

# Shortcomings in ecotoxicity data to provide a testset for an ECETOC project

Paul Thomas (CEHTRA), Faizan Sahigara (CEHTRA), James Dawick (Shell Health), Mark Lampi (ExxonMobil Petroleum & Chemical) , Roger van Egmond (Unilever), Malyka Galay Burgos (ECETOC)

## Background

ECETOC recently published its technical report no 120 (ECETOC 2013) which examined the relationship between Chemical Activities (as defined by phase equilibrium thermodynamics)) and toxicity of narcotic chemicals previously hypothesised (e.g. Mackay *et al.* 2009).

The activity coefficient can be regarded as the inverse expression of solubility, where a solute that is only sparingly soluble in a solvent (e.g. water), will have a high activity coefficient. When expressed as a mole fraction, the activity coefficient is the reciprocal of the solubility. As the relationship uses sub-cooled liquid solubility, an adjustment is made using the fugacity ratio for solid substances (at STP).

Thus, for the purposes of this exercise, for each substance used in the test set, high quality endpoint data were required for solubility, melting point and acute and chronic ecotoxicity for unicellular algae, aquatic invertebrates and fish (each requiring information on E/LC50s, species, test temperature, method [semi-static, flow-through, closed...], analytical verification and study duration).

Accurate estimations of Modes of Action (MOA) were also required and the scheme according to the Verhaar *et al.* (1992) and with modifications proposed by Enoch *et al.* (2008).

Several sources of information were identified but as narcotic modes of action were considered as the main driver for this exercise, the ECHA disseminated database was a primary source. The ECETOC task force felt that the advantage of using the ECHA database was that the studies contained therein relate to industrial chemicals and are therefore more appropriate for a relationship with baseline narcotics. Moreover, they have previously been validated and so less time would be required in the data location process. For this purpose, only studies with a Klimisch score of 1 or 2 (valid without restriction or valid with restrictions, respectively, for the purposes of risk assessment) were used in this exercise.

Table 1. Example of dataset marking (red: does not meet basic criteria e.g. study duration; orange: data removed due to secondary criterion e.g. technical deficiencies during study)

LC50	Species	duration	T°C	Mol/L	LC50	Species	Duration	T°C	Mol/L
99	Pimephales promelas	96	12	0,001165721	109	Palaemonetes pugio	48	20	0,00128347
97	Fundulus heteroclitus	48	20	0,001142171	1941	Daphnia magna	24		0,022855192
502	Pimephales promelas	96	25	0,005911028	1682	Daphnia magna	48		0,019805478
330	Pimephales promelas	96		0,003885736	510	Artemia salina	24		0,006005226
220	Lepomis macrochirus	96		0,002590491	220	Daphnia magna	48	22	0,002590491
					122	Artemia salina	24		0,001438545
18,2	Oncorhynchus mykiss	96		0,00015247	152,5	Crassostrea gigas	48	20	0,001277562
102	Poecilia reticulata	336	22	0,0008545	79	Daphnia magna	48	23	0,000661819
103	Pimephales promelas	96	22	0,000862878	29	Daphnia magna	48	22	0,000242946
51	Micropterus salmoides	96		0,00042725	79	Daphnia magna	24		0,000661819
75	Ictalurus punctatus	96		0,000628309					
70,7	Pimephales promelas	96	25	0,000592286	35	Daphnia magna	48		0,000227553
121	Danio rerio	48	23	0,001013672	97	Daphnia magna	24		0,000630648
24,3	Danio rerio	96	23	0,000157987	14	Americamysis bahia	96	22	0,000106464
					20,8	Daphnia magna	48	22	0,000158175
16	Limanda limanda	96		0,000121673	85,2	Daphnia magna	48		0,000647909
28,3	Jordanella floridae	96	25	0,000215209	24	Gammarus pulex	48		0,00018251
42	Oncorhynchus mykiss	48	15	0,000319392	56	Daphnia cucullata	48		0,000425856
54,8	Poecilia reticulata	96		0,00041673	47	Daphnia magna	24		0,000577947
52	Cyprinodon variegatus	96	22	0,000395437	27	Daphnia magna	96		0,000205323
47	Pimephales promelas	48		0,000357414					
40,7	Pimephales promelas	96		0,000309506					

Figure 4. Graphs of values and ranges for individual substances for MOA 1 In the acute regression slope for fish

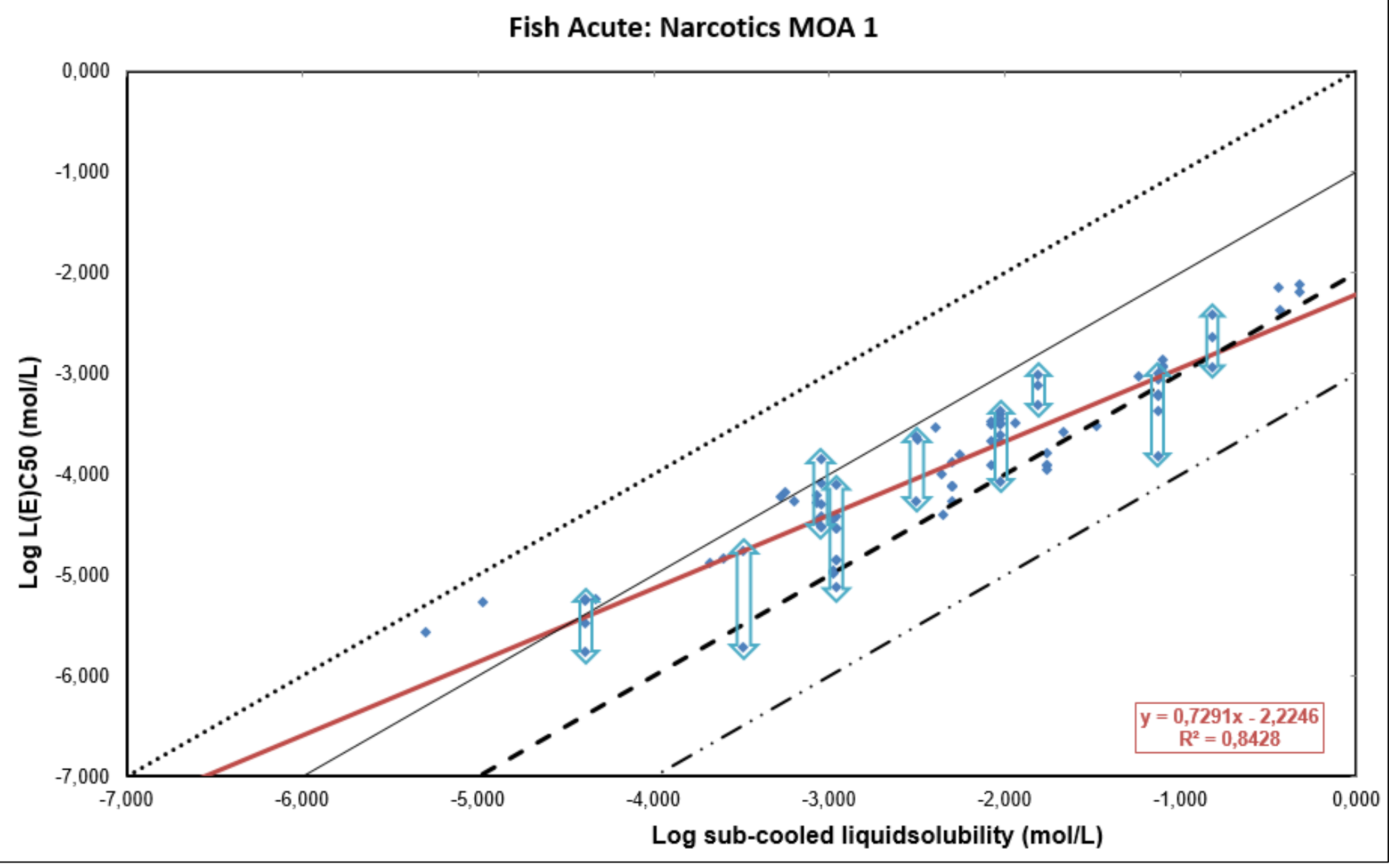


Table 2. Variability within the ectoxicity data for MOA1 substances (Fish LC50 96h)

Substance name	Fish Log L(E)C50 (mol/L)						
	n	Min	Max	Range	Mean	Variance	Standard deviation
1,2-dichloroethane	3	-2.935	-2.862	0.073	-2.907	0.002	0.039
tert-butyl methyl ether	2	-2.186	-2.118	0.068	-2.225	0.002	0.048
1,1,2,2-tetrachloroethane	4	-3.958	-3.915	0.042	-3.897	0.005	0.069
1,3 dichlorobenzene	3	-4.468	-4.207	0.261	-4.319	0.018	0.135
diethylphthalate	5	-4.268	-3.884	0.383	-4.103	0.019	0.138
Nitrobenzene	3	-3.319	-3.015	0.304	-3.153	0.024	0.154
Trichloroethylene	4	-3.915	-3.474	0.440	-3.641	0.040	0.201
1,1,1-trichloroethane	8	-4.080	-3.377	0.703	-3.574	0.051	0.226
1,2 dichlorobenzene	5	-4.985	-4.451	0.534	-4.869	0.054	0.233
Dibutyl Phthalate	4	-5.763	-5.240	0.523	-5.435	0.060	0.245
Dichloromethane	3	-2.933	-2.411	0.523	-2.662	0.069	0.262
Tetrachloroethylene	5	-4.521	-3.843	0.678	-4.255	0.085	0.291
Chloroform	6	-3.817	-2.994	0.823	-3.279	0.087	0.294
1,4 dichlorobenzene	7	-5.118	-4.099	1.019	-4.631	0.112	0.335
3-nitrotoluene	2	-4.268	-3.630	0.638	-3.949	0.203	0.451
1,2,3-trichlorobenzene	2	-5.715	-4.754	0.961	-5.234	0.462	0.680

## Results and Conclusion:

Over 2000 datapoints were evaluated in this project ( for MOA 1, 2 and 3 substances), selected from studies previously validated as K1 and 2. Of these, only 660 datapoints were retained for 123 substances and classed as valid for use in this project, less than 1/3 of the data originally collected. These data were organised in a database and classified according to the colour scheme in Table 1.

Solubility posed a specific problem: limited details of the method were available; much data used for the ECHA database were from the secondary literature; even when the method was reported, the normalised methods for solubility (e.g. OECD 105) are prone to high degrees of error especially for liquids at low solubility (e.g. <10 mg/L). Figures 1 and 2 highlight this. For MOA 1, solubility (and Kow values become more erratic once log Kow >5 suggesting that the sub-cooled liquid solubility values used in this exercise may be unreliable for such substances). This is further evidenced from the residuals plot in Figure 3a. A very poor relationship between log Kow and solubility was observed for MOA 2 substances (probably mostly due to their different ionised state in the two experiments). As a result, the residuals associated with most of the MOA2 substances were quite high as shown in Figure 3b. A better relationship would be expected for these substances if membrane/water partition coefficients were used rather than octanol/water partition coefficients.

Often, multiple ecotoxicity endpoints on the same trophic level were available and associated with high variability (up to one order of magnitude) (Figure 4). According to the thermodynamic principles on which this study was based, no difference is expected between E/LC50s for any species assuming the organisms reached steady state within the duration of the study. The variation observed between certain studies must therefore be related to experimental factors such as technical method, health status of the organisms, demonic intrusions and statistical treatment of the dataset. The substances with multiple measured LC50 values are reported in Table 2 in order of their increasing variances and therefore the overall confidence in these data is lower.

Overall, the thermodynamic relationship between solubility and ecotoxicity proposed by Mackay *et al.* holds true for this dataset. However, the variability in the E/LC50 values for several substances prevents the direct use of this dataset to create a reliable QSAR model. The authors recommend this dataset is improved with additional high quality data.

### References

ECETOC 2013. Technical report no. 120  
Enoch SJ, Hewitt M, Cronin MT, Azam S, Madden JC. 2008. *Chemosphere* 73(3):243-248.  
Klimisch HJ, Andreae M, Tillmann U. 1997. *Regul Toxicol Pharmacol* 25(1):1-5.  
Mackay D, Arnot JA, Petkova EP, Wallace KB, Call DJ, Brooke LT, Veith GD. 2009. *SAR QSAR Environ Res* 20(3-4):393-414  
Verhaar HJM, van Leeuwen CJ, Hermens JLM. 1992. *Chemosphere* 25(4):471-491.

## Issues/Methods

Once compiled, it became clear that the dataset needed further refinement and a series of rules were introduced to remove unreliable/unverifiable data:

- Solubility values significantly outside of the range 20-25°C or for which no information existed were excluded
- All ecotoxicity studies were evaluated on a case by case basis that an appropriate technical method had been applied (e.g. static tests acceptable only if substance know to be stable, soluble and non-volatile; Substances with volatility >1 Pa in closed system accompanied by analytical measurements)
- endpoints based on too short study duration were removed to avoid including datapoints which were not at steady state.
- Endpoints for which no standardised method exists were evaluated on a case by case basis for acceptability (eg. Publications using a non-standard method or species).

### Variability in the dataset

After carrying out the data refinement, statistical analysis was performed to evaluate the variability in the retained dataset.

Residuals plot: Measured Log Octanol-Water Partition Coefficient (LogKow) of dataset substances (MOA1 and 2) against their corresponding sub cooled liquid solubility values to obtain a linear regression. A residual plot was derived from this regression to identify substances which deviated from the expected regression line (potential outliers with unreliable solubility values).

Since several substances in the retained dataset had multiple entries (i.e. with measured LC50 values retrieved from multiple literature sources), the variance and standard deviation of the underlying data were reported. Data associated with higher dispersion were identified indicating that one or multiple measured values added noise/error to the results.

Figure 1 & 2. Graphs of log Kow v sub-cooled liquid solubility with indication of data quality

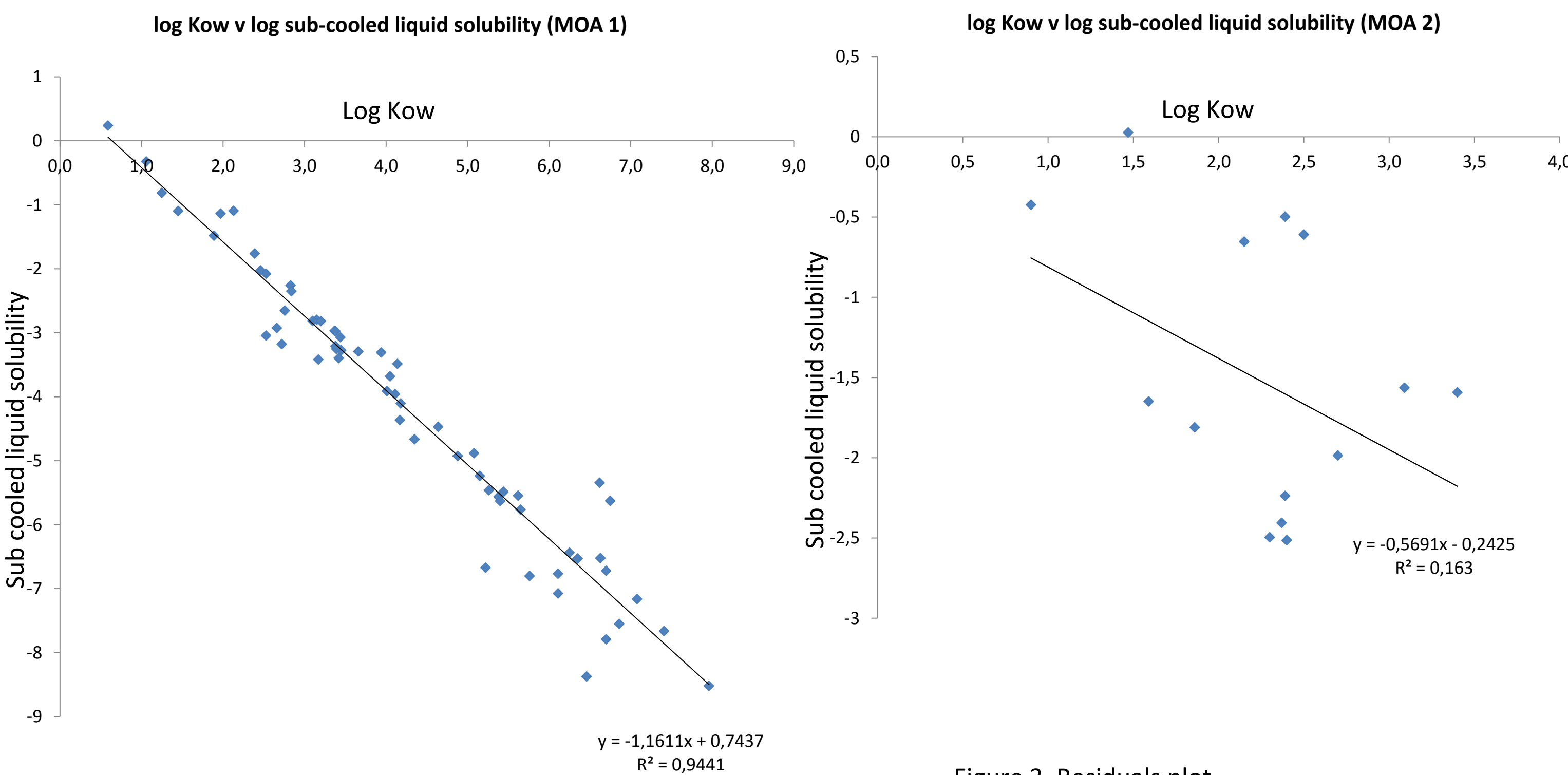


Figure 3. Residuals plot

