Assessment and Expression of Measurement Uncertainty

Antonio Possolo

Statistical Engineering Division
Information Technology Laboratory

September 8, 2011

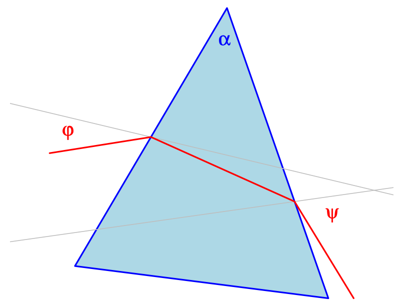
National Institute of Standards and Technology

Outline

- Refractive Index
  — Measurement & Measurement Uncertainty
- Arsenic in Oyster Tissue
  — Interlaboratory Study & Meta-Analysis
- Resistance and Reactance
  — Multivariate Measurand & Copulas
- Greenhouse Gases
  — Model Uncertainty & Cross-Validation
- Deepwater Horizon
  — Harmonizing Expert Opinion

Refractive Index

- Apex angle α, refractive index n, immersed in medium with refractive index m
- Light enters prism at angle φ, traverses prism's body, and exits at angle ψ
- As prism rotates about light's entrance point, $\delta = 2(\varphi + \psi - \alpha)$ decreases, reaches minimum δ_M, then increases



$$\frac{n}{m} = \frac{\sin((\alpha + \delta_M)/2)}{\sin(\alpha/2)}$$

Refractive Index

MEASUREMENT EQUATION

$$n = m\,\frac{\sin(\varphi + \psi - \alpha/2)}{\sin(\alpha/2)}$$

UNCERTAINTY ASSESSMENT

- Delta Method (GUM) — Analytic or numerical derivatives, variances and correlations of *input quantities*
- Monte Carlo Method (GUM S1) — Joint probability distribution of input quantities

Refractive Index — Estimate of Measurand (n)

$$n = m\,\frac{\sin(\varphi + \psi - \alpha/2)}{\sin(\alpha/2)}$$

$$= 1.0 \times \frac{\sin\left((48.6 + 30.0 - \frac{60.0}{2}) \times \frac{\pi}{180}\right)}{\sin\left(\frac{60.0}{2} \times \frac{\pi}{180}\right)}$$

$$\approx \mathbf{1.50}$$

Refractive Index — Delta Method ($u(n)$)

$$u(n) \approx \Big[\big(\dot{f}_m(m,\phi,\psi,\alpha)u(m)\big)^2 + \big(\dot{f}_\phi(m,\phi,\psi,\alpha)u(\phi)\big)^2 + \big(\dot{f}_\psi(m,\phi,\psi,\alpha)u(\psi)\big)^2 + \big(\dot{f}_\alpha(m,\phi,\psi,\alpha)u(\alpha)\big)^2\Big]^{\frac{1}{2}}$$

$$\approx \Big[(1.500222 \times 0.01)^2 + (1.322624 \times 0.486 \times \tfrac{\pi}{180})^2 + (1.322624 \times 0.300 \times \tfrac{\pi}{180})^2 + (-1.960542 \times 0.600 \times \tfrac{\pi}{180})^2\Big]^{\frac{1}{2}}$$

$$\approx \mathbf{0.029}$$

GUMMER for GUM

```
gummer = function (f, x, u,
                   r=diag(length(u)), ...)
{
  require(numDeriv)
  g = grad(f, x, ...)
  y = f(x, ...)
  uy = sqrt(matrix(g, nrow=1) %*%
            (outer(u, u, "*")*r) %*%
            matrix(g, ncol=1))
  return(c(y=y, "u(y)"=uy))
}
```


Refractive Index — R Function for n

- Function n computes prism's refractive index as function of m, φ, ψ, and α

```
n = function (x)
{
  m = x[1]
  phi = x[2]; psi = x[3]
  alpha = x[4]
  n = m * sin(phi + psi - alpha/2) /
      sin(alpha/2)
  return(n)
}
```

Refractive Index — Ingredients

```
m.mu     = 1
phi.mu   = 48.6 * (pi/180)
psi.mu   = 30.0 * (pi/180)
alpha.mu = 60.0 * (pi/180)

m.u      = 0.01
phi.u    = 0.486 * (pi/180)
psi.u    = 0.3   * (pi/180)
alpha.u  = 0.6   * (pi/180)
```


Refractive Index — GUMMER

```
gummer(f=n,
       x=c(m.mu, phi.mu, psi.mu, alpha.mu),
       u=c(m.u, phi.u, psi.u, alpha.u))
   y     u(y)
1.50    0.029
```

GUM Approximation — Shortcomings

- If some first order partial derivatives of f are zero at values of input quantities, then GUM's approximation is faulty
  - Radiant power $W = \kappa \cos(A)$
- GUM's approximation will be poor when f is markedly non-linear in neighborhood of values of input quantities
  - Need to study f's curvature is extra burden
  - If curvature is appreciable and influential, need higher-order approximation

GUM Approximation — More Shortcomings

- Expanded uncertainty $U = ku$

  Coverage factor k depends on generally unverifiable assumption that output quantity has Gaussian or Student's *t* distribution
- Even when $Y \sim t_{\nu}$, Welch-Satterthwaite formula often yields inappropriate value for ν

Example (GUM H.1 End-gauge calibration)

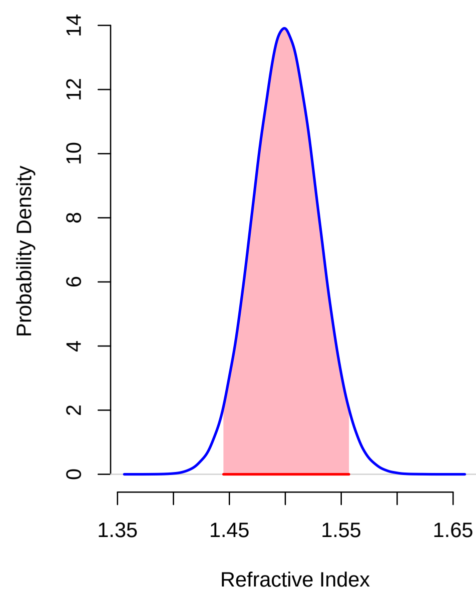
- $l = l_S + d - l_S(1 - \theta\,\delta\alpha - \alpha_S\,\delta\theta)$
- Welch-Satterthwaite formula yields $\nu_{\text{eff}} = 16$
- GUM's 99 % coverage interval **17 % longer** than it needs to be

GUM Supplement 1 — Refractive Index

```
K = 50000
m = rnorm(K, mean=m.mu, sd=m.u)
phi = rmv(K, mean=phi.mu, k=phi.kappa)
psi = rmv(K, mean=psi.mu, k=psi.kappa)
alpha = rmv(K, mean=alpha.mu, k=alpha.kappa)
n.values = apply(cbind(m,phi,psi,alpha), 1, n)
c(n=mean(n.values), "u(n)"=sd(n.values))
   n  u(n)
1.50 0.029
quantile(n.values, probs=c(0.025, 0.975))
2.5%    97.5%
1.44    1.56
```


GUM Supplement 1 — Refractive Index (CI)

- Kernel density estimate based on replicates $\{y_1, \ldots, y_K\}$ fully characterizes state of knowledge about prism's refractive index
- Shaded region includes 95 % of total area under curve: footprint is a 95 % *coverage interval* for measurand



Measurement & Measurement Uncertainty

References

GUM

Joint Committee for Guides in Metrology (2008)

Evaluation of measurement data — Guide to the expression of uncertainty in measurement

www.bipm.org/en/publications/guides/gum.html

VIM

Joint Committee for Guides in Metrology (2008)

International vocabulary of metrology — Basic and general concepts and associated terms

www.bipm.org/en/publications/guides/vim.html

Measurement

DEFINITION

- Experimental process that produces measurement result comprising
  - Measured value
  - Assessment of measurement uncertainty
- Measurand = quantity intended to be measured
- Measurement involves comparison of *measurand* against a reference value

Measurement — Input Quantities

All participating quantities that are required to assign a value to the measurand (*output quantity*)

- Experimental Data: Instrumental indications, readings, etc., that are informative about the measurand
  - Measuring temperature with thermocouple involves reading voltages
- Imported Input Values: Not estimated in the course of a particular measurement, have to be obtained elsewhere
  - Measuring length involves use of thermal expansion coefficients

Uncertainty — Meaning

MEANING

- Uncertainty is the condition of being *uncertain* (unsure, doubtful, not possessing complete or fully reliable knowledge)
  - Also a qualitative or quantitative expression of the degree or extent of such condition

*It is a subjective condition because it pertains to the perception or understanding that **you** have of the object of interest*

Uncertainty — Interpretation (GUM)

INTERPRETATION (GUM)

- *The uncertainty of the result of a measurement reflects the lack of exact knowledge of the value of the measurand* — GUM [3.3.1]
  - State of knowledge about measurand best described by a probability distribution over set of possible values
  - This probability distribution expresses how well one believes one knows the measurand's true value

Measurement Uncertainty — Definition

DEFINITION

- *Measurement uncertainty* is a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used — VIM 2.26
  - For scalar measurands, *non-negative parameter* typically chosen to be standard deviation of probability distribution describing that dispersion of values
  - For vectorial measurands, suitable multivariate counterpart

Measurement Uncertainty — Sources

- Definition of measurand — Aging, drifting
- Modeling — Mismatch between mathematical model and physical measurement situation, and plurality of alternative models that can reasonably be entertained
- Standards & calibration — Uncertainty of values of reference materials and instrument calibrations
- Uncontrolled environmental conditions
- Temporal drift of instruments and processes
- Differences between operators, laboratories, or measurement methods
- Data reduction methods that ultimately produce estimate of measurand

Measurement Uncertainty — Evaluations

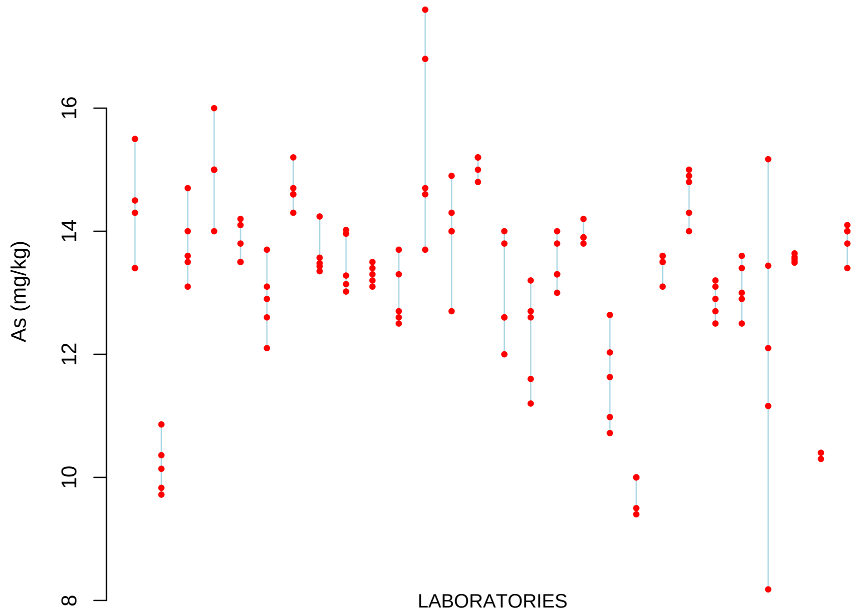
TYPE A

- Based on statistical scatter of measured values obtained under comparable measurement conditions (*repeatability*, *reproducibility*)

TYPE B

- Based on other evidence, including information
  - Published in compilations of quantity values
  - Obtained from a calibration certificate, or associated with certified reference material
  - Obtained from experts

Arsenic Interlaboratory Study
— Data & Uncertainty Components



- Within-lab variability
- Between-lab variability

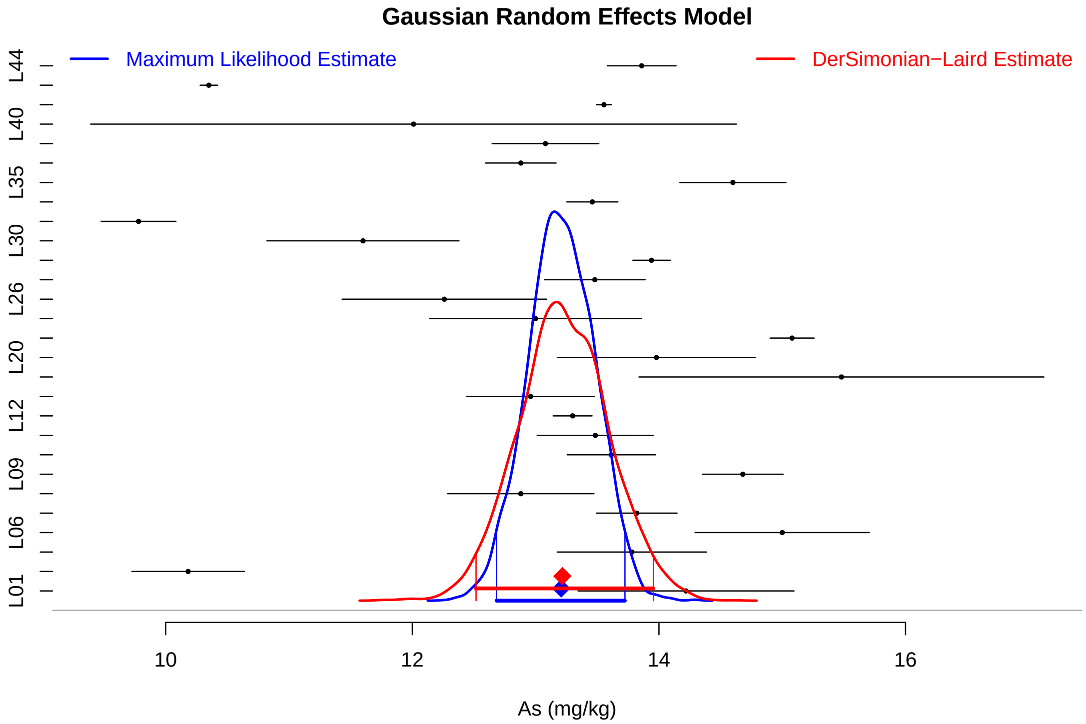
Arsenic Interlaboratory Study
— References & Model

- S. Willie & S. Berman (1995)
  - Intercomparison exercise for trace metals in marine sediments and biological tissues
  - NIST SRM 1566a Oyster tissue (14.0 ± 1.2mg/kg of arsenic): analysed by 28 laboratories, with 5 replicates per lab, but for one that produced 2 replicates only
- Heteroscedastic Gaussian mixed effects model
  Replicate i produced by laboratory j

$$y_{ij} = m + b_j + e_{ij}$$

Arsenic Interlaboratory Study

— MLE & DerSimonian-Laird Estimation



Arsenic Interlaboratory Study

Hierarchical Model with Adaptive Tail Heaviness (HMATH)

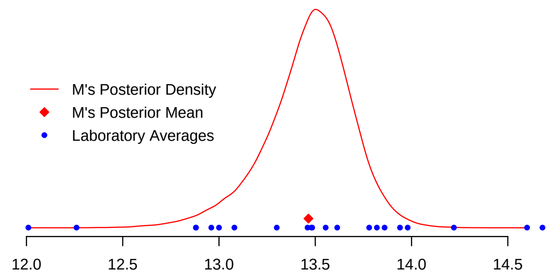
- $Y_{ij} = M + B_j + E_{ij}$
- $B_1, \ldots, B_n$ ~ Student's t_N with scale T
- $E_{1j}, \ldots, E_{m_j j} \sim \mathrm{GAU}(0, S_j^2)$
- M, T, $S_1, \ldots, S_n$, and N independent *a priori*
- M, $1/T^2$, $1/S_1^2, \ldots, 1/S_n^2$ have pretty flat prior Gaussian or gamma distributions
- $\Pr(N = n) \propto 1/n$ for $2 \leqslant n \leqslant 100$

Arsenic Interlaboratory Study

— HMATH Results

- N's posterior mean 11

	$\widehat{M}$	$U(\widehat{M})$
HMATH	13.44	0.49
SRM	*14.0*	*1.2*



- Incomplete decomposition of arsenobetaine in some labs

GUM Example H.2

— Simultaneous resistance and reactance measurement

INPUT QUANTITIES

- Amplitude V of sinusoidally-alternating potential difference across electrical circuit's terminals
- Amplitude I of alternating current
- Phase-shift angle ϕ of alternating potential difference relative to alternating current
- V, I, and ϕ are correlated

GUM Example H.2

— Simultaneous resistance and reactance measurement

OUTPUT QUANTITIES

- Resistance $R = \frac{V}{I} \cos \phi$
- Reactance $X = \frac{V}{I} \sin \phi$
- Impedance's magnitude $Z = \frac{V}{I}$

Joint distribution of (R, X, Z) concentrated on manifold $r^2 + x^2 - z^2 = 0$

GUM Example H.2: Problem & Solutions

— Simultaneous resistance and reactance measurement

PROBLEM

- Given marginal probability distributions for input quantities, and their correlations, manufacture *joint distribution* consistent with these

SOLUTIONS

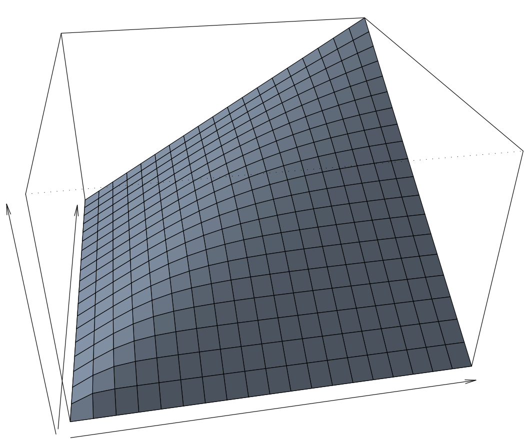
- There are infinitely many joint distributions consistent with given marginal distributions and correlations
- *Copulas* join univariate probability distributions into multivariate distributions and impose dependence structure

Copulas are not Cupolas

Manufacturer mentioned solely to acknowledge image source, with no implied recommendation or endorsement by NIST that cupola portrayed is the best available for any particular purpose

Copula — Definition

- A copula is the cumulative distribution function of a multivariate distribution on the unit hypercube all of whose margins are uniform



- Clayton copula inducing Kendall's $\tau = 0.6$

GUM Example H.2 — Data

V (V)	I (mA)	ϕ (rad)
5.007	19.663	1.045 6
4.994	19.639	1.043 8
5.005	19.640	1.046 8
4.990	19.685	1.042 8
4.999	19.678	1.043 3

GUM Example H.2 — Evaluations

- Input quantity values estimated by averages $\overline{V}$, $\overline{I}$, and $\overline{\phi}$, of sets of five observations
- Uncertainties and correlations of input quantities assessed by Type A evaluations
  - $u(\overline{V}) = \text{SD}(5.007, 4.994, 5.005, 4.990, 4.999)/\sqrt{5}$
  - Similarly for $u(\overline{I})$ and $u(\overline{\phi})$
  - $\text{cor}(V, I)$, $\text{cor}(V, \phi)$, and $\text{cor}(I, \phi)$ estimated by correlations between paired sets of five indications

GUM Example H.2 — GUM Supplement 1

CHALLENGES

- $u(V)$, $u(I)$, $u(\phi)$, $\mathrm{cor}(V, I)$, $\mathrm{cor}(V, \phi)$, and $\mathrm{cor}(I, \phi)$ estimated from very small numbers of observations
- Assign marginal distributions to V, I and ϕ, and link them using copula
- Reproduce correlations taking into account their uncertainty

GUM Example H.2 — GUM Supplement 1

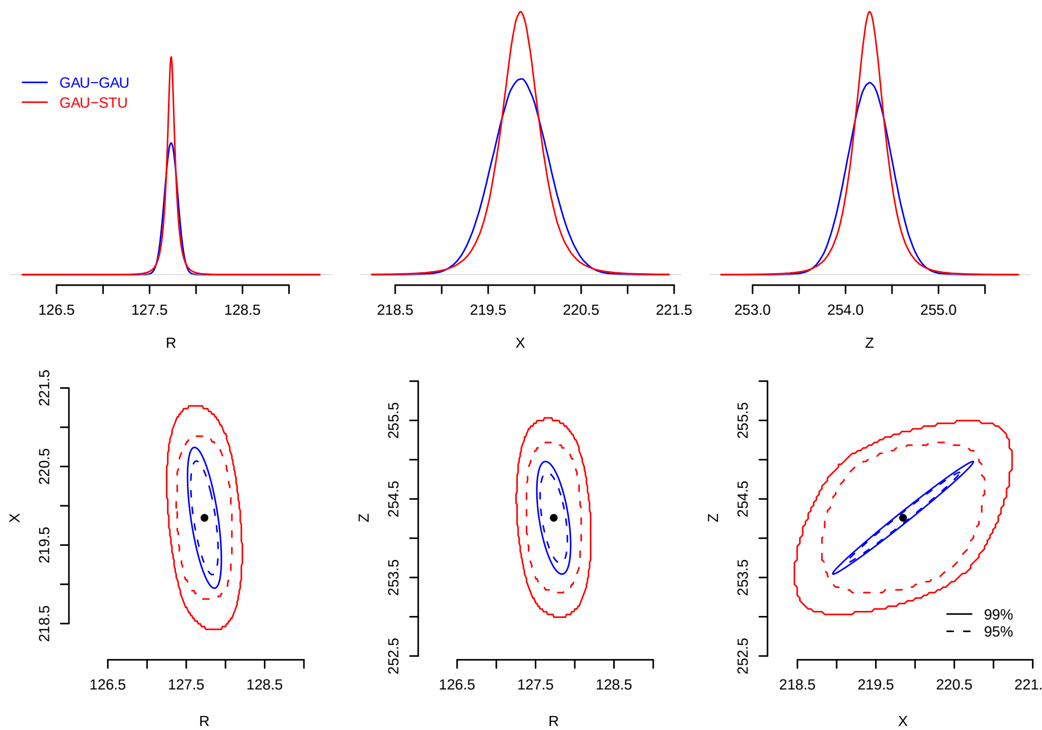
SOLUTION

- Employ Student t_4 distributions for

$$\frac{\overline{V} - \mu_V}{u(\overline{V})}, \ \frac{\overline{I} - \mu_I}{u(\overline{I})}, \text{ and } \frac{\overline{\phi} - \mu_\phi}{u(\overline{\phi})}$$

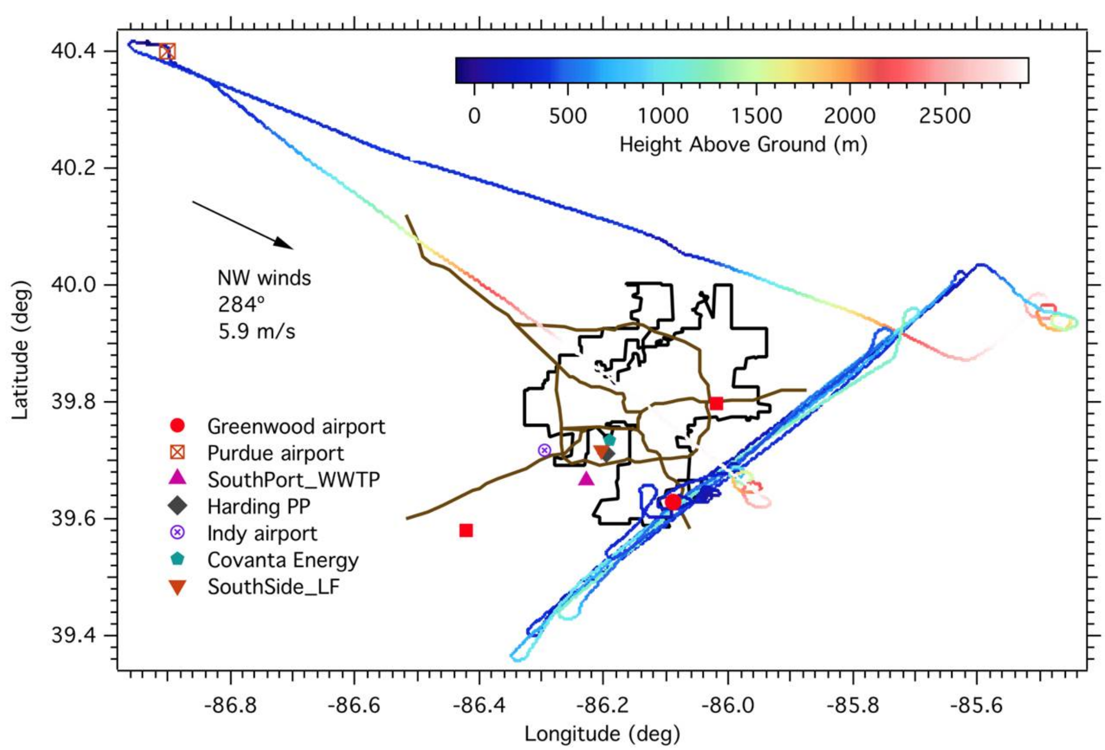
- *Tune* Gaussian copula using *correlation supplicant*
- Apply copula
  - To sample correlation matrix
  - To use sampled correlation matrix to produce sample from joint distribution of output quantities

GUM Example H.2 — Results



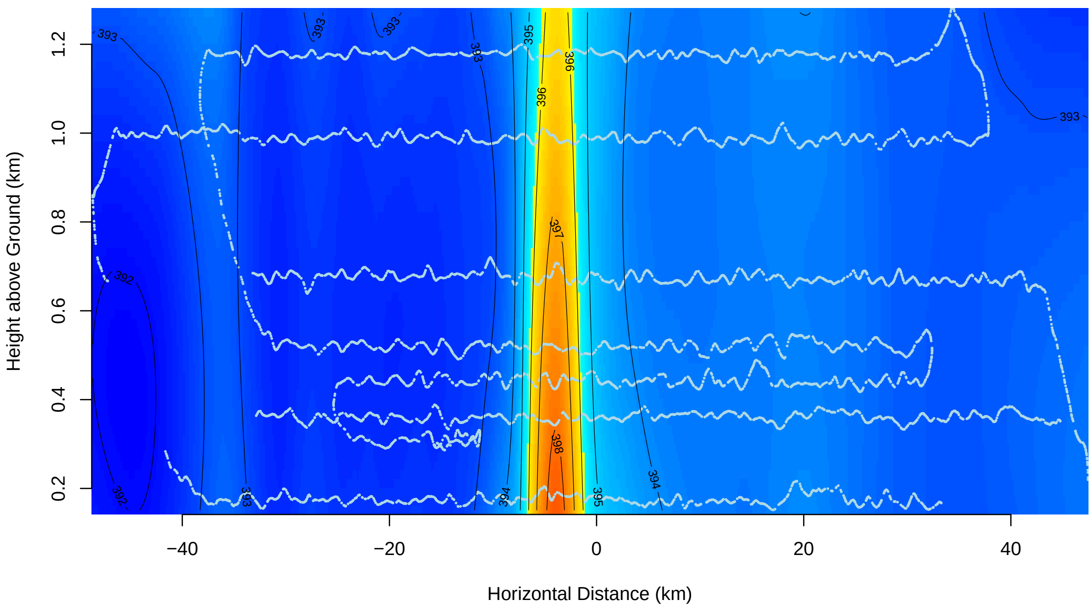
INFLUX Experiment (Indianapolis, IN)

— FLIGHT PATH



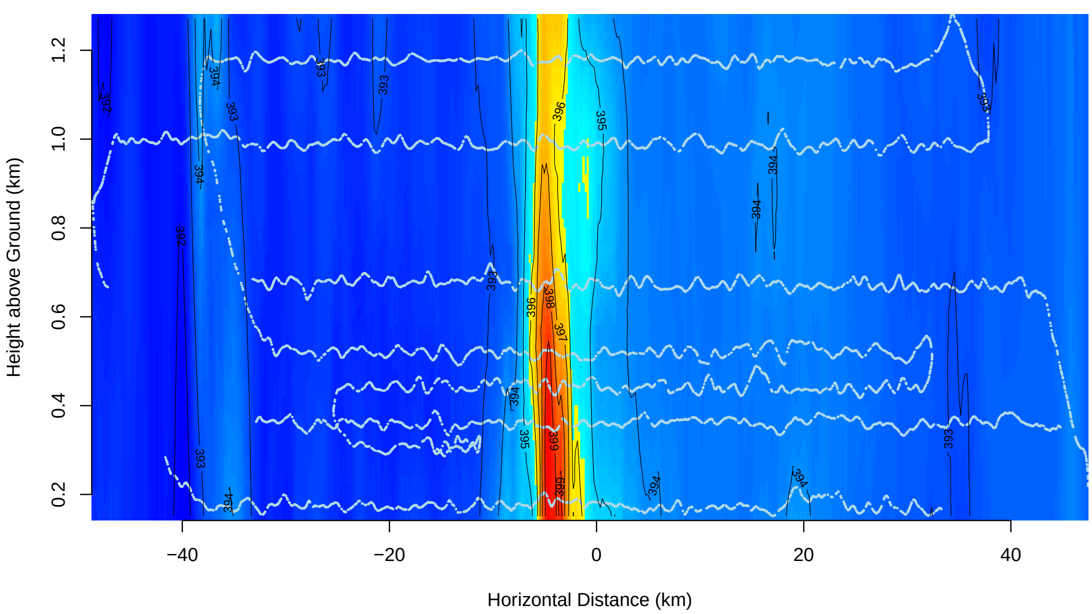
Local Regression Interpolation

— INFLUX EXPERIMENT: CO_2



Kriging Interpolation

— INFLUX EXPERIMENT: CO_2

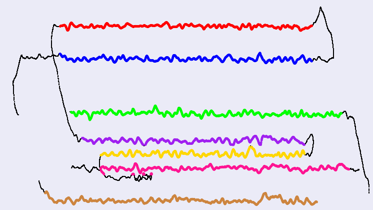


Cross-Validation & Model Uncertainty

— INFLUX EXPERIMENT: CO_2

CROSS-VALIDATION

- Partition data into training and testing subsets: fit models using former, assess performance on latter
- Partition may be random, or may include consideration for particulars of situation



MODEL UNCERTAINTY

- Compare predictions made by different models

Uncertainty Budget

— INFLUX EXPERIMENT: CO_2

SOURCE	EVALUATION	STD. UNCERT.
Model selection	CV	0.36
Interpolation	CV	0.91
Instr. calibration	LAB+CERT	0.034
Instr. repeatability	MANUF *	0.2
Instr. drift	MANUF *	0.2
Atmospheric temperature	MANUF *	0.0075
Atmospheric pressure	MANUF *	0.7
Expanded Uncertainty	$U_{95\%} = 2.5$ **ppmv**	

* Picarro G2301-m Flight

$$2.5 = 2\sqrt{0.36^2 + \cdots + 0.7^2}$$

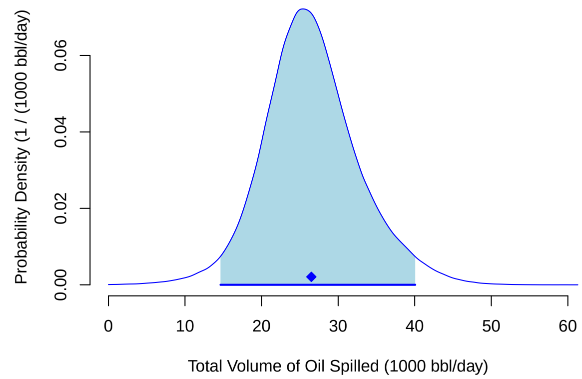
Deepwater Horizon Oil Spill

Plume Team — Seattle, WA (Jun 13, 2010)

Pooling Expert Assessments

PLUME TEAM — JUN 8, 2010

	low	high
A	20	30
B	20	34
C	20	30
D	20	30
E	25	40
F	15	34



- Linear Poll — Dennis Lindley (1983) Reconciliation of Probability Distributions

Summation

RECAP

Refractive Index	Measurement. . .
Arsenic in Oyster Tissue	Interlaboratory Study. . .
Resistance and Reactance	Multivariate Measurand. . .
Greenhouse Gases	Model Uncertainty. . .
Deepwater Horizon	Harmonizing. . .

SUGGESTIONS

- Probability and statistics well suited for the evaluation, production, and interpretation of uncertainty statements
- Increasingly complex measurands — in medical imaging, environmental remote sensing, meteorology, climatology, etc. — pose new challenges and call for methodological advances