

# 機械学習を用いたダム湖上流における 有害藻類ブルームの発生予測モデルの 開発と発生要因分析への応用

○森 雅也<sup>1</sup>, 山城 広周<sup>1</sup>, 山本 雄太<sup>1</sup>, 森 楓<sup>1</sup>, 鈴木祥広<sup>2</sup>, 平岡 透<sup>3</sup>, 野中 尋史<sup>1</sup>

1: 長岡技術科学大学

2: 宮崎大学

3: 長崎県立大学

## 川や湖，池などで発生する有害な藻類の 異常発生(有害藻類ブルーム)による環境汚染が問題



(出典: Wikipedia)

1. 水中の溶存酸素の欠乏や原核光合成生物が産生する毒素による，魚介類のへい死や生態系の破壊[1]
2. 有害藻類の毒素による人的影響(肝不全，肝臓ガン)[2]
3. 青潮と呼ばれる，藻類の死滅による悪臭や猛毒(硫化水素)の発生



有害藻類ブルームの抑制やメカニズムの解明が  
環境科学の分野で非常に重要な課題となっている[3][4]

## ダム湖上流やダム湖内，ダム湖下流での 有害藻類ブルームが注目

- ダム湖上流：毒素，悪臭，生態系の破壊
- ダム湖内：水力発電の発電効率低下
- ダム湖下流：飲水への影響，漁業などへの経済的影響



ダム湖周辺での有害藻類ブルームの  
抑制・メカニズムの解明が重要



川上が川下へ及ぼす影響を排除するために…

ダム湖上流における有害藻類ブルームの  
発生予測モデルの開発と発生要因分析

# 実験方法

2クラス分類により, 3ヶ月後に『有害藻類ブルームか否か』を予測

特徴量エンジニアリング

Step 1

特徴量空間の構築・評価

Step 2

サポートベクターマシン(SVM)による分類

Step 3

# 実験方法

2クラス分類により, 3ヶ月後に『有害藻類ブルームか否か』を予測

Step 1

## 特徴量エンジニアリング

一ツ瀬ダム の 2003 年 1 月から 2019 年 10 月 までの 3 ヶ月 おきの データ (67 プロット)

説明変数(特徴量)(79次元)

- ダム上流水質計測データ
- 1年間の分布の統計量

目的変数

- ダム上流藻類発生量計測データ
- 閾値未満を正常, 以上を異常

Step 2

## 特徴量空間の構築・評価

Step 3

## サポートベクターマシン(SVM)による分類

# 実験方法

## 2クラス分類により, 3ヶ月後に『有害藻類ブルームか否か』を予測

Step 1

Step 2

Step 3

### 特徴量エンジニアリング

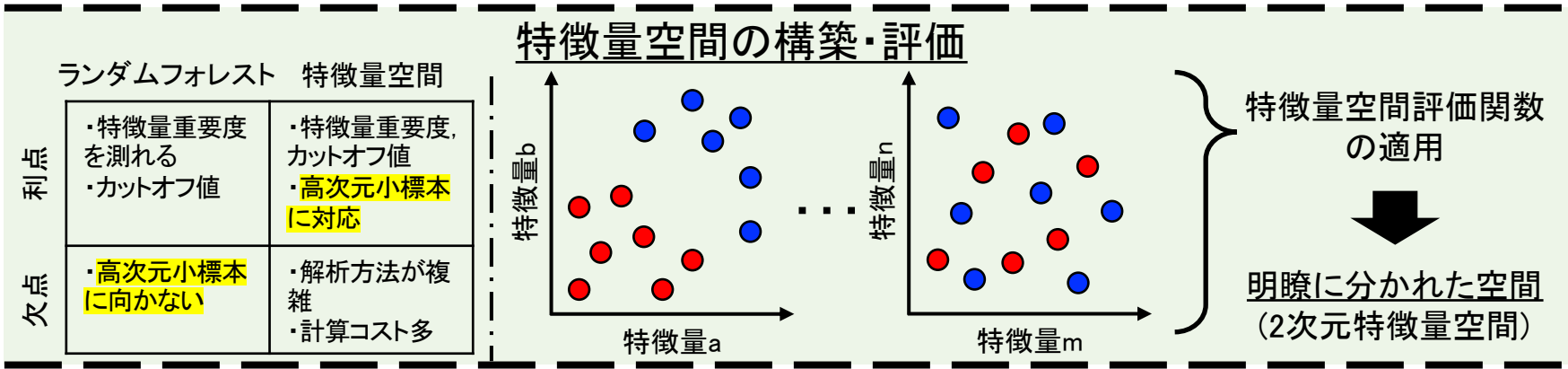
一ツ瀬ダムの2003年1月から2019年10月までの3ヶ月おきのデータ(67プロット)

説明変数(特徴量)(79次元)

- ダム上流水質計測データ
- 1年間の分布の統計量

目的変数

- ダム上流藻類発生量計測データ
- 閾値未満を**正常**, 以上を**異常**



### サポートベクターマシン(SVM)による分類

# 実験方法

## 2クラス分類により, 3ヶ月後に『有害藻類ブルームか否か』を予測

Step 1

Step 2

Step 3

### 特徴量エンジニアリング

一ツ瀬ダムの2003年1月から2019年10月までの3ヶ月おきのデータ(67プロット)

説明変数(特徴量)(79次元)

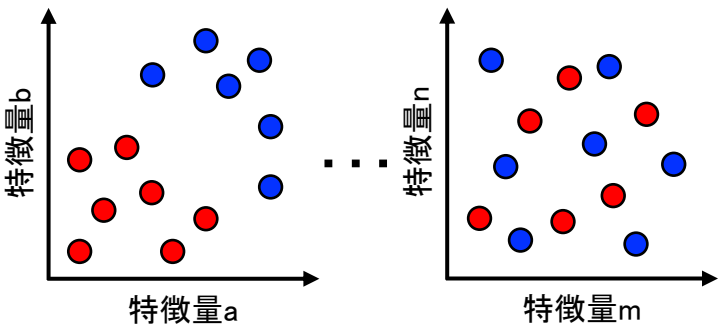
- ダム上流水質計測データ
- 1年間の分布の統計量

目的変数

- ダム上流藻類発生量計測データ
- 閾値未満を**正常**, 以上を**異常**

### 特徴量空間の構築・評価

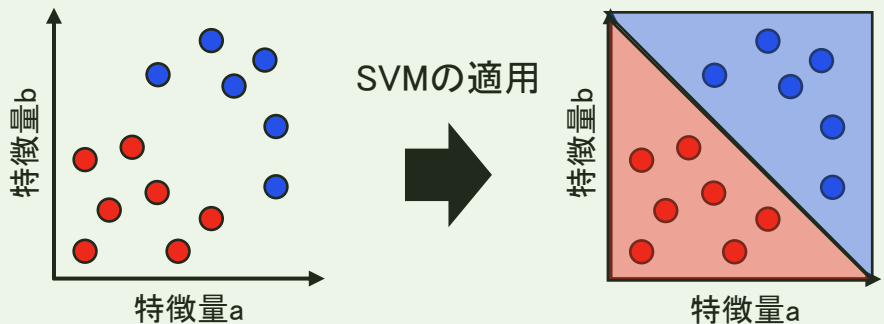
ランダムフォレスト 特徴量空間	
利点	<ul style="list-style-type: none"><li>特徴量重要度を測れる</li><li>カットオフ値</li><li>高次元小標本に対応</li></ul>
欠点	<ul style="list-style-type: none"><li>高次元小標本に向かない</li><li>解析方法が複雑</li><li>計算コスト多</li></ul>



特徴量空間評価関数の適用

明瞭に分かれた空間 (2次元特徴量空間)

### サポートベクターマシン(SVM)による分類



- パラメータチューニング
- 層化k分割交差検証
- 分類精度, F値
- 特徴量重要度(発生要因分析)

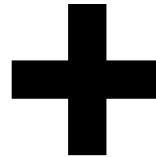
# Step 1: 特徴量エンジニアリング(説明変数)

2クラス分類により, 3ヶ月後に『有害藻類ブルームか否か』を予測

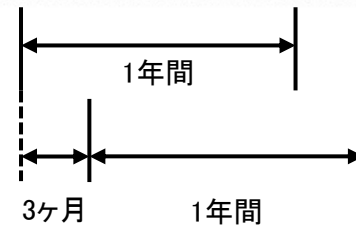
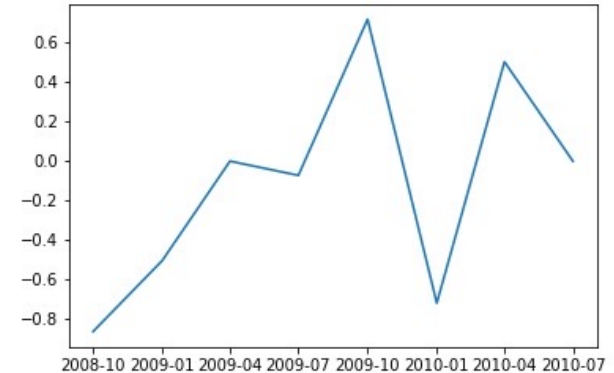
2004年1月から2019年7月までの  
3ヶ月おきのデータ(63プロット)

1. 水温
2. 透明度
3. 水色
4. 溶存酸素量
5. 溶存酸素飽和率
6. 伝導率
7. 濁度
8. 水素イオン濃度
9. 浮遊物質量
10. 化学的酸素要求量
11. リン酸
12. 全リン
13. アンモニア
14. 亜硝酸
15. 硝酸
16. 全窒素
17. クロロフィルa

平均: 0  
標準偏差: 1



2003年1月から2019年7月までの  
1年間の分布の統計量(63プロット)



窓幅: 1年間(4プロット)

スライド幅: 3ヶ月(1プロット)

平均, 標準偏差, 尖度, 歪度



全データ数: 63プロット

説明変数: 79次元

(リン酸, アンモニア, クロロフィルaの尖度と歪度) 8



# Step 1: 特徴量エンジニアリング(目的変数)

2クラス分類により, 3ヶ月後に『有害藻類ブルームか否か』を予測

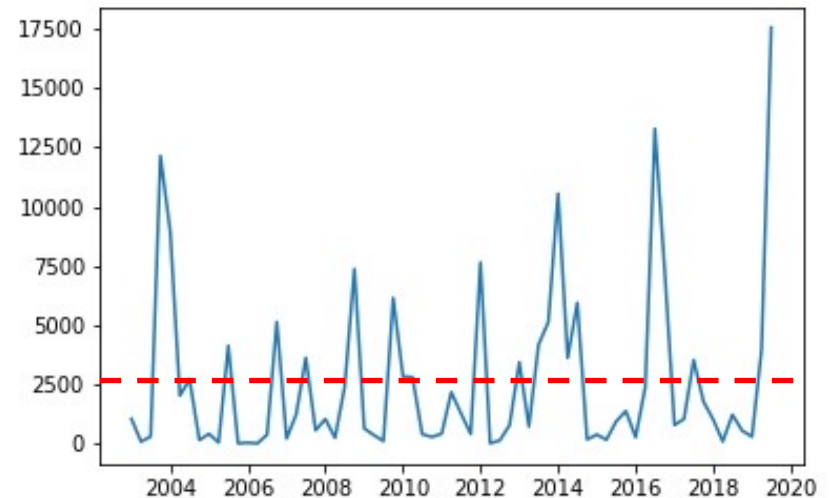
---

2004年4月から2019年10月までの3ヶ月おきのデータ(63プロット)

ダム湖上流藻類発生量計測データ

1. ラン藻
2. 緑藻
3. 珪藻
4. ミドリムシ藻類
5. 黄色鞭毛藻類
6. 渦鞭毛藻類
7. 褐色鞭毛藻類

各藻類の発生量の総和(平均: 2573)



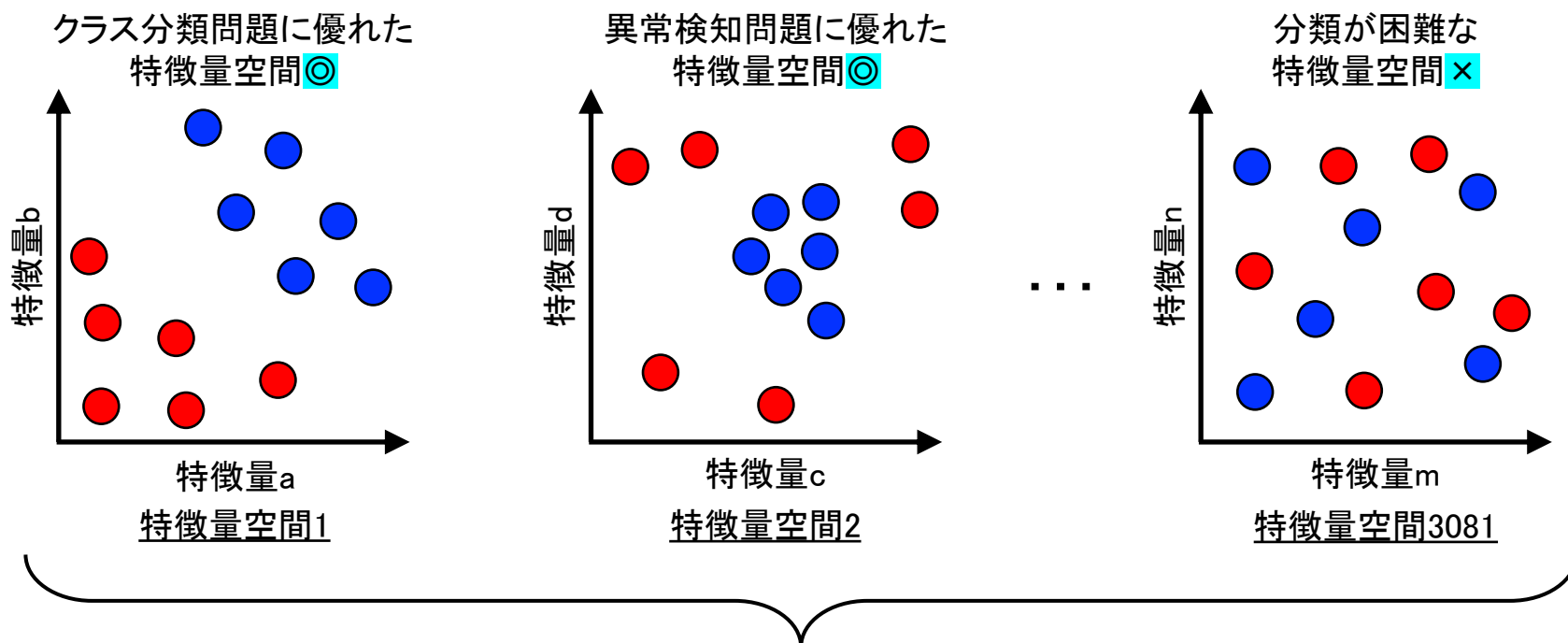
平均より発生量が少ない場合: 正常クラス(42プロット)

平均より発生量が多い場合: 異常クラス(21プロット)

# Step 2: 特徴量空間の構築・評価

2クラス分類により, 3ヶ月後に『有害藻類ブルームか否か』を予測

Step 1で用意した79の特徴量から, 2次元特徴量空間を構築( ${}_{79}C_2 = 3081$ 空間)



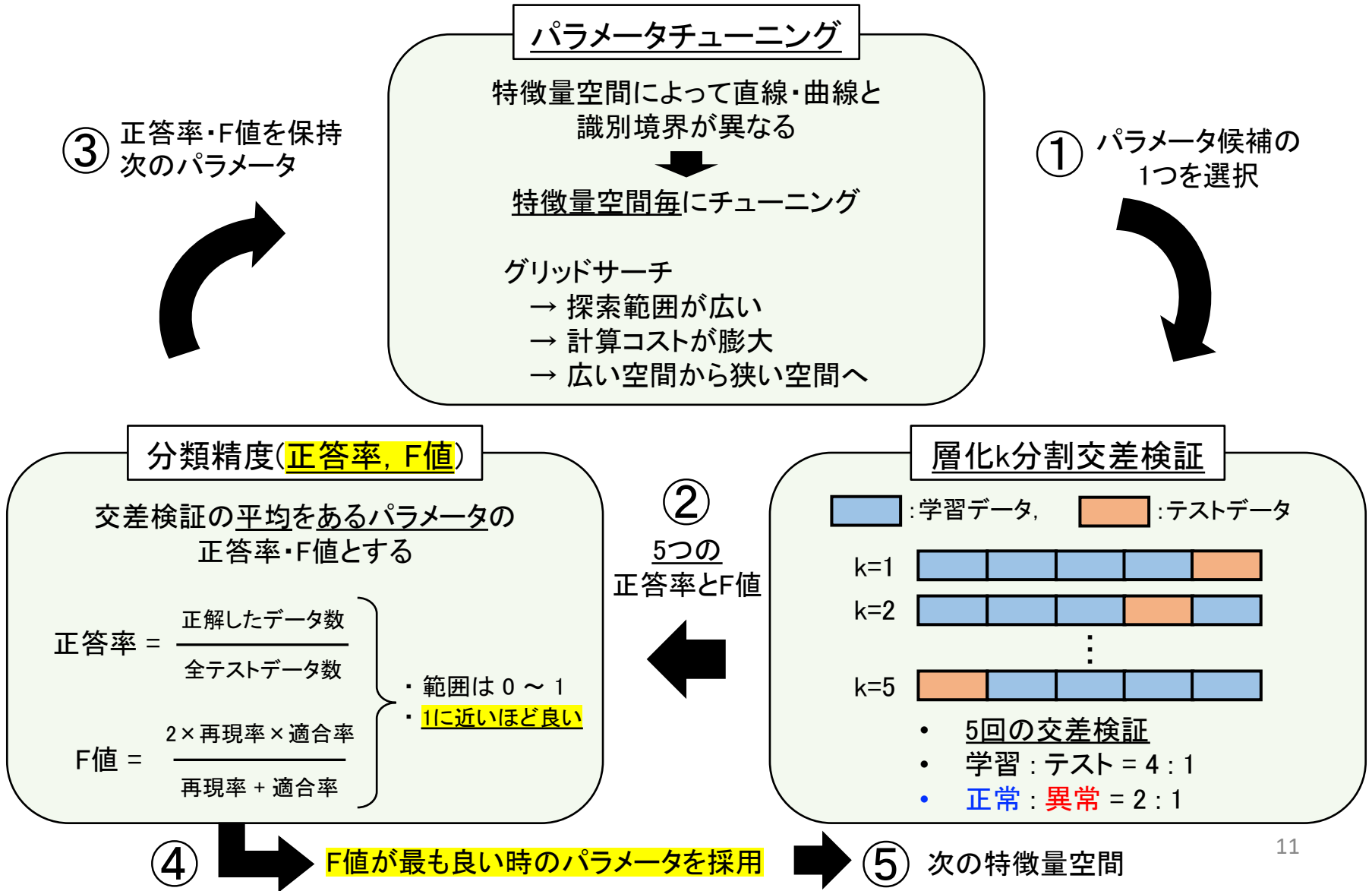
Minimum Reference Set(MRS)[5] による順位付け



上位20種の特徴量空間を解析対象へ

# Step 3: SVMによる分類

選出された上位20種の特徴量空間に対しSVMを適用  
(カーネル関数: 動径基底関数カーネル; RBF )



# 結果:2次元特徴量空間の場合

MRSによって選定された上位20種の中で  
F値が最も良い上位3種の特徴量空間の**特徴量と分類精度**

上位3種	特徴量:X軸	特徴量:Y軸	分類精度(正答率)		分類精度(F値)	
			trainデータ	testデータ	trainデータ	testデータ
rank 1	硝酸 - 平均	溶存酸素量 - 尖度	.730	.715	.656	.676
rank 2	硝酸 - 平均	溶存酸素量 - 歪度	.790	.717	.720	.638
rank 3	全窒素 - 平均	水温 - 尖度	.813	.808	.656	.628

## 考察

モデルの予測性能は、**正答率が70%以上**、**F値が0.6以上**

モデルの特徴量は

『硝酸 - 平均』『全窒素 - 平均』『溶存酸素量 - 尖度』『溶存酸素量 - 歪度』『水温 - 尖度』

水中の**栄養素・酸素量・水温**がダム湖上流での有害藻類ブルームの**発生要因**

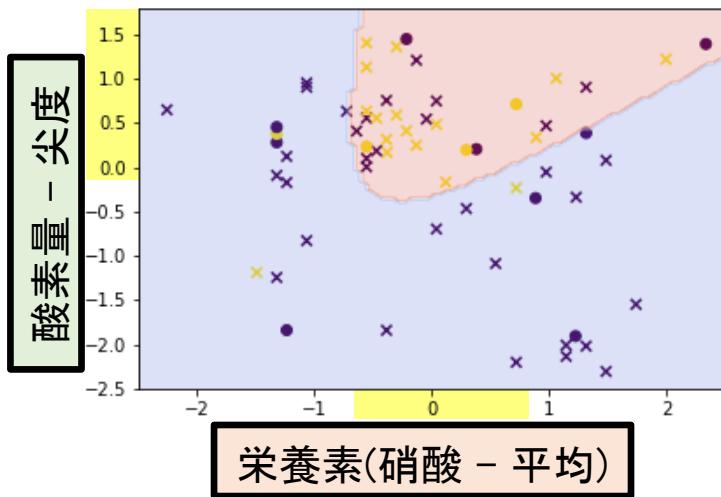
➡ **湖や海を対象とした先行研究でも示されている[4][6]**

[4] Derot, J., Yajima, H., & Jacquet, S. (2020). Advances in forecasting harmful algal blooms using machine learning models: A case study with Planktothrix rubescens in Lake Geneva. *Harmful Algae*, 99, 101906.  
[6] 金相暉, 藤井直紀, & 濱田孝治. (2019). 機械学習を用いた沿岸海域における赤潮予測. *土木学会論文集 B1 (水工学)*, 75(2), 1781-1786.

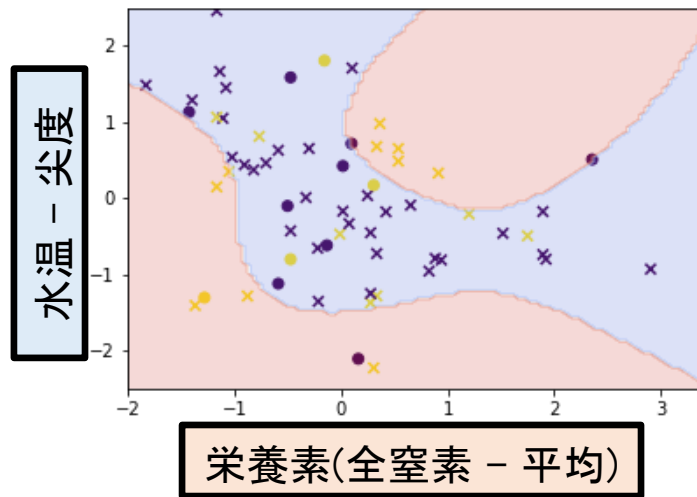
# 結果:2次元特徴量空間の場合

## rank 1とrank 3の特徴量空間の可視化

rank 1の特徴量空間



rank 3の特徴量空間



### 凡例

- ✕: 正常プロット(学習データ)
- : 正常プロット(テストデータ)
- : 正常領域
- ✕: 異常プロット(学習データ)
- : 異常プロット(テストデータ)
- : 異常領域

### 考察

#### rank 1の場合

- 酸素量の尖度が0以上 → 1年を通して安定した酸素量が必要(尖度0 = 正規分布)
- 栄養素の平均が $0 \pm 0.7$  → 少ないと増えない, 多いと増えすぎてしまう

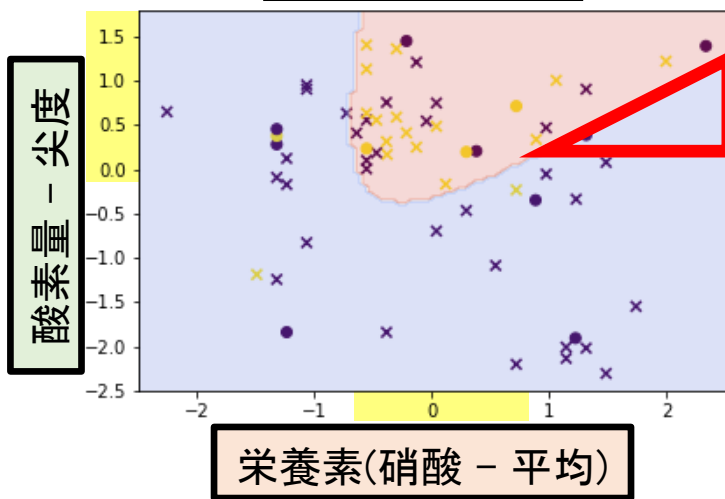
#### rank 3の場合

- 水温と栄養素が逆数の関係の時に正常クラス → 1年間の水温の分布がばらけていると✕  
→ 栄養素がある程度多い(0~1だ)と✕
- 栄養素が多すぎると正常クラス → 藻類が急激に増加, 溶存酸素量の低下
- 別の要因の影響が存在 → 多次元に拡張することで分類精度向上の可能性

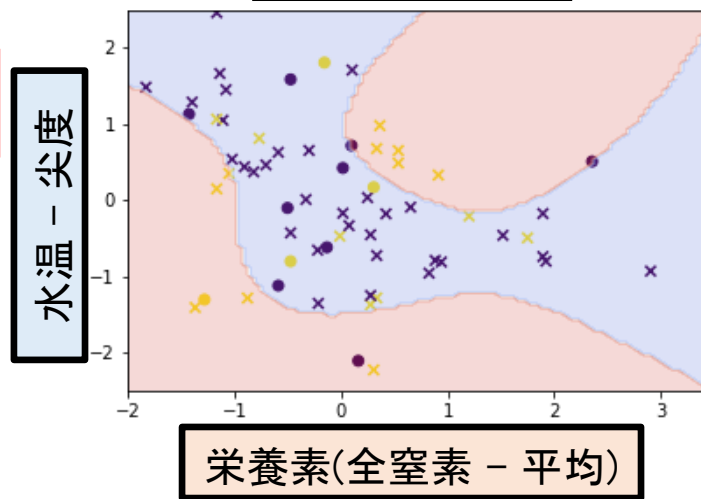
# 結果:2次元特徴量空間の場合

## rank 1とrank 3の特徴量空間の可視化

rank 1の特徴量空間



rank 3の特徴量空間



### 凡例

- ✕: 正常プロット(学習データ)
- : 正常プロット(テストデータ)
- : 正常領域
- ✕: 異常プロット(学習データ)
- : 異常プロット(テストデータ)
- : 異常領域

### 考察

#### rank 1の場合

- 酸素量の尖度が0以上 → 1年を通して安定した酸素量が必要(尖度0 = 正規分布)
- 栄養素の平均が $0 \pm 0.7$  → 少ないと増えない, 多いと増えすぎてしまう

- ➡
1. 栄養が多いため藻類が急激に増加
  2. 溶存酸素量が急激に減少
  3. 3ヶ月経つ前に増殖, 死滅

#### rank 3の場合

- 水温と栄養素が逆数の関係の時に正常クラス → 1年間の水温の分布がばらけていると✕  
→ 栄養素がある程度多い(0~1だ)と✕
- 栄養素が多すぎると正常クラス → 藻類が急激に増加, 溶存酸素量の低下
- 別の要因の影響が存在 → 多次元に拡張することで分類精度向上の可能性

# 追加研究：多次元特徴量空間の場合

3次元, 4次元, 5次元に拡張することによる, 分類精度の向上を確認する

${}_5C_3 + {}_5C_4 + {}_5C_5 = 16$ 空間の構築 + SVMによる識別境界

過学習していない状態で, 最もF値が良い特徴量空間

特徴量				分類精度 (正答率)			分類精度 (F値)		
1軸	2軸	3軸	4軸	2次元	train	test	2次元	train	test
溶存酸素量 尖度	溶存酸素量 歪度	全窒素 平均	水温 尖度	.715	.813	.808	.676	.718	.717

考察

2次元特徴量空間に対し, 正答率が約10%向上, F値が0.7台に

発生要因は栄養素・酸素・水温

川上が川下へ及ぼす影響を排除するために…

## ダム湖上流における有害藻類ブルームの発生予測モデルの開発と発生要因分析

### 手続き

1. 特徴量エンジニアリング
2. 特徴量空間の構築・評価
3. 上位20空間に対してSVMの適用
4. 分類精度の算出

### 結果

- 2次元の場合, 正答率が70%台, F値が0.6台
- 多次元の場合, 正答率が80%台, F値が0.7台
- 発生要因は栄養素・酸素量・水温
- 一定の栄養素, 安定した酸素量, 水温の高低差

### 今後の展望

- 今回の結果を元に, ダム湖上流での実験的検証
  - 選出された特徴量に関するより深い検討
- 本研究で, 特徴量空間による解析方法の有効性を確認
  - ダム湖内を対象に解析する
- 特徴量空間×ランダムフォレスト
  - 次元削減により過学習の抑制, 多次元におけるカットオフ値