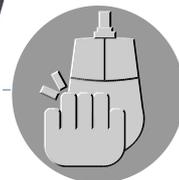


生態毒性予測QSARモデル KATEについて

 独立行政法人国立環境研究所
環境リスク研究センター 蓮沼 和夫

生態毒性予測システム KATE(KAshinhou Tool for Ecotoxicity)概要

詳しくは
<https://kate.nies.go.jp/>



KATE開発の経緯

2011年3月

KATE 2011公開

2009年3月

インターネット版「KATE on NET」に加え
スタンドアロン版「KATE on PAS」を公開

2008年1月

試用版 (KATE Ver0.1) 公開

2007年7月

三省合同審議会※に対し、
予測結果の提供を開始

注)※:薬事・食品衛生審議会
薬事分科会化学物質安全対策
部会化学物質調査会、化学物
質審議会審査部会、中央環境
審議会環境保健部会化学物質
審査小委員会

2004年

環境省の請負業務として研究・開発
開始(2004年度～2012年度)

生態毒性予測システムKATE

Quantitative Structure-Activity Relationship (QSAR)
定量的構造活性相関

化学物質の

- ・構造上の特徴
- ・物理化学定数

★KATEの場合

- ・化学物質の部分構造
- ・logP
(水-オクタノール分配係数)

相関

生物学的活性 (毒性等)

★KATEの場合

・魚類毒性値

(慢性は開発中)



・ミジンコ毒性値

(慢性は開発中)



・藻類毒性値

(開発中)

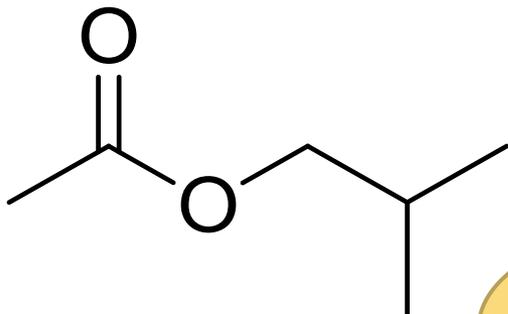


KATEの予測の流れ

ユーザーの操作

化学物質の構造入力

酢酸イソブチル



logP代入
1.6

結果の解釈

(構造C判定及びlogP判定
に注意)

魚類 LC₅₀: 36mg/L

▶ 5 ミジンコ EC₅₀: 46mg/L

KATE内部での動作

フラグメント(部分構造)抽出

-C(=O)O[脂肪族]	1個
脂肪族 O	2個
脂肪族 C	6個

クラス分類

esters aliphaticクラスの条件

-C(=O)O[脂肪族]	1個以上
芳香族 原子	含まない
チオール、アミン	含まない

~~esters aromaticクラスの条件~~

-C(=O)O[芳香族]	1個以上
チオール、アミン	含まない

毒性値の予測

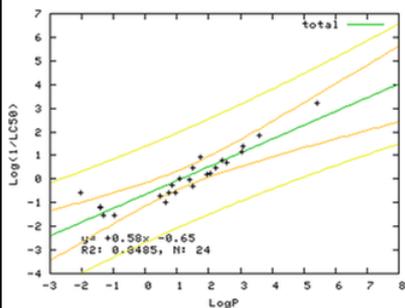
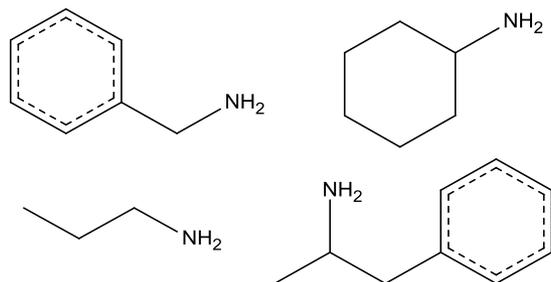
QSAR式を用いて毒性値を予測

・魚類	: $\log(1/LC_{50}) = 0.67 \times \log P - 0.59$
・ミジンコ	: $\log(1/EC_{50}) = 0.69 \times \log P - 0.72$

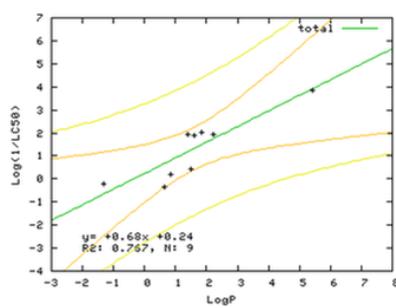
クラス分類

- ▶ フラグメントをもとに、クラスを分類(約45分類)

primary amines aliphatic/aromatic



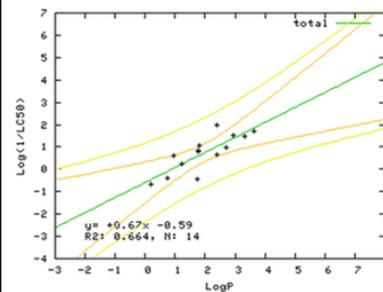
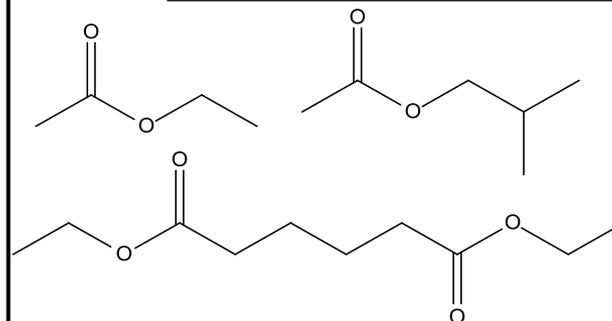
魚類



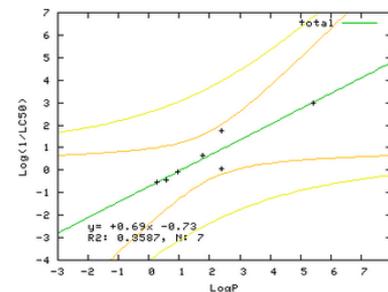
甲殻類



esters aliphatic

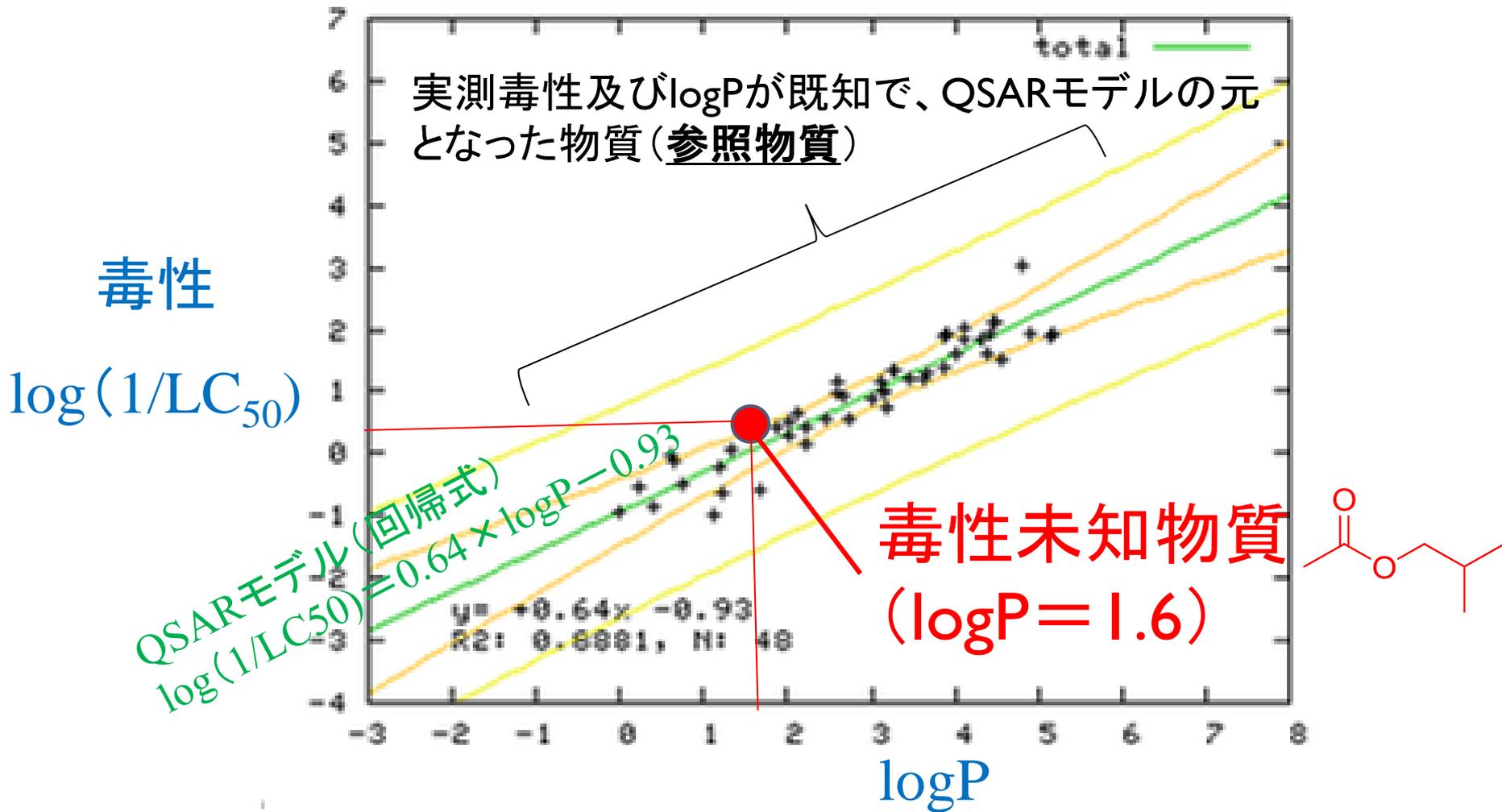


魚類

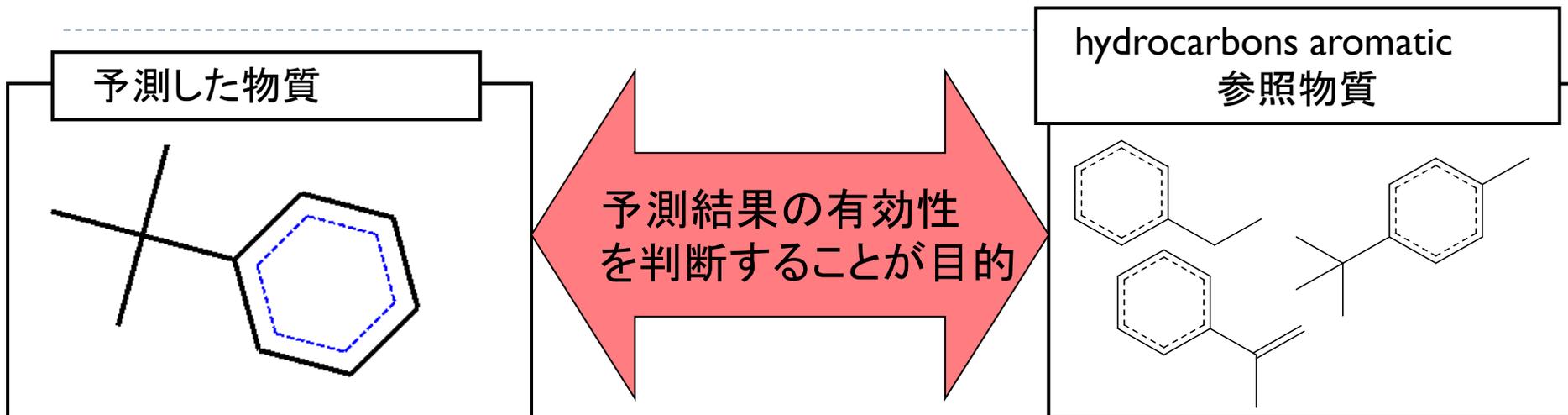


甲殻類

毒性値の予測



予測結果の適用範囲について（判定）



- **構造C判定**: 予測する化学物質のもつ部分構造すべてが、
 - ○:[そのクラス]の参照物質にも含まれる。
 - △:[そのクラス]または[Neutral Organicクラス]の参照物質にも含まれる。
 - ×:[そのクラス]や[Neutral Organicクラス]の参照物質には含まれない部分構造がある。として評価される。
- **logP判定**: 予測した物質のlogPがQSAR回帰式の有効範囲内に入っているか(内挿であるか)で評価される。
 - 8 ■ スタンドアロン版: 有効範囲外の場合、『>P』又は『<P』と評価
 - インターネット版: 有効範囲内の場合『○』、範囲外は『×』と評価

他の生態毒性予測QSARとの比較

比較に用いたQSARモデル、生態毒性データ

▶ 比較に用いたQSARモデルの特徴

QSARモデル	開発元	記述子	エンドポイント
KATE 2011年3月版	環境省、 国立環境研究所	logP	魚類96時間LC ₅₀ 甲殻類48時間EC ₅₀
TIMES Version 2.26.1	ブルガリア ブルガス大学	logBCF _{tox} 、 LUMO	急性毒性 (<i>Rana japonica</i> , <i>Lymnaea stagnalis</i> , <i>Carassius auratus</i> , <i>Oryzias latipes</i> , <i>Leuciscus idus</i> , <i>Pimephales promelas</i> , <i>Daphnia magna</i> , <i>Daphnia pulex</i> , <i>Ceriodaphnia dubia</i> , <i>Escherichia coli</i> , <i>Bacillus subtilis</i> , <i>Tetrahymena pyriformis</i> 等)
ECOSAR Version 1.0	米国EPA	主にlogP	魚類96時間LC ₅₀ 甲殻類48時間LC ₅₀ 藻類96時間LC ₅₀ 魚類・甲殻類・藻類ChV (ChV: Chronic Value、NOECとLOECの幾何平均)

▶ 検証に用いた生態毒性データ

▶ KATE2011年3月版の参照物質毒性データ

○魚類:582物質

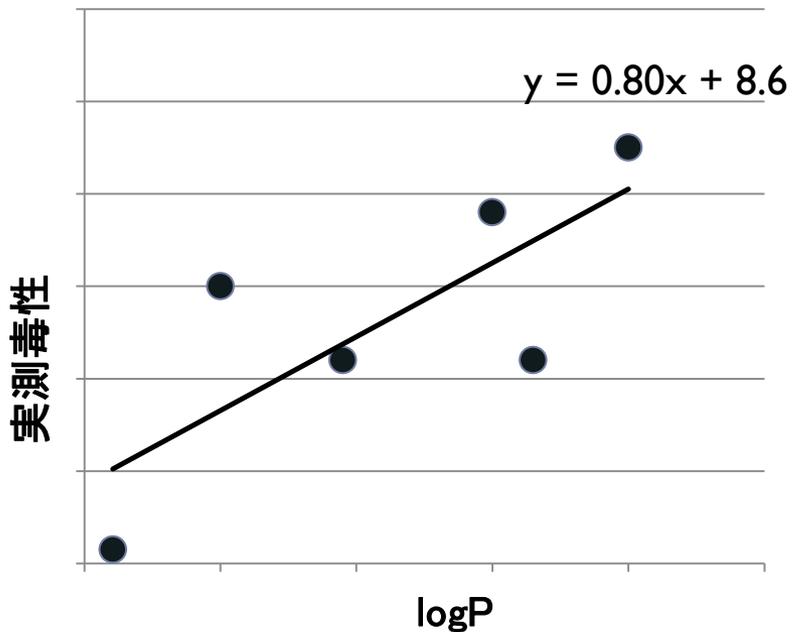
- ・環境省の生態毒性試験結果(メダカ)
- ・米国EPAファットヘッドミノールデータベース

○甲殻類:314物質

- ・環境省の生態毒性試験結果(オオミジンコ)

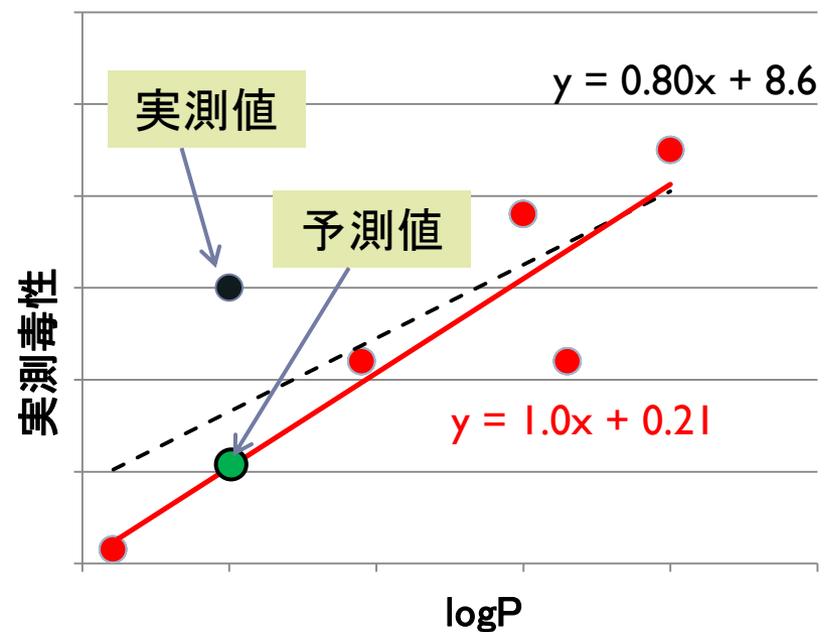
【補足】 leave-one-out法について

- ▶ KATE: 参照物質の毒性を予測しても意味がない
- ▶ 実測値と予測値を比較する1物質を除いてQSAR式を作成し、その1物質の毒性値を予測 (leave-one-out)



QSAR式

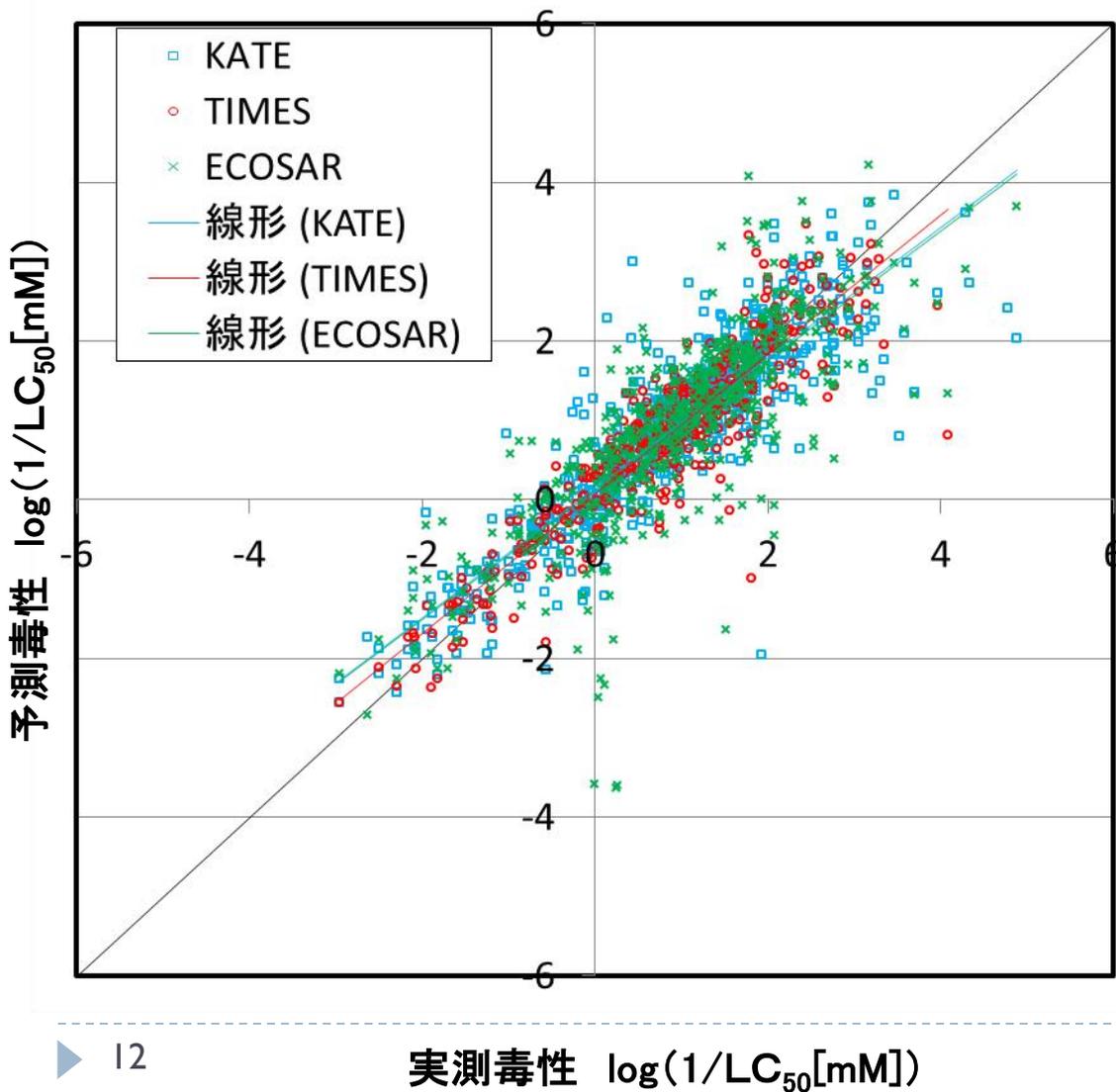
▶ II (実測毒性がある6物質より作成)



leave-one-out QSAR式

(1物質除いた5物質(赤丸)より作成)

比較結果 (魚類) ①



自由度調整済決定係数(R^2_{adj})

KATE	TIMES	ECOSAR
0.78	0.83	0.64

2乗平均平方根誤差(RMSE)

KATE	TIMES	ECOSAR
0.60	0.49	1.1

○KATE (leave-one-outの結果)

魚類96時間 LC_{50}

○TIMES (Reactive unspecifiedは除く)

Pimephales promelas (ファッドヘッドミノー)

魚類96時間 LC_{50}

○ECOSAR (Baseline Toxicityは除く)

魚類96時間 LC_{50}

備考(甲殻類も同様)

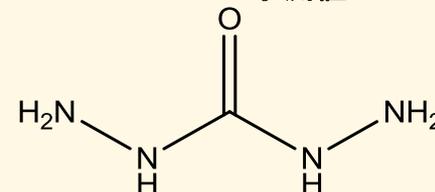
ECOSARはBaseline Toxicity以外全てのクラスの予測値を使用。ある物質が2つ以上のクラスに含まれる場合、両方の予測値を用いている。

例)

○実測毒性値: 52mg/L

○ECOSAR Hydrazones 予測値: 143 mg/L

○ECOSAR Amides 予測値: 395000mg/L



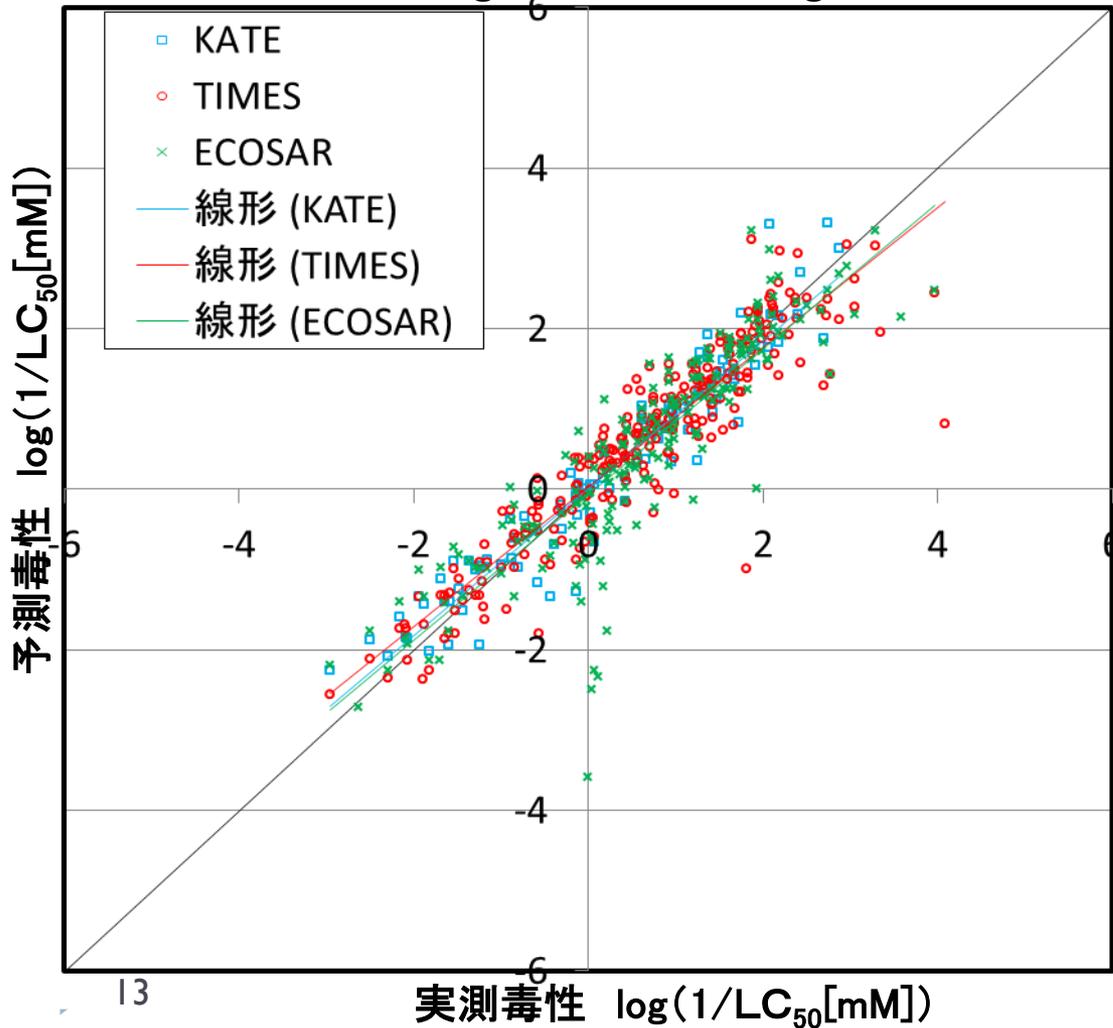
比較結果（魚類）②

麻酔作用によるクラスのみ限定

KATE: Neutral Organics

TIMES: basesurface narcotics, narcotic amines

ECOSAR: Neutral Organics, Neutral Organics-acid



自由度調整済決定係数(R^2_{adj})

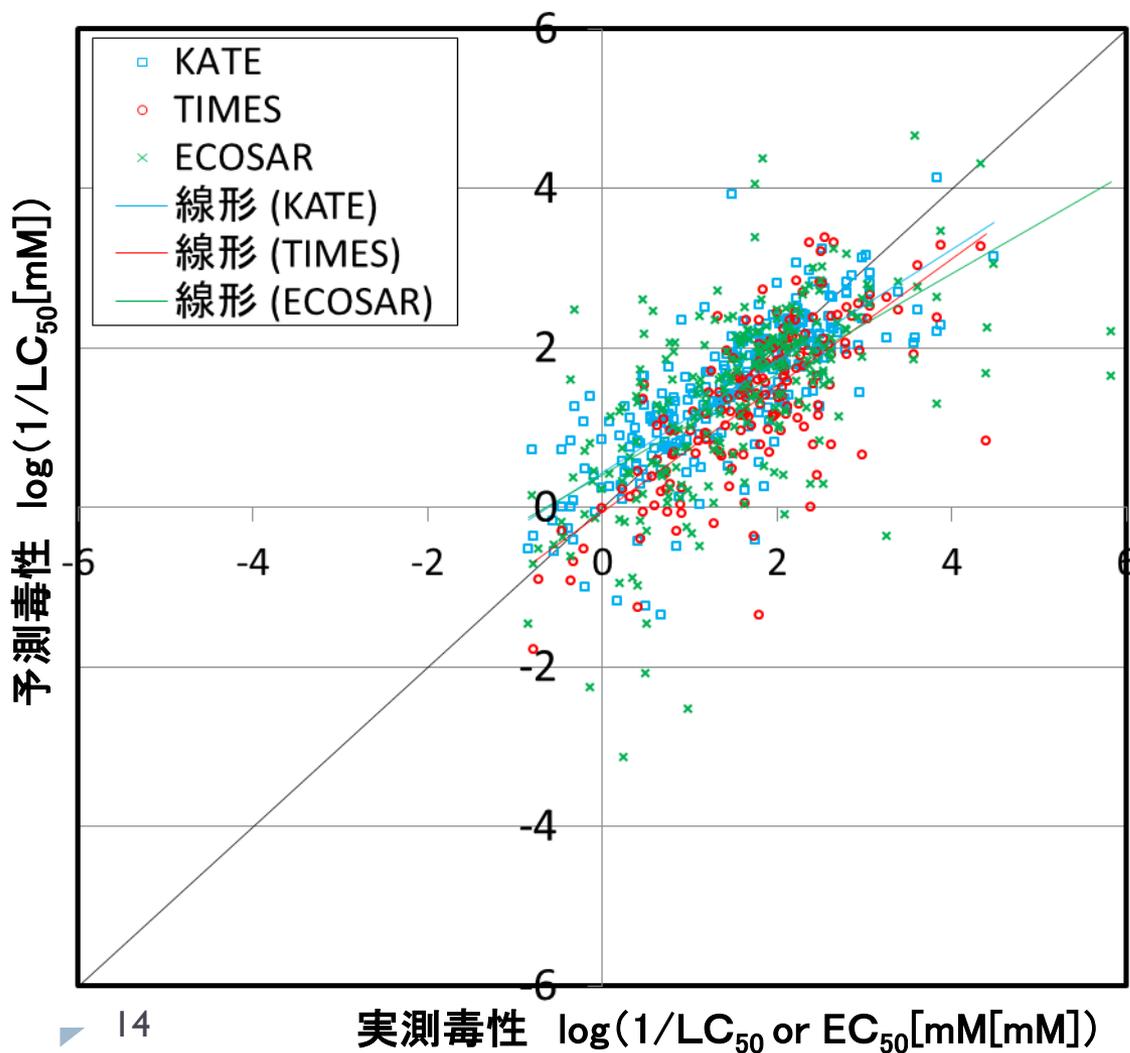
KATE	TIMES	ECOSAR
0.91	0.85	0.77

2乗平均平方根誤差(RMSE)

KATE	TIMES	ECOSAR
0.41	0.50	0.79

- KATE (leave-one-outの結果)
魚類96時間 LC_{50}
- TIMES (Reactive unspecifiedは除く)
Pimephales promelas (ファッドヘッドミノー)
魚類96時間 LC_{50}
- ECOSAR (Baseline Toxicityは除く)
魚類96時間 LC_{50}

比較結果 (甲殻類) ①



自由度調整済決定係数(R^2_{adj})

KATE	TIMES	ECOSAR
0.60	0.35	0.36

2乗平均平方根誤差 (RMSE)

KATE	TIMES	ECOSAR
0.62	0.53	1.4

- KATE (leave-one-outの結果)
甲殻類48時間 EC_{50}
- TIMES (Reactive unspecifiedは除く)
Daphnia magna 48時間 EC_{50}
- ECOSAR (Baseline Toxicityは除く)
甲殻類48時間 LC_{50}

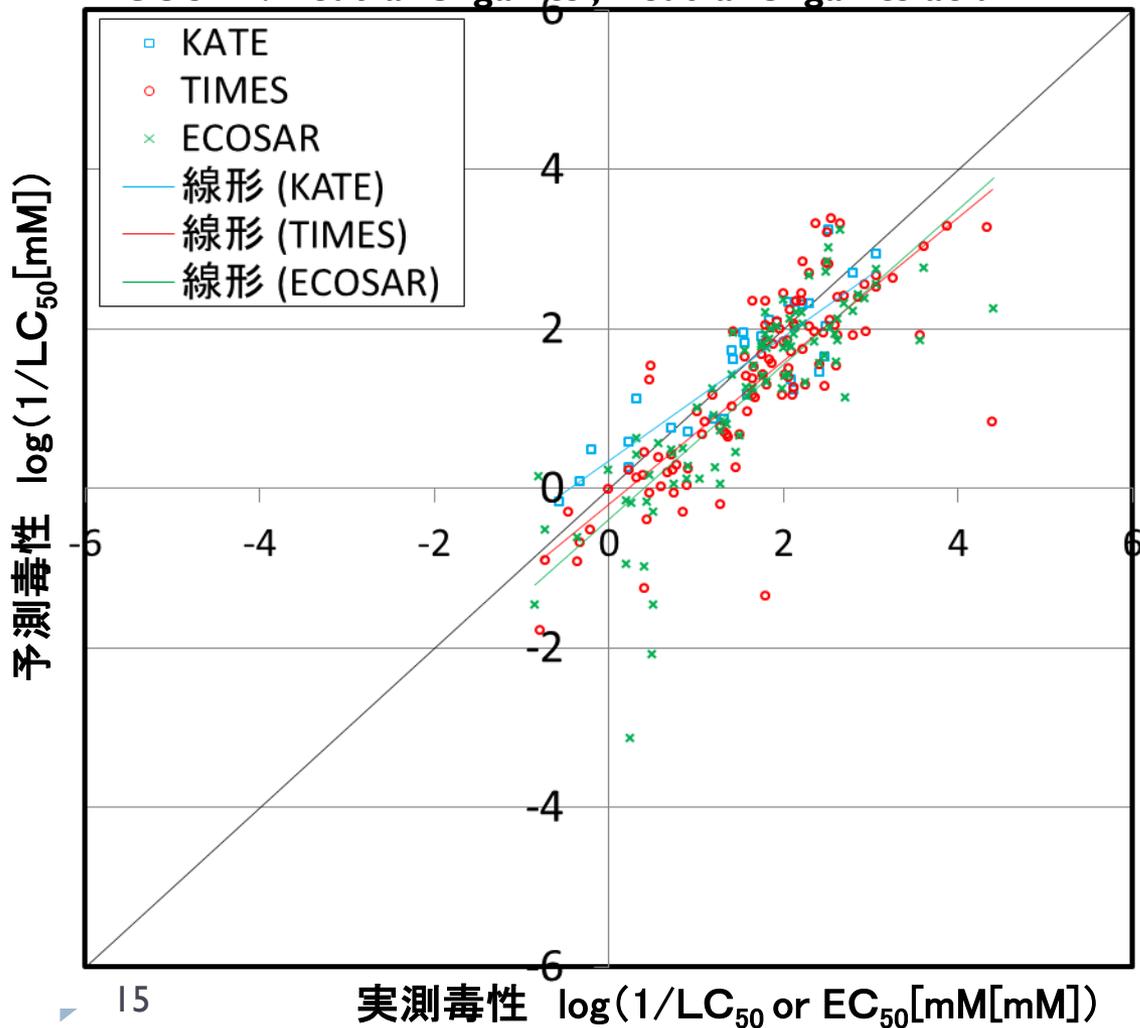
比較結果（甲殻類）②

麻酔作用によるクラスのみ限定

KATE: Neutral Organics

TIMES: basesurface narcotics

ECOSAR: Neutral Organics, Neutral Organics-acid



自由度調整済決定係数(R^2_{adj})

KATE	TIMES	ECOSAR
0.76	0.65	0.68

2乗平均平方根誤差 (RMSE)

KATE	TIMES	ECOSAR
0.46	0.79	0.46

○KATE (leave-one-outの結果)

甲殻類48時間 EC_{50}

○TIMES

Daphnia magna 48時間 EC_{50}

○ECOSAR

甲殻類48時間 LC_{50}

KATE2013(仮称)に向けた検討

KATE2013（仮称）に向けた検討事項

▶ 参照物質の追加

- ▶ 環境省生態毒性試験として新たに得られた結果を追加

▶ フラグメント・ルールの見直し

- ▶ 予測精度向上の為の見直し
 - ▶ クラス1: クラス2の為のフラグメント(化学物質の部分構造)
 - ▶ クラス2: クラスを定義するルール
 - ▶ クラス3: 構造C判定の為のフラグメント(化学物質の部分構造)

▶ 記述子の追加

- ▶ logPで説明できないクラスに対し、新たな記述子を追加
- ▶ 例: PEOE(Partial Equalization of Orbital Electronegativity)等

▶ エンドポイントの追加

- ▶ 慢性毒性(甲殻類、(魚類))
- ▶ 藻類(急性、慢性)

フラグメント・ルールの見直し①

▶ クラス1の修正例(不飽和の三級アルコール)

クラス1 定義修正

クラス1 (#6005: アリルアルコール)

KATE2011	F/1111/C=CCO/3M0,4H1,/
KATE2013	F/1111/C=CCO/3M0,3H>0,4H1,/

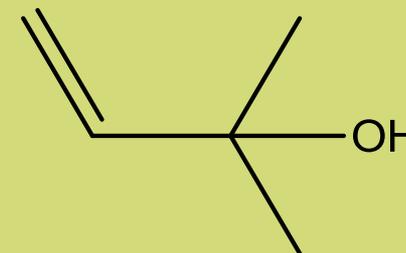
クラス1 フラグメント抽出結果

KATE2011		KATE2013	
4902(脂肪族 O)	1個	4902(脂肪族 O)	1個
4909(脂肪族 C)	5個	4909(脂肪族 C)	5個
4945(C-OH(脂肪族 alcohol))	1個	4945(C-OH(脂肪族 alcohol))	1個
6005(アリルアルコール)	1個		

KATE2011
conjugated systems2



KATE2013
alcohols aliphatic



クラス2 定義

conjugated systems2	alcohols aliphatic
6005(アリルアルコール)>0個	4945(C-OH(脂肪族 alcohol))>0個
:	4922(芳香族 n s o) =0個
	:

フラグメント・ルールの見直し②

▶ クラス2の修正例 (halidesの整理)

KATE2011

クラス名	p(slope)	RMSE	R ²	N
halides1	0.52	0.97	0.11	6
halides2	0.00092	0.33	0.95	6
halides3	2.6×10^{-20}	0.26	0.88	43

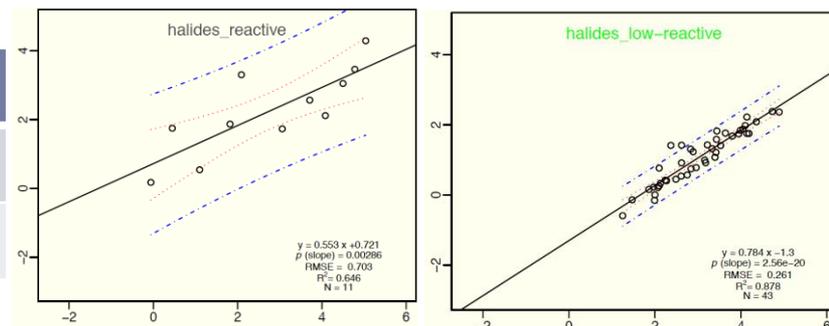
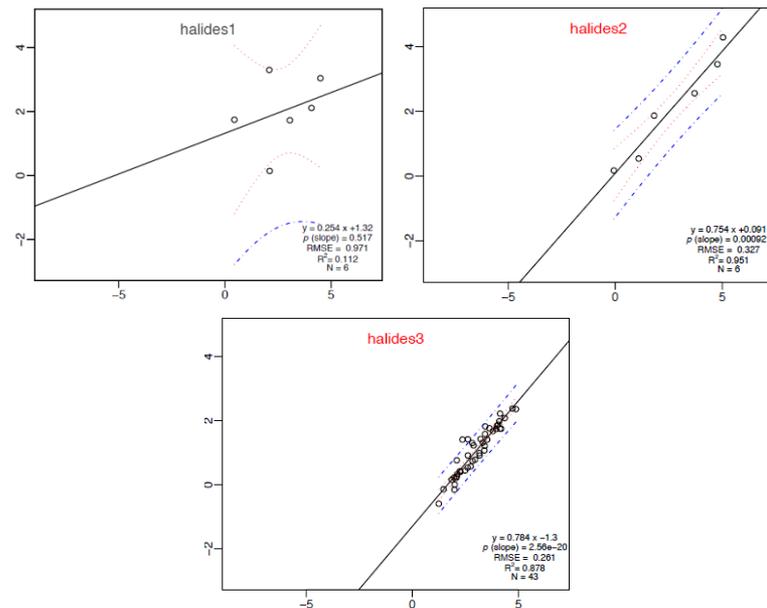
注) 魚類急性毒性



KATE2013(暫定)

クラス名	p(slope)	RMSE	R ²	N
halides_reactive	0.0029	0.70	0.65	11
halides_low-reactive	2.6×10^{-20}	0.26	0.88	43

注) 魚類急性毒性



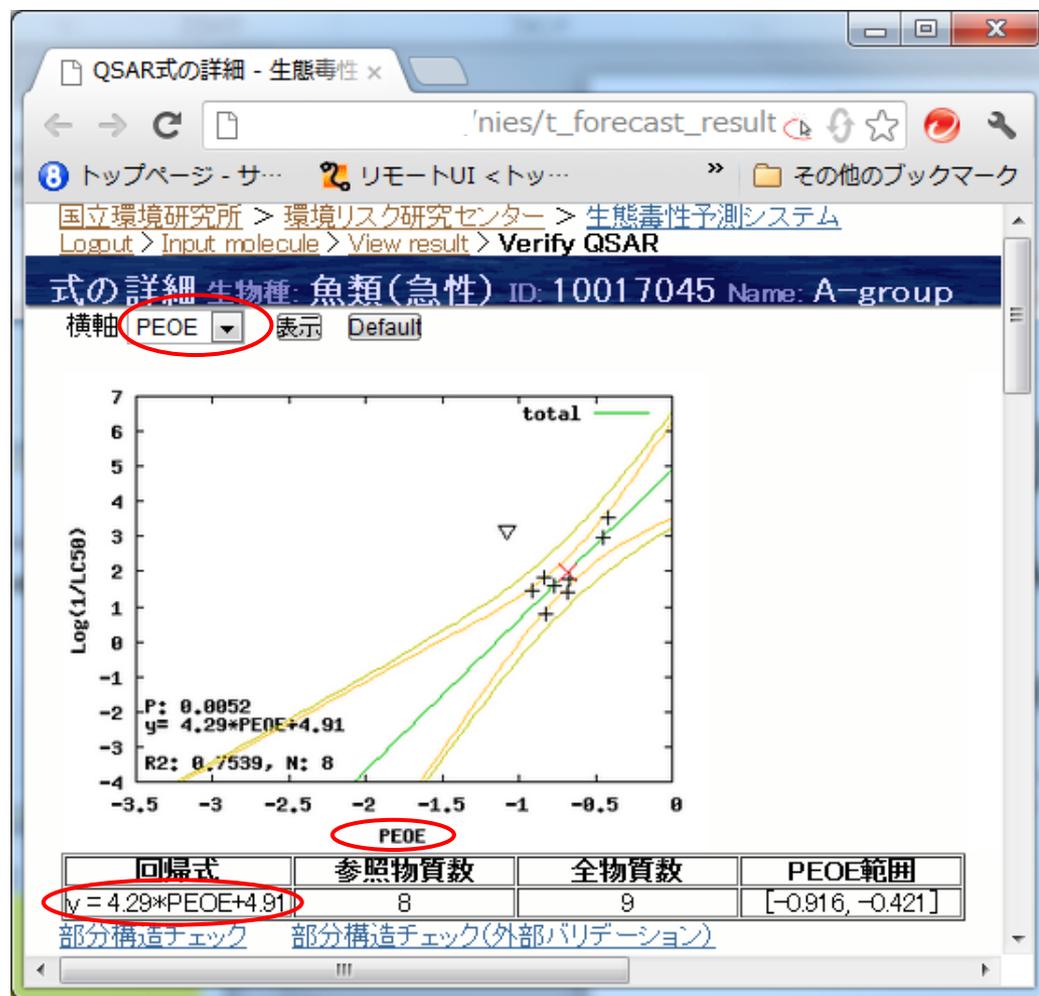
QSAR式

縦軸: 実測毒性 (1/LC₅₀[mM])

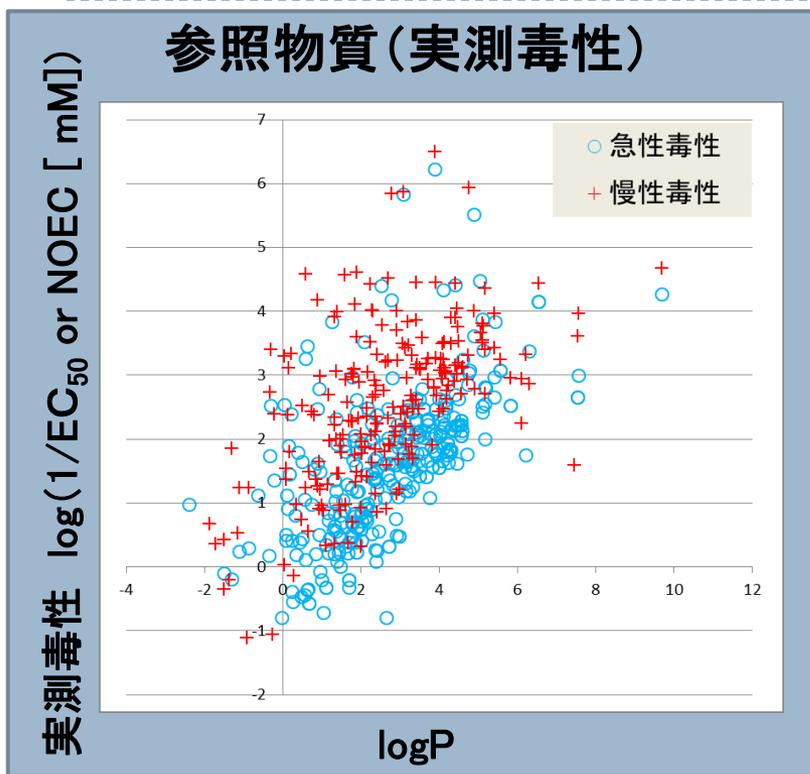
横軸: logP

記述子の追加

- ▶ logPを用いて毒性を説明できないクラスについて、PEOE, logBCFtox※の導入を検討
- ▶ KATE on NET(インターネット版)について、任意の記述子が使えるよう改良

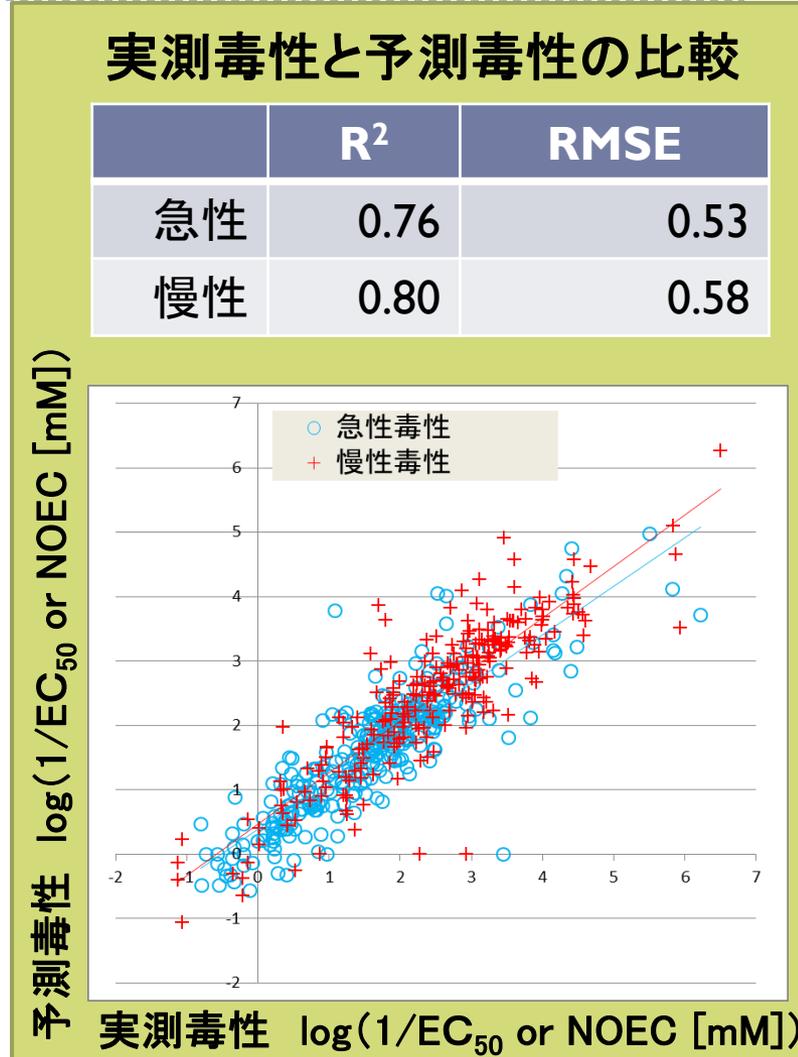


エンドポイントの追加 (甲殻類慢性①)



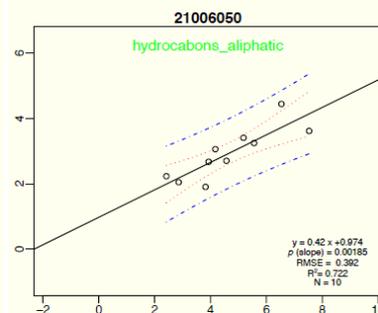
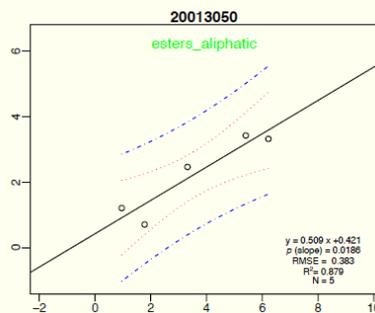
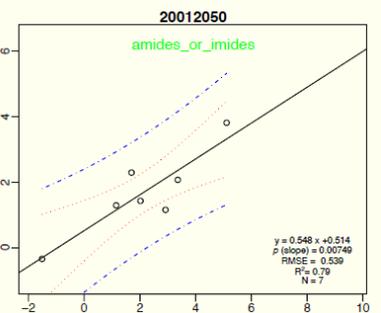
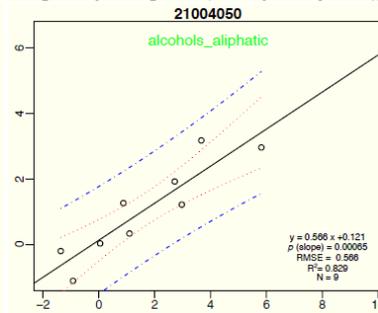
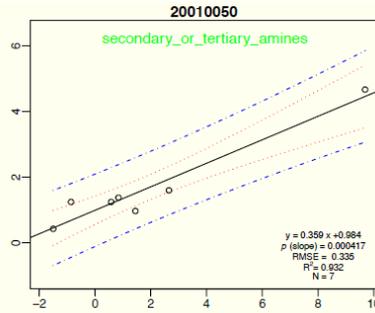
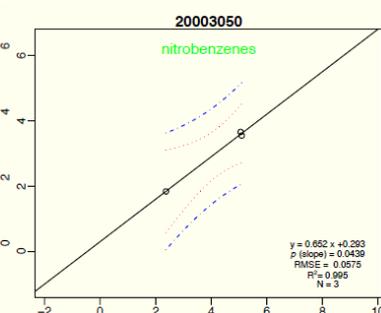
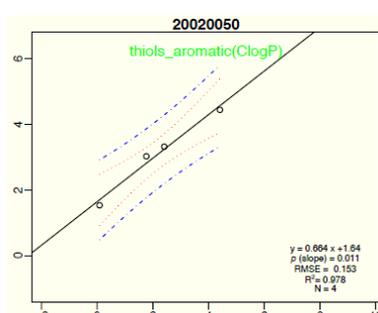
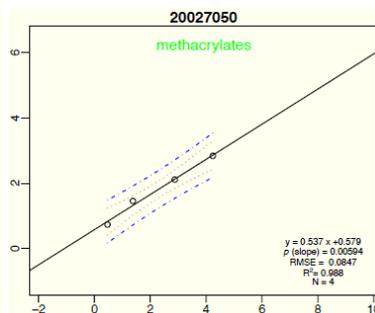
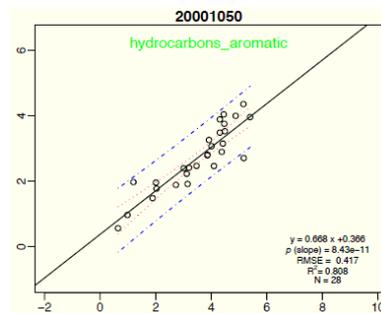
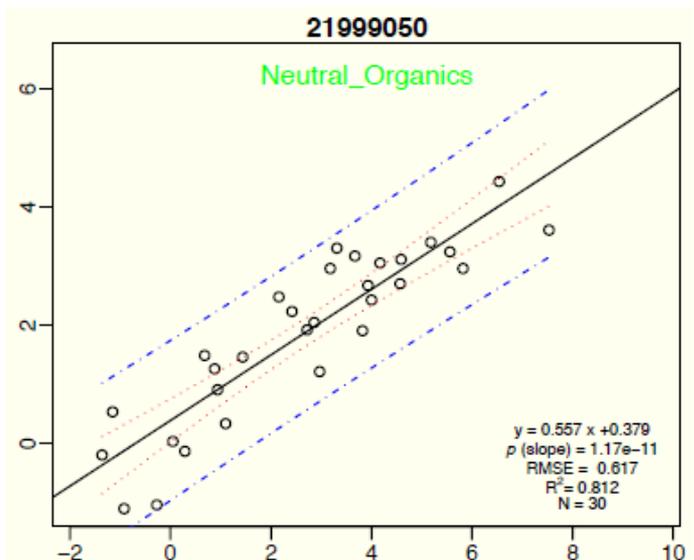
慢性QSAR式作成

KATE2013(暫定)急性毒性の
ルール・フラグメント



エンドポイントの追加（甲殻類慢性②）

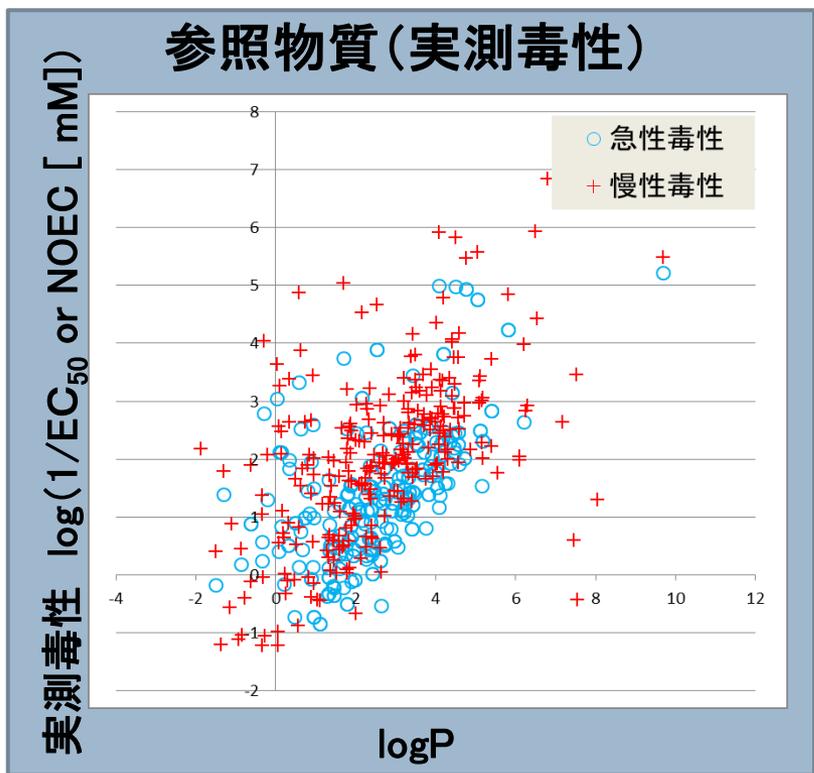
▶ $R^2 > 0.7$ 及び $p(\text{slope}) < 0.05$ のQSAR式



▶ $p(\text{slope})$

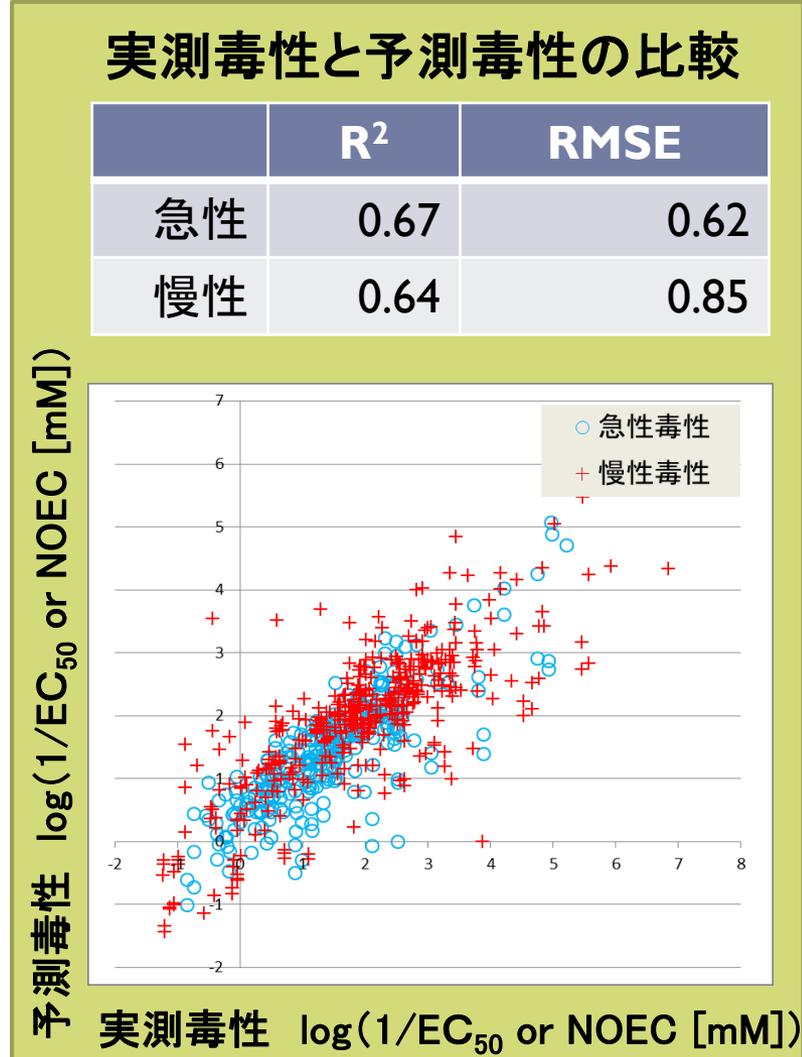
▶ QSAR式の回帰係数(傾き)が統計的に有意に0より異なる有意水準

エンドポイントの追加 (藻類①)



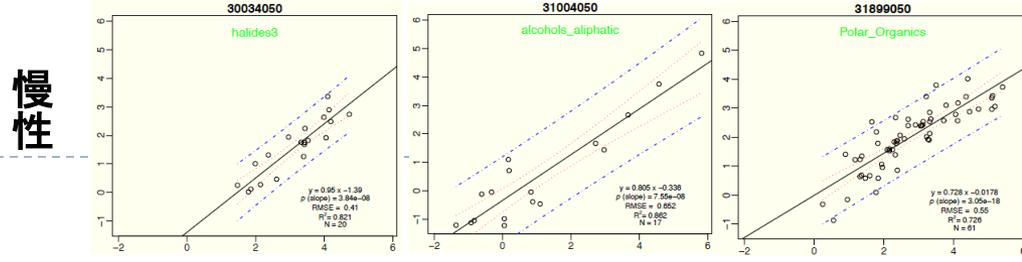
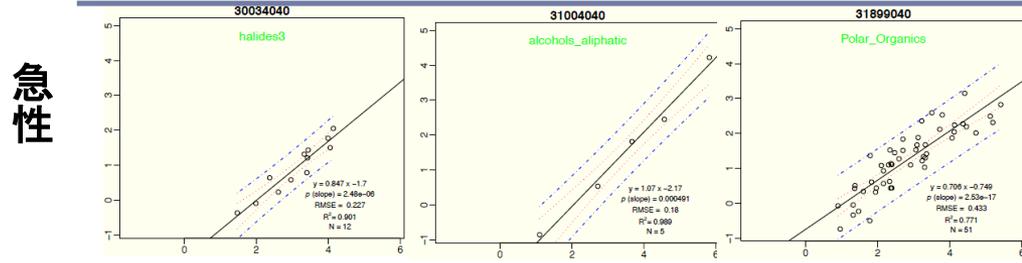
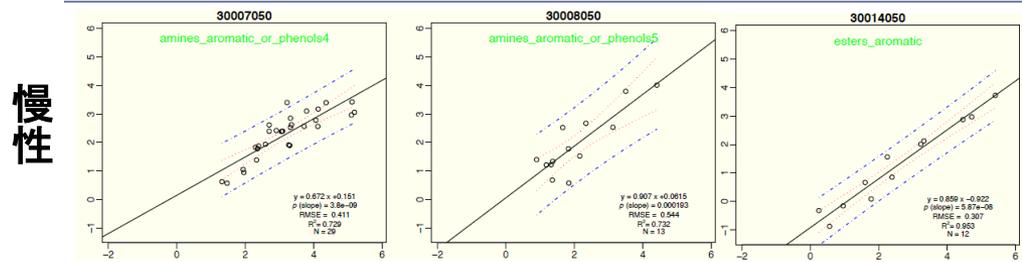
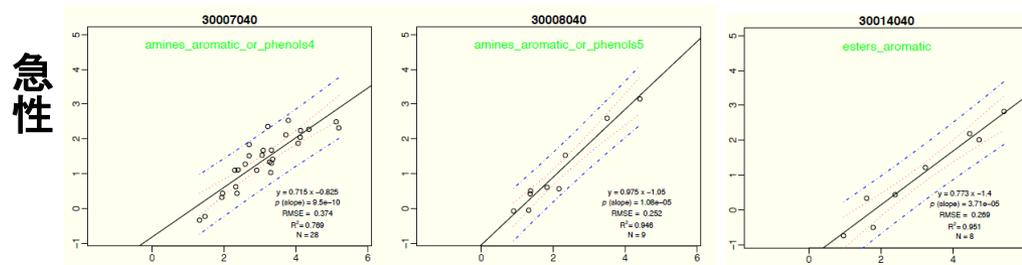
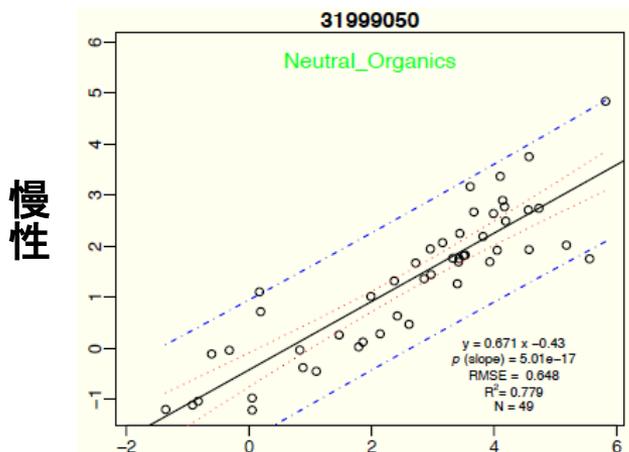
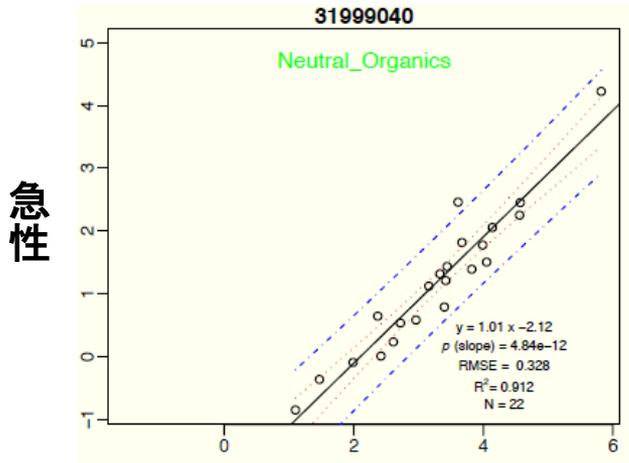
慢性QSAR式作成

検討中のルール・フラグメント
(魚類・甲殻類とは異なる)



エンドポイントの追加 (藻類②)

▶ 急性・慢性両者が $R^2 > 0.7$ 及び $p(\text{slope}) < 0.05$ のQSAR式



KATE2013(仮称)の外部バリデーション結果

外部バリデーション手法

▶ 参照物質以外の有害性データを用いて、予測精度の確認を行う

- ▶ 外部バリデーションに用いた有害性情報
 - ▶ 新規化学物質審査に使用された生態毒性試験結果
- ▶ 外部バリデーションに用いた統計量
 - ▶ External explained variance (Q^2_{ext})

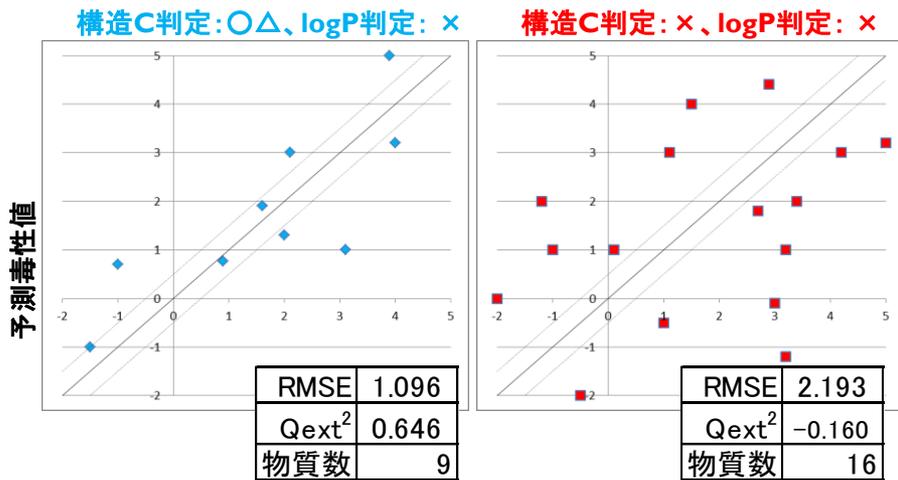
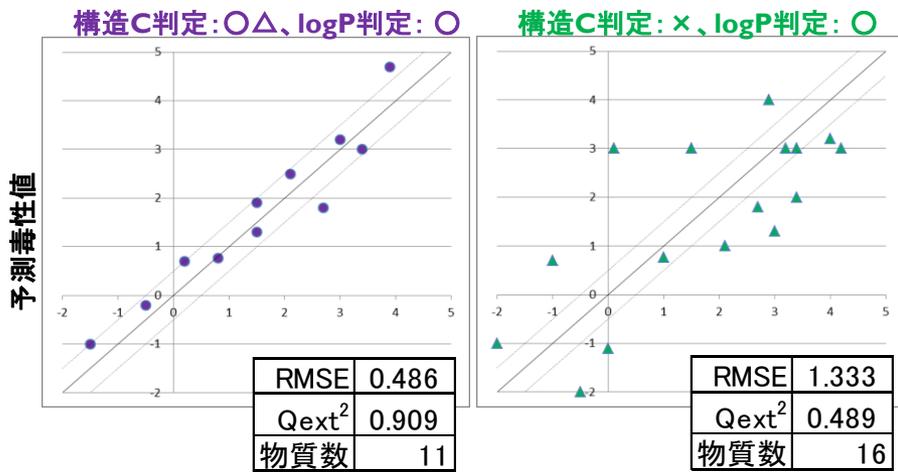
$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{n_{ext}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{ext}} (y_i - \bar{y})^2}, \quad -\infty < Q^2_{ext} \leq 1$$

y_i = observed response for the i -th object
 \hat{y}_i = predicted response for the i -th object
 \bar{y} = average response value of y_i

OECD Environment health and safety publications series on testing and assessment No. 69, "Guidance document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models", ENV/JM/MONO(2007)2, Paris, 2007

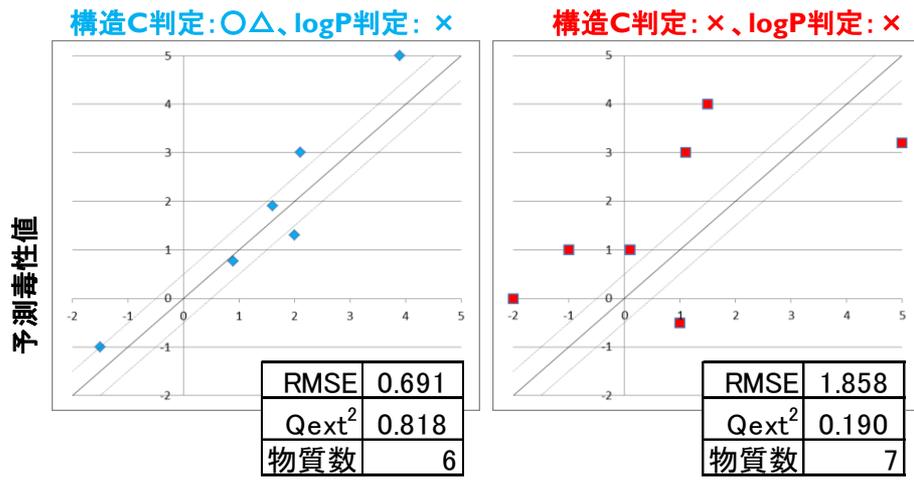
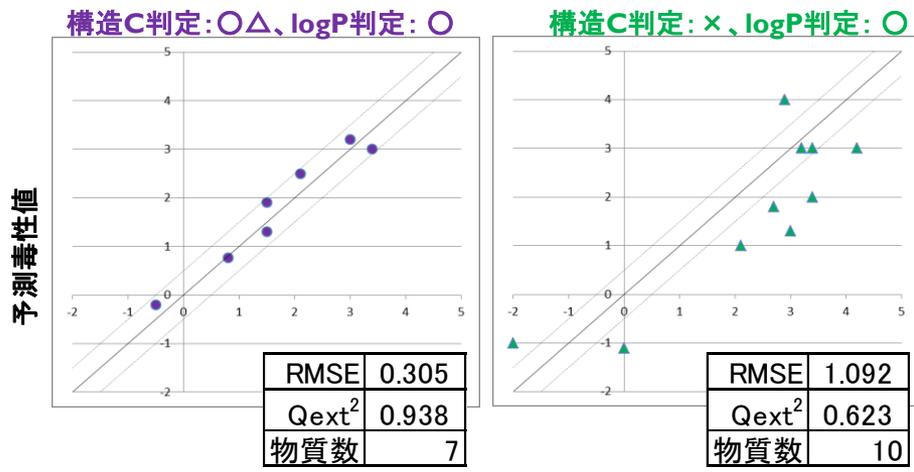
外部バリデーションのイメージ

全クラス



実測毒性値

有効なクラス ($R^2 > 0.7$ かつ $p(\text{slope}) < 0.05$)



実測毒性値

外部バリデーション結果①

魚類急性

	RMSE		Qext ²		物質数	
	全クラス	有効なクラス	全クラス	有効なクラス	全クラス	有効なクラス
構造C判定: ○△ logP判定: ○	0.499	0.361	0.731	0.859	19	12
構造C判定: ○△ logP判定: ×	1.9	1.95	-5.14	-5.49	9	7
構造C判定: × logP判定: ○	1.51	1.51	-1.22	-1.13	43	33
構造C判定: × logP判定: ×	2.36	2.25	-1.41	-0.61	10	7

甲殻類急性

	RMSE		Qext ²		物質数	
	全クラス	有効なクラス	全クラス	有効なクラス	全クラス	有効なクラス
構造C判定: ○△ logP判定: ○	0.441	0.473	0.864	0.851	18	13
構造C判定: ○△ logP判定: ×	1.51	1.76	-7.25	-338	7	4
構造C判定: × logP判定: ○	1.08	1.22	-0.39	-0.79	52	31
構造C判定: × logP判定: ×	1.97	2.21	-1.27	-0.47	17	7

有効なクラス

$R^2 > 0.7$

かつ

$p(\text{slope}) < 0.05$

外部バリデーション結果②

藻類急性

	RMSE		Qext ²		物質数	
	全クラス	有効なクラス	全クラス	有効なクラス	全クラス	有効なクラス
構造C判定: ○△ logP判定: ○	0.564	0.752	0.599	0.315	14	7
構造C判定: ○△ logP判定: ×	0.365	-	-47.2	-	2	-
構造C判定: × logP判定: ○	0.972	1.29	-0.31	-0.96	36	18
構造C判定: × logP判定: ×	1.65	2.03	-3.05	-699	18	6

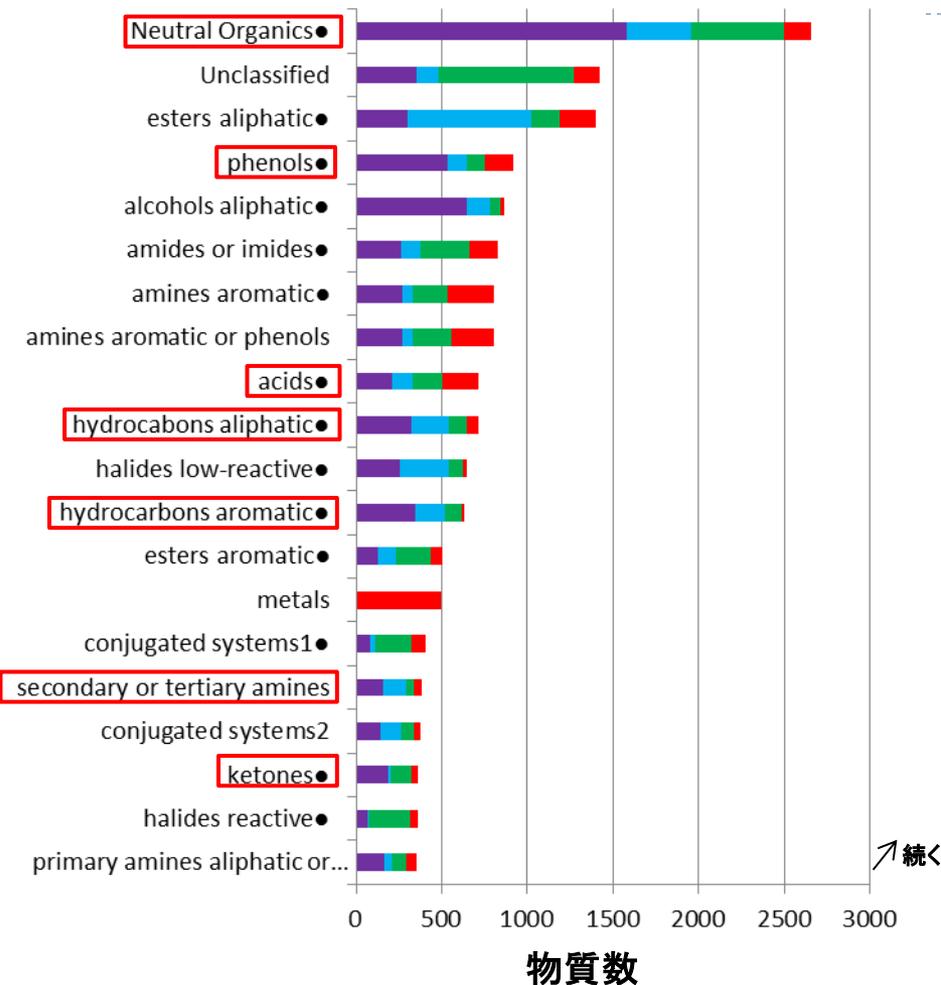
藻類慢性

	RMSE		Qext ²		物質数	
	全クラス	有効なクラス	全クラス	有効なクラス	全クラス	有効なクラス
構造C判定: ○△ logP判定: ○	0.679	0.545	0.624	0.728	27	18
構造C判定: ○△ logP判定: ×	-	-	-	-	-	-
構造C判定: × logP判定: ○	1.33	1.43	-0.01	-0.07	68	51
構造C判定: × logP判定: ×	2.11	2.51	-4.24	-13.3	16	9

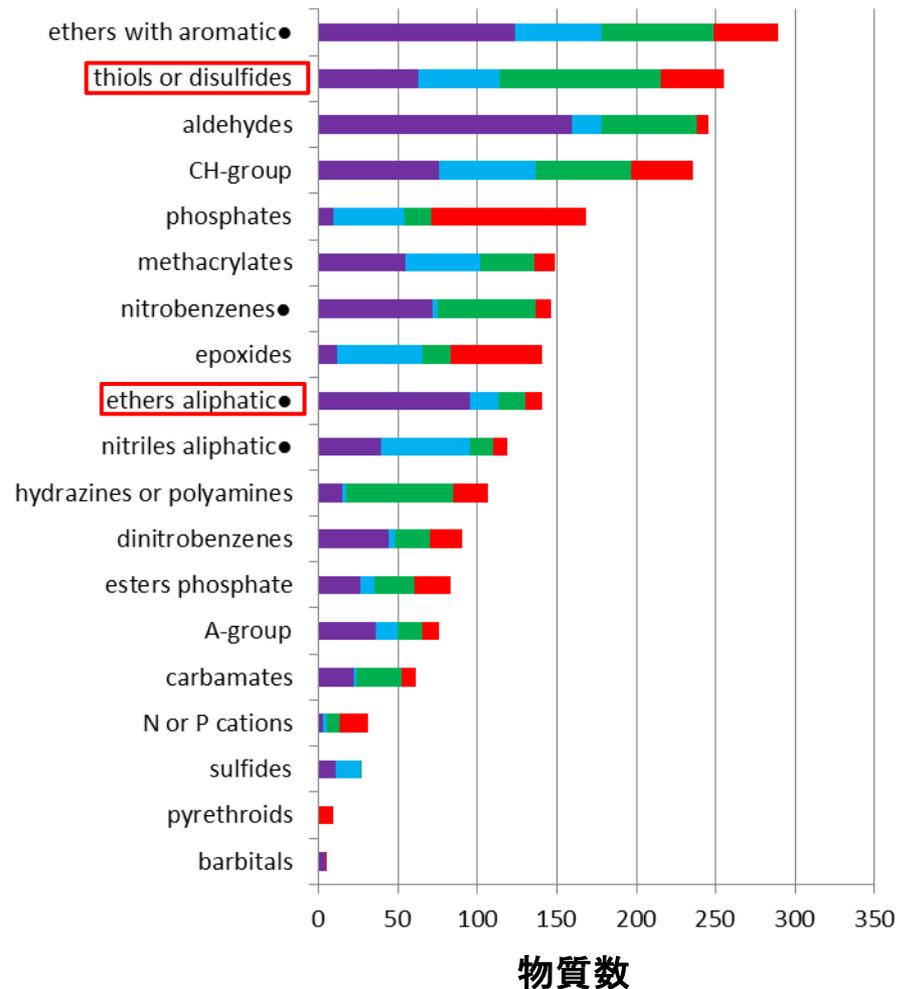
有効なクラス
 $R^2 > 0.7$
 かつ
 $p(\text{slope}) < 0.05$

既存化学物質のクラス分類（魚類急性の例）

構造C判定: ○△、logP判定: ○ 構造C判定: ○△、logP判定: × 構造C判定: ×、logP判定: ○ 構造C判定: ×、logP判定: ×



↑ 続く

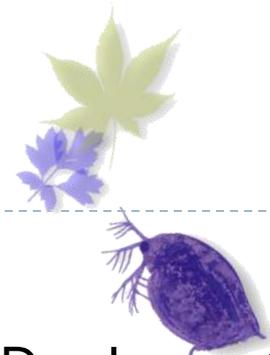


●: QSAR式が $R^2 > 0.7$ かつ $p(\text{slope}) < 0.05$ の条件を満たすクラス
 □: 外部バリデーションの際に、構造C判定が○又は△、logP判定が○の物質を含むクラス

まとめ

- ▶ ECOSAR、TIMESとKATEで、麻酔作用によるクラスは同等の予測結果が得られた。
- ▶ 新たな記述子や参照物質の導入、構造分類の方法の見直しと、外部バリデーションを行った。
- ▶ 甲殻類慢性及び藻類急性・慢性のエンドポイントについて、logPを記述子としたQSARが一部のクラスに関し構築可能であった。
- ▶ R^2 及び $p(\text{slope})$ を用いて有効なクラスに限定すれば、予測の信頼性が向上した。

ご静聴ありがとうございました



参考文献

- 1 A. Furuhamu, T. Toida, N. Nishikawa, Y. Aoki, Y. Yoshioka, H. Shiraishi, Development of an ecotoxicity QSAR model for the KAshinhou Tool for Ecotoxicity (KATE) system, March 2009 version, SAR QSAR Environ. Res., 21 (2010), pp. 403-413.
- 2 A. Furuhamu, K. Hasunuma, Y. Aoki, Y. Yoshioka and H. Shiraishi, Application of chemical reaction mechanistic domains to an ecotoxicity QSAR model, KAshinhou Tool for Ecotoxicity (KATE), SAR QSAR Environ. Res., 22 (2011), 505-523.
- 3 A. Furuhamu, Y. Aoki, and H. Shiraishi, Development of ecotoxicity QSAR models based on partial charge descriptors for acrylate and related compounds, SAR QSAR Environ. Res., in press.
- 4 吉岡義正「QSAR用プラットフォーム(PAS)作成のノウハウ-1 SMILES式からの要素の抽出-」環境毒性学会誌 Vol. 11 (2008), pp. 33-40 .
- 5 吉岡義正「QSAR用プラットフォーム(PAS)作成のノウハウ-2 構造図の描画-」環境毒性学会誌 Vol. 12 (2009) , pp. 107-112.
- 6 吉岡義正「QSAR用プラットフォーム(PAS)作成のノウハウ-3 部分構造の取得法-」環境毒性学会誌 Vol. 12 (2009), pp. 113-122.
- 7 古濱彩子、白石寛明「化学物質の生態毒性予測システムKATEとQSAR」日本化学会情報化学部会誌 Vol. 30 (2012), pp.42-45.