Applicability domain assessment

The applicability domain assessment in OCHEM is based on the concept of *distance to model*, which was introduced in **Tetko, I.V.** et al, *J. Chem. Inf. Mod.* **2008**, 48(9), 1733-1746.

The following excerpt was taken from a doctoral work by Sushko (Sushko I., Applicability Domain of QSAR models. Doctroral work. **2011**. http://mediatum.ub.tum.de/node?id=1004002)

"Distance to model" definition

The key abstract concept used in this work for assessment of AD is distance to model (DM), defined as follows:

Distance to a model is any numerical measure of the prediction uncertainty for a given compound by the model.

A distance to model assesses how “far” is the compound from the model. The compounds that are “further from the model”, which have larger values of DM, are by definition expected to have lower prediction accuracy than compounds that have smaller values of DM. It should be clearly stated that prediction accuracy correlates

with DM only in average: for example, compounds with DM in range [0.5, 0.6] will on average have higher prediction accuracy than compounds with DM in range [0.6, 0.7] but, nonetheless, the prediction errors for some compounds from the first interval can be bigger than for some compounds from the second interval. In other words, the key property of a DM is the discriminating ability, i.e. the ability to discriminate predictions of high and low accuracy. Importantly, DMs estimate the reliability of predictions. While accuracy is an objective measure that has a rigid calculation procedure, reliability is subjective and can be estimated in numerous ways. Therefore, there is a number of different DMs that assess the reliability of predictions from different perspectives. Here, we briefly overview the DMs used for the AD assessment in this work.

http://mediatum.ub.tum.de/node?id=1004002

On the OCHEM web site the DM which covers 95% of compounds from the training set is used to define applicability domain of the model.