Simple Guide
for the Evaluation and Expression of
the Uncertainty of NIST Measurement Results

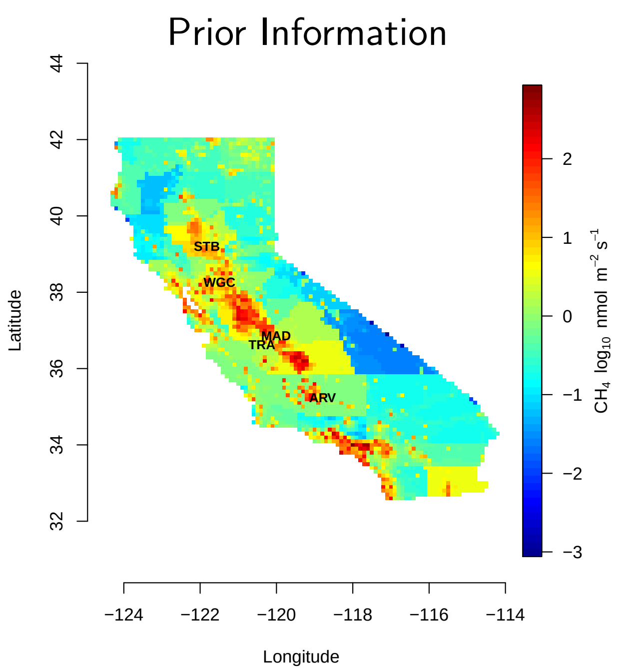
Antonio Possolo
NIST Fellow & Chief Statistician

Chair, Working Group on Statistics & Uncertainty
Sistema Interamericano de Metrología (SIM)

June 15, 2015

NIST
National Institute of Standards and Technology

California CH_4 Emissions



- Central Valley's five-site GHG network
- Bayesian inverse modeling framework
- Prior information from inventory of sources

Jeong *et al.* (2013) *JGR Atmospheres 11: 11339–11351*

Personalized Medicine

Treatment of uncertainty is **currently limited in scope***, due to the difficulties inherent in modeling different sources and types of uncertainty that often arise in personalized genomics-based analyses*

In cases of inherited genetic disease, even seemingly trivial **uncertainty calculations may make an important difference**

O'Rawe *et al.* (2015) *Trends in Genetics*

Simple Guide — Purposes

- Qualify all measured values with evaluations of measurement uncertainty that are *fit-for-purpose*
  - From atoms to galaxies, including the earth and its climate, and the health and nutrition of human populations
  - Expand application of best practices in uncertainty evaluation to areas of science and technology where they are still ignored

Simple Guide — Purposes

- Provide inclusive, comprehensive, and accessible guidance
  - Preserving gains and advances of past 20 years
  - Mindful of needs and capabilities of producers and consumers of measurement results worldwide
  - Progressively enriching toolkit for uncertainty evaluations to address emerging needs and challenges

Simple Guide — Structure

8	Sections	(10 pages)
30	Examples	(60 pages)

Weighing	Thermal Expansion Coefficient	Yeast Cells
Surface Temperature	Characteristic Strength of Alumina	Refractive Index
Falling Ball Viscometer	Voltage Reflection Coefficient	Ballistic Limit of Body Armor
Pitot Tube	Gas Analysis	Atomic Ionization Energy
DNA Sequencing	Sulfur Dioxide in Nitrogen	Forensic Glass Fragments
Thermistor Calibration	Thrombolysis	Gauge Blocks
Molecular weight of CO_2	Thermal Bath	Milk
Cadmium Calibration Standard	PCB in Sediment	Load Cell Calibration
Microwave Step Attenuator	Newtonian Constant of Gravitation	Atmospheric CO_2
Lead Standard Solution	Copper in Wholemeal Flour	Colorado Uranium

Simple Guide — Orientation

- Widen concept of *measurement*

  Informative assignment of value to **quantitative** *or* **qualitative** *properties*

- Adopt pervasive *probabilistic framework*

  Property values that are **imperfectly known** *modeled as random variables whose* **probability distributions** *describe states of knowledge about their true values*

Simple Guide — Procedure

- **Measurand**
- **Measurement Model**
  - Measurement Equations
  - Observation Equations
- **Inputs for Measurement Models**
- **Uncertainty Evaluation**
  - Bottom-Up (Uncertainty Budget)

    NIST Uncertainty Machine
  - Top-Down (e.g., Proficiency Test)

    Mathematical Statistics
- **Measurement Result**

Grandfathering

- **All uncertainty evaluations published** as part of measurement results produced in the delivery of NIST measurement services (reference materials, calibrations, and interlaboratory studies that NIST has participated in, including key comparisons) **remain valid and need not be redone**
- **Conventional procedures** for uncertainty evaluation described in NIST Technical Note 1297 (and in original version of the GUM) **may continue to be used** when there is no compelling reason to question their applicability

Measurement

Experimental or computational process that produces

- Estimate of true value of a property of a material or virtual object or collection of objects, or of a process, event, or series of events
- Evaluation of uncertainty surrounding that estimate, intended for use in support of *decision-making* (White, 2011)

Measurand

Property intended to be measured

- *Qualitative*

  Identity of nucleobase at particular location of DNA strand

- *Quantitative*

  Mass concentration of 25-hydroxyvitamin D_3 in NIST SRM 972a, Level 1

Measurement Uncertainty
HEAT CAPACITY OF AMMONIA

We think our reported value is good to 1 *part in* 10 000

We are willing to bet our own money at even odds that it is correct to 2 *parts in* 10 000

Furthermore, if by any chance our value is shown to be in error by more than 1 *part in* 1000, *we are prepared to eat the apparatus and drink the ammonia*

— C. H. Meyers, 1930s

Told by D. P. Johnson, reported by H. Ku, 1973
Quoted by T. Doiron & J. Stoup, 1997

Measurement Uncertainty

Doubt about true value of measurand that remains
after making a measurement

- Characterize *margin of doubt* by answering two questions (Bell, 1999):
  - *How big is the margin?*
  - *How bad is the doubt?*
- Probability distribution (on possible values of measurand) describes measurement uncertainty completely (Thompson, 2011; O'Hagan, 2014)

Measurement Models

Describe relationship between value of measurand (*output*) and values of qualitative or quantitative properties (*inputs*) that determine or inform its value

- *Measurement equations*
- *Observation equations (statistical models)*

Measurement Models
MEASUREMENT EQUATIONS

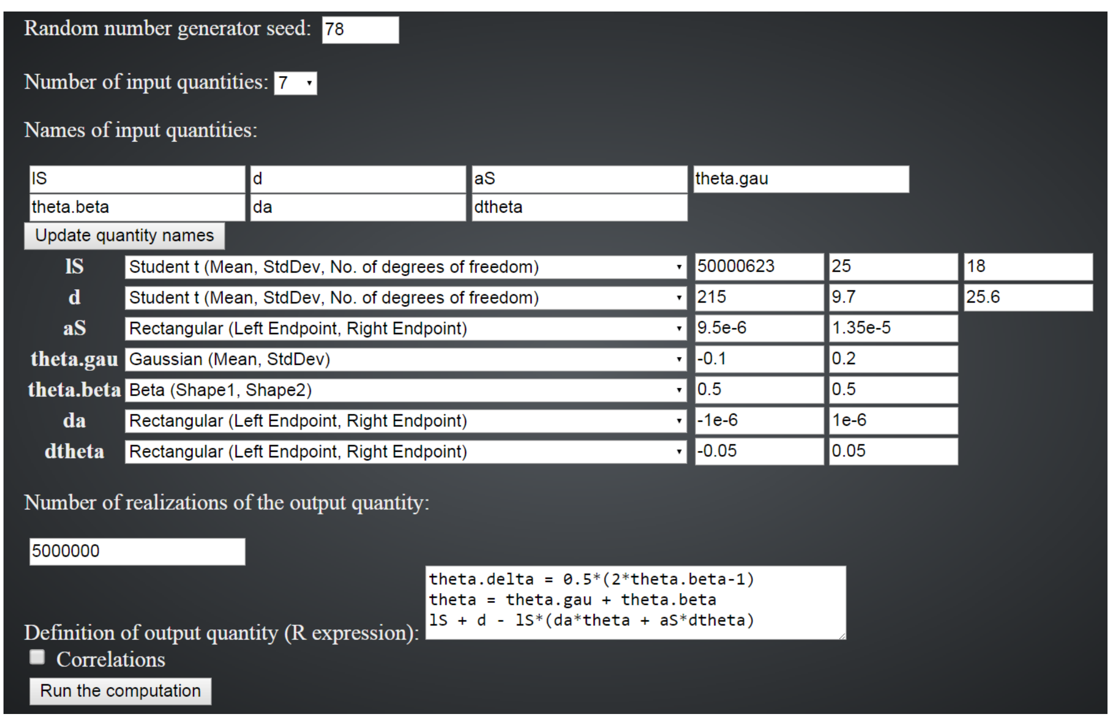
- *Measurement equation* expresses measurand as known function of input variables for which estimates and uncertainty evaluations are available

EXAMPLE

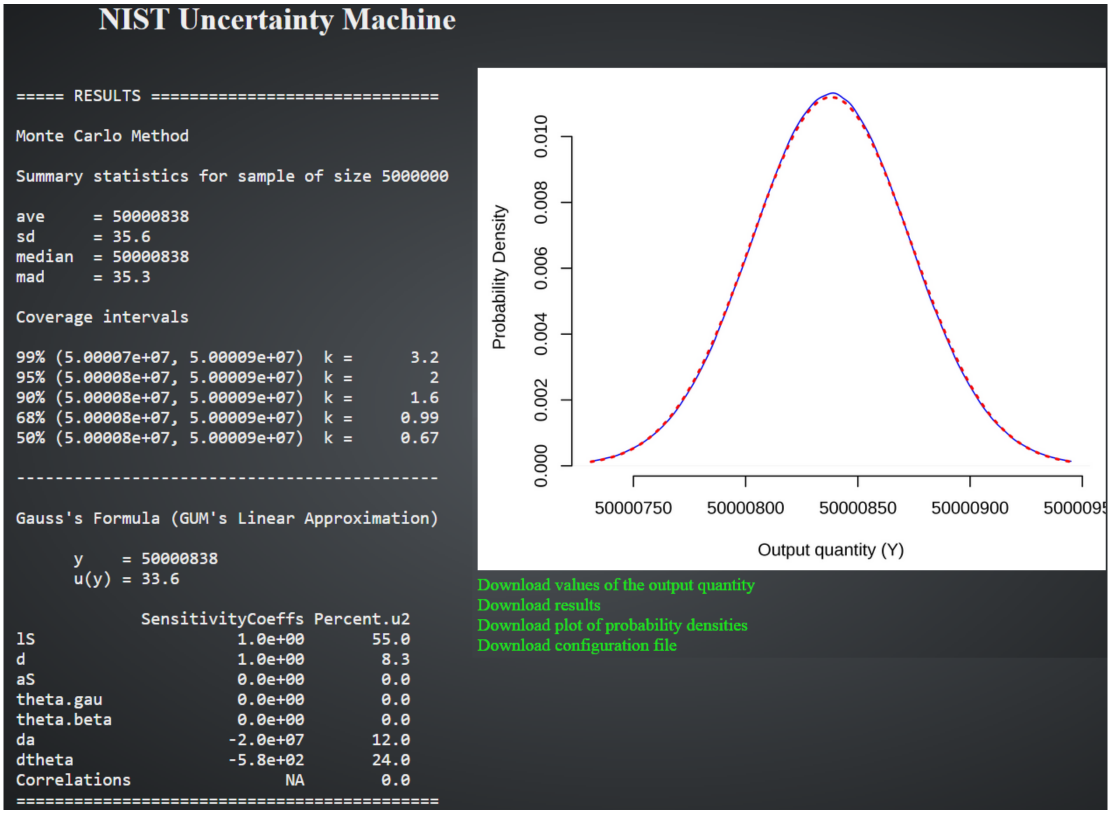
- End-Gage Calibration (GUM H-1)

$$\ell = \ell_S + d - \ell_S \times (\delta\alpha \times \theta + \alpha_S \times \delta\theta)$$

uncertainty.nist.gov



uncertainty.nist.gov



Uncertainty Budget

SENSITIVITY COEFFICIENTS *vs.* ANOVA

	GUM	ANOVA	(%)
ℓ_S	62	55	
d	9	8	
a_S	0	0	
θ	0	0	
$\delta\alpha$	1	1	
$\delta\theta$	28	24	
Residual		**12**	

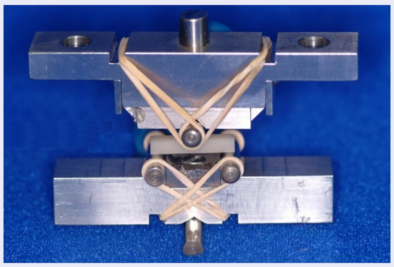
Lafarge & Possolo (2015) *NCLSI Measure*

Measurement Models

OBSERVATION EQUATIONS

- *Observation equation* expresses measurand as known function of parameters of probability distribution of inputs

EXAMPLE



- **Tensile Strength of Alumina**
  - Expected value of Weibull distribution of rupture stress in flexure test
  - Determined by Weibull scale and shape parameters

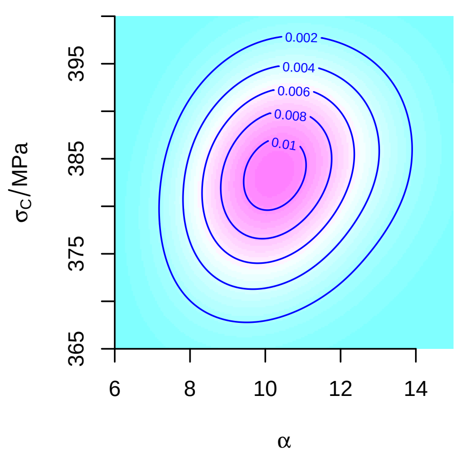
Tensile Strength of Alumina

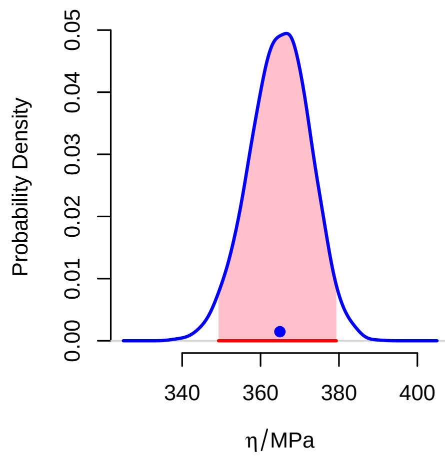
- Rupture stress S of 32 coupons: 265 MPa, ..., 435 MPa
- Observation equation: $\log S_i = \log \sigma_C + (1/\alpha) \log Z_i$
  - α Shape parameter
  - σ_C Scale parameter (*characteristic strength*)
  - $Z_1, \dots, Z_{32}$ Exponential(1) "errors"
- Measurand: Tensile strength η
- Measurement equation: $\eta = \sigma_C \Gamma(1 + 1/\alpha)$

Tensile Strength of Alumina

RESULTS

- Maximum likelihood estimates $\widehat{\alpha} = 10.1$, $\widehat{\sigma}_C = 383\,\mathrm{MPa}$
- $\widehat{\eta} = \widehat{\sigma}_C \Gamma(1 + 1/\widehat{\alpha}) = 365\,\mathrm{MPa}$
- Uncertainty evaluation (parametric statistical bootstrap):
  95 % coverage interval for η: $(349\,\mathrm{MPa}, 379\,\mathrm{MPa})$





Observation Equations

COMMON STATISTICAL MODELS

- Measurement Error Model $\varphi(x_i) = \psi(y) + \epsilon_i$
  - *Indications of resistance x of semiconductor at temperature y, obtained under conditions of repeatability*
- Random Effects Model $\varphi(x_i) = \psi(y) + \lambda_i + \epsilon_i$
  - *NIST SRM 3128 (Lead Standard Solution), blending gravimetric and ICP-OES determinations*
  - *Inter-laboratory studies (including KCs)*
- Regression Model $\varphi(x_i) = y(r_i) + \epsilon_i$
  - *NIST SRM 1693a (Sulfur dioxide in nitrogen)*
  - *NIST force transducer calibrations*

Uncertainty Evaluation for Inputs

- *Bottom-Up* and *Top-Down* approaches
- *Measurement Equations* and *Observation Equations*

A. Possolo and C. Elster (2014) Evaluating the uncertainty of input quantities in measurement models *Metrologia* 51

Prior Knowledge

- When there is relevant knowledge in hand about measurand before measurement, then use it together with fresh measurement data
  - Bayesian procedure employs Bayes's rule
  - Using informative prior distributions
  - Usually via Markov Chain Monte Carlo sampling

Prior Knowledge

EXAMPLES

GREENHOUSE GASES

Measure fluxes by updating prior information from inventories with current observations

HEART ATTACKS

Measure effect of early *vs.* late administration of anistreplase by blending field data with cardiologist's expert opinion

SRM 54D (TIN-BASE BEARING METAL, 1957)

Exploit relevant historical data and expert knowledge to develop prior distribution for $u(w_{\mathrm{Sn}})$, then blend it with original determinations of w_{Sn}

DNA Sequencing

PROBLEM & DATA

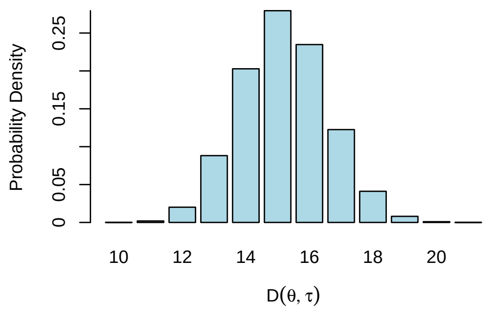
- Measurand θ is sequence of nucleobases in DNA strand
- Measure distance $D(\theta, \tau)$ to target sequence $\tau = (\mathrm{GGATTTTATTATAAATGGGTATACAATTTTTAAAATTTT})$
- Sequencing procedure yields $\widehat{\theta} = (\mathrm{TTTTTATAATTGGTTAA\mathbf{TCA}TTTTTTTTTAATTTTT})$
- *Quality scores* translated into probabilities

LOC	Q(A)	Q(C)	Q(G)	Q(T)	BASE	Pr(A)	Pr(C)	Pr(G)	Pr(T)
...	...	...	...	...	...	...	...	...	...
18	−25	−40	−12	11	**T**	0.00	0.00	0.06	0.94
19	−15	12	−36	−15	**C**	0.03	0.94	0.00	0.03
20	12	−16	−40	−14	**A**	0.94	0.02	0.00	0.04
...	...	...	...	...	...	...	...	...	...

DNA Sequencing

RESULTS

- Draw sample sequences $\theta_1^*, \ldots, \theta_K^*$ from distribution of measurand
- Compute and summarize corresponding distances $D(\theta_1^*, \tau), \ldots, D(\theta_K^*, \tau)$ to target sequence



Measured value of Damerau-Levenshtein distance
$D(\widehat{\theta}, \tau) = 13$

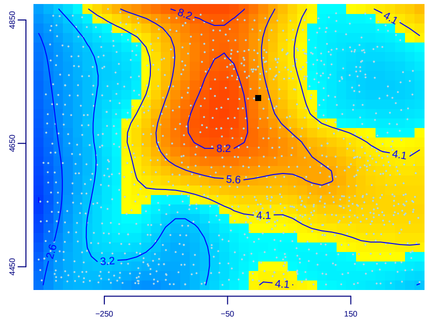
Colorado Uranium

- Mass fraction of uranium $w(u, v)$ (mg/kg) across Colorado
- 1150 samples, primarily of stream sediments
  National Uranium Resource Evaluation (NURE, 1975–1980)
- Measured using delayed neutron counting
- Observation equation

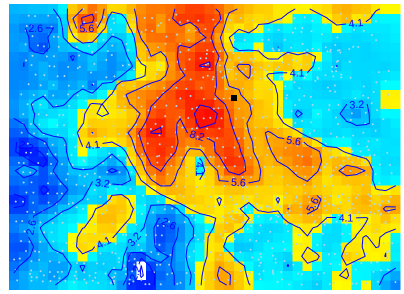
$$\frac{w(u, v)^{\lambda} - 1}{\lambda} = \theta(u, v) + \epsilon(u, v)$$
$$\theta \sim \text{Measurand}$$
$$\epsilon \sim \text{White noise}$$

Colorado Uranium
MODEL UNCERTAINTY

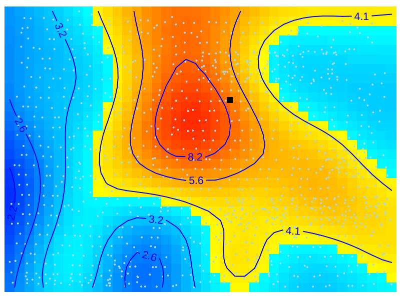
Locally quadratic regression



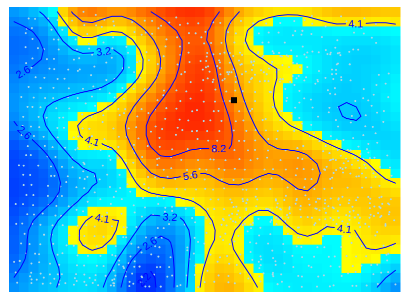
Ordinary kriging, Matérn



Generalized additive model

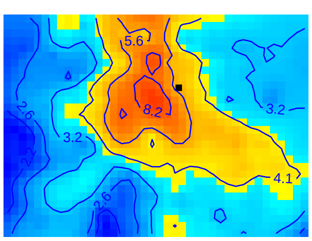


Multi-resolution Gaussian process

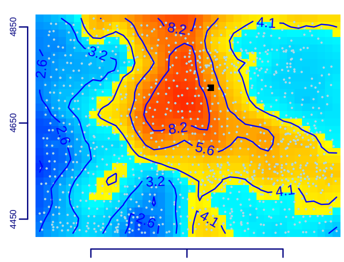


Colorado Uranium
MODEL AVERAGING

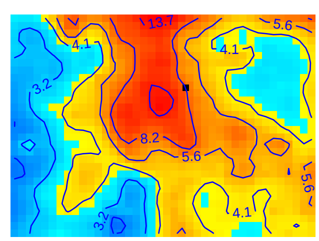
LOWER 95TH



ESTIMATE



UPPER 95TH



Summation

- NIST's *Simple Guide*
  - Bridges past accomplishments and future needs, mindfully of current capabilities and challenges (technical, educational, accreditation, compliance, etc.)
  - Opens door toward new approaches and methods for uncertainty evaluation (including Bayesian methods)
    - Intended for use when needed
    - Without requiring that techniques currently in use be abandoned
- *NIST Uncertainty Machine*, accessible universally as Web service, removes barriers to application of Gauss's and Monte Carlo uncertainty evaluation methods for all conventional measurement models