

Application: Uncertainties in Quantum Chemistry

- The ISO *Guide* can be used to obtain uncertainties for virtual measurements
- The CCCBDB contains estimated biases for many virtual measurements of many properties of many molecules
- Let's combine them

Bias in Quantum Chemistry

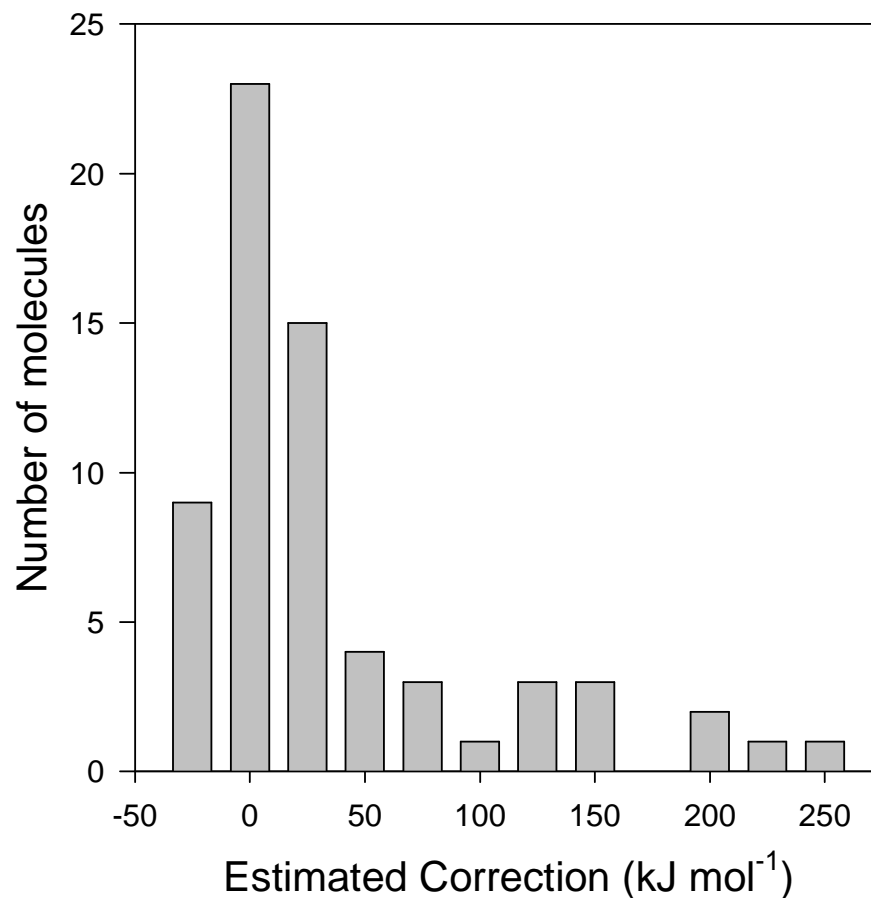
- There are two dominant contributions
 - Theoretical approximation (bias B_t)
 - Basis set truncation (bias B_b)
- Some popular theories and basis sets can be ordered
 - Let N_t be the ordinal number for a **theory**
 - Let N_b be the ordinal number for a **basis set**
- Correlation
 - B_t, B_b independent for large enough N_t, N_b
 - B_t, B_b **not** independent for typical N_t, N_b
- So for typical methods, consider only **aggregate bias $B_{(t,b)}$**

Uncertainty: Additive Bias

- Correcting the virtual measurement
 - Model output plus estimated correction for bias:
$$y_0 = x_{(t,b)} + c_{(t,b)}$$
 - Uncertainty: $u(y_0) = u(c_{(t,b)})$ since $u(x_{(t,b)}) \approx 0$
- Estimated values of $c_{(t,b)}$ and $u(c_{(t,b)})$ are mean and standard deviation of corrections for similar molecules in the CCCBDB
- Relies upon good **classification** of “similar” molecules

Example: Atomization Energy

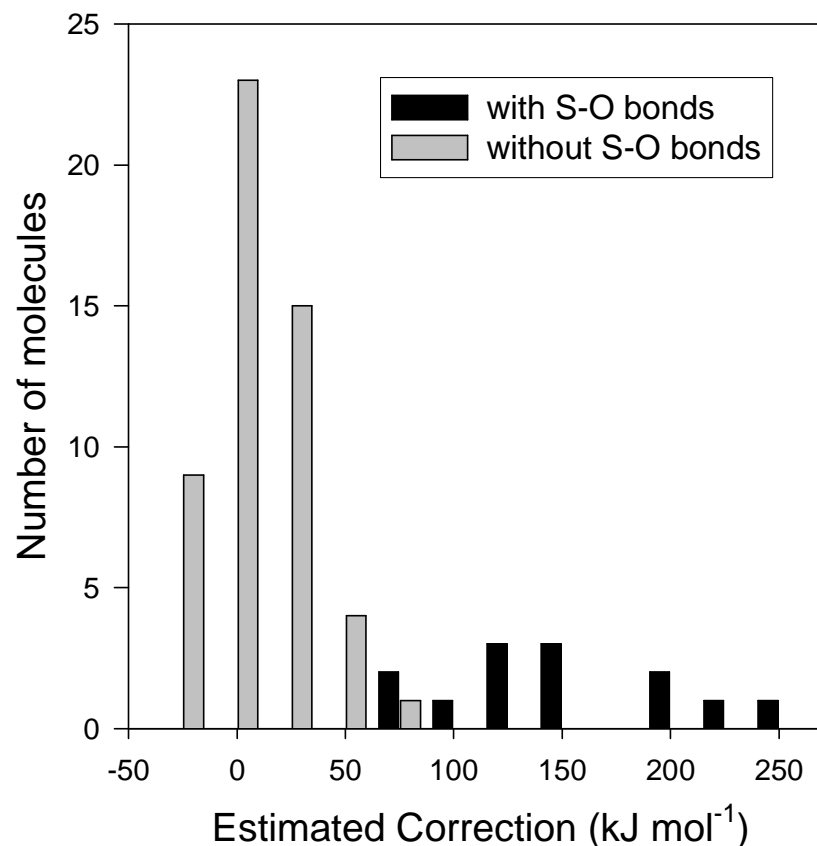
- Atomization: shredding a molecule into its constituent atoms, e.g., $\text{H}_2\text{O} \rightarrow 2\text{H} + \text{O}$
- Consider all sulfur-containing molecules in the CCCBDB (plot at right)
- Skewed distribution suggests poor classification



<i>m</i>	Mean	Standard deviation	Skewness
65	50.5	64.2	1.7

Example: Classification

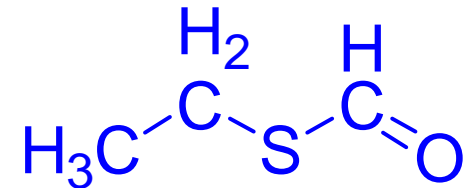
- Two classes here: with or without S-O bonds
- New distributions better characterized by mean and standard deviation



Set	<i>m</i>	Mean	Standard deviation	Skewness
All	65	50.5	64.2	1.7
With S-O	13	165.2	52.0	0.5
No S-O	52	21.8	19.0	0.6

Example: Atomization of ethyl thioformate (C₃H₆OS)

- Contains no S-O bonds
- Choose a **model**
 - Theory: mPW1PW91
 - Basis set: 6-31G(d)
- Virtual measurement:
(4116 ± 38) kJ mol⁻¹
- Physical measurement:
(4129 ± 5) kJ mol⁻¹



$$x = 4093.8 \text{ kJ mol}^{-1}$$

$$c = 21.8 \text{ kJ mol}^{-1}$$

$$y = x + c = 4115.6 \text{ kJ mol}^{-1}$$

$$2u(c) = 38.0 \text{ kJ mol}^{-1}$$

Uncertainty: Fractional Bias

- Correcting the virtual measurement

- Multiplicative correction for bias:

$$y_0 = x_{(t,b)} \times c_{(t,b)}$$

- Uncertainty: $u(y_0) = x_0 \times u(c_{(t,b)})$ since $u(x_{(t,b)}) \approx 0$

- Recall weights a_i :

$$c_{(t,b)} = \sum a_i c_i / \sum a_i, \text{ where } c_i = z_i / x_i$$

$$u(c_{(t,b)}) = \sum a_i [c_i - c_{(t,b)}]^2 / \sum a_i$$

Example: Vibrational Frequencies

- Multiplicative scaling is established practice
 - most cited paper is by Scott and Radom, 1996 (1700 citations)
- But **no uncertainties** available (yet)
- Least-squares corresponds to weighting:

$$a_i = x_i^2$$

TABLE 1: Frequency Scaling Factors Suitable for Fundamental Vibrations and rms_{ov} (cm^{-1}) Derived from a Least-Squares Fit of Frequencies^a

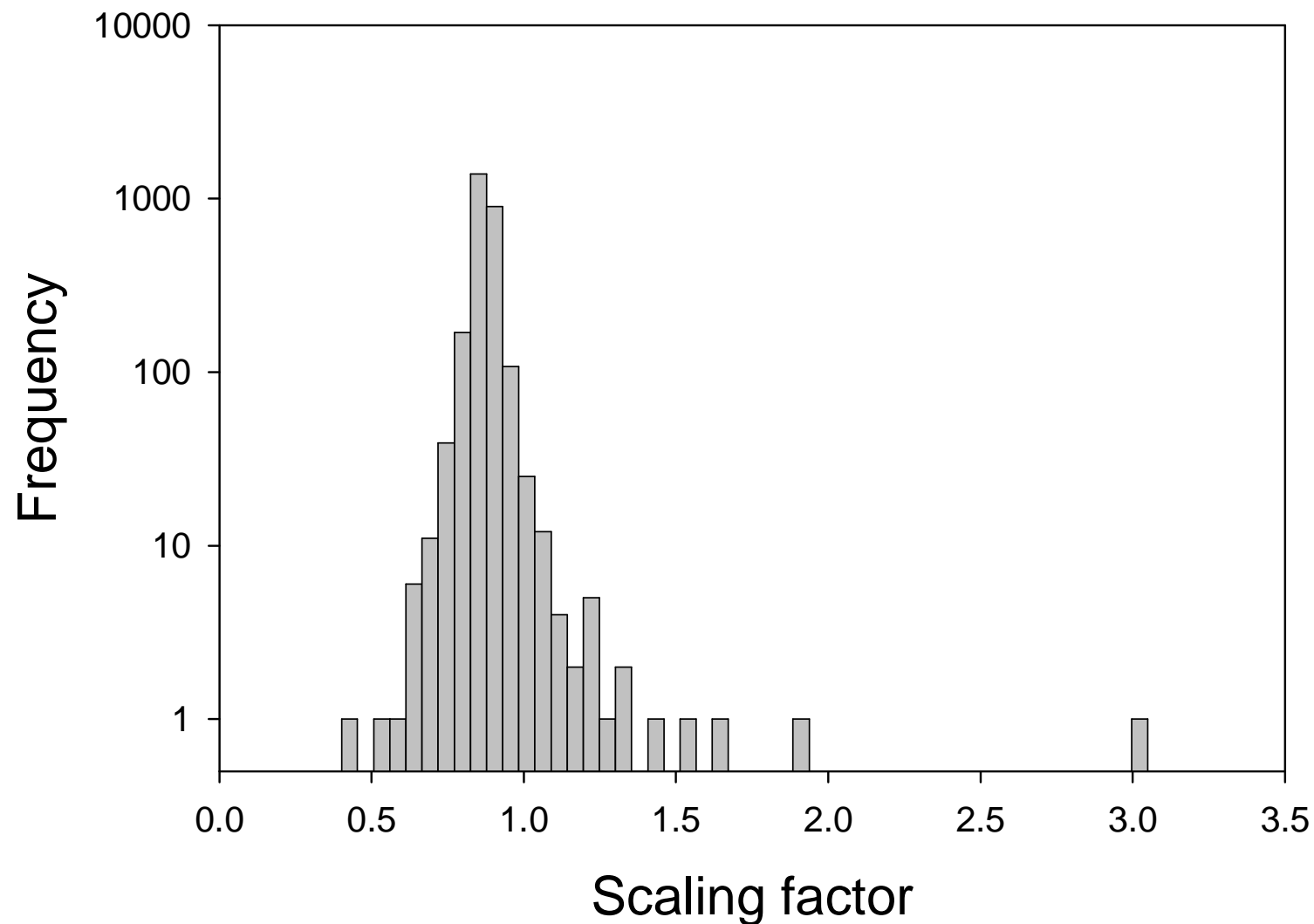
method	scale factor ^b	rms_{ov} ^c
AM1	0.9532	126
PM3	0.9761	159
HF/3-21G	0.9085	87
HF/6-31G(d)	0.8953	50
HF/6-31+G(d)	0.8970	49
HF/6-31G(d,p)	0.8992	53
HF/6-311G(d,p)	0.9051	54
HF/6-311G(df,p)	0.9054	56
MP2-fu/6-31G(d)	0.9427	61
MP2-fc/6-31G(d)	0.9434	63
MP2-fc/6-31G(d,p)	0.9370	61
MP2-fc/6-311G(d,p)	0.9496	60
QCISD-fc/6-31G(d)	0.9537	37
B-LYP/6-31G(d)	0.9945	45
B-LYP/6-311G(df,p)	0.9986	42
B-P86/6-31G(d)	0.9914	41
B3-LYP/6-31G(d)	0.9614	34
B3-P86/6-31G(d)	0.9558	38
B3-PW91/6-31G(d)	0.9573	34

^a Using F1 set of 1066 frequencies. ^b As defined by eq 10. ^c Overall root-mean-square error, as defined by eq 13.

Preliminary Results

Model	S&R	Ours	Uncert.
AM1	0.9532	0.9530	0.1176
PM3	0.9761	0.9730	0.1553
HF/3-21G	0.9085	0.9044	0.0812
HF/6-31G(d)	0.8953	0.8982	0.0476
HF/6-31G(d,p)	0.8992	0.9025	0.0492
MP2-fu/6-31G(d)	0.9427	0.9414	0.0518
MP2-fc/6-31G(d)	0.9434	0.9423	0.0532
MP2-fc/6-31G(d,p)	0.9370	0.9365	0.0643
QCISD-fc/6-31G(d)	0.9537	0.9512	0.0433
B-LYP/6-31G(d)	0.9945	0.9910	0.0553
B3-LYP/6-31G(d)	0.9614	0.9591	0.0556
B3-PW91/6-31G(d)	0.9573	0.9561	0.0438

Distribution for HF/6-31G(d)



Vibrational Frequencies: Conclusions

- Uncertainties in scaling factors are much larger than implicit
 - Only two significant figures, not four
- Uncertainty in scaling factor can be trivially propagated to the frequencies
 - No uncertainties were available previously

Summary

- Extension of ISO *Guide* to virtual measurements
- Application to quantum chemistry enabled by CCCBDB: <http://srdata.nist.gov/cccbdb>
- Initial application reveals uncertainties larger than expected
 - Work is in early stage and is ongoing