## Targeted, Ensemble-Based QSPR Prediction of Thermodynamic Properties using a Back-End Database of Critically-Evaluated Experimental Data

Andrei Kazakov <sup>C, S</sup> and Vladimir Diky Applied Chemicals and Materials Division, NIST, Boulder, CO, U.S.A. akazakov@nist.gov

We present an approach for prediction of thermodynamic properties based on the Quantitative Structure-Property Relationships (QSPR) model ensemble consensus. The model ensemble is generated using different machine learning techniques using a small subset of compounds that are most chemically similar to the target for which the prediction is to be made. Compound sampling for model development is performed from the back-end database of criticallyevaluated experimental data generated by the NIST/TRC ThermoData Engine. As with all empirical methods, the results are very sensitive to the data coverage of chemical space near the target compound. A combination of different similarity metrics used during sampling was found to be very beneficial, vastly improving the performance. For the typical application (i.e., using only two-dimensional structure definitions for QSPR descriptors), the method can be used for rapid, "on-the-fly" estimation of properties. Applications for prediction of normal boiling points and critical constants for pure compounds will be discussed.