Mathematical gnostics as an alternative to statistical data analysis

Magdalena Bendová

E. Hála Laboratory of Separation Processes
Research Group of Thermodynamics of Task-Specific Materials

Outline

1. Introduction
   - Data uncertainty treatment
   - Advanced data analysis by means of mathematical gnostics
2. Advanced data analysis by means of mathematical gnostics
   - Marginal analysis
   - Robust regression along a gnostic influence function

Non-statistical data analysis

"All models are wrong. But some of them are useful." G.E.P. Box

NIST Technical Note 1297 (1994): widely used guidelines for uncertainty treatment

Statistics requires an *a priori* knowledge of the data distribution function

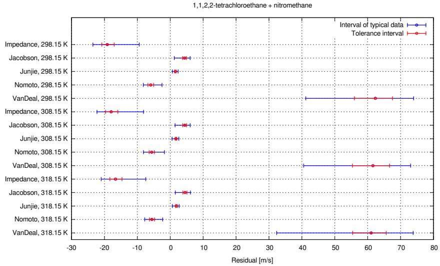
Mathematical gnostics derived based on the concept of "entropy of datum" – no *a priori* assumption of the data distribution function required

Z. Wagner et al. Advanced Analysis of Isobaric Heat Capacities by Mathematical Gnostics. submitted to J. Solut. Chem., Jan. 2017

A. Andresova et al. Influence of the alkyl side chain length on the thermophysical properties of chiral ionic liquids with a (1R,2S,5R)-(–)-menthol substituent and data analysis by means of mathematical gnostics, submitted to J. Mol. Liq., March 2017

Advanced data analysis by means of mathematical gnostics (MG)

- Derived from fundamental principles of the theory of measurement
- Provides several types of robustness, is valid for small datasets

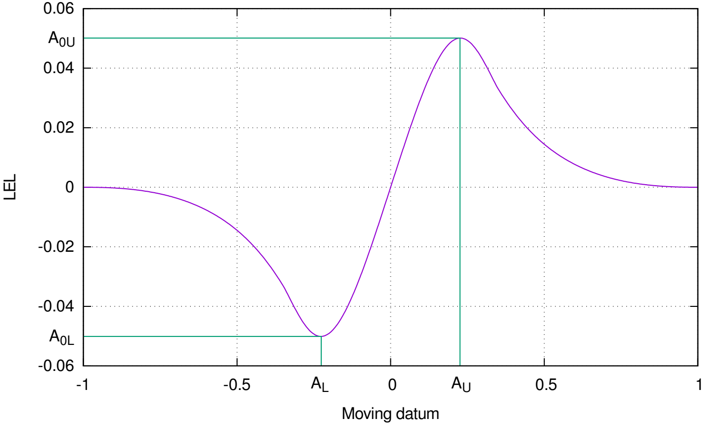


- Critical assessment of data
- Marginal analysis, robust linear and nonlinear data regression
- Comparison of predictive models
- Real-time data analysis and estimate of particle size distribution in atmospheric aerosol and timely detection of measurement device defects

Comparison of the speed-of-sound models in binary mixtures

Dragoescu, D.; Gheorghe, D.; Bendová, M.; Wagner, Z. Fluid Phase Equilibria 2015, 385, 105-119

Marginal analysis: Local estimate of location

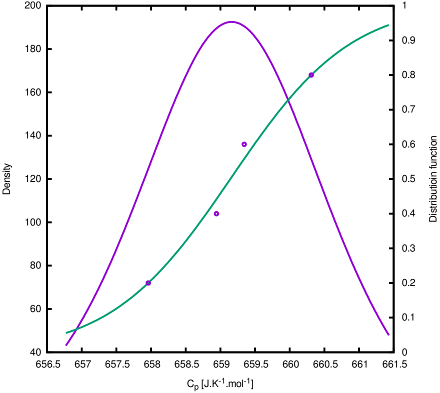


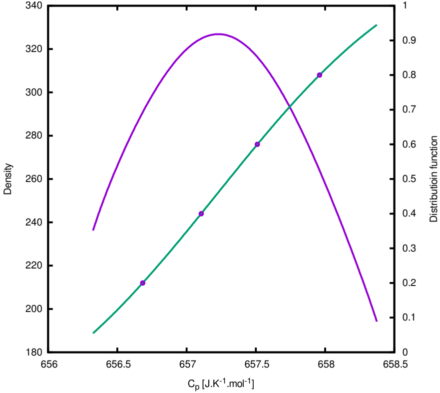
LEL – Local Estimate of Location (maximum probability density)

(A_L,A_U) – interval of typical data

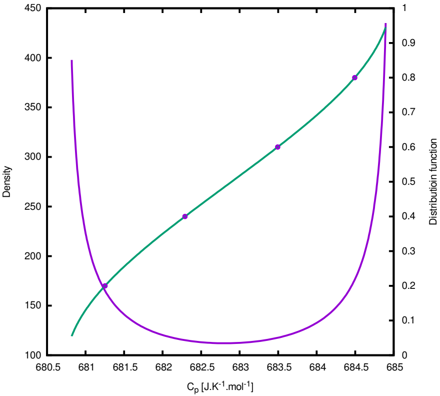
(A_{0L},A_{0U}) – tolerance interval

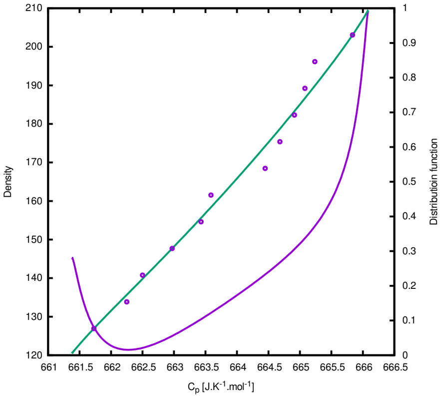
Examples of distribution functions



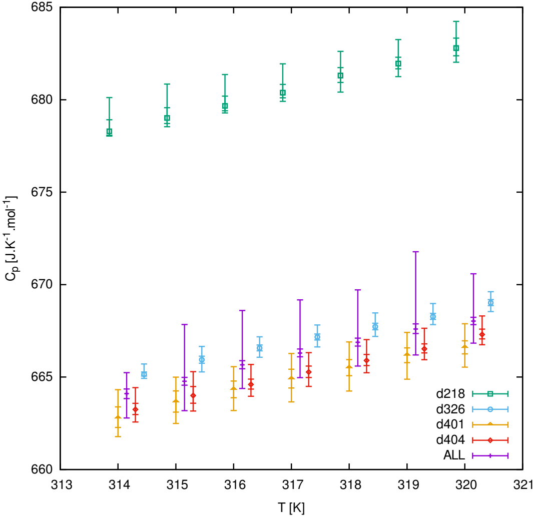


Examples of distribution functions





Heat capacity of [$C_4(C_1OC_2)$im][Tf_2N]



d218 0.49352 g

d326 1.02973 g

d401 1.15465 g, recycled (evaporation of DCM)

d404 1.04197 g, decolorized (active carbon, alumina)

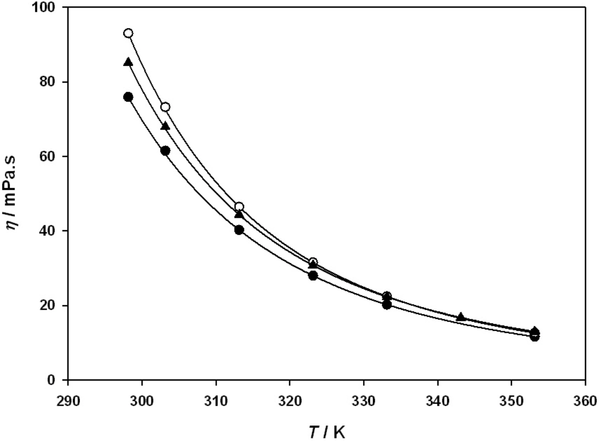
ALL homogeneous subsample of all data

3-butyl-1-(2-methoxyethyl)imidazolium bis{(trifluoromethyl)sulfonyl}imide

Example of a robust non-linear regression

Regression of experimental viscosity data for $[C_4xC_5im][Tf_2N]$ ILs using the Vogel-Tamman-Fulcher equation

$$\eta = \eta_0 \exp \frac{D}{T - T_0}$$



Example of a robust non-linear regression

Regression of experimental viscosity data for [C_4xC_5im][Tf_2N] ILs using the Vogel-Tamman-Fulcher equation

Deviations δ and weights w of the regression

	[C_4C_5im][Tf_2N]		[C_4iC_5im][Tf_2N]		[C_4cC_5im][Tf_2N]	
T [K]	δ	w	δ	w	δ	w
298.15	0.00152	0.99648	−0.00159	0.99897	−0.00120	0.99871
303.15	−0.89824	0.00003	−0.66502	0.00002	−0.52910	0.00001
313.15	−0.00637	0.99996	0.00834	0.99672	0.01122	0.96408
323.15	−0.00820	0.99996	0.00535	0.99900	−0.01357	0.98746
333.15	0.03090	0.93779	−0.03484	0.89896	−0.00792	0.99801
343.15					0.01943	0.92105
353.15	−0.01548	0.99689	0.01865	0.97691	−0.00411	1.00000

A problem with the measuring device detected at 303.15 K

Summary

- Data analysis of small datasets by means of statistics is meaningless
- Mathematical gnostics offers an alternative approach to uncertainty treatment
- Analysis of "bugs" in experimental procedure as well as more subtle influences in measurements is possible
- Robust regression along a gnostic influence function enables us to find the optimum data fit

http://ttsm.icpf.cas.cz