

The parameters uncertainty inflation fallacy

Supporting Information

Pascal PERNOT*

*Laboratoire de Chimie Physique,
UMR8000, CNRS / Univ. Paris-Sud,
F-91405 Orsay, France*

E-mail: pascal.pernot@u-psud.fr

Contents

The data and codes to reproduce two applications of the article are provided:

- R scripts¹ used for data extraction, cleanup and treatment of vibrational frequency data;
- R scripts¹ and data for the analysis of temperature-dependent Ar viscosity data.

Example 1

Two scripts are provided in `Example1/code`:

- `1_scrap_CCCBDB.R`: This script reads vibrational frequency data from the CCCBDB² for a specified level of theory/basis-set. The data are plotted and should be inspected for aberrant points (Fig. 1). The names of the molecules for which aberrant points are observed should be listed in variable `probList` and the script run again, until the plot(s) seems satisfactory. The cleaned data are saved in a file in the `Example1/data` directory for further treatment. **Important:** the observed issues should also be reported to the CCCBDB curator through the error form at <http://cccbdb.nist.gov/errorformx.asp>.

Command: `R CMD BATCH code/1_scrap_CCCBDB.R` in the `Example1` directory, or run interactively in `RStudio`.

- `2_generate_figures.R`: This script reads the data generated by the first script and generates the figures used for the article in the `Example1/figures` directory.

Command: `R CMD BATCH code/2_generate_figures.R`

The provided code is configured to run with `method = 6` (CCD) and `basis = 1` (6-31G*). For other combinations, the adequate codes have to be found in the CCCBDB at <http://cccbdb.nist.gov/vibscalejustx.asp> (hover the mouse over a table cell and note the codes appearing in the URL).

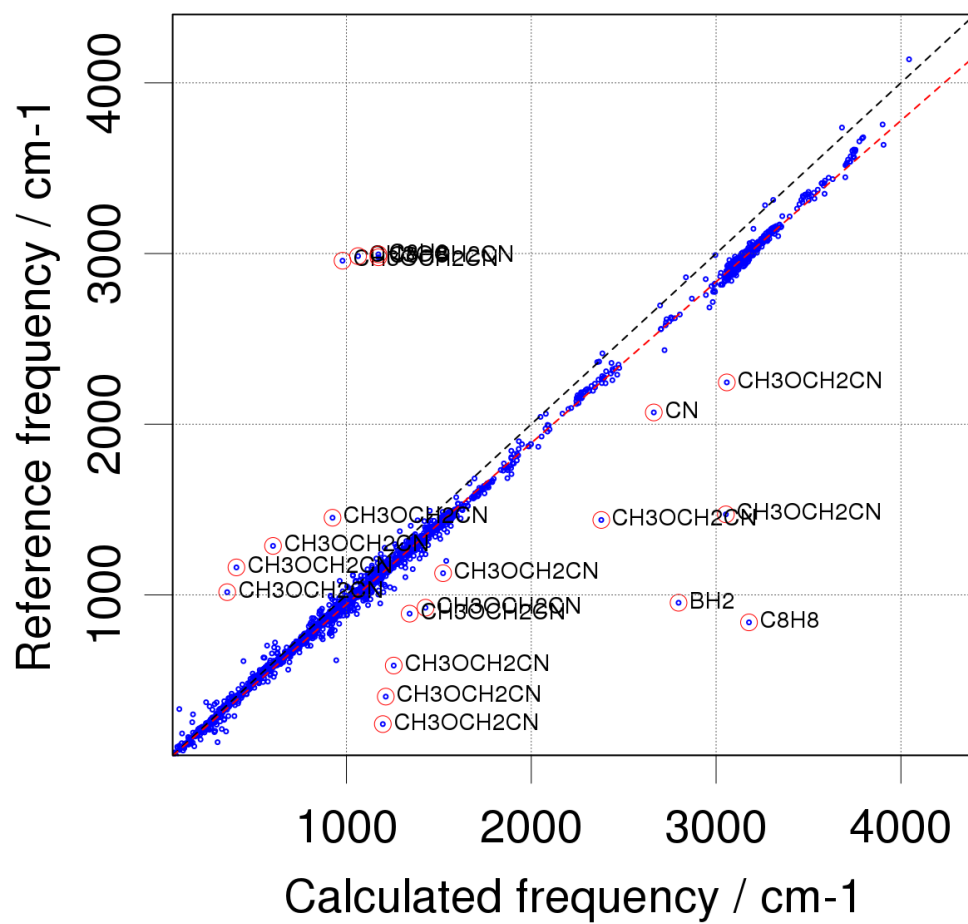


Figure 1: Figure output from `1_scrap_CCCBDB.R` showing the labelled aberrant points in the original CCD/6-31G* dataset. Errors include wrong symmetry attributions for CH₃OCH₂CN and C₈H₈.

Example 3

The main script is located in `Example3/code`. To run it, launch `R CMD BATCH code/Ar.R` in the `Example3` directory, or run interactively in `RStudio`.

The code might take some time to run. It takes advantage of multicore architectures to run (4) parallel Markov chains. The progress can be followed in the `Example3/simulation/Ar/diagnostics` directory. Note that, due to the stochastic sampling method, the numerical results might be slightly different from the ones in the article.

The results are generated in `Example3/simulation/Ar`:

- files with extension `.rda` contain the outputs of the models and can be loaded to further processing. To rerun one of the models, delete the corresponding file.
- directory `diagnostics` contains diagnostic plots and summaries to check the sample generation.
- directories `figures` and `tables` contain the data and plots from which the article's results are drawn.

Session Info

R version 3.3.0 (2016-05-03)

Platform: x86_64-redhat-linux-gnu (64-bit)

Running under: CentOS Linux 7 (Core)

locale:

[1] LC_CTYPE=fr_FR.utf8 LC_NUMERIC=C LC_TIME=fr_FR.utf8 LC_COLLATE=fr_FR.utf8

[5] LC_MONETARY=fr_FR.utf8 LC_MESSAGES=fr_FR.utf8 LC_PAPER=fr_FR.utf8 LC_NAME=C

[9] LC_ADDRESS=C LC_TELEPHONE=C LC_MEASUREMENT=fr_FR.utf8 LC_IDENTIFICATION=C

attached base packages:

[1] tcltk parallel stats graphics grDevices utils datasets methods base

other attached packages:

[1] aplpack_1.3.0 rstan_2.10.1 StanHeaders_2.10.0-2 ggplot2_2.1.0 knitr_1.13

loaded via a namespace (and not attached):

[1] Rcpp_0.12.5 codetools_0.2-14 digest_0.6.9 grid_3.3.0 plyr_1.8.4 gtable_0.2.0

[7] magrittr_1.5 stats4_3.3.0 evaluate_0.9 scales_0.4.0 KernSmooth_2.23-15 highr_0.6

[13] stringi_1.1.1 labeling_0.3 tools_3.3.0 stringr_1.0.0 munsell_0.4.3 inline_0.3.14

[19] colorspace_1.2-6 gridExtra_2.2.1

References

- (1) R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria, 2015.
- (2) Johnson III, R. NIST Computational Chemistry Comparison and Benchmark Database, Release 14; NIST Standard Reference Database Number 101. 2006; <http://cccbdb.nist.gov/>.