

Microsymposium

MS19.O04

Experimental charge density studies: Discard valid data and overfit?

R. Herbst-Irmer¹

¹*Georg-August University of Göttingen, Institute of Inorganic Chemistry, Göttingen, Germany*

In experimental charge density investigation it is indispensable to use the highest possible quality of data. Therefore the multiplicity should be as high as possible, but poor data should be omitted. To decide about resolution limit and discarding outlier data often limits for R_{int} or $I/\sigma(I)$ are used. A better approach is the 'paired refinement method' [1] comparing two data sets by the fit of the models derived by the same refinement protocol to both data sets. For macromolecular data sets it could be shown that a higher resolution should be used than normally derived from the above mentioned criteria. First results for charge density data seem to show the same tendency but of course on a different level. The paired refinement strategy can also be used to investigate the influence of different scaling methods. In a recent version of SADABS [2] a new error model and a 3λ correction is implemented. With the paired refinement strategy the improvement in data quality gets obvious. A further concern in charge density investigation is the question of overfitting. In macromolecular refinement this is answered by the R_{free} concept [3]. Here a refinement protocol is developed by refining against a work set of reflections, e.g. 90 % of the data. The remaining reflections are untouched in the whole refinement process but an R_{free} value is calculated using only this test set of reflections. An overfitting can clearly be identified by a decrease in R_{work} but an increase in R_{free} . This refinement protocol is then used for a final refinement against all data. It will be discussed how this method could support charge density studies.

[1] P. A. Karplus, K. Diederichs, *Science*, 2012, 336, 1030-1033, [2] G. M. Sheldrick, *SADABS*, Göttingen 2014, [3] A. T. Brünger, *Nature*, 1992, 355, 472-475

Keywords: charge density, data quality, SADABS