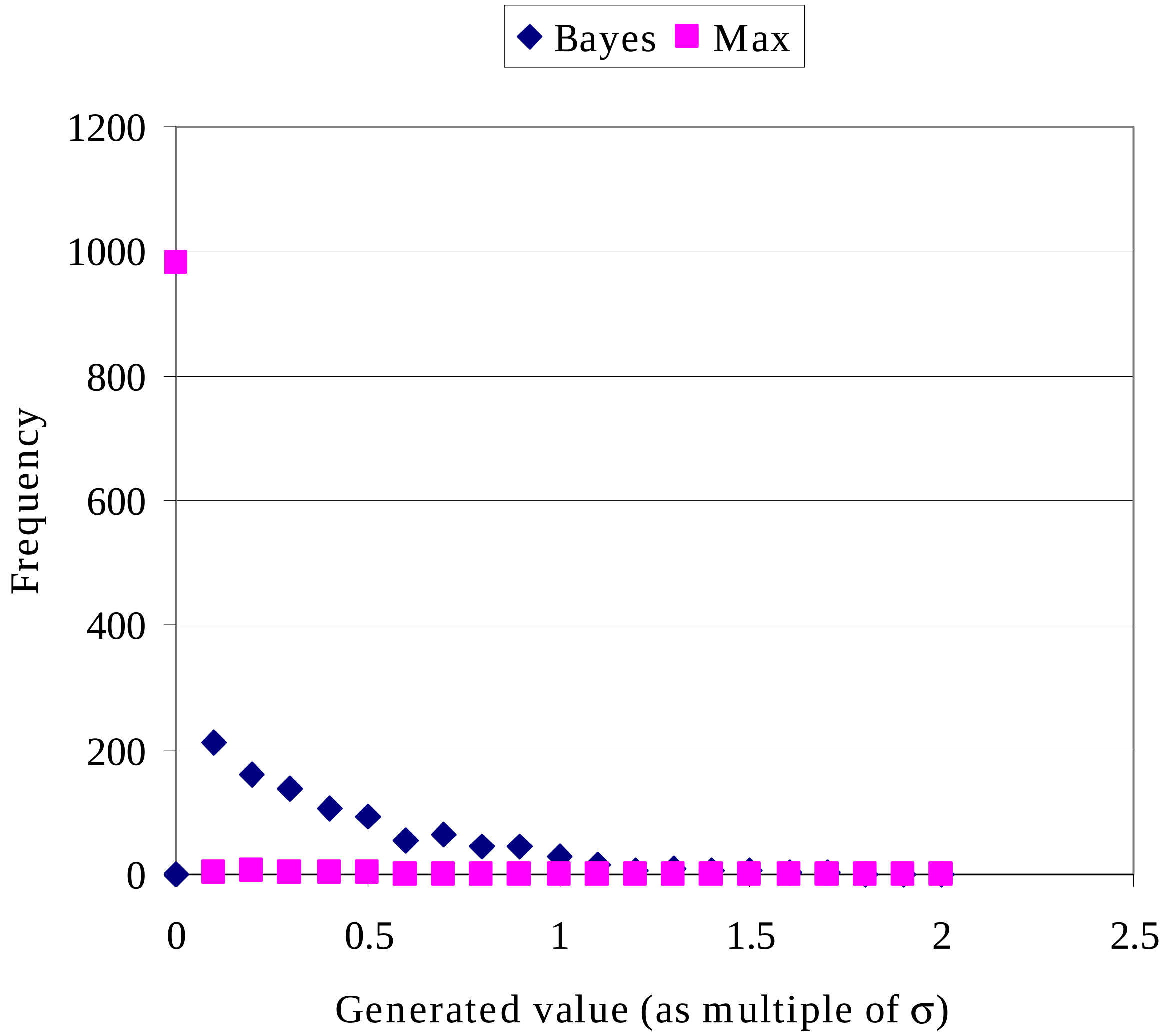
Confidence intervals

- As the pdf has only non-zero values within [0,1], CI cannot lie outside the interval
- As the CI is not symmetric, there is no
  - expanded uncertainty
  - coverage factor

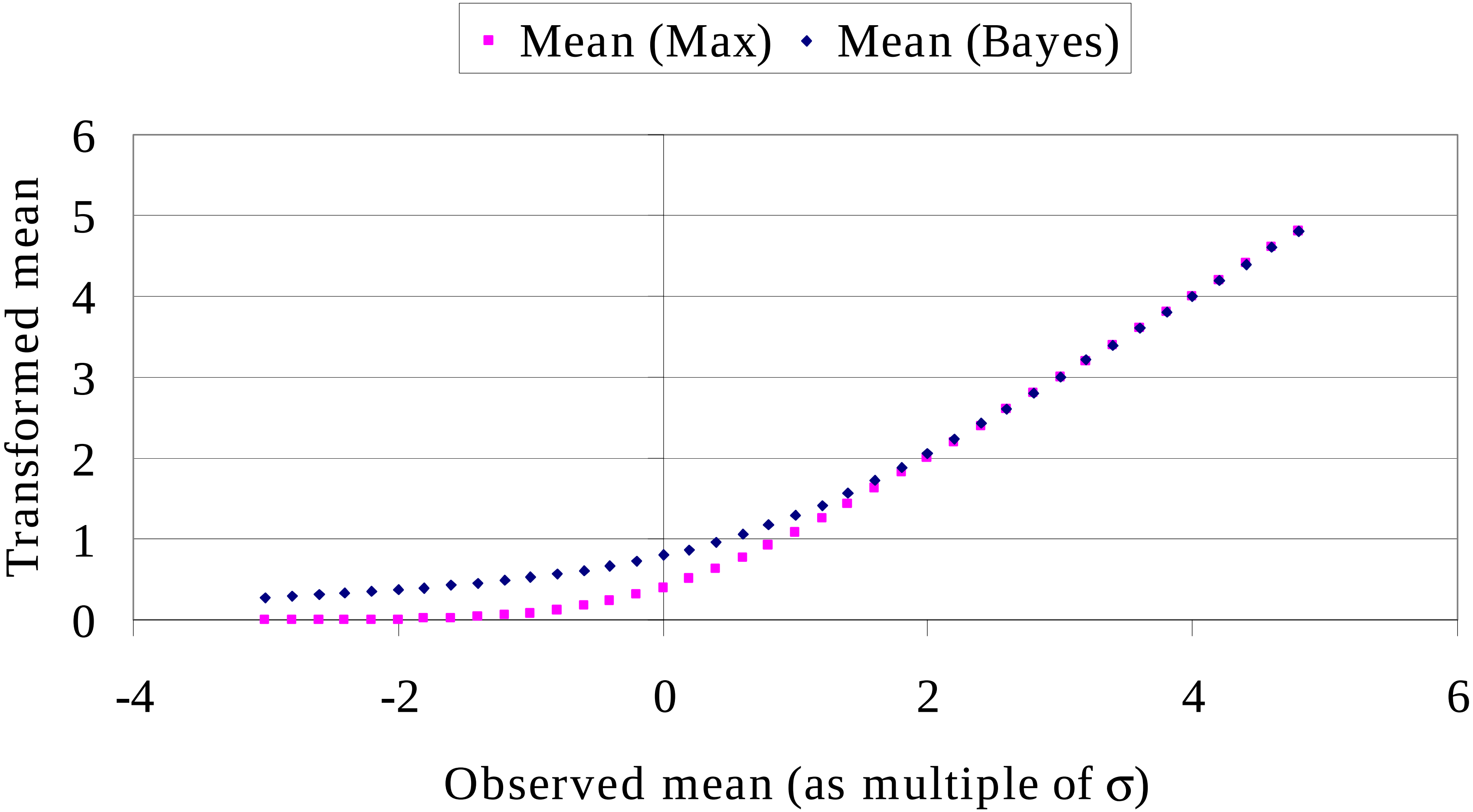
Experimental

- Impurities of o-xylene
- All impurities spiked
- Curves of spikes extrapolated to zero
- Uncertainty evaluated using common Gaussian statistics
- Results: x_i^{obs} and $u(x_i^{obs})$

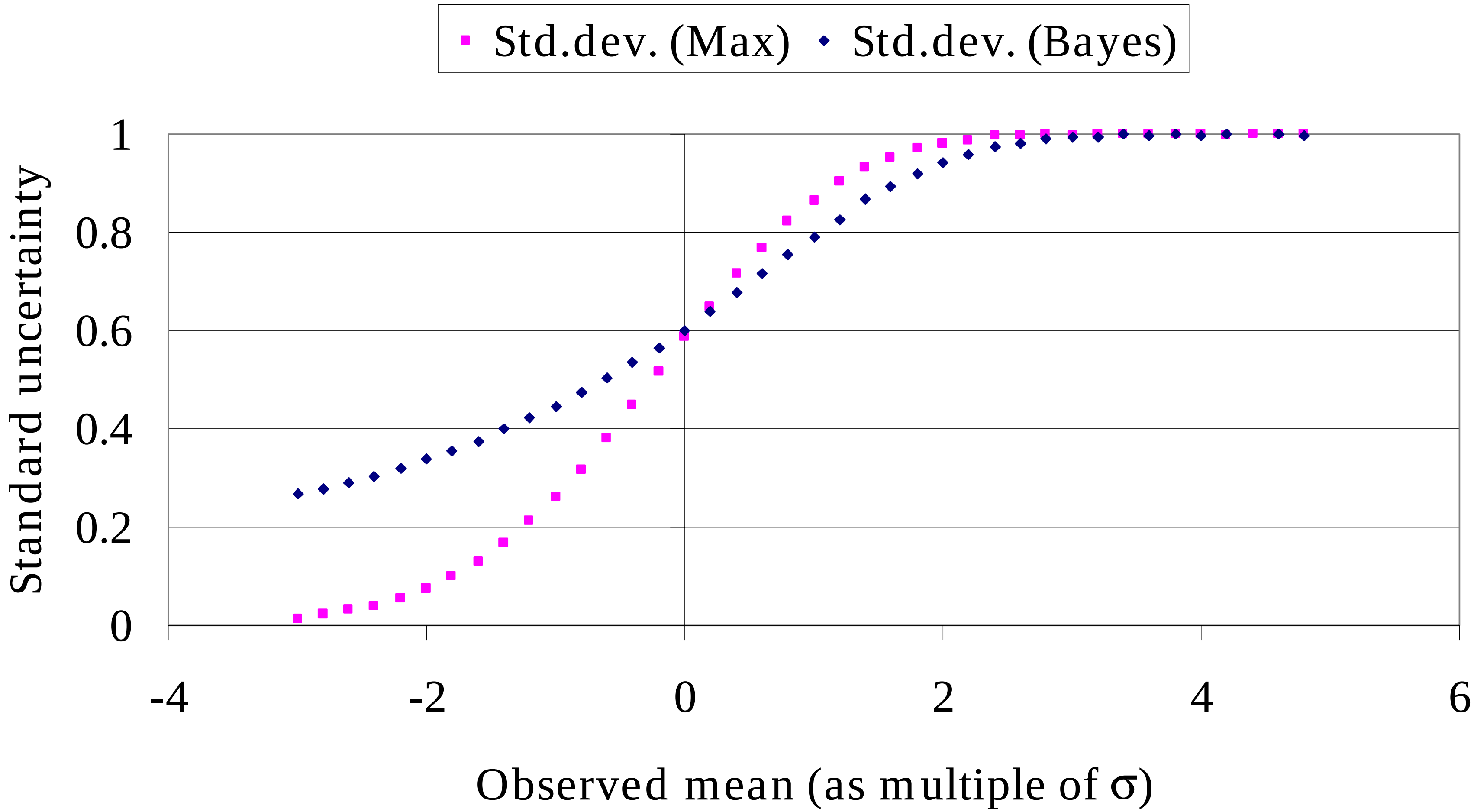
Output distributions (μ = -2σ)



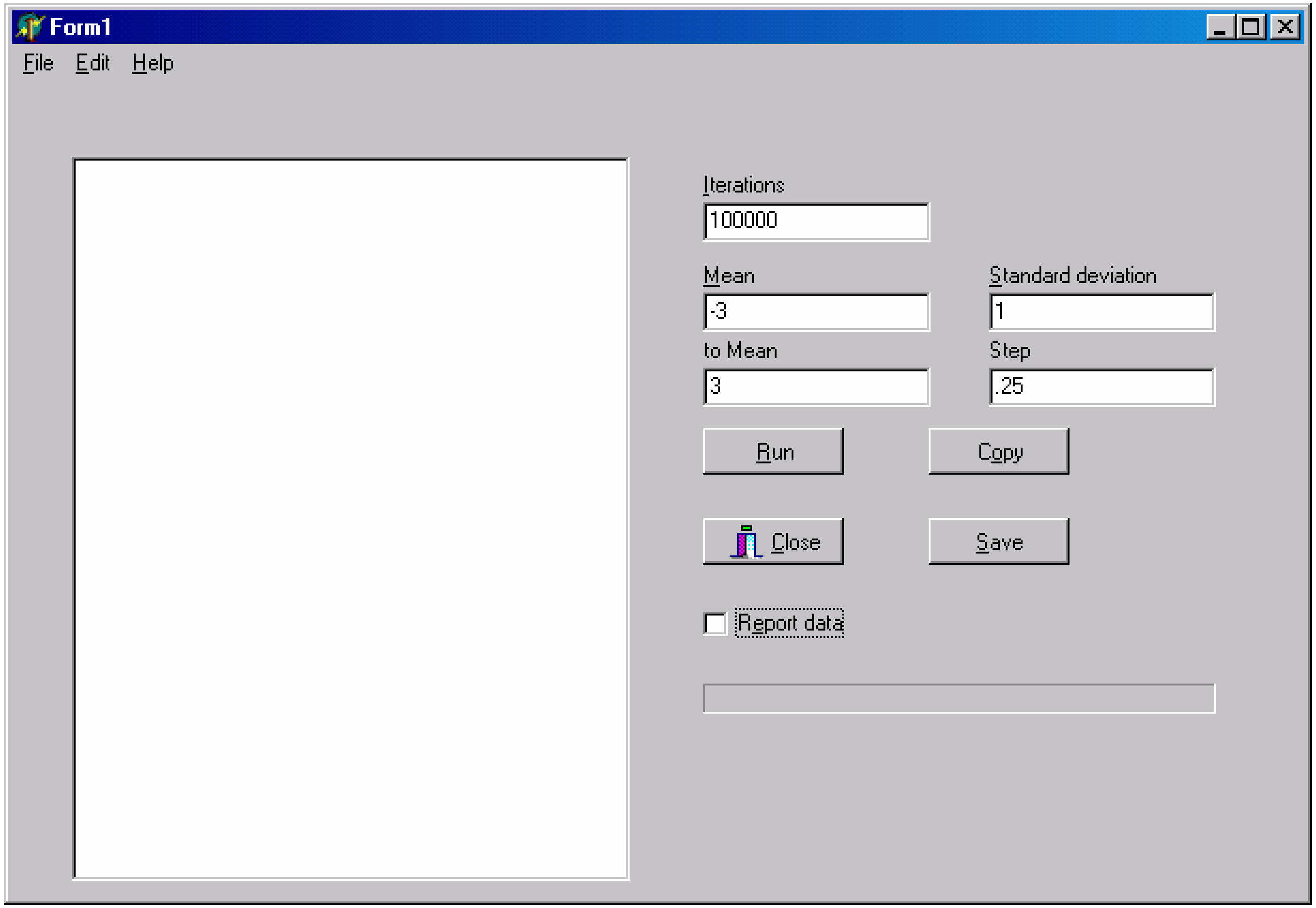
Max() versus Bayes (I)



Max() versus Bayes (II)



GUI


Form1
File Edit Help
Iterations
100000
Mean
-3
Standard deviation
1
to Mean
3
Step
.25
Run
Copy
Close
Save
Report data

Results

(mg/g)	Experimental		LOD		Bayes		Max	
	x	u	x	u	x	u	x	u
Benzene	***-0.016***	***0.012***	**0.046**	**0.027**	**0.006**	**0.005**	**0.0005**	**0.0022**
Toluene	***-0.010***	***0.010***	**0.009**	**0.005**	**0.005**	**0.004**	**0.0008**	**0.0026**
Ethylbenzene	0.021	0.003	0.021	0.003	0.021	0.003	0.021	0.003
p-Xylene	0.338	0.005	0.338	0.005	0.338	0.005	0.338	0.005
m-Xylene	2.801	0.045	2.801	0.045	2.801	0.045	2.801	0.045
Styrene	***-0.027***	***0.020***	**0.011**	**0.006**	**0.009**	**0.008**	**0.0008**	**0.0036**
n-Nonane	0.414	0.007	0.414	0.007	0.414	0.007	0.414	0.007
Cumene	0.775	0.010	0.775	0.010	0.775	0.010	0.775	0.010

Utilisation of results

- Two basic ways
  - propagation of distributions using pdf of x_i (e.g. by MCS)
  - propagation of uncertainties (a.k.a. 'mainstream GUM')
- For the user of the model, no specific knowledge of the interior of the model is required ('black box')

Concluding remarks

- New approaches have better properties than use of LoD or manufacturers’ specifications
- Critical evaluation of the alternatives is needed
- Uncertainty budgets should proof their validity in comparisons