

Quality Enhancement of Large Pure Chemical Databases

Richard Rowley^{C, S}, W. Vincent Wilding, and Neil Giles

Brigham Young University, Department of Chemical Engineering, Provo, UT, U.S.A.

richardlrowley@gmail.com

Addition of new chemicals and property values to chemical databases is such a demanding task that the crucial work of improving the accuracy and quality of the extant data may be overlooked. In fact, the addition of new data can be a valuable tool for upgrading the quality of current values in the database through interrelationships between newly added and extant compounds and their properties. Sponsors and principal investigators of the DIPPR® 801, critically-evaluated, pure-chemical database used this concept to launch a targeted quality enhancement project on existing property values. This presentation will discuss the tools developed that use the growing amounts of inter-related data in the DIPPR database to accomplish continual quality enhancement. Examples of improvements made by use of these tools will be discussed. A few of the tools developed include: (1) a Quality Index (QI) for the database that provides both a quality metric for the data and the database itself and identification of specific values or correlations that may merit revision; (2) inter-property and inter-chemical relationships that are used to identify and correct inconsistent and/or less reliable property values; and (3) an Impact Factor (IF) index that identifies properties in need of update through what is called a matrix review in DIPPR parlance.