

# Scaling factors for *ab initio* vibrational frequencies: comparison of uncertainty models for quantified prediction

Pascal Pernot<sup>1,2</sup>

January 5, 2009

<sup>1</sup>Laboratoire de Chimie Physique, CNRS, UMR8000, Orsay, F-91405

<sup>2</sup>Univ Paris-Sud, Orsay, F-91405

## Abstract

Bayesian Model Calibration is used to revisit the problem of scaling factor calibration for semi-empirical correction of *ab initio* calculations. A particular attention is devoted to uncertainty evaluation for scaling factors, and to their effect on prediction of observables involving scaled properties. We argue that linear models used for calibration of scaling factors are generally not statistically valid, in the sense that they are not able to fit calibration data within their uncertainty limits. Uncertainty evaluation and uncertainty propagation by statistical methods from such invalid models are doomed to failure. To relieve this problem, a stochastic function is included in the model to account for model inadequacy, according to the Bayesian Model Calibration approach. In this framework, we demonstrate that standard calibration summary statistics, as optimal scaling factor and root mean square, can be safely used for uncertainty propagation only when large calibration sets of precise data are used. For small datasets containing a few dozens of data, a more accurate formula is provided which involves scaling factor calibration uncertainty. For measurement uncertainties larger than model inadequacy, the problem can be reduced to a weighted least squares analysis. For intermediate cases, no analytical estimators were found, and numerical Bayesian estimation of parameters has to be used.

Keywords: *Bayesian data analysis, Model calibration, Scaling factor, Vibrational frequency*

# 1 Introduction

The final stage in the development of a model, after calibration on an experimental dataset and proper validation, consists in its use for prediction: "If the experimental dataset is sufficiently broad, there is a reasonable expectation that the results will be accurate to something like the target accuracy".<sup>1</sup> The estimation of uncertainty for the results of computational chemistry is indeed of paramount importance, notably to their use in multiscale modeling.<sup>2,3</sup> The concept of Virtual Measurement has been introduced by Irikura<sup>4</sup> to take advantage of the standardized procedures defined by the Guide to the Expression of Uncertainty in Measurement (also known as "the GUM"),<sup>5</sup> and to apply them in the case of quantum chemistry modeling. Random uncertainties have been shown to be very small, and the major uncertainty factors are biases due to basis-set and/or theory limitations. For quantum chemistry to be predictive, i.e. to be able to predict observables with confidence intervals, one has to correct for these biases, which commonly requires semi-empirical corrections. Uncertainty attached to these corrections have to be considered in the final uncertainty budget, to which they constitute often a major contribution.

Scaling of harmonic vibrational frequencies is an important example of calibration method in computational chemistry, where estimation of a vibrational frequency  $\nu$  is obtained by multiplying the corresponding harmonic vibrational frequency  $\omega$ , routinely calculated by standard computational chemistry codes, by an empirical scaling factor  $s$  (Fig. 1)

$$\nu = \omega s \quad (1)$$

Optimal scaling factors have been computed for extensive sets of theory/basis-set combinations.<sup>6,7,8,9</sup> In the majority of publications about scaling factors, two summary statistics are provided for each theory/basis-set combination: the optimal scaling factor and the root mean squares. From a reference dataset of experimental  $\{\nu_i\}_1^N$  and calculated vibrational frequencies  $\{\omega_i\}_1^N$  the optimal scaling factor obtained by the least-squares procedure is

$$\hat{s} = \frac{\sum \omega_i \nu_i}{\sum \omega_i^2} \quad (2)$$

and the quality of the correction is estimated by the root mean squares (rms) value

$$\gamma = \left( \frac{1}{N} \sum (\nu_i - \hat{s}\omega_i)^2 \right)^{1/2} \quad (3)$$

To our knowledge, these values are not explicitly used for uncertainty propagation, but the rms provides an estimate of the residual uncertainty resulting from the scaling correction ("something like the target accuracy",<sup>1</sup> or "a surrogate for uncertainty" according to<sup>10</sup>).

In a recent article (hereafter referred to as Paper I), Irikura *et al.*<sup>10</sup> treated the problem of uncertainty propagation from scaled frequencies, and, using procedures advocated by the GUM,<sup>5</sup> insisted that the scaling factor itself is subject to uncertainty, and proposed that this uncertainty is the major contribution to prediction uncertainty. On the practical side, prediction uncertainty would be proportional to the calculated harmonic frequency; on the factual side, these authors declare that scaling factors for vibrational frequencies are accurate to only two significant figures, whereas all other studies overstate their precision by reporting them with four figures. This approach has been adopted by the National Institute of Standards and Technology (NIST) and put into practice in the Computational Chemistry Comparison and Benchmark DataBase (CCCBDB),<sup>9</sup> section XIII.C.2, where all scaling factors are provided with uncertainties derived according to the procedure of Paper I.

In the present paper, we revisit the problem of scaling factor calibration and we recast it in the Bayesian Model Calibration framework, reputed for providing consistent uncertainty evaluation and propagation.<sup>11,12,13</sup> Section 2 presents the methodological elements used for calibration and validation procedures, which are applied to representative vibrational frequency and zero point vibrational energy datasets in Section 3. Bayesian calculations used in this study are fairly standard, but for the sake of completeness and for readers unfamiliar with statistical modeling, details are provided in the Appendix .

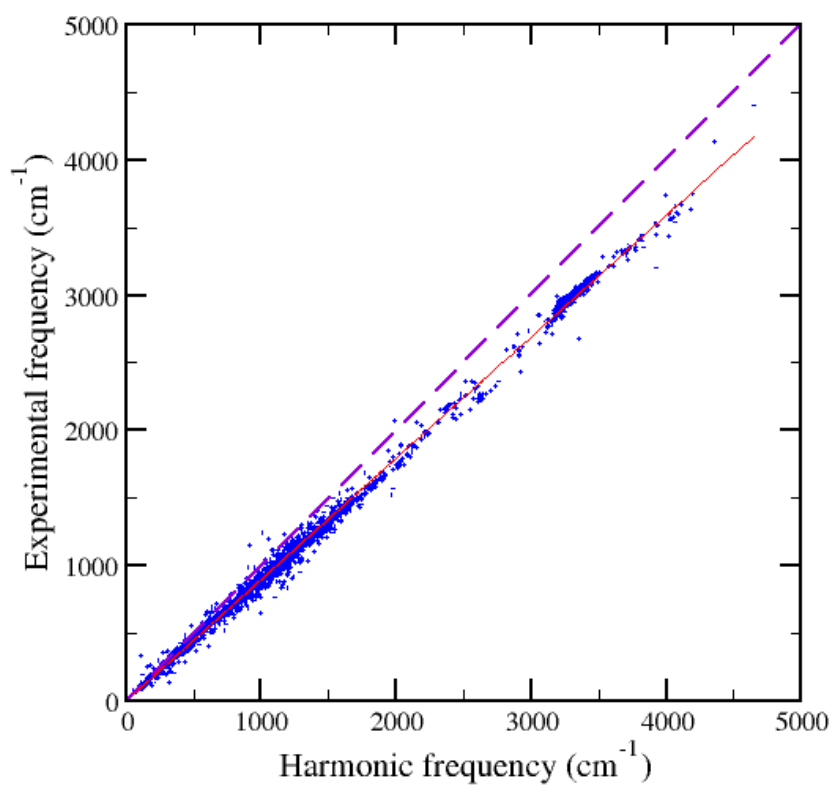


Figure 1: Correlation plot between calculated harmonic frequencies  $\omega_i$  and measured frequencies  $\nu_i$  for a set of vibrations extracted from the CCCBDB for the HF/6-31G\* combination of theory/basis-set (dots). The full line is the regression line  $\nu = s\omega$ ; the dashed line is a visual aid to appreciate the bias.

## 2 Methods

### 2.1 Statistical model for scaling factor calibration

Considering a measured frequency  $\nu^{obs}$ , one can assume that it is related to the true value  $\nu^{true}$  by

$$\nu^{obs} = \nu^{true} + \epsilon \quad (4)$$

where  $\epsilon \sim N(0, \rho^2)$  is a normal random variable centered at zero, with variance  $\rho^2$ , representing the measurement uncertainty in the hypothesis of additive white noise. In the following, we assume that  $\rho$  is known beforehand.

Calculated harmonic vibrational frequencies  $\omega$  are affected by random errors, related to numerical convergence defined by non-zero thresholds and the choice of starting point in geometry optimization, and to non-zero thresholds in wave-function optimization.<sup>4</sup> It has been shown that the associated uncertainties are negligible when compared to the measurement uncertainties.<sup>4</sup> In the following, one can thus assume that the harmonic vibrational frequencies, being deterministically calculated, have fixed values.

If one makes the hypothesis of a linear relationship between  $\nu^{true}$  and  $\omega$ , the calibration model resulting from this analysis is

$$\nu^{obs} = s\omega + \epsilon \quad (5)$$

For a single frequency, there is an optimal scaling parameter  $\hat{s} = \nu^{obs}/\omega$ . As  $\nu^{obs}$  is uncertain, with standard uncertainty  $\rho$ , the value of  $s$  cannot be known exactly and has a standard uncertainty  $u_s = \rho/\omega$ . For a calibration dataset with uniform measurement uncertainty  $\rho$ , it can be shown that the optimal value for  $s$  is still given by Eq. 2, and its standard uncertainty by  $u_s = \rho/\sqrt{\sum_{i=1}^N \omega_i^2}$  (*c.f.* Appendix A.3). Applicability of this formula is subject to one condition: the model (Eq. 5) should be statistically valid. This means that the residuals  $\{\nu_i^{obs} - \hat{s}\omega_i\}$  should have a normal distribution with variance  $\rho^2$ . Normality is not always verified, but most important, the variance condition is typically violated when precise data are used for calibration. The linear model (Eq. 5) is typically unable to reproduce a given set of measured frequencies *within their uncertainty range*. In such a case, the width of the

distribution of residuals is dominated by model misfit, not by measurement uncertainty, and the model is invalid.

An option would be to search for a better model than the linear scaling, which is beyond the purpose of this study. Instead, we consider here that the misfit is not deterministically predictable. The solution to preserve the linear scaling model in a statistically valid model, is thus to introduce a stochastic variable  $\xi$  to represent the discrepancy between model and observations

$$\nu^{obs} = s\omega + \xi + \epsilon \quad (6)$$

This equation is similar to the basic statistical model introduced by Kennedy and O'Hagan for Bayesian calibration of model outputs.<sup>11</sup> The discrepancy variable could formally depend on  $\omega$ , but we observed that the residuals between model and observations are not notably frequency dependent (Fig. 2), and the same is assumed for  $\xi$  which is considered null in average with variance  $\sigma^2$

$$\xi \sim N(0, \sigma^2) \quad (7)$$

The calibration equation depends thus on two unknown parameters  $s$  and  $\sigma$ .

## 2.2 Uncertainty propagation

The calibration model (Eq. 6) being linear with regard to uncertain variables  $s$ ,  $\xi$  and  $\epsilon$ , and in the hypothesis of normally distributed uncertainties, we can use the standard uncertainty propagation rules<sup>5</sup> to estimate the value and uncertainty on predicted vibrational frequencies:

$$\bar{\nu} = \bar{s}\omega \quad (8)$$

$$Var(\nu) = \left(\frac{\partial \nu}{\partial s}\right)_{s=\bar{s}}^2 Var(s) + \left(\frac{\partial \nu}{\partial \xi}\right)_{\xi=0}^2 Var(\xi) + \left(\frac{\partial \nu}{\partial \epsilon}\right)_{\epsilon=0}^2 Var(\epsilon) \quad (9)$$

$$= \omega^2 Var(s) + \sigma^2 + \rho^2 \quad (10)$$

The terms due to covariances have been omitted from this equation as the three variables are independent by hypothesis. In order to provide evaluated predictions of vibrational frequencies, we need therefore to estimate  $\bar{s}$ ,  $u_s^2 = Var(s)$  and  $\sigma^2$  from the calibration dataset. This is done in the next section, using Bayesian model calibration.

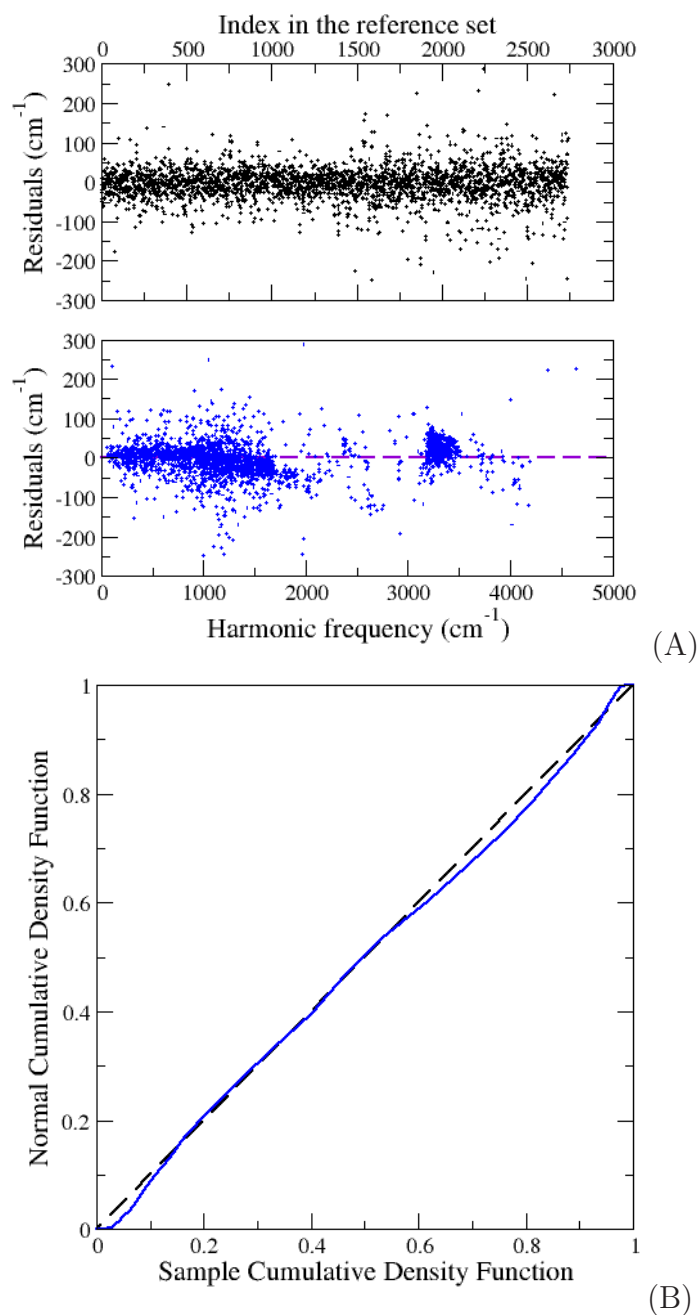


Figure 2: Residuals between calculated harmonic frequencies  $\omega_i$  and measured frequencies  $\nu_i$  for a set of vibrations extracted from the CCCBDB for the HF/6-31G\* combination of theory/basis-set (dots). (A) bottom: residuals as a function of  $\omega$ . In order to suppress the grouping effect linked with frequencies, the residuals were also plotted as a function of their order in the reference set (A)-top. (B): plot of the cumulative density function (CDF) for the residuals against a normal CDF shows that globally there is very little deviation from normality.

## 2.3 Bayesian Model Calibration (BMC)

### 2.3.1 General case

Starting from the statistical calibration model (Eq. 6), one derives the following expression for the posterior probability density function of the parameters, given a set of  $N$  measured and calculated frequencies (details are reported in Appendix A.1)

$$p(s, \sigma | \{\nu_i^{obs}, \rho_i, \omega_i\}_{i=1}^N) \propto \frac{1}{\sigma \prod_{i=1}^N (\sigma^2 + \rho_i^2)^{1/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{(\nu_i^{obs} - s\omega_i)^2}{\sigma^2 + \rho_i^2}\right) \quad (11)$$

Inferences from this pdf have to be done generally by numerical methods,<sup>12</sup> as will be applied later to zero point vibrational energies. For vibrational frequencies, we first derive an adapted, simplified, model.

### 2.3.2 The case of negligible measurement uncertainties

In the commonly met situation where the amplitude of  $\xi$  is much larger than the other sources of uncertainty ( $\sigma \gg \rho$ ), we can consider the approximate measurement model

$$\nu_i^{obs} = s\omega_i + \xi \quad (12)$$

for which the posterior pdf can be simplified to

$$p(s, \sigma | \{\nu_i^{obs}, \omega_i\}_{i=1}^N) \propto \frac{1}{\sigma^{N+1}} \exp\left(-\frac{N\gamma^2}{2\sigma^2}\right) \exp\left(-\frac{(s - \bar{s})^2 \sum_{i=1}^N \omega_i^2}{2\sigma^2}\right) \quad (13)$$

This expression is amenable to analytical derivation of the estimates for the parameters (see Appendix A.2):

- the average value for  $s$  is identical to the optimal value provided by least-squares analysis  $\bar{s} = \hat{s}$ ;
- the standard uncertainty on  $s$ ,  $u_s$  is related to the rms  $\gamma$  by the formula  $u_s = \gamma \sqrt{N / ((N - 3) \sum \omega_i^2)}$ ;
- the estimate of  $\sigma^2$  is related to  $\gamma$  according to  $\overline{\sigma^2} = \gamma^2 N / (N - 3)$ .

Using these estimates, we can establish an explicit expression for the standard uncertainty of  $\nu$ :

$$u_\nu = \gamma \sqrt{\frac{N}{N-3} \left( \frac{\omega^2}{\sum_i \omega_i^2} + 1 \right)} \quad (14)$$

Confidence intervals can be defined for prediction purpose, *e.g.* the 95% confidence interval for  $\nu$ , assuming the normality of uncertainty is given by

$$CI_{95}(\nu) = [\hat{s}\omega - 1.96 u_\nu, \hat{s}\omega + 1.96 u_\nu] \quad (15)$$

It can be seen that for large calibration sets of few hundreds of frequencies  $\sqrt{N/(N-3)} \simeq 1$  and  $\omega^2/\sum_i \omega_i^2 \ll 1$ , and thus

$$u_\nu \simeq \gamma \quad (16)$$

In such conditions, it is thus possible to derive directly confidence intervals from the the summary calibration statistics  $\hat{s}$  and  $\gamma$  provided by the reference literature.<sup>6,7,8</sup>

## 2.4 The Multiplicative Uncertainty (MU) method

Irikura *et al.*,<sup>10</sup> after a thorough analysis of the uncertainty sources, established that the major contribution to uncertainty propagation would be the uncertainty on the scaling factor  $\hat{s}$ . They estimate  $u_s$  from the weighted variance of  $s$  with weights  $a_i = \omega_i^2$ . This weighting scheme is derived in two steps: (1) they propose that the probability density function (pdf) for the scaling factor is a linear combination of pdf's for individual scaling factors in the reference set; and (2) from the comparison of the expression of the average value derived from this proposition with the least-squares solution Eq.2. This way, they obtain a standard uncertainty

$$u_s^* \simeq \left( \frac{1}{\sum \omega_i^2} \sum \omega_i^2 (s_i - \hat{s})^2 \right)^{1/2} \quad (17)$$

which can be related to  $\gamma$  by  $u_s^* \simeq \gamma \sqrt{N/\sum \omega_i^2}$ . This uncertainty is different, and generally smaller, than the dispersion of  $s$  values within the calibration set

$$\sigma_s = \left( \frac{1}{N} \sum (s_i - \hat{s})^2 \right)^{1/2} \quad (18)$$

The uncertainty on a scaled frequency is then reduced to

$$u_\nu \simeq \omega u_s^* \quad (19)$$

hence the name of "Multiplicative Uncertainty" (MU) used hereafter.

The salient feature of Eq. 19 is that uncertainty should be proportional to the calculated harmonic frequency. But, if one observes correlation plots for reference datasets (*e.g.* Fig. 1,  $\omega > 2000 \text{ cm}^{-1}$ ), this is not the case. When contrasted with the BMC approach, one understands that the multiplicative approach "absorbs", at least partially, model inadequacy in  $u_s^*$ . It is thus implicitly assumed in this approach that the model (Eq. 5) is in a statistical regime, although it is not. These points will be illustrated and discussed in the next section.

## 3 Applications and discussion

### 3.1 Vibrational frequencies

We illustrate the BMC and MU approaches on the HF/6-31G\* combination of theory/basis-set. The reference dataset, plotted in Fig. 1, has been downloaded from the NIST/CCCBDB in July 2008.<sup>9</sup>

#### 3.1.1 Calibration

Using the BMC model and estimators on the reference set, we obtain  $\hat{s} = 0.8984 \pm 0.0005$ , and  $\hat{\sigma} = 45.3 \pm 0.6 \text{ cm}^{-1}$  (Table 1), which is consistent with the value of the rms obtained by Merrick *et al.*<sup>8</sup> for the same theory/basis-set combination. For this dataset, the CCCBDB proposes  $\hat{s} = 0.899 \pm 0.025$ . We cross-checked this value using Eq. 17 (Table 1). The standard uncertainty on  $\hat{s}$  evaluated by both methods differ thus by a factor 50, which is expected to have noticeable effect on prediction uncertainty. In order to visualize this effect, we plotted the 95 percent confidence intervals in both cases (Fig. 3). It is immediately visible that the the MU approach has a tendency to underestimate uncertainty at low frequencies and to overestimate it at high frequencies. In comparison, the BMC approach is more balanced.

	Summary stat.			MU		BMC		
	$\hat{s}$	$\sigma_s$	$\gamma(\text{cm}^{-1})$	$u_s$	%CI <sub>95</sub>	$u_s$	$\sigma(\text{cm}^{-1})$	%CI <sub>95</sub>
<b>All frequencies (2738)</b>								
Full set	0.8984	0.069	45.3	0.024	-	0.0005	45.3	-
Calibration set	0.8986		45.3	0.024	-	0.0007	45.3	-
Validation set	-	-	-	-	83.0	-	-	94.6
<b>Frequencies between 3180 and 3500 cm<sup>-1</sup> (479)</b>								
Full set	0.9050	0.0087	28.7	0.0087	-	0.0004	28.8	-
Calibration set	0.9052		23.3	0.0071	-	0.0005	23.4	-
Validation set	-	-	-	-	95.0	-	-	95.4

Table 1: Statistical estimates and validation for MU and BMC models for vibrational frequencies extracted from the CCCBDB for the HF/6-31G\* combination of theory/basis-set.

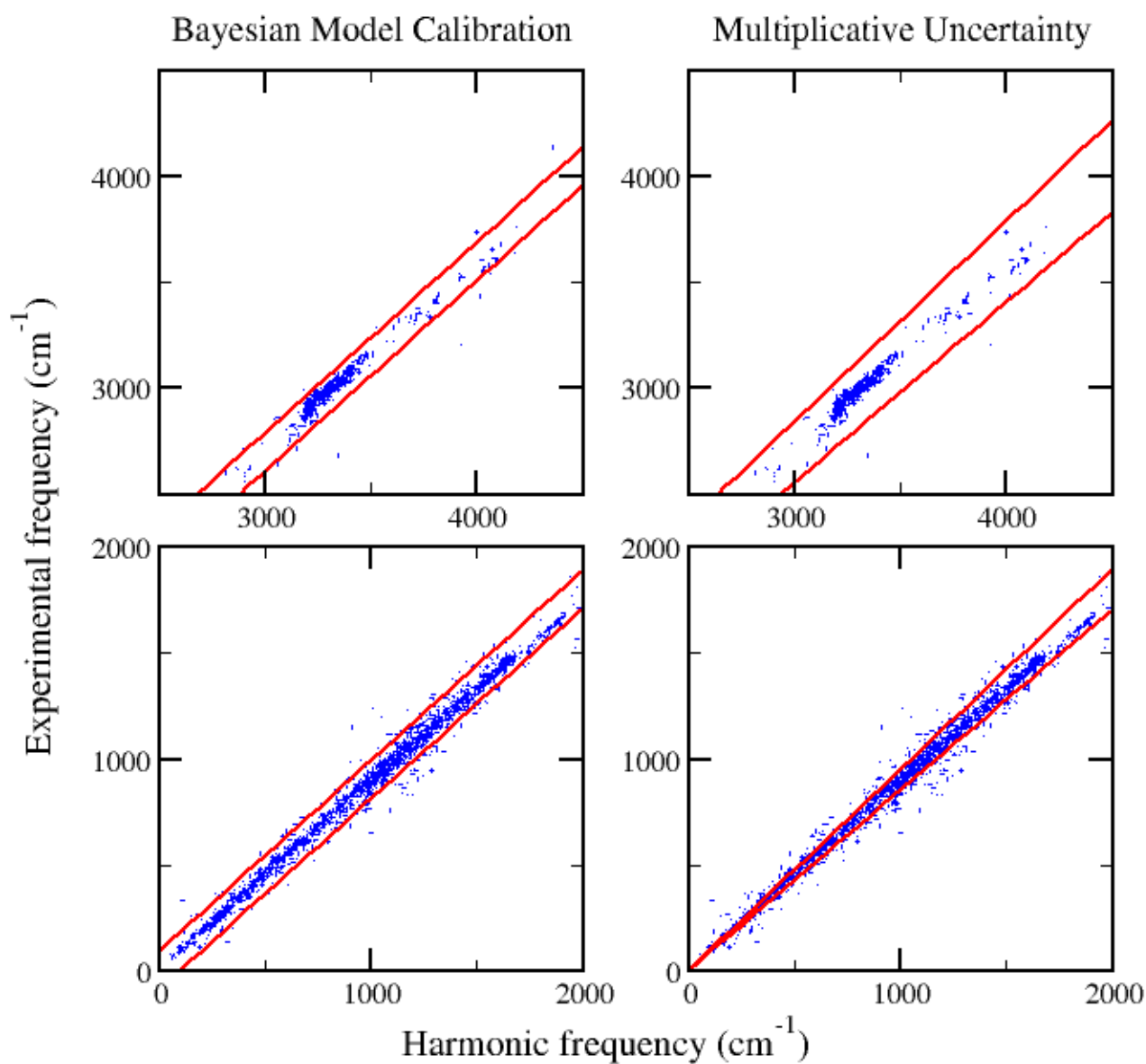


Figure 3: Confidence intervals at 95% level calculated with Bayesian Model Calibration (left column) and Multiplicative Uncertainty (right column) methods, in two representative frequency regions from the calibration dataset for the HF/6-31G\* combination of theory/basis-set (dots).

### 3.1.2 Validation

In order to comfort the observations of the previous section (3.1.1), we split the calibration dataset in two subsets by using the odd indexes for calibration, and the even ones for validation (the frequencies being ordered per molecule, this provides a quasi-random selection with regard to frequency value). In this case, one gets slightly different values of the parameters, as reported in Table 1. Using these values, we generate 95% confidence intervals and calculate the frequency of inclusion of the experimental values of the validation subset within these intervals (Fig. 3). For a consistent estimator, one should find a frequency close to 95%. BMC succeeds for 94.7%, whereas the MU model succeeds for only 83% of the frequencies in the validation set (Table 1). Considering the size of the samples, the difference is significant, and the consistency of the multiplicative uncertainty approach in this context can be questioned.

### 3.1.3 Test on a restricted frequency scale

As stated in Paper I, "to apply the fractional bias correction, it is important to select a class of frequencies similar to the ones to be estimated". For instance, if one selects in the reference set only those frequencies between 3180 and 3500  $\text{cm}^{-1}$ , one gets a much more uniform picture than with the full reference set. The calibration procedure was done with this limited set of 479 frequencies, providing  $\hat{s} = 0.9050 \pm 0.0087$  (Table 1). In this case, we note that the uncertainty factor for  $s$  is practically identical to the standard deviation calculated from the sample (0.00869 vs. 0.00871):  $u_s^* \simeq \sigma_s$ . Due to the restricted frequency range, one has indeed  $\omega_i^2 / \sum_i \omega_i^2 \simeq 1/N$  (*cf.* Eqs 17 and 18).

The restricted set has been split in two using index parity, as before. The scaling factor obtained by MU from the calibration subset is now  $\hat{s} = 0.9052 \pm 0.0071$ , and 95.0% of the validation frequencies fall within the 95% confidence interval. This result is now identical to the one obtained with BMC (Table 1), the confidence intervals obtained by both methods being indistinguishable.

It appears thus that in restrictive conditions, the MU method can be valid for reference sets where the individual scaling factors are uniformly distributed with regard to the harmonic frequencies. In such case however, the uncertainty is reduced to a conventional unweighted

standard deviation.

### 3.1.4 Uncertainty propagation

The relative importance of both factors  $u_s^2$  and  $\sigma^2$  ( $\rho = 0$  in this example) in Eq. 10 can be evaluated on the example of a calculated harmonic frequency in the higher range  $\omega = 3000$   $\text{cm}^{-1}$  (Table 2). In this case, the uncertainty on the scaling factor contributes only to one thousandth of the total prediction variance.

For any practical purpose, the uncertainty on the scaling factor of vibrational frequencies can therefore be neglected. The uncertainty on  $\sigma$  is also much too small to be relevant for confidence intervals calculation. One can thus safely apply the uncertainty propagation formula (Eq. 16), using the rms provided by most reference articles dealing with scaling factors calibration.<sup>6,7,8</sup>

## 3.2 Zero Point Vibrational Energies

We consider ZPVE as a an additional test because the reference datasets are considerably smaller than for the vibrational frequencies (e.g. 39 molecules in the Z1 set of Merrick *et al.*<sup>8</sup>), which is expected to enhance the role of the uncertainty on the scaling factor. In the absence of a systematic review of measurement errors for ZPVE of polyatomic molecules, we consider in the following that they can be neglected. The uncertainties reported by Irikura<sup>14</sup> for diatomic molecules are indeed very small (on the order of  $0.01$   $\text{cm}^{-1}$ ), but transposition to larger molecules is not straightforward.

Using our measurement model and estimators on the reference set, one gets  $\hat{s} = 0.9135 \pm 0.0027$  and  $\hat{\sigma} = 0.73 \pm 0.09$   $\text{kJ mol}^{-1}$  (Table 3), which is consistent with the rms obtained by Merrick *et al.*<sup>8</sup> for the HF/6-31G\* theory/basis-set combination. Relative uncertainties on these parameters have been increased by one order of magnitude, compared with the vibrational frequencies case. The validation exercise shows once more that the Multiplicative Uncertainty model fails to provide correct confidence intervals, with a score of only 0.63 for  $\text{CI}_{95}$ .

In such a case of small reference dataset, it is interesting to check if the approximate formula (Eq. 16) for uncertainty propagation which was validated for large sets of vibrational frequencies begins to break down, *i.e.* the contribution of the multiplicative term involving  $u_s$  stays

negligible or not, for the larger ZPVE values. For instance, if one considers a calculated ZPVE of  $100 \text{ kJ mol}^{-1}$ , one has  $u_\nu = \sqrt{(100 * 0.0027)^2 + 0.73^2} = 0.78 \text{ kJ mol}^{-1}$ , to be compared to  $\gamma = 0.71 \text{ kJ mol}^{-1}$ . It is to be noted also that the uncertainty on  $\sigma$  might contribute at the same level,  $u_\sigma = 0.09 \text{ kJ mol}^{-1}$ . Taking all uncertainty sources into account through Eq. 51 by Monte Carlo Uncertainty Propagation, one gets  $u_\nu = 0.77 \text{ kJ mol}^{-1}$ . Apparently, the fluctuations of  $\sigma$  do not have an effect and only the uncertainty on the scaling factor has a noticeable effect, albeit quite small.

In the same conditions, for the combination B3LYP/6-31G\*, one gets  $\gamma = 0.423 \text{ kJ mol}^{-1}$  and  $u_\nu = 0.472 \text{ kJ mol}^{-1}$ , to be compared with an exact value of  $u_\nu = 0.466 \text{ kJ mol}^{-1}$  (Table 3). There is globally only a 10% increase compared to the rms  $\gamma$ . In this range of ZPVEs,  $\gamma$  still provides a good approximation of the uncertainty factor (Table 2). However, the amplitude of the discrepancy between  $\gamma$  and  $u_\nu$  will probably increase with the size of the molecule. In consequence, for uncertainty propagation with ZPVEs, notably for large molecules, it would be safer to use the full UP formula (Eq. 14), involving the multiplicative uncertainty factor. Compilations of scaling factors should thus report the easily calculated value of  $u_s = \gamma \sqrt{N / ((N - 3) \sum \omega_i^2)}$ , in addition to the rms  $\gamma$ .

### 3.2.1 Non-negligible experimental uncertainties

When the measurement uncertainty becomes comparable to the rms, model discrepancy should be vanishing, and confidence intervals for prediction should include the measurement uncertainty, *i.e.*  $u_\nu^2 = \omega^2 u_s^2 + \sigma^2 + \rho^2$ . In the absence of an exhaustive compilation of experimental uncertainties on measured ZPVE, we performed simulations assuming a uniform uncertainty distribution over the full dataset. In order to test the sensitivity of the model parameters to this uncertainty, we repeated the estimations of previous section, using Eq. 27, for different values of  $\rho$  between  $0.1$  and  $1.0 \text{ kJ mol}^{-1}$ . The results are reported in Fig. 4.

As expected from the properties of the posterior pdf, the average/optimal value of the scaling factor is insensitive to the amplitude of  $\rho$ . Moreover, we observe only a slight increase of  $u_s$  from  $0.002$  to  $0.004$ . A transition from a constant  $u_s$ , defined by the  $\rho = 0$  limit, to a linear increase consistent with the weighted least squares limit (Eq. 50) is observed around  $\rho = \gamma$ , where both limits intersect. A closer look shows that the transition occurs indeed at values of

$\rho$  slightly smaller than  $\gamma$ , in a region ( $\rho \simeq 0.35$ ) where  $u_s$  displays a minimum.

The evolution of  $\sigma$  is more dramatic: it displays a sharp decrease and falls down to zero as soon as the measurement uncertainty reaches and overpasses the value of the rms  $\gamma$ . For values of  $\rho$  below 0.25,  $\sigma$  obeys to the  $\sigma^2 + \rho^2 = \gamma^2$  law (represented as a dashed line in Fig. 4), but the calculated decrease becomes much faster in the transition zone. The uncertainty on  $\sigma$  increases notably in the transition region observed for  $s$ .

In the limit of large experimental uncertainties, the uncertainty propagation formula can be written as

$$u_\nu^2 = \omega^2 u_s^2 + \rho^2 \quad (20)$$

$$= \rho^2 \left( 1 + \omega^2 / \sum \omega_i^2 \right) \quad (21)$$

In this region of large measurement uncertainty, the inadequacy variable  $\xi$  becomes useless, as the calibration with the scaling factor alone enters a statistical regime.

## 4 Conclusion

A reanalysis of scaling factors calibration in the aim of providing quantified predictions shows that for large validation sets of accurate data, as used for vibrational frequencies, faithful prediction uncertainties can be derived from the standardly reported optimal scaling factor and rms. For much smaller datasets of a few dozens of data, as in the case of ZPVEs, uncertainty on the scaling factor should also be reported. The following limit formulas have been validated and are proposed for uncertainty propagation from a given calculated harmonic frequency  $\omega$ :

- large calibration sets of precise data ( $\rho \ll \gamma$ ):  $u_\nu(\omega) = \gamma$ ;
- small calibration sets of precise data ( $\rho \ll \gamma$ ):  $u_\nu(\omega) = \gamma \sqrt{\frac{N}{N-3} (1 + \omega^2 / \sum_i \omega_i^2)}$ ;
- sets of data with large uniform uncertainty ( $\rho \geq \gamma$ ):  $u_\nu(\omega) = \rho \sqrt{1 + \omega^2 / \sum_i \omega_i^2}$ .

A general estimation framework, based on Bayesian Model Calibration, has been defined for those cases where the limit conditions defined above are not met.

The Multiplicative Uncertainty method proposed by Irikura *et al.*<sup>10</sup> ( $u_\nu \simeq \omega \gamma \sqrt{N / \sum \omega_i^2}$ ) has been shown to be inconsistent when large frequency ranges are considered. It could not

	$\omega$ (HF/6-31G*)	$\omega^2 u_s^2$	$\sigma^2$	$\nu \pm u_\nu$
Freq.	3000	2.25	2052.09	$2695 \pm 45 \text{ cm}^{-1}$
ZPVE	$100 \text{ kJ mol}^{-1}$	0.029	0.194	$98.12 \pm 0.47 \text{ kJ mol}^{-1}$

Table 2: Uncertainty propagation with calibrated model for a set of 2738 vibrational frequencies extracted from the CCCBDB for the HF/6-31G\* combination of theory/basis-set and for a set of 39 ZPVE of the Z1 set for the B3LYP/6-31G\* combination.

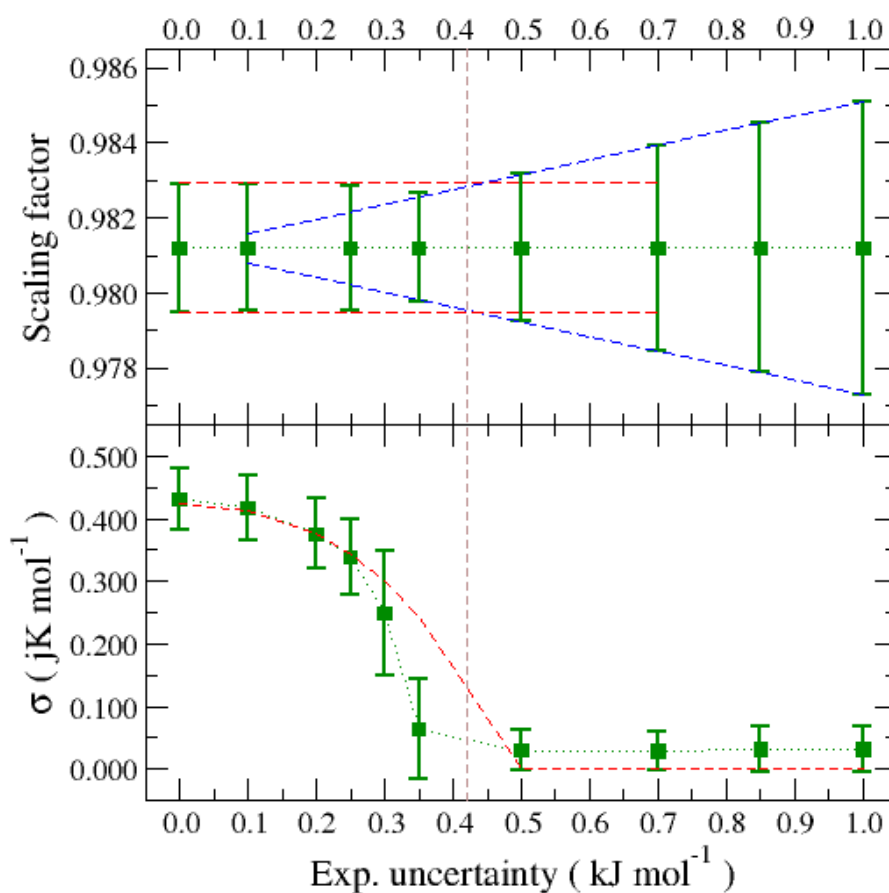


Figure 4: Evolution of measurement model parameters with the amplitude of an hypothetical uniform experimental measurement uncertainty  $\rho$  on ZPVE; B3LYP/6-31G\* combination of theory/basis-set (green squares with error bars). The brown vertical dashed line indicates the value of the rms  $\gamma$ . Top panel: the red dashed lines represent the  $1\sigma$  confidence interval in the limit of null experimental uncertainty; the blue dashed line represent the  $1\sigma$  confidence interval in the weighted least squares limit. Bottom panel: the red dashed line represents the  $\sigma^2 + \rho^2 = \gamma^2$  law, truncated to positive values of  $\sigma$ .

	Summary stat.			MU		BMC		
	$\hat{s}$	$\sigma_s$	$\gamma(\text{kJ mol}^{-1})$	$u_s$	%CI <sub>95</sub>	$u_s$	$\sigma(\text{kJ mol}^{-1})$	%CI <sub>95</sub>
<b>HF/6-31G*</b>								
Full set	0.9135	0.0607	0.71	0.0161	-	0.0027	0.73±0.09	-
Calibration set	0.9078		0.77	0.0214		0.0052	0.83±0.14	
Validation set					0.63			0.95
<b>B3LYP/6-31G*</b>								
Full set	0.9812	0.0375	0.42	0.0103	-	0.0017	0.44±0.05	-
Calibration set	0.9825		0.45	0.0134		0.0032	0.48±0.08	
Validation set					0.68			1.00

Table 3: Statistical estimates and validation for MU and BMC models for a set of 39 ZPVEs of the Z1 set for the HF/6-31G\* and B2LYP/6-31G\* combinations of theory/basis-set.

be recovered by the Bayesian approach, except when the dataset spans a restricted frequency range, in which case it is reduced to a trivial, unweighted, standard deviation. The use of the scaling factor uncertainties as reported presently (November 2008) in the CCCBDB<sup>9</sup> cannot therefore be recommended.

I would like to stress out that the validity of the formulas for uncertainty propagation depends to some extent on the normality of the residuals  $\nu_i - \hat{s}\omega_i$  of the linear regression. Inspection of histograms of residuals (see *e.g.* Fig.1 in Ref.<sup>7</sup>) shows that this is not always the case. The customary approach to choose an optimal theory/basis-set combination is to assess their performance by the rms alone, maybe weighted by computational cost considerations.<sup>6,7,8</sup> One could also consider a "normality criterion" in order to reject theory/basis-set combinations providing non-normal residuals and from which the summary statistics cannot be used faithfully for uncertainty propagation. Analysis of restricted ranges of data as presently done by some authors for vibrations<sup>8</sup> is one way to improve the normality of residuals.

Semi-empirical correction of ab-initio results by scaling is presently very common and efficient for many observables. It certainly would be a large step towards the general applicability of uncertainty propagation, if statistically pertinent estimators were systematically reported in the literature devoted to the calibration of these correction models.

## Acknowledgments

I would like to thank Prof. Leo Radom for providing me with the Z1 ZPVE dataset. B. Lévy is warmly acknowledged for helpful discussions.

## A Appendix

### A.1 Bayesian analysis of scaling factor calibration model

We consider the measurement model

$$\nu_i^{obs} = s\omega_i + \xi + \epsilon_i \quad (22)$$

where  $\epsilon_i \sim N(0, \rho_i^2)$  is the measurement uncertainty of  $\nu_i^{obs}$ , and  $\xi \sim N(0, \sigma^2)$  is a discrepancy function between the linear model and the observations. This model has two unknown parameters,  $s$  and  $\sigma$ , to be estimated on a calibration dataset consisting of  $N$  calculated harmonic frequencies  $\{\omega_i\}_{i=1}^N$ , and their corresponding observed frequencies  $\{\nu_i^{obs}\}_{i=1}^N$ .

In the Bayesian data analysis framework,<sup>15,12</sup> all information about parameters can be obtained from the joint posterior pdf  $p(s, \sigma | \{\nu_i^{obs}, \rho_i, \omega_i\}_{i=1}^N)$ . In order to simplify the notations, we will omit in the following the list indices when they are not necessary. This pdf is obtained through Bayes theorem

$$p(s, \sigma | \{\nu^{obs}, \rho, \omega\}) \propto p(\{\nu\} | s, \sigma, \{\rho, \omega\}) p(s, \sigma) \quad (23)$$

where  $p(\{\nu^{obs}\} | s, \sigma, \{\rho, \omega\})$  is the likelihood function and  $p(s, \sigma)$  is the prior pdf.

As the difference between observation and corrected frequency has a normal distribution

$$\nu_i^{obs} - s\omega_i \sim N(0, \sigma^2 + \rho_i^2) \quad (24)$$

the likelihood function for a single observed frequency is

$$p(\nu_i^{obs} | s, \sigma, \rho_i, \omega_i) = (2\pi (\sigma^2 + \rho_i^2))^{-1/2} \exp\left(-\frac{1}{2} \frac{(\nu_i^{obs} - s\omega_i)^2}{\sigma^2 + \rho_i^2}\right) \quad (25)$$

Considering that all frequencies are measured independently (with uncorrelated uncertainty) the joint likelihood is the product of the individual ones, *i.e.*

$$p(\{\nu^{obs}\} | s, \sigma, \{\rho, \omega\}) = \prod_{i=1}^N (2\pi (\sigma^2 + \rho_i^2))^{-1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{(\nu_i^{obs} - s\omega_i)^2}{\sigma^2 + \rho_i^2}\right) \quad (26)$$

As there is a priori no correlation between  $s$  and  $\sigma$ , we use a factorized prior pdf  $p(s, \sigma) = p(s)p(\sigma)$ . In the absence of a priori quantified information on  $s$ , a uniform pdf  $p(s) = cte$  is used. For  $\sigma$ , we have to consider a positivity constraint, and we use a Jeffrey's prior,  $p(\sigma) \propto \sigma^{-1}$ . The posterior pdf is finally defined up to a proportionality factor which is irrelevant for the following developments

$$p(s, \sigma | \{\nu^{obs}, \omega, \rho\}) \propto \sigma^{-1} \prod_{i=1}^N (\sigma^2 + \rho_i^2)^{-1/2} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{(\nu_i^{obs} - s\omega_i)^2}{\sigma^2 + \rho_i^2}\right) \quad (27)$$

## A.2 Case of negligible measurement uncertainties

For the analysis of vibrational frequencies, it is generally considered that experimental uncertainties are negligible ( $\rho_i = 0$ ). The general expression for the posterior pdf (Eq. 27) can be simplified accordingly to

$$p(s, \sigma | \{\nu^{obs}, \omega\}) \propto \sigma^{-N-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N (\nu_i^{obs} - s\omega_i)^2\right) \quad (28)$$

Using the identity

$$\sum_{i=1}^N (\nu_i^{obs} - s\omega_i)^2 = (s - \hat{s})^2 \sum \omega_i^2 + N\gamma^2 \quad (29)$$

it can be written in a convenient factorized form

$$p(s, \sigma | \{\nu^{obs}, \omega\}) \propto \sigma^{-N-1} \exp\left(-\frac{N\gamma^2}{2\sigma^2}\right) \exp\left(-\frac{(s - \bar{s})^2 \sum \omega_i^2}{2\sigma^2}\right) \quad (30)$$

from which we can derive analytical estimates for the parameters and their uncertainties.

### A.2.1 Estimation of $s$

The marginal density for  $s$  is obtained by integration over  $\sigma$

$$p(s | \{\nu^{obs}, \omega\}) = \int_0^\infty d\sigma p(s, \sigma | \{\nu^{obs}, \omega\}) \quad (31)$$

$$\propto \int_0^\infty d\sigma \sigma^{-N-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^N (\nu_i^{obs} - s\omega_i)^2\right) \quad (32)$$

$$\propto \left(\sum_{i=1}^N (\nu_i^{obs} - s\omega_i)^2\right)^{-N/2} \quad (33)$$

which can be rewritten as

$$p(s | \{\nu^{obs}, \omega\}) \propto \left(1 + \frac{(s - \hat{s})^2 \sum \omega_i^2}{N\gamma^2}\right)^{-N/2} \quad (34)$$

which has the shape of a Student's distribution<sup>16</sup>

$$\text{Stt}(x) \propto \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \quad (35)$$

Posing  $n = N - 1$  and  $x^2 = (N - 1)/N (s - \hat{s})^2 \sum \omega_i^2 / \gamma^2$ , we can directly use the properties of the Student's distribution

$$\mathbb{E}[x] = 0; \text{Var}[x] = n/(n - 2) \quad (36)$$

to derive

$$\bar{s} = \hat{s} \quad (37)$$

$$u_s^2 = \gamma^2 \frac{N}{(N - 3) \sum \omega_i^2} \quad (38)$$

### A.2.2 Estimation of $\sigma$

The marginal density for the standard uncertainty of the stochastic variable  $\xi$  is

$$p(\sigma | \{\nu^{obs}, \omega\}) = \int_{-\infty}^{\infty} ds p(s, \sigma | \{\nu^{obs}, \omega\}) \quad (39)$$

$$\propto \frac{1}{\sigma^{N+1}} \exp\left(-\frac{N\gamma^2}{2\sigma^2}\right) \int_{-\infty}^{\infty} ds \exp\left(-\frac{(s - \hat{s})^2 \sum \omega_i^2}{2\sigma^2}\right) \quad (40)$$

$$\propto \frac{1}{\sigma^N} \exp\left(-\frac{N\gamma^2}{2\sigma^2}\right) \quad (41)$$

Using the formula

$$\int_0^{\infty} dx x^{-n} e^{-a/x^2} = \frac{1}{2} \Gamma\left(\frac{n-1}{2}\right) / a^{(n-1)/2} \quad (42)$$

one obtains readily the following estimates

$$\hat{\sigma} = \gamma \quad (43)$$

$$\bar{\sigma} = \gamma \sqrt{\frac{N}{2} \frac{\Gamma[(n-2)/2]}{\Gamma[(n-1)/2]}} \quad (44)$$

$$\overline{\sigma^2} = \frac{N}{N-3} \gamma^2 \quad (45)$$

$$u_{\sigma} = \gamma \sqrt{\frac{N}{N-3} - \frac{N}{2} \left(\frac{\Gamma[(n-2)/2]}{\Gamma[(n-1)/2]}\right)^2} \quad (46)$$

## A.3 Case of very large measurement uncertainties

When model discrepancy is negligible compared to measurement uncertainties, *i.e.* when the standard linear model is in a statistical regime, one recovers standard statistical analysis, the

Bayesian analog to weighted least squares. The posterior pdf for  $s$  is then

$$p(s | \{\nu^{obs}, \omega, \rho\}) \propto \prod_{i=1}^N \rho_i^{-1} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{(\nu_i^{obs} - s\omega_i)^2}{\rho_i^2}\right) \quad (47)$$

from which one obtains

$$\hat{s} = \sum (\omega_i \nu_i^{obs} / \rho_i^2) / \sum (\omega_i^2 / \rho_i^2) \quad (48)$$

$$u_s^2 = 1 / \sum (\omega_i^2 / \rho_i^2) \quad (49)$$

For uniform experimental uncertainty over the dataset, the scaling factor uncertainty varies linearly with  $\rho$

$$u_s = \rho / \sqrt{\sum \omega_i^2} \quad (50)$$

## A.4 Uncertainty propagation

In the Bayesian framework, the posterior pdf  $p(s, \sigma | \{\nu^{obs}, \rho, \omega\})$  can be used to estimate the uncertainty of predicted frequencies

$$p(\nu' | \rho', \omega', \{\nu^{obs}, \rho, \omega\}) = \int ds d\sigma p(\nu' | s, \sigma, \rho', \omega') p(s, \sigma | \{\nu^{obs}, \rho, \omega\}) \quad (51)$$

where

$$p(\nu' | s, \sigma, \rho', \omega') \propto (\sigma^2 + \rho'^2)^{-1/2} \exp\left(-\frac{(\nu' - s\omega')^2}{2(\sigma^2 + \rho'^2)}\right) \quad (52)$$

results from the stochastic model (Eq. 12). This integral has generally to be evaluated numerically.

## References

- [1] John A. Pople. Nobel lecture: Quantum chemical models. *Rev. Mod. Phys.*, 71(5):1267–1274, 1999.
- [2] Carmen Pancerella, James D. Myers, Thomas C. Allison, Kaizar Amin, Sandra Bittner, Brett Didier, Michael Frenklach, Jr. William H. Green, Yen-Ling Ho, John Hewson, Wendy Koegler, Carina Lansing, David Leahy, Michael Lee, Renata McCoy, Michael Minkoff, Sandeep Nijsure, Gregor von Laszewski, David Montoya, Reinhardt Pinzon, William Pitz, Larry Rahn, Branko Ruscic, Karen Schuchardt, Eric Stephan, Al Wagner, Baoshan Wang, Theresa Windus, Lili Xu, and Christine Yang. Metadata in the Collaboratory for Multi-Scale Chemical Science. In *2003 Dublin Core Conference: Supporting Communities of Discourse and Practice-Metadata Research and Applications*, Seattle, WA, 28 September – 2 October 2003.
- [3] M. Frenklach. Transforming data into knowledge. Process Informatics for combustion chemistry. *Proceedings of the Combustion Institute*, 31:125–140, 2007.
- [4] K.K. Irikura, R.D. Johnson III, and R.N. Kacker. Uncertainty associated with virtual measurements from computational quantum chemistry models. *Metrologia*, 41:369–375, 2004.
- [5] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML. Guide to the expression of uncertainty in measurement (GUM). Technical Report ISBN 92-67-10188-9, International Organization for Standardization, Geneva, 1995. Corrected and reprinted.
- [6] A.P. Scott and L. Radom. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Möller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.*, 100(41):16502–16513, 1996.
- [7] Ming Wah Wong. Vibrational frequency prediction using density functional theory. *Chemical Physics Letters*, 256(4-5):391–399, 1996.
- [8] J.P. Merrick, D. Moran, and L. Radom. An evaluation of harmonic vibrational frequency scale factors. *J. Phys. Chem. A*, 111(45):11683–11700, 2007.

- [9] R.D. Johnson III. NIST Computational Chemistry Comparison and Benchmark Database, Release 14; NIST Standard Reference Database Number 101, September 2006. <http://cccbdb.nist.gov/>.
- [10] K.K. Irikura, R.D. Johnson, and R.N. Kacker. Uncertainties in scaling factors for ab initio vibrational frequencies. *J. Phys. Chem. A*, 109(37):8430–8437, 2005.
- [11] Marc C. Kennedy and Anthony O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.
- [12] P. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, 2005.
- [13] BIPM, IEC, IFCC, ISO, IUPAC, IUPAP, and OIML. Evaluation of measurement data - Supplement 1 to the GUM: propagation of distributions using a Monte Carlo method. Technical report, BIPM, September 2006. Final draft.
- [14] Karl K. Irikura. Experimental vibrational zero-point energies: Diatomic molecules. *J. Phys. Chem. Ref. Data*, 36(2):389–397, 2007.
- [15] D.S. Sivia. *Data Analysis: A Bayesian Tutorial*. Clarendon (Oxford Univ. Press), Oxford, 1996.
- [16] M. Evans, N. Hastings, and B. Peacock. *Statistical Distributions*. Wiley-Interscience, 3rd edition, 2000.