COSMOlogic
Predicting Solutions

Method benchmark on published reference data of infinite dilution activity coefficients (IDAC)

COSMO*therm* versus UNIFAC and COSMO-SAC

Setup

The comparison is made along a paper by Gerber and Soares[1], who compare the prediction of infinite dilution activity coefficients (IDAC) of various UNIFAC variants and two COSMO-SAC versions.

Two different data set have been used by Gerber and Soares:

- A non-water set with organic solvents in other organic solvents
- A water set with organic solvents and water

Overview

	non-water set		water set	
Total data points	386		359	
Temperature range / [K]	293 K	393 K	283 K	373 K
Substances	50		257	
IDAC range	0.54	76.7	2.51	9.74E+09
Chemical diversity	very low		low to medium	

1) Renan P. Gerber and Rafael de P. Soares, *Ind. Eng. Chem. Res.*, Article ASAP, **DOI:** 10.1021/ie901947m

Experimental Data

Chemical diversity of the two sets

- Important chemical functionalities are not included in both sets: amine, aniline, thiol, imine, cyano, nitro, and so on)
- Bi- or poly functional molecules are not included (amino-acids, diols, glycerols , halogenated acids, and so on)

Chemical diversity

Groups	Subgroup	Substances in DB		Groups	Subgroup	Substances in DB	
		non-water	water			non-water	water
alkanes	linear	16	8	halogen-alkanes	chloro	2	18
	iso		11		bromo		11
	neo		5		iodo		3
cyclo alkanes	simple	2	4	carbonlys	aldehyde		8
	branched	3	4		ketone	5	18
alkenes	1,2-alkenes	5	7		carbon acids	2	9
	2,3-alkenes	1	3		esther		24
	dienes		2		amide	1	
cyclo alkenes	1,2	4	4	alcohols	primary	6	21
	diene		4		secondary		10
	triene		1		tertiary		11
alkynes			6		phenols	1	
aromatics	simple	2	22	ether			13
	halogen subst.		5	epoxids			1
	polyaromatics		23	water			1

Methods

Overview of applied methods

- COSMO-SAC and UNIFAC results have been taken from a paper of R. P. Gerber, R. de P. Soares[1], only the best UNIFAC variants of that paper are presented here.
- Two COSMO-SAC variants are included in the comparison, the published literature variant (lit.) and a variant with adjusted (adj.) parameters by Gerber and Soares.
- COSMO*therm* results have been calculated with parameterization FINE1501

Method overview

Methods	# data points		uses molecule conformations	adjusted to the two sets	Comment
	non-water	water			
UNIFAC (1)	374	294	no	no	Group contribution method
UNIFAC (2)	386	114	no	no	Group contribution method
COSMO-SAC-D (lit.)	361	311	no	no	
COSMO-SAC-D (adj.)	361	311	no	yes	Adjusted to the used sets.
COSMO*therm*	386	358	yes	no	FINE1501 parameterization

1) Renan P. Gerber and Rafael de P. Soares, *Ind. Eng. Chem. Res.*, Article ASAP, **DOI:** 10.1021/ie901947m

Results

- The best method for the organic set is UNIFAC.
- The best method for the water set and both sets altogether is COSMO*therm*
- COSMO-SAC (even the version especially fitted to both sets) has higher AAD and RMSE values than COSMO*therm*.

Methods	Combined IDAC results for both DB				
	# points non-water	# points water	AAD	RMSE	Max dev
UNIFAC (1)	374	294	0.8	1.26	6.08
UNIFAC (2)	386	114	0.5	1.11	5.12
COSMO-SAC-D (lit.)	386	311	1.27	1.52	7.18
COSMO-SAC-D (adj.)	361	311	0.63	0.86	3.67
COSMOtherm	386	359	0.48	0.67	3.42

AAD = absolute average deviation, RMSE = root mean square error
The UNIFAC (2) AAD/RMSE is biased towards organic solvent, because only 114 water-DB points are included.

Methods	IDAC results for the non-water DB					IDAC results for the water DB				
	# points	AAD	RMSE	Max dev	mean dev	# points	AAD	RMSE	Max dev	mean dev
UNIFAC (1)	374	0.28	0.39	1.74	0.05	294	1.47	1.84	6.08	0.25
UNIFAC (2)	386	0.12	0.17	0.88	-0.05	114	1.8	2.31	5.12	-1.15
COSMO-SAC-D (lit.)	361	0.59	0.81	2.77	-0.36	311	2.05	2.41	7.18	-1.41
COSMO-SAC-D (adj.)	361	0.48	0.57	3.10	-0.18	311	0.81	1.11	3.67	0.39
COSMO*therm*	386	0.35	0.53	3.42	0.08	359	0.61	0.79	2.20	-0.30

The COSMO*therm* calculations include all data points, whereas the COSMO-SAC data have only been published for a reduced set

Summary

- COSMO*therm* outperforms UNIFAC and COSMO-SAC with respect to overall accuracy
- UNIFAC shows the best RMSE among the compared methods when used on mono-functional organic solvents with limited chemical diversity.