

グラファイトカーボンブラックベースの 吸着剤を前処理に用いた GC/MS、LC/MS による食品残留農薬の 回収率測定結果に対する、分子記述子を 説明変数としたランダムフォレストによる 考察

著者

芹野 武、山下 和之、
杉立 久仁代、澤田 有司

アジレント・テクノロジー
株式会社

要旨

食品中残留農薬規制では対象となる農薬が多岐に渡り、多様な化学的特性を有します。GC/MS や LC/MS の測定では食品のマトリクスを除去し、測定対象となる残留農薬を保持する固相抽出が用いられています。食品中残留農薬の前処理で使用される吸着剤にはさまざまな種類がありますが、グラファイトカーボンブラックベースの吸着剤が残留農薬分析の前処理で用いられた場合、農薬によっては分子内の平面構造がグラファイトの結晶に強く捕捉されることで回収率が非常に低下することが経験上知られています¹⁾。そのような回収率の低下を化合物の分子の特性をもとに理解する手法として定量的構造物性相関 (QSPR : Quantitative Structure-Property Relationship) があります。QSPR は液体クロマトグラフィーのリテンションタイム予測²⁾ やガスクロマトグラフィーによる化合物の物理化学特性の予測³⁾、食品残留農薬分析⁴⁾ などで用いられている手法です。このアプリケーションノートでは食品中の残留農薬分析の QSPR の適用事例として、溶媒スタンダードおよびほうれん草のマトリクスを添加した試料の農薬回収率のデータに対して、農薬の分子記述子を用いて、グラファイトカーボンブラックベースの吸着剤に吸着されて回収率が低くなる農薬の化学的特性を分子記述子の観点から考察を行いました。

実験

農薬名、測定手法（GC/MS あるいは LC/MS）および農薬の Canonical SMILES を表 1 に示します。

GC/MS, LC/MS の農薬回収率のデータ

農薬混合標準液（関東化学社製農薬混合標準液 63、GC/MS 用、農薬混合標準品 76、LC/MS 用）をそれぞれアセトニトリルで希釈し、100 ng/mL 濃度とした溶媒スタンダード（SS：Solvent Standard）およびほうれん草マトリクス（SP：SPinach matrix）の 2 種類の試験液を作成しました。GC/MS, LC/MS の測定条件はアジレント・テクノロジー株式会社アプリケーションノート 5994-6055JAJP に記載した条件で行いました。

表 1. 測定農薬リスト

農薬	測定手法 G : GC/MS L : LC/MS	SMILES
Dichlorvos	G	<chem>COP(=O)(OC)OC=C(Cl)Cl</chem>
Butylate	G	<chem>CCSC(=O)N(CC(C)C)CC(C)C</chem>
Isoprocarb	G	<chem>CC(C)C1=CC=CC=C1OC(=O)NC</chem>
Ethoprophos	G	<chem>CCCSP(=O)(OCC)SCCC</chem>
Bendiocarb	G	<chem>CC1(OCC2=C(O1)C(=CC=C2)OC(=O)NC)C</chem>
Terbufos	G	<chem>CCOP(=S)(OCC)SCSC(C)(C)C</chem>
Tefluthrin	G	<chem>CC1=C(C(=C(C=C1F)F)COC(=O)C2C(C2(C)C)C=C(C(F)(F)F)Cl)F)F</chem>
BHC (d)	G	<chem>C1(C(C(C(C(C1Cl)Cl)Cl)Cl)Cl)Cl</chem>
Ethiofencarb	G	<chem>CCSCC1=CC=CC=C1OC(=O)NC</chem>
Tolclofos-methyl	G	<chem>CC1=CC(=C(C(=C1)Cl)OP(=S)(OC)OC)Cl</chem>
Pirimiphos-methyl	G	<chem>CCN(CC)C1=NC(=CC(=N1)OP(=S)(OC)OC)C</chem>
Methiocarb	G	<chem>CC1=CC(=CC(=C1SC)C)OC(=O)NC</chem>
Malathion	G	<chem>CCOC(=O)CC(C(=O)OCC)SP(=S)(OC)OC</chem>
Metolachlor	G	<chem>CCC1=CC=CC(=C1N(C(C)COC)C(=O)CC)Cl</chem>
Diethofencarb	G	<chem>CCOC1=C(C=C(C=C1)NC(=O)OC(C)C)OCC</chem>
Dimethylvinphos (Z)	G	<chem>COP(=O)(OC)O/C(=C\Cl)/C1=C(C=C(C=C1)Cl)Cl</chem>
Isofenphos oxon	G	<chem>CCOP(=O)(NC(C)C)OC1=CC=CC=C1C(=O)OC(C)C</chem>
Chlorfenvinphos (a) (E)	G	<chem>CCOP(=O)(OCC)O/C(=C/Cl)/C1=C(C=C(C=C1)Cl)Cl</chem>
Isofenphos	G	<chem>CCOP(=S)(NC(C)C)OC1=CC=CC=C1C(=O)OC(C)C</chem>
Chlorfenvinphos (b) (Z)	G	<chem>CCOP(=O)(OCC)O/C(=C\Cl)/C1=C(C=C(C=C1)Cl)Cl</chem>
Quinalphos	G	<chem>CCOP(=S)(OCC)OC1=NC2=CC=CC=C2N=C1</chem>
Triadimenol	G	<chem>CC(C)(C)C(C(N1C=NC=N1)OC2=CC=C(C=C2)Cl)O</chem>
Triadimenol II	G	<chem>CC(C)(C)C(C(N1C=NC=N1)OC2=CC=C(C=C2)Cl)O</chem>
Chinomethionate	G	<chem>CC1=CC2=C(C=C1)N=C3C(=N2)SC(=O)S3</chem>

Paclobutrazol	G	CC(C)(C)C(C(C1=CC=C(C=C1)Cl)N2C=NC=N2)O
Flutolanil	G	CC(C)OC1=CC=CC(=C1)NC(=O)C2=CC=CC=C2C(F)(F)F
Pretilachlor	G	CCCOCN(C1=C(C=CC=C1CC)CC)C(=O)CCl
DDE (4,4')	G	C1=CC(=CC=C1C(=C(Cl)Cl)C2=CC=C(C=C2)Cl)Cl
flusilazole	G	C[Si](CN1C=NC=N1)(C2=CC=C(C=C2)F)C3=CC=C(C=C3)F
fensulfotion	G	CCOP(=S)(OCC)OC1=CC=C(C=C1)S(=O)C
Propiconazole I	G	CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl
lenacil	G	C1CCC(CC1)N2C(=O)C3=C(CCC3)NC2=O
Propiconazole II	G	CCCC1COC(O1)(CN2C=NC=N2)C3=C(C=C(C=C3)Cl)Cl
thenylchlor	G	CC1=C(C(=CC=C1)C)N(CC2=C(C=CS2)OC)C(=O)CCl
captafol	G	C1C=CCC2C1C(=O)N(C2=O)SC(C(Cl)Cl)(Cl)Cl
Acetamidrid	G	CC(=NC#N)N(C)CC1=CN=C(C=C1)Cl
phosalone	G	CCOP(=S)(OCC)SCN1C2=C(C=C(C=C2)Cl)OC1=O
Cyhalothrin I	G	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)/C=C(\C(F)(F)F)/Cl)C
mefenacet	G	CN(C1=CC=CC=C1)C(=O)COC2=NC3=CC=CC=C3S2
Cyhalothrin II	G	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)/C=C(\C(F)(F)F)\Cl)C
fenarimol	G	C1=CC=C(C=C1)C(C2=CC=C(C=C2)Cl)(C3=CN=CN=C3)O)Cl
Bitertanol I	G	CC(C)(C)C(C(N1C=NC=N1)OC2=CC=C(C=C2)C3=CC=CC=C3)O
Bitertanol II	G	CC(C)(C)C(C(N1C=NC=N1)OC2=CC=C(C=C2)C3=CC=CC=C3)O
pyridaben	G	CC(C)(C)C1=CC=C(C=C1)CSC2=C(C(=O)N(N=C2)C(C)(C)Cl)Cl
Cypermethrin I	G	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(Cl)Cl)C
Cypermethrin II	G	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(Cl)Cl)C
Flucythrinate	G	CC(C)C(C1=CC=C(C=C1)OC(F)F)C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3
Cypermethrin III	G	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(Cl)Cl)C
Cypermethrin IV	G	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(Cl)Cl)C
Flucythrinate II	G	CC(C)C(C1=CC=C(C=C1)OC(F)F)C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3
pyrimidifen	G	CCC1=C(C(=NC=N1)NCCOC2=C(C=C(C=C2)CCOCC)C)C)Cl
Fluvalinate-tau-I	G	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl
Fluvalinate-tau-II	G	CC(C)C(C(=O)OC(C#N)C1=CC(=CC=C1)OC2=CC=CC=C2)NC3=C(C=C(C=C3)C(F)(F)F)Cl
deltamethrin	G	CC1(C(C1C(=O)OC(C#N)C2=CC(=CC=C2)OC3=CC=CC=C3)C=C(Br)Br)C
Acephate	L	CC(=O)NP(=O)(OC)SC
Pymetrozine	L	CC1=NNC(=O)N(C1)/N=C/C2=CN=CC=C2
Omethoate	L	CNC(=O)CSP(=O)(OC)OC
Carbendazim	L	COC(=O)NC1=NC2=CC=CC=C2N1
Sethoxydim	L	CCC/C(=N\OCC)/C1=C(CC(CC1=O)CC(C)SCC)O
Carbetamide	L	CCNC(=O)[C@@H](C)OC(=O)NC1=CC=CC=C1
Ametryn	L	CCNC1=NC(=NC(=N1)SC)NC(C)C
Pyrimethanil	L	CC1=CC(=NC(=N1)NC2=CC=CC=C2)C
Oxabetrinil	L	C1COC(O1)CON=C(C#N)C2=CC=CC=C2

Propachlor	L	<chem>CC(C)N(C1=CC=CC=C1)C(=O)CCl</chem>
Metominostrobin(E,Z)	L	<chem>CNC(=O)/C(=N/OC)/C1=CC=CC=C1OC2=CC=CC=C2</chem>
Demeton-S-methyl sulfone	L	<chem>CCS(=O)(=O)CCSP(=O)(OC)OC</chem>
Dimethenamid	L	<chem>CC1=CSC(=C1N(C(C)COC)C(=O)CCl)C</chem>
Tepraloxydim	L	<chem>CC/C(=N\OC/C=C/C)/C1=C(CC(CC1=O)C2CCOCC2)O</chem>
Fentrazamide	L	<chem>CCN(C1CCCCC1)C(=O)N2C(=O)N(N=N2)C3=CC=CC=C3Cl</chem>
Coumaphos	L	<chem>CCOP(=S)(OCC)OC1=CC2=C(C=C1)C(=C(C(=O)O2)Cl)C</chem>
Amitraz	L	<chem>CC1=CC(=C(C=C1)N=CN(C)C=NC2=C(C=C(C=C2)C)C)C</chem>
Benfuracarb	L	<chem>CCOC(=O)CCN(C(C)C)SN(C)C(=O)OC1=CC=CC2=C1OC(C2)(C)C</chem>
Fluazifop-buthyl	L	<chem>CCCCOC(=O)C(C)OC1=CC=C(C=C1)OC2=NC=C(C=C2)C(F)(F)F</chem>
Flumiclorac pentyl	L	<chem>CCCCCOC(=O)COC1=C(C=C(C(=C1)N2C(=O)C3=C(C2=O)CCCC3)F)Cl</chem>
Cinosulfuron	L	<chem>COCCOC1=CC=CC=C1S(=O)(=O)NC(=O)NC2=NC(=NC(=N2)OC)OC</chem>
Chlorfluazuron	L	<chem>C1=CC(=C(C(=C1)F)C(=O)NC(=O)NC2=CC(=C(C(=C2)Cl)OC3=C(C=C(C(=N3)C(F)(F)F)Cl)Cl)F</chem>
Carbosulfan	L	<chem>CCCCN(CCCC)SN(C)C(=O)OC1=CC=CC2=C1OC(C2)(C)C</chem>

測定条件

データ解析

農薬の回収率の吸着特性が SS と SP で異なるため、階層型クラスタ分析にて SS、SP の回収率のプロファイルを階層型クラスタ分析で分類しました。本アプリケーションノートでは試料由来のマトリクス効果などを排除した農薬で考察を行うため、SS、SP でいずれも吸着、非吸着となる農薬のクラスタをそれぞれ“吸着群”と“非吸着群”とし、解析対象の農薬としました。

SS および SP のデータセットに含まれる 72 農薬の Canonical SMILES より、R プログラムの rcdk パッケージを用いて分子記述子を入手しました⁴⁾。得られた分子記述子の中から、吸着、非吸着の現象を特徴づける分子記述子を機械学習の手法で探索し、分子記述子の定義より解釈をおこないました。

本アプリケーションノートでは、機械学習の変数選択手法としてランダムフォレストを用いました。ランダムフォレストは弱学習器を複数組み合わせたアンサンブル学習の手法で、多くの決定木と呼ばれる分類器を組み合わせたものです。決定木を作成する手法である CART は Classification And Regression Trees の略で、Breiman などが 1984 年に導入した統計手法です⁵⁾。CART はデータを変数のある基準で分割して分類していく方法で決定木を作成していきます。例として、R プログラムの rpart パッケージを用いて、今回測定した 72 農薬を 165 の分子記述子を用いて CART で分類した結果を図 1 に示します。

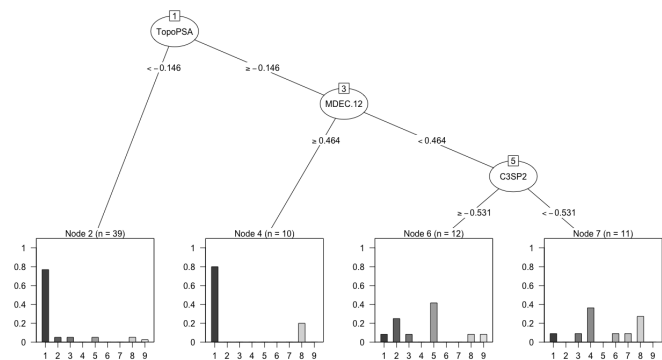


図 1. 分子記述子を用いた決定木の例

ランダムフォレストは「決定木」と「アンサンブル学習 (バギング)」の 2 つの手法を組み合わせたアルゴリズムで、決定木単体で使うよりも高い精度を出せる特徴があるだけでなく、予測モデル作成で使用した変数の重要度である Variable Importance (VI) が得られます。VI は説明変数が観測値に影響する度合いを数値にしたもので、線形モデルなどのような事前に想定するモデルが不要で、多数の変数があるデータに対して探索的に変数の重要度を求めるのに理想的な手法です。

本アプリケーションでは農薬の回収率をもとにした“吸着群”と“非吸着群”のグループを目的変数としました。農薬グループのそれぞれの分子記述子を説明変数としました。R プログラムの VSURF パッケージ⁶⁾ によってランダムフォレストにて分類モデルを構築し、分子記述子の重要度である VI の高い分子記述子を選択し、農薬の吸着特性を考察した結果を紹介します。

結果

農薬の回収率

SS および SP の 72 農薬の回収率の分布を図 2 に示します。SS は 1 % ~144 %、SP は 1 % ~188 % の範囲でした。

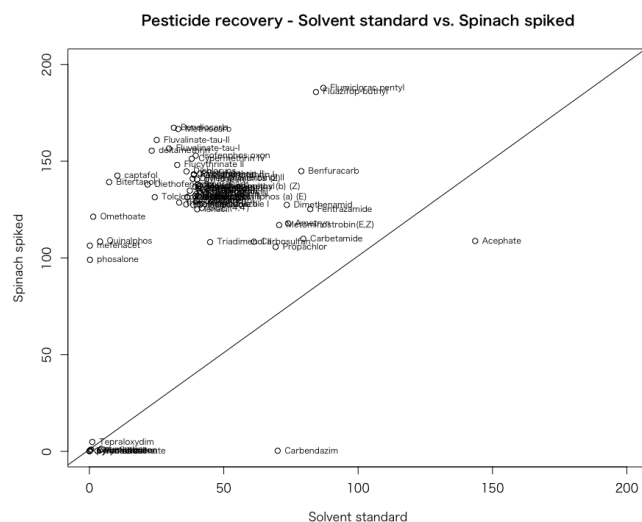


図 2. 72 農薬の溶媒スタンダード (横軸) およびほうれん草マトリクス (縦軸) の回収率

階層型クラスタ分析と最適なクラスタ分割数

SS および SP の回収率の結果に対して、農薬の回収率の変動パターン類似度で階層型クラスタ分析を行った結果を図 3 に、その結果に対して、Gap 関数を適用した結果を図 4 に示します。

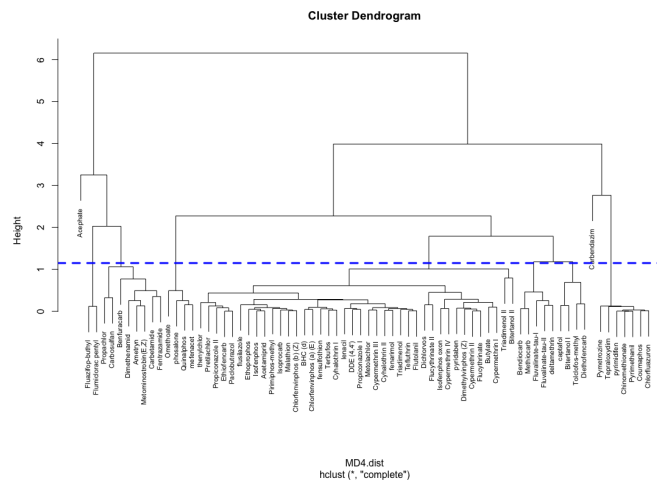


図 3. 階層型クラスタ分析のデンドログラム

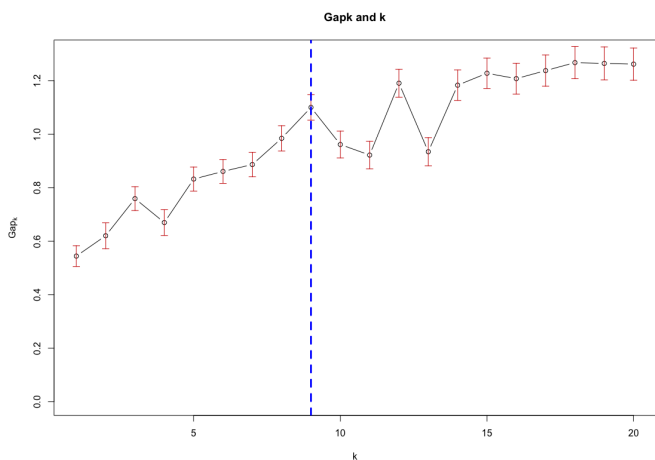


図 4. Gap 関数を適用した結果

Gap 関数は Gap 統計量を用いて数値的に分割するグループの個数を求める目安を求めることができる手法です。Gap 関数において Gap 統計量の変化が大きいたところが最適クラスターとなります。クラスターの分割数 k が大きいほど Gap 統計量は高くなり、1 つのクラスター内の均一性が増えるものの、 k が増えるほど解釈性が低下するというトレードオフが存在します。Gap 統計量が高く、なおかつできるだけ少ない分割数 (k) を選択することにより、このトレードオフを考慮した最適なクラスター分割数を選択します。本データでは 図 4 の結果より $k=9$ を吸着群と非吸着群のクラスター分割数としました。その結果を図 3 および図 4 の青破線に示します。クラスターごとにプロットで色分けした SS および SP の回収率の散布図を図 5 に示します。

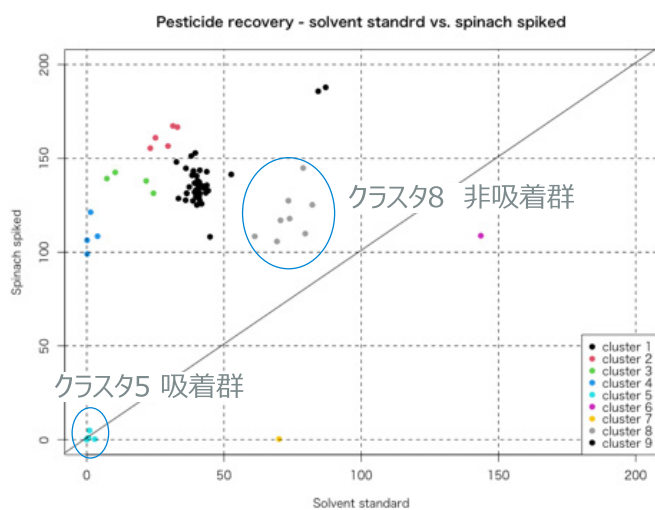


図 5. SS および SP の回収率の散布図 (クラスターで色分け)

図 5 より、クラスター 5 の 7 農薬 (図 6) は SS および SP において回収率がほぼ 0% 付近であり、吸着剤に吸着されて非常に低い回収率を示しています。一方で、クラスター 8 の 8 農薬 (図 7) は SS および SP いずれにおいても良好な回収率が得られています。本アプリケーションノートでは以降、クラスター 5 の農薬グループを吸着群、クラスター 8 の農薬グループを非吸着群と表記します。

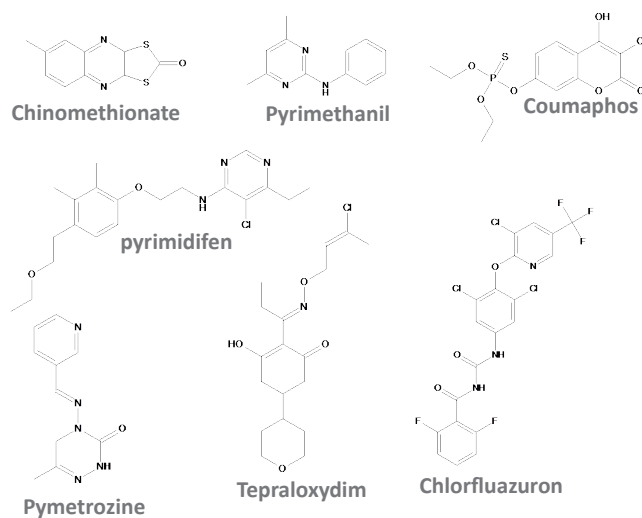


図 6. Cluster 5 (吸着群) の農薬

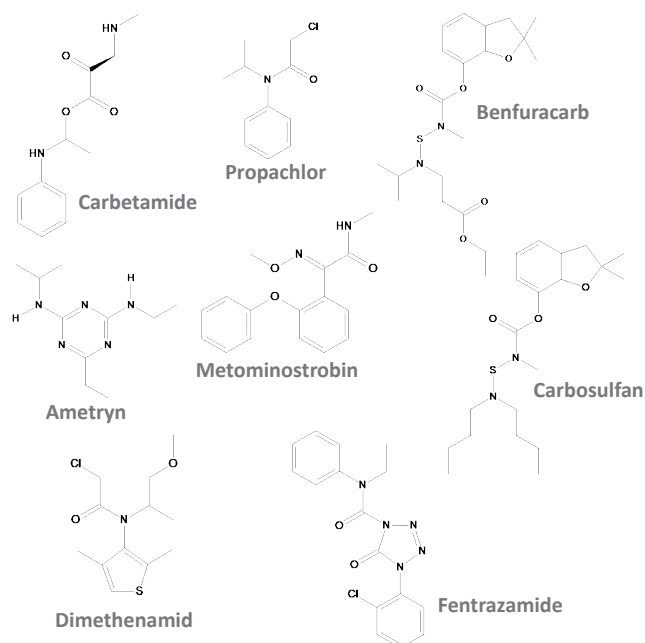


図 7. Cluster 8 (非吸着群) の農薬

ランダムフォレストを用いた分子記述子の選択

吸着群と非吸着群の農薬の違いを分子記述子の観点から考察を行う方法として、機械学習のランダムフォレスト法による変数選択を行いました。72 農薬の Canonical SMILES から rcdk パッケージを用いて 166 の分子記述子を取得しました。目的変数をクラスタ（吸着群または非吸着群）とし、166 の分子記述子を説明変数として、分類の機械学習モデルをランダムフォレストにて作成し、変数の重要度である VI およびその VI をランダムな値にした時に増加する分類誤差の Out of Bag error (OOB エラー) によって変数選択をおこなった結果を図 8 に示します。また、VSURF の各ステップにおける選択された分子記述子の数を表 2 に示します。

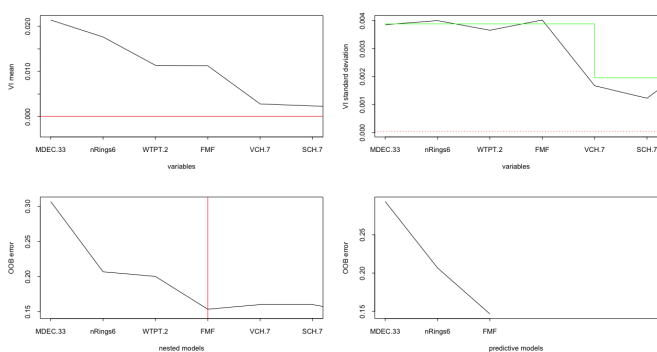


図 8. VSURF パッケージによる分子記述子の選択結果 (上位 6 つの VI の分子記述子を表示)

表 2. VSURF によって選択された分子記述子数

Step	分子記述子数
VSURF に供した変数	166
Threshold	29
Interpretation	4
Prediction	3

VSURF はランダムフォレストにより各変数の VI の大きさおよび、変数を追加した際の OOB エラーを元に変数を選択する R プログラムのパッケージです。⁵⁾

VSURF では最初に Threshold のステップで 50 個の判別モデルの決定木をランダムフォレストで作成し、著しく VI の小さい (分類に与える影響が小さい) 分子記述子を除外するための VI の Threshold を設定し、重要度の低い変数を除外します。本データセットでは 29 の分子記述子が Threshold の段階で選択され、残りの 136 の分子記述子が Threshold の段階で除外されました。

図 8 の左上の図では 165 の分子記述子の中から VI の上位 6 つが、左から VI の大きな順で配置されています。VI の値が大きな分子記述子は、吸着群と非吸着群の分類に大きな影響を与えます。この図中に Threshold が赤横線で示されており、この赤線以下の VI の分子記述子は Threshold のステップで除外されます。

図 8 の右上の図はランダムフォレストでモデルを構築した際の各分子記述子の VI の標準偏差が示されています。一般的に VI が小さくなるとランダムフォレストにより算出される VI の標準偏差も小さくなることから、標準偏差も VI 選択の目安となります。VI の標準偏差の変化の目安としてステップ関数の結果が黄緑色の線で表示されています。

さらに、VSURF では結果の解釈 (Interpretation) で使用する変数と予測 (Prediction) で使用する変数の結果が得られます。この Interpretation のステップでは 25 の判別モデルの決定木がランダムフォレストで作成されます。

図 8 の左下の図は Interpretation の分子記述子が表示されています。Interpretation のステップでは VI を降順に追加していき、最も低い OOB エラーとなる VI の分子記述子まで選択します。OOB エラーはランダムフォレストの分類モデル (決定木) を作成する過程において、Bootstrap で選択されなかったサンプルで分類モデルを作成して得られた誤差 (誤分類率) です。当てはまりの良い予測モデルほど OOB エラーは低い値となります。本データセットでは 4 つの分子記述子が Interpretation で選択されました。

図 8 の右下の図は Prediction (予測) で用いる変数が選択されています。Prediction では Interpretation で選択された変数に対して、VI の降順に OOB エラーの減少率を指標にして追加していきます。変数を追加した場合の OOB エラーが、全くランダムな変数を追加した時よりも減少するという考えで変数が選択されます。本データセットでは Prediction のステップにおいて 3 つの分子記述子が選択されました

本データの Prediction で選択された 3 つの分子記述子の定義およびその説明を Table 3 に、各クラスタの分子記述子の分布の箱ひげ図を図 9 に示します。

表 3. VSURF の Prediction ステップで選択された分子記述子の定義と説明

分子記述子	定義	説明
MDEC.33	molecular distance edge between all tertiary carbons	分子内の四級炭素の距離
nRings6	Number of 6-member rings	分子内の 6 員環の数
FMF	The descriptor is the ratio of heavy atoms in the Murcko framework to the total number of heavy atoms in the molecule.	分子内で Murcko フレームワーク (側鎖を取り払って環構造およびそれらをつないでいるリンカーのみとした構造) を構成している原子の割合

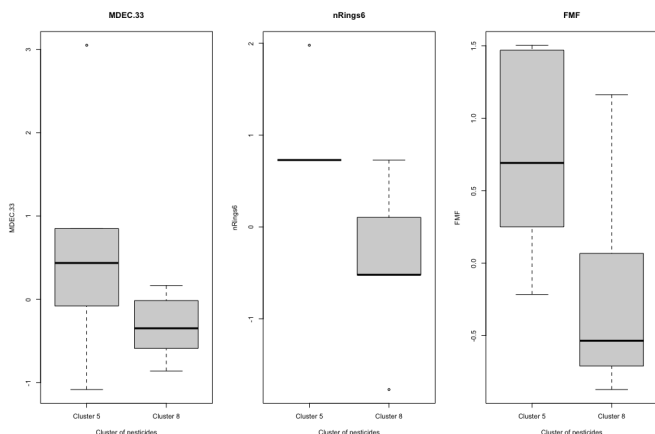


図 9. 分子記述子と農薬回収率の箱ひげ図。クラスタ 5 (左) が吸着群、クラスタ 8 (右) が非吸着群

表 3 の結果から、吸着群の農薬は以下の 2 つの影響によることが推察されます。

- 分子内の環による平面構造：nRings6, FMF
- 立体構造を持つ四級炭素の分子内の距離：MDEC.33

吸着群の農薬は経験的に解釈されていた平面構造を持つ環構造 (nRings6、FMF) が大きい値を取り、経験則を裏付けることが確認されました。さらに、吸着群の農薬ではその平面構造の影響を阻害する正四面体構造の SP3 炭素原子間の距離 (MDEC.33) が長い傾向にあることが確認できました。

すなわち、MDEC.33 の数値が吸着群の方で非吸着群に比べて相対的に大きな値を取り、その間にある 6 員環などの環構造 (nRings6、FMF) による平面構造による影響を受けやすくなり、その結果としてグラファイトカーボンブラックベースの吸着剤にトラップされていると考えられます。

まとめ

ランダムフォレストにより、吸着特性の違いを表す分子記述子を選択し、その結果からグラファイトカーボンブラックベースの吸着剤で回収率が著しく低くなる農薬について分子構造の観点で考察を行いました。

吸着群の農薬は、経験的に解釈されていた平面構造を持つ環構造 (nRings6、FMF) の存在が確認され、また、その影響を阻害する SP3 炭素原子間の距離が長いものが多い傾向にある (MDEC.33) ことが QSPR の手法で観測されました。今回得られた知見を元にしてその他の農薬の回収率の予測などへの活用が期待されます。

参考文献

- 1: 三重保環研年報 第13号(通巻第56号), 44 - 59頁(2011)
- 2: Andrew D. McEachrana et. al. Talanta Vol. 182, 371-379 (2018)
- 3: Alan R. Katritzky et. al. Chem. Soc. Rev., 24, 279-287 (1995)
- 4: Takeshi Serino et. al. Journal of Computer Aided Chemistry, Vol.20, 92-103 (2019)
- 5: Robin Genuer, Jean-Michel Poggi and Christine Tuleau-Malot. The R Journal Vol. 7/2 (2015)

ホームページ

www.agilent.com/chem/jp

カスタムコンタクトセンター

0120-477-111

email_japan@agilent.com

本製品は一般的な実験用途での使用を想定しており、医薬品医療機器等法に基づく登録を行っておりません。本文書に記載の情報、説明、製品仕様等は予告なしに変更されることがあります。

DE87957426

アジレント・テクノロジー株式会社
© Agilent Technologies, Inc. 2024
Printed in Japan, March 11, 2024
5994-7238JAJP