UNCERTAINTY FROM BIAS IN VIRTUAL MEASUREMENTS AND THE NIST CCCBDB

Karl Irikura, Russell Johnson III, Raghu Kacker
NIST, Gaithersburg, MD 20899-8910

Mathematical and Computational Sciences Division
APRIL 20, 2004

CONTENTS

- Generic approach based on ISO *Guide* to quantify uncertainty from bias in a virtual measurement (Raghu)

- Computational Chemistry Comparison and Benchmark Database (CCCBDB) developed by Russ Johnson of NIST
  - Estimated biases in virtual measurements from computational quantum chemistry models for many properties of many molecules

- Our approach to quantify uncertainties in quantum chemistry based on ISO *Guide* and CCCBDB (Karl)

ISO GUIDE ON EXPRESSION OF UNCERTAINTY

- Virtual measurement: output of a computational model
  - Virtual = Calculated, Physical = Measured = Experimental
- In computational chemistry, uncertainty arises mainly from bias relative to value of measurand
- Before ISO *Guide* no generally accepted approach to quantify and incorporate uncertainty arising form bias
- Also, some applications require both virtual and physical measurements
- For physical measurements, ISO *Guide* de facto standard
- In summary, ISO *Guide* is suitable for both virtual and physical measurements

ACCOUNTING FOR UNCERTAINTY FROM BIAS

- Y is value of measurand, x is virtual measurement, X is expected value and $u(x)$ is standard deviation of x
- Additive bias: $X - Y$, Fractional bias: X / Y
- ISO *Guide*: correct x for its bias and include uncertainty associated with correction in combined uncertainty
- Bias is a constant, but correction C for bias is a variable with state-of-knowledge distribution representing belief about required correction (negative/reciprocal of bias)
- Expected value and standard deviation of C are c and $u(c)$
- Corrected virtual measurement y for Y and uncertainty $u(y)$ determined from x, c, $u(x)$, and $u(c)$

CORRECTION FOR BIAS

- A measurement equation is required to apply correction
  - All quantities are variables with state-of-knowledge distributions
- Measurement equation for additive bias: $Y = X + C$
- $y = x + c$
- $u(y) = \sqrt{[u^2(x) + u^2(c)]}$
- Measurement equation for fractional bias: $Y = X \times C$
- $y = x \times c$
- $u_r(y) = \sqrt{[u_r^2(x) + u_r^2(c)]}$
- $u_r(y) = u(y)/y, \quad u_r(x) = u(x)/x, \quad u_r(c) = u(c)/c$

UNCERTAINTY IN VIRTUAL PREDICTIONS

- Repeat evaluations of x give same result
- So $u(x) = 0$
- Additive bias: $u(y) = \sqrt{[u^2(x) + u^2(c)]} = u(c)$
- Fractional bias: $u_r(y) = \sqrt{[u_r^2(x) + u_r^2(c)]} = u_r(c)$
- Thus $u(y) = y\, u_r(y) = y\, u(c)/c = x\, u(c)$
- Entire uncertainty arises from correction for bias

USE OF CCCBDB FOR SPECIFYING c AND $u(c)$

- CCCBDB is large collection of estimated biases in virtual measurements from quantum chemistry models
  - Estimated bias = $x_i - z_i$ or x_i / z_i, where z_i is high quality physical measurement corresponding to virtual measurement x_i
- Bias for target molecule is unknown
- Suppose it is possible to identify a class of molecules in CCCBDB for which biases are believed to be similar to the bias in the target molecule
  - Scientific judgment
- Then estimated biases for the matching class may be used to specify $E(C) = c$ and $S(C) = u(c)$

CORRECTION c AND UNCERTAINTY $u(c)$

- Suppose $c_1,\ldots, c_m$ are estimated corrections in CCCBDB for matching class of molecules
- Estimated correction for additive bias $c_i = z_i - x_i$
- Estimated correction for fractional bias $c_i = z_i / x_i$
- $c = \sum a_i c_i / \sum a_i$
- $u(c) = \sqrt{\{\sum a_i [c_i - c]^2 / \sum a_i\}}$
- Many ways to specify $a_1, \ldots, a_m$ depending on application
- Simple case: $a_i = 1$, $c = c_A = \sum c_i /m$
- $u(c) = s = \sqrt{\{\sum [c_i - c]^2 /m \}}$