

原 著

機械学習回帰モデルによる養豚排水中窒素化合物濃度の高精度予測

田中康男

(一財) 畜産環境整備機構畜産環境技術研究所, 福島県西白河郡, 961-8061

要 約 養豚農場の汚水処理においては、水質汚濁防止法の規制項目である「アンモニア、アンモニウム化合物、亜硝酸化合物及び硝酸化合物」(以下「硝酸性窒素等」と略す)の遵守が求められる。窒素除去を的確に進めるうえで、硝酸性窒素等の濃度と硝化の進行程度とを日常的に把握することが重要である。しかし、依頼分析は経費と時間を要するので、現場で自ら状況把握できる手法が望まれる。pHと電気伝導率(EC)を計測して、重回帰法で硝酸性窒素等濃度を予測する手法が報告されているが、精度向上が課題であった。そこで、pH、ECに加えて、試料採取日の月数、試料採取農場の緯度と経度、試料採取地域における採取月の平均気温、処理施設の日排水量の合計7つの特徴量を設定し、機械学習法(サポートベクターマシン法、ランダムフォレスト法、ニューラルネットワーク法、キュビスト法)の回帰モデルを利用することで予測精度を向上させることを試みた。この結果、pHとECに加えて他の特徴量も用いて機械学習法による予測を行うと精度が向上することが確認された。また、7特徴量を用いたキュビスト法は予測誤差(平均絶対誤差)が最低になったことから、最適な予測法であると推測された。

キーワード: 養豚汚水、浄化処理、硝酸性窒素等、アンモニア性窒素、機械学習回帰モデル

受領日: 22.09.2025. 受理日: 23.01.2026.

日本畜産環境学会会誌

緒 言

水質汚濁防止法で「アンモニア、アンモニウム化合物、亜硝酸化合物及び硝酸化合物」(以下「硝酸性窒素等」と略す)の規制が開始されて以降、窒素除去性能の向上が養豚汚水処理の重要課題となっている。

浄化施設の窒素除去性能を向上するためには、頻繁に処理水の硝酸性窒素等濃度と硝化の進行程度とを把握しつつ処理施設を調整することが必要である。しかし、公定法による硝酸性窒素等濃度測定は分析会社でないと対応できず、経費面から測定頻度には限界がある。現場において管理者自ら容易に硝

酸性窒素等濃度および硝化の進行程度の予測が可能になれば、汚水処理施設をよりの確に管理できるようになり、基準値の確実な遵守に資すると考えられる。

現場での測定が容易なpHと電気伝導率(EC)を特徴量とする重回帰法で硝酸性窒素等濃度が予測可能という知見がすでに報告されているものの[13]、予測精度が十分とはいえなかった。

そこで今回、pHとECに加えて試料採取日の月数、農場の緯度と経度、農場所在地域と日付から予測した気温、処理施設日排水量の合計7特徴量を回帰モデルの作成に利用する

機械学習による窒素化合物濃度の予測

ことで予測精度の向上が可能かどうかを検討した。また、重回帰法に加えて機械学習法の4種類のアルゴリズム(サポートベクターマシン法、ランダムフォレスト法、ニューラルネットワーク法、キュビスト法)についての予測精度の比較検討も行った。これらのアルゴリズムは原理が大きく異なり[1]、予測精度はデータ特性に左右される。そのため、性能比較には実データでの検証が不可欠である。

以上の背景にもとづき、既存の水質データを利用して予測法の基礎検討を行ったのでここに報告する。

材料および方法

1. 試料採取

養豚農場の活性汚泥法污水处理施設の処理水を放流槽または沈殿槽で採取した。回分式の処理施設で、沈殿槽や放流槽が設置されていない場合は、曝気停止時間に曝気槽の上澄水を処理水として採取した。無機凝集剤による凝集沈殿処理が行われている農場は、pH、EC値が凝集剤の影響を受けるため採取対象外とした。採取処理水は水質分析までは冷蔵保管した。

なお、試料採取および分析は、畜産環境技術研究所が独自に実施したものと、中央畜産会が調査事業で実施したものと2種類に分かれる。

2. 水質分析方法

畜産環境技術研究所が採取した試料の分析は以下の方法によった。pHは下水試験方法(日本下水道協会[10])によりガラス電極法(YOKOGAWA Personal PH/ORP Meter (MODEL-PH72))で測定した。電気伝導率(EC)はJIS-K0130:2008に基づき交流2極法(堀場製作所製コンパクト電気伝導率計B-771)で測定した。アンモニア性窒素(NH₄-N)、亜硝酸性窒

素(NO₂-N)および硝酸性窒素(NO₃-N)は日本下水道協会[10]によりイオンクロマトグラフ法(島津製作所製SCL-10A_{MP})により測定した。中央畜産会事業で採取した試料の水質分析は、JIS K0102により行った。

硝酸性窒素等の濃度は、水質汚濁防止法に基づき式1で算出した。

$$\text{硝酸性窒素等 (mg/L)} = \text{NH}_4\text{-N (mg/L)} \times 0.4 + \text{NO}_2\text{-N (mg/L)} + \text{NO}_3\text{-N (mg/L)} \cdots \text{式1}$$

3. データ解析

データ解析は、統計解析用のプログラミング言語であるR[12]と、Rの統合開発環境であるRStudio[11]によって行なった。解析に用いたパッケージおよび関数の全ては、Rのインターネットサイトから無料ダウンロードして利用した。

硝酸性窒素等およびアンモニア性窒素予測用の特徴量としては、pH、ECの他に下記5種を候補とした。

- ・試料採取日の月数:鹿熊[9]によると、全国の月別肉豚集荷頭数は120~155万頭の間で周年変動し、3~4月と10~12月に上昇すると報告されている。この変動が污水处理施設に負荷変動をもたらし、処理水質にも影響する可能性がある。したがって、月数(1~12)も処理水質と関連を持つ可能性がある。

- ・農場の緯度と経度:農場の位置は気温と関係を有することから間接的に水温および水質と関連する可能性がある。

- ・農場所在地域と日付から予測した気温(推定気温):温度データとしては本来であれば曝気槽水温が適切であるが、採水時の曝気槽水温が記録されていない試料も多い。そこで、全ての試料が保有する情報である採水地域(都道府県名)と採水日の属する月から推定する気温(推定気温)を特徴量に用いた。なお、推定気温と曝気槽水温は当然ながら一致するわけではなく、また両者の相関関

係も地域や処理施設の状況により多様と推測される。したがって、推定気温は曝気槽水温に代わりえるものではないが、採水時の温度条件の相対指標としては意味を有すると思われる。推定気温の具体的な算出法は、気象庁データベースにより県庁所在地における2024年の日平均気温の月平均値の表を作成し、各農場の都道府県名と採水月から表によって気温を特定した。

・処理施設日排水量 (m³/日) : 日排水量は曝気槽容積の相対指標となる。一般に、曝気槽容積が小さいほど低温期の水温が低下する傾向があり、このため小規模ほど年間の水質変動幅が大きくなる。よって、日排水量は水質に関連する可能性がある。

Rによるデータ解析は以下のように行った。

① 学習データとテストデータの分割

データをまず学習データとテストデータに7対3の割合で分割した。この分割にはcaretパッケージのcreateDataPartition関数を用い、目的変数の分布が学習データとテストデータで一致するように層化抽出法で行った。

② 回帰モデル作成

回帰モデルは学習データを用い、重回帰法はlm関数、サポートベクターマシン法はksvm関数、ランダムフォレスト法はrandomForest関数、ニューラルネットワーク法はneuralnet関数、キュビスト法はcubist関数によってそれぞれ作成した。

モデル作成に利用する特徴量の選択は、ラッパー法の一種である交差検証付き再帰的特徴量削減 (RFECV) 法で行った。具体的にはcaretパッケージのrfe関数を用い、特徴量の数と二乗平均平方根誤差 (RMSE) の関係から適正特徴量数を決定した。交差検証は10分割を基本としたが、ニューラルネットワーク法の場合10分割では計算負荷が過大になるため5分割とした。さらに、caretパッケー

ジのpredictors関数によって最適特徴量の種類を特定した。なお、キュビスト法の場合rfe関数が利用できなかったため、cubist関数の実行結果として表示される特徴量利用頻度を参考にして特徴量を選択した。

各回帰モデルのハイパーパラメータの最適値の決定 (チューニング) は、caretパッケージのtrain関数によりリサンプリング法で行った。

③ 予測値の算定

作成した各回帰モデルとテストデータを用いて、predict関数により予測値を算出した。この予測値と実測値の関係を示す散布図を作成するとともに、予測精度を平均絶対誤差 (MAE) で評価した。また、モデル間の予測精度の詳細な比較はMAEの確率密度分布図により行った。

結 果

1. 試料採取農場の所在地と硝酸性窒素等濃度

試料採取農場の所在地、試料数および硝酸性窒素等濃度測定値を表1に示した。試料採取農場数は48で、地域は北海道から宮崎までの16道県に及んだ。農場により試料数は異なり、総試料数は943であった。これらの中で畜産環境技術研究所が独自に採取した試料は847件、中央畜産会事業で採取された試料は96件[2~7]であった。硝酸性窒素等濃度の平均値は12 mg/Lから438 mg/Lまで大きくばらついた。また、同一農場でも採取時期によって大きく変動する場合も見られた。

各農場の硝酸性窒素等濃度平均値の頻度分布を図1に示した。平均値は0~500 mg/Lの範囲で大きくばらついたが、大部分は2026年1月現在養豚に適用されている暫定基準値の400 mg/Lより低かった。最頻値は50~100 mg/Lであった。

機械学習による窒素化合物濃度の予測

表 1. 調査農場の所在地、試料数と硝酸性窒素等濃度の最小、最大および平均値

所在地	農場略称	試料数	硝酸性窒素等(mg/L)		
			最小値	最大値	平均値
北海道	HOK	4	7	34	20
北海道	I	5	15	473	372
北海道	OO	1	12	12	12
青森	A	13	0	244	82
宮城	HY	1	22	22	22
宮城	MIG	3	23	164	97
福島	C	13	25	162	96
福島	SF	304	0	775	177
福島	SFK	20	51	262	177
茨城	MAR	3	20	56	38
栃木	F	8	27	416	172
栃木	NAS	5	81	434	197
栃木	SC	2	206	224	215
栃木	TY	275	5	1,285	386
群馬	ANA	4	27	61	38
群馬	ANO	1	52	52	52
群馬	B	7	80	615	319
群馬	CKM	4	41	138	79
群馬	DA	1	12	12	12
群馬	FRI	47	9	87	51
群馬	G	7	148	402	312
群馬	GUN	4	8	192	62
群馬	KZH	4	309	543	438
群馬	KZM	2	69	72	70
群馬	MAE	14	12	327	96
群馬	MSY	2	128	228	178
群馬	PFI	1	17	17	17
千葉	KO	5	53	165	113
千葉	NA	9	191	425	253
千葉	NAK	13	46	384	233
千葉	SH	22	14	261	94
千葉	TF	1	261	261	261
千葉	TO	31	46	257	166
神奈川	H	6	115	174	140
長野	D	13	3	190	58
岐阜	J	5	12	110	57
三重	KOB	1	23	23	23
京都	KI	9	125	322	196
京都	KIS	9	149	338	234
奈良	OK	9	90	548	296
奈良	OKU	9	115	554	321
熊本	SA	13	64	710	281
宮崎	E	4	230	460	337
宮崎	E2	6	49	166	74
宮崎	FA	1	253	253	253
宮崎	L	6	179	388	266
宮崎	MIZ	3	68	259	156
宮崎	SA	13	57	674	265

2. 特徴量間の相関関係

各特徴量間の相関行列を図2に示した。硝酸性窒素等およびアンモニア性窒素のどちらもECとの相関係数が最も高く、それぞれ

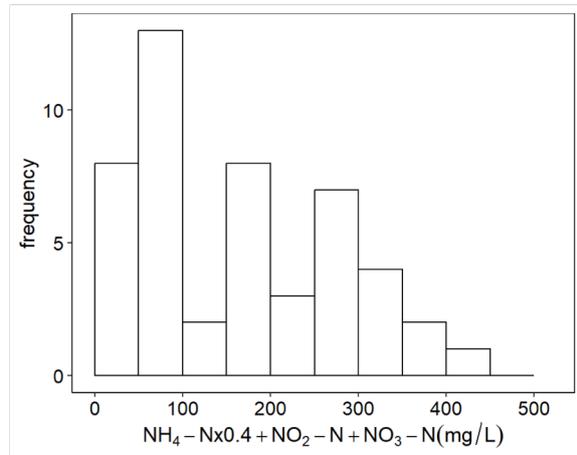


図 1. 硝酸性窒素等濃度平均値の頻度分布

0.69 および 0.86 を示した。また、硝酸性窒素等およびアンモニア性窒素のいずれも排水量が大きい場合は低濃度で安定しているのに対して、排水量が低下するにつれて濃度変動幅が拡大する傾向が見られた。

3. 硝酸性窒素等予測用回帰モデルの作成

①重回帰法による硝酸性窒素等濃度予測用回帰モデル

pH と EC のみを特徴量とする重回帰法での予測値と実測値の関係を図3に示した。MAE は 83 であった。

次に、rfe 関数による適正特徴量評価結果を図4-左に示した。RMSE は 7 特徴量で最低値を示したことから、全特徴量がモデルに有効と評価された。7 特徴量による重回帰法での予測値と実測値の関係を図4-右に示した。MAE は 7 特徴量でも 80 で、2 特徴量の場合との差は小さかった。重回帰式を式2に示した。相関係数は 0.844 ($p < 2.2e-16$) であった。

$$\text{硝酸性窒素等 (mg/L)} = 572.1 - 92.0 \times \text{pH} + 64.9 \times \text{EC (mS/cm)} + 1.8 \times \text{月} - 24.1 \times \text{緯度} + 6.3 \times \text{経度} - 1.6 \times \text{気温} - 0.37 \times \text{排水量 (m}^3/\text{日)}$$

・・・式2

②サポートベクターマシン法による硝酸性窒素等濃度予測用回帰モデル

ハイパーパラメータのチューニングの結

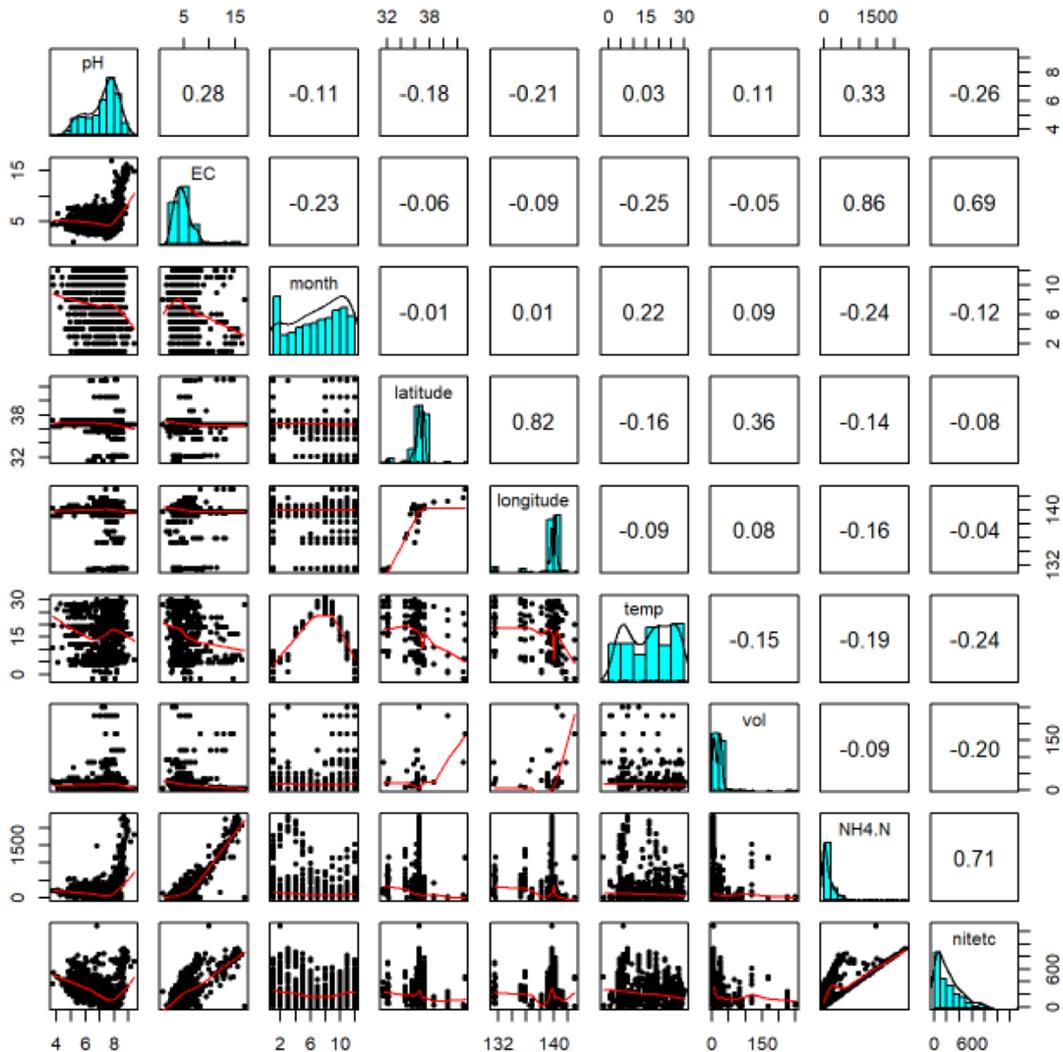


図2. 各特徴量間の相関行列

(temp: 推定気温°C, EC: 電気伝導率 mS/cm, vol: 日排水量 m³/日, NH₄.N: アンモニア性窒素 mg/L, nitetc: 硝酸性窒素等 mg/L)

果、pH と EC のみでの予測の場合、シグマ (sigma) が 1、コストパラメータ (C) は 20 が適正とされた。このハイパーパラメータ値での2特徴量による予測値と実測値の関係を図5に示した。MAEは68であった。

次に、rfe 関数により特徴量の数と予測誤差との関係を解析すると、図6-左のように7特徴量で RMSE が最低になったことから全特徴量がモデルに有効と評価された。7特徴量モデルでのチューニングの結果、ハイパーパラメータは sigma が 0.1、C は 20 が適正と

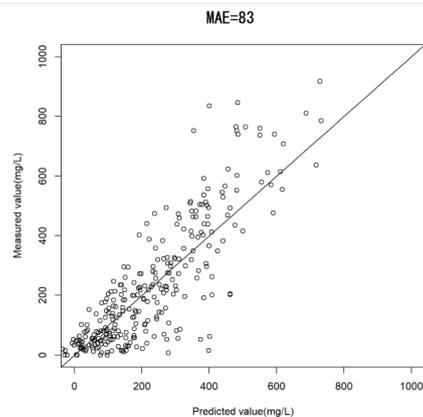


図3. 重回帰法による硝酸性窒素等濃度の予測値 (Predicted value)と実測値 (Measured value)の関係 (pH、ECのみでの予測)

機械学習による窒素化合物濃度の予測

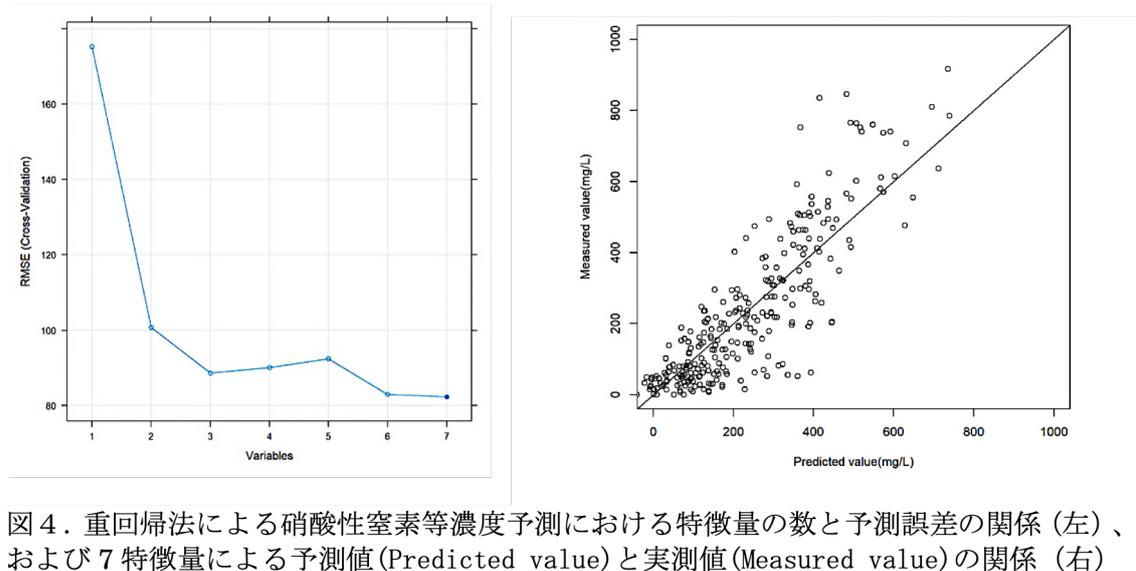


図4. 重回帰法による硝酸性窒素等濃度予測における特徴量の数と予測誤差の関係(左)、および7特徴量による予測値(Predicted value)と実測値(Measured value)の関係(右)

された。7特徴量による予測値と実測値の関係を図6-右に示した。MAEは56に低下し、2特徴量に比較して精度の向上が確認された。

③ランダムフォレスト法による硝酸性窒素等濃度予測用回帰モデル

pHとECの2特徴量での予測モデルでは、チューニングの結果ハイパーパラメータのmtryは2が適正值であった。この値を使った2特徴量モデルによる予測値と実測値の関係を図7に示した。MAEは75であった

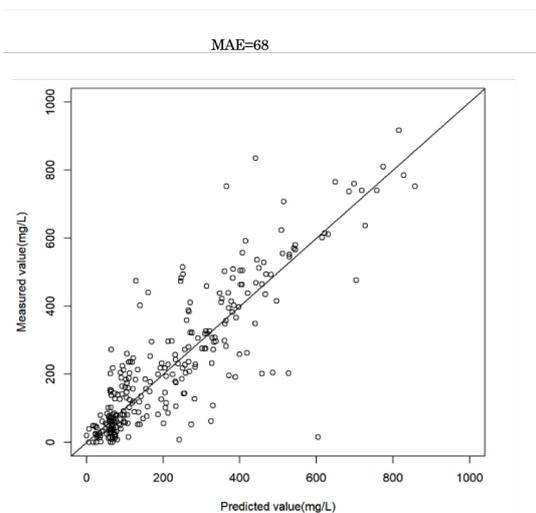


図5. サポートベクターマシン法による硝酸性窒素等濃度の予測値(Predicted value)と実測値(Measured value)の関係 (pHとECのみでの予測)

次に、7特徴量モデルのチューニングではmtryは3が適正と予測された。この条件でrfe関数により特徴量の数とモデル精度との関係を解析すると図8-左のように7特徴量でRMSEが最低になった。よって全特徴量がモデルに有効と評価された。7特徴量モデルでの予測値と実測値の関係を図8-右に示した。MAEは51に低下し、精度向上が確認された。

④ニューラルネットワーク法による硝酸性窒素等濃度予測用回帰モデル

ニューラルネットワーク法では、各特徴量は正規化処理後に解析に利用し、MAEを算出する際に非正規化を行った。活性化関数にはlogisticを用いた。また、事前試行により隠れ層は3層とした。

pHとECのみを特徴量とする場合の各層の適正ユニット数はチューニングの結果9、8、3となった。このユニット数による2特徴量での予測値と実測値の関係を図9に示した。MAEは74であった。

pH、ECに加えて他の特徴量を利用する場合の適正特徴量数は5となり(図10-左)、predictors関数により月と経度が不要な特徴量と認定された。5特徴量を用いたチュー

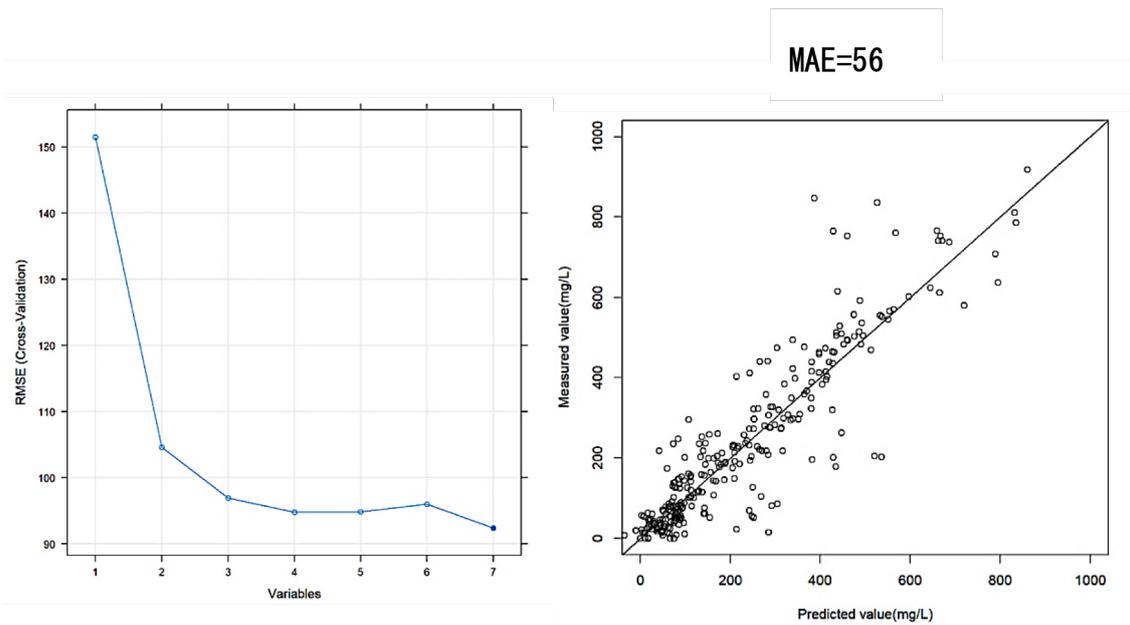


図 6. サポートベクターマシン法による硝酸性窒素等濃度の予測における特徴量の数と予測誤差の関係 (左)、および 7 特徴量による予測値(Predicted value)と実測値(Measured value)の関係 (右)

ニングの結果、各層の 適正ユニット数は 10、8、2 となった。このユニット数で 5 特徴量による予測値と実測値の関係を図 10-右に示した。MAE は 57 となり精度向上が確認された。

⑤キュビスト法による硝酸性窒素等濃度予測用回帰モデル

pH と EC による予測の場合、チューニングによりハイパーパラメータの適正值は committees が 25、neighbors が 7 となった。なお、rules は事前試行により 12 とした。これらの適正值を用いた予測値と実測値の関係を図 11 に示した。MAE は 70 であった。

pH、EC に加えて他の特徴量も利用する場合においては、まず各特徴量がどのくらいの頻度でルール条件に使用されたかを示す Conds と、どのくらいの頻度で線形モデルに使用されたかを示す Model の 2 種類の指標値を求めた。この結果、図 12-左に示すように、Conds には気温 (temp) を除く 6 特徴量が、Model には 7 特徴量のすべてが利用されていた。さらに、predictors 関数で必要特徴量選択を行っ

たところ 7 特徴量のすべてが選択された。よって、気温も含む 7 特徴量の全てがモデルに有効と考えられた。7 特徴量モデルについてチューニングを行ったところ、ハイパーパラメータの committees は 25、neighbors は 5 が適正值と予測された。また rules 数は事前試行により 12 とした。7 特徴量の場合の予測値と実測値の関係を図 12-右に示した。MAE は

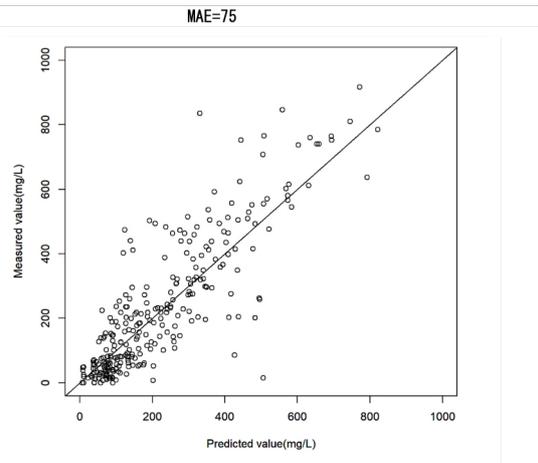


図 7. ランダムフォレスト法による硝酸性窒素等濃度の予測値(Predicted value)と実測値(Measured value)の関係 (pH と EC のみでの予測)

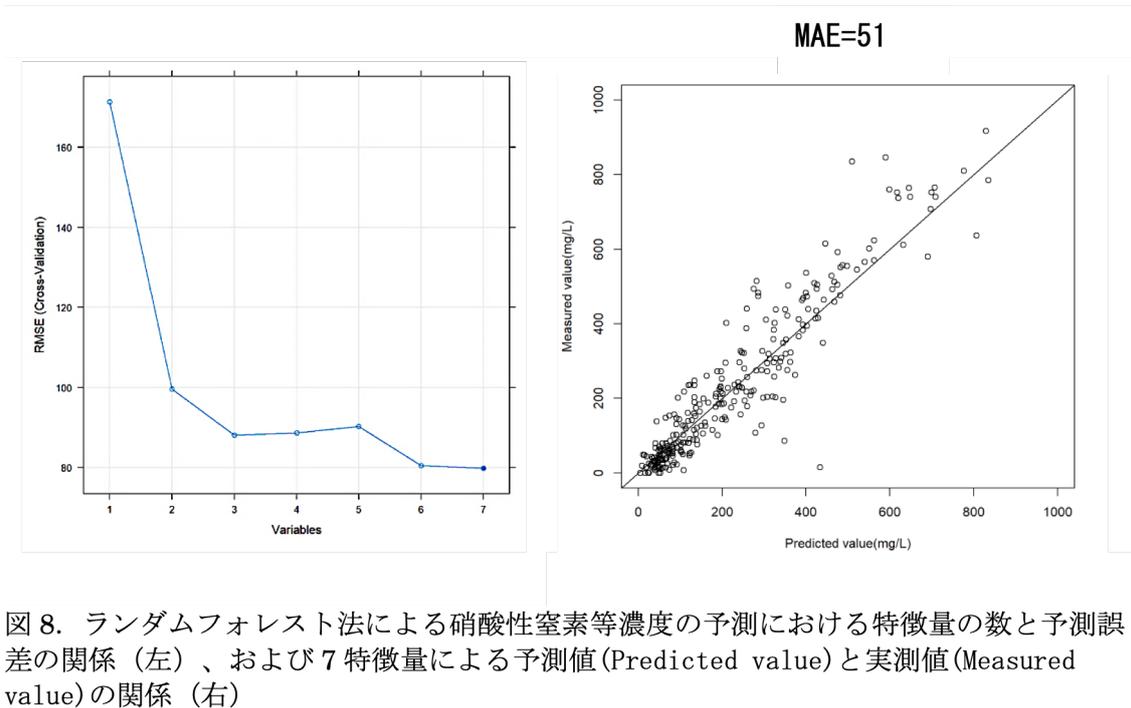


図 8. ランダムフォレスト法による硝酸性窒素等濃度の予測における特徴量の数と予測誤差の関係 (左)、および 7 特徴量による予測値 (Predicted value) と実測値 (Measured value) の関係 (右)

43 となり、2 特徴量の場合に比較して約半減した。

4. 硝酸性窒素等予測用回帰モデルの精度比較

各モデルの MAE を比較した結果を図 13 に示した。2 特徴量の場合、重回帰法より機械学習法の方が MAE が小さく、中でもサポートベクターマシン法の MAE が最小であった。た

だし、重回帰法の MAE が 83 に対してサポートベクターマシン法でも 68 であり、その差は小さかった。

一方、pH、EC 以外にも特徴量を利用した場合、重回帰法の MAE が 80 であったのに対して、ランダムフォレスト法は 51、キュビスト法は 43 にまで顕著に低下し、機械学習法の有効性が確認された。

精度の高かったランダムフォレスト法とキュビスト法について、予測誤差の確率密度曲線 (図 14) を比較すると、ランダムフォレスト法は裾野の広がりが大きいためキュビスト法の方が精度の高いことが示された。

5. アンモニア性窒素濃度予測用回帰モデルの作成

浄化施設の管理においては硝酸性窒素等の濃度に加えて、アンモニア性窒素濃度の把握も硝化の進行度合いを把握するために重要である。そこで、アンモニア性窒素予測モデルの検討も行った。既述のように硝酸性窒

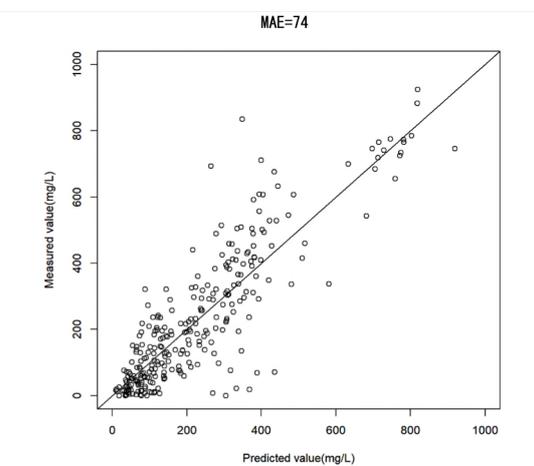


図 9. ニューラルネットワーク法による硝酸性窒素等濃度の予測値 (Predicted value) と実測値 (Measured value) の関係 (pH と EC のみでの予測)

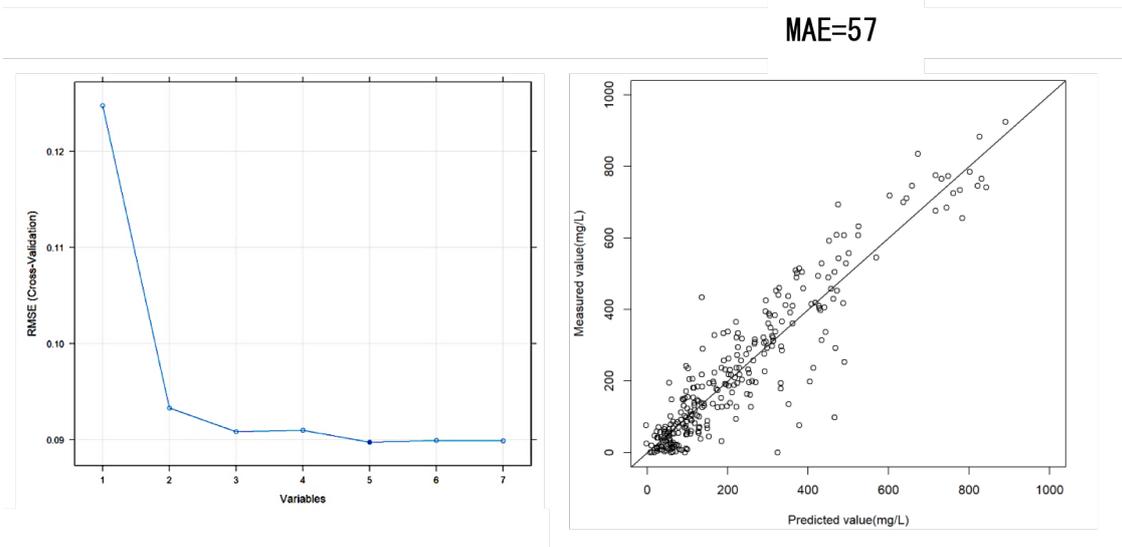


図 10. ニューラルネットワーク法による硝酸性窒素等濃度の予測における特徴量の数と予測誤差の関係（左）、および 5 特徴量による予測値 (Predicted value) と実測値 (Measured value) の関係（右）

素等の予測では pH と EC のみでは予測精度が低かったことから、アンモニア性窒素濃度の予測では多特徴量モデルのみを検討対象とした。

なお、硝酸性窒素等濃度の算出式（式 1）ではアンモニア性窒素には 0.4 を乗ずるが、これ以降に示すアンモニア性窒素濃度は 0.4 を乗じない本来の濃度値である。なお、アンモニア性窒素の実測最高値の約 2000 mg/L は硝酸性窒素等に換算すると 800 mg/L となる。

①重回帰法によるアンモニア性窒素濃度予測用回帰モデル

重回帰法によるアンモニア性窒素の予測においては、最適特徴量の数は 7 と推測された（図 15-左）。7 特徴量の場合の重回帰式を式 3 に示した。相関係数は 0.859 ($p < 2.2 \times 10^{-16}$) であった。

$$\text{アンモニア性窒素 (mg/L)} = 855.9 + 32.1 \times \text{pH} + 121.7 \times \text{EC (mS/cm)} - 3.5 \times \text{月} - 0.88 \times \text{緯度} - 10.4 \times \text{経度} + 0.21 \times \text{温度} - 0.59 \times \text{排水量 (m}^3/\text{日)} \dots \text{式 3}$$

図 15-右に上記重回帰式による予測値と実

測値との関係を示した。MAE は 142 であった。500 mg/L 以下の領域では、実測値がほぼゼロであるにも関わらず数百の予測値になる場合も多く実用的な精度は得られなかった。

②サポートベクターマシン法によるアンモニア性窒素濃度予測用回帰モデル

ハイパーパラメータのチューニングの結果、sigma は 0.1、C は 20 が適正と推測された。特徴量の数と予測誤差との関係を解析す

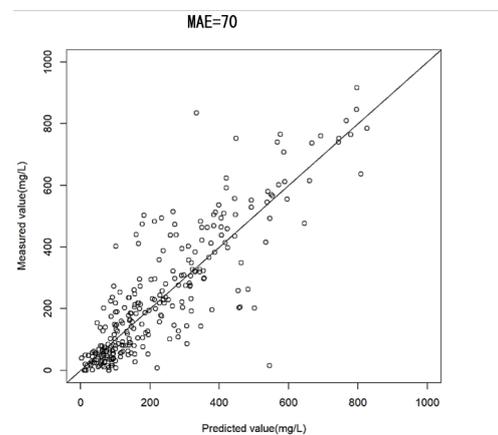


図 11. キュビスト法による硝酸性窒素等濃度の予測値 (Predicted value) と実測値 (Measured value) の関係（pH と EC のみでの予測）

機械学習による窒素化合物濃度の予測

特徴量	Conds (%)	Model (%)
電気伝導率 (EC)	73	95
pH	63	88
緯度 (katitude)	17	60
軽度 (longitude)	17	46
排水諒 (%)	16	30
月 (month)	16	40
気温 (temp)	—	33

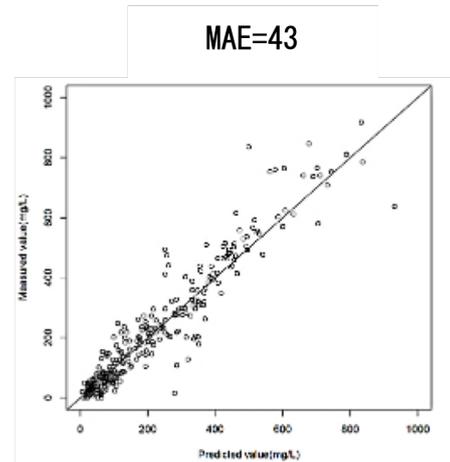
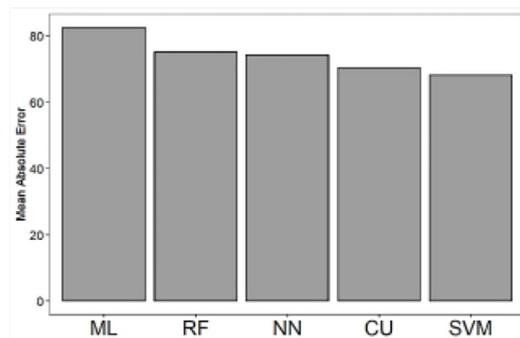


図 12. キュビスト法による硝酸性窒素等濃度の予測における各特徴量の利用頻度評価 (左) および 7 特徴量での予測値 (Predicted value) と実測値 (Measured value) の関係 (右)

pH、EC による場合



追加特徴量を含む場合

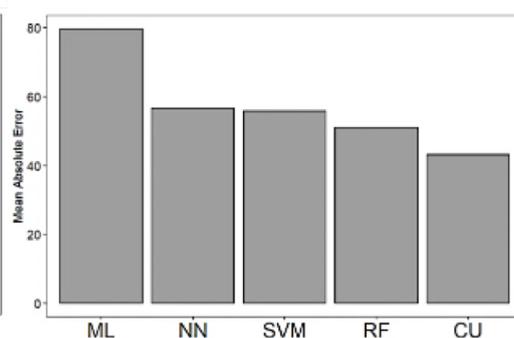


図 13. 硝酸性窒素等濃度予測用回帰モデルの平均絶対誤差 (Mean absolute error) の比較 (ML:重回帰法、SVM:サポートベクターマシン法、RF:ランダムフォレスト法、NN:ニューラルネットワーク法、CU:キュビスト法)

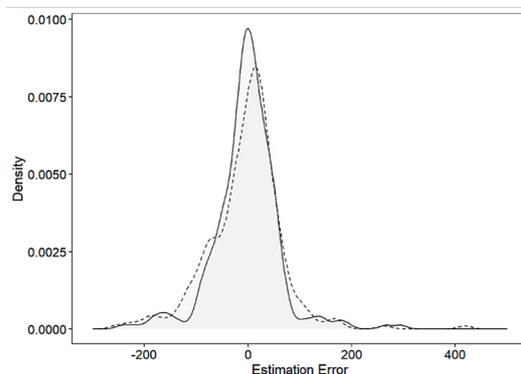


図 14. ランダムフォレスト法とキュビスト法の硝酸性窒素等濃度予測誤差の確率密度分布 (実線:キュビスト法、破線:ランダムフォレスト法)

ると、図 16-左のように 6 特徴量で最も精度が高くなることが示唆された。さらに、predictors 関数で特徴量選択を行ったところ排水量(vol)を除く 6 特徴量が選択された。この 6 特徴量による予測値と実測値の関係を図 16-右に示した。MAE は 80 となり、重回帰法に比べて精度が向上した。

③ランダムフォレスト法によるアンモニア性窒素濃度予測用回帰モデル

適正特徴量数を検討したところ、図 17-左に

機械学習による窒素化合物濃度の予測

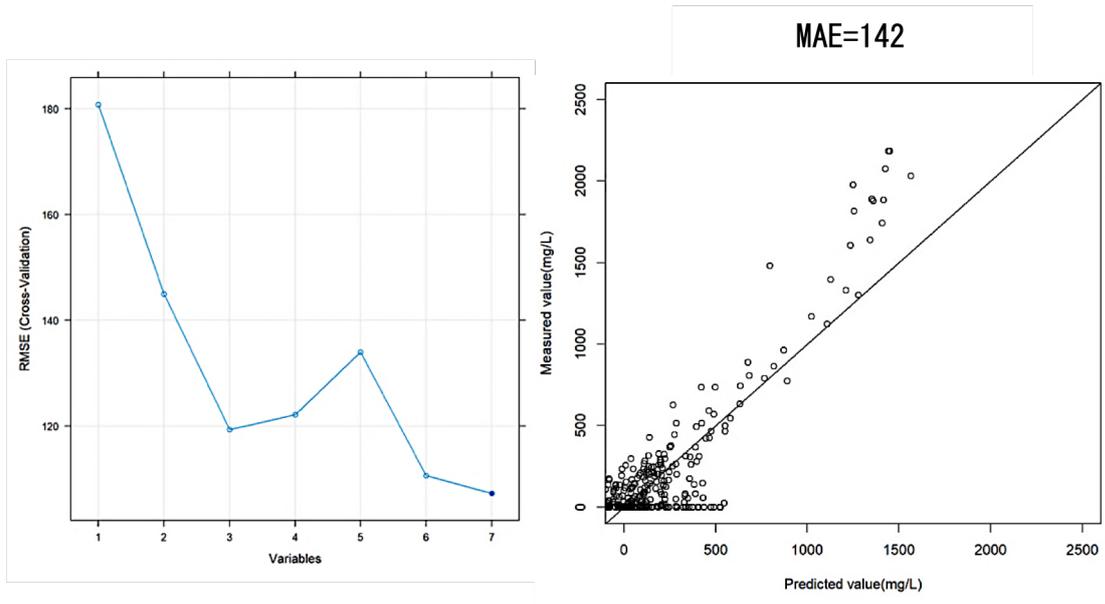


図 15. 重回帰法によるアンモニア性窒素濃度の予測における特徴量の数と予測誤差の関係（左）、および 7 特徴量による予測値 (Predicted value) と実測値 (Measured value) の関係（右）

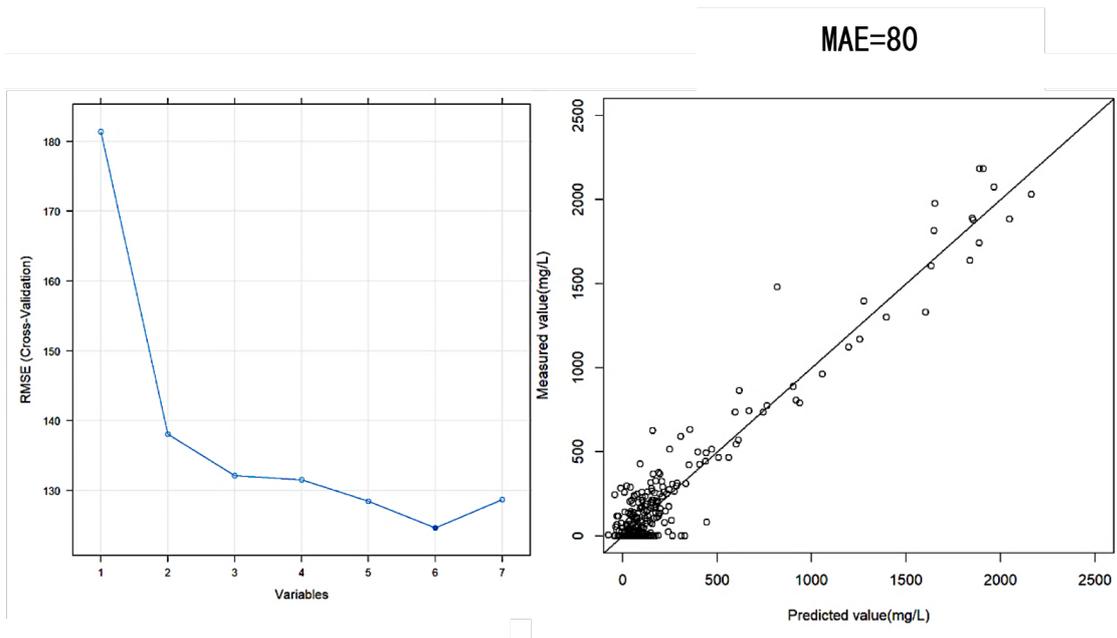


図 16. サポートベクターマシン法でのアンモニア性窒素濃度の予測における特徴量の数と予測誤差の関係（左）、および 6 特徴量による予測値 (Predicted value) と実測値 (Measured value) の関係（右）

示したように 7 特徴量が適正と推測された。また、チューニングの結果、mtry の適正值は 3 になった。特徴量 7、mtry 3 での予測値と実

測値の関係は図 17-右となり、MAE は 63 であった。500mg/L 以下の領域での精度向上が顕著であった。

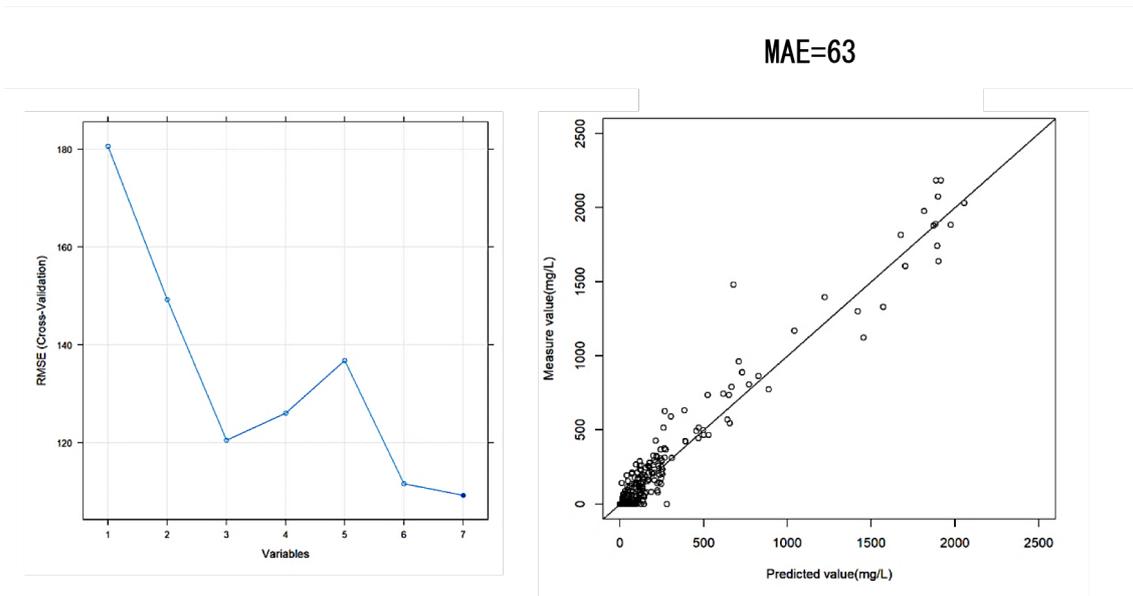


図 17. ランダムフォレスト法によるアンモニア性窒素濃度の予測における特徴量の数と予測誤差の関係 (左)、および 7 特徴量による予測値 (Predicted value) と実測値 (Measured value) の関係 (右)

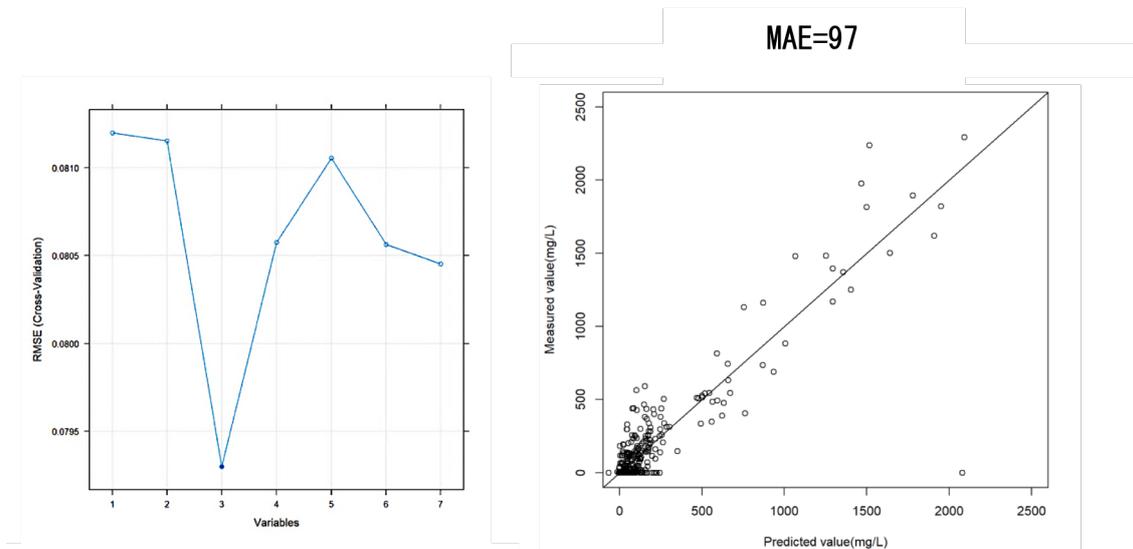


図 18. ニューラルネットワーク法によるアンモニア性窒素濃度の予測における特徴量の数と予測誤差の関係 (左)、および 3 特徴量による予測値 (Predicted value) と実測値 (Measured value) の関係 (右)

④ニューラルネットワーク法によるアンモニア性窒素濃度予測用回帰モデル

最適特徴量の選択では、pH、EC、経度が選択された (図 18-左)。また、チューニングにより 3 層のそれぞれの最適ユニット数を求め

たところ 9、7、2 が適正と予測された。これら適正值での予測値と実測値の関係は図 18-右のようになった。MAE は 97 で、ランダムフォレスト法に比較すると精度が低かった。

特徴量	Conds (%)	Model (%)
電気伝導率 (EC)	73	92
pH	42	62
緯度 (latitude)	6	24
経度 (longitude)	61	75
排水量(vol)	25	42
月 (month)	19	36
気温 (temp)	3	46

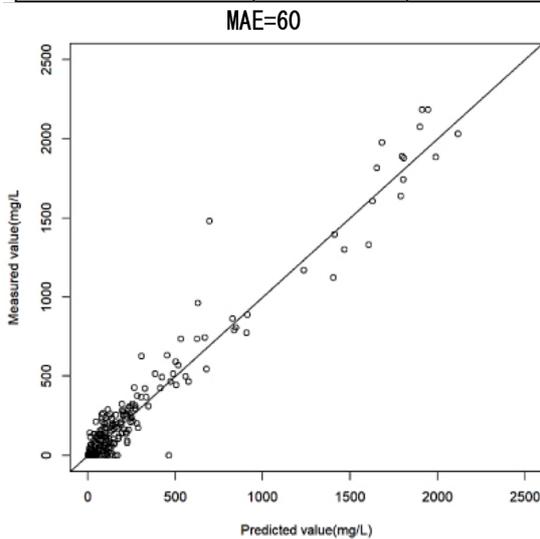


図 19. キュビスト法によるアンモニア性窒素濃度の予測における各特徴量の利用頻度評価 (上) および 7 特徴量での予測値 (下) (Predicted value) と実測値 (Measured value) の関係

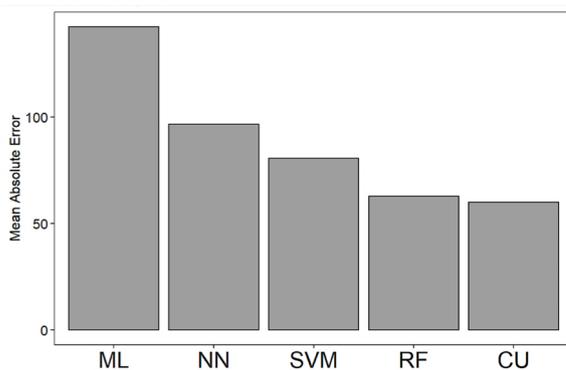


図 20. アンモニア性窒素濃度予測用回帰モデルの平均絶対誤差 (Mean absolute error) の比較

(ML:重回帰法、SVM:サポートベクターマシン法、RF:ランダムフォレスト法、NN:ニューラルネットワーク法、CU:キュビスト法)

⑤キュビスト法によるアンモニア性窒素濃度予測用回帰モデル

Condos と Model の指標値を求めたところ 7 特徴量のすべてが利用されていた (図 19-上)。また、チューニングによって committees および neighbors の最適値は 75 および 5 となった。この条件での予測値と実測値の関係を図 19-下に示した。MAE は 60 となり検討した機械学習法の中で最も精度が高かった。また、ランダムフォレスト法と同様に 500mg/L 以下の領域での精度向上が顕著であった。

6. アンモニア性窒素濃度予測用回帰モデルの精度比較

図 20 にアンモニア性窒素濃度予測における各アルゴリズムの MAE の比較を示した。重回帰法に比べてランダムフォレスト法とキュビスト法では MAE が半分以下にまで低下した。図 21 にキュビスト法とランダムフォレスト法の予測誤差の確率密度分布を示した。両者には大きな相違は見られなかったことから、予測精度はほぼ同等と推測される。したがって、アンモニア性窒素濃度予測にはキュビスト法またはランダムフォレスト法を用いることが最適と考えられる。

考 察

一般にモデルの作成において、理解志向型モデリングでは、何に対応するか説明できない変数はできるだけ排除し、応用志向型モデリングでは、似た変数でもその中の有用な情報を少しでも活用できるようにしてできるだけモデルに含めるべきとされている [8]。今回行ったモデリングは応用志向型であるため、特徴量選択はモデルの解釈性よりも予測性能を重視することが妥当と考えられた。この結

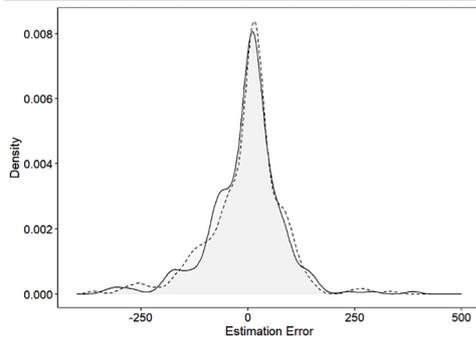


図 21. ランダムフォレスト法とキュビスト法のアンモニア濃度予測誤差の確率密度分布
(実線：キュビスト法、破線：ランダムフォレスト法)

果、重回帰法では硝酸性窒素等予測およびアンモニア性窒素どちらも特徴量は7種が適正であることが示唆された。機械学習法での特徴量は、ニューラルネットワーク法での硝酸性窒素等予測が5種、サポートベクターマシン法でのアンモニア性窒素予測が6種、ニューラルネットワーク法でのアンモニア性窒素予測が3種となったが、他はすべて全特徴量(7種)が適正であることが示唆された。よって、今回選定した7種の特徴量は、いずれも予測に有効であったと言える。

今回ニューラルネットワーク法において選択された特徴量が少なかったことの理論的理解は困難である。一般的に、ニューラルネットワーク法は、決定に至るまでの過程が複雑になるため、理解・解釈が困難な場合が多いとされている。

硝酸性窒素等濃度の予測において、重回帰法では特徴量が2種の場合と多種の場合とでMAEに大差なかった。一方で、機械学習法では多種の特徴量を利用することでMAEが約1/2にまで低下した。線形モデルである重回帰法では特徴量が増えて関係が複雑化すると対応できなかったのに対して、機械学習法

は非線形モデルであるため複雑な関係にも対応できたものと推測される。

硝酸性窒素等濃度の予測においては7特徴量を用いたキュビスト法の予測精度が最も高かった。アンモニア性窒素濃度の予測の場合、7特徴量を用いたキュビスト法とランダムフォレスト法が最適であった。特にアンモニア性窒素濃度が500 mg/L未満の領域では、重回帰法では予測が困難であったのに対して、キュビスト法とランダムフォレスト法ではほぼ妥当な予測ができた。よって、処理施設の日常管理においては7種の特徴量を用いたキュビスト法により硝酸性窒素等とアンモニア性窒素の濃度を予測するのが妥当と結論される。現在、畜産環境技術研究所は処理施設の適正管理支援用アプリを公開しており、その中にキュビスト法による予測アルゴリズムを組み込んでいる。今後さらに学習データを充実させ、予測精度を向上させていくこととしている。

曝気槽水温は硝化活性に強く影響することから、本来は重要な特徴量である。しかし、収集したデータの多くに曝気槽水温が含まれていなかったことから、今回は推定気温で代替した。この推定気温でも特徴量として有効であったことから、曝気槽水温が特徴量に利用できれば予測精度はさらに向上する可能性がある。よって、今後は曝気槽水温が含まれるデータを蓄積して機械学習の学習データに利用できるようにしていきたい。

近年多分野で応用の進んでいるディープラーニング法を利用するとさらに予測を高精度化できる可能性がある。このため、Keras3 パッケージを使用したディープラーニングによる回帰モデル作成も試みたが、現在のところキュビストを上回る精度は得られていない。ディープラーニング法用ハイパーパラメータのチューニングを適正化する

ことでさらに精度が高まる可能性があることから、今後さらに検討を進める予定である。

謝 辞

調査事業での収集データを開示いただいた中央畜産会に感謝いたします。また、中央畜産会の事業においてデータ収集にあたられた、畜産環境アドバイザー薬師堂謙一氏、畜産環境アドバイザー安富政治氏、千葉県畜産総合研究センター長谷川輝明氏、一般財団法人畜産環境整備機構道宗直昭氏に深謝いたします。調査にご協力いただいた農家の方々、および水質分析を担当された畜産環境技術研究所の矢田部枝恵子さんに感謝いたします。

本検討は日本中央競馬会畜産振興事業の助成（汚水浄化施設最適管理支援技術開発普及事業）によるものであることを記し謝意を表します。

文 献

[1] Brett Lantz (2023) Machine learning with R (Fourth edition): packt, Birmingham-Mumbai.

[2] 中央畜産会 (2020) 令和元年度家畜排せつ物処理に係わる新技術情報等の収集・提供事業－畜産経営における排水実態状況調査分析結果－:1-41、公益社団法人中央畜産会、東京.

[3] 中央畜産会 (2021) 令和2年度家畜排せつ物処理に係わる新技術情報等の収集・提供事業－畜産経営における排水実態状況調査分析結果－:1-76、公益社団法人中央畜産会、東京.

[4] 中央畜産会 (2022) 堆肥舎等長寿命化推進事業（畜産環境対策の推進支援）－令和3年度畜産経営における排水実態状況調査分析結果－:1-67、公益社団法人中央畜産会、東京.

[5] 中央畜産会 (2023) 堆肥舎等長寿命化推進事業（畜産環境対策の推進支援）－令和4年度畜産経営における排水実態状況調査分析結果－:1-75、公益社団法人中央畜産会、東京.

[6] 中央畜産会 (2024) 堆肥舎等長寿命化推進事業（畜産環境対策の推進支援）－令和5年度畜産経営における排水実態状況調査分析結果－:1-63、公益社団法人中央畜産会、東京.

[7] 中央畜産会 (2025) 堆肥舎等長寿命化推進事業（畜産環境対策の推進支援）－令和6年度畜産経営における排水実態状況調査分析結果－:1-53、公益社団法人中央畜産会、東京.

[8] 江崎貴裕 (2020) データ分析のための数理モデル入門 本質をとらえた分析のために:222-224、ソシム株式会社、東京.

[9] 鹿熊俊明 (2018) 四季の影響を受ける養豚経営:日豚会誌 55 (2) 48-55.

[10] 日本下水道協会 (2012) 下水試験方法 (上巻)－2012年版－:1-819、公益社団法人日本下水道協会、東京.

[11] Posit team (2025) RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL <http://www.posit.co/>.

[12] R Core Team (2023) R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

[13] 田中康男ら (2021) 養豚農場排水中硝酸性窒素等およびアンモニア性窒素の簡易濃度推定法:日豚会誌58(4):165-177.

Original Paper

High-accuracy prediction of nitrogen compound concentrations in purified swine wastewater using machine learning regression models

Yasuo Tanaka

Institute of Livestock Industry's Environmental Technology,
Nishi-Shirakawa, Fukushima, 961-8061

In wastewater treatment at pig farms, compliance with the Water Pollution Control Act requires regular monitoring of “ammonia, ammonium compounds, nitrite compounds, and nitrate compounds” (hereafter referred to as “nitrate-etc.”). However, the frequency of outsourced analyses is limited due to cost. It has been known that nitrate-etc. can be predicted from pH and electrical conductivity (EC) using multiple regression, but improving prediction accuracy has remained a challenge. Therefore, in addition to pH and EC, we incorporated variables such as the month of sampling, the latitude and longitude of the farm, the mean air temperature of the sampling month at the sampling site, and the daily discharge volume. We then attempted to improve prediction accuracy by using regression models based on machine learning methods (support vector machines, random forest, neural networks, and Cubist). As a result, it was confirmed that using machine learning methods with additional variables beyond pH and EC improves prediction accuracy. In particular, the Cubist regression model using all variables yielded the lowest mean absolute error, suggesting that it is the optimal prediction method.

Key words: swine wastewater, purified effluent, nitrogen compounds, machine learning,
regression model

Corresponding: ytanaka@chikusan-kankyo.jp

Receipt of Ms: 22.09.2025. Accepted: 23.01.2026.

Journal of Animal Production Environment Science