第Ⅱ部門　　機械学習を用いた水質定点自動観測データの異常検知とソフトセンサーに関する検討

大阪大学大学院工学研究科　学生会員　○二口　陽日
大阪大学大学院工学研究科　正会員　永野　隆紀
大阪大学大学院工学研究科　正会員　入江　政安

## 1. はじめに

水質汚濁が進行している大阪湾および東京湾では，これまでに定点自動観測(**図-1**)が実施され，貧酸素水塊の動態解析などに貢献してきた．しかし，自動観測データには,センサーの不具合等に起因する判別困難な異常値が含まれる場合がある (**図-2**)．本研究では，大阪湾を対象として開発された異常検知システム[1)]について，擬似観測の溶存酸素 (DO) を用いて推定性能低下の要因を検討した．さらに，東京湾における水質観測データへ適用し，本システムの汎用性を検討した．

## 2. 手法

**図-3** に本研究で用いる信頼性評価システムの概要図を示す[1)]．信頼性評価システムではまず，正規化した観測データから主成分分析により，主成分および再構成に用いる固有値行列を抽出する．続いて主成分を予測する機械学習 (ML) モデルを機械学習手法である勾配ブースティング決定木を用いて構築する．最後に，寄与率が高い上位数個の主成分を用いて再構成することで，推定値を得る．ML モデルの説明変数には，気温，風などの気象データ，河川流量，および天文潮位を用いた．また， DO 変動の時間差や気象の影響を考慮してタイムラグと累積時間を説明変数に導入した．

異常検知の可否および ML モデルの推定精度を評価するためには実際の観測値と異なる DO のデータセットが必要となる．そこで，3 次元流動水質モデル(ROMS) から算出した 5 地点（**図-1**）における 2012 年 ~ 2016 年の 5 年間の DO を使用した．擬似観測値で漸減を表現する場合には，7 月の計算 DO に 100%から 0%に線形に減少する係数を乗じたものを使用した．

異常値を検知する評価指標を以下に示す．

$$\alpha = \frac{(e_i - \mu_e)^2}{\sigma_e^2}$$

ここで，$e_i$ は推定値と擬似観測値のミスフィットを表し．$\alpha$ を異常度指標と定義する．

## 3. テストデータの選定

ML モデルでは，漸減を含まない擬似観測 DO を学習させた場合であっても，異常度上位数%で異常値を抽出するため、

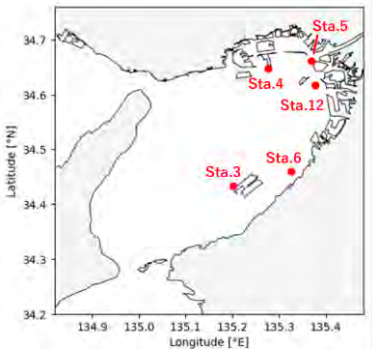



**図-1**　学習に使用した観測地点

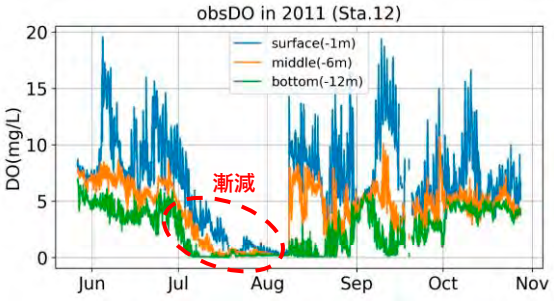



**図-2**　大阪湾波浪観測塔における漸減

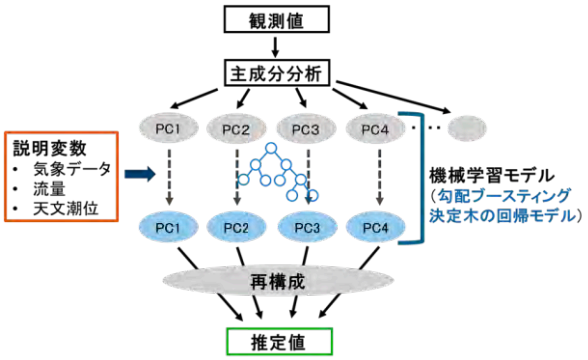



**図-3**　信頼性評価システムの概要図

正常値であっても誤って異常値と判定されてしまう場合がある．その場合，正確に漸減を検知できない恐れがあることから，漸減区間における DO の推定の検証に適したテストデータを選定する必要がある．そこで，漸減を含まない擬似観測 DO を学習させた ML モデルから得た推定 DO と漸減を含む擬似観測 DO を用いて F 値による異常度を検討し，テス

Haruhi FUTAKUCHI, Masayasu IRIE, Takanori NAGANO
futakuchi_h@civil.eng.osaka-u.ac.jp

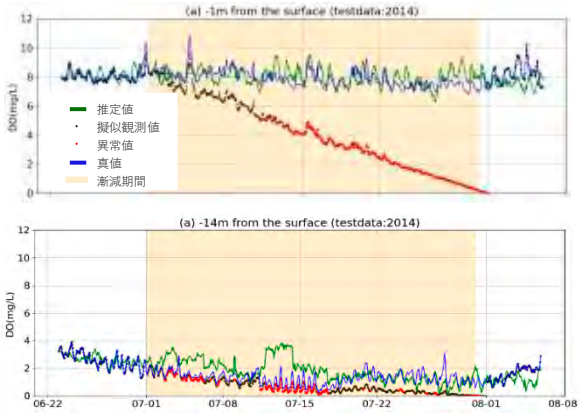



**図-4**　事前処理なしの場合の擬似観測値と推定値の時系列．上段は水深 1 m，下段は 14 m の結果を示す

トデータの選定を行った．F 値とは，漸減を異常と検知した適合率と，漸減の見逃しの少なさを表す再現率のバランスを評価するスコアであり，1 に近いほどモデルの精度が高いと評価される．表層および全層平均において F 値を算出したところ，異常度上位 6%のときに最も F 値が高く，ML モデルで推定した年が 2014 年のとき最も漸減を異常と検知した．

## 4. 事前処理が DO 推定精度に及ぼす影響

本システムでは，ML モデルに入力する事前処理として観測データに見られた漸減等の異常値を人力で省く必要があり，DO に関する予備知識が必要という課題がある．そこで，異常値の事前処理がない場合でも ML モデルの推定精度が維持されるかを検証するため，事前処理の有無が異常検知および ML モデルの推定精度に及ぼす影響について検討した．ML モデルの学習データには，2014 年を除いた 4 年間の計算 DO のうち，2012 年 7 月に与えた漸減を事前処理しない場合を(a)，事前処理した場合を(b)とし，2014 年の DO を推定した．さらに，推定 DO と 7 月に漸減を含む擬似観測 DO を比較し，異常度上位 6%の閾値で異常検知を試みた．**図-4** に (a) 事前処理なしの場合の異常検知結果を示す．事前処理の有無で異常検知の性能に差はなかった．いずれの条件でも表層の異常検知精度は高かった一方，底層では異常検知の精度が低下した．事前処理の有無で推定値に明瞭な変化は見られず，ML モデルの推定に及ぼす影響は小さいと考えられた．

## 5. 他水域への適用性に関する検討

さらに，異常検知システムの汎用性を検証するために，東京湾で稼働中の 4 地点 (浦安沖・川崎人工島・千葉港口第一号灯標・検見川沖) の DO モニタリングデータに適用した．東京湾モニタリングシステムは 2010 年に漸減異常が発生している [2]．そこで，2010 年を除く，2011 年から 2020 年の観測 DO を学習データとし，2010 年 4 月 1 日 ~2010

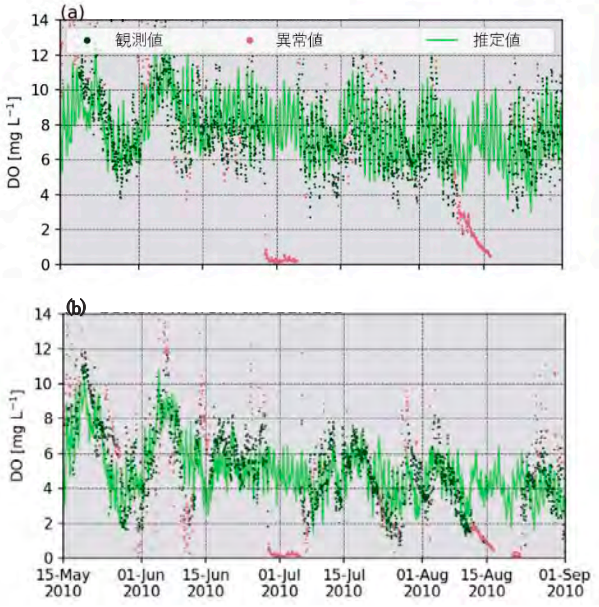



**図-5**　浦安沖における異常検知
(a) は表層，(b) は底層の DO の時系列を示す

年 12 月 31 日の DO を推定した．推定値と実観測値の差に基づき，異常検知を実施した結果を**図-5** に示す．異常発生区間における観測値（黒点）が異常値（赤点）として検出されていることが分かる．同図 (b)ではそれ以外の区間（例えば，7 月 19 日 ~7 月 25 日）においても異常値とされた観測値が存在しており，数日の間に発生する DO 低下の異常値判定には課題があるものと考えられる．

## 6. まとめ

主成分分析および機械学習を用いて，水質定点自動観測システムの異常検知システムを構築した．事前処理の有無に関係なく，表層においては十分な精度で異常検知できたものの，底層では異常検知の精度が低下した．学習データの事前処理による異常検知システムの推定精度への影響は小さく，事前処理は不要であると考えられた．また，本システムを東京湾の定点水質観測データに適用し，2010 年夏季に発生した異常値を検出することに成功した．一方で，誤検知とみられる期間が散見されたため，説明変数の見直しやより高度な深層学習を利用した更なる検討が必要であると考えられる．

**参考文献**

1) Nagano, T., Irie, M., & Wells, J. C. (2025). Development of a reliability Gevaluation system for dissolved oxygen monitoring data using machine learning. *Hydrological Research Letters*, *19*(3), 142-148.
2) 国土交通省関東地方整備局，東京湾口航路事務所，東京湾水質連続観測データ，補正済みデータに関する説明書