

# Uncertainty Associated with Virtual Measurements from Computational Quantum Chemistry Models

Speaker: Raghu N. Kacker

Authors: Karl K. Irikura, Russell D. Johnson III, Raghu N. Kacker

National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

Email: raghu.kacker@nist.gov

## Abstract

This paper is a review, from the viewpoint of methodology, of the following two papers: “Uncertainty Associated with Virtual Measurements from Computational Quantum Chemistry Models,” *Metrologia*, 41 (2004) pp. 369-375 and “Uncertainties in scaling factors for ab initio vibrational frequencies,” *Journal of Physical Chemistry A*, 109 (2005) pp. 8430-8437. Both are authored by Karl K. Irikura, Russell D. Johnson III, and Raghu N. Kacker of the US National Institute of Standards and Technology.

## Introduction

By a virtual measurement we mean a prediction determined from a computational model for the value of the measurand. A virtual measurement is an alternative to a physical measurement which is determined from laboratory experiments. Interest in virtual measurement is growing for economics and safety. Making high-quality physical measurements is an expensive and slow time-consuming process. The demand for property data from commercial, academic, and governmental researchers exceeds the national capacity for making high-quality physical measurements. Virtual measurements are comparatively fast, avoid handling of harmful materials, and are useful for determining properties of materials not yet synthesized. Thus virtual measurements have become critical for research and development of chemical processes, materials development, and drug discovery.

An important application is virtual measurements determined from computational quantum chemistry models. Chemists have achieved remarkable success in predicting properties of isolated gas-phase molecules from computational quantum chemistry models. Until now uncertainties associated with virtual measurements from computational quantum chemistry models have not been reported. We are the first to do so. The uncertainty associated with a virtual measurement arises largely from its bias with respect to the value of the measurand. The bias supposedly results from the lack of rigor of the theory and the basis sets used for computations.

In this paper, we describe an approach based on the ISO-GUM for quantifying the uncertainties associated with virtual measurements from computational quantum chemistry models. We note that before the publication of the ISO-GUM, there was no generally accepted approach to quantify the uncertainty arising from bias. The ISO-GUM requires specification of a correction for bias and its associated uncertainty. Until recently there was no practical way to specify a correction for bias and its associated

uncertainty. Specification of correction and uncertainty has recently been made possible by the NIST Computational Chemistry Comparison and Benchmark Database (CCCBDB) developed by Russell Johnson III. Thus our approach is based on the ISO-GUM and the CCCBDB.

### **Correction for bias and its associated uncertainty**

Following the ISO-GUM, we use the symbol  $Y$  for both the unknown value of the measurand, which is a molecular property of interest, and a variable with a state-of-knowledge probability distribution about the value of the measurand. We use the symbol  $x$  for the raw uncorrected virtual measurement and the symbol  $u(x)$  for the standard uncertainty associated with  $x$ . We use the symbol  $X$  for both the unknown expected value of the sampling distribution of  $x$  and a variable with a state-of-knowledge probability distribution about the expected value. The additive (or subtractive) bias in  $x$  is  $B = X - Y$ . The multiplicative (fractional) bias in  $x$  is  $B = X / Y$ .

The ISO-GUM recommends that a correction for additive bias or a correction factor for multiplicative bias must be applied to the raw result  $x$ . Since bias is an unknowable quantity, the correction carries uncertainty. The uncertainty associated with the correction or the correction factor must be quantified and included in the combined uncertainty associated with the corrected result. A measurement equation is required to apply a correction for additive bias or apply a correction factor for multiplicative bias.

The measurement equation that corresponds to the additive bias is  $Y = X + C$ , where  $Y, X, C$  are variables with state-of-knowledge distributions and  $C$  represents state-of-knowledge about the negative of bias  $B = X - Y$ . The uncorrected result  $x$  and uncertainty  $u(x)$  are identified with the expected value and standard deviation of the state-of-knowledge distribution for  $X$ . Thus  $E(X) = x$  and  $S(X) = u(x)$ . Suppose  $E(C) = c$  and  $S(C) = u(c)$ ;  $E(Y) = y$  and  $S(Y) = u(y)$ . The corrected result is  $y = x + c$ . The combined uncertainty associated with the corrected result is  $u(y) = \sqrt{u^2(x) + u^2(c)}$ .

The measurement equation that corresponds to the multiplicative bias is  $Y = X \square C$  where  $C$  represents the state-of-knowledge about the reciprocal of bias  $B = X / Y$ . The corrected result is  $y = x \square c$  and the uncertainty associated with the corrected result is  $u_r(y) = \sqrt{u_r^2(x) + u_r^2(c)}$ , where  $u_r(y) = u(y)/y$ ,  $u_r(x) = u(x)/x$ , and  $u_r(c) = u(c)/c$  are the relative standard uncertainties associated with  $y, x$ , and  $c$ , respectively. The combined uncertainty associated with the corrected result is  $u(y) = y \square u_r(y)$ .

Repeated determinations from a computational quantum chemistry model yield identical or approximately identical results. Thus the uncertainty  $u(x)$  associated with uncorrected result  $x$  is approximately zero. Thus the combined uncertainty associated with the corrected result with additive bias is approximately  $u(c)$ . The combined uncertainty associated with the corrected result with multiplicative bias is approximately  $x \square u(c)$ . Thus almost the entire uncertainty comes from correction for bias.

## Computational chemistry comparison and benchmark database (CCCBDB)

The CCCBDB is a large database of thermo-chemical gas properties containing results from over 140 computational models and the corresponding high-quality physical measurements. Thus it provides estimated biases in virtual measurements. The estimated additive bias  $b_i$  is  $x_i - z_i$  and the estimated multiplicative bias  $b_i$  is  $x_i / z_i$ , where  $z_i$  is high quality physical measurement corresponding to a virtual result  $x_i$  in the database. Suppose a class of results in CCCBDB can be identified for which biases are believed to be similar in sign and magnitude to the unknown bias in the target molecular property. Then estimated biases for the matching class may be used to specify the correction  $E(C) = c$  and the uncertainty  $S(C) = u(c)$ . Let  $c_1, \dots, c_m$  be estimated corrections in the matching class where  $c_i = z_i - x_i$  or  $c_i = z_i / x_i$ . Then  $c = [1/\sum_i a_i][\sum_i a_i \square c_i]$  and  $u(c) = \{[1/\sum_i a_i][\sum_i a_i u^2(c_i) + \sum_i a_i (c_i - c)^2]\}^{1/2}$  where  $a_1, \dots, a_m$  are suitably chosen weights. These formulas are based on linear mixture probability distributions.

### Atomization enthalpy

The following example of additive correction for bias is from the Metrologia paper. The expressions for the correction  $c$  and the uncertainty  $u(c)$  based on equal weight  $a_i = 1$  are as follows:  $c = [1/m][\sum_i c_i]$  and  $u(c) = \{(1/m)[\sum_i u^2(c_i)] + (1/m)[\sum_i (c_i - c)^2]\}^{1/2}$ . The expression for  $u(c)$  consists of the average uncertainty associated with the estimated corrections and variance between the estimated corrections for the matching class.

Consider the atomization enthalpy for ethyl thioformate ( $C_3H_6OS$ ). The uncorrected computed result  $x$  determined from computational quantum chemistry model is  $x = 4093.8 \text{ kJmol}^{-1}$ . The correction determined from a matching class of molecules in the CCCBDB is  $c = 21.8 \text{ kJmol}^{-1}$  with  $u(c) = 19.2 \text{ kJmol}^{-1}$ . Thus the virtual measurement is  $y = x + c = 4115.6 \text{ kJmol}^{-1}$  and the combined standard uncertainty associated with the virtual measurement is  $u(y) \approx 19.2 \text{ kJmol}^{-1}$ . The uncertainty expressed as a two-standard uncertainty interval is  $[4115.6 \pm 38.4] \text{ kJmol}^{-1}$ . A high-quality physical measurement of the atomization enthalpy for ethyl thioformate is  $[4129.2 \pm 5] \text{ kJmol}^{-1}$ . This agrees with the prediction based on computational model.

### Scaling factors for vibrational frequencies

The following example of multiplicative correction for bias is from the Journal of Physical Chemistry paper. The expressions for the correction  $c$  and uncertainty  $u(c)$  based on weight  $a_i$  determined from least-squares theory are as follows: scaling factor  $c = [1/\sum_i x_i^2][\sum_i x_i^2 \square c_i]$  and  $u(c) \approx \{[1/\sum_i x_i^2][\sum_i x_i^2 (c_i - c)^2]\}^{1/2}$ .

My coauthors have assembled a large table of scaling factors and their associated uncertainties. The table displays scaling factors and their associated uncertainties for eight theoretical methods and five basis sets. Each of forty scaling factors and uncertainties are based on a over 3000 independent vibrational frequencies from the CCCBDB. Values of scaling factors from literature are given when available. This is the first table to quantify the uncertainties.

A typical entry in the table is the scaling factor for the theoretical method; Hartree-Fock (HF) and the basis set: 6-311G(d, p). The scaling factor is  $c = 0.9085$  and its associated uncertainty is  $u(c) = 0.0246$ . These are based on 3478 independent vibrational frequencies in the database. The literature value for this scaling factor is 0.9051 (Scott and Radom (1996) Journal of Physical Chemistry). Uncertainty is not stated in the literature.

Scaling factors are reported in the literature to four decimal places. We have determined that scaling factors are accurate to only two places. We also determined that scaling factors depend only weakly on the basis sets. Thus scaling factors may be used for basis sets not included in Table 1. This investigation is currently being extended to molecular properties that depend on vibrational partition function (for example, zero point energies).

## Summary

In summary, uncertainty associated with a virtual measurement arises primarily from its bias. We have proposed an approach to quantify this uncertainty based on the ISO-GUM and the NIST CCCBDB. Using this approach we determined scaling factors and uncertainties for ab initio vibrational frequencies. Our scaling factors are similar to the literature values; however, we are first to quantify corresponding uncertainties. Uncertainties show that the scaling factors are accurate to only two places, not four as often reported.