**Atmospheric Chemistry and Physics Discussions**

Interactive Comment

# *Interactive comment on* "Technical Note: Review of methods for linear least-squares fitting of data and application to atmospheric chemistry problems" *by* C. A. Cantrell

**T. Brauers**

th.brauers@fz-juelich.de

Received and published: 20 April 2008

The manuscript by C.A. Cantrell elucidates some of the problems which appear when analyzing measured data by linear regression. This is important since many papers in atmospheric sciences apply linear regression without clear justification. However, I think that important aspects are not considered in the manuscript. Three missing points I will explicate here.

1. The manuscript lacks a clear discussion of the errors, both as input and as output. It is important to mention, that the weights must be based on statistical errors (precision) only, not on accuracy. Contributions from the accuracy (calibration

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

error) to the weights can significantly change the estimated slope and intercept for a given dataset and therefore false conclusion could be drawn.

It is mentioned that the regression procedures can calculate errors of slope and intercept. However, these are not reported for the data presented in the paper (i.e. Table 2 and 3). Especially for intercomparison of instruments it is vital to compare the slope of the regression (with its error) to the combined (a-priori) accuracies of both instruments (e.g. see discussion in Brauers et al., 1996).

2. The text book by Press et al. (1992) contains in Chapter 15.1 and 15.2 a clear and very practical discussion of maximum likelihood estimation (by minimizing $\chi^2$) to obtain regression parameters. Press et al. discuss how $\chi^2$ can be used to also assess the quality of the regression (*goodness-of fit*, $q$) by comparing the value of $\chi^2$ to its probability distribution for the given degree of freedom. This is a general feature of any regression model which uses correct statistical weights. On the contrary, if no statistical errors are available, the precision of the measurements can be estimated assuming a linear relation.

Press et al. (in Chap. 15.3) also provide a procedure (`fitexy`) which allows the estimation of the linear regression parameters for data sets where both coordinates have considerable contributions to the statistical error of the single data points. This routine is a fast and easy to use replacement for any standard linear regression. It provides the estimated $\chi^2$ at the minimum and the goodness-of-fit parameter $q$.

3. The use of `fitexy` routine was suggested by Brauers and Finlayson-Pitts (1997) for the analysis of relative rate experiments. In the 1997 paper, the authors performed a Monte-Carlo simulation of different distributions for the statistical errors of the coordinates and show the systematic influence on the relative rates calculated. A computer program was provided with the publication and is widely used by a number of different groups when analyzing relative rate data or laboratory

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

| | | | | | | |
|---|---|---|---|---|---|---|
| err(X) | data[a] | 0.00 | data[a] | 0.00 | 0.573 | 0.278 |
| err(Y) | data[a] | data[a] | 0.00 | 0.316 | 0.00 | 0.278 |
| $b$ | -0.480 | -0.611 | -0.631 | -0.540 | -0.566 | -0.546 |
| $\sigma(b)$ | 0.057 | 0.030 | 0.009 | 0.042 | 0.044 | 0.042 |
| $a$ | 5.478 | 6.106 | 5.949 | 5.761 | 5.862 | 5.785 |
| $\sigma(a)$ | 0.242 | 0.149 | 0.040 | 0.188 | 0.201 | 0.190 |
| $\chi^2$ | 11.9 | 34.4 | 544. | 8.00 | 8.00 | 8.00 |
| $q$ | 0.157 | 0.0004 | 0.000[b] | 0.433 | 0.433 | 0.433 |
| Case[c] | 5 | 2 | 4 | 1 | 3 | - |

Table 1: Linear regression analysis ($Y = a + b \times X$) of the Pearson (1901) test dataset (taken from the Cantrell paper) using different weighting models (see text for details). $\sigma(a)$ and $\sigma(b)$ are estimted standard deviations of fitted paramaters $a$ and $b$, respectively. [a]: Individual weights for each data point taken from Cantrell's Table 1. [b]: less than $3\times10^{-12}$. [c]: Numbering of cases in Brauers and Finlayson-Pitts (1997).

experiments.

We also applied the `fitexy` routine when analyzing our OH radical data in a number of papers (i.e. Brauers et al., 1996; Hofzumahaus et al., 1998; Brauers et al., 2001; Schlosser et al. 2007). In these publications the goodness-of-fit $q$ is used for indication of situations when the linear model is appropriate or not.

Using different weights, the Pearson (1901) dataset was analyzed in order to demonstrate the `fitexy` approach. The first data column of Table 1 is the result of `fitexy` with the errors (weights) given in Cantrell's Table 1. The result is close to the "zero

difference" results in his Table 2. It is important to mention, that the $\chi^2$ is slightly larger than the degrees of freedom (8), but for an experimental dataset one can conclude, that the scatter is well presented by the error bars. The second and third columns indicate results when X-errors and Y-errors were set to zero. In these cases, the statistical errors are to small to explain the scatter, the linear model should be rejected. The columns four and five were obtained by unit-weighted (*standard*) linear regression. Here, the value of the uniform errors was adjusted that $\chi^2$ is equal to the degrees of freedom. This corresponds to the assumption, the linear relation is valid. Thus, the (average) statistical error of the data points can be estimated (0.316 and 0.573). The last column shows the results for the bivariate case with equal weights in both coordinates for each point. Again, the error (0.278) was adjusted that $\chi^2 = 8$. Figures and additional material can be downloaded from our SAPHIR web site at `http://saphir.fz-juelich.de/public/acpd-8-6409-2008`.

## References

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B. P.: Numerical Recipes in C, 2nd ed., Cambridge University Press, 1992.

Brauers, T., Aschmutat, U., Brandenburger, U., Dorn, H.-P., Hausmann, M., Hofzumahaus, A., Holland, F., Plass-Dülmer, C., Ehhalt, D.H.: Intercomparison of tropospheric OH radical measurements by multiple folded long-path laser absorption and laser induced fluorescence, Geophys. Res. Lett., 23(18), 2545–2548, 1996.

Brauers, T. and Finlayson-Pitts B.J.: Analysis of relative rate measurements, Int. J. Chem. Kin., 29(9), 665–672, 1997.

Brauers, T., Hausmann, M., Bister, A., Kraus, A., Dorn, H.-P.: OH radicals in the boundary layer of the Atlantic Ocean 1. Measurements by long-path laser absorption spectroscopy. J. Geophys. Res. 106, 7399–7414, 2001.

Hofzumahaus, A., Aschmutat, U., Brandenburger, U., Brauers, T., Dorn, H.-P., Haus-

mann, M., Hessling, M., Holland, F., Plass-Dülmer, C.: Intercomparison of tropospheric OH measurements by different laser techniques during the POPCORN campaign 1994, J. Atmos. Chem. 31, 227–246, 1998.

Schlosser, E., Bohn, B., Brauers, T., Dorn, H.-P., Fuchs, H., Häseler, R., Hofzumahaus, A., Holland, F., Rohrer, F., Rupp, L. O., Siese, M., Tillmann, R., and Wahner, A.: Intercomparison of Two Hydroxyl Radical Measurement Techniques at the Atmosphere Simulation Chamber SAPHIR, J. Atmos. Chem., 56(2), 187–205, 2007.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper